



**REDE DE BIODIVERSIDADE E BIOTECNOLOGIA DA AMAZÔNIA LEGAL
PROGRAMA DE PÓS-GRADUAÇÃO EM BIODIVERSIDADE E BIOTECNOLOGIA
DA REDE BIONORTE**

CLEYDSON BRENO RODRIGUES DOS SANTOS

**DESENVOLVIMENTO RACIONAL DE FÁRMACOS ANTIMALÁRICOS
DERIVADOS DA ARTEMISININA USANDO MÉTODOS COMPUTACIONAIS SAR
E QSAR**

**Macapá - AP
2014**



**REDE DE BIODIVERSIDADE E BIOTECNOLOGIA DA AMAZÔNIA LEGAL
PROGRAMA DE PÓS-GRADUAÇÃO EM BIODIVERSIDADE E BIOTECNOLOGIA
DA REDE BIONORTE**

**DESENVOLVIMENTO RACIONAL DE FÁRMACOS ANTIMALÁRICOS
DERIVADOS DA ARTEMISININA USANDO MÉTODOS COMPUTACIONAIS
SAR E QSAR**

CLEYDSON BRENO RODRIGUES DOS SANTOS

Tese apresentada ao Programa de Pós Graduação em
Biodiversidade e Biotecnologia da Rede BIONORTE
como requisito para obtenção do título de doutor em
Biotecnologia

Universidade Federal do Amapá.

Orientador:

Prof. Dr. José Carlos Tavares Carvalho

Coorientador:

Prof. Dr. Williams Jorge da Cruz Macêdo

**Macapá – AP
2014**

Dados Internacionais de Catalogação na Publicação (CIP)
Biblioteca Central da Universidade Federal do Amapá

615.1
S237d

Santos, Cleydson Breno Rodrigues dos.

Desenvolvimento racional de fármacos antimaláricos derivados da artemisinina usando métodos computacionais SAR e QSAR / Cleydson Breno Rodrigues dos Santos -- Macapá, 2014.

251 f.

Orientador: Prof. Dr. José Carlos Tavares Carvalho

Co-orientador: Prof.Dr.Williams Jorge da Cruz Macêdo

Tese (Doutorado) – Fundação Universidade Federal do Amapá,
Programa de Mestrado em Biodiversidade e Biotecnologia.

1. Plasmodium falciparum W-2. 2. Artemisinina. 3. Química quântica – Métodos. 4. Antimaláricos. I. Sotil, José Walter Cárdenas. III. Fundação Universidade Federal do Amapá. IV. Título.



**REDE DE BIODIVERSIDADE E BIOTECNOLOGIA DA AMAZÔNIA LEGAL
PROGRAMA DE PÓS-GRADUAÇÃO EM BIODIVERSIDADE E BIOTECNOLOGIA
DA REDE BIONORTE**

**DESENVOLVIMENTO RACIONAL DE FÁRMACOS ANTIMALÁRICOS
DERIVADOS DA ARTEMISININA USANDO MÉTODOS COMPUTACIONAIS
SAR E QSAR**

CLEYDSON BRENO RODRIGUES DOS SANTOS

Banca Examinadora

Presidente da Banca Examinadora

1º Membro da Banca Examinadora

2º Membro da Banca Examinadora

3º Membro da Banca Examinadora

4º Membro da Banca Examinadora

Julgado em: ____/____/____ **Avaliação:** _____

- Mil poderão cair ao teu lado, e dez mil à tua direita; mas tu não serás atingido (Sl 91:7).
- Ouvi, filhos, a instrução do pai, e estai atentos para conhecerdes o entendimento (Pv 4:1).
- Agora sabem que tudo quanto me deste provém de ti (Jo 17:7).
- O temor do senhor é o princípio da sabedoria; e o conhecimento do santo é prudência (Pv 9:10).
- Todo aquele, pois, que ouve estas minhas palavras e as pratica será comparado a um homem prudente que edificou a sua casa sobre a rocha (Mt 7:24).

AGRADECIMENTOS

A **DEUS** Pai todo poderoso, por sempre ter me concedido ajuda ao longo da pós-graduação, pois sem ele jamais teria conseguido alcançado mais uma vitória.

Ao meu orientador **Prof. Dr. José Carlos Tavares Carvalho** e **Coorientador** Prof. Dr. Williams Jorge da Cruz Macêdo por acreditarem na realização deste propósito, e pelas inúmeras discussões, ensinamentos e dedicação nos momentos que mais precisei.

À **Universidade Federal do Amapá** e ao **Programa de Pós Graduação em Biodiversidade e Biotecnologia da Rede BIONORTE** pela oportunidade de realizar o curso de pós graduação em nível de doutorado.

Aos meus pais **Cláudio Cardoso** e **Regina Lúcia**, e as minhas **avós**, as minhas tias **Regina Bastos** e **Raimunda Braga** (*in memoriam*) e tantos outros familiares pelo meu aprendizado durante toda a minha vida.

Ao meu irmão **Cláudio Bruno Santos** que apesar de distante sempre esteve comigo.

Aos amigos, alunos, técnicos e professores do **Laboratório de Modelagem e Química Computacional (LMQC)**, do **colegiado do curso de Ciências Farmacêuticas e do colegiado do curso de Ciências Biológicas** da Universidade Federal do Amapá com quem dividi meu sonho e aflições. Sem a ajuda de vocês teria sido mais difícil. Obrigado pela ajuda incondicional.

À minha esposa (Luanny Vidal) por ser o melhor acontecimento de toda a minha vida, me dando apoio e incentivos nos momentos de dificuldades e aflições, e por ser essa companheira maravilhosa em vários momentos de minha ausência, e os meus filhos (Davi Santos e Brendo Samuel Santos). Obrigado por entender que esse era um sonho e um objetivo que tracei para minha vida, mas vocês entenderam, e me apoiaram e ficaram sempre do meu lado.

A todos que de alguma forma tiveram participação direta ou indiretamente nesse momento.

A todos vocês a minha eterna gratidão, e o meu muito obrigado.

SUMÁRIO

1 INTRODUÇÃO	15
2 OBJETIVOS	21
2.1 OBJETIVO GERAL	21
2.2 OBJETIVOS ESPECÍFICOS	21
3 REVISÃO BIBLIOGRÁFICA	22
3.1 MALÁRIA	22
3.2. CICLO EVOLUTIVO DO AGENTE ETIOLÓGICO DA MALÁRIA	22
3.3 TRATAMENTO	24
3.3.1 Classificação das principais drogas antimaláricas	25
3.4 ASPECTOS EPIDEMIOLÓGICOS DA MALÁRIA NA AMAZÔNICA LEGAL COM DESTAQUE PARA OS ESTADOS DO PARÁ E AMAPÁ	28
3.5 A IMPORTÂNCIA E APLICAÇÃO DOS MÉTODOS DE QUÍMICA QUÂNTICA PARA A MODELAGEM MOLECULAR DE COMPOSTOS BIOATIVOS	37
3.6 A ORIGEM DOS MÉTODOS DE QUÍMICA QUÂNTICA	38
3.7 MÉTODOS COMPUTACIONAIS UTILIZADOS EM CÁLCULOS DE PROPRIEDADES MOLECULARES	40
3.7.1 Métodos Semi-empíricos	41
3.7.2 Métodos <i>ab initio</i>	42
3.7.3 Teoria do Funcional de Densidade	42
3.8 SOFTWARES DE QUÍMICA COMPUTACIONAL PARA CÁLCULO DE PROPRIEDADES MOLECULARES	43
3.8.1 Desenho e visualização tridimensional	44
3.8.2 Análise conformacional e minimização de energia	45
3.9 APLICAÇÃO DO MÉTODO HARTREE-FOCK USANDO FERRAMENTAS COMPUTACIONAIS PARA A OBTENÇÃO DE PROPRIEDADES MOLECULARES	47
3.10 DOCKING MOLECULAR.....	48
3.10.1 Docking molecular no estudo de compostos antimaláricos derivados da artemisinina	50
3.11 POTENCIAL ELETROSTÁTICO MOLECULAR	51
3.11.1 Mapas de potencial eletrostático molecular no estudo de compostos antimaláricos derivados da artemisinina	51
3.12 MÉTODOS DE ANÁLISE MULTIVARIADA	53

3.12.1 Análise de componentes principais (PCA)	53
3.12.1.1 Escalonamento	54
3.12.2 Análise de agrupamento hierárquico (HCA)	55
3.12.3 Método dos mínimos quadrados parciais (PLS) e regressão dos componentes principais (PCR)	55
3.12.3.1 Método dos mínimos quadrados parciais (PLS)	57
3.12.3.2 Regressão dos componentes principais (PCR)	57
3.12.4 Avaliação dos modelos PLS e PCR	58
3.12.4.1 Avaliação do grau de ajuste dos modelos PLS e PCR	58
3.12.4.2 Avaliação do grau de significância dos modelos PLS e PCR	59
3.12.4.3 Avaliação do grau de previsibilidade do modelo PLS e PCR	60
4 MATERIAL E MÉTODOS	62
4.1 EQUAÇÕES DE HARTREE-FOCK	62
4.2 MÉTODO HARTREE-FOCK-ROOTHAAN	66
4.3 LIMITE HARTREE-FOCK	68
4.4 CONJUNTOS DE BASES DE VALÊNCIA SEPARADA UTILIZADOS NOS CÁLCULOS DAS PROPRIEDADES MOLECULARES	69
4.4.1 Funções tipo Slater	69
4.4.2 Conjunto de base mínima	71
4.4.3 Funções do tipo gaussianas	71
4.4.3.1 Funções gaussianas contraídas	71
4.4.3.2 Conjunto de base de valência separada	73
4.4.4 Função de polarização	74
4.4.5 Funções difusas	75
4.5 MODELAGEM MOLECULAR DA ARTEMISININA E DOS DERIVADOS ANTIMALÁRICOS ESTUDADOS	76
4.5.1 Validação do método computacional aplicado na modelagem molecular da artemisinina	80
4.5.2 Análise multivariada	80
4.5.2.1 Análise de componentes principais (PCA)	81
4.5.2.2 Análise de agrupamento hierárquico (HCA)	81
4.5.3 Compostos estudados	82
4.5.4 Cálculos de obtenção dos descritores moleculares ou propriedades moleculares	89
4.5.5 Interação entre artemisininas e heme	91

4.5.6 Mapas de potencial eletrostático molecular (MEP)	91
4.5.7 Seleção de variáveis e construção dos modelos QSAR (PLS e PCR)	92
5 RESULTADOS E DISCUSSÃO	95
5.1 OTIMIZAÇÃO DA GEOMETRIA DA ARTEMISININA EM DIFERENTES MÉTODOS E CONJUNTOS DE BASE PARA VALIDAR O MÉTODO COMPUTACIONAL APLICADO NA MODELAGEM MOLECULAR DOS COMPOSTOS ESTUDADOS	95
5.1.1 Aplicação do método PCA para selecionar os principais parâmetros geométricos da artemisinina	99
5.1.2 Aplicação do método HCA	106
5.2 DOCKING MOLECULAR	107
5.3 MAPA DE POTENCIAL ELETROSTÁTICO MOLECULAR (MEP)	110
5.4 RESULTADOS DA PCA.....	113
5.5 RESULTADOS DA HCA.....	117
5.6 RESULTADOS DOS MODELOS MÍNIMOS QUADRADOS PARCIAIS (PLS) E REGRESSÃO DE COMPONENTES PRINCIPAIS (PCR)	118
5.7 COMPOSTOS DO CONJUNTO TESTE	121
5.7.1 Mapa de Potencial Eletrostático Molecular (Conjunto Teste)	124
5.7.2 Aplicação dos modelos PLS e PCR aos compostos do conjunto teste	127
6 CONSIDERAÇÕES FINAIS	129
REFERÊNCIAS	131
ANEXOS	146

RESUMO

Artemisinina (qinghaosu) tem uma estrutura única tendo uma lactona endoperóxido estável (1,2,13-trioxano) totalmente diferente dos antimaláricos anteriores na sua estrutura e modo de ação. A artemisinina é isolada a partir da *Artemisia annua* L (Asteraceae), sendo um composto antimalárico com boa eficácia contra o *Plasmodium falciparum*. Neste trabalho propõem-se uma combinação de métodos de química quântica e análise multivariada para estudar os parâmetros geométricos da artemisinina na região do anel endoperóxido, a fim de validar a escolha do método e o conjunto de base quando comparado com os dados cristalográficos de raios-X. Os parâmetros geométricos mais importantes selecionados pela análise de componentes principais (PCA) foram O13C12, O1O2C3, C3O13C12C12a e C12C12aO1O2. Os resultados de PCA mostraram que o modelo foi construído com três componentes principais (3PCs), explicando 97.0861% da variância total. O conjunto de base HF/6-31G** mostrou elevada similaridade com os dados experimentais, assumindo que a combinação do método Hartree-Fock pode ser usada na modelagem molecular da artemisinina e seus derivados antimaláricos com mecanismo de ação na região do anel endoperóxido. Após a validação do método Hartree-Fock com o conjunto de base de valência 6-31G** foram calculados as propriedades moleculares da artemisinina e 20 derivados com atividade antimalárica. Mapas de potencial eletrostático molecular (MEP) e docking molecular foram utilizados para investigar a interação entre os ligantes (novos derivados da artemisinina) e o receptor (heme). A PCA e análise de agrupamento hierárquico (HCA) foram utilizadas para selecionar as propriedades moleculares mais importantes relacionadas com a atividade antimalárica. As correlações entre atividade biológica e as propriedades moleculares foram obtidas utilizando os métodos dos mínimos quadrados parciais (PLS) e regressão dos componentes principais (PCR). Os modelos construídos de regressão PLS e PCR foram também usados para prever a atividade antimalárica de 30 novos compostos derivados da artemisinina com atividade desconhecida. Os modelos obtidos mostraram significância estatística, e também a capacidade preditiva. As propriedades moleculares significativas relacionadas com atividade antimalárica dos compostos foram: energia de hidratação (EH), a carga sobre o átomo de oxigênio O11 (QO11), o ângulo de torção O1-O2-Fe-N2 (D2), e o índice máximo de R/Eletronegatividade de Sanderson (R_{Te}⁺). Estas variáveis levam a explicação física e estrutural das propriedades moleculares que devem ser selecionadas ao desenvolver novos ligantes para serem usados como agentes antimaláricos.

Palavras chaves: *Plasmodium falciparum* W-2. Artemisinina. Métodos de química quântica. SAR. QSAR.

ABSTRACT

Artemisinin (qinghaosu) has a unique structure having a stable lactone endoperoxide (1,2,13-trioxane) totally different from the previous antimalarials in its structure and mode of action. Artemisinin is isolated from *Artemisia annua* L (Asteraceae), being a compound with good antimalarial efficacy against *Plasmodium falciparum*. In this work, it is proposed a combination of quantum chemistry methods and multivariate analysis to study geometrical parameters of artemisinin in the region of the ring endoperoxide in order to validate the choice of method and the basis set when compared with the data of X-ray crystallographic. The most important geometrical parameters selected by principal component analysis (PCA) were O13C12, O1O2C3, C3O13C12C12a and C12C12aO1O2. The results of PCA showed that the model was constructed with three main components (3PCs), explaining 97.0861% of the total variance. The HF/6-31G** basis set showed high similarity with the experimental data, assuming that the combination of the Hartree-Fock method can be used in molecular modeling of artemisinin and its derivatives with antimalarial mechanism of action in the ring region endoperoxide. After validation of the Hartree-Fock method the valence basis set 6-31G ** were calculated molecular properties of artemisinin and 20 derivatives with antimalarial activity. Maps of molecular electrostatic potential (MEP) and molecular docking were used to investigate interaction between the ligands (new artemisinin derivatives) and receptor (heme). The PCA and hierarchical cluster analysis (HCA) were used to select the most important molecular properties related to antimalarial activity. Correlations between biological activity and molecular properties were obtained using partial least squares (PLS) and principal component regression (PCR) methods. The regression PLS and PCR models built in this study were also used predict the antimalarial activity of 30 new artemisinin compounds with unknown activity. The models obtained showed statistical significance and also predictive ability. The molecular properties significant related with antimalarial activity of the compounds were: hydration energy (HE), charge on the oxygen atom O11 (QO11), torsion angle O1-O2-Fe-N2 (D2), and the maximum index of R/Sanderson Electronegativity (RTE+). These variables lead to physical and structural explanation of molecular properties that must be selected to developing new ligands for use as antimalarial agents.

Keywords: *Plasmodium falciparum* W-2. Artemisinin. Quantum Chemical Methods. SAR. QSAR.

LISTA DE ILUSTRAÇÕES

Figura 1.	Estrutura da artemisinina	16
Figura 2.	Etapas do desenho racional de fármacos baseados em estrutura	18
Figura 3.	Ciclo de vida das espécies de <i>Plasmodium</i> causadoras de malária em humanos	23
Figura 4.	Principais drogas antimaláricas.....	25
Figura 5.	Mapa do Brasil destacando as áreas de risco para malária.....	29
Figura 6.	Linha de comparação de casos de malária referente ao ano de 2008. Brasil, 1960 a 2008.....	34
Figura 7.	Planta <i>Artemisia annua</i> L	76
Figura 8.	Estrutura da proteína heme.....	77
Figura 9.	Mecanismo de ação artemisinina e heme proposto por Posner (1995)	79
Figura 10.	Mecanismo de ação artemisinina	79
Figura 11.	Estrutura e atividade biológica da artemisinina e seus derivados.....	84
Figura 12.	Compostos do conjunto teste derivados da artemisinina com atividade antimalárica desconhecida contra <i>Plasmodium falciparum</i> do tipo W-2.....	86
Figura 13.	Plot dos scores PC1-PC2 para os 05 (cinco) métodos e 06 (seis) conjunto de base, correlacionados com os parâmetros geométricos do anel 1,2,13-trioxano da artemisinina.....	102
Figura 14.	Plot dos loading de PC1-PC2 usando quarto (04) parâmetros geométricos selecionados pela análise de componentes principais para cinco métodos (05) e seis (06) conjuntos de base, correlacionados com os parâmetros geométricos do anel 1,2,13-trioxano da artemisinina.....	103
Figura 15.	Dendograma do HCA para os 05 (cinco) métodos e 06 (seis) conjunto de base, correlacionado com os parâmetros geométricos experimental do anel 1,2,13-trioxano.....	106
Figura 16.	Interação entre heme-artemisinina dos compostos mais ativos do conjunto treinamento (1, 3, 4, 10, 11, 15, 16, 19 e 20)	109
Figura 17.	Mapas de potencial eletrostático da artemisinina e seus derivados com atividade antimalárica contra o <i>Plasmodium falciparum</i> (W-2).....	111
Figura 18.	Plot dos scores de PC1-PC2 para a artemisinina (1) e seus derivados com atividade antimalárica contra cepas de <i>P. falciparum</i> (W-2). Valores positivos indicam análogos mais potentes, e valores negativos indicam análogos menos potentes.....	116

Figura 19.	Plot dos loadings de PC1–PC2 com quatro descritores selecionados para construir os modelos PLS e PCR da artemisinina e seus derivados com atividade biológica contra cepas de <i>P. falciparum</i> (W-2).....	116
Figura 20.	Dendrograma da HCA da artemisinina e derivados com atividade biológica contra cepas de <i>P. falciparum</i> (W-2). Os valores positivos indicam análogos mais potentes, e os valores negativos indicam compostos menos ativos.....	118
Figura 21.	Plot dos valores preditos versus experimental para logAR através dos modelos (a) PLS e (b) PCR	121
Figura 22.	Mapas do Potencial Eletrostático Molecular para os compostos do conjunto teste mostrados na Figura 10 com atividade antimalárica desconhecida.....	125
Figura 23.	Compostos promissores para futuras sínteses e ensaios biológicos	127

LISTA DE TABELAS

Tabela 1.	Casos de malária, segundo estados na Amazônia Legal, janeiro a abril de 2008 a 2009.....	30
Tabela 2.	Porcentagem de variação casos de <i>Plasmodium falciparum</i> e casos de internação na Amazônia Legal, janeiro a abril de 2008 a 2009.....	31
Tabela 3.	Municípios que concentram 90% dos casos de malária no Estado do Pará, janeiro a maio de 2008.....	32
Tabela 4.	Casos de malária nos municípios do Marajó, Estado do Pará, janeiro a maio de 2008 a 2009.....	33
Tabela 5.	Dados de lâminas positivas no Amapá no período referente de 1970 a 2003.....	36
Tabela 6.	Municípios que concentram 80% dos casos de malária no estado do Amapá	37
Tabela 7.	Parâmetros teóricos e experimentais do anel 1, 2, 13-trioxano da artemisinina.....	98
Tabela 8.	Parâmetros geométricos selecionados pela análise de componentes principais, desvio padrão e matriz de correlação de Pearson.....	100
Tabela 9.	Análise dos componentes principais da seleção do modelo dos métodos e conjunto de base.....	101
Tabela 10.	Parâmetros geométricos selecionados pela análise de componente principais, métodos e conjunto de base, e a variação dos parâmetros geométricos com os respectivos dados experimentais (Δ e $\Delta\%$).....	105
Tabela 11.	Parâmetros calculados pelo docking molecular da heme-artemisinina e dos compostos derivados mais ativos.....	108
Tabela 12.	Propriedades moleculares selecionadas pela análise de componentes principais, valores experimentais ($\log AR$), IC_{50} e a matriz de correlação.....	114
Tabela 13.	Análise de componentes principais do modelo SAR, contribuição das propriedades moleculares selecionadas com a etapa de análise multivariada.....	115
Tabela 14.	Resultado predito dos modelos PLS e PCR, o erro de validação para $\log AR$ (experimental).....	120
Tabela 15.	Propriedades moleculares selecionadas pela análise de componentes principais do conjunto teste com atividade antimalárica desconhecida.....	123
Tabela 16.	Atividade antimalárica predita ($\log AR$) pelos modelos PLS e PCR para os compostos do conjunto teste e resíduos de previsão entre os modelos.....	128

LISTA DE SIGLAS

AR	Atividade Relativa
GTF	Gaussian Type Functions
GTO	Gaussian Type Orbital
HCA	Hierarchical Cluster Analysis
HF	Hartree-Fock
HFR	Hartree-Fok-Roothaan
MEP	Molecular Electrostatic Potentials
MO-LCAO	Molecular Orbital Linear Combination Atomic Orbital
OMS	Organização Mundial da Saúde
OPAS	Organização Pan-Americana da Saúde
PC	Principal Component
PCA	Principal Component Analysis
PCR	Principal Component Regression
PLS	Partial Least Squares
PRESS	Predicted Residual Error Sum of Squares
QSAR	Quantitative Structure-Activity Relationship
SAR	Structure-Activity Relationships
SCF	Self Consistent Field
SEP	Standard Error of Prediction
S_{PRESS}	Standard Deviation
STF	Slater Type Functions
STO	Slater Type Orbital
PDB	Protein Data Bank

1 INTRODUÇÃO

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 fêmea do gênero *Anopheles*. Os transmissores da malária humana são insetos da ordem dos dípteros, da família *Culicidae* e do gênero *Anopheles*. Este gênero compreende cerca de 400 espécies, das quais apenas reduzido número tem importância para a epidemiologia da malária em cada região (BRASIL, 2013; COSTA, 2013).

A malária é uma das maiores endemias que assolam o mundo e está concentrada nas áreas tropicais e sub-Saharianas, causando a morte de milhares de pessoas todos os dias. Na Amazônia legal, é registrada a maioria dos casos da doença, os aspectos endêmicos estão relacionados ao clima e as características socioculturais e econômicas da população, principalmente nas áreas de ocupações e de assentamentos (CESARIO; CESARIO, 2006; WORLD HEALTH ORGANIZATION, 2010; BRASIL, 2013; COSTA, 2013).

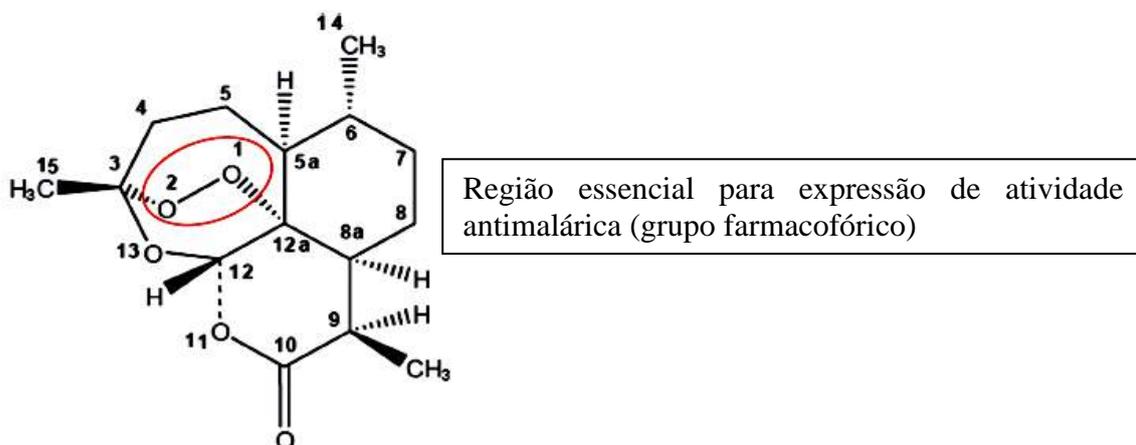
Na década de 50, grande parte das tentativas de erradicar esta endemia falhou, principalmente, devido ao desenvolvimento de resistência aos inseticidas e aos fármacos antimaláricos. Desde 1960 a transmissão da malária aumentou em muitas regiões onde a infecção é endêmica. Cepas de *Plasmodium falciparum* resistentes à cloroquina e a múltiplos agentes disseminaram-se, e o grau de resistência medicamentosa aumentou. Recentemente, cepas de *Plasmodium vivax* resistentes à cloroquina também foram identificadas. O principal problema no tratamento da malária é a multirresistência do *P. falciparum* a fármacos antimaláricos (FERREIRA, 1964; TRACY; WEBSTER JR, 1996; OLLIARO, 2001; COSTA, 2013).

Apesar de vários estudos, que vêm sendo feitos há muitos anos, ainda hoje não existe uma vacina que confira proteção satisfatória contra a malária. É comum confundir a vacina contra a febre amarela (antiamarílica) como se fosse contra a malária (antimalárica), porém somente a primeira existe, é eficaz e fundamental tomá-la quando se viaja para áreas endêmicas da febre amarela, a qual, de modo geral superpõe-se com as áreas malarígenas. A impossibilidade de remoção do vetor de transmissão da malária exige a necessidade de novos agentes que possam ter um novo mecanismo de ação. Portanto, a busca de novos compostos com ação efetiva no combate ao *P. falciparum*, tem elevada prioridade, devido a sua gravidade torna-se uma missão urgente de programas de pesquisas global (BOWMAN et al., 1999; BRASIL, 2013; COSTA, 2013).

Neste contexto, plantas medicinais são usadas para o tratamento de diversas doenças, dentre elas a malária. Este tratamento foi empregado pela medicina tradicional chinesa há mais de 2000 mil anos. O medicamento utilizado artemisinina (qinghaosu) é extraído da planta *Artemisia annua* L, utilizada no combate de 52 espécies de doenças na República Popular da China (KLAYMAN, 1985).

A artemisinina ou *qinghaosu* (Figura 1) é uma lactona sesquiterpênica que contém o anel 1, 2, 13 trioxano que é totalmente diferente dos antimaláricos anteriores em sua estrutura e modo de ação (LIU, 1981; KLAYMAN, 1985). O uso prático como um agente antimalárico deve-se a sua baixa solubilidade em solventes polares e apolares, além de apresentar uma baixa toxicidade e baixa eficácia por administração oral em tratamentos de pacientes (CHINA COOPERATIVE RESEARCH GROUP ON QINGHAOSU AND ITS DERIVATIVE AS ANTIMARIALS, 1982).

Figura 1. Estrutura da artemisinina



Fonte: SANTOS et al. (2014b)

A ligação endoperóxido (C-O-O-C) da artemisinina é acreditada ser a chave para o modo de ação da droga. O ferro no estado de oxidação +2 (Fe^{+2}) catalisa a quebra desta ligação, originando radicais livres altamente reativos, estes radicais livres derivados da artemisinina modificam e inibem uma variedade de moléculas do parasita, causando-lhes a morte. Uma rica fonte de Fe^{+2} intracelular é o heme que é um componente essencial da hemoglobina, e também responsável por ativar a artemisinina no interior do parasita (RIDLEY, 2003).

O processo para a obtenção de novos fármacos tem mudado muito através dos anos. Até aproximadamente a década de 80, novos medicamentos eram descobertos através de testes do tipo *caixa – preta* em células ou modelos – animais. Essa metodologia tradicional de desenvolvimento de fármacos, também chamada de “*triagem cega*”, consiste em testar aleatoriamente várias micromoléculas, em ensaios biológicos, sem nenhum conhecimento dos mecanismos de ação e/ou interação da molécula ligante. Embora a maioria dos fármacos terapêuticamente úteis disponíveis hoje tenham sido descobertos através desse processo, aos poucos essa metodologia mostrou-se ineficiente devido à probabilidade cada vez menor de se encontrar um novo fármaco, sendo necessário testar milhões de compostos em alguns casos. Esse fato resultou em um grande aumento no tempo e nos custos envolvidos para a descoberta e o desenvolvimento de novos fármacos (SUN; COHEN, 1993; LOMBARDINO; LOWE, 2004).

Para planejar um fármaco recorre-se uma das seguintes estratégias: a modificação molecular e o planejamento racional. A modificação molecular consiste em considerar um composto, de estrutura química conhecida e ação biológica comprovada, como modelo ou protótipo, sintetizar e ensaiar novos compostos que sejam congêneres, homólogos ou análogos estruturais do fármaco matriz. Este método quando utilizado requer elevado tempo e alto investimento financeiro na obtenção de um novo fármaco. Para a concretização do planejamento racional informações de diferentes áreas do conhecimento humano são utilizadas, sobretudo aquelas relacionadas aos níveis eletrônicos do fármaco, atividade biológica, parâmetros físico-químicos, tais como: hidrofóbicos, estéricos e eletrônicos, relacionados com a atividade biológica (KOROLKOVAS, 1989; COHEN et al., 1990).

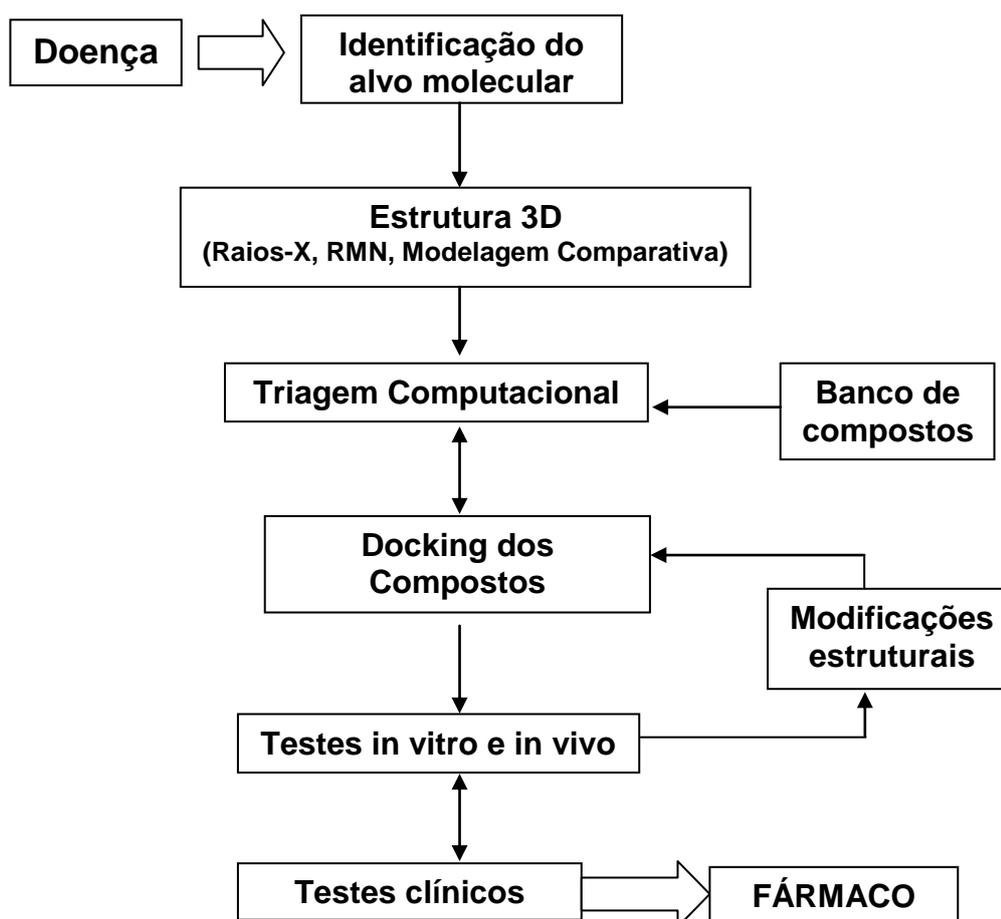
A Figura 2 apresenta um fluxograma com as várias etapas do Desenho Racional de Fármacos Baseado em Estrutura (DRBE). A primeira etapa constitui-se na escolha adequada do alvo terapêutico. O alvo molecular, ou biorreceptor pode ser uma proteína, relacionada à doença para a qual se deseja desenvolver um tratamento quimioterápico, cuja função deve ser bloqueada ou ativada. Após a identificação do alvo molecular, a estrutura molecular tridimensional do biorreceptor precisa ser obtida.

As estruturas moleculares podem ser obtidas através de técnicas experimentais, como difração de raios-X em cristais e a ressonância magnética nuclear (RMN), ou por métodos teóricos como modelagem comparativa (SALI; BLUNDELL, 1993).

Nesta etapa, as metodologias de docking (interação fármaco–receptor) precisam ser mais rápidas do que as utilizadas para o refinamento de compostos protótipos. Grandes

bancos de estruturas moleculares de ligantes, podendo conter milhões de moléculas, são testados contra o alvo molecular com a utilização dessas metodologias, a fim de se identificar compostos protótipos. Quando um composto promissor é encontrado, o químico medicinal analisa as modificações que podem ser efetuadas na molécula, para que a resposta biológica desejada seja potencializada. Nessa fase, entram em cena metodologias de docking mais acuradas, que visam tanto identificar a conformação de ligação quanto quantificar com maior precisão a afinidade de ligação receptor – ligante, objetivando uma otimização das moléculas promissoras selecionadas. Testes *in vitro* e *in vivo* são realizados para guiar e validar a otimização dos compostos protótipos e para que outras características, tal como toxicidade, sejam analisadas (SALI; BLUNDELL, 1993).

Figura 2. Etapas do desenho racional de fármacos baseados em estrutura



Fonte: SANTOS et al. (2014a)

Dentre as várias técnicas usadas no planejamento racional de fármacos, pode-se relacionar: técnica de planejamento com o auxílio de computação, utilizando principalmente parâmetros físico-químicos envolvidos na atividade biológica e métodos da química quântica, para determinar os compostos mais promissores de uma série (KOKPOL, 1988; COHEN et al., 1990; AVERY et al., 1993; BERNARDINELLI et al., 1994).

Leite et al. (2010) realizaram estudos de 18 compostos naturais oriundos da flora brasileira, que possuem um grupo peróxido e são caracterizados para atuar na proteína heme, levando uma redução na ligação peróxido e a produção de radicais que podem matar os agentes etiológicos da malária (cepas de *Plasmodium falciparum*). Estes resultados motivaram os pesquisadores a estudar as interações entre os 18 peróxidos naturais, realizando uma busca conformacional utilizando o método MM3 para cada molécula e a conformação mais estável foi otimizada pelo método PM3(tm). As interações entre os grupos peróxido-heme foi então avaliada, e os resultados indicaram que quatro dos compostos podem apresentar atividade antimalárica desejável.

Recentes estudos em 51 peróxidos foram realizados para identificar correlações entre parâmetros *in silico* e dados experimentais para identificação de novos agentes antimaláricos de fontes naturais. A interação do grupo foi estudado pelo docking molecular seguido por análise conformacional usando método paramétrico semi-empírico 6 (PM6). Os resultados indicaram que dois destes compostos são promissores antimaláricos (LEITE et al, 2012).

Recentemente, Cristino et al. (2012) usaram o método B3LYP/6-31G* para modelar a artemisinina e derivados da deoxiartemisinina 19, 10-substituída, com diferentes graus de atividade contra cepas *Plasmodium falciparum* D-6 da Serra Leoa. Métodos quimiométricos (PCA, HCA, KNN, SIMCA, and SDA) foram empregados para reduzir a dimensionalidade e para determinar que subconjuntos de descritores são responsáveis pela classificação entre os compostos mais e menos ativos.

O projeto de pesquisa propõe-se realizar a modelagem molecular da artemisinina e 20 derivados com diferentes atividades antimaláricas reportados da literatura, que foram testados *in vitro* contra *Plasmodium falciparum* (W-2), usando uma metodologia combinando métodos de química quântica e análise multivariada para estudar as relações estrutura-atividade qualitativa (SAR) e quantitativa (QSAR) entre a estrutura química e a atividade biológica, visando classificá-las, correlacioná-las com as suas propriedades moleculares e propor novos compostos com alta atividade que possam ser sintetizados e utilizados no tratamento da

malária (LIN; KLAYMAN; MILHOUS, 1987; LIN; LI; ANDERSEN, 1992; LIN; MILLER, 1995; LIN; ZIKRY; KYLE, 1997).

Na seção 2 serão mostrados os objetivos geral e os específicos. Na seção 3 encontra-se apresentado a revisão bibliográfica do trabalho de pesquisa. Na seção 4 introduziu-se os materiais e métodos utilizados no desenvolvimento desta pesquisa. A seção 5 traz os resultados e discussão deste trabalho. O último tópico contém as conclusões de perspectivas para futuros trabalhos.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Estudar e propor novos derivados da artemisinina com atividade antimalárica utilizando métodos de química quântica e análise multivariada.

2.2 OBJETIVOS ESPECÍFICOS

- a) Validar o método computacional aplicado na modelagem molecular da artemisinina com atividade antimalárica;
- b) Realizar a modelagem molecular da artemisinina e 20 derivados utilizando o método computacional validado;
- c) Determinar as propriedades moleculares da artemisinina e seus derivados com atividade antimalárica utilizando o método computacional validado;
- d) Construir *Docking Molecular* dos compostos estudados;
- e) Construir os mapas de potencial eletrostático molecular (MEP) dos compostos estudados;
- f) Construir a matriz de dados (PCA e HCA) para selecionar as propriedades moleculares mais importantes relacionadas com atividade antimalárica;
- g) Construir modelos QSAR (PLS e PCR);
- h) Desenvolver computacionalmente novos derivados da artemisinina com atividade antimalárica;
- i) Propor valores de atividade antimalárica para compostos derivados da artemisinina não sintetizados para realização de futuras sínteses e ensaios biológicos.

3 REVISÃO BIBLIOGRÁFICA

3.1 MALÁRIA

Anualmente mais de 250 milhões de pessoas são acometidas pela malária em todo o mundo, destas cerca de 2 milhões evoluem a óbito. Estes dados demonstram que a malária ainda é a mais importante das doenças tropicais. Todos os transmissores da malária humana são insetos da ordem dos dípteros, família *Culicidae* e gênero *Anopheles*. Este gênero compreende cerca de 400 espécies, das quais apenas um reduzido número tem importância para a epidemiologia da malária em cada região (WORLD HEALTH ORGANIZATION, 2010; COSTA, 2013).

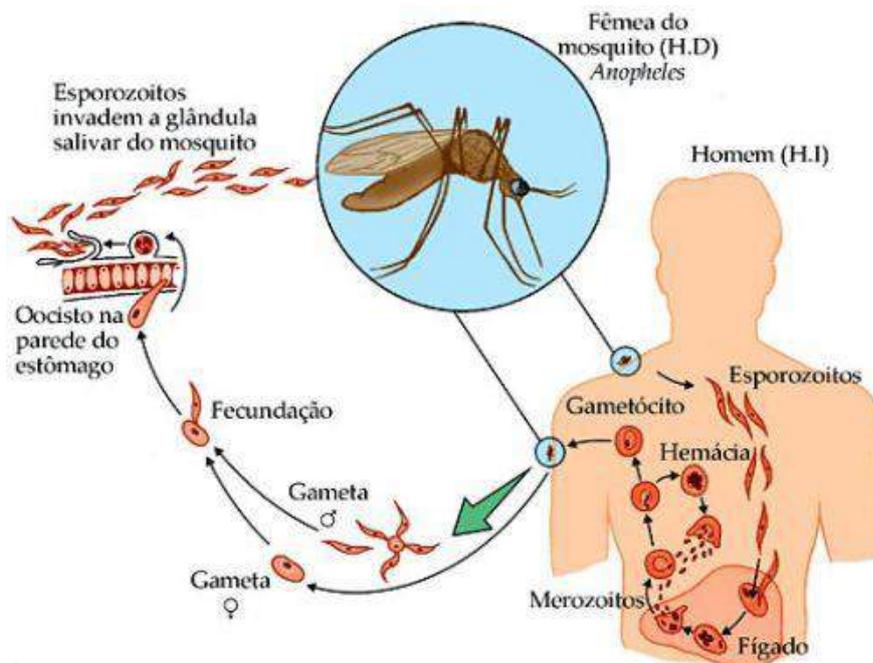
Os anofelinos são um grupo de dípteros hematófagos, com ampla distribuição mundial. O gênero *Anopheles*, contém várias espécies vetoras de plasmódios da malária. No contexto mundial, entre os vetores mais importantes, acham-se os grupos *A. maculipennis* na Europa, *A. culicifacies* na Índia, *A. minimus* em Ásia, China e Filipinas, *A. gambiae* e *A. funestus* na África. No Neotrópico, *A. albimanus* no México, *A. nuñeztovari* na Venezuela e Colômbia, *A. darlingi* e *A. aquasalis* no Brasil, são conhecidas como vetores dos plasmódios causadores da doença (FERREIRA, 1964; GARCIA, 1972; ALECRIM, 1979; DIXON; ROBERTS; LLEWELLYN, 1979; ROBERTS et al., 1981; LIU, 1981; ALBUQUERQUER, 1982; TADEI; MASCARENHAS; PODESTÁ, 1983; ZAHAR, 1984; COSTA, 2013).

Outras espécies importantes na transmissão da malária no Brasil são *A. bellator* e *A. cruzzi*, os equivalente ecológicos das duas últimas, no sul do Brasil, região atualmente livre desta enfermidade. De menor importância é o *A. albitarsis*, vetor potencial da malária nas áreas onde ocorre. Existem quatro espécies de *Plasmodium* causadores de malária humana: *falciparum*, *vivax*, *malariae* e *ovale*. A forma mais grave desta doença é causada pelo *P. falciparum* que pode ser fatal (CORREA; RAMOS, 1942; COSTA, 2013).

3.2 CICLO EVOLUTIVO DO AGENTE ETIOLÓGICO DA MALÁRIA

Em 1898, Giovanni Battista Grassi, Amico Bignami e Guiseppe Bastianelli documentaram a transmissão de parasitas da malária humana em *Anopheles claviger* (STANLEY et al., 1991) e logo depois descreveram os estágios de desenvolvimento do *Plasmodium falciparum* e do *Plasmodium vivax*, mostrados na Figura 3.

Figura 3. Ciclo de vida das espécies de *Plasmodium* causadoras de malária em humanos.



Fonte: Disponível em: <<http://nossomeioporinteiro.wordpress.com/2011/12/01/protozoarios/>>

Acesso em: 25/03/2014.

O ciclo assexuado do plasmódio, denominado esquizogônico, inicia-se após a picada do anofelino, com a inoculação dos esporozoítos infectantes no homem. A seguir, os esporozoítos circulam na corrente sanguínea durante alguns minutos e rapidamente penetram nas células do fígado (hepatócitos), dando início ao ciclo pré-eritrocítico ou esquizogonia tecidual, que dura seis dias para a espécie *P. falciparum*, oito dias para a *P. vivax* e 12 a 15 dias para a *P. malariae*. Durante esta fase, o *P. vivax* e o *P. ovale* apresentam desenvolvimento lento de alguns dos seus esporozoítos, formando os hipnozoítos, formas latentes (dormentes) do parasito responsáveis pelas recaídas da doença meses ou anos após (BRASIL, 2005a; COSTA, 2013).

Ao final do ciclo tecidual, os esquizontes rompem o hepatócito, liberando milhares de elementos-filhos na corrente sanguínea, chamados merozoítos. Ressalte-se que cada hepatócito rompido libera cerca de 2.000 merozoítos quando a infecção é devida ao *P. malariae*, 10.000, pelo *P. vivax* e 40.000 pelo *P. falciparum*. Os merozoítos irão invadir as hemácias, dando início ao segundo ciclo de reprodução assexuada dos plasmódios: o ciclo sanguíneo ou eritrocítico. O *P. malariae* só invade hemácias adultas (0,1% do total), o *P. vivax* invade preferencialmente as hemácias jovens e o *P. falciparum*, hemácias em qualquer fase evolutiva (BRASIL, 2005b; COSTA, 2013).

Durante um período que varia de 48 a 72 horas, o parasito se desenvolve no interior da hemácia até provocar a sua ruptura, liberando novos merozoítos que irão invadir novas hemácias. A ruptura e consequente liberação dos parasitos na corrente sanguínea traduz-se clinicamente pelo início do paroxismo malárico, que se repetirá com o término do novo ciclo (em dois dias, quando a infecção for devido ao *P. falciparum* ou *P. vivax* e em três dias, quando devido ao *P. malariae*).

Inicialmente, no ciclo sanguíneo, o parasito sofre uma série de transformações morfológicas – sem divisão celular – até chegar a fase de esquizonte, quando se divide e origina novos merozoítos que serão lançados na corrente sanguínea, após a ruptura do eritrócito. Assim, no exame microscópico do sangue pode-se observar variada morfologia do parasito – trofozoítos jovens (anéis), trofozoítos maduros, formas irregulares, esquizontes jovens e esquizontes maduros. Após um período de replicação assexuada, alguns merozoítos se diferenciam em gametócitos machos e fêmeas, que amadurecem sem divisão celular e tornam-se infectantes aos mosquitos. A função desses gametócitos é reprodutiva, isto é, garantir a perpetuação da espécie. Eles podem ser diferenciados microscopicamente em microgametócitos (masculinos) e os macrogametócitos (femininos) (BRASIL, 2005a; COSTA, 2013).

A reprodução sexuada (esporogônica) do parasito da malária ocorre no estômago do mosquito, após a diferenciação dos gametócitos em gametas e a sua fusão, com formação do ovo (zigoto). Este se transforma em uma forma móvel (ocineto) que migra até a parede do intestino médio do inseto, formando o oocisto, no interior do qual se desenvolverão os esporozoítos. O tempo requerido para que se complete o ciclo esporogônico nos insetos varia com a espécie de *Plasmodium* e com a temperatura, situando-se geralmente em torno de 10 a 12 dias. Os esporozoítos produzidos nos oocistos são liberados na hemolinfa do inseto e migram até as glândulas salivares, de onde são transferidos para o sangue do hospedeiro humano durante o repasto sanguíneo (BRASIL 2005b; COSTA, 2013).

3.3 TRATAMENTO

A quimioterapia da malária tem como objetivos: interromper a esquizogonia sanguínea responsável pela patogenia e manifestações clínicas da infecção; proporcionar a erradicação das formas latentes do parasito (hipnozoítas) das espécies *P. vivax* e *P. ovale* no ciclo tecidual, evitando as recaídas; e reduzir as fontes de infecção para os mosquitos, eliminando

as formas sexuadas dos parasitos. O tratamento adequado e oportuno da malária previne o sofrimento humano, a ocorrência de casos graves, o óbito e elimina a fonte de infecção (BRASIL, 2005b; COSTA, 2013).

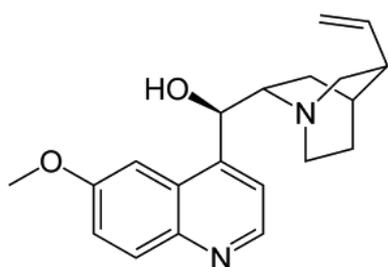
3.3.1 Classificação das principais drogas antimaláricas

O tratamento da malária é feito com o emprego de quimioterápicos, sendo o diagnóstico correto e o tratamento adequado dos pacientes, considerado primordial para o controle da doença. No Brasil, o Ministério da Saúde, por meio de uma política nacional de tratamento da malária, orienta a terapêutica e disponibiliza gratuitamente os medicamentos antimaláricos utilizados em todo o território nacional. Abaixo segue a classificação das principais drogas antimaláricas pelo grupo químico e alvo de ação no ciclo biológico do parasito.

- a) Pelo grupo químico – quinolinometanóis (quinina e mefloquina); 4-aminoquinolinas (cloroquina); 8-aminoquinolinas (primaquina); peróxido de lactona sesquiterpênica (derivados da artemisinina) são mostrados na Figura 4.

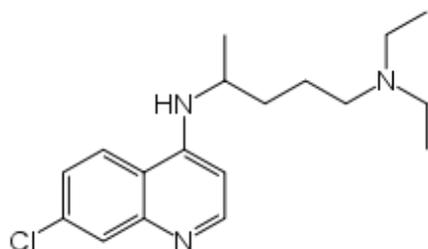
Figura 4. Principais drogas antimaláricas

Quinina



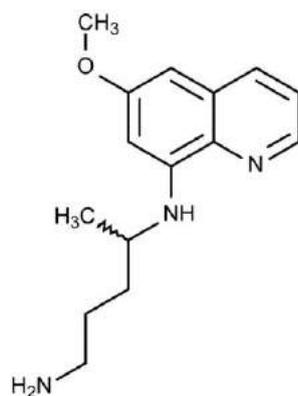
Isolada em 1982 da casca da *Chiconia officinalis* L. Rubiaceae (Quina), ativa contra formas eritrocíticas de *P.falciparum* e *P.vivax*.

Cloroquina



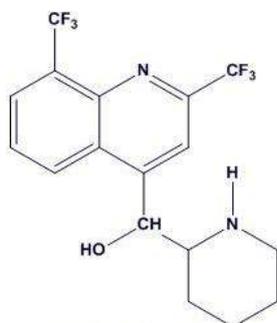
Isolada a partir da quinina. Possui atividade esquizonticida para todas as espécies gametocitocida para *P.vivax* e *P. malariae*

Primaquina



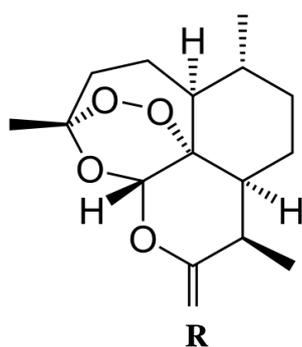
Sintetizada pela primeira vez nos Estados Unidos em 1946. Ativa contra gametócitos de todas as espécies de malária humana e hipnozoítos do *P. vivax*.

Mefloquina



É quimicamente relacionada com a quinina. Potente esquizonticida sanguíneo, de ação prolongada contra o *P. falciparum* resistente a cloroquina.

Artemisinina e seus derivados (dihidroartemisinina, artemeter, arteeter e artesunato de sódio)



São esquizonticidas sanguíneos potentes e de ação rápida, provocando a eliminação do parasito e melhora dos sintomas em menos tempo que a cloroquina ou a quinina.

Fonte: FRANÇA, SANTOS e FIGUEROA (2008)

- b) Pelo alvo de ação no ciclo biológico do parasito – esquizonticidas teciduais ou hipnozoiticidas (cura radical do *P. vivax* e *P. ovale*); esquizonticidas sanguíneos (promovem a cura clínica); gametocitocidas (bloqueiam a transmissão).

A decisão de como tratar o paciente com malária deve estar de acordo com o *Manual de Terapêutica da Malária* (BRASIL, 2008a) e ser precedida de informações sobre os seguintes aspectos:

- a) gravidade da doença – pela necessidade de drogas injetáveis de ação mais rápida sobre os parasitos, visando reduzir a letalidade;
- b) espécie de plasmódio – deve ser diferenciada, em face do perfil variado de resposta do *P. falciparum* aos antimaláricos. Caso não seja possível determinar a espécie do parasito, deve-se optar pelo tratamento do *P. falciparum*, pelo risco de evolução grave, devido à alta parasitemia;
- c) idade do paciente – pelo pior prognóstico na criança e no idoso;
- d) história de exposição anterior à infecção – indivíduos não-ímmunes (primoinfectados) tendem a apresentar formas clínicas mais graves.
- e) susceptibilidade dos parasitos aos antimaláricos convencionais – para indicar tratamento com drogas sabidamente eficazes para área de ocorrência do caso, evitando atraso no efeito terapêutico e agravamento do quadro clínico;
- f) gravidez – a gravidez aumenta o risco de gravidade da malária e de morte. As gestantes não-ímmunes correm risco de aborto, parto prematuro e natimortalidade. Estão mais propensas à malária cerebral, a hipoglicemia e edema agudo do pulmão (BRASIL, 2005a).

Na década de 50, houve uma grande expectativa de erradicação da malária, baseada em programas de controle do vetor, no entanto, a doença continuou progredindo e atingindo milhões de pessoas em todo mundo (COSTA, 2013).

O controle da doença com base no emprego em massa de antimaláricos tem se mostrado problemático e falho, pois, cepas de *P. falciparum* resistentes à cloroquina são encontradas em quase todas as partes do mundo e os resultados de vários trabalhos realizados no Brasil mostram um alto grau de resistência a este antimalárico comprometendo o seu emprego como droga de primeira linha no tratamento da malária por esta espécie de plasmódio (WORLD HEALTH ORGANIZATION, 2010; COSTA, 2013).

A resistência também ocorreu rapidamente com o uso dos sulfamídicos associados à pirimetamina e existem dados que demonstram a resistência de drogas mais recentes, como a mefloquina e artemisinina e estudos evidenciam que, ainda que de forma lenta, o *P. falciparum* vem apresentando níveis decrescentes de sensibilidade ao tratamento com quinino no Brasil e em várias partes do mundo (BRASIL, 2005a; NOEDL et al., 2008; COSTA, 2013).

O aumento expressivo da resistência do *P. falciparum* aos antimaláricos tem estimulado, em todas as partes do mundo, a realização de estudos para otimização de esquemas terapêuticos atualmente usados, inclusão de drogas já conhecidas e novas associações e, principalmente, a descoberta e/ou o desenvolvimento de novas drogas antimaláricas. Entre os fármacos mais utilizados atualmente destacam-se a cloroquina, a mefloquina, as associações artesunato/mefloquina e artemeter/lumefantrina que, pelas suas características farmacocinéticas, possibilitam esquemas posológicos de fácil aderência, no entanto, as artemisininas e seus derivados são as drogas que negativam mais rapidamente a parasitemia da malária, com rápido desaparecimento da febre, usadas principalmente na malária com quadro clínico grave e multirresistente, porém se os sintomas forem adequadamente tratados com fármacos modernos dificilmente serão reincidentes (BRASIL, 2010; COSTA, 2013).

3.4 ASPECTOS EPIDEMIOLÓGICOS DA MALÁRIA NA AMAZÔNICA LEGAL COM DESTAQUE PARA OS ESTADOS DO PARÁ E AMAPÁ

A malária é reconhecida como grave problema de saúde pública no mundo, ocorrendo em cerca de 40% da população em 100 países e territórios. Sua estimativa é de 300 a 500 milhões de novos casos e 1 milhão de mortes por ano.

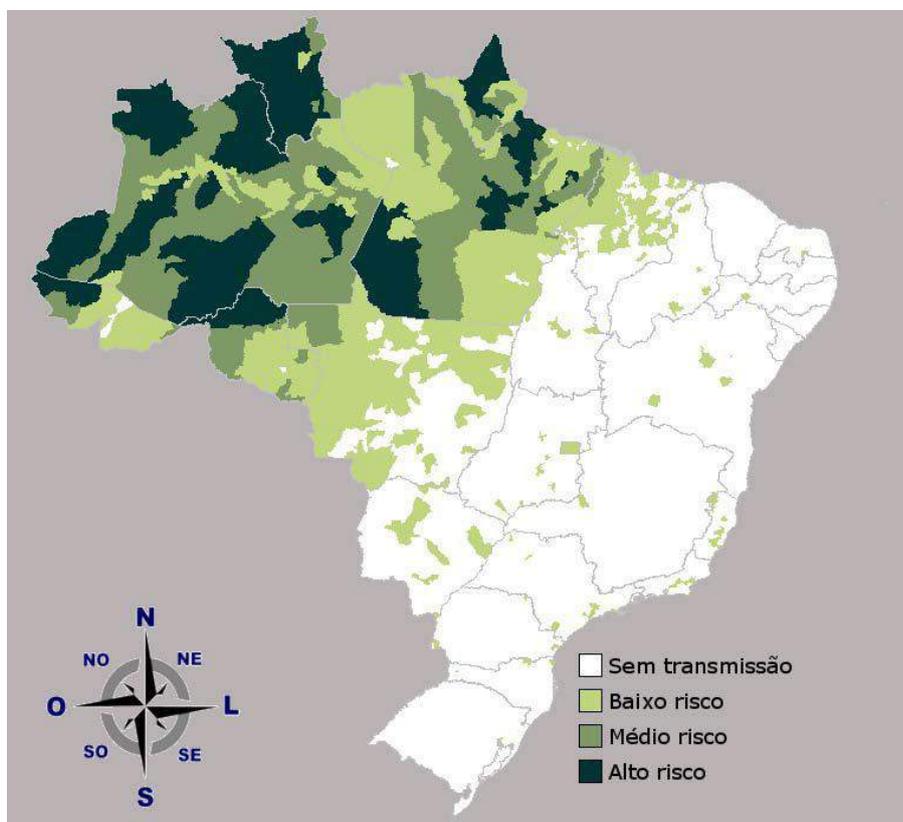
Os estados das regiões Norte e Nordeste do Brasil, por apresentarem grande desigualdade sociocultural e econômica, quando comparados aos estados do centro-sul, enfrentam enormes problemas de saneamento básico e habitacional, que acarretam a manifestações de inúmeras doenças características dos países nos quais a miséria assola. Neste contexto, a malária assume papel predominante nestas regiões com aspectos epidemiológicos bastante complexos, tanto para identificação e notificação do número de casos quanto para seu tratamento e a sua erradicação (INSTITUTO BRASILEIRO DE GEOGRAFIA ESTATÍSTICA, 2010; BRASIL, 2013; COSTA, 2013).

Na região extra-amazônica, 92% dos casos registrados são importados dos estados pertencentes à área endêmica, destacam-se os municípios localizados às margens do lago da usina hidrelétrica de Itaipu, áreas cobertas pela Mata Atlântica nos estados do Espírito Santo, Minas Gerais, Rio de Janeiro, São Paulo e Bahia; a região Centro-Oeste, nos estados de Goiás e Mato Grosso do Sul, e a região Nordeste, nos estados do Piauí, Pernambuco e Ceará (SANTOS, 2005).

Os casos de malária no Brasil têm aproximadamente 99% concentrados na região amazônica, composta pelos estados do Acre, Amapá, Amazonas, Maranhão, Mato Grosso, Pará, Rondônia, Roraima e Tocantins. Esta região é considerada endêmica no país (BRASIL, 2008b; BRASIL, 2013).

No Brasil a transmissão da malária não é completamente estável, de acordo com a incidência parasitária anual (IPA) costuma-se classificar as áreas endêmicas como de alto risco (IPA > 50/1.000 hab.), médio risco (IPA entre 10-49/1.000 hab.) e baixo risco (IPA < 10/1.000 hab.), conforme mostrado na Figura 5. Nesta Figura destacamos o estado do Amapá com alto risco no extremo norte do estado e nas áreas adjacentes com o estado do Pará e Guiana Francesa. Entretanto, no estado do Pará destaca-se três regiões de alto risco, sendo uma localizada nas ilhas do Pará (arquipélago do Marajó), próximas às cidades de Macapá-AP e Santana-AP, e outra região localizada no centro do estado e a terceira localizada na região sul do estado realizando fronteiras com os estados do Amazonas e Mato Grosso (BRASIL, 2011).

Figura 5. Mapa do Brasil destacando as áreas de risco para malária.



Fonte: BRASIL (2011)

Dentre os fatores que contribuíram para os altos índices da doença na região amazônica, nas décadas de 70 e 80, destaca-se a ocupação desordenada da região, incentivada por diversos órgãos governamentais. Esse desenvolvimento intensificado acelerou o processo migratório, atraindo moradores de outras regiões do país, devido aos projetos de colonização, expansão da fronteira agrícola, construção de estradas e hidrelétricas, projetos agropecuários, extração de madeira e mineração e a instalação de inúmeros garimpos. As precárias condições socioeconômicas da população migrante determinaram a rápida expansão da doença (FRANÇA; SANTOS; FIGUEROA-VILLAR, 2008).

Na Tabela 1, verifica-se que o número de casos de malária, de acordo com os estados da federação. Na Amazônia Legal, a variação dos números de casos de malária, em porcentagem (%), de janeiro a abril de 2008 a 2009 por unidade federada, foi maior nos estados do Amapá (16,6%), Pará (11,8%), Mato Grosso (4,5%) e Maranhão (1,7%), ocorrendo uma diminuição nos estados Tocantins (-52,1%), Amazonas (-41,9%), Acre (-17,6%), Rondônia (-15,0%) e Roraima (-11,1%), sendo que a Amazônia Legal teve uma redução dos números de casos da malária de -18,8%.

Em função dos programas de monitoramento, ocorreu o desenvolvimento de laboratórios sentinelas, com consequente alimentação da base de dados do Ministério da Saúde (MS) para seguida da tomada de providências ao combate da malária. Deste modo, alguns estados conseguiram diminuir a taxa de prevalência e mortalidade da doença.

Tabela 1. Casos de malária, segundo estados na Amazônia Legal, janeiro a abril de 2008 a 2009.

UF	Total de casos				% de Variação de casos	% por UF 2009
	2008	%	2009	%		
PA	20.770	21,06	23.225	29,01	11,8	29,0
AM	40.230	40,79	23.389	29,22	-41,9	29,2
RO	15.430	15,64	13.112	16,38	-15,0	16,4
AC	11.870	12,03	9.778	12,21	-17,6	12,2
AP	3.770	3,82	4.394	5,49	16,6	5,5
RR	3.965	4,02	3.524	4,40	-11,1	4,4
MT	1.297	1,31	1.355	1,69	4,5	1,7
MA	1.215	1,23	1.236	1,54	1,7	1,5
TO	94	0,10	45	0,06	-52,1	0,1
AMAZÔNIA	98.646	100,0	80.058	100,0	-18,8	100,0

Fonte: SIVEP/SVS/MS - Adaptado em 22.01.2014

Na Tabela 2, verifica-se que o número de casos de internação de malária, de acordo com os estados da Amazônia Legal. A porcentagem de variação de casos de *Plasmodium Falciparum*, de janeiro a abril de 2008 a 2009 por unidade federada, foi maior nos estados do Maranhão (65,4%) e Pará (22,5%). Nesse mesmo período, ocorreu uma diminuição nos estados de Tocantins (-64,5%), Amazonas (-47,0%), Mato Grosso (-39,7%) e Roraima (-34,7%). De um modo geral, Amazônia Legal teve uma redução dos números de casos *Plasmodium Falciparum* de -20,8%.

Ainda na Tabela 2, pode ser visto o aumento na porcentagem de variação de internação para os estados do Amapá (66,7%), Acre (4,2%) e Pará (2,2%). Já para a redução na porcentagem de variação de internação, destacam-se os estados de Tocantins (-50,0%), Amazonas (-47,0%), Roraima (-24,8%) e Mato Grosso (-21,1%). Na Amazônia Legal, houve uma redução de -13,9%.

Tabela 2. Porcentagem de variação de casos de *Plasmodium Falciparum* e casos de internação na Amazônia Legal, janeiro a abril de 2008 a 2009.

UF	Casos de <i>falciparum</i> + (f+V)		% de variação de <i>falciparum</i>	% <i>falciparum</i>		Internação		% de variação de internação
	2008	2009		2008	2009	2008	2009	
PA	3.747	4.589	22,5	18,0	19,8	231	236	2,2
AM	6.591	3.496	-47,0	16,4	14,9	202	107	-47,0
RO	2.250	1.470	-34,7	14,6	11,2	234	176	-24,8
AC	1.925	1.444	-25,0	16,2	14,8	168	175	4,2
AP	757	984	-30,0	20,1	22,4	33	55	66,7
RR	666	481	-27,8	16,8	13,6	92	85	-7,6
MT	348	210	-39,7	26,8	15,5	57	45	-21,1
MA	283	468	65,4	23,3	37,9	30	27	-10,0
TO	31	11	-64,5	33,0	24,4	14	7	-50,0
AMAZÔNIA	16.598	13.153	-20,8	16,8	16,4	1061	913	-13,9

Fonte: SIVEP/SVS/MS - Adaptado em 22.01.2014.

O Estado do Pará apresenta uma área de 1.247.689,515 Km², com população estimada de 7.588.078 habitantes, formado por 143 municípios, onde alguns deles localizam-se em um estuário delta chamado de Ilha do Marajó, que se encontra desfavorecido no que concerne à acessibilidade, notificação e controle e tratamento de endemias, onde muito desses pacientes oriundos do Pará se deslocam para a capital amapaense em busca de tratamento fazendo com que este número de caso aumente no estado do Amapá (INSTITUTO BRASILEIRO DE GEOGRAFIA ESTATÍSTICA, 2010).

Na Tabela 3, observam-se os municípios que concentram 90% dos casos de malária, no estado do Pará e nos municípios da ilha do Marajó próximo da capital amapaense, podendo se destacados os municípios com maiores porcentagens de casos acumulados no ano de 2009: o município de Anajás-PA (27,1%), o município do Afuá-PA (71,4%), o município de Portel-PA (75,1%), o município de Breves-PA (78,3%), o município de Chaves-PA (79,8%) e o município de Almeirim-PA (89,4%).

O estado do Amapá e o arquipélago da ilha do Marajó pode-se associar o grande índice de notificação de casos de malária ocorrido nesse período ao surgimento de grandes áreas de ocupação com conglomerados de assentamentos, cujas estruturas de saneamento básico são inexistentes, onde ocorreu um desmatamento desenfreado com exposição em massa do homem, que passou a disseminar a malária no estado do Pará e áreas adjacentes.

Tabela 3. Municípios que concentram 90% dos casos de malária no Estado do Pará, janeiro a maio de 2008.

Nº	Municípios	Número de casos		% de Participação	% Acumulado
		2008	2009		
1	ANAJÁS	4.782	7.906	27,1	27,1
2	ITAITUBA	2.306	2.174	7,4	34,5
3	GOIANÉSIA DO PARÁ	1.580	1.881	6,4	40,9
4	PACAJÁ	2.539	1.797	6,2	47,1
5	IPIXUNA DO PARÁ	459	1.584	5,4	52,5
6	JACAREACANGA	1.677	1.433	4,9	57,4
7	NOVO PROGRESSO	733	1.274	4,4	61,8
8	CACHOEIRA DE PIRIA	594	875	3,0	64,8
9	PARAGOMINAS	751	704	2,4	67,2
10	CURRALINHO	386	643	2,2	69,4
11	AFUÁ	113	585	2,0	71,4
12	ALTAMIRA	556	557	1,9	73,3
13	PORTEL	307	515	1,8	75,1
14	TUCURUÍ	478	492	1,7	76,8
15	BREVES	161	445	1,5	78,3
16	CHAVES	169	440	1,5	79,8
17	AURORA DO PARÁ	118	381	1,3	81,1
18	ORIXIMINÁ	524	338	1,2	82,3
19	MOJÚ	281	336	1,2	83,5
20	SANTA CRUZ DO ARARI	76	315	1,1	84,6
21	ANAPU	355	287	1,0	85,6
22	PRAINHA	398	283	1,0	86,6
23	SENADOR JOSÉ PORFIRIO	280	249	0,9	87,5
24	BREU BRANCO	307	243	0,8	88,3
25	NOVO REPARTIMENTO	531	189	0,6	88,9
26	ALMEIRIM	196	188	0,6	89,5
27	PONTA DE PEDRAS	107	188	0,6	90,1

Fonte: SISMAL/SIVEP/SVS/MS – atualizado em 22.01.2014

Na Tabela 4, são mostrados número de casos de malária no arquipélago da ilha do Marajó, no período de janeiro a maio de 2008 a 2009. Observa-se nesta Tabela que o número de casos de malária na Ilha do Marajó em 2008 foi de 6.234 casos, e em 2009 foi de 11.287 casos, tendo um aumento de 81,1% nesta região. Os municípios que apresentaram maiores números de casos em porcentagem de variação entre 2008 e 2009 foram os municípios de: São Sebastião da Boa Vista-PA (594,1%), Afuá-PA (417,7%), Santa Cruz do Arari-PA (314,5%), Breves-PA (176,4%), Chaves-PA (160,4%) e Gurupá-PA (160,0%).

Tabela 4. Casos de malária no arquipélago da ilha do Marajó, Estado do Pará, janeiro a maio de 2008 a 2009.

Municípios	Número de casos		% de Variação de casos	(% Acumulado	% de participação (Pará e Ilha do Marajó)
	2008	2009			
Anajás	4.782	7.906	65,3	70,0	70,0
Currálinho	386	643	66,6	75,7	5,7
Afuá	113	585	417,7	80,9	5,2
Portel	307	515	67,8	85,5	4,6
Breves	161	445	176,4	89,4	3,9
Chaves	169	440	160,4	93,3	3,9
Santa Cruz do Arari	76	315	314,5	96,1	2,8
Ponta de Pedras	107	188	75,7	97,8	1,7
S. Bastião B. Vista	17	118	594,1	98,8	1,0
Muaná	38	89	134,2	99,6	0,8
Cachoeira do Arari	61	14	-77,0	99,7	0,1
Gurupá	5	13	160,0	99,9	0,1
Bagre	4	10	150,0	99,9	0,1
Melgaço	3	3	0,0	100,0	0,0
Salvaterra	3	2	-33,3	100,0	0,0
Soure	2	1	-50	100,0	0,0
MARAJÓ	6.234	11.287	81,1	-	36,6
ESTADO DO PARÁ	25.200	29.210	15,9	-	-

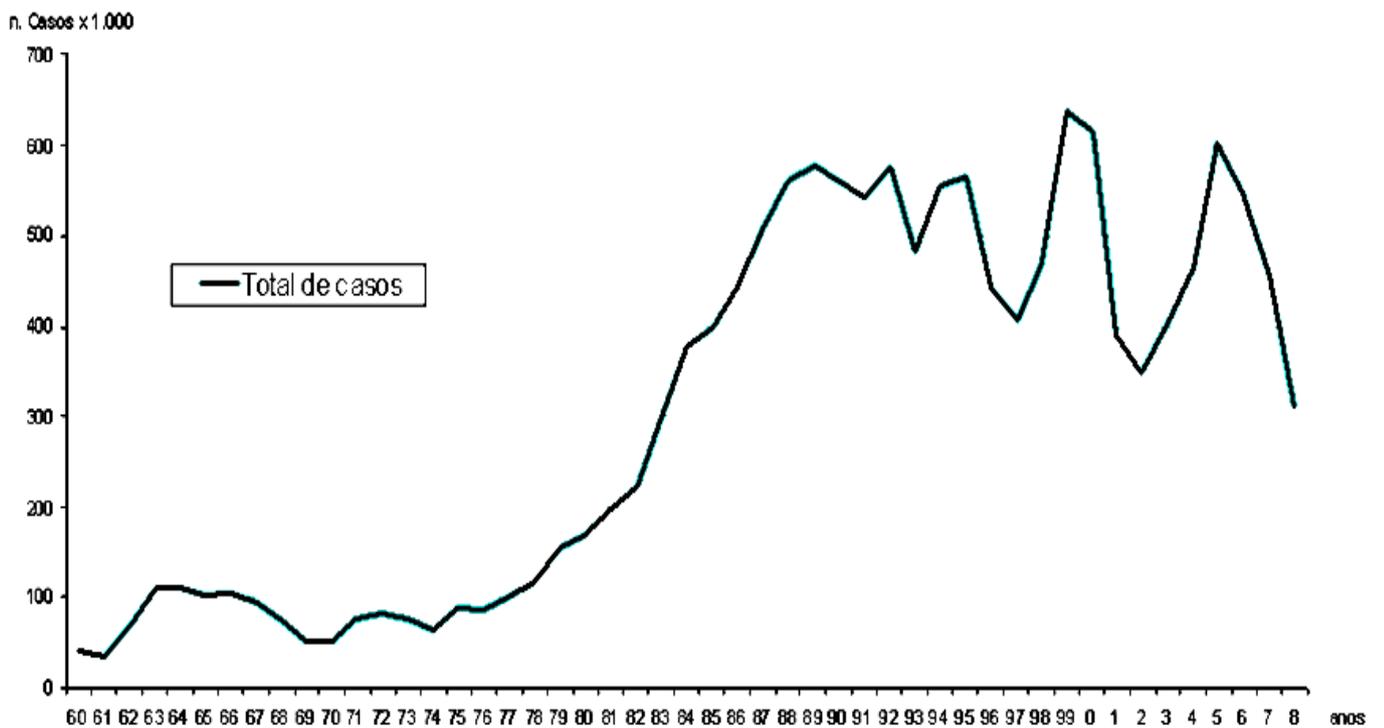
Fonte: SISMAL/SIVEP/SVS/MS – Adaptado em 22.01.2014

Na Figura 6 é mostrada a linha de comparação do número de casos de malária no Brasil no período de 1960 até 2008. Nesta Figura observa-se que houve um aumento no número de casos desta doença.

Na década de 70, com criação da Transamazônica e devido ao processo de desbravamento das florestas nativas, e introdução do homem em ambientes cujas condições

precárias de moradia e saúde, deixando como herança um quadro desolador da manifestação da malária nas áreas endêmicas que perpetuam até os dias atuais. Embora o Ministério da Saúde (MS), Organização Mundial da Saúde (OMS) e Organização Pan-Americana da Saúde (OPAS), apresentem políticas de redução da morbimortalidade da malária, somente no período de 2008-2010, os planos de ações globais vêm surtindo os efeitos desejáveis, com tendência de erradicação da malária (COSTA, 2013).

Figura 6. Linha de comparação de casos de malária referente ao ano de 2008. Brasil, 1960 a 2008.



Fonte: SISMAL/SIVEP/SVS/MS – Adaptado em 22.01.2014

Estudos realizados por Cardoso e Goldenberg (2007), descreveram a trajetória da malária no Estado do Amapá entre 1970-2003, nota-se uma carência de dados no Estado em períodos anteriores a 1970, devido o fato de ainda não haver um sistema eficiente de registro de informações na Secretaria de Saúde do Estado. Evidenciou-se ainda, através dos dados de lâminas positivas, a presença de todas as espécies de plasmódios existentes no país.

De maneira geral, na década de 70, a influência do desenvolvimento da região amazônica através da construção de estradas e usinas hidrelétricas trouxe, não somente um

fluxo migratório de pessoas em busca da oferta de emprego, mas também a difusão de endemias, e entre elas a malária (COSTA, 2013).

O pico de 1971 ocorre ainda na vigência da Superintendência de Campanhas de Saúde Pública (SUCAM), que segundo Loiola (2002) apud Cardoso e Goldenberg (2007), não possuía uma estrutura adequada para atender o aumento populacional decorrente da política de desenvolvimento da Amazônia, atraídos pelos grandes projetos regionais.

O estudo descreve ainda que, no ano de 1974, a malária volta a subir, constituindo um elevado pico em 1976. Registra-se, nesse período, a inauguração da Hidroelétrica Coaracy Nunes e a instalação do Projeto Jarí na região de fronteira entre os estados do Amapá e Pará. Esse desenvolvimento ocasionou além de desmatamento, o afluxo de migrantes, configurando o ambiente favorável à transmissão da doença (COSTA, 2013).

A partir dos anos 80, observa-se novamente um aumento no número de casos no estado, período que coincide com a introdução dos garimpos para extração de ouro, entre eles, um núcleo isolado no Lourenço, no município de Calçoene-AP. Avaliações epidemiológicas da época revelavam que os garimpos respondiam por 79% do total de casos de malária do Estado (COSTA, 2013).

Na década de 90, destaca-se a criação da Área de Livre Comércio de Macapá e Santana (ALCMS), em 1991, que registrou nos anos seguintes, segundo o Instituto Brasileiro de Geografia e Estatística (IBGE), mais da metade do crescimento populacional do Estado, devido à imigração proveniente, principalmente, dos estados do Pará, Maranhão e Ceará.

Nesse período várias famílias em busca de oportunidades de emprego chegaram ao Estado e passaram a criar novos bairros sem infraestrutura adequada e a invadir áreas de ressaca (termo regional usado para titular áreas alagadas, ricas em biodiversidade, comuns no estado do Amapá) criando com isso condições propícias para a existência de focos de vetores (COSTA, 2013).

Em 2000, observa-se novamente um aumento nos casos de malária não somente no Estado, mas em toda a Amazônia Legal, fazendo com que os três níveis de Governo (federal, estadual e municipal) somassem esforços para combater a doença. Diante desta grave situação, houve a implementação de diversas frentes, dentre elas a implantação do PIACM (no Amapá a partir do segundo semestre de 2001), que gerou uma redução da incidência da malária na ordem de 30,6%, no mesmo ano.

Na Tabela 5 são demonstradas as lâminas positivas de malária por espécie de plasmódio e índice parasitário anual – IPA (por mil habitantes) no Estado do Amapá, no

período de 1970 a 2003. Nesta Tabela, observa-se que os maiores números de lâminas positivas nos anos foram: 2001 (n=6.289), 2003 (n=6.027), 2000 (n=5.108) e 1990 (n=4.362). Em relação ao índice parasitário anual (IPA) em 2000 (n=30.074), 2001 (n=18.075), 1994 (n=10.450) e 2003 (n=10.309) apresentaram os maiores IPA.

Tabela 5. Dados de lâminas positivas no Amapá no período referente de 1970 a 2003.

Ano	Lâminas positivas	IPA	<i>P.falciparum</i>	<i>P.vivax</i>	<i>P. falciparum</i> + <i>P.vivax</i>	<i>P. malariae</i>
1970	1.104	650	3	1	1.758	116.000
1971	2.990	1.338	6	2	4.363	116.000
1972	1.603	1.018	15	0	2.636	130.102
1973	592	682	5	0	1.279	135.000
1974	1.652	1.443	17	0	3.112	137.000
1975	3.529	2.357	53	1	5.940	132.066
1976	3.683	3.282	65	0	7.030	133.131
1978	3.864	2.911	42	1	6.818	169.779
1980	611	1.106	10	0	1.729	184.853
1984	3.071	4.164	25	0	7.260	244.720
1986	3.849	4.330	40	0	8.219	265.498
1990	4.362	6.272	43	0	10.677	269.633
1994	1.778	10.450	21	5	12.254	317.597
2000	5.108	30.074	92	4	35.278	477.032
2001	6.289	18.075	104	19	24.487	498.735
2003	6.027	10.309	218	87	16.643	534.835

Fonte: CARDOSO e GOLDENBERG (2007) – Adaptado em 22.01.14

Em 2012, de acordo com dados do SIVEP na Tabela 6, dos dezesseis municípios do estado do Amapá, seis contribuíram com 80% dos casos, sendo eles: Macapá (n=3.485) Oiapoque (n=3.413), Mazagão (n=1.813), Porto Grande (n=1.445), Santana (n=1.304), Calçoene (n=1.208) e Pedra Branca do Amapari (n=704).

Os municípios de Santana-AP, Oiapoque-AP e Macapá-AP, estão entre os considerados prioridades para as ações de combate à malária, devido ao volume de casos recebidos de outras áreas, o que aumenta a vulnerabilidade dos mesmos.

Os municípios de Calçoene-AP e Porto Grande-AP são considerados áreas de grande risco de transmissão da malária, em função da localização e funcionamento de garimpos nestas regiões. Em Calçoene-AP, está localizado o garimpo do Lourenço, considerado o maior garimpo a céu aberto do estado do Amapá (COUTO, 2001; COSTA, 2013).

Tabela 6. Municípios que concentram 80% dos casos de malária no Estado do Amapá.

Nº	Municípios	Número de casos 2012	% de Participação	% Acumulado
1	MACAPÁ	3.485	20,85	20,85
2	OIAPOQUE	3.413	20,42	41,27
3	MAZAGÃO	1.813	10,85	52,11
4	PORTO GRANDE	1.445	8,64	60,76
5	SANTANA	1.304	7,80	68,56
6	CALÇOENE	1.208	7,23	75,79
7	PEDRA BRANCA	704	4,21	80,00

Fonte: SISMAL/SIVEP/SVS/MS – Adaptado em 25.03.2014

3.5 A IMPORTÂNCIA E APLICAÇÃO DOS MÉTODOS DE QUÍMICA QUÂNTICA PARA A MODELAGEM MOLECULAR DE COMPOSTOS BIOATIVOS

A importância da química quântica está na obtenção de soluções da equação de Schrödinger para a determinação precisa de propriedades de sistemas atômicos e moleculares. Dentro da área de Físico-química a química quântica é usada para cálculo de várias propriedades termodinâmicas tais como: entropia, capacidade calorífica de gases, para a interpretação de espectros moleculares, cálculos de comprimento de ligação, ângulos de ligação, momento dipolo e para o entendimento das forças intermoleculares (DYER, 1965; BUNGE, 1979; SILVERSTEIN; WEBSTER; KIEMLE, 2005; SOUZA; FARIAS, 2007).

Na orgânica, a química quântica pode estimar a estabilidade de moléculas, calcular as propriedades de intermediários de reações, reproduzir a aromaticidade em composto orgânico e simular espectros de Ressonância Magnética Nuclear (RMN) (DYER, 1965; BUNGE, 1979; SILVERSTEIN; WEBSTER; KIEMLE, 2005; SOUZA; FARIAS, 2007).

Na química analítica é usada para a interpretação das intensidades das linhas espectrais. Na área da química inorgânica é utilizada para a teoria do campo ligante, onde se fazem previsões, e se justificam propriedades de íons complexos de metais de transição. Na Bioquímica os cálculos de química quântica são usados para estudos de conformações de moléculas biológicas, tais como: ligação enzima-substrato e solvatação de moléculas

biológicas (DYER, 1965; BUNGE, 1979; SILVERSTEIN; WEBSTER; KIEMLE, 2005; SOUZA; FARIAS, 2007).

A aplicação de métodos de química quântica no estudo e planejamento de compostos bioativos tem se tornado uma prática rotineira nos dias atuais. Do ponto de vista do planejamento, é importante destacar que, quando se fala do uso da modelagem molecular (um termo coletivo que se refere aos métodos teóricos e técnicas computacionais para modelar ou mimetizar o comportamento das moléculas), não se pretende chegar a uma molécula bioativa simplesmente através do uso de programas de computador (SANT'ANNA, 2009).

O processo de desenvolvimento dessas moléculas, devido à sua complexidade, envolve necessariamente trabalho multidisciplinar, que emprega um grande conjunto de métodos computacionais de modo sistemático de forma a *facilitar e otimizar* o processo de desenvolvimento de compostos bioativos, em uma constante troca de informações com grupos de síntese química e avaliação da atividade destes compostos. Esses métodos computacionais podem ser usados como ferramentas do *planejamento racional de compostos bioativos*, assim chamado porque é orientado por uma hipótese racional sobre o mecanismo de ação destes compostos (SANT'ANNA, 2009).

A ação das moléculas bioativas é um fenômeno bastante complexo, mas um dos paradigmas da química medicinal é que essas moléculas têm seus efeitos associados a interações ou até reações químicas com estruturas macromoleculares presentes no sistema vivo, proteínas, na sua grande maioria. Caso estas proteínas sejam receptores celulares, as moléculas bioativas são classificadas como agonistas ou antagonistas de receptores; no caso das enzimas, estas moléculas atuam como inibidores enzimáticos (BARREIRO; FRAGA, 2001).

3.6 A ORIGEM DOS MÉTODOS DE QUÍMICA QUÂNTICA

A história e os debates relacionados ao chamado “método semi-empírico” de London-Eyring-Polanyi estão relatados no trabalho de Nye. A abordagem de Eyring e Polanyi, com o objetivo de mesclar teoria com resultados experimentais para construir superfícies de energia potencial, mostrou que é possível obter-se os mecanismos de reações adiabáticas, levando a importantes conceitos relacionados com a dinâmica das reações químicas, como estado de transição e complexo ativado. Já naquela época, ocorriam debates comparando as abordagens baseadas em primeiros princípios com a semi-empírica (NYE, 2007).

Coolidge e James (1933) publicaram um cálculo *ab initio* quando obtiveram a energia de ligação da molécula de H₂ com exatidão de 98%. Este cálculo demorou um ano para ser feito, mas seu sucesso aumentou a confiança dos autores que então criticaram o método semi-empírico como sendo um “feliz cancelamento de erros”, que ocorre por não levar em consideração “termos de importância considerável”.

Esta pouco útil expressão “cancelamento de erros” permanece até hoje no imaginário dos pesquisadores, apesar de não ser possível identificar com precisão que erros seriam estes e, principalmente, como e porque se cancelariam. Polanyi em 1937 contra-argumentava: “pessoalmente, não atribuo importância a uma concordância exata entre teoria e experimental no atual estágio, mas acredito que a teoria [referindo-se ao método semi-empírico] pode reivindicar oferecer uma descrição razoável do mecanismo das reações químicas que, de outra forma, permaneceria obscurecida” (MARCUS, 1964).

Em 1941, o pesquisador Hirschfelder apresentou o cálculo de pelo menos cem energias de ativação para várias reações utilizando o método de London-Eyring-Polanyi, mostrando que apenas em alguns casos os valores não concordavam com o experimental, tendo um erro de 10 kcal/mol (HYRSCHFELDER, 1941).

O pesquisador Hirschfelder defendia o método semi-empírico como “um método suficientemente flexível de tal forma que poderia ser tornado consistente com qualquer conjunto de fatos químicos, mantendo concordância com os princípios básicos da mecânica quântica”. Segundo Nye, nesta declaração Hirschfelder já identificava corretamente que um caráter *ad hoc* é inerente às abordagens semi-empíricas. Este caráter *ad hoc* envolve risco, uma vez que pode ser perfeitamente possível forçar uma concordância entre o modelo semi-empírico e experimental, mesmo que o modelo seja baseado em uma teoria errada. Tal risco é ainda maior se o modelo for utilizado para extrapolação (NYE, 2007).

O triunfo final da abordagem de primeiros princípios para a reação $H + H_2 \rightarrow H_2 + H$, somente veio ocorrer em 2003, quando Mielke e colaboradores apresentaram novas constantes de velocidades experimentais e teóricas. Os cálculos teóricos foram obtidos usando dinâmica quântica em uma superfície de potencial adiabática exata incluindo correções de Born-Oppenheimer. O experimento e a teoria concordam agora na faixa de 167K a 2112K, dentro do erro experimental, fazendo com que o problema possa ser atualmente resolvido (MIELKE et al., 2003).

A expressão “semi-empírico” foi utilizada pela primeira vez em química teórica em 1931 por Michael Polanyi (1891 – 1976) e Henry Eyring (1901 – 1981) em sua tentativa de

combinar termodinâmica, cinética química, mecânica quântica e a teoria da ligação de elétrons de valência (ATKINSON; SUPPES, 1968; NYE, 2000).

Nestes 83 anos, o método semi-empírico de London-Eyring-Polanyi, para construir superfícies de energia potencial, propiciou uma ferramenta de trabalho útil para obter intuições sobre como os processos físicos e químicos ocorrem, e para motivar o surgimento de novas técnicas e experimentos para estudar a combinação temporária de átomos no chamado estado de transição como nos trabalhos de Ahmed Zewail (MIELKE et al., 2003).

A novidade dos mapas visuais de superfícies de energia potencial e a linguagem dos poços de potencial e barreiras de ativação tornou-se também um importante instrumento pedagógico para a química. Em contraposição, cálculos de primeiros princípios de dinâmica quântica, com exatidão equivalente ao de Mielke e colaboradores, permanecem proibitivos, praticamente impossíveis de serem realizados para a ampla maioria das reações químicas de interesse prático (MIELKE et al., 2003).

Sem dúvida: os métodos *ab initio*, baseados integralmente em primeiros princípios, eventualmente prevalecerão. Porém, se a reação do átomo de hidrogênio com a molécula de hidrogênio só pôde ser considerada definitivamente resolvida há 10 anos, o que dizer de todas as outras reações químicas e bioquímicas? (MIELKE et al., 2003; MORGON; COUTINHO, 2007).

3.7 MÉTODOS COMPUTACIONAIS UTILIZADOS EM CÁLCULOS DE PROPRIEDADES MOLECULARES

Os métodos de química quântica podem ser aplicados aos estudos de relação estrutura-atividade (SAR), estrutura-atividade quantitativa (QSAR) e estrutura-propriedade (QSPR) pela derivação direta dos descritores eletrônicos a partir da função de onda molecular. Em geral, o tratamento teórico mais rigoroso não faz uso de parâmetros empíricos e é denominado *ab initio* (SANTOS et al., 2014a).

O método *ab initio* apesar de proporcionar informações relativamente precisas sobre o comportamento eletrônico, este método em termos operacionais é mais lento e mais caro. Assim, vários métodos semi-empíricos têm sido desenvolvidos, que são baseadas em certas hipóteses que servem para simplificar os cálculos usando certos parâmetros obtidos a partir de dados experimentais.

As exatidões dos métodos estão relacionadas com os erros associados do conjunto de base de valência selecionado, e do nível de tratamento de correlação eletrônica. Portanto, há vários métodos computacionais utilizados nos cálculos de propriedades moleculares, entre elas podemos citar: semi-empírico, *ab initio* e Teoria do Funcional da Densidade (DFT) (SANTOS et al., 2014a).

3.7.1 Métodos Semi-empíricos

Os métodos semi-empíricos utilizam o mesmo formalismo mecânico-quântico empregando apenas os elétrons da camada de valência do sistema. A razão por trás desta aproximação é que os elétrons envolvidos numa reação química e em outros fenômenos intermoleculares são os elétrons da camada de valência. Assim, a grande vantagem dos métodos semi-empíricos em relação aos métodos *ab initio* é a maior velocidade de processamento, uma vez que os cálculos são simplificados, reduzindo os custos de memória e tempo computacional (LEACH, 2001a).

Diferente dos métodos *ab initio*, nos métodos semi-empíricos emprega-se parâmetros empíricos, ou seja, derivados de dados experimentais tais como geometrias de equilíbrio, calor de formação, momento de dipolo molecular e potenciais de ionização, ou previamente calculados pela equação de Schrödinger, que permitem algumas integrais presentes no método *ab initio* não sejam calculadas simplificando os cálculos computacionais (SANT'ANNA, 2002; HENRE, 2003).

O primeiro método a utilizar essa aproximação é o *Complete Neglect of Differential Overlap* (CNDO), no qual os orbitais atômicos são considerados esfericamente simétricos na avaliação das integrais de repulsão eletrônica. Outros métodos também utilizam essas aproximações tais como, *Intermediate Neglect of Differential Overlap* (INDO) e *Neglect of Diatomic Differential Overlap* (NDDO) (LEACH, 2001b).

Os métodos semi-empíricos mais comumente utilizados são *Austin Modell* (AM1) e *Parametric Method 3* (PM3), ambos os métodos incorporam aproximações muito semelhantes, mas diferem na parametrização (DEWAR et al., 1985; STEWART, 1989).

Recentemente, o método AM1 foi objeto de uma re-parametrização para os átomos de H, C, N, O, P, S, F, Cl, Br e I, resultando no método *Recife Modell* (RM1), com menores erros de cálculos do que os gerados pelos métodos AM1 e PM3 (ROCHA et al., 2006).

3.7.2 Métodos *ab initio*

O termo em Latim *ab initio* significa “a partir do início” ou “a partir dos princípios fundamentais”, ou seja, são cálculos realizados a partir de constantes físicas fundamentais, usando equações exatas, que envolvem uma população eletrônica total da molécula sem o uso de parâmetros experimentais e sem aproximações adicionais.

O primeiro método de cálculo da estrutura eletrônica foi o método de Hartree-Fock (HF) que emprega a equação de Schrödinger completa para tratar todos os elétrons de um sistema químico. Esse modelo emprega conjuntos de funções de base (*basis set*) nos cálculos tais como as funções do tipo Slater (STO) e as funções Gaussianas (GTO *e.g.* 3-21G, 6-31G) (SANT’ANNA, 2002).

Essas bases mínimas apresentam diversas deficiências e para aprimorá-las faz-se a inclusão de função de polarização (i.e., orbitais p representado por *). Assim, 6-31G* refere-se ao conjunto de base 6-31G com função de polarização para átomos pesados (ou seja, átomos diferentes de hidrogênio), 6-31G** refere-se à inclusão de função de polarização para os átomos de hidrogênio e hélio. A base 6-31G** é particularmente útil onde ocorrem ligações hidrogênio. Funções de base com polarização parcial também foram desenvolvidas, por exemplo, 3-21G* que é a mesma base mínima 3-21G com funções de polarização parcial (LEACH, 2001a).

Apesar dos métodos *ab initio* darem uma predição quantitativa de alta qualidade para uma grande variedade de sistemas, eles são demorados e de alto custo computacional. Um recurso comumente empregado é otimizar a geometria com um conjunto de base mais simples e em seguida executar cálculos de “ponto único” (*Single Point*) com um conjunto de base mais completo permitindo determinar a energia e outras propriedades de um sistema molecular, usando uma base de cálculo mais sofisticada (LEACH, 2001b).

3.7.3 Teoria do Funcional de Densidade

A Teoria do Funcional de Densidade (*Density Functional Theory*– DFT) é um formalismo muito bem sucedido, onde o principal objetivo é substituir a função de onda usada para escrever os elétrons em métodos como o Hartree-Fock, pela densidade eletrônica. Os cálculos HF consideram uma densidade eletrônica média, já os cálculos DFT consideram as interações instantâneas de pares de elétrons com spins opostos (LEACH, 1996). Trata-se de

uma aproximação baseada na teoria de Hohenberg e Kohn que afirma que todas as propriedades de um sistema são funções da densidade de cargas. Assim, o teorema de Hohenberg-Kohn permite escrever a energia eletrônica total como uma função da densidade eletrônica ρ :

$$E(\rho) = E_{KE}(\rho) + E_C(\rho) + E_H(\rho) + E_{xc}(\rho) \quad (1)$$

onde $E_{KE}(\rho)$ é a energia cinética, $E_C(\rho)$ é o termo de interação núcleo-elétron, $E_H(\rho)$ é a energia de Coulomb e $E_{xc}(\rho)$ contém as contribuições de troca e correlação.

Os orbitais moleculares nos cálculos de funcional de densidade são geralmente escritos como uma expansão linear dos orbitais atômicos (ou seja, funções de base) que pode ser representado usando funções do tipo Gaussiana, orbitais de Slater ou orbitais numéricos (LEACH, 1996).

Os modelos de Funcional de densidade, assim como os modelos Hartree-Fock, são aplicáveis em moléculas de 50-100 átomos (HEHRE, 2003). O funcional exato não é conhecido, logo, existe um variado conjunto de funcionais diferentes que podem fornecer resultados diferentes para o mesmo problema. O método B3LYP (Becke, Lee, Yang e Parr) é um método híbrido amplamente aplicado, onde parte do funcional é obtida por mecânica quântica (combina energia de troca HF com o termo de troca DFT) e parte é parametrizado (adiciona funcionais de correlação) (FORESMAN; FRISCH 1993).

3.8 SOFTWARES DE QUÍMICA COMPUTACIONAL PARA CÁLCULO DE PROPRIEDADES MOLECULARES

Vários softwares de química computacional disponibilizam cálculos de propriedades moleculares através de métodos semi-empíricos e Hartree-Fock. Podemos citar: MOPAC2002 (STEWART, 1999), VAMP (CLARK, 2002), MNDO97 (THIEL, 1998), Gaussian98 (FRISCH et al., 1998), Spartan (SPARTAN), GAMESS (GORDON), Chem3D (CAMBRIDGESOFT CORPORATION, 1999), CAche (STEWART, 1999), HyperChem (HYPERCUBE), ZINDO (ZERNER, 1998), GEOMOS (RANALDI et al., 2003).

A maioria dos grandes avanços obtidos na pesquisa de métodos semi-empíricos foi essencialmente incorporada em uma série de programas intitulados MOPAC. MOPAC deve ser entendido como uma abreviação de Molecular Orbital PACKage. Iniciado em 1983, no grupo do Prof. M. Dewar, com a participação do Prof. J.J.P. Stewart, o projeto MOPAC tem atualmente 30 anos, e os principais marcos desta história são:

- 1983: MOPAC → Otimização de geometria, análise vibracional e estados de transição.
- 1984: MOPAC 2 → Concertos de erros.
- 1985: MOPAC 3 → Método AM1, cálculo de polímeros e CI (*Configuration interaction*).
- 1987: MOPAC 4 → Concerto de erros.
- 1988: MOPAC 5 → Método de PM3 e caminho de reação.
- 1989: MOPAC 6 → Concerto de erros, cálculo de potencial eletrostático e extensão do PM3 para os 28 elementos.
- 1993: MOPAC 7/93 → Modelo de solvente, propriedades óticas não-lineares e análise de simetria.
- 1997: MOPAC 97 → Método MNDO-d
- 1999: MOPAC2000 → Alocação dinâmica de memória, moléculas grandes, sólidos, metais de transição, melhoramento do algoritmo de busca do estado de transição.
- 2001: MOPAC2002 → Melhoramento do pacote de solvatação, todos os átomos do grupo principal, método PM5 e parametrização para mais metais de transição.
- 2004: MOPAC2005 → Cálculo de complexos de Eu(III) (ROCHA et al., 2004) e Th(IV) (YOUNG, 2001) através do modelo *Sparkle*.
- 2006: MOPAC 7.1 → Uma tradução do MOPAC 6 para Fortran 90 é disponibilizado na Internet, tendo incorporado o modelo Sparkle/AM1 para o cálculo de todos os íons lantanídeos trivalentes (STEWART, 2007);
- 2007: MOPAC2007 → uma aplicação do MOPAC7.1 incorporando os novos métodos do RM1, PM6 e Sparkle/AM1 e Sparkle/PM3. Este programa está disponibilizado gratuitamente para a academia e para aplicações sem fins lucrativos (STEWART, 2007).

3.8.1 Desenho e visualização tridimensional

Diversos programas de desenho de moléculas bidimensionais são disponíveis e de fácil utilização, como ChemWindow, Isis Draw, Chemdraw (CAMBRIDGE SOFT CORPORATION, 1998) e Chem3D (CAMBRIDGE SOFT CORPORATION, 1999). Eles permitem a elaboração de figuras e diagramas com qualidade e exatidão desejadas e facilitam o trabalho de documentação e comunicação científica.

O software ChemSketch 12.00 (ACD, 2010) é uma ferramenta avançada de desenho químico que fornece às propriedades moleculares, otimização e visualização 3D, capacidade

de nomear as moléculas, conforme IUPAC, e ainda possui um grande banco de dados com estruturas químicas e materiais de laboratório.

O software calcula automaticamente a valência de cada átomo e restringe a construção da molécula com base na regra do octeto, a não ser que seja instruído a não fazer esta limitação. Em seguida, é possível solicitar a construção da forma espacial 3D da espécie estudada, o que aciona outra janela onde o acadêmico pode rotacionar tridimensionalmente a espécie estudada, além de observar estas espécies em diferentes visualizações com possibilidade de visualizar ligações e arranjo espacial das espécies de forma realçada em cada uma destas representações. Alguns programas permitem o cálculo e a representação de várias propriedades moleculares, incluindo fórmula e massa moleculares, massa exata e análise teórica elementar.

Programas mais completos, como ChemDraw Ultra (CAMBRIDGESOFT CORPORATION, 1999), fornecem, adicionalmente, o nome químico correto (IUPAC) dos compostos químicos e podem prever os correspondentes deslocamentos químicos de RMN ^1H e ^{13}C , pontos de fusão e congelamento, log P, refratividade molar e calor de formação (CARVALHO et al., 2003).

O desenho e a visualização de fármacos 3D, com fatores estéricos relevantes para a atividade biológica, são importantes para análise do tamanho, volume e formato das moléculas (CARVALHO et al., 2003).

O Molekel é um software livre multiplataforma de visualização molecular, e foi originalmente desenvolvida na Universidade de Genebra por Peter F. Flükiger na década de 1990 para Silicon Graphics Computadores. Em 1998, Stefan Portmann assumiu a responsabilidade e lançou a versão 3.0. A versão 4.0 foi uma versão da plataforma quase independente. Outros desenvolvimentos levaram a versão 4.3, antes de Stefan Portmann se mudar e deixar de desenvolver os códigos. Em 2006, o Centro de Supercomputação Nacional Suíço (CSCS) reiniciou o projeto e a versão 5.0 foi lançada em 21 de dezembro do mesmo ano (FLUKIGER, 2002).

3.8.2 Análise conformacional e minimização de energia

Para a obtenção da análise conformacional e minimização de energia, podemos citar: o programa Chem3D muito empregado em estudos com esse objetivo, no entanto, outros

programas como Molecular Modeling Pro (CHEMSW® Inc, 2001), ChemSite (CHEMSW® Inc, 2001), Alchemy, Sybyl, ChemX, CAChe e Weblab Viewer são também disponíveis.

Na área de modelagem molecular, construção de gráficos e projetos de fármacos, o programa Hyperchem por ser uma ferramenta em 3D especializada em estruturas de interesse para a medicina, indústria farmacêutica e química orgânica (CHEMPLUS, 2000).

O programa Hyperchem permite desenhar complicadas moléculas. Este software é, também, uma alternativa na área de espectroscopia, que além da capacidade para a simulação a priori de espectros RMN através de métodos quânticos, contém uma base de dados com cerca de 10.000 moléculas, aplicável a macromoléculas assim como moléculas pequenas. O software também possui animações, além de cálculos de química quântica e mecânica molecular (CHEMPLUS, 2000).

A análise conformacional de uma molécula é realizada pela rotação de uma ligação, com mudança paralela dos ângulos torsionais ou diédricos das ligações, e cálculos correspondentes de energia estérica, decorrente da sobreposição espacial de átomos não-ligados e barreiras torsionais de rotação (PATRICK, 2001; SANTOS et al., 2014a).

As moléculas desenhadas de forma tridimensional não estão, necessariamente, na conformação mais estável. Durante a geração de uma determinada estrutura, ocorrem distorções na molécula, com formação desfavorável de comprimentos e ângulos de ligações e ângulos diédricos. Átomos não-ligados também interagem em uma mesma região do espaço e provocam repulsão estérica e eletrostática. Para corrigir estas distorções as moléculas são otimizadas pelo processo de minimização de energia, a partir de dois modelos matemáticos (i) mecânica molecular ou (ii) mecânica quântica. Interações não previsíveis, relacionadas à sobreposição de orbital molecular, distribuição de densidade eletrônica ou interferências estéricas podem ser solucionadas pelos métodos computacionais. A minimização de energia e a análise conformacional são usadas interativamente para otimizar a geometria de uma molécula (PATRICK, 2001; SANTOS et al., 2014a).

A escolha do método de minimização de energia depende de fatores relacionados ao tamanho da molécula, disponibilidade de parâmetros e dados armazenados e recurso computacional. Os modelos moleculares gerados pelo computador são resultantes de equações matemáticas que estimam as posições e propriedades dos elétrons e núcleos; os cálculos realizados exploram, de forma experimental, as características de uma estrutura, fornecendo uma molécula sob nova perspectiva (PATRICK, 2001; SANTOS et al., 2014a).

A representação estrutural tridimensional dos fármacos nos programas computacionais e a construção de modelos moleculares são importantes para o aprendizado das características geométricas e moleculares essenciais para a atividade biológica de algumas classes de fármacos.

Na etapa do desenho e visualização tridimensional devem-se reunir conhecimentos fundamentais de química orgânica, como estereoquímica, reatividade e nomenclatura, para entendimento da relação estrutura-atividade de fármacos (SANTOS et al., 2014a).

Na química quântica os softwares mais utilizados são GaussView e o Gaussian que utiliza as leis da mecânica quântica para prever as energias, estruturas e propriedades e a frequência vibracional de sistemas moleculares (FRISCH et al., 2004).

O GaussView 5.0 é um programa capaz de trabalhar no Windows e responsável pela construção das estruturas em estudo, pela visualização destas bem como pela geração do input das espécies em estudo para o programa de cálculos – Gaussian 03W. Este inclui um avançado modelador molecular, que pode ser usado para a construção e exame das três dimensões moleculares (FRISCH et al., 2004).

O Gaussian 03W é um programa capaz de trabalhar em sistemas operacionais Windows e Linux que realiza cálculos computacionais usados no estudo de mecanismos de reações, geometrias de equilíbrio de moléculas neutras, radicais e íons, e na determinação de parâmetros físico-químicos. Aprecia estrutura, reatividade, propriedades termodinâmicas, barreiras energéticas (estados de transição), análises conformacionais, empregando na otimização das moléculas e nos cálculos teóricos de espectros vibracionais. A partir da otimização obtém-se a estrutura mais adequada para a molécula, considerando os comprimentos e os ângulos de ligação e a energia de estabilização calculada (FRISCH et al., 2004).

3.9 APLICAÇÃO DO MÉTODO HARTREE-FOCK USANDO FERRAMENTAS COMPUTACIONAIS PARA A OBTENÇÃO DE PROPRIEDADES MOLECULARES

A obtenção das propriedades moleculares depende do método e conjunto de base, e representa um meio de informação química contida na estrutura molecular do composto estudado. Esta informação é transformada e codificada para muitos problemas químicos, farmacológicos e toxicológicos sobre Relação Estrutura Atividade (SAR), Relação Estrutura Atividade-Quantitativa (QSAR) e Relação Estrutura Propriedade-Quantitativa

(QSPR) (TODESCHINI; GRAMATICA, 1997a; TODESCHINI; GRAMATICA, 1998; GUHA, 2012).

As relações SAR, QSAR e QSPR representam um aspecto central da química medicinal. O fato de que uma pequena mudança na estrutura leva a uma pequena alteração na atividade biológica, e assim permitindo que os químicos orgânicos ou sintéticos possam racionalizar as substituições em posições específicas, dando-lhes a liberdade para modificar uma molécula para melhorar várias propriedades moleculares, tais como a lipofilicidade, a biodisponibilidade, e assim por diante, sem sacrificar atividade biológica e a potência do futuro fármaco (JOHNSON; MAGGIORA, 1990).

A perspectiva dos químicos teóricos que trabalham com modelagem molecular de novos fármacos é o princípio de que estruturas semelhantes apresentam atividades biológicas semelhantes, isto é a ideia geral para quaisquer estudos QSAR (LEACH, 2001a; MARTIN, 2010).

3.10 DOCKING MOLECULAR

A consideração da flexibilidade molecular do receptor e do ligante implica no tratamento de centenas de milhares de graus de liberdade, por parte dos algoritmos de “*docking*”. O reconhecimento molecular é um processo dinâmico e altamente complexo, envolvendo um grande número de interações intermoleculares entre o ligante, a molécula receptora e o solvente.

Devido a sua complexidade, o problema de “*docking*” é geralmente dividido em dois subproblemas: (i) desenvolvimento de um algoritmo que investigue de maneira eficaz uma complexa hipersuperfície de energia para prever a conformação e a orientação de uma molécula ligante relativa ao sítio ativo do receptor; (ii) predição da afinidade de ligação de um complexo receptor – ligante, ou seja, o desenvolvimento de um modelo de avaliação da energia livre de ligação (normalmente chamado na literatura de função de “*scoring*”), que seja viável computacionalmente, para discriminar corretamente entre diferentes modos de ligação do mesmo ligante e/ou para determinar, entre dois ligantes distintos, aquele com maior afinidade de ligação para um mesmo receptor (MORGON; COUTINHO, 2007).

A complexidade do problema de “*docking*” requer métodos computacionais com potencial para investigar efetivamente um grande número de soluções possíveis. Os primeiros algoritmos de *docking* molecular começaram a ser desenvolvido no início da década de 80.

Desde então, vários métodos têm sido implementados, constituindo uma área de pesquisa bastante ativa (KITCHEN et al., 2004). Além de diferirem na função de avaliação “*scoring*” utilizada. Os métodos de “*docking*” se diferenciam principalmente em relação a: (1) grau de flexibilidade das moléculas biológicas interagentes; e (2) método de otimização empregado na investigação da hipersuperfície de energia.

O primeiro programa de “*docking*” utilizando algoritmo Genético (AG) foi implementado por Judson (1993). O programa de “*docking*” GOLD utiliza um algoritmo genético para evoluírem múltiplas subpopulações de ligantes, onde a migração entre as populações é permitida (JONES et al., 1997).

O programa AutoDock possui implementado um Algoritmo Genético Lamarckiano (AGL). O AGL é um AG híbrido com busca local (BL). Os programas de “*docking*” mais amplamente utilizados atualmente, e que geralmente são utilizados para efeito de comparação com novos métodos propostos são: DOCK, FlexX, AutoDock e GOLD (SANTOS et al., 2014a).

A habilidade de um determinado programa em predizer o modo de ligação de um complexo fármaco-receptor é analisada em termos do desvio médio quadrático das posições dos átomos (RMSD – “Root-Mean-Square Deviation”), entre a estrutura do ligante encontrada pelo programa e a obtida experimentalmente. O valor padrão de RMSD utilizado para considerar estruturas que foram corretamente preditas é de 2.0 Å, embora um RMSD de até 2.5 Å seja considerado um bom resultado, principalmente no estudo de ligantes com muitos ângulos diedrais.

Várias publicações avaliam o desempenho dos vários métodos de “*docking*” existentes pela quantidade de complexos corretamente preditos. Além da habilidade de predizer o modo de ligação correto, a determinação acurada da afinidade de ligação de um ligante para um determinado receptor e o custo computacional do método, também são fatores importantes que influenciam na avaliação de um programa de docking (KONTOYIANNI; MCCLELLAN; SOKOL, 2004).

Atualmente, existem grandes bancos de estruturas moleculares de acesso público como, por exemplo, o *Protein Data Bank* (PDB), onde várias estruturas são depositadas e podem ser obtidas, e vários bancos de estruturas moleculares de ligantes, como o *Maybridge* e o *Cambridge Structural Database* (BERMAN et al., 2000).

No problema de *docking* proteína – ligante o objetivo é encontrar e quantificar o modo de ligação correto de uma molécula ligante no sítio ativo de uma macromolécula receptora, de maneira que a função do receptor possa ser inibida ou ativada (BARREIRO et al., 2002).

3.10.1 Docking molecular no estudo de compostos antimaláricos derivados da artemisinina

Tonmunphean et al. (2001) pesquisou sobre *docking* molecular automatizado da artemisinina com o heme. Os efeitos de cargas atômicas e as estruturas do ligante com o heme foram investigados. O *docking* molecular mostrou que a artemisinina aproxima-se do heme na posição oxigênio 1 (O1) na ligação endoperóxido para o centro do ferro, este resultado difere daquele relatado por Shukla, Gund e Meshnick (1995), e outros sugeridos que o heme ligasse com a artemisinina na posição oxigênio 2 (O2). Sendo que, o resultado de *docking* depende das estruturas da artemisinina na interação com o heme. Além disso, as cargas atômicas do heme têm efeito significativo nas configurações de *docking* molecular.

Cheng et al. (2002) realizou um estudo de *docking* molecular e QSAR-3D, relação quantitativa tridimensional da estrutura atividade, a fim de compreender o mecanismo antimalárico e a relação entre as propriedades físico-química e as atividades antimaláricas de análogos da artemisinina, através de simulações de *docking* molecular para sondar as interações destes análogos com o heme.

O QSAR-3D foi baseado nos modelos de *docking* empregando análise molecular comparativa campo de força (COMFA) e a análise molecular comparativa dos índices da similaridade (COMSIA). A análise parcial subsequente dos mínimos quadrados (PLS) indicou que as energias obrigatórias do cálculo correlacionaram bem com os valores experimentais da atividade. Os modelos COMFA e COMSIA foram conformações biativos que provaram ter uma boa habilidade de previsão. Por sua vez combinando bem os resultados de *docking*.

Pinheiro et al. (2003) planejaram derivados da artemisinina com atividade antimalárica com a ajuda de química quântica e método dos mínimos quadrados parciais (PLS). Construíram um modelo QSAR baseado em cinco descritores moleculares utilizados para prever a atividade antimalárica de dez compostos com atividade desconhecida, e destes um composto foi predito como mais ativo do que os compostos

estudados. Realizaram também a construção de gráficos moleculares e estudos de modelagem entre a artemisinina e o heme. Propuseram através de *docking* molecular da artemisinina e alguns derivados com a hemoglobina A, uma visão do modo de ligação entre artemisinina-heme.

3.11 POTENCIAL ELETROSTÁTICO MOLECULAR

Um dos temas frequentes da química teórica é a pesquisa visando melhorar métodos de modo a elucidar o comportamento reativo de moléculas e outras espécies químicas. Dentre os inúmeros índices de reatividade existentes o potencial eletrostático molecular $V(r)$ que é gerado em torno de uma molécula por seus núcleos e elétrons, se destaca por ser uma propriedade física real podendo ser determinada experimentalmente por métodos de difração, bem como computacionalmente (POLITZER; MURRAY, 1991).

Em termos de eletrostática clássica o potencial eletrostático molecular (MEP) é de fácil interpretação. Uma molécula gera um potencial ao redor de si mesma que é vista por uma carga de prova positiva próxima ou evitando regiões onde o MEP é negativo ou positivo, respectivamente (NÁRAY-SZABÓ; FERENCZY, 1995).

Existe amplo reconhecimento da importância da eletrostática clássica devido, proporcionar um importante modelo qualitativo e ferramenta quantitativa em biologia estrutural, bioquímica e química (HONIG; NICHOLLS, 1995).

O MEP pode ser obtido diretamente da função de onda, o cálculo é relativamente rápido e é possível obter valores das localizações espaciais ao redor da molécula e representar o MEP como um mapa em duas ou três dimensões, sendo esta uma característica chave do MEP. Os valores do MEP podem ser representados em planos por linhas de contorno ou por código de cores na superfície da molécula, obtendo assim da distribuição do MEP (NÁRAY-SZABÓ; FERENCZY, 1995).

3.11.1 Mapas de potencial eletrostático molecular no estudo de compostos antimaláricos derivados da artemisinina

Um dos primeiros estudos da relação estrutura-atividade da artemisinina e moléculas antimaláricas usando mapas de potencial eletrostático molecular foi realizado por Thomson, Cory e Zerner (1991), através do estudo teórico da estrutura da artemisinina e derivados

utilizando métodos semi-empíricos (AM1, PM3) e Hartree-Fock (3-21G, 6-31G). Os MEP obtidos com AM1 e PM3, quando comparados mostrou diferenças distintas entre moléculas ativas e inativas, tal como deoxiartemisinina, pois moléculas ativas têm uma ampla banda de potencial negativo em torno da parte da molécula contendo a ligação endoperóxido.

Bernardinelli et al. (1994) utilizaram mapas de potencial eletrostático molecular objetivando identificar as características chaves que são necessárias para a atividade antimalárica da artemisinina e alguns derivados. Os MEP mostraram que as moléculas ativas têm uma região de potencial negativo de forma similar próximo ao anel trioxano, porém esta região é deslocada em compostos inativos. Os MEP foram utilizados para fazer previsões qualitativas de novas e mais efetivas moléculas antimaláricas, dessa forma Bernardinelli et al., concluíram que qualquer nova molécula ativa irá ter MEP que é similar ao da artemisinina.

Suter (1995) construíram mapas de potencial eletrostático molecular para artemisinina com o método DFT BLYP/6-31G* e usando redes neurais projetaram os mapas 3-D sobre uma superfície 2-D tornando assim mais fácil à comparação das principais características dos MEP para as moléculas estudadas e correlacionaram com a atividade biológica, predizendo assim quais são os compostos antimaláricos mais potentes dentre uma dada série.

Jefford (2000) correlacionou o potencial eletrostático molecular de alguns peróxidos orgânicos (artemisinina, yingzhaosu A e alguns análogos sintéticos) com suas atividades antimalárica. Utilizando rede neural Kohonen a dimensionalidade dos mapas MEP foram reduzidas facilitando a comparação, revelando que peróxidos orgânicos com alta atividade antimalárica foram caracterizados por uma faixa contínua de potencial negativo ao redor da molécula, enquanto peróxidos orgânicos menos ativos mostraram uma faixa descontínua de potencial negativo.

Cardoso et al. (2007) realizaram um estudo da artemisinina e derivados com atividade antimalárica contra *Plasmodium falciparum*, unindo técnicas de química quântica, métodos quimiométricos (PCA, HCA, PLS) e estudo da relação estrutura atividade através da construção de mapas de potencial eletrostático molecular (MEP). Através do modelo PLS baseado em cinco descritores, propuseram um novo derivado, de onze testados, predito como mais potente que os compostos do conjunto estudado. Com a análise dos MEP de três compostos propostos, observaram que estes possuem um MEP similar próximo do anel trioxano com o MEP da artemisinina, os resultados dos mapas MEP deram suporte aos resultados obtidos com o modelo PLS.

3.12 MÉTODOS DE ANÁLISE MULTIVARIADA

Pinheiro, Ferreira e Romero (2001) combinaram técnicas de química quântica (Hartree-Fock 3-21G) e métodos de análise multivariada (PCA, HCA, KNN e SIMCA) para estudar e propor derivados da diidroartemisinina. Através da técnica PCA e HCA selecionaram sete descritores que foram responsáveis para a classificação dos compostos em duas classes distintas, e com construção dos modelos qualitativos KNN e SIMCA propuseram dois compostos de um conjunto de doze testados preditos como sendo de alta atividade.

Quando medidas são feitas sobre um número de objetos, os resultados são usualmente organizados em uma matriz, que é chamada matriz de dados. As medidas (em nosso estudo os descritores moleculares e a atividade antimalárica) são colocadas nas colunas, e os objetos (em nosso caso os compostos estudados) estão associados com as linhas.

Para uma matriz de dados multidimensionais, métodos de estatística multivariada são necessários para compreender tais dados em sua totalidade, com destaque para a análise de componentes principais que tem sido amplamente utilizada em problemas biológicos e químicos, o principal objetivo da análise é mostrar os dados multidimensionais em um espaço de baixa dimensionalidade com o mínimo de perda de informação total (FRANKE; GRUSKA, 1996; SANTOS et al., 2014a).

3.12.1 Análise de componentes principais (PCA)

Análise de componentes principais é um método de compressão de dados baseado na correlação entre as variáveis, essa compressão de dados gera um pequeno conjunto de variáveis as quais são denominadas componentes principais, sendo que estas são ortogonais entre si, de modo que a correlação entre as variáveis não limita sua aplicação, diferentemente da regressão linear múltipla que é sensível à presença de variáveis altamente correlacionadas, pois torna os coeficientes de regressão instáveis e sem significado (FERREIRA, 2002; FERREIRA; MONTANARI; GAUDIO, 2002).

Do ponto de vista matemático o modelo obtido com a análise de componentes principais é escrito como:

$$y_{ij} = \sum_{k=1}^p P_{ik} a_{kj} + \sum_{k=p+1}^m P_{ik}^{(0)} a_{kj}^{(0)} \quad (3.1)$$

onde P_{ik} são chamadas componentes principais (PCs), e também são referidas como “scores” e correspondem as características dos compostos. P_{ik} são vetores ortogonais, e são determinados de maneira que a matriz de dados seja reproduzida. a_{kj} representa o peso, ou seja, uma medida da contribuição da k -ésima PC com a j -ésima variável. Um alto valor de $|a_{kj}|$ demonstra uma alta importância da k -ésima PC para a j -ésima variável. Dessa forma a j -ésima variável possui alta contribuição na k -ésima PC, portanto a_{kj} também são chamados “loadings” (FRANKE; GRUSKA, 1996).

O número de PCs que pode ser extraído da matriz de dados é igual ao número de variáveis originais e com este número de componentes a matriz de dados pode ser reproduzida exatamente, porém este não é o resultado desejado, visto que não leva a redução da dimensionalidade da matriz. O que se almeja é encontrar um número de PCs de modo que as variáveis originais sejam representadas com o mínimo de perda de informação relevante (FRANKE; GRUSKA, 1996).

3.12.1.1 Escalonamento

Um modo de tratar todas as variáveis com a mesma importância é padronizá-las por autoescalamento de acordo com equação abaixo:

$$Y'_{jm} = \frac{(y_{jm} - \bar{y}_j)}{s_j} \quad (3.2)$$

onde \bar{y}_j é a média da variável j para o objeto m . Variáveis autoescaladas têm uma média igual a zero e variância unitária representadas pelas equações (3.3) e (3.4) respectivamente.

$$\bar{y}_j = \frac{\sum_{j=1}^n y'_{jm}}{n} = 0 \quad (3.3)$$

$$s_j^2 = \frac{\sum_{k=1}^n (y'_{jm} - y'_i)^2}{n-1} = 1 \quad (3.4)$$

3.12.2 Análise de agrupamento hierárquico (HCA)

Outro método de análise multivariada de grande importância é análise de agrupamento hierárquico, seu objetivo é mostrar os dados de tal modo a acentuar seus agrupamentos naturais e padrões. Como na análise de componentes principais, os resultados da análise de agrupamento hierárquico são qualitativos, sendo dispostos na forma de um dendograma permitindo assim visualizar as amostras (os compostos estudados em nosso caso) ou variáveis (os descritores moleculares em nosso estudo) em um espaço bi-dimensional (FERREIRA; MONTANARI; GAUDIO, 2002; SANTOS et al., 2014b).

Em análise de agrupamento hierárquico a distância entre as amostras (aqui os compostos estudados) é calculada e transformada em uma matriz de similaridade S , cujos elementos são os índices de similaridade. Então para duas amostras m e n , o índice de similaridade é escrito como:

$$S_{mn} = 1 - \frac{d_{mn}}{d_{\max}} \quad (3.5)$$

sendo S_{mn} um elemento de S , d_{\max} é a distância máxima para um par de amostras, d_{mn} vem a ser a distância euclidiana entre as amostras m e n calculadas como:

$$d_{mn} = \left[(x_{m1} - x_{n1})^2 + (x_{m2} - x_{n2})^2 + \dots + (x_{mh} - x_{nh})^2 \right]^{1/2} \quad (3.6)$$

com x_{ij} sendo um elemento da matriz de dados originais (FERREIRA; MONTANARI; GAUDIO, 2002).

3.12.3 Método dos mínimos quadrados parciais (PLS) e regressão dos componentes principais (PCR)

O objetivo de qualquer estudo QSAR é construir uma correlação entre as propriedades moleculares, tais como físico-química, estruturais com as atividades biológicas dos compostos estudados. As atividades biológicas podem ser representadas de dois modos: ou elas são continuamente avaliadas ou atribuídas categorias distintas. No caso, delas serem contínuas, é possível desenvolver modelos de regressão utilizando PLS ou PCR (FERREIRA, 2002; SANTOS et al., 2014c).

Molfetta et al. (2007) estudaram a relação da estrutura-atividade quantitativa (QSAR) de 19 compostos quinona com atividade tripanocida usando os métodos dos mínimos quadrados parciais (PLS) e regressão dos componentes principais (PCR) com o procedimento de validação cruzada leave-one-out para construir os modelos de regressão. A atividade tripanocida dos compostos foi relacionada com o seu primeiro potenciais catódica (EPC1). Os modelos de regressão PLS e PCR construídos foram também utilizados para prever o EPC1 de seis novos compostos de quinona.

O modelo PLS foi construído com três componentes principais que descreveu 96.50% da variância total e apresentou $Q^2 = 0.83$ e $R^2 = 0.90$. Os resultados obtidos com o modelo PCR foram semelhantes aos obtidos com o modelo PLS. O modelo de PCR também foi construído com três componentes principais que descreveu 96.67% da variância total, com $Q^2 = 0.83$ e $R^2 = 0.90$. Os descritores mais importantes para os modelos PLS e PCR foram HOMO-1 (energia do orbital molecular HOMO abaixo), Q4 (carga atômica na posição 4), MAXDN (diferença eletrotológica negativa máxima) e HYF (índice hidrofiliçidade).

Um dos mais promissores métodos de análise multivariada é os do mínimo quadrado parciais (PLS) e o de regressão dos componentes principais (PCR), pois centenas ou milhares de variáveis independentes (o bloco X, em nosso estudo os descritores moleculares) pode ser correlacionadas com uma ou várias variáveis dependentes (o bloco Y, em nosso caso a atividade antimalárica). Na análise PLS os vetores resultantes são levemente deslocados de suas posições originais, de tal modo que a correlação dos correspondentes vetores derivados do bloco X e Y seja otimizada (KUBINYI, 1993).

Através dos métodos PLS e PCR é possível obtermos uma descrição da variável dependente Y (atividade biológica) como uma combinação linear dos descritores moleculares, por meio das componentes principais, sendo que elas não são correlacionadas entre si. Para garantir que os componentes principais obtidos sejam importantes para a atividade biológica, o bloco Y (atividade antimalárica em nosso estudo) é utilizado para encontrar um padrão dentro do bloco X (descritores moleculares em nosso caso) que esteja correlacionado com o bloco Y, ou seja, as componentes principais são otimizadas para descrever melhor a relação entre o bloco X e Y,

simultaneamente, sendo agora os componentes principais chamadas de variáveis latentes e são utilizadas para modelar o bloco Y (FERREIRA; MONTANARI; GAUDIO, 2002).

3.12.3.1 Método dos mínimos quadrados parciais (PLS)

O modelo PLS encontra novas variáveis, K variáveis latentes, que são também chamadas “scores”, sendo representadas como t_K ($K=1, 2, 3, \dots, K$). Os “scores” são combinações lineares das variáveis originais j_s com pesos dos coeficientes W^*_{vk} ($k=1, 2, 3, \dots, K$).

$$t_{ik} = \sum_k W^*_{jk} x_{ik} \quad (3.7)$$

Os “scores” são bons preditores do bloco Y, então o modelo PLS pode ser escrito como:

$$Y_{im} = \sum aC_{ma} t_{ia} + f_{im} \quad (3.8)$$

na forma matricial possui a seguinte forma:

$$Y = TC' + F \quad (3.9)$$

combinando as equações (3.7) e (3.8), obtemos o modelo na forma de um modelo de regressão:

$$y_{im} = \sum aC_{ma} \sum kW^*_{jk} + f_{im} \quad (3.10)$$

onde o termo f_{im} representa os desvios entre os dados observados e preditos (em nosso estudo a atividade antimalárica) (WOLD, 1996).

Na análise PLS os resultados podem ser transformados em coeficientes de regressão, também chamados vetores de regressão, das variáveis do bloco X (os descritores moleculares em nosso estudo), estes indicam quais os descritores mais importantes na construção do modelo (KUBINYI, 1993; FERREIRA; MONTANARI; GAUDIO, 2002).

3.12.3.2 Regressão dos componentes principais (PCR)

A regressão dos componentes principais utiliza os passos semelhantes aos descritos na PCA para decompor a matriz de dados em componentes principais (BEEBE et al., 1998), que descrevem a estrutura latente de X, seguido de uma etapa de regressão. Com base na técnica

de decomposição em valores singulares introduzidos análise de PCA, pode-se escrever os modelos de PCR como:

$$[y] = [x][b] + e_y \Rightarrow [y] = [U][S][V^T][b] + e_y [\hat{b}] = [VS^{-1}U^T]^A [y] \quad (3.11)$$

onde o expoente A é o número de componentes principais incluídos no modelo ($V(A \times A)$, $S^{-1}(A \times A)$ e $UT(A \times J)$). Desde que os componentes principais sejam ortogonais entre si, eles são adequados como regressores para y usando regressão linear múltipla – RLM (FERREIRA et al., 1999). Neste caso, a solução dos mínimos quadrados para a estimativa dos coeficientes de regressão **a**, quando a matriz de dados original, X, é substituída com um subconjunto das colunas de A do arranjo da matriz T é dado por:

$$y = Ta + e = a_0 + a_1t_1 + \dots + a_A t_A + e \Rightarrow \hat{a} = (T^T T)^{-1} T^T y \quad (3.12)$$

A interpretação do modelo de PCR seria muito simplificada se **a** fosse dado em termos de coeficientes de regressão **b**. A partir das equações, a seguir:

$$X\hat{b} = T\hat{a} \quad (3.13)$$

Além disso, $T=XL$ ($X = TLT$; L é uma matriz ortogonal). Substituindo **T** com **XL** na Equação acima, é evidente que uma possível solução para \hat{b} é $L\hat{a}$. Desde que XTX possam estar perto do singular, a solução para \hat{b} é praticamente única. Quando todos os componentes principais são mantidos no modelo (sem compressão de dados; $A=J$, para $J < I$), o modelo de PCR corresponderia ao RLM (MARTENS et al., 1989).

3.12.4 Avaliação dos modelos PLS e PCR

Após construir os modelos PLS e PCR é necessário avaliá-lo, e esta avaliação consiste em verificar se a especificação do modelo se adéqua convenientemente aos dados preditos. Tal avaliação é dividida em três partes: (1) avaliação do grau de ajuste, (2) avaliação do grau de significância e (3) avaliação do grau de previsibilidade (GAUDIO; ZANDONADE, 2001).

3.12.4.1 Avaliação do grau de ajuste dos modelos PLS e PCR

A avaliação do grau de ajuste é feita calculando os seguintes parâmetros estatísticos: (a) coeficiente de correlação *R*, (b) coeficiente de correlação ajustado R^2_{ajust} , (c) desvio padrão *s* e (d) análise dos resíduos, com as expressões matemáticas sendo mostradas abaixo respectivamente:

$$R^2 = 1 - \frac{\sum (y_{\text{predito}} - y_{\text{experimental}})^2}{\sum (y_{\text{predito}} - y_{\text{médio}})^2} \quad (3.14)$$

$$R^2_{\text{ajust}} = R^2 - \left(\frac{k-1}{n-k} \right) (1 - R^2) \quad (3.15)$$

onde n , na equação (3.15) e (3.16), é o número de amostras incluídas no modelo (em nosso caso o número de compostos incluídos), k é o número de variáveis (descritores moleculares em nosso estudo) incluídas no modelo.

$$s^2 = \frac{\sum (y_{\text{experimental}} - y_{\text{predito}})^2}{n-2} \quad (3.16)$$

$$(y_{\text{experimental}} - y_{\text{predito}}) \quad (3.17)$$

Para um modelo QSAR ser aceito o coeficiente R deve ser maior que 0,9 para ensaios da atividade biológica *in vitro* e maior 0,8 para ensaios *in vivo* (KUBINYI, 1993). O coeficiente de correlação ajustado, que considera o número correções para o número de variáveis e o número de compostos utilizados, deve ser o maior possível e é esperado que o desvio padrão e os resíduos sejam o mais próximo possível de zero (GAUDIO; ZANDONADE, 2001). Porém o desvio padrão s não pode ser maior que o desvio padrão dos dados biológicos, em torno de 0,3 que é o erro médio de muitos dados biológicos, sendo que para ensaios biológicos *in vivo* esse valor deve ser menor (KUBINYI, 1993).

3.12.4.2 Avaliação do grau de significância dos modelos PLS e PCR

O grau de significância é avaliado através da execução de testes de validação, em nosso caso testou-se apenas a significância estatística do coeficiente de correlação R^2 através do teste de hipótese chamado teste F que verifica o quanto da variabilidade do bloco Y (atividade antimalárica) pode ser explicada pelo bloco X (descritores moleculares incluídos nos modelos PLS e PCR). Para validar o coeficiente de correlação através do teste F, com a expressão matemática sendo mostrada abaixo, é necessário comparar o valor de F obtido na construção dos modelos PLS e PCR com o valor tabelado (GAUDIO; ZANDONADE, 2001).

$$F_{(k,n-k-1)} = \frac{R^2(n-k-1)}{(k(1-R^2))} \quad (3.18)$$

onde n na equação (3.18) corresponde ao número de amostras incluídas o modelo (em nosso caso, o número de compostos incluídos), k é o número de variáveis utilizadas para construir o modelo (em nosso estudo, os descritores moleculares). O valor de F deve ser de quatro a cinco vezes o valor tabelado, demonstrando assim que o modelo é estatisticamente significativo e útil para fins preditivos (BARROS NETO; SCARMINIO; BRUNS, 1996).

3.12.4.3 Avaliação do grau de previsibilidade do modelo PLS e PCR

O procedimento de validação cruzada (cross-validation) deve ser usado para avaliar o grau de previsibilidade do modelo QSAR, bem como selecionar o modelo que possuir maior habilidade preditiva. A validação cruzada é realizada de modo que um ou vários objetos (compostos estudados) sejam eliminados do conjunto de dados, seja casualmente ou de maneira sistemática (KUBINYI, 1993). A validação cruzada é feita nas seguintes etapas: (1) excluir um dos compostos do modelo; (2) reconstruir o modelo sem esse composto; (3) utilizar o modelo para calcular o valor da atividade biológica do composto excluído; (4) obter o resíduo entre o valor experimental da atividade biológica e o valor predito para o composto excluído; (5) refazer as etapas (1)-(4) para os demais compostos, um por vez (GAUDIO; ZANDONADE, 2001; SANTOS et al., 2014c).

A avaliação do grau de previsibilidade é feita calculando alguns parâmetros estatísticos que mostram a qualidade preditiva dos modelos PLS e PCR: soma dos quadrados dos desvios ($PRESS$), desvio padrão da validação cruzada (S_{PRESS}), erro padrão de previsão (SEP) e o coeficiente de correlação da validação cruzada (Q^2) com as respectivas equações sendo mostradas abaixo:

$$PRESS = \sum (y_{\text{experimental}} - y_{\text{predito}})^2 \quad (3.19)$$

o ideal é que o valor do $PRESS$ não aumente com o número de variáveis latentes (STANTON, 2003).

$$S_{PRESS} = \frac{(PRESS)^{1/2}}{n-k-1} \quad (3.20)$$

onde n na equação (3.20) é o número de amostras incluídas no modelo (compostos) e K é o número de variáveis utilizadas para construir o modelo (descritores moleculares), S_{PRESS} é obtido como um critério para o número ótimo de variáveis latentes, o menor valor de S_{PRESS} indica o número ótimo de variáveis latentes (KUBINYI, 1993), um modelo com excelente capacidade preditiva apresentará S_{PRESS} próximo de zero (GAUDIO; ZANDONADE, 2001; SANTOS et al., 2014c).

$$SEP = \sqrt{\frac{PRESS}{n}} \quad (3.21)$$

na equação (3.21) n corresponde ao número de amostras (compostos) incluídas no modelo QSAR, SEP é diferente do S_{PRESS} devido não considerar o grau de liberdade no cálculo (KUBINYI, 1993). É também utilizado como principal critério de verificação do grau de predição do modelo PLS (FERREIRA; MONTANARI; GAUDIO, 2002; SANTOS et al., 2014c).

$$Q^2 = \frac{PRESS}{\sum (y_{\text{experimental}} - y_{\text{médio}})^2} \quad (3.22)$$

este parâmetro é utilizado para avaliar a qualidade estatística dos modelos PLS e PCR (FERREIRA, 2002), ele descreve a quantidade de variância em y que pode ser predita, seu valor varia de zero a um (0 a 1), onde um (1) significa um modelo perfeito, e zero (0) um modelo sem relevância (SJÖSTRÖM; ERIKSSON, 1996). Para validar uma variável latente é necessário que o valor de (Q^2) não diminua com aumento de variáveis latentes e sim aumente com cada variável latente (STANTON, 2003). Um modelo PLS e PCR com elevada preditibilidade para objetos (compostos) não incluídos no modelo deverá apresentar o valor de Q^2 mais próximo de um (1) quanto possível (GAUDIO; ZANDONADE, 2001). Porém um valor acima de 0,3 ($Q^2 > 0,3$) é em geral satisfatório (PATRICK, 2006; SANTOS et al., 2014b).

4 MATERIAL E MÉTODOS

4.1 EQUAÇÕES DE HARTREE-FOCK

Neste trabalho os cálculos químicos-quânticos partem das equações de Hartree-Fock para sistemas de camadas fechadas. Por essa razão será feita uma apresentação dessas equações. Um desenvolvimento de grande importância na química quântica se deu a partir do cálculo de funções de onda precisas para muitas moléculas diatômicas e poliatômicas, através do método do campo autoconsistente (Self Consistent Field-SCF), desenvolvido por Douglas Hartree (SANTOS et al., 2014d). Em sua teoria Hartree (1928) considerava a função de onda como sendo formada por uma combinação linear antissimétrica de produtos de spins-orbitais. Então a função de onda Hartree-Fock para átomos ou moléculas obedecendo ao princípio da exclusão de Pauli, deve ser escrita como um produto antisimetrizado de spin-orbitais, chamado determinante de Slater, que na sua formula restrita e normalizada, para um sistema de camadas fechadas contendo $2m$ elétrons, é dado por:

$$\Psi(1,2,2m) = \frac{1}{\sqrt{2m!}} \begin{vmatrix} \Phi_1(1) & \bar{\Phi}_1(1) & \cdots & \bar{\Phi}_m(1) \\ \Phi_1(2) & \bar{\Phi}_1(2) & \cdots & \bar{\Phi}_m(2) \\ \vdots & \vdots & \cdots & \vdots \\ \Phi_1(2m) & \bar{\Phi}_1(2m) & \cdots & \bar{\Phi}_m(2m) \end{vmatrix} \quad (4.1)$$

no qual cada spin-orbital $\Phi_i(\mathbf{x},\alpha)$ é um produto de uma função espacial $\varphi_i(\mathbf{x})$ e uma função de spin α ou β

$$\Phi_i(\mathbf{x},\alpha) = \varphi_i(\mathbf{x})\alpha(i) \quad \bar{\Phi}_i(\mathbf{x},\beta) = \varphi_i(\mathbf{x})\beta(i) \quad (4.2)$$

As funções de um elétron $\varphi_i(\mathbf{x})$ são conhecidas como “orbitais”, termo este proposto por Mulliken (1932), sendo este o análogo mecânico-quântico da órbita clássica (BLINDER, 1965).

Considera-se que os orbitais sejam ortonormais

$$\langle \varphi_i(1) | \varphi_j(1) \rangle = \delta_{ij} \quad \delta_{ij} = \begin{cases} 0 & \text{se } i \neq j \\ 1 & \text{se } i = j \end{cases} \quad (4.3)$$

Todos os elementos que formam uma dada coluna do determinante de Slater envolvem o mesmo spin-orbital, já todos os elementos em uma mesma fila envolvem os mesmos

elétrons. Quando a configuração de um átomo ou molécula for de camada fechada esta é representada por um único determinante, diferentemente para uma configuração de camada aberta, pois neste caso utiliza-se uma soma de determinantes de Slater. A análise feita a seguir refere-se a sistemas de configurações de camada fechada. Configurações de camada aberta não serão discutidas aqui visto a complexidade das fórmulas. Objetivando mostrar um formato simples para as expressões, utiliza-se o sistema de unidades atômicas nas equações (SANTOS et al., 2014d).

A expressão referente à energia eletrônica total Hartree-Fock E para um sistema de $2m$ elétrons e n núcleos no estado fundamental é dada pelo teorema do método variacional.

$$E = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \quad (4.4)$$

onde Ψ vem a ser a função de onda na forma de um determinante de Slater e \hat{H} é o operador Hamiltoniano sem considerar as coordenadas de spin do sistema e não-relativístico. Em se tratando de moléculas que é o nosso caso o operador \hat{H} é o Hamiltoniano puramente eletrônico que é obtido após a separação de Born-Oppernheimer (LEVINE, 1991) que escrito em unidades atômicas tem a forma:

$$\hat{H} = -\frac{1}{2} \sum_i \nabla_i^2 - \sum_{\alpha} \sum_i \frac{Z_A}{r_{\alpha i}} + \sum_{i < j} \sum_j \frac{1}{r_{ij}} \quad (4.5)$$

O primeiro termo na equação (4.5) é o operador para a energia cinética dos elétrons, o segundo termo vem a ser a energia potencial de atração entre os elétrons e os núcleos e o último termo refere-se a energia potencial de repulsão entre os elétrons, Z_A é a carga do núcleo A , $r_{\alpha i}$ é a distância do elétron i ao núcleo α e r_{ij} é a distância do elétron i ao elétron j .

O operador \hat{H} pode ser separado em dois outros operadores

$$\hat{H} = \hat{H}_1 + \hat{H}_2 \quad (4.6)$$

O operador \hat{H}_1 é relativo à energia cinética dos elétrons e à energia potencial de interação dos elétrons com os núcleos:

$$\hat{H}_1 = \sum_i^{2m} \hat{h}(i) \quad (4.7)$$

o operador $\hat{h}(i)$ é o Hamiltoniano correspondente ao movimento de um elétron no campo gerado unicamente pelos núcleos

$$\hat{h}(i) = -\frac{1}{2}\nabla_i^2 - \sum_{\alpha} \frac{Z_A}{r_{\alpha j}} \quad (4.8)$$

O operador \hat{H}_2 está associado à repulsão intereletrônica

$$\hat{H}_2 = \sum_{i < j} \sum \frac{1}{r_{ij}} \quad (4.9)$$

Utilizando a função de onda (4.1) e o Hamiltoniano (4.5) na expressão para o funcional da energia (4.4), tem-se como resultado a expressão para a energia eletrônica total de um sistema representado por uma função de onda na forma de um determinante de Slater.

$$E = 2\sum_{i=1} h_{ii} + \sum_i \sum_j (2J_{ij} - K_{ij}) \quad (4.10)$$

Esta relação tem reduzido a integração de muitos elétrons ao conjunto de integrais de dimensões três (h_{ij}) e seis (J_{ij}) e (K_{ij}), escritas da seguinte forma.

$$h_{ii} = \left\langle \varphi_i(1) \left| \hat{h}(1) \right| \varphi_i(1) \right\rangle \quad (4.11)$$

$$J_{ij} = \left\langle \varphi_i(1)\varphi_j(2) \left| \frac{1}{r_{12}} \right| \varphi_i(1)\varphi_j(2) \right\rangle \quad (4.12)$$

$$K_{ij} = \left\langle \varphi_i(1)\varphi_j(2) \left| \frac{1}{r_{12}} \right| \varphi_j(1)\varphi_i(2) \right\rangle \quad (4.13)$$

A integral de um elétron h_{ii} representa a soma da energia cinética de um elétron no orbital $\varphi_i(1)$ e sua integral potencial devido à ação dos núcleos. As integrais de Coulomb J_{ij} e de troca K_{ij} estão associadas à interação entre um elétron no orbital $\varphi_i(\mathbf{x})$ e outro no orbital $\varphi_j(\mathbf{x})$. Na mecânica clássica a integral de Coulomb representa a energia de interação entre duas distribuições de cargas, já a integral de troca não possui análogo na mecânica clássica, sua presença surge devido à anti-simetria da função de onda (BLINDER, 1965).

Variando E na equação (4.4) para cada orbital e fazendo $\delta E = 0$ mantendo-se a ortonormalidade da função de onda chega-se às equações de 1-elétron que definem o método Hartree-Fock

$$\hat{F}(\mathbf{1})\varphi_i(\mathbf{1}) = \varepsilon_i\varphi_i(\mathbf{1}) \quad i = 1, 2, \dots, m \quad (4.14)$$

Onde operador de Fock $\hat{F}(\mathbf{1})$ é dado por

$$\hat{F}(\mathbf{1}) = \hat{h}(\mathbf{1}) + \sum_j^m \left[2\hat{J}_j(\mathbf{1}) - \hat{K}_j(\mathbf{1}) \right] \quad (4.15)$$

onde \hat{J}_j e \hat{K}_j são respectivamente os operadores de Coulomb e de troca

$$\hat{J}_j(\mathbf{1})\varphi_i(\mathbf{1}) = \left[\int \varphi_j^*(2) \frac{1}{r_{12}} \varphi_j(2) d\tau_2 \right] \varphi_i(\mathbf{1}) \quad (4.16)$$

$$\hat{K}_j(\mathbf{1})\varphi_i(\mathbf{1}) = \left[\int \varphi_j^*(2) \frac{1}{r_{12}} \varphi_i(2) d\tau_2 \right] \varphi_j(\mathbf{1}) \quad (4.17)$$

As equações de Hartree-Fock são resolvidas em geral por um procedimento iterativo (SCF). Ao final do procedimento, os ε_i são os autovalores monoelétrônicos **HF** do sistema. Cada um dos ε_i é freqüentemente chamado de energia orbital e é interpretado como sendo a energia de um elétron no orbital φ_i , resultante de sua energia cinética, sua energia de atração com os núcleos e sua energia de troca e de repulsão devido a todos os outros elétrons em suas densidades de carga $|\varphi_i|^2$ (SANTOS et al., 2014d).

A equação (4.14) fornece

$$\varepsilon_i = \langle \varphi_i(\mathbf{1}) | \hat{F}(\mathbf{1}) | \varphi_i(\mathbf{1}) \rangle \quad (4.18)$$

Substituindo a equação (4.15) em (4.18) podemos relacionar as energias orbitais com as integrais (4.11), (4.12) e (4.13).

$$\varepsilon_i = h_{ii} + \sum_{j=1}^m (2J_{ij} - K_{ij}) \quad (4.19)$$

Pode-se também achar a energia eletrônica total E a partir das energias orbitais. Mas esta energia total não é simplesmente igual à soma das energias de um elétron ε_i . Tal fato ocorre devido à soma das energias de um elétron incluir duas vezes cada interação elétron-elétron, ou seja, a repulsão entre os elétrons 1 e 2 contribui para as energias de um elétron associado com ambos os elétrons. Assim tem-se que o segundo termo na equação abaixo corrige esse problema

$$E = 2 \sum_i^m \varepsilon_i + \sum_i^m \sum_j^m (2J_{ij} - J_{ij}) \quad (4.20)$$

Assim substituindo (4.19) em (4.20) encontra-se

$$E = \sum_{i=1}^m \left[2h_{ii} + \sum_{j=1}^m (2J_{ij} - K_{ij}) \right] \quad (4.21)$$

que é igual a

$$E = \sum_{i=1}^m (\varepsilon_i - h_{ii}) \quad (4.22)$$

com h_{ii} sendo o mesmo da equação (4.11).

4.2 MÉTODO HARTREE-FOCK-ROOTHAAN

O problema de se resolver às equações de Hartree-Fock para átomos e moléculas deve-se ao fato da ausência de simetria central. Por isso se faz necessário usar aproximações para os melhores orbitais. Assim para sistemas que contenham muitos elétrons, uma maneira aproximada de resolver as equações de Hartree-Fock consiste em expandir o orbital de Hartree-Fock $\varphi_i(\mathbf{x})$ em uma combinação linear de K funções de base χ_μ , conforme foi proposto por Roothaan (1951), sendo este método denominado de método Hartree-Fock-Roothaan (HFR) ou também Molecular-Orbital Linear Combination Atomic Orbitals (MO-LCAO).

Assim o orbital pode ser expandido em uma combinação linear da seguinte maneira.

$$\varphi_i = \sum_{\mu}^K c_{\mu i} \chi_{\mu} \quad (4.23)$$

onde os $c_{\mu i}$ são os coeficiente de expansão, os quais serão tratados como parâmetros variacionais e as funções de base χ_{μ} são os orbitais atômicos tipo Slater ou tipo Gaussianos.

Para representar exatamente os orbitais, as funções deveriam formar um conjunto completo. Entretanto, esse requerer um número infinito dessas funções. O que se deve usar na realidade é um número finito de funções de base. Os orbitais devem obedecer à condição de ortonormalidade. Caso contrário é possível fazer uma transformação linear tornando-os assim ortonormais:

$$\sum_{\mu\nu} \mathbf{c}_{\mu i}^* \mathbf{c}_{\mu j} \mathbf{S}_{\mu\nu} = \delta_{ij} \quad (4.24)$$

onde $\mathbf{S}_{\mu\nu}$ é a integral de sobreposição (overlap)

$$\mathbf{S}_{\mu\nu} = \langle \chi_{\mu}(i) | \chi_{\nu}(i) \rangle \quad (4.25)$$

Para moléculas os $\varphi_i(\mathbf{x})$ são orbitais moleculares e os χ_{μ} são os orbitais atômicos. Assim existe uma melhora significativa nos cálculos computacionais quando as funções orbitais são expandidas em termos de um conjunto finito de funções de base. As equações integro-diferenciais são então transformadas em equações algébricas para os coeficientes de expansão (BLINDER, 1965; SANTOS et al., 2014d).

A energia eletrônica total é obtida quando a equação (4.23) é substituída em (4.21).

$$E = 2 \sum_i^{ocup} \sum_{\mu\nu} \mathbf{c}_{\mu i}^* \mathbf{c}_{\nu i} h_{\mu\nu} + \sum_{ij} \sum_{\mu\nu\lambda\sigma} \mathbf{c}_{\mu i}^* \mathbf{c}_{\lambda j}^* \mathbf{c}_{\nu i} \mathbf{c}_{\sigma j} [2\langle \mu\lambda | \nu\sigma \rangle - \langle \mu\nu | \lambda\sigma \rangle] \quad (4.26)$$

onde

$$h_{\mu\nu} = \langle \chi_{\mu}(1) | \hat{h}(1) | \chi_{\nu}(1) \rangle \quad (4.27)$$

com $\hat{h}(1)$ dado pela equação (4.8), substituindo-se i por $\mathbf{1}$.

As integrais $\langle \mu\lambda | \nu\sigma \rangle$ e $\langle \mu\nu | \lambda\sigma \rangle$ são integrais de interação de dois elétrons que são representadas por

$$\langle \mu\lambda | \nu\sigma \rangle = \iint \chi_{\mu}^*(1) \chi_{\lambda}^*(2) \left| \frac{1}{r_{12}} \right| \chi_{\nu}(1) \chi_{\sigma}(2) d\tau_1 d\tau_2 \quad (4.28)$$

$$\langle \mu\nu | \lambda\sigma \rangle = \iint \chi_{\mu}^*(1) \chi_{\nu}^*(2) \left| \frac{1}{r_{12}} \right| \chi_{\lambda}(1) \chi_{\sigma}(2) d\tau_1 d\tau_2 \quad (4.29)$$

Os melhores coeficientes $\mathbf{c}_{\mu i}$ são determinados variando-se a energia eletrônica dada pela equação (4.26) com relação a eles, obedecendo a condição de ortonormalidade.

Com isso chega-se as equações de Hartree-Fock-Roothaan

$$\sum_{\nu} \mathbf{c}_{\nu i} \left\{ h_{\mu\nu} + \sum_j \sum_{\nu\sigma} \mathbf{c}_{\nu j}^* \mathbf{c}_{\sigma j} [2\langle \mu\lambda | \nu\sigma \rangle - \langle \mu\sigma | \lambda\sigma \rangle] \right\} = \sum_j \sum_{\nu} \varepsilon_{ij} \mathbf{c}_{\nu j} \mathbf{S}_{\mu\nu} \quad (4.30)$$

onde o operador de Fock definido como

$$\mathbf{F}_{\mu\nu} = h_{\mu\nu} + \sum_j \sum_{\nu\sigma} \mathbf{c}_{\nu j}^* \mathbf{c}_{\sigma j} [2\langle \mu\lambda | \nu\sigma \rangle - \langle \mu\nu | \lambda\sigma \rangle] \quad (4.31)$$

Podemos substituir a equação (4.31) em (4.30) e dessa forma obter as equações de Hartree-Fock-Roothaan na seguinte notação

$$\sum_{\nu} c_{\nu i} F_{\mu\nu} - \sum_j \sum_{\nu} \varepsilon_j c_{\nu j} S_{\mu\nu} = 0 \quad \mu = 1, 2, \dots, k \quad (4.32)$$

Aplicando-se uma transformada unitária na equação (4.32) para diagonalizar a matriz ε , tem-se

$$\sum_{\nu} (F_{\mu\nu} - \varepsilon_i S_{\mu\nu}) c_{\nu i} = 0 \quad \mu = 1, 2, \dots, k \quad (4.33)$$

A equação (4.33) pode ser escrita na forma matricial dessa forma tem-se

$$FC = SCE \quad (4.34)$$

As equações de Hartree-Fock-Roothaan são resolvidas da mesma forma que as equações de Hartree-Fock, ou seja, por um processo iterativo (SCF-MO-LCAO). Os ε_i obtidos no final do procedimento são os autovalores de HFR do sistema (SANTOS et al., 2014d).

4.3 LIMITE HARTREE-FOCK

No método Hartree-Fock, utiliza-se uma função de onda eletrônica constituída somente por um determinante de Slater, fornecendo assim apenas uma descrição aproximada da função de onda exata devida esta não poder ser descrita por um único determinante de Slater. A solução exata da equação de Schrödinger não seria obtida caso os orbitais espaciais $\varphi_i(\mathbf{x})$ fossem expandidos em uma combinação linear de funções de base χ_{μ} , embora quanto maior e mais completo for o conjunto dessas funções maior será o grau de flexibilidade na expansão para os spin-orbitais e menor o valor esperado para a energia. Conjunto de bases maiores diminuirão a energia **HF** até um limite determinado. Esse limite é o mais baixo que pode ser obtido com uma função de onda de um determinante e é chamado de Limite Hartree-Fock (SANTOS et al., 2014d).

Mesmo assim, conforme o método variacional, essa energia limite E_{HF} ainda estará acima da energia exata não-relativística $E_{ex}(\mathbf{NR})$ devido à energia de correlação eletrônica E_{corr}

$$E_{ex}(\mathbf{NR}) = E_{HF} + E_{corr} \quad (4.35)$$

O método SCF é uma aproximação válida, porém gera erros na energia, pois descreve de modo aproximado às interações entre os elétrons. Devem-se considerar também as

interações instantâneas entre os elétrons. Os movimentos dos elétrons estão correlacionados uns com os outros, ou seja, existe uma correlação nas posições dos elétrons diferentes e precisa ser levada em conta. O valor exato da energia E incluindo a correlação eletrônica e os efeitos relativísticos E_{rel} são dados por:

$$E = E_{HF} + E_{corr} + E_{rel} \quad (4.36)$$

4.4 CONJUNTOS DE BASES DE VALÊNCIA SEPARADA UTILIZADOS NOS CÁLCULOS DAS PROPRIEDADES MOLECULARES

Para realizar cálculos atômicos e moleculares, determina-se um conjunto de base formado por funções matemáticas. Tais bases consistem em uma combinação linear das funções de onda de um elétron em termos de um número finito de funções de base, que contém um conjunto de parâmetros a serem otimizados. A denominação dos conjuntos de base depende dos vários tipos de funções de base, do número dessas funções empregadas na expansão das funções monoelétrônicas (orbitais) e das características dos parâmetros a serem otimizados. Dessa forma é de fundamental importância a escolha cuidadosa dessas funções quando se deseja obter resultados precisos (SANTOS et al., 2014d).

Neste trabalho, os cálculos foram realizados com o conjunto de base valência separada HF/6-31G** disponível no programa Gaussian 03 para realização de cálculos da Teoria do Orbital Molecular.

4.4.1 Funções tipo Slater

Sistemas atômicos que contenham apenas um elétron, átomos hidrogenóides, possuem como solução da equação de Schrödinger não-relativística, uma função do tipo

$$\Psi_{nlm}(r, \theta, \varphi) = R_{nl}(r) Y_{lm}^{(\theta, \varphi)} \quad (4.4.1)$$

onde n , l e m são os números quânticos principais, momento angular orbital e magnético, $R_{nl}(r)$ é a parte radial e $Y_{lm}^{(\theta, \varphi)}$ é a parte angular denominada harmônico esférico.

A parte radial possui a forma

$$R_{nl}(r) = N_{nl} \left(\frac{2Zr}{na} \right)^l \exp\left(-\frac{Zr}{na} \right) L_{n-l-1}^{2l+1} \left(\frac{2Zr}{na} \right) \quad (4.4.2)$$

onde N_{nl} vem a ser um fator de normalização, Z é a carga nuclear, r é o raio, a é uma constante e $L_{n+1}^{2l+1}\left(\frac{2Zr}{na}\right)$ representa de forma genérica um polinômio chamado polinômio associado de Laguerre.

A parte angular é igual a:

$$Y_{lm}^{(\theta,\varphi)} = N_{lm} P_l^{(m)}(\cos\theta) e^{im\varphi} \quad (4.4.3)$$

sendo N_{lm} um fator de normalização e $P_l^{(m)}$ o polinômio associado de Legendre. Por mais que os orbitais hidrogenóides sejam ortogonais eles não formam um conjunto completo de funções contínuas. Eles tiveram pouca aplicação, devido muitas das integrais requeridas em cálculos de orbitais moleculares serem um tanto difíceis de serem calculadas, principalmente quando se tem valores elevados do número quântico principal por causa da complexidade do polinômio em r .

Então Slater (1930) propôs uma forma analítica mais simples para a função radial, introduzindo então funções do tipo Slater (Slater Type Functions - STFs).

$$R_{nl}(r) = \frac{[2\zeta/a_0]^{n+1/2}}{[(2n)!]^{1/2}} r^{n-1} \exp(-\zeta r/a_0) \quad (4.4.4)$$

Sendo que o expoente orbital é escrito como

$$\zeta = \frac{Z-s}{n^*} \quad (4.4.5)$$

Onde s é uma constante relacionada ao efeito de blindagem dos elétrons das camadas internas dos átomos e n^* é um número quântico principal efetivo. Logo a forma geral das funções tipo Slater pode ser escrita da seguinte maneira, que é análoga a equação.

$$\Psi_{nlm}(\zeta, r, \theta, \varphi) = [(2n)!]^{1/2} (2\zeta)^{n+1/2} r^{n-1} e^{-\zeta r} Y_{lm}^{(\theta,\varphi)} \quad (4.4.6)$$

Slater (1930) avaliou por meio de regras empíricas a escolha dos parâmetros s e n^* , visando assim uma boa aproximação para os melhores orbitais atômicos desse tipo. Os valores de ζ foram determinados pelo método variacional por Clementi e Raimondi (1963), Clementi, Raimoni e Reinhart (1967) para átomos neutros até o rutênio em seus respectivos estados fundamentais, utilizando para isso o método SCF.

Os expoentes ζ são números positivos e ajustáveis nos métodos de cálculos. Esses expoentes determinam o tamanho do orbital, assim expoentes grandes caracterizam orbitais densos e expoentes pequenos caracterizam orbitais difusos.

Os chamados orbitais tipo Slater (Slater Type Orbitals - STOs) são formados pelo produto da parte angular, equação (4.4.3), pelo produto da parte radial, equação (4.4.4). Os STOs geram representações razoáveis dos orbitais atômicos. Porém, como eles substituem o polinômio em r nos orbitais hidrogenóides por uma simples potência de r , não possuem o número próprio de nós radiais, logo não representam bem a parte interior de um orbital.

Uma das limitações é que eles não são de modo algum ortogonais, por mais que isso possa ser corrigido utilizando um conjunto de STOs ortogonalizados. Além do mais o uso desses orbitais em cálculos moleculares SCF faz com que as integrais multicêntricas, que envolvem interações entre elétrons, torna-se um tanto quanto difícil de ser solucionada numericamente, aumentando assim o tempo computacional (SANTOS et al., 2014d).

4.4.2 Conjunto de base mínima

Conjunto de base mínima ou base “single-zeta” é o conjunto contendo uma única função para representar cada orbital atômico ocupado de diferentes números quânticos n e l da configuração eletrônica. Por isso, esse conjunto apresenta um tempo computacional reduzido para a execução dos cálculos, assim pode ser usado em cálculos envolvendo moléculas grandes que é o nosso caso. Nos cálculos onde um conjunto de base mínima é utilizado são usados em geral orbitais tipo Slater. Devido ao pequeno tamanho da base, os conjuntos de base mínima geram resultados apenas de forma qualitativa das propriedades. Mesmo assim os cálculos da estrutura eletrônica eram até por volta de 1960 todos realizados em termos de base mínima. Os primeiros métodos semi-empíricos mais utilizados foram baseados em um conjunto de base mínima STOs (HOFFMANN, 1963; POPLE; BEVERIDGE, 1970; SANTOS et al., 2014d).

4.4.3 Funções do tipo gaussianas

4.4.3.1 Funções gaussianas contraídas

O uso de STOs em cálculos de estrutura eletrônica gera problemas computacionais, devido o surgimento de integrais multicêntricas, embora esses orbitais descrevam bem o comportamento funcional dos orbitais moleculares. Visando simplificar os cálculos de

integrais multicêntricas, Boys (1950) propôs para os cálculos envolvendo moléculas o uso de funções do tipo Gaussianas (Gaussian Type Functions - GTFs)

$$R_{nl}(r) = \left[\frac{2^{2n+3/2}}{(2n-1)! \sqrt{\pi}} \right]^{1/2} \alpha^{2n+1/4} r^{n-1} \exp(-\alpha r^2) \quad (4.4.7)$$

onde n e l são os números quânticos principal e momento angular orbital e α é o expoente orbital que é um parâmetro variacional. O uso de GTFs em cálculos de integrais de interação eletrônica apresenta uma enorme vantagem, pois o produto de duas dessas funções entradas em centros diferentes é equivalente a uma única função centrada em um novo centro, assim existe a redução de integrais multicêntricas a integrais calculadas em termos de funções centradas em um mesmo ponto (SHAVITT, 1963).

As funções do tipo Gaussianas possuem um comportamento deficiente dos orbitais atômicos em relação às funções tipo Slater, a razão disso é que as GTFs não têm o “cusp” na região próximo ao núcleo. Assim as GTFs apresentam um comportamento funcional diferente daquele apresentado pelos orbitais moleculares, dessa forma é necessário utilizar duas a cinco GTFs para representar de forma adequada cada STO. Como o número de integrais de interação eletrônica oriundas da utilização de um conjunto de base de dimensão m aumenta na ordem de m^4 , a maior velocidade e simplificação dos cálculos das integrais de interação eletrônica em termos de GTFs compensa o grande número de integrais a serem calculadas, quando se compara utilizando STFs (HEHRE; STEWART; POPLE, 1969).

A solução das equações do método SCF nos cálculos HFR é também um processo que utiliza um enorme tempo computacional, este sendo também proporcional à quarta potência do número de funções da base. Porém o número de ciclos do processo iterativo do método SCF aumenta com o número de coeficientes a serem otimizados. Assim o uso de GTFs contraídas originadas de combinações lineares de Gaussianas primitivas é geralmente mais adequado (HEHRE; STEWART; POPLE, 1969).

As funções de base contraídas podem ser escolhidas de forma a se assemelhem com STOs, orbitais atômicos HF ou qualquer outro conjunto de funções. Um tipo de GTFs contraída é o conjunto STO-NG. Tal nomenclatura vem a ser um conjunto de GTFs contraídas que descreve uma STF por meio de N GTFs primitivas usadas na contração. Cada STO é então aproximado como uma combinação linear de N GTFs, no qual os coeficientes e expoentes contraídos são escolhidos de modo que a função de base aproxime-se de uma STF. Dessa forma o tempo computacional diminui, porém os resultados gerados pelo conjunto de GTFs contraídos não são bons. Os STOs ns e np são aproximados pelas respectivas funções

$$\varphi_{ns}(r) = \sum_{k=1}^m d_{ns,k} g_{1s}(\alpha_{nk}, r) \quad n = 1, 2, 3 \quad (4.4.8)$$

$$\varphi_{np}(r) = \sum_{k=1}^m d_{np,k} g_{2p}(\alpha_{nk}, r) \quad n = 1, 2, 3 \quad (4.4.9)$$

onde g_{1s} e g_{2p} são GTFs, α_{nk} são expoentes Gaussianos e $d_{ns,k}$ e $d_{np,k}$ são coeficientes da expansão, onde os três últimos parâmetros são determinados pelo método dos mínimos quadrados (STEWART, 1969).

4.4.3.2 Conjunto de base de valência separada

Com o intuito de se obter melhores resultados, muitos estudos têm sido feitos visando encontrar conjuntos de base com essa capacidade. Nos últimos anos, além dos conjuntos de base já descritos anteriormente, outros conjuntos estão sendo empregados nos cálculos de estrutura eletrônica. Dentre esses existem os conjuntos de base de valência separada. Conjuntos de base de valência separada são conjuntos estendidos de GTFs contraídas, onde os mais usuais são os referidos a seguir: 4-31G, 3-21G, 6-31G e 6-311G. Nesses conjuntos usam-se duas funções para os orbitais de valência e somente uma para os orbitais da camada interna, podendo cada uma dessas funções serem uma combinação linear ou não de Gaussianas primitivas (LEVINE, 1991).

A ocorrência disso é devido à camada interna contribuir pouco para as propriedades de interesse químico. Como as funções das camadas internas não duplicadas isso gera efeitos na energia total, porém representa pouco em se tratando de momentos dipolares, potencial de ionização de valência, densidade de carga, energia de dissociação entre outras propriedades químicas (LEVINE, 1991; SANTOS et al., 2014d).

Por exemplo, na base 4-31G cada orbital atômico das camadas internas é descrito por somente uma GTF contraída que é formada pela combinação linear de quatro GTFs primitivas. Para cada orbital atômico da camada de valência existem duas funções de base, sendo que uma delas é uma GTF contraída a partir da combinação linear de três GTFs primitivas, descrevendo a parte interna do orbital de valência, já a outra é uma única GTF primitiva, a qual descreve a parte externa do orbital de valência.

Considerando os átomos da primeira fila (Li até F) tem-se

$$\varphi_{1s}(r) = \sum_{k=1}^4 d_{k,1s} g_{1s}(\alpha_{k,1s}, r) \quad (4.4.10)$$

$$\varphi'_{2s}(r) = \sum_{k=1}^3 d'_{k,2s} g_{1s}(\alpha'_{k,2sp}, r) \quad (4.4.11)$$

$$\varphi''_{2s}(r) = g_{1s}(\alpha''_{2sp}, r) \quad (4.4.12)$$

$$\varphi'_{2p}(r) = \sum_{k=1}^3 d'_{k,2p} g_{2p}(\alpha'_{k,2sp}, r) \quad (4.4.13)$$

$$\varphi''_{2p}(r) = g_{2sp}(\alpha''_{2sp}, p) \quad (4.4.14)$$

Para o átomo de hidrogênio, que não possui camada interna, tem-se

$$\varphi'_{1s}(r) = \sum_{k=1}^3 d'_{k,1s} g_{1s}(\alpha'_{k,1s}, r) \quad (4.4.15)$$

onde as funções que aparecem com (') são funções internas enquanto que as que aparecem com (") são funções externas. De modo análogo à estrutura da base descrita anteriormente é a base 6-31G. Só que agora cada orbital das camadas internas é representado por uma GTF contraída oriunda da combinação linear de seis GTFs primitivas. No caso da base 6-311G, inclui-se uma GTF primitiva ao conjunto 6-31G para representar uma nova camada da valência externa.

4.4.4 Função de polarização

Todos os conjuntos de base discutidos até agora possuem uma característica peculiar, ou seja, eles compreendem funções restringidas a estarem centradas nos núcleos. Porém, existem evidências mostrando que os orbitais atômicos se distorcem ou polarizam quando formam uma molécula. Em razão disso deve-se levar em consideração a possibilidade de deslocamento não-uniforme de cargas elétricas fora do núcleo atômico, ou seja, a polarização. Desse modo é possível obter uma melhor descrição das mudanças bem como também das deformações nos orbitais atômicos dentro da molécula (SANTOS et al., 2014d).

Um modo de se considerar tal polarização é introduzindo na base em questão funções em que os valores de l (número quântico do momento angular orbital) maiores que aqueles do estado fundamental de um dado átomo. A esse tipo de funções dá-se o nome de funções de polarização. Para o átomo de hidrogênio a descrição do estado fundamental utiliza unicamente funções s . As funções p , d ,... centradas no H, para o cálculo molecular, são consideradas como

funções de polarização. De forma geral, com a inclusão de funções de polarização na base molecular existe uma maior possibilidade de obter melhores resultados para muitas das propriedades de interesse químico, tais como, energia de dissociação e momentos dipolares. Em termos reais observa-se que não é satisfatória a inclusão de funções de polarização de simetria d e f juntamente com os conjuntos de base s e p pequenos, isto é, as funções de polarização só devem ser adicionadas quando se trabalha com conjuntos de base ditos saturados (MULLIKEN; LIU, 1971).

Das bases de valência separada com funções de polarização mais utilizadas em cálculos moleculares estão as $STO-3G^*$, $3-1G^*$, $6-31G^*$, $6-31G^{**}$, $6-311G^*$ e $6-311G^{**}$ (HEHRE, 1986). As bases $6-31G^*$ e $6-31G^{**}$ são formadas pela inclusão de funções de polarização à base $6-31G$, sendo que a base $6-31G^*$ é construída pela adição de um conjunto de cinco funções de polarização do tipo GTFs de simetria d à base $6-31G$, para cada átomo diferente de hidrogênio e do hélio, já a base $6-31G^{**}$ é construída adicionando-se um conjunto de três funções de polarização do tipo GTFs de simetria p ao conjunto $6-31G^*$, para cada átomo de hidrogênio (MULLIKEN; LIU, 1971; SANTOS et al., 2014d).

4.4.5 Funções difusas

A utilização de conjunto de bases de valência com função de polarização em cálculos envolvendo ânions não gera bons resultados, devido à nuvem eletrônica dos sistemas aniônicos possuírem uma tendência à expansão. Dessa forma é necessário também adicionar funções difusas apropriadas, pois elas possibilitam uma maior ocupação da região do espaço pelos orbitais (SANTOS et al., 2014b; SANTOS et al., 2014d).

A importância das funções difusas em cálculos de metais de transição é devido os átomos metálicos apresentarem orbitais tipo d e estes têm características difusas. Então se torna necessário a inclusão de funções difusas à função de base associada à configuração do átomo metálico neutro com o intuito de se obter uma melhor descrição do complexo metálico. A grande importância das funções difusas deve-se ao fato de que elas descrevem melhor os orbitais moleculares mais afastados dos núcleos (SANTOS et al., 2014b; SANTOS et al., 2014d).

As bases $3-21+G$ e $6-21+G^*$ são formadas, respectivamente, a partir das bases $3-21G$ e $6-31G^*$ pela inclusão de quatro (4) funções difusas (s, p_x , p_y e p_z) para cada átomo diferente do hidrogênio. Já as bases $3-21++G$ e $6-21++G^*$ são formadas, respectivamente, a partir das

bases 3-21G e 6-21G* pela adição de uma função difusa s em cada átomo de hidrogênio (SANTOS et al., 2014b; SANTOS et al., 2014d).

4.5 MODELAGEM MOLECULAR DA ARTEMISININA E DOS DERIVADOS ANTIMALÁRICOS ESTUDADOS

A artemisinina (Figura 7) e seus derivados são os mais potentes compostos com ação antimalárica, podendo reduzir a biomassa do parasita da malária a 1.10^{-4} (NOEDL, 2005).

Figura 7. Planta *Artemisia annua* L.



Fonte: MACÊDO (2010).

A artemisinina tem uma ampla e extraordinária atividade contra parasitas na forma assexuada, matando todos os estágios, na malária do tipo causada por *Plasmodium falciparum* a artemisinina também mata os gametócitos, incluindo os quatro estágios dos gametócitos, que são sensíveis somente a primaquina (WORLD HEALTH ORGANIZATION, 2010).

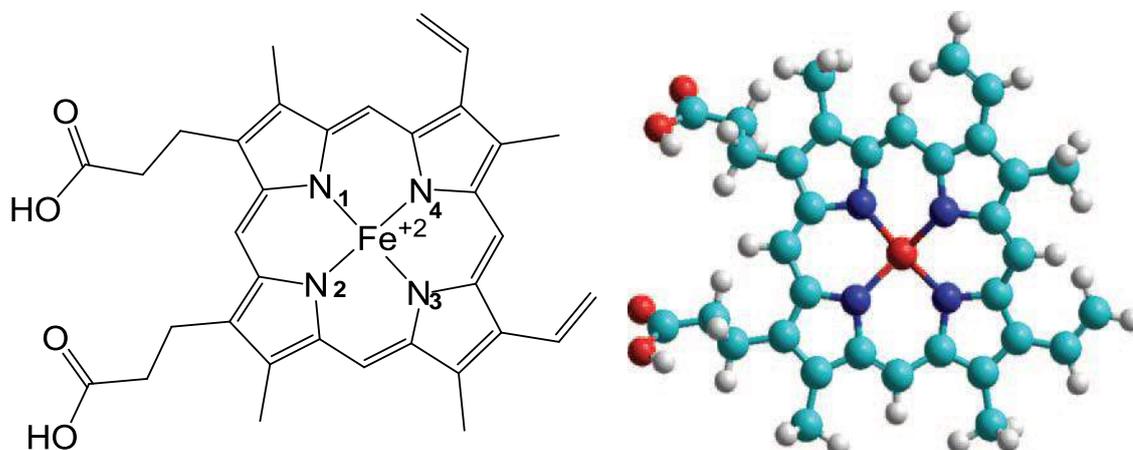
Ensaio biológico *in vitro* indicaram que alguma quantidade de ferro deve ser acrescentado à artemisinina para mostrar a atividade antimalárica. No homem, o composto heme (Figura 8), que é um produto oriundo de um processo digestivo é acreditado ser a fonte de ferro para a artemisinina. Em humanos os parasitas da malária digerem mais de 70% da hemoglobina dentro das células vermelhas do sangue infectado, tendo a globina e heme como produtos, a globina é hidrolisada, gerando aminoácidos que são utilizados na síntese de

proteínas pelo parasita, já o heme sofre um processo de polimerização tendo a hemozoína como produto (TONMUNPHEAN; PARASUK; KOKPOL, 2000; TONMUNPHEAN; PARASUK; KOKPOL, 2001).

A polimerização do heme é um alvo de algumas drogas antimaláricas, como a cloroquina, que inibe este processo. A artemisinina também inibe a polimerização do heme. Cepas de *Plasmodium berghei* resistentes a cloroquina que por falta da hemozoína, possivelmente por causa da polimerização do heme não ocorrer, são também resistentes a artemisinina. Isto reforça a visão de que a inibição da polimerização do heme é o modo de ação da artemisinina, indicando assim que a artemisinina muito possivelmente interage como o heme livre e em razão disso iniba o processo de polimerização (TONMUNPHEAN; PARASUK; KOKPOL, 2001).

A numeração do heme é sistemática para efeito de determinação dos parâmetros geométricos: comprimento de ligação, ângulo de ligação e ângulo de torção entre artemisinina e seus derivados (fármacos) com o heme (receptor).

Figura 8. Estrutura da proteína heme.



Fonte: SANTOS et al. (2014a)

O Fe⁺² tem sido proposto catalisar a clivagem redutiva do peróxido para dar origem aos radicais oxigênicos centrados ou alcoxil e então radicais carbonos centrados e produtos neutros. Estes intermediários reativos são sugeridos serem atuais agentes parasiticidas quando reagem com biomoléculas vitais ou sensíveis no parasita (HAYNES; KRISHNA, 2004).

A interação com a heme pode ser importante para a atividade antimalárica, pois quando adicionada à cultura de parasitas, a artemisinina tem mostrado associar-se com a

hemozoína e causar danos ao vacúolo digestivo do parasita. *In vitro*, tanto o ferro livre e heme catalisam a conversão da artemisinina em radicais livres que alquilam as proteínas do parasita e que forma adutos com o heme (FRANCIS; SULLIVAN; GOLDBERG, 1997).

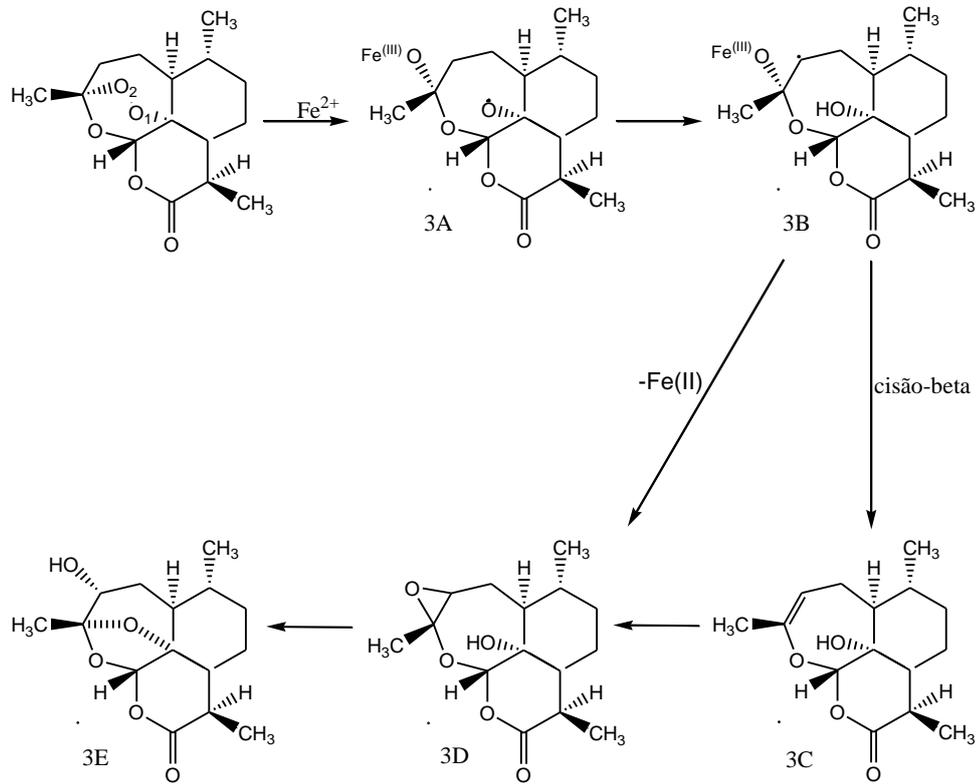
O mecanismo para ação antimalárica destes compostos é ainda sem conclusão, contudo investigações dão uma grande significância para o grupo endoperóxido na artemisinina para atividade biológica. Parasitas de malária em humanos degradam a hemoglobina e a célula vermelha do sangue em heme e globina. Subsequentemente, a globina é hidrolisada para produzir aminoácidos como fonte para síntese de proteínas, e a porção tóxica do heme será desintoxicar pelo processo de polimerização a hemozoína.

Na descoberta e desenvolvimento de drogas, conhecer o mecanismo de ação pode ajudar o desenvolvimento de novas e mais efetivas drogas. Para o mecanismo de ação da artemisinina, existem duas possibilidades para o ferro do heme acoplar nos compostos artemisínicos. Posner (1994) propôs o ataque do ferro ao composto, se dá pela posição O2 e produz o radical livre na posição O1 com posterior rearranjo para um radical livre em C4 (Figura 9). Este radical (3B) foi sugerido por ser uma importante substância para atividade biológica. 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 malárico e possivelmente causa danos aos parasitas (ASAWAMAHASAKDA, 1994; POSNER et al., 1995).

Por outro lado, Jefford et al., 1996 acreditam que o ataque do ferro aos compostos se dá na posição O1 e produz um radical livre na posição O2 (Figura 10). 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 (JEFFORD et al., 1996).

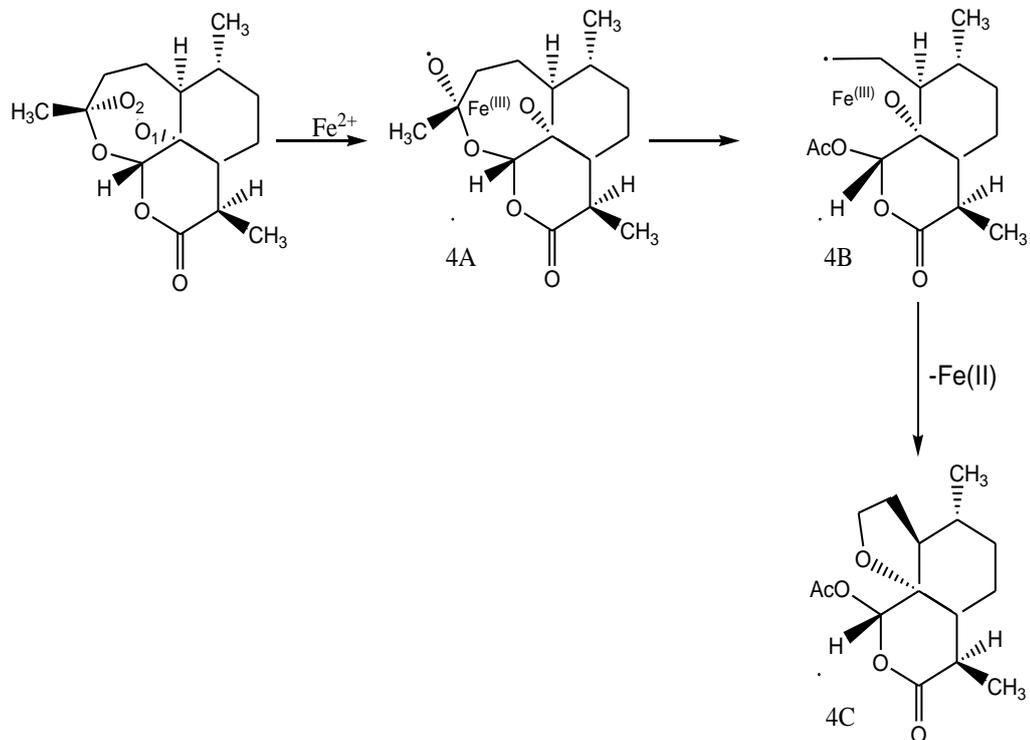
Na artemisinina a ligação peróxido nos oxigênios 1 e 2 (O1 e O2) está contida na estrutura farmacofórica do anel trioxano da artemisinina, sendo essencial para a expressão da atividade antimalárica. Peróxidos são conhecidos por sofrerem cisão redutiva por metais de transição de baixa valência, gerando radicais livres de oxigênios. Estes radicais livres de oxigênios são potentes agentes de abstração de hidrogênios, podem também gerar radicais livres de carbonos por abstrair átomos de hidrogênios intramolecular, estes radicais livres de carbonos são presumidos reagir com biomoléculas (OLLIARO, 2001).

Figura 9. Mecanismo de ação artemisinina e heme proposto por Posner (1995).



Fonte: POSNER (1995)

Figura 10. Mecanismo de ação da artemisinina.



Fonte: SANTOS et al. (2014b)

4.5.1 Validação do método computacional aplicado na modelagem molecular da artemisinina

O composto da artemisinina foi construído seguindo a estratégia descrita: inicialmente a estrutura da artemisinina, foi realizada com o programa GaussView 5.0 (FRISCH et al., 2004), e otimizada completamente com os diferentes métodos e conjunto de base – Semi-empíricos (AM1, PM3 e ZINDO), ab initio/Hartree-Fock (HF/6-31G, HF/6-31G* e HF/6-31G**) e DFT (B3LYP/3-21G, B3LYP/3-21G*, B3LYP/3-21G**) – com o implementado no programa Gaussian 03 (FRISCH et al., 2004). Estes cálculos foram realizados para encontrar o método e conjunto de base com o melhor ajuste entre o tempo computacional e a precisão da informação relativa aos dados experimentais (LISGARTEN et al., 1998).

Depois da otimização os parâmetros geométricos teóricos da artemisinina na região no anel 1,2,13-trioxano foram determinados (comprimento de ligação, ângulo de ligação e ângulo diedro ou ângulo de torção dos átomos que formam o anel), afim de avaliar a qualidade da função de onda molecular comparando os parâmetros geométricos teóricos com os dados experimentais.

A estrutura experimental da artemisinina foi retirada do Banco de Dados Cambridge Structural CSD, com REFCODES: QNGHSU10, fator R cristalográfico 3.6 (ALLEN, 2002). A numeração dos átomos adotada neste estudo é mostrada na Figura 11 (artemisinina). Os cálculos envolvendo química quântica foram realizados em um ambiente IBM RISC 6000.

4.5.2 Análise multivariada

Quando medidas são feitas sobre um número de objetos, os resultados são usualmente organizados em uma matriz, que é chamada matriz de dados. As medidas (neste estudo os parâmetros geométricos, relativo aos métodos em diferentes conjuntos de base) são colocadas nas colunas, e os objetos (neste estudo os métodos químico-quânticos aplicados) estão associados com as linhas. A etapa de análise multivariada foi efetivada com o programa Piroutte 3.10 (INFOMETRIX, 2001).

4.5.2.1 Análise de componentes principais (PCA)

É um método de compressão de dados baseado na correlação entre as variáveis, essa compressão de dados gera um pequeno conjunto de variáveis as quais são denominadas componentes principais, sendo que estas são ortogonais entre si, de modo que a correlação entre as variáveis não limita sua aplicação, diferentemente da regressão linear múltipla que é sensível à presença de variáveis altamente correlacionadas, pois torna os coeficientes de regressão instáveis e sem significado (FERREIRA, 2002; FERREIRA; MONTANARI; GAUDIO, 2002). Para uma matriz de dados multidimensionais, métodos de estatística multivariada são necessários empregar a padronização por autoescalamento, a fim de tratar todas as variáveis com o mesmo grau de importância para compreender tais dados em sua totalidade. O principal objetivo da análise é mostrar os dados multidimensionais em um espaço de baixa dimensionalidade com o mínimo de perda de informação total (FRANKE; GRUSKA, 1996).

A PCA foi realizada com processamento autoescalado, com máximo de três fatores (3 PCS) através do procedimento “leave-one-out e método de validação cross-validation”. A matriz de dados foi construída com dimensão 10x18, onde cada coluna foi relacionada com cinco métodos e seis conjunto de base, sendo uma das colunas relacionada com os parâmetros geométricos experimentais (LISGARTEN et al., 1998), e cada linha representa os 18 parâmetros geométricos do anel 1,2,13-trioxano (comprimentos de ligação, ângulos de ligação e ângulos de torção). O resultado final da PCA é a seleção de um pequeno número de parâmetros geométricos da artemisinina que pode ser melhor relacionada com a variável dependente, neste caso o desvio padrão dos diferentes métodos e conjunto de base.

4.5.2.2 Análise de agrupamento hierárquico (HCA)

Como em PCA os resultados do HCA são qualitativos, sendo dispostos na forma de um dendograma permitindo assim visualizar os métodos estudados ou a variável (parâmetros geométricos da artemisinina) em um espaço bi-dimensional, que ilustra as fusões ou divisões feitas em cada estágio sucessivo da análise. As amostras (métodos e conjunto de base) são representadas pelos ramos na parte inferior do dendograma. A

similaridade entre os aglomerados é dada pelo comprimento dos seus ramos, de modo que os métodos e conjunto de base apresentando baixa similaridade têm ramos longos, enquanto que tais métodos e conjunto de base de alta similaridade têm ramos curtos (FERREIRA, 2002). Em HCA a distância entre as variáveis estudadas é calculada e transformada em uma matriz de similaridade S. A análise de agrupamento hierárquico tem como objetivo mostrar os dados de tal modo a acentuar seus agrupamentos naturais e padrões. A análise estatística necessária neste estudo deve agrupar os métodos e conjunto de base de maneira semelhante nas respectivas categorias. HCA é um método estatístico desenvolvido para este fim. A HCA foi realizada com processamento autoescalado, com distância euclidiana e método incremental.

4.5.3 Compostos estudados

Inicialmente, artemisinina e 20 derivados da artemisinina foram selecionados da literatura (Figura 11). Os compostos 2-7 foram propostos por Lin et al. (1992), que verificou que a acetilação de derivados da diidroartemisinina, contendo açúcar conduz a uma atividade semelhante ou melhor comparado com a atividade da artemisinina. No entanto, a desacetilação de açúcares reduz consideravelmente a atividade antimalárica. Os compostos 8-13 foram escolhidos para examinar o impacto da estereoespecificidade da cadeia lateral alquil na propriedade biológica, e foram propostos por Lin et al. (1987) para se obter compostos com maior atividade biológica do que os antimaláricos do ácido artelinico, artemisinina, artemeter e arteether. No entanto, a conversão de ésteres dos seus ácidos correspondentes reduz drasticamente a sua atividade antimalárica. Os compostos 14-21 foram propostos por Lin et al. (1995, 1997), onde avaliaram que grandes substituintes no carbono α -metileno (*C) aumentam substancialmente a atividade antimalárica dos compostos com base nos efeitos eletrônicos e estéricos. Pois o aumento da lipofilicidade dos derivados da artemisinina resulta em aumento da atividade antimalárica. Eles foram escolhidos pelas suas bioatividades *in vitro* resistentes a drogas contra a malária de cepas *P. falciparum* (W-2), que é resistente a cloroquina, pirimetamina, sulfadoxina e quinina, porém sensível a mefloquina.

O valor da atividade relativa (AR) dos compostos mostrados na Figura 11 foi usado para nosso estudo e definido de acordo com a relação abaixo:

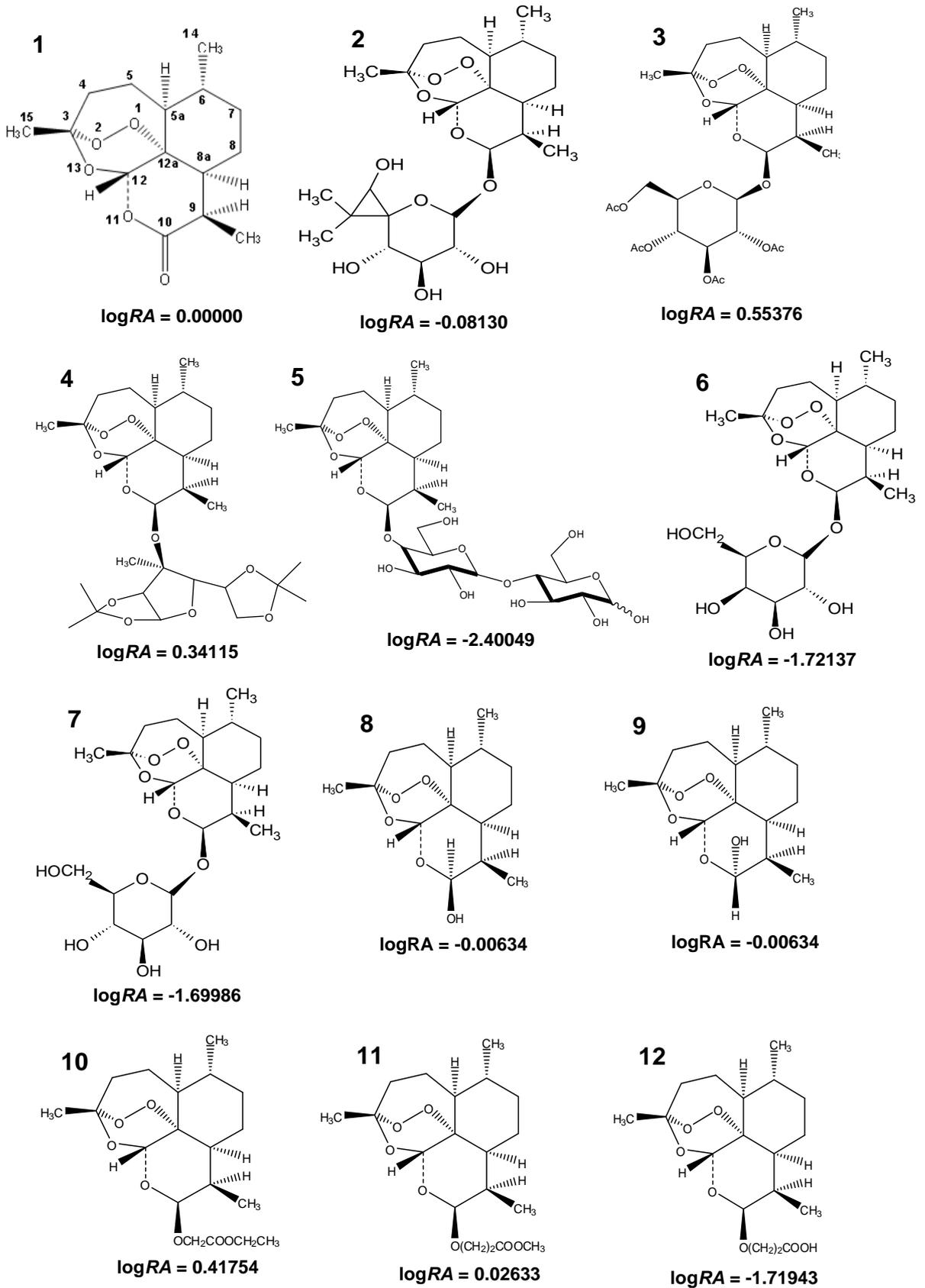
$$AR = \frac{IC_{50} \text{ artemisinina}}{IC_{50} \text{ derivado}}$$

Os compostos da Figura 11 serão chamados de compostos do conjunto de treinamento, foram construídos com o programa GaussView 5.0 (FRISCH et al., 2004), seguindo a estratégia descrita: inicialmente a estrutura da artemisinina, composto 1, Figura 11, foi construída e otimizada completamente com o método Hartree-Fock e com o conjunto de base de valência separada 6-31G** como implementado no programa Gaussian 03 (FRISCH et al., 2004). Com a obtenção da geometria mais estável da artemisinina, foram construídos os demais compostos (2-21), e posterior a isso se obteve suas estruturas mais estáveis. Para os compostos apresentados na Figura 12, compostos 22-51, que nesse estudo serão chamados compostos do conjunto teste, suas estruturas foram construídas a partir da estrutura mais estável da artemisinina, composto 1 da Figura 11. Como os dados biológicos foram obtidos a partir de diferentes fontes, o logaritmo do valor de IC_{50} da artemisinina em relação ao valor de IC_{50} dos compostos (logaritmo atividade relativa, $\log AR$) foi usada para reduzir as inconsistências causadas pelos ambientes experimentais:

$$\log AR = \log(IC_{50} \text{ da artemisinina} / IC_{50} \text{ do derivado})$$

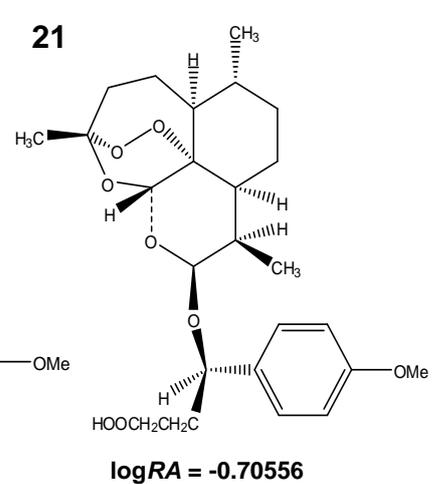
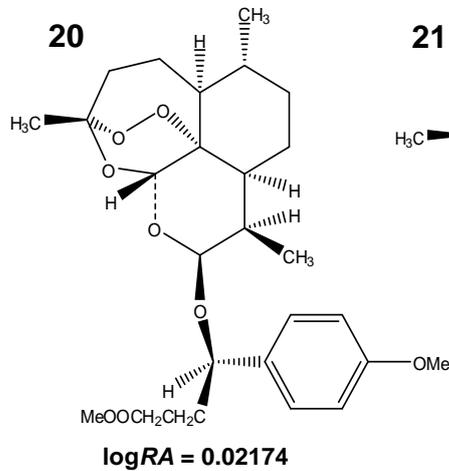
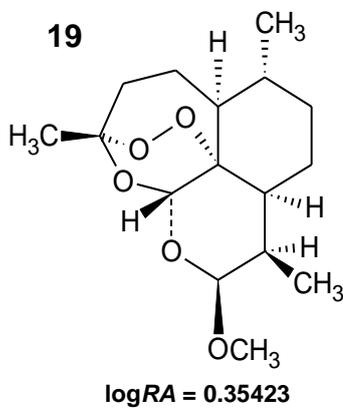
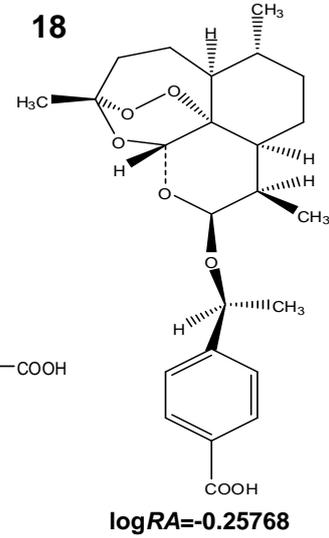
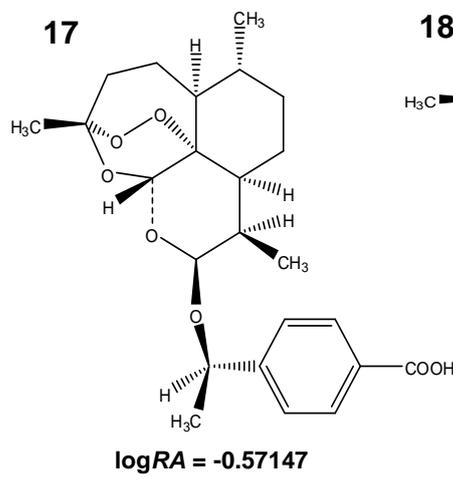
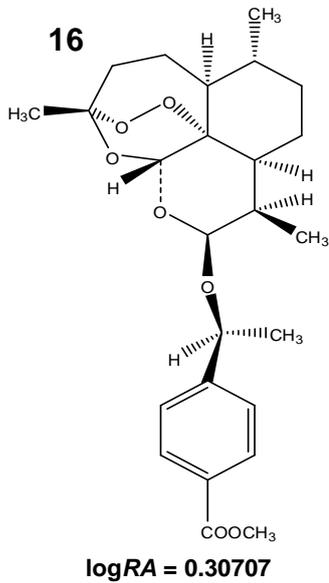
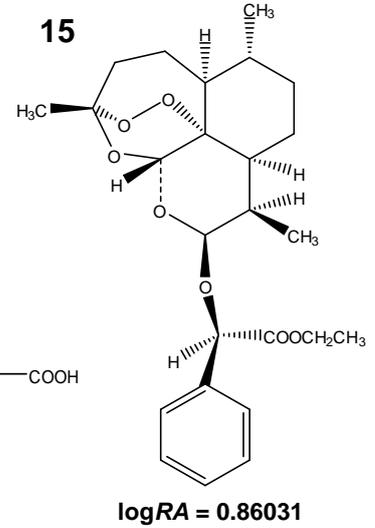
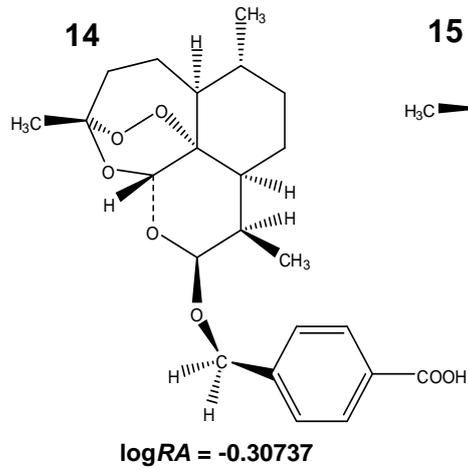
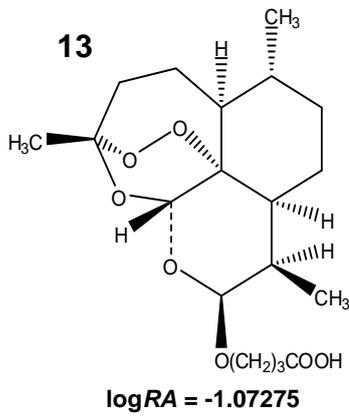
onde IC_{50} é a concentração inibitória de 50%. Neste estudo, a seguinte classificação baseada nas respostas antimaláricas foi adotada: compostos com $\log AR \geq 0,00$, variando de 0.00000 a 0.86031, foram consideradas como análogos mais potentes (1, 3, 4, 10, 11, 15, 16, 19 e 20), e aqueles com $\log AR < 0,00$, variando entre a -0,00634 a -2,40049, foram considerados análogos menos potentes (2, 5-9, 12 - 14, 17, 18 e 21). Com base nos valores relativos de atividade (AR), compostos de 3, 4, 10, 15, 16 e 19 são 2-7 vezes mais potente que a artemisinina. O composto 15 é o composto mais potente na série estudado.

Figura 11. Estrutura e atividade biológica da artemisinina e seus derivados.



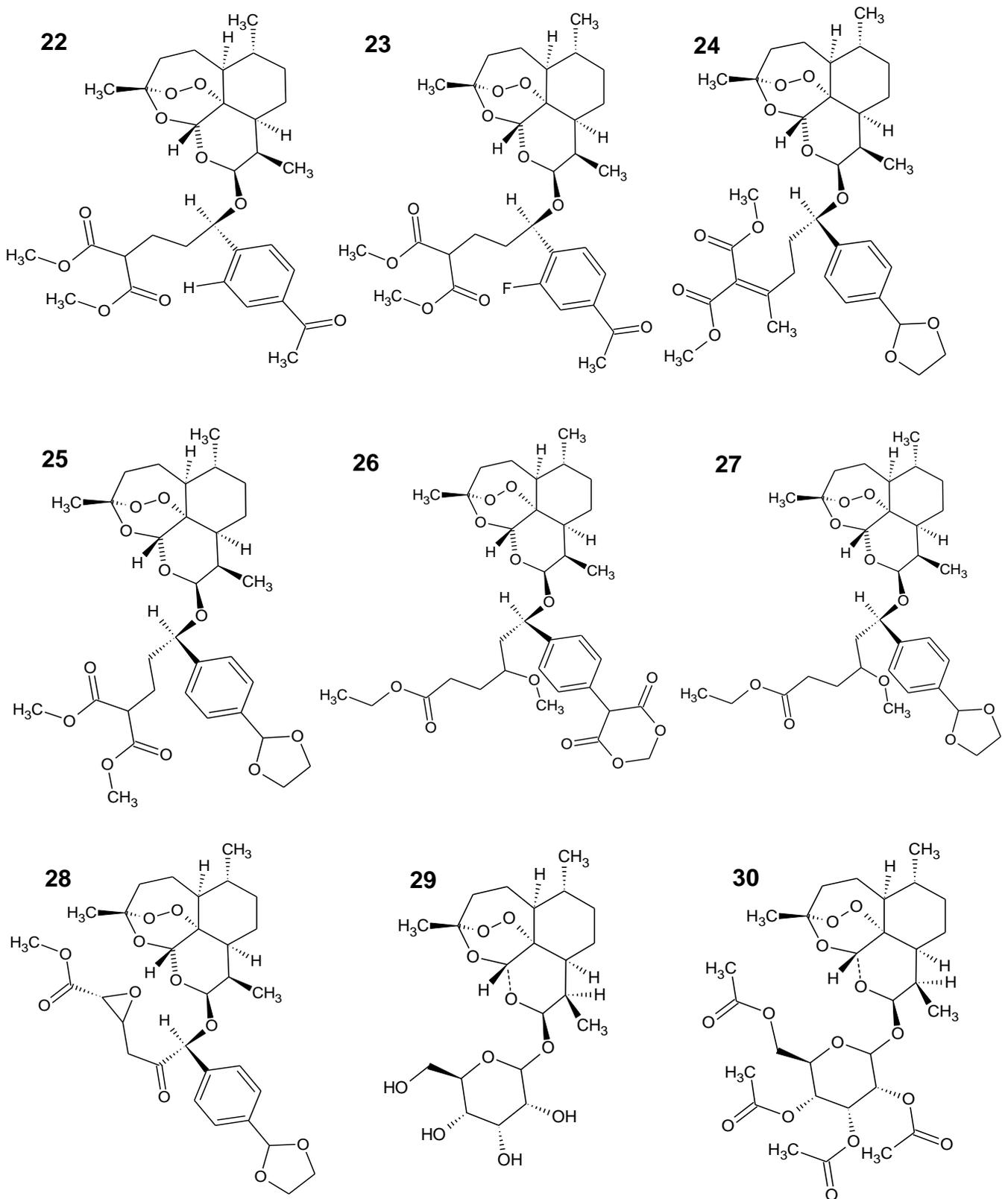
Fonte: AUTOR (2014)

Figura 11. Continuação.



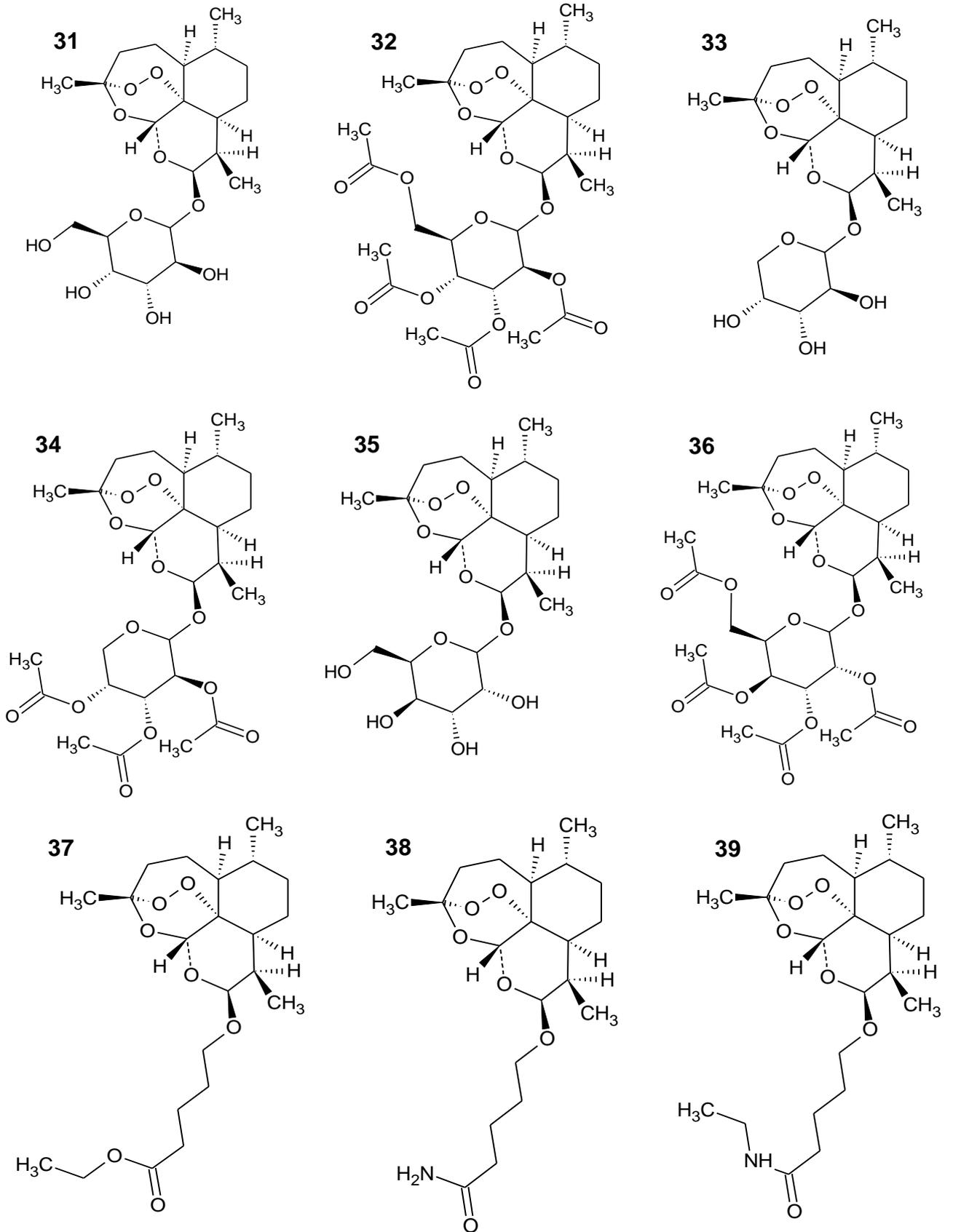
Fonte: AUTOR (2014)

Figura 12. Compostos do conjunto teste derivados da artemisinina com atividade antimalárica desconhecida contra *Plasmodium falciparum* do tipo W-2



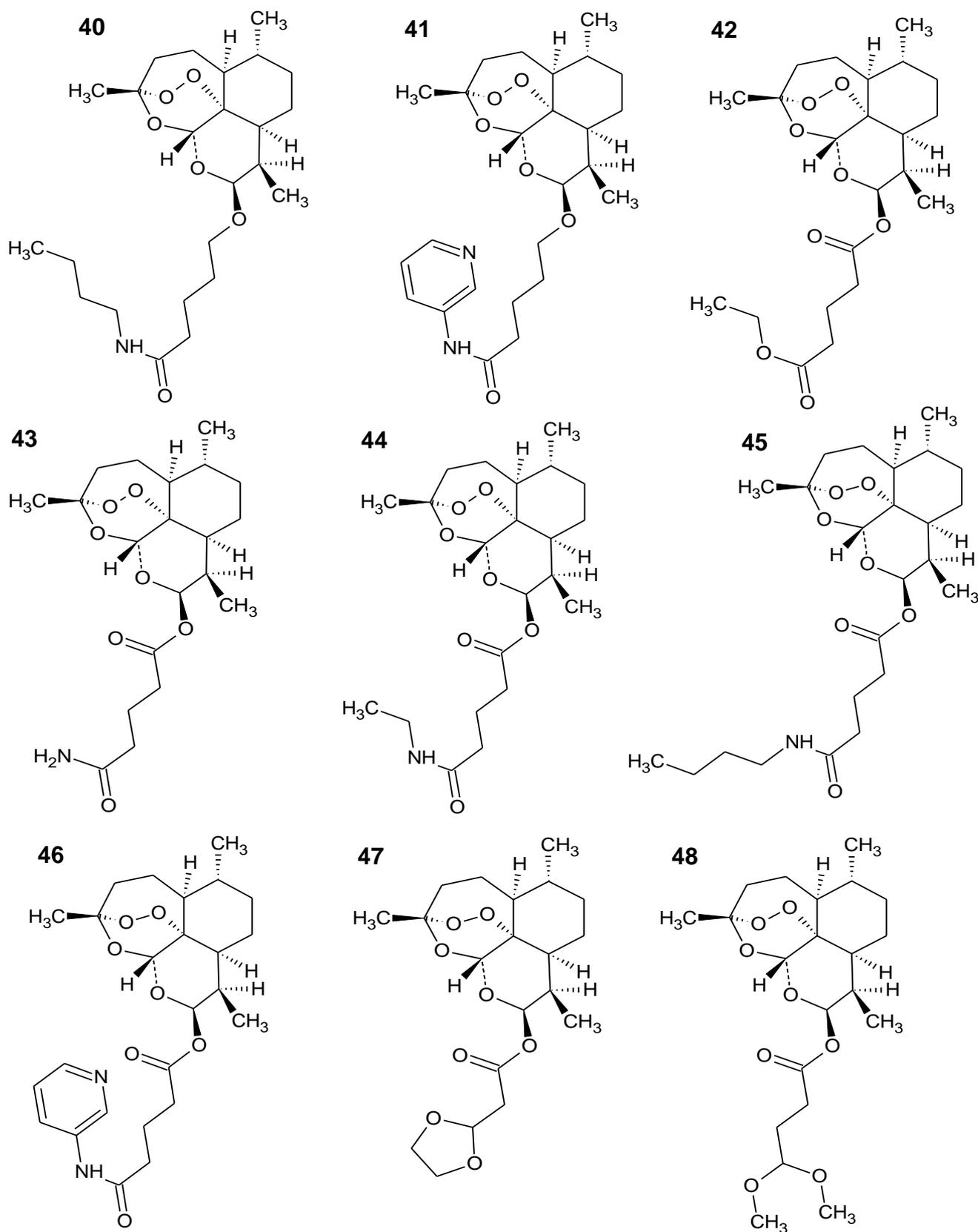
Fonte: AUTOR (2014)

Figura 12. *Continuação.*



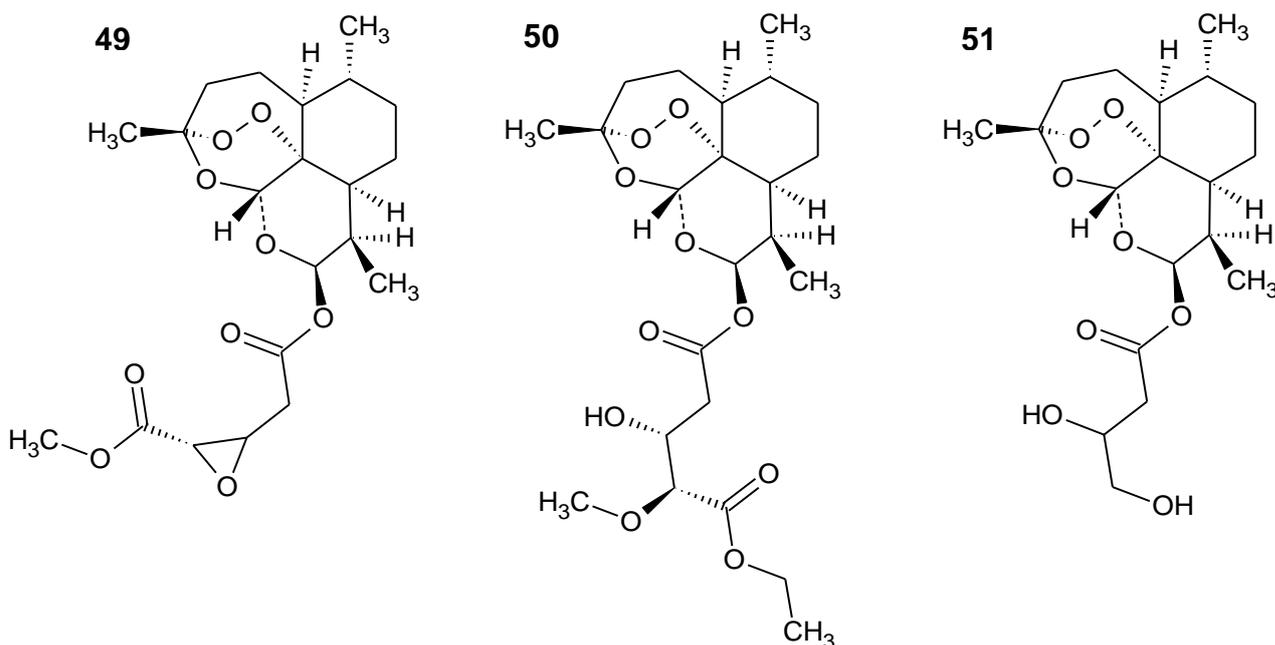
Fonte: AUTOR (2014)

Figura 12. Continuação.



Fonte: AUTOR (2014)

Figura 12. *Continuação.*



Fonte: AUTOR (2014)

4.5.4 Cálculos de obtenção dos descritores moleculares ou propriedades moleculares

Depois de determinadas às estruturas em três dimensões (3-D), foram calculados vários descritores moleculares para cada molécula do conjunto estudado. Eles representam diferentes fontes de informação química (características) sobre as moléculas e incluindo geometria, eletrônico, química quântica, descritores físico-químico e topológicos, entre outros. Eles são importantes para a descrição quantitativa da estrutura molecular e a descoberta de modelos preditivos adequados (ESTRADA; MOLINA, 2001). O cálculo dos descritores foi realizado utilizando os seguintes programas: programa Gaussian 03 (FRISCH et al., 2004), e- Dragon (VCCLAB, 2005; TETKO et al., 2005), AutoDock 4.0 (MORRIS et al., 1998; THE SCRIPPS RESEARCH INSTITUTE, 2007), Molekel (FLUKIGER, 2002) e HyperChem 6.02 (CHEMPLUS, 2000). Com o auxílio do programa de e-Dragon, 1666 descritores foram calculados e divididos nas seguintes 20 categorias: 48 descritores constitucionais; 47 descritores de quantidade e de trajetória; 47 índices de informação; 107 índices de adjacência; 21 índices de carga topológica; 41 perfis Radic moleculares; 150 descritores RDF; 154 contagens do grupo funcional; 14 descritores de carga, 33 índices de conectividade; 96 autocorrelações 2-D, 64 autovalores Burden; 44

índices baseado em autovalores; 74 descritores geométricos; 160 MORSE-3D; 120 fragmentos centrados no átomo; 31 descritores de propriedades moleculares; 119 índices topológicos, 99 descritores WHIM e 197 descritores de Getaway. Outros descritores foram obtidos tais como:

a) Descritores DOCKING MOLECULAR: Foram calculados com o objetivo de melhor representar a interação entre o fármaco-receptor com auxílio do programa AutoDock 4.0, e incluídos na matriz de dados 17 descritores do docking molecular: energia de ligação (EL); função partição (Q); energia livre de Gibbs (G); energia interna (U); energia eletrostática (EE); comprimento de ligação (O1-Fe, O2-Fe, O13-Fe, O11-Fe); ângulo de ligação (O2-O1-Fe, O1-O2-Fe, C4-O1-Fe, C5a-O1-Fe); ângulo diedro (O2-O1-Fe-N1, O2-O1-Fe-N2, O2-O1-Fe-N3, O2-O1-Fe-N4).

b) Descritores QUÍMICO-QUÂNTICOS: Em nosso estudo foram calculados os seguintes descritores químico-quânticos: energia total (E_T), energia do orbital molecular mais alto ocupado (HOMO), um nível abaixo da energia do orbital molecular mais alto ocupado (HOMO-1), energia do orbital molecular mais baixo desocupado (LUMO), um nível acima da energia do orbital molecular mais baixo desocupado (LUMO+1), diferença na energia entre HOMO e LUMO (GAP HOMO-LUMO), eletronegatividade de Mulliken (χ), dureza molecular (η), moleza molecular ($1/\eta$), carga no átomo n , onde $n=1, 2, 3, 4, 5, 5a, 6, 7, 8, 8a, 9, 10, 11, 12, 12a$ e 13, comprimento de ligação, ângulo de ligação e ângulo de torção dos átomos que constituem o anel 1, 2, 4-trioxano. As cargas atômicas utilizadas neste trabalho foram obtidas com a palavra chave POP=CHELPG através do potencial eletrostático (BRENEMAN; WINBERG, 1990). Com essa estratégia foi possível obter o melhor potencial molecular quântico de um conjunto de pontos definidos ao redor da molécula e as cargas atômicas apresentam a vantagem de em geral serem fisicamente mais satisfatórias do que as cargas Mulliken (SINGH; KOLLMAN, 1984).

C) Descritores relacionados às propriedades quantitativas de estrutura química e atividade biológica

Em nossa matriz de dados foram incluídos descritores QSAR, ou seja, a área total da superfície (TSA), volume molecular (MV), refratividade molar (MR), molar polarizabilidade (MP), coeficiente de lipofilicidade (logP), massa molecular (MM) e energia de hidratação (HE) de acordo com o programa HyperChem 6,02. Os descritores moleculares foram selecionados para fornecer informações valiosas sobre a influência das

características eletrônicas, estéricos, hidrofílicos e hidrofóbicos sobre a atividade antimalárica da artemisinina.

4.5.5 Interação entre artemisininas e heme

A interação entre os ligantes (artemisinina e seus derivados) e o receptor (heme) foi estudada com docking molecular para determinar a melhor geometria para a formação do complexo entre estas duas moléculas (ligante-receptor). A geometria da artemisinina e seus derivados (ligantes), foi realizado com HF/6-31G**, ao passo que a geometria do heme (receptor) foi obtida a partir da estrutura 1A6M no banco de dados de proteínas RCSB (PDB) de Vojtechovsky et al. (1999).

O arranjo do cálculo de docking molecular levou em conta a presença do resíduo de histidina próximo ao plano do anel de porfirina. Esta porção de histidina é usualmente coordenada perpendicularmente para Fe^{+2} através do átomo de nitrogênio sp^2 do seu anel imidazol. Tal disposição permite que o Fe^{+2} para atingir um arranjo octaédrico hexacoordenado após a ligação com a molécula de artemisinina (ARAÚJO et al., 2008). A orientação do ligante foi definida acima do plano do heme. Então, para cada interação ligante/receptor, foram calculadas 100 conformações, e a mais provável foi determinada com base nas mais baixas energias de interação. Os cálculos de docking foram realizados para desenvolver possíveis conformações para a formação do complexo empregando o Algoritmo Genético Lamarckiano implementado no pacote AutoDock 4.0. Este programa inicia o docking com a exibição do ligante em uma conformação e posição arbitrária, e procurando dockings favoráveis com o receptor usando tanto a simulação de anelamento e algoritmos genéticos. AutoDock utiliza um gerador de número aleatório para criar novas posições para o ligante durante a sua pesquisa, e estimativa da energia livre de ligação de um ligante ao seu alvo molecular. As conformações resultantes foram classificadas em ordem crescente de energia de ligação da menor conformação de energia de ligação em cada cluster.

4.5.6 Mapas de potencial eletrostático molecular (MEP)

Um importante conceito explorado no presente estudo foi a correlação da atividade da estrutura das espécies estudadas através das características do potencial eletrostático na região do anel de 1,2,13-trioxano, pois de acordo com a literatura a artemisinina e seus derivados

com atividade antimalárica apresentam padrões semelhantes em seus mapas MEP (BERNARDINELLI et al., 1994; CARDOSO et al., 2007; CARDOSO et al., 2008). Assim, o MEP permite a utilização de uma análise qualitativa para localizar sítios reativos numa molécula, e determinar os papéis desempenhados por ambos os efeitos eletrônicos e estéricos (tamanho/forma) sobre a sua potência. É de fundamental importância que a visualização do MEP forneça informações qualitativas sobre as moléculas, tais como o comportamento da interação entre um ligante e o receptor. O MEP num dado ponto (x, y, z) na vizinhança de uma molécula é definida em termos da energia de interação entre a carga elétrica gerada a partir dos elétrons das moléculas, núcleos e uma carga de teste positivo (um próton) situado em r . Para os compostos estudados, os valores $V(r)$ foram calculadas pela equação (4.5.6.1) abaixo, como descrita por Politzer e Murray (2002):

$$V(r) = \sum_A \frac{Z_A}{|R_A - r|} - \int \frac{\rho(r')}{|r' - r|} dr' \quad (4.5.6.1)$$

onde Z_A é a carga do núcleo A, localizada em R_A , $\rho(r')$ é a função da densidade eletrônica da molécula, e r' é variável de integração simulado.

Os mapas de MEP para a artemisinina e seus derivados, foram calculados a partir da carga atômica no nível HF/6-31G** utilizando o programa Gaussian 03, e os resultados são apresentados com o software Molekel.

4.5.7 Seleção de variáveis e construção dos modelos QSAR (PLS e PCR)

Após a determinação de todos os descritores moleculares, foi possível construir uma matriz de dados para desenvolver a etapa de análise multivariada. Antes de começar a análise multivariada, foi necessário fazer o autoescalamento ou padronização da matriz de dados $X=(n, m)$, constituído por vinte e um linhas (os compostos estudados) e um mil setecentos de trinta e três colunas (neste caso, os descritores calculados para cada molécula), onde n é o número de compostos estudados e m é o número de variáveis. O objetivo de usar a matriz de dados padronizada é dar a cada variável um peso igual em termos matemáticos, para cada variável foi centrada na média e escaladas para variância unitária. Para reduzir o conjunto de dados, variáveis foram selecionados com base na análise da matriz de correlação entre as variáveis (descritores ou propriedades moleculares) e o logaritmo da atividade relativa ($\log AR$). As variáveis com pouca ou nenhuma correlação (correlação abaixo de 0.30 – valor de corte) foram descartadas, com exceção dos descritores QSAR e químico quânticos,

resultando em apenas duzentos e trinta descritores restantes do conjunto inicial de 1733 descritores. Após esta compressão de dados, foram empregados dois métodos complementares para a análise exploratória de dados (PCA e HCA) para estudar as relações entre as amostras e entre as variáveis, e selecionar as propriedades que mais contribuem para a classificação dos compostos em dois grupos. Um grupo contendo os análogos mais potentes, e o outro contendo os análogos menos potentes. A PCA foi empregada para reduzir a dimensionalidade dos dados para encontrar descritores que poderiam ser úteis na caracterização do comportamento dos compostos que agem contra a malária, e analisar o agrupamento natural dos dados e amostras discrepantes. Durante a execução de PCA, foram feitas várias tentativas para obter uma boa classificação dos compostos. Em cada tentativa, os plots dos scores e loading foram baseados nas variáveis utilizadas na análise. O plot dos scores dá informações sobre os compostos (similaridade e diferenças). O plot dos loading dá informação sobre as variáveis (como elas são ligadas uns aos outros, e que melhor se descreve a variância nos dados originais). Os descritores selecionados pela PCA foram usados para realizar HCA, PLS e PCR. O objetivo da HCA foi apresentar os compostos distribuídos em grupos naturais, e os resultados confirmam os resultados da PCA. Assim, várias abordagens foram tentadas para estabelecer ligações entre as amostras/cluster. Todos eles eram de um tipo de aglomeração, pois cada amostra foi primeiro definido como o seu próprio cluster, e em seguida, outros foram agrupados para formar novos clusters até que todas as amostras fizessem parte de um único cluster.

Os modelos QSAR para os compostos da artemisinina estudados foram construídos pelos métodos PLS e PCR, com base nos dados autoescalados e o procedimento validação crossvalidation leave-one-out (FERREIRA, 2002; GELADI, 1988). O objetivo final da análise multivariada (PLS e PCR) foi à construção de um modelo matemático que pode ser utilizado para prever a atividade antimalárica. As amostras selecionadas para compor o conjunto de validação externa foram 2, 12 e 13. Os parâmetros estatísticos utilizados para avaliar a qualidade dos modelos foram o erro da soma dos quadrados do resíduo de previsão (PRESS), eq. (4.5.7.1), o erro padrão de validação (SEV), eq. (4.5.7.2), variância total explicativa, R^2 (correlação entre os valores estimados previstos pelo modelo construído com o conjunto completo de dados reais e os valores de y), Q^2 (coeficiente de correlação cruzada validado) e S_{PRESS} (desvio padrão de validação cruzada) dada pelas equações (4.5.7.3), (4.5.7.4) e (4.5.7.5), respectivamente (GELADI, 1988; NETO, 1996; CUSTÓDIO et al., 1997; PIMENTEL; GAUDIO; ZANDONADE, 2001; FERREIRA, 2002).

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4.5.7.1)$$

$$SEV = \sqrt{\frac{PRESS}{n}} \quad (4.5.7.2)$$

$$R^2 = 1 - \left[\frac{PRESS_{cal}}{\sum_{i=1}^n (y_i - \bar{y})^2} \right] \quad (4.5.7.3)$$

$$Q^2 = 1 - \left[\frac{PRESS_{val}}{\sum_{i=1}^n (y_i - \bar{y})^2} \right] \quad (4.5.7.4)$$

$$S_{PRESS} = \frac{\sqrt{PRESS}}{n - k - 1} \quad (4.5.7.5)$$

Nas equações (4.5.7.1) e (4.5.7.2), n é o número de compostos usados para o modelo de calibração ou de validação, y_i é o valor experimental da propriedade físico-química para a amostra e \hat{y}_i é o valor previsto pelo modelo de calibração ou validação. Nas equações (4.5.7.3) e (4.5.7.4), $PRESS_{cal}$ é a calibração de Previsão do erro da soma dos quadrados, e $PRESS_{val}$ é a validação de Previsão do Erro da soma dos quadrados. Em ambos, $PRESS_{cal}$ e $PRESS_{val}$ são avaliados a partir da equação (4.5.7.1) alterando \hat{y}_i para a calibração ou modelo de validação. Os valores da variância explicativa (R^2_A , isto é, $R^2_{ajustado}$), o desvio padrão (s) e F (teste de Fisher) foram determinados. As análises de dados multivariados (PCA, HCA, PLS e PCR) foram realizadas empregando o software Pirouette 3.01 (INFOMETRIX, 2001).

5 RESULTADOS E DISCUSSÃO

5.1 OTIMIZAÇÃO DA GEOMETRIA DA ARTEMISININA EM DIFERENTES MÉTODOS E CONJUNTOS DE BASE PARA VALIDAR O MÉTODO COMPUTACIONAL APLICADO NA MODELAGEM MOLECULAR DOS COMPOSTOS ESTUDADOS

Em todos os três conjuntos de bases (HF/6-31G, HF/6-31G*, HF/6-31G**), o método Hartree-Fock descreve todos os parâmetros estruturais muito bem em termos de magnitude e sinal quando comparado aos valores experimentais (ver Tabela 6). Isto está em contraste com os métodos AM1, PM3, ZINDO e DFT (B3LYP/3-21G, B3LYP/3-21G*, B3LYP/3-21G**), em que não há uma boa concordância entre os valores experimentais e teóricos para os ângulos de torção, especialmente o ângulo formado pelos átomos C3O13C12C12a, com desvios $<-13.900^\circ$ (AM1), $<-22.489^\circ$ (PM3), $<-7.880^\circ$ (ZINDO), $>0.020^\circ$ (HF/6-31G), $>2.132^\circ$ (HF/6-31G*), $>2.100^\circ$ (HF/6-31G**), $>-3.759^\circ$ (B3LYP/3-21G), $>-3.760^\circ$ (B3LYP/3-21G*) e $>-3.780^\circ$ (B3LYP/3-21G**) e desvios padrão de 4.776, 8.388, 4.372, 1.663, 2.484, 1.762, 1.915, 1.855 e 1.987, respectivamente. Ao comparar estes métodos com o método HF, encontramos que os conjuntos de bases HF/6-31G e HF/6-31G** têm baixos desvios padrão em relação aos métodos semiempíricos e DFT. A variação foi de ± 0.099 entre eles (HF/6-31G e HF/6-31G**).

A Tabela 7 mostra que os conjuntos de base HF/6-31G, HF/6-31G*, HF/6-31G** mostram excelentes resultados para comprimento de ligação em comparação com os dados experimentais. A base 6-31G descreve bem os ângulos de ligação, com valores próximos aos resultados experimentais. No entanto, as bases mínimas (6-31G e 3-21G) têm várias deficiências; assim, uma função de polarização foi incluída para melhorar estas bases (isto é, orbitais *p* representado por *). Esses orbitais seguem funções restritas que estão centradas em núcleos. No entanto, os orbitais atômicos se distorcem ou polarizam uma molécula quando é formada. Portanto, deve-se considerar a possibilidade de deslocamento não-uniforme de cargas elétricas fora do núcleo atômico, ou seja, a polarização. Assim, é possível obter uma melhor descrição das cargas e deformações de orbitais atômicas dentro de uma molécula.

Um modo de polarização pode ser considerado através da introdução de funções para as quais os valores de l (número quântico do momento angular orbital) são maiores do que as do estado fundamental de um dado átomo. Para estes tipos, os nomes dos conjuntos de base denotam as funções de polarização. Assim, 6-31G*, refere-se a um conjunto de base 6-31G com uma função de polarização para os átomos pesados (isto é, que não sejam átomos de hidrogênio), e 6-31G** refere-se à inclusão de uma função de polarização para os átomos de hidrogênio e hélio (LEACH, 2001).

Quando conjuntos de base com funções de polarização são usados em cálculos que envolvendo íons ânions, bons resultados não são obtidos devido à nuvem eletrônica de sistemas aniônicos que tendem a expandir-se. Assim, as funções difusas apropriadas devem ser incluídas, pois elas permitem uma maior ocupação no orbital em uma determinada região do espaço. As funções difusas são importantes nos cálculos de metais de transição porque os átomos de metal têm orbitais " d ", que tendem a ser difuso. Então, torna-se necessário incluir funções difusas na função de base associada com a configuração de um átomo de metal neutro para se obter uma melhor descrição do complexo com metal. A base 6-31G** é particularmente útil no caso das ligações de hidrogênios (LEVINE, 1991; LEACH, 2001; HEHRE, 2003; MULLIKEN; LIU, 2009; SANTOS et al., 2013).

Este estudo destacou o conjunto de base HF/6-31G**, que é o mais próximo dos resultados experimentais, e mostrou bom desempenho na descrição quando se comparou com os ângulos de ligação C3O13C12 e C12aO1O2. Os ângulos de torção ou o ângulo diedro mostraram também uma boa concordância com os valores experimentais relatados na literatura. Os ângulos de torção O1O2C3O13 e C13C12C12aO1 também estão próximos dos dados cristalográficos.

Derivados da artemisinina com atividade antimalárica contra o *Plasmodium falciparum*, que é resistente à mefloquina, já foram estudados usando métodos químicos quânticos (HF/6-31G*) e o método dos mínimos quadrados parciais (PLS), onde três componentes principais explicou 89.55 % da variância total, com $Q^2 = 0.83$ e $R^2 = 0.92$. A partir de um conjunto de 10 derivados da artemisinina propostos (derivados de artemisinina com atividade antimalárica desconhecido contra *Plasmodium falciparum*), e como resultado um novo composto foi proposto com atividade antimalárica superior em

comparação com os compostos anteriormente descritos na literatura (PINHEIRO et al., 2003).

Cardoso et al. (2007) usaram o conjunto de base HF/3-21G** e o método PLS para desenvolver novos derivados da artemisinina com atividade antimalárica contra o *P. falciparum*. O método PLS foi usado para construir um modelo de regressão multivariado, o que levou a propor a atividade antimalárica de novos derivados da artemisinina com atividade desconhecida. Além disso, os mapas MEP para os compostos estudados e propostos foram construídos e avaliados para identificar características chaves nas moléculas ativas.

Cardoso et al. (2008) estudaram artemisinina e seus derivados com atividade contra cepas D-6 de *Plasmodium falciparum*, usando o método HF/3-21G para verificar a confiabilidade da geometria obtida, comparando os parâmetros estruturais do anel trioxano da artemisinina com os valores teóricos e experimentais da literatura.

Ferreira et al. (2010) estudaram a artemisinina e 18 derivados com atividade antimalárica contra cepas de W-2 de *Plasmodium falciparum* através de química quântica e análise multivariada. A otimização da geometria das estruturas foi realizada utilizando o método de Hartree-Fock e conjunto de base 3-21G**.

Tabela 7. Parâmetros teóricos e experimentais do anel 1, 2, 13-trioxano da artemisinina

Parâmetros ^a	Semi-empírico			Hartree-fock/HF			DFT/B3LYP			Experimental ^[c]
	AM1	PM3	ZINDO	6-31G	6-31G*	6-31G** ^b	3-21G	3-21G*	3-21G**	
Comprimento de ligação (Å)										
O1O2	1.288	1.544	1.237	1.447	1.391	1.390	1.524	1.524	1.524	1.469
O2C3	1.447	1.403	1.400	1.435	1.393	1.396	1.455	1.455	1.454	1.416
C3O13	1.427	1.428	1.396	1.435	1.388	1.408	1.473	1.473	1.472	1.445
O13C12	1.416	1.403	1.392	1.403	1.400	1.376	1.430	1.430	1.430	1.379
C12C12a	1.537	1.555	1.513	1.533	1.533	1.532	1.535	1.535	1.535	1.523
C12aO1	1.468	1.426	1.416	1.469	1.429	1.429	1.504	1.504	1.504	1.461
Ângulo de ligação (°)										
O1O2C3	112.530	110.340	114.310	108.800	106.100	109.460	105.590	105.590	105.480	108.100
O2C3O13	103.600	104.810	105.370	106.760	110.800	107.800	108.220	108.220	108.250	106.600
C3O13C12	115.480	116.010	115.843	117.300	112.800	115.300	113.200	113.200	113.200	114.200
O13C12C12a	113.510	115.200	113.270	112.280	108.700	112.300	113.300	113.300	113.230	114.500
C12C12aO1	111.070	113.180	107.290	110.910	110.500	110.545	112.410	112.410	112.470	110.700
C12aO1O2	113.740	112.290	118.380	113.240	112.700	112.700	109.620	109.620	109.590	111.200
Ângulo de torção (°)										
O1O2C3O13	-77.800	-73.310	-70.403	-71.840	-73.369	-73.400	-76.610	-76.610	-76.740	-75.500
O2C3O13C12	42.070	52.700	36.370	33.390	31.034	31.100	33.750	33.750	33.720	36.000
C3O13C12C12a	11.400	2.811	17.420	25.320	27.432	27.400	29.059	29.060	29.080	25.300
O13C12C12aO1	-41.770	-40.510	-46.610	-49.410	-50.100	-50.143	-52.190	-52.190	-52.030	-51.300
C12C12aO1O2	12.050	19.940	18.110	12.510	10.900	10.924	9.060	9.600	9.340	12.700
C12aO1O2C3	47.050	35.630	40.130	46.700	48.700	48.674	51.060	51.060	51.320	47.800
Desvio padrão	4.776	8.388	4.372	1.663	2.484	1.762	1.915	1.855	1.987	-

^a Os átomos são numerados de acordo com a Figura 11.

^b Conjunto de base de valência validado para o cálculo de obtenção das propriedades moleculares.

^c J. N. Lisgarten, B. S. Potter, C. Bantuzeko and R. A. Palmer, "Structure, Absolute Configuration, and Conformation of the Antimalarial Compound, Artemisinin," *Journal of Chemical Crystallography*, Vol. 28, No. 7, 1998, pp. 539-543. <http://dx.doi.org/10.1023/A:1023244122450>

5.1.1 Aplicação do método PCA para selecionar os principais parâmetros geométricos da artemisinina

Os parâmetros teóricos e experimentais do anel 1, 2, 13-trioxano da artemisinina foram utilizados com o objetivo de identificar, através de análises de PCA e HCA, que otimização da geometria da artemisinina em diferentes métodos e conjuntos de bases define os resultados mais próximos aos dados experimentais. A vantagem em utilizar os métodos PCA e HCA neste estudo foi que todos os parâmetros estruturais são considerados simultaneamente, e que leva em conta as correlações entre eles. Os resultados de PCA mostraram que quatro descritores mais importantes relacionados ao anel trioxano foram O13C12, O1O2C3, C3O13C12C12a, C12C12aO1O2.

Os valores dos parâmetros geométricos selecionados pela análise de componentes principais, desvio padrão e matriz de correlação de Pearson são mostrados na Tabela 6. Esses parâmetros geométricos são responsáveis pela separação dos métodos e conjuntos de bases em três classes: métodos semiempíricos e métodos *ab initio*/Hartree-Fock e DFT. Os quatro parâmetros geométricos relacionados ao anel trioxano são identificados pelos átomos com as medidas O13C12 que é a distância interplanar entre estes dois átomos (comprimento de ligação), O1O2C3 é o ângulo de ligação entre esses três átomos, C3O13C12C12a e C12C12aO1O2 estão relacionados com ângulo diedro ou de torção.

A Tabela 8 mostra uma matriz de correlação entre os parâmetros geométricos e o desvio padrão, observa-se que a correlação entre os parâmetros geométricos é menor ou igual a 0.827, enquanto que a correlação entre os parâmetros geométricos e o desvio padrão é menor ou igual a 0.914. Os parâmetros geométricos selecionados com a técnica PCA representam as características necessárias para validar a escolha do melhor método e conjunto de base

Tabela 8. Parâmetros geométricos selecionados pela análise de componentes principais, desvio padrão e matriz de correlação de Pearson

MÉTODOS	O13C12	O1O2C3	C3O13C12C12a	C12C12aO1O2	Desvio Padrão
AM1	1.416	112.530	11.400	12.050	4.776
PM3	1.403	110.340	2.811	19.940	8.388
ZINDO	1.392	114.310	17.420	18.110	4.372
6-31G	1.403	108.800	25.320	12.510	1.663
6-31G*	1.400	106.100	27.432	10.900	2.484
6-31G**	1.376	109.460	27.400	10.924	1.762
3-21G	1.430	105.589	29.059	9.060	1.915
3-21G*	1.430	105.590	29.060	9.600	1.855
3-21G**	1.430	105.480	29.080	9.340	1.987
Experimental	1.379	108.100	25.300	12.700	0.000
O13C12		-0.465	0.140	-0.418	0.068
O1O2C3			-0.700	0.727	0.540
C3O13C12C12a				-0.827	-0.914
C12C12aO1O2					0.758

Os resultados do modelo de seleção são apresentados na Tabela 9, e mostram que o modelo foi construído com três componentes principais (3PCs), onde a primeira componente principal (PC1) descreve 24.3990% da informação total, a segunda componente principal (PC2) descreve 8.0043% e a terceira (PC3) 2.5477%. Ainda nesta tabela, observa-se que PC1 contém 67.7751% da informação original, as duas primeiras (PC1+PC2) 90.0093% e os três primeiros (PC1+PC2+PC3) conseguem explicar 97.0861% da informação total, perdendo apenas 2.9139% da informação original. Na mesma tabela, verifica-se que os parâmetros geométricos O1O2C3 e C12C12aO1O2 são os que mais contribuem em PC1, enquanto que os parâmetros geométricos O13C12 e C12C12aO1O2 são os que mais contribuem em PC2.

Tabela 9. Análise dos componentes principais da seleção do modelo dos métodos e conjunto de base

	Componente Principal		
	PC1	PC2	PC3
Variância (%)	24.3990	8.0043	2.5477
Variância Acumulada (%)	67.7751	90.0093	97.0861
Parâmetros geométricos	Contribuição		
	PC1	PC2	
O13C12	-0.3296	0.8771	
O1O2C3	0.5447	-0.0404	
C3O13C12C12a	-0.5239	-0.4628	
C12C12aO1O2	0.5657	0.1213	

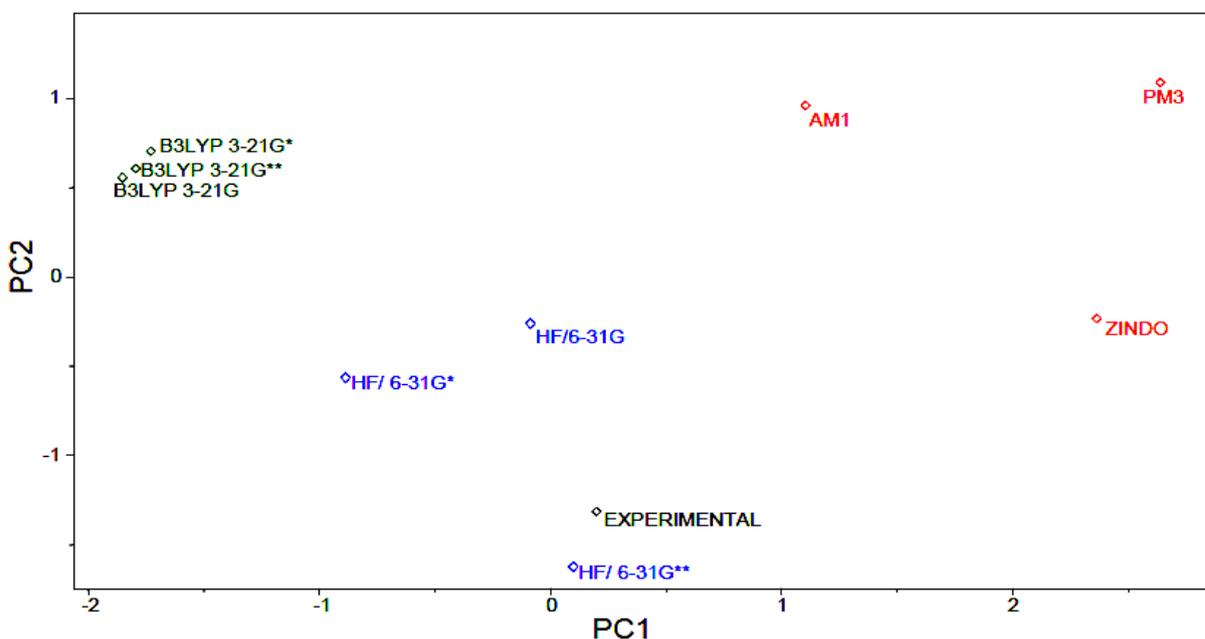
Como as componentes principais podem ser escritas como combinação linear dos parâmetros geométricos selecionados, as expressões matemáticas para PC1 e PC2 são mostradas abaixo:

$$PC1 = -0.3296(O13C12) + 0.5447(O1O2C3) - 0.5239(C3O13C12C12a) + 0.5657(C12C12aO1O2) \quad (1)$$

$$PC2 = 0.8771(O13C12) - 0.0404(O1O2C3) - 0.4628(C3O13C12C12a) + 0.1213(C12C12aO1O2) \quad (2)$$

A Figura 13 mostra os escores para cinco métodos e seis conjuntos de base, relacionando com o parâmetro geométrico experimental do anel 1,2,13-trioxano. De acordo com o escore observa-se que os métodos são discriminados em três grupos de acordo com PC1. O método semi-empírico (AM1, PM3 e ZINDO) estão localizados no lado direito, enquanto que o método *ab initio*/Hartree-Fock e o experimental estão localizados na parte central da Figura 13. Já o método DFT/B3LYP estão localizado na parte superior esquerda. Além disso, podemos ver que o método *ab initio* HF/6-31G, HF/6-31G* e HF/6-31G** são os mais próximos do parâmetro geométrico experimental, indicando que qualquer um deles pode ser utilizado no desenvolvimento de cálculos de propriedades moleculares.

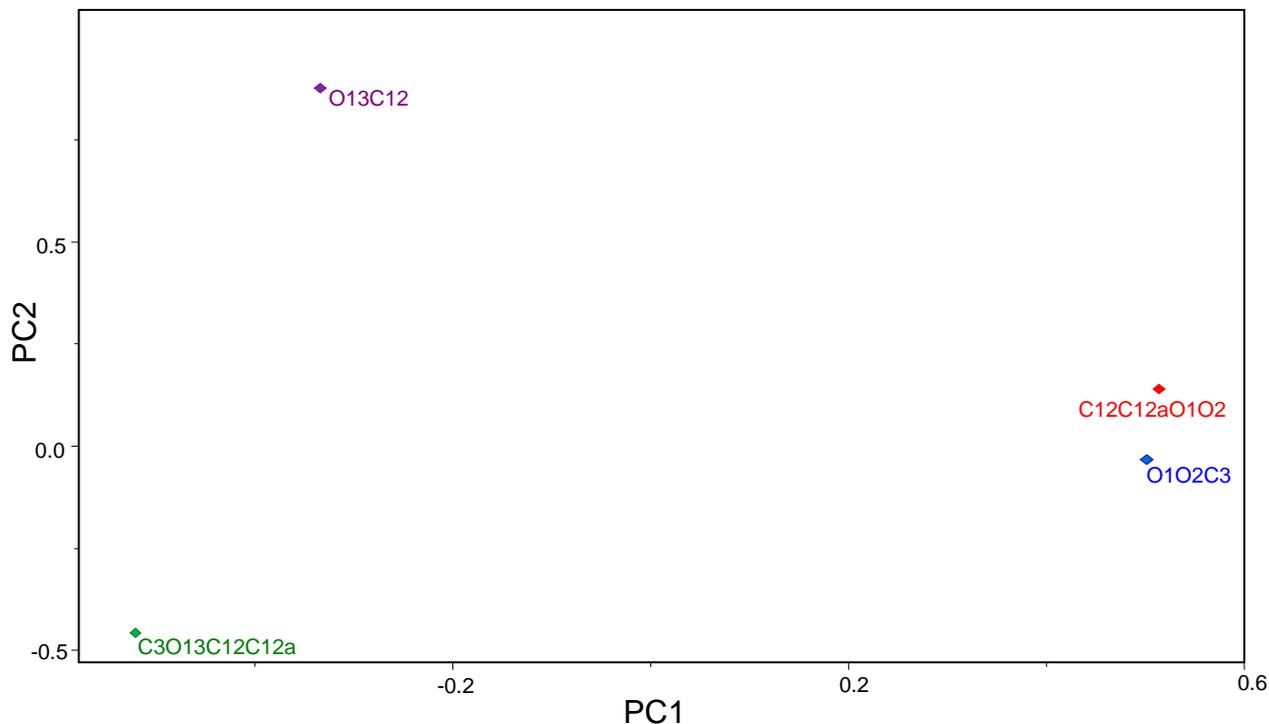
Figura 13. Plot dos scores PC1-PC2 para os cinco métodos e seis conjuntos de base, correlacionados com os parâmetros geométricos do anel 1,2,13-trioxano da artemisinina.



Fonte: AUTOR (2014)

A Figura 14 mostra os loadings para os quatro descritores mais importantes na classificação dos métodos e conjunto de base, observa-se no escore que os parâmetros geométricos em relação aos métodos e conjunto de base DFT (B3LYP/3-21G, B3LYP/3-21G*, B3LYP/3-21G**) e Semi-empíricos (AM1 e PM3), têm principal contribuição dos parâmetros geométricos O13C12 e C12C12aO1O2, e estes são responsáveis em deslocar os métodos e conjunto de base para a parte superior do escore. Enquanto que o método *ab initio* (HF/6-31G, HF/6-31G* e HF/6-31G**) e o Semi-empirico (ZINDO) têm alta contribuição dos parâmetros geométricos O1O2C3, C3O13C12C12a, sendo estes responsáveis em deslocar os métodos e conjunto de base para a parte inferior do escore. Ainda na Figura 14, nota-se que quanto maior a contribuição dos parâmetros geométricos O13C12 e C12C12aO1O2 na segunda componente principal maior será o valor do escore, e assim menos eficiente será o método e conjunto de base, pois se distância do experimental. Os parâmetros geométricos O1O2C3 e C3O13C12C12a contribuem em menor grau, por terem peso negativo em PC2, demonstrando que os métodos e conjunto de base, em geral têm maiores valores destes parâmetros geométricos.

Figura 14. Plot dos loading de PC1-PC2 usando quatro parâmetros geométricos selecionados pela análise de componentes principais para cinco métodos os conjuntos de base cinco e seis, correlacionados com os parâmetros geométricos do anel 1,2,13-trioxano da artemisinina.



Fonte: AUTOR (2014)

Os parâmetros geométricos O13C12, O1O2C3, C3O13C12C12a e C12C12aO1O2 são de grande importância em nosso estudo, pois de acordo com a proposta feita por Jefford e colaboradores o ferro do heme ataca a artemisinina na posição O1 e gera um radical livre na posição O2, após a ligação C3-C4 ser quebrada gerando um carbono radicalar em C4 (JEFFORD et al., 2001). Este radical livre em C4 tem sido sugerido como uma importante substância na atividade antimalárica (POSNER et al., 1994). Estudo de docking molecular entre a artemisinina e o seu receptor, o grupo heme, feitos por Tonmunphean, Parasuk e Kokpol em 2000, também indicaram que o ferro do grupo heme interage com O1 mais preferencialmente que O2 (TONMUNPHEAN et al., 1998; TONMUNPHEAN et al., 2001), o que configura a importância do parâmetro geométrico O1O2C3, selecionado neste modelo por estar associado ao mecanismo sugerido.

Na Tabela 9 mostra os parâmetros geométricos selecionados pela análise de componentes principais, método e conjunto de base, variação dos parâmetros geométricos em relação aos dados experimentais (Δ e $\Delta\%$). Nos métodos semi-empírico e DFT não

existem uma boa concordância entre os valores teóricos e experimentais para os ângulos de torção, especialmente os ângulos formado pelos átomos C3O13C12C12a e C12C12aO1O2.

No método semi-empírico AM1, PM3 e ZINDO apresentam desvios $\Delta = -13.900^\circ$ ($\Delta\% = -54.940$), $\Delta = -22.489^\circ$ ($\Delta\% = -88.889$) e $\Delta = -7.879^\circ$ ($\Delta\% = -31.146$) em relação ao ângulo C3O13C12C12a, respectivamente. Para o ângulo C12C12aO1O2 apresenta desvios $\Delta = -0.650^\circ$ ($\Delta\% = -5.118$), $\Delta = 7.240^\circ$ ($\Delta\% = 57.007$) e $\Delta = 5.410^\circ$ ($\Delta\% = 42.598$). Em relação ao DFT (3-21G, 3-21G* e 3-21G**) o ângulo formado pelos átomos C3O13C12C12a, têm desvios $\Delta = 3.759^\circ$ ($\Delta\% = 14.857$), $\Delta = 3.760^\circ$ ($\Delta\% = 14.861$) e $\Delta = 3.780^\circ$ ($\Delta\% = 14.940$), respectivamente. Para o ângulo C12C12aO1O2 apresenta desvios $\Delta = -3.640^\circ$ ($\Delta\% = -28.660$), $\Delta = -3.100^\circ$ ($\Delta\% = -24.409$) e $\Delta = -5.410^\circ$ ($\Delta\% = 42.598$). Já no método *ab initio*/Hartree-Fock é visto que os três níveis (HF/6-31G, HF/6-31G* e HF/6-31G**) mostram resultados excelentes para o comprimento de ligação O13C12, com desvios $\Delta = 0.024\text{\AA}$ ($\Delta\% = 1.740$) para HF/6-31G, $\Delta = 0.021\text{\AA}$ ($\Delta\% = 1.522$) para HF/6-31G* e $\Delta = -0.003\text{\AA}$ ($\Delta\% = -0.217$) para HF/6-31G**, conforme mostrado na Tabela 8, página 96.

Ainda na Tabela 10, destaca-se a comparação do método *ab initio* com os conjuntos de base HF/6-31G e HF/6-31G**, onde o conjunto de base de valência HF/6-31G obteve resultados satisfatórios, porém a base HF/6-31G** apresentou excelentes resultados em magnitude e sinal com relação ao comprimento de ligação O13C12 e o ângulo de ligação O1O2C3. Assim como, os ângulos de torções ou ângulo diedro apresentaram uma boa concordância com os valores experimentais reportados da literatura, mostrando que a base HF/6-31G** neste parâmetro estão próximos dos dados cristalográficos da artemisinina na região do anel endoperóxido.

Tabela 10. Parâmetros geométricos selecionados pela análise de componentes principais, métodos e conjunto de base, e a variação dos parâmetros geométricos com os respectivos dados experimentais (Δ e $\Delta\%$).

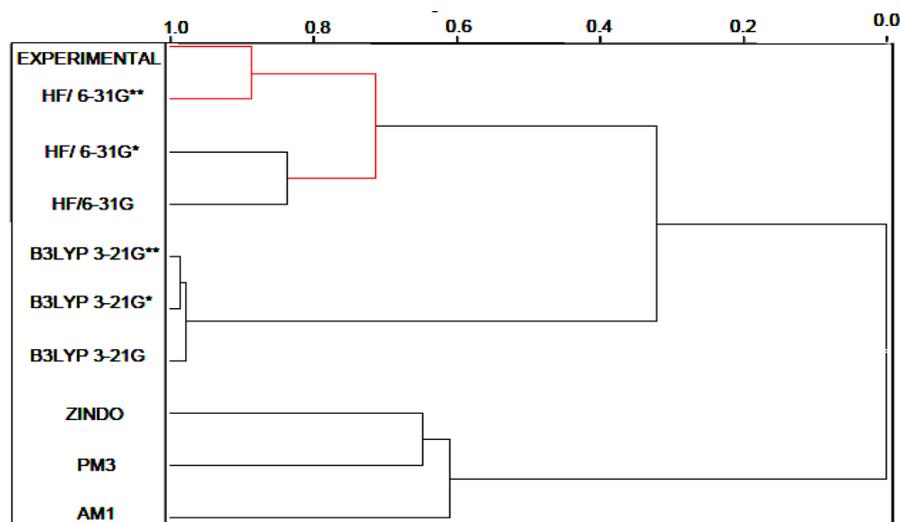
Parâmetros ^a	Semi-empírico									Experimental
	AM1	Δ_{AM1}	$\Delta(\%)$	PM3	Δ_{PM3}	$\Delta(\%)$	ZINDO	Δ_{ZINDO}	$\Delta(\%)$	
Comprimento de ligação (Å)										
O13C12	1.416	0.037	2.683	1.403	0.024	1.740	1.392	0.013	0.942	1.379
Ângulo de ligação (°)										
O1O2C3	112.530	4.430	4.098	110.340	2.240	2.072	114.310	6.210	5.744	108.100
Ângulo de torção (°)										
C3O13C12C12a	11.400	-13.900	-54.940	2.811	-22.489	-88.889	17.420	-7.880	-31.146	25.300
C12C12aO1O2	12.050	-0.650	-5.118	19.940	7.240	57.007	18.110	5.410	42.598	12.700
Parâmetros ^a	<i>ab initio</i> /Hartree-fock/HF									Experimental
	6-31G	Δ_{6-31G}	$\Delta(\%)$	6-31G*	Δ_{6-31G^*}	$\Delta(\%)$	6-31G**	$\Delta_{6-31G^{**}}$	$\Delta(\%)$	
Comprimento de ligação (Å)										
O13C12	1.403	0.024	1.740	1.400	0.021	1.522	1.376	-0.003	-0.217	1.379
Ângulo de ligação (°)										
O1O2C3	108.800	0.700	0.647	106.100	-2.000	-1.850	109.460	1.360	1.258	108.100
Ângulo de torção (°)										
C3O13C12C12a	25.320	0.020	0.079	27.432	2.132	8.426	27.400	2.100	8.300	25.300
C12C12aO1O2	12.510	-0.190	-1.496	10.900	-1.800	-14.173	10.924	-1.776	-13.984	12.700
Parâmetros ^a	DFT/B3LYP									Experimental
	3-21G	Δ_{3-21G}	$\Delta(\%)$	3-21G*	Δ_{3-21G^*}	$\Delta(\%)$	3-21G**	$\Delta_{3-21G^{**}}$	$\Delta(\%)$	
Comprimento de ligação (Å)										
O13C12	1.430	0.051	3.698	1.430	0.051	3.698	1.430	0.051	3.698	1.379
Ângulo de ligação (°)										
O1O2C3	105.589	-2.511	-2.322	105.590	-2.510	-2.321	105.480	-2.620	-2.423	108.100
Ângulo de torção (°)										
C3O13C12C12a	29.059	3.759	14.857	29.060	3.760	14.861	29.080	3.780	14.940	25.300
C12C12aO1O2	9.060	-3.640	-28.66	9.600	-3.100	-24.409	9.340	-3.360	-26.456	12.700

Δ = Teórico – Experimental; $\Delta(\%) = \Delta \times 100 / \text{Experimental}$

5.1.2 Aplicação do método HCA

A HCA foi usada para validar o método e conjunto de base mais apropriado para os futuros cálculos de propriedades moleculares com maior precisão nos seus resultados. A análise foi obtida na forma de um dendograma conforme mostra a Figura 15. Observam-se nesta Figura os cinco métodos: Semi-empíricos (AM1, PM3 e ZINDO), DFT (B3LYP/3-21G, B3LYP/3-21G*, B3LYP/3-21G**) e *ab initio*/Hartree-Fock (HF/6-31G, HF/6-31G* e HF/6-31G**). Pode-se notar que os métodos AM1, PM3 e ZINDO, apresentaram ramos longos caracterizando baixa similaridade com o experimental. Entretanto, o método DFT (B3LYP/3-21G, B3LYP/3-21G*, B3LYP/3-21G**), apresentou alta similaridade entre eles – conjunto de base, pelo fato de terem ramos curtos. Já o método *ab initio* (HF/6-31G, HF/6-31G* e HF/6-31G**) apresentou alta similaridade entre eles, e independente dos outros métodos e conjunto de base o HF/6-31G** tem alta similaridade com o experimental, destacado pelo ramo de coloração vermelho. Portanto, conforme os resultados obtidos a combinação do método *ab initio* com o conjunto de base HF/6-31G**, pode ser usado para a modelagem molecular da estrutura da artemisinina e de seus derivados antimaláricos, com mecanismo de ação na região do anel endoperóxido. Portanto, os resultados dos métodos teóricos e experimentais foram distribuídos de forma similares aos obtidos com PCA. A HCA confirma os resultados de PCA.

Figura 15. Dendograma do HCA para os cinco métodos e seis conjuntos de base, correlacionado com os parâmetros geométricos experimental do anel 1,2,13-trioxano.



Fonte: AUTOR (2014)

Recentemente, Santos et al. (2014b) validou o método computacional HF/6-31G** aplicado na modelagem molecular da artemisinina, propondo uma combinação de métodos de química quântica e da análise estatística para estudar os parâmetros geométricos de artemisinina na região do anel de endoperóxido (1,2,13-trioxano) comparando com os dados cristalográficos de raio-X. Na determinação das estruturas mais estáveis dos compostos estudados, bem como a obtenção das propriedades moleculares, o método Hartree-Fock com o conjunto de base de valência 6-31G** é adequado, em vez de abordagens tais como semiempíricas AM1, PM3, ZINDO e DFT (B3LYP/3-21G, B3LYP/3-21G*, B3LYP/3-21G**).

5.2 DOCKING MOLECULAR

Os cálculos de docking mostraram que toda a molécula do ligante é colocada paralelamente ao plano do anel de porfirina do heme, e a parte polar do ligante, que contém a ligação peróxido é dirigida para a parte polar do sistema heme contendo Fe^{2+} . Essa interação é visualizada na Figura 16, para os compostos ativos do conjunto treinamento (1, 3, 4, 10, 11, 15, 16, 19 e 20). Essas orientações foram assumidas para ser o mais favorável e, portanto, para representar o sistema real investigado, uma vez que eles foram escolhidos com base na menor energia livre de ligação (energia de interação). Para os compostos no conjunto estudado, os valores de $d(Fe-O1)$ variou 2.310 a 2.727Å; no entanto, este intervalo para a distância $d(Fe-O2)$ variou de 2.760 a 3.808Å. A distância $d(Fe-O13)$ variou de 4.811 a 5.434, e a distância $(Fe-O11)$ variou 4.897 a 5.525, como mostrado na Tabela 11.

Para artemisinina (1), a distância calculada $d(Fe-O1)$ foi 2.542Å, que é muito próximo ao valor reportado (2,7 Å) em outros estudos teóricos (COSTA et al., 2007; TONMUNPHEAN et al., 2001). Existe uma clara tendência interatômica envolvendo a separação entre Fe^{2+} e o átomo de oxigênio no anel trioxano, pois as distâncias são mais curtas para o átomo O1 do que para o átomo de O2. Este resultado reforça a ideia de que o átomo de O1 da artemisinina se liga preferencialmente ao Fe^{2+} do heme, em vez do átomo de O2.

Os compostos 4, 10, 11 e 20 têm atividade maior do que a artemisinina e também valores mais elevados de $d(Fe-O1)$. Eles têm uma grande substituinte que, certamente,

devido à repulsão provoca efeitos estéricos, o que os impede de se ligar mais perto do heme. Os compostos 5 e 6 foram desenvolvidos para aumentar a lipofilicidade, pois observou-se que quanto maior a lipofilicidade da artemisinina maior será a correlação com a atividade biológica. Os compostos 15, 16 e 20 apresentam grandes grupos substituintes no carbono α -metileno (*C), que aumentam substancialmente a atividade antimalárica dos compostos, devido aos seus efeitos eletrônicos e estéricos, respectivamente. O composto 3 demonstrou que a acetilação dos derivados da diidroartemisinina que contenham açúcar têm atividades similares ou melhores do que a artemisinina. No entanto, a desacetilação de açúcares reduz consideravelmente a atividade antimalárica.

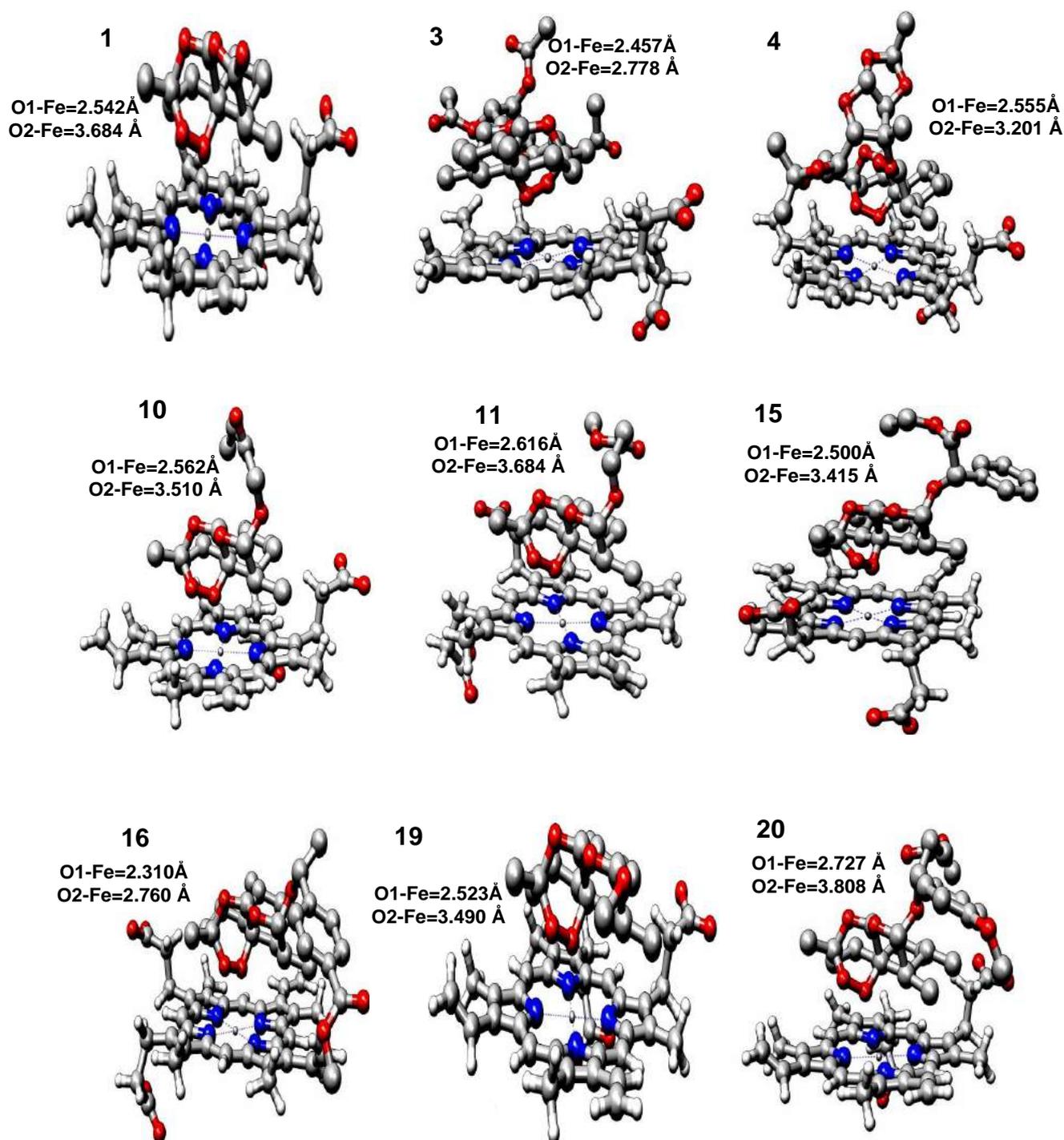
Tabela 11. Parâmetros calculados pelo docking molecular do heme-artemisinina e dos compostos derivados mais ativos.

Compostos	$E_{Complex}$ (Kcal mol ⁻¹)	Fe-O1 distância (Å)	Fe-O2 distância (Å)	Fe-O13 distância (Å)	Fe-O11 distância (Å)	logAR
1	-6.06	2.542	3.684	5.153	5.525	0.00000
3	-5.09	2.457	2.778	4.811	5.202	0.55376
4	-6.54	2.555	3.201	4.982	5.448	0.34115
10	-5.27	2.562	3.510	5.184	5.404	0.41754
11	-5.37	2.616	3.684	5.300	5.364	0.02633
15	-4.70	2.500	3.415	5.127	5.351	0.86031
16	-5.53	2.310	2.760	4.874	4.897	0.30707
19	-5.99	2.523	3.490	5.158	5.357	0.35423
20	-5.03	2.727	3.808	5.434	5.475	0.02174
$E_{Complex}$		0.06551	0.01761	0.19250	-0.20162	0.38917
Fe-O1			0.84202	0.85273	0.83598	-0.44984
Fe-O2				0.94792	0.81259	-0.48039
Fe-O13					0.65135	-0.48864
Fe-O11						-0.27755

A energia de interação para o complexo ligante/receptor mostrou boa correlação linear com atividade antimalárica ($r=0.38917$), e variou de -6.54 a -5.03 kcal.mol⁻¹, ver Tabela 11. Na verdade, apesar de algumas orientações estarem associadas com a menor energia de interação, elas pareciam ter forte atividade contra a malária, pois apresentou a ligação endoperóxido próximo do Fe²⁺. Atualmente, os mecanismos mais aceitos de ação antimalárica envolve a formação de um complexo entre o heme e derivados de artemisinina, em que o ferro do heme interage com O1 da ligação endoperóxido. Além disso, os substituintes e os efeitos da conformação podem afetar a distribuição de carga no

oxigênio e até mesmo a ligação Fe-O1 (PINHEIRO et al., 2003). Um aumento na área polar da artemisinina aumenta as interações polares entre heme, o ligante e a globina.

Figura 16. Interação entre heme-artemisininas dos compostos mais ativos do conjunto treinamento (1, 3, 4, 10, 11, 15, 16, 19 e 20).

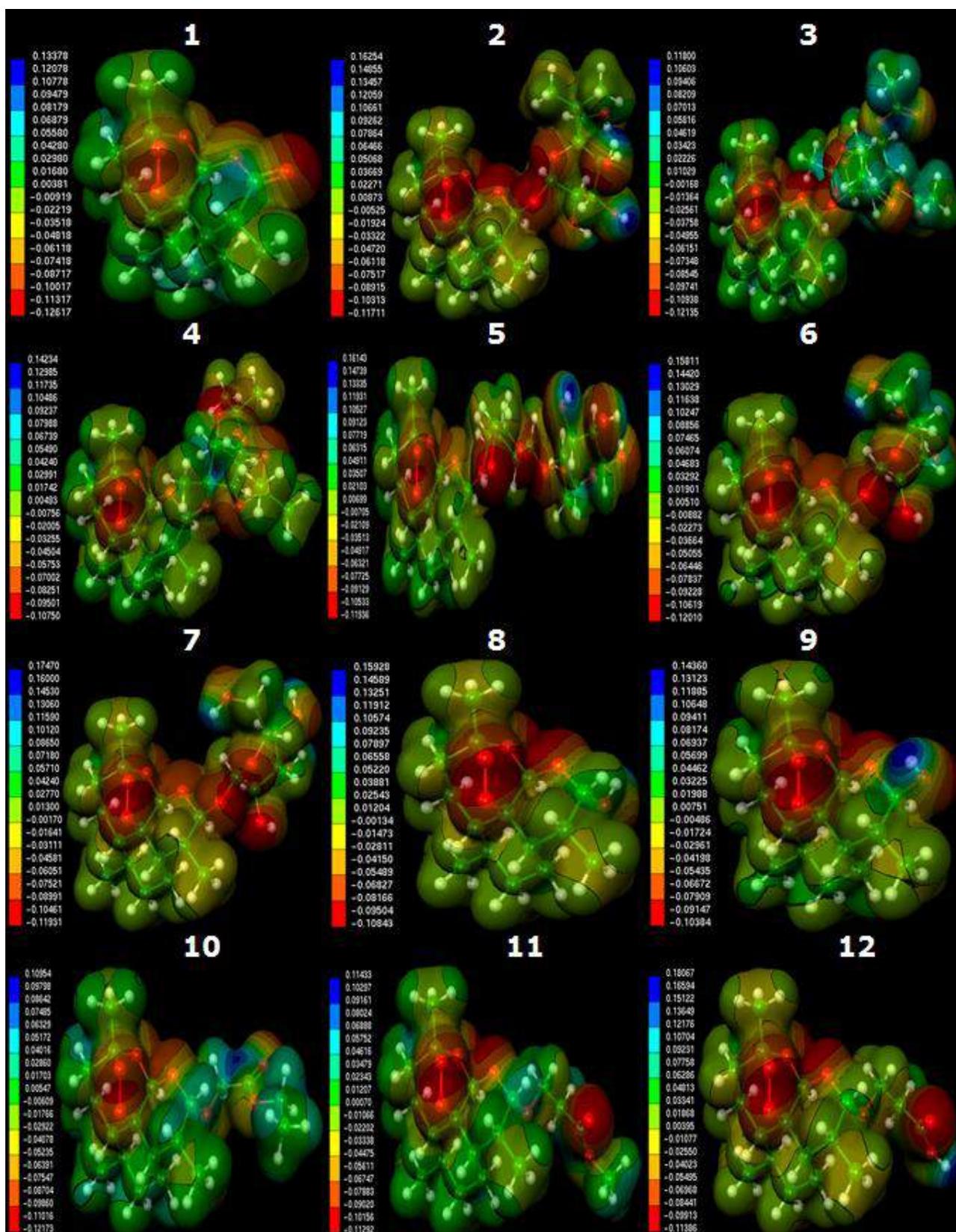


Fonte: AUTOR (2014)

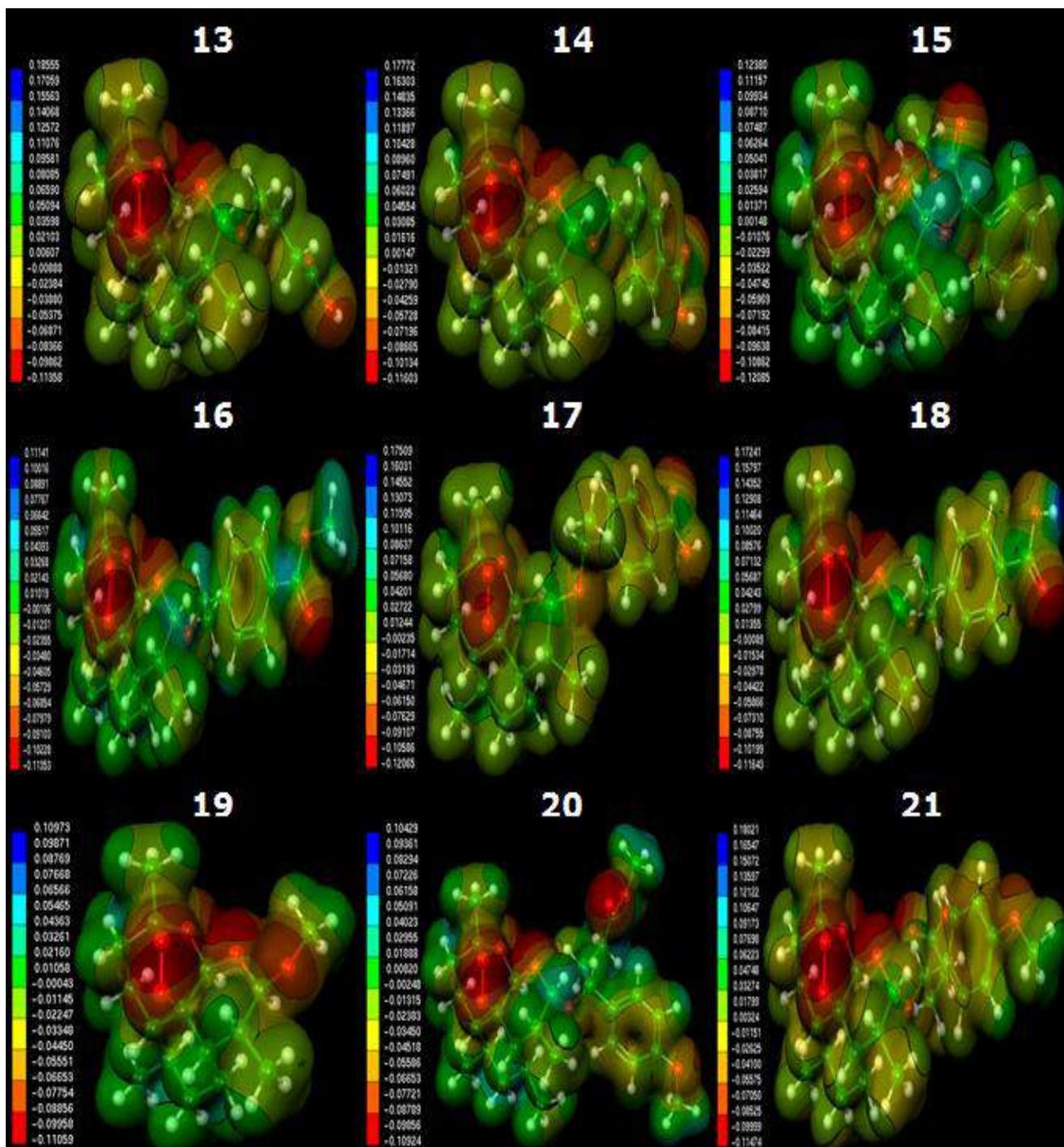
5.3 MAPA DE POTENCIAL ELETROSTÁTICO MOLECULAR (MEP)

Para identificar as características-chaves dos compostos derivados da artemisinina, os MEP foram avaliados e utilizados para comparações qualitativas na região do anel de 1,2,13-trioxano da artemisinina e seus derivados. A forma geométrica do potencial na região do anel de 1,2,13-trioxano é semelhante para todos os compostos ativos, e é caracterizada pelo potencial eletrostático negativo (região vermelha) de acordo com a literatura (GUHA; JURIS, 2004). A visualização do MEP é mostrada na Figura 17. Os compostos (2-21) tem uma região de potencial negativo próximo ao anel trioxano, semelhante ao MEP da artemisinina (composto 1), que tem um potencial eletrostático máximo de 0.13378u.a (região azul), e um mínimo de -0,12617u.a (região vermelha). O MEP máximo positivo (região azul) variou de 0.14234u.a a 0.10429u.a para os compostos ativos, enquanto que para os compostos menos ativos variou de 0.18555u.a a 0.14360u.a. Os valores correspondentes ao potencial eletrostático negativo mínimo (região vermelha) para os compostos mais ativos variaram de -0.10750u.a a -0.12617u.a, com valores próximos dos potenciais eletrostáticos da artemisinina. O potencial eletrostático negativo mínimo (região vermelha) para os compostos menos ativos variou de -0.10384u.a a -0.12065u.a, que são mais altos do que aqueles da artemisinina. A região de potencial eletrostático negativa é devido à ligação do endoperóxido (-C-O-O-C-), que é a característica mais notável do MEP da artemisinina e seus derivados. A distribuição da densidade eletrônica em torno do anel trioxano é considerada ser a responsável pela atividade contra a malária, isso é acreditado pelo fato de que a complexação da artemisinina com o heme envolva uma interação entre a ligação peróxido na zona mais carregada negativamente no ligante, e Fe^{2+} , a zona mais carregada positivamente no heme (a molécula receptora) (JEFFORD, 2001; CHENG et al., 2002). A presença de uma superfície negativa próximo ao anel trioxano sugere que estes compostos possuem um sítio reativo para o ataque eletrofílico, e deve possuir potência antimalárica; consequentemente, eles estão a ser investigados. Assim, no caso de um ataque eletrofílico do ferro do heme contra uma zona eletronegativa, há uma preferência para que possa ocorrer através da ligação endoperóxido. Através da análise dos mapas de MEP, a seleção de compostos inativos pode ser evitada.

Figura 17. Mapas de potencial eletrostático da artemisinina e seus derivados com atividade antimalárica contra o *Plasmodium falciparum* (W-2).



Fonte: AUTOR (2014)

Figura 17. *Continuação.*

Fonte: AUTOR (2014)

5.4 RESULTADOS DA PCA

Os resultados dessa análise mostraram que os descritores mais importantes foram os seguintes: a energia de hidratação (EH), carga sobre o átomo de oxigênio O11 (QO11), ângulo de torção D2 (O2-O1-Fe-N2) e o índice máximo de R/eletronegatividade de Sanderson (R Te^+). A energia de hidratação é a energia liberada quando as moléculas de água são separadas uma das outras e são atraídas por moléculas ou íons de um soluto que está se dissociando na água. A energia de hidratação compreende as interações solvente-solvente e soluto-solvente (WULFSBERG, 1987). A carga sobre o átomo O11 (QO11) é a medida da força com a qual uma partícula pode interagir eletrostaticamente com outra partícula (KARELSON et al., 1996). O R Te^+ é um tipo de descritor GETAWAY (geometria, topologia e conjunto dos pesos-atômicos), este descritor está associado com a forma, simetria do tamanho e distribuição molecular do átomo (CONSONNI et al., 2002).

O ângulo de torção D2 (O1-O2-Fe-N2), é de grande importância em nosso estudo, pois de acordo com a proposta de Jefford e colaboradores, o ferro do heme ataca a artemisinina na posição O1, e gera um radical livre na posição O2, após a ligação C3-C4 ser quebrada gerando um radical de carbono em C4 (JEFFORD et al., 1996). Este radical livre em C4 tem sido sugerido como uma importante substância na atividade antimalárica (POSNER, 1994). Estudos de docking molecular entre artemisinina e seu receptor, o grupo heme, feito por Tonmunphean, Parasuk e Kokpol também indicaram que o ferro do grupo heme interage preferencialmente com O1 em vez de O2 (TONMUNPHEAN et al., 2001).

Os valores dos descritores de cada composto selecionado com a técnica da PCA, bem como os valores de logAR, atividade relativa (AR) e IC₅₀ (concentração inibitória de 50%) são apresentados na Tabela 12. A Tabela 12 mostra a matriz de correlação de Pearson entre os descritores e logAR, observa-se que a correlação entre os pares de descritores é inferior a 0.70, enquanto que a correlação entre os descritores e logAR é inferior a 0.87. Os descritores selecionados com a técnica PCA representam as características necessárias para quantificar a atividade antimalárica desses compostos contra *Plasmodium falciparum* W-2.

Tabela 12. Propriedades moleculares selecionadas pela análise de componentes principais, valores experimentais ($\log AR$), IC_{50} e a matriz de correlação.

Compostos	EH	QO11	RTe+	D2	$\log AR$	AR	IC_{50} (ng/ml)
1 ⁺	-2.820	-0.605	0.105	120.868	0.00000	1.000000	0.6800
2 ⁻	-13.330	-0.516	0.127	27.480	-0.08130	0.829268	0.8200
3 ⁺	-4.190	-0.567	0.066	-62.834	0.55376	3.578947	0.1900
4 ⁺	-2.970	-0.558	0.069	94.272	0.34115	2.193548	0.3100
5 ⁻	-26.220	-0.558	0.079	-143.766	-2.40049	0.003977	171.0000
6 ⁻	-16.200	-0.547	0.123	-163.237	-1.72137	0.018994	35.6000
7 ⁻	-16.640	-0.517	0.135	-158.396	-1.69986	0.019959	34.0700
8 ⁻	-3.690	-0.631	0.187	111.395	-0.00634	0.985507	0.6900
9 ⁻	-5.670	-0.676	0.126	113.465	-0.00634	0.985507	0.6900
10 ⁺	-1.680	-0.599	0.100	13.716	0.41754	2.615385	0.2600
11 ⁺	-3.330	-0.662	0.080	124.375	0.02633	1.062500	0.6400
12 ⁻	-8.530	-0.666	0.131	105.696	-1.71943	0.019080	35.6400
13 ⁻	-8.210	-0.648	0.132	138.353	-1.07275	0.084577	8.0400
14 ⁻	-7.420	-0.651	0.131	127.415	-0.30737	0.492754	1.3800
15 ⁺	-2.930	-0.675	0.068	-17.600	0.86031	7.249467	0.0938
16 ⁺	-2.540	-0.637	0.081	100.819	0.30707	2.028035	0.3353
17 ⁻	-6.900	-0.754	0.150	95.974	-0.57147	0.268245	2.5350
18 ⁻	-6.980	-0.645	0.118	107.153	-0.25768	0.552486	1.2308
19 ⁺	-1.870	-0.501	0.105	114.392	0.35423	2.260638	0.3008
20 ⁺	-5.560	-0.641	0.082	153.908	0.02174	1.051330	0.6468
21 ⁻	-11.120	-0.651	0.141	10.910	-0.70556	0.196987	3.4520
EH		-0.329	-0.156	0.694	0.860	-	-
QO11			-0.203	-0.509	-0.127	-	-
RTe+				0.128	-0.333	-	-
D2					0.485	-	-

Os resultados do modelo SAR são apresentados na Tabela 13. O modelo foi construído com três componentes principais (3PCs). A primeira componente principal (PC1) descreve 40.8865% da informação total, a segunda componente principal (PC2) descreve 22.7045%, e a terceira (PC3) 11.5660%. Ainda nesta Tabela, observa-se que PC1 contém 51.1081% dos dados originais, e a combinação das duas primeiras componentes (PC1+PC2) contém 79.4887%, e todas as três (PC1+PC2+PC3) explicam 93.9461% da informação total, perdendo apenas 6.0539% da informação original. Os descritores EH, D2 e QO11 contribuir mais para PC1, enquanto no PC2, o descritor RTe+ é o principal contribuinte. Os componentes principais podem ser escrito como uma combinação linear dos descritores selecionados, as expressões matemáticas para o PC1 e PC2 são mostradas abaixo:

$$PC1 = 0.5705HE - 0.5088QO11 + 0.0925R\text{Te}^+ + 0.6381D2 \quad (1)$$

$$PC2 = -0.3847HE - 0.2987QO11 + 0.8731R\text{Te}^+ - 0.0207D2 \quad (2)$$

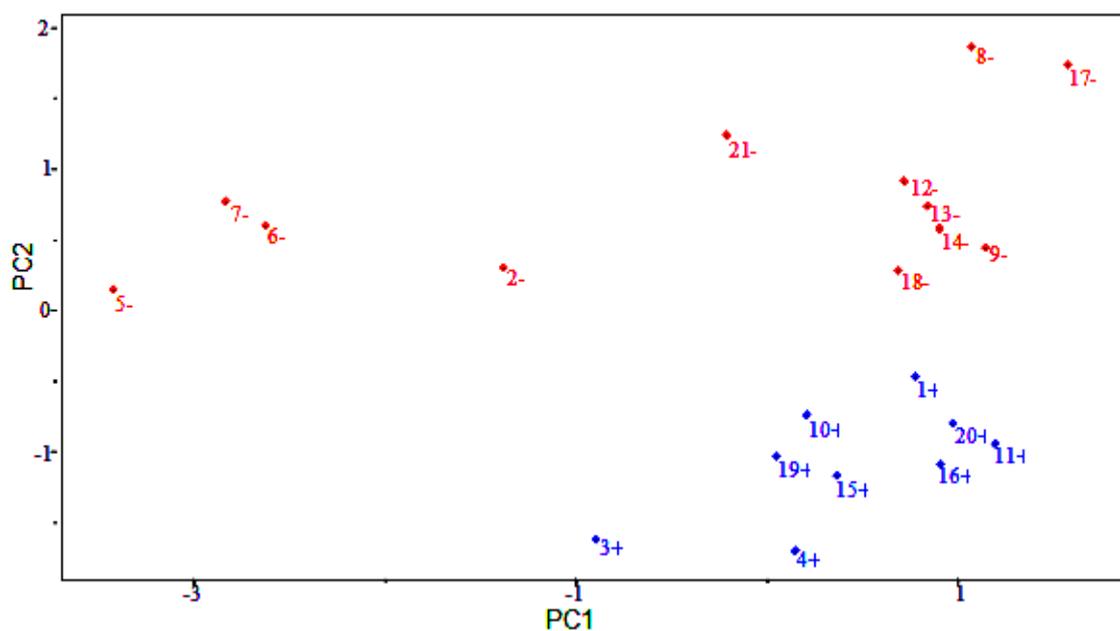
Tabela 13. Análise de componentes principais do modelo SAR, contribuição das propriedades moleculares selecionadas com a etapa de análise multivariada.

	Componente Principal		
	PC1	PC2	PC3
Variância (%)	40.8865	22.7045	11.5660
Variância acumulada (%)	51.1081	79.4887	93.9461
Propriedades Moleculares	Contribuição		
	PC1	PC2	
EH	0.5705	-0.3847	
QO11	-0.5088	-0.2987	
R Te^+	0.0925	0.8731	
D2	0.6381	-0.0207	

A Figura 18 mostra os scores dos 21 compostos estudados. Com base no gráfico, PC2 distingue entre os compostos que são mais potentes e menos potentes. Os compostos mais potentes encontram-se na parte inferior (1, 3, 4, 10, 11, 15, 16, 19 e 20), enquanto que os compostos menos potentes estão localizados na parte superior do gráfico (2, 5, 6, 7, 8, 9, 12, 13, 14, 17, 18 e 21).

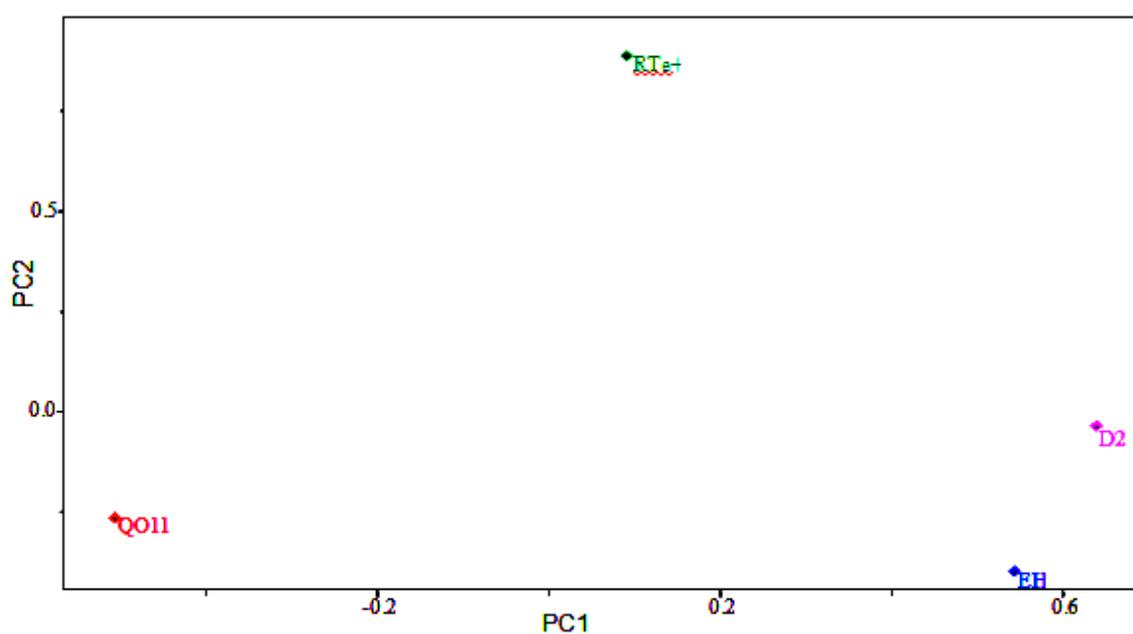
A Figura 19 mostra os loadings para os quatro descritores que são mais importantes para a classificação dos compostos. Os compostos mais potentes têm grandes contribuições dos descritores QO11, HE e D2, enquanto os compostos menos potentes têm uma elevada contribuição do descritor R Te^+ . Assim, os descritores QO11, HE e D2 são responsáveis em deslocar os compostos mais potentes na parte inferior do gráfico. Já o descritor R Te^+ desloca os compostos menos potentes na parte superior do gráfico. A Figura 18 mostra também que a maior contribuição do descritor R Te^+ na segunda componente principal, isto é, quanto maior for o valor do índice máximo de R/eletro-negatividade de Sanderson para um determinado composto, maior o valor do score, e assim indicando que o composto é menos potente do que os outros. Os outros descritores contribuem em menor grau. Por exemplo, o descritor de EH tem peso negativo em PC2, demonstrando que os compostos mais potentes têm geralmente valores mais elevados deste descritor.

Figura 18. Plot dos scores de PC1–PC2 para a artemisinina (1) e seus derivados com atividade antimalárica contra cepas de *P. falciparum* (W-2). Valores positivos indicam análogos mais potentes, e valores negativos indicam análogos menos potentes.



Fonte: AUTOR (2014)

Figura 19. Plot dos loadings de PC1–PC2 com quarto descritores selecionados para construir os modelos PLS e PCR da artemisinina e seus derivados com atividade biológica contra cepas de *P. falciparum* (W-2).



Fonte: AUTOR (2014)

Costa et al. (2007) demonstraram que a presença da água muda o ângulo diedro envolvido na complexação heme-artemisinina na região C-Fe-O1-O2. Assim, esta complexação influencia o processo de reconhecimento molecular entre artemisinina e seus derivados com o heme em sistemas biológicos aquosos. Portanto, a seleção do ângulo de torção D2 (O1-O2-Fe-N2) sugerem que a ação de novos medicamentos contra a malária dependa do ataque eletrofílico na ligação endoperóxido, particularmente sobre o átomo de O1. Este resultado foi confirmado pela análise dos mapas de MEP, docking molecular e na seleção das propriedades moleculares com a técnica PCA, como discutido anteriormente.

5.5 RESULTADOS DA HCA

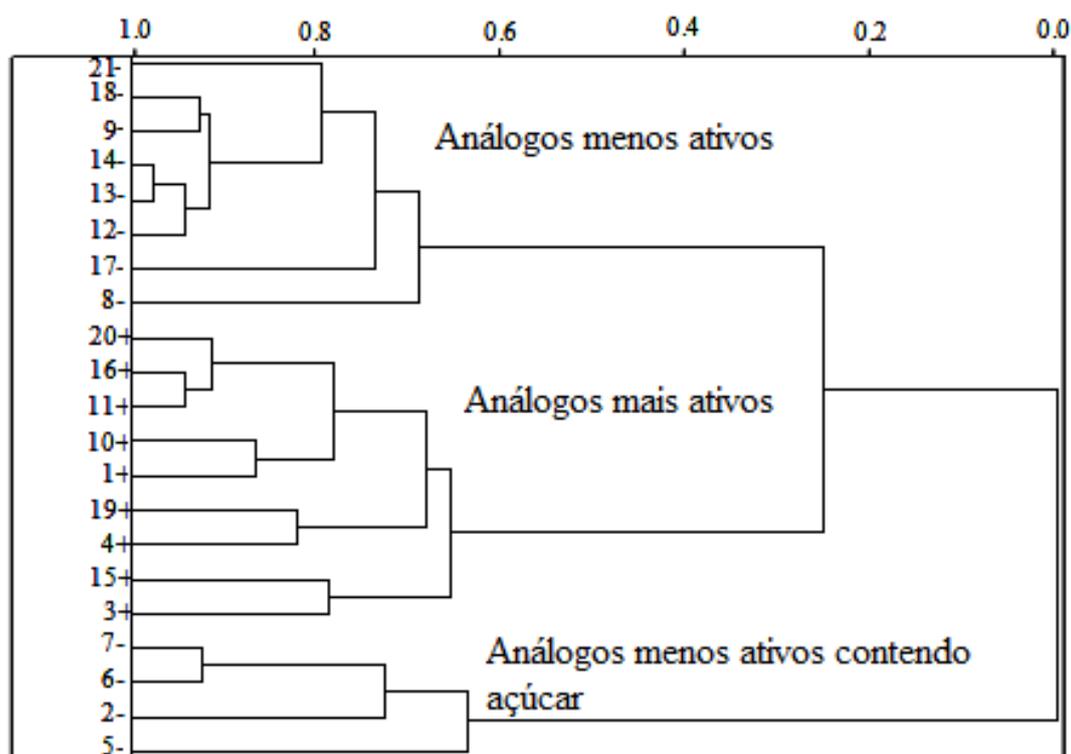
A análise estatística utilizada neste estudo deve agrupar compostos similares em categorias. As categorias são representadas por um diagrama de duas dimensões conhecido como dendrograma que ilustra as fusões ou divisões realizadas em cada fase sucessiva da análise. As amostras individuais (compostos) são representadas pelas ramificações na parte inferior do dendrograma. A similaridade entre os clusters é dada pelo comprimento de seus ramos, assim compostos que apresentam baixa similaridade têm ramos longos enquanto que os compostos de alta similaridade têm ramos curtos. O método HCA classificou os compostos em três classes (mais ativo, menos ativo e menos ativo contendo açúcar), e foi baseado na distância Euclidiana e método incremental (INFOMETRIX, 2001). No método incremental, a distância entre dois grupos é a distância máxima entre uma variável num aglomerado e uma variável no outro conjunto. Os descritores utilizados para realizar o HCA foram às mesmas que as utilizadas para a PCA, isto é, EH, QO11, D2 (O1-O2-Fe-N2) e RTe+.

Na técnica de HCA, as distâncias entre os pares de amostras são calculados e comparados. As pequenas distâncias implicam que os compostos são semelhantes, enquanto que as amostras diferentes serão separadas por distâncias relativamente grandes. O dendrograma na Figura 20 mostra o gráfico de HCA, bem como os compostos são separados em três categorias principais. A escala de similaridade varia de 0 (zero) para as amostras com nenhuma similaridade a 1 para as amostras com semelhança idêntica. Ao analisar o dendrograma, algumas conclusões podem ser tiradas, embora os compostos apresentem alguma diversidade estrutural. A HCA mostraram resultados semelhantes aos obtidos com PCA.

Os compostos são agrupados de acordo com suas atividades biológicas. Os compostos mais potentes são 1, 3, 4, 10, 11, 15, 16, 19 e 20. Os compostos menos potentes são agrupados

em dois grupos, um dos quais contém compostos 8, 9, 12, 13, 14, 17, 18 e 21, e o outro cluster contém derivados da artemisinina contendo açúcar (2, 5, 6 e 7).

Figura 20. Dendrograma da HCA da artemisinina e derivados com atividade biológica contra cepas de *P. falciparum* (W-2). Os valores positivos indicam análogos mais potentes, e os valores negativos indicam compostos menos ativos.



Fonte: AUTOR (2014)

5.6 RESULTADOS DOS MODELOS MÍNIMOS QUADRADOS PARCIAIS (PLS) E REGRESSÃO DE COMPONENTES PRINCIPAIS (PCR)

A qualidade estatística (SNEDECOR et al., 1967) dos modelos PLS e PCR foram avaliadas por parâmetros como o coeficiente de correlação e coeficiente de correlação ao quadrado (R^2), variância explicativa (R^2_{ajust} , ou seja, R^2 ajustado), desvio padrão (s), razão de variância (F), coeficiente de correlação de validação cruzada (Q^2), erro padrão de validação (SEV), erro da soma dos quadrados dos resíduos preditos (PRESS) e desvio padrão de validação cruzada (S_{press}) (CHATTERJEE et al., 2000; DIUDEA, 2000;

BIKASH et al., 2003). Os melhores modelos de regressão foram selecionados com base em altos valores de R^2 , R^2_{ajust} , Q^2 e F (estatística de avaliação do significado conjunta), e baixos valores de s , SEV , $PRESS$ e S_{press} .

As propriedades calculadas e os valores de atividades experimentais para os compostos do conjunto treinamento (Tabela 13) foram utilizados para construir os modelos de regressão. Os modelos construídos usando os métodos PLS e PCR basearam-se em três variáveis latentes, e 18 compostos, sendo 3 compostos (2, 12 e 13) para o conjunto de validação externa.

As equações de regressão obtidas para os modelos PLS (Eq. 3) e PCR (Eq. 4) que se relacionam com os descritores e atividade biológica são os seguintes:

$$\text{Log AR} = 0.520565\text{EH} - 0.151382\text{QO11} - 0.158294\text{RTe}^+ + 0.353200\text{D2} \quad (3)$$

$n = 18$, $R^2 = 0.9468$, $R^2_{ajust} = 0.9354$, $s = 0.2211$, $F_{(4,11)} = 57.8889$, $Q^2 = 0.8566$,
 $SEV = 0.3202$, $PRESS = 0.6847$, $S_{PRESS} = 0.0636$.

$$\text{Log AR} = 0.304816\text{EH} - 0.258614\text{QO11} + 0.046881\text{RTe}^+ + 0.340163\text{D2} \quad (4)$$

$n = 18$, $R^2 = 0.8488$, $R^2_{ajust} = 0.8164$, $s = 0.3729$, $F_{(4,11)} = 18.2454$, $Q^2 = 0.7313$,
 $SEV = 0.4438$, $PRESS = 1.9476$, $S_{PRESS} = 0.1073$.

Os resultados obtidos com os modelos PLS e PCR foram muito próximos, com variação de $R^2 = \pm 0.098$, $R^2_{ajust} = \pm 0.119$, $s = \pm 0.1518$, $F_{(4,11)} = \pm 39.6435$, $Q^2 = \pm 0.1253$, $SEV = \pm 0.1236$, $PRESS = \pm 1.2629$ e $S_{press} = \pm 0.0437$ (entre PLS e PCR). A qualidade dos modelos PLS e PCR pode ser demonstrada através da comparação da medida e das atividades preditas. Os erros de validação obtidos pelo método leave-one-out cross-validation são mostrados na Tabela 14. No modelo PLS apenas cinco compostos (2, 3, 12, 13 e 15) apresentaram erros elevados de validação, e o modelo de PCR resultou em cinco compostos (3, 12, 13, 15 e 17) com valores residuais elevados. Os modelos PLS e PCR construídos apresentaram o melhor ajuste para os compostos com alta atividade, pois os compostos com baixa atividade apresentaram valores residuais elevados.

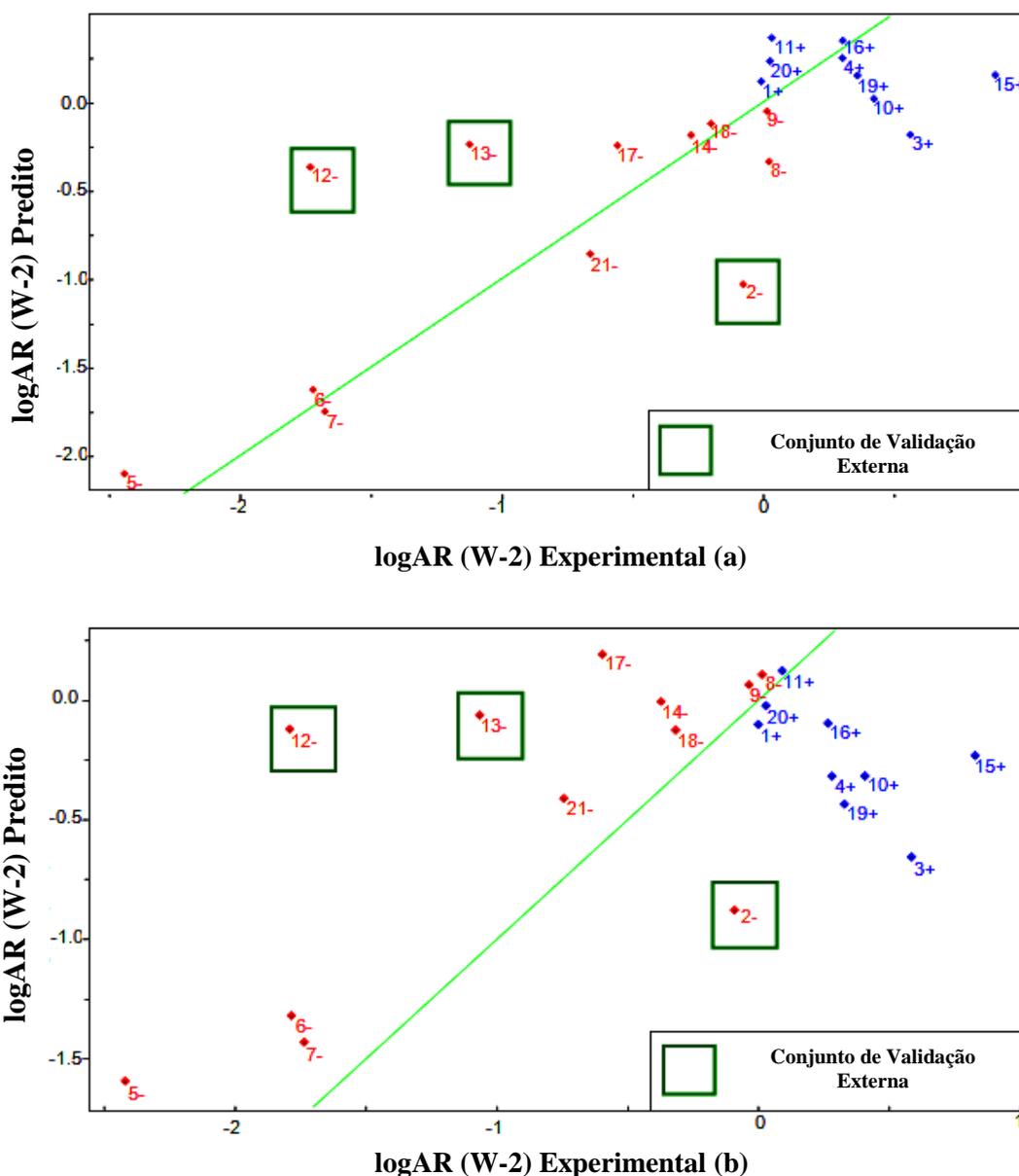
Tabela 14. Resultado predito dos modelos PLS e PCR, o erro de validação para logAR (experimental)

Compostos	Predito		Erro de validação		Experimental
	PLS	PCR	PLS	PCR	LogRA
1+	0.2548	0.0902	0.2548	0.0902	0.0000
2- ^[a]	-1.0163	-0.8805	-0.935	-0.7992	-0.0813
3+	-0.2855	-0.6548	-0.8392	-1.2085	0.5537
4+	0.2199	-0.1981	-0.1212	-0.5392	0.3411
5-	-2.1192	-1.7899	0.2812	0.6105	-2.4004
6-	-1.6837	-1.4214	0.0376	0.2999	-1.7213
7-	-1.8113	-1.5164	-0.1115	0.1834	-1.6998
8-	-0.1174	0.2199	-0.1111	0.2262	-0.0063
9-	0.0948	0.2284	0.1011	0.2347	-0.0063
10+	0.0342	-0.1872	-0.3833	-0.6047	0.4175
11+	0.4495	0.2485	0.4232	0.2222	0.0263
12- ^[a]	-0.1670	0.1032	1.5524	1.8226	-1.7194
13- ^[a]	-0.0920	0.1144	0.9807	1.1871	-1.0727
14-	-0.0583	0.1144	0.249	0.4217	-0.3073
15+	0.1452	-0.0974	-0.7151	-0.9577	0.8603
16+	0.3812	0.1289	0.0742	-0.1781	0.3070
17-	0.0203	0.4326	0.5917	1.004	-0.5714
18-	-0.0448	0.0392	0.2128	0.2968	-0.2576
19+	0.0888	-0.2535	-0.2654	-0.6077	0.3542
20+	0.3283	0.1684	0.3066	0.1467	0.0217
21-	-0.6913	-0.3463	0.0142	0.3592	-0.7055

[a] Compostos para o conjunto de validação externa

Os valores medidos versus predito utilizando os modelos PLS e PCR construídos são apresentados na Figura 21(a) e 21(b), respectivamente. Nos plots de PLS e PCR os compostos com maior atividade são identificados pela coloração azul, e compostos com menor atividade são identificados pela coloração vermelha, incluindo os compostos do conjunto de validação externa, destacado pela coloração verde. De acordo com os modelos PLS e PCR, as quatro variáveis apresentam diferentes magnitudes dos coeficientes de regressão (em valor absoluto). Os modelos revelam que os compostos com elevada potência biológica contra *P. falciparum* tem uma combinação de valores mais elevados de EH e D2, e valores inferiores de QO11 e RTe+ para o modelo PLS. Porém, para o modelo de PCR, os compostos têm valores mais elevados de EH e D2, baixo valor de QO11 e valores positivos para RTe+. Os parâmetros de validação suportam o fato de que os modelos são eficientes e, portanto satisfatório dado à complexidade dos mecanismos antimaláricos, e do pequeno número de descritores (quatro) selecionados para construir o modelo QSAR.

Figura 21. Plot dos valores preditos versus experimental para logAR através dos modelos (a) PLS e (b) PCR.



Fonte: AUTOR (2014)

5.7 COMPOSTOS DO CONJUNTO TESTE

Os compostos do conjunto teste ainda não sintetizados são mostrados na Figura 12, compostos de 22-51, que foram planejados de acordo com três propostas:

a) Lin et al. (1989) planejaram com intuito de aumentar a lipofilicidade, pois se observou que quanto maior a lipofilicidade da artemisinina maior a atividade biológica.

b) A proposta de Lin et al. (1992) considerou que acetilação de derivados da diidroartemisinina contendo açúcar possuem atividades comparavelmente similares ou melhores que artemisinina. Contudo, a desacetilação de açúcares reduz consideravelmente a atividade antimalárica.

c) Lin e Miller (1995) consideraram que grandes grupos substituintes no carbono α -metileno (*C) aumentam substancialmente a atividade antimalárica dos compostos em função dos efeitos eletrônicos e estéricos, respectivamente.

Os compostos do conjunto teste foram modelados a partir da estrutura mais estável da artemisinina, composto 1 da Figura 12, e construídos usando o programa GaussView 5.0 (FRISCH et al., 2004), realizando a otimização completa da geometria de cada composto com o conjunto de base de valência separada 6-31G** utilizando o método Hartree-Fock como implementado no programa Gaussian 03 (FRISCH et al., 2004). Após obter-se a geometria mais estável de cada composto, determinou-se apenas os descritores selecionados em PCA e utilizados na construção dos modelos QSAR (PLS e PCR), sendo eles EH, QO11, D2 e RTe+, mostrados na Tabela 15.

Tabela 15. Propriedades moleculares selecionadas pela análise de componentes principais do conjunto teste com atividade antimalárica desconhecida.

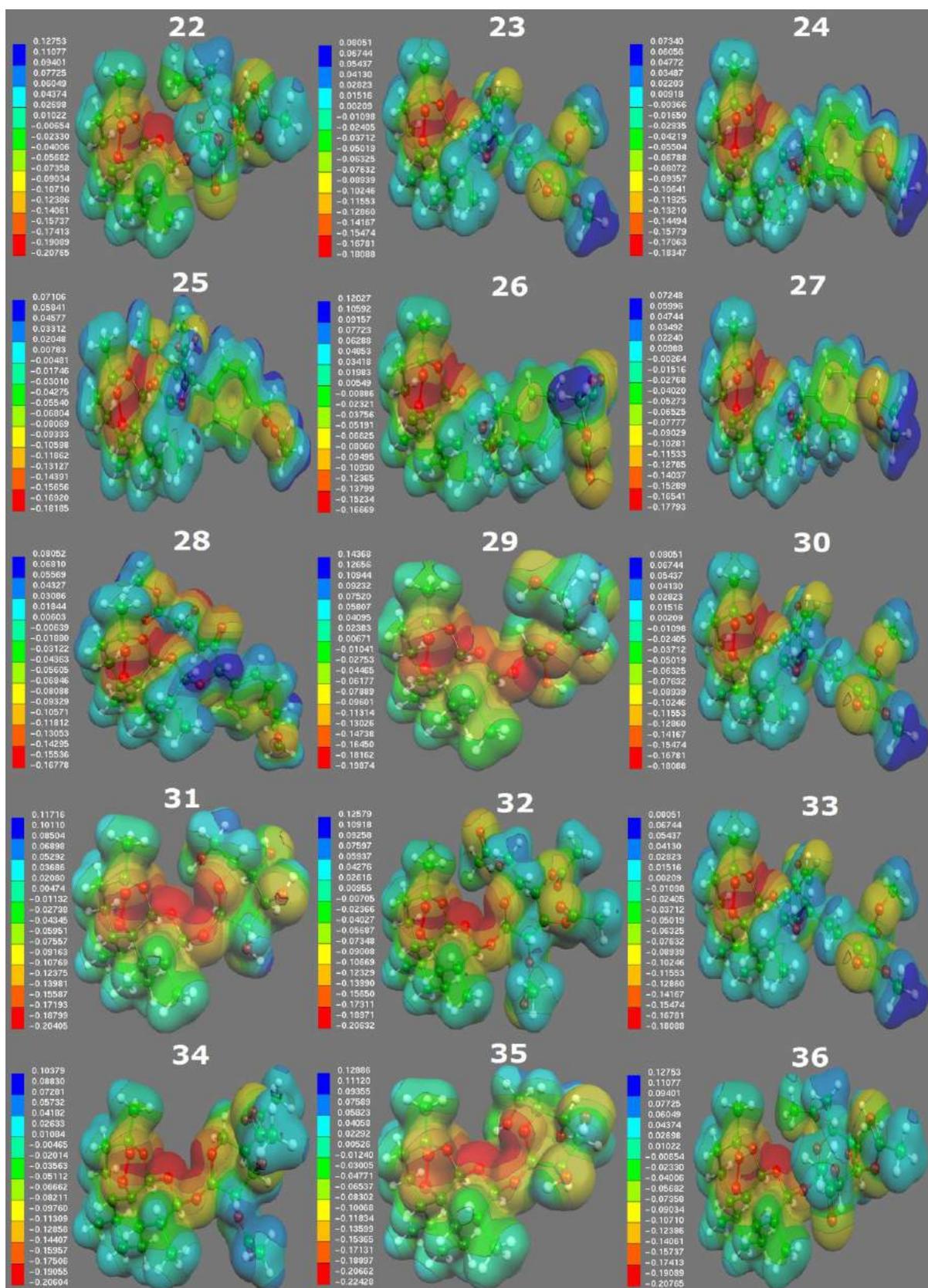
Compostos Testes	EH	QO11	RTe ⁺	D2
22	-3.460	-0.663	0.076	7.585
23	-3.370	-0.664	0.077	141.065
24	-4.790	-0.556	0.069	130.453
25	-5.780	-0.675	0.077	98.153
26	-8.070	-0.603	0.076	-76.018
27	-4.650	-0.602	0.073	-4.170
28	-7.440	-0.575	0.066	-9.051
29	-15.920	-0.482	0.100	73.480
30	-4.470	-0.594	0.070	125.875
31	-15.240	-0.601	0.106	9.276
32	-4.500	-0.532	0.063	-37.529
33	-13.680	-0.578	0.126	-83.125
34	-4.550	-0.572	0.071	8.222
35	-13.620	-0.523	0.121	32.018
36	-4.280	-0.584	0.071	-27.718
37	-2.740	-0.650	0.105	152.098
38	-2.850	-0.673	0.081	101.819
39	-2.680	-0.603	0.068	-13.617
40	-3.290	-0.577	0.064	-65.438
41	-10.210	-0.615	0.122	10.190
42	-7.044	-0.557	0.062	-13.671
43	-7.841	-0.654	0.131	127.514
44	-2.910	-0.657	0.072	-25.670
45	-2.870	-0.670	0.069	-19.115
46	-7.020	-0.745	0.155	95.479
47	-4.240	-0.600	0.066	122.578
48	-8.120	-0.684	0.123	131.353
49	-8.350	-0.665	0.134	105.669
50	-5.676	-0.667	0.126	113.564
51	-3.640	-0.636	0.067	7.855

5.7.1 Mapa de Potencial Eletrostático Molecular (Conjunto Teste)

Os MEP dos compostos do conjunto teste foram construídos, com o programa Molekel (FLUKIGER, 2002), na tentativa de identificar-se características chaves que são responsáveis pela atividade antimalárica dos compostos. Os mapas MEP foram calculados a partir da estrutura mais estável de cada composto, utilizando o método Hartree-Fock e o conjunto de base de valência separada 6-31G** como implementado no programa Gaussian 03 (FRISCH et al., 2004). Os mapas MEP foram computados pelas cargas de Mullikan utilizando as palavras chaves POP=FULL.

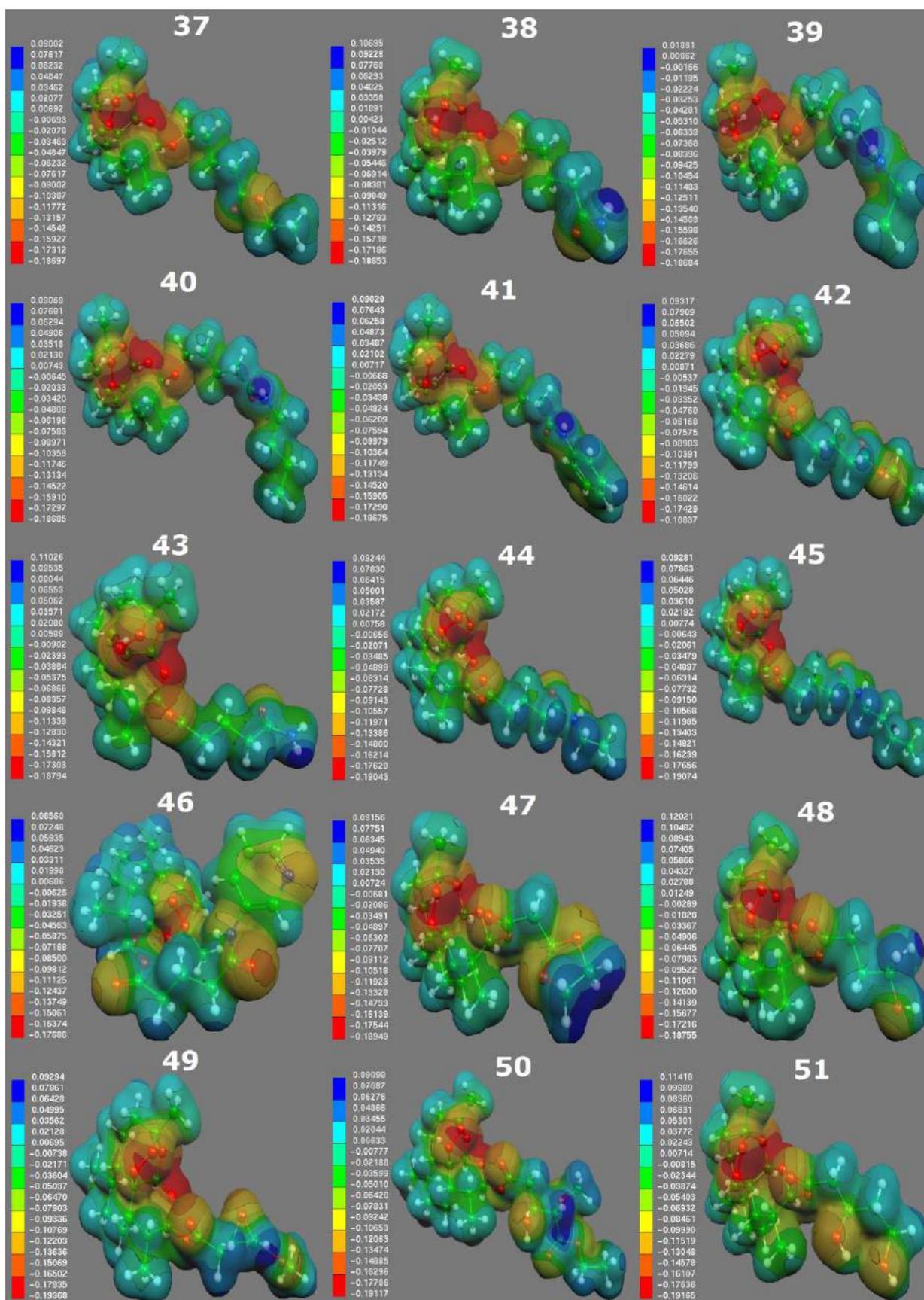
A Figura 22 dispõem os mapas MEP de cada composto do conjunto teste, observa-se que todos os compostos têm uma região de potencial negativo de forma similar próximo do anel trioxano, semelhante à região de potencial negativo da artemisinina. A região de potencial negativo é atribuída à ligação endoperóxido (C-O-O-C) que é essencial para a expressão da atividade antimalárica dos compostos. Os mapas de potencial eletrostático molecular dos compostos do conjunto teste dão suporte qualitativo aos resultados das predições da atividade antimalárica obtidos com os modelos QSAR (PLS e PCR).

Figura 22. Mapas do Potencial Eletrostático Molecular para os compostos do conjunto teste mostrados na Figura 12 com atividade antimalárica desconhecida.



Fonte: AUTOR (2014)

Figura 22. Continuação



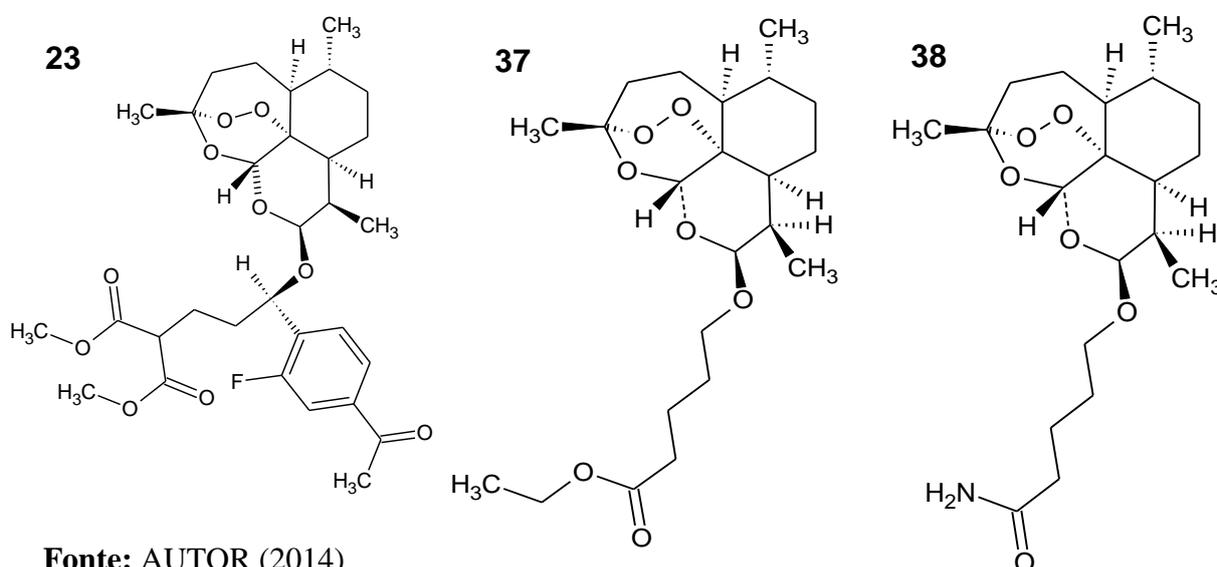
Fonte: AUTOR (2014)

5.7.2 Aplicação dos modelos PLS e PCR aos compostos do conjunto teste

Os modelos QSAR (PLS e PCR) foram construídos para prever a atividade antimalárica desconhecida de 30 novos derivados da artemisinina mostrados na Figura 12, compostos 22-51. Na Tabela 16 são mostrados os resultados do $\log AR$ pelos modelos PLS e PCR. De acordo com Tabela 16 o modelo PLS mostrou que 15 compostos do conjunto teste (22, 23, 27, 30, 32, 34, 36-40, 44, 45, 47 e 51) são preditos como sendo mais ativos, pois apresentaram valores de $\log AR$ maiores que zero ($\log AR > 0$).

O modelo PCR apenas nove compostos do conjunto teste (23, 25, 37, 38, 43, 46, 48-50) foram preditos como mais ativos, onde mostraram valores de $\log AR$ maiores que zero ($\log AR > 0$), totalizando 24 compostos propostos como mais ativo dos 30 compostos sugeridos. Porém, os compostos 23, 37 e 38, mostrados na Figura 23, foram os únicos que apresentaram valores de $\log AR$ maiores que zero ($\log AR > 0$), em ambos os modelos (PLS e PCR), com resíduos de previsão variando entre 0.028951 à -0.1351, sugerindo que esses novos compostos nos dois modelos (PLS e PCR) são mais potentes que a artemisinina podendo ser sintetizados e testados contra atividade antimalárica, leishmaniose como realizado por Carvalho et al. (2011) que propuseram um conjunto de 13 artemisininas (7 menos ativo e 6 que não tinham sido testados) e desses seis, um foi esperado para ser mais ativo contra *Leishmania donovani*. Além disso, podem ser testados em atividade anticancer contra carcinoma hepatocelular humano como proposto por Barbosa et al. (2011) que realizaram a modelagem molecular e estudos quimiométricos envolvendo artemisinina e seus derivados.

Figura 23. Compostos promissores para futuras sínteses e ensaios biológicos.



Fonte: AUTOR (2014)

Tabela 16. Atividade antimalárica predita (log AR) pelos modelos PLS e PCR para os compostos do conjunto teste e resíduos de previsão entre os modelos.

Compostos do conjunto teste	Predito (logAR)		Resíduos de previsão (PLS - PCR)
	PLS	PCR	
22	0.28515	-0.08153	0.36669
23	0.16134	0.29643	-0.13510
24	-0.02669	-0.17894	0.15225
25	-0.13236	0.11823	-0.25059
26	-0.27509	-0.71065	0.43556
27	0.13050	-0.37989	0.51040
28	-0.25528	-0.60997	0.35469
29	-1.52738	-1.01052	-0.51686
30	0.02281	-0.04421	0.06701
31	-1.36654	-0.73626	-0.63028
32	0.18562	-0.72371	0.90933
33	-1.05361	-0.98490	-0.06871
34	0.13185	-0.44910	0.58095
35	-1.16288	-0.86252	-0.30036
36	0.20640	-0.49559	0.70198
37	0.23829	0.33770	-0.09941
38	0.27428	0.24533	0.02895
39	0.41609	-0.32859	0.74468
40	0.38362	-0.59287	0.97649
41	-0.66296	-0.46071	-0.20225
42	-0.19509	-0.67464	0.47955
43	-0.45107	0.10809	-0.55916
44	0.39618	-0.17717	0.57335
45	0.39508	-0.11549	0.51058
46	-0.30336	0.40099	-0.70434
47	0.05839	-0.02784	0.08623
48	-0.49407	0.20276	-0.69682
49	-0.50004	0.06913	-0.56917
50	-0.13357	0.19691	-0.33048
51	0.25967	-0.19372	0.45339

Os compostos mais potentes apresentam $\log AR \geq 0$.

Os compostos menos potentes apresentam $\log AR < 0$.

6 CONSIDERAÇÕES FINAIS

A utilização da metodologia combinando métodos da química quântica e análise multivariada foram úteis para validar o método e nível de teoria (HF/6-31G**), quando comparado com os dados cristalográficos. O uso de cálculos semi-empíricos, *ab initio*/Hartree-Fock e DFT foi possível pelo fato de ter moléculas formadas entre 50-100 átomos, e dessa forma os cálculos de química quântica foram desenvolvidos em diferentes métodos e conjunto de base.

A técnica de análise multivariada PCA e HCA foram importantes, devido possibilitar a classificação dos métodos e níveis de teoria em três grupos distintos, uma contendo os métodos semi-empíricos, DFT e *ab initio*/Hartree-Fock. Dentre os parâmetros analisados e selecionados os mais importantes na classificação dos métodos e conjunto de base foram quatro parâmetros geométricos relacionados ao anel endoperóxido (1, 2, 13-trioxano), e identificados pelos átomos: O13C12 que é a distância interplanar entre estes dois átomos (comprimento de ligação); O1O2C3 que é o ângulo de ligação entre esses três átomos; C3O13C12C12a e C12C12aO1O2 estão relacionados com ângulo diedro ou de torção. Os resultados de HCA foram semelhantes aos obtidos com PCA. Após as análises de PCA e HCA verificou-se que o método Hartree-Fock com o conjunto de base de valência separada 6-31G** é adequada para estudos de modelagem molecular da estrutura da artemisinina e seus derivados.

Os docking moleculares puderam auxiliar no estudo da atividade biológica dos novos compostos do conjunto teste com atividade antimalárica desconhecida. Os mapas de potencial eletrostático molecular (MEP) dos compostos do conjunto de treinamento revelaram que os compostos 2-21 apresentam região de potencial negativo de forma similar próximo do anel trioxano, e esta região de potencial negativo é semelhante da artemisinina. Os MEPs forneceram informações qualitativas, sendo úteis para o planejamento dos compostos do conjunto teste, e deram suporte aos resultados das previsões da atividade biológica antimalárica feita com os modelos QSAR (PLS e PCR).

As análises multivariadas PCA e HCA foram importantes, devido possibilitar a classificação dos compostos do conjunto de treinamento em duas classes distintas, uma contendo os compostos mais ativos e a outra os menos ativos, onde o total de descritores calculados foi de 1.187, dentre os descritores analisados e selecionados os mais importantes na classificação dos compostos após a análise da matriz de correlação entre os descritores e o logaritmo da atividade relativa ($\log AR$) foram a energia de hidratação (EH), carga sobre o

átomo O11 (QO11), ângulo de torção O2O1FeN2 (D2) e o índice máximo de R/eletronegatividade de Sanderson.

Os resultados de HCA foram semelhantes aos obtidos com PCA, e os métodos PLS e PCR possibilitaram a obtenção de significativos modelos QSAR. As avaliações dos modelos demonstraram de forma adequada e convenientemente aos dados preditos, e os parâmetros que verificaram o grau de ajuste dos modelos (resíduos, R^2 , R^2_{ajust} , s) são muito significativos, e o grau de significância executado através do teste F foi bastante satisfatório, e o grau de preditividade avaliado através dos parâmetros Q^2 , SEP , $PRESS$ e S_{PRESS} demonstraram que os modelos QSAR possuem elevada capacidade preditiva. Dessa forma pode-se dizer que os modelos obtidos são bem ajustados e possuem excelentes capacidades preditivas.

Na análise dos trinta novos derivados propostos da artemisinina (conjunto teste), os compostos 23, 37 e 38 apresentaram valores de $\log AR$ maiores que zero ($\log AR > 0$), em ambos os modelos (PLS e PCR), com resíduos de previsão variando entre 0.028951 à -0.1351, não descartando a hipótese de que os 24 novos derivados ativos (15 compostos propostos pelo modelo PLS e 9 compostos sugeridos pelo modelo PCR), não possam descrever a conformação antimalárica com mecanismo de ação na região do anel endoperóxido. Portanto, neste trabalho foram usadas técnicas diferenciadas para se obter informações consistentes a um posterior trabalho das estruturas estudadas. Como propostas para futuros trabalhos destacam-se:

- a) Síntese dos compostos que foram preditos mais ativos do conjunto teste;
- b) Realizações de ensaios biológicos destes novos compostos;
- c) Aplicação da metodologia deste trabalho no estudo de novos derivados com atividade biológica.
- d) Comprovando as atividades biológicas dos compostos do conjunto teste, originam-se novos produtos biotecnológicos, podendo ser patenteados.
- e) Os compostos patenteados com as suas devidas atividades biológicas poderão ser utilizados no Sistema Único de Saúde para o tratamento da malária no Brasil.

REFERÊNCIAS

- ACD/CHEMSKETCH FREEWARE, version 12.00: Advanced Chemistry Development, Inc, Toronto, ON, Canada, 2010.
- ALBUQUERQUER, C. **Foco residual de malária na Rodovia Manaus-Boa Vista**. Tese de Mestrado. Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, 1982.
- ALECRIM, W.D. **Estudos clínicos e epidemiológicos da malária no Rio Amazonas**. Brasília. Tese de Mestrado. Faculdade de Ciências Médicas da Universidade de Brasília, Brasília, DF, 1979.
- ALLEN, F. H. The Cambridge Structural Database: A Quarter of a Million Crystal Structures and Rising. **Acta Crystallographica**. Section B, v.58, n.01, p.380-388, 2002., <http://dx.doi.org/10.1107/S0108768102003890>
- ARAÚJO, J. Q.; CARNEIRO, J. W. M.; ARAÚJO, M. T.; LEITE, F. H. A.; TARANTO, A. G. Interaction between artemisinin and heme. A Density Functional Theory study of structures and interaction energies. **Bioorganic & Medicinal Chemistry**, v.16, p.5021-5029, 2008.
- ASAWAMAHASAKDA, W.; ITTARAT, I.; PU, Y.-M.; ZIFFER, H.; MESHNICK, S.R. Reaction of Antimalarial Endoperoxides with Specific Parasite Proteins. **Antimicrobial Agents Chemotherapy**, v.38, p.1854-1858, 1994.
- ATKINSON, R. C.; SUPPES, P. Program in computer-assisted instruction: Final report, Washington, DC. Office of Education (DGEW), edited by M. R. Simonson and A. Thompson, **Educational Computing Foundations**, Prentice Hall, Upper Saddle River, New Jersey, 1968.
- AVERY, M. A.; GAO, F.; CHONG, W. K. M. ; MEHROTRA, S.; MILHOUS, W. K. Structure-activity relationships of the antimalarial agent artemisinin. 1. Synthesis and comparative molecular field analysis of C-9 analogs of artemisinin and 10-deoxoartemisinin. **Journal of Medicinal Chemistry**, v.36, p.4264-4275, 1993.
- BARBOSA, J. P.; FERREIRA, J. E. V.; FIGUEIREDO, A. F.; ALMEIDA, R. C. O.; SILVA, O. P. P.; CARVALHO, J. R. C.; CRISTINO, M. G. G.; PINHEIRO, J. C.; VIEIRA, J. L. F.; SERRA, R. T. A. Molecular Modeling and Chemometric Study of Anti- cancer Derivatives of Artemisinin. **Journal of the Serbian Chemical Society**, v.76, n.9, p. 1263-1282, 2011. Disponível em : <<http://dx.doi.org/10.2298/JSC111227111B>>.
- BARREIRO, E. J.; FRAGA, C. M. A. **Química Medicinal – As bases moleculares da ação dos fármacos**. Artmed: Porto Alegre, 2001.
- _____. **Química Medicinal, as Bases Moleculares da Ação dos Fármacos**. Artmed: Porto Alegre, 2002.
- BARROS, NETO, B.; SCARMINIO, I. S.; BRUNS, R. E. **Planejamento e otimização de experimentos**, 2 ed. p.152-298. Campinas: Editora da UNICAMP, 1996.

BEEBE, K. R.; PELL, R. J.; SEASHOLTZ, M. B. **Chemometrics: A Practical Guide**. Wiley: New York, 1998.

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

BERNARDINELLI, G.; JEFFORD, C. W.; MARIĆ, D.; THOMSON, C.; WEBER, J. *Int. J. Quantum Chem. Quantum Biology Symposium*, v.21, p.117-131, 1994.

BIKASH, D.; SHOVALAL, G.; SUBRATA, B.; SOMA, S.; TARUN, J. QSAR study on some pyridoacridine ascididemin analogues as anti-tumor agents. **Bioorganic & Medicinal Chemistry**, v.11, p.5493-5499, 2003.

BLINDER, S. M. Basic Concepts of Self-Consistent-Field Theory American. **Journal of Physics**, v.33, p.431-520, 1965.

BOWMAN, D.; LAWSON, D.; BASHAN, D.; BROWN, T.; CHILLINGWORTH, C. M.; CHURCHER, A.; CRAIG, R. M.; DAVIES, K.; DEULIN, T. F.; GENTHES, R.; GWILLIAM, N. H. D.; HARNIS, S.; HOLROYD, T.; I-IORNSBY, P.; HORROCKS, K.; JAGEIS, B.; JASSAL, S. J.; KYES, M.; MOULE, K.; MUNGAIL, L.; MURPHY, K.; OLIVER, M. A.; QUALI, M. A.; RAJADREAM, S.; RUTTER, J.; SKELTON, R.; SQUARES, S.; SQUARES, J. E.; SU, S.; WHITEHEAD, J. R.; WOODWARD, C.; NAWBOLD, B. G. B. The complete nucleotide sequence of Chromosome 3 of *Plasmodium falciparum*, **Nature**, v.400, p. 532-538, 1999.

BOYS, S. F. A General Method of Calculation for the Stationary States of Any Molecular System. **Proceedings of the Royal Society A**, v.200, p.542-554, 1950.

BRASIL. Ministério da Saúde. **Manual de terapêutica da malária**. Brasília, FUNASA, 2001.

_____. Ministério da Saúde. Secretaria de Vigilância em Saúde. **Guia de vigilância epidemiológica**. 6. ed. Brasília, p.816, 2005a.

_____. Ministério da Saúde. Secretaria de Vigilância em Saúde. **Manual de diagnóstico laboratorial da malária**. Brasília, p.112, 2005b.

_____. Ministério da Saúde. Secretaria de Vigilância em Saúde. **Guia para profissionais de saúde sobre prevenção da malária em viajantes**. Brasília, p.24, 2008a.

_____. Ministério da Saúde. **Situação epidemiológica da malária no Brasil, 2008**. Brasília, p.330-334, 2008b.

_____. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica. **Guia prático de tratamento da malária no Brasil / Ministério da Saúde**, Secretaria de Vigilância em Saúde, Departamento de Vigilância Epidemiológica. – Brasília : Ministério da Saúde, 2010. p. 36: il. color. – (Serie A. Normas e Manuais Técnicos).

_____. Ministério da Saúde. **Relatório Situação Epidemiológica da malária na Amazônia**. 2010. Brasília –Set/2011

_____. Secretaria de Vigilância em Saúde – Ministério da Saúde. **Boletim Epidemiológico**, Brasília. v. 44, n. 1, 2013.

BRENEMAN, C. M.; WINBERG, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. **Journal of Computational Chemistry**, v.11, p.367-373, 1990.

BUNGE, A. V. **Introdução a Química Quântica**. Edgard Blücher, São Paulo, 1979.

CAMBRIDGE SOFT CORPORATION, CS. Chem3D 5.0: Molecular Modeling and Analysis, Cambridge, 1998.

_____. ChemDraw 5.0: Chemical Structure Drawing Standard, Cambridge, 1998.

_____. Chem3D. 100 Cambridge Park Drive, Cambridge, MA 02140: Cambridge, 1999.

CARDOSO, F. J. B.; COSTA, R. B.; FIGUEIREDO, A. F.; BARBOSA, J. P.; NAVA-JUNIOR, I.; PINHEIRO, J. C.; ROMERO, O. A. S. Modeling Artemisinin Derivatives with Potent Activity against *P. falciparum* Malária with Ab Initio and PLS Methods. **Internet Electronic Journal of Molecular Design**, v.6, p.122-134, 2007.

_____.; FIGUEIREDO, A. F.; LOBATO, M. S.; MIRANDA, R. M.; ALMEIDA, R. C. O.; PINHEIRO, J. C. A study on antimalarial artemisinin derivatives using MEP maps and multivariate QSAR. **Journal Molecular Modeling**, v.14, p.39-48, 2008.

CARDOSO, R. F.; GOLDENBERG, P. Malária no Estado do Amapá, Brasil, de 1970 a 2003: trajetória e controle. **Cadernos de Saúde Pública**, Rio de Janeiro, v.23, n.6, p.1339-1348, 2007.

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. **Química Nova**, v.26, p.428, 2003.

CARVALHO, C. J. R.; FERREIRA, V. E. J.; BARBOSA, P. J.; LOBATO, S. M.; MENESES, F. C. C.; SOEIRO, M. M.; FARIAS, S. M.; ALMEIDA, O. R. C.; VENTURA, C. K.; PINHEIRO, C. J.; PINHEIRO, R. L. A. Computational Modeling of Artemisinins with Antileishmanial Activity. **Journal Computational and Theoretical Nanoscience**, v.8, n.11, p. 2193-2203, 2011.

CESARIO, M.; CESARIO, R. R. Malária, Amazônia e desenvolvimento. **Scientific American Brasil**, São Paulo ano 4, n.46, p.54-55, mar. 2006.

CHATTERJEE, S.; HADI, A. S.; PRICE, B. Regression Analysis by Examples. New York: **Wiley VCH**. 3.ed., 2000.

CHEMPLUS. Modular Extensions for HyperChem Release 6.02: Molecular Modeling for Windows, HyperClub, Inc, Gainesville, 2000.

- CHEMSITE. Interactive 3d molecular modeling: ChemSW® Inc., Fairfield, 2001.
- CHENG, F.; SHEN, J.; LUO, X.; ZHU, W.; GU, J.; JI, R.; JIANG, H.; CHEN, K. Molecular Docking and 3D-QSAR Studies on the Possible Antimalarial Mechanism of Artemisinin Analogues. **Bioorganic & Medicinal Chemistry**, v.10, p.2883-2891, 2002.
- CHINA COOPERATIVE RESEARCH GROUP ON QINGHAOSU AND ITS DERIVATIVE AS ANTIMARIALS. Clinical Studies On The Treatment Of Malária With Qinghaosu And Its Derivatives. **Journal of Traditional Chinese Medicine**, v.2, p.45-50, 1982.
- CLARK, T. VAMP: Computer Cahemie Centrum, Friedrich-Alexander-Universität Erlangen-Nürnberg, NögelsbachstraBe 25, D-91052 Erlangen, Germany, 2002.
- CLEMENTI, E.; RAIMONDI, D. L. Atomic Screening Constants from SCF Functions. **The Journal of Chemical Physics**, v.38, p.2686-2689, 1963.
- _____.; REINHARDT, W. P. Constantes Triagem Atômicas de SCF Funções II Átomos com 37-86 Elétrons. **The Journal of Chemical Physics**, v.47, p. 1300-1307, 1967.
- CRISTINO, M. G. G.; MENESES, C. C. F.; SOEIRO, M. M.; Ferreira, J. E. V.; FIGUEIREDO, A. F.; BARBOSA, J. P.; ALMEIDA, R. C. O.; PINHEIRO, J. C.; PINHEIRO, A. L. R. Computational Modeling of Antimalarial 10 Substituted Deoxoartemisinins. **Journal of Theoretical and Computational Chemistry**, v.11, n.2, p.241-263, 2012. Disponível em: <<http://dx.doi.org/10.1142/S0219633612500162>>.
- COHEN, N.C; J. M. BLANEY; C. HUMBLET; P. GUND; D. C. BANY. Molecular Modeling Software and Methods for Medicinal Chemistry. **Journal of Medicinal Chemistry**, v.33, p.883-894, 1990.
- COMPUTATIONAL CHEMISTRY PROGRAM: Molecular Modeling ProTM4.0, ChemSW®, Inc., Fairfield, 2001.
- CONSONNI, V.; TODESCHINI, R.; PAVAN, M. Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors, 1. Theory of the Novel 3D Molecular Descriptors. **Journal of Chemical Information and Computer Science**, v.42, p. 682-692, 2002.
- _____.; GRAMATICA, P. Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors, 2. Application of the Novel 3D Molecular Descriptors to QSAR/QSPR Studies. **Journal of Chemical Information Computer Science**, v.42, p.693-705, 2002.
- CORREA, R.; RAMOS, A. S. Os anofelinos da região meridional do Estado de São Paulo. **Arquivos de Higiene e Saúde Pública**, v.7, p.37-57, 1942.
- COSTA, M. S.; KIRALJ, R.; FERREIRA, M. M. C. Estudo teórico da interação existente entre a artemisinina e o heme. **Química Nova**, v.30, p.25-31, 2007.

COSTA, E. V. M. **Estudo etnobotânico sobre plantas utilizadas como antimaláricas no estado do amapá, avaliação da atividade antimalárica e toxicidade aguda por via oral de *amazonia campestris* (aubl.) moldenke**. Macapá-AP, Brasil, 142f. Tese de Doutorado - Programa de Pós-Graduação em Biodiversidade Tropical - PPGBio, UNIFAP / EMBRAPA-AP / IEPA / CI-BRASIL, 2013.

COUTO, A. A. et al. Controle da transmissão da malária em área de garimpo no estado do Amapá com participação da iniciativa privada. **Caderno de Saúde Pública**, Rio de Janeiro, v.17, n.4, p.897-907, jul.-ago., 2001.

CUSTÓDIO, R.; ANDRADE, J. C.; AUGUSTO, F. O Ajuste de Funções Matemáticas a Dados Experimentais. **Química Nova**, v.20, p. 219-225, 1997.

DEWAR, M. J. S.; ZOEBISCH, E. G.; HEALY, E. F.; STEWART, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. **Journal of the American Chemical Society**, v.107, p.3902-3909, 1985.

DIUDEA, M. V. **QSPR/QSAR Studies for Molecular Descriptors**. Nova Science: Huntingdon, New York, 2000.

DIXON, K.E.; ROBERTS, D.R., LLEWELLYN, C.H. Contribuição ao estudo epidemiológico da malária em trecho da Rodovia Transamazônica, Brasil. **Revista do Instituto de Medicina. Tropical São Paulo**, v.21:287-292, 1979

DYER, J. R. **Applications of Absorption Spectroscopy of Organic Compounds**. Prentice-Hall, New Jersey, 1965.

ESTRADA, E.; MOLINA, E. Novel local (fragment-based) topological molecular descriptors for QSPR/QSAR and molecular design. **Journal Molecular Graphics Modelling**, v.20, p. 54-64, 2001.

FERREIRA, E. Distribuição geográfica dos anofelinos no Brasil e sua relação com o estado atual da erradicação da malária. **Revista Brasileira de Malariologia e Doenças Tropicais**, v.16, p.329-348, 1964.

FERREIRA, P. Malária. 2006. [On-line]. Disponível em: <<http://www.fiocruz.br/ccs/cgi/cgilua.exe/sys/start.htm?infoid=191&sid=6>>. Acesso em: 08 mar. 2013.

FERREIRA, J. E. V.; FIGUEIREDO, A. F.; BARBOSA, J. P.; CRISPINO, M. G. G.; MACEDO, W. J. C.; SILVA, O. P. P.; MALHEIROS, B. V.; SERRA, R. T. A.; PINHEIRO, J. C. A study of new antimalarial artemisinins through molecular modeling and multivariate analysis. **Journal of Serbian Chemical Society**, v.75, p.1533-1548, 2010.

FERREIRA, M. M. C.; ANTUNES, A. M.; MELGO, M. S.; VOLPE, P. L. O. Quimiometria I: Calibração Multivariada, um Tutorial. **Química Nova**, v.22, n.5, p.724, 1999.

_____. Multivariate QSAR. **Journal of Brazilian Chemical Society**, v.13, n.6, p. 742-753, 2002.

_____.; MONTANARI, C. A.; GAUDIO, A. C. Seleção de Variáveis em QSAR. **Química Nova**, v.25, n.3, p.439-448, 2002.

FLUKIGER, P. MOLEKEL 4.2: Swiss Center for Scientific Computing, Manó, Switzerland, 2002.

FORESMAN, J. B.; FRISCH, A. E. **Exploring Chemistry with Electronic Structure Methods: A Guide to Using Gaussian**. Gaussian, Inc, USA, 1993.

FRANÇA, T. C. C.; SANTOS, M. G.; FIGUEROA-VILLAR, J. D. Malária: aspectos históricos e quimioterapia. **Química Nova**, São Paulo, v.31, n.5, p.1271-1278, 2008.

FRANCIS, S. E.; SULLIVAN JR., D. J.; GOLBERG, D. E. Hemoglobin metabolism in the malaria parasite plasmodium falciparum. **Annual Review of Microbiology**, v.51, p.97-123, 1997.

FRANKE, R.; GRUSKA, A. Multivariate data analysis of chemical and biological data in: van de WATERBEEMD, H. (Eds.). **Chemometrics Methods in Molecular Design**. New York: VCH, 1996. p.115.

FRISCH, M. J., G. W. TRUCKS, H. B. SCHLEGEL, G. E. SCUSERIA, M. A. ROBB, J.R. CHEESEMAN, V. G. ZAKRZEWSKI, J. A. MONTGOMERY, JR., R.E. STRATMANN, J. C. BURANT, S. DAPPRICH, J. M. MILLAM, A.D. DANIELS, K. N. KUDIN, M. C. STRAIN, O. FRAKAS, J. TOMASI, V. BARONE, M. COSSI, R. CAMMI, B. MENNUCCI, C. POMELLI, C. ADAMO, S. CLIFFORD, J. OCHTERSKI, G. A. PETERSSON, P. Y. AYALA, Q. CUI, K. MOROKUMA, D. K. MALICK, A. D. RABUCK, K. RAGHAVACHARI, J. B. FORESMAN, J. CIOSLOWSKI, J. V. ORTIZ, A. G. BABOUL, B. B. STEFANOV, G. LIU, A. LIASHENKO, P. PISKORZ, I. KOMAROMI, R. GOMPERTS, R. L. MARTIN, D. J. FOX, T. KEITH, M. A. AL-LAHAM, C. Y. PENG, A. NANAYAKKARA, C. GONZALEZ, M. CHALLACOMBE, P. M. W. GILL, B. JOHNSON, W. CHEN, M. W. WONG, J. L. ANDRES, C. GONZALEZ, M. HEAD-GORDON, E. S. REPLOGLE, AND J. A. POPLE. Gaussian 98: revision A. 7 Gaussian, Inc., Pittsburgh PA, 1998.

FRISCH, M. J., G. W. TRUCKS, H. B. SCHLEGEL, G. E. SCUSERIA, M. A. ROBB, J. R. CHEESEMAN, J. A. MONTGOMERY, JR, T. VREVEN, K. N. KUDIN, J. C. BURANT, J. M. MILLAM, S. S. IYENGAR, J. TOMASI, V. BARONE, B. MENNUCCI, M. COSSI, G. SCALMANI, N. REGA, G. A. PETERSSON, H. NAKATSUJI, M. HADA, M. EHARA, K. TOYOTA, R. FUKUDA, J. HASEGAWA, M. ISHIDA, T. NAKAJIMA, Y. HONDA, O. KITAO, H. NAKAI, M. KLENE, X. LI, J. E. KNOX, H. P. HRATCHIAN, J. B. CROSS, C. ADAMO, J. JARAMILLO, R. GOMPERTS, R. E. STRATMANN, O. YAZYEV, A. J. AUSTIN, R. CAMMI, C. POMELLI, J. W. OCHTERSKI, P. Y. AYALA, K. MOROKUMA, G. A. VOTH, P. SALVADOR, J. J. DANNENBERG, V. G. ZAKRZEWSKI, S. DAPPRICH, A. D. DANIELS, M. C. STRAIN, O. FARKAS, D. K. MALICK, A. D. RABUCK, K. RAGHAVACHARI, J. B. FORESMAN, J. V. ORTIZ, Q. CUI, A. G. BABOUL, S. CLIFFORD, J. CIOSLOWSKI, B. B. STEFANOV, G. LIU, A. LIASHENKO, P. PISKORZ, I. KOMAROMI, R. L. MARTIN, D. J. FOX, T. KEITH, M. A. AL-LAHAM, C. Y. PENG, A. NANAYAKKARA, M. CHALLACOMBE, P. M. W. GILL, B. JOHNSON, W. CHEN, M. W. WONG, C. GONZALEZ, AND J. A. POPLE, Gaussian 03: Revision C.02, Gaussian, Inc. Wallingford, CT, 2004.

GARCIA-MARTINS, G. Status of malária eradication in the Americas. **The American Journal Tropical Medicine and Hygiene**, v.21, n.23, p.617, 1972.

GAUDIO, A. C.; ZANDONADE, E. Proposição, validação e análise dos modelos que correlacionam estrutura química e atividade biológica. **Química Nova**. v.24, p.658-671, 2001.

GELADI, P. Notes on the history and nature of partial least squares (PLS) modeling. **Journal Chemom**, v.2, p.231-246, 1988.

GORDON, M. GAMESS, 1605 Gilman Hall, Iowa State University, Ames, IA 50011-3111.

GUHA, R. Exploring structure-activity data using the landscape paradigm. **WIREs Computational Molecular Science**, v.2, n.6, p. 829-841, 2012.

_____.; JURIS, P. C. Development of QSAR models to predict and interpret the biological activity of artemisinin analogues. **Journal Chemical Information and Computer Science**, v.44, p.1440-1449, 2004.

HARTREE, D. R. The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods. **Mathematical Proceedings of the Cambridge Philosophical Society**, v.24, p.89-110, 1928.

HAYNES, R. K.; KRISHNA, S. Artemisinins: activities and actions. **Microbes and Infection**, v.6, n.14, p.1339-1346, 2004

HEHRE, W. J. Ab Initio Molecular Orbital Theory. **John Wiley & Sons**, New York, p. 678-682, 1986.

_____. **A Guide to Molecular Mechanics and Quantum Chemical Calculations**: Wavefunction, Inc. (2003).

_____.; STEWART, R. F.; POPLER, J. A. Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals. **The Journal of Chemical Physics**, v.51, p.2657-2664, 1969

HIRSCHFELDER, J. O. Semi-Empirical Calculations of Activation Energies. **The Journal of Chemical Physics**, v.9, n.8, p. 645, 1941.

HOFFMANN, R. An Extended Huckel Theory. I. Hydrocarbons. **The Journal of Chemical Physics**, v.39, p.1397-1412, 1963.

HONIG, B.; NICHOLLS, A. Classical electrostatics in biology and chemistry. **Science**, v.268, p.1144-1149, 1995.

HSU, E. The history of qing hao in the Chinese materia medica. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v.100, n.6, p.505-508, 2006.

HYPERCUBE. Hyperchem: Program Version 4.5, Gainesville, Florida, USA.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Disponível em: <<http://www.ibge.gov.br/estadosat/perfil.php?sigla=pa>>. Acesso em: 10 dez. 2010.

INFOMETRIX, Inc.: Pirouette 3.01, Woodinville, WA, 2001.

JAMES, H. M.; COOLIDGE, A. S. The Ground State of the Hydrogen Molecule. **The Journal of Chemical Physics**, v.1, n.12, p.825, 1933.

JEFFORD, C. W. Correlating the Molecular Electrostatic Potentials of Some Organic Peroxides with Their Antimalarial Activities. **Journal of Chemical Information and Computer Science**, v.40, n.2, p.354-357, 2000.

_____. Why artemisinin and certain synthetic peroxides are potent antimalarials. Implications for the mode of action. **Current Medicinal Chemistry**, v.8, p.1803-1826, 2001

_____.; 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. **Helvetica Chimica Acta**, v.79, p.1475-1487, 1996.

JOHNSON, M.; MAGGIORA, G. Concepts and Applications of Molecular Similarity. **John Wiley & Sons**. New York, 1990.

JONES, G.; WILLET, P.; GLEN, R. C.; LEACH, A. R.; TAYLOR R. Development and validation of a genetic algorithm for flexible docking. **Journal of Molecular Biology**, v. 267, n.3, p.727-748, 1997.

JUDSON, R. S.; JAEGAR, E. P.; TREASURYWALA, A. M.; PETERSON, M. L. **Journal Computational Chemistry**, v.14, p.1407, 1993.

KARELSON, M.; LOBANOV, V. S.; KATRITZKY, A. R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. **Chemical Reviews**, v.96, p.1027-1044, 1996.

KITCHEN, D. B.; DECORNEZ H.; FURR J. R.; BAJORATH. Docking and scoring in virtual screening for drug discovery: methods and applications. **Journal of Nature Reviews – Drug Discovery**, v.3, n.11, p.935-949, 2004.

KLAYMAN, D. L. Qinghaosu (Artemisinin): An Antimalarial Drug from China. **Science**. v.228, p.1049-1055, 1985.

KOKPOL, S. K. Analysis of Structure-Activity Relation for Primaquine Antimalarial Drugs by a Quantum Pharmacological Approach. **Analytical Sciences**, v.4, n.6, p.565-568, 1988.

KONTOYIANNI, M.; MCCLELLAN, L. M.; SOKOL, G. S. Evaluation of Docking Performance: Comparative Data on Docking Algorithms **Journal of Medicinal Chemistry**, v.47, n.3, p.558-564, 2004.

KOROLKOVAS, A. Planejamento de Fármacos. **Ciência e Cultura**, v.41, p.528-537, 1989.

KUBINYI, H. QSAR: Hansch analysis and related approaches in: MANNHOLD, R.; KROGSGAARD-LARSEN, P.; TIMMERMAN, H. (Eds.). **Methods and Principles in Medicinal Chemistry**. New York: VCH, v.1, p. 21-106, 1993,

LEACH, A. R., *Molecular Modelling Principles and Applications*. **Longman**, Essex, 1996.

_____. *Molecular Modelling: Principles and Applications*. **Prentice Hall**. Englewood Cliffs. NJ, 2001a.

_____. *Molecular Modelling—Principles and Applications*. **Pearson Education Limited**, 2001b.

LEITE, F. H. A.; TARANTO, A. G.; SANTOS JUNIOR, M. C.; BRANCO, A.; ARAUJO, M. T.; CARNEIRO, J. W. M. Search for New Antimalarial Compounds Obtained From Natural Sources by Molecular Modeling. **International Journal of Quantum Chemistry**, v.110, n.11, p. 2057-2066, 2010.

_____.; CARNEIRO, J. W. M.; ARAUJO, M. T.; COMAR Jr, M.; TARANTO, A. G. Docking Between Natural Peroxides and Heme Group by Parametric Method 6. **International Journal of Quantum Chemistry**, v.112, n.20, p.3390-3397, 2012. <http://dx.doi.org/10.1002/qua.24247>

LEVINE, I. N. **Quantum Chemistry**. 4 ed. New York: Prentice-Hall, p. 343-464, 1991.

LIN, A. J.; KLAYMAN, D. L.; MILHOUS, W. K. Antimalarial activity of new water-soluble dihydroartemisinin derivatives. **Journal of Medicinal Chemistry**, v.30, p.2147-2150, 1987

_____.; LI, L. Q.; ANDERSEN, S. L. Antimalarial activity of new dihydroartemisinin derivatives. 5. Sugar analogues. **Journal of Medicinal Chemistry**, v.35, p.1639-1642, 1992.

_____.; MILLER, R. E.; Antimalarial activity of new dihydroartemisinin derivatives. 6. alpha-Alkylbenzylic ethers. **Journal of Medicinal Chemistry**, v.38, 764-770, 1995.

_____.; ZIKRY, A. B.; KYLE, D. E. Antimalarial activity of new dihydroartemisinin derivatives. 7. 4-(p-substituted phenyl)-4(R or S)-[10(alpha or beta)-dihydroartemisininoxy] butyric acids. **Journal of Medicinal Chemistry**, v.40, p.1396-1400, 1997.

LISGARTEN, J. N.; POTTER, B. S.; BANTUZEKO, C.; PALMER, R. A. Structure, Absolute Configuration, and Conformation of the Antimalarial Compound, Artemisinin. **Journal of Chemical Crystallography**, v.28, n.7, p.539-543, 1998. <http://dx.doi.org/10.1023/A:1023244122450>

LIU, SY. Problemas asociados con el vector: resistencia fisiologica y complejos de especies. In: **Malária en las Americas**. Washington, Publicações Científicas, v.405, 1981.

LOMBARDINO, J. G.; LOWE, J. A. A guide to drug discovery: The role of the medicinal chemist in drug discovery — then and now. **Nature Reviews – Drug Discovery**, v.3, n.10, p.853-862, 2004.

MACÊDO, W. J. C. **Estudo da relação estrutura-atividade (SAR) de derivados da diidroartemisinina com atividade antimalárica.** Tese de Mestrado. Universidade Federal do Pará, Belém, PA, 2010.

MARCUS, R. A. Generalization of the Activated Complex Theory of Reaction Rates. I. Quantum Mechanical Treatment. **The Journal of Chemical Physics**, v.41, n.9, p.2614-2623, 1964.

MARTENS, H.; NAES, T. **Multivariate Calibration.** New York: Wiley, 1989.

MARTIN, Y. **Quantitative Drug Design: A Critical Introduction.** Boca Raton. 2.ed. FL: CRC Press, 2010.

MIELKE, S. L.; PETERSON, K. A.; SCHWENKE, D. W.; GARRETT, B. C.; TRUHLAR, D. G.; MICHAEL, J. V.; SU, M. C.; SUTHERLAND, J. W. H+H₂ Thermal Reaction: A Convergence of Theory and Experiment. **Physical Review Letters**. v.91, n.6, p.63201-63204, 2003.

MOLFETTA, F. A.; BRUNI, A. T.; ROSSELLI, F. P.; SILVA, A. B. F. A partial least squares and principal component regression study of quinone compounds with trypanocidal activity. **Structural Chemistry**, v.18, p.49–57, 2007.
doi:10.1007/s11224-006-9120-3

MORGON, N. H.; COUTINHO, K., eds. Em Métodos de Química Teórica e Modelagem Molecular; Editora Livraria da Física: São Paulo, cap.9, p.491-527, 2007.

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, 1998.

MULLIKEN, R. S. Electronic Structures of Polyatomic Molecules and Valence. II. General Considerations. **Physical Review**, v.41, p.49-71, 1932.

_____.; LIU, B. Self-Consistent-Field Wave Functions of P2 and PO, and the Role of d Functions in Chemical Bonding and of s -p Hybridization in N2 and P2. **Journal of the American Chemical Society**, v.93, p.6738-6744, 1971.

NÁRAY-SZABÓ, G.; FERENCZY, G. G. Molecular Electrostatics. **Chemical Reviews**. v.95, p.829-847, 1995.

NOEDL, H. Artemisinin resistance: how can we find it? **Trends in parasitology**, v.21, n.9, 404-405, 2005.

NOEDL, H.; SE, Y.; SCHAECHER, K.; SMITH, B.; SOCHEAT, D., FUKUDA, M. Evidence of Artemisinin-Resistant Malária in Western Cambodia. **N Engl J Med**, v.359, n.24, p.2619-2620, 2008.

NYE, M. J. **Instruments and Experimentation in the History of Chemistry.** Cambridge: MIT Press, 2000.

_____. Working tools for theoretical chemistry: Polanyi, Eyring, and debates over the 'semiempirical method'. **Journal of Computational Chemistry**, v.28, n.1, p.98-108, 2007.

OLLIARO, P. L. Possible modes of action of the artemisinin-type compounds. **Trends in Parasitology**, v.17, n.3, p.122-126, 2001,

OSTLUND, N. HyperChem, Hypercube Inc., 1115 NW 4th St., Gainesville, FL, USA 32601.

PATRICK, G. L. An Introduction to Medicinal Chemistry, 2nd edn., Oxford University Press, Oxford (2001).

_____. An Introduction to Medicinal Chemistry. 3. ed. New York: Oxford, 2006. p. 293.

PIMENTEL, M.; NETO, B. B. Calibração: Uma Revisão para Químicos Analíticos. **Química Nova**, v.19, p.268-277, 1996.

PINHEIRO, J. C.; FERREIRA, M. M. C.; ROMERO, O. A. S. Antimalarial Activity of Dihydroartemisinin Derivatives against *P. falciparum* Resistant to Mefloquine: A Quantum Chemical and Multivariate Study. **Journal of Molecular Structure: THEOCHEM**, v.572, 35-44, 2001.

_____.; KIRALJ, R. Artemisinin Derivatives with Antimalarial Activity against Plasmodium falciparum Designed with the Aid of Quantum Chemical and Partial Least Squares Methods. **QSAR & Combinatorial Science**, v.22, 830, 2003.

POLITZER, P.; MURRAY, J. S. Molecular Electrostatic Potentials and Chemical Reactivity. **Reviews in Computational Chemistry**, v.2, p.273-312, 1991.

_____.; The fundamental nature and role of the electrostatic potential in atoms and molecules. **Theoretical Chemistry Accounts**. v.108, p.134-142, 2002.

POPLE, J. A.; BEVERIDGE, D. L. **Approximate Molecular Orbital Theory**. New York: McGraw-Hill, 1970. 271p.

POSNER, G. H. 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. **Journal of Medicinal Chemistry**, v.37, p.1256-1258, 1994.

_____.; CUMMINGS, J.N.; PLOYPRADITH, P.; OH, C.O. Evidence for Fe(IV).O in the Molecular Mechanism of Action of the Trioxane Antimalarial Artemisinin, **Journal of American Chemical Society**, v.117, p.5885-5886, 1995.

RANALDI, D.; HOGGAN, P. E.; CARTIER, A.; BAKA, K.; MORNARD, G.; THERY, V.; LOOS, M.; MOKRANE, A.; DILLET. **GEOMOS: Manual, Laboratoire de Chimie Théorique**, Université de Nancy I, Nancy, 2003.

RIDLEY, R. G. Malaria: to kill a parasite. **Science**, v.424, p.887-889, 2003.

ROBERTS, D. R.; HOCH, A. L.; PETERSON, M. E.; PINHEIRO, F. P. Programa multidisciplinario de vigilância de las enfermedades infecciosas en zonas colindantes con la carretera transamazonica en Brasil. IV. Estudio Entomologico. **Boletim Oficina Sanitaria Panamá**. v.91, p.379-398, 1981.

ROCHA, G. B.; FREIRE, R. O.; N.B. DA COSTA-JR; G. F. DE SÁ; A. M. SIMAS. Sparkle Model for AM1 Calculation of Lanthanide Complexes: Improved Parameters for Europium. **Inorganic Chemistry**. v.43, n.7, p. 2346-2354, 2004.

_____.;SIMAS, A. M.; STEWART, J. J. P. RM1: A reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. **Journal of Computational Chemistry**, v.27, n.10, p.1101, 2006.

ROOTHAAN, C. C. J. New Developments in Molecular Orbital Theory. **Reviews of Modern Physics**, v.23, p.69-89, 1951.

SALI, A.; BLUNDELL, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. **Journal Molecular Biology**, v.234, n.3, p.779-815, 1993

SANT'ANNA, C. M. R. Glossário de Termos Usados no Planejamento de Fármacos (Recomendações da IUPAC para 1997). **Química Nova**, v.25, p.505, 2002.

_____. Métodos de modelagem molecular para estudo e planejamento de compostos bioativos: Uma introdução. **Revista Virtual de Química**. v. 01, n. 01, p.49-57, 2009.

SANTOS, V. Uma antiga inimiga. Hebron, Abril, 2005. Disponível em: <<http://www.hebron.com.br>>. Acesso em: 17 out. 2013.

SANTOS, C. B. R.; LOBATO, C. C.; VIEIRA, J. B.; BRASIL, D. S. B.; BRITO, A. U.; MACÊDO, W. J. C.; CARVALHO, J. C. T.; PINHEIRO, J. C. Evaluation of Quantum Chemical Methods and Basis Sets Applied in the Molecular Modeling of Artemisinin. **Computational Molecular Bioscience**, v.3, n.3, p.66-79, 2013.

_____.; LOBATO, C. C.; SOUSA, M. A. C.; MACÊDO, W. J. C.; CARVALHO, J. C. T. Molecular Modeling: Origin, Fundamental Concepts and Applications Using Structure-Activity Relationship and Quantitative Structure-Activity Relationship. **Reviews in Theoretical Science**, v.2, n.2, p.91-115, 2014a.

_____.; VIEIRA, J. B.; FORMIGOSA, A. S.; COSTA, E. V. M.; PINHEIRO, M. T.; SILVA, J. O.; MACÊDO, W. J. C.; CARVALHO, J. C. T. Validation of Computational Methods Applied in Molecular Modeling of Artemisinin with Antimalarial Activity. **Journal of Computational and Theoretical Nanoscience**, v.11, n.3, p.1-9, 2014b.

_____.; VIEIRA, J. B.; LOBATO, C. C.; HAGE-MELIM, L. I. S.; SOUTO, R. N. P.; LIMA, C. S.; COSTA, E. V. M.; BRASIL, D. S. B.; MACÊDO, W. J. C.; CARVALHO, J. C. T. A SAR and QSAR Study of New Artemisinin Compounds with Antimalarial Activity. **Molecules**, v.19, n.1, p.367-399; doi:10.3390/molecules19010367, 2014c.

_____.; LOBATO, C. C.; BRAGA, F. S.; MORAIS, S. S. S.; SANTOS, C. F.; FERNANDES, C. P.; BRASIL, D. S. B.; HAGE-MELIM, L. I. S.; MACÊDO, W. J. C.; CARVALHO, J. C. T. Application of Hartree-Fock Method for modeling of bioactive

molecules using SAR and QSPR. **Computational Molecular Bioscience**, v.4, p.1-24, 2014d; Disponível em: <<http://dx.doi.org/10.4236/cmb.2014.41001>>

SHAVITT, I. The gaussian function in calculations of statistical mechanics and quantum mechanics in: ALDER, B.; FERNBACH, S.; ROTENBERG, M. (Eds.). **Methods in Computational Physics**. New York: Academic Press, 1963, p. 1-45.

SHUKLA, K. L.; GUND, T. M.; MESHNICK, S. R. Molecular modeling studies of the artemisinin (qinghaosu)-hemin interaction: Docking between the antimalarial agent and its putative receptor **Journal of Molecular Graphics**, v.13, n.4, p.215-222. 1995.

SILVERSTEIN, R. M.; WEBSTER, F. X.; KIEMLE, D. **Spectrometric Identification of Organics Compounds**. New Jersey: Jhon Wiley & Sons, 2005

SINGH, U. C.; KOLLMAN, P. A. An approach to computing electrostatic charges for molecules. **Journal of Computational Chemistry**, v.5, n.2, p.129-145, 1984

SJÖSTRÖM, M; ERICKSSON, L. Applications of statistical experimental design and pls modeling in qsar in: van de WATERBEEMD, H. **Chemometrics Methods in Molecular Design**. New York: VCH, 1996. p. 68.

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

SNEDECOR, G. W.; COCHRAN, W. G. **Statistical Methods**. New Delhi: Oxford and IBH.1967, p.381.

SOUZA, A. A.; FARIAS, R. F. **Elementos de Química Quântica**. Campinas: Átomo, 2007.

SPARTAN, Wavefunction, Inc., 18401, Von Karman Avenue, Suite 370, Irvine CA 92612 USA.

STANLEY, C.; OAKS, J. R.; VIOLAINE, S.; MITCHELL, G. W.; CHARLES, C.J., eds.; **Malária – Obstacles and opportunities A report of the Committee for the Study on Malarial Prevention and Control: Status Review and Alternative Strategies**, Division of International Health, Institute of Medicine, National Academy Press: Washington D.C., 1991.

STANTON, D. T. On the Physical Interpretation of QSAR Models. **Journal Chemical Information and Computer Science**, v.43, p.1423-1433, 2003.

STEWART, R. F. Small Gaussian Expansions of Atomic Orbitals. **The Journal of Chemical Physics**, v.50, p.2485-2495, 1969.

_____. Optimization of parameters for semiempirical methods I. Method. **Journal Computational Chemistry**, v.10, n.2, p.209-220, 1989.

_____. MOPAC 2002. Fujitsu Limited, Tokyo, Japan, 1999. Disponível em: <<http://www.cachesoftware.com/>>.

_____. MOPAC 2007: Stewart Computational Chemistry EUA, 2007. Disponível em: <<http://openmopac.net>>.

SUN E.; COHEN F. E. Computer-assisted drug discovery — a review. **Gene**, v.137, n.1, p.127-132, 1993.

SUTER, H. U.; MARIC, D. M.; WEBER, J.; THOMSON, C. Quantum Chemistry and Drug Design. **CHIMIA International Journal for Chemistry**, v.49, n.5, p.125. 1995.

TADEI, W.P.; MASCARENHAS, B.M.; PODESTÁ, M.G. Biologia de anofelinos amazônicos. VIII. Conhecimentos sobre a distribuição de espécies de Anopheles na região de Tucuruí-Marabá (Pará). **Acta Amazônica**, v.13, p.103-140, 1983.

TETKO, I. V.; GASTEIBER, 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. Virtual computational chemistry laboratory—design and description. **Journal of Computer-Aided Molecular Design**, v.19, p.453-463, 2005.

THE SCRIPPS RESEARCH INSTITUTE. Auto-Dock 4.0. Department of Molecular Biology, MB-5, La Jolla, CA, 2007.

THIEL, W. MNDO97: Organisch-chemisches Institut, Universitaet Zuerich, interthurerstrasse 190, CH-8057 Zuerich, Switzerland, 1998.

THOMSON, C.; CORY, M.; ZERNER, M. Theoretical Studies of Some New Anti-Malarial Drugs. **International Journal of Quantum Chemistry: Quantum Biology Symposium**, v.18, p.231-245, 1991.

TODESCHINI, R.; GRAMATICA, P. WHIM-3D 3.3, Milano, 1997a

_____. 3D-modelling and prediction by WHIM descriptors. Part 6. Applications of WHIM descriptors in QSAR studies. **Quantitative Structure-Activity Relationships**, v.16, p.120-125, 1997b

TODESCHINI, R.; GRAMATICA, P. New 3D molecular descriptors: the WHIM and QSAR applications. **Perspectives in Drug Discovery and Design**. v.9-11, n.26, p.355-380, 1998.

TONMUNPHEAN, S.; PARASUK, V.; KOKPOL, S.; WOLSCHANN, P.; WINGER, R.H., LIEDL, K.R.; RODE, B.M, Comparative molecular field analysis of artemisinin derivatives: Ab initio versus semiempirical optimized structures. **Journal of Computer-Aided Molecular Design**, v.12, p.397, 1998.

_____. QSAR Study of Antimalarial Activities and-Heme Binding Properties Obtained from Docking Claculatons. **Quantitative Structure-Activity Relationships**, v.19, p.475-483, 2000.

_____. Automated calculation of docking of artemisinin to heme. **Journal of Molecular Modeling**, v.7, p.26-33, 2001.

TRACY, J. W.; WEBSTER JR., L. T. Fármacos usados no tratamento das protozooses: malária. In: HARDMAN, J. G.; LIMBIRD, L. E.; MOLINOFF, P. B.; RUDDON, R. W.; GILMAN, A. G. **As bases farmacológicas da terapêutica**. 9 ed. Rio de Janeiro: MacGraw-

Hill, 1996. p. 709-724.

VCCLAB 2005: Virtual Computational Laboratory. Disponível em: <<http://www.vcclab.org>>
Acesso em: jan. 2013.

VOJTECHOVSKY, J.; CHU, K.; BERENDZEN, J.; SWEET, R. M.; SCHLICHTING, I. Crystal structures of myoglobin-ligand complexes at near-atomic resolution. **Biophysical Journal**. v.77, p.2153-2174, 1999.

WATERBEEMD, H. (Eds.). **Chemometrics Methods in Molecular Design**. New York: VCH. p.202, 1996.

WOLD, S. **Chemometrics Methods in Molecular Design**, p.202, 1996.

WORLD HEALTH ORGANIZATION. **Guidelines for the treatment of malaria**, 2010.

WULFSBERG, G. Principles of Descriptive Chemistry. **Brooks/Cole Publishing**. Monterey CA, p.23, 1987.

YOUNG, D. C. Computational Chemistry – A Practical Guide for Applying Techniques to Real-World Problems, Cytoclonal Pharmaceuticals Inc. **John Wiley & Sons, Inc.** Publication New York, 2001.

ZAHAR, A. P. Vector control operations in the African Context. **Bulletin of the World Health Organization**. v.62 (supl), p.89-100, 1984.

ZERNER, M. C. ZINDO Manual.QTP, University Of Florida, Gainesville. p.32611-8435, 1998.

ANEXOS

Evaluation of Quantum Chemical Methods and Basis Sets Applied in the Molecular Modeling of Artemisinin

Cleydson B. R. dos Santos^{1,2,3*}, Cleison C. Lobato¹, Josinete B. Vieira¹,
Davi S. B. Brasil^{1,4}, Alaan U. Brito¹, Williams J. C. Macêdo^{1,3},
José Carlos T. Carvalho^{1,2}, José C. Pinheiro³

¹Laboratory of Modelling and Computational Chemistry, Federal University of Amapá, Macapá, Brazil

²Postgraduate Program in Biotechnology and Biodiversity-Network BIONORTE, Campus Universitário Marcos Zero, Macapá, Brazil

³Laboratory of Theoretical and Computational Chemistry, Institute of Exact and Natural Sciences, Federal University of Pará, Belém, Brazil

⁴Institute of Technology, Federal University of Pará, Belém, Brazil
Email: breno@unifap.br

Received July 28, 2013; revised August 29, 2013; accepted September 7, 2013

Copyright © 2013 Cleydson B. R. dos Santos *et al.* This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

In this paper, we evaluate semiempirical methods (AM1, PM3, and ZINDO), HF and DFT (B3LYP) in different basis sets to determine which method best describes the sign and magnitude of the geometrical parameters of artemisinin in the region of the endoperoxide ring compared to crystallographic data. We also classify these methods using statistical analysis. The results of PCA were based on three main components, explaining 98.0539% of the total variance, for the geometrical parameters C3O13, O1O2C3, O13C12C12a, and O2C3O13C12. The DFT method (B3LYP) corresponded well with the experimental data in the hierarchical cluster analysis (HCA). The experimental and theoretical angles were analyzed by simple linear regression, and statistical parameters (correlation coefficients, significance, and predictability) were evaluated to determine the accuracy of the calculations. The statistical analysis exhibited a good correlation and high predictive power for the DFT (B3LYP) method in the 6-31G** basis set.

Keywords: Artemisinin; Molecular Modeling; Quantum Chemical Methods; Statistical Analysis; B3LYP/6-31G**

1. Introduction

Artemisinin (or Qinghaosu, QHS, **Figure 1**) represents the most relevant advance in the treatment of malarial disease for the last 20 years [1]. Artemisinin is a sesquiterpene lactone with an endoperoxide group, which has been used in traditional Chinese medicine for many centuries as a natural product for fever and malarial treatment. This drug was isolated by Chinese chemists in the early 1970s from the ancient *Artemisia annua* L. Nowadays, artemisinin and derivatives are widely used around the world because of their potent antimalarial activity, fast action, and low toxicity. As a result, artemisinin and its derivatives have become recognized as a new generation of antimalarial drugs [2].

Many studies on the rational design of new antimalarial drugs have been performed using molecular modeling.

However, complex molecular systems containing external and internal transition atoms, proteins, polymers, or compounds with a higher molecular weight overestimate the ability of these methods and basis sets to obtain molecular properties, which can lead to inaccurate results when compared with experimental data. Therefore, an evaluation of these methods and basis sets will be strongly dependent on the system under consideration [3].

Costa *et al.* studied the interaction between heme and artemisinin using the PM3 method, which exhibited a potential energy barrier for the relative rotation of the artemisinin-heme complex being studied both *in vacuo* and partially solvated. The authors observed that the inclusion of water molecules did not significantly affect the stability of the heme-artemisinin complex [4].

Leite *et al.* conducted studies of 18 natural compounds from Brazilian flora, which possess a peroxide group and

*Corresponding author.

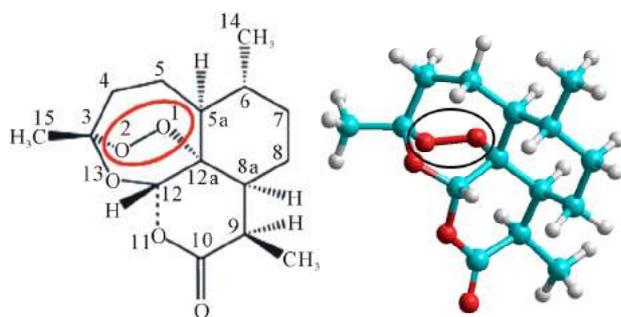


Figure 1. Artemisinin (structure) and region essential for the expression of biological activity (pharmacophore).

are assumed to act in heme protein, leading to a reduction in peroxide binding and the production of radicals that can kill the etiological agents of malaria (*Plasmodium falciparum* strains). These findings motivated the researchers to study the interactions among 18 natural peroxides. They initially performed a conformational search using the MM3 method for each molecule, and the most stable conformers were optimized by the PM3 (tm) method. The interactions between the peroxide-heme groups were then evaluated, and the results indicated that four of the compounds may exhibit desirable antimalarial activity [5].

Recent studies on 51 peroxides were conducted to identify correlations between *in silico* parameters and experimental data for identifying new antimalarial agents from natural sources. The interaction of the heme group was studied by molecular docking refinement followed by conformational analysis using semiempirical parametric method 6 (PM6). The results indicated that two of these compounds are promising antimalarials [6].

Artemisinin derivatives with antimalarial activity against *Plasmodium falciparum*, which is resistant to mefloquine, were studied using quantum chemical methods (HF/6-31G*) and the partial least-squares (PLS) method. Three main components explained 89.55% of the total variance, with $Q^2 = 0.83$ and $R^2 = 0.92$. From a set of 10 proposed artemisinin derivatives (artemisinin derivatives with unknown antimalarial activity against *Plasmodium falciparum*), a novel compound was produced with superior antimalarial activity compared to the compounds previously described in the literature [7].

Recently, Cristino *et al.* [8] used the B3LYP/6-31G* method to model artemisinin and 19,10-substituted deoxyartemisinin derivatives, with different degrees of activity against the *Plasmodium falciparum* D-6 strains of Sierra Leone. Chemometric methods (PCA, HCA, KNN, SIMCA, and SDA) were employed to reduce the dimensionality and to determine which subset of descriptors is responsible for the classification between more and less active agents.

Figueiredo *et al.* [9] conducted studies using the

B3LYP/6-31G* method for antimalarial compounds against *Plasmodium falciparum* K1. These studies led to multivariate models for artemisinin derivatives and series of dispiro-1,2,4-trioxolanes. The application of these models has enabled the prediction of activity for compounds designed without known biological activity. Moreover, a new series of antimalarial compounds is currently in the study phase.

Araújo *et al.* [10] used density functional theory (6-31G*) to verify the performance of a basis set in reproducing experimental data, particularly geometrical parameters, and to calculate the interaction energies, electronic states, and geometrical arrangements for complexes composed of a heme group and artemisinin. The results demonstrated that the interaction between artemisinin and the heme group occurs at long distances through a complex in which the iron atom of the heme group retains its electronic characteristics, with the quintet state being the most stable. These results suggest that the interaction between artemisinin and heme is thermodynamically favorable.

In this paper, we propose to identify the best method and basis set for molecular modeling of the pharmacophoric group of artemisinin and its derivatives. There is currently a diversity of methods and basis sets that can be applied to reproduce experimental data and to elucidate the biological significance of this compound. An identification of the best theoretical method for data acquisition is critical in achieving credible *in silico* results with respect to biological action. Artemisinin was modeled using five methods and seven basis sets (6-31G, 6-31G*, 6-31G**, 3-21G, 3-21G*, 3-21G**, and 6-311G) based on the properties of the endoperoxide group present in artemisinin, which is responsible for the biological activity, and the results were evaluated via principal component analysis (PCA), hierarchical cluster analysis (HCA), and statistical analysis using simple linear regression.

2. Experimental

2.1. Molecular Modeling of Artemisinin

The artemisinin compound was constructed as follows: initially, the structure of artemisinin was established with the Gauss View 3.0 program [11] and optimized using different methods and basis sets: semiempirical (AM1, PM3, and ZINDO), Hartree-Fock (HF/6-31G, HF/6-31G*, HF/6-31G**, HF/3-21G, HF/3-21G*, HF/3-21G**, and HF/6-311G), and DFT (B3LYP/6-31G, B3LYP/6-31G*, B3LYP/6-31G**, and B3LYP/3-21G) implemented with the Gaussian 03 program [12]. These calculations were performed to determine which methods and basis sets provide the best compromise between computational time and accuracy compared to the experimental data [13]. The experimental structure of artemisinin was re-

moved from the Database Cambridge Structural CSD with REFCODES: QNGHSU10, crystallographic R factor 3.6 [14]. The numbering of atoms adopted in this study is shown in **Figure 1** (artemisinin).

When measurements are made on a number of objects, the results are typically organized in a data matrix. The measures in this study (geometrical parameters) were organized in rows, and the objects (quantum chemical methods and basis sets) were organized in columns. Statistical analysis was conducted with the Piroutte 3.10 and Statistica 6.2 programs [15,16].

The statistical analysis of the geometrical parameters was based on the studies of Silva *et al.*, in which density functional theory (DFT) calculations (B3PW91/DGD-ZVP) were employed to determine ^{13}C and ^1H nuclear magnetic resonance (NMR) chemical shifts for the two dihydrochalcones: 3,4,5-tetramethoxydihydrochalcone and 2,3,4,4-trimethoxydihydrochalcone. The theoretical and experimental NMR data were analyzed by simple linear regression, and the most relevant parameters were selected. In addition, other statistical parameters (correlation coefficients, significance, and predictability) were employed to verify the accuracy of the calculations. The statistical analysis indicated a good correlation between the NMR experiment results and the theoretical data, with high predictive power [17].

2.2. Principal Component Analysis (PCA)

PCA was performed with autoscaled processing for a maximum of three factors (3PCS), using the “leave-one-out validation method and cross-validation” procedure. The data matrix was constructed from the combinations of methods and basis sets, resulting in a 15×18 matrix. Each column was related to five methods and seven basis sets, and one column was related to the experimental geometrical parameters [13]. Each row represents 18 geometrical parameters of the 1,2,13-trioxane ring (bond lengths, bond angles, and torsion angles). The final PCA results led to the selection of a small number of geometrical parameters that were most strongly related to the dependent variable, which was the standard deviation of the methods and basis sets in this case.

2.3. Hierarchical Cluster Analysis (HCA)

Similar to the PCA results, the HCA results are qualitative and are arranged in the form of a dendrogram to display the methods studied and the variables (geometrical parameters of artemisinin) in a two-dimensional space. The results illustrate the combinations or divisions made in each successive stage of analysis. The samples (methods and basis sets) are represented by the bottom branch of the dendrogram. The similarity between agglomerates is given by the length of each branch such that methods

and basis sets with low levels of similarity have long branches and methods and basis sets with high similarity have short branches [18]. In HCA, the distance between these variables is calculated and transformed into a similarity matrix S . A hierarchical cluster analysis aims to display data in a manner that accentuates the natural groupings and patterns. Statistical analysis was required in this study to group similar methods and basis sets in their respective categories. HCA is a statistical method that was developed for this purpose.

2.4. Statistical Analysis

Statistical analysis via simple linear regression can correlate data from more relevant parameters to estimate and predict values using a model built with a full dataset and actual values of y_i (r), the explained variance (R_A^2 , *i.e.*, adjusted R^2), Fisher ratio values (F), and the standard error of estimation (SEE). The equations were also tested for their predictive power using a cross-validation procedure. Cross-validation is a practical and reliable method for verifying the predictive power. In the so-called “leave-one-out” approach, a number of models are developed with one sample omitted in each step. After developing each model, the omitted data are predicted and the differences between the actual and predicted y values are calculated. The sum of squares of these differences is computed, and finally, the performance of the model (its predictive ability) can be given by the prediction residual error sum of squares (PRESS) and standard deviation of the cross-validation (S_{PRESS}), as shown in Equations (1) and (2) [19]:

$$\text{PRESS} = \sum_{i=1}^n (y_i - \hat{y})^2 \quad (1)$$

$$S_{\text{PRESS}} = \frac{\sqrt{\text{PRESS}}}{n - k - 1} \quad (2)$$

where y_i is the experimental value, \hat{y} is the predicted value, n is the number of samples used to build the model, and k is the number of geometrical parameters. The predictive ability of the model can also be quantified in terms of the cross-validated correlation coefficient (Q^2), which is defined as [19]

$$Q^2 = 1.0 - \frac{\text{PRESS}}{\sum_{i=1}^n (y_i - \bar{y})^2}; \text{ where, } \bar{y} = y_{\text{mean}} \quad (3)$$

In addition, other statistical parameters are available to verify the accuracy of the calculation, but none of these parameters are fully satisfactory if taken alone. For each system, we present parameters a and b from linear regression $\delta_{\text{calcd}} = a + b\delta_{\text{exp1}}$; the mean absolute error (MAE) defined as $\text{MAE} = \sum_n |\delta_{\text{calcd}} - \delta_{\text{exp1}}| / n$, and the corrected mean absolute error, CMAE [20], defined as

$CMAE = \sum_n |\delta_{corr} - \delta_{expt}| / n$, where $\delta_{corr} = (\delta_{calcd} - a) / b$ to correct for systematic errors.

3. Results and Discussion

3.1. Determination of the Theoretical Geometrical Parameters of Artemisinin

We determined the geometrical parameters for the 1,2,13-trioxane ring of artemisinin (bond length, bond angle, and torsion angle of atoms in this ring), as shown in **Table 1**. **Table 1** illustrates that for the DFT method, all four basis sets (B3LYP/6-31G, B3LYP/6-31G*, B3LYP/6-31G**, and B3LYP/3-21G) can accurately describe all of the structural parameters with respect to their magnitude and sign when compared with the experimental values. Meanwhile, the semiempirical (AM1, PM3, and ZINDO) and Hartree-Fock (HF/6-31G, HF/6-31G*, HF/6-31G**, HF/3-21G, HF/3-21G*, HF/3-21G**, and HF/6-311G) methods exhibited standard deviations of 4.776, 8.388, and 4.372 and 1.663, 2.484, 1.762, 1.722, 1.714, 1.797, and 1.658, respectively. By comparing these methods with the DFT method, we find that all of the basis sets (B3LYP/6-31G, B3LYP/6-31G*, and B3LYP/6-31G**) have low standard deviations in relation to the semiempirical and Hartree-Fock methods at 0.843 (B3LYP/6-31G), 1.227 (B3LYP/6-31G*), and 1.103 (B3LYP/6-31G**). The variation was ± 0.384 be-

tween B3LYP/6-31G and B3LYP/6-31G*, ± 0.260 between B3LYP/6-31G and B3LYP/6-31G**, and ± 0.124 between B3LYP/6-31G* and B3LYP/6-31G**.

3.2. PCA and HCA

The advantage of the PCA and HCA methods in this study is that all structural parameters are considered simultaneously and all of their correlations are considered.

Figure 2 presents the PC1-PC2 scores for the five methods and seven basis sets and one score related to the experimental geometrical parameters of the 1,2,13-trioxane ring. The methods are divided into two groups according to PC2, where the semiempirical and Hartree-Fock methods are located at the bottom and the experimental data and the DFT/B3LYP method and basis sets (B3LYP/6-31G, B3LYP/6-31G*, B3LYP/6-31G**, and B3LYP/3-21G) are located at the top.

Figure 3 presents the PC1-PC2 loading for the four most important geometrical parameters related to the trioxane ring, namely, C3O13, O1O2C3, O13C12C12a, and O2C3O13C12. These geometrical parameters are responsible for the separation of the methods and basis sets into three groups, namely, semiempirical, Hartree-Fock, and DFT methods, as shown in **Figure 2**. The four geometrical parameters related to the trioxane ring are identified by atoms C3O13, which provides the interplanar distance between these two atoms (bond length),

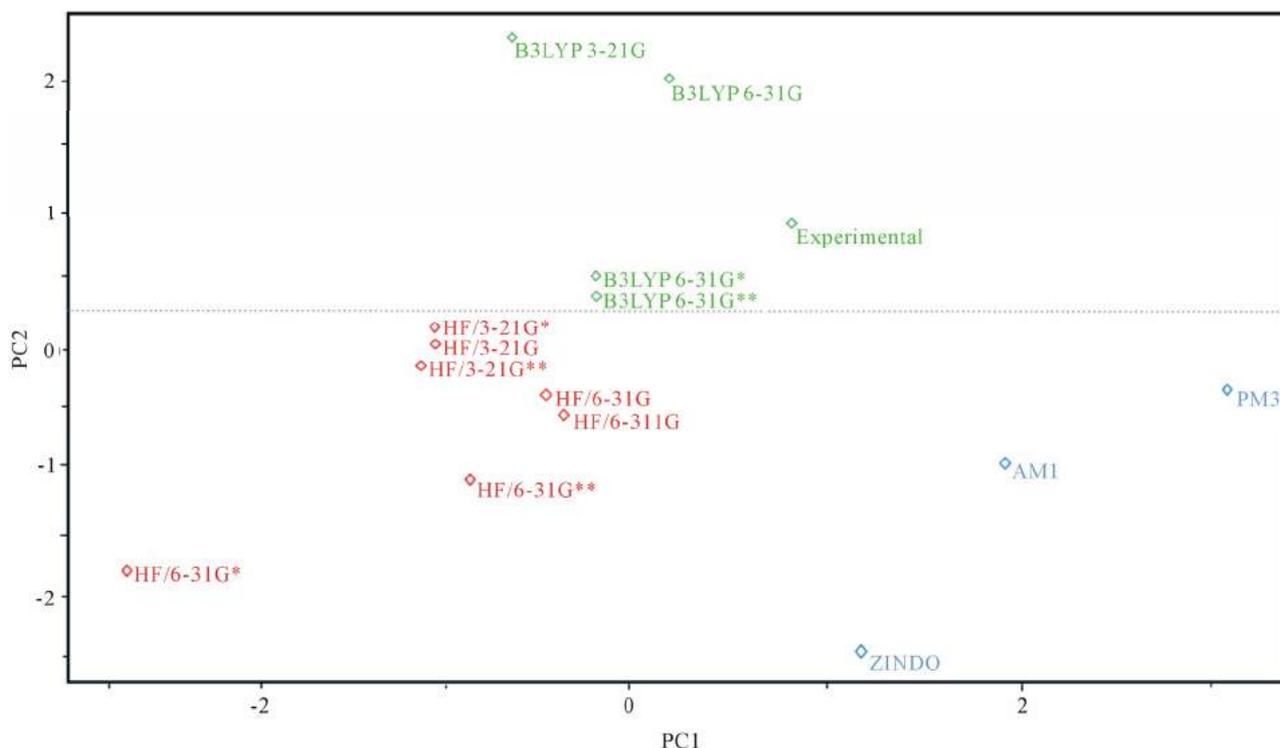


Figure 2. Plot of PC1-PC2 scores for the five methods and seven basis sets and the experimental geometrical parameters of the 1,2,13-trioxane ring.

Table 1. Theoretical and experimental parameters for the 1,2,13-trioxane ring in artemisinin.

Parameters ^a	Semiempirical			Hartree-Fock/HF					DFT/B3LYP			EXP [13]			
	AM1	PM3	ZINDO	6-31G	6-31G*	6-31G**	3-21G	3-21G*	3-21G**	6-311G	6-31G		6-31G*	6-31G** ^b	3-21G
Bond length/Å															
O1O2	1.288	1.544	1.237	1.447	1.391	1.390	1.461	1.461	1.462	1.429	1.524	1.459	1.459	1.524	1.469
O2C3	1.447	1.403	1.400	1.435	1.393	1.396	1.440	1.440	1.439	1.432	1.452	1.413	1.414	1.455	1.416
C3O13	1.427	1.428	1.396	1.435	1.388	1.408	1.436	1.436	1.435	1.434	1.473	1.441	1.441	1.473	1.445
O13C12	1.416	1.403	1.392	1.403	1.400	1.376	1.407	1.407	1.407	1.401	1.425	1.395	1.396	1.430	1.379
C12C12a	1.537	1.555	1.513	1.533	1.533	1.532	1.529	1.529	1.530	1.530	1.438	1.539	1.539	1.535	1.523
C12aO1	1.468	1.427	1.416	1.469	1.429	1.429	1.477	1.477	1.477	1.438	1.499	1.455	1.455	1.504	1.461
Bond angle/°															
O1O2C3	112.530	110.340	114.310	108.800	106.100	109.460	107.100	107.080	107.060	109.210	107.300	108.280	108.280	105.590	108.100
O2C3O13	103.600	104.810	105.370	106.760	110.800	107.800	107.270	107.285	107.300	106.670	107.730	108.490	108.490	108.220	106.600
C3O13C12	115.480	116.010	115.843	117.300	112.800	115.300	115.670	115.680	115.710	116.960	114.990	114.080	114.060	113.200	114.200
O13C12C12a	113.510	115.200	113.270	112.280	108.700	112.300	112.080	112.080	112.030	112.360	113.640	113.250	113.240	113.300	114.500
C12C12aO1	111.070	113.180	107.290	110.910	110.500	110.545	111.570	111.600	111.600	110.760	111.740	111.290	111.280	112.410	110.700
C12aO1O2	113.740	112.290	118.380	113.240	112.700	112.700	111.290	111.290	111.290	113.360	111.400	111.600	111.590	109.620	111.200
Torsion angle/°															
O1O2C3O13	-77.800	-73.310	-70.403	-71.840	-73.369	-73.400	-74.670	-74.700	-74.690	-71.940	-73.460	-73.900	-73.910	-76.610	-75.500
O2C3O13C12	42.070	52.700	36.370	33.390	31.034	31.100	32.300	32.360	32.180	33.010	34.970	32.800	32.780	33.750	36.000
C3O13C12C12a	11.400	2.811	17.420	25.320	27.432	27.400	28.290	28.190	28.330	25.380	26.260	27.460	25.500	29.060	25.300
O13C12C12aO1	-41.770	-40.510	-46.610	-49.410	-50.100	-50.143	-50.860	-50.770	-50.700	-49.470	-51.200	-51.270	-51.340	-52.190	-51.300
C12C12aO1O2	12.050	19.940	18.110	12.510	10.900	10.924	9.989	9.940	9.750	12.480	12.740	11.730	11.780	9.060	12.700
C12aO1O2C3	47.050	35.630	40.130	46.700	48.700	48.674	50.330	50.350	50.530	46.870	46.900	47.850	47.830	51.060	47.800
Standard Deviation	4.776	8.388	4.372	1.663	2.484	1.762	1.722	1.714	1.797	1.658	0.843	1.227	1.103	1.915	-

^aThe atoms are numbered according to Figure 1; ^bValence basis set separately validated for calculating the molecular properties.

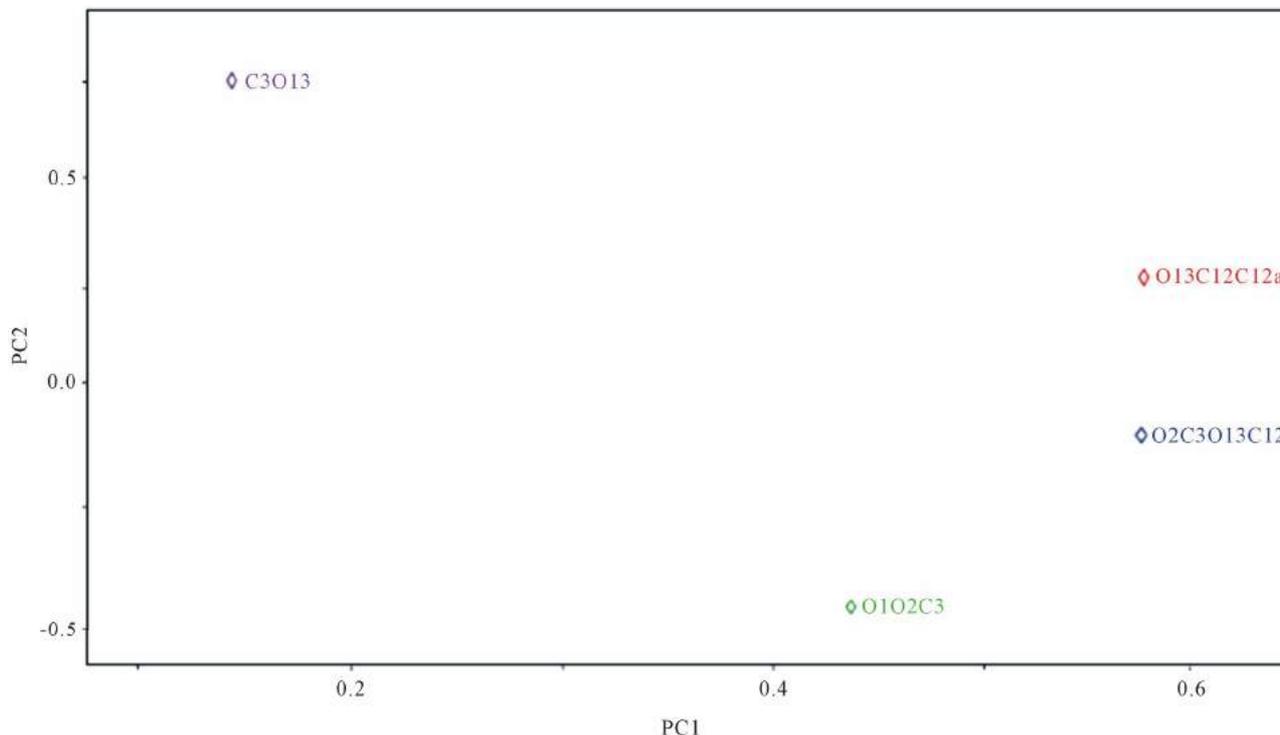


Figure 3. Plot of PC1-PC2 loadings using the four geometrical parameters selected by PCA.

O1O2C3 and O13C12C12a, corresponding to the bond angle between three atoms, and O2C3O13C12, which is related to the dihedral angle or torsion. The semiempirical (AM1, PM3, and ZINDO) and Hartree-Fock methods (HF/6-31G, HF/6-31G*, HF/6-31G**, HF/3-21G, HF/3-21G*, HF/3-21G**, and HF/6-311G) have a greater contribution from geometrical parameters O1O2C3 and O2C3O13C12, which are responsible for the low scores of these methods and basis sets. In contrast, the DFT method (B3LYP/6-31G, B3LYP/6-31G*, B3LYP/3-21G**, and B3LYP/6-31G) has a high contribution from the geometrical parameters C3O13 and O13C12C12a, which are responsible for the higher scores of these methods and basis sets. As shown in **Figure 3**, the given method and basis set have larger values as the contribution of geometrical parameters O3C13 and O13C12C12a in the second main component increase, resulting in improved scores and a closer agreement with the crystallographic experimental data. The geometrical parameters O1O2C3 and O2C3O13C12 contribute to a lesser degree, with a negative weight in PC2, demonstrating that the methods and basis sets generally have higher values for these geometrical parameters.

The values of the geometrical parameters selected based on PCA, their standard deviation, and the Pearson correlation matrix are shown in **Table 2**, which presents the correlation matrix between the geometrical parameters and their standard deviations. The correlations between geometrical parameters are less than or equal to

0.639, whereas the correlations between the geometrical parameters and standard deviations are less than or equal to 0.863. The geometrical parameters selected using PCA represent characteristics required to evaluate the methods and basis sets applied in the molecular modeling of artemisinin and its derivatives.

The results of the selection model in **Table 3** illustrate that the model was built with three main components (3PCs), where the first principal component (PC1) describes 28.3156% of the total information, the second principal component (PC2) describes 21.0067%, and the third (PC3) describes 5.5879%. Furthermore, the table illustrates that PC1 contains 50.5635% of the original data, the first two (PC1 + PC2) contain 88.0755%, and the first three (PC1 + PC2 + PC3) can explain 98.0539% of the total information, losing only 1.9461% of the original information. In the same table, descriptors O13C12C12a (0.6125) and O2C3O13C12 (0.6152) appear to be the main contributors to PC1, whereas descriptors C3O13 (0.7947) and O13C12C12a (0.3424) are the main contributors to PC2.

Recently, Santos *et al.* [21] validated computational methods applied in the molecular modeling of artemisinin, proposing a combination of chemical quantum methods and statistical analysis to study the geometrical parameters of artemisinin in the region of the endoperoxide ring (1,2,13-trioxane). The PCA results indicated that their model was built with three main components (3PCs), explaining 97.0861% of the total variance.

Table 2. Geometrical parameters selected by PCA, standard deviations, and the Pearson correlation matrix.

METHODS	Bond length/Å	Bond angle/°	Bond angle/°	Torsion angle/°	Standard deviation
	C3O13	O1O2C3	O13C12C12a	O2C3O13C12	
AM1	1.427	112.530	113.510	42.070	4.776
PM3	1.428	110.340	115.200	52.700	8.388
ZINDO	1.396	114.310	113.270	36.370	4.372
HF/6-31G	1.435	108.800	112.280	33.390	1.663
HF/6-31G*	1.388	106.100	108.700	31.034	2.484
HF/6-31G**	1.408	109.460	112.300	31.100	1.762
HF/3-21G	1.436	107.100	112.080	32.300	1.722
HF/3-21G*	1.435	107.080	112.080	32.360	1.714
HF/3-21G**	1.435	107.060	112.030	32.180	1.797
HF/6-311G	1.434	109.210	112.360	33.010	1.658
B3LYP/6-31G	1.473	107.300	113.640	34.970	0.843
B3LYP/6-31G*	1.441	108.280	113.250	32.800	1.227
B3LYP/6-31G**	1.441	108.280	113.240	32.780	1.103
B3LYP/3-21G	1.473	105.590	113.300	33.750	1.915
Experimental	1.445	108.100	114.500	36.000	0.000
C3O13		-0.453	0.522	-0.002	-0.360
O1O2C3			0.360	0.483	0.564
O13C12C12a				0.639	0.281
O2C3O13C12					0.863

Table 3. PCA of selection model for the methods and basis sets.

	Principal component		
	PC1	PC2	PC3
Variance/%	28.3156	21.0067	5.5879
Cumulative Variance/%	50.5635	88.0755	98.0539
Geometrical Parameters	Contribution		
	PC1	PC2	
C3O13	0.0806	0.7947	
O1O2C3	0.4898	-0.4989	
O13C12C12a	0.6125	0.3424	
O2C3O13C12	0.6152	-0.0478	

The main components can be written as linear combinations of the four geometrical parameters selected by PCA. The mathematical expressions for PC1 and PC2 are shown:

$$\begin{aligned} \text{PC1} = & 0.0806(\text{C3O13}) + 0.4898(\text{O1O2C3}) \\ & + 0.6125(\text{O13C12C12a}) \\ & + 0.6152(\text{O2C3O13C12}) \end{aligned} \quad (4)$$

$$\begin{aligned} \text{PC2} = & 0.7947(\text{C3O13}) - 0.4989(\text{O1O2C3}) \\ & + 0.3424(\text{O13C12C12a}) \\ & - 0.0478(\text{O2C3O13C12}) \end{aligned} \quad (5)$$

The geometrical parameters C3O13, O1O2C3, O13C12C12a, and O2C3O13C12 are of great importance in this study because according to the proposal made by Jefford *et al.*, the iron in heme attacks artemisinin at position O1 and generates a free radical at position O2. After the ligation at C3-C4 is broken, a heme carbon radical is generated at C4 [22]. This free radical at C4 has been suggested as an important substance in antimalarial activity [23]. The study of molecular docking between artemisinin and its receptor, heme, conducted by Tonmunpuean *et al.* also indicated that heme iron preferentially interacts with O1 rather than O2 [24]. This phenomenon

leads to the importance of geometry parameters C3O13 and O1O2C3, which were selected to be associated with the suggested mechanism in this model.

Table 4 presents the geometrical parameters selected by PCA, the methods, the basis sets, and the variations in the geometrical parameters with respect to the experimental data (Δ and $\Delta\%$). The semiempirical and Hartree-Fock methods do not exhibit a good agreement between the theoretical and experimental values for the bond angles and torsion angles, particularly for the angles formed by atoms O13C12C12a and O2C3O13C12, respectively.

The semiempirical methods (AM1, PM3, and ZINDO) present deviations of $\Delta = 0.990^\circ$ ($\Delta\% = 0.864$), $\Delta = -0.700^\circ$ ($\Delta\% = -0.611$), and $\Delta = 1.230^\circ$ ($\Delta\% = 1.074$), respectively, in relation to the angle O13C12C12a. For angle O2C3O13C12, deviations of $\Delta = -6.070^\circ$ ($\Delta\% = -16.861$), $\Delta = -16.700^\circ$ ($\Delta\% = -46.388$), and $\Delta = -0.370^\circ$ ($\Delta\% = -1.027$) were observed. For the Hartree-Fock methods (HF/6-31G, HF/6-31G*, HF/6-31G**, HF/3-21G, HF/3-21G*, HF/3-21G**, and HF/6-311G), the angle formed by atoms O13C12C12a have deviations of $\Delta = 2.220^\circ$ ($\Delta\% = 1.938$), $\Delta = 5.800^\circ$ ($\Delta\% = 5.065$) and $\Delta = 2.200^\circ$ ($\Delta\% = 1.921$), $\Delta = 2.420^\circ$ ($\Delta\% = 2.113$), $\Delta = 2.420^\circ$ ($\Delta\% = 2.113$), $\Delta = 2.470^\circ$ ($\Delta\% = 2.157$), and $\Delta = 2.140^\circ$ ($\Delta\% = 1.868$), respectively. For the angle O2C3O13C12, deviations of $\Delta = 2.610^\circ$ ($\Delta\% = 7.250$), $\Delta = 4.966^\circ$ ($\Delta\% = 13.794$), $\Delta = 4.900^\circ$ ($\Delta\% = 13.611$), $\Delta = 3.700^\circ$ ($\Delta\% = 10.277$), $\Delta = 3.640^\circ$ ($\Delta\% = 10.111$), $\Delta = 3.820^\circ$ ($\Delta\% = 10.611$), and $\Delta = 2.990^\circ$ ($\Delta\% = 8.305$) were obtained. For the DFT/B3LYP method, the four levels (6-31G, 6-31G*, 6-31G**, and 3-21G) exhibited excellent results for the bond angle O1O2C3, with deviations of $\Delta = -0.800^\circ$ ($\Delta\% = -0.740$) for DFT/B3LYP 6-31G, $\Delta = -0.180^\circ$ ($\Delta\% = -0.166$) for DFT/B3LYP 6-31G*, $\Delta = 0.180^\circ$ ($\Delta\% = -0.166$) for DFT/B3LYP 6-31G**, and $\Delta = 2.510^\circ$ ($\Delta\% = 2.321$) for DFT/B3LYP 3-21G, as shown in **Table 4**.

As also shown in **Table 4**, by accounting for the features of the basis sets B3LYP/6-31G* and B3LYP/6-31G**, excellent results were obtained in relation to the bond length C3O13 and bond angle O1O2C3, which exhibited good agreement with the experimental values reported in the literature. Thus, for the 6-31G* and 6-31G** bases, these parameters are close to the crystallographic experimental data in the region of the endoperoxide ring of artemisinin.

HCA was used to explore and more appropriately group the methods and basis sets according to their similarities. A dendrogram was obtained with autoscaled processing based on the Euclidean distance and incremental method, as shown in **Figure 4**. This figure illustrates the grouping of the three classes: semiempirical (AM1, PM3, and ZINDO), Hartree-Fock (HF/6-31G,

HF/6-31G*, HF/6-31G**, HF/3-21G, HF/3-21G*, HF/3-21G**, and HF/6-311G), and DFT (B3LYP/6-31G, B3LYP/6-31G*, B3LYP/6-31G**, and B3LYP/3-21G). The semiempirical method has long branches, indicating a low similarity with the experimental data. However, the Hartree-Fock method exhibits similarity between the basis sets, indicated by short branches, for the HF/3-21G**, HF/3-21G*, and HF/3-21G bases. Cardoso *et al.* [25] studied artemisinin and some of its derivatives with activity against D-6 strains of *Plasmodium falciparum* using the HF/3-21G method. To verify the reliability of the geometry obtained, Cardoso *et al.* compared the structural parameters of the artemisinin trioxane ring with theoretical and experimental values from the literature. Ferreira *et al.* [26] also studied artemisinin and 18 derivatives with antimalarial activity against W-2 strains of *Plasmodium falciparum* through quantum chemistry and multivariate analysis. The geometry optimization of structures was realized using the Hartree-Fock method and the 3-21G** basis set.

The DFT/B3LYP method exhibits high similarity between the basis sets, particularly for the B3LYP/6-31G* and B3LYP/6-31G** basis sets, as indicated by the short branches between them, shown in **Figure 4**.

The results of the theoretical methods and experimental data exhibited a distribution similar to that obtained with PCA. Thus, HCA confirmed the results achieved by PCA.

3.3. Statistical Analysis of Parameters

In this step, regression models with high values of r (%), R_A^2 (%), Q^2 , and F (a statistic assessing the overall significance) and low values of MAE, CMAE, SEE, PRESS, and S_{PRESS} were selected.

The most relevant statistical parameters are given in **Table 5**. When comparing the semiempirical methods, the best values were found for ZINDO ($r = 99.7039\%$, $R_A^2 = 99.3496\%$, $Q^2 = 99.3909$, $F = 1681.28$, $\text{SEE} = 5.3983$, $\text{MAE} = 4.335$, $\text{CMAE} = 4.0998$, $\text{PRESS} = 305,773$, and $S_{\text{PRESS}} = 1.7486$). Among the basis sets for the Hartree-Fock method, the best values were observed for HF/6-31G ($r = 99.9609\%$, $R_A^2 = 99.9142\%$, $Q^2 = 99.9118\%$, $F = 12811.33$, $\text{SEE} = 1.9641$, $\text{MAE} = 1.491$, $\text{CMAE} = 1.4715$, $\text{PRESS} = 44.286$, $S_{\text{PRESS}} = 0.6654$), HF/6-311G ($r = 99.9606\%$, $R_A^2 = 99.9134\%$, $Q^2 = 99.9124\%$, $F = 12705.14$, $\text{SEE} = 1.9733$, $\text{MAE} = 1.492$, $\text{CMAE} = 1.4626$, $\text{PRESS} = 43.985$, $S_{\text{PRESS}} = 0.6632$), and HF/3-21G* ($r = 99.9532\%$, $R_A^2 = 99.8970\%$, $Q^2 = 99.9063\%$, $F = 10679.82$, $\text{SEE} = 2.1638$, $\text{MAE} = 1.647$, $\text{CMAE} = 1.6311$, $\text{PRESS} = 47.017$, $S_{\text{PRESS}} = 0.6856$).

The DFT method using B3LYP, with the valence-separate basis sets B3LYP/6-31G, B3LYP/6-31G*, and B3LYP/6-31G**, achieved the best results among all of the methods and basis sets studied herein (semiempirical

Table 4. Geometrical parameters selected by PCA, the methods, the basis sets, and the variation in geometrical parameters with respect to the experimental data (Δ and $\Delta\%$).

Parameters	Semiempirical													
	AMI	Δ_{AMI}	$\Delta\%$	PM3	Δ_{PM3}	$\Delta\%$	ZINDO	Δ_{ZINDO}	$\Delta\%$	Exp				
Bond length/Å														
C3O13	1.427	0.018	1.245	1.428	0.017	1.176	1.396	0.049	3.391	1.445				
Bond angle/°														
O1O2C3	112.530	-4.430	-4.098	110.340	-2.240	-2.072	114.310	-6.210	-5.744	108.100				
O13C12C12a	113.510	0.990	0.864	115.200	-0.700	-0.611	113.270	1.230	1.074	114.500				
Torsion angle/°														
O2C3O13C12	42.070	-6.070	-16.861	52.700	-16.700	-46.388	36.370	-0.370	-1.027	36.000				
Parameters	<i>ab initio</i> /Hartree-Fock/HF													
	6-31G	Δ_{6-31G}	$\Delta\%$	6-31G*	Δ_{6-31G^*}	$\Delta\%$	6-31G**	$\Delta_{6-31G^{**}}$	$\Delta\%$	3-21G	Δ_{3-21G}	$\Delta\%$	Exp	
Bond length/Å														
C3O13	1.435	0.010	0.692	1.388	0.057	3.944	1.408	0.037	2.560	1.436	0.010	0.692	1.445	
Bond angle/°														
O1O2C3	108.800	-0.700	-0.647	106.100	2.000	1.850	109.460	-1.360	-1.258	107.100	1.000	0.952	108.100	
O13C12C12a	112.280	2.220	1.938	108.700	5.800	5.065	112.300	2.200	1.921	112.080	2.420	2.113	114.500	
Torsion angle/°														
O2C3O13C12	33.390	2.610	7.250	31.034	4.966	13.794	31.100	4.900	13.611	32.300	3.700	10.277	36.000	
Parameters	<i>ab initio</i> /Hartree-Fock/HF													
	3-21G*	Δ_{3-21G^*}	$\Delta\%$	3-21G**	$\Delta_{3-21G^{**}}$	$\Delta\%$	6-311G	Δ_{6-311G}	$\Delta\%$	Exp				
Bond length/Å														
C3O13	1.435	0.010	0.692	1.435	0.010	0.692	1.434	0.011	0.761	1.445				
Bond angle/°														
O1O2C3	107.080	1.020	0.943	107.060	1.040	0.943	109.210	-1.110	-1.026	108.100				
O13C12C12a	112.080	2.420	2.113	112.030	2.470	2.157	112.360	2.140	1.868	114.500				
Torsion angle/°														
O2C3O13C12	32.360	3.640	10.111	32.180	3.820	10.611	33.010	2.990	8.305	36.000				
Parameters	DFT/B3LYP													
	6-31G	Δ_{6-31G}	$\Delta\%$	6-31G*	Δ_{6-31G^*}	$\Delta\%$	6-31G**	$\Delta_{6-31G^{**}}$	$\Delta\%$	3-21G	Δ_{3-21G}	$\Delta\%$	Exp	
Bond length/Å														
C3O13	1.473	-0.028	-1.937	1.441	0.004	0.276	1.441	0.004	0.276	1.473	-0.028	-1.937	1.445	
Bond angle/°														
O1O2C3	107.300	-0.800	-0.740	108.280	-0.180	-0.166	108.280	-0.180	-0.166	105.590	2.510	2.321	108.100	
O13C12C12a	113.640	-0.860	-0.751	113.250	1.250	1.091	113.240	1.260	1.100	113.300	1.200	1.048	114.500	
Torsion angle/°														
O2C3O13C12	34.970	1.030	2.861	32.800	3.200	8.888	32.780	3.220	8.944	33.750	2.250	6.250	36.000	

Δ = theoretical - experimental; $\Delta\%$ = $\Delta \times 100/\text{experimental}$.

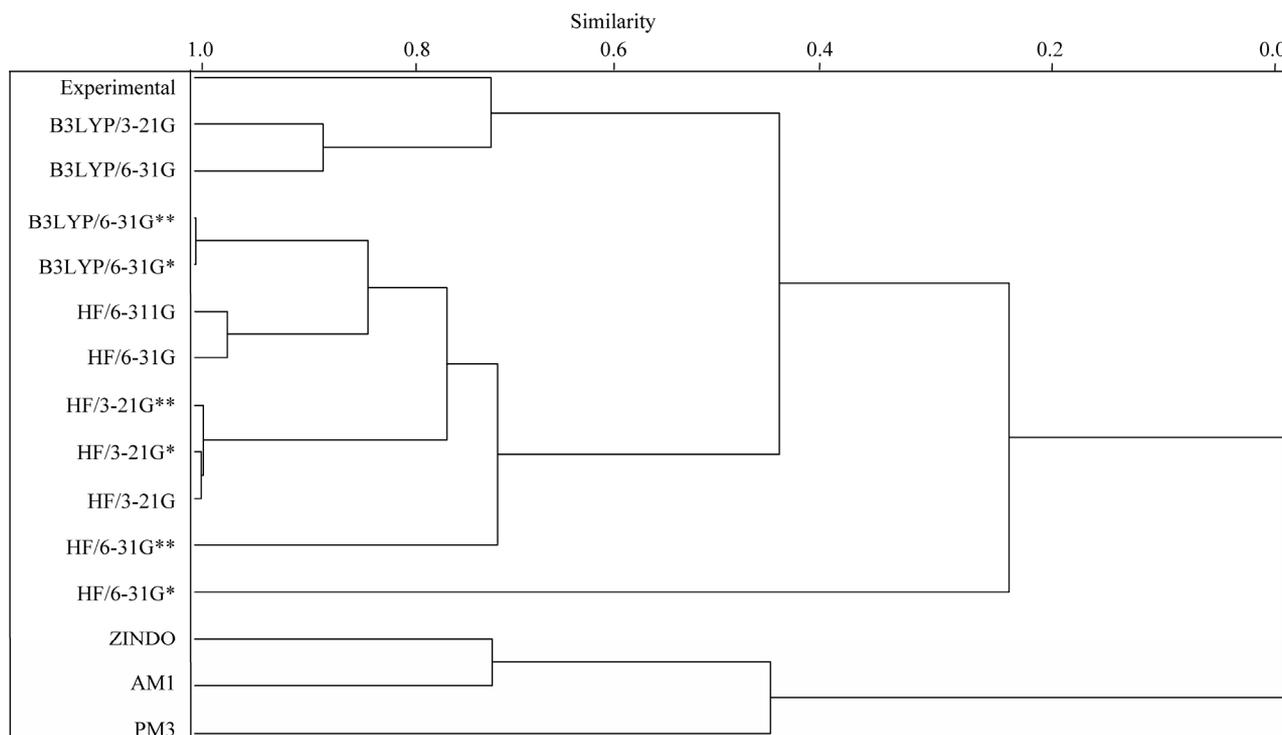


Figure 4. Dendrogram for five methods, seven basis sets, and one factor related to the experimental geometrical parameters of the 1,2,13-trioxane ring.

and Hartree-Fock), as shown in **Table 5**. However, these minimum bases (6-31G and 3-21G) have several deficiencies; thus, a polarization function was included to improve upon these bases (*i.e.*, p orbitals represented by *). These orbitals follow restricted functions that are centered at the nuclei. However, it was found that the atomic orbitals become distorted or polarized when a molecule is formed. For this reason, one must consider the possibility of a non-uniform displacement of electric charges outside of the atomic nucleus, *i.e.*, polarization. Thus, it is possible to obtain a better description of the charges and deformations of atomic orbitals within a molecule. A mode of polarization can be considered by introducing functions for which the values of l (quantum number of the orbital angular momentum) are larger than those of the fundamental state of a given atom. For these types, the basis set names denote the polarization functions. Thus, 6-31G* refers to basis set 6-31G with a polarization function for heavy atoms (*i.e.*, atoms other than hydrogen), and 6-31G** refers to the inclusion of a polarization function for hydrogen and helium atoms [27].

When basis sets with polarization functions are used in calculations involving anions, good results are not obtained due to the electronic cloud of anionic systems, which tend to expand. Thus, appropriate diffuse functions must be added because they allow for a greater orbital occupancy in the region of space. Diffuse functions are important in the calculations of transition metals be-

cause metal atoms have “ d ” orbitals, which tend to diffuse. It then becomes necessary to include diffuse functions in the basis function associated with the configuration of a neutral metal atom to obtain a better description of the metal complex. The 6-31G** basis is particularly useful in the case of hydrogen bonds [27-30].

Pereira *et al.* (2008) studied four structures of artemisinin by reductive decomposition A, B1, B2, and B3 with 13 species (QHS, 1/2, 3, 4, 5, 5a, 6, 7, 18, 18a, 19, 20, and 21), and the structures of the studied species were analyzed in terms of geometrical parameters, Löwdin bond orders, atomic partial charges, spin densities, electronic energies, free energies, and entropy. These studies were carried out at the B3LYP/6-31G** level [31].

Barbosa *et al.* (2011) performed molecular modeling and chemometric studies involving artemisinin and 28 derivatives exhibiting anticancer activity against human hepatocellular carcinoma HepG2. The calculations of the studied compounds were performed at the B3LYP/6-31G** level [32].

Carvalho *et al.* (2011) used the B3LYP/6-31G** method to study artemisinin and 31 analogues with anti-leishmanicidal activity against *Leishmania donovani*. The authors proposed a set of 13 artemisinins, seven of which are less active and six of which that have not been tested; of these six, one is expected to be more active against *L. donovani* [33].

The statistical analysis revealed good correlations ($r >$

Table 5. Calculated correlation and fitting parameters for different methods and basis sets.

Method and basis set	<i>a</i>	<i>b</i>	<i>r</i> /%	R_A^2 /%	MAE ^b	CMAE ^c	<i>F</i>	SEE	PRESS	S_{PRESS}	Q^2 /%	
Semiempirical	°/AM1	0.3527 (±2.2869)	0.9975 (±0.0269)	99.6375	99.2040	3.817	3.8668	1372.07	6.0341	364.982	1.9104	99.2729
	°/PM3	1.7477 (±3.9789)	0.9815 (±0.0468)	98.8797	97.5491	6.807	6.6778	438.82	10.4985	1125.718	3.3551	97.7576
	°/ZINDO	1.4298 (±2.0459)	0.9879 (±0.0240)	99.7039	99.3496	4.335	4.0998	1681.28	5.3983	305.773	1.7486	99.3909
Hartree-Fock/HF	°/6-31G	0.8997 (±0.7444)	0.9922 (±0.0087)	99.9609	99.9142	1.491	1.4715	12811.33	1.9641	44.286	0.6654	99.9118
	°/6-31G*	0.2600 (±1.1455)	0.9890 (±0.0134)	99.9071	99.7957	2.352	2.1988	5375.55	3.0225	98.767	0.9938	99.8033
	°/6-31G**	0.3649 (±0.8369)	0.9969 (±0.0098)	99.9511	99.8925	1.701	1.6137	10231.47	2.2083	49.7003	0.7049	99.9010
	°/3-21G	0.1135 (±0.8242)	0.9980 (±0.0097)	99.9527	99.8960	1.643	1.6281	10572.08	2.1748	47.494	0.6891	99.9054
	°/3-21G*	0.1156 (±0.8200)	0.9980 (±0.0096)	99.9532	99.8970	1.647	1.6311	10679.82	2.1638	47.017	0.6856	99.9063
	°/3-21G**	0.1206 (±0.8601)	0.9979 (±0.0101)	99.9485	99.8868	1.720	1.7042	9707.47	2.2694	51.715	0.7191	99.8970
DFT/B3LYP	°/6-311G	0.8464 (±0.7478)	0.9927 (±0.0088)	99.9606	99.9134	1.492	1.4626	12705.14	1.9733	43.985	0.6632	99.9124
	°/6-31G	0.4455 (±0.3781)	0.9960 (±0.0044)	99.9900	99.9780	0.824	0.8457	50025.59	0.9977	11.367	0.3371	99.9774
	°/6-31G*	0.2160 (±0.5843)	0.9981 (±0.0068)	99.9762	99.9477	1.036	1.0427	21041.00	1.5417	24.099	0.4909	99.9520
	°/6-31G**	-0.0257 (±0.5282)	0.9993 (±0.0062)	99.9806	99.9574	0.870	0.8856	25814.64	1.3936	19.485	0.4414	99.9612
	°/3-21G	-0.4588 (±0.9309)	1.0025 (±0.0109)	99.9402	99.8686	2.044	1.9031	8363.10	2.4562	61.878	0.7866	99.8767
	°/3-21G*	-0.3886 (±0.9058)	1.0020 (±0.0106)	99.9433	99.8754	1.999	1.8794	8824.32	2.3901	58.246	0.7631	99.8840
°/3-21G**	-0.3994 (±0.9435)	1.0021 (±0.0111)	99.9385	99.8649	2.069	1.9459	8134.51	2.4895	63.162	0.7947	99.8742	

Notes: ^aLinear fitting parameters refer to $\delta_{\text{calcd}}(n) = a + b\delta_{\text{exptl}}$ for $n = 12$; ^bMean average error: $\text{MAE} = \sum_n |\delta_{\text{calcd}} - \delta_{\text{exptl}}|/n$; ^cCorrected mean average error: $\text{CMAE} = \sum_n |\delta_{\text{corr}} - \delta_{\text{exptl}}|/n$ (see text); Where $\delta_{\text{corr}} = (\delta_{\text{calcd}} - a)/b$ to correct for systematic errors.

98% and $R_A^2 > 99\%$) for all models with respect to the angles observed in crystallographic data. **Figure 5** presents the predicted and residuals values that were constructed for the two best linear regression models using the DFT (B3LYP/6-31G* and B3LYP/6-31G**) method. The linear regression models for each basis set and method exhibited results that were similar to the experimental values (see **Figures 5(a)-(d)**).

In these figures, the residues are randomly distributed about the zero value of the line, and therefore, there is strong evidence supporting a lack of fit for these models. Currently, the linearity of a straight calibration line is often verified using a correlation coefficient (*r*) computer

program to perform regression. Unfortunately, there is a statistical test that can be applied to this coefficient to check linearity of straight line at a given confidence level [34].

In **Figure 5(c)**, the maximum and minimum residue values obtained for the B3LYP/6-31G* method are shown to be 1.9911 and -3.3488, respectively. As shown in **Figure 5(d)**, the maximum and minimum residue values for the B3LYP/6-31G** method were 1.9820 and -3.1718, respectively.

The statistical parameter results for the methods and basis sets exhibited a distribution similar to that obtained with HCA (see **Figure 4**). However, the results of the

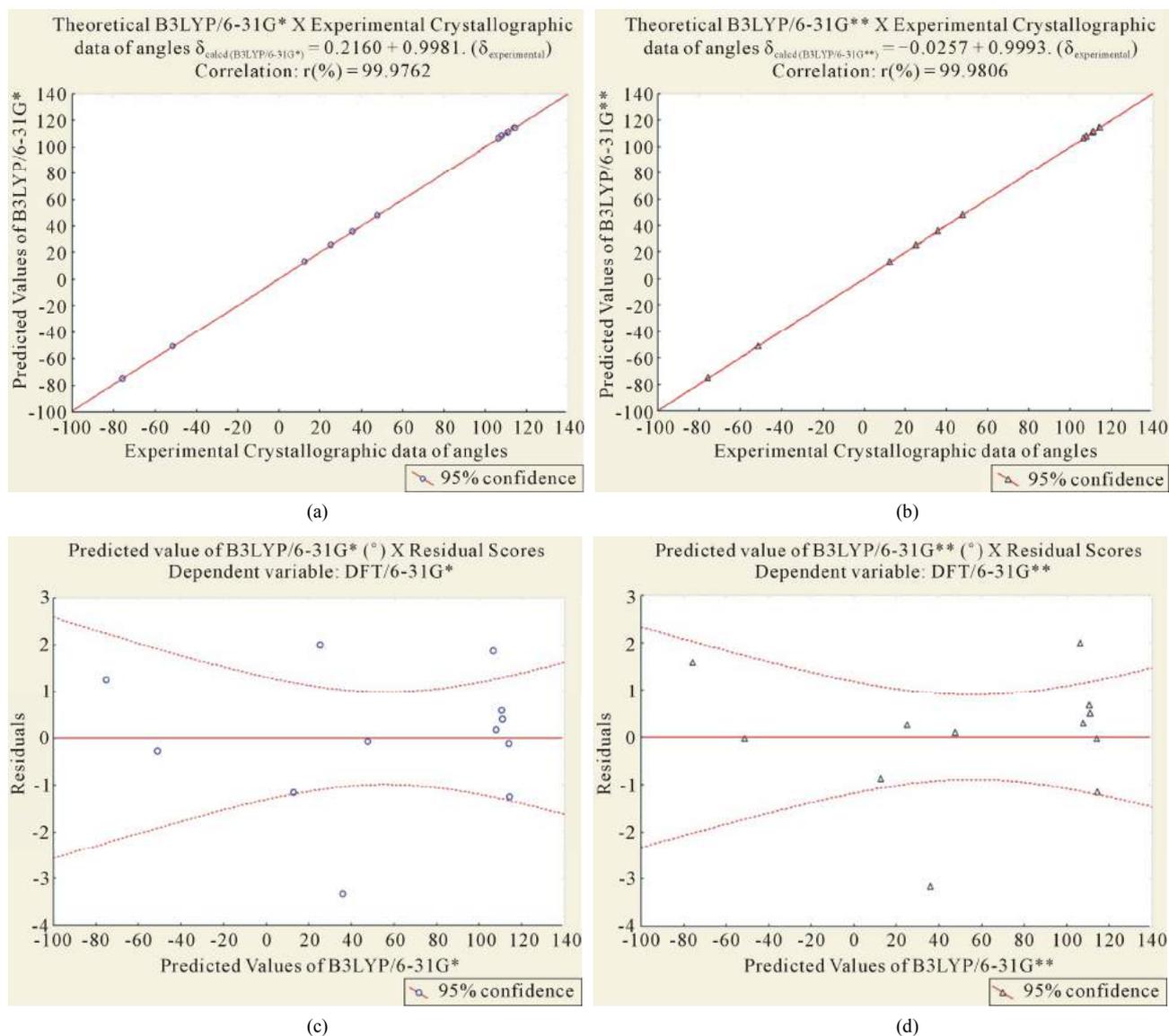


Figure 5. Correlation between the calculated and experimental angle data; predicted values for the DFT/B3LYP method and the 6-31G* and 6-31G** basis sets (a) and (b) and residual values (c) and (d).

DFT method for the B3LYP/6-31G** basis set exhibited the lowest values for MAE, CMAE, SEE, PRESS, and S_{PRESS} ; in comparison with B3LYP/6-31G*, the two sets have a variation of MAE = ± 0.166 , CMAE = ± 0.1571 , SEE = ± 0.1481 , PRESS = ± 4.614 , and S_{PRESS} = ± 0.0495 .

The molecular properties obtained depend on the method and basis set used, which represent a number of functions used in the expansion mono-electronics (orbital) and parameter characteristics that must be optimized. Therefore, it is of fundamental importance that these functions be carefully evaluated to obtain accurate results. Many studies have been performed to develop methods and basis sets that can provide more accurate results. In recent years, other sets have been employed in electronic structure calculations in addition to the methods and ba-

sis sets described previously [30]. Generally, the inclusion of polarization functions in the molecular basis allows for a greater probability of better results for many chemical properties of interest, such as the dissociation energy and dipole moment. In practice, the inclusion of polarization functions with d and f symmetry for small s and p basis sets has been shown to be unsatisfactory; thus, polarization functions should only be added to saturated basis sets.

These results confirm the relevance of theoretical data used to calculate the angles of crystallographic data for this compound (artemisinin) for the method and basis set DFT/B3LYP 6-31G**. HCA confirmed the results obtained by statistical analysis. Therefore, we conclude that the DFT method combined with the B3LYP/6-31G**

basis set can be used for the calculation of molecular properties and for the molecular modeling of artemisinin and its derivatives with biological activity based on the mechanism of action in the region of the endoperoxide ring (1,2,13-trioxane).

4. Conclusions

The statistical analysis techniques of PCA and HCA were vital in enabling the classification of methods and basis sets into three separate groups. The geometrical parameters that were found to be the most important in the classification of methods and basis sets were related to the trioxane ring: C3O13, O1O2C3, O13C12C12a, and O2C3O13C12. The HCA results were similar to those obtained with PCA.

The DFT/B3LYP method with valence-separate basis set 6-31G** exhibited the best results and high predictive ability compared to the methods and basis sets studied herein (semiempirical, Hartree-Fock, and DFT). This method is suitable for molecular modeling studies of artemisinin and for determining the conformation of its derivatives along with their biological activity and mechanism of action in the region of the endoperoxide ring. The DFT/B3LYP 6-31G** method can be used for future calculations of molecular properties, which represent a means of obtaining chemical information contained in the molecular structure of a compound for chemical, pharmacological, and toxicological studies on quantitative structure-activity and structure-property relationships.

5. Acknowledgements

We gratefully acknowledge the support provided by the Brazilian Agency National Council of Scientific and Technological Development and the Institute of Exact and Natural Sciences of Federal University of Para for use of the GaussView and Gaussian software. The authors would like to thank the Postgraduate Program in Biotechnology and Biodiversity-Network BIONORTE, the Laboratory of Modeling and Computational Chemistry, Federal University of Amapá for computational support, and especially Professor Dr. José Walkimar de Mesquita Carneiro of Federal University Fluminense for his contributions.

REFERENCES

- [1] S. R. Meshnick, C. W. Jefford, G. H. Posner, M. A. Avery and W. Peters, "Second-Generation Antimalarial Endoperoxides," *Parasitology Today*, Vol. 12, No. 2, 1996, pp. 79-82. [http://dx.doi.org/10.1016/0169-4758\(96\)80660-0](http://dx.doi.org/10.1016/0169-4758(96)80660-0)
- [2] D. L. Klayman, "Qinghaosu (Artemisinin): An Antimalarial Drug from China," *Science*, Vol. 228, No. 4703, 1985, pp. 1049-1055.
- [3] A. A. C. Braga and N. H. Morgon, "Cálculos Teóricos de Afinidades por Próton De N-Alquilaminas Usando o Método Oniom," *Química Nova*, Vol. 29, No. 2, 2009, pp. 187-193. <http://dx.doi.org/10.1126/science.3887571>
- [4] M. S. Costa, R. Kiralj and M. M. C. Ferreira, "Estudo Teórico da Interação Existente Entre a Artemisinina e o Heme," *Química Nova*, Vol. 30, No. 1, 2007, pp. 25-31. <http://dx.doi.org/10.1590/S0100-40422007000100006>
- [5] F. H. A. Leite, A. G. Taranto, M. C. Santos Junior, A. Branco, M. T. Araujo and J. W. M. Carneiro, "Search for New Antimalarial Compounds Obtained From Natural Sources by Molecular Modeling," *International Journal of Quantum Chemistry*, Vol. 110, No. 11, 2010, pp. 2057-2066.
- [6] F. H. A. Leite, J. W. M. Carneiro, M. T. Araujo, M. Comar Jr. and A. G. Taranto, "Docking Between Natural Peroxides and Heme Group by Parametric Method 6," *International Journal of Quantum Chemistry*, Vol. 1012, No. 20, 2012, pp. 3390-3397. <http://dx.doi.org/10.1002/qua.24247>
- [7] J. C. Pinheiro, R. Kiralj, M. M. C. Ferreira and O. A. S. Romero, "Artemisinin Derivatives with Antimalarial Activity against Plasmodium falciparum Designed with the aid of Quantum Chemical and Partial Least Squares Methods," *QSAR & Combinatorial Science*, Vol. 22, No. 8, 2003, pp. 830-842. <http://dx.doi.org/10.1002/qsar.200330829>
- [8] M. G. G. Cristino, C. C. F. Meneses, M. M. Soeiro, J. E. V. Ferreira, A. F. Figueiredo, J. P. Barbosa, R. C. O. Almeida, J. C. Pinheiro and A. L. R. Pinheiro, "Computational Modeling of Antimalarial 10-Substituted Deoxoartemisinins," *Journal of Theoretical and Computational Chemistry*, Vol. 11, No. 2, 2012, pp. 241-263. <http://dx.doi.org/10.1142/S0219633612500162>
- [9] A. F. Figueiredo, J. E. V. Ferreira, J. P. Barbosa, W. J. C. Macêdo, M. G. G. Cristino, M. S. Lobato, J. C. Pinheiro and R. T. A. Serra, "A Computational Study on Antimalarial Dispiro-1,2,4-Trioxolanes," *Journal of Computational and Theoretical Nanoscience*, Vol. 8, No. 9, 2011, pp. 1847-1856. <http://dx.doi.org/10.1166/jctn.2011.1892>
- [10] J. Q. Araújo, J. W. M. Carneiro, M. T. Araújo, F. H. A. Leite and A. G. Taranto, "Interaction between Artemisinin and Heme. A Density Functional Theory Study of Structures and Interaction Energies," *Bioorganic & Medicinal Chemistry*, Vol. 16, No. 9, 2008, pp. 5021-5029. <http://dx.doi.org/10.1016/j.bmc.2008.03.033>
- [11] GaussView 3.07, Gaussian, Inc.; Pittsburgh, PA, 1997.
- [12] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Och-

- terski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, "Gaussian 03", Revision C.02; Gaussian, Inc., Wallingford, 2004.
- [13] J. N. Lisgarten, B. S. Potter, C. Bantuzeko and R. A. Palmer, "Structure, Absolute Configuration, and Conformation of the Antimalarial Compound, Artemisinin," *Journal of Chemical Crystallography*, Vol. 28, No. 7, 1998, pp. 539-543. <http://dx.doi.org/10.1023/A:1023244122450>
- [14] F. H. Allen, "The Cambridge Structural Database: A Quarter of a Million Crystal Structures and Rising," *Acta Crystallographica Section B*, Vol. 58, No. 01, 2002, pp. 380-388. <http://dx.doi.org/10.1107/S0108768102003890>
- [15] Pirouette Software, Version 3.01, Infometrix Inc., 2001.
- [16] STATISTICA (Data Analysis Software System); Version 6.1, StatSoft, Inc., 2004. <http://www.statsoft.com>
- [17] S. O. Silva, M. J. C. Corrêa, H. R. Bitencourt, W. R. Monteiro, J. Lameira, L. S. Santos, G. M. S. P. Guilhon and D. S. B. Brasil, "Density Functional Theory Calculations of the Nuclear Magnetic Resonance Parameters for Two Dihydrochalcones," *Journal of Computational and Theoretical Nanoscience*, Vol. 9, No. 7, 2012, pp. 953-956. <http://dx.doi.org/10.1166/jctn.2012.2123>
- [18] M. M. C. Ferreira, "Multivariate QSAR," *Journal of Brazilian Chemical Society*, Vol. 13, No. 6, 2002, pp. 742-753. <http://dx.doi.org/10.1590/S0103-50532002000600004>
- [19] A. C. Gaudio and E. Zandonade, "Proposição, Validação e Análise dos Modelos que Correlacionam Estrutura Química e Atividade Biológica," *Química Nova*, Vol. 24, No. 5, 2001, pp. 658-671. <http://dx.doi.org/10.1590/S0100-40422001000500013>
- [20] P. Cimino, P. L. D. Gomez, R. R. Duca and G. Bifulco, "Comparison of Different Theory Models and Basis Sets in the Calculation of ^{13}C NMR Chemical Shifts of Natural Products," *Magnetic Resonance in Chemistry*, Vol. 42, No. S1, 2004, pp. S26-S33. <http://dx.doi.org/10.1002/mrc.1410>
- [21] C. B. R. Santos, J. B. Vieira, A. S. Formigosa, E. V. M. Costa, M. T. Pinheiro, J. O. Silva, W. J. C. Macêdo and J. C. T. Carvalho, "Validation of Computational Methods Applied in Molecular Modeling of Artemisinin with Antimalarial Activity," *Journal of Computational and Theoretical Nanoscience*, Vol. 11, No. 3, 2014, pp. 1-9.
- [22] C. W. Jefford, "Why Artemisinin and Certain Synthetic Peroxides are Potent Antimalarials. Implications for the Mode of Action," *Current Medicinal Chemistry*, Vol. 8, No. 15, 2001, pp. 1803-1826. <http://dx.doi.org/10.2174/0929867013371608>
- [23] G. H. Posner, A. J. McRiner, I. H. Paik, S. Sur, K. Borstnik, S. Xie, T. A. Shapiro, A. Alagbala and B. Foster, "Anticancer and Antimalarial Efficacy and Safety of Artemisinin-Derived Trioxane Dimers in Rodents," *Journal of Medicinal Chemistry*, Vol. 47, No. 5, 2004, pp. 1299-1301. <http://dx.doi.org/10.1021/jm0303711>
- [24] S. Tonmuphean, V. Parasuk and S. Kokpol, "Automated Calculation of Docking of Artemisinin to Heme," *Journal of Molecular Modeling*, Vol. 7, No. 4, 2001, pp. 26-33.
- [25] F. J. B. Cardoso, A. F. Figueiredo, M. S. Lobato, R. M. Miranda, R. C. O. Almeida and J. C. Pinheiro, "A Study on Antimalarial Artemisinin Derivatives Using MEP Maps and Multivariate QSAR," *Journal of Molecular Modeling*, Vol. 14, No. 1, 2008, pp. 39-48. <http://dx.doi.org/10.1007/s00894-007-0249-9>
- [26] J. E. V. Ferreira, A. F. Figueiredo, J. P. Barbosa, M. G. G. Cristino, W. J. C. Macêdo, O. P. P. Silva, B. V. Malheiros, R. T. A. Serra and J. C. Pinheiro, "A Study of New Antimalarial Artemisinins Through Molecular Modeling and Multivariate Analysis," *Journal of the Serbian Chemical Society*, Vol. 75, No. 11, 2010, pp. 1533-1548. <http://dx.doi.org/10.2298/JSC100126124F>
- [27] A. Leach, "Molecular Modelling—Principles and Applications," 2nd Edition, Pearson Education Limited, Upper Saddle River, 2001.
- [28] W. J. Hehre, "A Guide to Molecular Mechanics and Quantum Chemical Calculations," Wavefunction, Inc., Irvine, 2003.
- [29] R. S. Mulliken and B. Liu, "Self-Consistent-Field Wave Functions of P2 and PO, and the Role of d Functions in Chemical Bonding and of s-p Hybridization in N2 and P2," *Journal of the American Chemical Society*, Vol. 93, No. 25, 1971, pp. 6738-6744. <http://dx.doi.org/10.1021/ja00754a004>
- [30] I. N. Levine, "Quantum Chemistry," 4th Edition, Prentice-Hall, New York, 1991.
- [31] M. S. C. Pereira, R. Kiralj and M. M. C. Ferreira, "Theoretical Study of Radical and Neutral Intermediates of Artemisinin Decomposition," *Journal Chemical Information and Modeling*, Vol. 48, No. 1, 2008, pp. 85-98. <http://dx.doi.org/10.1021/ci700011f>
- [32] J. P. Barbosa, J. E. V. Ferreira, A. F. Figueiredo, R. C. O. Almeida, O. P. P. Silva, J. R. C. Carvalho, M. G. G. Cristino, J. C. Pinheiro, J. L. F. Vieira and R. T. A. Serra, "Molecular Modeling and Chemometric Study of Anticancer Derivatives of Artemisinin," *Journal of the Serbian Chemical Society*, Vol. 76, No. 9, 2011, pp. 1263-1282. <http://dx.doi.org/10.2298/JSC111227111B>
- [33] J. R. C. Carvalho, J. E. V. Ferreira, J. P. Barbosa, M. S. Lobato, C. C. F. Meneses, M. M. Soeiro, M. S. Farias, R. C. O. Almeida, K. C. Ventura, J. C. Pinheiro and A. L. R. Pinheiro, "Computational Modeling of Artemisinins with Antileishmanial Activity," *Journal of Computational and Theoretical Nanoscience*, Vol. 8, No. 11, 2011, pp. 2193-2203. <http://dx.doi.org/10.1166/jctn.2011.1943>
- [34] L. M. Z. G. Passari, P. K. Soares, R. E. Bruns and I. S. Scarminio, "Estatística Aplicada à Química: Dez Dúvidas Comuns," *Química Nova*, Vol. 34, No. 5, 2011, pp. 888-892. <http://dx.doi.org/10.1590/S0100-40422011000500028>

Molecular Modeling: Origin, Fundamental Concepts and Applications Using Structure-Activity Relationship and Quantitative Structure-Activity Relationship

Cleydson Breno Rodrigues dos Santos^{1,2,3,*}, Cleison Carvalho Lobato³, Marcos Alexandre Costa de Sousa³, Williams Jorge da Cruz Macêdo⁴, and José Carlos Tavares Carvalho^{1,2,3}

¹Postgraduate Program in Biotechnology and Biodiversity-Network BIONORTE, Federal University of Amapá, Rod JK km 2, Macapá–Amapá, 68902-280, Brazil

²Laboratory of Modelling and Computational Chemistry, Federal University of Amapá, Rod JK km 2, Macapá, Amapá 68902-280, Brazil

³Laboratory of Drug Research, School of Pharmaceutical Sciences, Federal University of Amapá, Rod JK km 2, Macapá–Amapá, 68902-280, Brazil

⁴Secretary of State for Public Health, Central Laboratory of the State of Pará—LACENIPA, Rod Augusto Montenegro, km 10, Icoaraci, 66823-010, Belém-Pará-Amazon-Brazil

Molecular modeling is an important tool to aid the understanding of the fundamental concepts of structure-activity relationships, and to elucidate the mechanism of action of drugs (drug-receptor interaction), used in the teaching-research-extension. The physico-chemical properties as well as three-dimensional visualization of electronic and steric molecular properties elucidation of the interaction between drugs and macromolecules target can be calculated and/or suggested by molecular modeling programs. In this work we show that studies of structure-activity relationships are of great importance in modern chemistry, biochemistry, molecular biology, and other fields of knowledge of health sciences. In order to obtain a significant correlation, it is essential that the descriptors are used appropriately. Thus, the quantum chemical calculations are an attractive source of new molecular descriptors that can, in principle, express all the geometric and electronic properties of molecules and their interactions with biological receptor.

KEYWORDS: Molecular Modeling, Quantum Chemical Methods, Quantum Chemical Descriptors, Multivariate Analysis, Structure-Activity Relationships, Artemisinin.

CONTENTS

1. Origin of Computational Chemistry	3	4.4. Molecular Dipole Moment (μ)	10
2. Computational Methods Used in the Calculation of Molecular Properties	4	4.5. Principal Quantum-Chemical Descriptors Utilized in Studies SAR/QSAR	10
2.1. Semi-Empirical Methods	4	5. Origin of Quantitative Structure-Activity Relationship	11
2.2. <i>Ab Initio</i> Methods	5	6. Multivariate Analysis Methods	13
2.3. Density Functional Theory	5	6.1. Analysis of Principal Components (PCA)	13
3. The Evolution of Computational Chemistry	5	6.2. Hierarchical Cluster Analysis (HCA)	13
3.1. Design and Conformational Analysis	6	6.3. Method of Partial Least Squares (PLS)	14
4. Molecular Descriptors	7	6.4. K-Nearest Neighbor (KNN)	15
4.1. Map of Molecular Electrostatic Potential	8	6.5. Stepwise Discriminant Analysis (SDA)	15
4.2. Energy Frontier Orbital (HOMO and LUMO)	9	6.6. Soft Independent Modeling of Class Analogy (SIMCA)	16
4.3. Descriptors Polarizability (α), Hardness (η) and Molecular Softness (S)	9	7. Case Study on Malaria	16
		7.1. Design and Development of New Antimalarial Drugs	17
		7.2. The Therapy of Malaria	17
		8. Some Examples of Use Quantum-Chemical Parameters in Studies SAR/QSAR	20
		9. Final Considerations	22
		Acknowledgments	22
		References and Notes	23

* Author to whom correspondence should be addressed.

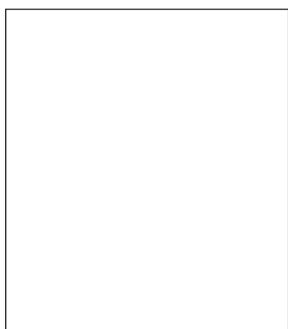
Email: breno@unifap.br

Received: 30 March 2013

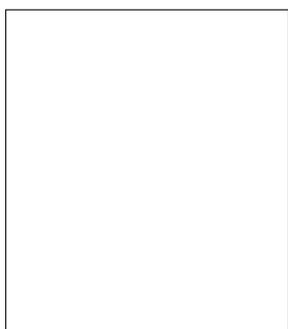
Accepted: 17 April 2013



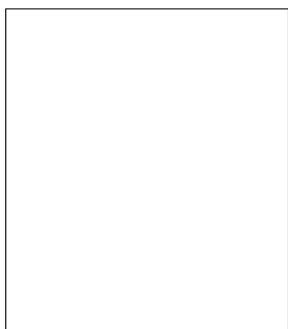
Cleydson Breno Rodrigues dos Santos Has graduation in Bachelors Chemistry in Federal University of Pará–UFPA (2006), Masters in Chemistry from the UFPA (2008) and Doctorate by Bionorte Network in Biotechnology in Federal University of Amazonas–UFAM. Is professor Level II and is part of the research group on Drugs and Applied Computational Chemistry, Federal University of Amapá–UNIFAP. Performs works in the area of Computational Chemistry and Molecular Modeling, aiming to simulate drug-receptors interactions with aim of clarifying the possible mechanisms of action of new drugs of biodiversity Brazilian Amazon. Has experience in the area of Theoretical Chemistry, Computational Chemistry and Molecular Modeling, with an emphasis on Teaching-Research-Extension.



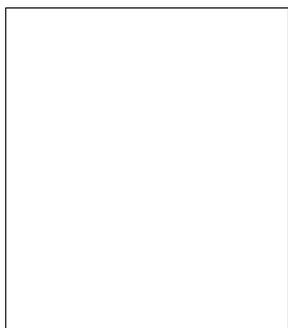
Cleison Carvalho Lobato



Marcos Alexandre Costa de Sousa



Williams Jorge da Cruz Macêdo



José Carlos Tavares Carvalho

1. ORIGIN OF COMPUTATIONAL CHEMISTRY

The name of computational chemistry can be understood, in general, the field of computational methods applied chemistry and related fields. Since there is a large number of computational procedures, computational chemistry presents itself as an interdisciplinary field, by branching out into different areas that, traditionally, it is customary to divide the chemical (Physical Chemistry, Organic Chemistry, Pharmaceutical Chemistry, Biochemistry, Inorganic Chemistry, Analytical Chemistry, chemical Technology and other knowledge areas of human health, as well as the Molecular Biology and Computer Science). Over the past 25 years many application areas of computational chemistry suffered a significant development due to the emergence of new computers, sophisticated software and a better understanding of the basic principles.¹

The potential offered by current technology hardware and software led to the development of a variety of techniques for numerical and symbolic computations. These methods have opened many application areas and the spectacular increase the potential of this technology to make computational chemistry one of the most promising interdisciplinary toward the twenty-first century.¹

The Computational Chemistry is a vector in the direction of unification, by identifying himself as an interdisciplinary field based on a ubiquitous laboratory tool—the computer—it is a tool capable of treating both quantum and classical modeling, geometry and chemical information. This essential characteristic and perhaps only introduced a new scientific community based on the ability of the computer to solve chemical problems and the meeting of scientists from different fields traditionally separated. The interaction between them has led to scientific and technological achievements of great importance as well as the encouragement of new generations of young researchers.¹

Improving the understanding of the students about the concepts of chemistry has been a major goal of researchers in the Teaching of Chemistry (and Science in general) during the last decades. A resource that has been used since the decade of 60 as a tool for learning is the computer, as can be seen in the pioneering work of Atkinson (1968),² Suppes Morningstar (1968).³ The possibilities of using this technology are very large, and over the years it has evolved and changed, as we can see in some articles selected to investigate the presence of this theme in the world of chemistry.⁴

At the beginning of the decade 70, at the University of Lancaster, England, a course in Quantum Chemistry has been carefully organized by chemical Duke⁵ in order to regain lost motivation due to failed attempts of computer scientists that by introducing computational techniques in teaching chemistry, not taking matters relevant to chemistry. This placement reveals that, in 1972, the use of

computers in chemical education was already the focus of research, but was not being conducted in a way to encourage students. In their experiment, Duke a program used to calculate the properties of aromatic compounds by the molecular orbital method.

Fortunately the framework of dissatisfaction was overcome notes as Duke (1972)⁵ “This practical experiment proved successful in that most students learned a substantial body of knowledge on the application of molecular orbital theory to chemistry organic and seems rather interested.” Years later, the same topic was addressed in the University Chemical Laboratory, Cambridge, UK, where microcomputers were used as a teaching tool for Molecular Orbital Theory. “We believe that such programs are of great help in teaching theoretical chemistry.”⁶

The Royal Swedish Academy of Sciences awarded the Nobel Prize in Chemistry 1998 researchers: Walter Kohn (University of California, Santa Barbara, California, USA) for his contribution to the development of Density Functional Theory and Pople (North Western University, Evanston, Illinois, USA) for his contribution to the development of computational methods in quantum chemistry. With these developments that were started from the 1960s, the chemistry reaffirms itself as an exact science, computable.⁷

The IUPAC defines computational chemistry as follows: “Molecular aspects of research as made practical by the use of computers.”⁸ Since then, progress in development of software and hardware combined with a steady reduction of cost of materials informatics, computational chemistry makes one of the most promising areas of this new century. More recently supramolecular chemistry (compounds formed by several molecules) was approached simultaneously by researchers from the Université Louis Pasteur in France and Novosibirsk State University in Russia in 2000, through a course-based model CAI, instruction aided by computer (Computer Aided Instruction), which, according to the researchers Varnek et al.⁹ “allows the visualization of complex structures and performing calculations modest occurs while concurrently reading the text, and computer-assisted courses are easily upgradeable, which is especially important for fields (science) that expand and develop rapidly.”

In the same year, the Department of Chemistry and Biochemistry Brigham Young University in Utah, USA, has developed a method of teaching that included a package with animations for understanding Molecular Orbital in organic reactions. The purpose of the use of simulations was to facilitate the visualization and understanding of this topic. With this method the authors Fleming, Hart and Savage (2000)¹⁰ concluded that “students may benefit from three-dimensional computer representations of chemical events.”

The development and subsequent use of software in the classroom helps solving chemical problems, and versatility

of computational chemistry not only allows its application in teaching chemistry as well as in research and development laboratories and industries. The molecular modeling, for example, is an important tool in developing pharmaceuticals and can be used in rational design of new drugs.

According to Rodrigues (2001)¹¹ “Molecular modeling provides important information for the process of drug discovery. It allows to obtain specific properties of a molecule that can influence the interaction with the receptor.”

The development of software for computational chemistry followed two separate paths, with the current complement of a third. Initially, we worked with the development of codes capable of solving the equations of quantum mechanics to atomic systems strictly based on “first principles.” These calculations involve the resolution of a large amount of integrals, even after applying the usual simplifications, such as the Born-Oppenheimer approximation (roughly speaking can be seen as the separation of electronic and nuclear motion) and the absence of relativistic effects, among others. From this, we have reached an initial commitment resulting in a method called SCF (Self-Consistent Field), virtually omnipresent in all computational chemistry programs. This method leads to a new approximation that ignores the electron–electron interaction (which is “re-introduced” in several ways, the most popular is via perturbation theory, methods abbreviated MP (n), where the electron correlation effects (interaction between electrons) is introduced via TP-RS (Perturbation Theory Rayleigh-Schrödinger).

The second development path followed not from first principles, but the approaches and results of calculations parameterized by atomic groups in order to save the computational steps. Aiming to “embrace” fast calculation of larger systems, and therefore greater chemical interest. The semi-empirical methods parameterize several calculations that the *ab initio* methods necessarily repeated for each new system. So make up scales of power calculation, with a natural loss of accuracy.

Among the most popular methods, are the INDO, CNDO, AM1 and PM3 methods longer used. Finally, in recent decades, a new scale of calculation was introduced, with the MD simulation methods (Molecular Dynamics) and Monte Carlo,¹² which basically does not treat the chemical system as obeying rules of quantum mechanics, but classics like atomic systems using classical potentials to model the behavior of these systems. Applies to systems with thousands of atoms, typically for modeling systems of biological interest or solvation, where this approach is justified by its chemical interest. It is common for software currently existing computational chemistry methods or methods using these products, which we will not discuss here. Some of these methods can be used both for research but also for teaching chemistry.

2. COMPUTATIONAL METHODS USED IN THE CALCULATION OF MOLECULAR PROPERTIES

The methods of quantum chemistry can be applied to quantitative structure-activity relationships (QSAR) for the direct derivation of the electronic descriptors from the molecular wave function. In general, the more rigorous theoretical treatment does not use empirical parameters and is called *ab initio*. Although this type of method provides relatively accurate information about the electronic behavior, he is, in operational terms, slower and more expensive. Therefore, several semi-empirical methods have been developed which are based on certain assumptions which serve to simplify the calculations and use certain parameters obtained from experimental data. Note that the accuracy of these methods is related to the error associated with the selected basis set and the level of treatment of electron correlation. Therefore, there are several computational methods used in the calculations of molecular properties, among them are: the semi-empirical, *ab initio* and Density Functional Theory (DFT).

2.1. Semi-Empirical Methods

The semi-empirical methods using the same formalism as mechanical-like employing basis sets including only the electrons of the valence shell of the system.

The reason behind this approach is that the electrons involved in chemical reactions and other phenomena are the intermolecular electron from the valence shell.¹³

Thus, the great advantage of the semi-empirical methods compared to *ab initio* methods is the higher processing speed, since the calculations are simplified, reducing the cost of memory and computational time. Unlike *ab initio* methods, methods employing semi-empirical is empirical parameters, i.e., derived from experimental data such as geometry of equilibrium heat of formation, molecular dipole moment and ionization potentials, or previously calculated by the Schrödinger equation, allowing some integrals present in the *ab initio* method are not calculated simplifying computations.^{14, 15}

The first method using this approach is the CNDO (Complete Neglect of Differential Overlap), in which the atomic orbitals are considered in evaluating spherically symmetrical electron repulsion integral. Other methods also use these approaches such as INDO (Intermediate Neglect of Differential Overlap), and NDDO (Neglect of Diatomic Differential Overlap).¹³

The semi-empirical methods most commonly used are AM1 (Austin Model1)¹⁶ and PM3 (Parametric Method 3),¹⁷ both methods incorporate approaches very similar, but differ in the parameterization.

Recently, the AM1 method was subjected to a re-parameterization for the atoms of H, C, N, O, P, S, F, Cl, Br and I, resulting in method RM1 (Recife Model1), with minor miscalculations than those generated by AM1 and PM3.¹⁸

2.2. *Ab Initio* Methods

The term Latin means *ab initio* “from the top” or “from the fundamental principles,” i.e., calculations are performed from fundamental physical constants using exact equations, involving a total electronic population of molecule without the use experimental parameters and without additional approaches. The first method for calculating the electronic structure was the Hartree-Fock (HF) which employs the full Schrödinger equation to treat all the electrons in a chemical system.¹⁴

This model employs sets of basis functions (basis set) calculations such as the functions of the Slater Type (STO) and Gaussian Functions (GTO 3-21G, 6-31G). These bases have several deficiencies minimum and to enhance them is the inclusion of polarization function (i.e., p orbitals represented by *).¹³

Thus, 6-31G* refers to basis set 6-31G with polarization function to heavy atoms (i.e., atoms different from hydrogen), 6-31G** refers to the inclusion of function for the polarization atoms hydrogen and helium. The 6-31G** basis is particularly useful where there are hydrogen bonds. Basis functions with partial polarization have also been developed, for example, 3-21G* which is the same minimum 3-21G basis functions with partial polarization.¹³

Although the *ab initio* methods give a quantitative prediction of high quality for a wide variety of systems, they are time consuming and of high cost computation. A resource is commonly used to optimize the geometry with a set of base simplest and then perform calculations to “single point” with a more complete set of base allowing to determine the energy and other properties of a molecular system, using a more sophisticated calculation basis.¹³

2.3. Density Functional Theory

The Density Functional Theory (DFT) is a very successful formalism, where the main objective is to replace the wave function, used to write the electrons in methods like Hartree-Fock, the electron density. The HF calculations consider an average electron density, since the DFT calculations consider instant interactions of pairs of electrons with opposite spins.¹⁹ It is an approach based on the theory of Hohenberg and Kohn which states that all properties of a system are functions of the charge density.

Thus, the Hohenberg-Kohn theorem allows to write the total electronic energy as a function of the electron density ρ :

$$E(\rho) = E_{KE}(\rho) + E_C(\rho) + E_H(\rho) + E_{xc}(\rho) \quad (1)$$

where $E_{KE}(\rho)$ is the kinetic energy, $E_C(\rho)$ is the interaction term nucleus-electron, $E_H(\rho)$ is the Coulomb energy and $E_{xc}(\rho)$ contains the contributions of exchange and correlation.

The molecular orbital calculations of density functional are usually written as a linear expansion of atomic orbitals (ie basis functions) that can be represented using Gaussian

type functions, Slater orbitals or orbital numeric.¹⁹ Functional models density as well as the Hartree-Fock models are applicable in molecules of 50–100 atoms.¹⁵ The exact function is not known, so there is a varied range of different functional that can provide different results for the same problem. The method B3LYP (Becke, Lee, Yang and Parr) is a hybrid method widely applied, where part of the functional is obtained by quantum mechanics (HF combines energy exchange with DFT exchange term) and part is parameterized (adds functional correlation).²⁰

3. THE EVOLUTION OF COMPUTATIONAL CHEMISTRY

One of the most important advances in the design and discovery of new drugs has been the use of Molecular Modeling (MM). Currently, the MM is an indispensable tool not only in the process of drug discovery, but also in optimizing existing prototypes and the rational design of drug candidates.¹⁴⁻²³

According to IUPAC, the MM is the investigation of molecular structures and properties by the use of computational chemistry and graphical visualization techniques, aiming to provide a three-dimensional representation, under a given set of circumstances.¹⁴ The nature of the molecular properties used and the extent to which they describe the structural characteristics of the molecules may be related to biological activity, which is an important part of any QSAR study.

The great development of MM in recent years is due largely to the advancement of computational resources in terms of hardware (calculation speed) and software (computer programs), in addition to advances in computational chemistry, nuclear magnetic resonance, ray crystallography X-biochemistry and molecular biology. This allowed large contribution in the discovery of drug candidates, leading to rapid progress in research and attracting the interest of both academic environment and pharmaceutical industries.^{21, 23}

Silva et al.²⁴ used density functional theory (DFT) calculations (B3PW91/DGDZVP) to determine ¹³C and ¹H nuclear magnetic resonance (NMR) chemical shifts for the two dihydrochalcones: 3,4,5-trimethoxydihydrochalcone and 2,3,4,4-tetramethoxydihydrochalcone. The experimental and theoretical NMR data were analyzed by simple linear regression and the more relevant parameters were selected. In additional, other statistical parameters (correlation coefficients, significance and predictability) were available to judge the quality of the calculations. Finally, the statistical analysis show good correlation experimental and theoretical NMR data with high predictive power.

The computers increase considerably the possibilities for scientific research in drug discovery and thereby can allow chemists collect, store, manipulate, analyze and visualize data. QSAR analysis may be used for planning of biological properties such as potency, efficacy, selectivity and bioavailability of a drug.²⁵

Due to recent advances in computational area and the development of efficient algorithms for calculation, a great advance was also verified in the development of quantum chemical calculations. The *ab initio* and semi-empirical methods quantum chemical molecular parameters provide realistic in a short period of time. Quantum-chemical calculations are great source of molecular descriptors that can, in principle, express many geometric and electronic properties of molecules and their interactions. In fact, several recent studies on SAR and QSAR employing quantum-chemical descriptors alone or combined with conventional descriptors.^{26,27} The Quantum Chemistry provides a more accurate and detailed description of the electronic effects when compared to empirical methods.

Ibrahim et al.²⁸ used Semiempirical molecular modeling technique is applied to assess the interaction of amino acids (alanine, asparagines, aspartic, arginine, cysteine, glutamine, glycine and tryptophan) with chitosan. Results indicate the selectivity of chitosan furthermore we introduce the site whereas amino acid could interact with chitosan. Chitosan is interacting with amino acid through NH₂ group. It is concluded that chitosan is acting with amino acids like protein interaction which dedicate chitosan for many applications in the biological system.

Bayuelo et al.²⁹ proposed methodology to the alternative way to determining of local reactivity indexes using MQS based on the Hirshfeld partitioning. In addition contribution was postulated in this news perspectives in the field such as chemical reactivity, chemical potential, hardness and electrophilicity relative, alternatives to the traditional (chemical potential, hardness and electrophilicity) chosen in the conceptual DFT which allowed us to report the local reactivity indexes proposed with the global that considering the pursuit of local descriptors of reactivity on supported ideas of MQS in Cycloaddition Reactions.

Methods of Quantum Chemistry and molecular modeling techniques allow the definition of a large number of atomic and molecular properties characterizing properties related to reactivity, form and mode of binding fragments and molecular substituents. Due to the large information content contained in many molecular descriptors, the use of quantum-chemical descriptors in QSAR studies has two main advantages: the compounds and their substituents and several fragments can be directly characterized based only on their molecular structures; mechanism proposed action can be justified directly in terms of the chemical reactivity of the compounds studied. Consequently, the obtained QSAR models include information about the nature of intermolecular forces involved in determining the biological activity of the compounds under study.³⁰

3.1. Design and Conformational Analysis

3.1.1. Design and Three-Dimensional Visualization

Several programs two-dimensional of design of molecules are available and easy to use, as ChemWindow, Isis Draw,

ChemDraw³¹ and Chem3D.³² They allow the preparation of figures and diagrams with desired quality and accuracy and facilitate the documentation and scientific communication.

The software ChemSketch 12.00³³ is an advanced design that provides chemical molecular properties, optimization and 3D visualization, ability to name the molecules, as IUPAC, and still has a large database of chemical structures and laboratory materials. The software automatically calculates the valence of each atom and restricts the construction of the molecule based on the octet rule, unless instructed to do this restriction. Then is possible to request the construction of 3D spatial form of the species studied, which triggers another window where the academic can rotate tridimensionally the species studied, in addition to observing these species in different visualizations with possibility to visualize bonds and spatial arrangement of species prominently in each of these representations.

Some programs allow the calculation and representation of various molecular properties, including formula and molecular mass, exact mass and elementary theoretical analysis. More complete programs such as ChemDraw Ultra,³⁴ provide additionally the correct chemical name (IUPAC) of chemical compounds and can predict the corresponding chemical shifts of ¹H and ¹³C NMR, melting points and freezing, log P, molar refractivity and heat of formation.³⁵

The design and visualization of 3D drugs, with steric factors relevant to biological activity, are important for analysis of the size, volume and shape of the molecules.³⁵

The Molekel is a free software multiplatform molecular visualization. It was originally developed at the University of Geneva by Flükiger in the 1990s for Silicon Graphics computers. In 1998, Stefan Portmann took responsibility and released version 3.0. The version 4.0 was almost one version of the platform independent. Other developments lead version 4.3, before Stefan Portman moved and stopped developing the codes. In 2006, the Swiss National Supercomputing Centre (CSCS) restarted the project and version 5.0 was released on December 21 of the same year (FLUKIGER, 2001).³⁶

3.1.2. Conformational Analysis and Energy Minimization

To obtain the conformational analysis and energy minimization, we can cite: the program Chem3D³² widely used in studies with this objective.³⁵ However, other programs like Molecular Modeling Pro,³⁷ ChemSite (ChemSW),³⁸ Alchemy, Sybyl, ChemX, cache and WebLab Viewer are also available.

In the area of molecular modeling, graphics construction and projects of drugs, the program Hyperchem³⁹ for being a tool specializing in 3D structures of interest to the medical, pharmaceutical and organic chemistry. The program lets you design complicated molecules. This software

is also an alternative in the field of spectroscopy, which besides the ability to simulate a priori by the NMR spectra quantum methods, contains a database of approximately 10.000 molecules applicable to macromolecules as well as small molecules. The software also includes animations, and quantum chemical calculations and molecular mechanics.

Conformational analysis of a molecule is performed by rotating a binding with parallel change of torsional angles, and calculation corresponding of the steric energy, due to spatial overlap of atoms unlinked and barriers torsional rotation.⁴⁰

Molecules designed three-dimensionally are not necessarily the most stable conformation. During the generation of a particular structure, distortion occurs in the molecule, with formation unfavorable of lengths, bond of angles and torsional angles. Atoms do not interact also-linked in the same region of space and cause steric demand and electrostatic. To correct these distortions molecules are optimized by energy minimization process, from two mathematical models (i) molecular mechanics or (ii) quantum mechanics. Interactions occur unpredictably related to overlapping molecular orbital, the electron density distribution or steric interference can be solved by computational methods. The energy minimization and conformational analysis are used interactively to optimize the geometry of a molecule.⁴⁰

The choice of method for energy minimization depends on factors related to the size of the molecule, parameters of availability and stored data and computational resources. Molecular models generated by the computer are the result of mathematical equations that estimate the positions and properties of the electrons and nuclei, the calculations exploit experimentally, the characteristics of a structure, providing a new perspective on the molecule.⁴⁰

The three-dimensional structural representation of drugs into computer programs and the construction of molecular models are important for learning the geometric characteristics and molecular essential for the biological activity of some drug classes. In the stage of design and three-dimensional visualization should be gather fundamental knowledge of organic chemistry such as stereochemistry, nomenclature and reactivity to understanding the structure-activity relationship of drugs.

In quantum chemistry the softwares more used are Gaussian and GaussView that uses the laws of quantum mechanics to predict the energies, structures and properties and vibrational frequency of molecular systems (FRISCH, 2003).⁴¹

The GaussView 5.0 is a program that can work on Windows and responsible for building the structures under study, by viewing these as well as for generating the input of the species under study for the program calculations–Gaussian 03W. This includes an advanced molecular modeler, which can be used for construction and molecular dimensions of the three test.⁴¹

The Gaussian 03W is a program that can work on Windows and Linux that performs computations used in the study of reaction mechanisms, equilibrium geometries of neutral molecules, radicals and ions, and the determination of physicochemical parameters. Appreciates structure, reactivity, thermodynamic properties, energy barriers (transition states), conformational analysis, employing the optimization of molecules and theoretical calculations of vibrational spectra. From the optimization is obtained the most appropriate structure to the molecule, whereas the lengths and bond angles and power stabilization calculated by E(RB = HF-LYP) ua.⁴¹

The density functional theory (DFT) is based on the electron density. She gives the distribution of charges in a molecule to assume an approximate Hamiltonian with interaction between pairs exclusively under the Born-Oppenheir approximation and after neglect relativistic effects. The DFT has great computational speed while providing good accuracy.¹⁹

There are several functional in the DFT, in this work we used the B3LYP, which is a functional hybrid. In 1988, the model BLYP was proposed by Becke–Lee–Yang, but with the increasing complexity of the compounds to be studied today, the trend is the use of B3LYP, this is a model that reflects the combination of term LYP correlation with the density functional exchange Becke, B3. In this introduction there are three parameters calculated by Becke and determined by settings using the Hartree-Fock method, which is based on corrections heat of formation for a series of molecules.²⁰

The DFT reputation as conjecture that the exact function is not known, therefore the calculated energy is only a good approximation. Nevertheless, the existing functional not give satisfactory results in cases of weak interactions, anions (situations with scattering electronic cloud) and electron delocalization.

4. MOLECULAR DESCRIPTORS

Obtaining properties (descriptors) depends on the molecular level theory and method, and represents a means of chemical information contained in the molecular structure of the compound studied. This information is transformed and encoded for lots of problems chemical, pharmacological and toxicological studies on the relationship between quantitative structure-activity and structure-property (QSPR and QSAR). The molecular properties take into account different aspects of chemical information, this information can be through experiments or theoretical calculations simple counting, consider the entire molecule, fragments or functional groups, knowledge of the 3D structure of the molecule or molecular graphics his or her simply formula, information defined by scalar values, vectors or scalar fields.⁴²

Recently Santos et al. (2013),⁴³ published in the Journal of Computational and Theoretical Nanoscience,

a prestigious journal in the interdisciplinary area, where the impact factor is 0.932 (2011) an article entitled “Validation of computational methods applied in molecular modeling of artemisinin with antimalarial activity.” Where we propose a combination of chemical quantum methods and multivariate analysis to study the geometric parameters of artemisinin in the ring endoperoxide of the ring (1, 2, 13-trioxane), in order to be effective in selecting the method and level of theory when compared with data crystallographic, aiming to classify and correlate. The most important geometrical parameters selected by principal component analysis (PCA) were O13C12, O1O2C3, C3O13C12C12a and C12C12aO1O2. The results of PCA showed that the model was built with three main components (3PCs), explains 97.0861% of the total variance. The level of theory HF/6-31G** show high similarity with the experimental data assuming that the combination of *ab initio* method can be used for modeling the molecular and for futures calculations of molecular properties structure of artemisinin and its derivatives antimalarial drugs with mechanism of action in the ring region endoperoxide.

4.1. Map of Molecular Electrostatic Potential

The charge density of a chemical species describes the distribution of electrons responsible for the chemical behavior of each species. To estimate the magnitude of this charge is not an easy task, considering that a molecule is a dynamic system and not a simple arrangement of protons and electrons with positive charges with negative charges. Thus, classical mechanics is unable to explain this type of system. There are several methods of calculating atomic charges available computer programs.⁴⁴ The difficulty lies in the fact that calculation of the loads are not obtained directly from the wave function.⁴⁴ The method of calculation of atomic charges most popular is the Mulliken population analysis,⁴⁵ but it is a method for designating arbitrary loads, since, for performing these calculations, the charge density between two atoms is split evenly, not taking into account the electronegativity of these atoms. Another method to evaluate the distribution of the load is to adjust the molecular electrostatic potential, which is a property directly obtained from a calculation SCF (Self Consistent-Field) at a series of points located at the centers atoms. Therefore, we define a set of points around the molecule for calculating the electrostatic potential and a further adjustment is made to model the point loads. From this reasoning, there are many methods derived from the electrostatic potential for the calculation of atomic charges.⁴⁴ One method used to calculate the loads arising from the electrostatic potential was developed by Chirlian and Franci.⁴⁶ According to this method, the potential is determined for a selected number of dots arranged spherically around the molecule. The interactions between ligand and receptor are closely linked to biological mechanisms electrostatic—attraction, repulsion, load

transfers. Thus, charges are calculated at important positions and substituent groups in the molecules.

The MEP is one of the descriptors used in most studies and aims to reveal the size and location of the total molecular electrostatic potential in the molecule. The surfaces of three-dimensional maps of molecular electrostatic potential (MEPs) after superimposition is generated in the molecule of a positively charged particle that under the contact surface of the molecule van der Waals repulsion shows a region representing a potential positive, blueness and the region of negative potential in the molecule, represented by red color.

The electronic parameters are one of the main factors that govern drug-receptor interaction, in this sense, the map of molecular electrostatic potential (MEP) may be an alternative approach in order to understand the electrostatic contribution of these derivatives for biological activity.

To construct the MEP requires three steps: the construction of the surface electron density of the molecule, the construction of the surface electrostatic potential and applying colors to denote the surface obtained potential values.

One of the frequent topics of theoretical chemistry is research to improve methods to elucidate the behavior of molecules and other reactive chemical species. Among the numerous existing reactivity indices the molecular electrostatic potential $V(r)$ that is generated around a molecule by its nuclei and electrons, is known for being a real physical property can be determined experimentally by diffraction methods, as well as computationally.²⁵

The MEP at a given point (x, y, z) in the vicinity of a molecule is defined in terms of the interaction energy between the electrical charge generated from the molecule's electrons and nuclei and a positive test charge (a proton) located at r . For the studied compounds, the $V(r)$ values were calculated as described previously using Eq. (2)

$$V(r) = \sum_A \frac{Z_A}{|R_A - r|} - \int \frac{\rho(r')}{|r' - r|} dr' \quad (2)$$

where Z_A is the charge of nucleus A , located at R_A , $\rho(r')$ is the electronic density function of the molecular, and r' is the dummy integration variable.

The molecular electrostatic potential (Fig. 1) has been an important tool for analyzing processes of recognition of a molecule by other types of interactions such as drug-receptor and enzyme-substrate because of its potential by being a kind that sees another in a process biological recognition.

In Figure 1 the MEP intended to identify and key features evaluate the compound from qualitative comparisons of the form of molecular electrostatic potential in the region of the ring 1, 2, 4-trioxane artemisinin (red color). Since the geometric shape of the electrostatic potential in the region of the ring 1, 2, 4-trioxane is similar for

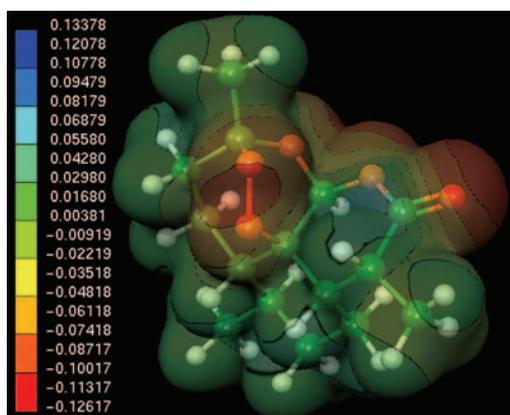


Fig. 1. Electrostatic Potential Map of artemisinin was calculated using the Hartree-Fock (HF) method and HF/6-31G** level of theory. The MEP was realized by the Molekel program.³⁸

all active compounds, that is characteristic according to the literature.⁴⁷ Therefore, compounds having some structural similarity may have electrostatic potentials that allow a being recognized by another, with similar biological activities.²⁵

Structure-activity relationship (SAR) indicates the molecular structure modifications that increase the drug effectiveness. In general, reports show that these modifications are made throughout small changes in the leading compound structure, followed by trials in laboratory to quantify the variations in the biological activity due to changes in the molecular structure.⁴⁸

4.2. Energy Frontier Orbital (HOMO and LUMO)

A second category of quantum-chemical descriptors widely used in studies SAR/QSAR is related to energy of the frontier orbitals (HOMO and LUMO). The reason for this relates to the fact that these properties provide information about the character electron donor and/or electron-acceptor compound and thereby forming a charge transfer complex (CTC).⁴⁹

The energy of Molecular Orbital Highest Occupied Energy (HOMO) and the Molecular Orbital Energy of Lowest Unoccupied (LUMO) are quantum-chemical descriptors, which play an important role in chemical reactions and the formation of several complexes of charge transfer.

In Figure 2 we can see the region bounded by the HOMO orbital measuring the character electron-donor compound, and the LUMO measuring the electron acceptor character. From these definitions, two important features can be observed: the higher the energy of the HOMO, the greater the electron-donor capacity and the lower the energy of the LUMO lower the resistance to accept electrons. In this figure we observe that the HOMO is located in the region of the ring which is the trioxane pharmacophore of the active molecules. When the rings are aromatic substituent or have high electron density, such as

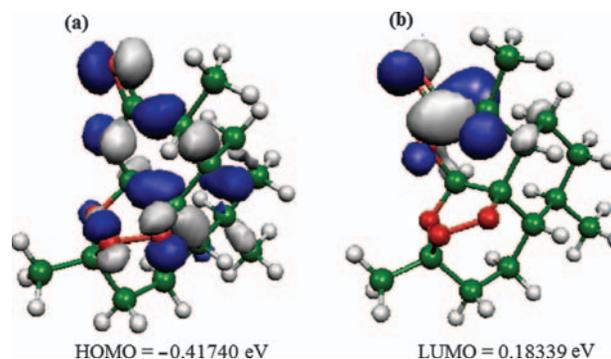


Fig. 2. Orbital energy HOMO (a) and LUMO (b) were calculated using the Hartree-Fock (HF) method and HF/6-31G** level of theory. The HOMO and LUMO orbitals were realized by the Molekel program.³⁸

in carbonyls, amines and amides, the most pronounced HOMO will be strongly influenced to conduct stereo electronics side effects that might impair the pharmacological activity of the compound.

The energies of HOMO and LUMO have been used for some decades as indices of chemical reactivity and are commonly correlated with other indices, such as electron affinity and ionization potential.⁵⁰⁻⁵⁴

The energy of HOMO is directly related to the ionization potential of the compound and characterizes the ability of the molecule to perform nucleophilic attacks. The LUMO energy is directly related to electron affinity, characterized by the susceptibility of the compound in relation to attacks by nucleophiles.⁵⁵ The difference between the orbital energies of HOMO-LUMO (gap) is an important indicator of molecular stability. Molecules with low band gap value are generally reactive, while molecules with a high value of gap indicating high stability of the molecule in the sense of low reactivity in chemical reactions.⁵⁶

$$\text{gap} = E_{\text{HOMO}} - E_{\text{LUMO}} \quad (3)$$

Lobato et al. (2012),⁵⁷ computational calculations performed with the aim of studying the reactivity and stability of isomeric products in reactions of halidification in alkenes by analyzing the orbital border (HOMO and LUMO), index of chemical reactivity, affinity electronics and ionization potential. Where the product more stable, among the investigated was 2-Iodo-2-methylpropane having a more stable variation between normal and branched chain 1.067% stability.

4.3. Descriptors Polarizability (α), Hardness (η) and Molecular Softness (S)

Descriptors are very important in studies of SAR/QSAR, it can be correlated with lipophilicity, molar volume and impediments stereos, aiding the interpretation of the mechanisms of interaction between a compound and their respective biological receptor. The concept of chemical hardness and softness (or softness) molecular was formulated in accordance with the concept of acids and Lewis

bases.⁵⁸ For a basic “soft” (or softer) the donor atom has high polarizability and low electronegativity, and can be easily oxidized, or is associated with occupied orbitals (HOMO) of high energy, making it more effective interaction with the LUMO of soft acids (the energy difference between the HOMO and LUMO is small).

In the interaction between a basic “hard” and an acid “hard” there is a large energy difference between the frontier orbitals HOMO and LUMO, making the electronic transition. For both categories of acids, have the following characteristics: the acids “soft” (or soft), the acceptor atom has small positive charge, a large volume, and various other excitable electrons easily, in acids “hard,” the atom acceptor has high positive charge, small size and has no other electrons easily excitable. This classification was made according to the following rule: acids “soft” with bases “soft” (character covalent interactions) and acids “hard” with bases “hard.” The difficulty of the principle of acids and bases (hard and soft) is in relation to quantification. For this reason we developed a way to calculate molecular hardness (η) and softness (S) from measurements of ionization potential (IP) and electron affinity (difference of total energy between a neutral and anion species, AE) or from the energies of HOMO and LUMO:⁵⁹

$$\eta = \frac{1}{2}(PI - AE) \quad (4)$$

$$\eta = \frac{1}{2}(E_{\text{LUMO}} - E_{\text{HOMO}}) \quad (5)$$

$$S = \frac{1}{\eta} \quad (6)$$

The ionization potential of an atom is a measure of the force with which an electron is bonded to an atom. The first ionization potential (IP) of an atom is the energy needed to remove an electron from that atom at an infinite distance ($A \rightarrow A^+ + e^-$). Low ionization potential values for active compounds may indicate possible mechanisms for transferring charges in the ligand-receptor interaction, and may also indicate that the ionic form of the substance it shows biological activity.⁶⁰

The electronegativity of an element is a measure of the strength of an atom to attract electrons to itself involved in a binding which this atom is also involved.⁴⁵ This property can be used to estimate the ability of a molecule to attract electrons to another, when there is an interaction between these two molecules. Mulliken electronegativity (χ) Calculated as:⁴⁸

$$\chi = \frac{1}{2}(-E_{\text{LUMO}} - E_{\text{HOMO}}) \quad (7)$$

It is possible therefore estimate the hardness or softness of a molecule. These values are expressed in terms of the ionization energy of the atom and its neutral anion. Therefore, molecules that have a high ionization potential and

high electronegativity have high absolute hardness, and the higher the hardness, the lower the smoothness of the molecule. Thus, it can be said that the hardness represents the resistance of a molecule to deformation and softness represents the ease with which a molecule is deformed. The smaller the higher the hardness or softness, the lower the amount of energy required for the transition of an electron of the HOMO to LUMO.⁶¹

Lobato et al. (2011)⁶² carried out studies of isomeric products in chain reactions of halidification in alkenes using computational methods at the HF/3-21G level of theory for the determination of molecular properties, including: polarizability (α), hardness (η), molecular softness (S), bond length (C2-X), total energy and the construction of maps of molecular electrostatic potential of the substrates, reagents and products. Aiming to analyze the stability of isomeric products studied.

4.4. Molecular Dipole Moment (μ)

This is a property that measures the magnitude of charge when displaced atoms of different electronegativity are interconnected. The direction of the dipole moment of a molecule is based on the relative electronegativities of the atoms of this molecules and value is obtained by the vector resultant of the dipole moments of each bond present in the molecule. The presence of substituents with different electronegativity alters molecular properties as acidity and basicity of a compound so that the dipole moment can answer questions about the same reactivity.^{63, 64}

The polarity of a molecule is important for various physicochemical properties and thus, whereas the drug and receptor interaction occurs because of the differences in charges with opposite values, many descriptors have been proposed to quantify the effects of polarity, among which the dipole moment of the molecule (which reflects only the overall polarity thereof) is the most used. The electric dipole moment (m) calculated the $\mu = |m|$, where μ is given by:

$$\mu = \int \rho(r)rdr \quad (8)$$

and $r(r)$ stands for electrical charge density.

4.5. Principal Quantum-Chemical Descriptors Utilized in Studies SAR/QSAR

The quantum-chemical descriptors are fundamentally different from measures obtained experimentally, although there is some overlap naturally. A basic disadvantage of quantum-chemical descriptors is the failure to reproduce stereo effect.³⁰ A summary of the principal terms used in quantum-chemical studies SAR/QSAR is presented in Table I.

The energy parameters such as electron energy, total energy and heat of formation, are also widely used for correlating structure and activity. The electronic energy is determined by means the Born-Oppenheimer

Table I. Summary of the principal quantum-chemical descriptors used in studies SAR/QSAR.

Nome	Definição
Q_N	Liquid atomic charge on atom N
Q_{\min}, Q_{\max}	Atoms charges more negative and more positive
Q_{AB}	Total charge of the group containing the atoms A and B
$\sum q_A^2$	Sum of the squares of the charge densities in atoms of type A
$q_{E,A}, q_{N,A}$	Nucleophilic and electrophilic electronic charge calculated from the occupied and unoccupied orbitals
Q_B, Q_A	Sum of absolute values of the charges of all the atoms in a given molecule or functional group
Q_B^2, Q_A^2	Sum of squares of the charges of all the atoms in a given molecule or functional group
Q_m	Mean of the absolute values of the charges on all atoms
$\epsilon_{\text{HOMO}}, \epsilon_{\text{LUMO}}$	Occupied molecular orbital energy of the highest energy (ϵ_{HOMO}) and unoccupied molecular orbital energy of the lowest energy (ϵ_{LUMO})
$\Delta\eta = \eta_R - \eta_T$	Hardness activation states <i>R</i> and <i>T</i> represent the transition state and reagents
$q_{N\sigma}, q_{Np}$	Electron density σ and π in atom A
$Q_{A,\text{HOMO}}, Q_{A,\text{LUMO}}$	Electron density HOMO/LUMO in the atom A
$f_r^E = \sum (C_{\text{HOMO},n})^2$	Electron density of electrophilic frontier atomic, $C_{\text{HOMO},n}$ are the coefficients of the atomic orbitals X_n in HOMO
$f_r^N = \sum (C_{\text{LUMO},n})^2$	Electron density of nucleophilic frontier atomic, $C_{\text{LUMO},n}$ are the coefficients of the atomic orbitals X_n in LUMO
$F_r^E = f_r^E/\epsilon_{\text{HOMO}}, F_r^N = f_r^N/\epsilon_{\text{LUMO}}$	Índices of electron density of frontier
$\sum S_{E,A}, \sum S_{N,A}$	Sum of nucleophilic and electrophilic super displacement
$\sum \pi_{AA}$	Sum of auto polarizabilities atomic
A	Molecular polarizability
$\alpha = 1/3(a_{XX} + a_{YY} + a_{ZZ})$	Mean polarizability of the molecules
$\beta^2 = 1/2[(\alpha_{XX} - \alpha_{YY})^2 + (\alpha_{YY} - \alpha_{ZZ})^2 + (\alpha_{ZZ} - \alpha_{XX})^2]$	Anisotropy of the polarizability
$p = \sum_{A=1}^N Q_A /N$	Polarization of the molecule, the sum of the atomic charges of all atoms in the molecule
$\mu_{\text{char}}, \mu_{\text{hybr}}$	Components charge and hybridization dipole moment
D_x, D_y, D_z	Components of the dipole moment along the axes
Δ	Polarity submolecular parameters (greater difference in charge of electrons between two atoms)
$D = \sum_{AB} Q_A - Q_B /N_{AB}$	Indices dipole local sum over all pairs of bonded atoms
T	Quadrupole moment tensor
E_T	Total energy
$E_b = \sum_i E_{Ni} - E_T$	Bond energy
ΔH_f°	Enthalpy of formation
$\Delta(\Delta H_f)$	Heat of formation relative
E_{prot}	Protonation energy (energy difference between the protonated form and the sum of the energies of form neutral and of separate proton)
E_{Hidrat}	Hydration energy (energy related to the stability of different molecular conformations in aqueous solution)
RM	Molecular refractivity (property that depends on the structure of the bioactive substance and expresses the character lipophilic and electronic of substituent groups present in the molecule)

approximation, i.e., assuming a fixed position of the nuclei, the Schrödinger equation is resolved in order to find the electronic energy of the molecule. This procedure is repeated for different configurations of fixed of the nuclei (through interactions SCF). The nuclear configuration which corresponds to the minimum value of the energy geometry is the of equilibrium geometry of the molecule.⁶⁵ The total energy is often utilized to estimate the stability of a chemical species, and corresponds to the sum of repulsion energy nuclear with the electronic energy. As the total energy, heat of formation of a molecule is also used to estimate the chemical stability. From the total energy of the system, we calculate the energy of atomization from subtracting the total energies of the atoms in their stoichiometric ratios. The heat of formation (ΔH_f) can be calculated using the enthalpies of atomization of the atoms and the energy of atomization.⁶⁶

5. ORIGIN OF QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP

In 1963, Hansch and Fujita⁶⁷ observed that the biological activity of some series of compounds correlated with lipo-hydrophilicity of the molecules, then the technique developed was expanded to other classes of compounds,⁶⁸ exhibiting equally elevated correlation. These studies resulted in the quantitative analysis of chemical structure and biological activity (QSAR-Quantitative Structure-Activity Relationship), whose greatest interest is providing the rational development of a novel compound, particularly a better drug, avoiding random synthesis and biological testing of new onerous molecules. A mathematical model developed by Free and Wilson⁶⁹ also contributed to the QSAR studies of that era. Currently, most of the models constructed based on QSAR descriptors (parameters that correlate with the biological activities of molecules)

three-dimensional which encode the molecular properties, such as steric and electrostatic, based on the spatial structure of a class counterpart of molecules.

Studies of quantitative relationships between chemical structure and biological activity (QSAR), or between chemical structure and some type of physicochemical property (QSPR), are of great importance in modern chemistry and biochemistry. The central objective in studies of QSAR/QSPR is rationalizing the search for compounds with desired properties using chemical intuition and experience in a way mathematically quantified and computerized. Since the correlation between the structure/property and activity is found, large numbers of compounds, including those that have not yet been synthesized, can be readily examined on the computer in order to select structures with desired properties. Thus, it is possible to select the most promising compounds for synthesis and testing laboratories. It can be said that the studies involving QSAR/QSPR are considered excellent tools to accelerate and succeed in development of new molecules to be used as pharmaceuticals, materials, additives and other purposes. As is not easy to find correlations structure-activity-property, the exponential growth in the number of papers involving studies QSAR/QSPR clearly demonstrates the rapid progress in this area.⁷⁰⁻⁷⁵ To obtain a significant correlation, it is critical that appropriate descriptors are employed, whether theoretical, empirical or derived from experimental data. Many descriptors reflect simple molecular properties and can provide information on the physico-chemical nature of the activity/property under study.⁷⁶

Many studies employing quantum-chemical descriptors are realized in the area of QSAR more than QSPR, ie the descriptors have been correlated with biological activities such as inhibition of enzymatic activity and/or hallucinogenic activity.^{77,78} In part this is because, historically, the search for quantitative relationships with chemical structures began with the development of theoretical methods for drug discovery. The employment of quantum-chemical descriptors has great utility for correlating the reactivity of organic compounds, partition coefficient octanol/water chromatographic retention indices and various other physical properties of molecules.⁷⁶

In recent decades, many significant advances were observed in the area of QSAR, stand out: developing methods dimensional (2D QSAR)^{79,80} three-dimensional (3D QSAR),⁸¹ in addition to methods that involve more advanced molecular information (4D and 5D QSAR), with the main difference between these methods is related to the type of molecular information used for building of quantitative model.^{82,83} Analyzes using more advanced techniques such as QSAR 3D and 4D, require knowledge of the molecular conformations for three dimensional alignment, assuming that the biological response is directly associated with the interactions between the bioactive molecule and biological receptor, these being described

interactions and represented by their stereos fields, electrostatic and other three-dimensional fields. Thus, obtaining quantum-chemical properties is important to the fields of molecular generating and hence the development of QSAR models 3D and 4D of good quality.

The methods CoMFA (Comparative Molecular Field Analysis)⁸⁴ and CoMSIA (Comparative Molecular Similarity Indices Analysis)⁸⁵ stand out as 3D-QSAR methods due to the numerous published papers using these methodologies. These largely accepted methods that are the most widely discussed in 3D-QSAR studies and require three dimensional alignment of the ligands and therefore there is a need structural similarity for overlapping of the chemical structures, therefore, must meet a series congener. Formalisms 4D,⁸⁶ 5D⁸⁷ and 6D⁸⁸ have been applied to incorporate new degrees of freedom (dimensions), so that a more refined analysis on the adaptation of the active site of an enzyme to the topology of the ligand, and vice versa, can be better represented. However, molecular descriptors 2D, usually physico-chemical descriptors referred to in classical QSAR analyzes not have been shown to be lower than those 3D descriptors being extremely potent as the convenience and simplicity of the calculations.⁸⁹ In fact, the need for a conformation of the linker scanning and a comprehensive three-dimensional alignment of structures that may not correspond to the forms of bioactive molecules, reflects the main disadvantages of the techniques associated with the methodology nD, so are an approximation.

A predictive approach equally, but much faster, cheaper and simpler to operate, was developed in 2005 and named MIA-QSAR (Multivariate Image Analysis applied to QSAR).⁹⁰ The MIA descriptors have been successfully applied not only to correlate chemical structures with biological activities,⁹¹⁻⁹⁶ but also with physical properties such as boiling temperatures,⁹⁷ chemical shifts⁹⁸ and electrophoretic profiles.⁹⁹ The method is based on using pixels as image descriptors; how pixels can be processed numerically as binary, white color digit equals 765 and the black pixels digit 0, according to the RGB color system. In MIA-QSAR, the images correspond to the chemical structures drawn by some program to design molecules such as ChemDraw or ChemSketch. Structural modifications or changes in position of the substituents on a series congener molecules correspond to changes in the coordinates of the pixels of the image, and these changes explain the variance in the Y block, the block corresponding to the dependent variables (biological activities, for example).

There aren't advices of some rare manuscripts and reviews of examination boards show some skepticism about the existence of meaning physicochemical descriptors for MIA and therefore on them can be correlated with some property chemical, physical or biological. Some even linked the results of an MIA-QSAR analysis by chance that the correlation could exist between students' scores on a test with the alphabetical order of their names,

which would be a completely arbitrary assumption. We reinforce the assertion that MIA descriptors may encode the chemical, physical and biological, the chemical and physical description must be incorporated in any manner in which substituents are represented. For example, the MIA descriptors may encode steric effects (substituents of organic molecules occupying a large area in the space devoted to the design of structures), stereogenic centers (wedge or dashed lines to represent links for either forward or backward relative to a chiral carbon).¹⁰⁰

6. MULTIVARIATE ANALYSIS METHODS

When measurements are made on a number of objects, the results are usually arranged in a matrix, which is called data matrix. Measures (in our study the molecular descriptors and biological activity) are placed in columns, and objects (in our case the compounds studied) are associated with the lines. For a matrix of multidimensional data, multivariate statistical methods are needed to understand these data in its entirety.

After obtaining a given number of molecular parameters, it becomes necessary to use methods that allow the simultaneous analysis of all parameters obtained since, initially, the main factors responsible for the biological ligand-receptor interaction are not known. Thus, the multivariate methods of analysis are very useful tools in studies of this type, i.e., analysis of data sets with a high number of properties, which makes data interpretation.

Some methods which often use multivariate data analysis are: principal components analysis (PCA) has been widely used in chemical and biological problems, and the main aim of this analysis is to show the data in a multidimensional space of low dimensionality with the minimum of loss total information;¹⁰¹ hierarchical cluster analysis (HCA), partial least squares regression (PLS) method of *K*-Nearest Neighbor (KNN), stepwise discriminant analysis (SDA) and independent models of similarity using principal component (SIMCA–Soft Independent Modeling of Class Analogy).^{102–114}

6.1. Analysis of Principal Components (PCA)

Principal component analysis is a method of data compression based on the correlation between the variables, this data compression generates a small set of variables which are called principal components and these are orthogonal between Themselves, so that the correlation between variables does not limit its application, unlike the multiple linear regression that is sensitive to the presence of highly correlated variables, because it makes the regression coefficients unstable and without significance.^{115, 116}

From the mathematical point of view the model obtained with principal component analysis is written as:

$$y_{ij} = \sum_{k=1}^p P_{ik} a_{kj} + \sum_{k=p+1}^m P_{ik}^{(0)} a_{kj}^{(0)} \quad (9)$$

where P_{ik} are called principal components (PCs), and also are referred to as “scores” and correspond to the characteristics of the compounds. P_{ik} are orthogonal vectors, and are determined so that the data matrix is replicated. a_{kj} represents the weight, i.e., a measure of the contribution of the k PC with the j variable. A high value of a_{kj} demonstrates a high importance of k PC for the j variable. Thus the j variable has a high contribution to the k PC, so a_{kj} are also called “loadings.”¹⁰¹

The number of PCs that can be extracted from the data matrix equals the number of original variables and with this number of components the data matrix can be reproduced exactly, but this is not the desired result, since it does not lead to reduction of dimensionality the matrix. What aims is to find a number of PCs so that the original variables are represented with the minimum of loss of relevant information.¹⁰¹

6.1.1. Staggering

A way of treat all variables with the same importance is standardizing them by autoscale according to the following equation:

$$Y'_{jm} = \frac{(y_{jm} - \bar{y}_j)}{s_j} \quad (10)$$

where \bar{y}_{jm} is the mean of variable j to the object m . Variables autoscaled have an average equal to zero and unit variance represented by Eqs. (11) and (12) respectively.

$$\bar{y}_j = \frac{\sum_{j=1}^n y'_{jm}}{n} = 0 \quad (11)$$

$$s_j^2 = \frac{\sum_{k=1}^n (y'_{jm} - y'_i)^2}{n-1} = 1 \quad (12)$$

6.2. Hierarchical Cluster Analysis (HCA)

Another method of multivariate analysis is of great importance hierarchical cluster analysis, its aim is to show the data in such a way to accentuate their natural groupings and standards. As the principal component analysis, the results of hierarchical cluster analysis are qualitative, being arranged in the form of a dendrogram thus allowing displaying samples (the compounds studied in our case) or variables (molecular descriptors in our study) in a space two-dimensional.¹¹⁵

In hierarchical cluster analysis the distance between samples (the compounds studied here) is calculated and transformed into a similarity matrix S , whose elements are the similarity index. So for two samples m and n , the similarity index is written as:

$$S_{mn} = 1 - \frac{d_{mn}}{d_{\max}} \quad (13)$$

Being S_{mn} an element of S , d_{\max} is the maximum distance for a pair of samples, d_{mn} comes to be the Euclidean distance between samples m and n calculated as:

$$d_{mn} = [(x_{m1} - x_{n1})^2 + (x_{m2} - x_{n2})^2 + \dots + (x_{mh} - x_{nh})^2]^{1/2} \quad (14)$$

With x_{ij} being a matrix element of the original data.¹¹⁵

6.3. Method of Partial Least Squares (PLS)

One of the most promising methods of multivariate analysis is the partial least squares (PLS), because hundreds or thousands of independent variables (the bloc X, in our study the molecular descriptors) may be correlated with one or several dependent variables (bloc Y in our case the biological activity). In the PLS analysis the resulting vectors are slightly displaced from their original positions, so that the correlation of corresponding vectors derived from bloc X and Y be optimized.¹¹⁷

Through the PLS method is possible obtain a description of the dependent variable Y (biological activity) as a linear combination of molecular descriptors, through the principal components being they are not correlated. To ensure that the principal components obtained are important for biological activity, the bloc Y (biological activity in our study) is used for finding the standard within the bloc X (molecular descriptors in our case) that is correlated with the Y bloc, or is the principal components are optimized to best describe the relationship between the bloc X and Y both being the principal components now called latent variables and are used to model the bloc Y.¹¹⁶

The PLS model finds new variables, K latent variables, which are also called "scores" and are represented as t_k ($k = 1, 2, 3, \dots, K$). The "scores" are linear combinations of original variables j_s weighting coefficients W_{vk}^* ($k = 1, 2, 3, \dots, K$).

$$t_{ik} = \sum_k W_{jk}^* x_{ik} \quad (15)$$

The "scores" are good predictors of the bloc Y, then the PLS model can be written as:

$$Y_{im} \sum a C_{ma} t_{ia} + f_{im} \quad (16)$$

in matrix form has the following form:

$$Y = TC' + F \quad (17)$$

combining Eqs. (15) and (16) we obtain the model as a regression model:

$$y_{im} = \sum a C_{ma} \sum k W_{jk}^* + f_{im} \quad (18)$$

where the term f_{im} represents the deviation between the Observed and predicted data (in our study the biological activity).¹¹⁸

In PLS analysis results can be transformed into regression coefficients, also called vector regression, the variables of the bloc X (the molecular descriptors in our study), these indicate what the most important descriptors in model building.^{116, 117}

After building the PLS model is necessary to evaluate it and this review is to ascertain if the specification of the model fits the data adequately predicted. This review is divided into three parts: (1) evaluation of the degree of fit, (2) evaluation of degree of significance and (3) evaluation the degree of predictability.¹¹⁹

6.3.1. Evaluation of Degree of Fit of the Model PLS

The evaluation of the degree of adjustment is done by calculating the following statistical parameters: (a) correlation coefficient R , (b) correlation coefficient adjusted R_{adjusted}^2 , (c) standard deviation and (d) analysis of the waste, with mathematical expressions are shown respectively below:

$$R^2 = 1 - \frac{\sum (y_{\text{predicted}} - y_{\text{Experimental}})^2}{\sum (y_{\text{predicted}} - y_{\text{mean}})^2} \quad (19)$$

$$R_{\text{adjusted}}^2 = R^2 - \left(\frac{k-1}{n-k} \right) (1 - R^2) \quad (20)$$

where n , in Eqs. (20) and (21), is the number of samples included in the model (in our case the number of compounds included), k is the number of variables (molecular descriptors in our study) included in the model.

$$s^2 = \frac{\sum (y_{\text{Experimental}} - y_{\text{predicted}})^2}{n-2} \quad (21)$$

$$(y_{\text{Experimental}} - y_{\text{predicted}}) \quad (22)$$

For a QSAR model to be accepted coefficient R must be greater than 0.9 for tests of biological activity *in vitro* and greater than 0.8 for testing *in vivo*.¹¹⁷ The correlation coefficient adjusted which considers the number corrections to the number of variables and number of compounds used, must be the greatest possible and it is expected that the standard deviation and the waste is closest to zero.¹¹⁹ But the standard deviation s cannot be greater than the standard deviation of biological data, around 0.3 which is the mean error of many biological data whereas for *in vivo* biological test this value should be less.¹¹⁷

6.3.2. Evaluation of Degree of Significance of the PLS Model

The degree of significance is assessed by performing validation tests, in our case we tested only the statistical significance of the correlation coefficient R^2 through the hypothesis test called F test that checks how much of the variability of the Y bloc (biological activity) can be explained by the bloc X (molecular descriptors included in PLS model). Order to validate the correlation coefficient through the test F, the mathematical expression being shown below, it is necessary to compare the value of F obtained in constructing the PLS model with the tabulated value.¹¹⁹

$$F_{(k, n-k-1)} = \left[\frac{R^2(n-k-1)}{k(1-R^2)} \right] \quad (23)$$

where n in Eq. (23) corresponds to the number of samples included in the model (in our case, the number of compounds included), k is the number of variables used to construct the model (in our study, molecular descriptors). The value of F shall be four to five times the value tabulated, thus demonstrating that the model is statistically significant and useful for predictive purposes.¹²⁰

6.3.3. Evaluation the Degree of Predictability of the PLS Model

The cross-validation procedure should be used to evaluate the degree of predictability of the QSAR model, and select the model that has greater predictive ability. The cross-validation is performed so that one or several objects (compound studied) are eliminated from the set of data, either randomly or systematically.¹¹⁷ The cross-validation is done in the following steps:

- (1) delete one of the model compounds,
- (2) reconstruct the model without this compound,
- (3) use the model to calculate the value of the biological activity of the compound excluded,
- (4) obtain the residue between the experimental value of the biological activity and the predicted value for the compound excluded;
- (5) remake steps (1)–(4) for the other compounds, one at a time.¹¹⁹

The evaluation of the degree of predictability is done by calculating some statistical parameters that show the predictive quality of the PLS model: sum of squares of deviations (*PRESS*), standard deviation of the cross-validation (S_{PRESS}), standard error of prediction (*SEP*) e o correlation coefficient of the cross-validation (Q^2), with the respective equations are shown below:

$$PRESS = \sum (y_{\text{Experimental}} - y_{\text{predicted}})^2 \quad (24)$$

it is ideal that the value of the *PRESS* not increase with the number of latent variables.¹²¹

$$S_{PRESS} = \frac{(PRESS)^{1/2}}{n - k - 1} \quad (25)$$

where n in Eq. (25) is the number of samples included in the model (compounds) and K is the number of variables used to construct the model (molecular descriptors), S_{PRESS} is obtained as a criterion for the optimum number of latent variables, the S_{PRESS} lowest value indicates the optimal number of latent variables,¹¹⁷ a model with excellent predictive capability submit S_{PRESS} close to zero.¹¹⁹

$$SEP = \sqrt{\frac{PRESS}{n}} \quad (26)$$

in Eq. (26), n is the number of samples (compounds) included in the QSAR model, the *SEP* is different of S_{PRESS} because not consider the degree of freedom in the calculation.¹¹⁷ It is also used as the main criterion for checking the degree of PLS prediction model.¹¹⁶

$$Q^2 = \frac{PRESS}{\sum (y_{\text{Experimental}} - y_{\text{mean}})^2} \quad (27)$$

This parameter is used to evaluate the statistical quality of the PLS model,¹¹⁵ it describes the amount of variance in y can be predicted, its value ranges from zero

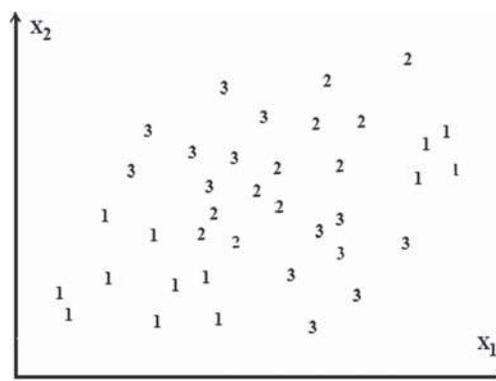


Fig. 3. Classification of K near neighbors belonging to three classes.

to one (0 to 1), where one (1) signifies a model perfect, and zero (0) a model without relevance.¹²² To validate a latent variable is necessary that the value Q^2 does not decrease with increased latent variables, but increase with each latent variable.¹²¹ A PLS model with high predictability for objects (compounds) are not included in the model should display the value of Q^2 nearest one (1) as possible,¹¹⁹ but a value above 0.3 ($Q^2 > 0.3$) is generally satisfactory.¹²³

6.4. K-Nearest Neighbor (KNN)

An unknown object can be classified¹²⁴ by investigating objects of their close neighbors of K which class the group is known. To find a close neighbor of the object is necessary to compute the distance (usually Euclidean distance) to every object in the dataset. An unknown object is assigned to the class that has the most among the K neighbors (Fig. 3). Frequently used values for K are between one and ten. This approach is similar to a spectral library search, although more complicated (nonlinear). Similar measures are used for research in the library and usually point to identify the object in place of the standings. The method KNN can also be compared as a product for estimating the location probability density for the class.

Moreover, this mathematical simplicity of KNN classification has some advantages: no ordering on the distribution of data is required, the class of objects need not be linearly separated, he is a multiclass method, and it is not necessary to classify the training.

A KNN classification consists for any vector object assembly training, this, however, requires a computational time rather large even for data set size medium, the KNN classification is often used as a reference method.

6.5. Stepwise Discriminant Analysis (SDA)

Stepwise discriminant analysis (SDA) is a method that can be used for discrimination (recognition or classification) and prediction samples. Its main purpose is to determine discriminant functions, which represent linear combinations of variables calculated.¹²⁵ The procedure used in the

SDA method is to construct the discriminant functions (a function for active compounds and one for inactive) adding one variable each time until obtaining the end discriminant functions based on the set of variables which best discriminates between the groups of compounds. This method is useful to select the most relevant variables to separate the compounds into different groups (often referred to as the discriminating power of the variables), since constructs discriminant functions using a variable time until the best discriminant function is obtained. After statistical validation of the model through this procedure, the discriminant functions can be used to make predictions with unknown compounds.

6.6. Soft Independent Modeling of Class Analogy (SIMCA)

It is a method for classification¹²⁴ that considers information of population distribution estimates a confidence level of classification and may provide new samples as belonging to one or more classes or no classes. To make the classification the SIMCA uses the space of the principal components of each class. Thus, the n class shall be represented by Eq. (28).

$$X_n = \bar{T}_n \bar{P}_n^t + E_n \quad (28)$$

where X_n are the class data, T_n the matrix containing coordinates in the principal component of the class n , P_n the linear transformation matrix and E_n the matrix of residues. In the construction of the SIMCA classification model calculates for each class separately, the standard deviation of residues. For the space described by the principal components are computed variances of samples in each axis. These two parameters are used for classification of new samples. The objective of SIMCA is to create a limited space for each class. This can be better understood for a class described by two principal components. In geometric terms, the residues of this class correspond to the distances of the samples to the plane of the main components. Thus, the calculation of the standard deviation of residues originates two planes parallel to these components, i.e., one above and one below. Considering the variance in each main component and planes, for the standard deviation of residues, one can say that the class is bounded by a box, a hyper box in case of three or more components and a cylinder for a main component.

The classification of a new sample is made by its projection on the main components of each class, where the variances are calculated and its residue. These two are compared by F tests with those already determined for constructing the model. Thus, those class where the residue is less than or equal to, the same is true for the variances, the sample is classified positively. With this, the sample may be placed in one or more classes. Otherwise, larger deviation or variance, the sample is classified as belonging to the class not.

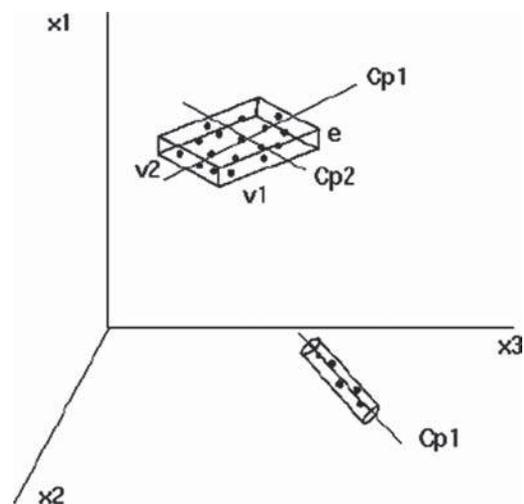


Fig. 4. Graphical representation of a matrix with three columns and divided into two classes. x_1 , x_2 and x_3 represent the columns of the matrix, the first principal component PC1, PC2, the second principal component v_1 variance in PC1, v_2 variance in PC2, the standard deviation of residues is represented by e .

The Figure 4 shows an example of a matrix with three columns, divided into two classes. The first is represented by two major components, where v_1 is the variance in the first principal component PC1, v_2 variance in the second component PC2 and the standard deviation of residues are represented by:

7. CASE STUDY ON MALARIA

Malaria is a potentially serious infectious disease caused by parasites (protozoan of the genus *Plasmodium*) that are transmitted from one person to another by the bite of female mosquitoes of the genus *Anopheles*. The transmitters of human malaria are insects of the order diptera, family *Culicidae* and the genus *Anopheles*. This genus comprises about 400 species, of which only few have importance for the epidemiology of malaria in each region. Among the different species of malaria, malaria caused by *Plasmodium falciparum* and *Plasmodium vivax* are the most widespread. The malaria produced by *Plasmodium vivax*, are rarely fatal, however, symptoms tend to recur periodically, even after long periods of treatment (BRASIL, 2012).

The malaria is one of the most common diseases in tropical and subtropical countries with more than 300 million infections and millions of deaths from malaria occur worldwide each year. The rapid spread of resistance to quinoline antimalarial made malaria a serious global problem, so it is essential to seek new drugs against malaria and understand its mechanism of action for the treatment of patients.¹²⁷ Where, pathways rule, institutions do not have the economic resources that can be used to access high technology which provides the adequate development of research aimed at solving the problem.¹²⁸

7.1. Design and Development of New Antimalarial Drugs

Despite several studies that have been done many years ago, today does not exist a vaccine that confers protection against malaria satisfactory. It is common to confuse the vaccine against yellow fever as though it were against the malaria (antimalarial), but only the first exists, is effective and fundamental to take it when you travel to yellow fever endemic areas, which generally overlaps with the malarial areas (BRASIL, 2012).

The processes for obtaining new drugs have changed much over the years. Until approximately the decade eighties, new drugs were discovered through tests black-box type in cells or models-animals. This traditional method of drug development, also called blind triage, consists of testing several micromolecules randomly in biological assays, without any knowledge of the mechanisms of action and/or interaction of the ligand molecule. Although most of the drug therapeutically useful available today have been discovered through this process gradually this methodology proved to be inefficient due to the ever smaller probability of finding a new drug, being necessary to test millions of compounds in some cases. This fact resulted in a great increase in the time and costs involved in the discovery and development of new drugs.^{129, 130}

The impossibility of removing the vector of malaria transmission requires the necessity of new agents that may have a novel mechanism of action. Therefore, the search for new compounds with effective action to combat *Plasmodiumfalciparum* has high priority due to its severity becomes an urgent mission of global research programs.¹³¹

In order to plan a drug applies one of the following strategies: the rational design and molecular modification.¹³² The molecular modification consists in considering a compound of known chemical structure and biological action proven as a model or prototype, synthesize and test new compounds which are congeners, homologues or structural analogues of the drug matrix.¹³² This method when used requires high time and high financial investment in obtaining a new drug.

The implementation of rational design requires information from different areas of human knowledge, especially those related to electronic levels of the drug, biological activity, physicochemical parameters, such as: hydrophobic, steric and electronic, related to biological activity.¹³³

The Figure 5 shows a flowchart with several steps of Rational Drug Design Based on Structure (RDBE). The first step constitutes the appropriate choice of therapeutic target. The target molecular bioreceptor or may be a protein related disease for which is desired develop a chemotherapy treatment, where function shall be blocked or activated. After identifying the molecular target, the three-dimensional molecular structure of the bioreceptor needs to be obtained. The molecular structures

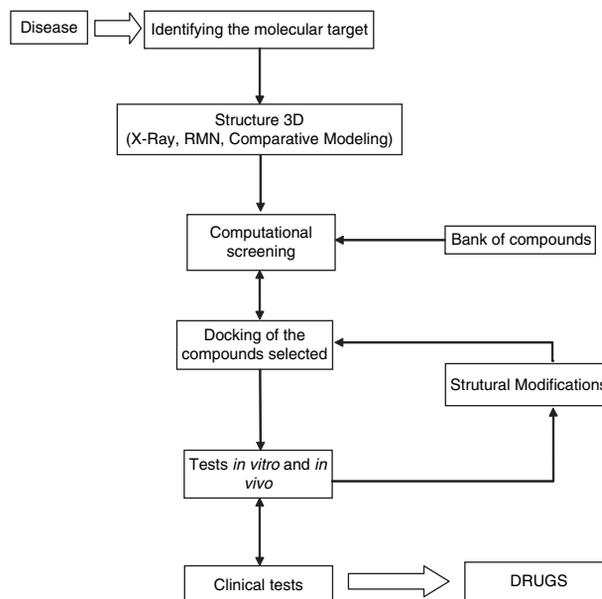


Fig. 5. Steps of rational drug design based on structure.

may be obtained using experimental techniques, such as X-ray diffraction in crystals and nuclear magnetic resonance (NMR), or by theoretical methods as comparative modeling.¹³⁴ In this step, the docking methodologies (drug-receptor interaction) need to be faster than those used for the refinement of compounds prototypes. Large banks of molecular structures of ligands may contain millions of molecules are tested against molecular target with the use these methods in order to identify compounds prototypes. When a promising compound is found, the medicinal chemist analyzes the modifications that can be made in the molecule, so that the desired biological response is potentiated. In this phase, enter the scene more accurate docking methodologies, which aim to identify both binding conformation as quantify more precisely the binding affinity receptor-ligand, aiming an optimization of promising molecules selected. *In vitro* and *in vivo* tests are conducted to validate and guide the optimization of compounds prototypes and so that other characteristics such as toxicity, are analyzed.¹³⁴

From among the various techniques used in rational drug design, we can relate: design technique with the aid of computing, principally using physicochemical parameters involved in biological activity and methods of quantum chemistry to determine the most promising compounds in a series.^{48, 133, 135, 136}

7.2. The Therapy of Malaria

In medicinal chemistry, the term “structure-activity relationship” comprises studying the effects that the chemical structure of a compound (ligand) may cause during their interaction with the biological receptor, and consequently rationalize the main factors that govern this interaction. The interactions of a drug with its biological receptor

are determined by intermolecular forces, i.e., interactions lipophilic, polar, electrostatic and steric. Therefore, substances showing therapeutic properties when interacting with a specific target (an enzyme, a receptor, an ion channel, a nucleic acid or other biological macromolecule) should have a three dimensional structure so that the provisions of their functional groups promote greater complementarities to the binding site. This can be summed up as follows: how much better the “Docking” and the complementarity of the surface properties of a drug, greater its affinity and greater may be its biological activity.^{117,137}

To describe the types of interactions between a biological receptor and its ligand may be used an extensive set of molecular properties, since these properties are directly related to intermolecular forces involved in ligand-receptor interaction, even as are related to transport properties and distribution of drugs.

The molecular descriptors represent an important tool for predicting properties of substances, classify chemical structures or search for similarities between them.^{138–140} Different descriptors have been introduced in recent years and the number continues growing, since it is believed that with this increase, significant problems in studies on structure-activity relationships (SAR) would be solved.¹⁴¹

The treatment of malaria was employed by traditional Chinese medicine for more than 2000 years. The medicine used artemisinin (qinghaosu) is extracted from the plant *Artemisia annua* L, used in combatting of 52 species of diseases in People’s Republic of China.¹⁴²

Artemisinin (qinghaosu) has a single structure (Fig. 6) having a lactone stable endoperoxide (1, 2, 4-trioxane) sesquiterpene totally different from the previous antimalarial in its structure and mode of action, being isolated from *Artemisia annua* is a compound of remarkable life and economic antimalarial effective against *Plasmodium falciparum* and cerebral malaria.¹²⁷

Artemisinin and its derivatives induce a rapid reduction in the number of parasites when compared with other known drugs. They are therefore of particular interest for severe malaria. The first reduction in the number of parasites is also beneficial for combination therapies. This led to an enormous interest in the mechanism of action, chemical¹⁴³ and drug development¹⁴⁴ of a new class of antimalarials.

The group endoperoxide is essential for antimalarial activity¹⁴² and is mediated by active oxygen (superoxide, H_2O_2 and/or radicals hidroxis) or carbon free radicals.¹⁴⁵

The artemisinin has a broad and extraordinary activity against parasites in asexual form, killing at all stages in malaria caused by *Plasmodium falciparum*. The artemisinin also kills the gametocytes, including the four stages of gametocytes, which are sensitive only to primaquine.¹⁴⁶

In vitro biological assays indicated that some amount of iron must be added artemisinin to show antimalarial activity. In humans, the heme compound (Fig. 7), which is a product originating from one digestive process the RBCs must be the source of iron for artemisinin. In human malaria parasites digest over 70% of the hemoglobin within the red blood cells infected, since the globin and heme as products, the protein is hydrolyzed, generating aminoacids that are utilized in protein synthesis by the parasite, since heme undergoes a polymerization process having as a product hemozoína.^{147–149}

The polymerization of heme is a target of some drugs antimalarials such as chloroquine, which inhibits this process. Strains of chloroquine-resistant *Plasmodium berghei* that for want of hemozoín are caused by the non-occurrence of polymerization of heme. This reinforces the view that the inhibition of heme polymerization is the mode of action of artemisinin, thus indicating that artemisinin very possibly interacts with the free heme and for this reason inhibits the polymerization process.¹⁴⁹

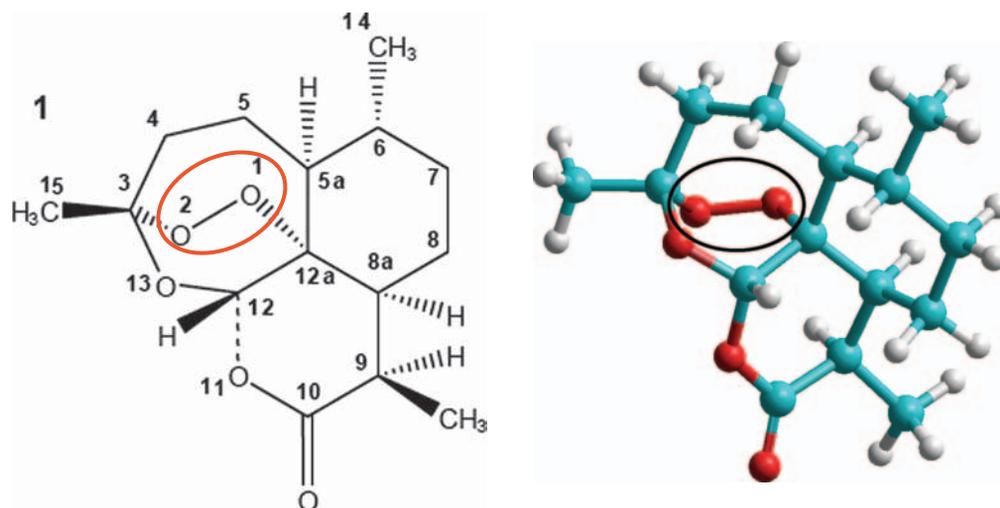


Fig. 6. Artemisinin (structure) and the region essential for expression of the biological activity (pharmacophore). The Structures were visualized using the ChemSketch 12.00³³ and Hyperchem³⁹ softwares, respectively.

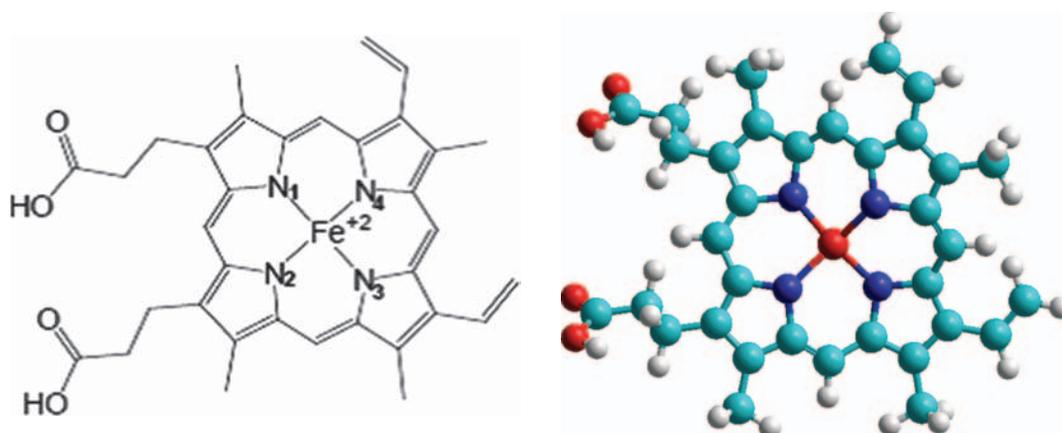


Fig. 7. The numbering the heme group is systematic for the purpose of determining the geometrical parameters. This structure was visualized by means of ChemSketch 12.00³³ and Hyperchem³⁹ softwares, respectively.

The numbering of the heme is systematic for purposes of determination of geometrical parameters: bond length, bond angle and torsion angle between artemisinin and derivatives (drugs) with heme (receptor).

The mechanism for antimalarial action of these compounds is still without conclusion, however investigations give great significance to the endoperoxide group in artemisinin for biological activity.

There are two possibilities for the heme iron attach in artemisinin derivatives. Posner et al. (1995)¹⁵⁰ have proposed the attack of the iron to the compound, occurs by the position O2 and produces radical free at position O1 with subsequent rearrangement to a free radical C4. This radical formed was suggested to be an important substance for biological activity. The compound formed is changed to another compound by a reaction of beta cleavage. Subsequently, the new compound is rearranged to form an epoxide compound. Alternatively, a direct intramolecular formation for the radical form another compound. This compound is capable of alkylated protein specific malarial parasite and possibly cause damage to parasites.¹⁵¹ Moreover, Jefford et al.¹⁵² believe that the attack of the iron the compounds takes place at the position O1 and produces a free radical at position O2. After the C3—C4 bond is broken to give a carbon radical C4. This radical can also be very harmful to the parasite. In the discovery and development of drugs, knowing the mechanism of action may help the development of new and more effective drugs.

In artemisinin peroxide binding in 1 and 2 oxygens (O1 and O2) is contained pharmacophoric structure of trioxane ring of artemisinin, is essential for the expression of antimalarial activity. Peroxides are known by suffer reductive cleavage transition metal of low valency, generating free radicals oxygens. These free radicals oxygens are potent agents of abstraction of hydrogen, can also generate free radicals carbon by abstracting intramolecular hydrogen atoms, these radicals carbons are assumed to react with biomolecules.¹⁵³

The consideration of molecular flexibility of receptor and ligand implies in treatment hundreds of thousands of degrees of freedom, on the part of algorithms of “Docking.” The molecular recognition is a dynamic process and highly complex, involving a large number of intermolecular interactions between the ligand, the receptor molecule and solvent. Due to its complexity, the problem of “docking” is generally divided into two subproblems: (i) development of an algorithm that investigate of way effective a complex energy hypersurface to predict the conformation and orientation of a ligand molecule relative to the active site of the receptor; (ii) predicting the binding affinity of a receptor complex–ligand, i.e., the development of a model for assessing binding free energy (usually called in the literature function of “scoring”), it is computationally viable to correctly discriminate between different binding modes of the same binder and/or to determine between two different ligands, those with the greatest binding affinity for a given receptor.¹⁵⁴

The first program of “docking” using Genetic Algorithm (GA) was implemented by Judson (1993).¹⁵⁵ The program “docking” GOLD uses a genetic algorithm to evolve multiple subpopulations of ligands, where migration between populations is permitted.¹⁵⁶ The program AutoDock has implemented a Lamarckian Genetic Algorithm (LGA). The LGA is a hybrid GA with local search (LS). The programs “docking” most widely used today, and that are generally used for comparison with new proposed methods are: DOCK, FLEXX, AutoDock and GOLD.

Currently, there large banks of molecular structures for public access, for example, the Protein Data Bank (PDB), where various structures are deposited and can be obtained, and several banks of molecular structures of ligands, as the Maybridge and the Cambridge Structural Database.¹⁵⁷ In the problem of docking protein–ligand the purpose is to find and quantify the mode of correct binding a ligand molecule in the active site of a macromolecule receptor such that the receptor function can be enabled or inhibited.¹³⁷

Cheng et al. (2002)¹⁵⁸ performed a study of molecular docking and 3D-QSAR, quantitative relationship three-dimensional structure activity, in order to understand the antimalarial mechanism and the relationship between physicochemical properties and antimalarial activity of artemisinin analogues by molecular docking simulations to probe the interactions of these analogues with heme. The 3D-QSAR was based docking models employing comparative molecular field force analysis (CoMFA) and comparative molecular similarity indices analysis (COMSIA). The subsequent analysis of partial least squares (PLS) indicated that the required energies of the calculation correlated well with the experimental values of the activity. The CoMFA and COMSIA models were active conformations that have demonstrated a good predictive ability. In turn combining together the results of docking.

The studies on the mechanism of action artemisinin have been conducted in search of providing guides for synthesis of new derivatives with improved efficiency and stability in combat malaria *falciparum*. In these studies the activities of artemisinin and its derivatives appear to be mediated by its interaction with iron in hemoglobin, see Figure 8, through endoperoxide function.¹⁵⁹

The endoperoxide binding (C—O—O—C) of artemisinin is believed to be the key to the mode of action of the drug. The iron in the +2 oxidation state (Fe^{+2}) catalyzes the breaking of this bond, resulting in highly reactive free radicals these free radicals from artemisinin derivatives modify and inhibit a variety of parasite molecules, causing them to die. A rich source of intracellular Fe^{+2} is the heme that is an essential component of hemoglobin, and also responsible for activating artemisinin within the parasite.¹⁶⁰

Malaria parasites in humans degrade hemoglobin and red blood cell within the heme and globin. Subsequently, globin is hydrolyzed to produce amino acids as the source for protein synthesis. The toxic portion of heme will be principally detoxify by the process of hemozoina polymerization.^{147, 148}

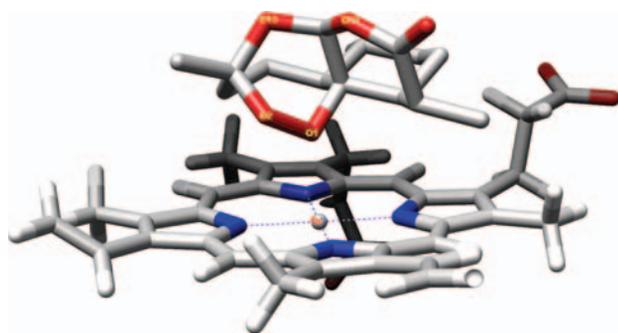


Fig. 8. The molecular docking can be performed using the GOLD software.¹⁵⁶ The structure of heme to the realization of interaction drug-receptor/artemisinin-heme (Molecular Docking) was obtained from the Protein Data Bank (PDB).¹⁵⁷

8. SOME EXAMPLES OF USE QUANTUM-CHEMICAL PARAMETERS IN STUDIES SAR/QSAR

Martins et al. related the geometric and electronic descriptors derivatives rutaecarpine analogues with biological activity against cancer of the central nervous system (CNS), where calculations of quantum chemistry used was the molecular level B3LYP/6-31 (d) and statistical analyzes were performed for 21 rutaecarpine derived analogues. Of (86) calculated molecular descriptors, (05) were selected for constructing the model of principal component analysis (PCA). The component PC1, which responds by 46.11% of the total variance, was able to completely discriminate compounds in two classes: active and inactive. All molecular descriptors selected by PCA model were electronic parameters. The hierarchical cluster analysis (HCA) was also applied to the descriptors selected by PCA model. Based on (05) descriptors selected were possible to suggest new derivatives assets of the rutaecarpine to be synthesized. Furthermore, a model of partial least squares for discriminant analysis (PLS-DA) was built supervised and successfully applied in discriminating similar to rutaecarpine, which was validated using an independent set of compounds.⁴⁸

One of the first studies of structure-activity relationship of artemisinin and antimalarial molecules using molecular electrostatic potential maps was performed by Thomson, Cory and Zerner (1991), through the theoretical study of the structure of artemisinin and its derivatives using semi-empirical methods (AM1 and PM3) and Hartree-Fock (3-21G, 6-31G). The MEP obtained with AM1 and PM3, when compared showed distinct differences between active and inactive molecules, such as 8-deoxyartemisinin because active molecules have a wide band around the negative potential of the molecule containing endoperoxide binding.¹⁶¹

Bernardinelli (1994) used maps of molecular electrostatic potential aiming to identify the key features that are necessary for the antimalarial activity of artemisinin and some derivatives. The MEP showed that the active molecules have a region of negative potential similarly near the trioxane ring, but this region is displaced in inactive compounds. The MEP were used to make predictions of quality new and more effective antimalarial molecules, thereby Bernardinelli et al. concluded that any further active molecule will have MEP similar to artemisinin.⁴⁷

Pinheiro, Ferreira and Romero (2001) techniques combined quantum chemical (Hartree-Fock 3-21G) and multivariate analyzes methods (PCA, HCA, KNN and SIMCA) to study and propose diidroartemisinin derivatives. Through the technique PCA and HCA selected seven (7) descriptors that were responsible for the classification of compounds into two distinct classes, and with construction of qualitative models KNN and SIMCA proposed two

(2) compounds of a set of twelve (12) tested predicted as high activity.¹⁶²

Pinheiro et al. (2003) planned artemisinin derivatives with antimalarial activity with the help of quantum chemistry and partial least squares method (PLS). They built a QSAR model based on five (5) molecular descriptors used to predict the antimalarial activity of ten (10) compounds with unknown activity, and of these one (1) compound was predicted to be more active than the compounds studied. Also performed constructing molecular graphics and molecular docking studies between artemisinin and heme. Proposed by molecular docking of artemisinin and some derivatives with the hemoglobin A, a view of the binding mode between artemisinin-heme.¹⁵⁹

Ferreira et al. (2010) studied artemisinin and 18 derivatives with antimalarial activity against strains of *Plasmodium falciparum* W-2, through quantum chemical and multivariate analysis. The optimization of the geometry of the structures was carried out with the theory of Hartree-Fock (HF) and HF/3-21G** basis set. Maps of molecular electrostatic potential (MEP) and molecular docking were used to investigate the interaction between the ligands and the receptor (heme). The principal component analysis (PCA) and hierarchical cluster analysis (HCA) were used to select the most important descriptors related activity.¹⁶³

Leite and colleagues (2010) conducted studies of 18 natural compounds Brazilian flora, which have the peroxide group and presumably act in heme protein, leading to a reduction of binding of peroxide and producing radicals which can kill the etiologic agent of malaria (*Plasmodium falciparum* strains), and may show antimalarial activity. These facts motivated to study the interaction of the 18 natural peroxides, initially performing a conformational search using the MM3 method for each molecule. The most stable conformers were optimized by PM3(tm) method. Then there was a docking between peroxide and of the heme group, followed again by a conformational search. In conclusion, the results showed that four of the compounds (10, 13, 14 and 15) may be desirable antimalarial activity.¹⁶⁴

Barbosa et al. (2011) performed molecular modeling and chemometric studies involving artemisinin and 28 derivatives with anticancer activity against human hepatocellular carcinoma HepG2. The calculation of the studied compounds were performed B3LYP/6-31G** level. The electrostatic potential maps were used in an attempt to identify key structural features of artemisinin and its derivatives, that are required for its activities, and to investigate its interaction with transferrin. The chemometric method (PCA, HCA, KNN, SIMCA and SDA) were used to reduce the dimensionality and investigate which subset of variables could be more effective for the classification of compounds according to their degree of anticancer activity. Furthermore, the molecular docking was used to investigate the interaction between the ligands and

receptor. The results showed that the approximation of the ligands to the receptor is through endoperoxide binding.¹⁶⁵

Recently Figueiredo et al. (2011) developed studies antimalarial compounds with biological activity against *Plasmodium falciparum* K1. These studies have led to obtaining multivariate models for artemisinin derivatives and a series of dispiro-1,2,4-trioxolanes. A predictive model was generated by PLS method, with three latent variables explaining 99.8% of the total variance, $Q^2 = 0.87$, $R^2 = 0.85$, obtained for 16/4 molecules in the training/external validation set. The descriptors selected for the model were the binding free energy, logarithm of octanol-water partition coefficient ($\log P$) and molecular volume. The application of these models has enabled the prediction of activities compounds designed with the information obtained from studies developed. Moreover, the studies of new series of antimalarial compounds are found in the study phase which also can be inferred about the activities of new compounds to be designed.¹⁶⁶

Carvalho et al. (2011) studied with B3LYP/6-31G** level of theory the artemisinin and 31 analogues with anti-leishmanicidal activity against *Leishmania donovani*, and proposed a set of 13 artemisinins, 7 less active and 6 that have not been tested, and of these six, one was expected to be more active against *L. donovani*. In this study, maps Electrostatic Potential (MEP) were used in an attempt to identify key structural features of artemisinin and analogs, and mode of interaction with its receptor (heme). The chemometric methods: PCA, HCA, SDA, KNN and SIMCA were used to reduce dimensionality and investigate which subset of descriptors are responsible for the classification of the activity anti-leishmanicidal as larger and smaller.¹⁶⁷

Recently Cristino e colleagues (2012) used the B3LYP/6-31G* theory to model artemisinin and Nineteen 10-substituted deoxyartemisinin derivatives, with different degrees activity against strains of *Plasmodium falciparum* D-6 of Sierra Leone. The chemometric methods (PCA), (HCA), (KNN), (SIMCA) and (SDA) were employed to reduce the dimensionality and investigate which subset of descriptors are responsible for classification between more and less active deoxyartemisinin antimalarial. The chemometric methods: Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA), K-Nearest Neighbor (KNN), Soft Independent Modeling of Class Analogy (SIMCA) and discriminant analysis (SDA) were used to reduce the dimensionality and investigate which subset of descriptors are responsible for classification between more and less active.¹⁶⁸

Recent studies of 51 peroxides were made to find correlations between *in silico* parameters and experimental data for identifying new antimalarials from natural sources. The interaction of heme was studied by molecular docking refinement followed by conformational analysis using semiempirical parametric method 6 (PM6). The results indicated that compounds 5 and 24 are promising antimalarials.¹⁶⁹

Zhao (2013) conducted a study about the natures general of Dynamic of proteins and thermodynamics which were analyzed based in the theory irreversible for protein folding and in the theory protein structure thermodynamic. The main contents included:

- (1) basic concepts of irreversible thermodynamics,
- (2) the irreversible thermodynamic theory to the folding of proteins that reveals the fundamental rules of movement coupled to a protein,
- (3) thermodynamic theory of the structure of protein
- (4) the dynamic nature and thermodynamics of proteins conformational change,
- (5) the role of protein dynamics in enzymatic reaction,
- (6) the thermodynamic relationship between the protein and biological function.

Many problems of enzymology and protein science are discussed. The analysis of Zhao showed that the properties and function of the proteins could be well explained by the application of concepts dynamics and thermodynamics proteins. Dynamic of proteins and thermodynamics are fundamental to all biological processes, among them we can mention the molecular docking simulations of anti-malarial compounds with heme protein.¹⁷⁰

In recent years we can realize that studies and research related to computational quantum chemistry and molecular modeling have been growing in an exponential manner. Khrennikov has reported about a series of recent debates about the bases-quantum and incompleteness of quantum mechanics and a new classical model which reproduces the bases quantum averages and correlations, classical field theory prequantum statistics.¹⁷¹

This article reinforces that theoretical studies contribute significantly to new discoveries and research involving various areas of human knowledge, as we mention:

(1) Chang et al. (2013) has investigated theoretically the encapsulation of amino acids of a carbon nanotube single wall zigzag, and the results revealed the stability of amino acids along the inner wall of the cavity. The essential structural features were observed about the modification of the properties of encapsulated molecules as well as the surface properties. The carbon nanotube experienced has associated geometric distortions with corresponding internal molecular structure.¹⁷²

(2) Kala et al. (2013) computational calculations performed using the semi-empirical method that is computationally flexible to capture various quantitative descriptions of molecules of moderate size, and these were investigated the transport properties for the quantum heterocyclic molecules such as pyrrole, furan and thiophene, based on Green's function nonequilibrium (NEGF) formalism combined with Extended Huckel theory (EHT). In these systems, molecular transport properties are strongly influenced by the geometry of the molecule, the chain length of the molecules and their bonding strength to the electrodes attached to the side.¹⁷³

(3) Li (2013) developed a three-dimensional model to investigate the formation of nanoparticles incorporated by ion beam implantation. The nucleation and growth process dynamics, including the known maturation Ostwald, were successfully reconstructed by a theoretical model. This theoretical model gives a remarkable insight into the formation mechanism, and makes it possible to control and optimize fully the nanostructure through ion implantation technology.¹⁷⁴

9. FINAL CONSIDERATIONS

This study aimed to demonstrate the variety of types of molecular descriptors based on quantum-chemical calculations derived from the wave functions and charge distribution, and how these descriptors have been used to explain the physico-chemical and/or biological activities in studies of structure-activity relationships.

The selection of the best set of descriptors for the activity under study can be achieved by careful selection and combination among many descriptors. In many cases the quantum-chemical descriptors have physical meaning and are used to unravel the complex mechanisms details involved in intra and intermolecular interaction. It should be highlighted that the quantum-chemical descriptors may be calculated from the structure of the molecules that is, starting from a geometry optimized or determined experimentally (for example, X-rays and nuclear magnetic resonance). However, these descriptors are not completely universal, because they are dependent on the structures and systems studied. Even though it is based on a minimum of energy, the calculated descriptors have values very close to their respective empirical values and indicate trends electronic systems under study. Based on the aspects highlighted in this work is evident, therefore, that the quantum-chemical descriptors have a wide range of applications in SAR and QSAR studies, as well as in many areas of integration of fundamental knowledge of Organic Chemistry, Biochemistry, Molecular Biology, Pharmacology and Pharmaceutical Chemistry.

Thus it is justified to study and develop new derivatives with more potent biological activity using quantum chemical methods (SAR and QSAR), multivariate analysis (PCA, HCA, PLS) analysis of pattern recognition KNN, SDA and SIMCA and Molecular Docking (interaction drug-receptor), using the resulting information as a guide to obtaining of new derivatives active most promising to be synthesized.

Acknowledgments: We gratefully acknowledge the final support of the Brazilian Agency Conselho Nacional de Desenvolvimento Científico e Tecnológico. We also thank the Instituto de Ciências da Saúde of Federal University of Pará for the use of the GaussView and Gaussian softwares. The authors would like to thank the Postgraduate Program in Biotechnology and Biodiversity—Network

BIONORTE and to the Laboratório de Modelagem e Química Computacional, Federal University of Amapá for the computational support.

References and Notes

1. F. M. S. S. Fernandes, *Química Computacional: Uma seta na direção do século XXI*, edited by H. V. Caetano and M. G. Santos, *Cadernos Didáticos de Ciências*, Lisboa (2001), Chap. 4, pp. 55–74.
2. R. C. Atkinson and P. Suppes, Program in computer-assisted instruction: Final report, Washington, DC. Office of Education (DGEW), edited by M. R. Simonson and A. Thompson, Educational Computing Foundations, Prentice Hall, Upper Saddle River, New Jersey (1968).
3. P. Suppes and M. Morningstar *Science* 166, 343 (1968).
4. A. A. Ribeiro and I. M. Greca, *Quim. Nova* 26, 542 (2003).
5. B. U. Duke, *J. Chem. Educ.* 49, 703 (1972).
6. S. Colwell and N. Handy, *J. Chem. Educ.* 65, 21 (1988).
7. L. C. Freitas, *Quim. Nova na Escola* 8, 3 (1998).
8. IUPAC, *Glossary of Terms Used in Medicinal Chemistry* Iupac recommendations (1998).
9. A. Varnek, *J. Chem. Educ.* 77, 222 (2000).
10. S. A. Fleming, G. R. Hart, and P. B. Savage, *J. Chem. Educ.* 77, 790 (2000).
11. C. Rodrigues, *Cadernos Temáticos de Química Nova na Escola* 3, 43 (2001).
12. M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press (1987).
13. A. Leach, *Molecular Modelling—Principles and Applications*, Pearson Education Limited (2001).
14. C. M. R. Sant’Anna, *Quim. Nova* 25, 505 (2002).
15. W. J. Hehre, *A Guide to Molecular Mechanics and Quantum Chemical Calculations*, Wavefunction, Inc. (2003).
16. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, *J. Am. Chem. Soc.* 107, 3902 (1985).
17. J. J. P. Stewart, *J. Comput. Chem.* 10, 209 (1989).
18. G. B. Rocha, R. O. Freire, A. M. Simas, and J. J. P. Stewart, *J. Comput. Chem.* 27, 1101 (2006).
19. A. R. Leach, *Molecular Modelling Principles and Applications*, Longman, Essex (1996).
20. J. B. Foresman and A. E. Frisch, *Exploring Chemistry with Electronic Structure Methods: A Guide to Using Gaussian*, Gaussian, Inc, USA (1993).
21. N. C. Cohen, *Guidebook on Molecular Modeling in Drug Design*, Academic Press, San Diego (1996).
22. I. Carvalho, A. D. L. Borges, and L. S. C. Bernardes, *J. Chem. Educ.* 82, 588 (2005).
23. C. G. Wermuth, *The Practice of Medicinal Chemistry*, 3rd edn., Academic Press, London (2009).
24. S. O. Silva, M. J. C. Corrêa, H. R. Bitencourt, W. R. Monteiro, J. Lameira, L. S. Santos, G. M. S. P. Guilhon, and D. S. B. Brasil, *J. Comput. Theor. Nanosci.* 9, 953 (2012).
25. P. Politzer and J. S. Murray, *Theor. Chem.* 108, 134 (2002).
26. K. C. Weber, K. M. Honório, A. T. Bruni, A. D. Andricopulo, and A. B. F. Silva, *Struct. Chem.* 17, 307 (2006).
27. M. Fernández and J. Caballero, *J. Mol. Model.* 13, 465 (2007).
28. M. A. Ibrahim and A. E. A. Gawad, *J. Comput. Theor. Nanosci.* 9, 1120 (2012).
29. A. M. Bayuelo and R. V. Reyes, *J. Math. Chem.* 51, 125 (2013).
30. A. Arroio, K. M. Honório, and A. B. F. Silva, *Quim. Nova* 33, 694 (2010).
31. Cambridge Soft Corporation, CS ChemDraw 5.0, Chemical Structure Drawing Standard, Cambridge (1998).
32. Cambridge Soft Corporation, CS Chem3D 5.0, Molecular Modeling and Analysis, Cambridge (1998).
33. ACD/Chemsketch Freware, version 12.00, Advanced Chemistry Development, Inc, Toronto, ON, Canada (2010).
34. Cambridge Soft Corporation, CS Chem3D Ultra 5.0, Ultimate Modeling, Visualization and Analysis, Cambridge (1998).
35. I. Carvalho, M. T. Pupo, A. D. L. Borges, and L. S. C. Bernardes, *Quim. Nova* 26, 428 (2003).
36. P. Flukiger, Molekel 4.2, Swiss Center for Scientific Computing, Manno, Switzerland (2002).
37. Molecular Modeling ProTM4.0, Computational Chemistry Program, ChemSW®, Inc., Fairfield (2001).
38. ChemSite: Interactive 3D Molecular Modeling, ChemSW® Inc., Fairfield (2001).
39. Chemplus, Modular Extensions for HyperChem Release 6.02, Molecular Modeling for Windows, HyperClub, Inc, Gainesville (2000).
40. G. L. Patrick, *An Introduction to Medicinal Chemistry*, 2nd edn., Oxford University Press, Oxford (2001).
41. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr, T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, *Gaussian 03*, Revision C.02, Gaussian, Inc. Wallingford, CT (2004).
42. R. Todeschini and P. Gramatica, *Perspectives in Drug Discovery and Design* 355 (1998).
43. C. B. R. Santos, J. B. Vieira, A. S. Formigosa, E. V. M. Costa, M. T. Pinheiro, J. O. Silva, W. J. C. Macêdo, and J. C. T. Carvalho, *J. Comput. Theor. Nanosci.* No Prelo (2013).
44. R. E. Bruns, P. H. Guadagnini, and A. A. Souza, *Quim. Nova* 19, 148 (1996).
45. R. S. Mulliken, *J. Chem. Phys.* 2, 782 (1934).
46. L. E. Chirlian and M. M. Francl, *J. Comput. Chem.* 8, 894 (1987).
47. Bernardinelli, *Int. J. Quant. Chem.* 21, 117 (1994).
48. G. R. Martins, H. B. Napolitano, L. T. F. M. Camargo, and A. J. Camargo, *J. Braz. Chem. Soc.* 23, 2183 (2012).
49. K. M. Honório and A. B. F. Silva, *Int. J. Quant. Chem.* 95, 126 (2003).
50. R. Contreras, L. R. Domingo, J. Andrés, P. Pérez, and O. Tapia, *J. Phys. Chem.* 103, 1367 (1999).
51. B. W. Clare, *J. Mol. Struct. (Theochem)* 331, 63 (1995).
52. B. W. Clare, *J. Mol. Struct. (Theochem)* 337, 139 (1995).
53. B. W. Clare, *Theor. Chim. Acta* 87, 415 (1994).
54. C. A. Heaton, A. K. Miller, and R. L. Powell, *J. Fluorine Chem.* 107, 1 (2001).
55. G. H. Grant and W. G. Richards, *Computational Chemistry*, Oxford Science Publications (1996).
56. G. Zhang and C. B. Musgrave, *J. Phys. Chem.* 111, 1554 (2007).
57. C. C. Lobato, E. M. Silva, J. B. Vieira, W. J. C. Macêdo, E. Viana, J. C. T. Carvalho, and C. B. R. Santos, *Estudo computacional da reatividade dos produtos isoméricos em reações de halidificação através da análise dos orbitais de fronteiras—HOMO e LUMO*. In 52º Congresso Brasileiro de Química—Química e Inovação: Caminho Para a Sustentabilidade, Recife-PE, Brasil (2012).
58. D. F. Shriver, P. W. Atkins, and C. H. Langford, *Inorganic Chemistry*, 2nd edn., Oxford University, Oxford (1994).

59. R. G. Parr and P. K. Chattaraj, *J. Am. Chem. Soc.* 113, 1854 (1991).
60. A. V. Bunge, *Introdução à Química Quântica*, Edgard Blucher Ltda, São Paulo (1977).
61. R. G. Pearson and W. E. Palke, *J. Phys. Chem.* 96, 3283 (1992).
62. C. C. Lobato, R. D. Trindade, J. E. M. Diniz, and C. B. R. Santos, *Análise Computacional dos Produtos Isoméricos de Cadeia da Reação de Halidificação em Alcenos*. In 51° Congresso Brasileiro de Química-Meio Ambiente e Energia, São Luís-MA, Brasil (2011).
63. J. A. Hirsch, *Concepts in Theoretical Organic Chemistry* Allynand Baco, Boston (1975).
64. E. J. Lien, Z. R. Guo, and C. T. Su, *J. Pharm. Sci.* 71, 641 (1982).
65. I. N. Levine, *Quantum Chemistry*, 4th edn., Prentice Hall, New York (1991).
66. A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronics Structure Theory*, Macmillan, New York (1982).
67. C. Hansch, R. M. Muir, T. Fujita, P. P. Maloney, C. F. Geiger, and M. J. Streich, *J. Am. Chem. Soc.* 85, 2817 (1963).
68. C. Hansch and T. Fujita, *J. Am. Chem. Soc.* 86, 1616 (1964).
69. S. M. Free and J. W. Wilson, *J. Med. Chem.* 7, 395 (1964).
70. F. A. Pasha, H. K. Srivastava, and P. P. Singh, *Bioorg. Med. Chem.* 13, 6823 (2005).
71. K. M. Honório, R. C. Garratt, I. Polikarpov, and A. D. Andricopulo, *J. Mol. Graph. Model* 25, 921 (2007).
72. K. C. Weber, K. M. Honório, A. D. Andricopulo, and A. B. F. Silva, *Med. Chem.* 4, 328 (2008).
73. J. Padmanabhan, R. Parthasarathi, V. Subramanian, and P. K. Chattaraj, *Bioorg. Med. Chem.* 14, 1021 (2006).
74. J. Ghasemi, S. Saaidpour, and S. D. Brown, *J. Mol. Struct. (Theochem)* 805, 27 (2007).
75. A. A. Toropov, A. P. Toropova, and E. Benfenati, *Chem. Phys. Lett.* 461, 343 (2008).
76. M. Karelson and V. S. Lobanov, *Chem. Rev.* 96, 1027 (1996).
77. K. C. Weber, K. M. Honório, S. L. Silva, R. Mercadante, and A. B. F. Silva, *Int. J. Quantum Chem.* 103, 731 (2005).
78. F. A. Molfetta, A. T. Bruni, K. M. Honório, and A. B. F. Silva, *Eur. J. Med. Chem.* 40, 329 (2005).
79. K. M. Honório, R. C. Garratt, and A. D. Andricopulo, *Bioorg. Med. Chem. Lett.* 15, 3119 (2005).
80. K. M. Honório, R. C. Garratt, I. Polikarpov, and A. D. Andricopulo, *Lett. Drug Des. Disc.* 3, 261 (2006).
81. K. M. Honório, R. C. Garratt, I. Polikarpov, and A. D. Andricopulo, *J. Mol. Graph. Model* 25, 921 (2007).
82. A. K. Debnath, *Mini-Rev. Med. Chem.* 1, 187 (2001).
83. A. Vedani and M. Dobler, *J. Med. Chem.* 45, 2139 (2002).
84. R. D. Cramer, D. E. Patterson, and J. D. Bunce, *J. Am. Chem. Soc.* 110, 5959 (1988).
85. G. Klebe, U. Abraham, and T. Mietzner, *J. Med. Chem.* 37, 4130 (1994).
86. A. J. Hopfinger, S. Wang, J. S. Tokarski, B. Jin, M. Albuquerque, P. J. Madhav, and C. Duraiswami, *J. Am. Chem. Soc.* 119, 10509 (1997).
87. A. Vedani and M. Dobler, *J. Med. Chem.* 45, 2139 (2002).
88. A. Vedani, M. Dobler, and M. A. Lill, *J. Med. Chem.* 48, 3700 (2005).
89. F. Tian, P. Zhou, and Z. Li, *J. Mol. Struct.* 871, 7140 (200).
90. M. P. Freitas, S. D. Brown, and J. A. Martins, *J. Mol. Struct.* 738, 149 (2005).
91. M. P. Freitas, *Org. Biomol. Chem.* 4, 1154 (2006).
92. M. P. Freitas, *Curr. Comput.-Aid. Drug Des.* 3, 235 (2007).
93. M. P. Freitas, *Chemom. Intell. Lab. Sys.* 91, 173 (2008).
94. M. Goodarzi and M. P. Freitas, *Chemom. Intell. Lab. Sys.* 96, 59 (2009).
95. M. Goodarzi and M. P. Freitas, *Mol. Simul.* 36, 267 (2010).
96. R. A. Cormanich, M. P. Freitas, and R. Rittner, *J. Braz. Chem. Soc.* 22, 637 (2011).
97. M. Goodarzi and M. P. Freitas, *J. Phys. Chem. A.* 112, 11263 (2008).
98. M. Goodarzi, M. P. Freitas, and T. C. Ramalho, *Spectrochim. Acta Part. A.* 74, 563 (2009).
99. M. Goodarzi and M. P. Freitas, *Separ. Purif. Technol.* 68, 363 (2009).
100. R. A. Cormanich, C. A. Nunes, and M. P. Freitas, *Quim. Nova* 35, 1157 (2012).
101. R. Franke and A. Gruska, *Chemometrics Methods in Molecular Design* 115 (1996).
102. A. J. Camargo, R. Mercadante, K. M. Honório, C. N. Alves, and A. B. F. Silva, *J. Mol. Struct. (Theochem)* 583, 105 (2002).
103. J. Souza, R. H. D. Santos, M. M. C. Ferreira, F. A. Molfetta, A. J. Camargo, K. M. Honório, and A. B. F. Silva, *Eur. J. Med. Chem.* 38, 929 (2003).
104. J. Souza, F. A. Molfetta, K. M. Honório, R. H. A. Santos, A. B. F. Silva, *J. Chem. Inf. Comput. Sci.* 44, 1153 (2004).
105. K. M. Honório and A. B. F. Silva, *J. Mol. Mod.* 11, 200 (2005).
106. A. K. Calgarotto, S. Miotto, K. M. Honório, A. B. F. Silva, S. Marangoni, J. L. Silva, M. Comar, K. M. T. Oliveira, and S. L. Silva, *J. Mol. Struct. (Theochem)* 808, 25 (2007).
107. K. C. Weber, K. M. Honório, A. T. Bruni, and A. B. F. Silva, *J. Mol. Mod.* 12, 915 (2006).
108. A. Arroio, K. M. Honório, and A. B. F. Silva, *J. Mol. Struct. (Theochem)* 709, 223 (2004).
109. F. A. Molfetta, K. M. Honório, C. N. Alves, and A. B. F. Silva, *J. Mol. Struct. (Theochem)* 674, 191 (2004).
110. A. J. Camargo, K. M. Honório, R. Mercadante, F. A. Molfetta, C. N. Alves, and A. B. F. Silva, *J. Braz. Chem. Soc.* 14, 809 (2003).
111. V. R. S. Malta, A. V. Pinto, F. A. Molfetta, K. M. Honório, C. A. Simone, M. A. Pereira, R. H. A. Santos, and A. B. F. Silva, *J. Mol. Struct. (Theochem)* 634, 271 (2003).
112. A. Arroio, E. F. Lima, K. M. Honório, and A. B. F. Silva, *Struct. Chem.* 20, 577 (2009).
113. C. N. Alves, J. C. Pinheiro, A. J. Camargo, M. M. C. Ferreira, and A. B. F. Silva, *J. Mol. Struct. (Theochem)* 530, 39 (2000).
114. C. N. Alves, J. C. Pinheiro, A. J. Camargo, M. M. C. Ferreira, R. A. F. Romero, and A. B. F. Silva, *J. Mol. Struct. (Theochem)* 541, 81 (2001).
115. M. M. C. Ferreira, *J. Braz. Chem. Soc.* 13, 742 (2002).
116. M. M. C. Ferreira, C. A. Montanari, and A. C. Gaudio, *Quim. Nova* 25, 439 (2002).
117. H. Kubinyi, *Methods and Principles in Medicinal Chemistry* 1, 21 (1993).
118. S. Wold, *Chemometrics Methods in Molecular Design* 202 (1996).
119. A. C. Gaudio and E. Zandonade, *Quim. Nova* 25, 658 (2001).
120. B. Barros Neto, I. S. Scarminio, and R. E. Bruns, *Planejamento e Otimização de Experimentos*, 2nd edn., Campinas, UNICAMP (1996), pp. 152–298.
121. D. T. Stanton, *J. Chem. Inf. Comput. Sci.* 43, 1423 (2003).
122. M. Sjöström and L. Ricksson, *Chemometrics Methods in Molecular Design* 68 (1996).
123. G. L. Patrick, *An Introduction to Medicinal Chemistry* 3, 293 (2006).
124. R. G. Brereton, *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*, John Wiley & Sons, Ltd. (2003).
125. W. P. Gardiner, *Statistical Analysis Methods for Chemists*, Royal Society Chemistry, Cambridge (1997).
126. Brasil, Ministerio da Saude, Secretaria de Vigilancia em Saude, Departamento de Vigilancia Epidemiologica, *Guia pratico de tratamento da malaria no Brasil*, Brasília (2012).
127. M. Srivastava, H. Singh, and P. K. Naik, *Curr. Res. J. Biol. Sci.* 2, 83 (2010).
128. A. R. Butler, and Y.-L. Wu, *Chem. Soc. Rev.* 21, 85 (1992).
129. E. Sun and F. E. Cohen, *Gene* 137, 127 (1993).
130. J. G. Lombardino and J. A. Lowe, III, *Nature Reviews-Drug Discovery* 3, 853 (2004).

131. Bowman, D. Lawson, D. Bashan, D. Brown, T. Chillingworth, C. M. Churcher, A. Craig, R. M. Davies, K. Deulin, T. Feltwell, S. Gentes, R. Gwilliam, N. Ham, D. Harnis, S. Holroyd, T. I-Iornsby, P. Horrocks, K. Jageis, B. Jassal, S. J. Kyes, S. McLean, S. Moule, K. Mungail, L. Murphy, K. Oliver, M. A. Quali, M.-A. Rajadream, S. Rutter, J. Skelton, R. Squares, S. Squares, J. E. Su, S. Whitehead, J. R. Woodward, C. Nawbold, and B. G. Barreil, *Nature* 400, 532 (1999).
132. A. Korolkovas, *Ciênc. e Cult.* 41, 528 (1989).
133. N. C. Cohen, J. M. Blaney, C. Humblet, P. Gund, and D. C. Bany, *J. Med. Chem.* 33, 883 (1990).
134. A. Sali and T. L. Blundell, *J. Mol. Biol.* 234, 779 (1993).
135. M. A. Avery, *J. Med. Chem.* 36, 4264 (1993).
136. S. K. Kokpol, *Analytical Sciences* 4, 565 (1998).
137. E. J. Barreiro and C. A. M. Fraga, *Química Medicinal: As bases moleculares da ação dos fármacos*, Artmed, Porto Alegre (2001).
138. R. A. Toledo, M. C. Santos, K. M. Honório, A. B. F. Silva, E. T. G. Cavalheiro, and L. H. Mazo, *Anal. Lett.* 39, 507 (2006).
139. R. A. Toledo, M. C. Santos, K. M. Honório, A. B. F. Silva, E. T. G. Cavalheiro, and L. H. Mazo, *Quim. Nova* 28, 456 (2005).
140. H. B. Suffredini, M. C. Santos, D. Souza, L. Codognoto, P. Homem-de-Mello, K. M. Honório, A. B. F. Silva, S. A. S. Machado, and L. A. Avaca, *Anal. Lett.* 38, 1587 (2005).
141. P. Labute, *J. Mol. Graphics Model* 18, 464 (2000).
142. D. L. Klayman, *Science* 228, 1049 (1985).
143. R. K. Haynes and S. C. Vonwiller, *Tetrahedron. Lett.* 37, 253 (1996).
144. M. Jung, *Curr. Med. Chem.* 1, 35 (1994).
145. J. N. Cumming, P. Ploypradith, and G. H. Posner, *A. Pharmacol* 37, 253 (1997).
146. World Health Organization, *Guidelines for the Treatment of Malaria* (2006).
147. S. Tonmuphean, V. Parasuk, and S. Kokpol, *Quant. Struct-Act. Relat.* 19, 475 (2000).
148. S. Tonmuphean, V. Parasuk, and S. Kokpol, *Quant. Struct-Act. Relat.* 19 (2000).
149. S. Tonmuphean, V. Parasuk, and S. Kokpol, *J. Mol. Model* 7, 26 (2001).
150. G. H. Posner, J. N. Cummings, P. Ploypradith, and C. O. Oh, *J. Am. Chem. Soc.* 117, 5885 (1995).
151. W. Asawamasakda, I. Ittarat, Y.-M. Pu, H. Ziffer, and S. R. Meshnick, *Antimicrob. Agents Chemother.* 38, 1854 (1994).
152. C. W. Jefford, *Curr. Med. Chem.* 8, 1803 (2001).
153. P. L. Olliaro, *Trends in Parasitology* 17, 122 (2001).
154. N. H. Morgon and K. Coutinho, *Métodos de química teórica e modelagem molecular*, Livraria da física (2007), pp. 491–527.
155. R. S. Judson, E. P. Jaegar, A. M. Treasurywala, and M. L. Peterson, *J. Comput. Chem.* 14, 1407 (1993).
156. G. Jones, *J. Mol. Biol.* 267, 727 (1997).
157. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, *The Protein Data Bank. Nucleic Acids Research* 28, 235 (2000).
158. F. Cheng, *Bioorg. Med. Chem.* 10, 2883 (2002).
159. J. C. Pinheiro, *QSAR Comb. Sci.* 22, 830 (2003).
160. R. G. Ridley, *Science* 424, 887 (2003).
161. C. Thomson, M. Cory, and M. Zerner, *Int. J. Quant. Chem.* 18, 231 (1991).
162. J. C. Pinheiro, M. M. C. Ferreira, and O. A. S. Romero, *J. Mol. Struct. (Theochem)* 572, 35 (2001).
163. J. E. V. Ferreira, A. F. Figueiredo, J. P. Barbosa, M. G. G. Cristino, W. J. C. Macêdo, O. P. P. Silva, B. V. Malheiros, R. T. Serra, and J. C. Pinheiro, *J. Serb. Chem. Soc.* 75, 1533 (2010).
164. F. H. A. Leite, A. G. Taranto, M. C. Santos Junior, A. Branco, M. T. Araujo, and J. W. M. Carneiro, *Int. J. Quant. Chem.* 110, 2057 (2010).
165. J. P. Barbosa, J. E. V. Ferreira, A. F. Figueiredo, R. C. O. Almeida, O. P. P. Silva, J. R. C. Carvalho, M. G. G. Cristino, J. C. Pinheiro, J. L. F. Vieira, and R. T. A. Serra, *J. Serb. Chem. Soc.* 76, 1263 (2011).
166. A. F. Figueiredo, J. E. V. Ferreira, J. P. Barbosa, W. J. C. Macedo, M. G. G. Cristino, M. S. Lobato, J. C. Pinheiro, R. T. A. Serra, *J. Comput. Theor. Nanosci.* 8, 1847 (2011).
167. J. R. C. Carvalho, J. E. V. Ferreira, J. P. Barbosa, M. S. Lobato, C. C. F. Meneses, M. M. Soeiro, M. S. Farias, R. C. O. Almeida, K. C. Ventura, J. C. Pinheiro, and A. L. R. Pinheiro, *J. Comput. Theor. Nanosci.* 8, 2193 (2011).
168. M. G. G. Cristino, C. C. F. Meneses, M. M. Soeiro, J. E. V. Ferreira, A. F. Figueiredo, J. P. Barbosa, R. C. O. Almeida, J. C. Pinheiro, and A. L. R. Pinheiro, *J. Theor. Comput. Chem.* 11, 241 (2012).
169. F. H. A. Leite, J. W. M. Carneiro, M. T. Araujo, M. Comar, and A. G. Taranto, *Int. J. Quant. Chem.* 112, 3390 (2012).
170. Q. Zhao, *Reviews in Theoretical Science* 1, 83 (2013).
171. A. Khrennikov, *Reviews in Theoretical Science* 1, 34 (2013).
172. C. M. Chang, H. L. Tseng, A. de Leon, A. Posada-Amarillas, and A. F. Jalbout, *J. Comput. Theor. Nanosci.* 10, 521 (2013).
173. C. P. Kala, P. A. Priya, and D. J. Thiruvadigal, *J. Comput. Theor. Nanosci.* 10, 213 (2013).
174. K. D.- Li, *J. Comput. Theor. Nanosci.* 10, 629 (2013).



Validation of Computational Methods Applied in Molecular Modeling of Artemisinin with Antimalarial Activity

Cleydson Breno Rodrigues dos Santos^{1,2,3,*}, Josinete Braga Vieira³,
Adriane da Silva Formigosa³, Elizabeth Viana Moraes da Costa¹,
Mayara Tânia Pinheiro^{1,2}, Jocivânia Oliveira da Silva^{1,3}, Williams Jorge
da Cruz Macêdo⁴, and José Carlos Tavares Carvalho^{1,2}

¹Laboratory of Drug Research, School of Pharmaceutical Sciences, Federal University of Amapá, Rod JK km 2, Macapá-Amapá, 68902-280, Brazil

²Postgraduate Program in Biotechnology and Biodiversity-Network BIONORTE, Federal University of Amapá, Rod JK km 2, Macapá-Amapá, 68902-280, Brazil

³Laboratory of Modelling and Computational Chemistry, Federal University of Amapá, Rod JK km 2, Macapá, Amapá 68902-280, Brazil

⁴Secretary of State for Public Health, Central Laboratory of the State of Pará-LACEN/PA, Rod Augusto Montenegro, km 10, Icoaraci, 66823-010, Belém-Pará-Amazon-Brazil

Artemisinin (qinghaosu) has a unique structure bearing a stable endoperoxide lactone (1,2,13-trioxane) totally different from previous antimalarial in its structure and mode of action, are isolated from *Artemisia annua* is a composed of a remarkable life and antimalarial effective against *Plasmodium falciparum* and cerebral malaria. We propose a combination of chemical quantum methods and multivariate analysis to study the geometric parameters of artemisinin in the region endoperoxide of the ring (1,2,13-trioxane), in order to be effective in selecting the method and level of theory when compared with data crystallographic, aiming to classify and correlate. The most important geometrical parameters selected by principal component analysis (PCA) were O13C12, O1O2C3, C3O13C12C12a and C12C12aO1O2. The results of PCA showed that the model was built with three main components (3PCs), explains 97.0861% of the total variance. The level of theory HF/6-31G** show high similarity with the experimental data assuming that the combination of *ab initio* method can be used for modeling the molecular structure of artemisinin and its derivatives antimalarial drugs with mechanism of action in the ring region endoperoxide.

Keywords: Artemisinin, HF/6-31G**, Molecular Modeling, Quantum Chemical Methods, Multivariate Analysis.

1. INTRODUCTION

The evolution of computational chemistry is one of the most important advances in the design and discovery of new drugs has been the use of molecular modeling (MM). Currently, MM is an indispensable tool not only in the process of drug discovery, but also the optimization of existing prototypes and the rational design of drug candidates.¹⁻⁴ According to IUPAC, the MM is the investigation of molecular structures and properties by the use of computational chemistry and graphical visualization techniques in

order to provide a three-dimensional representation under a given set of circumstances.² The nature of the molecular properties used and the extent to which they describe the structural features of molecules can be related to biological activity, which is an important part of any QSAR studies.

The molecular properties represent a means of chemical information, contained in the molecular structure of the compound. This information is transformed and codified for lot of problems chemical, pharmacological and toxicological of the relationship between quantitative structure-activity and structure-property studies (QSAR, QSPR). The molecular properties take into account different aspects

* Author to whom correspondence should be addressed.

of chemical information, this information may be through experiments, theoretical calculations or simple counting, considering the entire molecule, fragments or functional groups, knowledge of 3-D structure of the molecule or its molecular graphics or simply its formula, information defined by scalar values, vectors or scalar fields.⁵

The great development of MM in recent years was due largely to the advancement of computational resources in terms of hardware (speed of computation) and software (computer programs), in addition to advances in computational chemistry, nuclear magnetic resonance, ray crystallography-X, biochemistry and molecular biology. This allowed major contribution to the discovery of drug candidates, leading to rapid progress in research and attracting the interest of both academia and pharmaceutical industries.^{1,4}

Silva et al.⁶ used density functional theory (DFT) calculations (B3PW91/DGDZVP) to determine ¹³C and ¹H nuclear magnetic resonance (NMR) chemical shifts for the two dihydrochalcones: 3,4,5-trimethoxydihydrochalcone and 2,3,4,4-tetramethoxydihydrochalcone. The experimental and theoretical NMR data were analyzed by simple linear regression and the more relevant parameters were selected. In addition, other statistical parameters (correlation coefficients, significance and predictability) were available to judge the quality of the calculations. Finally, the statistical analysis show good correlation experimental and theoretical NMR data with high predictive power.

Ibrahim et al.⁷ used Semiempirical molecular modeling technique is applied to assess the interaction of amino acids (alanine, asparagines, aspartic, arginine, cysteine, glutamine, glycine and tryptophan) with chitosan. Results indicate the selectivity of chitosan furthermore we introduce the site whereas amino acid could interact with chitosan. Chitosan is interacting with amino acid through NH₂ group. It is concluded that chitosan is acting with amino acids like protein interaction which dedicate chitosan for many applications in the biological system.

The Malaria is one of the most common diseases in tropical countries. More than 300 million infections and millions of malaria deaths occur annually worldwide. The rapid spread of resistance to current quinoline antimalarial has made malaria a major global problem, so it is essential to seek new drugs against malaria and understand its mechanism of action for treating patient.⁸

The treatment of malaria has been used in traditional Chinese medicine for more than two million years. The medicine used is the artemisinin (qinghaosu) and is extracted from the plant *Artemisia annua* L, used to combat diseases of 52 species of the People's Republic of China. Artemisinin (qinghaosu) has a unique structure (Fig. 1) bearing a stable endoperoxide lactone (1,2,13-trioxane) totally different from previous antimalarial in its structure and mode of action, are isolated from *Artemisia annua* is a composed of a remarkable life and antimalarial effective against *Plasmodium falciparum* and cerebral malaria.^{8,9}

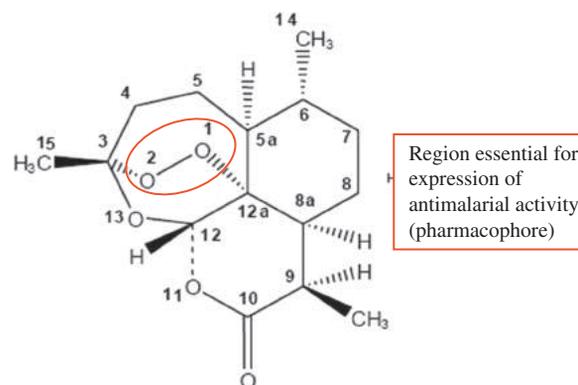


Fig. 1. Artemisinin (structure).

Artemisinin and its derivatives induce a rapid reduction of the number of parasites when compared with other known drugs. Consequently, they are of particular interest to severe malaria. The first decline in the number of parasites is also beneficial for combination therapies. This led to an enormous interest in the mechanism of action, chemistry and drug development of this new class of antimalarials. The group endoperoxide is essential for the antimalarial activity and is mediated by activated oxygen (superoxide, H₂O₂ and/or hydroxyl radicals) or carbon free-radicals.⁹⁻¹²

Carvalho et al.¹³ studied with B3LYP/6-31G** level of theory the artemisinin and 31 analogues with anti leishmanicidal activity against *Leishmania donovani*, and proposed a set of 13 artemisinins, 7 less active and 6 that have not been tested, and of these six, one was expected to be more active against *L. donovani*. In this study, maps Electrostatic Potential (MEP) were used in an attempt to identify key structural features of artemisinin and analogs, and mode of interaction with its receptor (heme). The chemometric methods: PCA, HCA, SDA, KNN and SIMCA were used to reduce dimensionality and investigate which subset of descriptors are responsible for the classification of the activity anti leishmanicidal as major and minor.

Figueiredo et al.¹⁴ studied by computational chemistry dispiro-1,2,4-trioxolanes with antimalarial activities against *K-1* strains of *Plasmodium falciparum*. Molecules were optimized with B3LYP/631G* method. A predictive model was generated by PLS method, with three latent variables explaining 99.8% of the total variance, $Q^2 = 0.87$, $R^2 = 0.85$, obtained for 16/4 molecules in the training/external validation set. The descriptors selected for the model were the binding free energy, logarithm of octanol-water partition coefficient ($\log P$) and molecular volume.

In this work, molecular modeling of the artemisinin with activities antimalarial, tested *in vitro* against human malaria *Plasmodium falciparum*, was modeled and constructed by following the strategy primarily based on the knowledge that the group endoperoxide the artemisinin present is responsible for its antimalarial activity. The artemisinin molecule was modeled in three levels of theory

and methods, were performed principal component analysis (PCA) and hierarchical cluster analysis (HCA), in order to evaluate which geometric parameters the artemisinin in the region of ring endoperoxide (1,2,13-trioxane), may be more effective in the choice of a method and theory levels studied, as compared to the crystallographic data.

2. METHODOLOGY

2.1. Molecular Modeling

The molecular modeling of artemisinin was built following the strategy described: initially the structure of artemisinin was performed with the program GaussView 3.0,¹⁵ and optimized with methods and levels of theory different-Semi-empirical (AM1, PM3 and ZINDO), *ab initio*/Hartree-Fock (HF/6-31G, HF/6-31G* and HF/6-31G**) and DFT (B3LYP/3-21G, B3LYP/3-21G*, B3LYP/3-21G**), implemented with the program Gaussian 03.¹⁶ These calculations were performed to find the method and level of theory with the best fit between the computational time and accuracy of the information on the experimental data.¹⁷ The experimental structure of artemisinin was taken from the Cambridge Structural Database CSD, with REFCODES: QNGHSU10, crystallographic *R* factor 3.6.¹⁸ The numbering of the atoms used in this study is shown in Figure 1 (artemisinin).

3. CHEMOMETRICS

3.1. Principal Component Analysis (PCA) e Hierarchical Cluster Analysis (HCA)

The multivariate analysis is when measurements are made on a number of objects, the results are usually arranged in a matrix, which is called the data matrix. The measures (in this study the geometrical parameters, concerning the methods and different levels of theory) are placed in columns, and objects (in this study the quantum-chemical methods applied) are associated with the lines. The step multivariate analysis was accomplished with the program Pirouette 3.10.¹⁹

Principal component analysis (PCA) is a method of data compression based on the correlation between the variables, the compression of data generates a small set of variables which are known as principal components, and they are mutually orthogonal, so that the correlation between the variables does not limit its application, unlike the multiple linear regression that is sensitive to the presence of highly correlated variables because the regression coefficients become unstable and meaningless.^{20,21} For a multidimensional data matrix, multivariate statistical methods are needed for standardization autoscaled employ in order to treat all variables with the same degree of importance for understanding such data in its entirety. The main objective of the analysis is to show the data in a

multidimensional space of low dimensionality with minimal loss of information overall.²²

The PCA was performed with autoscaled processing, with a maximum of three factors (3 PCS) using the procedure “leave-one-out method validation and cross-validation.” The data matrix was constructed with dimensions 10×18 , where each column was associated with three (03) methods and nine (09) level of theory, one (01) of columns relating to the geometric parameters experimental,¹⁷ and each line 18 represents the geometric parameters of 1,2,13-trioxane ring (bond lengths, bond angles, and torsion angles). The final result of PCA is to select a small number of geometric parameters which artemisinin can best be related to dependent variable, in this case the standard deviation of the various methods and levels of theory.

As in PCA results of hierarchical cluster analysis (HCA) are qualitative, being arranged in the form of a dendrogram thus view the methods studied or the variable (geometric parameters of artemisinin) in a two-dimensional space, which illustrates the merger or divisions made in each successive stage analysis. Samples (methods and levels of theory) are represented by the bottom branch of the dendrogram. The similarity between the agglomerates is given by the length of its branches, so that the methods and theory has low levels of similarity have long branches, while such methods and theory levels of high similarity have short branches.²⁰ In HCA the distance between these variables is calculated and transformed into a similarity matrix *S*. A hierarchical cluster analysis aims to show the data in such a way to accentuate their natural groupings and patterns. The statistical analysis required in this study to group the methods and theory levels similarly in their respective categories. HCA is a statistical method developed for this purpose. HCA was performed with autoscaled processing, with euclidean distance and the incremental method.

4. RESULTS AND DISCUSSION

4.1. Method and Basis Set for the Description of the Geometries of Artemisinin

We determined the theoretical geometric parameters of artemisinin in the region endoperoxide of the ring 1,2,13-trioxane (bond length, bond angle and angle of twist of forming this ring atoms), with objective of assess the quality of the wave function of the molecular theoretical and experimental geometrical parameters of the ring 1,2,13-trioxane artemisinin shown in Table I.

According to Table I shows that the method *ab initio*/Hartree-Fock all three sets of base (HF/6-31G**, HF/6-31G*, HF/6-31G) describe well all the structural parameters in magnitude and sign when compared with the experimental values, contrary the semi-empirical methods

Table I. Theoretical and experimental parameters of the 1,2,13-trioxane ring in artemisinin.

Parameters ^a	Semi-empirical			Hartree-fock/HF			DFT/B3LYP			Experimental ¹⁷
	AM1 ²³	PM3 ²³	ZINDO ²³	6-31G ²³	6-31G* ²⁴	6-31G** ^b	3-21G ²⁵	3-21G* ²⁵	3-21G** ²⁵	
Bond length (Å)										
O1O2	1.288	1.544	1.237	1.447	1.391	1.390	1.524	1.524	1.524	1.469
O2C3	1.447	1.403	1.400	1.435	1.393	1.396	1.455	1.455	1.454	1.416
C3O13	1.427	1.428	1.396	1.435	1.388	1.408	1.473	1.473	1.472	1.445
O13C12	1.416	1.403	1.392	1.403	1.400	1.376	1.430	1.430	1.430	1.379
C12C12a	1.537	1.555	1.513	1.533	1.533	1.532	1.535	1.535	1.535	1.523
C12aO1	1.468	1.426	1.416	1.469	1.429	1.429	1.504	1.504	1.504	1.461
Bond angle (°)										
O1O2C3	112.530	110.340	114.310	108.800	106.100	109.460	105.590	105.590	105.480	108.100
O2C3O13	103.600	104.810	105.370	106.760	110.800	107.800	108.220	108.220	108.250	106.600
C3O13C12	115.480	116.010	115.843	117.300	112.800	115.300	113.200	113.200	113.200	114.200
O13C12C12a	113.510	115.200	113.270	112.280	108.700	112.300	113.300	113.300	113.230	114.500
C12C12aO1	111.070	113.180	107.290	110.910	110.500	110.545	112.410	112.410	112.470	110.700
C12aO1O2	113.740	112.290	118.380	113.240	112.700	112.700	109.620	109.620	109.590	111.200
Torsion angle (°)										
O1O2C3O13	-77.800	-73.310	-70.403	-71.840	-73.369	-73.400	-76.610	-76.610	-76.740	-75.500
O2C3O13C12	42.070	52.700	36.370	33.390	31.034	31.100	33.750	33.750	33.720	36.000
C3O13C12C12a	11.400	2.811	17.420	25.320	27.432	27.400	29.059	29.060	29.080	25.300
O13C12C12aO1	-41.770	-40.510	-46.610	-49.410	-50.100	-50.143	-52.190	-52.190	-52.030	-51.300
C12C12aO1O2	12.050	19.940	18.110	12.510	10.900	10.924	9.060	9.600	9.340	12.700
C12aO1O2C3	47.050	35.630	40.130	46.700	48.700	48.674	51.060	51.060	51.320	47.800
Standard deviation	4.776	8.388	4.372	1.663	2.484	1.762	1.915	1.855	1.987	-

Notes: ^aThe atoms are numbered according to Figure 1; ^bValence basis set separately validated for obtaining the calculations of molecular properties; ¹⁷J. N. Lisgarten, et al., *Journal of Chemical Crystallography* 28, 539 (1998); ²³J. C. Pinheiro, et al., *Journal Molecular Structure (THEOCHEM)* 572, 35 (2001); ²⁴C. Thomsom, et al., *International Journal of Quantum Chemistry: Quantum Biology Symposium* 18, 231 (1991); ²⁵A. D. Becke, *J. Chem. Phys.* 98, 5648 (1993).

(AM1, PM3 and ZINDO) and DFT (B3LYP/3-21G, B3LYP/3-21G*, B3LYP/3-21G**), which show the standard deviation 4.776, 8.388 and 4.372 for the semi-empirical and 1.915, 1.855 and 1.987 for the DFT, respectively. When comparing the method *ab initio*/Hartree-Fock, we can see that both levels of theory (6-31G and 6-31G**) have the lowest standard deviations in relative to semi-empirical methods and DFT, with values of 1.663 and 1.762, respectively, having a variation of about 0.099 between them.

4.2. PCA Method

The theoretical and experimental parameters of the 1,2,13-trioxane ring of artemisinin were used with objective to identify, through PCA and HCA, which optimize the geometry of artemisinin in different methods and levels of theory defines the results closer to experimental data. The advantage of using PCA and HCA methods in the present study was that all structural parameters are considered simultaneously, and that takes into account the correlations between them.

The PCA results showed that descriptors four (04) most important related to ring trioxane were O13C12, O1O2C3, C3O13C12C12a and C12C12aO1O2. The values of geometric parameters selected by principal component analysis, standard deviation and Pearson correlation matrix are shown in Table II. These geometric parameters

are responsible for the separation of methods and levels in three classes: semi empirical, *ab initio*/Hartree-Fock and DFT. The four (04) geometrical parameters related to trioxane ring are identified by atoms with the measures O13C12, which is the interplanar distance between the two atoms (bond length), O1O2C3 is the angle of bond between these three atoms, C3O13C12C12a and C12C12aO1O2 are related or dihedral angle of torsion.

Table II shows the correlation matrix between the geometric parameters and the standard deviation, it is noted that the correlation between the geometric parameters is less or equal to 0.827, while the correlation between the geometric parameters and the standard deviation is less or equal to 0.914. The geometric parameters selected with the PCA technical represent characteristics necessary to validate the best method and level of theory.

The results of model selection are shown in Table III, and show that the model was constructed with three main components (3PCs), where the first principal component (PC1) describes 24.3990% of the total information, the second main component (PC2) describes 8.0043%, and third (PC3) 2.5477%. Further in this table, we observe that PC1 contains 67.7751% of the original data, the first two (PC1 + PC2) 90.0093% and first three (PC1 + PC2 + PC3) can explain 97.0861% of the total information, losing only 2.9139% of original information. In the same table, it is verified that the geometrical parameters O1O2C3 (0.5447) and C12C12aO1O2 (0.5657) are the main contributors

Table II. Geometric parameters selected by principal component analysis, standard deviation and Pearson correlation matrix.

Methods	O13C12	O1O2C3	C3O13C12C12a	C12C12aO1O2	Standard deviation
AM1	1.416	112.530	11.400	12.050	4.776
PM3	1.403	110.340	2.811	19.940	8.388
ZINDO	1.392	114.310	17.420	18.110	4.372
6-31G	1.403	108.800	25.320	12.510	1.663
6-31G*	1.400	106.100	27.432	10.900	2.484
6-31G**	1.376	109.460	27.400	10.924	1.762
3-21G	1.430	105.589	29.059	9.060	1.915
3-21G*	1.430	105.590	29.060	9.600	1.855
3-21G**	1.430	105.480	29.080	9.340	1.987
Experimental	1.379	108.100	25.300	12.700	0.000
O13C12		-0.465	0.140	-0.418	0.068
O13C3			-0.700	0.727	0.540
C3O13C12C12a				-0.827	-0.914
C12C12aO1O2					0.758

to PC1, while that the geometrical parameters O13C12 (0.8771) and C12C12aO1O2 (0.1213) are the main contributors to PC2.

As the principal components can be written as a linear combination of selected geometrical parameters, mathematical expressions for PC1 and PC2 are shown below:

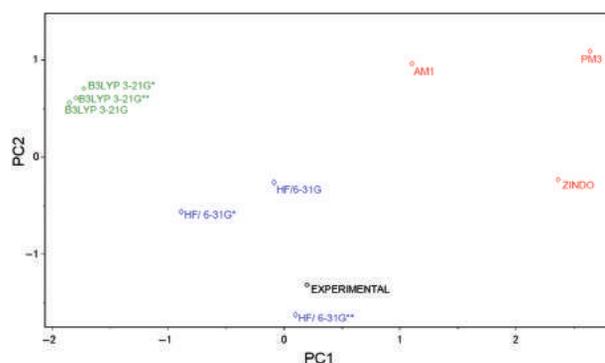
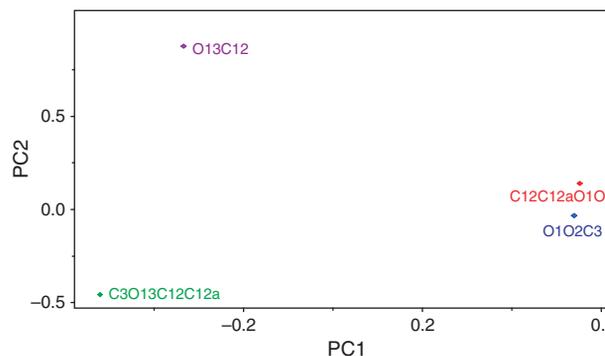
$$PC1 = -0.3296(O13C12) + 0.5447(O1O2C3) - 0.5239(C3O13C12C12a) + 0.5657(C12C12aO1O2) \quad (1)$$

$$PC2 = 0.8771(O13C12) - 0.0404(O1O2C3) - 0.4628(C3O13C12C12a) + 0.1213(C12C12aO1O2) \quad (2)$$

Figure 2 shows the scores for 03 (three) methods and 09 (nine) levels of theory, correlating with the experimental geometry parameter of 1,2,13-trioxane ring. According to the score shows that the methods are broken into three groups according PC1. The semi-empirical method (AM1 and PM3 ZINDO) are located on the right side, while the method *ab initio*/Hartree-Fock and experimental are located in the central part of Figure 2. Already the DFT/B3LYP

method are located in the upper left. Furthermore, we can see that the method *ab initio* HF/6-31G, HF/6-31G* and HF/6-31G** are the closest experimental geometry parameter indicating that any one of them can be used in the development of calculations of molecular properties.

Figure 3 shows the loading for four (4) major classification descriptors of the methods and theory levels

**Fig. 2.** Plot of the scores PC1–PC2 for 03 (three) methods and 09 (nine) levels of theory, correlating with the experimental geometry parameter of 1,2,13-trioxane ring.**Fig. 3.** Plot of PC1–PC2 loadings using four (04) with geometric parameters selected principal components analysis to three (3) methods and nine (09) levels of theory, correlating with the experimental geometry parameter ring 1,2,13-trioxane.**Table III.** Principal component analysis of the selection model of levels of theory.

	Principal component		
	PC1	PC2	PC3
Variance (%)	24.3990	8.0043	2.5477
Variance cumulative (%)	67.7751	90.0093	97.0861
Molecular descriptors	Contribution		
	PC1	PC2	
O13C12	-0.3296	0.8771	
O1O2C3	0.5447	-0.0404	
C3O13C12C12a	-0.5239	-0.4628	
C12C12aO1O2	0.5657	0.1213	

are observed in the score that the geometrical parameters regarding the methods and levels of theory DFT (B3LYP/3-21G, B3LYP/3-21G*, B3LYP/3-21G**) and semi-empirical (AM1 and PM3) have main contribution of the geometric parameters O13C12 and C12C12aO1O2, and these are responsible for moving the methods and theory levels for the upper score. While the method *ab initio* (HF/6-31G, HF/6-31G* and HF/6-31G**) and semi-empirical (ZINDO) have a high contribution of the geometric parameters O1O2C3 and C3O13C12C12a, which are responsible in displacing methods and theory levels to the bottom of the score. Also in Figure 3, note that the larger the contribution of the geometric parameters O13C12 and C12C12aO1O2 the second main component the greater the score, and thus the method and levels of theory are less efficient, because if the distance experiment. The geometric parameters O1O2C3 and C3O13C12C12a contribute in lesser degree, by weight have a negative PC2, demonstrating that the methods and theory levels in general have higher values of the geometric parameters.

The geometric parameters O13C12, O1O2C3, C3O13C12C12a and C12C12aO1O2 are of great importance in our study, since according to the proposal made by Jefford and colleagues of the heme iron attacks the artemisinin O1 position and generates a free radical in position O2 (Fig. 4), after bond C3–C4 be broken into a radical carbon at C4.²⁶ This free radical C4 has been suggested as an important substance in the antimalarial activity.²⁷ Study of molecular docking of artemisinin and its receptor, heme, made by Tonmunpuean, Parasuk and

Kokpol also indicated that the heme iron interacts with O1 more preferably to O2,²⁸ which sets the importance of geometric parameter O1O2C3, selected this model to be associated with mechanism of action suggested.

Elhaes et al.²⁹ reported that the possible interactions of nanomaterials with living cells must be of concern, and studies have shown that the interaction between Epoxides C60 (C60–O) and hemoglobin is more likely to happen to adsorb and complex state. Djemil et al.³⁰ studied the structural aspects for the complexation of dopamine (DA) and epinephrine (EP) to β -CD were Explored by using PM6, HF and ONIOM methods. The structures show the presence of several intermolecular hydrogen bond interactions that were studied on the basis of NBO analysis employed to quantify the donor–acceptor interactions between the guest molecules and β -CD.

Table IV shows the geometrical parameters selected by principal component analysis, method and theory levels, variation of geometrical parameters in relation to the experimental data (Δ and $\Delta\%$). In the semi-empirical and DFT methods there isn't good agreement between theoretical and experimental values for the torsion angles, especially the angles formed by atoms C3O13C12C12a and C12C12aO1O2.

In semi-empirical method AM1, PM3 and ZINDO shows deviations $\Delta = -13.900^\circ$ ($\Delta\% = -54.940$), $\Delta = -22.489^\circ$ ($\Delta\% = -88.889$) and $\Delta = -7.879^\circ$ ($\Delta\% = -31.146$) in relative to torsion angles C3O13C12C12a, respectively. For the torsion angles C12C12aO1O2 shows deviations $\Delta = -0.650^\circ$ ($\Delta\% = -5.118$), $\Delta = -7.240^\circ$ ($\Delta\% = 57.007$) and $\Delta = 5.410^\circ$ ($\Delta\% = 42.598$). In relation

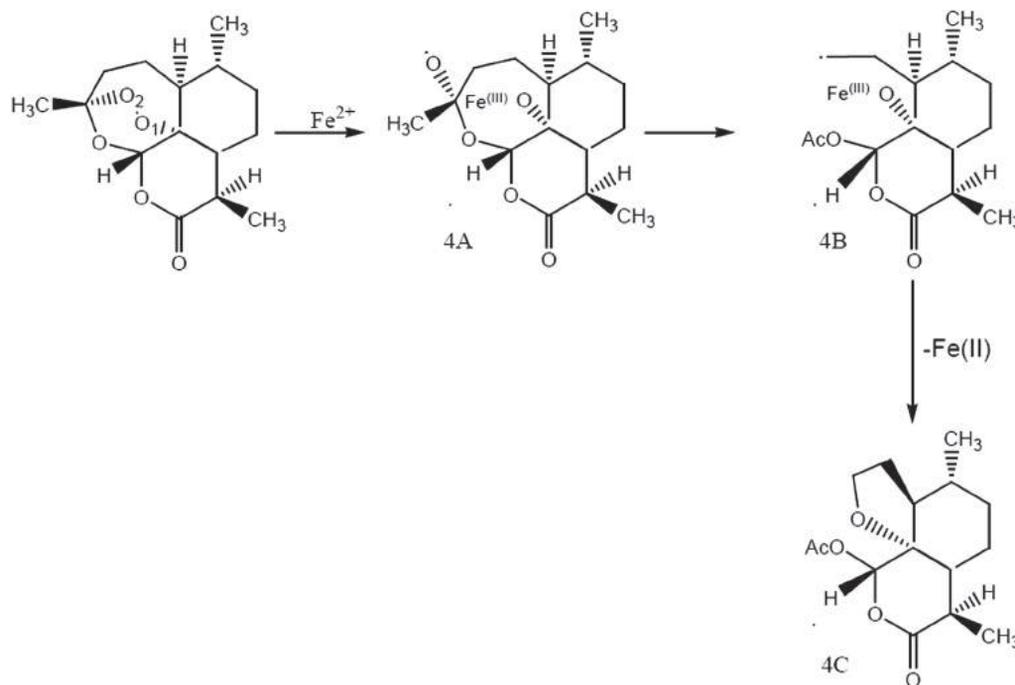


Fig. 4. Mechanism of action of artemisinin.

Table IV. Geometric Parameters selected by principal component analysis, method and theory levels, the variation of geometrical parameters with respect experimental data (Δ and $\Delta\%$).

Parameters ^a	Semi-empirical									Experimental
	AM1	Δ_{AM1}	Δ (%)	PM3	Δ_{PM3}	Δ (%)	ZINDO	Δ_{ZINDO}	Δ (%)	
Bond length (Å)										
O13C12	1.416	0.037	2.683	1.403	0.024	1.740	1.392	0.013	0.942	1.379
Bond angle (°)										
O1O2C3	112.530	4.430	4.098	110.340	2.240	2.072	114.310	6.210	5.744	108.100
Torsion angle (°)										
C3O13C12C12a	11.400	-13.900	-54.940	2.811	-22.489	-88.889	17.420	-7.880	-31.146	25.300
C12C12aO1O2	12.050	-0.650	-5.118	19.940	7.240	57.007	18.110	5.410	42.598	12.700
Parameters ^a	Ab initio/Hartree-fock/HF									Experimental
	6-31G	Δ_{6-31G}	$\Delta\%$	6-31G*	Δ_{6-31G^*}	Δ (%)	6-31G**	$\Delta_{6-31G^{**}}$	Δ (%)	
Bond length (Å)										
O13C12	1.403	0.024	1.740	1.400	0.021	1.522	1.376	-0.003	-0.217	1.379
Bond angle (°)										
O1O2C3	108.800	0.700	0.647	106.100	-2.000	-1.850	109.460	1.360	1.258	108.100
Torsion angle (°)										
C3O13C12C12a	25.320	0.020	0.079	27.432	2.132	8.426	27.400	2.100	8.300	25.300
C12C12aO1O2	12.510	-0.190	-1.496	10.900	-1.800	-14.173	10.924	-1.776	-13.984	12.700
Parameters ^a	DFT/B3LYP									Experimental
	3-21G	Δ_{3-21G}	Δ (%)	3-21G*	Δ_{3-21G^*}	Δ (%)	3-21G**	$\Delta_{3-21G^{**}}$	Δ (%)	
Bond length (Å)										
O13C12	1.430	0.051	3.698	1.430	0.051	3.698	1.430	0.051	3.698	1.379
Bond angle (°)										
O1O2C3	105.589	-2.511	-2.322	105.590	-2.510	-2.321	105.480	-2.620	-2.423	108.100
Torsion angle (°)										
C3O13C12C12a	29.059	3.759	14.857	29.060	3.760	14.861	29.080	3.780	14.940	25.300
C12C12aO1O2	9.060	-3.640	-28.66	9.600	-3.100	-24.409	9.340	-3.360	-26.456	12.700

Notes: Δ = Theoretical-experimental; Δ (%) = $\delta \times 100/\text{experimental}$.

to the DFT (3-21G, 3-21G* and 3-21G**) the torsion angles formed by atoms C3O13C12C12a, have deviations $\Delta = 3.759^\circ$ ($\Delta\% = 14.857$), $\Delta = 3.760^\circ$ ($\Delta\% = 14.861$) and $\Delta = 3.780^\circ$ ($\Delta\% = 14.940$), respectively. For torsion angles C12C12aO1O2 present deviations $\Delta = -3.640^\circ$ ($\Delta\% = -28.660$), $\Delta = -3.100^\circ$ ($\Delta\% = -24.409$) and $\Delta = -5.410^\circ$ ($\Delta\% = 42, 598$). Already in the method *ab initio*/Hartree-Fock is seen that the three levels (HF/6-31G, HF/6-31G* and HF/6-31G**) show excellent results for the bond length O13C12 with deviations $\Delta = 0.024 \text{ \AA}$ ($\Delta\% = 1.740$) to HF/6-31G, $\Delta = 0.021 \text{ \AA}$ ($\Delta\% = 1.522$) to HF/6-31G* and $\Delta = -0.003 \text{ \AA}$ ($\Delta\% = -0.217$) to HF/6-31G**, as shown in Table IV.

In Table IV, we highlight the comparison *ab initio* method with levels of theory HF/6-31G and HF/6-31G**, where the valence basis set HF/6-31G obtained satisfactory results, but the base HF/6-31G** showed excellent results in sign and magnitude with respect to the bond length and bond angle O13C12 O1O2C3. As the angles of twists or dihedral angle showed a good agreement with the experimental values reported in the literature, showing that the HF/6-31G** basis in this parameter are close

to the crystallographic data in the region of artemisinin endoperoxide ring.

4.3. HCA Method

A hierarchical cluster analysis (HCA) was used to validate the method and level of theory more appropriate for future calculations of molecular properties with greater accuracy in your results. The analysis was obtained in the form of a dendrogram as shown in Figure 5. It is observed in this figure three methods: Semi-empirical (AM1, PM3 and ZINDO), DFT (B3LYP/3-21G, B3LYP/3-21G*, B3LYP/3-21G**) and *ab initio*/Hartree-Fock (HF/6-31G, HF/6-31G* and HF/6-31G**). We note that the semi-empirical method (AM1, PM3 and ZINDO), has long branches featuring low similarity with the experimental. However, the levels of theory of DFT method (B3LYP/3-21G, B3LYP/3-21G*, B3LYP/3-21G**) have high similarity among them, because they have short branches. Since the levels of theory of *ab initio* method (HF/6-31G, HF/6-31G* and HF/6-31G**) shows high similarity among them and independent of other methods and levels of theory the HF/6-31G** has

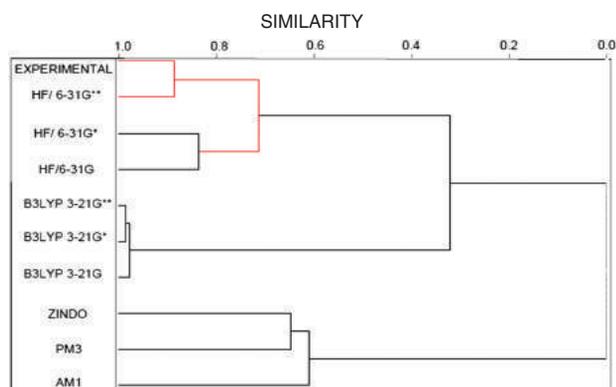


Fig. 5. HCA dendrogram for 03 (three) methods and 09 (nine) levels of theory, correlating with the experimental geometry parameter of 1,2,13-trioxane ring.

high similarity with the experimental, highlighted by red color branch. Therefore, as the results of combination of *ab initio* method with the level of theory 6-31G** can be used for molecular modeling of the structure of artemisinin and derivatives antimalarials, with mechanical action in the region of the endoperoxide ring.

The results of theoretical methods and experimental were distributed similar to those obtained with PCA. HCA confirms the results of PCA.

5. CONCLUSIONS

The multivariate analysis technique PCA and HCA were of great importance, because enable the classification of the methods and levels of theory in three groups: semi-empirical, *ab initio*/Hartree-Fock and DFT methods. Where the total geometrical parameters calculated was 18 (eighteen), among the parameters analyzed and selected the most important in the classification of methods and levels of theory, after analysis of the correlation matrix between the geometric parameters and standard deviation were four (04) parameters geometric related to trioxane ring, and identified by atoms: O13C12 which is the interplanar distance between the two atoms (bond length); O1O2C3 which is the angles length between these three atoms; C3O13C12C12a and C12C12aO1O2 are related dihedral angle or torsion. The results of HCA were similar to those obtained with PCA.

The utilization of semi-empirical calculations, *ab initio*/Hartree-Fock and DFT was possible because we have a molecule size is not very large, and thus the calculations involving quantum chemistry were developed at different levels of theory and methods. After the analyzes of PCA and HCA found that the Hartree-Fock method with the basis set separate valence 6-31G** is of good quality, and is suitable for molecular modeling studies of the structure of artemisinin, and to describe the conformation of derivatives antimalarial with mechanism of action in the ring region endoperoxide. The *ab initio* HF/6-31G** method

can be used for futures calculations of molecular properties which represent a means of chemical information, contained in the molecular structure of the compound on the development of chemical problems studies, pharmacological and toxicological studies in quantitative relationship between structure-activity and structure-property (QSAR, QSPR).

Acknowledgments: We gratefully acknowledge the final support of the Brazilian Agency Conselho Nacional de Desenvolvimento Científico e Tecnológico. We also thank the Instituto de Ciências da Saúde of Federal University of Pará for the use of the GaussView and Gaussian softwares. The authors would like to thank the Postgraduate Program in Biotechnology and Biodiversity—Network BIONORTE and to the Laboratório de Modelagem e Química Computacional, Federal University of Amapá for the computational support.

References

1. N. C. Cohen, Guidebook on Molecular Modeling in Drug Design, Academic Press, San Diego (1996).
2. C. M. R. Sant'anna, G. de termos usados no planejamento de fármacos, Recomendações da IUPAC para 1997, Quim. Nova (2002), Vol. 25, pp. 505–12.
3. I. Carvalho, A. D. L. Borges, and L. S. C. Bernardes, *J. Chem. Educ.* 82, 588 (2005).
4. C. G. Wermuth, The Practice of Medicinal Chemistry, 3rd edn., Academic Press, London (2009).
5. R. Todeschini and P. Gramatica, *Perspectives in Drug Discovery and Design* (1998).
6. O. S. de Silva, J. C. M. Corrêa, R. H. Bitencourt, R. W. Monteiro, J. Lameira, S. L. Santos, S. P. G. M. Guilhon, and S. B. D. Brasil, *J. Comput. Theor. Nanosci.* 9, 953 (2012).
7. A. M. Ibrahim and A. El-Din Alaa Gawad, *J. Comput. Theor. Nanosci.* 9, 1120 (2012).
8. M. Srivastava, H. Singh, and P. K. Naik, *Current Research Journal of Biological Sciences* 2, 83 (2010).
9. D. L. Klayman, *Science* 228, 1049 (1985).
10. S. C. Vonwiller and R. K. Haynes, *Tetrahedron. Lett.* 37, 253 (1996).
11. M. Jung, *Curr. Med. Chem.* 1, 35 (1994).
12. J. N. Cumming, P. Ploypradith, and G. H. Posner, *Adv. Pharmacol.* 37, 253 (1997).
13. C. de José Ribamar Carvalho, V. E. J. Ferreira, P. J. Barbosa, S. Da Maycon Lobato, F. C. C. Meneses, M. M. Soeiro, S. De Márcio Farias, O. R. C. De Almeida, C. K. Ventura, C. J. Pinheiro, and R. De Lourdes Andréia Pinheiro, *J. Comput. Theor. Nanosci.* 8, 2193 (2011).
14. F. A. Figueiredo, V. E. J. Ferreira, P. J. Barbosa, C. Da Jorge Williams Macedo, G. G. Da Maria Cristina, S. Da Maycon Lobato, C. J. Pinheiro, and A. T. R. Serra, *Journal of Computational and Theoretical Nanoscience* 8, 1847 (2011).
15. Gaussview, 1.0, Gaussian, Inc., Pittsburgh, PA (1997).
16. M. J. Frisch, Gaussian 98-Revision A.11, Gaussian, Inc., Pittsburgh, PA (2001).
17. J. N. Lisgarten, *Journal of Chemical Crystallography* 28, 539 (1998).
18. F. H. Allen, *Acta Cryst. B* 58, 380 (2002).
19. Infometrix, Inc., Pirouette 3.01, Woodinville, WA (2001).
20. M. M. C. Ferreira, *Journal of the Brazilian Chemical Society* 13, 742 (2002).

21. M. M. C. Ferreira, C. A. Montanari, and A. C. Gaudio, *Química Nova*. 25, 439 (2002).
22. R. Franke and A. Gruska, Multivariate data analysis of chemical and biological data, *Chemometrics Methods in Molecular Design*, edited by H. van de Waterbeemd, VCH, New York (1996), p. 115.
23. J. C. Pinheiro, M. M. C. Ferreira, and O. A. S. Romero, *Journal Molecular Structure (THEOCHEM)* 572, 35 (2001).
24. C. Thomson, M. Cory, and M. Zerner, *International Journal of Quantum Chemistry: Quantum Biology Symposium* 18, 231 (1991).
25. A. D. Becke, *J. Chem. Phys.* 98, 5648 (1993).
26. C. W. Jefford, *Helvetica Chimica Acta* 79, 1475 (1996).
27. G. H. Posner, J. N. Cummings, P. Ploypradith, and C. O. Oh, *J. Am. Chem. Soc.* 117, 5885 (1995).
28. S. Tonmunpheap, V. Parasuk, and S. Kokpol, *Journal of Molecular Modeling* 7, 26 (2001).
29. H. Elhaes, O. Osman, and M. Ibrahim, *J. Comput. Theor. Nanosci.* 9, 901 (2012).
30. R. Djemil and D. Khatmi, *J. Comput. Theor. Nanosci.* 9, 1571 (2012).

Received: 12 December 2012. Accepted: 7 January 2013.

Article

A SAR and QSAR Study of New Artemisinin Compounds with Antimalarial Activity

Cleydson Breno R. Santos ^{1,2,3,*}, Josinete B. Vieira ¹, Cleison C. Lobato ¹, Lorane I. S. Hage-Melim ¹, Raimundo N. P. Souto ², Clarissa S. Lima ³, Elizabeth V. M. Costa ³, Davi S. B. Brasil ⁴, Williams Jorge C. Macêdo ¹ and José Carlos T. Carvalho ^{2,3}

¹ Laboratory of Modeling and Computational Chemistry, Federal University of Amapá, Macapá 68902-280, Amapá, Amazon, Brazil; E-Mails: jnetbio.unifap2011.ap@gmail.com (J.B.V.); cleyson.cl@gmail.com (C.C.L.); lorane@unifap.br (L.I.S.H.-M.); williamsmacedo@yahoo.com.br (W.J.C.M.)

² Postgraduate Program in Biotechnology and Biodiversity-Network BIONORTE, Macapá 68902-280, Amapá, Amazon, Brazil; E-Mails: rnpsouto@unifap.br (R.N.P.S.); farmacos@unifap.br (J.C.T.C.)

³ Laboratory of Drug Research, School of Pharmaceutical Sciences, Federal University of Amapá, Macapá 68902-280, Amapá, Amazon, Brazil; E-Mails: lima.clarissa@gmail.com (C.S.L.); elizabethviana@unifap.br (E.V.M.C.)

⁴ Institute of Technology, Federal University of Pará, Av. Augusto Corrêa, 01, Belém 66075-900, Pará, Amazon, Brazil; E-Mail: davibb@ufpa.br

* Author to whom correspondence should be addressed; E-Mail: breno@unifap.br; Tel.: +55-96-4009-2920; Fax: +55-96-4009-2907.

Received: 21 October 2013; in revised form: 19 November 2013 / Accepted: 19 November 2013

Published: 30 December 2013

Abstract: The Hartree-Fock method and the 6-31G** basis set were employed to calculate the molecular properties of artemisinin and 20 derivatives with antimalarial activity. Maps of molecular electrostatic potential (MEPs) and molecular docking were used to investigate the interaction between ligands and the receptor (heme). Principal component analysis and hierarchical cluster analysis were employed to select the most important descriptors related to activity. The correlation between biological activity and molecular properties was obtained using the partial least squares and principal component regression methods. The regression PLS and PCR models built in this study were also used to predict the antimalarial activity of 30 new artemisinin compounds with unknown activity. The models obtained showed not only statistical significance but also predictive ability. The significant

molecular descriptors related to the compounds with antimalarial activity were the hydration energy (HE), the charge on the O11 oxygen atom (QO11), the torsion angle O1-O2-Fe-N2 (D2) and the maximum rate of R/Sanderson Electronegativity ($R\text{Te}^+$). These variables led to a physical and structural explanation of the molecular properties that should be selected for when designing new ligands to be used as antimalarial agents.

Keywords: artemisinin; antimalarial activity; HF/6-31G**; molecular docking; MEPs; SAR; QSAR

1. Introduction

Malaria is a very serious infectious disease caused by protozoans of the genus *Plasmodium* and is transmitted through the bite of infected female *Anopheles* mosquitoes. Every year, over one million people die from malaria, especially in tropical and subtropical areas. Most of the deaths are attributed to the parasite species *Plasmodium falciparum*. Many drugs have been investigated for their efficacy in the treatment of the disease, but strains of *P. falciparum* resistant to some of these drugs have appeared. Hence, the discovery of new classes of more potent compounds to treat the disease is necessary [1–6]. Artemisinin (qinghaosu) has been used in traditional Chinese medicine to treat disease for more than two million years. The medicine is extracted from the plant *Artemisia annua* L. and is used to combat 52 species of diseases in the People's Republic of China [7]. Artemisinin has a unique structure with a stable endoperoxide lactone (1, 2, 13-trioxane) that is totally different from previous antimalarials in its structure and mode of action. Artemisinin is remarkably effective against *Plasmodium falciparum* and cerebral malaria [8]. Currently, semi-synthetic artemisinin derivatives play an important role in the treatment of *P. falciparum* malaria [9–11]. Although the true mechanism of their biological activity against malaria has not been completely elucidated, various studies suggest that the trioxane ring is essential for antimalarial activity due to the properties displayed by the endoperoxide linkage. The literature also suggests that free heme could be the target of artemisinin in biological systems and that Fe^{2+} interacts with the peroxide when artemisinin reacts with heme [12–15]. Artemisinin and its derivatives induce a rapid reduction in the number of parasites when compared with other known drugs. Consequently, they are of particular interest for severe cases of malaria. The initial decline in the number of parasites is also beneficial for combination therapies. Therefore, there is an enormous interest in the mechanism of action, chemistry and drug development of this new class of antimalarials. The endoperoxide group is essential for the antimalarial activity and is mediated by activated oxygen (superoxide, H_2O_2 and/or hydroxyl radicals) or carbon free radicals [16–19].

In the evolution of computational chemistry, the use of molecular modeling (MM) has been one of the most important advances in the design and discovery of new drugs. Currently, MM is an indispensable tool in not only the process of drug discovery but also the optimization of existing prototypes and the rational design of drug candidates [20–23]. According to IUPAC, MM is the investigation of molecular structures and properties using computational chemistry and graphical visualization techniques to provide a three-dimensional representation of the molecule under a given set of circumstances [21]. The nature of the molecular properties used and the extent to which they

describe the structural features of molecules can be related to biological activity, which is an important part of any Structure-Activity Relationship (SAR) or Quantitative Structure-Activity Relationship (QSAR) study. QSAR studies use chemometric methods to describe how a given biological activity or a physicochemical property varies as a function of the molecular descriptors describing the chemical structure of the molecule. Thus, it is possible to replace costly biological tests or experiments using a given physicochemical property (especially those involving hazardous and toxically risky materials or unstable compounds) with calculated descriptors that can, in turn, be used to predict the responses of interest for new compounds [24]. Recently, Cristino *et al.* studied nineteen 10-substituted deoxyartemisinin derivatives and artemisinin with activity against D-6 strains of malarial *falciparum* in Sierra Leone. They used chemometric modeling to reduce dimensionality and determine which subset of descriptors are responsible for the classification between more active (MA) and less active (LA) artemisinins. A predictive study was performed with a new set of eight artemisinins using chemometric methods, and five of them were predicted to be active against D-6 strains of *falciparum* malaria [25].

In this paper, a SAR and QSAR study of artemisinin and 20 derivatives (see Figure 1) with different antimalarial activities, tested *in vitro* against *P. falciparum* (W-2), was performed. Initially, the structures were modeled, and many different molecular descriptors were computed. Maps of the molecular electrostatic potential (MEP) and molecular docking were employed to better understand the correlation between structure and activity and the interaction between the ligands (artemisinin and derivatives) and the receptor (heme). Multivariate analysis methods were used to deal with the large number of descriptors and generate a predictive model [26]. Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA) were employed to choose the molecular descriptors that are most related to the biological property investigated. Then, a QSAR model was elaborated through the Principal Component Regression (PCR) and Partial Least Square (PLS) methods that were used to perform predictions of 30 new artemisinin compounds with unknown antimalarial activity and to aid in future studies searching for other new antimalarial drugs [27–29].

2. Results and Discussion

2.1. Optimization of the Geometry of Artemisinin in Different Methods and Basis Sets

In all three basis sets (HF/6-31G, HF/6-31G*, HF/6-31G**), the Hartree-Fock method describes all structural parameters very well in terms of magnitude and sign when compared to the experimental values (see Table 1). This is in contrast to the AM1, PM3, ZINDO and DFT (B3LYP/3-21G, B3LYP/3-21G*, B3LYP/3-21G**) methods, in which there is not good agreement between the experimental and theoretical values for the torsion angles, especially the angle formed by atoms C3O13C12C12a, with deviations $<-13.900^\circ$ (AM1), $<-22.489^\circ$ (PM3), $<-7.880^\circ$ (ZINDO), $>0.020^\circ$ (HF/6-31G), $>2.132^\circ$ (HF/6-31G*), $>2.100^\circ$ (HF/6-31G**) $> -3.759^\circ$ (B3LYP/3-21G), $>-3.760^\circ$ (B3LYP/3-21G*) and $>-3.780^\circ$ (B3LYP/3-21G**) and standard deviations of 4.776, 8.388, 4.372, 1.663, 2.484, 1.762, 1.915, 1.855 and 1.987, respectively. By comparing these methods with the HF method, we find that the HF/6-31G and HF/6-31G** basis sets have low standard deviations in relation to the semiempirical and DFT methods. The variation was ± 0.099 between HF/6-31G and HF/6-31G**.

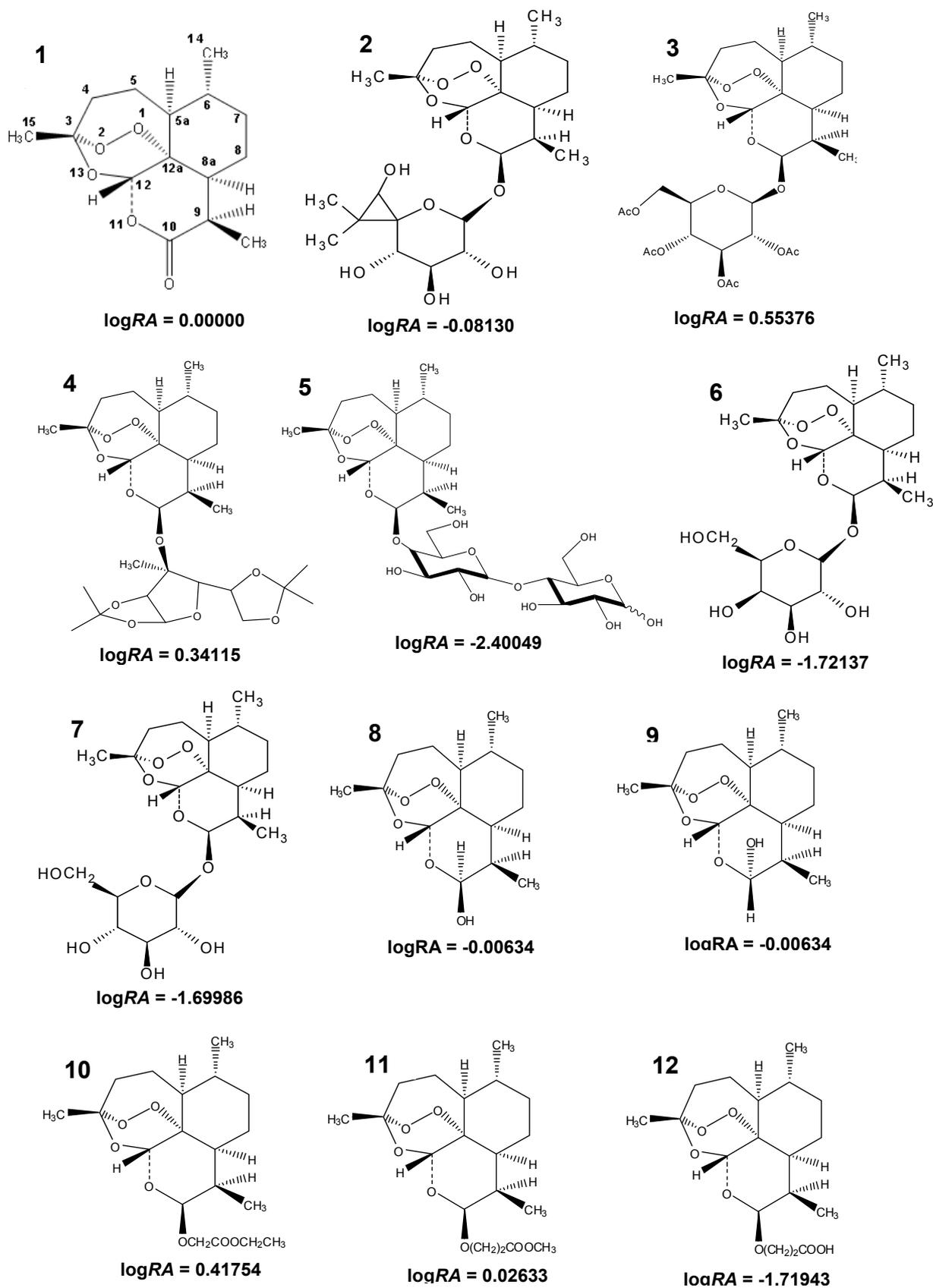
Figure 1. Structure and biological activity of artemisinin derivatives.

Figure 1. Cont.

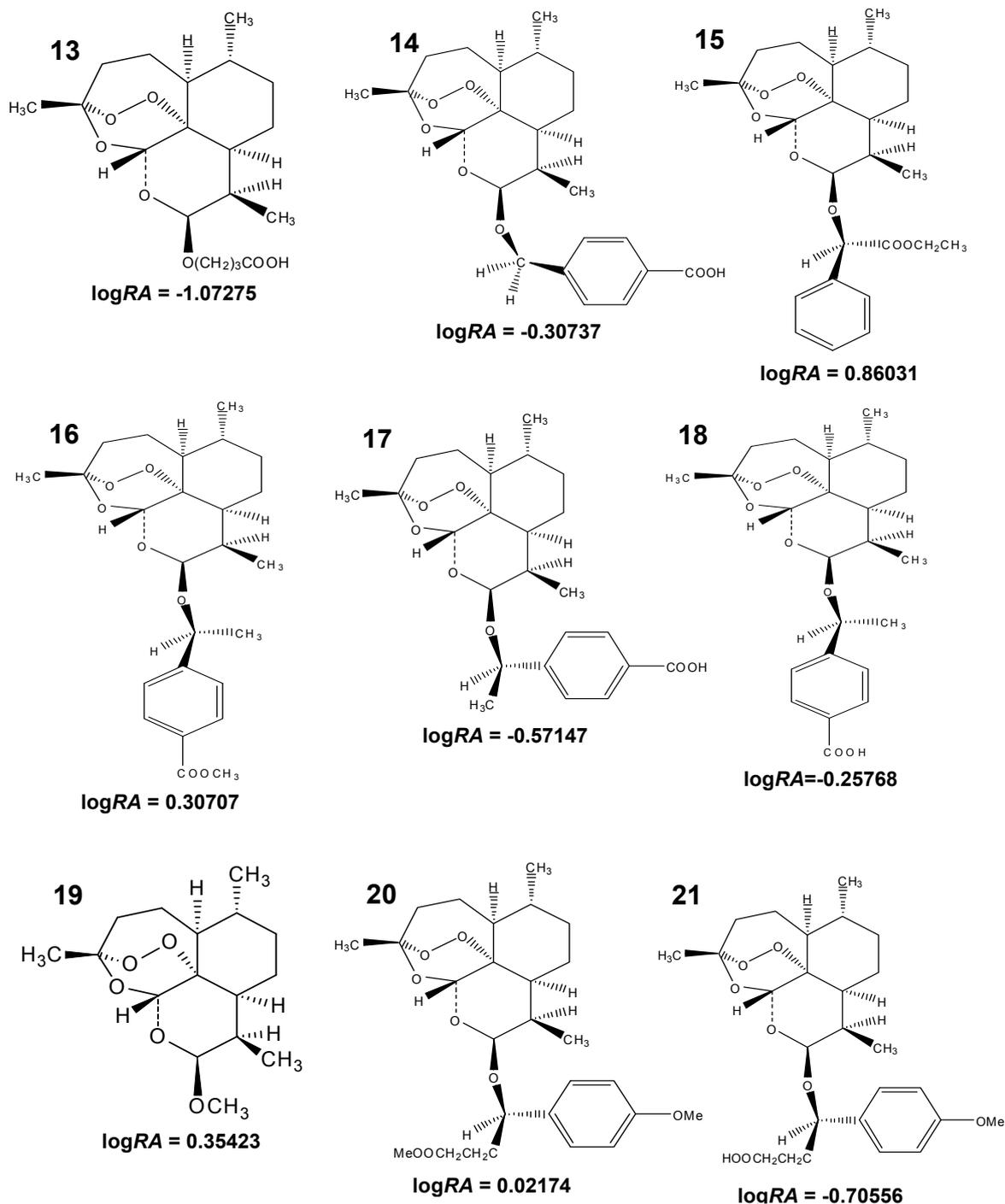


Table 1 shows that the HF/6-31G, HF/6-31G*, HF/6-31G** basis sets show excellent results for bond length compared to the experimental data. The 6-31G basis set described the bond angles well, with values close to the experimental results. However, the minimum bases (6-31G and 3-21G) have several deficiencies; thus, a polarization function was included to improve upon these bases (*i.e.*, *p* orbitals represented by *). These orbitals follow restricted functions that are centered at the nuclei. However, the atomic orbitals become distorted or polarized when a molecule is formed. Therefore, one must consider the possibility of non-uniform displacement of electric charges outside of the atomic nucleus, *i.e.*, polarization. Thus, it is possible to obtain a better description of the charges

and deformations of atomic orbitals within a molecule. A mode of polarization can be considered by introducing functions for which the values of l (quantum number of the orbital angular momentum) are larger than those of the fundamental state of a given atom. For these types, the basis set names denote the polarization functions. Thus, 6-31G* refers to basis set 6-31G with a polarization function for heavy atoms (*i.e.*, atoms other than hydrogen), and 6-31G** refers to the inclusion of a polarization function for hydrogen and helium atoms [30]. When basis sets with polarization functions are used in calculations involving anions, good results are not obtained due to the electronic cloud of anionic systems, which tend to expand. Thus, appropriate diffuse functions must be included because they allow for a greater orbital occupancy in a given region of space. Diffuse functions are important in the calculations of transition metals because metal atoms have “ d ” orbitals, which tend to be diffuse. It then becomes necessary to include diffuse functions in the basis function associated with the configuration of a neutral metal atom to obtain a better description of the metal complex. The 6-31G** basis is particularly useful in the case of hydrogen bonds [30–34].

This study highlighted that the HF/6-31G** basis set, which is closer to the experimental results and shows good performance in the description when comparing the C3O13C12 and C12aO1O2 bond angles. The torsion angles or dihedral angle also showed good agreement with the experimental values reported in the literature, showing that with the 6-31G** basis set, the torsion angles O1O2C3O13 and C13C12C12aO1 are closer to the crystallographic data. Artemisinin derivatives with antimalarial activity against *Plasmodium falciparum*, which is resistant to mefloquine, were studied using quantum chemical methods (HF/6-31G*) and the partial least-squares (PLS) method. Three main components explained 89.55% of the total variance, with $Q^2 = 0.83$ and $R^2 = 0.92$. From a set of 10 proposed artemisinin derivatives (artemisinin derivatives with unknown antimalarial activity against *Plasmodium falciparum*), a novel compound was produced with superior antimalarial activity compared to the compounds previously described in the literature [35]. Cardoso *et al.* [36] used HF/3-21G** *ab initio* and PLS methods to design new artemisinin derivatives with activity against *P. falciparum* malaria. The PLS method was used to build a multivariate regression model, which led to new artemisinin derivatives with unknown antimalarial activity. Additionally, MEP maps for the studied and proposed compounds were built and evaluated to identify common features in active molecules.

Cardoso *et al.* [37] studied artemisinin and some of its derivatives with activity against D-6 strains of *Plasmodium falciparum* using the HF/3-21G method. To verify the reliability of the geometry obtained, Cardoso *et al.* compared the structural parameters of the artemisinin trioxane ring with theoretical and experimental values from the literature. Ferreira *et al.* [16] studied artemisinin and 18 derivatives with antimalarial activity against W-2 strains of *Plasmodium falciparum* through quantum chemistry and multivariate analysis. The geometry optimization of structures was performed using the Hartree-Fock method and the 3-21G** basis set. Recently, Santos *et al.* [38] validated the HF/6-31G** computational methods applied in the molecular modeling of artemisinin, proposing a combination of chemical quantum methods and statistical analysis to study geometrical parameters of artemisinin in the region of the 1, 2, 13-trioxane endoperoxide ring. In determining the most stable structures of the studied compounds as well as the molecular properties, the Hartree-Fock method with the 6-31G** valence basis set separately has been used instead of semiempirical approaches such as AM1, PM3 and ZINDO, due to the number of relatively small compounds.

Table 1. Theoretical and experimental parameters of the 1, 2, 13-trioxane ring in artemisinin.

Parameters ^[a]	Semiempirical			Hartree-Fock/HF			DFT/B3LYP			Experimental ^[f]
	AM1 ^[b, c]	PM3 ^[b, c]	ZINDO ^[b, c]	6-31G ^[b, c]	6-31G* ^[b, c]	6-31G** ^[d]	3-21G ^[e]	3-21G* ^[e]	3-21G** ^[e]	
Bond Length (Å)										
O1O2	1.288	1.544	1.237	1.447	1.391	1.390	1.524	1.524	1.524	1.469
O2C3	1.447	1.403	1.400	1.435	1.393	1.396	1.455	1.455	1.454	1.416
C3O13	1.427	1.428	1.396	1.435	1.388	1.408	1.473	1.473	1.472	1.445
O13C12	1.416	1.403	1.392	1.403	1.400	1.376	1.430	1.430	1.430	1.379
C12C12a	1.537	1.555	1.513	1.533	1.533	1.532	1.535	1.535	1.535	1.523
C12aO1	1.468	1.426	1.416	1.469	1.429	1.429	1.504	1.504	1.504	1.461
Bond Angle (°)										
O1O2C3	112.530	110.340	114.310	108.800	106.100	109.460	105.590	105.590	105.480	108.100
O2C3O13	103.600	104.810	105.370	106.760	110.800	107.800	108.220	108.220	108.250	106.600
C3O13C12	115.480	116.010	115.843	117.300	112.800	115.300	113.200	113.200	113.200	114.200
O13C12C12a	113.510	115.200	113.270	112.280	108.700	112.300	113.300	113.300	113.230	114.500
C12C12aO1	111.070	113.180	107.290	110.910	110.500	110.545	112.410	112.410	112.470	110.700
C12aO1O2	113.740	112.290	118.380	113.240	112.700	112.700	109.620	109.620	109.590	111.200
Torsion Angle (°)										
O1O2C3O13	-77.800	-73.310	-70.403	-71.840	-73.369	-73.400	-76.610	-76.610	-76.740	-75.500
O2C3O13C12	42.070	52.700	36.370	33.390	31.034	31.100	33.750	33.750	33.720	36.000
C3O13C12C12a	11.400	2.811	17.420	25.320	27.432	27.400	29.059	29.060	29.080	25.300
O13C12C12aO1	-41.770	-40.510	-46.610	-49.410	-50.100	-50.143	-52.190	-52.190	-52.030	-51.300
C12C12aO1O2	12.050	19.940	18.110	12.510	10.900	10.924	9.060	9.600	9.340	12.700
C12aO1O2C3	47.050	35.630	40.130	46.700	48.700	48.674	51.060	51.060	51.320	47.800
Standard Deviation	4.776	8.388	4.372	1.663	2.484	1.762	1.915	1.855	1.987	-

^[a]: The atoms are numbered according to compound 1 in Figure 1; ^[b] Ref. [36]; ^[c] Ref. [37]; ^[d] Valence basis set separately validated to calculate the molecular properties;

^[e] Ref. [38]; ^[f]: Ref. [39].

2.2. Molecular Docking

Docking calculations showed that the entire ligand molecule is placed parallel to the plane of the porphyrin ring of heme, and the polar part of the ligand, which contains the peroxide bond, is directed toward the polar part of the heme system containing Fe^{2+} . This interaction is visualized in Figure 2 for most active compounds (1, 3, 4, 10, 11, 15, 16, 19 and 20). These orientations were assumed to be the most favorable and therefore to represent the real system under investigation, given that they were chosen based on the lowest free-energy of binding (interaction energy). For the compounds in the studied set, the values of $d(\text{Fe}-\text{O}1)$ ranged from 2.310 to 2.727 Å; however, this interval for the $d(\text{Fe}-\text{O}2)$ distances ranged from 2.760 to 3.808 Å. The $d(\text{Fe}-\text{O}13)$ distances ranged from 4.811 to 5.434, and the $d(\text{Fe}-\text{O}11)$ distances ranged from 4.897 to 5.525, as shown in Table 2.

Figure 2. Heme-artemisinin interactions of the most active compounds (1, 3, 4, 10, 11, 15, 16, 19 and 20).

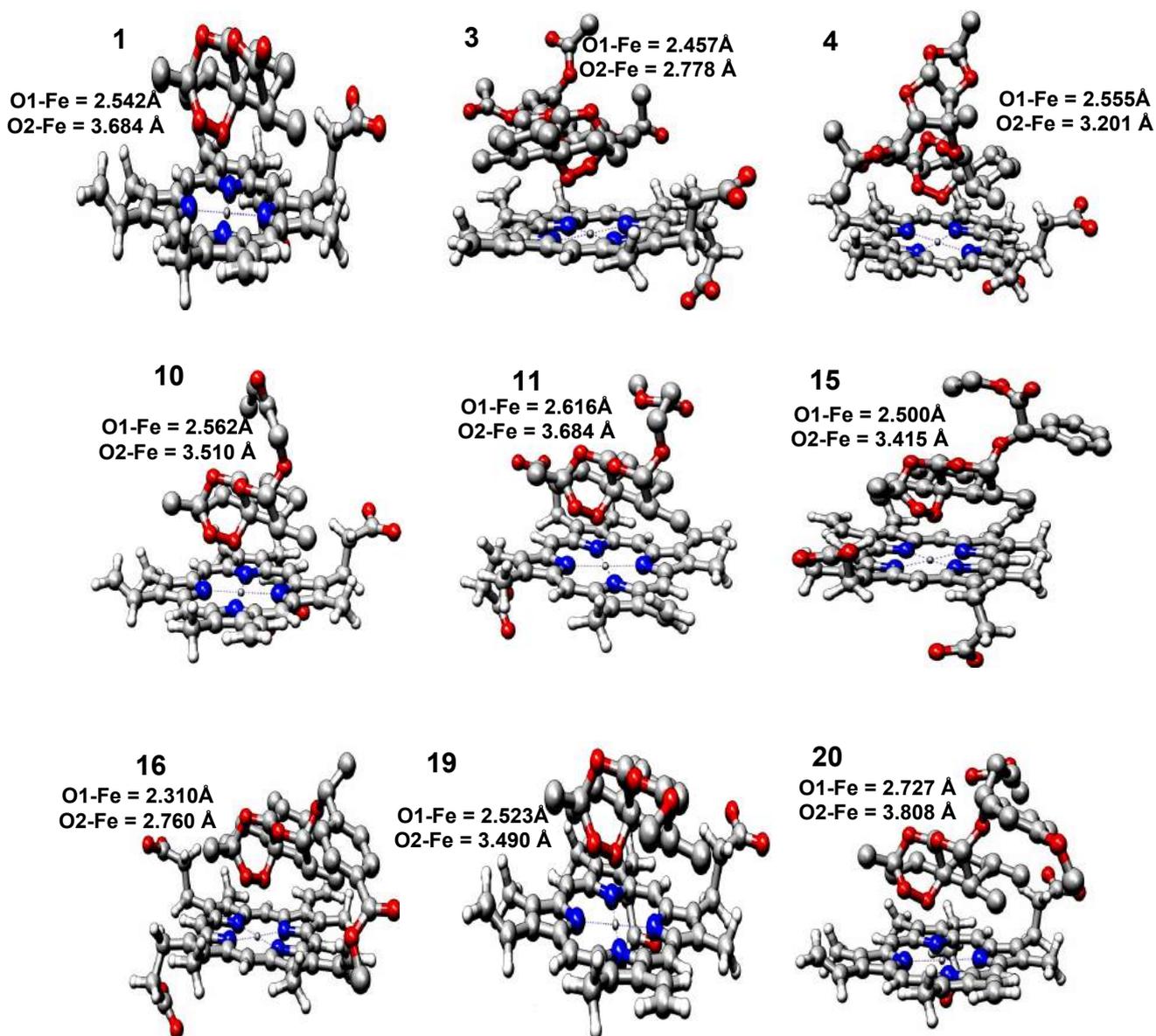


Table 2. Parameters calculated by molecular docking of heme-artemisinin and most active derivatives.

Compounds	E_{Complex} (Kcal mol ⁻¹)	Fe–O1 Distance (Å)	Fe–O2 Distance (Å)	Fe–O13 Distance (Å)	Fe–O11 Distance (Å)	logRA
1	−6.06	2.542	3.684	5.153	5.525	0.00000
3	−5.09	2.457	2.778	4.811	5.202	0.55376
4	−6.54	2.555	3.201	4.982	5.448	0.34115
10	−5.27	2.562	3.510	5.184	5.404	0.41754
11	−5.37	2.616	3.684	5.300	5.364	0.02633
15	−4.70	2.500	3.415	5.127	5.351	0.86031
16	−5.53	2.310	2.760	4.874	4.897	0.30707
19	−5.99	2.523	3.490	5.158	5.357	0.35423
20	−5.03	2.727	3.808	5.434	5.475	0.02174
E_{Complex}		0.06551	0.01761	0.19250	−0.20162	0.38917
Fe–O1			0.84202	0.85273	0.83598	−0.44984
Fe–O2				0.94792	0.81259	−0.48039
Fe–O13					0.65135	−0.48864
Fe–O11						−0.27755

For artemisinin (1), the $d(\text{Fe–O1})$ calculated distance was 2.542 Å, which is very close to the value reported (2.7 Å) in other theoretical studies [40,41]. There is a clear trend involving interatomic separation between Fe^{2+} and the oxygen atom in the trioxane ring because the distances are shorter for the O1 atom than for the O2 atom. This result reinforces the idea that the O1 atom from artemisinin preferentially binds to the Fe^{2+} from heme instead of the O2 atom.

Compounds **4**, **10**, **11** and **20** have higher activity than artemisinin and also higher values of $d(\text{Fe–O1})$. They have a large substituent that certainly causes repulsion due to steric effects, which prevents them from binding closer to the heme. Compounds **5** and **6** were designed to increase lipophilicity because it was observed that higher lipophilicity of artemisinin correlates with greater biological activity. Compounds **15**, **16** and **20** present large substituent groups on the α -methylene carbon (*C) that substantially increase the antimalarial activity of the compounds due to electronic and steric effects, respectively. Compound **3** demonstrated that the sugar-containing dihydroartemisinin acetylation derivatives have similar or better activities than artemisinin. However, the deacetylation of sugars reduces the antimalarial activity considerably.

The interaction energy for the ligand/receptor complex showed good linear correlation with activity ($r = 0.389177$) and ranged from −6.54 to −5.03 kcal·mol⁻¹ when compared with Fe–O1, Fe–O2, Fe–O13 and Fe–O11 distances (Å) (Table 2). In fact, even though some orientations were associated with the lowest interaction energy, they seemed to have strong activity against malaria because they presented the endoperoxide bond away from Fe^{2+} . Currently, the most accepted mechanisms of antimalarial action involve the formation of a complex between heme and artemisinin derivatives in which the iron of heme interacts with O1 of the endoperoxide. Moreover, substituent and conformation effects may affect the charge distribution at the oxygen and even the Fe–O1 bond [35]. An increase in the polar area of artemisinin increases the polar interactions between heme, the ligand and the globin.

2.3. Molecular Electrostatic Potential Maps

To identify key characteristics of compounds derived from artemisinin, maps of molecular electrostatic potential (MEPs) were evaluated and used for qualitative comparisons in the region of the 1, 2, 13-trioxane ring of artemisinin and its derivatives. The geometrical form of the potential in the region of the 1, 2, 13-trioxane ring is similar for all active compounds and is characterized by negative electrostatic potential (red region) according to the literature [42].

The MEP visualization is shown in Figure 3. Compounds **2–21** have a region of negative potential near the trioxane ring, similar to the MEP of artemisinin (compound 1), which has an electrostatic potential maximum of 0.13378 u.a. (blue region) and a minimum of -0.12617 u.a. (red region). The maximum positive MEP (blue region) varied from 0.14234 u.a. to 0.10429 u.a. for active compounds, while less active compounds ranged from 0.18555 u.a. to 0.14360 u.a. The values corresponding to the minimum negative electrostatic potential (red region) for the most active compounds ranged from -0.10750 u.a. to -0.12617 u.a., presenting potential values close to those of artemisinin. The minimum negative electrostatic potential (red region) for less active compounds ranged from -0.10384 u.a. to -0.12065 u.a., which are higher than those of artemisinin.

The region of negative electrostatic potential is due to the binding of the endoperoxide (C-O-O-C), which is the most notable feature of MEP. The distribution of the electron density around the trioxane ring is thought to be responsible for activity against malaria, a belief supported by the fact that the complexation of artemisinin with heme involves an interaction between the peroxide bond, the most negatively charged zone on the ligand, and Fe^{2+} , the most positively charged zone on heme (the receptor molecule) [15,43].

The presence of a negative surface close to the trioxane ring suggests that these compounds have a reactive site for electrophilic attack and must possess antimalarial potency; consequently they are being investigated. Thus, in the case of an electrophilic attack of the iron of heme against an electronegative zone, there is a preference for it to occur through the endoperoxide linkage. By analyzing *MEP* maps, the selection of inactive compounds can be avoided.

2.4. PCA Results

The PCA results showed that the most important descriptors were the following: the hydration energy (HE), charge on the oxygen atom O11 (QO11), torsion angle D2 (O2–O1–Fe–N2) and the maximum rate of R/Sanderson electronegativity (RTe^+). The hydration energy is the energy released when water molecules are separated from each other and are attracted by solute molecules or ions. Hydration energy comprises solvent-solvent and solute-solvent interactions [44]. The charge on the O11 atom (QO11) is a measure of the force with which a particle can electrostatically interact with another particle [45]. O RTe^+ is a GETAWAY (geometry, topology and set of atomic weights) type descriptor associated with the form, symmetry size and molecular distribution of the atom [46,47]. The torsion angle D2 (O2–O1–Fe–N2) is of great importance in our study; according to the proposal of Jefford and colleagues, the iron of heme attacks artemisinin at O1 and generates a free radical in position O2 after the C3–C4 bond is broken, generating a carbon radical at C4 [48]. This free radical at C4 has been suggested to be an important component of antimalarial activity [49]. Molecular docking

of artemisinin and its receptor, the heme group, performed by Tonmunphean, Parasuk and Kokpol also indicated that the iron of the heme group preferentially interacts with O1 rather than O2 [41].

Figure 3. Molecular electrostatic potential maps of the studied artemisinin derivatives with antimalarial activity against *Plasmodium falciparum* (W-2 clone).

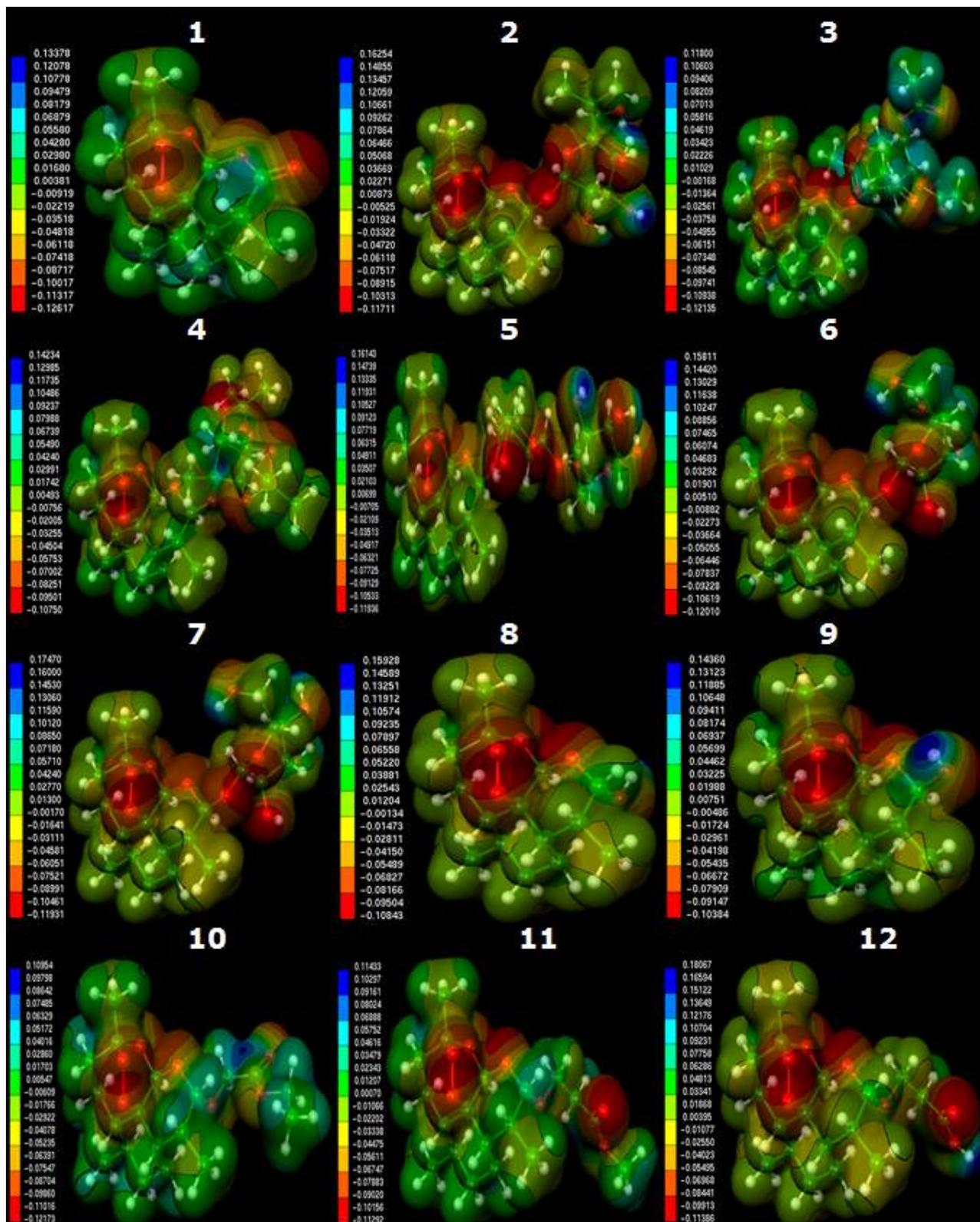
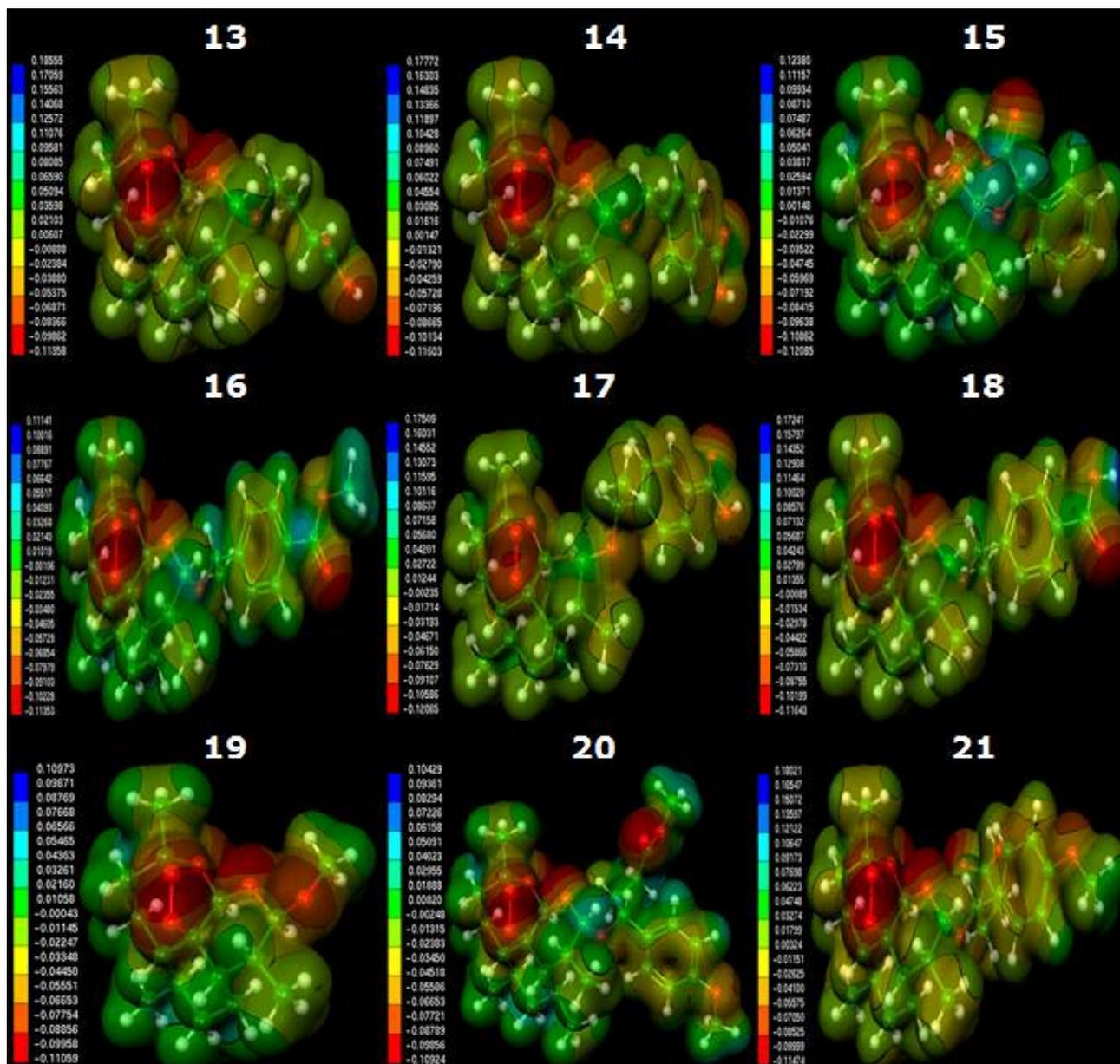


Figure 3. Cont.



The values of the important descriptors of each selected compound identified via PCA as well as the values of $\log RA$, relative activity (RA) and the IC_{50} is the 50% inhibitory concentration are shown in Table 3. The Table 3 shows the Pearson correlation matrix between the descriptors and $\log RA$, and the correlation between pairs of descriptors is less than 0.70, while the correlation between the descriptors and $\log RA$ is less than 0.87. The descriptors selected by PCA represent the characteristics necessary to quantify the antimalarial activity of these compounds against *Plasmodium falciparum* W-2.

The results of the SAR model are presented in Table 4. The model was constructed with three main components (3 PCs). The first principal component (PC1) describes 40.8865% of the total information, the second principal component (PC2) describes 22.7045%, and the third (PC3) 11.5660%. PC1 contains 51.1081% of the original data, and the combination of the first two components (PC1 + PC2) contains 79.4887%, and all three (PC1 + PC2 + PC3) explain 93.9461% of the total information, losing only 6.0539% of the original information. The descriptors HE, D2 and QO11

contribute the most to PC1, while in PC2, the descriptor RTe⁺ is the primary contributor. The main components can be written as a linear combination of the selected descriptors. Mathematical expressions for PC1 and PC2 are shown below.

$$PC1 = 0.5705HE - 0.5088QO11 + 0.0925RTe^+ + 0.6381D2 \quad (1)$$

$$PC2 = -0.3847HE - 0.2987QO11 + 0.8731RTe^+ - 0.0207D2 \quad (2)$$

Figure 4 shows the scores for the 21 compounds studied. Based on the graph, PC2 distinguishes between compounds that are more potent and less potent. The most potent compounds are located at the bottom (**1, 3, 4, 10, 11, 15, 16, 19** and **20**), while the less potent compounds are located in the upper portion of the graph (**2, 5, 6, 7, 8, 9, 12, 13, 14, 17, 18** and **21**).

Table 3. Physicochemical properties selected by principal component analysis, experimental logRA values, IC₅₀ and the correlation matrix.

Compounds	HE	QO11	RTe ⁺	D2	logRA	RA	IC ₅₀ (ng/mL)
1 ⁺	-2.820	-0.605	0.105	120.868	0.00000	1.000000	0.6800
2 ⁻	-13.330	-0.516	0.127	27.480	-0.08130	0.829268	0.8200
3 ⁺	-4.190	-0.567	0.066	-62.834	0.55376	3.578947	0.1900
4 ⁺	-2.970	-0.558	0.069	94.272	0.34115	2.193548	0.3100
5 ⁻	-26.220	-0.558	0.079	-143.766	-2.40049	0.003977	171.0000
6 ⁻	-16.200	-0.547	0.123	-163.237	-1.72137	0.018994	35.6000
7 ⁻	-16.640	-0.517	0.135	-158.396	-1.69986	0.019959	34.0700
8 ⁻	-3.690	-0.631	0.187	111.395	-0.00634	0.985507	0.6900
9 ⁻	-5.670	-0.676	0.126	113.465	-0.00634	0.985507	0.6900
10 ⁺	-1.680	-0.599	0.100	13.716	0.41754	2.615385	0.2600
11 ⁺	-3.330	-0.662	0.080	124.375	0.02633	1.062500	0.6400
12 ⁻	-8.530	-0.666	0.131	105.696	-1.71943	0.019080	35.6400
13 ⁻	-8.210	-0.648	0.132	138.353	-1.07275	0.084577	8.0400
14 ⁻	-7.420	-0.651	0.131	127.415	-0.30737	0.492754	1.3800
15 ⁺	-2.930	-0.675	0.068	-17.600	0.86031	7.249467	0.0938
16 ⁺	-2.540	-0.637	0.081	100.819	0.30707	2.028035	0.3353
17 ⁻	-6.900	-0.754	0.150	95.974	-0.57147	0.268245	2.5350
18 ⁻	-6.980	-0.645	0.118	107.153	-0.25768	0.552486	1.2308
19 ⁺	-1.870	-0.501	0.105	114.392	0.35423	2.260638	0.3008
20 ⁺	-5.560	-0.641	0.082	153.908	0.02174	1.051330	0.6468
21 ⁻	-11.120	-0.651	0.141	10.910	-0.70556	0.196987	3.4520
EH		-0.329	-0.156	0.694	0.860	-	-
QO11			-0.203	-0.509	-0.127	-	-
RTe ⁺				0.128	-0.333	-	-
D2					0.485	-	-

Table 4. Principal component analysis of the SAR model and contribution of selected descriptors based on step multivariate analysis.

Parameters	Main Component		
	PC1	PC2	PC3
Variance (%)	40.8865	22.7045	11.5660
Cumulative Variance (%)	51.1081	79.4887	93.9461
Molecular Descriptors	Contribution		
	PC1	PC2	
EH	0.5705	-0.3847	
QO11	-0.5088	-0.2987	
RTe ⁺	0.0925	0.8731	
D2	0.6381	-0.0207	

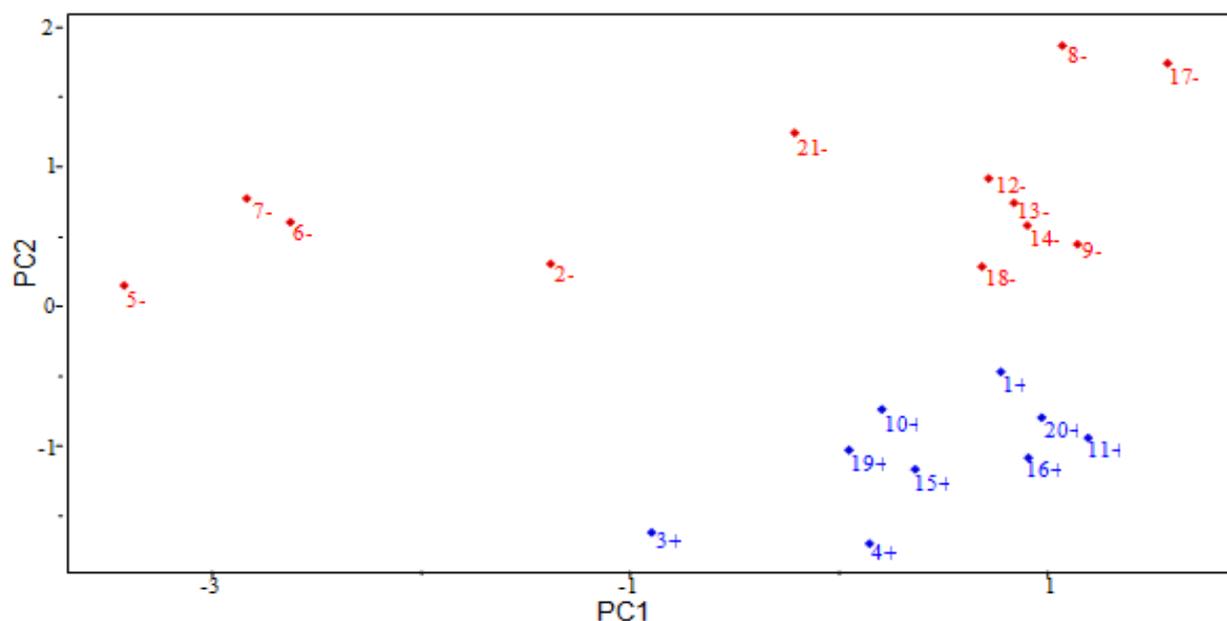
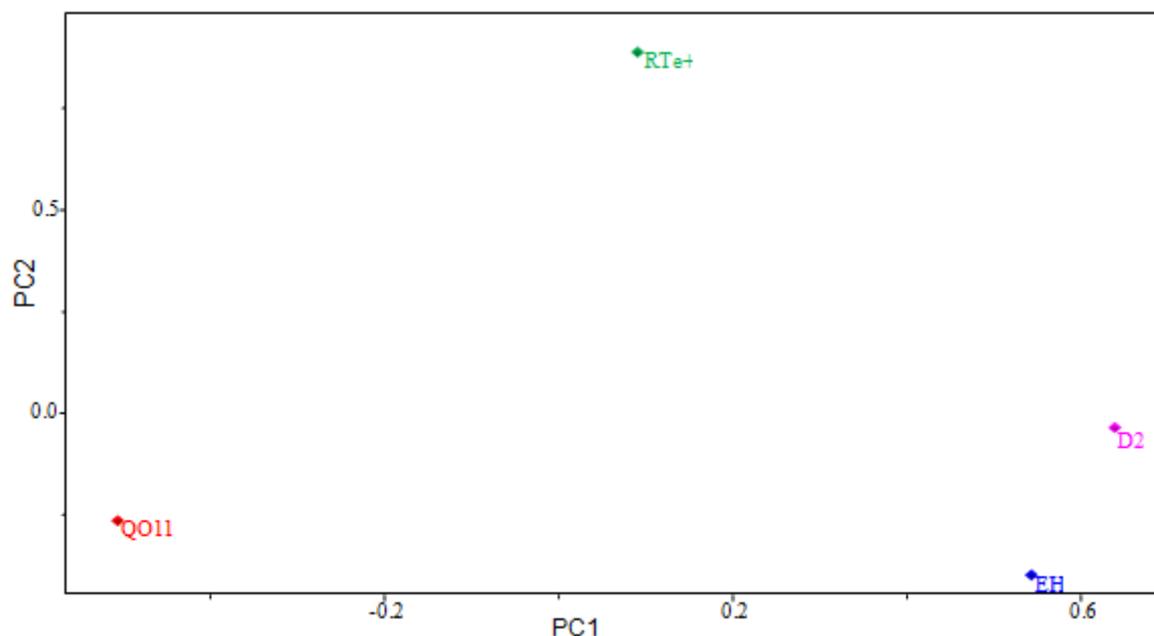
Figure 4. Plot of PC1–PC2 scores for artemisinin and derivatives with antimalarial activity against W-2 strains of *P. falciparum*. Positive values indicate more potent analogs, and negative values indicate less potent analogs.

Figure 5 shows the loadings for the four descriptors that are most important in the classification of compounds. More potent compounds have high contributions from the descriptors QO11, HE and D2, while less potent compounds have a high contribution from the descriptor RTe⁺. Thus, the descriptors QO11, HE and D2 are responsible for the location of more potent compounds at the bottom of the graph. The descriptor RTe⁺ places less potent compounds in the upper part of the graph. Figure 5 also shows that the higher the contribution of the descriptor RTe⁺ in the second principal component, *i.e.*, the higher the value of the maximum index of R/Sanderson electronegativity for a certain compound, the higher the score value will be, indicating that the compound is less potent than others. The other descriptors contribute to a lesser degree. For example, the descriptor HE has negative weight in PC2, demonstrating that the most potent compounds generally have higher values of this descriptor.

Figure 5. Plot of the PC1–PC2 loadings with the four descriptors selected to build the PLS and PCR models of artemisinin and derivatives with biological activity against W-2 strains of *P. falciparum*.



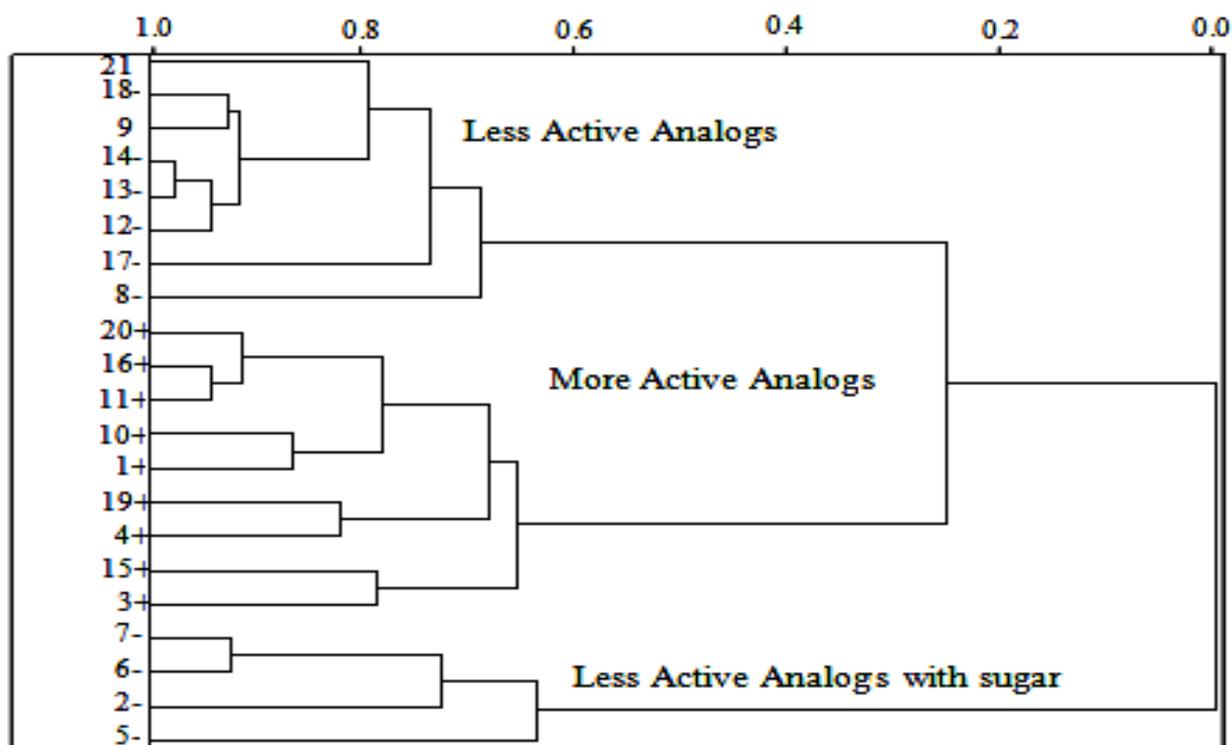
Costa *et al.* [40] showed that the presence of water changed the dihedral angle involved in the heme–artemisinin complex (C–Fe–O1–O2). Thus, this effect is believed to influence the process of molecular recognition between artemisinin and derivatives and heme in aqueous biological systems. The selection of the torsion angle D2 (O2–O1–Fe–N2) descriptor suggests that the action of drugs against malaria depends on electrophilic attack on the endoperoxide bond, particularly on the O1 atom. This result was confirmed by both an analysis of the *MEP* maps and by molecular docking as discussed previously.

2.5. HCA Results

The statistical analysis utilized in this study should group similar compounds into categories. The categories are represented by a two-dimensional diagram known as dendrogram that illustrates the fusions or divisions made at each successive stage of the analysis. Single samples (compounds) are represented by the branches on the bottom of the dendrogram. The similarity among the clusters is given by the length of their branches, so compounds presenting low similarity have long branches whereas compounds of high similarity have short branches. The HCA method classified the compounds into three classes (more active, less active and less active containing sugar) and was based on the Euclidean distance and the incremental method [50]. In the incremental linkage, the distance between two clusters is the maximum distance between a variable in one cluster and a variable in the other cluster. The descriptors employed to perform HCA were the same as those used for PCA, *i.e.*, HE, QO11, D2 (O2–O1–Fe–N2) and RTe⁺. In the HCA technique, the distances between pairs of samples are computed and compared. Small distances imply that compounds are similar, while dissimilar samples will be separated by relatively large distances. The dendrogram in Figure 6 shows the HCA graphic as well as the compounds separated into three main classes. The scale of similarity

varies from 0 for samples with no similarity to 1 for samples with identical similarity. By analyzing the dendrogram, some conclusions can be drawn even though the compounds present some structural diversity. HCA showed results similar to those obtained with PCA. The compounds are grouped according to their biological activities. The most potent compounds are **1, 3, 4, 10, 11, 15, 16, 19** and **20**. The less potent compounds are grouped into two clusters, one of which contains compounds **8, 9, 12, 13, 14, 17, 18** and **21**, and the other cluster contains artemisinin derivatives that possess a sugar (**2, 5, 6** and **7**).

Figure 6. HCA dendrogram for artemisinin and derivatives with biological activity against W-2 strains of *P. falciparum*. Positive values indicate more potent analogs, and negative values indicate less active compounds.



2.6. Partial Least Squares (PLS) and Principal Component Regression (PCR) Results

The statistical quality [51] of the PLS and PCR models was gauged by parameters such as correlation coefficient or squared correlation coefficient (R^2), explained variance (R^2_{adjust} , *i.e.*, adjusted R^2), standard deviation (s), variance ratio (F), cross-validated correlation coefficient (Q^2), standard error of validation (SEV), predicted residual error sum of squares ($PRESS$) and standard deviation of cross-validation (S_{PRESS}) [52–54]. The best regression models were selected based on high values of R^2 , R^2_{adjust} , Q^2 and F (a statistic of assessing the overall significance) and low values of s , SEV , $PRESS$ and S_{press} .

The calculated properties and the experimental activity values for the compounds studied (Table 5) were used to build the regression models. The models built using the PLS and PCR methods were based on three latent variables, 18 test compounds and 3 compounds (**2, 12** and **13**) from the external validation set.

Table 5. Predicted PLS and PCR results and validation errors for logRA (experimental).

Compounds	Predicted		Validation Error		Experimental
	PLS	PCR	PLS	PCR	logRA
1+	0.2548	0.0902	0.2548	0.0902	0.0000
2- ^[a]	-1.0163	-0.8805	-0.935	-0.7992	-0.0813
3+	-0.2855	-0.6548	-0.8392	-1.2085	0.5537
4+	0.2199	-0.1981	-0.1212	-0.5392	0.3411
5-	-2.1192	-1.7899	0.2812	0.6105	-2.4004
6-	-1.6837	-1.4214	0.0376	0.2999	-1.7213
7-	-1.8113	-1.5164	-0.1115	0.1834	-1.6998
8-	-0.1174	0.2199	-0.1111	0.2262	-0.0063
9-	0.0948	0.2284	0.1011	0.2347	-0.0063
10+	0.0342	-0.1872	-0.3833	-0.6047	0.4175
11+	0.4495	0.2485	0.4232	0.2222	0.0263
12- ^[a]	-0.1670	0.1032	1.5524	1.8226	-1.7194
13- ^[a]	-0.0920	0.1144	0.9807	1.1871	-1.0727
14-	-0.0583	0.1144	0.249	0.4217	-0.3073
15+	0.1452	-0.0974	-0.7151	-0.9577	0.8603
16+	0.3812	0.1289	0.0742	-0.1781	0.3070
17-	0.0203	0.4326	0.5917	1.004	-0.5714
18-	-0.0448	0.0392	0.2128	0.2968	-0.2576
19+	0.0888	-0.2535	-0.2654	-0.6077	0.3542
20+	0.3283	0.1684	0.3066	0.1467	0.0217
21-	-0.6913	-0.3463	0.0142	0.3592	-0.7055

^[a] Compounds from the external validation set.

The regression Equations obtained for PLS (Equation (3)) and PCR (Equation (4)) models that relate the descriptors and biological activity are the following:

$$\text{Log RA} = 0.520565\text{HE} - 0.151382\text{QO11} - 0.158294\text{RTe}^+ + 0.353200\text{D2} \quad (3)$$

$n = 18$, $R^2 = 0.9468$, $R^2_{\text{adjust}} = 0.9354$, $s = 0.2211$, $F_{(4,11)} = 57.8889$, $Q^2 = 0.8566$, $SEV = 0.3202$, $PRESS = 0.6847$, $S_{PRESS} = 0.0636$.

$$\text{Log RA} = 0.304816\text{HE} - 0.258614\text{QO11} + 0.046881\text{RTe}^+ + 0.340163\text{D2} \quad (4)$$

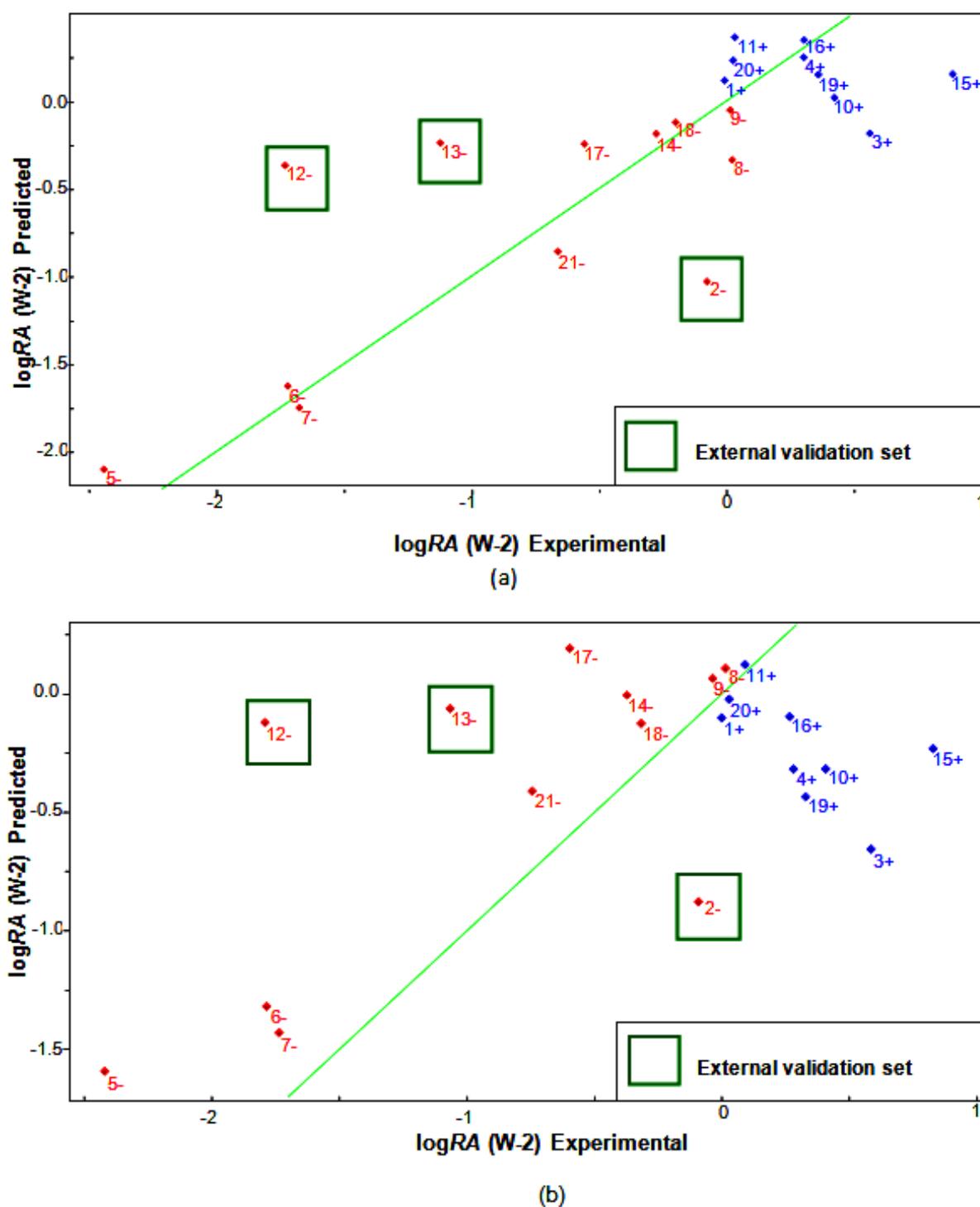
$n = 18$, $R^2 = 0.8488$, $R^2_{\text{adjust}} = 0.8164$, $s = 0.3729$, $F_{(4,11)} = 18.2454$, $Q^2 = 0.7313$, $SEV = 0.4438$, $PRESS = 1.9476$, $S_{PRESS} = 0.1073$.

The results obtained with the PLS and PCR models were very close, with variation of $R^2 = \pm 0.098$, $R^2_{\text{adjust}} = \pm 0.119$, $s = \pm 0.1518$, $F_{(4,11)} = \pm 39.6435$, $Q^2 = \pm 0.1253$, $SEV = \pm 0.1236$, $PRESS = \pm 1.2629$ and $S_{PRESS} = \pm 0.0437$ (between PLS and PCR). The quality of the PLS and PCR models can be demonstrated by comparing the measured and the predicted activities. The validation errors obtained by the leave-one-out cross-validation method are shown in Table 5. For the PLS, model only five compounds (**2**, **3**, **12**, **13** and **15**) had high validation errors, and the PCR model yielded five compounds (**3**, **12**, **13**, **15** and **17**) with high residual values. Our PLS and PCR models present the best fit for compounds with high activity because compounds with low activity showed high residuals values.

The measured versus predicted values using our PLS and PCR models are presented in (Figure 7a,b) respectively. The PLS and PCR plots identify compounds with higher activity (blue) and compounds with lower activity (red), including compounds from the external validation set. According to the PLS

and PCR models, the four variables present different magnitudes of regression coefficients (in absolute value). The models reveal that compounds with high biological potency against *P. falciparum* have a combination of higher values of HE and D2 and lower values of QO11 and R_{Te}^+ for the PLS model, but for the PCR model, compounds have higher values of HE and D2, lower QO11 values and positive values for R_{Te}^+ . The validation parameters support the fact that the models are efficient and hence satisfactory given the complexity of the antimalarial mechanisms and the small number of descriptors (four) selected to build the QSAR model.

Figure 7. Plot of experimental versus predicted values for $\log RA$ modeled by (a) PLS and (b) PCR.



The compounds of the set test were molded from the most stable structure of artemisinin, compound **1** of Figure 1, and constructed using GaussView 5.0 program, carrying the complete optimization of the geometry of each compound with the basis set of separated valence 6-31G** using the Hartree-Fock method as implemented in Gaussian 03 program. After obtain the most stable geometry of each compound was determined only selected descriptors in PCA and used in the construction of the QSAR (PLS and PCR) models, namely EH, QO11, RTe⁺ and D2, shown in Table 6.

The QSAR models (PLS and PCR) were built used to predict the unknown antimalarial activity of thirty new artemisinin derivatives shown in Figure 8, compounds **22–51**. Table 7 shows the results of the logRA by PCR and PLS models. According to Table 7 the PLS model showed that fifteen compounds of the test set (**22, 23, 27, 30, 32, 34, 36–40, 44, 45, 47, 51**) are predicted to be more active, they had values of logRA greater than zero (logRA > 0). However, the PCR model only nine compounds of all test sets (**23, 25, 37, 38, 43, 46, 48–50**) were predicted as most active, which showed values of logRA higher than zero (logRA > 0), a total of 24 compounds proposed as more active of thirty suggested compounds. However, compounds **23, 37** and **38** were the ones that had values of logRA greater than zero (logRA > 0) in both models (PLS and PCR) with residues of prediction ranging from 0.028951 to −0.1351, suggesting that these new compounds in the two models (PLS and PCR) are more potent than artemisinin may be synthesized and tested for antimalarial activity.

Table 6. Molecular properties selected by analysis of main components of test set with antimalarial activity unknown.

Test Set	EH	QO11	RTe ⁺	D2
22	−3.460	−0.663	0.076	7.585
23	−3.370	−0.664	0.077	141.065
24	−4.790	−0.556	0.069	130.453
25	−5.780	−0.675	0.077	98.153
26	−8.070	−0.603	0.076	−76.018
27	−4.650	−0.602	0.073	−4.170
28	−7.440	−0.575	0.066	−9.051
29	−15.920	−0.482	0.100	73.480
30	−4.470	−0.594	0.070	125.875
31	−15.240	−0.601	0.106	9.276
32	−4.500	−0.532	0.063	−37.529
33	−13.680	−0.578	0.126	−83.125
34	−4.550	−0.572	0.071	8.222
35	−13.620	−0.523	0.121	32.018
36	−4.280	−0.584	0.071	−27.718
37	−2.740	−0.650	0.105	152.098
38	−2.850	−0.673	0.081	101.819
39	−2.680	−0.603	0.068	−13.617
40	−3.290	−0.577	0.064	−65.438
41	−10.210	−0.615	0.122	10.190
42	−7.044	−0.557	0.062	−13.671
43	−7.841	−0.654	0.131	127.514
44	−2.910	−0.657	0.072	−25.670

Table 6. Cont.

Test Set	EH	QO11	RTe ⁺	D2
45	-2.870	-0.670	0.069	-19.115
46	-7.020	-0.745	0.155	95.479
47	-4.240	-0.600	0.066	122.578
48	-8.120	-0.684	0.123	131.353
49	-8.350	-0.665	0.134	105.669
50	-5.676	-0.667	0.126	113.564
51	-3.640	-0.636	0.067	7.855

Figure 8. Compounds of the test set artemisinin derivatives with unknown antimalarial activity against *Plasmodium falciparum* type W-2.

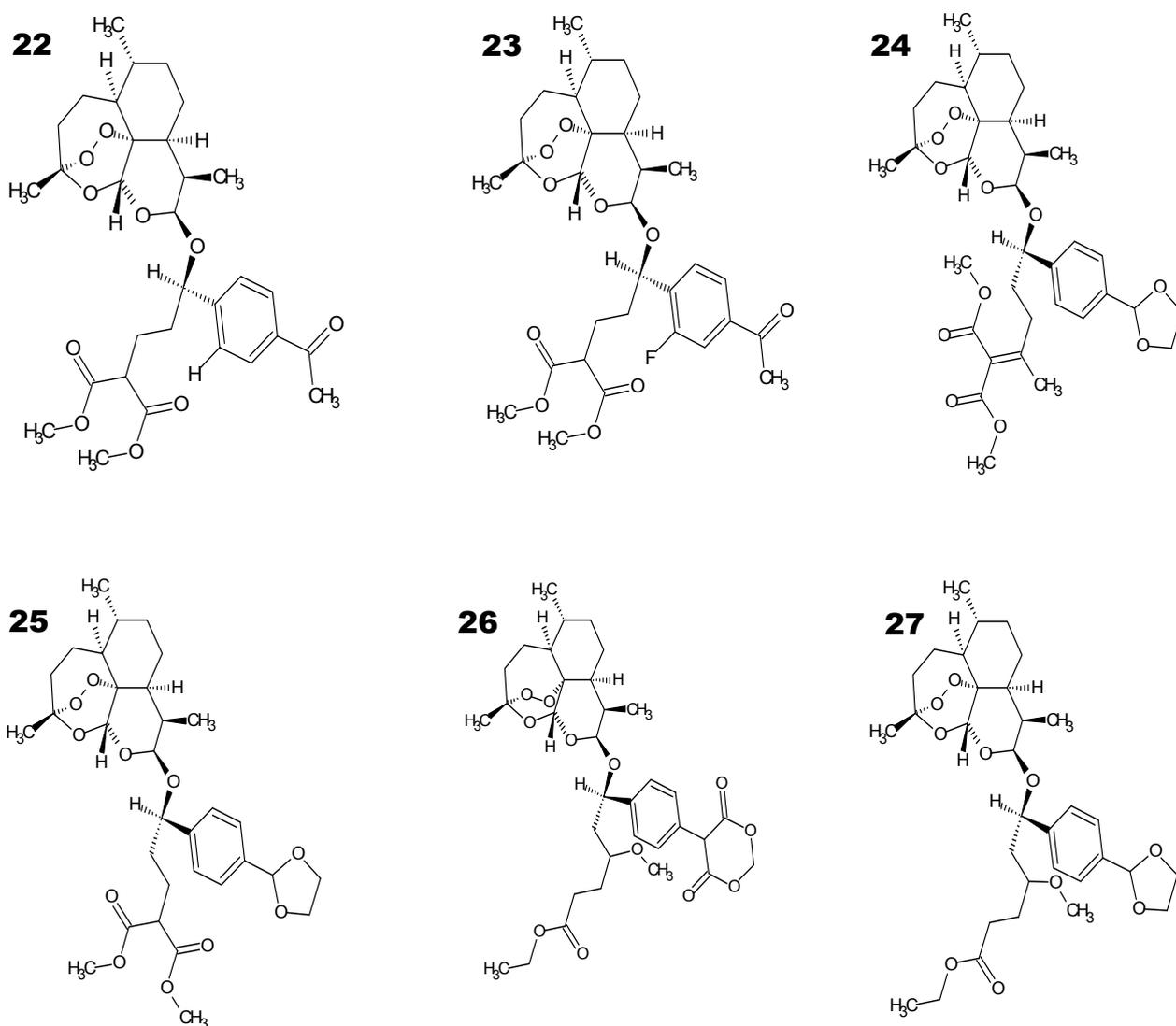


Figure 8. Cont.

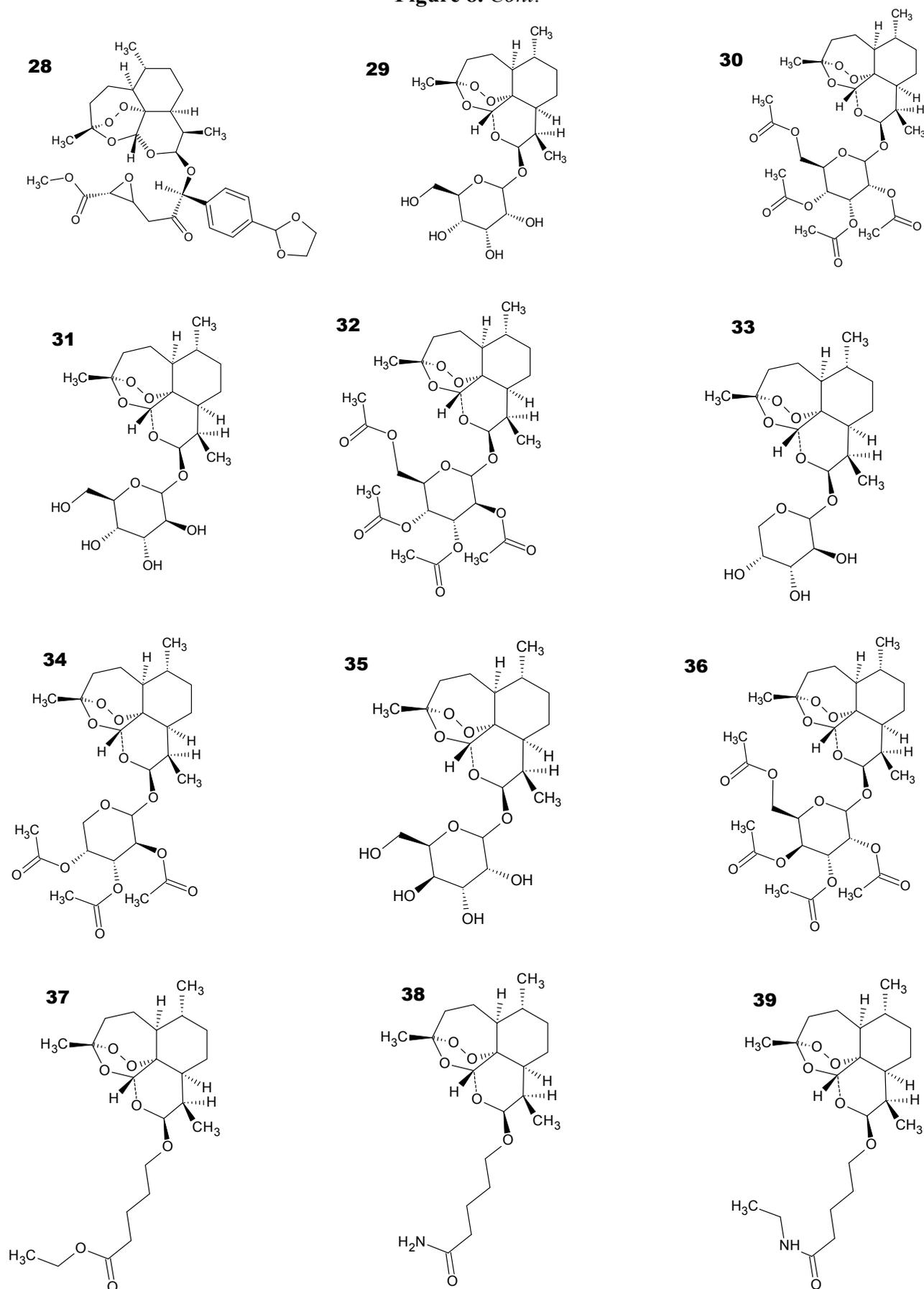


Figure 8. Cont.

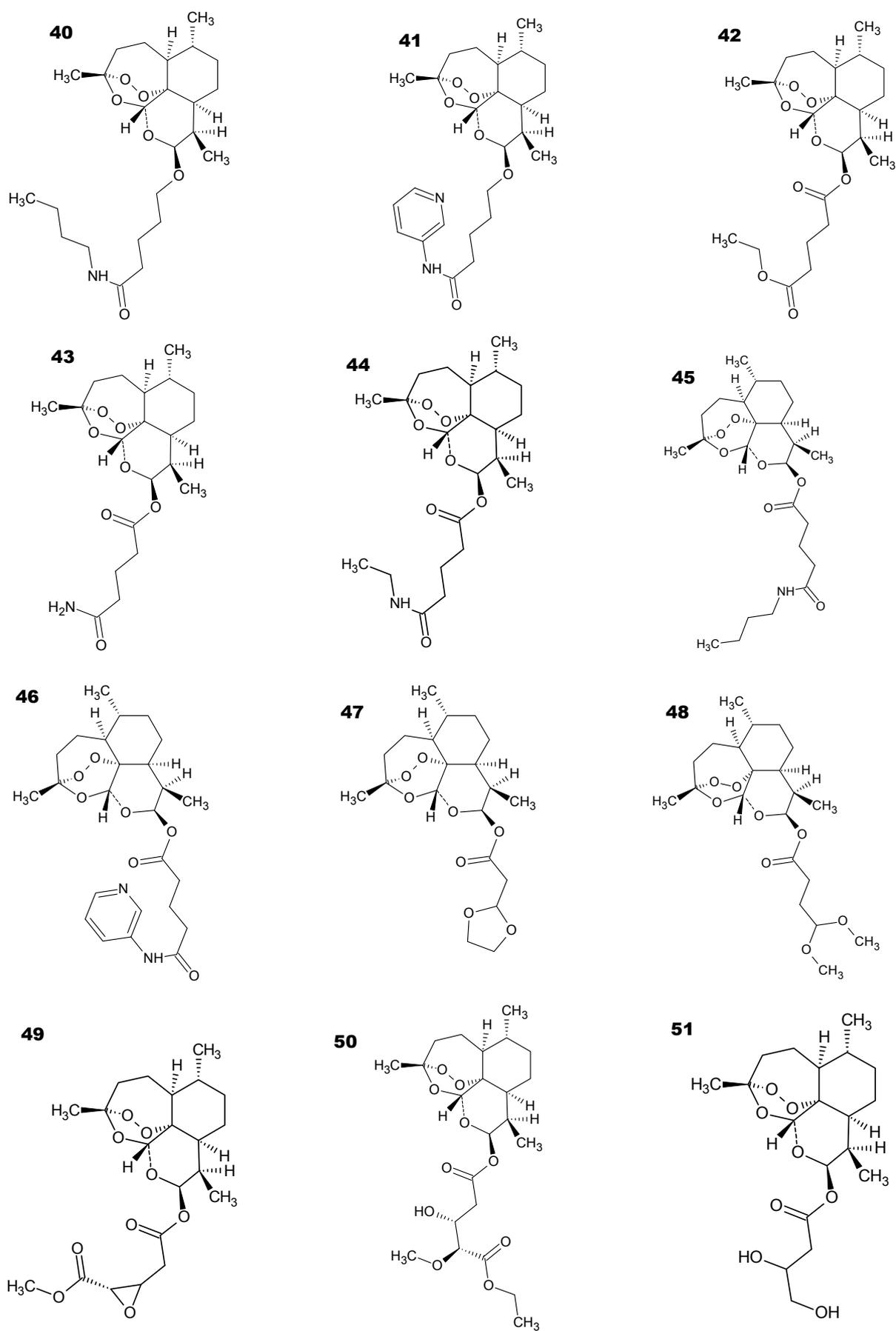


Table 7. Antimalarial activity predicted (log AR) by PCR and PLS models for the test set compounds and residues of prediction between models.

Test Set Compounds	Predicted (logRA)		Residues of Prediction (PLS–PCR)
	PLS	PCR	
22	0.28515	−0.08153	0.36669
23	0.16134	0.29643	−0.13510
24	−0.02669	−0.17894	0.15225
25	−0.13236	0.11823	−0.25059
26	−0.27509	−0.71065	0.43556
27	0.13050	−0.37989	0.51040
28	−0.25528	−0.60997	0.35469
29	−1.52738	−1.01052	−0.51686
30	0.02281	−0.04421	0.06701
31	−1.36654	−0.73626	−0.63028
32	0.18562	−0.72371	0.90933
33	−1.05361	−0.98490	−0.06871
34	0.13185	−0.44910	0.58095
35	−1.16288	−0.86252	−0.30036
36	0.20640	−0.49559	0.70198
37	0.23829	0.33770	−0.09941
38	0.27428	0.24533	0.02895
39	0.41609	−0.32859	0.74468
40	0.38362	−0.59287	0.97649
41	−0.66296	−0.46071	−0.20225
42	−0.19509	−0.67464	0.47955
43	−0.45107	0.10809	−0.55916
44	0.39618	−0.17717	0.57335
45	0.39508	−0.11549	0.51058
46	−0.30336	0.40099	−0.70434
47	0.05839	−0.02784	0.08623
48	−0.49407	0.20276	−0.69682
49	−0.50004	0.06913	−0.56917
50	−0.13357	0.19691	−0.33048
51	0.25967	−0.19372	0.45339

The most potent compounds have $\log RA \geq 0$; the less potent compounds have $\log RA < 0$.

3. Experimental

3.1. Compounds Studied

Initially, 21 molecules were selected from the literature (Figure 1). Compounds **2–7** were proposed by Lin *et al.* [55], who found that the acetylation of dihydroartemisinin derivatives containing a sugar leads to similar or better activity than that of artemisinin. However, the deacetylation of sugars considerably reduces the antimalarial activity. Compounds **8–13** were chosen to examine the impact of the stereospecificity of the alkyl side chain on the biological properties and were proposed by Lin *et al.* [56] to obtain compounds with better biological activity than the antimalarials artelinic acid, artemisinin,

artemether and arteether. However, the conversion of esters to their corresponding acids dramatically reduces their antimalarial activity. Compounds **14–21** were proposed by Lin *et al.* [57,58] because large substituents on the α -methylene carbon (*C) substantially increase the antimalarial activity of the compounds on the basis of electronic and steric effects and because the increased lipophilicity of artemisinin derivatives results in increased antimalarial activity. They were chosen for their *in vitro* bioactivity against the drug-resistant malarial strain *P. falciparum* (W-2 clone), which is chloroquine resistant but mefloquine sensitive. The numbering of the atoms used in this study is shown in Figure 1 (compound **1**—artemisinin). Because biological data were obtained from different sources, the logarithm of the IC_{50} value of artemisinin over the IC_{50} value of the compounds (logarithm of relative activity, $\log RA$) was used to reduce inconsistencies caused by individual experimental environments:

$$\log RA = \log (IC_{50} \text{ of artemisinin} / IC_{50} \text{ of an analog}) \quad (5)$$

where IC_{50} is the 50% inhibitory concentration. In this study, the following classification based on the antimalarial responses was adopted: compounds with $\log RA \geq 0.00$, ranging from 0.00000 to 0.86031, were assumed to be more potent analogs (**1, 3, 4, 10, 11, 15, 16, 19** and **20**), and those with $\log RA < 0.00$, ranging from -0.00634 to -2.40049 , were considered to be less potent analogs (**2, 5–9, 12–14, 17, 18** and **21**). Based on the relative activity (RA) values, compounds **3, 4, 10, 15, 16**, and **19** are 2–7 times more potent than artemisinin. Compound **15** is the most potent compound in the series studied.

3.2. Geometric Optimization and Descriptor Calculations

Molecular modeling started with the construction of the structure of artemisinin using GaussView 3.0 program [59], which was then optimized with different methods and basis sets—semiempirical (AM1, PM3 and ZINDO), ab initio/Hartree-Fock (HF/6-31G, HF/6-31G* and HF/6-31G**) and DFT (B3LYP/3-21G, B3LYP/3-21G* and B3LYP/3-21G**). These calculations were performed to find the method and basis sets with the best fit between the computational time and accuracy of the information compared to the experimental data [39]. After initial determination and structural optimization of artemisinin, the theoretical geometrical parameters of artemisinin in the region of the 1,2,13-trioxane ring (bond length, bond angle and torsion angle of the atoms that form this ring) were determined with the aim of evaluating the quality of the molecular wave function comparing the theoretical geometrical parameters with the experimental data (Table 1). The experimental structure of artemisinin was taken from the Cambridge Structural Database CSD, with REFCODES: QNGHSU10, crystallographic R factor 3.6 [60]. All the other structures (Figure 7) were built with the optimized structure of artemisinin using the Gaussian 03 program [61] with the Hartree-Fock (HF) method and 6-31G** basis set. After the structures were determined in three dimensions, various descriptors for each molecule of the set studied were calculated. They represent different sources of chemical information (features) regarding the molecules and include geometric, electronic, quantum-chemical, physical-chemical and topological descriptors, among others. They are important for the quantitative description of molecular structure and to finding appropriate predictive models [62]. The computation of the descriptors was performed employing the following software: Gaussian 03 program, [61] e-Dragon [63,64], Autodock 4.0 [65], Molekel [66] and HyperChem 6.02 [67]. With aid of the e-Dragon program, 1666 calculated descriptors were divided into the following 20 classes: 48 constitutional descriptors; 47 descriptors of

quantity and trajectory; 47 information indexes; 107 adjacency indexes; 21 topological charge indexes; 41 molecular Radic profiles; 150 RDF descriptors; 154 functional groups; 14 charge descriptors; 33 connectivity indexes; 96 2-D autocorrelations, 64 Burden eigenvalues; 44 indexes based on eigenvalues; 74 geometric descriptors; 160 MORSE-3D; 120 fragments centered in the atom; 31 molecular property descriptors; 119 topological indexes; 99 WHIM descriptors; and 197 Getaway descriptors. Other descriptors such as the following were obtained:

- (a) *MOLECULAR DOCKING* descriptors: These were calculated to better represent the interaction between the drug and receptor with the aid of the AutoDock 4.0 program. The following 17 molecular docking descriptors were included in the data matrix: binding energy (BE); partition function (Q), Gibbs free energy (G), internal energy (U), electrostatic energy (EE); bond length (Fe–O1, Fe–O2, Fe–O13 and Fe–O11), bond angle (O2–O1–Fe, O1–O2–Fe, C4–O1–Fe and C5a–O1–Fe); and dihedral angle (O2–O1–Fe–N1, O2–O1–Fe–N2, O2–O1–Fe–N3 and O2–O1–Fe–N4).
- (b) *QUANTUM CHEMICAL* descriptors: In our study, we calculated the following 25 quantum-chemical descriptors: total energy (ET), energy of the highest occupied molecular orbital (HOMO), a level below the energy of the highest occupied molecular orbital (HOMO-1), lowest unoccupied molecular orbital energy (LUMO), a level above the energy of the lowest unoccupied molecular orbital (LUMO + 1), difference in energy between HOMO and LUMO (GAP = HOMO-LUMO), Mulliken electronegativity (χ), molecular hardness (η), molecular softness ($1/\eta$), and charge on the atom n (where $n = 1, 2, 3, 4, 5, 5a, 6, 7, 8, 8a, 9, 10, 11, 12, 12a, 13$). The atomic charges used in this study were obtained with the key word POP = CHELPG using the electrostatic potential [68]. With this strategy, it was possible to obtain the best potential quantum molecular series of points defined around the molecule, and atomic charges offer the general advantage of being physically more satisfactory than Mulliken charges [69].
- (c) *Descriptors related to quantitative properties of chemical structure and biological activity*: In our data matrix, QSAR descriptors were included, *i.e.*, total surface area (TSA), molecular volume (MV), molar refractivity (MR), molar polarizability (MP), coefficient of lipophilicity (logP), molecular mass (MM) and hydration energy (HE) according to the HyperChem 6.02 program. The molecular descriptors were selected to provide valuable information about the influence of electronic, steric, hydrophilic and hydrophobic features on the antimalarial activity of artemisinin.

3.3. Interaction between Artemisinins and Heme

The interaction between the ligands (artemisinins) and the receptor (heme) was studied with molecular docking to determine the best geometry for the complex formed between these two molecules. The geometry of artemisinin and its derivatives (ligands) was designed with HF/6-31G**, whereas the geometry of heme (receptor) was obtained from the 1A6M structure in the RCSB protein data bank (PDB) from Vojtechovsky *et al.* [70]. The arrangement in the docking calculation took into account the presence of the proximal histidine residue under the plane of the porphyrin ring. This histidine moiety is, as usual, perpendicularly coordinated to Fe²⁺ through the sp² nitrogen atom of its

imidazole ring. Such an arrangement allows the Fe^{2+} to attain a nearly octahedral hexacoordinated arrangement after binding to the artemisinin molecule [14]. The orientation of the ligand was set just above the plane of the heme. Then, for each ligand/receptor interaction, 100 conformations were calculated, and the most probable one was determined based on the lowest energies of interaction. Automated docking calculations were performed to develop possible conformations for the complex employing the Lamarckian Genetic Algorithm implemented in the package Autodock 4.0. This program starts the docking by displaying the ligand in an arbitrary conformation and position and looking for favorable dockings with the receptor using both simulating annealing and genetic algorithms. AutoDock uses a random number generator to create new poses for the ligand during its search and estimates the free energy of binding of a ligand to its target. The resulting conformations were ranked in order of increasing binding energy of the lowest binding energy conformation in each cluster.

3.4. Molecular Electrostatic Potential Maps

An important concept explored in this research was the correlation of the structure activity of the species studied here through the characteristics of the electrostatic potential in the region of the 1,2,13-trioxane ring because in the literature, artemisinin and derivatives with antimalarial activities present similar patterns in their *MEP* maps [36,37,71]. Such a method enables the use of a qualitative analysis to locate reactive sites in a molecule and determine the roles played by both the electronic and steric (size/shape) effects on its potency. It is worthwhile to note that the visualization of *MEP* maps provides qualitative information on molecules, such as the behavior of the interaction between a ligand and a receptor. The *MEP* at a given point (x, y, z) in the vicinity of a molecule is defined in terms of the interaction energy between the electrical charge generated from the molecules electrons and nuclei and a positive test charge (a proton) located at r . For the studied compounds, the $V(r)$ values were calculated by Equation (6) as described previously (see [72]):

$$V(r) = \sum_A \frac{Z_A}{|R_A - r|} - \int \frac{\rho(r')}{|r' - r|} dr' \quad (6)$$

where Z_A is the charge of nucleus A, located at R_A , $\rho(r')$ is the electronic density function of the molecule, and r' is the dummy integration variable. The *MEP* maps for artemisinin and its derivatives were computed from the atomic charge at the HF/6-31G** level using the Gaussian 03 program, and the results are displayed with Molekel software.

3.5. Variable Selection and Model Building

After the determination of all molecular descriptors, it was possible to construct a data matrix to develop step multivariate analysis. Before we began the multivariate analysis, it was necessary to make the autoscale or standardizing data matrix $X = (n, m)$ consisting of 21 lines (the compounds studied) and 1,733 columns (in this case, the calculated descriptors for each molecule), where n is the number of studied compounds and m is the number of variables. The aim of using the standardizing matrix is to give each variable equal weight in mathematical terms, so each variable was centered on the mean and scaled to unit variance. To reduce the data set, variables were selected based on the analysis of the

correlation matrix between variables (descriptors) and the logarithm of the relative activity ($\log RA$). Those with small or no correlation (under the 0.30 correlation value cutoff) were discarded, except for QSAR and quantum chemical descriptors, resulting in only 230 descriptors remaining from the initial set of 1,733 descriptors. After this data compression, two complementary methods for exploratory data analysis were employed (PCA and HCA) to study intersample and intervariable relationships and to select the properties that contribute the most to the classification of the compounds into two groups. One group contained more potent analogs and the other less potent analogs. PCA was employed to reduce the dimensionality of the data, find descriptors that could be useful in characterizing the behavior of the compounds acting against malaria and look for natural clustering in the data and outlier samples. While performing PCA, several attempts to obtain a good classification of the compounds were made. At each attempt, the score and loading plots were analyzed based on the variables employed in the analysis. The score plot gives information about the compounds (similarity and differences). The loading plot gives information about the variables (how they are connected to each other and which best describe the variance in the original data). The descriptors selected by PCA were used to perform HCA, PLS and PCR. The objective of HCA was to present the compounds distributed in natural groups and the results confirm the PCA results. Thus, several approaches were attempted to establish links between samples/cluster. All of them were of an agglomerative type because each sample was first defined as its own cluster, and then others were grouped together to form new clusters until all the samples were part of a single cluster.

The QSAR models for the artemisinin compounds studied were constructed by the PCR and PLS methods based on the autoscaled data and the leave-one-out crossvalidation procedure [28,29]. The final purpose of the multivariate analysis (PLS and PCR) was the construction of a mathematical model that can be used to predict antimalarial activity. The samples selected to compose the external validation set were **2**, **12** and **13**. The statistical parameters used to assess the quality of the models were the Prediction Residual Error Sum of Squares ($PRESS$), Equation (7), the Standard Error of Validation (SEV), Equation (8), the total variance explained, R^2 (correlation between the estimated values predicted by the model built with the full data set and actual values of y), Q^2 (the cross-validated correlation coefficient) and S_{PRESS} (standard deviation of cross-validation) given by Equations (9)–(11), respectively [28,29,73–75]:

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (7)$$

$$SEV = \sqrt{\frac{PRESS}{n}} \quad (8)$$

$$R^2 = 1 - \left[\frac{PRESS_{cal}}{\sum_{i=1}^n (y_i - \bar{y})^2} \right] \quad (9)$$

$$Q^2 = 1 - \left[\frac{PRESS_{val}}{\sum_{i=1}^n (y_i - \bar{y})^2} \right] \quad (10)$$

$$S_{PRESS} = \frac{\sqrt{PRESS}}{n - k - 1} \quad (11)$$

In Equations (7) and (8), n is the number of compounds used for the calibration or validation model, y_i is the experimental value of the physicochemical property for the sample and \hat{y}_i is the value predicted by a calibration or validation model. In Equations (9) and (10), $PRESS_{cal}$ is the Calibration Prediction Error Sum of Squares and $PRESS_{val}$ is the Validation Prediction Error Sum of Squares. Both $PRESS_{cal}$ and $PRESS_{val}$ are evaluated from Equation (7) by changing \hat{y}_i for a calibration or validation model. The values of explained variance (R^2_A , *i.e.*, adjusted R^2), standard deviation (s) and F (Fisher test) were determined. The multivariate data analyses (PCA, HCA, PLS and PCR) were performed by employing the Pirouette 3.01 software [50].

4. Conclusions

The HF method and the 6-31G** basis set revealed themselves to be adequate to optimize the structures of artemisinin and derivatives for consequent study. The molecular docking studies reinforced the idea that the Fe^{2+} ion from heme preferentially binds the O1 atom from artemisinins rather than the O2 atom and that such a preference may be due to a greater steric hindrance at O2 than O1 and a more negative charge on the latter atom. Both factors are essential for intermolecular approach. MEP maps characterize the region of the 1,2,13-trioxane ring in artemisinin and derivatives as a region of negative electrostatic potential, and the use of MEP maps identified key structural features necessary for antimalarial activity. Investigation of the interaction with the molecular receptor (heme) showed that the presence of a red surface near the 1,2,13-trioxane ring suggested that these compounds have a reactive site for electrophilic attack. This attack preferentially occurs through the endoperoxide linkage. The predictive classification models for artemisinin derivatives were obtained with a set of molecular descriptors selected by chemometric approaches. PCA and HCA methods classified the studied compounds into groups according to their degree of antimalarial activity against *P. falciparum* (W-2 clone). The descriptors hydration energy (HE), charge on oxygen atom of O11 (QO11), torsion angle O1–O2–Fe–N2 (D2) and maximum rate of R/Sanderson Electronegativity (RTe^+) were responsible for distinguishing compounds with higher and lower antimalarial activity. The molecular features represented by these descriptors are in good agreement with previous SAR analysis performed on artemisinin derivatives. The combination of these structural attributes is believed to govern the antimalarial effects of the compounds studied in this work. The PLS and PCR models obtained here showed not only statistical significance but also predictive ability. Through this strategy and our findings, useful information was obtained that could be of use in experimental syntheses and biological evaluation to understand the molecular and structural requirements for designing new ligands to be used as antimalarial agents.

Acknowledgments

We gratefully acknowledge the financial support provided by the Brazilian Agency National Council of Scientific and Technological Development (CNPq Proc. 306676/2010-9) and Institute of Exact and Natural Sciences of Federal University of Pará for use of the GaussView and Gaussian software. The authors would like to thank the Postgraduate Program in Biotechnology and Biodiversity—Network BIONORTE, Laboratory of Modeling and Computational Chemistry of

Federal University of Amapá for computational support to the Francinaldo Sarges Braga and Adriana Maciel Ferreira for technical support.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Opsenica, D.M.; Solaja, B.A. Antimalarial peroxides. *J. Serb. Chem. Soc.* **2009**, *74*, 1155–1193.
2. Arav-Boger, R.; Shapiro, T.A. Molecular mechanisms of resistance in antimalarial chemotherapy: The unmet challenge. *Annu. Rev. Pharmacol. Toxicol.* **2005**, *45*, 565–585.
3. White, N.J. Antimalarial drug resistance. *J. Clin. Invest.* **2004**, *113*, 1084–1092.
4. Ridley, R.G. Malaria: To kill a parasite. *Nature* **2003**, *424*, 887–889.
5. O'Neill, P.M.; Searle, N.L.; Kan, K.W.; Storr, R.C.; Maggs, J.L.; Ward, S.A.; Raynes, K.; Park, B.K. Novel, potent, semisynthetic antimalarial carba analogues of the first-generation 1,2,4-trioxane artemether. *J. Med. Chem.* **1999**, *42*, 5487–5493.
6. Woolfrey, J.R.; Avery, M.A.; Doweiko, A.M. Comparison of 3D quantitative structure-activity relationship methods: Analysis of the *in vitro* antimalarial activity of 154 artemisinin analogues by hypothetical active-site lattice and comparative molecular field analysis. *J. Comput. Aided Mol. Des.* **1998**, *12*, 165–181.
7. Klayman, D.L. Qinghaosu (artemisinin): An antimalarial drug from China. *Science* **1985**, *228*, 1049–1055.
8. Srivastava, M.; Singh, H.; Naik, P.K. Molecular modeling evaluation of the antimalarial activity of artemisinin analogues: Molecular docking and rescoring using Prime/MM-GBSA Approach. *Curr. Res. J. Biol. Sci.* **2010**, *2*, 83–102.
9. Creek, D.J.; Chalmers, D.K.; Charman, W.N.; Duke, B.J. Quantum chemical study of the intermediate complex required for iron-mediated reactivity and antimalarial activity of dispiro-1,2,4-trioxolanes. *J. Mol. Graph. Model.* **2008**, *27*, 394–400.
10. O'Neill, P.M. The therapeutic potential of semi-synthetic artemisinin and synthetic endoperoxide antimalarial agents. *Expert Opin. Invest. Drugs* **2005**, *14*, 1117–1128.
11. Biagini, G.A.; O'Neill, P.M.; Bray, P.G.; Ward, S.A. Current drug development portfolio for antimalarial therapies. *Curr. Opin. Pharmacol.* **2005**, *5*, 473.
12. Haynes, R.K.; Vonwiller, S.C. The behaviour of Qinghaosu (Artemisinin) in the presence of Heme Iron (II) and (III). *Tetrahedron Lett.* **1996**, *37*, 253–256.
13. Posner, G.H.; Wang, D.; Cumming, J.N.; Oh, C.H.; French, A.N.; Bodley, A.L.; Shapiro, T.A. Further evidence supporting the importance of and the restrictions on a carbon-centered radical for high antimalarial activity of 1,2,4-trioxanes like artemisinin. *J. Med. Chem.* **1995**, *38*, 2273–2275.
14. Araújo, J.Q.; Carneiro, J.W.M.; Araújo, M.T.; Leite, F.H.A.; Taranto, A.G. Interaction between artemisinin and heme. A Density Functional Theory study of structures and interaction energies. *Bioorg. Med. Chem.* **2008**, *16*, 5021–5029.
15. Jefford, C.W. Why artemisinin and certain synthetic peroxides are potent antimalarials. Implications for the mode of action. *Curr. Med. Chem.* **2001**, *8*, 1803–1826.

16. Ferreira, J.E.V.; Figueiredo, A.F.; Barbosa, J.P.; Crispino, M.G.G.; Macedo, W.J.C.; Silva, O.P.P.; Malheiros, B.V.; Serra, R.T.A.; Pinheiro, J.C. A study of new antimalarial artemisinins through molecular modeling and multivariate analysis. *J. Serb. Chem. Soc.* **2010**, *75*, 1533–1548.
17. Meshnick, S.R.; Yang, Y.Z.; Lima, V.; Kuypers, F.; Kamchonwongpaisan, S.; Yuthavong, Y. Iron-dependent free radical generation from the antimalarial agent artemisinin (qinghaosu). *Antimicrob. Agents Chemother.* **1993**, *37*, 1108–1114.
18. Jung, M. Current developments in the chemistry of artemisinin and related compounds. *Curr. Med. Chem.* **1994**, *1*, 35–49.
19. Cumming, J.N.; Ploypradith, P.; Posner, G.H. Antimalarial activity of artemisinin (qinghaosu) and related trioxanes: Mechanism(s) of action. *Adv. Pharmacol.* **1997**, *37*, 253–299.
20. Cohen, N.C. *Guidebook on Molecular Modeling in Drug Design*; Academic Press: San Diego, CA, USA, 1996.
21. Sant’Anna, C.M.R. Glossário de Termos Usados no Planejamento de Fármacos (Recomendações da IUPAC para 1997). *Quim. Nova* **2002**, *25*, 505–512.
22. Carvalho, I.; Borges, A.D.L.; Bernardes, L.S.C. Medicinal chemistry and molecular modeling: An integration to teach drug structure-activity relationship and the molecular basis of drug action. *J. Chem. Educ.* **2005**, *82*, 588–596.
23. Wermuth, C.G. *The Practice of Medicinal Chemistry*, 3rd ed.; Academic Press: London, UK, 2009.
24. Ribeiro, F.A.L.; Ferreira, M.M.C. QSPR models of boiling point, octanol-water partition coefficient and retention time index of polycyclic aromatic hydrocarbons. *J. Mol. Struct. Theochem.* **2003**, *663*, 109–126.
25. Cristino, M.G.G.; Meneses, C.C.F.; Soeiro, M.M.; Ferreira, J.E.V.; Figueiredo, A.F.; Barbosa, J.P.; Almeida, R.C.O.; Pinheiro, J.C.; Pinheiro, A.L.R. Computational modeling of antimalarial 10-substituted deoxoartemisinins. *J. Theor. Comput. Chem.* **2012**, *11*, 241–263.
26. Gramatica, P. Principles of QSAR models validation: Internal and external. *QSAR Comb. Sci.* **2007**, *26*, 694–701.
27. Geladi, P.; Kowalski, B.R. Partial least-squares regression: A tutorial. *Anal. Chim. Acta* **1986**, *185*, 1–17.
28. Geladi, P. Notes on the history and nature of partial least squares (PLS) modeling. *J. Chemom.* **1988**, *2*, 231–246.
29. Ferreira, M.M.C. Multivariate QSAR. *J. Braz. Chem. Soc.* **2002**, *13*, 742–753.
30. Leach, A. *Molecular Modelling—Principles and Applications*, 2nd ed.; Pearson Education Limited: Upper Saddle River, NJ, USA, 2001.
31. Hehre, W.J. *A Guide to Molecular Mechanics and Quantum Chemical Calculations*; Wavefunction, Inc.: Irvine, CA, USA, 2003.
32. Mulliken, R.S.; Liu, B. Self-consistent-field wave functions of P2 and PO, and the role of d functions in chemical bonding and of s-p hybridization in N2 and P2. *J. Am. Chem. Soc.* **1971**, *93*, 6738–6744.
33. Levine, I.N. *Quantum Chemistry*, 4th ed.; Prentice-Hall: New York, NY, USA, 1991.
34. Santos, C.B.R.; Lobato, C.C.; Vieira, J.B.; Brasil, D.S.B.; Brito, A.U.; Macêdo, W.J.C.; Carvalho, J.C.T.; Pinheiro, J.C. Evaluation of quantum chemical methods and basis sets applied in the molecular modeling of artemisinin. *Comput. Mol. Biosci.* **2013**, *3*, 66–79.

35. Pinheiro, J.C.; Kiralj, R.; Ferreira, M.M.C.; Romero, O.A.S. Artemisinin derivatives with antimalarial activity against *Plasmodium falciparum* designed with the aid of *Quantum* chemical and partial least squares methods. *QSAR Comb. Sci.* **2003**, *22*, 830–842.
36. Cardoso, F.J.B.; Costa, R.B.; Figueiredo, A.F.; Barbosa, J.P.; Nava-Junior, I.; Pinheiro, J.C.; Romero, O.A.S. Modeling artemisinin derivatives with potent activity against *P. falciparum* malaria with *ab initio* and PLS methods. *Int. Electron. J. Mol. Des.* **2007**, *6*, 122–134.
37. Cardoso, F.J.B.; Figueiredo, A.F.; Lobato, M.S.; Miranda, R.M.; Almeida, R.C.O.; Pinheiro, J.C. A study on antimalarial artemisinin derivatives using MEP maps and multivariate QSAR. *J. Mol. Model.* **2008**, *14*, 39–48.
38. Santos, C.B.R.; Vieira, J.B.; Formigosa, A.S.; Costa, E.V.M.; Pinheiro, M.T.; Silva, J.O.; Macêdo, W.J.C.; Carvalho, J.C.T. Validation of computational methods applied in molecular modeling of Artemisinin with antimalarial activity. *J. Comput. Theor. Nanosci.* **2014**, *11*, 553–561.
39. Lisgarten, J.N.; Potter, B.S.; Bantuzeko, C.; Palmer, R.A. Structure, absolute configuration, and conformation of the antimalarial compound, Artemisinin. *J. Chem. Cryst.* **1998**, *28*, 539–543.
40. Costa, M.S.; Kiralj, R.; Ferreira, M.M.C. Estudo teórico da interação existente entre a artemisinina e o heme. *Quim. Nova* **2007**, *30*, 25–31.
41. Tonmunphean, S.; Parasuk, V.; Kokpol, S. Automated calculation of docking of artemisinin to heme. *J. Mol. Model.* **2001**, *7*, 26–33.
42. Guha, R.; Jurs, P.C. Development of QSAR models to predict and interpret the biological activity of artemisinin analogues. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1440–1449.
43. Cheng, F.; Shen, J.; Luo, X.; Zhu, W.; Gu, J.; Ji, R.; Jiang, H.; Chen, K. Molecular docking and 3D-QSAR studies on the possible antimalarial mechanism of artemisinin analogues. *Bioorg. Med. Chem.* **2002**, *10*, 2883–2891.
44. Wulfsberg, G. *Principles of Descriptive Chemistry*; Brooks Cole Publishing: Monterey, CA, USA, 1987.
45. Karelson, M.; Lobanov, V.S.; Katritzky, A.R. Quantum-chemical descriptors in QSAR/QSPR studies. *Chem. Rev.* **1996**, *96*, 1027–1044.
46. Consonni, V.; Todeschini, R.; Pavan, M. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors, 1. Theory of the novel 3D molecular descriptors. *Chem. Inf. Comput. Sci.* **2002**, *42*, 682–692.
47. Consonni, V.; Todeschini, R.; Pavan, M.; Gramatica, P. Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors, 2. Application of the novel 3D molecular descriptors to QSAR/QSPR studies. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 693–705.
48. 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.
49. Posner, G.H. 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.
50. *Pirouette 3.01*; Infometrix Inc.: Seattle, WA, USA, 2001.
51. Snedecor, G.W.; Cochran, W.G. *Statistical Methods*; Oxford and IBH: New Delhi, India, 1967; p. 381.

52. Chatterjee, S.; Hadi, A.S.; Price, B. *Regression Analysis by Examples*, 3rd ed.; Wiley VCH: New York, NY, USA, 2000.
53. Diudea, M.V. *QSPR/QSAR Studies for Molecular Descriptors*; Nova Science: Huntington, NY, USA, 2000.
54. Bikash, D.; Shovanlal, G.; Subrata, B.; Soma, S.; Tarun, J. QSAR study on some pyridoacridine ascididemin analogues as anti-tumor agents. *Bioorg. Med. Chem.* **2003**, *11*, 5493–5499.
55. Lin, A.J.; Li, L.Q.; Andersen, S.L.; Klayman, D.L. Antimalarial activity of new dihydroartemisinin derivatives. 5. Sugar analogues. *J. Med. Chem.* **1992**, *35*, 1639–1642.
56. Lin, A.J.; Klayman, D.L.; Milhous, W.K. Antimalarial activity of new water-soluble dihydroartemisinin derivatives. *J. Med. Chem.* **1987**, *30*, 2147–2150.
57. Lin, A.J.; Miller, R.E. Antimalarial activity of new dihydroartemisinin derivatives. 6. alpha-Alkylbenzylic ethers. *J. Med. Chem.* **1995**, *38*, 764–770.
58. Lin, A.J.; Zikry, A.B.; Kyle, D.E. Antimalarial activity of new dihydroartemisinin derivatives. 7. 4-(p-substituted phenyl)-4(R or S)-[10(alpha or beta)-dihydroartemisininoxy]butyric acids. *J. Med. Chem.* **1997**, *40*, 1396–1400.
59. *GaussView 3.07*; Gaussian Inc.: Pittsburgh, PA, USA, 1997.
60. Allen, F.H. The Cambridge Structural Database: A quarter of a million crystal structures and rising. *Acta Cryst. B* **2002**, *58*, 380–388.
61. Frisch, M.J. *Gaussian 98 Revision A.11*; Gaussian, Inc.: Pittsburgh, PA, USA, 2001.
62. Estrada, E.; Molina, E. Novel local (fragment-based) topological molecular descriptors for QSPR/QSAR and molecular design. *J. Mol. Graph. Modell.* **2001**, *20*, 54–64.
63. Virtual Computational Laboratory, VCCLAB 2005. Available online: <http://www.vcclab.org/> (accessed on 20 January 2013).
64. 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.; *et al.* Virtual computational chemistry laboratory—design and description. *J. Comput. Aided Mol. Des.* **2005**, *19*, 453–463.
65. Auto-Dock 4.0. The scripps research institute, department of molecular biology: La Jolla, CA, USA, 2007.
66. *Molekel 4.3*. Swiss Center for Scientific Computing: Manno, Switzerland, 2000.
67. ChemPlus. *Modular Extensions to HyperChem, Release 6.02, Molecular Modeling for Windows*; Hyper, Inc.: Gainesville, FL, USA, 2000.
68. Breneman, C.M.; Winberg, K.B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.* **1990**, *11*, 361–373.
69. Singh, U.C.; Kollman P.A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **1984**, *5*, 129–145.
70. Vojtechovsky, J.; Chu, K.; Berendzen, J.; Sweet, R.M.; Schlichting, I. Crystal structures of myoglobin-ligand complexes at near-atomic resolution. *Biophys. J.* **1999**, *77*, 2153–2174.
71. Bernardinelli, G.; Jefford, C.W.; Marić, D.; Thomson, C.; Weber, J. Computational studies of the structures and molecules. *Int. J. Quantum Chem.* **1994**, *21*, 117–131.
72. Politzer, P.; Murray, J.S. The fundamental nature and role of the electrostatic potential in atoms and molecules. *Theor. Chem. Acc.* **2002**, *108*, 134–142.

73. Custódio, R.; Andrade, J.C.; Augusto, F.O. Ajuste de Funções Matemáticas a Dados Experimentais. *Quim. Nova* **1997**, *20*, 219–225.
74. Pimentel, M.; Neto, B.B. Calibração: Uma Revisão para Químicos Analíticos. *Quim. Nova* **1996**, *19*, 268–277.
75. Gaudio, A.C.; Zandonade, E. Proposição, validação e análise dos modelos que correlacionam estrutura química e atividade biológica. *Quim. Nova* **2001**, *24*, 658–671.

Sample Availability: Not available.

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).

Application of Hartree-Fock Method for Modeling of Bioactive Molecules Using SAR and QSPR

Cleydson B. R. Santos^{1,2,3*}, Cleison C. Lobato¹, Francinaldo S. Braga¹, Sílvia S. S. Morais⁴, Cesar F. Santos¹, Caio P. Fernandes^{1,3}, Davi S. B. Brasil^{1,5}, Lorane I. S. Hage-Melim¹, Williams J. C. Macêdo^{1,2}, José C. T. Carvalho^{1,2,3}

¹Laboratory of Modeling and Computational Chemistry, Federal University of Amapá, Macapá, Brazil

²Postgraduate Program in Biotechnology and Biodiversity-Network BIONORTE, Campus Universitário Marco Zero, Macapá, Brazil

³Laboratory of Drug Research, School of Pharmaceutical Sciences, Federal University of Amapá, Macapá, Brazil

⁴Laboratory of Statistical and Computational Modeling, University of the State of Amapá, Macapá, Brazil

⁵Institute of Technology, Federal University of Pará, Belém, Brazil

Email: breno@unifap.br

Received 12 January 2014; revised 12 February 2014; accepted 20 February 2014

Copyright © 2014 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

The central importance of quantum chemistry is to obtain solutions of the Schrödinger equation for the accurate determination of the properties of atomic and molecular systems that occurred from the calculation of wave functions accurate for many diatomic and polyatomic molecules, using Self Consistent Field method (SCF). The application of quantum chemical methods in the study and planning of bioactive compounds has become a common practice nowadays. From the point of view of planning it is important to note, when it comes to the use of molecular modeling, a collective term that refers to methods and theoretical modeling and computational techniques to mimic the behavior of molecules, not intend to reach a bioactive molecule simply through the use of computer programs. The choice of method for energy minimization depends on factors related to the size of the molecule, parameters of availability, stored data and computational resources. Molecular models generated by the computer are the result of mathematical equations that estimate the positions and properties of the electrons and nuclei, the calculations exploit experimentally, the characteristics of a structure, providing a new perspective on the molecule. In this work we show that studies of Highest Occupied Molecular Orbital Energy (HOMO), Low Unoccupied Molecular Orbital Energy (LUMO) and Map of molecular electrostatic potential (MEP) using Hartree-Fock method with different basis sets (HF/3-21G*, HF/3-21G**, HF/6-31G, HF/6-31G*, HF/

*Corresponding author.

6-31G** and HF/6-311G), that are of great importance in modern chemistry, biochemistry, molecular biology, and other fields of knowledge of health sciences. In order to obtain a significant correlation, it is essential that the descriptors are used appropriately. Thus, the quantum chemical calculations are an attractive source of new molecular descriptors that can, in principle, express all the geometrical and electronic properties of molecules and their interactions with biological receptor.

Keywords

Molecular Modeling; Meps Maps; Orbital Frontier; Quantum Chemical Methods

1. Introduction

Brazil has the largest genetic diversity in plant species in the world, however it is estimated that less than 10% were evaluated as their biological characteristics and less than 5% were submitted the detailed phytochemical studies. Despite a recent increase in research in this area, plants still constitute a source relatively underused and potentially very valuable for discovery of new biologically active substances [1] [2]. The bioactive compounds present in the vegetable kingdom have important functions and biological actions and may be considered as promoters of human health. We have already recognized the association between intake of fruits and vegetables and decreased risk of development various disorder chronic degenerative, such as cancer, inflammation, cardiovascular disease, cataracts, macular degeneration and other being carotenoids and phenolic compounds, some of the groups of bioactive compounds to which are attributed such actions. The seeds and extracts of annatto (*Bixa orellana* L.) are used as colorants in food, pharmaceutical and cosmetic industries due to the predominant presence of the carotenoid bixin [3]. The efficiency of biological activity of bioactive compounds of vegetable origin depends on their structure and concentration. In turn, the amount of these substances in vegetables is largely influenced by genetic factors and environmental conditions, in addition of the degree of maturation and plant variety, among other things. It is known also that the biological capacity is influenced by the substrate used in the assay, by the solvent and by extraction technique used, as well as the binomial time-temperature. Regarding the organic solvents, methanol, by managing to extract high amount of bioactive compounds, it has been touted as the most effective [4]-[7].

The *Montrichardia linifera* (Arruda) Schott, Araceae family, popularly known as “aninga”, is an aquatic macrophyte that forms large clonal populations along the rivers and streams of the Amazon. Those bordering consider this poisonous plant because its sap causes skin burns and eye contact can cause blindness [8]. However, paradoxically, it is widely used in traditional Amazonian medicine, mainly due to the healing properties of sap and of the juice of this plant, which has already been mentioned in the literature since the nineteenth century for the treatment of wounds and ulcers [9]-[12], leading to the hypothesis that this species may contain biologically active substances. However, very little is known about their chemical composition and biological activities that eventually may be useful also against human infections caused by parasites, which are serious problems in tropical and subtropical in developing countries, despite the discovery of new anti-protozoal drugs [13]. Between antimalarial compounds isolated from plants, artemisinin is one of the most important discoveries nowadays [14]-[16]. Artemisinin (or Qinghaosu, QHS, **Figure 1**) represents the most relevant advance in the treatment of malarial disease for the last 20 years [17]. Artemisinin is a sesquiterpene lactone with an endoperoxide group, which has been used in traditional Chinese medicine for many centuries as a natural product for fever and malarial treatment. This drug was isolated by Chinese chemists in the early 1970s from the ancient *Artemisia annua* L. Nowadays, artemisinin and derivatives are widely used around the world because of their potent antimalarial activity, fast action, and low toxicity. As a result, artemisinin and its derivatives have become recognized as a new generation of antimalarial drugs [18].

Pinheiro, Ferreira and Romero (2001) techniques combined quantum chemical (Hartree-Fock 3-21G) and multivariate analyzes methods (PCA, HCA, KNN and SIMCA) to study and propose diidroartemisinin derivatives. Through the technique PCA and HCA, seven (7) descriptors that were responsible for the classification of compounds into two distinct classes were selected, and with construction of qualitative models KNN and SIMCA

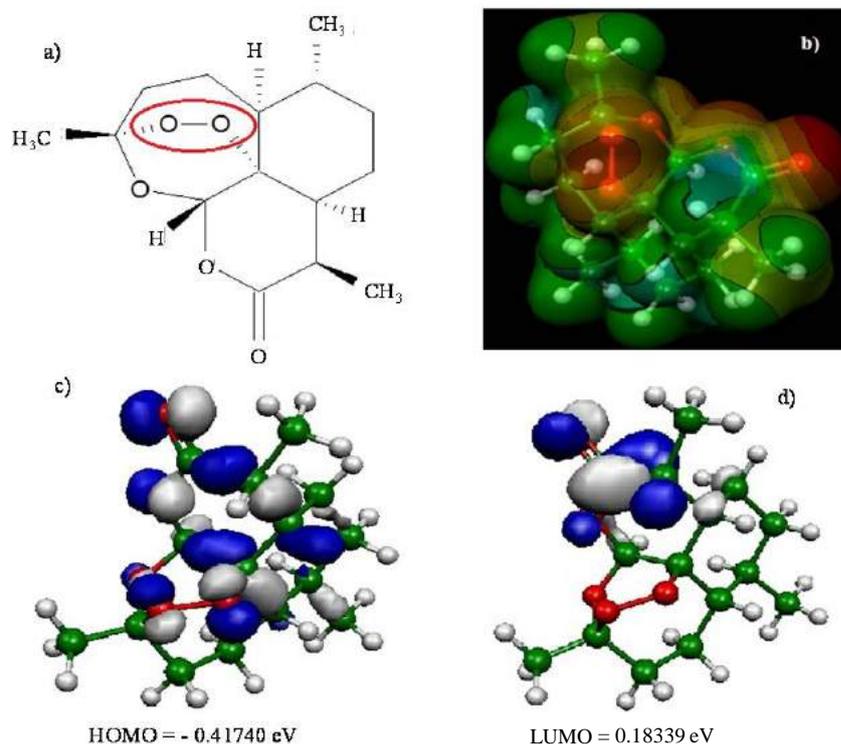


Figure 1. (a) The artemisinin structure and the region essential for expression of the biological activity (pharmacophore) was visualized using ChemSketch 12.00 program [23]. (b) Map of molecular electrostatic potential (MEP). ((c) and (d)) Orbital's energy HOMO and LUMO. The MEP, HOMO and LUMO were visualized and calculated using Hartree-Fock (HF) method and HF/6-31G** basis set by the Molekel program [24].

two (2) compounds of a set of twelve (12) test predicted as of high activity were proposed [19]. Artemisinin derivatives with antimalarial activity against *Plasmodium falciparum*, which is resistant to mefloquine, were studied using quantum chemical methods (HF/6-31G*) and the partial least-squares (PLS) method. Three main components explained 89.55% of the total variance, with $Q^2 = 0.83$ and $R^2 = 0.92$. From a set of 10 proposed artemisinin derivatives (artemisinin derivatives with unknown antimalarial activity against *Plasmodium falciparum*), a novel compound was produced with superior antimalarial activity compared with the compounds previously described in the literature [20]. Cardoso *et al.* (2008) studied artemisinin and some of its derivatives with activity against D-6 strains of *Plasmodium falciparum* using the HF/3-21G method. To verify the reliability of the geometry obtained, Cardoso *et al.* compared the structural parameters of the artemisinin trioxane ring with theoretical and experimental values from the literature. MEP was used in an attempt to identify key features of the compounds that are necessary for their activities, and they use those to propose new artemisinin derivatives [21]. There is however, the need for discovery of new antimalarial drugs using techniques of quantum chemistry which has been extensively used in the pharmaceutical industry, where predicting the probable activity of a drug from molecular orbital calculations is much less expensive than manufacturing it in order to perform expensive tests [22].

2. The Importance and Application of the Methods of Quantum Chemistry to Molecular Modeling

The central importance of quantum chemistry is to obtain solutions of the Schrödinger equation for the accurate determination of the properties of atomic and molecular systems. Within the area of physical chemistry, quantum chemistry is used to calculate various thermodynamic properties such as entropy, heat capacity of gases, for the interpretation of molecular spectra, calculations of bond length, bond angles, dipole moment and the understanding of intermolecular forces. In organic quantum chemistry can estimate the stability of molecules, calculate the properties of intermediate reactions, reproduce the aromaticity of organic compound and simulate spec-

tra of nuclear magnetic resonance (NMR). In analytical chemistry is used for the interpretation of intensities of the spectral lines. In the area of inorganic chemistry is used for the ligand field theory, where predictions are made, and properties justify complex ions of transition metals. In Biochemistry of quantum chemistry calculations are used to study the conformation of biological molecules such as enzyme-substrate binding and solvation of organic molecules [25]-[28].

The application of quantum chemical methods in the study and planning of bioactive compounds has become a common practice nowadays. From the point of view of planning, it is important to note that, when speaking of the use of molecular modeling (a collective term that refers to theoretical methods and computational techniques to model or mimic the behavior of molecules), it is not intended to reach a bioactive molecule simply through the use of computer programs. The development process of these molecules due to their complexity, necessarily involves multidisciplinary approach, which employs a large set of computational methods of systematic way form facilitate and optimize the development process of bioactive compounds, on a constant exchange of information with groups of chemical synthesis and evaluation of the activity of these compounds. These computational methods can be used as tools of rational design of bioactive compounds, so called because it is guided by a rational hypothesis about the mechanism of action of these compounds. The action of bioactive molecules is a very complex phenomenon, but one of the paradigms of medicinal chemistry is that these molecules are related to their effects or interactions to chemical reactions with macromolecular structures present in living systems, proteins, in its large majority. Cases these proteins are cell receptors, bioactive molecules are classified as receptor agonists or antagonists, in the case of enzymes, these molecules act as enzyme inhibitors [29] [30].

The Origin of Quantum Chemistry Methods

The history and debates related to the so called “semiempirical method” of London-Eyring-Polanyi are reported on a fascinating work of Nye [31]. The approach of Eyring and Polanyi, aiming to merge theory with experimental results to construct potential energy surfaces, showed that it is possible to gain insights into the mechanisms of adiabatic reactions, leading to important concepts related to the dynamics of chemical reactions, such as transition state and activated complex. Already at that time, there were debates comparing the approaches based on first principles with the semiempirical. In 1933 James and Coolidge [32] published an ab initio calculation when obtained the binding energy of H₂ molecule with an accuracy of 98%. This calculation took a year to be done, but its success increased confidence of authors that then criticized the semiempirical method as a “happy cancellation of errors” that occurs for not taking into account “terms of considerable importance” [32]. This expression “cancellation of errors” remains to this day in the imagination of researchers, although it is not possible to identify precisely that are these errors, and mainly, how and why they if would cancel. Polanyi in 1937 counter-argued: “Personally, I do not attach importance to an exact agreement between theory and experiment at this stage, but I believe that the theory [referring to the semiempirical method] can claim to provide a reasonable description of the mechanism of reactions chemical that would otherwise remain obscured” [33].

In 1941 Hirschfelder presented the calculation of at least one hundred activation energies for several reactions using the method of London-Eyring-Polanyi, showing that only in some cases the values did not agree with the experimental, with an error of 10 kcal/mol [34]. Hirschfelder defended the semiempirical method as “a method sufficiently flexible so that it could be made consistent with any set of chemical facts, maintaining consistent with the basic principles of quantum mechanics”. According to Nye [31], in this statement Hirschfelder has correctly identified that an ad hoc character is inherent to semi-empirical approaches. This ad hoc character involves risk, since it may well be possible to force an agreement between the semi-empirical and experimental model, even if the model is based on an erroneous theory. Such risk is even greater if the model is used for extrapolation. The final triumph of the first principles approach to the reaction $H + H_2 \rightarrow H_2 + H$, occur only came in 2003 when Mielke and colleagues [35] presented a new constants of experimental and theoretical speeds. The theoretical calculations were obtained using quantum dynamics in an exact adiabatic potential surface including Born-Oppenheimer corrections. The experiment and theory now agree in the range of 167 K to 2112 K, within experimental error, causing the problem can be solved now.

The “semiempirical” expression was first used in theoretical chemistry in 1931 by Michael Polanyi (1891-1976) and Henry Eyring (1901-1981) [36] [37] in their attempt to combine thermodynamics, chemical kinetics, quantum mechanics and the theory of binding electrons valence. In these 82 years, semiempirical method of London-Eyring-Polanyi to construct potential energy surfaces, provided a work tool useful to get insights about

how the physical and chemical processes occur, and to motivate the development of new techniques and experiments to study the temporary combination of atoms called transition state as in the works of Ahmed Zewail. The novelty of the visual maps of potential energy surfaces and the language of potential wells and barriers of activation has also become an important pedagogical tool for chemistry. In contrast, first-principles calculations of quantum dynamics, with equivalent accuracy to that of Mielke and colleagues [35], remain prohibitive, almost impossible to be performed for the vast majority of chemical reactions of practical interest. Undoubtedly: the *ab initio* methods, based entirely on first principles eventually prevail. However, if the reaction of the hydrogen atom with a hydrogen molecule could be regarded as definitely settled 10 years ago, what about all the other chemical and biochemical reactions? [38]

3. Hartree-Fock Equations

A development of great importance in quantum chemistry occurred from calculation functions accurate wave diatomic and for many polyatomic molecules, by the method of self-consistent field developed by Douglas Hartree [39]. In his theory Hartree considered the wave function as being formed by an antisymmetric linear combination of products of spin-orbitals. Then the wave function Hartree-Fock for atoms or molecules obeying the Pauli exclusion principle, should be written as a product antisimetrizado spin-orbitals, called Slater determinant, that in the formula restricted and normalized to a system of closed layers containing $2m$ electrons, is given by:

$$\Psi(1, 2, 2m) = \frac{1}{\sqrt{2m!}} \begin{vmatrix} \Phi_1(1) & \bar{\Phi}_1(1) & \cdots & \bar{\Phi}_m(1) \\ \Phi_1(2) & \bar{\Phi}_1(2) & \cdots & \bar{\Phi}_m(2) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_1(2m) & \bar{\Phi}_1(2m) & \cdots & \bar{\Phi}_m(2m) \end{vmatrix} \quad (1)$$

in which each spin-orbital $\Phi_i(x, \alpha)$ is a product of a spatial function $\varphi_i(x)$ and a spin function α or β

$$\Phi_i(x, \alpha) = \varphi_i(x)\alpha(i), \bar{\Phi}_i(x, \beta) = \varphi_i(x)\beta(i) \quad (2)$$

The functions of an electron $\varphi_i(x)$ are known as “orbitals”, a term proposed by Mulliken [40], which is the quantum mechanical analogue of the classical orbit [41]. It is considered that the orbitals are orthonormal

$$\langle \varphi_i(1) | \varphi_j(1) \rangle = \delta_{ij}, \delta_{ij} = \begin{cases} 0 & \text{se } i \neq j \\ 1 & \text{se } i = j \end{cases} \quad (3)$$

All elements forming a given column of Slater determinant involve the same spin-orbital already all elements in the same row involve the same electrons. When the configuration of an atom or molecule layer is closed this is represented by a single determinant, unlike a configuration layer to open, since in this case it is used a sum of Slater determinant. The analysis made the following refers the systems configurations closed layer. Open layer settings will not be discussed here, viewed the complexity of the formulas. Aiming to show a simple format for the expressions, it is used the system of atomic units in the equations.

The expression refers to the total electronic energy Hartree-Fock E for a system of electrons n nuclei in fundamental state is given by the theorem of the variational method.

$$E = \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \quad (4)$$

where Ψ comes to be wave function in the form of Slater determinant and \hat{H} is the Hamiltonian operator without considering the coordinates spin system and non-relativistic. In the case of molecules that in our case the operator \hat{H} is the Hamiltonian purely electronic which is obtained after the separation of Born-Oppenheimer [42] that written in atomic units has the form:

$$\hat{H} = -\frac{1}{2} \sum_i \nabla_i^2 - \sum_{\alpha} \sum_i \frac{Z_A}{r_{\alpha i}} + \sum_i \sum_{i < j} \frac{1}{r_{ij}} \quad (5)$$

The first term in Equation (5) is the operator for the kinetic energy of the electrons, the second term comes to be potential energy of attraction between electrons and nuclei, and the last term refers to the potential energy of repulsion between the electrons, Z_A is the charge of the nucleus A , $r_{\alpha i}$ is the distance of the electron i the nucleus α and r_{ij} is the distance of the electron i to the electron j .

The operator \hat{H} can be separated into two other operators

$$\hat{H} = \hat{H}_1 + \hat{H}_2 \quad (6)$$

The operator \hat{H}_1 is relative to the kinetic energy of the electrons and the potential energy of interaction of the electrons with the nuclei:

$$\hat{H}_1 = \sum_i^{2m} \hat{h}(i) \quad (7)$$

The operator $\hat{h}(i)$ is the corresponding Hamiltonian to the motion of an electron in the field generated by the nuclei only

$$\hat{h}(i) = -\frac{1}{2} \nabla_i^2 - \sum_{\alpha} \frac{Z_A}{r_{\alpha i}} \quad (8)$$

The operator \hat{H}_2 is associated with the interelectronic repulsion

$$\hat{H}_2 = \sum_{i < j} \sum \frac{1}{r_{ij}} \quad (9)$$

Utilizing the wave function (1) and the Hamiltonian (5) in the expression for the energy functional (4), has been as a result the expression for the total electronic energy of a system represented by a wave function in the form of Slater determinant.

$$E = 2 \sum_{i=1} h_{ii} + \sum_i \sum_j (2J_{ij} - K_{ij}) \quad (10)$$

This relationship has reduced the integration of many electrons the set of three dimensional integrals (h_{ij}) and six (J_{ij}) and (K_{ij}), written of following form:

$$h_{ii} = \left\langle \varphi_i(1) \left| \hat{h}(1) \right| \varphi_i(1) \right\rangle \quad (11)$$

$$J_{ij} = \left\langle \varphi_i(1) \varphi_j(2) \left| \frac{1}{r_{12}} \right| \varphi_i(1) \varphi_j(2) \right\rangle \quad (12)$$

$$K_{ij} = \left\langle \varphi_i(1) \varphi_j(2) \left| \frac{1}{r_{12}} \right| \varphi_j(1) \varphi_i(2) \right\rangle \quad (13)$$

The integral of an electron h_{ii} represents the sum of the kinetic energy of an electron in the orbital $\varphi_i(1)$ and its potential integral due to the action of the nuclei. The Coulomb integrals J_{ij} and of exchange K_{ij} are associated to interaction between an electron in the orbital $\varphi_i(x)$ and another in the orbital $\varphi_j(x)$. In classical mechanics the integral Coulomb is the interaction energy between two distributions of charges, already integral exchange does not have analogue in classical mechanics, their presence arises due to anti-symmetry of the wave function [41]. Varying E in Equation (4) for each orbital and making $\delta E = 0$ while keeping orthogonality of the wave function arrives at the 1-electron equations that define the Hartree-Fock method

$$\hat{F}(1) \varphi_i(1) = \varepsilon_i \varphi_i(1), i = 1, 2, \dots, m \quad (14)$$

where Fock operator $\hat{F}(1)$ is given by

$$\hat{F}(1) = \hat{h}(1) + \sum_j^m [2\hat{J}_j(1) - \hat{K}_j(1)] \quad (15)$$

where \hat{J}_i and \hat{K}_j are respectively operators of Coulomb and exchange

$$\hat{J}_i(1)\varphi_i(1) = \left[\int \varphi_j^*(2) \frac{1}{r_{12}} \varphi_j(2) d\tau_2 \right] \varphi_i(1) \quad (16)$$

$$\hat{K}_j(1)\varphi_i(1) = \left[\int \varphi_j^*(2) \frac{1}{r_{12}} \varphi_i(2) d\tau_2 \right] \varphi_j(1) \quad (17)$$

The Hartree-Fock equations are solved in general by an interactive procedure (SCF). At the end of the procedure, ε_i are the eigenvalues monoelétrônicos HF system. Each ε_i is often called orbital energy and is interpreted as being the energy of an electron in the orbital φ_i , resulting from its kinetic energy, energy of attraction with the nuclei and their energy exchange and repulsion due to all other electrons in their charge density $|\varphi_i|^2$.

The Equation (14) provides

$$\varepsilon_i = \langle \varphi_i(1) | F(1) | \varphi_i(1) \rangle \quad (18)$$

Substituting Equation (15) in (18) we can relate the orbital energies with the integrals (11), (12) and (13).

$$\varepsilon_i = h_{ii} + \sum_{j=1}^m (2J_{ij} - K_{ij}) \quad (19)$$

It may be also find the total electron energy E from the orbital energies. But this total energy is not simply equal the sum of the energies of an electron. This fact is due to the sum of the energies of an electron including twice each electron-electron interaction, *i.e.*, repulsion between electrons 1 and 2 contributes to the energy of an electron associated with both electrons. Thus it follows that the second term in the equation below corrects this problem

$$E = 2 \sum_i^m \varepsilon_i + \sum_i^m \sum_j^m (2J_{ij} - J_{ij}) \quad (20)$$

Thus substituting (19) in (20) have

$$E = \sum_{i=1}^m \left[2h_{ii} + \sum_{j=1}^m (2J_{ij} - K_{ij}) \right] \quad (21)$$

which equals

$$E = \sum_{i=1}^m (\varepsilon_i - h_{ii}) \quad (22)$$

with h_{ii} being the same Equation (11).

3.1. Hartree-Fock-Roothaan Method

The problem of solving the Hartree-Fock equations for atoms and molecules is due to the fact of absence of central symmetry. Therefore it is necessary to use approximations for the best orbital. Thus for systems containing many electrons, an approximated way of revolve the Hartree-Fock equations consists of expand the orbital Hartree-Fock $\varphi_i(x)$ on a linear combination of K basis functions χ_μ , in accordance proposed by Roothaan [43], and this method denominated of Hartree-Fock-Roothaan (HFR) or also Molecular-Orbital Linear Combination Atomic Orbitals (LCAO-MO). Thus the orbital can be expanded in a linear combination as follows.

$$\varphi_i = \sum_{\mu}^K c_{\mu i} \chi_{\mu} \quad (23)$$

where the $c_{\mu i}$ are the coefficient of expansion, which will be treated as variational parameters and basis functions χ_{μ} are Slater type atomic orbitals or Gaussian type. To accurately represent the orbitals, functions should form a complete set. However, this requires an infinite number of such functions. What you should use is actually a finite number of basis functions. Orbitals should obey the orthonormality condition. Otherwise you can make a linear transformation thereby making them orthonormal

$$\sum_{\mu\nu} c_{\mu i}^* c_{\mu j} S_{\mu\nu} = \delta_{ij} \quad (24)$$

where $S_{\mu\nu}$ is the overlap integral

$$S_{\mu\nu} = \langle \chi_{\mu}(i) | \chi_{\nu}(i) \rangle \quad (25)$$

For the molecules $\varphi_i(x)$ are molecular orbitals and χ_{μ} are the atomic orbitals. Thus there is a significant improvement in computational calculations when the orbital functions are expanded in terms of a finite set of basis functions. The integro-differential equations are then transformed into algebraic equations for the expansion coefficients [41]. The total electronic energy is obtained when Equation (23) is replaced in (21).

$$E = 2 \sum_i^{ocup} \sum_{\mu\nu} c_{\mu i}^* c_{\nu i} h_{\mu\nu} + \sum_{ij}^{ocup} \sum_{\mu\nu\lambda\sigma} c_{\mu i}^* c_{\lambda j}^* c_{\nu i} c_{\sigma j} [2\langle \mu\lambda | \nu\sigma \rangle - \langle \mu\nu | \lambda\sigma \rangle] \quad (26)$$

where

$$h_{\mu\nu} = \langle \chi_{\mu}(1) | \hat{h}(1) | \chi_{\nu}(1) \rangle \quad (27)$$

with $\hat{h}(1)$ given by Equation (8) by replacing i by 1.

The integrals $\langle \mu\lambda | \nu\sigma \rangle$ and $\langle \mu\nu | \lambda\sigma \rangle$ are integral interaction two electrons are represented by

$$\langle \mu\lambda | \nu\sigma \rangle = \iint \chi_{\mu}^*(1) \chi_{\lambda}^*(2) \left| \frac{1}{r_{12}} \right| \chi_{\nu}(1) \chi_{\sigma}(2) d\tau_1 d\tau_2 \quad (28)$$

$$\langle \mu\nu | \lambda\sigma \rangle = \iint \chi_{\mu}^*(1) \chi_{\nu}^*(2) \left| \frac{1}{r_{12}} \right| \chi_{\lambda}(1) \chi_{\sigma}(2) d\tau_1 d\tau_2 \quad (29)$$

The best coefficients $c_{\mu i}$ are determined by varying the electron energy given by Equation (26) with respect to them, obeying the condition of orthonormality. Thus we arrive at the equations of Hartree-Fock-Roothaan

$$\sum_{\nu} c_{\nu i} \left\{ h_{\mu\nu} + \sum_j^{ocup} \sum_{\nu\sigma} c_{\nu j}^* c_{\sigma j} [2\langle \mu\lambda | \nu\sigma \rangle - \langle \mu\sigma | \lambda\sigma \rangle] \right\} = \sum_j \sum_{\nu} \varepsilon_{ij} c_{\nu j} S_{\mu\nu} \quad (30)$$

where the operator Fock defined as

$$F_{\mu\nu} = h_{\mu\nu} + \sum_j^{ocup} \sum_{\nu\sigma} c_{\nu j}^* c_{\sigma j} [2\langle \mu\lambda | \nu\sigma \rangle - \langle \mu\nu | \lambda\sigma \rangle] \quad (31)$$

We can replace the Equation (31) in (30) and thus obtain the equations of Hartree-Fock-Roothaan the following notation

$$\sum_{\nu} c_{\nu i} F_{\mu\nu} - \sum_j \sum_{\nu} \varepsilon_{ij} c_{\nu j} S_{\mu\nu} = 0, \mu = 1, 2, \dots, k \quad (32)$$

Applying a unitary transform on Equation (32) to diagonalize the matrix ε , has been

$$\sum_{\nu} (F_{\mu\nu} - \varepsilon_i S_{\mu\nu}) c_{\nu i} = 0, \mu = 1, 2, \dots, k \quad (33)$$

Equation (33) can be written in matrix form in this way has been

$$FC = SCE \quad (34)$$

The equations of Hartree-Fock-Roothaan are resolved in the same way that the equations of Hartree-Fock, *i.e.*, by an interactive process (SCF-MO-LCAO). Those obtained at the end of the procedure are the eigenvalues of HFR system.

3.2. Hartree-Fock Limit

In Hartree-Fock method, is used a electronic wave function composed only by a Slater determinant, thereby providing only approximate description of exact wave function because this can not be described by a single

Slater determinant. The exact solution of the Schrödinger equation would not be obtained in case the spatial orbitals $\varphi_i(x)$ be expanded into a linear combination of basis functions χ_μ , although the larger and more complete is the set of those functions greater the degree of flexibility in expansion for spin-orbital and lowest expected value for energy. Larger bases sets decrease the HF energy up to a certain limit. This limit is the lowest that can be obtained from a wave function of a determinant and it is called of Hartree-Fock limit. Even so, as a variational method, this limit energy E_{HF} will still be above the exact nonrelativistic energy $E_{ex}(NR)$ due to the energy of electron correlation E_{corr}

$$E_{ex}(NR) = E_{HF} + E_{corr} \quad (35)$$

The SCF method is a valid approach, but generates errors in energy because it describes of approximate mode the interactions between the electrons. Should be also consider the instantaneous interactions between the electrons. The moves of the electrons are correlated with each other, *i.e.*, there is a correlation in the different positions of the electron and must be taken into account.

The exact value of the energy E including electron correlation and relativistic effects E_{rel} is given by:

$$E = E_{HF} + E_{corr} + E_{rel} \quad (36)$$

4. Application of Basis Sets Separate Valence Used in the Molecular Modeling

To perform atomic and molecular calculations, determines a basis set formed by mathematical functions. These bases consist of a linear combination of the wave functions of an electron in terms of a finite number of base functions, which contains a set of parameters to be optimized. The denomination of the base sets depends on various types of basis functions, of the number of those employed functions in the expansion of monoelétrônicas functions (orbital) and of characteristics of the parameters to be optimized. Thus it is of fundamental importance the careful choice these functions when you want to get accurate results. The main basis sets separate valence available in computational chemistry programs for performing calculations of molecular properties are presented below:

4.1. Slater Type Functions

The atomic systems containing only one electron, hydrogenic atoms, have as a solution of the Schrödinger non-relativistic equation, a function of type

$$\Psi_{nlm}(r, \theta, \varphi) = R_{nl}(r) Y_{lm}^{(\theta, \varphi)} \quad (37)$$

where n , l and m are the principal quantum numbers, orbital angular momentum and magnetic, $R_{nl}(r)$ is the radial part and $Y_{lm}^{(\theta, \varphi)}$ is the angular part denominated spherical harmonics. The radial part has the form

$$R_{nl}(r) = N_{nl} \left(\frac{2Zr}{na} \right)^l \exp\left(-\frac{Zr}{na}\right) \mathcal{L}_{n-1}^{2l+1} \left(\frac{2Zr}{na} \right) \quad (38)$$

where N_{nl} comes to be a normalization factor, Z is nuclear charge, r is radius, a is a constant and $\mathcal{L}_{n-1}^{2l+1} \left(\frac{2Zr}{na} \right)$ represents in a generic form a polynomial called associated Laguerre polynomial. The angular part is equal to

$$Y_{lm}^{(\theta, \varphi)} = N_{lm} P_l^{(m)}(\cos \theta) e^{im\varphi} \quad (39)$$

being N_{lm} a normalization factor and $P_l^{(m)}$ the associated Legendre polynomial.

For more that hydrogenic orbitals are orthogonal they do not form a complete set of continuous functions. They have limited application because many of the required integrals in molecular orbital calculations are somewhat hard to calculated, especially when you have high values of the principal quantum number because of the complexity of the polynomial in r . Then Slater (1930) proposed a form analytical simpler for the radial function, then introducing functions such as Slater (Slater Type Functions-STFs) [42].

$$R_{nl}(r) = \frac{[2\zeta/a_0]^{n+1/2}}{[(2n)!]^{1/2}} r^{n-1} \exp(-\zeta r/a_0) \quad (40)$$

Since the orbital exponent is written as

$$\zeta = \frac{Z - s}{n^*} \quad (41)$$

where s is a constant related to the shielding effect of the electrons of the internal layers of atoms and n^* is an effective principal quantum number. Therefore the general shape of Slater type functions can be written as follows, which is similar to the equation.

$$\Psi_{nlm}(\zeta, r, \theta, \varphi) = [(2n)!]^{1/2} (2\zeta)^{n+1/2} r^{n-1} e^{-\zeta r} Y_{lm}^{(\theta, \varphi)} \quad (42)$$

Slater evaluated through empirical rules to select of parameters s and n^* , thereby aiming at a good approximation for the atomic orbitals best of this type. The values of ζ were determined by the variational method by Clementi (1963 and 1967) for neutral atoms to the ruthenium in their fundamental states, using for this the SCF method [44] [45]. The exponents ζ are positive numbers and adjustable in the methods of calculation. These exponents determine the size of the orbital, thereby large exponents characterize dense orbital and small exponents characterize diffuse orbitals.

The so-called Slater type orbitals (STOs) are formed by the product of the angular part of Equation (39), by the product of the radial part of Equation (40). The STOs generate reasonable representations of atomic orbitals. However, as they replace the polynomial in r in the hydrogenic orbitals by a simple power r , do not have the proper number of radial nodes, just not well represent the internal part of an orbital.

One of the limitations is that they are not of mode some orthogonal, as much as it can be corrected using a set of STOs orthogonalized. In addition the use of these orbitals in molecular SCF calculations makes the integrals multicenter, that involve interactions between electrons, becomes somewhat difficult to be solved numerically, thereby increasing the computational time.

4.2. Minimum Base Sets

Minimum base sets or base “single-zeta” is set containing a single function to represent each occupied atomic orbital of different quantum numbers n and l of electron configuration. Thus, this set presents a reduced computational time to make these calculations, so can be used in calculations involving large molecules which are our case. In the calculations where a minimum base set is used are used in general Slater type orbitals. Due to the small size of the base, minimum base sets generate results only form qualitative of the properties. Nevertheless the calculations of the electronic structure were until around 1960 all performed in terms of minimum base. The first semiempirical methods more used were based in a set of minimum base STOs [46] [47].

4.3. Gaussian Type Functions

4.3.1. Contracted Gaussian Functions

The use of STOs in electronic structure calculations generates computational problems due to the emergence of multicenter integrals, although these orbitals well describe the functional behavior of molecular orbitals. In order to simplify the calculation of multicenter integrals, Boys (1950) proposed for calculations involving molecules using Gaussian type functions (GTFs) [48]

$$R_{nl}(r) = \left[\frac{2^{2n+3/2}}{(2n-1)! \sqrt{\pi}} \right]^{1/2} \alpha^{2n+1/4} r^{n-1} \exp(-\alpha r^2) \quad (43)$$

where n and l are the main quantum numbers and orbital angular momentum and α is the orbital exponent which is a variational parameter. The use of GTFs in calculations of electronic interaction integrals presents an enormous advantage, because the product two of these functions entries in different centers is equivalent to a single function centered on a new center [49], thereby there is a reduction of multicenter integrals to integrals calculated in terms of functions centered on the same point.

The functions of the type Gaussian have a poor behavior of atomic orbitals in relation to Slater type functions, the reason is that the GTFs do not have the “cusp” in the region near the nucleus. Thus the GTFs have a behavior different functional that seen for molecular orbitals, so it is necessary to use two (2) to five (5) GTFs to represent adequately each STO. As the number of electronic interaction integrals originated using of a basis set dimension m increases in the order of m^4 , fastest speed and simplify the calculations of the electronic interaction

integrals in terms of GTFs compensates the large number of integrals to be calculated when comparing using STFs.

The solution of the equations of SCF method in the calculation HFR is also a process using an enormous computational time, it is also proportional to the fourth power of the number of base functions. However, the number of cycles of the interactive process SCF method increases with the number of coefficients to be optimized. Thus the use of GTFs contracted originated from linear combinations of Gaussian primitive is generally more adequate [47] [50] [51].

The contracted basis functions can be chosen in order to resemble STOs, atomic orbitals HF or any other set of functions. A type of GTFs contracted is set STO-NG. This nomenclature comes being a set of GTFs contracted that describes a STF through N GTFs primitives used in contraction. Each STO is then approximated as a linear combination of N GTFs, in which the coefficients and exponents contracted are chosen so that the basis function approach is an STF. Thus the computational time decreases, but the results generated by the set of contracted GTFs are not good.

The STOs ns and np are approximated by their respective functions

$$\varphi_{ns}(r) = \sum_{k=1}^m d_{ns,k} g_{1s}(\alpha_{nk}, r), n = 1, 2, 3 \quad (44)$$

$$\varphi_{np}(r) = \sum_{k=1}^m d_{np,k} g_{2p}(\alpha_{nk}, r), n = 1, 2, 3 \quad (45)$$

where g_{1s} and g_{2p} are GTFs, α_{nk} are gaussian exponents and $d_{ns,k}$ and $d_{np,k}$ are coefficients of the expansion, where the last three parameters are determined by the least squares method [51].

4.3.2. Basis Sets Separate Valence

In order to obtain better results, many studies have been made aimed at finding basis sets with this capability. In recent years, besides the basis sets previously described, other sets are being used in electronic structure calculations. Among these, there are basis sets separate valence [42].

Basis sets separate valence are extended sets GTFs contracted, where the most usual are listed below: 4-31G, 3-21G, 6-31G e 6-311G. In such sets are used two functions for the valence orbitals and only one for the orbitals of the internal layer, may each of these functions is a linear combination or not primitive gaussian. The occurrence of this is due the internal layer contribute little to the chemical properties of interest. As the functions of the internal layers not duplicate it generates effects on total energy, however represents little it comes of dipole moments, ionization potentials of the valence, charge density, dissociation energy and other chemical properties. For example, the base 4-31G each atomic orbital of the internal layers is described by only one GTF contracted which is formed by the combination of four linear (4) GTFs primitive. For each atomic orbital of the valence shell there are two basic functions, being that of them is a GTF contracted from the linear combination of three (3) GTFs primitive, describing the internal part of valence orbital, since the other is a single GTF primitive, which describes the outer part of the valence orbital. Considering the first row atoms (Li to F) we have:

$$\varphi_{1s}(r) = \sum_{k=1}^4 d_{k,1s} g_{1s}(\alpha_{k,1s}, r) \quad (46)$$

$$\varphi'_{2s}(r) = \sum_{k=1}^3 d'_{k,2s} g_{1s}(\alpha'_{k,2sp}, r) \quad (47)$$

$$\varphi''_{2s}(r) = g_{1s}(\alpha''_{2sp}, r) \quad (48)$$

$$\varphi'_{2p}(r) = \sum_{k=1}^3 d'_{k,2p} g_{2p}(\alpha'_{k,2sp}, r) \quad (49)$$

$$\varphi''_{2p}(r) = g_{2p}(\alpha''_{2sp}, p) \quad (50)$$

For the hydrogen atom, which has no internal layer, it has been:

$$\varphi'_{1s}(r) = \sum_{k=1}^3 d'_{k,1s} g_{1s}(\alpha'_{k,1s}, r) \quad (51)$$

where the functions which appear with (') are internal functions while appearing with (") are external functions. Similarly to the base structure previously described is the basis 6-31G. Only now each orbital of internal layers is represented by a GTF contracted originated from linear combination of six (6) GTFs primitive. In the case of the base 6-311G, includes a GTF primitive the set 6-31G to represent a new layer of the outer valence.

4.4. Polarization Function

All basis sets discussed so far have a peculiar characteristic, *i.e.*, they comprise functions restricted to be centered in the nuclei. However, there is evidence showing that the atomic orbitals distort or polarize when they form a molecule. For this reason one should take into consideration the possibility of non-uniform displacement of electric charges outside the atomic nucleus, *i.e.* the polarization. Thus it is possible to obtain a better description of the changes as well as also deformations in the atomic orbitals within the molecule.

One way to consider such a polarization is introduced on the basis in question functions wherein the values of l (quantum number of the orbital angular momentum) larger than those of fundamental state of a given atom. In this type of functions gives the name of polarization functions. For the hydrogen atom of the description fundamental state uses only functions s . The functions p , d , ... centered on H , for molecular calculate, are considered as functions of polarization. Generally, with the inclusion of polarization functions in the molecular basis there is a higher possibility of obtaining better results for many of the properties of chemical interest, such as dissociation energy and dipole moments. In real terms, it is noted that it is not satisfactory to include polarization functions of symmetry d and f together with the basis sets s and p small, that is, the polarization functions should only be added when working with basis sets said saturated [52].

Of the bases valence separate with polarization functions most commonly used in molecular calculations are the STO-3G*, 3-1G*, 6-31G*, 6-31G**, 6-311G* and 6-311G** [53]. The base 6-31G* and 6-31G** are formed by the inclusion of polarization functions to the base 6-31G, being that the base 6-31G* is constructed by adding a set of five (5) polarization functions of type GTFs of symmetry d to the base 6-31G, for each different atom of hydrogen and helium, already the basis 6-31G** is constructed by adding a set of three (3) polarization functions of type GTFs of symmetry p to the set 6-31G*, for each hydrogen atom.

4.5. Diffuse Functions

The use of valence basis set with polarization function in calculations involving anions does not generate good results, because the electron cloud of the anionic systems possess a tendency to expansion. Thus it is necessary to also add appropriate diffuse functions because they allow a greater occupation of space by the orbital region. The importance of diffuse functions in calculations of transition metals is due to the metal atoms present orbital type d and these have diffuse characteristics. So it becomes necessary adding diffuse functions to basis function associated to configuration of the neutral metal atom in order to obtain a better description of the metal complex. The great importance of diffuse functions is due to the fact that they better describe the farthest molecular orbital of the nuclei [54].

5. Computational Chemistry Software for Calculating Molecular Properties

Several computational chemistry softwares make available calculations of molecular properties through semi-empirical and Hatree-Fock methods. We can cite: MOPAC2002 [55], VAMP [56], MNDO97 [57], Gaussian98 [58], Spartan [59], GAMESS [60], Chem3D [61], CAche [55], HyperChem [62], ZINDO [63], GEOMOS [64]. Most of the great advances obtained in the research of semi-empirical methods were essentially incorporated on a series of programs entitled MOPAC [62]. MOPAC should be understood as an abbreviation of Molecular Orbital PACKage. Initiated in 1983, on group of Prof. M. Dewar, with the participation of Prof. J. J., P. Stewart, the MOPAC project currently has 30 years, and the main landmarks of this story are:

- 1) 1983: MOPAC → Geometry optimization, vibrational analysis and transition states.
- 2) 1984: MOPAC 2 → Solution of errors.
- 3) 1985: MOPAC 3 → AM1 method, calculation of polymers and CI (*Configuration interaction*).

- 4) 1987: MOPAC 4 → Solution of errors.
- 5) 1988: MOPAC 5 → PM3 method and way of reaction.
- 6) 1989: MOPAC 6 → Repair errors, calculation of electrostatic potential and extent of PM3 for 28 elements.
- 7) 1993: MOPAC 7/93 → Model solvent, nonlinear optical properties and analysis of symmetry.
- 8) 1997: MOPAC 97 → MNDO-d method
- 9) 1999: MOPAC2000 → Dynamic memory allocation, large molecules, solids, transition metals, improving the search algorithm of the transition state.
- 10) 2001: MOPAC2002 → Improvement of the package solvate, all atoms of the main group, PM5 method and parameterization for more transition metals.
- 11) 2004: MOPAC2005 → Calculation of complex Eu(III) [65] and Th(IV) [66] by Sparkle Model.
- 12) 2006: MOPAC 7.1 → A translation of MOPAC 6 for Fortran 90 is available on the Internet, having incorporated the Sparkle/AM1 model for calculating all trivalent lanthanide ions [67];
- 13) 2007: MOPAC2007 → An application of MOPAC7.1 incorporating new methods of RM1, PM6, Sparkle/AM1 and Sparkle/PM3. This program is available free of charge to the academy and for applications nonprofit [67].

Several programs two-dimensional of design of molecules are available and easy to use, as ChemWindow, Isis Draw, ChemDraw [68] and Chem3D [69]. They allow the preparation of figures and diagrams with desired quality and accuracy and facilitate the documentation and scientific communication. The software ChemSketch 12.00 [23] is an advanced design that provides chemical molecular properties, optimization and 3D visualization, ability to name the molecules, as IUPAC, and still has a large database of chemical structures and laboratory materials. The software automatically calculates the valence of each atom and restricts the construction of the molecule based on the octet rule, unless instructed to do this restriction. Then is possible to request the construction of 3D spatial form of the species studied, which triggers another window where the academic can rotate tri-dimensionally the species studied, in addition to observing these species in different visualizations with possibility to visualize bonds and spatial arrangement of species prominently in each of these representations. The design and visualization of 3D drugs, with steric factors relevant to biological activity, are important for analysis of the size, volume and shape of the molecules [70]. The Molekel is a free software multiplatform molecular visualization. It was originally developed at the University of Geneva by Flükiger in the 1990s for Silicon Graphics computers. In 1998, Stefan Portmann took responsibility and released version 3.0. The version 4.0 was almost one version of the platform independent. Other developments lead version 4.3, before Stefan Portman moved and stopped developing the codes. In 2006, the Swiss National Supercomputing Centre (CSCS) restarted the project and version 5.0 was released on December 21 of the same year [24].

To obtain the conformational analysis and energy minimization, we can cite: the program Chem3D [69] widely used in studies with this objective [70]. However, other programs like Molecular Modeling Pro [71], ChemSite (ChemSW) [72], Alchemy, Sybyl, ChemX, cache and WebLab Viewer are also available.

In the area of molecular modeling, graphics construction and projects of drugs, the program Hyperchem [73] for being a tool specializing in 3D structures of interest to the medical, pharmaceutical and organic chemistry. The program lets you design complicated molecules. This software is also an alternative in the field of spectroscopy, which besides the ability to simulate a priori by the NMR spectra quantum methods, contains a database of approximately 10^4 molecules applicable to macromolecules as well as small molecules. The software also includes animations, and quantum chemical calculations and molecular mechanics.

The choice of method for energy minimization depends on factors related to the size of the molecule, parameters of availability and stored data and computational resources. Molecular models generated by the computer are the result of mathematical equations that estimate the positions and properties of the electrons and nuclei, the calculations exploit experimentally, the characteristics of a structure, providing a new perspective on the molecule [74].

In quantum chemistry the softwares more used are Gaussian and GaussView that uses the laws of quantum mechanics to predict the energies, structures and properties and vibrational frequency of molecular systems [75]. The GaussView 5.0 is a program that can work on Windows and responsible for building the structures under study, by viewing these as well as for generating the input of the species under study for the program calculations–Gaussian 03W. This includes an advanced molecular modeler, which can be used for construction and molecular dimensions of the three test [75].

The Gaussian 03W is a program that can work on Windows and Linux that performs computations used in the

study of reaction mechanisms, equilibrium geometries of neutral molecules, radicals and ions, and the determination of physicochemical parameters. Appreciates structure, reactivity, thermodynamic properties, energy barriers (transition states), conformational analysis, employing the optimization of molecules and theoretical calculations of vibrational spectra. From the optimization is obtained the most appropriate structure to the molecule, whereas the lengths and bond angles and power stabilization calculated by $E(RB = HF - LYP)$ ua [75].

6. Application of Hartree-Fock Method Using Computational Tools

Obtaining molecular properties depends of method and basis sets, and represents a means of chemical information contained in the molecular structure of the compound studied. The structure-activity relationships (SAR) represent a core aspect of medicinal chemistry. The fact that a small change in structure leads to a small change in biological activity and allows chemists to rationalize substitutions at specific positions, giving them the freedom to modify a molecule to improve various properties such as lipophilicity, bioavailability, and so on without sacrificing potency (to a large extent). From the modelers' perspective, the principle of similar structures having similar activities [76] is a cornerstone of quantitative structure-activity relationship (QSAR) modeling [77] [78]. This information is transformed and encoded for lots of problems chemical, pharmacological and toxicological studies on the relationship between structure-activity, quantitative structure-activity and structure-property (SAR, QSAR and QSPR) [79]-[81].

Ferreira *et al.* [82] studied artemisinin and 18 derivatives with antimalarial activity against W-2 strains of *Plasmodium falciparum* through quantum chemistry and multivariate analysis. The geometry optimization of structures was realized using the Hartree-Fock method and the 3-21G** basis set. Maps of molecular electrostatic potential and docking molecular were used to investigate the interaction between the ligands and the receptor (Heme).

Santos *et al.* [83] performed studies using Hartree-Fock method and the 6-31G** basis set were employed to calculate the molecular properties of artemisinin and 20 derivatives with antimalarial activity. Maps of molecular electrostatic potential (MEPs) and molecular docking were used to investigate the interaction between ligands and the receptor (heme). Principal component analysis and hierarchical cluster analysis were employed to select the most important descriptors related to activity. The correlation between biological activity and molecular properties was obtained using the partial least squares and principal component regression methods. The regression PLS and PCR models built in this study were also used to predict the antimalarial activity of 30 new artemisinin compounds with unknown activity. The models obtained showed not only statistical significance but also predictive ability. The significant molecular descriptors related to the compounds with antimalarial activity were the hydration energy (HE), the charge on the O11 oxygen atom (QO11), the torsion angle O1-O2-Fe-N2 (D2) and the maximum rate of R/Sanderson Electronegativity (R_{Te}⁺). These variables led to a physical and structural explanation of the molecular properties that should be selected for when designing new ligands to be used as antimalarial agents.

Case Study on Aspirin

Aspirin introduced in 1899 was one of the first drugs developed and is still one of the most widely used. Estimated 20 billion aspirin tablets are consumed each year in the United States. Originally planned to ease the pain and relieve sore muscles, proved to be a highly complex drug with the power and unexpected limitations. It turned out that it reduces the incidence of heart attacks and is effective in reducing the incidence of Alzheimer's disease and cancer of the digestive tract. At the same time, however, aspirin attacking the stomach lining, causing bleeding or even ulcers, and usually cause intestinal problems [84]. One of the forms of action of aspirin is blocking an enzyme (a type of protein) called COX-2, which promotes inflammation, pain and fever. Unfortunately, it also interferes with COX-1 a correlate enzyme that produces essential hormones to the health of stomach and kidney. An analgesic and anti-inflammatory agent is efficient in COX-1. **Figure 2** is shown the structure of aspirin which acts transferring part of its molecule known as acetyl group, for COX-2, disabling it. This drug-receptor interaction is irreversible nature due to the formation of a covalent bond resulting from the nucleophilic attack of the hydroxyl group of the amino acid serina₅₃₀ to the electrophilic acetyl grouping present in aspirin [85].

A substitute for aspirin (new drug) has to keep this aspect of the molecule, and the replacement should the general format and size of the molecule in such a way it clicks into place in molecular target the same manner

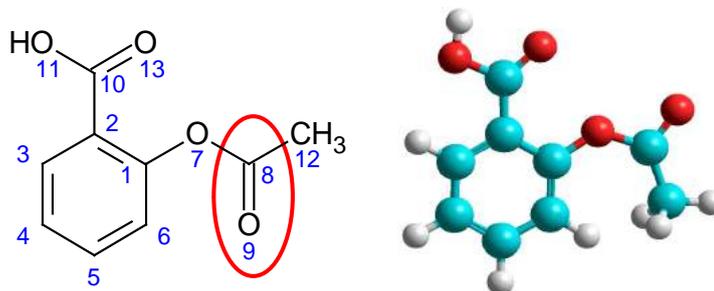


Figure 2. Aspirin (structure) and region essential for the expression of biological (pharmacophore) was visualized using ChemSketch 12.00 program [23] and HyperChem Release 6.02 [73].

that aspirin.

The use of maps of electrostatic potential (MEP) is a computational tool that aids in the process of recognition of one molecule by another, as interactions of types receptor drugs and enzyme substrates, because being by their potential that chemical species interact with other in biological recognition process. The electronic parameters are one of the main factors governing drug-receptor interaction, in this sense; the MEP can be considered an alternative approach in order to understand the electrostatic contribution of this drug and its new derivatives for biological activity [86]. In the construction of the MEP are necessary three steps: construction of the surface density of the molecule, construction of the electrostatic surface potential and applying colors the surface obtained to designate potential values. One of the frequent topics of theoretical chemistry is research to improve methods to elucidate the behavior of molecules and other reactive chemical species. Among the numerous existing reactivity indices the molecular electrostatic potential $V(r)$ that is generated around a molecule by its nuclei and electrons, is known for being a real physical property can be determined experimentally by diffraction methods, as well as computationally [87].

In **Figure 3** are shown the MEPs of aspirin with Hatree-Fock method in different basis sets (HF/3-21G*, HF/3-21G**, HF/6-31G, HF/6-31G*, HF/6-31G** and HF/6-311G). In this figure we can observe that in the HF/6-31G** method, showed the lowest positive electrostatic potential (region blue color) of 0.06715 au (atomic unit). However, HF/6-311G method, showed higher positive electrostatic potential 0.07607 au. The variation between them was ± 0.00892 au (HF/6-311G and HF/6-31G**); Also in this figure we observed that the HF/6-311G method, showed the lowest negative electrostatic potential (region red color) equal to -0.10235 au. However, HF/3-21G** method showed higher negative electrostatic potential of -0.09071 au. The variation between them was ± 0.01164 au (HF/6-311G and HF/3-21G**). Thus, the presence of a negative potential surface in carbonyl oxygen atom of the acetyl group is nucleophilic (affinity for positive nuclei), while the carbonyl carbon is electrophilic (electron affinity) in accordance with the literature [85]. Thus, we have shown that the use of different basis sets present values of different electrostatic potentials for the same case study. Therefore, obtaining molecular properties will depend on the method and basis set to reproduce experimental data with greater accuracy. In this case, the MEP of aspirin in different basis set was used to evaluate the key features of aspirin from qualitative comparisons in the region of the acetyl group, which according to Bernardinelli *et al.* [88] the geometric form of the electrostatic potential is similar for all active compounds. However, new derivatives of aspirin must have some structural similarity in terms of their electrostatic potentials that allow one to be recognized by the other, with similar biological activities [87] [89] [90].

Structure-activity relationship (SAR) indicates molecular structure modifications that increase the drug effectiveness. In general, reports show that these modifications are made throughout small changes in the leading compound structure, followed by trials in laboratory to quantify the variations in the biological activity due to changes in the molecular structure [91].

The quantum-chemical descriptors widely used in SAR, QSAR and QSPR studies are related to the energy of the frontier orbitals (HOMO and LUMO). The reason for this is related to the fact that these properties provide information about the character electron donor and/or electron-acceptor and a compound thus forming a charge transfer complex (CTC) [92]. The energy of Highest Occupied Molecular Orbital Energy (HOMO) and Lowest Unoccupied Molecular Orbital Energy (LUMO) are quantum-chemical descriptors, which play an important role in chemical reactions and the formation of many complex charge transfer [93]. In **Figure 4** is shown the fron-

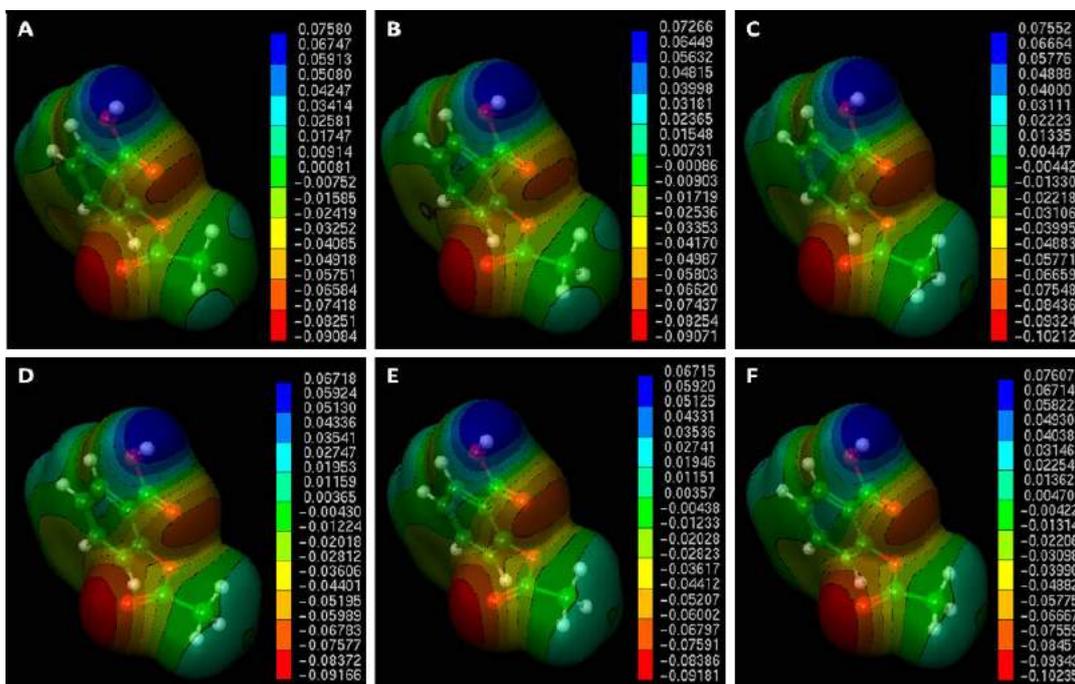


Figure 3. Molecular Electrostatic Potential Map of aspirin in different basis sets (A – HF/3-21 G*, B – HF/3-21G**, C – HF/6-31G, D – HF 6-31G*, E – HF 6-31G**, F – HF 6-311G). The MEPs were visualized and calculated by the Molekel program [24].

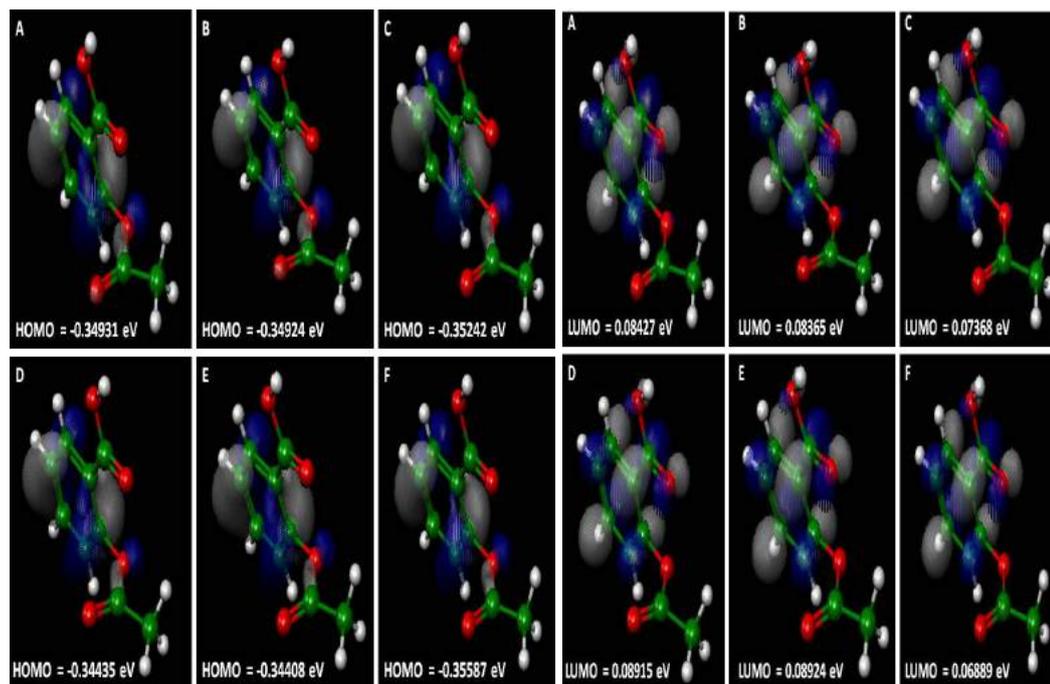


Figure 4. Orbital's energy HOMO and LUMO of aspirin in different basis sets (A – HF/3-21 G*, B – HF/3-21G**, C – HF/6-31G, D – HF 6-31G*, E – HF 6-31G**, F – HF 6-311G). The HOMO and LUMO were visualized and calculated by the Molekel program [24].

tiers orbital (HOMO and LUMO) with the respective values of energy (eV), using Hatree-Fock method at different basis sets (HF/3-21G*, HF/3-21G**, HF/6-31G, HF/6-31G*, HF/6-31G** and HF/6-311G) for aspirin,

and this figure we note the delimited region for the HOMO orbital which measures the electron-donor character of aspirin, and the LUMO which measures the electron-acceptor character. From these definitions, two important features can be observed: the higher the energy of the HOMO greater electron-donating ability, which can be observed in HF/6-31G** with a value of HOMO = -0.34408 eV, and the lower the energy of the LUMO will be lowest resistance to accept electrons that can be noticed in HF/6-311G with a value of LUMO = 0.06889 eV. In this figure the HOMO is located in the region around the benzene ring and the acetyl group, when the substituents are made in the aromatic ring or acetyl group, depending on the substituent will have a high electron density, such as the carbonyls, amines or amides, the more pronounced HOMO region is strongly influenced to perform electronic stereo secondary effects, which may compromise the pharmacological activity of the compound. The energy of HOMO is directly related to the ionization potential of the compound and characterizes the ability of the molecule to perform nucleophilic attacks. Also in **Figure 4**, the LUMO is located in close region to the benzene ring and of the carboxylic group, so the energy of LUMO is directly related to the electron affinity, characterized by the susceptibility of the compound in relation to attacks by nucleophiles [94].

The energies of HOMO and LUMO have been used for decades as indices of chemical reactivity and are commonly correlated with other indices, such as electron affinity and ionization potential [95]-[99]. The difference between the energies of the HOMO-LUMO orbitals (*gap*) is an important indicator of molecular stability. Molecules with low *gap* value are generally reactive, while molecules with higher *gap* value indicate high stability of the molecule, in the sense of low reactivity in chemical reactions [100]. In **Table 1** are shown some molecular properties obtained with Hartree-Fock method in different basis sets (HF/3-21G*, HF/3-21G**, HF/6-31G, HF/6-31G*, HF/6-31G** e HF/6-311G) and the correlation matrix of Pearson [101]-[103], considering how an independent variant the total energy (E_{total}) of aspirin, and other properties as the dependent variable, this treatment was performed using Stastica 6.2 program [104]. In this table it is observed that the correlation between the molecular properties of aspirin is less than or equal to 0.96890 (LUMO), while the correlation between the molecular properties and the total energy is less than or equal to 0.85448 (QC8), which represents the charge on the carbon atom 8, see **Figure 2**. Among these properties obtained those that had greater relevance to building a model QSPR in function of the total energy were the volume (-0.78677), hydration energy (-0.45775), GAP (-0.43419) and QC_8 (0.85448). Therefore, we can represent the QSPR model according to the values of the statistical parameters in Equation (52) below. The statistical quality [105] of the regression equations was gauged by parameters like correlation coefficient (r) or squared correlation coefficient (r^2), explained variance (R_A^2 , *i.e.*, adjusted R^2), standard error of estimate (SEE), and variance ratio (F) [106]-[108]. The better regression models were selected on the basis of the higher r , F value (a statistic of assessing the overall significance) and the lower SEE.

$$E_{\text{total}} = 286.1621 - 1.1850(V) + 8.7851(HE) + 615.3683(GAP) + 39.8875(QC_8) \quad (52)$$

This type of treatment shows that QSPR studies are important to obtain of molecular properties that take into account different aspects of chemical information, this information can be through experiments or theoretical calculations simple counting, consider the entire molecule, fragments or functional groups, knowledge of the 3D structure of the molecule or molecular graphics his or her simply formula, information defined by scalar values, vectors or scalar fields [109].

In the discovery of a superaspirin Jacob *et al.* (2012) [110] synthesized by conjugation aspirin-glucose (**Figure 5**) in order to study the solubility in water and anticancer activity as compared to aspirin and evaluated that aspirin-glucose was seven times more soluble in water than aspirin, and about 8 to 9 times more active in inhibiting cell growth than aspirin in its anticancer activity in cell culture breast, pancreatic, and cell lines prostate, while the activity was similar in a line benign non-cancer cells. According to the computer calculations performed to aspirin and aspirin-glucose in HF/6-31G** verified that the anticancer activity of aspirin-glucose was increased by the fact of presenting high value $GAP = -0.43232$ eV, having greater molecular stability and low chemical reactivity, being that the value of GAP to aspirin in HF/6-31G** was -0.43332 eV, having a variation of $GAP = \pm 10^{-3}$ eV (between aspirin and aspirin-glucose).

In **Figure 5** is shown the map of electrostatic potential for aspirin-glucose, and observed the region of negative electrostatic potential, characterized by red color, which is the region essential for the expression of biological activity (pharmacophore). Therefore, these results are directly related with the conclusion that Jacob *et al.* indicated that the hydrolysis of the aspirin-glucose in human serum is at a relatively slower rate compared to aspirin where there has been significant anticancer activity at the doses studied under the experimental conditions.

Table 1. Molecular properties obtained at different basis sets of aspirin and Pearson correlation matrix.

Method and basis set	Total energy (u.a)	Surface Area (Grid) (\AA^2)	Volume (\AA^3)	Hydration Energy (kcal/mol)	HOMO (eV)	LUMO (eV)	GAP (eV)	QC ₈ (C)	QO ₉ (C)	C8-O9 (\AA)
HF/3-21G*	-641.3663	345.70	527.07	-8.00	-0.34931	0.08427	-0.43358	0.856	-0.584	1.19329
HF/3-21G**	-641.4183	345.02	526.10	-8.04	-0.34924	0.08365	-0.43289	0.825	-0.585	1.19312
HF/6-31G	-644.6781	347.46	530.81	-8.04	-0.35242	0.07368	-0.42610	0.778	-0.541	1.20473
HF/6-31 G*	-644.9649	345.45	528.40	-7.89	-0.34435	0.08915	-0.43350	0.783	-0.544	1.18114
HF/6-31G**	-644.9837	345.33	528.31	-7.89	-0.34408	0.08924	-0.43332	0.772	-0.545	1.18127
HF/6-311	-644.8264	345.99	529.85	-8.03	-0.35587	0.06889	-0.42476	0.723	-0.490	1.20201
Total energy	1.00000									
Surface Area (Grid)	-0.36056	1.00000								
Volume	-0.78677	0.83299	1.00000							
Hydration Energy	-0.45775	-0.42636	-0.09622	1.00000						
HOMO	-0.05826	-0.53886	-0.44601	0.86149	1.00000					
LUMO	0.18092	-0.65181	-0.64085	0.75742	0.96890	1.00000				
GAP	-0.43419	0.72982	0.81092	-0.58584	-0.86310	-0.96123	1.00000			
QC ₈	0.85448	-0.27889	-0.73507	-0.09622	0.33329	0.52087	-0.69165	1.00000		
QO ₉	0.34835	0.28008	0.03220	-0.61938	-0.54970	-0.45390	0.31382	0.18617	1.00000	
C ₈ -O ₉	0.10478	0.71376	0.48478	-0.91231	-0.94599	-0.92477	0.83323	-0.18682	0.55693	1.00000

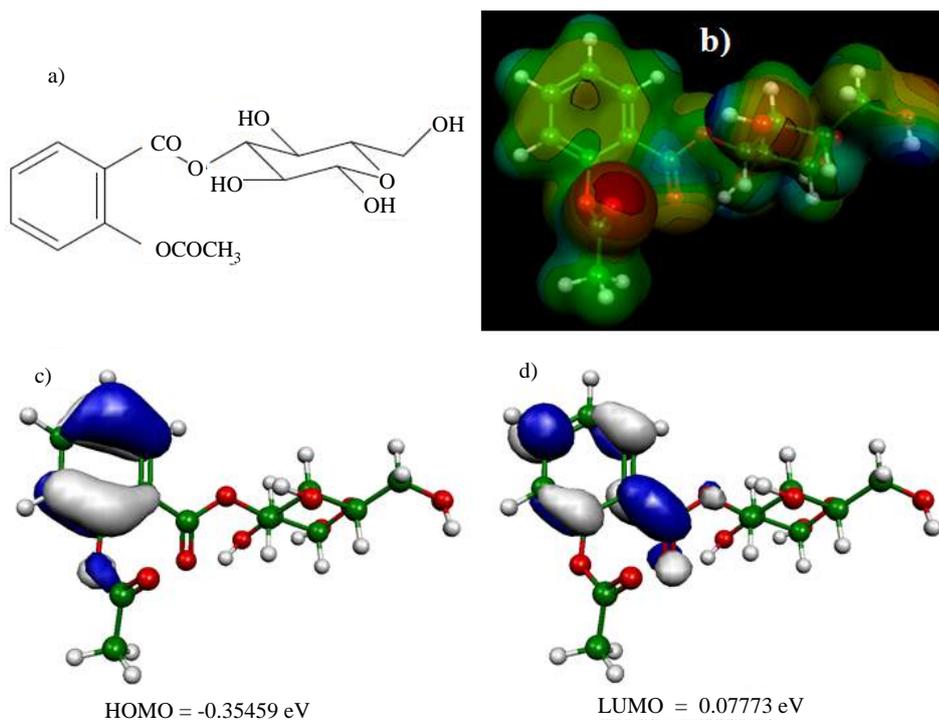


Figure 5. (a) The glucose-aspirin structure was visualized using ChemSketch 12.00 program [23]. (b) Map of molecular electrostatic potential (MEP). ((c) and (d)) Orbital's energy HOMO and LUMO. The MEP, HOMO and LUMO were visualized and calculated using Hartree-Fock (HF) method and HF/6-31G** basis set by the Molekel program [24].

The anticancer activity *in vitro* was much stronger for aspirin-glucose compared to aspirin in cancer cell lines. Therefore, further studies are needed among them we can mention the use of maps of electrostatic potential as an indicator of site of chemical reactivity to confirm this finding in a system *in vivo*. The high solubility of the conjugated glucose aids the development of a form of injection of aspirin.

This new derivative can be considered in the future a “superaspirin” that must pass safety tests for long periods before being placed on pharmacy shelves, but the time to replace aspirin and other non-steroidal anti-inflammatory drugs.

7. Final Considerations

The computational chemistry is considered as one of the greatest intellectual realizations of the twentieth century, and has been the conceptual basis that allows the understanding of chemistry of a way much deeper than that existing before age 20, epoch wherein were launched the bases of quantum theory. The impact of this theory in chemistry can be verified for its practical implications in various fields such as spectroscopy, electron microscopy, molecular modeling, among others.

The molecular modeling provides important information for the process of drug discovery, because it facilitates obtaining of specific molecular properties of a molecule that can influence the interaction with the receptor and biological activity.

The Hartree-Fock method provides a quantitative prediction of high quality for a wide variety of chemical and biological systems, but the calculations are time consuming and high computational cost. A resource commonly employed is optimizing geometry with a basis set simpler, and then perform calculations “Single Point” with a basis set most complete allowing the determination of energy and other molecular properties of a system, using a base of more sophisticated calculation.

The application of the Hartree-Fock method depends on the basis set and of system to be studied to obtain different molecular properties related to the biological activity of bioactive molecules. The computer calculations should describe the necessary characteristics related to the experimental data of the molecule or set of molecules under study, for that molecular modeling can be represented of effective and/or efficient form. However, these descriptors are not completely universal, because they are dependent on the structures and systems studied. Even though it is based on a minimum of energy, the calculated descriptors have values very close to their respective empirical values and indicate trends electronic systems under study. The aspects highlighted in this work are evident, therefore, the quantum-chemical descriptors have a wide range of applications in SAR, QSAR and QSPR studies, as well as in many areas of integration of fundamental knowledge of Organic Chemistry, Biochemistry, Molecular Biology, Pharmacology and Pharmaceutical Chemistry.

Acknowledgements

We gratefully acknowledge the financial support of the Brazilian Agency National Council of Scientific and Technological Development (CNPq Proc. 306676/2010-9). The authors would like to thank the Postgraduate Program in Biotechnology and Biodiversity—Network BIONORTE and to the Laboratory of Modeling and Computational Chemistry, Federal University of Amapá for the computational support.

References

- [1] Luna, J.S., Santos, A.F., Lima, M.R.F., Omena, M.C., Mendonça, F.A.C., Bieber, L.W. and Sant’Ana, A.E.G. (2005) A Study of the Larvicidal and Molluscicidal Activities of Some Medicinal Plants from Northeast Brazil. *Journal of Ethnopharmacology*, **97**, 199-206. <http://dx.doi.org/10.1016/j.jep.2004.10.004>
- [2] Amarante, C.B., Müller, A.H., Póvoa, M.M. and Dolabela, M.F. (2011) Estudo Fitoquímico Biomonitorado Pelos Ensaaios de Toxicidade Frente à Artemia Salina e de Atividade Antiplasmódica do Caule de Aninga (*Montrichardia linifera*). *Acta Amazônica*, **41**, 431-434. <http://dx.doi.org/10.1590/S0044-59672011000300015>
- [3] Robards, K., Prenzler, P.D., Tucker, G., Swatsitang, P. and Glover, W. (2003) Phenolic Compounds and Their Role in Oxidative Processes in Fruits. *Food Chemistry*, **66**, 401-436. [http://dx.doi.org/10.1016/S0308-8146\(99\)00093-X](http://dx.doi.org/10.1016/S0308-8146(99)00093-X)
- [4] Moure, A., Cruz, J., Franco, D., Dominguez, J., Sineiro, J., Dominguez, H., Nnunez, M. and Parajó, J. (2001) Natural Antioxidants from Residual Sources. *Food Chemistry*, **72**, 145-171. [http://dx.doi.org/10.1016/S0308-8146\(00\)00223-5](http://dx.doi.org/10.1016/S0308-8146(00)00223-5)
- [5] Huang, D., Ou, B. and Prior, R.L. (2005) The Chemistry behind Antioxidant Capacity Assays. *Journal of Agricultural*

- and *Food Chemistry*, **53**, 1841-1856. <http://dx.doi.org/10.1021/jf030723c>
- [6] Zibadi, S., Farid, R., Moriguchi, S., Lu, Y., Foo, L., Tehrani, P., Lreich, J. and Watson, R. (2007) Oral Administration of Purple Passion Fruit Peel Extract Attenuates Blood Pressure in Female Spontaneously Hypertensive Rats and Humans. *Nutrition Research*, **27**, 408-416. <http://dx.doi.org/10.1016/j.nutres.2007.05.004>
- [7] Sun, J., Chu, Y.F., Wu, X. and Liu, R.H. (2002) Antioxidant and Antiproliferative Activities of Common Fruits. *Journal of Agricultural and Food Chemistry*, **50**, 7449-7454. <http://dx.doi.org/10.1021/jf0207530>
- [8] Amarante, C.B., Silva, J.C.F., Solano, F.A.R., Nascimento, L.D., Moraes, L.G., Silva, G.F. and Uno, W.S. (2009) Spectrometric Study of the Leaves of Aninga (*Montrichardia linifera*) Collected from the Guama River, Campus of UFPA, Belém-PA. A Contribution to the Chemical Study of the Araceae Family. *Revista Científica da UFPA*, **7**, 1-19.
- [9] D'Oliveira, H.V. (1854) System of Brazilian Vegetable Materia Medica. Casa de Eduardo & Henrique Laemmert, Rio de Janeiro.
- [10] Moreira, N.J. (1862) Dictionary of Brazilian Medicinal Plants. Typographia do Correio Mercantil, Rio de Janeiro.
- [11] Catsro, J.M. (1878) Purgatives Brazil's Indigenous. Faculdade de Medicina do Rio de Janeiro.
- [12] Fenner, R., Betti, A.H., Mentz, L.A. and Rates, S.M.K. (2006) Plantas Utilizadas na Medicina Popular Brasileira com Potencial Atividade Antifúngica. *Revista Brasileira de Ciências Farmacêuticas*, **42**, 369-394. <http://dx.doi.org/10.1590/S1516-93322006000300007>
- [13] Moo-Puc, R., Robledo, D. and Freile-Pelegrin, Y. (2008) Evaluation of Selected Tropical Seaweeds for *in Vitro* Anti-Trichomonal Activity. *Journal of Ethnopharmacology*, **120**, 92-97. <http://dx.doi.org/10.1016/j.jep.2008.07.035>
- [14] Osorio, D.E.J., Montoya, P.G.L. and Arango, A.G.J. (2006) Alkaloidal Natural Products with Antiprotozoal Activity. *Vitae*, **13**, 61-84.
- [15] Vikas, D.K., Vishweshwar, R. and Lakshmi, N.M. (2000) Current Status of Artemisinin and Its Derivatives as Antimalarial Drugs. *Life Science*, **66**, 279-300.
- [16] Amarante, C.B., Silva, J.C.F., Müller, R.C.S. and Müller, A.H. (2011) Avaliação da Composição Mineral Do Chá da Folha Senescente de *Montrichardia linifera* (Arruda) Schott (Araceae) por Espectrometria de Absorção Atômica com Chama (FAAS). *Química Nova*, **34**, 419-423. <http://dx.doi.org/10.1590/S0100-40422011000300010>
- [17] Meshnick, S.R., Jefford, C.W., Posner, G.H., Avery, M.A. and Peters, W. (1996) Second-Generation Antimalarial Endoperoxides. *Parasitology Today*, **12**, 79-82. [http://dx.doi.org/10.1016/0169-4758\(96\)80660-0](http://dx.doi.org/10.1016/0169-4758(96)80660-0)
- [18] Klayman, D.L. (1985) Qinghaosu (Artemisinin): An Antimalarial Drug from China. *Science*, **228**, 1049-1055. <http://dx.doi.org/10.1126/science.3887571>
- [19] Pinheiro, J.C., Ferreira, M.M.C. and Romero, O.A.S. (2001) Antimalarial Activity of Dihydroartemisinin Derivatives against *P. falciparum* Resistant to Mefloquine: A Quantum Chemical and Multivariate Study. *Journal of Molecular Structure: THEOCHEM*, **572**, 35-44. [http://dx.doi.org/10.1016/S0166-1280\(01\)00522-X](http://dx.doi.org/10.1016/S0166-1280(01)00522-X)
- [20] Pinheiro, J.C., Kiralj, R., Ferreira, M.M.C. and Romero, O.A.S. (2003) Artemisinin Derivatives with Antimalarial Activity against *Plasmodium falciparum* Designed with the Aid of Quantum Chemical and Partial Least Squares Methods. *QSAR & Combinatorial Science*, **22**, 830-842. <http://dx.doi.org/10.1002/qsar.200330829>
- [21] Cardoso, F.J.B., Figueiredo, A.F., Lobato, M.S., Miranda, R.M., Almeida, R.C.O. and Pinheiro, J.C. (2008) A Study on Antimalarial Artemisinin Derivatives Using MEP Maps and Multivariate QSAR. *Journal of Molecular Modeling*, **14**, 39-48. <http://dx.doi.org/10.1007/s00894-007-0249-9>
- [22] Atkins, P. and Jones, L. (2002) Chemical Principles: The Quest for Insight. Freeman, New York.
- [23] ACD/Chemsketch Freware, version 12.00. (2010) Advanced Chemistry Development, Inc, Toronto.
- [24] Flukiger, P. (2002) Molekel 4.2. Swiss Center for Scientific Computing, Mano.
- [25] Dyer, J.R. (1965) Applications of Absorption Spectroscopy of Organic Compounds. Prentice-Hall, New Jersey.
- [26] Silverstein, R.M., Webster, F.X. and Kiemle, D. (2005) Spectrometric Identification of Organics Compounds. Jhon Wiley & Sons, New Jersey.
- [27] Bunge, A.V. (1979) Introdução a Química Quântica. Edgard Blücher, São Paulo.
- [28] Souza, A.A. and Farias, R.F. (2007) Elementos de Química Quântica. Átomo, Campinas.
- [29] Sant'Anna, C.M.R. (2009) Métodos de Modelagem Molecular Para Estudo e Planejamento de Compostos Bioativos: Uma Introdução. *Revista Virtual de Química*, **1**, 49-57.
- [30] Barreiro, E.J. and Fraga, C.A.M. (2001) Química Medicinal, as Bases Moleculares da Ação dos Fármacos. Artmed, Porto Alegre.
- [31] Nye, M.J. (2007) Working Tools for Theoretical Chemistry: Polanyi, Eyring, and Debates over the "Semiempirical Method". *Journal of Computational Chemistry*, **28**, 98-108. <http://dx.doi.org/10.1002/jcc.20527>

- [32] James, H.M. and Coolidge, A.S. (1933) The Ground State of the Hydrogen Molecule. *The Journal of Chemical Physics*, **1**, 825. <http://dx.doi.org/10.1063/1.1749252>
- [33] Marcus, R.A. (1964) Generalization of the Activated Complex Theory of Reaction Rates. I. Quantum Mechanical Treatment. *The Journal of Chemical Physics*, **41**, 2614-2623. <http://dx.doi.org/10.1063/1.1726329>
- [34] Hirschfelder, J.O. (1941) Semi-Empirical Calculations of Activation Energies. *The Journal of Chemical Physics*, **9**, 645. <http://dx.doi.org/10.1063/1.1750966>
- [35] Mielke, S.L., Peterson, K.A., Schwenke, D.W., Garrett, B.C., Truhlar, D.G., Michael, J.V., Su, M.C. and Sutherland, J.W. (2003) H+H₂ Thermal Reaction: A Convergence of Theory and Experiment. *Physical Review Letters*, **91**, 63201-63204. <http://dx.doi.org/10.1103/PhysRevLett.91.063201>
- [36] Atkinson, R.C. and Suppes, P. (1968) Program in Computer-Assisted Instruction: Final Report. In: Simonson, M.R. and Thompson, A., Eds., *Educational Computing Foundations*, Prentice Hall, Upper Saddle River.
- [37] Nye, M.J. (2000) *Instruments and Experimentation in the History of Chemistry*. MIT Press, Cambridge.
- [38] Morgon, N.H. and Coutinho, K. (2007) *Métodos de Química Teórica e Modelagem Molecular*. Livraria da Física, São Paulo.
- [39] Hartree, D.R. (1928) The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods. *Mathematical Proceedings of the Cambridge Philosophical Society*, **24**, 89-110. <http://dx.doi.org/10.1017/S0305004100011919>
- [40] Mulliken, R.S. (1932) Electronic Structures of Polyatomic Molecules and Valence. II. General Considerations. *Physical Review*, **41**, 49-71. <http://dx.doi.org/10.1103/PhysRev.41.49>
- [41] Blinder, S.M. (1965) Basic Concepts of Self-Consistent-Field Theory. *American Journal of Physics*, **33**, 431-520. <http://dx.doi.org/10.1119/1.1971665>
- [42] Levine, I.N. (1991) *Quantum Chemistry*. Prentice-Hall, New York.
- [43] Roothaan, C.C.J. (1951) New Developments in Molecular Orbital Theory. *Reviews of Modern Physics*, **23**, 69-89. <http://dx.doi.org/10.1103/RevModPhys.23.69>
- [44] Clementi, E. and Raimondi, D.L. (1963) Atomic Screening Constants from SCF Functions. *The Journal of Chemical Physics*, **38**, 2686-2689. <http://dx.doi.org/10.1063/1.1733573>
- [45] Clementi, E., Raimondi, D.L. and Reinhardt, W.P. (1967) Constantes Triagem Atômicas de SCF Funções. II. Átomos com 37-86 Elétrons. *The Journal of Chemical Physics*, **47**, 1300-1307. <http://dx.doi.org/10.1063/1.1712084>
- [46] Hoffmann, R., (1963) An Extended Huckel Theory. I. Hydrocarbons. *The Journal of Chemical Physics*, **39**, 1397-1412. <http://dx.doi.org/10.1063/1.1734456>
- [47] Pople, J.A. and Beveridge, D.L. (1970) *Aproximate Molecular Orbital Theory*. McGraw-Hill, New York.
- [48] Boys, S.F. (1950) Electronic Wave Functions. I. A General Method of Calculation for the Stationary States of Any Molecular System. *Proceedings of the Royal Society A*, **200**, 542-554. <http://dx.doi.org/10.1098/rspa.1950.0036>
- [49] Shavitt, I. (1963) The Gaussian Function in Calculations of Statistical Mechanics and Quantum Mechanics. In: Alder, B., Fernbach, S. and Rotenberg, M., Eds., *Methods in Computational Physics*, Academic Press, New York, 1-45.
- [50] Hehre, W.J., Stewart, R.F. and Pople, J.A. (1969) Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals. *The Journal of Chemical Physics*, **51**, 2657-2664. <http://dx.doi.org/10.1063/1.1672392>
- [51] Stewart, R.F. (1969) Small Gaussian Expansions of Atomic Orbitals. *The Journal of Chemical Physics*, **50**, 2485-2495. <http://dx.doi.org/10.1063/1.1671406>
- [52] Mulliken, R.S. and Liu, B. (1971) Self-Consistent-Field Wave Functions of P₂ and PO, and the Role of d Functions in Chemical Bonding and of s-p Hybridization in N₂ and P₂. *Journal of the American Chemical Society*, **93**, 6738-6744. <http://dx.doi.org/10.1021/ja00754a004>
- [53] Hehre, W. (1986) *Ab Initio Molecular Orbital Theory*. John Wiley & Sons, New York, 678-682.
- [54] Santos, C.B.R., Lobato, C.C., Vieira, J.B., Brasil, D.S.B., Brito, A.U., Macêdo, W.J.C., Carvalho, J.C.T. and Pinheiro, J.C. (2013) Evaluation of Quantum Chemical Methods and Basis Sets Applied in the Molecular Modeling of Artemisinin. *Computational Molecular Bioscience*, **3**, 66-79. <http://dx.doi.org/10.4236/cmb.2013.33009>
- [55] Stewart, J.J.P. (1999) MOPAC2002. Fujitsu Limited, Tokyo. <http://www.cachesoftware.com/>
- [56] Clark, T. (2002) VAMP, Computer Chemie Centrum. Friedrich-Alexander-Universität Erlangen-Nürnberg, Nögelsbachstra Be 25, D-91052, Erlangen.
- [57] Thiel, W. (1998) MNDO97. Organisch-Chemisches Institut, Universitaet Zuerich, Winterthurerstrasse 190, CH-8057, Zurich.

- [58] Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Zakrzewski, V.G., Montgomery Jr., J.A., Stratmann, R.E., Burant, J.C., Dapprich, S., Millam, J.M., Daniels, A.D., Kudin, K.N., Strain, M.C., Frakas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Pettersson, G.A., Ayala, P.Y., Cui, Q., Morokuma, K., Malick, D.K., Rabuck, A.D., Raghavachari, K., Foresman, J.B., Cioslowski, J., Ortiz, J.V., Baboul, A.G., Stefanov, B.B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Gomperts, R., Martin, R.L., Fox, D.J., Keith, T., Al-Laham, M.A., Peng, C.Y., Nanayakkara, A., Gonzalez, C., Challacombe, M., Gill, P.M.W., Johnson, B., Chen, W., Wong, M.W., Andres, J.L., Gonzalez, C., Head-Gordon, M., Replogle, E.S. and Pople, J.A. (1998) Gaussian 98, Revision A. 7, Gaussian, Inc., Pittsburgh.
- [59] SPARTAN, Wavefunction, Inc., 18401, Von Karman Avenue, Suite 370, Irvine CA 92612 USA.
- [60] Gordon, M., GAMESS, Gilman Hall, Iowa State University (Ames).
- [61] CambridgeSoft Corporation (1999) Chem3D. Cambridge.
- [62] Ostlund, N. (1993) HyperChem. Hypercube Inc., Gainesville
- [63] Zerner, M.C. (1998) ZINDO Manual. QTP, University Of Florida, Gainesville, 32611-38435.
- [64] Ranaldi, D., Hoggan, P.E., Cartier, A., Baka, K., Mornard, G., They, V., Loos, M., Mokrane, A. and Dillet, V. (2003) GEOMOS Manual. Laboratoire de Chimie Théorique, Université de Nancy I, Nancy.
- [65] Rocha, G.B., Freire, R.O., da Costa Jr., N.B., de Sá, G.F. and Simas, A.M. (2004) Sparkle Model for AM1 Calculation of Lanthanide Complexes: Improved Parameters for Europium. *Inorganic Chemistry*, **43**, 2346-2354. <http://dx.doi.org/10.1021/ic034882p>
- [66] Young, D.C. (2001) Computational Chemistry—A Practical Guide for Applying Techniques to Real-World Problems. Cytoconal Pharmaceuticals Inc., John Wiley & Sons, Inc., New York.
- [67] Stewart, J.J.P. (2007) MOPAC2007, Stewart Computational Chemistry EUA. <http://openmopac.net>
- [68] Cambridge Soft Corporation (1998) CS ChemDraw 5.0, Chemical Structure Drawing Standard. Cambridge.
- [69] Cambridge Soft Corporation (1998) CS Chem3D 5.0, Molecular Modeling and Analysis. Cambridge.
- [70] Carvalho, I., Pupo, M.T., Borges, A.D.L. and Bernardes, L.S.C. (2003) Introdução A Modelagem Molecular de Fármacos no Curso Experimental de Química Farmacêutica. *Quimica Nova*, **26**, 428-438. <http://dx.doi.org/10.1590/S0100-40422003000300023>
- [71] ChemSW[®], Inc., (2001) Molecular Modeling ProTM4.0, Computational Chemistry Program. ChemSW[®], Inc., Fairfield.
- [72] ChemSW[®] Inc., (2001) ChemSite: Interactive 3D Molecular Modeling. ChemSW[®], Inc., Fairfield.
- [73] Chemplus, Modular Extensions for HyperChem Release 6.02, Molecular Modeling for Windows. (2000) HyperClub, Inc, Gainesville.
- [74] Patrick, G.L. (2001) An Introduction to Medicinal Chemistry. Oxford University Press, Oxford.
- [75] Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., Montgomery Jr., J.A., Vreven, T., Kudin, K.N., Burant, J.C., Millam, J.M., Iyengar, S.S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G.A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J.E., Hratchian, H.P., Cross, J.B., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R.E., Yazyev, O., Austin, A.J., Cammi, R., Pomelli, C., Ochterski, J.W., Ayala, P.Y., Morokuma, K., Voth, G.A., Salvador, P., Dannenberg, J.J., Zakrzewski, V.G., Dapprich, S., Daniels, A.D., Strain, M.C., Farkas, O., Malick, D.K., Rabuck, A.D., Raghavachari, K., Foresman, J.B., Ortiz, J.V., Cui, Q., Baboul, A.G., Clifford, S., Cioslowski, J., Stefanov, B.B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R.L., Fox, D.J., Keith, T., Al-Laham, M.A., Peng, C.Y., Nanayakkara, A., Challacombe, M., Gill, P.M.W., Johnson, B., Chen, W., Wong, M.W., Gonzalez, C. and Pople, J.A. (2004) Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford.
- [76] Johnson, M. and Maggiora, G. (1990) Concepts and Applications of Molecular Similarity. John Wiley & Sons, New York.
- [77] Leach, A. (2001) Molecular Modelling: Principles and Applications. Prentice Hall, Englewood Cliffs.
- [78] Martin, Y. (2010) Quantitative Drug Design: A Critical Introduction. 2nd Edition, CRC Press, Boca Raton. <http://dx.doi.org/10.1201/9781420071009>
- [79] Todeschini, R. and Gramatica, P. (1997) WHIM-3D 3.3. Milano.
- [80] Todeschini, R. and Gramatica, P. (1998) New 3D Molecular Descriptors: The WHIM and QSAR Applications. *Perspectives in Drug Discovery and Design*, **9-11**, 355-380. <http://dx.doi.org/10.1023/A:1027284627085>
- [81] Guha, R. (2012) Exploring Structure-Activity Data Using the Landscape Paradigm. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **2**, 829-841. <http://dx.doi.org/10.1002/wcms.1087>
- [82] Ferreira, J.E.V., Figueiredo, A.F., Barbosa, J.P., Cristino, M.G.G., Macêdo, W.J.C., Silva, O.P.P., Malheiros, B.V., Serra, R.T.A. and Pinheiro, J.C. (2010) A Study of New Antimalarial Artemisinins through Molecular Modeling and

- Multivariate Analysis. *Journal of the Serbian Chemical Society*, **75**, 1533-1548. <http://dx.doi.org/10.2298/JSC100126124F>
- [83] Santos, C.B.R., Vieira, J.B., Lobato, C.C., Hage-Melim, L.I.S., Souto, R.N.P., Lima, C.S., Costa, E.V.M., Brasil, D.S.B., Macêdo, W.J.C. and Carvalho, J.C.T. (2014) A SAR and QSAR Study of New Artemisinin Compounds with Antimalarial Activity. *Molecules*, **19**, 367-399. <http://dx.doi.org/10.3390/molecules19010367>
- [84] Flower, R. (2003) What Are All the Things that Aspirin Does? *British Medical Journal*, **327**, 572-573. <http://dx.doi.org/10.1136/bmj.327.7415.572>
- [85] Fraga, C.A.M. (2001) Razões da Atividade Biológica: Interações Micro- e Biomacromoléculas. *Química Nova na Escola*, São Paulo-SP, 33-42.
- [86] Santos, C.B.R., Lobato, C.C., Sousa, M.A.C., Macêdo, W.J.C. and Carvalho, J.C.T. (2014) Molecular Modeling: Origin, Fundamental Concepts and Applications Using Structure-Activity Relationship and Quantitative Structure-Activity Relationship. *Reviews in Theoretical Science*, **2**, 91-115. <http://dx.doi.org/10.1166/rits.2014.1016>
- [87] Politzer, P. and Murray, J.S. (2002) The Fundamental Nature and Role of the Electrostatic Potential in Atoms and Molecules. *Theoretical Chemistry Accounts*, **108**, 134-142. <http://dx.doi.org/10.1007/s00214-002-0363-9>
- [88] Bernardinelli, G., Jefford, C.W., Maric, D., Thomson, C. and Weber, J. (1994) Computational Studies of the Structures and Molecules. *International Journal Quantum Chemistry*, **52**, 117-131. <http://dx.doi.org/10.1002/qua.560520710>
- [89] Murray, J.S. and Politzer, P. (2011) The Electrostatic Potential: An Overview. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **1**, 153-163. <http://dx.doi.org/10.1002/wcms.19>
- [90] Galabov, B., Ilieva, S., Koleva, G., Allen, W.D., Schaefer III, H.F. and Schleyer, P.V.R. (2013) Structure-Reactivity Relationships for Aromatic Molecules: Electrostatic Potentials at Nuclei and Electrophile Affinity Indices. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **3**, 37-55. <http://dx.doi.org/10.1002/wcms.1112>
- [91] Martins, G.R., Napolitano, H.B., Camargo, L.T.F.M. and Camargo, A.J. (2012) Structure-Activity Relationship Study of Rutaecarpine Analogous Active against Central Nervous System Cancer. *Journal of the Brazilian Chemical Society*, **23**, 2183-2190. <http://dx.doi.org/10.1590/S0103-50532013005000004>
- [92] Honório, K.M. and Silva, A.B.F. (2003) An AM1 Study on the Electron-Donating and Electron-Accepting Character of Biomolecules. *International Journal of Quantum Chemistry*, **95**, 126-132. <http://dx.doi.org/10.1002/qua.10661>
- [93] Hirao, H. (2011) Correlation Diagram Approach as a Tool for Interpreting Chemistry: An Introductory Overview. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **1**, 337-349. <http://dx.doi.org/10.1002/wcms.20>
- [94] Grant, G.H. and Richards, W.G. (1996) *Computational Chemistry*. Oxford Science Publications, Oxford.
- [95] Contreras, R., Domingo, L.R., Andrés, J., Pérez, P. and Tapia, O. (1999) Nonlocal (Pair Site) Reactivity from Second-Order Static Density Response Function: Gas- and Solution-Phase Reactivity of the Acetaldehyde Enolate as a Test Case. *The Journal of Physical Chemistry A*, **103**, 1291-1436. <http://dx.doi.org/10.1021/jp982801i>
- [96] Clare, B.W. (1995) The Relationship of Charge Transfer Complexes to Frontier Orbital Energies in QSAR. *Journal of Molecular Structure: Theochem*, **331**, 63-78. [http://dx.doi.org/10.1016/0166-1280\(94\)03783-H](http://dx.doi.org/10.1016/0166-1280(94)03783-H)
- [97] Clare, B.W. (1995) Charge Transfer Complexes and Frontier Orbital Energies in QSAR: A Congeneric Series of Electron Acceptors. *Journal of Molecular Structure: Theochem*, **337**, 139-150. [http://dx.doi.org/10.1016/0166-1280\(95\)04135-S](http://dx.doi.org/10.1016/0166-1280(95)04135-S)
- [98] Clare, B.W. (1994) Frontier Orbital Energies in Quantitative Structure-Activity Relationships: A Comparison of Quantum Chemical Methods. *Theoretical Chemistry Accounts*, **87**, 415-430. <http://dx.doi.org/10.1007/BF01127805>
- [99] Heaton, C.A., Miller, A.K. and Powell, R.L. (2001) Predicting the Reactivity of Fluorinated Compounds with Copper Using Semi-Empirical Calculations. *Journal of Fluorine Chemistry*, **107**, 1-3. [http://dx.doi.org/10.1016/S0022-1139\(00\)00324-9](http://dx.doi.org/10.1016/S0022-1139(00)00324-9)
- [100] Zhang, G. and Musgrave, C.B. (2007) Comparison of DFT Methods for Molecular Orbital Eigenvalue Calculations. *The Journal of Physical Chemistry*, **111**, 1554-1561. <http://dx.doi.org/10.1021/jp061633o>
- [101] Comelli, N.C., Duchowicz, P.R., Lobayan, R.M., Jubert, A.H. and Castro, E.A. (2012) QSPR Study of Valproic Acid and Its Functionalized Derivates. *Molecular Informatics*, **31**, 181-188. <http://dx.doi.org/10.1002/minf.201100119>
- [102] Tropsha, A. (2010) Best Practices for Qsar Model Development, Validation and Exploitation. *Molecular Informatics*, **29**, 476-488. <http://dx.doi.org/10.1002/minf.201000061>
- [103] Barbosa, J.P., Ferreira, J.E.V., Figueiredo, A.F., Almeida, R.C.O., Silva, O.P.P., Carvalho, J.R.C., Cristino, M.G.G., Pinheiro, J.C., Vieira, J.L.F. and Serra, R.T.A. (2011) Molecular Modeling and Chemometric Study of Anticancer Derivatives of Artemisinin. *Journal of the Serbian Chemical Society*, **76**, 1263-1282. <http://dx.doi.org/10.2298/JSC111227111B>
- [104] STATISTICA (Data Analysis Software System). (2004) Version 6.1. StatSoft, Inc., Tulsa. <http://www.statsoft.com>

- [105] Snedecor, G.W. and Cochran, W.G. (1967) *Statistical Methods*. Oxford and IBH, New Delhi, 381-418.
- [106] Chatterjee, S., Hadi, A.S. and Price, B. (2000) *Regression Analysis by Examples*. 3rd Edition, Wiley VCH, New York.
- [107] Diudea, M.V. (2000) *QSPR/QSAR Studies for Molecular Descriptors*. Nova Science, Huntington, New York.
- [108] Debnath, B., Gayen, S., Bhattacharya, S., Samanta, S. and Jha, T. (2003) QSAR Study on Some Pyrodoacridine Ascidiemin Analogues as Anti-Tumor Agents. *Bioorganic & Medicinal Chemistry*, **11**, 5493-5499. <http://dx.doi.org/10.1016/j.bmc.2003.09.022>
- [109] Santos, C.B.R., Vieira, J.B., Formigosa, A.S., Costa, E.V.M., Pinheiro, M.T., Silva, J.O., Macêdo, W.J.C. and Carvalho, J.C.T. (2014) Validation of Computational Methods Applied in Molecular Modeling of Artemisinin with Antimalarial Activity. *Journal of Computational and Theoretical Nanoscience*, **11**, 1-9.
- [110] Jacob, J.N. and Tazawa, M.J. (2012) Glucose-Aspirin: Synthesis and *in Vitro* Anti-Cancer Activity Studies. *Bioorganic & Medicinal Chemistry Letters*, **22**, 3168-3171. <http://dx.doi.org/10.1016/j.bmcl.2012.03.053>