

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE QUÍMICA

DEPARTAMENTO DE FÍSICO-QUÍMICA

Estudo Teórico das Relações Estrutura - Atividade Biológica de uma série de derivados de Chalcona (1,3-difenil-2-propen-1-ona) como Agentes Anti-*Plasmodium falciparum* (Agentes Anti-maláricos).

Dissertação de Mestrado
apresentada ao Instituto
de Química da UNICAMP
para obtenção do título
de Mestre em Química.

200413522



Aluno
Luiz Frederico Motta

Orientador
Prof. Dr. Yuji Takahata

Campinas, SP.

Janeiro/2004.

UNICAMP
BIBLIOTECA CENTRAL
SEÇÃO CIRCULANTE

VIDADE	7C
2 CHAMADA	
	T/UNICAMP
	M 858e
	EX
OMBO BC/	59503
ROC.	16.117.04
C	☐
D	☒
PREÇO	11,00
DATA	13.9.04
Nº CPD	

Bib Id 322412

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE QUÍMICA
UNICAMP**

M858e

Motta, Luiz Frederico.

Estudo teórico das relações estrutura – atividade biológica de uma série de derivados de Chalcona (1,3-difenil-2-propen-1-ona) como agentes anti – *Plasmodium falciparum* (agentes antimaláricos) / Luiz Frederico Motta. -- Campinas, SP: [s.n], 2004.

Orientador: Yuji Takahata.

Dissertação (Mestrado) – Universidade Estadual de Campinas, Instituto de Química.

1. Chalcona. 2. QSAR. 3. Modelo linear de Hansch. 4. Malária. I. Takahata, Yuji.
II. Universidade Estadual de Campinas. III. Título.

Dedico esta pesquisa:
A meus pais, Luiz Carlos e
Clara Lúcia (*in memoriam*);
A meus filhos,
Frederico e Fernando,
E a minha querida esposa,
Fabiana.

Agradecimentos

Meu orientador, Dr. Yuji Takahata, um ilustre e iluminado homem! Homem que me ensinou: “Teoria e prática são termos que sempre deverão estar lado a lado, mas se por algum motivo a teoria vier a faltar, é preciso aprender a trabalhar com a máquina mesmo sem conhecer a teoria do funcionamento desta máquina”.... Muito obrigado, Takahata, pelo enorme apoio, carinho e compreensão. Posso até afirmar que o estimo como um segundo pai!

Ao Prof. Dr. Anderson Coser que gentilmente me ensinou a trabalhar com o programa de sua autoria BuildQSAR, que foi de grande utilidade em minha pesquisa QSAR;

Ao Prof. Dr. Fernando Coelho que por intermédio de suas aulas passei a simpatizar e a compreender melhor certos mecanismos fundamentais da química orgânica, e por explicar de forma objetiva e abrangente os diversos tópicos da química medicinal. Muito obrigado também pelo aceite em colaborar com a validação de minha pesquisa contribuindo com a síntese e o planejamento de novos fármacos, objeto de pesquisa no meu Doutorado;

À Prof. Dra. Marcia Miguel de Castro, pelas valiosas contribuições na parte estatística de minha pesquisa e pelo carinho e respeito para com a minha pessoa;

À minha maravilhosa esposa, Fabiana Motta, que soube ter o “dom da paciência” por vários momentos..... momentos que serão lembrados eternamente! Nesta trajetória de vida, aprendemos a caminhar juntos.... e aprendemos também “que nada está perdido e que sempre existe uma solução”. Obrigado pelo enorme carinho e amor que você nos dedicou durante esses dez anos...

Meus filhos, Fredinho e Fernando, que tanto se esforçaram para não fazer barulho, pois quantas vezes ouviram “o pai precisa estudar meus filhos”.

E finalmente,

Aos funcionários do Instituto de Química, em especial a Bel, que sempre estiveram prontos a atender às necessidades da pesquisa e da pós-graduação;

Muito obrigado!

Curriculum vitae

Luiz Frederico Motta

Natural de Uberlândia (MG)

Estado civil: Casado

Filho de Luiz Carlos Motta e Clara Lúcia de Paula Motta

➤ **Formação Acadêmica**

- ◆ Bacharel em Farmácia pela Faculdade de Ciências Farmacêuticas da Universidade São Francisco (1998);
- ◆ Bioquímico com habilitação em Análises Clínicas pela Faculdade de Ciências Farmacêuticas da Universidade São Francisco (1999);
- ◆ Licenciado em Química pelo Programa Especial de Formação Pedagógica (Resolução CNE 2/97) pela Fundação Instituto de Ensino de Osasco (2000);
- ◆ Pós-graduado em Química (*Lato Sensu*) pelas Faculdades Oswaldo Cruz (2002).

➤ **Atividade Profissional:**

❖ **Docente de química**

- ◆ Colégio e Curso Objetivo: Uberlândia – MG (1988-1990)
- ◆ Sistema Integrado de Ensino (Objetivo): Patos de Minas - MG (1990-1991)
- ◆ Sistema Integrado de Ensino (Objetivo): Araguari – MG (1990-1991)
- ◆ Colégio e Curso Anglo: Ituiutaba – MG (1990-1991)
- ◆ Colégio Oswaldo Cruz (COC): Patrocínio – MG (1988-1991)
- ◆ Colégio Estadual Pré-vestibular de Goiânia: Goiânia – GO (1992-1995)
- ◆ Colégio Carlos Chagas: Goiânia –GO (1992-1995)
- ◆ Sistema Anglo de Ensino: Goiânia – GO (1993-1995)
- ◆ Colégio Sigma: Goiânia – GO (1994 –1995)
- ◆ Curso Método Vestibulares: Goiânia – GO (1994-1995)
- ◆ Colégio e Curso Campus Vestibulares: Goiânia – GO (1994-1996)
- ◆ Colégio Disciplina: Goiânia – GO (1995)
- ◆ Colégio Ateneu Dom Bosco: Goiânia – GO (1995-1997)
- ◆ Colégio e Curso Domus: Goiânia – GO (1995-1996)
- ◆ Colégio Máximo – Sistema Positivo: Goiânia – GO (1996)
- ◆ Colégio e Curso Prisma Vestibulares: Goiânia – GO (1996-1997)
- ◆ Colégio Interativo: Goiânia – GO (1996-1997)
- ◆ Curso LFM Vestibulares: Goiânia – GO (1996-1997)
- ◆ Colégio Guimarães Rosa (Curso Integral): Bragança Paulista – SP (1998-2001)
- ◆ Sistema Anglo de Ensino: Pouso Alegre – MG (1998-2002)
- ◆ Sistema Anglo de Ensino: Ouro Fino – MG(1999-2002)
- ◆ Universidade São Camilo: São Paulo - Curso: Farmácia; Disciplina: Físico-química (Professor Convidado) – 1º Semestre de 2002.
- ◆ Universidade São Camilo: São Paulo - Curso: Nutrição; Disciplina: Bioquímica Básica (Professor Convidado) – 2º Semestre de 2002.
- ◆ Colégio e Curso Integral: Atibaia – SP (2001- 2003)
- ◆ Sistema Pueri Domus: Bragança Paulista – SP (2003)
- ◆ Colégio e Curso Integral: Jundiaí – SP (2003)

❖ **Estágios**

- ◆ Farmácia Cisne de Ouro (Pouso Alegre – MG) - 20 a 31 de julho de 1998; Período Integral.

❖ **Participação em cursos científicos**

- ◆ Curso de Computação Básica e Programação (Universidade São Francisco – 1998).
- ◆ Métodos Cromatográficos na Análise de fármacos e alimentos (Universidade São Francisco – 1998).
- ◆ Cosmetologia: “Tecnologia avançada, Matérias primas e Marketing” (Universidade São Francisco – 1998).
- ◆ Plantas utilizadas como drogas ilícitas: “Psicotrópicos e Eufomânicas” (Universidade São Francisco – 1998).
- ◆ Formas farmacêuticas sólidas (Universidade São Francisco – 1999).
- ◆ Assistência farmacêutica na farmácia hospitalar (Universidade São Francisco – 1999).
- ◆ Imunodeficiência e HIV – “Diagnóstico imunológico e biologia molecular” (Universidade São Francisco – 1999).
- ◆ Relações entre Estrutura Química e Atividade Biológica (SAR e QSAR) (Universidade Estadual de Campinas – UNICAMP – 2000).
- ◆ Introdução à Química Computacional (CENAPAD – Universidade Estadual de Campinas – UNICAMP – 2001).
- ◆ Fundamentos de Cosmetologia (Instituto de Química da Universidade Estadual de Campinas – UNICAMP – 2002).
- ◆ O Universo da Química Computacional e seu Impacto Científico (Instituto de Química da Universidade Estadual de Campinas – UNICAMP – 2002).
- ◆ Câncer: “Uma visão clínica, uma visão farmacêutica” (Faculdade de Ciências Farmacêuticas da Universidade de São Paulo – USP – 2002).
- ◆ Bioquímica do Câncer (Faculdade de Ciências Farmacêuticas da Universidade de São Paulo – USP – 2002).

❖ **Participação em simpósios e congressos**

- ◆ XI Semana Científica Farmacêutica (Universidade Federal de Goiás – UFG – 1992).
- ◆ XI Jornada Universitária Farmacêutica e II Simpósio de Farmácia (Universidade São Francisco – USF – 1998).
- ◆ XII Jornada Universitária Farmacêutica (Universidade São Francisco – USF – 1999).
- ◆ I Workshop de Modelagem Molecular de Ribeirão Preto (Faculdade de Ciências Farmacêuticas – USP – 2002).
- ◆ II Semana de Química (Instituto de Química da Universidade Estadual de Campinas – UNICAMP – 2002).
- ◆ XXXVII Semana Universitária Paulista de Farmácia e Bioquímica e VII Semana Farmacêutica de Ciência e Tecnologia (Faculdade de Ciências Farmacêuticas da Universidade de São Paulo – USP – 2002).

❖ **Disciplinas cursadas como aluno especial**

- Local: Universidade São Francisco
- Curso: Mestrado em Ciências Farmacêuticas
- ◆ Métodos Instrumentais de Análise Química Aplicados a Insumos e Medicamentos.
- ◆ Quimiometria: Análise Multivariada de Dados Experimentais

Estudo Teórico das Relações Estrutura - Atividade Biológica de uma série de derivados de Chalcona (1,3-difenil-2-propen-1-ona) como Agentes Anti - Plasmodium falciparum (Agentes Antimaláricos).

Luiz Frederico Motta

Orientador: Prof. Dr. Yuji Takahata

Palavras- chave: Chalcona, Malária, QSAR, RLM, Modelo Linear de Hansch.

Resumo

O presente trabalho teve como objetivo investigar teoricamente as relações quantitativas entre estrutura química e atividade biológica (QSAR) de uma série de derivados de Chalcona (1,3-difenil-2-propen-1-ona). Os dados observados de atividade biológica correspondem à inibição *in vitro* da parasitemia ocasionada pelas cepas cloroquina-resistente de *Plasmodium falciparum*, agente causador da malária. A primeira etapa da investigação consistiu em determinar a geometria molecular das estruturas químicas realizando minimização energética e otimização da geometria para cada estrutura pelo método Mecânico Molecular MM2 e semi-empírico PM3. A estrutura resultante foi submetida a uma minimização simplex utilizando o valor da constante dielétrica em meio aquoso. A seguir, inúmeras propriedades moleculares de natureza lipofílica, eletrônica, estereoquímica, termodinâmica e de biodisponibilidade foram calculadas. Primeiramente, selecionamos quatorze propriedades moleculares importantes. Estas variáveis foram submetidas à análise estatística Regressão Linear Múltipla (RLM) no intuito de investigarmos a correlação existente entre a estrutura química e atividade biológica, sendo que algumas variáveis explicativas foram selecionadas para estimativa de um modelo linear de Hansch. O modelo proposto revelou a contribuição de quatro descritores independentes (Log P, Hydrogen bond donor, parâmetro estérico B4 de Verloop e comprimento molecular). Concluímos também que a carga atômica parcial no oxigênio da carbonila parece estar relacionada com a atividade biológica destes derivados.

Theoretical study of Structure-Activity Relationship of a series of Chalcone derivatives (1,3-diphenyl-2-propen-1-one), Anti *Plasmodium falciparum* Agents (Anti malaria agents).

Luiz Frederico Motta

Supervisor: Prof. Dr. Yuji Takahata

Keywords: Chalcone, Malaria, QSAR, MLR, Hansch Linear Model

Abstract

The object of the present work is to conduct quantitative structure activity relationships (QSAR) studies on a series of Chalcone derivatives (1,3-diphenyl-2-propen-1-one). The compounds have inhibitory *in vitro* activity against chloroquine-resistant *Plasmodium falciparum*, which causes malaria. The first step of the investigation consisted of determination of molecular geometry calculating the lowest energy employing Molecular Mechanics MM2 and PM3. Environment of aqueous solution was assumed by adopting the value of dielectric constant (ϵ) of water. Then a large number of molecular properties such as lipophilic, electronic, steric, thermodynamic, and bioavailability were calculated. We first selected about fourteen important molecular properties. They were used the Multiple Linear Regression analysis to establish QSAR model of Hansch type. Four independent descriptors were finally selected: Log P, Hydrogen Bond Donor Capacity, a steric parameter B4 of Verloep and Molecular length. There is an indication also that the atomic charge of carbonyl oxygen seems to be related with biological activity of the compounds.

Índice de Abreviaturas

3D	Tridimensional
ACCVIP	Australian Computational Chemistry via the Internet Project
ANOVA	Análise de Variância
ASH	Área Superficial Hidrofílica
ATP	Adenosina Trifosfato
CADD	Computer - Assisted Drug Design
CAMD	Computer - Aided Molecular Design
C _{aq}	Concentração em fase aquosa
CNDO	Compleat Neglect of Differential Overlap
C _{org}	Concentração em fase orgânica
CSP	ChemSite Pro
d	Distância interatômica
DAS	Divisão de Estrutura Química e Atividade Biológica
DHFR	Dihidrofolato Redutase
DHPS	Dihidropterato sintetase
DMD	Divisão de Química Medicinal
DNA	Ácido Desoxirribonucleico
FDA	Food and Drug Administration
FH ₂	Dihidrofolato
FH ₄	Tetrahidrofolato
HBN	Número de Ligações de Hidrogênio
HFR	Hartree-Fock-Roothaan
HLB	Balanço hidrofílico-lipofílico
HOMO	Orbital Molecular Ocupado de maior Energia
IL-12	Interleucina 12
INDO	Intermediate Neglect of Differential Overlap
INF- γ	Interferon gama
LCAO	Linear Combination of Atomic Orbitals
Log S.	Logarítimo da Solubilidade

LUMO	Orbital Molecular Desocupado de menor Energia
L_x	Comprimento Molecular
L_y	Largura Molecular
L_z	Profundidade Molecular
MINDO	Modified Intermediate Neglect of Differential Overlap
MM	Mecânica Molecular
MMP	Molecular Modeling Pro
MQ	Mecânica Quântica
MR	Refratividade molar
MSP-1	Merozoite Surface Protein-1
NH	Número de Hidratação
NK	Natural Killer
NOX	Número de oxidação
OMS	Organização Mundial de Saúde
P	Coeficiente de partição
PA	Parachor
PABA	Ácido paraminobenzóico
PC	Pressão Crítica
PDH	Ligação de Hidrogênio
PE	Ponto de Ebulição
PES	Energia Potencial de Superfície
PM	Peso Molecular
PM3	Parametric Method 3
PNCM	Programa Nacional de Prevenção e Controle da Malária
PS	Parâmetro de Solubilidade
q	Carga Atômica Parcial
QM/MM	Quantum Mechanics/ Molecular Mechanics
QSAR	Quantitative Structure-Activity Relationships
RBM	Roll Back Malaria
REDOX	Oxi-redução
RHF	Restricted Hartree-Fock

RLM	Regressão Linear Múltipla
RNA	Ácido Ribonucleico
ROI	Reagentes Intermediários Oxigenados
SAR	Structure-Activity Relationships
SCF	Self-Consistent Field
SNC	Sistema Nervoso Central
SSP-2	Sporozoite Surface Protein-2
SvdW	Superfície de van der Waals
SvdWsol	Superfície de van der Waals acessível ao solvente
tau	Número efetivo de Ligações Torsionais
TC	Temperatura Crítica
TNF α	Fator de Necrose Tumoral alfa
UHF	Unrestricted Hartree-Fock
UMP	Uridina Monofosfato
VM	Volume Molecular
XN	Xanthohumol
ZDO	Zero differential overlap
α	Ângulo entre ligações
β	Ângulo Diedral
γ	Tensão Superficial
ε	Parâmetro de Verloop
μ	Momento de Dipolo
ρ	Densidade Molecular
σ	Constante de Hammett

Índice de Tabelas

Tabela 01 – Derivados de Chalcona e suas respectivas atividades biológicas: concentração (μM) necessária para inibição <i>in vitro</i> de 50% da parasitemia provocada pela cepa W2 de <i>P. falciparum</i> _____	41
Tabela 02 – Energia em Kcal/mol dos derivados de Chalcona durante o processo de minimização energética e otimização da geometria molecular _____	110
Tabelas 03 e 04 – Parâmetros Hidrofóbicos _____	111 e 112
Tabela 05 – Parâmetros eletrônicos empíricos _____	113
Tabelas 06 a 08 – Parâmetros eletrônicos quânticos _____	114 a 116
Tabela 09 – Parâmetros estereoquímicos _____	117
Tabelas 10 e 11– Parâmetros estereoquímicos de Verloop _____	118 e 119
Tabela 12 – Parâmetros termodinâmicos _____	120
Tabela 13 – Parâmetros de biodisponibilidade _____	121
Tabelas 14 e 15 – Outros parâmetros _____	122 e 123
Tabela 16 – Regressão Linear: $-\text{Log IC}_{50}(\text{W2}) = f$ (parâmetros hidrofóbicos) ____	124
Tabela 17 – Regressão Linear: $-\text{Log IC}_{50}(\text{W2}) = f$ (parâmetros eletrônicos empíricos) _____	125
Tabela 18 – Regressão Linear: $-\text{Log IC}_{50}(\text{W2}) = f$ (parâmetros eletrônicos quânticos) _____	125
Tabela 19 – Regressão Linear: $-\text{Log IC}_{50}(\text{W2}) = f$ (parâmetros estereoquímicos) _____	125
Tabela 20 – Regressão Linear: $-\text{Log IC}_{50}(\text{W2}) = f$ (parâmetros estereoquímicos) – (SvdW) _____	126

Tabela 21 – Regressão Linear: $-\text{Log IC}_{50}(\text{W2}) = f$ (parâmetros estereoquímicos de Verloop) (ə)	126
Tabela 22 – Regressão Linear: $-\text{Log IC}_{50}(\text{W2}) = f$ (parâmetros termodinâmicos)	126
Tabela 23 – Regressão Linear: $-\text{Log IC}_{50}(\text{W2}) = f$ (parâmetros de biodisponibilidade)	126
Tabela 24 – Regressão Linear: $-\text{Log IC}_{50}(\text{W2}) = f$ (distância interatômica)	127
Tabela 25 – Regressão Linear: $-\text{Log IC}_{50}(\text{W2}) = f$ (ângulo entre ligações e ângulo diedral)	127
Tabela 26 – Matriz de correlação entre os parâmetros hidrofóbicos	130
Tabela 27 – Matriz de correlação entre os parâmetros eletrônicos empíricos	130
Tabela 28 – Matriz de correlação entre os parâmetros eletrônicos quânticos	131
Tabela 29 – Matriz de correlação entre os parâmetros estereoquímicos	131
Tabela 30 – Matriz de correlação entre os parâmetros termodinâmicos	132
Tabela 31 – Matriz de correlação entre os parâmetros de biodisponibilidade	132
Tabela 32 – Matriz de correlação entre distâncias interatômicas, ângulos entre ligações e ângulos diedrais	132
Tabela 33 - Matriz de correlação entre as variáveis selecionadas (22 variáveis) na primeira etapa seletiva	133
Tabela 34 - Matriz de correlação entre as variáveis descritoras selecionadas (14 variáveis)	134
Tabelas 35 a 37 – Seleção de variáveis descritoras por intermédio de algoritmo genético: combinações com 3, 4 e 5 variáveis descritoras	136 e 137

Tabelas 38 a 40 – Seleção de variáveis descritoras por intermédio de busca sistemática: combinações com 3, 4 e 5 variáveis descritoras _____ 138 e 139

Tabelas 41 a 43 – Três melhores combinações apresentando 3, 4 e 5 variáveis descritoras _____ 140

Tabela 44 – Matriz de correlação entre as variáveis participantes do modelo linear proposto _____ 151

Tabela 45 – Análise de variância (ANOVA) da regressão para o modelo proposto _____ 152

Tabela 46 – Matriz de correlação de alguns parâmetros eletrônicos _____ 165

Índice de Figuras

Figura 1: Mosquito <i>Anopheles</i> sp _____	09
Figura 2: Distribuição geográfica da malária _____	10
Figura 3: Ciclo evolutivo do <i>Plasmodium</i> sp _____	12
Figura 4: Atividade enzimática das enzimas DHPS e DHFR _____	17
Figura 5: Endocitose do citosol da hemácia _____	18
Figura 6: Tipos de clivagem executada pelas proteases _____	19
Figura 7: Microfotografia eletrônica da invasão do eritrócito pelo merozoíta ____	21
Figura 8: Degradação semi-ordenada da hemoglobina no vacúolo digestivo do <i>Plasmodium</i> sp _____	24
Figura 9: Células de <i>P. falciparum</i> lisando eritrócitos. Microscopia eletrônica de varredura, artificialmente colorida _____	25
Figura 10: Citoaderência endotelial _____	25
Figura 11: Dimerização do grupo Heme em β -Hematina _____	29
Figura 12: Biocristalização da Hemozoína _____	29
Figura 13: Mecanismo de ação de anti-folatos _____	34
Figura 14: Distribuição da malária e resistência do <i>Plasmodium falciparum</i> à cloroquina _____	36
Figura 15: Estrutura química dos derivados de Chalcona _____	39 -41
Figura 16: Energia de estiramento de ligação _____	53
Figura 17: Energia de deformação angular _____	53
Figura 18: Energia de deformação torsional _____	54
Figura 19: Energia de van der Waals _____	55

Figura 20: Estrutura protótipo dos derivados de Chalcona	84
Figura 21: Estrutura química base dos derivados de Chalcona com a numeração fornecida via utilização do programa Molecular Modeling Pro	87
Figura 22: Análise gráfica do modelo 2 (74) com 3 variáveis descritoras obtida no programa BuildQSAR: atividade observada x atividade prevista (sup.); resíduos dos valores previstos (inf.)	145
Figura 23: Análise gráfica do modelo 1 (76) com 4 variáveis descritoras obtida no programa BuildQSAR: atividade observada x atividade prevista (sup.); resíduos dos valores previstos (inf.)	147
Figura 24: Análise gráfica do modelo 2 (77) com 4 variáveis descritoras obtida no programa BuildQSAR: atividade observada x atividade prevista (sup.); resíduos dos valores previstos (inf.)	148
Figura 25: Análise gráfica do modelo 3 (81) com 5 variáveis descritoras obtida no programa BuildQSAR: atividade observada x atividade prevista (sup.); resíduos dos valores previstos (inf.)	149 e 150
Figura 26: Caráter lipofílico-hidrofílico da Chalcona 07, executado no programa MMP	161
Figura 27: Densidade eletrônica superficial obtida via cargas parciais no programa ChemSite Pro pelo método Combinação Linear de Orbitais Atômicos (LCAO) para a Chalcona 07	163
Figura 28: Cargas atômicas parciais obtidas pelo programa MMP para a Chalcona 07	164
Figura 29: Conformação da Chalcona 07	166

Índice

1.0 Parte I: Introdução	01
1.1 Desenvolvimento de fármacos e a indústria farmacêutica	01
1.2 A importância da Química Medicinal	03
1.3 Planejamento racional de fármacos - "Drug Design"	04
1.4 Química teórica computacional	07
1.5 Malária: uma doença nos países em desenvolvimento	09
1.6 Malária: um ciclo evolutivo bem adaptado	12
1.7 Bioquímica do <i>Plasmodium falciparum</i>	15
1.8 Proteases: enzimas necessárias no metabolismo do <i>Plasmodium falciparum</i>	20
1.8.1 Proteases mediando invasão e ruptura do eritrócito	21
1.8.2 Proteases mediando degradação da hemoglobina	22
1.9 Malária: doença com severas complicações	24
1.10 Aspectos imunológicos e mecanismo de escape do <i>Plasmodium</i>	26
1.11 Toxicidade do grupo heme	28
1.12 Quimioterapia da malária	30
1.12.1 Classificação dos anti-maláricos	31
1.12.2 Mecanismo de ação de anti-maláricos	32
1.12.3 Resistência aos anti-maláricos	34
1.13 Chalconas: um <i>Lead Compound</i>	36
2.0 Parte II: Introdução	42
2.1 O mundo Quântico no desenvolvimento de fármacos	42
2.2 Métodos computacionais para sistemas biológicos	46
2.3 Métodos Mecânicos Moleculares (MM)	51
2.4 Métodos Mecânicos Quânticos (MQ)	59
2.4.1 Aproximação de Born-Oppenheimer	62
2.4.2 Aproximação do elétron independente	64
2.4.3 Aproximação LCAO	68
2.4.4 Métodos semi empíricos	70

3.0 Objetivo geral	72
3.1 Objetivos específicos	72
4.0 Metodologia utilizada	73
4.1 Minimização energética e otimização da geometria	73
4.2 Cálculo de parâmetros físico-químicos	77
4.2.1 Parâmetros hidrofóbicos	78
4.2.2 Parâmetros eletrônicos	82
4.2.3 Parâmetros estereoquímicos	89
4.2.4 Parâmetros termodinâmicos	92
4.2.5 Parâmetros de biodisponibilidade	94
4.2.6 Outros parâmetros	95
4.3 Análise estatística	96
4.3.1 Regressão linear simples	98
4.3.2 Matriz de correlação e seleção de variáveis em QSAR	100
4.3.3 Análise do modelo escolhido	105
5.0 Resultados	109
5.1 Minimização energética e otimização da geometria	109
5.2 Cálculo de parâmetros físico-químicos	110
5.3 Análise estatística	124
5.3.1 Regressão linear simples	124
5.3.2 Matriz de correlação e seleção de variáveis	128
5.3.3 Seleção de variáveis descritoras	135
5.3.4 Regressão linear múltipla	139
5.3.5 Modelos multidimensionais lineares selecionados	141
5.3.6 Análise dos modelos selecionados	143
5.3.7 Modelo multidimensional linear proposto	150
6.0 Discussão	154
7.0 Conclusões	169
8.0 Referências bibliográficas	172

Parte I: Introdução

1.1 Desenvolvimento de fármacos e a indústria farmacêutica

A Saúde é um parâmetro de suma importância no quadro econômico, político, cultural e social de um país, sendo então caracterizada como um dos pilares fundamentais no processo de desenvolvimento social de uma nação. Uma enorme gama de pesquisas sociais têm demonstrado significativa influência na relação entre a Saúde e os indicadores de qualidade de vida das populações.

No capítulo da Ordem Social da Constituição Federal do Brasil, consta o princípio básico: “A Saúde é direito de todos e dever do Estado, assegurado mediante políticas sociais e econômicas que visam à redução do risco de doenças e de outros agravantes e ao acesso universal e igualitário às ações e serviços para a sua promoção, proteção e recuperação” (Constituição Federal, 1988). Poderíamos enfatizar que o termo SAÚDE é empregado como sendo um bom e normal estado de desenvolvimento físico e psíquico do indivíduo apresentando um perfeito equilíbrio funcional de seus órgãos e sistemas, mas a Organização Mundial de Saúde (OMS), vai mais além nessa caracterização, inserindo também como parâmetro de saúde a situação do indivíduo no contexto social em que vive. Sem qualquer questionamento duvidoso, a Saúde “estado de completo bem-estar físico, mental e social”, é uma necessidade básica do ser humano e ao longo dos tempos tem sido alvo de constante preocupação.

No intuito de conservar e manter uma boa Saúde, o ser humano vem utilizando os mais variados recursos para prevenir e combater as doenças através de medidas profiláticas e terapêuticas. Entende-se por *medida terapêutica* o conjunto de medidas incluindo a utilização de medicamentos e/ou outros recursos que visam o tratamento ou cura de uma determinada doença já instalada no organismo do indivíduo (Silva, 1994).

Os produtos utilizados como medida terapêutica geralmente constituem os medicamentos e correspondem a uma variedade de fármacos com efeitos biológicos diversos. O enorme avanço nas tecnologias de desenvolvimento de

fármacos verificado nos últimos anos deve-se em parte ao extraordinário papel que os produtos naturais vem desempenhando na terapêutica.

A trajetória de desenvolvimento de um fármaco potencialmente ativo resultando o mínimo de efeitos adversos é longa, complexa e onerosa. A química medicinal tem contribuído de modo decisivo na descoberta e no desenvolvimento de novos agentes terapêuticos.

Os elevados custos no desenvolvimento de novos fármacos estão associados a uma série de fatores que levam em consideração as várias fases dispendiosas envolvidas, desde a descoberta do fármaco até a introdução do medicamento no mercado (Cohen, 1996). Segundo centros de pesquisa de desenvolvimento de fármacos nos E.U.A., o custo do processo de desenvolvimento de um novo fármaco está na ordem de dois bilhões de dólares em um período de dez a quinze anos de pesquisa (Patrick, 2001).

Para que um novo medicamento seja comercializado de modo seguro é necessário um longo período de pesquisas e elevados investimentos financeiros, seja na indústria farmacêutica, em universidades ou em laboratórios de pesquisa. Nos últimos vinte anos, mais de noventa por cento dos fármacos lançados no mercado foram desenvolvidos em indústrias farmacêuticas (Korolkovas & Burckhalter, 1988).

Na busca de novos agentes terapêuticos, inúmeras substâncias são sintetizadas e testadas *in vitro* e *in vivo* em indústrias farmacêuticas, sendo que para uma média aproximada de cada lote de dez mil substâncias que entram no processo de desenvolvimento de um fármaco candidato, apenas uma é aprovada pelo FDA (INTERFARMA, 2002).

As perspectivas de crescimento no ramo da indústria farmacêutica são boas, visto que várias doenças conhecidas poderão ser tratadas no futuro e, conseqüentemente, haverá um aumento na expectativa de vida do homem. Este fato estimula a elevada competitividade entre indústrias farmacêuticas e acarreta um acréscimo cada vez maior no que diz respeito ao investimento financeiro e tecnológico por parte das indústrias na tentativa em visar lucros elevados (Krogsgaard-Larsen *et al*, 1996).

1.2 A importância da química medicinal

Nos últimos anos, a química medicinal esteve voltada para o estudo das razões moleculares da ação dos fármacos e para o estudo das relações entre a estrutura química e a atividade biológica destes fármacos, por intermédio do planejamento racional de novas substâncias bioativas (Barreiro & Fraga, 2001). As substâncias devem apresentar propriedades terapêuticas úteis, para que possam ser classificadas como um novo fármaco. A química medicinal é uma área científica marcada pela interdisciplinaridade e multidisciplinaridade e com o nítido envolvimento de inúmeras áreas de conhecimento, o que é perfeitamente justificável pelo amplo espectro de fatores envolvidos por se tratar de sistemas biológicos (Amaral & Montanari, 2002).

A química medicinal está intimamente associada à farmacologia molecular no sentido em compreender cada vez mais as razões moleculares das interações fármaco-receptor (Leitão *et al*, 2000).

O entendimento das interações fármaco-receptor em sistemas biológicos é uma tarefa minuciosa e extremamente complexa, pois envolve uma multiplicidade de fatores que estão relacionados com a resposta terapêutica. Deste modo a resposta biológica desejada em um organismo humano é resultado do reconhecimento molecular do fármaco pelo biorreceptor (Florence & Attwood, 2003).

A pesquisa e o estudo das interações, dos possíveis mecanismos de ação dos fármacos e de seus efeitos biológicos são alvos da *farmacodinâmica*. A química medicinal não se restringe apenas ao mecanismo de ação dos fármacos, leva também em consideração outros fatores como a absorção, distribuição, biotransformação e a eliminação destes, aspectos relacionados à *farmacocinética*. A *biodisponibilidade* do fármaco e o seu *perfil de toxicidade* também deverão ser analisados e fazem parte da química medicinal (Barreiro & Fraga, 2001).

Entre as inúmeras atribuições da química medicinal a grande colaboração está no planejamento racional de novas substâncias bioativas incluindo síntese e modificação molecular, isolamento de princípios ativos naturais, identificação ou elucidação de estruturas químicas, relações entre estrutura química e atividade

biológica através de cálculo de propriedades físico-químicas de moléculas, compreensão em nível molecular de processos bioquímicos, farmacodinâmicos, farmacocinéticos, toxicológicos e finalmente, propor e validar modelos matemáticos que correlacionem a estrutura química de substâncias com a sua atividade biológica (Amaral & Montanari, 2002). Esta área de conhecimento utiliza diferentes estratégias metodológicas sendo bastante útil no planejamento racional de fármacos. No entanto o sucesso no planejamento racional de um fármaco requer um amplo conhecimento multidisciplinar e uma interatividade contínua de pesquisadores em cooperações mútuas (Amaral & Montanari, 2002).

No intuito em buscar uma interatividade nesta área científica, surgiu no Brasil, a Divisão de Estrutura Química e Atividade Biológica (DAS), precursora da química medicinal, que seis anos depois viria a ser renomeada Divisão de Química Medicinal (DMD). Percebe-se que num futuro próximo existirá no Brasil uma enorme tendência em formação de grupos de pesquisadores e/ou pequenas empresas, que estarão envolvidos no desenvolvimento de fármacos e serão financiados por agentes financeiros de capital privado e/ou público ou pelas indústrias. Deste modo, doenças consideradas como negligenciadas em países em desenvolvimento, serão alvos de estudo objetivando o desenvolvimento de novos fármacos potencialmente ativos (Amaral & Montanari, 2002).

Por intermédio da química medicinal, moléculas candidatas a serem bioativas poderão ser inter-relacionadas estruturalmente visando planejar análogos com base no tipo de interação fármaco-biorreceptor. Através de pesquisas das Relações entre Estrutura Química e Atividade Biológica, algumas estruturas químicas otimizadas e validadas serão escolhidas para estudos *in vivo* e serão posteriormente selecionadas para ensaios pré-clínicos e clínicos (Krogsgaard-Larsen *et al*, 1996).

1.3 Planejamento racional de fármacos – “Drug Design”

Várias estratégias estão disponíveis para o desenho estrutural de novos fármacos. A principal estratégia utilizada no planejamento da arquitetura molecular de substâncias bioativas baseia-se no mecanismo de ação dos fármacos, visando

a identificação da estrutura do protótipo (Lead Compound) ou de um análogo ativo, principalmente quando a topografia do biorreceptor envolvido não é conhecida (Barreiro & Fraga, 2001).

O planejamento da estrutura molecular destas substâncias que serão candidatas a novos fármacos fundamenta-se no mecanismo fisiopatológico e escolha do biorreceptor envolvido. Os receptores são biomacromoléculas que apresentam uma estrutura química tridimensional estéreo-específica para com os fármacos (Silverman, 1992).

A atividade biológica de um fármaco é atribuída aos diversos tipos de interações deste com as estruturas químicas do receptor, formando o complexo fármaco-receptor. Os fármacos podem ser estruturalmente específicos, apresentando assim um elevado grau de complementaridade para com a biomacromolécula. Para compreender a estereoespecificidade dos fármacos é de suma importância conhecer as forças de interação envolvidas no processo de complexação (Kalant & Roschlau, 1991).

Através de técnicas da química teórica computacional via métodos da mecânica molecular e mecânica quântica é possível prever os prováveis tipos de interações fármaco-biorreceptor. A compreensão das interações químicas entre o fármaco com os grupos funcionais químicos do receptor é muito importante visto que resulta numa seqüência de alterações conformacionais na biomacromolécula receptora. Deve-se levar em consideração que a capacidade do fármaco em se adaptar ao receptor está intimamente associada às características estruturais e conformacionais de ambos (Kubinyi, 1993).

Levando em consideração a “teoria da perturbação macromolecular”, proposta por Belleau (Korolkovas & Burckhalter, 1988), onde a formação do complexo fármaco-biorreceptor é acompanhada de variações de energia livre de Gibbs, fica fácil compreender que tal complexação ocorrerá simultaneamente com alterações conformacionais no receptor. Esta teoria oferece um fundamento termodinâmico plausível para a explicação de processos físico-químicos que ocorrem em termos de interações intermoleculares na região ativa do biorreceptor.

Muitos fármacos apresentam como alvo terapêutico enzimas. Os fármacos que atuam em enzimas poderão desencadear um mecanismo bioquímico de ativação ou inibição. Em termos bioquímicos, a inibição enzimática refere-se ao efeito provocado por um fármaco em enzimas, inibindo então a trajetória metabólica. A inibição enzimática provocada por fármacos poderá ser reversível ou irreversível (Voet *et al*, 2002). Assim é importante ressaltar que muitos parasitas necessitam de uma extensa e complexa cascata enzimática para o seu metabolismo. A descoberta de novos fármacos inibidores enzimáticos resulta no bloqueio e interferência no metabolismo desses parasitas.

Quando a estrutura tridimensional da região ativa do biorreceptor é conhecida, a previsão de possíveis interações químicas entre o fármaco e o receptor é concretizada via processos de complementaridade molecular, identificando assim os grupos farmacofóricos das substâncias bioativas que são responsáveis pelas propriedades moleculares e o mecanismo de ação. Este processo envolve a identificação de um bioligante eficiente, cuja estrutura química é planejada com base na estrutura do biorreceptor. Não necessariamente um bioligante identificado *in vitro* possuirá perfil de biodisponibilidade adequado, sendo necessário então efetuar *modificações moleculares* em sua estrutura química a fim de melhorar sua eficiência farmacocinética, desde que a região farmacofórica seja preservada para que possa ocorrer então o reconhecimento molecular do fármaco pelo biorreceptor (Barreiro & Fraga, 2001).

Se a topografia da biomacromolécula receptora não for conhecida, devemos levar em consideração o mecanismo fisiopatológico e bioquímico do processo em estudo. Vários fármacos foram descobertos por planejamento racional tendo como base uma abordagem fisiopatológica. Neste caso, elege e planeja-se o padrão estrutural original através de variações no índice de similaridade molecular de forma a identificar um análogo ativo. Após ser comprovada a atividade biológica *in vivo* deste análogo, tem-se então um novo protótipo. De modo semelhante, o novo protótipo terá também sua eficácia farmacocinética melhorada por intermédio de modificações na estrutura química desde que os grupos farmacofóricos sejam conservados (Kubinyi, 1995).

Percebe-se então que o planejamento de fármacos inclui não apenas o planejamento da arquitetura molecular do bioligante, mas também a sua farmacocinética e o seu perfil de toxicidade, características que estão na maioria das vezes além das possibilidades do planejamento auxiliado pelo computador. Todavia, ferramentas quimiométricas apropriadas, são extremamente úteis e valiosas no planejamento e na avaliação de experimentos e resultados farmacocinéticos e toxicológicos.

O processo de desenvolvimento de fármacos envolve a identificação do receptor, a descoberta do protótipo (Lead Compound), testes da fase pré-clínica (em animais), testes da fase clínica (seres humanos) e testes clínicos pós-mercado. Na fase de descoberta do protótipo, pesquisas SAR e QSAR têm sido muito úteis para identificação de grupos farmacofóricos do protótipo (bioligante). Este estágio é conhecido como planejamento de fármacos auxiliado por computador, CADD, "Computer-Assisted Drug Design" (Cohen, 1996).

O planejamento de fármacos auxiliado por computador (CADD) é um conjunto de técnicas computacionais utilizadas para descobrir, planejar e otimizar compostos biologicamente ativos (Krogsgaard-Larsen *et al*, 1996). O planejamento de fármacos via computador é um procedimento necessário, eficiente e útil na tentativa em reduzir gastos e o tempo de confecção de um medicamento. O universo em CADD amplia-se a cada momento por intervenção de uma crescente tecnologia computacional e de novas metodologias aplicadas aos cálculos em química teórica computacional (Cohen, 1996). Através do CADD, propriedades físico-químicas moleculares de substâncias bioativas são calculadas e exploradas, levando em consideração dados experimentais de atividade biológica e os princípios de química teórica computacional.

1.4 Química teórica computacional

Sabe-se que a estrutura química de substâncias bioativas pode ser correlacionada com a resposta biológica. As relações entre estrutura química e atividade biológica de compostos podem ser expressas por intermédio de equações matemáticas que apresentam determinados parâmetros capazes de

representar as propriedades físico-químicas da estrutura destas substâncias que serão candidatas a novos fármacos potencialmente ativos. Levando em consideração que a maioria dos fármacos são estruturalmente específicos, ou seja, o seu efeito biológico resulta da interação específica com os receptores, então o reconhecimento molecular destes pelos respectivos biorreceptores depende da estrutura química de ambos, incluindo o arranjo espacial dos grupos funcionais que deverão ser complementares. Portanto, é de extrema importância que nestas equações matemáticas apareçam parâmetros de solubilidade (lipofílicos), descritores eletrônicos e descritores estéricos para uma melhor compreensão das estruturas químicas envolvidas, das possíveis interações fármaco-receptor e conseqüentemente planejar e otimizar o desenvolvimento de novos fármacos potencialmente ativos (Thomas, 2003).

A exploração e análise estatística de parâmetros qualitativos e quantitativos de uma série de estruturas químicas, correlacionando-os com a atividade biológica, pode ser realizada pela química teórica computacional por intermédio de um conjunto de métodos computacionais (Amaral & Montanari, 2002). O conjunto de procedimentos utilizados para correlacionar a estrutura química e atividade biológica de compostos de modo qualitativo e quantitativo chama-se respectivamente SAR (*Structure-Activity Relationships*) e QSAR (*Quantitative Structure-Activity Relationships*).

A química computacional é muito útil no estudo destas relações, pois utiliza métodos matemáticos para executar cálculos de propriedades moleculares simulando o comportamento molecular. Deste modo, a química computacional estende além dos limites tradicionais das ciências, permitindo explorar moléculas por intermédio de computadores e atua como um suplemento para a química experimental. “Quanto mais as ciências físicas progridem, mais elas tendem a entrar no domínio da matemática, que é um tipo de centro para onde elas convergem. Nós podemos julgar o grau de perfeição que a ciência tem alcançado, pela facilidade com que ela pode ser submetida a CÁLCULO”.- (A. Quetelet) (Morgon, 2001).

Uma das inúmeras vantagens da química teórica computacional é a de atuar como ferramenta de apoio na análise e interpretação de dados experimentais, através de informações que muitas vezes não são possíveis de serem obtidas diretamente dos experimentos.

1.5 Malária: uma doença nos países em desenvolvimento

A malária é uma doença sistêmica causada por protozoários unicelulares do gênero *Plasmodium* que parasitam e destroem as hemácias ou eritrócitos (glóbulos vermelhos do sangue) e são introduzidos no organismo humano por meio da picada de fêmeas de mosquito do gênero *Anopheles* (figura 1), por injeção e transfusão de sangue de pessoas infectadas ou por seringas hipodérmicas (Neves, 2000). Estes protozoários são caracterizados como esporozoários, ou seja, parasitas intracelulares obrigatórios de vertebrados e que não apresentam organelas de locomoção (Levental & Cheadle, 1997). O gênero *Plasmodium* engloba uma categoria de parasitas que infectam o sangue e os tecidos.

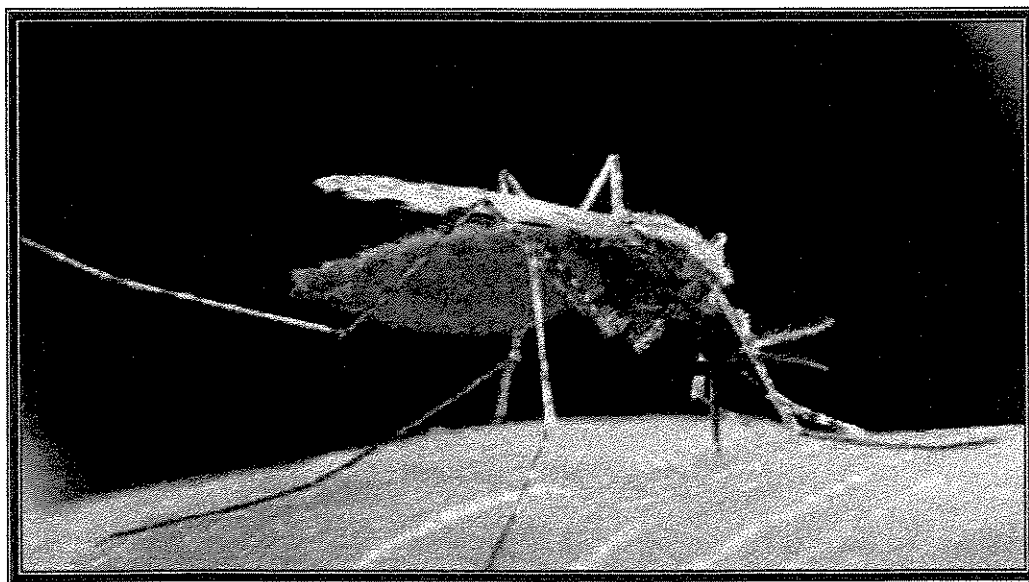


Figura 1: Mosquito *Anopheles* sp; Fonte: www.hanzeplein.nl/tropenzorg/malaria.html

A malária humana é causada pelas seguintes espécies de *Plasmodium*:
Plasmodium falciparum (Welch; 1897),
Plasmodium vivax (Grassi & Feletti; 1890),
Plasmodium malariae (Laveran; 1881) e
Plasmodium ovale (Stephens; 1922), sendo que no Brasil, as três primeiras espécies são as causadoras de malária (Katz *et al*, 1988).

O nome *malária* deriva do italiano “mal ária” e literalmente significa *mau ar*, pois escritores italianos defendiam a tese de que a doença era causada por vapores nocivos exalados dos pântanos. Nomes alternativos como, paludismo, impaludismo, são de origem francesa e têm o mesmo significado. No Brasil é também conhecida por outros nomes populares: febre palustre, impaludismo, febre intermitente, febre terçã benigna, febre terçã maligna, febre quartã, febre palúdica, maleita e sezão. Apesar de todas estas designações, o nome mais utilizado, inclusive no meio médico é malária (Neves, 2000).

A malária é uma das parasitoses que maior dano já causou a milhões de pessoas, principalmente nas regiões tropicais e subtropicais do globo (RBM) (figura 2). Segundo a Organização Mundial de Saúde, a malária é reconhecida como grave problema de Saúde Pública no mundo, atingindo aproximadamente 40% da população de mais de cem países. Estima-se que ocorrem no mundo cerca de 300 a 500 milhões de novos casos. Atualmente, é considerada como uma das principais causas de óbito mundial, levando a um milhão de mortes ao ano (FUNASA, 2002).

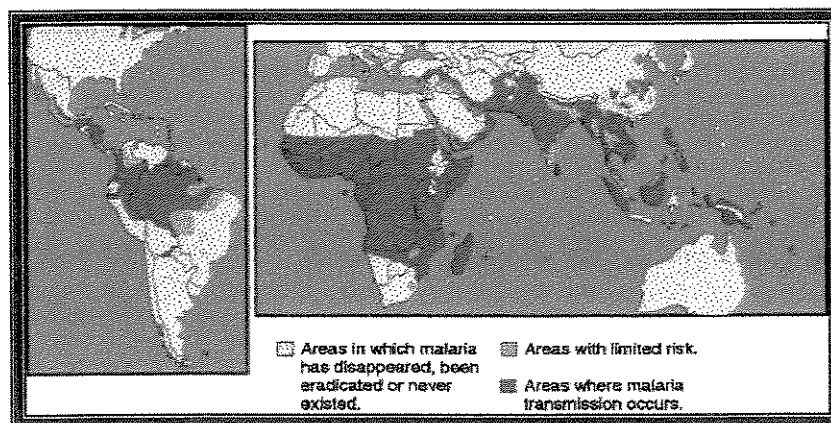


Figura 2: Distribuição geográfica da malária; Fonte: www.cdc.gov/

Várias pesquisas indicam que o *Plasmodium falciparum* é o responsável por mais de 40% dos casos de malária registrados no mundo e por mais de 90% de todas as mortes causadas por essa doença (RBM).

No Brasil, aproximadamente 99% dos casos se concentram na região amazônica sendo considerada endêmica em nove estados (FUNASA, 2002), onde as condições sócio-econômicas e ambientais favorecem a proliferação do mosquito anofelino (vetor da doença), e, conseqüentemente, a exposição de grandes contingentes populacionais. A doença causa óbito, sofrimento, perdas sociais e elevada perda econômica. Grandes investimentos empresariais são prejudicados em função da doença e a exploração do potencial turístico da região também é afetada (FUNASA, 2002).

Apesar do esforço do Ministério da Saúde em combater a malária no Brasil, através do Programa Nacional de Prevenção e Controle da Malária (PNCM), o sucesso do controle é ainda muito pequeno. Este programa é alicerçado em uma série de componentes que fazem parte das estratégias de intervenção a serem implementadas e fortalecidas de forma integrada no sentido em combater a malária (FUNASA, 2002).

Inúmeros fatores epidemiológicos (população dispersa, migrações constantes dos habitantes, moradia inadequada, falta de controle em zonas de grande concentração humana), fatores sócio-econômicos, o aumento na resistência do mosquito frente aos inseticidas e o gradativo aumento de cepas de *Plasmodium falciparum* resistentes aos fármacos constituem grandes empecilhos no controle mundial da doença. É preocupante o aumento no número de novos casos de malária no Brasil nos últimos anos, demonstrando que a parasitose não está controlada (FUNASA, 2002). Considerando o número alarmante de casos de malária no mundo e no Brasil, percebe-se que o controle da transmissão nos países em desenvolvimento continua sendo um grande desafio e um projeto a ser concretizado a longo prazo.

1.6 Malária: um ciclo evolutivo bem adaptado

O ciclo evolutivo da malária (figura 3) tem início com a picada da fêmea do mosquito (vetor) do gênero *Anopheles* ao hospedeiro vertebrado, onde ocorre a inoculação de formas infectantes (esporozoítas) juntamente com a saliva (agente anticoagulante) do mosquito. Uma vez inoculados, os esporozoítas deixam o tecido subcutâneo e entram nos vasos sanguíneos da pele caindo na corrente circulatória. Estas formas desaparecem da circulação alcançando as células do fígado (hepatócitos), efetuando assim, a invasão das células hepáticas através de receptores de superfície na membrana da célula hospedeira (Rey, 1991).

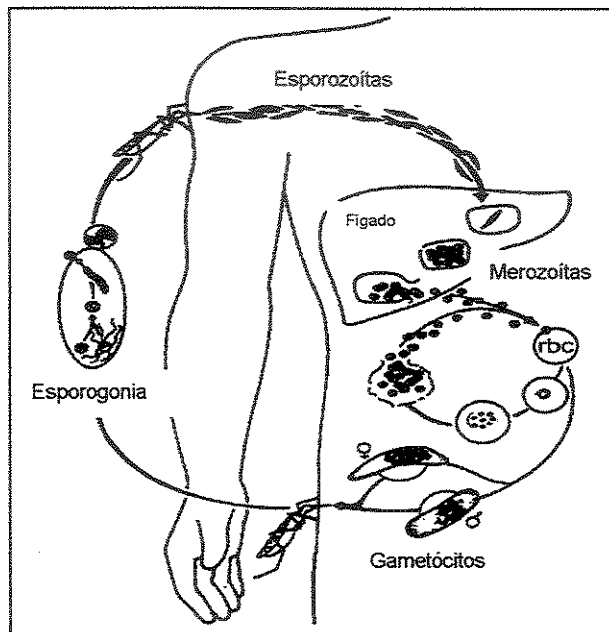


Figura 3: Ciclo evolutivo do *Plasmodium* sp

O mecanismo de invasão dos hepatócitos é complexo, havendo a participação de proteínas de superfície como a proteína SSP-2 (*Sporozoite Surface Protein-2*) dos esporozoítas, que desencadeiam mecanismos de interação molecular e modificações celulares justificando a interiorização do esporozoíta na célula hospedeira (Ferreira, 1999). Após invadir os hepatócitos, os esporozoítas diferenciam-se em formas criptozoítas, que se multiplicam por intermédio de um processo assexuado de reprodução chamado de esquizogonia resultando os

esquizontes teciduais e posteriormente as inúmeras formas merozoítas, determinando assim o rompimento dos hepatócitos e que mais tarde ganharão a circulação para invadir as hemácias. Esta primeira etapa do ciclo endógeno no homem é denominada de etapa exo-eritrocítica ou pré-eritrocítica, pois não ocorreu ainda a invasão da célula vermelha do sangue (Rey, 1991).

A segunda etapa do ciclo endógeno no homem (ciclo eritrocítico) inicia-se com a invasão das hemácias por formas merozoítas que se diferenciam em formas trofozoítas (forma de anel). O mecanismo de invasão da hemácia também é complexo e pode ser compreendido por intermédio de sinais e interações bioquímicas que justificam o reconhecimento específico entre as proteínas de superfície dos merozoítas e os receptores de membrana nas células vermelhas (Murray *et al*, 2002). O principal receptor de membrana na hemácia envolvido no processo de invasão corresponde as glicoforinas que são glicoproteínas transmembrana. Existem glicoforinas do tipo A, B e C, sendo que a glicoforina A é a principal glicoproteína transmembrana e possui aproximadamente 90% do ácido siálico da membrana do eritrócito. A glicoforina A contém sítios de ligação para as proteínas de superfície das formas merozoítas (Murray *et al*, 2002; Lorenzi, 1999). Considera-se atualmente, que a principal proteína de superfície dos merozoítas é a MSP-1 (*Merozoite Surface Protein-1*) (Ferreira, 1999).

No interior das hemácias, as formas trofozoítas crescem e se desenvolvem, originando o esquizonte sanguíneo que finalmente ao atingir seu estágio maduro, irá resultar inúmeras formas merozoítas, novamente pelo processo multiplicativo assexuado chamado esquizogonia e que serão liberadas no plasma sanguíneo após a lise da hemácia, havendo então a possibilidade destas formas parasitar novas hemácias (Rey, 1991).

Os trofozoítas e os esquizontes sanguíneos utilizam um complexo aparato bioquímico da célula hospedeira. O crescimento dos trofozoítas e dos esquizontes se processa às custas da célula hospedeira utilizando suas biomoléculas para obter metabólitos de pequena massa molecular necessários para o seu metabolismo construtivo. Vários componentes metabólicos são utilizados pelos parasitas durante o estágio intra-eritrocítico, sendo que a hemoglobina é

incorporada pelo protozoário por intermédio de uma organela chamada citóstoma e posteriormente é degradada no interior do vacúolo digestivo destes parasitas como uma fonte principal de aminoácidos. Outros metabólitos como a biotina, glicose, metionina, ácido paraminobenzóico (PABA), purinas, pirimidinas e fosfato são oriundos do plasma sanguíneo (Rey, 1991).

Indiscutivelmente, a compreensão de rotas bioquímicas e a interconversão destas biomoléculas é muito importante, pois através de mecanismos bioquímicos, o parasita retira nutrientes da hemácia para realizar o seu catabolismo como fonte de obtenção de moléculas necessárias às suas necessidades nutricionais e como fonte de energia (Wiser, 1999). A co-evolução do hospedeiro e do parasita ao longo dos tempos determinou, por consequência, o aparecimento de novas rotas bioquímicas únicas que distinguem a célula do parasita da célula hospedeira. A compreensão destas rotas pode levar à exploração de diferenças entre o parasita e o hospedeiro e por final no desenvolvimento de novos agentes terapêuticos, principalmente agentes que atuam inibindo enzimas do parasita, biocatalisadores necessários e fundamentais ao seu metabolismo.

Mais tarde, após fases sucessivas de esquizogonia, algumas formas merozoítas sofrem um processo de diferenciação no interior das hemácias, originando os gametócitos, estruturas que não se multiplicam mais no interior das hemácias. Os gametócitos permanecem no interior das hemácias até serem ingeridos pelo mosquito do gênero *Anopheles*. No interior do mosquito os gametócitos rompem a membrana das hemácias através de movimentos de contração e expansão. No intestino médio do mosquito, ocorre o processo de gametogênese onde o gametócito feminino se transforma em macrogameta e o gametócito masculino se transforma em microgametas, sendo que o macrogameta será fecundado por um microgameta originando o zigoto (processo sexuado). Após 24 horas da fecundação o zigoto que agora é denominado oocineto começa a se movimentar por mecanismos de contração do corpo, atingindo a parede do intestino médio do mosquito e encistando então, no tecido epitelial, sendo neste caso chamado de oocisto (oocineto encistado). É nesta fase em que presenciamos o processo de multiplicação esporogônica, onde inúmeros

esporozoítas são originados no oocisto por intermédio de meiose do zigoto e sucessivas divisões mitóticas e que posteriormente escapam gradualmente pela parede do oocisto através de pequenas perfurações (Sinden, 1974). Estes esporozoítas se disseminam por todo o corpo do mosquito atingindo as glândulas salivares, esperando o momento oportuno para serem inoculados no homem através da picada do mosquito e, finalmente, determinar o início de um novo ciclo evolutivo do *Plasmodium falciparum* (Neves, 2000).

1.7 Bioquímica do *Plasmodium falciparum*

O *Plasmodium* é um parasita intracelular obrigatório de extrema dependência da célula hospedeira, pois utiliza biomoléculas da célula parasitada com a finalidade em obter nutrientes necessários ao seu metabolismo e para a obtenção de energia (Wiser, 1999). Um equilíbrio homeostático do parasita é necessário para manutenção de seu crescimento, desenvolvimento e para que possa se reproduzir originando inúmeros parasitas e assim dar continuidade ao seu ciclo de vida. Deste modo, é necessário compreender as rotas metabólicas nestas espécies parasitas, as interconversões entre biomoléculas e as fases de seu ciclo de vida, objetivando explorar alvos do parasita para ação de novos agentes terapêuticos. A compreensão das rotas envolvidas no ciclo da vida destas espécies parasitas é apenas uma das etapas necessárias no processo de descoberta e desenvolvimento de novos protótipos como agentes antimaláricos potenciais.

Para a manutenção do equilíbrio fisiológico da célula parasita, o *Plasmodium* necessita adquirir biomoléculas do hospedeiro efetuando a decomposição destas em moléculas menores e energia (catabolismo). Estas moléculas menores e a energia obtida são utilizadas na síntese de nutrientes necessários ao metabolismo do parasita (anabolismo). Sejam nos processos catabólicos ou anabólicos, inúmeras enzimas parasitas catalisam as diversas reações bioquímicas durante o ciclo intra-eritrocítico (Murray *et al*, 2002). Em contrapartida, estas enzimas podem ser inibidas por fármacos potenciais (inibidores enzimáticos) em qualquer uma das etapas do metabolismo do parasita,

bloqueando por consequência a reprodução do *Plasmodium* no interior da hemácia.

Durante o desenvolvimento do parasita nota-se um aumento na demanda de nucleotídeos púricos e pirimídicos, aminoácidos, carboidratos, ácidos graxos e outras espécies de biomoléculas. Para que ocorra no interior da hemácia o processo reprodutivo esquizogônico, o conteúdo de ácidos nucleicos no parasita deverá ser aumentado visando a realização de sua divisão nuclear e conseqüente obtenção de novas formas parasitas, necessitando então de nucleotídeos púricos e pirimídicos (Wiser, 1999). Os nucleotídeos púricos são obtidos por intermédio do sangue do hospedeiro, pois os plasmódios não conseguem sintetizar o anel purínico devendo obtê-lo no meio em que se encontram, enquanto que os nucleotídeos pirimídicos são obtidos por síntese a partir da glutamina, bicarbonato e aspartato resultando a uridina-monofosfato (UMP) que por final é convertida em pirimidinas (Rey, 1991). Isto ocorre também, pelo fato da concentração de nucleotídeos púricos no sangue do hospedeiro ser elevada, sendo que a mesma não é evidenciada com os nucleotídeos pirimídicos. Esta limitação no metabolismo de nucleotídeos pirimídicos sugere possibilidade de exploração de rota bioquímica, sendo esta considerada alvo de ação farmacológico de agentes antimalariais (Wiser, 1999).

Alguns agentes antimalariais conhecidos (pirimetamina e sulfonamidas) atuam inibindo a Dihidrofolato redutase (DHFR) (figura 4), enzima encontrada no parasita e no homem, biocatalisador crucial no metabolismo de folatos. O folato é um cofator essencial na síntese de nucleotídeos, aminoácidos, timidina e na catálise de ácido tetraidrofólico (Kalant & Roschlau, 1991). A inibição da DHFR resultaria em bloqueio no metabolismo celular do parasita, pois ocasionaria a inibição da síntese de nucleotídeos, evitando assim a síntese de ácidos nucleicos e proteínas, e conseqüente bloqueio na multiplicação celular do parasita.

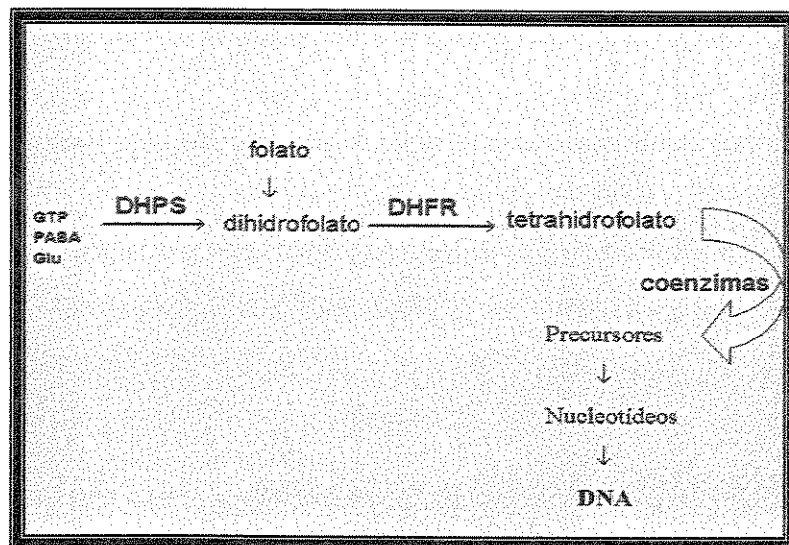


Figura 4: Atividade enzimática das enzimas DHPS e DHFR.

O mais interessante seria obter fármacos que pudessem agir seletivamente, ou seja, fármacos que sejam específicos às enzimas dos parasitas e não às enzimas humanas. Uma das soluções para este problema está intrinsicamente focada não apenas na descoberta e no desenvolvimento de fármacos potencialmente ativos, mas também seletivos contra as enzimas dos parasitas, evitando efeitos adversos e assim provocando uma toxicidade mínima no organismo humano. Em contrapartida, inúmeras proteases malarias já foram identificadas e caracterizadas por intermédio de técnicas imunológicas e bioquímicas (Banerje *et al*, 2002; Rosenthal, 1998).

Durante o crescimento do parasita, o aumento em seu tamanho determina também um aumento na demanda por aminoácidos para que possa manter o mecanismo de síntese protéica. Os aminoácidos podem ser obtidos através de sínteses novas, seqüestro de aminoácidos livres do plasma do hospedeiro e pela degradação da hemoglobina do hospedeiro. A hemoglobina é uma proteína extremamente abundante no eritrócito, correspondendo a aproximadamente 95% do total das proteínas, sendo que em média 60 a 80% da hemoglobina intracelular é degradada pelo parasita (Lorenzi, 1999).

A degradação da hemoglobina que ocorre no eritrócito do hospedeiro é de fundamental importância para o metabolismo do parasita e tal processo pode ser

evidenciado através de algumas análises observadas experimentalmente (McKerrow *et al*, 1993):

- O teor de hemoglobina em eritrócitos infectados diminui sensivelmente durante o ciclo de vida dos parasitas;
- Ao infectar eritrócitos com hemoglobinas radiomarcadas notou-se posteriormente o aparecimento de aminoácidos marcados em proteínas do parasita;
- A concentração de aminoácidos livres em eritrócitos infectados é maior que em hemácias não infectadas.

A hemoglobina e as biomoléculas necessárias ao metabolismo do parasita são interiorizados pelo plasmódio por micropinocitose ou através de uma organela especial chamada de citóstoma, onde pequenas vesículas contendo o pigmento hemoglobina serão originadas e que posteriormente se fundirão com o vacúolo digestivo do parasita (figura 5) (Eggleson *et al*, 1999).

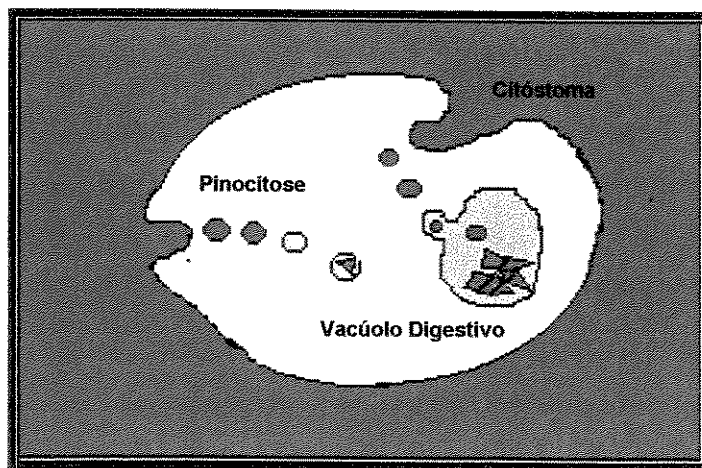


Figura 5: Endocitose do citosol da hemácia.

O vacúolo digestivo dos plasmódios é um compartimento ácido semelhante aos lisossomos das células humanas com pH na faixa de 5,0 - 5,4 onde ocorre a atividade proteolítica da hemoglobina através de enzimas denominadas de proteases. Várias atividades de proteases foram identificadas no vacúolo digestivo dos plasmódios. Proteases são enzimas que promovem a clivagem de ligações

peptídicas (amídicas) das proteínas resultando numa desorganização e desestabilização da estrutura tridimensional protéica.

Proteases que catalisam a clivagem de uma ligação peptídica na porção interna da cadeia de uma proteína são às vezes caracterizadas como proteinases ou endopeptidases. Proteases que clivam aminoácidos em região amino-terminal ou carboxila-terminal de um peptídeo ou proteína são referidas com exopeptidases (figura 6). Estas denominações em proteases são importantes, pois indicam se a clivagem da ligação peptídica irá ocorrer nas extremidades da cadeia polipeptídica ou em região interna da cadeia e deste modo dão indícios da função da protease (Barrett, 1986).

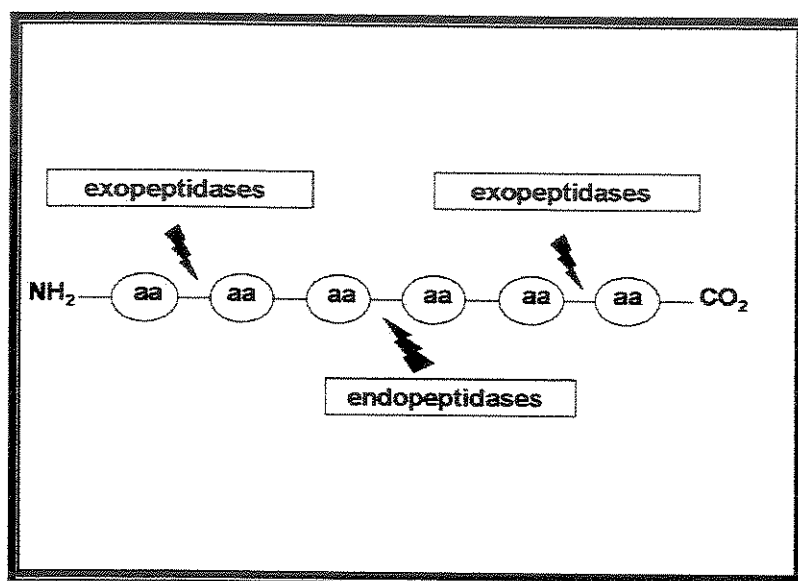


Figura 6: Tipos de clivagem executada pelas proteases.

A cinética das proteases e os aspectos de interações químicas entre as proteases e o substrato específico explicitam a capacidade desta enzima em interagir com o respectivo substrato, sendo um indicativo da formação do estado de transição (complexo substrato-enzima). A formação do estado de transição ocorre por intermédio de condições geométricas e energéticas adequadas (Marzzoco & Torres, 1999).

É imprescindível que se tenha um conhecimento bioquímico a respeito da estrutura química tridimensional e do mecanismo catalítico das proteases para melhor entender a cinética enzimática destas, e então, o mecanismo de inibição enzimático exercido por alguns fármacos que atuam como inibidores de proteases.

Avanços recentes na bioquímica, biologia molecular e na bioinformática estão permitindo caracterizar e identificar cada vez mais inúmeras proteases em vários microorganismos parasitas focando uma atenção maior nestas biomacromoléculas específicas dos parasitas, minimizando deste modo, vários efeitos adversos e tóxicos em células humanas.

As proteases maláricas serão a chave futura para o alvo de novos agentes terapêuticos, visto que, são enzimas específicas encontradas nas diversas espécies de parasitas da malária (Tamta & Mukhopadhyay, 2003). Estas enzimas estão intrinsecamente associadas a diversos processos metabólicos do parasita, desde a invasão de células vermelhas, replicação de seu material genético, ruptura da membrana dos eritrócitos por espécies maduras, eventos relacionados com a hidrólise de proteínas do citoesqueleto do eritrócito, proteólise da hemoglobina e muitos outros.

Está claro que o estudo de fármacos que atuam em proteases de microorganismos será alvo de pesquisa por muitos anos no intuito em bloquear rotas metabólicas dos parasitas.

1.8 Proteases: enzimas necessárias no metabolismo do *Plasmodium falciparum*

Embora a relação, parasita-hospedeiro no ciclo evolutivo da malária seja muito complexa, a atribuição de funções a proteases maláricas é de fundamental importância na compreensão dos mecanismos envolvidos e no desenvolvimento de novos agentes inibidores enzimáticos (Rosenthal *et al*, 2002; Rosenthal, 1998).

Durante os estágios do ciclo de vida do *Plasmodium falciparum*, evidencia-se a participação de proteases maláricas. Atualmente, vários pesquisadores envolvidos no desenvolvimento de antimaláricos, estão preferindo focar os estágios de invasão do eritrócito pelo parasita e o processo de degradação da

hemoglobina da célula hospedeira como alvos promissores (Sabnis *et al*, 2003; Shenai *et al*, 2002; Hanspal *et al*, 2002; Humphreys *et al*, 1999).

1.8.1 Proteases mediando invasão e ruptura do eritrócito

A invasão da célula vermelha é explicada pelo acoplamento (figura 7) da forma merozoíta do *Plasmodium* aos receptores de membrana do eritrócito e através do alinhamento de organelas do parasita à membrana da célula vermelha (Mckerrow *et al*, 1993). O citoesqueleto da célula vermelha mantém uma morfo-funcionalidade importante para a estrutura da membrana plasmática e atua como uma barreira importante contra a invasão de células hospedeiras. Pesquisas recentes sugestionam que a proteólise do citoesqueleto da hemácia parasitada está relacionada à presença de determinadas proteases maláricas (Debrabant & Delplace, 1989), o que também indica o papel de proteases no mecanismo de invasão e da lise do eritrócito.

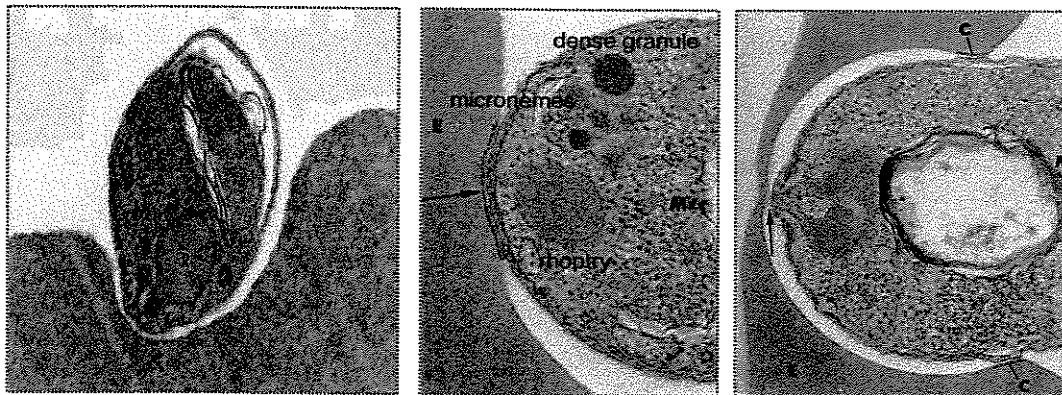


Figura 7: Microfotografia eletrônica da invasão do eritrócito pelo merozoíta;

- A) Reorientação e deformação da hemácia;
- B) Junção *Plasmodium falciparum* - hemácia;
- C) Entrada do *Plasmodium falciparum*; Fonte: www.tulane.edu/%7Ewiser/

Diferentes proteases foram identificadas como enzimas potencialmente envolvidas na invasão e ruptura dos eritrócitos. Uma protease cisteína de 68kDa foi identificada em esquizontes e em formas merozoítas do *Plasmodium falciparum* (Grellier *et al*, 1989; Bernard & Schrevel, 1987; Mayer *et al*, 1991; Wernsdorfer,

1988). Pesquisas com merozoítas levaram à conclusão que esta protease está localizada no ápice do parasita sendo liberada por uma organela chamada roptria durante o processo de invasão (Bernard & Schrevel, 1987). Uma protease serina foi identificada em formas merozoítas de *Plasmodium falciparum* (Rosenthal *et al*, 1998).

Uma outra protease serina de 76kDa foi também identificada em esquizontes e em formas merozoítas de *Plasmodium falciparum* ligada à membrana numa forma inativa ancorada por *glicosil-fosfatidilinositol*. A ativação dessa protease pelo *fosfatidilinositol-específico fosfolipase C*, sugere a importância desta protease no processo de invasão da célula vermelha (Braun-Breton *et al*, 1988).

Diversos inibidores de proteases de *Plasmodium* estão sendo pesquisados. Sabe-se que o Chymostatin (inibidor da protease serina), bloqueou especificamente a invasão dos eritrócitos pelos merozoítas, o mesmo não acontecendo com o Leupeptin (inibidor de protease cisteína), sugerindo então o papel das proteases serina com atividade fundamental no processo de invasão das células vermelhas (Braun-Breton *et al*, 1992).

Estruturas aproximadas de fármacos inibidores de proteases podem ser projetadas para atuarem como agentes anti-maláricos, levando em consideração a estrutura de modelos de proteases já identificadas. Embora a função específica de diversas categorias de proteases ainda não esteja clara, fica evidente que a participação de inibidores de proteases serina e cisteína bloqueiam de modo consistente a ruptura e a invasão das células vermelhas do sangue do hospedeiro (McKerrow *et al*, 1993).

1.8.2 Proteases mediando degradação da hemoglobina

Extensivas evidências sugerem a ocorrência da degradação da hemoglobina do hospedeiro no vacúolo digestivo do parasita (Klemba & Goldberg, 2002). A degradação da hemoglobina é um dos mecanismos que o parasita possui para requerer aminoácidos para o seu crescimento e desenvolvimento no interior das células vermelhas. Isto pode ser justificado pelo fato de tais parasitas

apresentarem uma capacidade limitada na síntese de aminoácidos livres, e reside também na circunstância da concentração de aminoácidos livres no citoplasma do eritrócito ser reduzida, ou seja, insuficiente para a necessidade do *Plasmodium* (McKerrow *et al*, 1993).

A proteólise da hemoglobina ocorre por intermédio de processos ordenados com a atuação sequencial de diferentes proteases. Dentre as diversas atividades de proteases envolvidas no processo de degradação da hemoglobina, três categorias principais já foram identificadas no vacúolo digestivo dos trofozoítas do *Plasmodium falciparum* (Klemba & Goldberg, 2002):

- Plasmepsins – proteases aspárticas (ácidas);
- Falcipains – proteases cisteínicas (tióis);
- Falcilysins – metalopeptidases.

O mecanismo de degradação da hemoglobina envolve a participação de plasmepsin-1 e plasmepsin-2 que são proteases capazes de clivar ligações peptídicas da hemoglobina não desnaturada (nativa) entre os resíduos de aminoácidos fenilalanina e leucina localizados nas posições 33 e 34 da cadeia polipeptídica de α -globina, resultando em um desdobramento da hemoglobina em subunidades de globina e grupo heme (Klemba & Goldberg, 2002). Estes resíduos estão localizados em regiões conhecidas como região de dobramento da proteína em domínios conservados. A clivagem destas ligações promove a desestabilização da estrutura global da hemoglobina, expondo outros sítios que estão localizados no interior das cadeias polipeptídicas de globina à ação de proteases.

Posteriormente, nota-se a participação de plasmepsins e falcipains, proteases que degradam fragmentos grandes de globina em polipeptídeos médios (até 20 aminoácidos). Finalmente, estes polipeptídeos médios serão clivados em fragmentos pequenos de aminoácidos (6 a 8 aminoácidos) pelas falcilysins. Portanto, a digestão da hemoglobina ocorre por intermédio de uma sequência semi-ordenada (figura 8) que vai desde a degradação inicial da hemoglobina em grandes fragmentos de cadeias polipeptídicas seguido por subseqüentes degradações em peptídeos menores. Deste modo, a ação combinada das

enzimas plasmepsins, falcipains e falcilysins origina pequenos peptídeos que mais tarde serão bombeados do vacúolo digestivo do parasita para o seu citoplasma através de proteínas carreadoras *Pfmdr-1* (dependente de ATP), existentes na membrana do vacúolo digestivo. No citoplasma do *Plasmodium*, estes peptídeos menores serão degradados em aminoácidos livres através de aminopeptidases (Wiser, 1999).

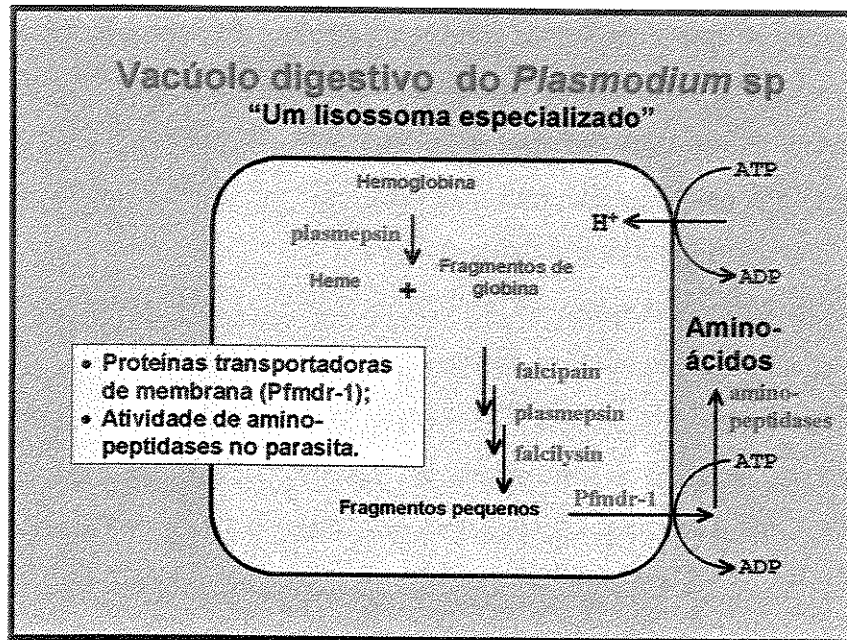


Figura 8: Degradação semi-ordenada da hemoglobina no vacúolo digestivo do *Plasmodium* sp; Fonte: www.tulane.edu/%7Ewiser/

1.9 Malária: doença com severas complicações

O indivíduo infectado apresenta um quadro típico de sintomas. Inicialmente apresenta mal-estar, cefaléia e indisposição (sintomas prodrômicos). Em seguida, nota-se a ocorrência de febre periódica elevada dando origem ao chamado acesso malárico (Neves, 2000). O acesso malárico é caracterizado primeiramente por calafrios, seguido de sensação de calor e finaliza com a sudorese. Estas manifestações clínicas coincidem com a ruptura das hemácias. Os parasitas provocam uma série de modificações fisiológicas e morfológicas nas células vermelhas resultando em aglutinação e lise (figura 9), liberando assim os

parasitas, diversos metabólitos e toxinas na circulação induzindo a uma resposta do hospedeiro (Rey, 1991).

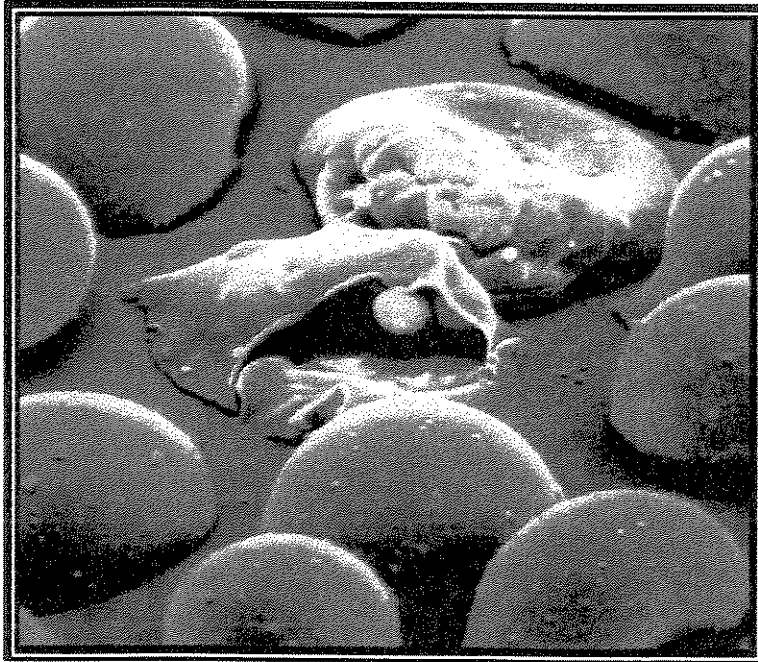


Figura 9: Células de *Plasmodium falciparum* lisando eritrócitos. Microscopia eletrônica de varredura, artificialmente colorida; Fonte: <http://bioinformatics.leeds.ac.uk/>

Uma característica biológica interessante presenciada em indivíduos parasitados pelo *Plasmodium falciparum* é a ocorrência de aderência dos eritrócitos parasitados ao endotélio de capilares e vênulas (figura 10). Este mecanismo denominado citoaderência endotelial é mediado por proteínas do parasita (adesinas) expressas na superfície dos eritrócitos infectados. A citoaderência endotelial compromete o fluxo sanguíneo obstruindo os capilares de órgãos como cérebro, rins, fígado e outros órgãos afetados (Cimerman & Cimerman, 1999).

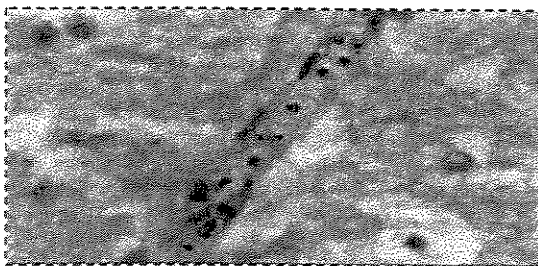


Figura 10: Citoaderência endotelial; Fonte: www.tulane.edu/~%7Ewiser/

É importante ressaltar que a diminuição do fluxo sanguíneo (isquemia) provocada pela citoaderência endotelial reduz o fluxo de oxigênio (anóxia) nos tecidos induzindo a um metabolismo anaeróbio tecidual, pois a limitação no suprimento de oxigênio determina uma depleção na síntese de ATP por falência no transporte de elétrons na cadeia respiratória mitocondrial. Deste modo haverá um nítido aumento na concentração de íons cálcio no citoplasma levando a uma perda da integridade da membrana plasmática e da função mitocondrial, que finalmente resulta em necrose coagulativa (Rubin & Farber, 1990).

Em decorrência da hemólise, ocorre liberação dos parasitas e de seus metabólitos, estimulando os macrófagos à ação fagocitária. A hemólise, também induz à ocorrência de anemia pela destruição dos eritrócitos parasitados e não parasitados pelos macrófagos, e pela destruição destes por mecanismo imune e sequestro destas células no baço (Roitt *et al*, 1999).

Enfim, a infecção pelo *Plasmodium falciparum* resulta em um quadro clínico muito mais complexo (febre alta, hemólise maciça, danos cerebrais, falência renal, hepatoesplenomegalia, anemia intensa, etc) podendo resultar em morte. Deve-se levar em consideração que a doença pode evoluir apresentando variações clínicas como malária cerebral. Neste caso, o indivíduo pode chegar ao coma profundo.

A malária é potencialmente grave e fatal para crianças em áreas endêmicas nos primeiros cinco anos de vida. A malária pode representar uma complicação muito séria na gravidez, causando a morte materna, aborto, morte fetal ou prematuridade devido ocorrência do desenvolvimento e seqüestramento dos parasitas na placenta, levando este órgão à falência e conseqüente sofrimento fetal (Rey, 1991).

1.10 Aspectos imunológicos e mecanismo de escape do *Plasmodium*

Ao longo de vários anos de evolução, os parasitas desenvolveram mecanismos imunológicos na tentativa de se adaptar ao hospedeiro, pois não seria interessante matá-lo sem que a doença seja transmitida a outro hospedeiro (Roitt *et al*, 1999). A capacidade desses microorganismos em sobreviver no

interior dos hospedeiros reflete as adaptações evolutivas que lhes permitiram resistir aos mecanismos imunológicos (Abbas *et al*, 1995).

A presença de antígenos parasitários circulantes no organismo do hospedeiro, desencadeia um mecanismo de resposta imunológica:

a) Os linfócitos B fazem o reconhecimento específico de antígenos solúveis parasitários e se diferenciam em plasmócitos, que ativamente, secretam anticorpos (imunoglobulinas);

b) Os linfócitos T-auxiliares reconhecem antígenos parasitários nas superfícies de células acessórias e secretam hormônios denominados citocinas ($\text{IFN}\gamma$, interleucinas e $\text{TNF}\alpha$), que por sua vez ativam os linfócitos B e os macrófagos teciduais (Abbas *et al*, 1995).

Os macrófagos podem ser também ativados por mecanismos T-independentes:

a) Macrófagos produzem interleucina-12 (IL-12) que estimulam as células NK (Natural Killer) a secretarem Interferon gama ($\text{IFN}\gamma$);

b) Macrófagos secretam o Fator de Necrose Tumoral alfa ($\text{TNF}\alpha$) em resposta aos antígenos fosfolipídicos do *Plasmodium*, que por sua vez ativa outros macrófagos.

Os principais mecanismos no qual os anticorpos controlam as infecções pelo *Plasmodium* são (Roitt *et al*, 1999):

a) As imunoglobulinas produzidas pelos plasmócitos ativam a via clássica do complemento provocando dano direto no parasita;

b) As imunoglobulinas neutralizam diretamente as formas merozoítas pelo bloqueio de seu sítio de ligação, evitando assim a fixação do merozoíta com o receptor de superfície na hemácia;

c) As imunoglobulinas potencializam a ação fagocitária mediada por macrófagos.

As formas parasitas da malária desenvolveram mecanismos de escape das respostas imunológicas do hospedeiro. Os mecanismos de evasão verificados são:

- a) Distração imunológica – Os antígenos solúveis do *Plasmodium falciparum* (antígenos S ou termoestáveis) diluem os anticorpos circulantes agindo com uma “cortina de fumaça” e assim prejudicando a resposta do hospedeiro;
- b) Variação antigênica - Alteração dos antígenos de seu revestimento externo como forma de se proteger do hospedeiro (Roitt *et al*, 1999).

1.11 Toxicidade do grupo Heme

Durante o processo de degradação da hemoglobina ocorre liberação do grupo Heme e fragmentos de globina. Heme livre é tóxico ao parasita, determina a desestabilização de membranas e inibe a atividade de várias enzimas. Durante o catabolismo da hemoglobina, ocorre alteração no NOX do elemento ferro, ou seja, nota-se evidência de um processo de oxi-redução. O ferro da hemoglobina se encontra sob o estágio de íon ferroso (Fe^{+2}) e após a ocorrência da proteólise a forma resultante corresponde ao íon férrico (Fe^{+3}). Podemos então afirmar que durante o catabolismo da Hemoglobina que ocorre no vacúolo digestivo do *Plasmodium falciparum* mediante mecanismo de oxi-redução, haverá transferência de elétrons que participarão na formação de reagentes intermediários.

A oxidação do ferro promove a formação de *reagentes intermediários oxigenados* (ROI), que são considerados como a principal fonte de Stress oxidativo do parasita. Os intermediários reativos formados (ânion superóxido, radical hidroxila, peróxido de hidrogênio e oxigênio), são capazes de causar danos celulares ao parasita. Como mecanismo de defesa à formação de ROI, o parasita converte os intermediários oxigenados em outras espécies químicas menos tóxicas.

No vacúolo digestivo do *Plasmodium falciparum*, o grupo Heme proveniente da hemoglobina é convertido através de Heme polimerases em pigmento malarial (Hemozoína). Ocorre primeiramente a dimerização do grupo Heme em β -Hematina (Pagola, 2000) (figura 11), que posteriormente será polimerizada em massas insolúveis de pigmento malarial (Hemozoína) visível ao microscópio. Os dímeros β -Hematina interagem por intermédio de ligações de hidrogênio, formando cristais insolúveis do pigmento Hemozoína (figura 12). Este mecanismo

de formação e deposição do pigmento é denominado biocristalização ou biomineralização (Hempelmann & Egan, 2002; Mann, 1989). O mecanismo de biocristalização não está ainda muito esclarecido, mas sabe-se que enzimas ricas em histidinas podem participar no processo, facilitando a biocristalização (Sullivan, 1996; Sullivan, 2002; Choi, 1999).

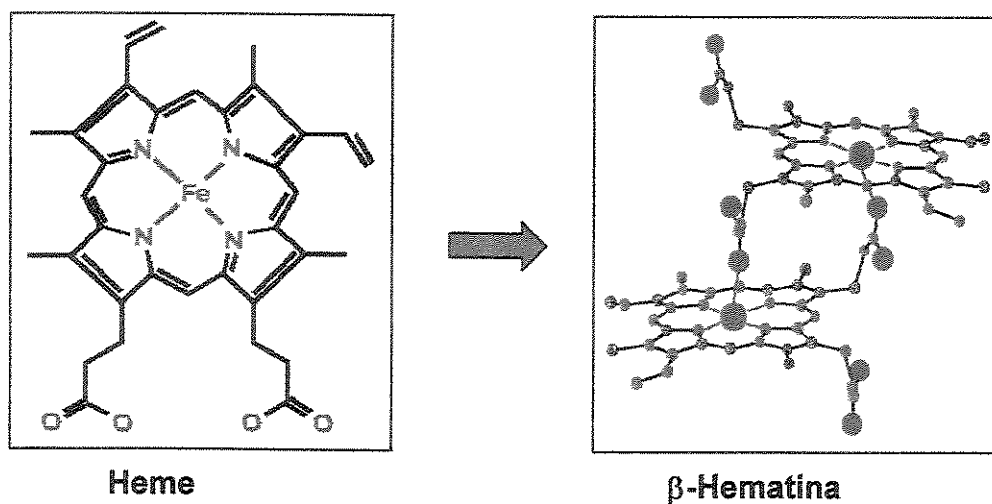


Figura 11: Dimerização do grupo Heme em β -Hematina.

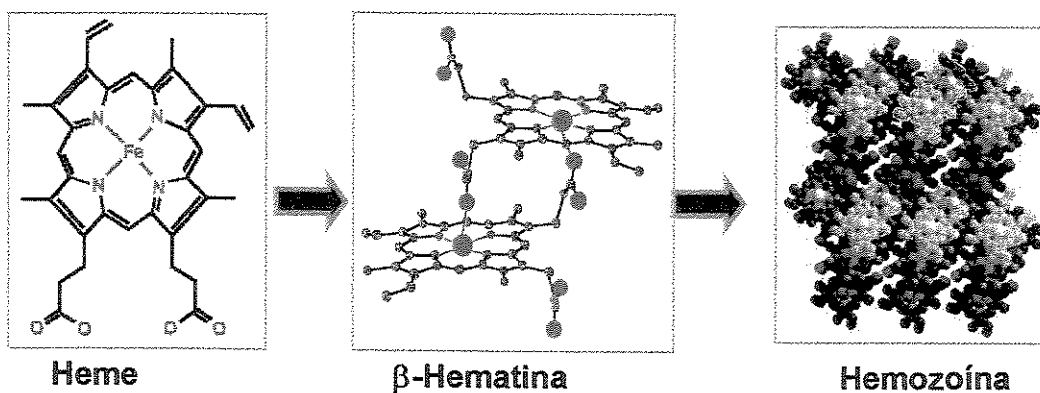


Figura 12: Biocristalização da Hemozoína.

Alguns quimioterápicos (cloroquina e 4-aminoquinolinas) atuam inibindo a conversão do grupo Heme em Hemozoína (Ginsburg *et al*, 1999; Egan *et al*, 1999), aumentando de forma considerável os níveis de Heme no Vacúolo digestivo, determinando assim a morte do parasita.

Com relação ao mecanismo REDOX, o intermediário superóxido produzido durante a oxidação do ferroso em férrico é convertido pela superóxido dismutase no interior do vacúolo do parasita em peróxido de hidrogênio. Posteriormente, o peróxido de hidrogênio é degradado em água e oxigênio pela enzima catalase.

1.12 Quimioterapia da malária

Mesmo diante de elevado grau de desenvolvimento científico e tecnológico alcançado pela humanidade, os países tropicais e subtropicais não conseguiram, ainda, controlar a malária.

Uma quimioterapia adequada durante o tratamento da malária pode reduzir a sintomatologia e evitar a ocorrência de complicações e até a morte do hospedeiro, permitindo deste modo um maior controle da doença.

O sucesso da quimioterapia depende em grande parte da exploração das diferenças metabólicas existentes entre o parasita e o hospedeiro. Entretanto, as autoridades de saúde enfrentam o grande problema da resistência às drogas pelos parasitas dessa doença (Rang & Dale, 1993).

Desde 1960, a transmissão da malária vem aumentando em regiões endêmicas, notando-se uma aumento considerável em cepas de *Plasmodium falciparum* resistentes à cloroquina e a outros agentes quimioterápicos. As drogas utilizadas atualmente no tratamento da malária estão ficando cada vez mais obsoletas, o grau de resistência ao medicamento está aumentando, constatando-se um relativo aumento nos índices de resistência para espécies do gênero *Plasmodium* que não eram resistentes a cloroquina (*Plasmodium vivax*) (Segurado *et al*, 1997).

A maioria dos agentes anti-maláricos possui como mecanismo farmacológico, ação contra as formas eritrocíticas dos parasitas. Os agentes de maior eficácia contra os parasitas da malária são: cloroquina, quinidina, mefloquina e a halofantrina. Outros agentes como a pirimetamina, sulfonamidas, sulfonas e as tetraciclina, são agentes terapêuticos de ação mais lenta e geralmente administrados sob forma combinada.

Uma lactona sesquiterpênica (artemisinin) extraída de *Artemisia annua* L. conhecida dos chineses a anos e isolada em 1972, vem sendo utilizada contra cepas resistentes de *Plasmodium falciparum* (Silva, 1994; Rang & Dale, 1993; Rosenthal, 1998). O artemisinin e seus derivados são endoperóxidos que apresentam ação rápida contra as formas sanguíneas.

1.12.1 Classificação dos antimaláricos

As principais drogas antimaláricas podem ser classificadas levando em consideração a sua estrutura química ou o seu alvo de ação no ciclo de vida do parasita (FUNASA – manual de terapêutica da malária, 2002):

Quanto ao grupo químico:	Quanto ao alvo de ação:
• Arilaminoálcoois - quinina, mefloquina, halofantrina; • 4-aminoquinolinas - cloroquina, amodiaquina; • 8-aminoquinolinas -primaquina; • Diaminopirimidinas - pirimetamina, trimetoprima; • Biguanidas e triazinas - proguanila, cicloguanila; • Sulfonamidas e sulfonas - sulfadiazina, sulfametoxazol, sulfadoxina, dapsona • Naftoquinonas - atovaquona; • Peróxido de lactona sesquiterpênica derivados do artemisinin; • Antibióticos - tetraciclina, clindamicina, doxiciclina;	• Esquizonticida tecidual ou agente profilático causal - atua sobre as formas assexuadas do parasita nos tecidos; • Esquizonticida sanguíneo ou agente supressor - atua nas formas assexuadas dos parasitas na corrente sanguínea; • Gametocitocida -atua sobre as formas sexuadas (gametócitos), bloqueando a transmissão; • Esporozoitocida - atua sobre as formas infectantes (esporozoítas) - nenhum fármaco disponível no mercado;

1.12.2 Mecanismo de ação de antimaláricos

A descoberta de técnicas para manutenção contínua de *Plasmodium falciparum* humano *in vitro* (Trager & Jensen, 1976) levou a ensaios práticos de sensibilidade destes parasitas aos agentes anti-maláricos. Este importante avanço ainda vem sendo explorado em termos de atividade biológica com o objetivo em compreender os mecanismos de ação e a resistência dos parasitas frente aos agentes quimioterápicos.

Os fármacos antimaláricos agem especificamente interferindo no metabolismo do parasita, apresentando então atividade biológica nos chamados *targets* (alvos). O desenho racional de drogas visa a exploração de rotas metabólicas do parasita objetivando desenvolver fármacos com toxicidade seletiva, ou seja, drogas que atuam especificamente em receptores ou interferindo em vias metabólicas do parasita causando uma toxicidade mínima no ser humano. O planejamento racional de um anti-malárico deve buscar primeiramente uma toxicidade seletiva ao *Plasmodium falciparum* e toxicidade mínima ao hospedeiro. Os principais mecanismos de ação dos fármacos antimaláricos estão relacionados com:

- A) Inibição no processo de formação da Hemozoína;
- B) Ação no metabolismo de folatos;
- C) Mecanismo REDOX.

A) Ação no vacúolo digestivo do parasita:

Alguns fármacos atuam inibindo a conversão do grupo Heme em pigmento malarial Hemozoína, ocorrendo um acúmulo deste grupo no vacúolo digestivo do *Plasmodium*, e assim ocasionando a lise de membranas do parasita. Neste tipo de ação biológica, ocorre inibição da biocristalização pela interferência na rota de conversão do grupo Heme em Hemozoína.

Como exemplo de inibição da biocristalização temos a droga cloroquina que em meio ácido do vacúolo digestivo do parasita, sofre protonação originando uma estrutura química que interage com a enzima Heme polimerase, inibindo a conversão do grupo Heme em Hemozoína, e conseqüentemente, aumenta os

níveis de Heme livre no vacúolo (Ginsburg *et al*, 1999; Egan *et al*, 1999). Não podemos esquecer que a estrutura protonada da cloroquina realiza mecanismo de complexação com a ferriprotoporfirina IX, um metabólito oriundo da degradação da hemoglobina. Este complexo formado, lesa as membranas e acarreta lise tanto do parasita como do eritrócito (Harvey *et al*, 1998).

B) Ação no metabolismo de folatos

O folato é necessário a biossíntese de aminoácidos e dos precursores do DNA e RNA tanto para o *Plasmodium* quanto para o homem. O homem obtém o ácido fólico a partir da alimentação e não precisa sintetizá-lo, enquanto que os parasitas necessitam desta síntese (Kalant & Roschlau, 1991).

A incapacidade do parasita em utilizar folato externo torna a biossíntese do folato um bom alvo para drogas. O folato é sintetizado a partir de três componentes básicos: GTP, ácido p-aminobenzoico (PABA) e glutamato, e com participação de enzimas das quais podemos destacar a ação da DHPS e da DHFR. Fármacos que atuam inibindo a síntese de folatos são denominados agentes anti-folatos.

A dihidropterato sintetase (DHPS) participa catalisando a síntese do dihidrofolato, agente precursor de folatos (figura 13). A sulfadoxina e o dapsona são exemplos de dois agentes antimaláricos comuns cujo alvo é a DHPS.

A dihidrofolato redutase (DHFR) é uma enzima de suma importância que reduz o dihidrofolato (FH_2) a tetrahidrofolato (FH_4), que por sua vez funciona como um cofator específico em uma das etapas da síntese de DNA (figura 13). A inibição da DHFR determina bloqueio na síntese de DNA, evitando então o crescimento e replicação do parasita conduzindo-o à morte. Pirimetamina e proguanil são exemplos de fármacos que atuam inibindo DHFR.

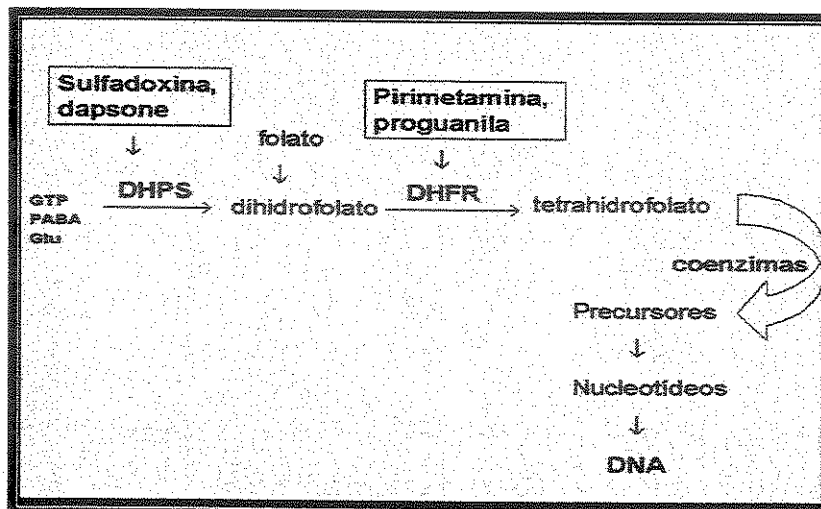


Figura 13: Mecanismo de ação de anti-folatos.

C) Drogas que atuam no mecanismo REDOX

O mecanismo de oxi-redução ocorrido na degradação da hemoglobina do hospedeiro no interior do parasita acarreta a produção de *reagentes intermediários oxigenados* (ROI), conduzindo o protozoário a um *stress oxidativo*, vindo a prejudicar componentes celulares, tais como lipídeos, proteínas e ácidos nucléicos.

Algumas drogas atuam aumentando especificamente o *stress oxidativo* do parasita conduzindo-o à morte. O nível de *stress oxidativo* pode ser aumentado por drogas que atuam como oxidantes diretos.

Alguns fármacos que atuam, via mecanismo REDOX precisam ser ativados para que sejam efetivos contra seus alvos, como é o caso do metronidazol e outros nitroimidazóis. Essas drogas são ativadas por uma redução do grupo nitro em um radical ânion que é altamente reativo e interage com proteínas e DNA, conduzindo essas moléculas à inatividade de suas funções e conseqüentemente levando à morte celular (Wiser, 2003).

1.12.3 Resistência aos antimaláricos

O aumento da incidência de cepas resistentes de *Plasmodium* representa o maior obstáculo à quimioterapia antimalárica. Os parasitas da malária apresentam mecanismos de escape no sentido de mascarar a ação imunológica do hospedeiro

e assim dificultando a defesa natural contra estes agentes. O mecanismo da resistência pode ser explicado por inúmeros processos seja de procedência natural ou adquirida. A resistência natural de certas espécies é decorrente de características inerente ao parasita e que já existem antes da exposição ao medicamento. A resistência adquirida é o mecanismo que surge quando cepas resistentes emergem de populações sensíveis geralmente pela exposição a medicamentos.

O mecanismo de resistência às drogas pode ser explicado por alterações genéticas nas cepas do parasita, onde geralmente envolve mutações no alvo da droga evitando interações da droga com os alvos receptores. A resistência também pode ser desenvolvida por acúmulo de mutações em um mesmo alvo ou em alvos diferentes.

A diminuição no acúmulo de droga ou metabolização da droga em produtos não-tóxicos resulta em menor interação droga-receptor, contribuindo também para aumento da resistência à droga. A classificação moderna das drogas antimaláricas baseia-se no estágio do ciclo de vida do *Plasmodium* sobre o qual atuam. A cloroquina é a droga de escolha para o tratamento da malária, exceto onde ocorrem cepas resistentes a este fármaco.

A malária provocada pelo *Plasmodium falciparum* resistente a Cloroquina vem sendo encontrada com frequência crescente na América do Sul e Central, Ásia, Sudeste da Ásia, Oceania e na África. A resistência a Cloroquina está associada à diminuição do acúmulo da droga no vacúolo digestivo do *Plasmodium*. Nestes casos, exige-se uma quimioterapia associativa de fármacos (pirimetamina/sulfadoxina), no entanto já existem relatos recentes que documentam uma resistência crescente de *Plasmodium falciparum* a esta associação (Lüllmann & Mohr, 2204).

O futuro da farmacoterapia no controle da malária é sombrio. O uso substancial de antimaláricos acelerou seletivamente cepas fármaco-resistentes de *Plasmodium falciparum* (figura 14), que é atualmente a espécie responsável pelas formas graves de malária humana.

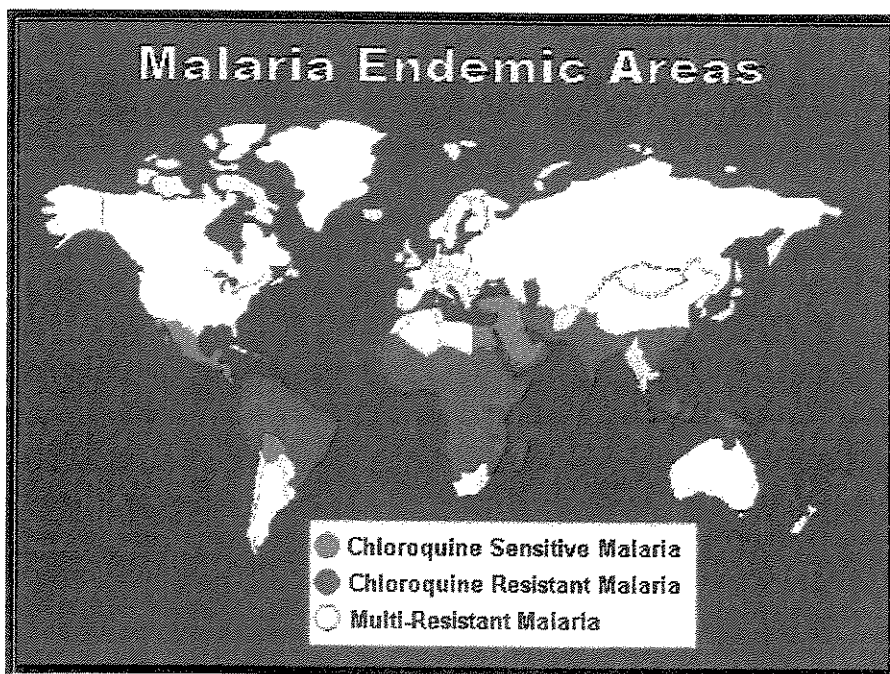


Figura 14: Distribuição da malária e resistência do *Plasmodium falciparum* à cloroquina;
 Fonte: www.traveldoctor.co.uk/malaria.thm

1.13 Chalconas: um *lead compound*

Atualmente o problema da resistência do *Plasmodium falciparum* frente aos fármacos existentes no mercado, tornou-se o mais importante obstáculo ao tratamento. O mecanismo da resistência de cepas de *Plasmodium* às drogas como a cloroquina, mefloquina e às combinações sulfadoxina-pirimetamina expôs milhares de pessoas à doença (Range & Dale, 1993). Diante este quadro torna-se emergencial o prosseguimento em pesquisas e o desenvolvimento de tecnologias para o surgimento de novos protótipos (Lead Compound).

Muitas categorias de substâncias com atividade anti-malária de estruturas químicas variadas estão sendo pesquisadas sob o ponto de vista farmacológico. Entre as mais promissoras encontramos: Floxacrina, Piperaquina, Artemisinin, Xantonas, Bisquinolinas e Chalconas. Destas, é de suma importância destacar alguns derivados de Chalcona (1,3-difenil-2-propen-1-ona), por apresentarem mecanismo inibidor enzimático de proteases malárias (Rosenthal, 2002).

Chalconas são precursores biossintéticos de flavonóides. Uma das primeiras Chalconas a demonstrar atividade antimalária foi a diidrochalcona

Phlorizidin encontrada em *Micromelum tephrocarpum* (Rutaceae) (Kayser *et al*, 2003).

Em destaque, podemos citar o composto Licochalcona A por apresentar potencial atividade biológica em termos de atividade antiparasitária, sendo considerada como um dos compostos mais promissores da série. A Licochalcona A foi isolada de *Glycyrrhiza glabra* (Fabaceae) e tem sido alvo de inúmeros estudos pré-clínicos. Sua atividade biológica contra diferentes parasitas (*Plasmodium falciparum*, *Leishmania donovani* e *L. major*) é bem documentada *in vivo* e *in vitro*. O seu mecanismo de ação anti-leishmanial está relacionado com a inibição da cadeia transportadora de elétrons nas mitocôndrias das espécies parasitas (Kayser *et al*, 2001).

Recentemente dez derivados de Chalcona foram sintetizados e testados contra *Leishmania brasiliensis* e *Trypanosoma cruzi*, sendo que todos os derivados testados inibiram o crescimento *in vitro* destes parasitas (Lunardi *et al*, 2003).

Além de sua potente atividade antiparasitária (*Plasmodium*, *Trypanosoma*, *Leishmania*), grande número de pesquisas tem demonstrado que as Chalconas apresentam também outras atividades biológicas, entre as quais anti-inflamatória, antiviral, antitumoral e antimitótico.

Segundo Gerhauser *et al* 2002, a Xanthohumol (XN), uma Chalcona de *Humulus lupulus* L., possui mecanismo inibitório nos estágios de iniciação, promoção e progressão da carcinogênese, podendo ser utilizada como agente quimio-preventivo tumoral. Sua habilidade está associada à capacidade de remover radicais peroxi e hidroxila, além de inibir o radical ânion superóxido e a produção de óxido nítrico.

Mecanismos antiproliferativos da XN para prevenir carcinogênese na fase de progressão incluem inibição da síntese de DNA e suspensão da indução do ciclo celular na fase S, apoptose e diferenciação celular.

Herencia *et al*, 1998 sintetizaram derivados de Chalcona que apresentaram considerável atividade anti-inflamatória sendo que cinco dos derivados

sintetizados possuíram potencial atividade inibitória da 5-Lipoxigenase e quatro dos derivados inibiram a enzima Ciclooxigenase-2.

Chen *et al*, 2001 relataram derivados de Chalcona com potente atividade *in vitro* e *in vivo* contra espécies de *Leishmania major* e *Leishmania donovani* com mecanismo inibitório da enzima fumarato redutase.

Recentemente inúmeros derivados de Chalcona hidroxilados a alcoxilados foram sintetizados e testados *in vitro* e *in vivo* contra cepas de *Plasmodium falciparum*, onde os derivados hidroxilados se mostraram ser menos ativos (Liu, *et al*, 2001).

Ensaio para protease cisteína demonstraram que derivados de Chalcona são inibidoras reversíveis da enzima CRUZAIN, uma protease cisteínica do *Trypanosoma cruzi*, uma enzima bastante homóloga à protease cisteínica malarial (Li *et al*, 1995).

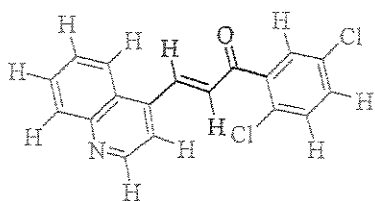
As Chalconas apresentam uma similaridade estrutural com as Acil-hidrazidas, substâncias que apresentam atividade contra cepas de *Plasmodium falciparum*. Através de pesquisa computacional em DATABASE, pesquisadores do grupo do Prof. Cohen (<http://www.cmpharm.ucsf.edu/cohen/>) identificaram uma Acil-hidrazida como protótipo antimalarial. Chalconas quando comparadas com as Acil-hidrazidas, são mais ativas e apresentam estabilidade metabólica no vacúolo digestivo do parasita do que as Acil-hidrazidas, que são instáveis em meio ácido.

Li, *et al* 1995 sintetizaram uma série de derivados de Chalcona através de reações de condensação Claisen-Schmidt de aldeídos com cetonas e os condensados obtidos foram testados *in vitro* contra cepas de *Plasmodium falciparum*, CDC/Indochina III (W2) e CDC/Sierra Leone I (D6) mostrando-se serem ativos, onde a cepa W2 é considerada resistente a Cloroquina, Quinina e Pirimetamina (Tabela 1) (Li *et al*, 1995).

A figura 15 relata a estrutura química dos derivados de Chalcona que foram abordadas nesta pesquisa QSAR. Os derivados envolvidos nesta pesquisa apresentam a mesma estrutura protótipo, se diferenciando apenas quanto aos substituintes A e B. A atividade biológica destes derivados foi apresentada na tabela 1.

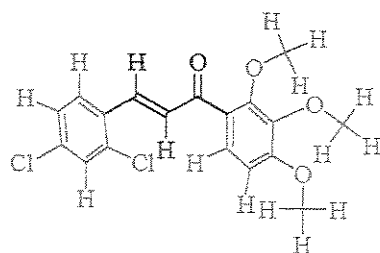
Figura 15: Estrutura química dos derivados de Chalcona.

Chalcona 7



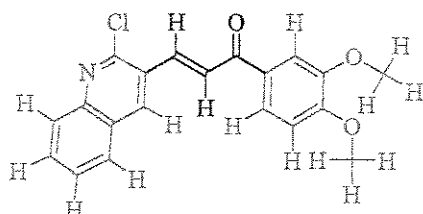
1-(2,5-diclorofenil)-3-(quinolin-4-il)-prop-2-en-1-ona

Chalcona 8



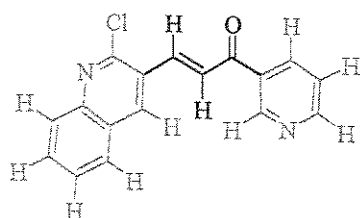
3-(2,4-diclorofenil)-1-(2,3,4-trimetoxifenil)-prop-2-en-1-ona

Chalcona 9



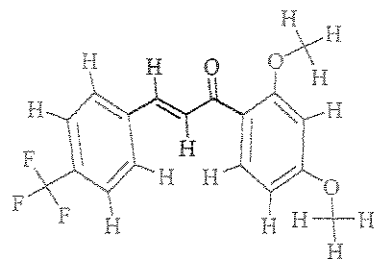
3-(2-cloroquinolin-3-il)-1-(3,4-dimetoxifenil)-prop-2-en-1-ona

Chalcona 10



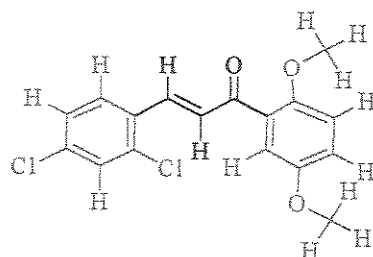
3-(2-cloroquinolin-3-il)-1-(piridin-3-il)-prop-2-en-1-ona

Chalcona 11



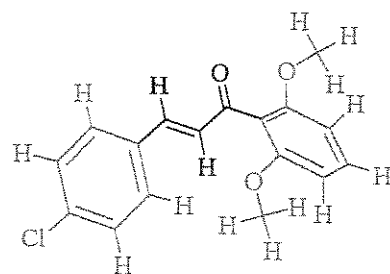
1-(2,4-dimetoxifenil)-3-(4-(trifluormetil)fenil)-prop-2-en-1-ona

Chalcona 12



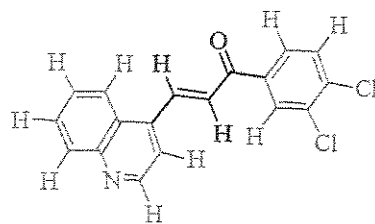
3-(2,4-diclorofenil)-1-(2,5-dimetoxifenil)-prop-2-en-1-ona

Chalcona 13



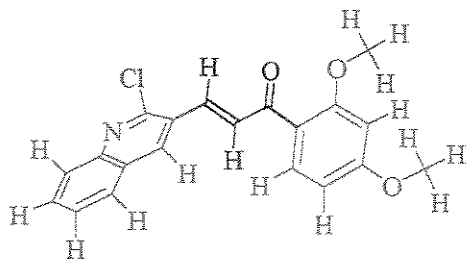
3-(4-clorofenil)-1-(2,6-dimetoxifenil)-prop-2-en-1-ona

Chalcona 14



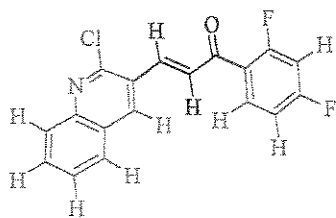
1-(3,4-diclorofenil)-3-(quinolin-4-il)-prop-2-en-1-ona

Chalcona 15



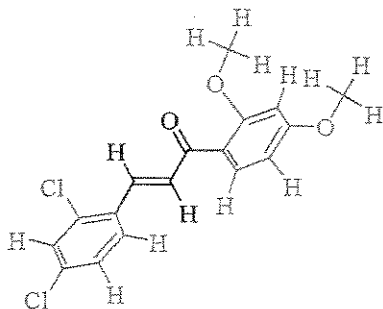
3-(2-cloroquinolin-3-il)-1-(2,4-dimetoxifenil)-prop-2-en-1-ona

Chalcona 16



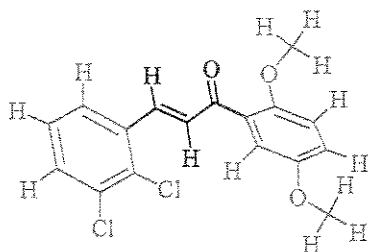
3-(2-cloroquinolin-3-il)-1-(2,4-difluorfenil)-prop-2-en-1-ona

Chalcona 17



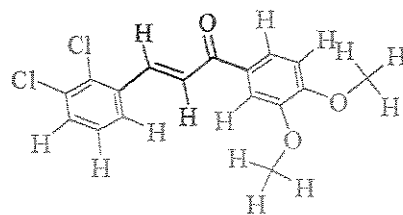
3-(2,4-diclorofenil)-1-(2,4-dimetoxifenil)-prop-2-en-1-ona

Chalcona 18



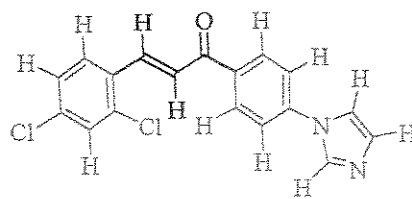
3-(2,3-diclorofenil)-1-(2,5-dimetoxifenil)-prop-2-en-1-ona

Chalcona 19



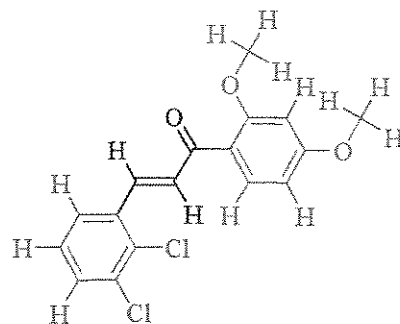
3-(2,3-diclorofenil)-1-(3,4-dimetoxifenil)-prop-2-en-1-ona

Chalcona 20



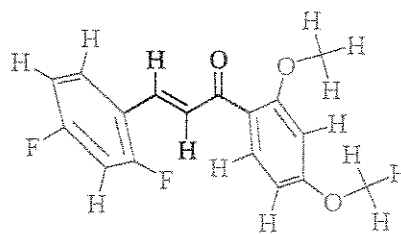
1-(4-(1-imidazol-1-il)fenil)-3-(2,4-diclorofenil)-prop-2-en-1-ona

Chalcona 21



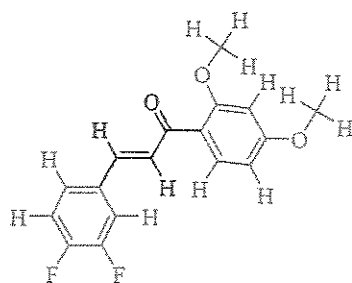
3-(2,3-diclorofenil)-1-(2,4-dimetoxifenil)-prop-2-en-1-ona

Chalcona 22



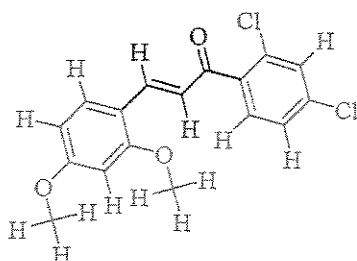
3-(2,4-difluorfenil)-1-(2,4-dimetoxifenil)-prop-2-en-1-ona

Chalcona 23



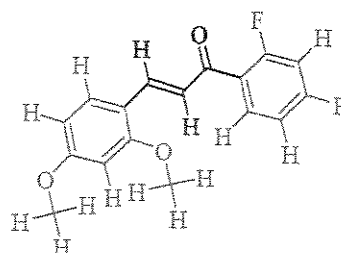
3-(3,4-difluorfenil)-1-(2,4-dimetoxifenil)-prop-2-en-1-ona

Chalcona 24

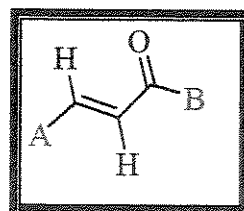


1-(2,4-diclorofenil)-3-(2,4-dimetoxifenil)-prop-2-en-1-ona

Chalcona 25



1-(2,4-difluorfenil)-3-(2,4-dimetoxifenil)-prop-2-en-1-ona



*Estrutura protótipo dos derivados de Chalcona com substituintes A e B.

Tabela 1: Derivados de Chalcona e suas respectivas atividades biológicas: concentração (μM) necessária para inibição *in vitro* de 50% da parasitemia provocada pela cepa W2* de *Plasmodium falciparum*, CDC/Indochina III, segundo Li *et al*, 1995.

W2* - Cepa de *Plasmodium falciparum* resistente a cloroquina, quinina e pirimetamina.

Chalconas	IC ₅₀ (μM) - cepa W2*	- Log IC ₅₀ - cepa W2*
Chalcona 7	0,23	0,63826
Chalcona 8	0,38	0,44368
Chalcona 9	0,51	0,29241
Chalcona 10	0,71	0,14872
Chalcona 11	0,75	0,12492
Chalcona 12	0,75	0,12492
Chalcona 13	0,77	0,11350
Chalcona 14	0,80	0,09891
Chalcona 15	1,00	0,00000
Chalcona 16	1,00	0,00000
Chalcona 17	0,80	0,04575
Chalcona 18	0,92	0,03819
Chalcona 19	1,14	-0,05690
Chalcona 20	1,18	-0,07191
Chalcona 21	1,38	-0,01399
Chalcona 22	1,90	-0,27870
Chalcona 23	2,09	-0,32020
Chalcona 24	3,80	-0,57968
Chalcona 25	11,1	-1,04527

Parte II: Introdução

2.1 O mundo quântico no desenvolvimento de fármacos

Quase todas as propriedades químicas dos sistemas moleculares podem ser explicadas em termos de propriedades atômicas. É preciso entender a estrutura eletrônica dos átomos e as interações elétron-elétron e elétron-núcleo para que possamos compreender as ligações químicas existentes nas moléculas e as possíveis interações intermoleculares.

Até o final do século XIX, a matéria era descrita em função da mecânica clássica ou mecânica Newtoniana, da termodinâmica e do eletromagnetismo. Na mecânica clássica, os cientistas descreviam a estrutura eletrônica dos átomos em função das leis de movimento de Newton propostas no século XVII. No final do século XIX, notava-se que a mecânica clássica explicava bem os fenômenos macroscópicos e falhava nos microscópicos, como por exemplo, na compreensão dos processos envolvidos à estrutura eletrônica dos átomos e moléculas, onde um grande número de indícios experimentais demonstrava cada vez mais o fracasso da mecânica clássica aplicada ao movimento de partículas muito pequenas.

No início do século XX, novas leis que ficaram conhecidas como leis da mecânica quântica foram desenvolvidas e aplicadas aos diversos sistemas atômicos e moleculares. Com a falência da mecânica clássica para movimento de sistemas microscópicos, houve uma estruturação conceitual sobre alguns aspectos da matéria e assim deu-se o aparecimento de uma nova abordagem sobre partículas, a unificação dos termos clássicos de “partícula” e de “onda”.

Na mecânica clássica, as energias podem variar continuamente, enquanto pelos conceitos da mecânica quântica, a energia da radiação eletromagnética e da matéria não pode ser variada continuamente, encarando que os termos onda e partícula são entidades de caráter duplo exibido pela matéria. Temos aí o surgimento da **dualidade onda-partícula** da matéria, admitindo que as partículas em vez de se deslocarem ao longo de uma trajetória, se distribuem no espaço como uma onda. Considerando que as partículas de um determinado sistema apresentam propriedades de onda, não podemos esperar que se comportem

como objetos pontuais, ou seja, movimentando-se em trajetórias precisas. Cientistas do século XX reorganizaram a descrição da matéria levando em conta a dualidade onda-partícula. O austríaco Erwin Schrödinger, substituiu o termo *trajetória precisa para partícula* por uma **função de onda**, representada por Ψ , sugerindo assim uma equação para determinar a função de onda de sistemas.

Os métodos quânticos baseiam-se nas soluções da equação de Schrödinger e na aplicação destas aos sistemas biológicos. Esses modelos recorrem à química teórica computacional para a realização de diversos cálculos devido a enorme quantidade de parâmetros considerados. A química quântica leva em consideração alguns parâmetros:

- ⇒ O núcleo e os elétrons são distinguidos um do outro;
- ⇒ Interações de caráter elétron-elétron e elétron-núcleo são explícitas;
- ⇒ As interações são caracterizadas pelas cargas nucleares e eletrônicas (energia potencial) e pelo movimento dos elétrons;
- ⇒ As interações determinam a distribuição espacial do núcleo e dos elétrons e suas respectivas energias;
- ⇒ Não é possível especificar a localização precisa de uma partícula se ela se comporta como onda. A dualidade onda-partícula elimina a possibilidade de descrever a localização se o momento linear da partícula estiver precisamente definido, ou seja, não podemos especificar a trajetória de partículas, temos então o conhecido “**Princípio da Incerteza de Heisenberg**”.

Para verificar as propriedades de onda de um elétron, L. de Broglie (1923) introduziu a idéia que para cada partícula de massa m existe um comprimento de onda associado fornecido pela equação matemática:

$$\lambda = h/mv = h/p \quad (1)$$

onde λ é o comprimento de onda, h é a constante de proporcionalidade ou constante de Planck ($6,63 \times 10^{-34}$ J.s), m é a massa da partícula, v é a velocidade da partícula e p é o momento linear da partícula;

⇒ Baseia-se na determinação de soluções aproximadas para a equação de onda de Schrödinger:

$$\mathbf{H} \Psi = \mathbf{E} \Psi \quad (2)$$

onde, o operador Hamiltoniano, $\mathbf{H}$, possui informação de descrição dos elétrons e núcleo em um sistema, a função de onda eletrônica, Ψ , descreve o estado dos elétrons em termos de seu movimento e posição, enquanto $\mathbf{E}$ representa a energia associada ao estado particular do elétron. Ao calcularmos Ψ , podemos prever a probabilidade da partícula ser encontrada em uma dada região do espaço;

⇒ Soluções exatas para a equação de Schrödinger são possíveis, somente para sistemas núcleo-1-elétron (átomos hidrogenóides). As soluções descrevem um conjunto de estados admissíveis para um elétron. A quantidade observável para estes estados é descrita como uma função de probabilidade, Ψ^2 , sendo que esta função é o quadrado da função de onda, e quando propriamente normalizada, descreve a probabilidade de encontrar um elétron naquele estado. A probabilidade de encontrar um elétron em uma região do espaço é proporcional ao quadrado de Ψ , sendo o termo Ψ^2 considerado como uma densidade de probabilidade. Este significado da função de onda foi incorporado pelo alemão Max Born (Atkins & Jones, 2001).

$$\int \Psi^2 d\tau = 1 \quad (3)$$

onde $d\tau = dx \, dy \, dz$ (volume)

Deste modo a interpretação da função de onda de Born, opera com o quadrado da função de onda e impõe severas restrições às funções de onda aceitáveis. A função de onda Ψ deve ser contínua, ter a primeira derivada contínua, ser unívoca e finita em todos os pontos do seu domínio. Uma função de onda não pode ser sempre nula, pois a partícula que descreve tem que estar num

certo ponto. Podemos então afirmar que a partícula tem que apresentar certas energias para que a sua função de onda seja fisicamente aceitável, portanto, a energia da partícula é quantizada. A interpretação de Born nos fornece a informação sobre as propriedades dinâmicas da partícula (Atkins, 1999);

⇒ Funções de onda de elétrons em átomos são chamadas de orbitais atômicos e suas energias são denominadas de energias de orbitais. As equações para os orbitais atômicos são obtidas como soluções da equação de Schrödinger;

⇒ Para uma molécula com muitos elétrons e núcleos o objetivo é ser capaz de descrever os orbitais moleculares e suas energias de modo análogo à equação de Schrödinger, à medida que possível.

Sabemos que a compreensão da estrutura eletrônica de um dado composto químico é um fator de extrema importância não apenas em termos de propriedades químicas, mas também é algo relevante na análise e no estudo da atividade biológica destes, nos receptores dos organismos vivos (Bunge, 1977). Conseqüentemente, existem relações entre índices mecânico-quânticos e atividade biológica. Estas relações devem ser consideradas no desenvolvimento de novos fármacos, no estudo de possíveis mecanismos de ação de drogas, na determinação da topografia de receptores biológicos e outros aspectos que favorecerão a otimização de processos no desenvolvimento de fármacos.

O estabelecimento de relações entre estrutura química e atividade biológica é considerado no contexto atual como uma temática de grande relevância em química medicinal e na farmacologia molecular. “O grande sonho dos químicos farmacêuticos e dos farmacologistas têm sido o de obter racionalmente medicamentos sob medida, isto é, medicamentos que tenham ação farmacológica específica. Vários meios foram e têm sido utilizados para lograr este diserto, mas as probabilidades de êxito continuam remotas: não raro é preciso sintetizar e depois ensaiar milhares de novos compostos químicos antes que um dado composto venha a ser introduzido terapeuticamente como medicamento” (Korolkovas) (Bunge, 1977). As perspectivas atuais são mais brilhantes que há algumas décadas. Tornaram-se mais evidentes, a partir do momento que alguns pesquisadores da área de desenvolvimento de fármacos passaram a recorrer aos

conhecimentos modernos em diversas áreas correlatas com aplicação da química teórica computacional. Com a utilização de métodos utilizados em diversas áreas de pesquisa, como a físico-química, química quântica, bioquímica, biofísica, parasitologia, imunologia, genética, farmacologia, estatística, microbiologia, biologia molecular, fisiopatologia e outras, o arsenal terapêutico foi sendo enriquecido resultando assim uma inserção cada vez maior de novos medicamentos no mercado mundial.

Sem dúvida nenhuma, a química quântica tem contribuído de forma bastante significativa para o descobrimento e desenvolvimento de novos fármacos. A teoria dos orbitais moleculares, aplicada a sistemas químicos e biológicos fornece informação rápida e barata sobre um número grande de moléculas de interesse farmacológico. Não obstante, inúmeras universidades e grandes indústrias farmacêuticas, estão necessitando cada vez mais de químicos quânticos para atuar em pesquisa na área de química medicinal.

Devemos levar em consideração que qualquer índice de atividade biológica é certamente resultado de um grande número de fatores específicos, como a estrutura eletrônica, variações estruturais, conformação espacial da estrutura química, distribuição eletrônica na molécula (Krogsgaard-Larsen *et al*, 1996). Para estudar e pesquisar de modo racional estruturas químicas de substâncias bioativas, a química quântica estará sempre contribuindo através de métodos computacionais teóricos, seja por intermédio de métodos *ab initio*, semi-empírico ou funcional de densidade.

2.2 Métodos computacionais para sistemas biológicos

Em épocas anteriores, o estudo das relações entre estrutura química e atividade biológica (SAR) era realizado levando em consideração a síntese de diversos análogos e a concretização posterior de inúmeros ensaios visando verificar a atividade biológica de substâncias químicas. O que sabemos hoje é fruto de muitas pesquisas e investigações realizadas na área de química medicinal e em áreas correlatas, o que nos permitiu com o passar das décadas estabelecer

uma vasta biblioteca de informações a respeito das propriedades moleculares de compostos derivados de um protótipo.

Experimentos foram realizados por várias décadas e ainda continuam sendo um meio de garantia do conhecimento, pois o semi-empírico e o empírico caminham juntos nesta jornada evolucionista do conhecimento. Técnicas tradicionais de modificação estrutural nos indicou ao longo de vários anos a influência de modificações moleculares específicas em estruturas químicas relacionando-as com os efeitos biológicos. Isto permitiu a exploração de características das moléculas, como as propriedades: dimensões moleculares, propriedades hidrofóbicas, propriedades eletrônicas e características relacionadas aos aspectos farmacocinéticos e farmacodinâmicos, juntamente com os fatores estereoquímicos envolvidos no reconhecimento molecular fármaco-receptor. Os mecanismos de metabolização das moléculas em organismos vivos também foram analisados objetivando explorar os processos envolvidos na ação farmacológica dos compostos.

Estudos empíricos tradicionais a respeito das relações entre estrutura química e atividade biológica envolvem técnicas onerosas e que requer um longo período de investigação para que possam ser bem sucedidos. No entanto, o rápido avanço em tecnologias computacionais tem facilitado cada vez mais as investigações empíricas tradicionais.

Nos últimos vinte anos, com as diversas conquistas obtidas na química quântica, biotecnologia, bioquímica, biofísica, biologia molecular e o aprimoramento em tecnologias de caracterização e identificação de estruturas químicas por métodos físicos, bioquímicos, imunológicos e através da biologia molecular, o mecanismo de descoberta e desenvolvimento de novos fármacos tornou-se uma trajetória mais racional, menos dispendiosa e exaustiva. Deste modo, a modelagem molecular tem desempenhado papel importante no contexto atual da química de fármacos, sendo utilizada por estudantes e pesquisadores em várias instituições de pesquisa e em universidades como uma ferramenta moderna.

A modelagem molecular consiste na investigação de estruturas químicas e no estudo das propriedades moleculares de estruturas utilizando um conjunto de técnicas da química computacional que aliam os métodos da química teórica a dados experimentais, juntamente com o auxílio de técnicas de visualização gráfica, no intuito de fornecer aspectos geométricos tridimensionais de estruturas químicas (Cohen, 1996).

Atualmente, percebemos que o cenário da pesquisa em biomoléculas tornou-se um universo de interatividade envolvendo químicos teóricos e experimentais, notando-se um nítido aumento no número de computadores em locais de trabalho e de pesquisa.

A disponibilidade de programas computacionais em química voltado ao estudo de biomoléculas em adição ao número crescente de banco de dados em rede, permite desenvolver pesquisas com aplicação de uma abordagem racional voltada para a descoberta de protótipos e estruturas análogas com uma atividade biológica (Morgon, 2001).

Através de métodos teóricos pode-se avaliar e prever propriedades moleculares que são de suma importância no entendimento das correlações estrutura-atividade. A química teórica é basicamente fundamentada em conceitos da mecânica clássica e mecânica quântica. São diversos os aspectos abordados em química computacional incluindo a modelagem molecular, métodos computacionais, Computer-Aided Molecular Design (CAMD), bancos de dados químicos e outros (Cohen, 1996).

Os métodos computacionais envolvidos na química teórica correspondem aos métodos mecânicos moleculares (MM) e mecânicos quânticos (QM). Os métodos mecânicos moleculares (mecânica dinâmica – determinístico e monte Carlo – estocástico) utilizam as leis da física clássica a núcleos atômicos das moléculas sem considerar explicitamente os efeitos eletrônicos e confiam em um campo de força com parâmetros empíricos embutidos (Morgon, 2001). São métodos que posam da vantagem de serem computacionalmente menos intensivos, rápidos e úteis em computadores limitados de recursos. Outra vantagem reside no fato de serem utilizados para macromoléculas e enzimas

(sistemas de milhares de átomos). Estes métodos possuem desvantagens tais como o fato do campo de força ser aplicável para uma limitada classe de moléculas, não realizar cálculos de propriedades eletrônicas, requerer dados experimentais para parâmetros e pelo fato de serem utilizados para sistemas ou processos que não possam ocorrer nenhum rompimento ou formação de ligações químicas.

Os métodos mecânicos quânticos (*ab initio*, semi-empírico e funcional de densidade) confiam na equação de Schrödinger para descrição de moléculas levando em consideração o tratamento explícito da estrutura eletrônica.

Os métodos semi-empíricos são métodos quânticos utilizados para sistemas contendo centenas de átomos e que requerem parâmetros empíricos experimentalmente derivados. Possuem como vantagens o fato de computacionalmente serem menos exigentes que os métodos *ab initio* e serem capazes de calcular estados de transição e estados excitados. Estes métodos posam de desvantagens como o fato de exigirem dados experimentais para parâmetros e de serem menos rigorosos que os métodos *ab initio* (Ratner & Schatz, 2001; Rogers, 1994; Levine, 1991).

Os métodos *ab initio* são métodos quânticos utilizados para sistemas contendo dezenas de átomos e não requerem nenhum parâmetro empírico. Estes métodos apresentam vantagens de serem utilizados para uma ampla gama de sistemas, não depender de dados experimentais, serem matematicamente rigorosos e capazes de determinar não só estados fundamentais como também estados de transição e estados excitados. A desvantagem destes métodos reside em serem computacionalmente muito intensivos (Ratner & Schatz, 2001; Rogers, 1994; Levine, 1991).

Os métodos computacionais podem executar cálculo de energia potencial de superfície (PES) para uma estrutura molecular, que pode ser entendido como, a concretização das forças de interação entre átomos em uma molécula com finalidade em obter informações estruturais desta molécula. O cálculo da energia potencial de superfície para um determinado conjunto de coordenadas atômicas é denominado *Single Point Energy Calculation*. O principal foco de interesse em

termos de energia potencial de superfície está voltada para a busca de extremos energéticos de coordenadas atômicas. Estes extremos são caracterizados como Global Minimum (mínimo global) e Local Minimum (mínimo local).

Os métodos computacionais também executam cálculos de minimização energética de estruturas químicas e de otimização da geometria molecular (*Geometry Optimization*). Tais cálculos objetivam determinar a geometria molecular da estrutura mais estável possível (Carvalho *et al.*, 2003). Isto é realizado por intermédio de modificações sistemáticas nas coordenadas atômicas de um dado modelo inicial desenhado tridimensionalmente, obtendo uma geometria molecular onde o campo de forças resultante na estrutura tende a zero. Este procedimento é denominado minimização de energia e baseia-se no rearranjo das posições atômicas no sentido em minimizar os overlaps atômicos (sobreposições atômicas) e assim obter uma geometria 3D mais estável possível. O critério fundamental da minimização energética baseia-se na obtenção da estrutura química com geometria molecular apresentando a menor tensão energética.

Outros diversos tipos de cálculos, de propriedades moleculares, tais como momento dipolar, calor de formação, cargas parciais, comprimento de ligações químicas, ângulos entre as ligações químicas, ângulos diedrais e outros mais, podem ser executados via métodos computacionais.

O importante é salientar que nem todos os tipos de cálculos são executados por todos os métodos e nenhum método é melhor o suficiente para todos os propósitos. Assim, cada método apresentará vantagens e desvantagens para uma determinada aplicação, e por fim, caberá ao pesquisador a escolha do método mais eficaz, levando em consideração uma série de critérios como a natureza da molécula, o tipo de informação a ser obtida, recursos computacionais e a disponibilidade em usar determinados parâmetros que foram determinados experimentalmente (Morgon, 2001).

Atualmente, um novo cenário vem sendo traçado, fruto do “casamento” entre os métodos mecânicos moleculares (MM) e os métodos mecânicos quânticos (QM). Esta combinação deu origem aos métodos híbridos clássico-

quântico, uma técnica mais satisfatória, conhecida por *Quantum Mechanics/Molecular Mechanics* (QM/MM) (Mundim, 2003).

Visando a exploração das interações fármaco-receptor através de técnicas computacionais, é de fundamental importância a determinação da geometria molecular das estruturas químicas em estudo para posteriormente executar os devidos cálculos de propriedades moleculares e assim correlacionar a estrutura química dos compostos análogos estudados com a resposta biológica.

Seja qual for o método computacional utilizado para determinar a geometria molecular, é fundamental fornecer o modelo inicial da estrutura a ser pesquisada. O modelo inicial pode ser obtido a partir de informação estrutural via métodos físicos ou então pode ser construído por intermédio de programas computacionais de desenho de estruturas químicas em 3D. Uma vez construído o modelo inicial da estrutura química, é necessário explorar as conformações moleculares permitidas. Geralmente, as estruturas moleculares de menor energia são as que mais interessam ao químico medicinal, embora nem sempre a conformação bioativa seja a de mínima energia.

2.3 Métodos Mecânicos Moleculares (MM)

Os métodos MM utilizam equações que obedecem as leis da física clássica (equações de Newton) aplicando-as para os núcleos dos átomos nas moléculas não incluindo os elétrons em seus cálculos. Pode-se dizer que são métodos que consideram as moléculas como um conjunto de átomos ligados entre si por intermédio de ligações químicas covalentes do tipo “mola” ou potenciais harmônicos, onde as ligações são representadas através de funções potenciais de energia. Nestes métodos, um conjunto de funções de energia potencial é utilizado na tentativa em descrever as interações intramoleculares da estrutura em estudo (Costa, 1998).

Em mecânica molecular, calcula-se a energia potencial do sistema molecular utilizando-se uma função de energia empírica que tenta descrever as ligações covalentes através de funções potenciais (Clark, 1985). Este conjunto de funções potenciais é denominado campo de força e contém parâmetros ajustáveis

que são otimizados para obter o melhor conjunto de propriedades das moléculas, computando os movimentos nucleares e dos elétrons, como um potencial efetivo; a energia potencial total da molécula em seu estado fundamental é obtida em função da posição nuclear (Burket & Allinger, 1982).

A mecânica molecular é rápida, menos intensiva em termos computacionais se comparada à mecânica quântica, simples de introduzir a evolução temporal e pode incorporar a temperatura como um fator externo, mas não faz cálculos de propriedades eletrônicas já que seus elétrons não são considerados nos cálculos. Mesmo não levando em consideração os elétrons, métodos MM são úteis para realização de alguns tipos de cálculos, como (Patrick, 2001):

- ⇒ Minimização energética das moléculas visando a obtenção de conformações estáveis;
- ⇒ Cálculo de *Single Point Energy* para comparar conformações da mesma molécula;
- ⇒ Estudo dos movimentos moleculares usando a dinâmica molecular;
- ⇒ Pesquisas de espaço conformacional obtidas pela variação de único ângulo de diedro.

Em mecânica molecular o átomo é representado com um corpo esférico que possui uma massa particular e a energia potencial total da molécula é resultado de um conjunto de contribuições de energias potenciais (ACCVIP, 2003; Espírito Santo, 2001; Costa, 1998). As contribuições de energia potencial do sistema molecular incluem os seguintes termos:

- **V_{lig} (*Bond Stretching Energy*)** – termo associado ao estiramento da ligação (figura 16) entre átomos diretamente ligados, ou seja, é a contribuição de energia potencial necessária para comprimir ou alongar uma ligação em relação ao seu comprimento original. A equação da energia potencial de estiramento de ligação baseia-se na lei de Hooke, onde os termos K_r e r_0 são parâmetros ajustáveis:

$$V_{lig} = 1/2 \sum K_r (r_n - r_0)^2 \quad (4)$$

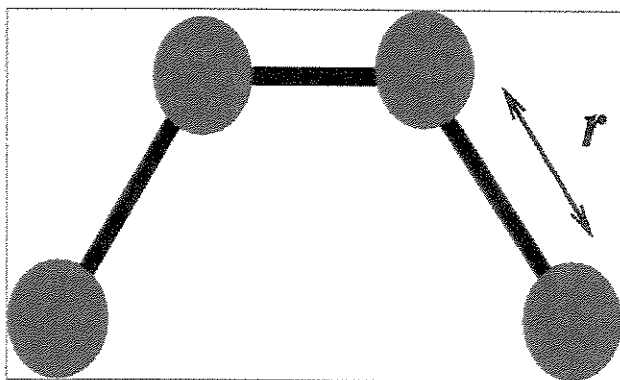


Figura 16: Energia de estiramento de ligação

sendo que K_r é a constante de estiramento da ligação e r_0 é o comprimento da ligação de referência, onde o somatório (Σ) é realizado para todas as ligações.

- V_{ang} (*Angle Bending Energy*) – termo associado ao ângulo entre ligações químicas (figura 17), ou seja, é contribuição de energia potencial envolvida na deformação angular da ligação química em relação ao ângulo natural da ligação.

$$V_{ang} = \Sigma K_{\theta} (\theta_n - \theta_0)^2 \quad (5)$$

sendo que K_{θ} é a constante da força de deformação do ângulo e θ_0 é o ângulo natural da ligação, onde o somatório (Σ) é realizado para todos os ângulos.

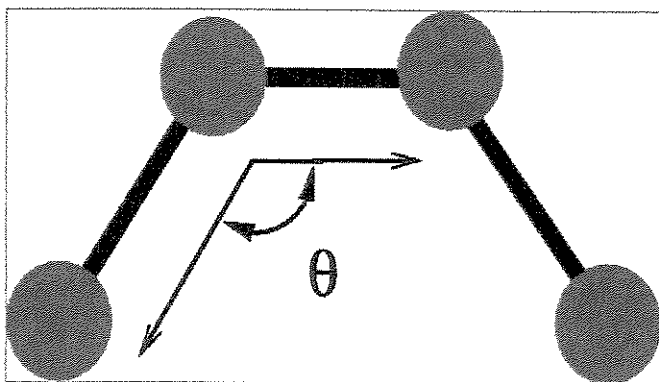


Figura 17: Energia de deformação angular

- V_{tor} (**Torsion Energy**) – termo associado às rotações internas das ligações simples (figura 18), ou seja, é a contribuição da energia potencial de torção (giro) de ligação simples entre átomos que são vicinais para dois outros contribuindo com um diedro.

$$V_{\text{tor}} = \sum V_n/2 [1 + \cos (n\delta - \delta_0)] \quad (6)$$

sendo que o parâmetro $V_n/2$ é a constante de força torsional, n indica a sua periodicidade e ϕ é o ângulo de torção.

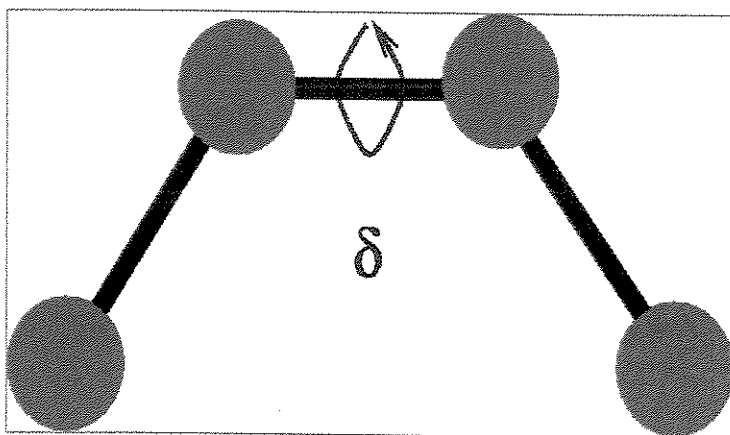


Figura 18: Energia de deformação torsional

- $V_{\text{non-bonded}}$ (**Non-bonded Energy**) – termo utilizado para designar as interações entre átomos não ligados, ou seja, representa a contribuição de energia potencial envolvida neste tipo de interação. Geralmente, as interações entre átomos não ligantes incluem efeitos de repulsão entre átomos que estão muito próximos e uma atração para longas distâncias proveniente das forças de dispersão. A contribuição de energia potencial para átomos não ligantes, pode ser subdividida em dois tipos principais:

A) $V_{\text{van der Waals}}$ (**van der Waals Energy**) – termo utilizado para descrever as interações atrativas ou repulsivas entre átomos não ligados (figura 19), ou seja, é a contribuição de energia potencial envolvida nas interações entre átomos não ligados que estão muito próximos ou muito afastados. Quando a

distância entre átomos não ligantes for ligeiramente menor que a soma dos raios de van der Waals destes, as forças de repulsão serão dominantes. Efeito semelhante acontece quando os átomos estão muito afastados, determinando o aparecimento de forças atrativas entre eles. Na equação de contribuição da energia potencial de van der Waals nota-se a combinação entre uma repulsão exponencial com uma interação de dispersão atrativa.

$$V_{vdW} = \sum \sum \epsilon [B_{ij}/r_{ij}^{12} - A_{ij}/r_{ij}^6] \quad (7)$$

Sendo que A_{ij} e B_{ij} são parâmetros que controlam a profundidade e a posição do poço de energia potencial em função da distância (r_{ij}) entre os pares de átomos (i,j) não ligados, enquanto ϵ determina a intensidade de energia potencial atrativa. O parâmetro A_{ij} pode ser obtido através de medidas de polarizabilidade atômica ou pode ser calculada através da mecânica quântica enquanto o parâmetro B_{ij} é derivado de dados cristalográficos (NIH - Center for molecular modeling).

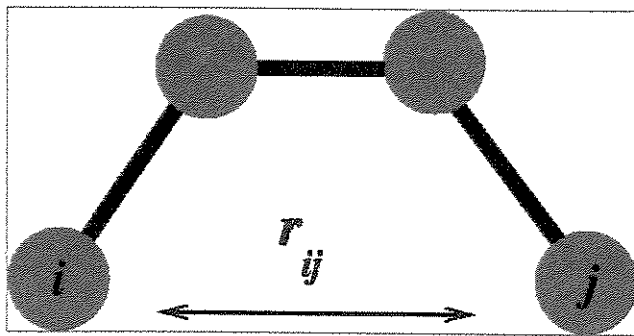


Figura 19: Energia de van der Waals

B) V_{ele} (Electrostatic Energy) – termo associado às interações eletrostáticas entre átomos não ligados, ou seja, é a contribuição de energia potencial elétrica oriunda de cargas dos átomos não ligantes. Este tipo de interação elétrica é também denominada *interação eletrostática coulombiana*.

$$V_{ele} = \sum \sum q_i \cdot q_j / \epsilon \cdot r_{ij} \quad (8)$$

sendo que, q_i e q_j são cargas pontuais de átomos não ligantes, r_{ij} é a distância entre os átomos não ligantes (i,j), enquanto ϵ é a constante dielétrica em um determinado meio e pode assumir valores constantes entre 1 e 5. Deve-se levar em consideração que existem três possibilidades de interação eletrostática, onde cada tipo de interação usa diferentes formas de equação eletrostática:

- 1) Contribuição do tipo interação carga/carga;
- 2) Contribuição do tipo interação dipolo/dipolo e
- 3) Contribuição do tipo interação carga/dipolo.

Outros termos de energia potencial podem aparecer contribuindo com a energia total do sistema molecular como V_{hyb} e V_{PdH} .

- **V_{hyb} (Hybridization Energy)** – termo utilizado para representar a contribuição de energia potencial de hibridação necessária para manter átomos de carbono tensionados e aumentar automaticamente o ângulo entre os apêndices dos anéis tensionados.

- **V_{PdH} (Bonds Hydrogen Energy)** – termo que representa a contribuição de energia potencial envolvida em interações do tipo ligações de hidrogênio (PdH). Este termo é importante para moléculas onde as ligações de hidrogênio fazem parte da estabilização da estrutura química.

A energia potencial total do sistema molecular é representada matematicamente pela somatória de todos os termos do campo de força e é denominada energia estérica da molécula, podendo ser representada pela equação a seguir (Areas *et al*, 1994; Areas *et al*, 1995):

$$V_{Total} = V_{lig} + V_{ang} + V_{tor} + V_{non-bonded} + V_{hyb} + V_{PdH} \quad (9)$$

onde $V_{non-bonded} = V_{vdW} + V_{ele}$ (Σ da contribuição de van der Waals e contribuição eletrostática).

Os programas que executam cálculos por métodos MM determinam as interações resultantes do estiramento das ligações, deformação angular, deformação torsional, interação entre átomos não ligantes e outros tipos de interações e assim calculam a energia da molécula de partida de forma comparativa em relação a um determinado padrão.

Um método mecânico molecular promove a modificação dos comprimentos de ligações químicas entre os átomos e alteração nos ângulos entre as ligações dos átomos na molécula de partida, fornecendo conformações com suas respectivas energias potenciais. Os programas reconhecem as mudanças realizadas de forma a obter estruturas mais estáveis, estruturas moleculares com menor energia potencial total (Carvalho *et al*, 2003).

O valor absoluto da energia estérica do sistema molecular em estudo, não tem significado físico, e a diferença de energia entre os confôrmeros indica a variação de energia necessária para transformar uma conformação em outra (Mundim, 2003).

Em 1973, foi desenvolvido o método MM1 (Allinger, 1976), surgindo em 1977 a segunda geração dos métodos mecânicos moleculares – MM2 (Allinger, 1977), onde o campo de força em MM1 foi consideravelmente modificado em relação à contribuição de energia potencial torsional. Posteriormente o método MM2 foi reparametrizado aparecendo um novo campo de força – MM2' reajustando as falhas existentes no método MM2. O campo de força MM2 falhava na reprodução de barreiras de rotação de ligações simples, superestimava as distâncias entre hidrogênios, os parâmetros torsionais eram pequenos e por último, os parâmetros de van der Waals eram muito grandes.

Mais tarde, o método MM2' foi reajustado por modificação no termo de contribuição de energia potencial das ligações de hidrogênio, pois estas ligações eram mais curtas e mais intensas na versão MM2. A correção foi feita através da redução dos raios de van der Waals dos átomos de hidrogênio envolvidos nas interações do tipo PdH e também pelo incremento na componente atração de van der Waals entre o hidrogênio e os tipos de átomos eletronegativos (Allinger *et al*, 1988).

Em resumo, os métodos mecânicos moleculares levam em consideração os seguintes princípios:

- Consideram apenas os núcleos atômicos, não tendo sentido o cálculo de propriedades eletrônicas;
- Os átomos são tratados tipicamente como esferas;
- Ligações entre os átomos em uma molécula são tratadas como osciladores harmônicos;
- As interações entre átomos não ligantes são tratadas usando funções potenciais via mecânica clássica;
- Funções potenciais individuais contribuem para descrever as diferentes interações envolvidas: Interações de estiramento de ligações, interações de deformação dos ângulos entre as ligações, deformação em ângulo diedral devido a ocorrência de torsões em ligações simples, interações entre átomos não ligados seja do tipo van der Waals ou eletrostática e outras interações;
- A somatória das contribuições de todas as interações determina a distribuição espacial dos átomos em uma molécula, ou seja, a sua conformação espacial;
- As energias estéricas são utilizadas para comparar as energias de duas ou mais conformações oriundas de uma mesma molécula.

Em nossa pesquisa, utilizamos método MM2 modificado (Allinger, 1980) inserido no programa computacional Molecular Modeling Pro (MMP) para concretizar minimização energética das estruturas químicas, objetivando encontrar conformações estáveis para cada derivado de Chalcona. Mais tarde, no item metodologia utilizada, comentaremos em detalhes a abordagem utilizada na minimização energética para cada estrutura pesquisada.

2.4 Métodos Mecânicos Quânticos (MQ)

Muitos aspectos da estrutura molecular podem ser modelados utilizando mecânica clássica (mecânica molecular). O campo de força clássico é baseado em resultados empíricos calculados em função de uma média sob um grande número de moléculas. O campo de força clássico é bom para macromoléculas, não podendo ser utilizado na necessidade em determinar características eletrônicas a respeito de uma estrutura química. Propriedades estruturais que dependem diretamente da distribuição da densidade eletrônica necessitam da utilização de métodos que se baseie nos postulados da mecânica quântica (métodos quânticos) (Levine, 1991).

Os métodos quânticos descrevem as moléculas analisando as interações entre o núcleo e os elétrons. O sistema é descrito por uma função de onda que possa ser encontrada resolvendo a equação de Schrödinger. A equação de Schrödinger relaciona os estados estacionários e a energia do sistema ao operador Hamiltoniano. Este procedimento é a estratégia utilizada para relacionar a energia do sistema com a função de onda que descreve as posições dos núcleos e dos elétrons deste sistema. Na prática a equação de Schrödinger não pode ser resolvida exatamente, havendo a necessidade o uso de algumas aproximações (McQuarrie & Simon, 1997).

As energias e as funções de onda dos estados estacionários de um determinado sistema são caracterizadas pelas soluções da equação de Schrödinger:

$$\hat{H}\Psi = E\Psi \quad (10)$$

Na equação acima, $\hat{H}$ representa o operador Hamiltoniano, que fornece as energias cinéticas e potenciais dos núcleos e dos elétrons em um dado sistema. Ψ é uma função de onda que descreve as propriedades dinâmicas do sistema, dependente das coordenadas dos elétrons e dos núcleos e E é a energia do estado.

O átomo de hidrogênio é o único sistema que permite uma solução exata para a equação de Schrödinger sendo que a sua resolução pode ser feita observando a expressão (Hehre *et al*, 1998):

$$\left[\frac{-\hbar^2}{8\pi^2m} \nabla^2 - \frac{Ze^2}{r} \right] \Psi(x,y,z) = E \Psi(x,y,z) \quad (11)$$

Na equação acima, em parênteses, temos a energia cinética e a energia potencial de um elétron de massa m a uma distância r de uma carga nuclear $Z=1$ (Hidrogênio), $\hbar$ é a constante de Planck e e é a carga eletrônica. E na equação, representa a energia do elétron e Ψ representa a função de coordenadas atômicas. Como vimos anteriormente, Ψ^2 , indica a probabilidade de encontrar o elétron em um dado volume do espaço, melhor afirmando, corresponde à densidade eletrônica.

Tentativas para resolver a equação de Schrödinger para sistemas multieletrônicos levaram à proposição de uma equação na sua forma independente do tempo, que introduz simplificações permitindo o tratamento matemático para estes sistemas. Para um sistema molecular com N núcleos e e elétrons, as aproximações necessárias podem ser compreendidas analisando a equação de Schrödinger na sua forma independente do tempo (Hehre *et al*, 1998; www.chemkeys.com/bra/md/mdhtm):

$$\hat{H} = \hat{T}_N + \hat{T}_e + \hat{V}_{ee} + \hat{V}_{ee} + \hat{V}_{eN} \quad (12)$$

sendo que os dois primeiros termos refere-se à energia cinética do núcleo e dos elétrons, onde $\hat{T}_N$ indica o operador de energia cinética dos núcleos e $\hat{T}_e$ o operador de energia cinética dos elétrons

$$\hat{T}_N = \sum_{A=1}^N \frac{\hat{p}_A^2}{2M_A} = -\frac{1}{2} \sum_{A=1}^N \frac{\nabla_A^2}{M_A} \quad (13)$$

$$\hat{T}_e = \sum_{i=1}^N \frac{\hat{p}_i^2}{2} = -\frac{1}{2} \sum_{i=1}^N \nabla_i^2 \quad (14)$$

Os três últimos termos da equação 12 refere-se a energia potencial das atrações e repulsões ocorridas entre núcleos e elétrons, sendo então, $\hat{V}_{Ne}$ o operador de energia potencial de atração núcleo-elétron, $\hat{V}_{ee}$ o operador de energia potencial de repulsão elétron-elétron e $\hat{V}_{NN}$ o operador de energia potencial de repulsão núcleo-núcleo (www.chemkeys.com/bra/md/mdhtm):

$$\hat{V}_{Ne} = - \sum_{A=1}^N \sum_{i=1}^N \frac{Z_A}{R_{iA}} \quad (15)$$

$$\hat{V}_{ee} = \sum_{i=1}^N \sum_{j < i}^N \frac{1}{r_{ij}} \quad (16)$$

$$\hat{V}_{NN} = \sum_{A=1}^N \sum_{B=1}^N \frac{Z_A Z_B}{R_{AB}} \quad (17)$$

sendo que: $\hat{P}_A$ - operador diferencial do momentum do núcleo;

$\hat{p}_i$ - operador diferencial dos elétrons;

Z_A - carga nuclear do núcleo A;

R_{iA} - distância entre o elétron i e o núcleo A;

R_{AB} - distância entre o núcleo A e o núcleo B;

r_{ij} - distância entre os elétrons i e j.

Quatro aproximações geralmente são realizadas na química quântica (Pilar, 1990; Hehre *et al*, 1998):

- 1 – Independência do tempo;
- 2 – Negligência dos efeitos relativísticos;
- 3 – Aproximação de Born-Oppenheimer → Separação do movimento dos núcleos e dos elétrons;
- 4- Aproximação orbital → Os elétrons são confinados a determinadas regiões do espaço.

2.4.1 Aproximação de Born-Oppenheimer

Existe uma grande diferença entre a massa do elétron e do núcleo atômico. Por apresentar esta diferença na massa, o núcleo apresenta movimento com intensidade muito menor que os elétrons. Um caminho de simplificar a equação de Schrödinger para sistemas moleculares envolve a separação das energias nucleares e eletrônicas. Esta aproximação chamada de *Aproximação de Born-Oppenheimer* considera o movimento eletrônico com um núcleo fixo, permitindo assim que a energia eletrônica seja descrita separadamente da energia nuclear. Descobre-se como varia a energia da molécula em função do comprimento de ligação e levanta-se uma curva de energia potencial molecular. Esta curva leva em consideração o termo energia potencial, pois a energia cinética dos núcleos estacionários é nula (Hehre *et al*, 1998).

Escrevendo o Hamiltoniano total para o sistema molecular como sendo a soma da parte nuclear (H_n) com a parte eletrônica (H_e), a função de onda total do sistema (18) que é o produto das funções de onda nuclear (χ_n) e eletrônica (ψ_e) e a equação de Schrödinger para o sistema (19) teriam a seguinte forma:

$$\Psi_{e,n}(r,R) = \chi_n(R) \psi_e(R,r) \quad (18)$$

$$H\Psi = E\Psi \Rightarrow (H_n + H_e) \Psi_n \psi_e = E \chi_n \psi_e \quad (19)$$

sendo H_n – Hamiltoniano nuclear;

H_e – Hamiltoniano eletrônico;

Como o Hamiltoniano eletrônico (H_e) é função de $\mathbf{r}$, uma vez que os $\mathbf{R}'_s$ são tratados como constantes, podemos assim escrever a equação de Schrödinger eletrônica (20) para uma configuração nuclear específica, $\mathbf{R}$, da seguinte forma:

$$H_e \psi_e = E_e \psi_e \quad (20)$$

e a equação de Schrödinger nuclear (21) assume a forma:

$$H_n \Psi_n \psi_e \cong \sum \psi_e \nabla^2 \chi_n \Rightarrow H_n \Psi_n \cong E_n \chi_n \quad (21)$$

onde E_n e E_e refere-se a energia cinética do núcleo e eletrônica respectivamente. O Hamiltoniano eletrônico contém três termos: energia cinética, interação eletrostática entre os elétrons e núcleos e repulsão eletrostática entre os elétrons (Espírito Santo, 2001).

$$\hat{H}_e = -\frac{1}{2} \sum_{i=1}^n \Delta_i - \sum_{i=1}^n \sum_{A=1}^N \frac{Z_A}{|R_A - r_i|} + \sum_{i < j}^n \frac{1}{r_{ij}} \quad (22)$$

onde o símbolo ∇ corresponde ao operador de Laplace:

$$\Delta_i = \frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \quad (23)$$

Assim a energia total do modelo de Born-Oppenheimer é obtida somando a energia nuclear com a energia eletrônica:

$$E_{tot} = E_e + E_n \quad (24)$$

$$E_n = \sum_{A < B}^N \frac{Z_A Z_B}{|R_A - R_B|} \quad (25)$$

Dessa forma, a aproximação de Born-Oppenheimer separa o movimento nuclear do movimento eletrônico.

2.4.2 Aproximação do elétron independente

Lembrando que a equação de Schrödinger eletrônica não permite soluções exatas devido às interações de repulsão intereletrônicas, torna-se necessário então uma outra aproximação denominada de *aproximação do elétron independente*. Esta aproximação aceita o elétron ser independente, ou seja, cada elétron do sistema é influenciado por uma média de campo oriundo de todos os outros elétrons. Neste caso assumimos que a energia cinética e potencial de cada elétron sejam independentes do movimento dos demais elétrons, de forma que a Hamiltoniana eletrônica total (H_e) possa ser aproximadamente a soma das Hamiltonianas efetivas de 1-elétron (H_i^{efe}) (26) (Espírito Santo, 2001; Pilar, 1990; www.caos.kun.nl/):

$$H_e = \sum H_i^{\text{efe}} \quad (26)$$

Esta aproximação considera que, a função de onda total (Ψ) que descreve o sistema molecular, deve ser escrita como o produto das funções de onda 1-elétron (monoeletrônica). Esta função é conhecida pelo nome de *produto de Hartree* e sugere que a função que descreve um elétron qualquer é completamente independente dos outros elétrons. A equação 27 corresponde ao modelo de partículas independentes para um sistema contendo $2n$ elétrons (www.chemkeys.com/bra/md/mdhtm):

$$\Psi(1,2,\dots,2n) = \phi_1(1)\phi_2(2)\dots\phi_{2n-1}(2n-1)\phi_{2n}(2n) \quad (27)$$

Como vimos acima, a uma simplificação para a equação de Schrödinger, envolve uma separação de variáveis, que é a substituição da função de onda multieletrônica por um produto de funções de onda 1-elétron (monoeletrônicas).

Podemos resolver a equação de Schrödinger para o elétron independente (28) que se estende por toda a molécula e que descreve o elétron de modo individual pela equação:

$$H_i^{\text{efe}} \phi_i = \varepsilon_i \phi_i \quad (28)$$

A equação acima possui múltiplas soluções ϕ_i e ε_i , sendo as funções ϕ_i são chamadas de funções orbitais monoeletrônicas do i-ésimo elétron com energia ε_i .

Lembrando que cada orbital depende unicamente das coordenadas cartesianas do respectivo elétron, então as funções orbitais correspondem a:

$$\begin{aligned} \phi_1(1) &= \phi_1(x_1, y_1, z_1) \\ \phi_2(2) &= \phi_2(x_2, y_2, z_2) \\ &\vdots \\ \phi_{2n}(2n) &= \phi_{2n}(x_{2n}, y_{2n}, z_{2n}) \end{aligned} \quad (29)$$

Para resolver a equação de Schrödinger de modo aproximado, Hartree, desenvolveu uma expressão para o termo de repulsão eletrônica. Esta aproximação simples representa o modelo de Hartree, também conhecido como modelo do campo autoconsistente ou SCF (*Self-Consistent Field*). De acordo com Hartree, se um dado sistema apresenta dois elétrons, ao colocar o elétron 1 em determinada região do espaço, a energia potencial de repulsão deste elétron em relação ao campo médio produzido pelo elétron 2 será:

$$\hat{v}_1(ee) = \int \frac{\phi_2^* \phi_2}{r_{12}} d\tau_2 \quad (30)$$

Deste modo, para um sistema que apresenta um determinado número de elétrons, a energia de repulsão do i -ésimo elétron em relação aos outros elétrons será:

$$\hat{v}_i(ee) = \sum_{j=2} \int \frac{\phi_j^* \phi_j}{r_{ij}} d\tau_j \quad (31)$$

Sendo a energia eletrônica total de um sistema (22) apresentando $2n$ elétrons equivalente à soma de todas as ε_i , ou seja:

$$E_{el} = \sum_{i=1}^{2n} \varepsilon_i \quad (32)$$

então, determina-se a energia total eletrônica (33) pelo método de Hartree executando:

$$E_{el} = \sum_{i=1}^{2n} \varepsilon_i - \sum_{i=1}^{2n} v_i(ee) \quad (33)$$

Percebemos então que a aproximação de Hartree consiste em incluir os efeitos das repulsões intereletrônicas no modelo orbital como um valor médio (energia média de interação de repulsão).

As funções de onda apresentadas até agora levam em consideração apenas as coordenadas espaciais dos elétrons e não as propriedades de spin. Para considerar os efeitos de spin eletrônico, na função de onda deverá aparecer, a coordenada de spin (ξ). Assim a função de onda deverá ser reescrita com a presença de coordenadas de spin (www.chemkeys.com/bra/md/mdhtm):

$$\Psi(1,2,\dots,2n) = \phi'_1(x_1, y_1, z_1, \xi_1) \phi'_2(x_2, y_2, z_2, \xi_2) \dots \phi'_{2n}(x_{2n}, y_{2n}, z_{2n}, \xi_{2n}) \quad (34)$$

As funções ϕ' denominadas de *spin-orbitais* incluem as coordenadas espaciais e de spin. Observações experimentais sugerem que só existem duas

coordenadas de spin, e que podem assumir duas configurações: spin *up* (α) ou spin *down* (β). Os modelos propostos anteriormente desprezam o *princípio da exclusão de Pauli*, ou seja, dois elétrons em um mesmo orbital deverão apresentar momentos de spin antiparalelos, devendo a função de onda ser considerada *função de onda anti-simétrica*. Um procedimento utilizado para obter funções de onda anti-simétricas consiste em construir a função de onda pelo determinante abaixo:

(35)

$$\Psi = \frac{1}{\sqrt{(2n)!}} \begin{vmatrix} \phi'_1(1) & \phi'_1(2) & \cdots & \phi'_1(2n) \\ \phi'_2(1) & \phi'_2(2) & \cdots & \phi'_2(2n) \\ \vdots & \vdots & \ddots & \vdots \\ \phi'_{2n}(1) & \phi'_{2n}(2) & \cdots & \phi'_{2n}(2n) \end{vmatrix}$$

A sistemática de construção de funções de onda via determinante (anti-simetria), foi proposta por Slater e a função de onda obtida pela equação 35 ficou conhecida como determinante de Slater.

A aplicação do determinante de Slater ao método de Hartree comentado anteriormente, foi sugerida por V. Fock. O tratamento do campo de força autoconsistente de Hartree considerando os spin-orbitais de Fock é caracterizado como Método de Hartree-Fock (Levine, 1991).

Uma das conseqüências do Método de Hartree-Fock consiste no determinante de Slater não permitir que dois spin-orbitais sejam idênticos, ou seja, dois elétrons não podem apresentar as mesmas coordenadas espaciais e de spin. Este princípio é conhecido como *Princípio da exclusão de Pauli*.

Para um sistema molecular, uma matriz Hamiltoniana 1-elétron é construída para descrever as interações 1-elétron entre esse elétron e o centro do núcleo. Então temos a equação para dois orbitais atômicos ϕ_u e ϕ_v :

$$H_{uv} = \int \phi_u H_i^{\text{efe}} \phi_v d\tau \quad (36)$$

Porém para sistemas moleculares, esse Hamiltoniano não existe em termos de interação entre elétrons com dois ou mais diferentes centros de interação ou interação entre dois elétrons. Deste modo, o Hamiltoniano é mais uma vez modificado, sendo que esta modificação renomeia o operador Hamiltoniano como operador Fock.

O operador Fock é composto de uma série de Hamiltonianos 1-elétron que é suplementado por termos que descrevem a interação entre 2-elétrons. Estes termos incluem a densidade de matriz, integral coulombiana e integral de permutação (Espírito Santo, 2001; Levine, 1991).

$$F_{uv} = H_{uv} + \sum P_{\lambda\sigma} [\text{Coulomb} + \text{Permuta}] \quad (37)$$

A matriz Fock assume duas formas:

⇒ Restricted (RHF) – Requer que spin *up* e spin *down* dos elétrons tenham a mesma energia e ocupem o mesmo orbital.

⇒ Unrestricted (UHF) – Permite que os elétrons ocupem diferentes orbitais e tenham diferentes energias.

Com o uso das funções de onda anti-simétricas o termo de repulsão eletrônica apresentará uma componente Coulombiana e uma outra de natureza puramente quântica, denominada termo de troca. Pode-se dizer que esse termo de troca constitui-se na primeira correção dos efeitos de correlação eletrônica no modelo de Hartree.

2.4.3 Aproximação LCAO

As equações de Hartree-Fock eram possíveis de serem resolvidas numericamente para átomos, mas inviáveis para moléculas, pois este procedimento demonstrava ser computacionalmente inadequado. Geralmente, realiza-se uma aproximação conhecida como LCAO (*Linear Combination of Atomic Orbitals*), para computar orbitais moleculares (Levine, 1991).

Roothaan sugeriu uma forma eficiente de combinar orbitais atômicos, onde as funções utilizadas para representar orbitais moleculares poderiam ser obtidas

em termos de funções que representassem os orbitais atômicos de sistemas multieletrônicos, já que estes são funções aproximadas. Este método ficou conhecido como o método da Combinação Linear de Orbitais Atômicos (*Linear Combination of Atomic Orbitals* - LCAO) (McQuarrie & Simon, 1997). Inicialmente deve-se considerar que um orbital atômico ou molecular ϕ_i pode ser escrito através da expansão:

$$\phi_i = \sum_{k=1}^m c_{ik} \chi_k \quad (38)$$

onde c_{ik} corresponde a coeficientes de combinação linear que representarão os orbitais ϕ_i como misturas de m funções de base χ_k previamente escolhidas. A sua aplicação na equação:

$$\hat{H} \phi_i = \varepsilon_i \phi_i \quad (39)$$

resulta a relação:

$$\hat{H} \sum_k c_{ik} \chi_k = \varepsilon_i \sum_k c_{ik} \chi_k \quad (40)$$

Considerando-se que tem-se um conjunto de m funções do tipo χ , inicia-se o processo de solução multiplicando-se a equação acima por cada uma das funções χ_k e integra-se sobre todo o espaço de variáveis, resultando:

$$\sum c_{ik} (F_{kk} - \varepsilon_i S_{kk}) = 0 \quad (41)$$

A equação acima (41) quando passada para a forma matricial será transformada na equação de Hartree-Fock-Roothaan (HFR) (Okamoto, 2002):

$$\mathbf{FC} = \varepsilon \mathbf{SC} \quad (42)$$

onde $\mathbf{F}$ = matriz Fock;

$\mathbf{C}$ = matriz de coeficientes de orbital molecular;

$\mathbf{S}$ = matriz de superposição

ε = matriz de energia diagonalizada de orbital molecular.

Desde que, as equações de Fock sejam uma função de orbitais moleculares, elas serão linearmente independentes. Tais equações devem ser resolvidas usando freqüentemente o método do campo autoconsistente.

O maior obstáculo à resolução das equações de Hartree-Fock-Roothaan, está no grande número de integrais de dois elétrons a serem calculadas, que depende do número de centros atômicos (Espírito Santo, 2001).

Os métodos *ab-initio* caracterizam-se pelo cálculo de todas as integrais mencionadas acima, restringindo a sua aplicação para pequenos sistemas moleculares. Os métodos semiempíricos utilizam algumas aproximações para tentar viabilizar os cálculos dessas integrais de dois elétrons. As aproximações adotadas tornam estes métodos mais rápidos computacionalmente em comparação com os métodos *ab-initio*, permitindo assim que tais métodos sejam utilizados para sistemas moleculares maiores (Levine, 1991).

2.4.4 Métodos semiempíricos

Vimos que os métodos *ab-initio* usam a forma completa do operador Fock para a construção da equação de onda. O mesmo não acontece com os métodos semiempíricos, que usam operadores Fock simplificados para viabilizar a computação.

Os métodos semiempíricos se dão com dois tipos de aproximação. A primeira aproximação refere-se aos elementos de matriz da Hamiltoniana efetiva de 1-elétron (36). Estes elementos são fornecidos por intermédio de valores parametrizados empíricos ou semiempíricos, que tentam correlacionar o resultado dos cálculos aos dados experimentais existentes. A segunda aproximação dá-se em alguns termos relativos às integrais eletrônicas que são desprezados.

Os métodos semiempíricos são baseados na aproximação de Hartree-Fock. Uma matriz Fock é construída e as equações de Hartree-Fock são resolvidas de modo iterativo. As aproximações ocorrem na construção da matriz Fock, ou seja, nas expressões de energia.

Os métodos semiempíricos apresentam em comum, a característica de todas as integrais de superposição, que envolvem dois elétrons em átomos diferentes, serem desprezadas (ZDO – *zero differential overlap*).

As principais diferenças entre os diversos métodos semiempíricos se encontram nas integrais que são desprezadas na resolução das equações de Hartree-Fock-Roothaan, no conjunto de funções de base empregado, e nos parâmetros utilizados.

Os métodos semiempíricos consideram apenas os elétrons de valência. O primeiro método a surgir, o Método de Hückel, despreza completamente as integrais de 2-elétrons. A seguir veio o Método de Huckel Estendido, onde alguns efeitos das interações elétron-elétron foram incluídos através de parametrizações, posteriormente foram desenvolvidos os métodos CNDO (*complet neglect of differential overlap*), INDO (*intermediate neglect of differencial overlap*), MINDO (*modified intermediate neglect of differencial overlap*), NDDO (*Neglect of diatomic differential overlap*) e MNDO (*Modified Neglect of Differential Overlap*). A descrição detalhada de cada método pode ser encontrada na monografia de Pople *et al*, 1970 (Pople & Beveridge, 1970).

3.0 Objetivo geral

O presente trabalho tem como objetivo geral, estudar teoricamente as relações quantitativas entre estrutura química e atividade biológica (**QSAR**) a partir dos dados de atividade biológica obtidos experimentalmente para uma determinada série de derivados de Chalcona, cujo fenômeno biológico observado é a inibição *in vitro* da parasitemia ocasionada pelas cepas W2 de *Plasmodium falciparum*, CDC/Indochina III (cloroquina-resistente), agente causador da malária.

3.1 Objetivos específicos

A concretização desta pesquisa ocorrerá objetivando-se especificamente:

- 1) Calcular diversas propriedades moleculares (parâmetros físico-químicos) para todos os derivados de Chalcona utilizando métodos da química teórica computacional;
- 2) Identificar entre as propriedades moleculares calculadas quais poderão ser tomadas como descritores moleculares, e assim avaliar quais desses descritores estão diretamente correlacionados com a atividade biológica observada utilizando a técnica estatística regressão linear múltipla para estabelecer a relação entre os dados experimentais e os dados calculados;
- 3) Selecionar os melhores modelos matemáticos lineares apresentando três, quatro e cinco variáveis independentes;
- 4) Propor uma estimativa de modelo matemático multidimensional linear que correlacione as propriedades físico-químicas e estruturais dos derivados de Chalcona com a atividade biológica observada;
- 5) Formular hipóteses em nível atômico-molecular na tentativa de explicar o significado dos resultados obtidos, levando em consideração as propriedades moleculares nos modelos selecionados, para que posteriormente possamos prever um provável mecanismo de ação destas substâncias químicas.

4.0 Metodologia utilizada

A primeira etapa concretizada neste trabalho visando o estudo das relações quantitativas entre estrutura química e atividade biológica (**QSAR**) dos derivados de Chalcona consistiu em determinar a geometria molecular de estruturas químicas estáveis (menor energia potencial), para que posteriormente na próxima etapa desta pesquisa, sejam executados os cálculos de propriedades físico-químicas e estruturais.

4.1 Minimização energética e otimização da geometria

Seja qual for o programa utilizado no desenho tridimensional das estruturas químicas, as moléculas não estão inicialmente na geometria mais adequada, isto é, em sua conformação mais estável. Durante o processo de desenho da estrutura química, ocorrerão distorções em ligações químicas, ângulos entre as ligações e ângulos diedrais (Horvath, 1992). Além das distorções, ocorrerá também interação entre átomos não ligantes, competindo para com a mesma região no espaço, justificando assim, as elevadas repulsões estéricas, aumentando o valor absoluto da componente função potencial $V_{\text{non-bonded}}$, seja por interação de van der Waals ou interação eletrostática Coulombiana. Deste modo, a molécula desenhada não está em conformação estável ou mais estável, necessitando de alterações em sua estrutura inicial de forma a reduzir a energia estérica (V_{total}). A minimização energética e a otimização da geometria da molécula inicial podem ser realizadas por métodos mecânicos moleculares, quânticos ou combinação dos dois métodos (Carvalho *et al*, 2003). No processo de otimização da geometria de uma molécula ocorrerão ajustes nas coordenadas atômicas no sentido em obter uma conformação com menor energia potencial possível.

A minimização energética e a otimização da geometria molecular de uma dada estrutura química pode ser realizada computacionalmente executando três procedimentos:

1) Cálculos das derivadas da função energia potencial – determinação das estruturas estáveis pelo método dos gradientes;

2) Mapeamento da hiper-superfície de energia – determinação das estruturas estáveis através de métodos estocásticos;

3) Evolução temporal do sistema molecular – determinação das estruturas estáveis através de dinâmica molecular.

Os procedimentos utilizados na minimização da energia das estruturas moleculares correspondem aos métodos de cálculo das derivadas de energia potencial que é realizado pelos métodos gradientes, que correspondem:

- Método gradiente *steepest descents* ou método do máximo declive;
- Método Newton-Raphson ou método gradiente conjugado;
- Método Fletcher-Powell;
- Método simplex ou
- Combinação de métodos citados acima.

Os procedimentos mais comuns utilizados na minimização da energia estérica das moléculas são os métodos de cálculo das derivadas da função energia potencial (métodos dos gradientes) (Mundim, 2003).

A informação química de uma estrutura está intimamente associada ao arranjo tridimensional desta e às propriedades eletrônicas de regiões específicas de uma determinada molécula. Sabe-se que a minimização da energia estérica de um modelo inicial resulta conformações estáveis. Entretanto, a estrutura originada pela minimização energética não necessariamente é a mais estável, pois ao executar o cálculo de minimização, o programa pára ao encontrar mínimos de energia estérica, mas não necessariamente será um mínimo correspondente à conformação mais estável (Carvalho *et al*, 2003).

Os mínimos de energia encontrados representam conformações estáveis e são chamados de *mínimos locais*. Os *mínimos locais* fornecem informações a respeito do arranjo espacial de uma das muitas conformações encontradas.

A estabilidade molecular é explicada em termos de energia potencial. Uma molécula pode apresentar várias conformações com baixa energia potencial. Se o programa encontrar o estágio molecular de menor energia estérica, então, ter-se-á

a informação a respeito do arranjo espacial mais estável. Este estado energético da molécula é denominado de *mínimo global* sendo que a estrutura química só pode apresentar apenas um estado assim. Entre um *mínimo local* e o *mínimo global* existe uma barreira de energia potencial que conduz a um estágio molecular de mais alta energia potencial. Este estágio molecular é considerado estado de transição molecular e pode também ser chamado de *saddle point* (Hehre *et al*, 1998).

O grande dilema consiste em saber qual será a verdadeira conformação biologicamente ativa. Por mais que façamos uma análise conformacional levando em consideração as ligações com possibilidade de giro rotacional em cada estrutura química obtida pós-minimização energética, nada garante que a estrutura química resultante pós-análise conformacional seja a conformação biologicamente ativa.

Neste trabalho, preocupamos em obter uma sistemática metodológica consistente no sentido em minimizar a energia estérica das estruturas químicas para cada derivado de Chalcona. Procuramos então, encontrar um procedimento onde cada estrutura química resultante da minimização energética apresentasse a menor energia estérica possível, pois o processo de minimização da energia potencial permite obter estruturas químicas estáveis e com uma maior probabilidade de se assemelhar à conformação bioativa.

Após realizar uma intensa busca metodológica, reconhecemos uma abordagem de otimização geométrica que resultava estruturas químicas com uma baixa energia potencial para todas os derivados de Chalcona. A abordagem mencionada a seguir foi utilizada na etapa de minimização energética e de otimização da geometria molecular para todos os derivados de Chalcona, sendo concretizada com auxílio dos programas computacionais:

- a) Molecular Modeling Pro (MMP) (versão 4.0) (ChemSW® Inc., 2001) e
- b) Hyperchem 6.0 (Hypercube® Inc., 2000).

A abordagem metodológica envolvida neste trabalho no intuito em otimizar a geometria molecular para cada derivado de Chalcona consistiu primeiramente em desenhar e visualizar tridimensionalmente as estruturas químicas de todos os

derivados, utilizando o programa computacional MMP. Após ter efetuado o desenho, executou-se uma minimização inicial da energia estérica para cada derivado pelo método mecânico molecular (MM2) via programa MMP. Posteriormente, a geometria molecular de cada estrutura química foi otimizada utilizando método semi-empírico PM3 (Parametric Method 3) com auxílio do programa Hyperchem. As opções utilizadas para todos os derivados de Chalcona no programa Hyperchem via método semi-empírico PM3 foram:

Menu Setup: Semi-empirical PM3	Charge and Spin: Total charge = 0 Spin multiplicity = 1 Spin pairing = RHF	SCF controls: Accelerate convergence State: Lowest
Menu compute: Geometry optimization	Algoritmo: Steepest descent	RMS gradiente: 0,1 Kcal/ Å mol ou 2000 ciclos in vácuo

Finalmente, executamos uma minimização simplex utilizando constante dielétrica (ϵ), 78,54 (meio aquoso). Este último procedimento foi realizado com auxílio do programa MMP com a seguinte abordagem:

Menu Geometry: Minimize geometry	Minimization: Método simplex Maximum iterations = 1000	Strain unbonded: Potencial 6-12 Lennard-Jones Constante dielétrica = 78,54 Use H-bond function
--	---	--

O procedimento da minimização da energia total para cada estrutura química baseou-se no método gradiente *steepest descents*, seja executando mecânica molecular MM2 em MMP ou então executando mecânica quântica semi-empírica PM3 em HyperChem. O método alternativo de minimização chamado de *d downhill simplex* ou *método simplex* implementado no programa MMP foi também utilizado no intuito de diminuir ainda mais a energia total das estruturas químicas.

Durante o procedimento de concretização da geometria molecular para cada derivado de Chalcona anotou-se a energia total de cada estrutura química obtida após a realização de cada passo mencionado e as estruturas resultantes foram armazenadas em arquivos do tipo “*MDL Molfiles*” no programa MMP. A conformação final obtida pós-minimização simplex em MMP foi utilizada posteriormente para execução dos cálculos de propriedades físico-químicas.

4.2 Cálculo de parâmetros físico-químicos

Visto que um número enorme de compostos análogos, podem ser sintetizados levando em consideração os protótipos com possíveis combinações estruturais de substituintes, então é claramente vantajosa a utilização de uma abordagem racional para decidir qual (ais) substituinte (es) exercem importância na atividade biológica dos compostos pesquisados (Thomas, 2003). Deste modo, a pesquisa QSAR tem provado ser muito útil em tentar correlacionar quantitativamente propriedades moleculares dos compostos com a atividade biológica.

A segunda etapa deste trabalho deu-se no cálculo das diversas propriedades moleculares focando a estrutura química com a sua geometria já otimizada. Com a disponibilidade da estrutura química de um conformero estável de cada derivado de Chalcona, diversas propriedades moleculares foram calculadas e estimadas computacionalmente.

No trabalho em questão, utilizamos a estrutura resultante obtida após termos executado a minimização simplex com constante dielétrica 78,54 (meio aquoso) no cálculo das diversas propriedades moleculares, visto que as interações intermoleculares que justificam a atividade biológica ocorrerão em meio

aquoso. Deve-se levar em consideração que os valores da energia total de cada molécula calculado *in vácuo* e em meio aquoso são diferentes (Kubinyi, 1993). Os programas utilizados na execução dos cálculos de propriedades moleculares foram:

- a) Molecular Modeling Pro (MMP) – Versão 4.0 (ChemSW® Inc., 2001) e
- b) ChemSite Pro (CSP) – Versão 5.0 (ChemSW® Inc., 2001)

Em QSAR, as propriedades moleculares calculadas são denominadas de parâmetros físico-químicos e, posteriormente, na terceira etapa desta pesquisa (análise estatística) serão chamadas de *variáveis independentes* ou *variáveis explicativas*.

Os parâmetros físico-químicos utilizados em QSAR podem ser subdivididos em parâmetros de natureza lipofílica, eletrônica, estereoquímica, polar, termodinâmica, biodisponibilidade e outros. No item 4.2.1 indicaremos as propriedades moleculares calculadas, qual foi o método de cálculo envolvido, e o programa utilizado para estimar os valores das propriedades moleculares.

4.2.1 Parâmetros hidrofóbicos

São parâmetros relacionados à solubilidade e ao transporte do fármaco em meio biológico. Os parâmetros lipofílicos constituem medidas do caráter hidrofóbico dos fármacos e estão associados à passagem destas substâncias através de membranas biológicas, podendo também ser importante em algumas interações fármaco-receptor. A mudança nos substituintes de um protótipo pode alterar significativamente o caráter hidrofóbico e consequentemente a sua atividade biológica. Não podemos esquecer que substâncias com elevado caráter lipofílico atravessam membranas biológicas com facilidade, interagindo com maior intensidade às proteínas plasmáticas e podem passar pela barreira hematoencefálica, atingindo o Sistema Nervoso Central (SNC). Deste modo, é importante termos uma previsão quantitativa do caráter hidrofóbico dos fármacos para correlacionar a sua estrutura química com a atividade biológica.

Nesta pesquisa QSAR, várias propriedades moleculares relacionadas à solubilidade e hidrofobicidade foram previstas com o intuito de associar

quantitativamente a atividade biológica (-Log IC₅₀) dos derivados de Chalcona com a (s) propriedade (s) escolhida (s).

As propriedades comentadas a seguir foram calculadas com auxílio do programa computacional Molecular Modeling Pro.

1) Log do coeficiente de partição (P) n-octanol/água (Log K_{ow} = Log P_{Hansch}) – Cálculo da estimativa do Log P utilizando constantes de fragmento - Método modificado de Hansch & Leo (Hansch & Leo, 1979). O cálculo do Log P pelo método de Hansch é dado pela soma das contribuições hidrofóbicas dos fragmentos que constitui a molécula do composto (Wang *et al*, 1997).

O coeficiente de partição (P) é atualmente descrito pela razão entre a concentração da droga em fase orgânica (C_{org}) pela concentração em fase aquosa (C_{aq}). Assim, fármacos com elevado coeficiente de partição, apresentam maior afinidade por meios orgânicos, atravessando com facilidade biomembranas. O coeficiente de partição é tradicionalmente determinado empregando o *n*-octanol como fase orgânica devido a sua semelhança estrutural com os fosfolípídeos de membrana. Emprega-se o Log P para correlacionar a atividade biológica de substâncias com o coeficiente de partição.

$$\text{Coeficiente de partição (P)} = C_{\text{org}}/C_{\text{aq}} \quad (43)$$

Se $P > 1$, o composto é mais lipofílico, pois $C_{\text{org}} > C_{\text{aq}} \rightarrow$ Maior afinidade em meios hidrofóbicos;

Se $P < 1$, o composto é mais hidrofílico, pois $C_{\text{org}} < C_{\text{aq}} \rightarrow$ Maior afinidade em meios hidrofílicos.

2) Log P_{Ghose & Crippen} – Cálculo da estimativa do Log P baseado na soma de contribuições hidrofóbicas de tipos de átomos – Método de Ghose & Crippen (Viswanadhan *et al*, 1989).

O cálculo do Log P pelo método de Ghose & Crippen é baseado em contribuições hidrofóbicas aditivas atômicas (Wang *et al*, 1997). As contribuições levam em consideração os tipos de átomos que constituem a molécula e seus ambientes topológicos, ou seja, características como tipos de átomos, os ligantes

destes átomos, a eletronegatividade dos elementos e o tipo de hibridação são fatores que contribuem diferentemente no valor global do Log P.

3) HLB (Balanço hidrofílico-lipofílico) – Indica a razão entre o peso molecular da porção hidrofílica e o peso molecular total multiplicado pelo fator 20 – Método de Griffin (www.norgwyn.com/hsa.zip).

$$\text{HLB} = (\text{PM}_{\text{hidrofílico}} / \text{PM}_{\text{total}}) \times 20$$

onde PM – peso molecular (44)

Esta medida determina a proporção da molécula em termos de massa que é hidrofílica em uma escala de 0 a 20. A grande contribuição do HLB reside na interpretação da relação entre a natureza da estrutura química e a densidade de energia coesiva. Moléculas não polares apresentam baixa densidade de energia coesiva, enquanto moléculas polares apresentam valores intermediários e substâncias que apresentam ligações de hidrogênio possuem altos valores de densidade de energia coesiva (Meyers, 1999).

Estruturas que apresentam HLB entre 7-18 são consideradas hidrofílicas, enquanto as que possuem HLB entre 3-6 são hidrofóbicas e substâncias com um HLB superior a 14 serão solúveis em meio aquoso.

4) Parâmetro de solubilidade (P.S) – Parâmetro determinado empiricamente e que indica a solubilidade das estruturas em $\text{delta/MPa}^{0.5}$. É calculado em função dos parâmetros de solubilidade de Hansen 3D, ou seja, raiz quadrada da soma dos quadrados da Dispersão Hansen 3D, Polaridade Hansen 3D e Ligação de Hidrogênio Hansen 3D (Barton, 1983).

5) Parâmetros de solubilidade de Hansen 3D – É determinado em função da somatória de constantes dos fragmentos, sendo que os valores das constantes dos fragmentos são dados tabelados e obtidos experimentalmente por Hansen.

5.1) Dispersão de Hansen 3D (D. Hansen 3D) – Medida das forças repulsivas entre as moléculas, ou seja, mede a tendência para dispersão molecular (Barton, 1983).

5.2) Polaridade de Hansen 3D (δ Hansen 3D) – Medida da polaridade das moléculas (Barton, 1983).

A polaridade da molécula é uma propriedade elétrica relacionada com as ações dos núcleos que competem pelas cargas dos elétrons levando à existência de um momento de dipolo elétrico (μ). Moléculas polares são as que apresentam momento de dipolo elétrico permanente. O momento de dipolo permanente é proveniente das cargas atômicas parciais nas moléculas ocasionadas pelas diferenças de eletronegatividade entre os átomos ou por características de ligação química (Atkins, 1999). A polaridade é uma propriedade molecular importante na compreensão dos processos de solubilidade e nas interações intermoleculares entre fármaco e o meio que este encontra disperso.

5.3) Ligação de Hidrogênio de Hansen 3D (H.B. Hansen 3D) – Medida da tendência de uma molécula em formar ligações de hidrogênio (Barton, 1983).

A ligação de hidrogênio é uma interação atrativa que ocorre entre moléculas polares e que apresentam elevado momento de dipolo elétrico. A ligação de hidrogênio é estabelecida entre um átomo de hidrogênio de uma molécula com um átomo muito eletronegativo de outra molécula que apresenta par isolado de elétrons.

6) Área superficial hidrofílica (ASH) – Medida da área superficial ($\text{cm}^2/\text{mol} \cdot 10^9$) da molécula que possui afinidade com moléculas de água. Utiliza um algoritmo para determinar quais átomos são hidrofílicos e compara com a área superficial total da molécula – Método de Griffin (www.norgwyn.com/hsa.zip).

O programa calcula a área superficial de cada átomo através do raio de van der Waals (Bondi, 1964). A área de overlap entre átomos ligados também é determinada e posteriormente subtraída da área superficial total. Posteriormente, o programa utiliza um valor inicial de hidrofilicidade (tabelado) para cada átomo, realizando os devidos ajustes de hidrofilicidade para os átomos vizinhos. Assim, usando valores de hidrofilicidade ajustados para cada átomo, os átomos são caracterizados como lipofílicos ou hidrofílicos. Finalmente, as áreas superficiais dos átomos classificados como hidrofílicos são somadas para resultar a área superficial hidrofílica da molécula (Quinn, 2003).

7) % Área superficial hidrofílica (% ASH) – Indica a porcentagem de área superficial da molécula que possui afinidade por moléculas de água. Calcula a razão entre a área superficial hidrofílica e a área superficial total da molécula multiplicado pelo fator 100 – Método de Griffin (www.norgwyn.com/hsa.zip).

8) Log solubilidade em água (Log S. H₂O) – Calcula uma previsão para o Log da solubilidade (mol/m³) do composto em água - Método de Klopman *et al* (Klopman *et al*, 1992).

9) Número de Hidratação (NH) – Determina o número de moléculas de água carregadas à molécula em solução - Método da adição de fragmentos de McGowan (McGowan, 1990).

10) Refratividade molar (MR_{Ghose & Crippen}) – Determina a refatividade molar da estrutura em função da somatória dos valores de contribuição da refração atômica de cada átomo - Método de Ghose & Crippen (Viswanadhan *et al*, 1989).

A refatividade molar está relacionada com o volume molecular (VM) através da equação:

$$MR = \frac{n^2 - 1}{n^2 + 2} \cdot VM \quad (45)$$

onde, n = índice de refração.

4.2.2 Parâmetros eletrônicos

São parâmetros relacionados à distribuição eletrônica sobre uma molécula, à facilidade com que essa distribuição pode ser modificada e estão associados às forças de interação fármaco-meio e fármaco-receptor.

Nas interações fármaco-meio e fármaco-receptor estão envolvidos uma série de mecanismos que dependem da distribuição dos elétrons na molécula, que por sua vez estão relacionados à distribuição de cargas atômicas parciais nestas moléculas. Não podemos esquecer que as drogas em meio aquoso estão sujeitas

a mecanismos de ionização, portanto a passagem dos fármacos por membranas biológicas dependerá do pH do meio e do pK destas substâncias.

O mecanismo de passagem dos fármacos pelas biomembranas está intimamente associado à constante de ionização destes, visto que, espécies químicas não ionizadas atravessam as membranas com maior facilidade por serem mais lipofílicas. Espécies químicas que sofrerão ionização no meio aquoso são carregadas eletricamente e normalmente encontram-se solvatadas por moléculas de água, o que dificulta o processo de difusão passiva pelas membranas (Barreiro & Fraga, 2001).

Os parâmetros eletrônicos podem ser subdivididos em duas categorias:

A) Parâmetros eletrônicos clássicos ou empíricos: São medidas determinadas experimentalmente e que caracterizam os efeitos provocados pelos substituintes na molécula. Neste trabalho, foram estimados os valores das constantes de Hammett (σ) para os substituintes A e B em todos os derivados de Chalcona, com auxílio do programa Molecular Modeling Pro.

11) Constante de Hammett (σ) – Parâmetro eletrônico empírico que mede a habilidade do substituinte na molécula quanto ao poder em ceder ou atrair elétron. A distribuição dos elétrons em uma molécula depende da natureza dos grupos elétron-retirantes e dos grupos elétron-doadores encontrados em uma estrutura química. Hammett utilizou os conceitos de equilíbrio e deslocamento do equilíbrio químico para calcular as constantes de Hammett (σ_x) para vários substituintes no ácido benzóico (padrão de referência) (Thomas, 2003). As constantes de Hammett dos substituintes foram determinadas pela equação:

$$\sigma_x = \text{Log } K_x/K_a \Leftrightarrow \sigma_x = \text{Log } K_x - \text{Log } K_a \quad (46)$$

onde K_x = constante de ionização do composto substituído;

K_a = constante de ionização do ácido benzóico(padrão).

Portanto, podemos perceber que:

➤ Se $\sigma_x < 0$ (negativo) $\Leftrightarrow \text{Log } K_x < \text{Log } K_a \Leftrightarrow K_x < K_a$, então o substituinte X é um grupo doador de elétron;

➤ Se $\sigma_x > 0$ (positivo) $\Leftrightarrow \text{Log } K_x > \text{Log } K_a \Leftrightarrow K_x > K_a$, então o substituinte X é um grupo retirante de elétron.

Como a constante de Hammett leva em consideração os efeitos de ressonância e os efeitos indutivos, então os valores de σ irá depender da natureza do substituinte e da sua posição no anel aromático.

Nesta pesquisa foram determinados, Hammett sigma para (σ_p), Hammett sigma meta (σ_m), Hammett sigma star (σ^*) e Hammett sigma induction (σ_I), para os substituintes A e B (figura 20).

11.1) Hammett sigma para (σ_p) – mede a contribuição do efeito eletrônico que ocorre com substituintes aromáticos substituídos na posição para;

11.2) Hammett sigma meta (σ_m) – mede a contribuição do efeito eletrônico ocorrido com substituintes aromáticos substituídos na posição meta;

11.3) Hammett sigma star (σ^*) – mede a contribuição do efeito eletrônico ocorrido com substituintes alifáticos;

11.4) Hammett sigma induction (σ_I) – mede a contribuição do efeito eletrônico indutivo para componentes doadores de elétrons.

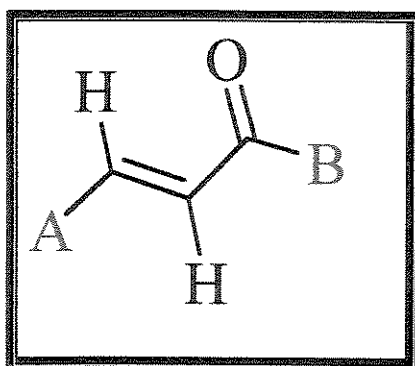


Figura 20: Estrutura protótipo dos derivados de Chalcona.

B) Parâmetros eletrônicos quânticos: São propriedades calculadas por intermédio de métodos mecânicos-quânticos. As propriedades de natureza eletrônica quântica foram calculadas no programa Molecular Modeling Pro (MMP) utilizando o método semi-empírico CNDO (*Compleat Neglect of Differential Overlap closed shell*).

Este método despreza as integrais de troca, retém as integrais de 2-elétrons do tipo $(\mu\mu|vv)$ e considera uma simetria esférica para os orbitais atômicos. Cálculos pelo método CNDO, usam aproximações Hartree-Fock para resolver a equação de Schrödinger e a repulsão eletrônica entre elétrons em diferentes orbitais depende apenas da natureza dos átomos envolvidos e não do tipo de orbital envolvido. O método CNDO usa duas aproximações:

- a) Aproximação do núcleo e
- b) Aproximação ZDO (*Zero differential overlap*).

Em Molecular Modeling Pro, ao rodar CNDO *closed shell*, o programa acompanhará o desenvolvimento das cargas atômicas parciais e do momento de dipolo. Ao final, obtivemos também os valores da contribuição doadora de ligação de hidrogênio (H. bond donor), contribuição aceptora de ligação de hidrogênio (H. bond acceptor), E_{Homo} e E_{lumo} .

12) Momento de dipolo (μ) – Foi utilizado método semi-empírico CNDO/2 *closed shell* para calcular o momento de dipolo em debyes da molécula (Pople & Beveridge, 1970).

A diferença de eletronegatividade relativa entre átomos em uma ligação química resulta cargas elétricas puntiformes $+q$ e $-q$ separadas a uma dada distância (r). Esta configuração de cargas é representada por um vetor denominado de momento de dipolo de cargas pontuais (μ_p), onde o módulo deste vetor é determinado matematicamente pela equação:

$$\mu_p = q \cdot r$$

(47)

O momento dipolar total (μ_T) em uma molécula é calculado pela somatória do momento de dipolo de carga pontual (μ_p) com o momento dipolar de hibridação (μ_H):

$$\mu_T = \mu_p + \mu_H \quad (48)$$

O momento de dipolo de carga pontual é determinado pela somatória do produto entre cada carga atômica (q_n) e o seu respectivo vetor posição (r_n):

$$\mu_p = \sum q_n \cdot r_n \quad (49)$$

13) Cargas atômicas parciais (q) – Foi utilizado método semi-empírico CNDO/2 *closed shell* para determinar a carga atômica líquida (Pople & Beveridge, 1970).

Funções de onda moleculares escritas como a combinação linear de orbitais atômicos (CLOA) dentro da aproximação ZDO, os elementos diagonais da matriz densidade, $P_{\mu\mu}^\alpha$ e $P_{\mu\mu}^\beta$, resulta a população eletrônica no orbital atômico em questão. A somatória das populações de todos os orbitais centrados em um dado átomo B resulta a população eletrônica total do átomo B (P_{bb}), sendo então:

$$P_{bb} = \sum (P_{\mu\mu}^\alpha + P_{\mu\mu}^\beta) \quad (50)$$

onde α e β são os spins eletrônicos.

A carga atômica líquida de cada átomo é determinada pela subtração entre a sua carga nuclear (Z_b) e a sua população eletrônica (P_{bb}):

$$q_b = Z_b - P_{bb} \quad (51)$$

Devido a dificuldade existente em obter valores mais precisos de cargas atômicas parciais, recomenda-se que este parâmetro seja utilizado como estudo

comparativo entre compostos de uma mesma série. É de grande importância escolhermos como parâmetro a carga parcial de átomos em comum a todos os compostos da série. Nesta pesquisa foram calculadas cargas atômicas parciais para o núcleo α - β insaturado de todos os derivados de Chalcona, ou seja, para os átomos: $q(C_1)$, $q(C_2)$, $q(C_3)$, $q(C_4)$, $q(C_8)$ e $q(O_7)$ (figura 21). Preocupamos em determinar as cargas atômicas parciais apenas para os átomos da estrutura protótipo por tal região apresentar relativa contribuição de ressonância.

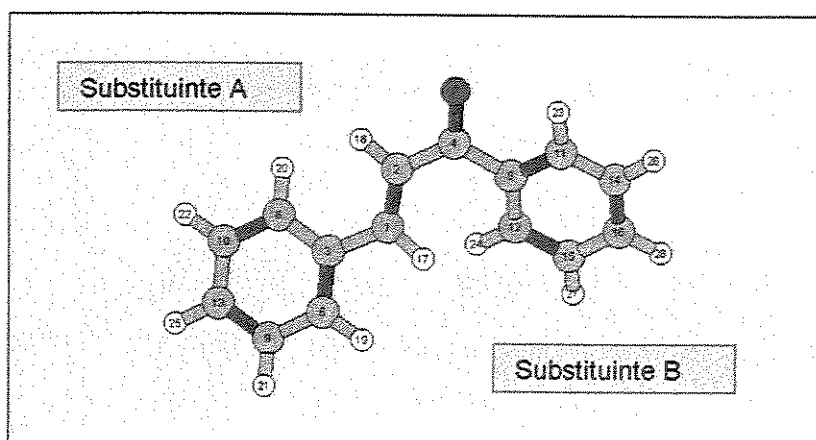


Figura 21 - Estrutura química base dos derivados de Chalcona com a numeração fornecida via utilização do programa Molecular Modeling Pro

- Ligação mais escura – Insaturação.

14) Hydrogen bond acceptor (H. B. acceptor) – Calcula a contribuição que a molécula possui em receptionar ligações de hidrogênio pelo método CNDO/2 *closed shell*. Determinado pela soma das cargas atômicas parciais menores que $-0,15$.

$$\text{H. B. acceptor} = \sum q (q < -0,15)$$

(52)

15) Hydrogen bond donor (H. B. donor) – Calcula a contribuição que a molécula apresenta em fornecer ligações de hidrogênio pelo método CNDO/2 *closed shell*. Determinado pela soma das cargas atômicas parciais maiores que $+0,03$.

$$\text{H.B. donor} = \sum q (q > +0,03)$$

(53)

16) Energia dos orbitais de Fronteira (E_{HOMO} e E_{LUMO}) – Os orbitais HOMO (orbital molecular ocupado de maior energia) e LUMO (orbital molecular desocupado de menor energia) foram denominados orbitais de fronteira (Fukui *et al*, 1954). Estes orbitais apresentam importância relevante em reações químicas de oxi-redução, pois nestas reações ocorre um mecanismo de transferência de elétrons, ou seja, espécie química fornecendo elétron e uma outra recepcionando elétron. Em reações do tipo REDOX, a energia envolvida para remover um elétron da molécula (E_{HOMO}) é proporcional ao potencial de ionização, enquanto a energia envolvida para que uma molécula receba um elétron (E_{LUMO}) é proporcional a eletroafinidade da molécula.

A aproximação entre moléculas em processos químicos desencadeia interações entre os orbitais destas moléculas que por consequência são acompanhadas com variações de energia. Quando as variações de energia ocorrem no sentido em diminuir a energia global do sistema em estudo, a estabilidade deste sistema tende a aumentar.

A interação entre o orbital HOMO de uma molécula com orbital HOMO de outra molécula conduz a um estágio de mais alta energia visto que este processo é endotérmico. A interação entre o orbital HOMO de uma molécula e o orbital LUMO de uma outra molécula conduz a um estágio de menor energia, pois este mecanismo é exotérmico. Deste modo, a combinação entre orbitais ocupados (HOMO) e desocupados (LUMO), resulta em um maior abaixamento energético no sistema molecular devido a proximidade entre os dois orbitais.

16.1) E_{HOMO} (*Highest occupied molecular orbital*) – Calcula o conteúdo energético do *orbital molecular ocupado de mais alta energia*. Representa a energia necessária para arrancar um elétron da camada de valência da molécula, ou seja, esta propriedade está intrinsecamente focada na facilidade com que a molécula pode doar elétron.

16.2) E_{LUMO} (*Lowest unoccupied molecular orbital*) – Calcula o conteúdo energético do *orbital molecular desocupado de mais baixa energia*. Representa a facilidade com que uma molécula pode aceitar um elétron.

4.2.3 Parâmetros estereoquímicos

Para que um fármaco possa interagir com um determinado receptor, é necessário entender quais propriedades dimensionais da molécula são relevantes na interação com a região ativa do receptor. Grupos substituintes volumosos podem dificultar a interação fármaco-receptor através do impedimento estérico, como podem também ajudar na orientação de interações com região ativa do receptor, intensificando assim a atividade biológica. Pode-se dizer que os parâmetros estereoquímicos estão diretamente relacionados com as características topológicas da molécula (Patrick, 2001).

O reconhecimento molecular ligante-receptor está associado às interações entre a micromolécula (droga) e a biomacromolécula (receptor). Desta forma, o volume molecular, os parâmetros dimensionais do ligante e dos substituintes presente na molécula do ligante, juntamente com o arranjo espacial são características estruturais de extrema importância na elucidação das forças relevantes envolvidas no reconhecimento molecular ligante-receptor.

Os parâmetros estereoquímicos mencionados a seguir foram calculados utilizando os programas: Molecular Modeling Pro (MMP) e ChemSite Pro (CSP).

As propriedades moleculares calculadas em MMP foram: volume molecular, densidade molecular, comprimento molecular, largura molecular, profundidade molecular e parâmetros estéricos de Verloop.

17) Volume molecular (VM) – Indica o volume da molécula em cm^3/mol – Método de Bondi (Bondi, 1964). O programa calcula o volume molecular pelo princípio elementar de geometria 3D. Levando em consideração os raios de van der Waals determina-se o volume das esferas que posteriormente é subtraído dos volumes relativos aos overlaps atômicos.

O volume molecular permite avaliar o efeito estérico ao modificar os substituintes da molécula. O tamanho dos substituintes inter-relaciona com o volume total da molécula devido às diversas interações ocorridas entre os átomos na molécula. Ao utilizarmos a estrutura química com a sua geometria já otimizada para o cálculo do volume molecular, estamos contando com uma estrutura

química com uma menor intensidade em termos de tensão energética no que diz respeito aos overlaps atômicos.

18) Densidade molecular (ρ) – A densidade da molécula é determinada efetuando a divisão do peso molecular (PM) pelo volume molecular (VM) e com a realização dos devidos ajustes para os fragmentos orgânicos e para as contribuições da superfície de cada átomo (Bondi, 1964).

19) Parâmetros dimensionais (L_x , L_y e L_z) - Uma vez caracterizada a geometria molecular para cada derivado de Chalcona, com auxílio do programa MMP, calculamos as propriedades dimensionais para cada estrutura. Para calcular as dimensões de cada molécula, o programa gira a molécula em incremento de 5 graus ao longo dos eixos Y e Z por 360 graus, encontrando as dimensões de máximo e mínimo, que logo são informadas. O programa oferece a opção de orientar as moléculas em dimensões de máximo ou mínimo ao longo dos eixos X, Y ou Z. Uma vez perguntado qual seria a opção dimensional (máximo ou mínimo), preferimos então escolher as opções de máximo dimensional ao longo dos eixos X, Y e Z. Este procedimento foi realizado para todos os derivados de Chalcona, fixando suas estruturas em um máximo comprimento molecular, largura molecular e também profundidade molecular. Assim, foram calculadas as propriedades dimensionais:

19.1) Comprimento molecular (L_x) – maior extensão da molécula ao longo do eixo X;

19.2) Largura molecular (L_y) – maior extensão da molécula ao longo do eixo Y;

19.3) Profundidade molecular (L_z) – maior extensão da molécula ao longo do eixo Z.

20) Parâmetros estereoquímicos de Verloop (ϑ) – Cálculo das medidas dos valores estéricos dos substituintes – Método de Verloop (Verloop *et al*, 1976).

Os parâmetros estéricos de Verloop são propriedades dimensionais dos substituintes, melhor dizendo, são propriedades correspondentes ao comprimento (L_1) e aos raios (B1-B4) dos grupos substituintes.

O comprimento e os raios dos substituintes são características estruturais importantes para uma melhor compreensão da complexação droga-receptor, visto que o mecanismo de ação das drogas baseia-se em um processo de complementaridade espacial entre fármaco-receptor. Esta estereo-especificidade entre os fármacos e os seus respectivos alvos é justificada por intermédio de interações bioquímicas que ocorrem em locais específicos dos receptores denominados de centros ativos. É compreensível pesquisarmos as características dimensionais dos grupos substituintes nas moléculas dos compostos pesquisados para que possamos descobrir a possível região farmacofórica na molécula e assim correlacionar quais parâmetros estéricos se responsabilizam pela atividade biológica.

Os parâmetros estereoquímicos de Verloop (L_1 , B_1 , B_2 , B_3 e B_4) foram estimados para os substituintes A e B em todos os derivados de Chalcona.

20.1) (ə) L_1 – indica o comprimento do substituinte ao longo do eixo formado pela ligação entre este substituinte e o átomo que este se encontra conectado.

Os parâmetros B_1 , B_2 , B_3 e B_4 correspondem aos raios extendidos perpendicularmente em relação ao eixo formado pelo parâmetro L_1 .

20.2) (ə) B_1 – indica menor raio do substituinte;

20.3) (ə) B_2 – indica o segundo menor raio do substituinte;

20.4) (ə) B_3 – indica o segundo maior raio do substituinte;

20.5) (ə) B_4 – indica o maior raio do substituinte.

Outras propriedades estereoquímicas foram calculadas utilizando o programa CSP:

21) Superfície de van der Waals (SvdW) – Calcula a estimativa da área superficial ($\text{\AA}^2$) da molécula em função do raio de van der Waals das esferas. O princípio do cálculo é baseado na determinação espacial de uma região tridimensional limite pelas regiões não interseccionais das esferas com um raio de van der Waals pré-fixado, onde o programa sugere o raio desejado. Ao especificar um raio igual a zero, o programa irá calcular a área superficial do envelope de van der Waals.

22) Superfície de van der Waals acessível ao solvente (SvdWsol) – Determina a estimativa da área superficial ($\text{\AA}^2$) da molécula acessível ao solvente. O princípio do cálculo é o mesmo utilizado na execução do cálculo da área superficial de van der Waals, sendo que a diferença está na indicação do raio que deverá ser especificado. O raio especificado no programa para execução do cálculo da superfície de van der Waals acessível ao solvente água foi de 1,4 $\text{\AA}$.

4.2.4 Parâmetros termodinâmicos

Os parâmetros termodinâmicos foram estimados utilizando o programa Molecular Modeling Pro.

- Ponto de ebulição (PE) e Parachor (PA) – Estas propriedades foram estimadas pelo método do “Handbook of Chemical Property Estimation” (Lyman *et al*, 1982);

23) Ponto de ebulição (PE) de um determinado composto refere-se à temperatura que a fase líquida desta substância entra em fervura, ou seja, é o estágio em que a pressão de vapor deste líquido torna-se igual à pressão externa.

O ponto de ebulição de uma substância é uma grandeza que está relacionada com o tipo de interação intermolecular da substância em questão. Quanto maior for a interação intermolecular em uma substância, maior será o ponto de ebulição normal. Compostos que apresentam ligações de hidrogênio entre suas moléculas apresentarão uma tendência a possuir um maior PE se comparado a estruturas químicas de mesma massa molar, mas com interações intermoleculares mais fracas. Compostos que apresentam menor PE são mais voláteis e ganham a fase gasosa com mais facilidade, pois sua pressão máxima de vapor é menor. A temperatura de ebulição (TE) se relaciona com a temperatura crítica (TC) pela equação:

$$TE / TC = 0,584 + 0,965 \sum \Delta T - (\sum \Delta T)^2$$

sendo que ΔT é uma contribuição tabelada; (54)

24) Parachor (PA) é uma propriedade físico-química relacionada com a tensão superficial, peso molecular e a densidade molecular. A expressão para cálculo do parachor é:

$$PA = \gamma^{1/4} \cdot PM / \rho \quad \text{onde } PA = \text{Parachor}; \gamma = \text{tensão superficial}; \quad (55)$$

PM = peso molecular e ρ = densidade molecular.

Alguns autores preferem (Kubinyi, 1993) considerar o parâmetro físico-químico parachor como sendo um parâmetro estereoquímico por estar relacionado com o volume molecular. Não podemos esquecer que o quociente entre o peso molecular (PM) e o volume molecular (VM) representa a densidade molecular, portanto:

$$\rho = PM / VM \quad \Leftrightarrow \quad VM = PM / \rho \quad \Leftrightarrow \quad PA = \gamma^{1/4} \cdot VM \quad (56)$$

Outros autores preferem (Lyman *et al*, 1982) considerar o parâmetro físico-químico parachor como sendo uma propriedade termodinâmica por estar relacionado com a tensão superficial.

25) Número efetivo de ligações torsionais (tau) e número de ligações de hidrogênio (HBN) – Propriedades estimadas utilizando o método de Yalkowsky e co-autores (Myrdal *et al*, 1996);

26) Temperatura crítica (TC), Pressão crítica (PC): Propriedades estimadas pelo método termodinâmico de Joback & Reid (Joback & Reid, 1987).

As propriedades termodinâmicas críticas são propriedades físico-químicas empíricas dos compostos que foram estimadas pelo método de Joback & Reid, que provém de uma modificação do método de Lydersen (Horvath, 1992). As expressões utilizadas para cálculo destas propriedades são:

$$T_{cr} (K) = TE (K) / 0,584 + 0,965 \sum \Delta T - (\sum \Delta T)^2 \quad (57)$$

$$P_{cr} (bar) = 1 / 0,113 + 0,0032 n_a - (\sum \Delta P) \quad (58)$$

sendo que ΔT e ΔP são contribuições tabeladas;

n_a = número de átomos na molécula;

T_{cr} = temperatura crítica (K);

P_{cr} = pressão crítica (bar);

Este método é extremamente simples, mas é muito bom para estimar as propriedades críticas da matéria.

4.2.5 Parâmetros de biodisponibilidade

Os parâmetros de biodisponibilidade foram estimados utilizando o programa Molecular Modeling Pro. A biodisponibilidade de uma substância refere-se à percentagem da droga que após ser administrada no organismo, estará disponível à circulação sistêmica. A parcela de droga disponível depende das seguintes características:

- A) Liberação do fármaco da forma farmacotécnica (forma de apresentação);
- B) Absorção pelo trato gastrointestinal;
- C) Eliminação pré-sistêmica.

A biodisponibilidade absoluta relaciona a quantidade da droga disponível sistemicamente/dose administrada. Deste modo, os parâmetros de biodisponibilidade de um fármaco apresentam relativa importância na compreensão da farmacocinética da droga. Como a velocidade de absorção depende da via de administração da droga, quanto mais rápida ocorrer a absorção, menor será o tempo (t_{max}) necessário para alcançar a concentração plasmática máxima (C_{max}). Quanto maior for a C_{max} , mais cedo o nível plasmático irá cair novamente. A determinação da biodisponibilidade pode ser realizada via construção da curva da concentração plasmática do fármaco em função do tempo, definindo então a área sob a curva de concentração plasmática.

27) Peso molecular (PM) – O peso molecular para cada estrutura foi calculado levando em consideração os respectivos pesos atômicos (Veber *et al*, 2002). O peso molecular aceitável para uma droga em termos de sistema nervoso central (SNC) deverá ser menor que 1000 u.a

28) Área superficial polar (AS_{polar}) – A área superficial ($\text{\AA}^2$) polar para cada estrutura foi calculada (Veber *et al*, 2002). A AS_{polar} aceitável em termos de SNC deverá ser menor que $140 \text{\AA}^2$.

4.2.6 Outros Parâmetros

Os valores das propriedades, distância interatômica (d), ângulo entre as ligações químicas (α) e ângulos diedrais (β) foram anotados e tabelados após realizarmos a minimização simplex em MMP. Preocupamos em obter valores destas propriedades focando apenas a região da estrutura que contém os átomos envolvidos na contribuição de ressonância (estrutura protótipo). Os valores destas propriedades foram extraídos da conformação com a sua geometria já otimizada.

29) Comprimento de Ligação (d) – As distâncias interatômicas envolvidas foram $d(C_1C_2)$, $d(C_2C_4)$, $d(C_1C_3)$, $d(C_4O_7)$ e $d(C_4C_8)$;

30) Ângulo entre ligações (α) – Os ângulos entre ligações químicas envolvidos foram $\alpha(C_3C_1C_2)$, $\alpha(C_1C_2C_4)$, $\alpha(C_2C_4O_7)$, $\alpha(C_2C_4C_8)$ e $\alpha(C_8C_4O_7)$;

31) Ângulo diedral (β) – Os ângulos diedrais envolvidos foram $\beta(C_3C_1C_2C_4)$, $\beta(C_1C_2C_4O_7)$ e $\beta(C_1C_2C_4C_8)$.

4.3 Análise estatística

A atividade biológica de um fármaco é função de sua capacidade em interagir com regiões específicas dos receptores. Deste modo, a atividade biológica de compostos está associada a uma gama de fatores que vão desde a entrada do fármaco no organismo até encontrar o alvo em que ocorrerá a ação farmacológica. Durante este trajeto, a atividade biológica dependerá basicamente do transporte do fármaco até o sítio de ação e da sua complexação com a biomacromolécula alvo.

Através de pesquisas QSAR, é possível identificar e quantificar propriedades físico-químicas e estruturais de uma droga com intuito em caracterizar quais propriedades são relevantes para o entendimento de mecanismos bioquímicos que justifiquem a sua atividade biológica. Se uma determinada droga é ativa, uma relação matemática pode ser construída no sentido de prever a atividade biológica desta droga e de análogos estruturais em função de algumas propriedades moleculares. É possível então obter uma equação matemática que relacione a atividade biológica de uma série de substâncias em função de propriedades moleculares destas substâncias.

O atual sucesso no desenvolvimento de novos fármacos está focado no imenso esforço por parte de alguns pesquisadores e numa rápida e crescente reciclagem computacional. O sucesso depende em grande parte da força de vontade de pesquisadores em estar associado a grupos que estejam imbuídos em desenvolver e aprender sempre novas técnicas, garantindo um crescimento mais racional e assim estar aptos a participar desta *seleção natural do conhecimento*.

A pesquisa QSAR é uma tentativa de retirar o elemento sorte do desenvolvimento e planejamento racional de novas espécies químicas ativas. O estabelecimento de relações que focam um conhecimento multidisciplinar facilita grandiosamente o aprimoramento de técnicas e um avanço frenético em busca de compostos que possam combater inúmeras doenças e assim solucionar de modo eficiente parte dos problemas sociais no mundo.

Uma vez estabelecida a geometria molecular de uma série de compostos, torna-se necessário a realização de diversos cálculos de propriedades moleculares para que posteriormente, possamos por intermédio da análise estatística efetuar uma correlação estrutura química-atividade biológica.

A terceira etapa realizada nesta pesquisa QSAR consistiu em correlacionar as diversas propriedades calculadas anteriormente com a atividade biológica de uma série de derivados de Chalcona como agentes anti-maláricos. É importante lembrar que em QSAR, a atividade biológica corresponde ao fenômeno observado, enquanto as propriedades moleculares de natureza lipofílica, eletrônica, esteroquímica, polar ou de qualquer outra natureza são denominadas variáveis independentes ou explicativas (Gaudio & Zandonadi, 2001).

As pesquisas na área QSAR têm como principal objetivo a construção de modelos matemáticos que correlacionem as propriedades moleculares com a atividade biológica de uma série de compostos análogos. Geralmente estes compostos diferem quanto a presença de um ou mais substituintes em posições definidas de uma determinada estrutura química (Ferreira *et al*, 2002).

A construção de modelos matemáticos requer a elaboração de uma matriz de dados contendo a medida quantitativa experimental da atividade biológica e os parâmetros físico-químicos calculados. Este conjunto de dados é a matéria prima para a construção de modelos matemáticos capazes de justificar a atividade biológica dos compostos (Ferreira *et al*, 2002).

A construção de um modelo pode ser concretizada através da técnica estatística regressão linear múltipla (RLM). Levando em consideração o método de Hansch-Fujita (Hansch & Fujita, 1964; Kubinyi, 1993) conhecido como modelo extratermodinâmico, os valores de atividade biológica são correlacionados aos valores calculados das variáveis independentes através de regressão linear, regressão linear múltipla e regressão não linear (Gaudio & Zandonade, 2001).

O modelo linear de Hansch é uma combinação linear de variáveis independentes, capaz de explicar de melhor maneira possível valores experimentais da atividade biológica de uma série de compostos. Isto pode ser justificado pelo fato de que o modelo extratermodinâmico (QSAR clássico) é uma

estimativa de modelo que tenta relacionar as propriedades moleculares dos compostos com a variação da energia livre de Gibbs envolvida no processo de interação fármaco-biorreceptor (Kubinyi, 1993). A equação clássica de Hansch de um dado modelo multidimensional linear pode ser descrita com o termo dependente ($\text{Log } 1/\text{IC}_{50}$ ou $-\text{Log } \text{IC}_{50}$) em função das variáveis independentes:

$$-\text{Log } \text{IC}_{50} = a (X_{\text{lipofílico}}) + b (X_{\text{eletrônico}}) + c (X_{\text{estéreo}}) + d (X_{\text{polar}}) + e \quad (59)$$

sendo que, a representação X na equação refere-se às variáveis independentes (propriedades moleculares) enquanto os símbolos a - e correspondem aos coeficientes de ajuste obtidos pela RLM e o termo $-\text{Log } \text{IC}_{50}$ representa os valores previstos da resposta biológica (Gaudio & Zandonade, 2001).

Neste trabalho, a atividade biológica está representada pela medida do IC_{50} (μM), ou seja, é a concentração molar de cada derivado de Chalcona capaz de inibir *in vitro* 50% da parasitemia provocada pelas cepas resistentes (W2) de *Plasmodium falciparum* (Li *et al*, 1995).

Não devemos esquecer que um dos objetivos da análise multivariada em QSAR é desenvolver modelos com o intuito em prever a atividade biológica de compostos levando em consideração propriedades moleculares selecionadas como relevantes, no sentido em explicar um provável mecanismo de ação farmacológico destes compostos.

4.3.1 Regressão linear simples

Para que possamos construir um modelo matemático multidimensional linear da atividade biológica em função de variáveis explicativas ou variáveis independentes, é necessário selecionar quais variáveis estão correlacionadas à ação biológica dos compostos que estão sendo pesquisados. Deste modo, a escolha de variáveis independentes úteis na construção do modelo é uma tarefa que deve ser realizada minuciosamente.

A primeira etapa da análise estatística realizada neste trabalho deu-se na execução da regressão linear simples levando em consideração o conjunto de

dados dos valores de cada propriedade molecular calculada (variável explicativa) e os valores observados da atividade biológica (IC_{50}) da série de compostos derivados de Chalcona. Como sugerido na literatura, os valores iniciais observados de atividade biológica (IC_{50}) foram convertidos em $1/\text{Log } IC_{50} = -\text{Log } IC_{50}$ (Kubinyi, 1993). Posteriormente, os valores de cada propriedade molecular calculada juntamente com a atividade biológica de todos os derivados foram tabelados, constituindo então a nossa matriz de dados. Cada conjunto de dados (atividade biológica-propriedade molecular) foi armazenado no programa estatístico Analyse-it para Microsoft Excel (AME) versão 1,68 (Analyse-it Software Ltd, 2000). Com auxílio deste programa, executamos a regressão linear para todos os conjuntos de dados armazenados, obtendo valores de r^2 , r , F e p .

r^2 = coeficiente de determinação ou explicação;

r = coeficiente de correlação;

F = valor de Fischer;

p = nível de confiabilidade do modelo (Fonseca *et al*, 1985; Kubinyi, 2003).

Após realizar a regressão linear simples para todos os conjuntos de dados atividade-propriedade, os valores de r^2 , r , F e p foram anotados e posteriormente tabelados.

Com a realização do procedimento estatístico mencionado acima, algumas propriedades moleculares foram excluídas por apresentar um coeficiente de explicação muito baixo (pequeno r^2). Tais propriedades, não participarão da equação matemática de previsão da atividade biológica, ou seja, são consideradas propriedades que não explicam a atividade biológica observada e deste modo ao serem acrescentadas no modelo não deverão prever bem a atividade de novos compostos.

O procedimento inicial de exclusão de variáveis via regressão linear simples analisando os valores de r^2 , r , F e p visa descartar propriedades que não contribuem no processo de construção de modelos. As variáveis independentes candidatas a participarem do modelo foram selecionadas levando em consideração o seguinte critério: maior r^2 , maior F e menor p obtidos por intermédio da regressão simples.

Com a realização desta etapa, das 69 propriedades calculadas, conseguimos selecionar 22 variáveis candidatas, ou seja, a concretização desta regressão excluiu 47 variáveis que não iriam contribuir de modo relevante para a nossa análise e na construção do modelo.

A segunda etapa da análise estatística realizada nesta pesquisa consistiu em construir matrizes de correlação para cada categoria de parâmetros e no cálculo dos coeficientes de correlação entre as variáveis para finalmente executar a regressão linear múltipla.

4.3.2 Matriz de correlação e seleção de variáveis em QSAR

A regressão linear múltipla, que também é conhecida como Regressão por Quadrados Mínimos inversos (RLM), consiste na projeção da atividade biológica em função das propriedades moleculares. Este método de regressão multivariada necessita que o número de variáveis independentes seja menor do que o número de amostras, portanto a seleção de variáveis é um procedimento necessário para a concretização da regressão, sendo que as variáveis deverão ser cuidadosamente escolhidas, a fim de obtermos o melhor modelo multivariado.

Outro detalhe importante reside no fato de que as variáveis independentes que participarão da equação não deverão apresentar entre si elevada correlação, ou seja, um elevado coeficiente de correlação (Fonseca *et al*, 1985). A colinearidade entre variáveis faz com que o problema seja matematicamente instável, pois variáveis muito correlacionadas resultarão em determinante próximo de zero, não sendo então possível determinar a inversa da matriz de dados.

Visando selecionar variáveis explicativas para a construção de um modelo multidimensional linear de Hansch – Fujita, realizamos:

- Construção de uma matriz de correlação para cada categoria de variáveis;
- Seleção de variáveis por intermédio de busca sistemática;
- Seleção de variáveis por intermédio de algoritmo genético.

A construção da matriz de correlação e os cálculos dos coeficientes de correlação entre variáveis foram realizados com auxílio do pacote computacional Build QSAR versão Beta - 2001 (Oliveira & Gaudio, 2001).

Considerando que algumas classes de variáveis são extremamente correlacionadas, torna-se necessário utilizarmos uma metodologia seletiva de variáveis para que possamos então excluí-las, visto que variáveis altamente correlacionadas implica em redundância, pois apresentam essencialmente o mesmo tipo de informação.

Não podemos esquecer que a metodologia de construção de uma matriz de correlação pode ser utilizada com intuito em facilitar a exclusão de algumas variáveis que são muito correlacionadas entre si. Para verificar se duas variáveis X_i e X_j são altamente correlacionadas, basta calcular o coeficiente de correlação (R_{ij}) entre elas de acordo com a equação:

$$R_{ij} = \frac{S_{ij}}{\sqrt{S_{ii} \cdot S_{jj}}} \quad (60)$$

Em que S_{ij} , S_{ii} e S_{jj} são definidos de acordo com as equações:

$$S_{ij} = \sum X_i \cdot X_j - \sum X_i \cdot \sum X_j / n \quad (61)$$

$$S_{ii} = \sum X_i^2 - (\sum X_i)^2 / n \quad (62)$$

$$S_{jj} = \sum X_j^2 - (\sum X_j)^2 / n \quad (63)$$

A matriz de correlação (M) é simétrica em relação à diagonal de elementos r_{ij} , deste modo, se $i = j$, percebemos que o valor dos termos da diagonal principal será igual a 1. Em QSAR, costuma-se evitar as combinações em que R_{ij} seja maior ou igual a 0,6 (Kubinyi, 1993; Ferreira *et al*, 2002).

$$M = \begin{pmatrix} r_{11} & r_{12} & r_{13} & \dots & r_{1m} \\ r_{21} & r_{22} & r_{23} & \dots & r_{2m} \\ r_{31} & r_{32} & r_{33} & \dots & r_{3m} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ r_{m1} & r_{m2} & r_{m3} & \dots & r_{mm} \end{pmatrix} \quad (64)$$

A abordagem seletiva de variáveis foi realizada com auxílio dos programas:

- a) Analyse-it para Microsoft Excel (AME) – versão 1,68 (Analyse-it Software Ltd, 2000);
- b) BuildQSAR versão beta – 2001(Oliveira & Gáudio, 2001).

Como visto anteriormente, a primeira etapa seletiva de variáveis foi realizada analisando os valores dos coeficientes de explicação (r^2) pós-execução da regressão linear simples no programa AME e também pela análise dos coeficientes de correlação (R_{ij}) entre variáveis. As matrizes de correlação (tabelas 26-32) para todas as categorias de parâmetros físico-químicos calculados nesta pesquisa QSAR foram construídas: Hidrofóbicos, eletrônicos empíricos, eletrônicos quânticos, estereoquímicos, termodinâmicos, biodisponibilidade e outros parâmetros.

Com a execução da regressão linear simples e observação dos valores dos coeficientes de correlação entre as variáveis ($R_{ij} < 0,6$) (tabelas 26-32), selecionamos 22 variáveis (primeira etapa seletiva). Ao executar os cálculos dos coeficientes de correlação para estas 22 variáveis (tabela 33), selecionamos apenas 14 variáveis (segunda etapa seletiva), sendo agora utilizado como critério um $R_{ij} < 0,40$. Esta nova matriz de dados (contendo as 14 variáveis) foi submetida a um novo cálculo de coeficiente de correlação (tabela 34), com intuito em verificar se existe elevada correlação entre as variáveis selecionadas.

Finalmente, a matriz de dados contendo a atividade biológica observada e os valores das 14 variáveis candidatas foi submetida aos métodos de seleção de variáveis: algoritmo genético e busca sistemática.

O algoritmo genético é um método seletivo de variáveis encontrado no pacote BuildQSAR onde as variáveis são combinadas entre si, montando-se modelos com um número pré-fixado de variáveis escolhidas. O procedimento utilizado na busca seletiva de variáveis descritoras pelo método algoritmo genético foi realizado com a finalidade em obter modelos com três, quatro e até cinco variáveis. Esta sistemática metodológica utilizada na busca de modelos com 3, 4 e 5 descritores foi desenvolvida executando em BuildQSAR sempre o mesmo conjunto de dados (14 variáveis) e os mesmos parâmetros de busca.

A abordagem metodológica utilizada neste trabalho pelo método do algoritmo genético foi desenvolvida utilizando os seguintes parâmetros:

- 35% de mutação;
- 30 modelos por geração;
- 200 gerações;
- O ajuste foi realizado em função do coeficiente de correlação entre variáveis;
- O indicador de atividade biológica foi (- Log IC₅₀).

Posteriormente, realizamos uma busca sistemática de variáveis utilizando também as 14 variáveis escolhidas. Nesta abordagem, utilizamos como critério para escolha de variáveis um limite mínimo de coeficiente de correlação 0,9 e um limite mínimo de Q² 0,85 (quadrado do coeficiente de correlação das previsões).

Após executar algoritmo genético e busca sistemática de variáveis, o programa BuildQSAR selecionou automaticamente 10 melhores modelos em cada método apresentando três, quatro e cinco variáveis. Destes 10 modelos selecionados, escolhemos os três melhores (com 3, 4 e 5 descritores) levando em consideração *r*, *s* e *F* objetivando assim a realização da regressão linear múltipla.

Analisando um conjunto de variáveis independentes $X_1, X_2, \dots, X_j, \dots, X_k$, (propriedades moleculares) e a variável dependente *Y* (atividade biológica), podemos afirmar que realizar a regressão linear múltipla de *Y* em função das variáveis $X_1, X_2, \dots, X_j, \dots, X_k$, significa construir um modelo do tipo:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (65)$$

sendo β_0 a constante de ajuste do modelo;

$\beta_1, \beta_2 \dots \beta_k$ são os coeficientes das variáveis independentes;

ε é o erro de ajuste linear da equação.

A determinação exata dos valores de β só poderá ser feita se todos os valores possíveis de *y* forem incluídos no modelo. Em QSAR, significaria incluir no modelo todos os compostos com alguma atividade sobre o sistema biológico em estudo. Na prática este procedimento é muito complicado, visto que não é possível

saber quantos compostos conhecidos e desconhecidos, apresentam atividade sobre um dado sistema. Outro fator importante reside na determinação do erro ε ser considerado tarefa extremamente difícil, pois os fatores que contribuem para seu valor são irregulares e dependentes de um grande número de parâmetros. Na prática, os valores de β permanecerão desconhecidos e o que podemos então obter é a *estimativa de modelo* via regressão linear múltipla. Em QSAR, a estimativa de modelo é uma equação capaz de fornecer *valores previstos* para Y que geralmente serão representados por y_{pred} :

$$y_{\text{pred}} = b_0 + b_1X_1 + b_2X_2 + \dots b_kX_k + \varepsilon \quad (66)$$

onde $b_0, b_1, b_2, \dots b_k$ são estimativas para $\beta_0, \beta_1, \beta_2, \dots \beta_k$. A construção da estimativa é realizada pela concretização da RLM, ou seja, através da regressão por quadrados mínimos inverso.

Aplicar modelo de dependência linear sobre um conjunto de dados significa acreditar que tal dependência existe, sendo a variável Y dependente da combinação linear das variáveis independentes (X_i).

Podemos representar a equação (65) na seguinte forma matricial:

$$y = \beta \cdot X + \varepsilon \quad (67)$$

A análise regressão linear múltipla (RLM) envolve os seguintes vetores e matriz:

$$y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & X_{11} & X_{12} & \dots & X_{1k} \\ 1 & X_{21} & X_{22} & \dots & X_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \dots & X_{nk} \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix} \quad 1 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \quad (68)$$

onde n é o número de amostras (derivados de Chalcona) e k é o número de variáveis independentes (propriedades moleculares).

Matematicamente, executar RLM significa determinar o vetor de parâmetros β e o vetor de erros ε . Como vimos acima, o vetor b é considerado valor estimado de β . O vetor procurado b é obtido a partir da equação:

$$b = (X^tX)^{-1}X^ty \quad (69)$$

onde $(X^tX)^{-1}$ é o inverso da matriz resultante do produto da matriz transposta X^t por X .

Após realizar RLM para os três melhores modelos com 3, 4 e 5 descritores, obtivemos nove modelos matemáticos sendo que um deles foi proposto como estimativa de modelo multidimensional linear de Hansch-Fujita.

4.3.3 Análise do modelo escolhido

Para que uma equação resultante de uma regressão linear múltipla seja promovida a um modelo matemático é necessário verificar a qualidade deste modelo. A qualidade de um modelo matemático reside na verificação de algumas regras propostas por Unger & Hansch (1973). A proposição de um modelo matemático leva em consideração a análise de parâmetros que estão relacionados com o seu grau de ajuste, grau de significância e grau de previsibilidade (Gáudio & Zandonadi, 2001). Segundo Unger & Hansch, regras gerais devem ser obedecidas para obtenção de modelos matemáticos de relações estrutura-atividade:

- **Seleção de variáveis independentes** – grande número de variáveis devem ser testadas para que possam ser selecionadas a participarem da equação matemática;
- **Validação estatística das variáveis independentes** – as variáveis selecionadas por métodos estatísticos a participarem da equação devem ser submetidas a métodos estatísticos apropriados de validação;
- **Princípio da parcimônia** – quando houver dúvida na escolha de um modelo matemático entre muitos outros, devemos escolher o mais simples;

- **Número de variáveis descritoras presentes no modelo** – a média é de cinco ou seis compostos para cada variável explicativa incluída no modelo matemático para que se possa evitar a ocorrência de correlação por coincidência (critério de Topliss) (Topliss & Costello, 1972);
- **Modelo qualitativo para o mecanismo de ação dos compostos** – o modelo quantitativo deve ser consistente com o mecanismo de ação em nível molecular.

A proposição de modelos quantitativos não é uma tarefa trivial, pois uma equação obtida via RLM deverá sobreviver às regras de proposição devendo ser validada estatisticamente e havendo a possibilidade, deverá ser validada em termos de sua capacidade de explicação quanto ao mecanismo de ação.

A validação estatística de uma equação linear pode ser realizada através da avaliação do modelo escolhido. A avaliação de um modelo consiste em verificar se a equação se adapta convenientemente aos dados observados e pode ser subdividida em três etapas:

- a) Avaliação do grau de ajuste;
- b) Avaliação do grau de significância estatística;
- c) Avaliação do grau de previsibilidade da equação.

A avaliação do grau de ajuste está associada à capacidade do modelo em reproduzir o valor das atividades observadas. Esta avaliação pode ser realizada via Análise de Variância (ANOVA) da regressão verificando o coeficiente de correlação (r), o coeficiente de correlação ajustado (r_{ajust}), o desvio padrão (s) e análise de resíduos ($y - y_{\text{pred}}$). Modelos com excelente grau de ajuste, deverão apresentar elevado coeficiente de correlação (próximo de 1), desvio padrão baixo (próximo de zero) e resíduos da regressão em função da atividade prevista distribuídos em torno de zero. Compostos com elevados resíduos num modelo de regressão serão considerados derivados anômalos (*outliers*) e deverão ser excluídos da análise. As amostras *outliers* são as que aparecem distribuídas nos diagramas dos resíduos em função da atividade prevista, fora da região do limite de confiança (95%), ou seja, aparecem em uma região do diagrama em que o valor previsto da atividade difere em 5% do valor observado da atividade.

O programa BuildQSAR plota estes diagramas diferenciando as amostras (derivados de Chalcona) com cores diferentes (azul, amarelo e vermelho), visando a identificação de um possível derivado *outlier*. As amostras que melhor se ajustam ao modelo de regressão são codificadas de azul, enquanto aquelas que se apresentam codificadas de vermelho serão consideradas *outliers*. As amostras codificadas de amarelo estão na região do limite de confiança (95%).

A avaliação do grau de significância estatística é realizada através da razão das variâncias conhecido como teste de Fischer ou teste-F. O teste-F é um teste de hipótese que tem por finalidade verificar a significância estatística de r^2 . O valor de F pode ser obtido também a partir da tabela de Análise de Variância da regressão, sendo que o modelo deverá apresentar um elevado valor de F. Outro parâmetro que deve ser observado na tabela ANOVA é o valor p, que indica o nível de confiabilidade da regressão. Um valor de p menor que 0,0001, significa que o valor de r^2 da regressão é estatisticamente significativa, onde o erro envolvido é menor que 0,01% (Arango, 2001).

A avaliação do grau de previsibilidade de um modelo é realizada através de um procedimento estatístico denominado *cross validation* (validação cruzada). A validação cruzada mede a capacidade do modelo em fazer previsões em relação a amostras não incluídas na análise. Uma das formas de realizar o processo de validação cruzada consiste em excluir um dos derivados fazendo a regressão com um derivado a menos (n-1), utilizando a regressão obtida para prever a atividade do derivado ($y_{i \text{ pred}}$) deixado fora do modelo (*leave one out*). O processo é repetido para todos os derivados (amostras), obtendo o parâmetro **PRESS** (*Predictive Residual Error Sum of Squares*).

$$\text{PRESS} = \sum (y_i - y_{i \text{ pred}})^2 \quad (70)$$

onde $(y_i - y_{i \text{ pred}})$ é o erro de previsão.

O **PRESS** é a soma dos resíduos entre o valor atividade observada e o valor da atividade prevista do derivado deixado fora do processo de validação cruzada, ou seja, é também a soma dos quadrados dos erros de previsão. Deste

modo, o valor **PRESS** deverá ser o menor possível para que o modelo apresente uma boa capacidade de previsão. A partir do **PRESS**, podemos obter outros parâmetros de verificação da capacidade preditiva (Q^2 e **SPRESS**). O Q^2 que também é chamado de quadrado do coeficiente de correlação das previsões ou da validação cruzada é obtido pela equação:

$$Q^2 = 1 - \frac{\text{PRESS}}{\sum (y_i - y_{i \text{ mean}})^2} \quad (71)$$

O desvio padrão da validação cruzada é representado pelo **SPRESS** e pode ser determinado pela equação:

$$\text{SPRESS} = \frac{\sqrt{\text{PRESS}}}{n - k - 1} \quad (72)$$

Um modelo que apresenta elevado grau de previsibilidade deverá possuir Q^2 próximo de um e **SPRESS** próximo de zero.

5.0 Resultados

A primeira etapa desta pesquisa QSAR fundamentou-se na busca de uma metodologia consistente que resultasse estruturas químicas com a menor energia total possível para todos os derivados de Chalcona. Esta etapa foi caracterizada pela minimização energética para cada estrutura inicial (desenho 3D em MMP) dos compostos envolvidos com posterior otimização da geometria molecular.

5.1 Minimização energética e otimização da geometria

A energia total de cada estrutura química dos derivados de Chalcona resultante do processo de minimização energética e da otimização da geometria molecular levando em consideração a metodologia comentada anteriormente (4.1) foi anotada e posteriormente tabelada (tabela 2). Não podemos esquecer que E_4 representa o valor da energia em Kcal/mol de cada derivado após ter realizado a minimização simplex utilizando o opcional constante dielétrica correspondente para meio aquoso com potencial 6-12 Lennard Jones e função Hydrogen bond. A constante dielétrica é uma grandeza que varia com a temperatura, sendo que o programa MMP utiliza o valor 78,54 como constante dielétrica em meio aquoso a 37°C.

Tabela 2 – Energia em Kcal/mol dos derivados de Chalcona durante o processo de minimização energética e otimização da geometria molecular; E₁ – energia pós-desenho 3D (MMP), E₂ – energia pós-minimização via método MM2 (MMP), E₃ – energia pós-otimização geometria molecular via método PM3 (Hyperchem), E₄ – energia pós-minimização simplex (MMP).

Chalconas	E ₁	E ₂	E ₃	E ₄
Chalcona 7	5193,216	72,32564	38,07842	22,06407
Chalcona 8	11327,12	313,5485	44,5623	21,2326
Chalcona 9	511,2683	79,29557	43,00481	25,62006
Chalcona 10	413,7483	78,96114	37,29515	22,34175
Chalcona 11	458,7544	58,74694	40,09446	19,41005
Chalcona 12	2856,746	60,98282	34,07318	19,50561
Chalcona 13	2742,005	56,08046	33,12215	18,83625
Chalcona 14	451,5143	69,20435	38,0844	22,13786
Chalcona 15	6781,766	77,73465	45,05817	26,10589
Chalcona 16	2238,347	68,33436	36,91638	22,47113
Chalcona 17	7120,973	56,14112	34,85855	20,21091
Chalcona 18	2856,922	61,03864	34,2207	18,85694
Chalcona 19	8992,541	288,2372	34,59664	20,37552
Chalcona 20	469,3787	115,92	54,59203	41,18779
Chalcona 21	2856,704	60,75869	34,71777	19,21926
Chalcona 22	3606,643	55,53101	34,11912	19,71325
Chalcona 23	2742,176	55,33726	32,41883	19,58587
Chalcona 24	5095,438	60,08042	32,90902	18,1939
Chalcona 25	2263,683	58,44695	32,15776	18,23991

5.2 Cálculo de parâmetros físico-químicos

Após definirmos a geometria molecular para cada derivado de Chalcona, propriedades moleculares estruturais e físico-químicas foram calculadas com o intuito de relacionar a atividade biológica com as propriedades calculadas para que posteriormente possamos selecionar por intermédio de métodos estatísticos quais destas propriedades serão capazes de justificar a atividade biológica observada. Os cálculos de parâmetros foram executados para todas as estruturas químicas apresentando a menor energia (E₄), ou seja, conformação obtida após ter realizado a minimização simplex em MMP. As tabelas 3 – 15 representam os valores das propriedades moleculares calculadas.

Tabela 3 - Parâmetros Hidrofóbicos: Log P_{Hansch}, Log P_{Ghose e Crippen}, Balanço Hidrofílico-Lipofílico (HLB), Parâmetro de solubilidade (P.S), Dispersão de Hansen 3D (D. Hansen 3D), Polaridade de Hansen 3D (δ Hansen 3D).

Chalconas ($\epsilon = 78,54$)	Log P Hansch	Log P Ghose e Crippen	HLB	P.S	D. Hansen 3D	δ Hansen 3D
Chalcona 07	4,057000	5,0175	2,602514	24,97928	23,55290	6,384537
Chalcona 08	3,342001	3,9539	0,344027	23,50839	21,83283	5,282660
Chalcona 09	2,792001	4,0243	1,366989	24,08010	22,38016	5,578986
Chalcona 10	1,867000	4,0232	5,532198	25,10479	23,06764	7,333460
Chalcona 11	3,660001	4,0534	0,473424	20,99382	19,55086	5,084429
Chalcona 12	4,203001	4,2066	0,522446	23,38213	21,87482	5,429868
Chalcona 13	3,295001	3,6886	0,622303	22,37047	20,92691	4,969752
Chalcona 14	4,057000	5,0175	2,601827	24,97928	23,55290	6,384537
Chalcona 15	2,987000	4,0243	1,386589	24,08010	22,38016	5,578986
Chalcona 16	3,630000	4,8087	1,470242	25,18922	23,67493	6,820307
Chalcona 17	4,203001	4,2066	0,519241	23,38213	21,87482	5,429868
Chalcona 18	4,203001	4,2066	0,512150	23,38213	21,87482	5,429868
Chalcona 19	4,008001	4,2066	0,515115	23,38213	21,87482	5,429868
Chalcona 20	4,284000	4,3873	4,476958	25,23376	23,16554	7,090923
Chalcona 21	4,203001	4,2066	0,506328	23,38213	21,87482	5,429868
Chalcona 22	3,063000	3,4496	0,644264	22,47912	20,95706	5,379371
Chalcona 23	3,063000	3,4496	0,646536	22,47912	20,95706	5,379371
Chalcona 24	4,203001	4,2066	0,136053	23,38213	21,87482	5,429868
Chalcona 25	3,063000	3,4496	0,242115	22,47912	20,95706	5,379371

Tabela 4 - Parâmetros Hidrofóbicos: Ligação de Hidrogênio de Hansen 3D (H. B. Hansen 3D), Área superficial hidrofílica (ASH), % Área superficial hidrofílica (% ASH), Log solubilidade em água (Log S. H₂O), Número de hidratação (NH), Refratividade molar (MR_{Ghose e Crippen})

Chalconas ($\epsilon = 78,54$)	H. B. Hansen 3D	ASH	% ASH	Log S. H ₂ O	NH	MR _{Ghose e Crippen}
Chalcona 07	5,335073	3,415308	17,663110	-4,7331	4,5	90,2262
Chalcona 08	6,932905	1,826872	7,843600	-3,2518	4,5	95,8762
Chalcona 09	6,917716	2,756186	12,291260	-4,1715	6,5	99,5157
Chalcona 10	6,659943	5,465337	30,400860	-3,9631	7,5	84,2943
Chalcona 11	5,714285	1,810238	8,406195	-3,2765	3,5	85,7771
Chalcona 12	6,223536	1,826144	8,619334	-3,1559	3,5	89,4130
Chalcona 13	6,148478	1,818892	9,053493	-2,4025	3,5	84,6082
Chalcona 14	5,335073	3,415715	17,660120	-4,7331	4,5	90,2262
Chalcona 15	6,917716	2,785283	12,376480	-4,1715	6,5	99,5157
Chalcona 16	5,241939	2,428754	12,740180	-4,5827	4,5	87,0221
Chalcona 17	6,223536	1,824269	8,605396	-3,1559	3,5	89,4130
Chalcona 18	6,223536	1,814830	8,574568	-3,1559	3,5	89,4130
Chalcona 19	6,223536	1,816444	8,587459	-3,1559	3,5	89,4130
Chalcona 20	7,058247	5,301311	25,812860	-4,2183	7,5	93,8589
Chalcona 21	6,223536	1,809766	8,549256	-3,1559	3,5	89,4130
Chalcona 22	6,097107	1,821550	9,148975	-2,2521	3,5	80,2362
Chalcona 23	6,097107	1,812555	9,158853	-2,2521	3,5	80,2362
Chalcona 24	6,223536	1,469911	6,939362	-3,1559	3,5	89,4130
Chalcona 25	6,097107	1,469250	7,400501	-2,2521	3,5	80,2362

Tabela 5 - Parâmetros eletrônicos empíricos: constante de Hammett (σ) para os substituintes A e B.

σ_p – Hammett sigma para, σ_m – Hammett sigma meta, σ^* - Hammett sigma star, σ_i – Hammett sigma induction.

Chalconas ($\epsilon = 78,54$)	σ_p (A)	σ_p (B)	σ_m (A)	σ_m (B)	σ^* (A)	σ^* (B)	σ_i (A)	σ_i (B)
Chalcona 07	0,505	0,217	0,6790000	0,2460000	0,817	1,467000	0,226	0,286
Chalcona 08	0,122	0,140	0,1200000	0,0440000	1,211	1,229000	0,246	0,246
Chalcona 09	0,701	0,130	1,2460000	0,0610000	2,789	1,004000	0,390	0,261
Chalcona 10	0,701	0,409	1,2460000	0,5700000	2,789	0,790000	0,390	0,181
Chalcona 11	0,025	0,125	-0,0860000	0,0260000	0,694	1,145000	0,201	0,229
Chalcona 12	0,122	0,130	0,1200000	0,0229999	1,211	1,175000	0,246	0,226
Chalcona 13	0,025	0,125	-0,0860000	-0,0120000	0,694	1,316000	0,201	0,194
Chalcona 14	0,505	0,130	0,6790000	0,0610000	0,817	1,004000	0,226	0,261
Chalcona 15	0,701	0,125	1,2460000	0,0260000	2,789	1,145000	0,390	0,229
Chalcona 16	0,701	0,196	1,2460000	0,1650000	2,789	1,377000	0,390	0,308
Chalcona 17	0,122	0,125	0,1200000	0,0260000	1,211	1,145000	0,246	0,229
Chalcona 18	0,127	0,130	0,1170000	0,0229999	1,241	1,175000	0,243	0,226
Chalcona 19	0,127	0,130	0,1170000	0,0610000	1,241	1,004000	0,243	0,261
Chalcona 20	0,122	0,115	0,1200000	0,0430000	1,211	0,920000	0,246	0,244
Chalcona 21	0,127	0,125	0,1170000	0,0260000	1,241	1,145000	0,243	0,229
Chalcona 22	0,106	0,125	0,0360000	0,0260000	1,151	1,145000	0,265	0,229
Chalcona 23	0,040	0,125	-0,0680000	0,0260000	0,778	1,145000	0,218	0,299
Chalcona 24	0,035	0,212	-0,1030000	0,2490000	0,919	1,437000	0,186	0,289
Chalcona 25	0,035	0,196	-0,1030000	0,1650000	0,919	1,377000	0,186	0,308

Tabela 6 - Parâmetros eletrônicos quânticos: Momento de dipolo (μ), Hydrogen bond acceptor (H. B. acceptor), Hydrogen bond donor (H. B. donor), E_{HOMO} , E_{LUMO} .

Chalconas ($\epsilon = 78,54$)	μ	H. B. acceptor	H. B. donor	E_{HOMO}	E_{LUMO}
Chalcona 07	2,776738	0,1509322	0,0064960	-0,4054409	0,0406347
Chalcona 08	6,021101	0,2871797	0,0000000	-0,4310551	0,0627521
Chalcona 09	3,180744	0,2474873	0,0000000	-0,4150968	0,0295800
Chalcona 10	2,781697	0,1218783	0,0000000	-0,4121257	0,0271000
Chalcona 11	3,743521	0,4684491	0,0000000	-0,4284725	0,0786462
Chalcona 12	6,552880	0,2506073	0,0073037	-0,4205667	0,0656528
Chalcona 13	4,788819	0,2629606	0,0000000	-0,4346325	0,0672191
Chalcona 14	1,723343	0,1463352	0,0011166	-0,4034062	0,0362228
Chalcona 15	3,184682	0,2699979	0,0000518	-0,4120781	0,0267397
Chalcona 16	2,596465	0,2072395	0,0197635	-0,4189104	0,0314058
Chalcona 17	2,591375	0,2692726	0,0016736	-0,4318215	0,0581758
Chalcona 18	8,430083	0,2516165	0,0053430	-0,4253062	0,0641182
Chalcona 19	7,053628	0,2299472	0,0000000	-0,4351138	0,0623057
Chalcona 20	2,678296	0,1996031	0,0004691	-0,4045944	0,0599702
Chalcona 21	9,001968	0,2557490	0,0029316	-0,4281165	0,0654286
Chalcona 22	3,892121	0,3366489	0,0129435	-0,4274311	0,0805034
Chalcona 23	6,017005	0,3313825	0,0036802	-0,4342023	0,0923625
Chalcona 24	5,204623	0,2447222	0,0115017	-0,4242770	0,0526500
Chalcona 25	3,702250	0,3449589	0,0215632	-0,4278709	0,0765902

Tabela 7 - Parâmetros eletrônicos quânticos: Cargas atômicas parciais (q) sobre os átomos C₁, C₂ e C₃.

Chalconas ($\epsilon = 78,54$)	q (C ₁)	q (C ₂)	q (C ₃)
Chalcona 07	0,0108350	-0,0508978	0,0638794
Chalcona 08	0,0110592	-0,0435886	0,0596666
Chalcona 09	0,0137039	-0,0528557	0,0177402
Chalcona 10	0,0175568	-0,0566820	0,0208672
Chalcona 11	-0,0017994	-0,0514257	0,0560831
Chalcona 12	0,0036947	-0,0385947	0,0638461
Chalcona 13	0,0129890	-0,0482149	0,0448094
Chalcona 14	0,0127480	-0,0561976	0,0634271
Chalcona 15	0,0081044	-0,0475198	0,0214626
Chalcona 16	0,0149052	-0,0502809	0,0237476
Chalcona 17	0,0021110	-0,0361712	0,0632994
Chalcona 18	0,0059708	-0,0421181	0,0575785
Chalcona 19	0,0078089	-0,0449892	0,0508683
Chalcona 20	0,0088708	-0,0479600	0,0604783
Chalcona 21	-0,0004273	-0,0405700	0,0545179
Chalcona 22	0,0188501	-0,0494848	-0,0299627
Chalcona 23	-0,0013781	-0,0449820	0,0435425
Chalcona 24	0,0409438	-0,0683702	-0,0376122
Chalcona 25	0,0191780	-0,0592652	-0,0362397

Tabela 8 - Parâmetros eletrônicos quânticos: Cargas atômicas parciais (q) sobre os átomos C₄, C₈ e O₇.

Chalconas ($\epsilon = 78,54$)	q (C ₄)	q (C ₈)	q (O ₇)
Chalcona 07	0,2603298	-0,0006194	-0,2679864
Chalcona 08	0,2542730	-0,0899717	-0,2701172
Chalcona 09	0,2555919	-0,0281393	-0,2740448
Chalcona 10	0,2516232	-0,0539130	-0,2694947
Chalcona 11	0,2627521	-0,0957345	-0,2749034
Chalcona 12	0,2663692	-0,0390846	-0,2734942
Chalcona 13	0,2704643	-0,1150234	-0,2722905
Chalcona 14	0,2516701	-0,0147402	-0,2644040
Chalcona 15	0,2688261	-0,0845744	-0,2879161
Chalcona 16	0,2695003	-0,0808306	-0,2804427
Chalcona 17	0,2622516	-0,0784532	-0,2856748
Chalcona 18	0,2661625	-0,0420179	-0,2705345
Chalcona 19	0,2431669	-0,0347759	-0,2647831
Chalcona 20	0,2477945	-0,0662009	-0,2693616
Chalcona 21	0,2720728	-0,0842610	-0,2811883
Chalcona 22	0,2586650	-0,0951060	-0,2763566
Chalcona 23	0,2674494	-0,0807521	-0,2881262
Chalcona 24	0,2608554	0,0001748	-0,2660569
Chalcona 25	0,2763236	-0,0777351	-0,2840539

Tabela 9 - Parâmetros estereoquímicos: Volume molecular (VM), Densidade molecular (ρ), Comprimento molecular (L_x), Largura molecular (L_y), Profundidade molecular (L_z), Superfície de van der Waals (SvdW), Superfície de van der Waals acessível ao solvente (SvdW_{sol}).

Chalconas ($\epsilon = 78,54$)	VM	ρ	L_x	L_y	L_z	SvdW	SvdW _{sol}
Chalcona 07	165,2515	1,180419	11,553200	13,7285	8,02210	208,5	105,1
Chalcona 08	186,7981	1,260866	10,077000	15,8380	9,40550	225,5	123,4
Chalcona 09	186,6909	1,171354	9,655001	17,3179	5,83600	232,7	130,9
Chalcona 10	153,5890	1,134220	9,788700	15,6241	5,42020	196,4	109,8
Chalcona 11	172,6587	1,197011	11,390000	16,4127	7,69690	219,2	120,9
Chalcona 12	171,5456	1,235219	9,002701	16,0490	6,37520	212,3	121,8
Chalcona 13	162,6785	1,160833	10,574400	14,1902	6,21820	200,2	110,6
Chalcona 14	165,3037	1,180051	11,619500	14,6143	6,70720	207,9	118,2
Chalcona 15	187,2802	1,167570	9,718800	17,2375	6,88080	222,7	123,9
Chalcona 16	162,5305	1,206203	9,833000	16,3316	6,07500	205,3	116,4
Chalcona 17	171,6219	1,234646	9,269899	16,1618	6,85700	210,8	118,6
Chalcona 18	171,3853	1,236425	8,811399	15,7934	7,61120	211,4	121,4
Chalcona 19	171,2892	1,237149	9,443800	16,6691	5,87880	213,8	120,2
Chalcona 20	172,6486	1,176705	9,302600	18,0037	5,36920	220,0	126,4
Chalcona 21	171,4109	1,236232	9,122200	16,2655	6,69760	210,5	120,4
Chalcona 22	160,5819	1,183203	9,078300	15,3741	6,61370	197,4	106,2
Chalcona 23	159,7285	1,189761	9,098499	15,6064	6,40130	197,4	109,2
Chalcona 24	171,5207	1,235406	9,087100	16,9126	7,59880	218,4	123,6
Chalcona 25	160,1784	1,186296	9,141100	16,2847	6,31370	203,0	111,0

Tabela 10 - Parâmetros estereoquímicos de Verloop (ε).

- ε L1 (A) e (B) – Verloop L1 para os substituintes A e B,
- ε B1 (A) e (B) – Verloop B1 para os substituintes A e B,
- ε B2 (A) – Verloop B2 para o substituinte A.

Chalconas (ε = 78,54)	ε L1 (A)	ε L1 (B)	ε B1 (A)	ε B1 (B)	ε B2 (A)
Chalcona 07	6,541162	6,470678	1,770000	1,779290	3,197927
Chalcona 08	7,842712	8,661812	1,770000	1,896983	4,482778
Chalcona 09	8,576578	8,618576	1,770000	2,192573	4,487854
Chalcona 10	8,598517	6,447505	1,770000	1,770000	4,490911
Chalcona 11	7,895205	8,646161	2,068402	2,168087	3,125317
Chalcona 12	7,842075	7,342750	1,770000	2,259595	3,228987
Chalcona 13	7,844814	6,497647	1,770000	2,257009	3,192143
Chalcona 14	6,565150	7,864295	1,770000	1,771919	5,717386
Chalcona 15	8,588703	8,711747	1,770000	1,770000	4,492193
Chalcona 16	8,602199	7,180599	1,770000	1,774558	4,515000
Chalcona 17	7,859898	8,691294	1,770000	1,785500	3,199387
Chalcona 18	6,459242	7,307997	1,770000	2,267795	4,507043
Chalcona 19	6,448768	8,661847	1,770000	1,883410	4,526188
Chalcona 20	7,842031	9,698078	1,792507	1,896774	4,506215
Chalcona 21	6,464288	8,701036	1,770000	1,772731	3,249254
Chalcona 22	7,205367	8,698980	1,782479	1,770000	3,212538
Chalcona 23	7,150577	8,663775	1,778216	1,786232	3,292225
Chalcona 24	8,622450	7,840757	1,770000	1,770000	5,281165
Chalcona 25	8,638336	7,166421	1,786517	1,770000	5,349347

Tabela 11 - Parâmetros estereoquímicos de Verloop (ϑ).

ϑ B2 (B) – Verloop B2 para o substituinte B,

ϑ B3 (A) e (B) – Verloop B3 para os substituintes A e B,

ϑ B4 (A) e (B) – Verloop B4 para os substituintes A e B.

Chalconas ($\epsilon = 78,54$)	ϑ B2 (B)	ϑ B3 (A)	ϑ B3 (B)	ϑ B4 (A)	ϑ B4 (B)
Chalcona 07	4,595470	5,700418	4,518767	1,858154	1,868668
Chalcona 08	5,031200	3,252031	3,297416	1,792460	3,065203
Chalcona 09	5,180004	4,410695	3,107664	1,961689	3,116829
Chalcona 10	3,185579	4,428206	3,184035	2,014877	1,860177
Chalcona 11	5,486969	3,124650	3,101983	2,851562	2,902363
Chalcona 12	5,001549	4,497711	5,437396	1,793886	2,316163
Chalcona 13	5,432764	3,206546	5,500188	1,882035	2,292890
Chalcona 14	4,476333	3,192786	3,214087	1,839970	1,809869
Chalcona 15	5,481412	4,441313	3,262521	1,845351	3,084818
Chalcona 16	3,230817	4,344332	3,883639	1,896770	1,795815
Chalcona 17	5,506016	4,502231	3,250907	1,815932	3,027209
Chalcona 18	5,087956	3,302691	5,398762	1,789911	2,314198
Chalcona 19	5,367317	3,232959	3,209377	1,921321	3,028946
Chalcona 20	3,185370	3,190463	3,190449	1,865551	2,591658
Chalcona 21	5,426905	4,500305	3,262843	1,802636	3,026906
Chalcona 22	5,560515	3,871483	3,242798	1,782715	3,029820
Chalcona 23	5,377376	3,776243	3,258753	1,788104	3,064308
Chalcona 24	4,504360	3,211792	3,239859	3,223830	1,831798
Chalcona 25	3,873280	3,220666	3,266833	3,173289	1,847932

Tabela 12 - Parâmetros termodinâmicos: Ponto de ebulição (PE), Parachor (PA), Número efetivo de ligações torsionais (tau), Número de Ligações de Hidrogênio (HBN), Temperatura Crítica (TC) e Pressão Crítica (PC).

Chalconas ($\epsilon = 78,54$)	PE	PA	tau	HBN	TC	PC
Chalcona 07	498,9547	668,9003	1,5	0,0030469	1158,6460	23,98217
Chalcona 08	499,8613	755,6005	4,5	0,0054462	1126,4740	17,81845
Chalcona 09	596,1644	746,3005	3,5	0,0048954	1198,5720	19,82351
Chalcona 10	519,5225	613,2002	1,5	0,0033928	1155,1950	27,90608
Chalcona 11	455,1099	674,7003	4,5	0,0051501	994,5330	17,89390
Chalcona 12	483,4796	695,6003	3,5	0,0051365	1081,2990	19,91206
Chalcona 13	456,1308	656,0002	3,5	0,0057209	1035,9450	20,86933
Chalcona 14	451,0547	668,9003	1,5	0,0030469	1158,6460	23,98217
Chalcona 15	613,0645	746,3005	3,5	0,0048954	1198,5720	19,82351
Chalcona 16	513,8548	646,5002	1,5	0,0030327	1102,9250	22,46131
Chalcona 17	483,4796	695,6003	3,5	0,0051365	1081,2990	19,91206
Chalcona 18	435,5796	695,6003	3,5	0,0051365	1081,2990	19,91206
Chalcona 19	418,6795	695,6003	3,5	0,0051365	1081,2990	19,91206
Chalcona 20	620,3239	695,0004	2,0	0,0029136	1157,3440	22,61109
Chalcona 21	435,5796	695,6003	3,5	0,0051365	1081,2990	19,91206
Chalcona 22	434,1432	636,6002	3,5	0,0056920	982,3971	19,63075
Chalcona 23	432,5431	636,6002	3,5	0,0056920	982,3971	19,63075
Chalcona 24	483,4798	695,6003	3,5	0,0051365	1081,2990	19,91206
Chalcona 25	434,1432	636,6002	3,5	0,0056920	982,3971	19,63075

Tabela 13 - Parâmetros de biodisponibilidade: Peso molecular (PM), Área superficial polar (AS_{polar}); Outros parâmetros: Ângulos diedrais (β) – C₃C₁C₂C₄, C₁C₂C₄O₇ e C₁C₂C₄C₈.

Chalcona ($\epsilon = 78,54$)	PM	AS _{polar}	β C ₃ C ₁ C ₂ C ₄	β C ₁ C ₂ C ₄ O ₇	β C ₁ C ₂ C ₄ C ₈
Chalcona 07	328,1967	33,12	-176,605	110,804	-69,581
Chalcona 08	367,2281	47,92	177,507	110,930	-70,784
Chalcona 09	353,8047	51,58	178,859	-86,568	93,534
Chalcona 10	294,7397	46,01	178,296	-102,448	77,965
Chalcona 11	336,3105	38,69	-176,883	110,166	-73,302
Chalcona 12	337,2018	38,69	180,000	-90,366	88,070
Chalcona 13	302,7570	38,69	177,610	133,678	-48,803
Chalcona 14	328,1967	33,12	175,836	-96,463	80,556
Chalcona 15	353,8047	51,58	178,621	-108,026	71,856
Chalcona 16	329,7328	33,12	178,266	-104,670	74,973
Chalcona 17	337,2018	38,69	-178,902	109,010	-69,789
Chalcona 18	337,2018	38,69	179,553	91,658	-91,284
Chalcona 19	337,2018	38,69	-176,241	-101,710	81,086
Chalcona 20	343,2115	37,56	177,974	-102,882	76,078
Chalcona 21	337,2018	38,69	-179,292	-89,602	91,016
Chalcona 22	304,2932	38,69	179,292	-108,498	59,034
Chalcona 23	304,2932	38,69	-177,717	97,622	-82,079
Chalcona 24	337,2018	38,69	-178,355	67,794	-110,288
Chalcona 25	304,2932	38,69	178,584	-72,377	106,707

Tabela 14 - Outros parâmetros - Distância interatômica (d): C₁C₂, C₂C₄, C₁C₃, C₄O₇, C₄C₈.

Chalconas ($\varepsilon = 78,54$)	d C ₁ C ₂	d C ₂ C ₄	d C ₁ C ₃	d C ₄ O ₇	d C ₄ C ₈
Chalcona 07	1,341	1,464	1,522	1,222	1,509
Chalcona 08	1,328	1,466	1,489	1,220	1,500
Chalcona 09	1,333	1,452	1,493	1,218	1,489
Chalcona 10	1,345	1,465	1,465	1,221	1,500
Chalcona 11	1,355	1,456	1,515	1,221	1,490
Chalcona 12	1,344	1,454	1,494	1,223	1,511
Chalcona 13	1,321	1,450	1,500	1,220	1,500
Chalcona 14	1,337	1,463	1,497	1,221	1,512
Chalcona 15	1,339	1,472	1,501	1,221	1,496
Chalcona 16	1,341	1,462	1,496	1,220	1,507
Chalcona 17	1,343	1,459	1,497	1,229	1,523
Chalcona 18	1,338	1,458	1,488	1,215	1,516
Chalcona 19	1,334	1,454	1,490	1,223	1,517
Chalcona 20	1,340	1,456	1,510	1,222	1,507
Chalcona 21	1,345	1,455	1,511	1,219	1,496
Chalcona 22	1,342	1,451	1,501	1,220	1,512
Chalcona 23	1,345	1,458	1,499	1,231	1,505
Chalcona 24	1,335	1,456	1,491	1,221	1,494
Chalcona 25	1,349	1,451	1,532	1,220	1,498

Tabela 15 - Outros parâmetros - Ângulos entre ligações (α): $C_1C_2C_4$, $C_2C_1C_3$, $C_2C_4C_8$, $C_2C_4O_7$, $O_7C_4C_8$.

Chalconas ($\varepsilon = 78,54$)	α $C_1C_2C_4$	α $C_2C_1C_3$	α $C_2C_4C_8$	α $C_2C_4O_7$	α $O_7C_4C_8$
Chalcona 07	120,393	120,145	120,234	119,723	120,040
Chalcona 08	121,177	118,307	121,000	119,383	119,592
Chalcona 09	119,795	121,329	116,891	121,009	122,101
Chalcona 10	119,122	121,002	118,601	120,372	121,026
Chalcona 11	119,979	119,241	117,253	121,264	121,390
Chalcona 12	121,299	119,145	119,771	119,989	120,222
Chalcona 13	120,911	118,249	118,122	120,661	121,170
Chalcona 14	119,162	119,939	119,402	119,834	120,696
Chalcona 15	120,318	119,589	118,982	120,346	120,671
Chalcona 16	120,127	120,446	119,292	120,164	120,544
Chalcona 17	120,376	119,108	120,318	119,101	120,570
Chalcona 18	120,685	119,445	119,195	120,106	120,631
Chalcona 19	119,594	120,166	117,919	120,168	121,851
Chalcona 20	118,659	121,437	118,330	120,648	121,016
Chalcona 21	121,248	119,858	118,194	120,596	121,208
Chalcona 22	120,330	118,994	119,842	119,306	120,804
Chalcona 23	121,063	120,283	120,269	119,888	119,842
Chalcona 24	118,903	120,057	118,582	120,435	120,956
Chalcona 25	120,150	120,516	118,458	120,652	120,885

5.3 Análise estatística

A proposição de um modelo que correlacione a estrutura química das substâncias pesquisadas com a atividade biológica observada é feita através de Regressão Linear Múltipla (RLM). A análise de regressão linear é uma técnica estatística empregada para prever a atividade biológica de compostos utilizando o método dos quadrados mínimos inverso. A atividade biológica é uma função dependente de variáveis que são as propriedades moleculares calculadas e que na análise estatística são denominadas de variáveis independentes.

5.3.1 Regressão linear simples

O primeiro procedimento efetuado nesta pesquisa QSAR em termos de análise estatística deu-se na realização da regressão linear simples no programa Analyse-it, levando em consideração a atividade biológica observada e as propriedades moleculares calculadas, não esquecendo que a atividade é uma função dependente das propriedades. A regressão linear simples é uma etapa que deve ser cumprida com a finalidade em buscar a primeira fatia de variáveis a serem candidatas a participarem do modelo linear. Neste primeiro processo seletivo de variáveis levamos em consideração os valores de r^2 , F e p marcando então as variáveis que após executar a regressão simples originaram maior r^2 e F e menor p. Estes valores foram anotados e tabelados (tabelas 16 – 25).

Tabela 16 – Regressão Linear: - Log IC₅₀ (W2) = f (parâmetros hidrofóbicos)

Propriedade	r^2	r	F	P
Log P _{Ghose e Crippen}	0,42	0,6480	12,55	0,0025
Log S. H ₂ O	0,42	0,6480	12,54	0,0025
D. Hansen 3D	0,23	0,4795	4,94	0,0401
MR _{Ghose e Crippen}	0,22	0,4690	4,74	0,0438
P.S.	0,19	0,4358	4,03	0,0609
ASH	0,13	0,3605	2,64	0,1229
% ASH	0,12	0,3464	2,30	0,1474
HLB	0,12	0,3464	2,30	0,1474
δ Hansen 3D	0,08	0,2828	1,42	0,2490
NH	0,06	0,2449	1,08	0,3139
H. B. Hansen 3D	0,03	0,173	0,59	0,4527
Log P _{Hansch}	0,02	0,1414	0,31	0,5859

Tabela 17 – Regressão Linear: - Log IC₅₀ (W2) = f (parâmetros eletrônicos empíricos) – Constantes de Hammett (σ)

Propriedade e	r^2	r	F	p
σ_p (A)	0,20	0,447	4,37	0,0518
σ_m (A)	0,17	0,4123	3,51	0,0783
σ_i (B)	0,09	0,30	1,59	0,2249
σ^* (B)	0,07	0,2645	1,30	0,2704
σ_i (A)	0,04	0,20	0,75	0,3988
σ^* (A)	0,01	0,10	0,11	0,7490
σ_m (B)	0,00	0,00	0,04	0,8479
σ_p (B)	0,00	0,00	0,05	0,8273

Tabela 18 – Regressão Linear: - Log IC₅₀ (W2) = f (parâmetros eletrônicos quânticos)

Propriedade	r^2	r	F	p
q (C ₃)	0,50	0,7071	16,72	0,0008
H. B. donor	0,34	0,5831	8,71	0,0089
E _{HOMO}	0,22	0,4690	4,76	0,0434
E _{LUMO}	0,22	0,4690	4,88	0,0412
H. B. acceptor	0,21	0,4582	4,58	0,0471
q (O ₇)	0,19	0,4358	4,06	0,0601
q (C ₄)	0,18	0,4242	3,69	0,0717
q (C ₁)	0,09	0,30	1,75	0,2037
q (C ₈)	0,07	0,2645	1,33	0,2655
q (C ₂)	0,06	0,2449	1,03	0,3243
μ	0,05	0,2236	0,81	0,3801

Tabela 19 – Regressão Linear: - Log IC₅₀ (W2) = f (parâmetros estereoquímicos)

Propriedade	r^2	r	F	p
L _x	0,47	0,6855	15,01	0,0012
L _y	0,17	0,4123	3,60	0,0748
L _z	0,06	0,2449	1,01	0,3288
VM	0,06	0,2449	1,02	0,3260
ρ	0,01	0,10	0,10	0,7579
AS	0,00	0,00	0,02	0,9010

Tabela 20 – Regressão Linear: - Log IC₅₀ (W2) = f (parâmetros esteroquímicos) – (SvdW)

Propriedade	r ²	r	F	p
SvdW	0,05	0,2236	0,88	0,3624
SvdWsol	0,01	0,10	0,24	0,6283

Tabela 21 – Regressão Linear: - Log IC₅₀ (W2) = f (parâmetros estereoquímicos de Verloop) (α)

Propriedade	r ²	r	F	p
α B4 (A)	0,34	0,5830	8,79	0,0087
α L1 (A)	0,14	0,374	2,85	0,1094
α B3 (A)	0,08	0,2828	1,43	0,2479
α B1 (B)	0,03	0,1732	0,54	0,4727
α B3 (B)	0,03	0,1732	0,48	0,50
α B2 (A)	0,01	0,10	0,12	0,7305
α L1 (B)	0,01	0,10	0,13	0,7228
α B2 (B)	0,01	0,10	0,14	0,7138
α B1 (A)	0,00	0,00	0,00	0,9905
α B4 (B)	0,00	0,00	0,00	0,9776

Tabela 22 – Regressão Linear: - Log IC₅₀ (W2) = f (parâmetros termodinâmicos)

Propriedade	r ²	r	F	p
TC	0,35	0,5916	9,35	0,0071
HBN	0,25	0,50	5,75	0,0283
tau	0,12	0,3464	2,31	0,1472
PC	0,12	0,3464	2,27	0,1502
PA	0,08	0,2828	1,40	0,2535
PE	0,03	0,1732	0,61	0,4439

Tabela 23 – Regressão Linear: - Log IC₅₀ (W2) = f (parâmetros de biodisponibilidade)

Propriedade	r ²	r	F	p
PM	0,11	0,3316	2,15	0,1604
AS _{Polar}	0,00	0,00	0,00	0,9495

Tabela 24 – Regressão Linear: - Log IC₅₀ (W2) = f (distância interatômica)

Propriedade	r ²	r	F	p
d (C ₂ -C ₄)	0,20	0,447	4,37	0,0520
d (C ₁ -C ₂)	0,08	0,2828	1,54	0,2320
d (C ₁ -C ₃)	0,07	0,2645	1,21	0,2862
d (C ₄ -C ₈)	0,03	0,1732	0,52	0,4792
d (C ₄ -O ₇)	0,01	0,10	0,17	0,6892

Tabela 25 – Regressão Linear: - Log IC₅₀ (W2) = f (ângulo entre ligações e ângulo diedral)

Propriedade	r ²	r	F	p
α (C ₂ -C ₄ -O ₇)	0,05	0,2236	0,82	0,3766
α (C ₂ -C ₁ -C ₃)	0,03	0,1732	0,40	0,5352
α (C ₂ -C ₄ -C ₈)	0,03	0,1732	0,58	0,4569
α (O ₇ -C ₄ -C ₈)	0,02	1,414	0,30	0,5902
β (C ₁ -C ₂ -C ₄ -O ₇)	0,01	0,10	0,13	0,7245
β (C ₃ -C ₁ -C ₂ -C ₄)	0,01	0,10	0,16	0,6957
α (C ₁ -C ₂ -C ₄)	0,00	0,00	0,01	0,9345
β (C ₁ -C ₂ -C ₄ -C ₈)	0,00	0,00	0,03	0,8637

Analizando as tabelas de regressão simples fornecidas anteriormente pudemos concluir que algumas propriedades moleculares apresentavam de modo bastante relevante um elevado coeficiente de explicação (r^2). Estas propriedades foram selecionadas (vermelho) a participar da próxima etapa estatística de busca de variáveis descritoras. As demais propriedades foram descartadas por apresentarem baixo coeficiente de correlação (r) para com a atividade biológica e elevado valor p . Antes de descartarmos algumas variáveis construímos a chamada matriz de correlação com finalidade em confrontar as variáveis entre si para checar as possíveis correlações diretas entre elas e assim facilitar a exclusão de variáveis na primeira etapa seletiva.

5.3.2 Matriz de correlação e seleção de variáveis

A segunda etapa estatística realizada em nossa pesquisa QSAR fundamentou-se na busca de variáveis descritoras para construção do modelo linear de Hansch – Fujita (Hansch & Fujita, 1964). Para afirmarmos que a nossa proposta de modelo é uma proposta plausível e que justifica a atividade dos compostos da série, utilizamos devida cautela na escolha das variáveis, seja selecionando descritores ou no processo de descarte das variáveis que não farão parte da equação. Para que tal procedimento fosse concretizado de modo racional preferimos subdividir a nossa pesquisa em etapas:

Primeira etapa seletiva de variáveis: Seleção de variáveis levando em consideração a regressão linear (Analyse-it) e os coeficientes de correlação elevados ($R_y > 0,6$) entre variáveis (BuildQSAR). Foram construídas as matrizes de correlação para cada categoria de parâmetros (tabelas 26 - 32) com critério limite de correlação entre variáveis 0,6. As variáveis destacadas em vermelho são altamente correlacionadas entre si, não devendo então fazer parte na equação do modelo linear.

Os parâmetros selecionados (22 variáveis) na primeira etapa foram:

A) Hidrofóbicos:

- $\log P_{\text{Ghose e Crippen}}$ – Log do coeficiente de partição octanol/água (Ghose e Crippen);
- $\log S_{\text{H}_2\text{O}}$ – Log da solubilidade em água;
- D. Hansen 3D – Dispersão de Hansen 3D;
- $MR_{\text{Ghose e Crippen}}$ – Refratividade molar (Ghose e Crippen);
- P. S. – Parâmetro de solubilidade.

B) Eletrônicos empíricos:

- $\sigma_p(A)$ – Hammett sigma para no substituinte A;
- $\sigma_m(A)$ – Hammett sigma meta no substituinte A.

C) Eletrônicos quânticos:

- $q(C_3)$ – Carga atômica parcial no carbono C_3 ;
- H. B. donor – Contribuição doadora de ligação de hidrogênio;
- E_{HOMO} – Energia do orbital molecular mais alto ocupado;

- E_{LUMO} – Energia do orbital molecular mais baixo desocupado;
- H. B. acceptor – Contribuição receptora de ligação de hidrogênio;
- $q(O_7)$ – Carga atômica parcial no oxigênio O_7 ;
- $q(C_4)$ – Carga atômica parcial no carbono C_4 .

D) Estereoquímicos:

- L_x – Comprimento molecular;
- L_y – Largura molecular.

E) Estereoquímicos de Verloop:

- $\ni B4(A)$ – Verloop B4 para o substituinte A;
- $\ni L1(A)$ – Verloop L1 para o substituinte A.

F) Termodinâmicos:

- TC – Temperatura crítica;
- HBN – Número de ligações de hidrogênio;
- tau – Número de ligações torsionais.

G) Outros parâmetros:

- $d(C_2-C_4)$ – Distância interatômica entre carbonos C_2 e C_4 .

Segunda etapa seletiva de variáveis: Seleção de variáveis descritoras levando em consideração as 22 variáveis escolhidas na primeira etapa seletiva, com o limite do coeficiente de correlação entre variáveis 0,4 (tabela 33). Os cálculos dos coeficientes de correlação foram executados, que por consequência resultou em uma seleção de apenas 14 variáveis para a realização da RLM. As 14 variáveis descritoras selecionadas (tabela 34) a participar da Regressão Linear Múltipla foram:

Hidrofóbicos: $\log P_{\text{Ghose e Crippen}}$ e $MR_{\text{Ghose e Crippen}}$; **Eletrônicos:** E_{HOMO} , E_{LUMO} , H. B. acceptor, H. B. donor, $q(C_3)$, $q(O_7)$ e $q(C_4)$; **Estereoquímicos:** L_x e L_y ; **Estereoquímicos de Verloop:** $\ni B4(A)$ e $\ni L1(A)$; **Outros parâmetros:** $d(C_2-C_4)$.

Tabela 26: Matriz de correlação entre parâmetros hidrofóbicos

	Log P Ochiai & Coppes	Log S. H&O	D. Hansen 3D	MR Ochiai & Coppes	P. S.	ASH	% ASH	HLB	δ Hansen 3D	NH	H. B. Hansen 3D	Log P Hansch
Log P Ochiai & Coppes	1	0,727	0,575	0,171	0,48	0,143	0,135	0,135	0,304	0,033	0,206	0,276
Log S. H&O		1	0,891	0,372	0,88	0,413	0,374	0,374	0,606	0,384	0,017	0,012
D. Hansen 3D			1	0,187	0,88	0,424	0,415	0,415	0,86	0,318	0,006	0,011
MR Ochiai & Coppes				1	0,2	0,051	0,014	0,014	0,008	0,253	0,232	0,034
P. S.					1	0,537	0,519	0,519	0,725	0,442	0,002	0,001
ASH						1	0,931	0,981	0,785	0,767	0,054	0,088
%ASH							1	1	0,822	0,874	0,024	0,086
HLB								1	0,822	0,874	0,024	0,086
δ Hansen 3D									1	0,406	0,004	0,016
NH										1	0,295	0,195
H. B. Hansen 3D											1	0,097
Log P Hansch												1

Tabela 27: Matriz de correlação entre parâmetros eletrônicos empíricos

	σ_p (A)	σ_m (A)	σ^+ (A)	α (A)	σ_p (B)	σ_m (B)	σ^+ (B)	α (B)
σ_p (A)	1	0,988	0,677	0,736	0,184	0,198	0,054	0
σ_m (A)		1	0,768	0,818	0,18	0,187	0,071	0
σ^+ (A)			1	0,961	0,149	0,135	0,109	0,014
α (A)				1	0,116	0,101	0,147	0,028
σ_p (B)					1	0,974	0,012	0,009
σ_m (B)						1	0,01	0
σ^+ (B)							1	0,271
α (B)								1

Nota: Os valores em vermelho correspondem aos coeficientes de correlação entre variáveis maiores que 0,6

Tabela 28: Matriz de correlação entre parâmetros eletrônicos quânticos

	μ	E HOMO	E LUMO	H. B. acceptor	H. B. donor	q (C ₁)	q (C ₂)	q (C ₃)	q (C ₄)	q (C ₅)	q (O ₇)
μ	1	0,339	0,24	0,047	0,006	0,065	0,175	0,027	0,044	0,004	0,001
E _{HOMO}		1	0,486	0,399	0,006	0,029	0,1	0,012	0,095	0,267	0,093
E _{LUMO}			1	0,498	0,018	0,088	0,076	0,001	0,064	0,185	0,048
H. B. acceptor				1	0,015	0,072	0,013	0,043	0,163	0,328	0,216
H. B. donor					1	0,204	0,119	0,411	0,241	0	0,058
q (C ₁)						1	0,666	0,572	0,013	0,147	0,142
q (C ₂)							1	0,461	0,009	0,143	0,106
q (C ₃)								1	0,052	0	0,027
q (C ₄)									1	0,166	0,469
q (C ₅)										1	0,353
q (O ₇)											1

Tabela 29: Matriz de correlação entre parâmetros estereoquímicos

	ϕ L1 (A)	ϕ B1 (A)	ϕ B2 (A)	ϕ B3 (A)	ϕ B4 (A)	ϕ L1 (B)	ϕ B1 (B)	ϕ B2 (B)	ϕ B3 (B)	ϕ B4 (B)	SvdW _{int}	SvdW	L ₁	L ₂	L ₃	VM	p
ϕ L1(A)	1	0,006	0,065	0	0,204	0,007	0	0,145	0,045	0,024	0,047	0,042	0,025	0,205	0,039	0,02	0,088
ϕ B1(A)		1	0,072	0,076	0,187	0,040	0,083	0,028	0,036	0,03	0,008	0,03	0,170	0,014	0,041	0	0,003
ϕ B2(A)			1	0,179	0,125	0,002	0,067	0,267	0,094	0,194	0,109	0,053	0,001	0,090	0,006	0,01	0,002
ϕ B3(A)				1	0,149	0,068	0,045	0	0,016	0	0,056	0,005	0,005	0,046	0,002	0	0,028
ϕ B4(A)					1	0,014	0,008	0,037	0,070	0,121	0,002	0,008	0	0,057	0,014	0,01	0
ϕ L1(B)						1	0,02	0,098	0,369	0,600	0,289	0,252	0,071	0,419	0	0,28	0,074
ϕ B1(B)							1	0,093	0,413	0,014	0,117	0,082	0,001	0	0,001	0,07	0,008
ϕ B2(B)								1	0,011	0,479	0,005	0,03	0	0,023	0,134	0,15	0,107
ϕ B3(B)									1	0,106	0,028	0,044	0	0,217	0,009	0,02	0,011
ϕ B4(B)										1	0,109	0,145	0,067	0,134	0,008	0,3	0,057
SvdW _{int}											1	0,776	0,037	0,558	0,002	0,56	0,140
SvdW												1	0,002	0,358	0,095	0,58	0,101
L ₁													1	0,282	0,082	0	0,115
L ₂														1	0,081	0,26	0,029
L ₃															1	0,15	0,281
VM																1	0,135
p																	1

Nota: Os valores em vermelho correspondem aos coeficientes de correlação entre variáveis maiores que 0,6

Tabela 30: Matriz de correlação entre parâmetros termodinâmicos

	TC	HBN	tau	PC	PE	PA
TC	1	0,394	0,22	0,213	0,56	0,282
HBN		1	0,62	0,627	0,199	0,027
tau			1	0,828	0,068	0,193
PC				1	0,061	0,242
PE					1	0,22
PA						1

Tabela 31: Matriz de correlação entre parâmetros de biodisponibilidade

	PM	AS _{polár}
PM	1	0,131
AS _{polár}		1

Tabela 32: Matriz de correlação entre distâncias interatômicas, ângulos entre ligações e ângulos diedrais

	d (C-C)	d (C-C)	d (C-C)	d (C-C)	d (C-O)	α (C-C-C)	α (C-C-C)	α (C-C-C)	α (C-C-O)	α (O-C-C)	β (C-C-C)	β (C-C-C)	β (C-C-O)
d(C-C)	1	0,123	0	0,001	0,057	0,002	0,083	0,002	0,018	0,002	0,023	0,033	0,085
d(C-C)		1	0,071	0,023	0,001	0,024	0	0,007	0,043	0,001	0,001	0,003	0,048
d(C-C)			1	0,001	0,006	0,003	0	0,194	0,09	0,237	0,010	0	0,005
d(C-C)				1	0,109	0,005	0,031	0,298	0,531	0,090	0,003	0	0,002
d(C-O)					1	0,007	0	0,166	0,153	0,119	0,052	0,053	0,26
α (C-C-C)						1	0,396	0,171	0,081	0,212	0,045	0,117	0,002
α (C-C-C)							1	0,167	0,162	0,136	0,209	0,264	0
α (C-C-C)								1	0,811	0,85	0,114	0,084	0
α (C-C-O)									1	0,439	0,051	0,026	0,001
α (O-C-C)										1	0,18	0,139	0,002
β (C-C-C-C)											1	0,501	0,214
β (C-C-C-O)												1	0,181
β (C-C-C-O)													1

Nota: Os valores em vermelho correspondem aos coeficientes de correlação entre variáveis maiores que 0,6.

Tabela 33: Matriz de correlação entre as variáveis selecionadas (22 variáveis) na primeira etapa seletiva; Critério adotado: Limite do coeficiente de correlação = 0,4.

	Log P Ghose e Crippen	Log S H ₂ O	D. Hansen 3D	MR Ghose e Crippen	P.S.	E _{HOMO}	E _{LUMO}	H.B. acceptor
Log P _{Ghose e Crippen}	1	0,727	0,675	0,171	0,482	0,415	0,370	0,405
Log S _{H₂O}		1	0,581	0,372	0,575	0,714	0,714	0,428
D. _{Hansen 3D}			1	0,187	0,378	0,537	0,821	0,827
MR _{Ghose e Crippen}				1	0,202	0,161	0,355	0,088
P.S.					1	0,912	0,622	0,815
E _{HOMO}						1	0,486	0,400
E _{LUMO}							1	0,498
H.B. acceptor								1
H.B. donor								
q (C ₂)								
q (C ₄)								
q (O ₂)								
α _p (A)								
α _m (A)								
d (C ₂ -C ₄)								
TC								
HBN								
tau								
L _x								
L _y								
≥ L1 (A)								
≥ B4 (A)								

	H.B. donor	q (C ₂)	q (C ₄)	q (O ₂)	α _p (A)	α _m (A)	d (C ₂ -C ₄)	TC
Log P _{Ghose e Crippen}	0,005	0,22	0,088	0,203	0,219	0,175	0,187	0,374
Log S	0,039	0,11	0,14	0,103	0,82	0,575	0,366	0,720
D. _{Hansen 3D}	0	0,04	0,121	0,093	0,507	0,468	0,271	0,832
MR _{Ghose e Crippen}	0,232	0,108	0,097	0,031	0,189	0,204	0,184	0,898
P.S.	0,002	0,033	0,156	0,09	0,512	0,485	0,277	0,889
E _{HOMO}	0,008	0,012	0,095	0,093	0,457	0,400	0,210	0,541
E _{LUMO}	0,019	0,001	0,064	0,048	0,823	0,810	0,341	0,766
H.B. acceptor	0,015	0,043	0,163	0,216	0,331	0,297	0,141	0,515
H.B. donor	1	0,411	0,241	0,058	0,004	0,006	0,059	0,172
q (C ₂)		1	0,062	0,027	0	0	0,072	0,086
q (C ₄)			1	0,489	0,025	0,022	0,016	0,184
q (O ₂)				1	0	0,001	0,004	0,097
α _p (A)					1	0,988	0,347	0,556
α _m (A)						1	0,334	0,551
d (C ₂ -C ₄)							1	0,354
TC								1
HBN								
tau								
L _x								
L _y								
≥ L1 (A)								
≥ B4 (A)								

	HBN	tau	L _x	L _y	≥ L1 (A)	≥ B4 (A)
Log P _{Ghose e Crippen}	0,551	0,428	0,251	0,022	0,073	0,006
Log S	0,758	0,473	0,245	0,005	0,002	0,005
D. _{Hansen 3D}	0,723	0,706	0,027	0,001	0,001	0,320
MR _{Ghose e Crippen}	0,056	0,001	0,011	0,181	0,013	0,025
P.S.	0,728	0,687	0,013	0,002	0,001	0,303
E _{HOMO}	0,708	0,559	0,143	0,001	0,006	0,019
E _{LUMO}	0,434	0,353	0,085	0,007	0,096	0,098
H.B. acceptor	0,522	0,645	0,016	0,041	0,017	0,444
H.B. donor	0,001	0,023	0,082	0,001	0,054	0,022
q (C ₂)	0,078	0,009	0,131	0,043	0,268	0,014
q (C ₄)	0,157	0,063	0,035	0,022	0,04	0,001
q (O ₂)	0,139	0,076	0,125	0,040	0,031	0
α _p (A)	0,393	0,373	0,088	0	0,075	0,050
α _m (A)	0,347	0,319	0,054	0,007	0,106	0,057
d (C ₂ -C ₄)	0,214	0,111	0,102	0,004	0,005	0,013
TC	0,394	0,22	0,054	0,036	0,014	0,129
HBN	1	0,823	0,191	0,004	0,001	0,010
tau		1	0,094	0,067	0,003	0,057
L _x			1	0,282	0,025	0,170
L _y				1	0,205	0,014
≥ L1 (A)					1	0,006
≥ B4 (A)						1

Tabela 34: Matriz de correlação entre as variáveis descritoras selecionadas (14 variáveis);
Critério adotado: Limite do coeficiente de correlação = 0,4

	Log P Ghose e Crippen	MR Ghose e Crippen	E _{HOMO}	E _{LUMO}	H. B. acceptor	H. B. donor	q (C ₃)
Log P Ghose e Crippen	1	0,171	0,415	0,37	0,405	0,005	0,22
MR Ghose e Crippen		1	0,181	0,355	0,088	0,232	0,106
E _{HOMO}			1	0,486	0,399	0,006	0,012
E _{LUMO}				1	0,498	0,018	0,001
H. B. acceptor					1	0,015	0,043
H. B. donor						1	0,411
q (C ₃)							1
q (C ₄)							
q (O ₇)							
d (C ₂ -C ₄)							
L _x							
L _y							
Δ L1 (A)							
Δ B4 (A)							

	q (C ₄)	q (O ₇)	d (C ₂ -C ₄)	L _x	L _y	Δ L1 (A)	Δ B4 (A)
Log P Ghose e Crippen	0,088	0,203	0,187	0,251	0,022	0,073	0,044
MR Ghose e Crippen	0,097	0,031	0,184	0,011	0,181	0,013	0,071
E _{HOMO}	0,095	0,093	0,21	0,143	0,001	0,006	0,022
E _{LUMO}	0,064	0,048	0,341	0,085	0,007	0,096	0,03
H. B. acceptor	0,163	0,216	0,141	0,015	0,041	0,017	0,164
H. B. donor	0,241	0,058	0,059	0,082	0,001	0,064	0,183
q (C ₃)	0,052	0,027	0,072	0,131	0,043	0,268	0,367
q (C ₄)	1	0,469	0,016	0,035	0,022	0,04	0,048
q (O ₇)		1	0,004	0,125	0,04	0,081	0,001
d (C ₂ -C ₄)			1	0,102	0,004	0,006	0,073
L _x				1	0,282	0,025	0
L _y					1	0,205	0,057
Δ L1 (A)						1	0,204
Δ B4 (A)							1

5.3.3 Seleção de variáveis descritoras

Uma vez selecionado as variáveis que participarão da regressão linear múltipla com objetivo em estabelecer uma estimativa de modelo, executamos um novo cálculo dos coeficientes de correlação entre as 14 variáveis para melhor compreender quais variáveis apresentam entre si uma maior correlação.

A próxima etapa da nossa pesquisa QSAR consistiu em selecionar quais são os possíveis descritores a participarem da equação proposta para o modelo de Hansch (Hansch & Fujita, 1964). Este mecanismo seletivo de variáveis realizado no programa BuildQSAR, foi desenvolvido por intermédio de duas metodologias: algoritmo genético e busca sistemática, visando uma maior segurança na proposição de uma estimativa de modelo e também por ser um procedimento comparativo entre resultados obtidos.

A) Algoritmo genético

Neste trabalho, o conjunto completo de dados compreende 69 variáveis, um valor de atividade biológica (-Log IC₅₀) e 19 derivados análogos de Chalcona com valores de IC₅₀ determinados experimentalmente. Levando em consideração as regras gerais em QSAR (Unger & Hansch, 1973) para evitar a ocorrência de *overfitting*, ou seja, ajuste forçado pelo excesso de variáveis descritoras incluídas no modelo, adota-se que cada variável descritora presente no modelo deverá acomodar cinco ou seis compostos.

Visando encontrar variáveis descritoras, realizamos algoritmo genético para as 14 variáveis selecionadas com características pré-determinadas como mencionado na metodologia anteriormente (4.3.2), tendo como indicador de atividade biológica o termo - Log IC₅₀. O conjunto de dados não foi alterado, mantendo-se sempre as 14 variáveis escolhidas. A variação na metodologia ocorreu apenas no que diz respeito ao número de variáveis descritoras (3, 4 ou 5) a serem escolhidas como candidatas ao modelo. Ao término do processo, foram selecionados os descritores relevantes e uma sequência de 10 modelos (tabelas 35, 36 e 37) contendo estes descritores levando em conta a ordem decrescente do coeficiente de correlação.

Tabela 35 - Seleção de variáveis descritoras por intermédio de algoritmo genético:

Conjunto de dados: todos os parâmetros selecionados (14 variáveis); modelo com 3 variáveis descritoras; 35% de mutação; 30 modelos por geração; 200 gerações; ajuste - coeficiente de correlação; indicador de atividade biológica: - Log IC₅₀

Var 1	Var 2	Var 3	r	s	F
MR _{Ghose e Crippen}	L _x	↗ B4 (A)	0,938	0,158	58,127
q (O ₇)	L _x	↗ B4 (A)	0,933	0,164	53,739
Log P _{Ghose e Crippen}	L _x	↗ B4 (A)	0,930	0,167	50,883
E _{LUMO}	L _x	↗ B4 (A)	0,921	0,176	45,007
q (C ₄)	L _x	↗ B4 (A)	0,920	0,178	44,138
H.B. donor	L _x	↗ B4 (A)	0,917	0,181	42,261
H.B. acceptor	L _x	↗ B4 (A)	0,917	0,181	42,130
q (C ₃)	L _x	↗ B4 (A)	0,914	0,184	40,780
E _{HOMO}	L _x	↗ B4 (A)	0,913	0,185	40,288
Log P _{Ghose e Crippen}	H. B. donor	L _y	0,912	0,186	39,465

Tabela 36 - Seleção de variáveis descritoras por intermédio de algoritmo genético:

Conjunto de dados: todos os parâmetros selecionados (14 variáveis); modelo com 4 variáveis descritoras; 35% de mutação; 30 modelos por geração; 200 gerações; ajuste - coeficiente de correlação; indicador de atividade biológica: - Log IC₅₀

Var 1	Var 2	Var 3	Var 4	r	s	F
MR _{Ghose e Crippen}	q (O ₇)	L _x	↗ B4 (A)	0,957	0,136	54,470
Log P _{Ghose e Crippen}	H. B. donor	L _x	↗ B4 (A)	0,953	0,142	49,227
Log P _{Ghose e Crippen}	MR _{Ghose e Crippen}	L _x	↗ B4 (A)	0,947	0,151	43,330
E _{LUMO}	q (O ₇)	L _x	↗ B4 (A)	0,945	0,154	41,455
MR _{Ghose e Crippen}	q (C ₄)	L _x	↗ B4 (A)	0,944	0,154	41,099
Log P _{Ghose e Crippen}	q (O ₇)	L _x	↗ B4 (A)	0,944	0,155	40,919
MR _{Ghose e Crippen}	H. B. acceptor	L _x	↗ B4 (A)	0,943	0,156	40,444
MR _{Ghose e Crippen}	q (C ₃)	L _x	↗ B4 (A)	0,942	0,157	39,525
E _{LUMO}	q (C ₃)	L _x	↗ B4 (A)	0,942	0,157	39,464
MR _{Ghose e Crippen}	L _x	L _y	↗ B4 (A)	0,942	0,158	39,301

Tabela 37 - Seleção de variáveis descritoras por intermédio de algoritmo genético:

Conjunto de dados: todos os parâmetros selecionados (14 variáveis); modelo com 5 variáveis descritoras; 35% de mutação; 30 modelos por geração; 200 gerações; ajuste - coeficiente de correlação; indicador de atividade biológica: - Log IC₅₀

Var 1	Var 2	Var 3	Var 4	Var 5	r	s	F
MR _{Ghose e Crippen}	q (C ₃)	q (O ₇)	L _x	⇒ B4 (A)	0,961	0,135	41,783
Log P _{Ghose e Crippen}	MR _{Ghose e Crippen}	q (O ₇)	L _x	⇒ B4 (A)	0,960	0,136	41,064
Log P _{Ghose e Crippen}	H. B. donor	q (O ₇)	L _x	⇒ B4 (A)	0,959	0,138	39,798
E _{LUMO}	q (C ₃)	q (O ₇)	L _x	⇒ B4 (A)	0,960	0,136	40,993
MR _{Ghose e Crippen}	q (O ₇)	L _x	L _x	⇒ B4 (A)	0,958	0,139	39,488
MR _{Ghose e Crippen}	q (O ₇)	d (C ₂ -C ₄)	L _x	⇒ B4 (A)	0,958	0,139	39,034
MR _{Ghose e Crippen}	E _{LUMO}	q (O ₇)	L _x	⇒ B4 (A)	0,958	0,140	38,686
MR _{Ghose e Crippen}	H. B. donor	q (O ₇)	L _x	⇒ B4 (A)	0,958	0,140	38,604
MR _{Ghose e Crippen}	q (C ₄)	q (O ₇)	L _x	⇒ B4 (A)	0,958	0,140	38,529
MR _{Ghose e Crippen}	q (O ₇)	L _x	⇒ L1 (A)	⇒ B4 (A)	0,957	0,141	38,216

B) Busca sistemática

A busca sistemática é o método mais seguro de seleção de variáveis descritoras utilizado na construção de modelos e consiste em combinar as variáveis de forma a analisar as possíveis equações de regressão em função de um número de descritores pré-fixados (Ferreira *et al*, 2002). A seleção sistemática retorna todas as combinações encontradas, tendo como resultado desta busca um grande número de combinações, de modo que para um elevado número de variáveis o mecanismo seletivo se torna inviável computacionalmente. Levando em consideração o fato da busca sistemática ser tão onerosa computacionalmente, é que lançamos mão dos artifícios mencionados anteriormente como o da exclusão de variáveis altamente correlacionadas entre si.

O procedimento busca sistemática de variáveis nos gerou um elevado número de combinações, onde selecionamos 10 modelos com 3, 4 e 5 variáveis (tabelas 38, 39 e 40) de acordo com o critério ordem decrescente em coeficiente de correlação.

Tabela 38 - Seleção de variáveis descritoras por intermédio de busca sistemática (3 variáveis); critérios: limite mínimo de coeficiente de correlação = 0,90; limite mínimo de $Q^2 = 0,85$.

Var 01	Var 02	Var 03	n	r	s	F
MR _{Ghose e Crippen}	Lx	э - B4 (A)	19	0,938	0,163	36,329
q (O ₇)	Lx	э - B4 (A)	19	0,933	0,169	33,587
Log P _{Ghose e Crippen}	Lx	э - B4 (A)	19	0,930	0,173	31,802
E _{LUMO}	Lx	э - B4 (A)	19	0,921	0,182	28,129
q (C ₄)	Lx	э - B4 (A)	19	0,920	0,184	27,586
H. B. acceptor	Lx	э - B4 (A)	19	0,917	0,187	26,331
H. B. donor	Lx	э - B4 (A)	19	0,917	0,187	26,413
q (C ₃)	Lx	э - B4 (A)	19	0,914	0,190	25,488
E _{HOMO}	Lx	э - B4 (A)	19	0,913	0,191	25,180
Log P _{Ghose e Crippen}	H. B. donor	Ly	19	0,912	0,193	24,666

Tabela 39 - Seleção de variáveis descritoras por intermédio de busca sistemática (4 variáveis); Critérios: limite mínimo de coeficiente de correlação = 0,90; limite mínimo de $Q^2 = 0,85$.

Var 01	Var 02	Var 03	Var 04	n	r	s	F
MR _{Ghose e Crippen}	q (O ₇)	Lx	э - B4 (A)	19	0,957	0,141	38,129
Log P _{Ghose e Crippen}	H. B. donor	Lx	э - B4 (A)	19	0,953	0,147	34,459
Log P _{Ghose e Crippen}	MR _{Ghose e Crippen}	Lx	э - B4 (A)	19	0,947	0,156	30,331
E _{LUMO}	q (O ₇)	Lx	э - B4 (A)	19	0,945	0,159	29,019
Log P _{Ghose e Crippen}	q (O ₇)	Lx	э - B4 (A)	19	0,944	0,160	28,643
MR _{Ghose e Crippen}	q (C ₄)	Lx	э - B4 (A)	19	0,944	0,160	28,770
MR _{Ghose e Crippen}	H. B. acceptor	Lx	э - B4 (A)	19	0,943	0,161	28,311
MR _{Ghose e Crippen}	q (C ₃)	Lx	э - B4 (A)	19	0,942	0,163	27,667
MR _{Ghose e Crippen}	Lx	Ly	э - B4 (A)	19	0,942	0,163	27,511
E _{LUMO}	q (C ₃)	Lx	э - B4 (A)	19	0,942	0,163	27,625

Tabela 40 - Seleção de variáveis descritoras por intermédio de busca sistemática (5 variáveis);
Critérios: limite mínimo de coeficiente de correlação = 0,90; limite mínimo de $Q^2 = 0,85$.

Var 01	Var 02	Var 03	Var 04	Var 05	n	r	s	F
MR _{Ghose e} Crippen	q (C ₃)	q (O ₇)	Lx	∑ - B4 (A)	19	0,961	0,140	31,039
Log P _{Ghose e} Crippen	MR _{Ghose e} Crippen	q (O ₇)	Lx	∑ - B4 (A)	19	0,960	0,141	30,505
Log P _{Ghose e} Crippen	H. B. donor	q (O ₇)	Lx	∑ - B4 (A)	19	0,959	0,143	29,564
E _{LUMO}	q (C ₃)	q (O ₇)	Lx	∑ - B4 (A)	19	0,960	0,141	30,452
MR _{Ghose e} Crippen	E _{LUMO}	q (O ₇)	Lx	∑ - B4 (A)	19	0,958	0,145	28,738
MR _{Ghose e} Crippen	H. B. donor	q (O ₇)	Lx	∑ - B4 (A)	19	0,958	0,145	28,681
MR _{Ghose e} Crippen	q (O ₇)	d (C ₂ -C ₄)	Lx	∑ - B4 (A)	19	0,958	0,145	28,997
MR _{Ghose e} Crippen	q (O ₇)	Lx	Ly	∑ - B4 (A)	19	0,958	0,144	29,334
MR _{Ghose e} Crippen	E _{HOMO}	q (O ₇)	Lx	∑ - B4 (A)	19	0,957	0,146	28,338
MR _{Ghose e} Crippen	H. B. acceptor	q (O ₇)	Lx	∑ - B4 (A)	19	0,957	0,146	28,365

5.3.4 Regressão linear múltipla

Após a realização da busca de variáveis sistematicamente e por algoritmo genético, a próxima etapa concretizada em nossa pesquisa QSAR visou a criação de modelos através de análise de regressão linear múltipla. Esta técnica consistiu em estimar uma equação através de combinações lineares das variáveis independentes (propriedades moleculares) de forma que tais variáveis fossem capazes de reproduzir de melhor maneira os valores de atividade biológica obtidos experimentalmente.

Realizamos a RLM para as dez combinações de 3, 4 e 5 variáveis no programa BuildQSAR obtendo então trinta modelos matemáticos. Analisando os valores de r, s e F apresentados nas tabelas 35 – 40, escolhemos as três melhores combinações de 3, 4 e 5 variáveis (tabelas 41, 42 e 43). Analisando as

tabelas 37 e 40, percebemos que o quarto modelo apresentado possui maior r , menor s e maior F . Este modelo não foi selecionado entre os três melhores pelo fato de apresentar um menor Q^2 (0,818) e maior **SPRESS** (0,215), sendo que os modelos selecionados em vermelho apresentaram maior grau de previsibilidade (maior Q^2 e menor **SPRESS**) – veja modelos selecionados (equações 79, 80 e 81). O pacote BuildQSAR oferece várias possibilidades de análise do conjunto de dados, inclusive a realização de regressão múltipla para construção de modelos parabólicos e bi-lineares. Neste trabalho, preocupamos em estimar equações multidimensionais lineares segundo a proposta de Hansch-Fujita (Hansch & Fujita, 1964). Escolhemos os modelos obtidos por intermédio de busca sistemática, visto que este processo levanta modelos levando em consideração os critérios: Coeficiente de correlação (r) e quadrado do coeficiente de correlação da previsão (Q^2).

Tabela 41 – Três melhores combinações apresentando 3 descritores.

Var 1	Var 2	Var 3	r	s	F
MR _{Ghose e Crippen}	L _x	↗ B4 (A)	0,938	0,163	36,329
q (O ₇)	L _x	↗ B4 (A)	0,933	0,169	33,587
Log P _{Ghose e Crippen}	L _x	↗ B4 (A)	0,930	0,173	31,802

Tabela 42 – Três melhores combinações apresentando 4 descritores.

Var 1	Var 2	Var 3	Var 4	r	s	F
MR _{Ghose e Crippen}	q (O ₇)	L _x	↗ B4 (A)	0,957	0,141	38,129
Log P _{Ghose e Crippen}	H. B. donor	L _x	↗ B4 (A)	0,953	0,147	34,459
Log P _{Ghose e Crippen}	MR _{Ghose e Crippen}	L _x	↗ B4 (A)	0,947	0,156	30,331

Tabela 43 – Três melhores combinações apresentando 5 descritores.

Var 1	Var 2	Var 3	Var 4	Var 5	r	s	F
MR _{Ghose e Crippen}	q (C ₃)	q (O ₇)	L _x	↗ B4 (A)	0,961	0,140	31,039
3Log P _{Ghose e Crippen}	MR _{Ghose e Crippen}	q (O ₇)	L _x	↗ B4 (A)	0,960	0,141	30,505
Log P _{Ghose e Crippen}	H. B. donor	q (O ₇)	L _x	↗ B4 (A)	0,959	0,143	29,564

5.3.5 Modelos multidimensionais lineares selecionados

Com o objetivo em estimarmos uma equação que correlacione as propriedades moleculares dos compostos com a atividade biológica, escolhemos uma trajetória de QSAR clássico, onde as variáveis utilizadas nas equações obtidas relacionam-se linearmente com a variação de energia livre envolvida no mecanismo de interação fármaco-receptor.

Os modelos lineares obtidos por RLM foram:

- Modelos com três variáveis descritoras:

Modelo 1

$$\begin{aligned} -\text{Log IC}_{50} = & + 0,01956 (\pm 0,01497) [\text{MR}] + 0,31852 (\pm 0,09207) [\text{Lx}] - 0,47444 \\ & (\pm 0,18016) [\text{B4 (A)}] - 3,84709 (\pm 1,641617) \\ (n = 019; r = 0,938; s = 0,163; F = 36,329; Q^2 = 0,779; \text{SPRESS} = 0,220); \end{aligned} \quad (73)$$

Modelo 2

$$\begin{aligned} -\text{Log IC}_{50} = & + 13,73306 (\pm 11,72024) [q(\text{O}_7)] + 0,28988 (\pm 0,10122) [\text{Lx}] - \\ & 0,54485 (\pm 0,17969) [\text{B4 (A)}] + 2,08822 (\pm 3,717744) \\ (n = 019; r = 0,933; s = 0,169; F = 33,587; Q^2 = 0,791; \text{SPRESS} = 0,215); \end{aligned} \quad (74)$$

Modelo 3

$$\begin{aligned} -\text{Log IC}_{50} = & + 0,24131 (\pm 0,22452) [\text{Log P}] + 0,26918 (\pm 0,11311) [\text{Lx}] - 0,48714 \\ & (\pm 0,18983) [\text{B4 (A)}] - 2,59968 (\pm 1,117187) \\ (n = 019; r = 0,930; s = 0,173; F = 31,802; Q^2 = 0,763; \text{SPRESS} = 0,228). \end{aligned} \quad (75)$$

Nota: O modelo selecionado acima com uma cor diferenciada (74) foi considerado por nós como sendo o modelo com três variáveis descritoras que apresentou maior grau de previsibilidade.

• Modelos com quatro variáveis descritoras:

Modelo 1

$$- \text{Log IC}_{50} = + 0,01701 (\pm 0,01316) [\text{MR}] + 11,58850 (\pm 9,96216) [q(O_7)] + 0,28493 (\pm 0,08492) [L_x] - 0,48869 (\pm 0,15675) [\rightarrow B4(A)] - 0,07859 (\pm 3,538740) \\ (n = 019; r = 0,957; s = 0,141; F = 38,129; Q^2 = 0,842; \text{SPRESS} = 0,193); \quad (76)$$

Modelo 2

$$- \text{Log IC}_{50} = + 0,29576 (\pm 0,19766) [\text{Log P}] - 15,78087 (\pm 13,06190) [\text{H. B. donor}] + 0,21999 (\pm 0,10515) [L_x] - 0,37751 (\pm 0,18630) [\rightarrow B4(A)] - 2,49014 (\pm 0,961884) \\ (n = 019; r = 0,953; s = 0,147; F = 34,459; Q^2 = 0,852; \text{SPRESS} = 0,187); \quad (77)$$

Modelo 3

$$- \text{Log IC}_{50} = + 0,15920 (\pm 0,22031) [\text{Log P}] + 0,01529 (\pm 0,01556) [\text{MR}] + 0,28013 (\pm 0,10330) [L_x] - 0,45485 (\pm 0,17546) [\rightarrow B4(A)] - 3,79185 (\pm 1,581366) \\ (n = 019; r = 0,947; s = 0,156; F = 30,331; Q^2 = 0,792; \text{SPRESS} = 0,221). \quad (78)$$

• Modelos com cinco variáveis descritoras:

Modelo 1

$$- \text{Log IC}_{50} = + 0,01585 (\pm 0,01339) [\text{MR}] + 1,46784 (\pm 2,94726) [q(C_3)] + 11,30517 (\pm 9,99361) [q(O_7)] + 0,26557 (\pm 0,09352) [L_x] - 0,42625 (\pm 0,20091) \\ [\rightarrow B4(A)] - 0,04384 (\pm 3,544843) \\ (n = 019; r = 0,961; s = 0,140; F = 31,039; Q^2 = 0,844; \text{SPRESS} = 0,199); \quad (79)$$

Modelo 2

$$- \text{Log IC}_{50} = + 0,09455 (\pm 0,21182) [\text{Log P}] + 0,01481 (\pm 0,01417) [\text{MR}] + 10,05966 (\pm 10,62478) [q(O_7)] + 0,26656 (\pm 0,09510) [L_x] - 0,47518 (\pm 0,16112) \\ [\rightarrow B4(A)] - 0,54296 (\pm 3,721016) \\ (n = 019; r = 0,960; s = 0,141; F = 30,505; Q^2 = 0,837; \text{SPRESS} = 0,204); \quad (80)$$

Modelo 3

$$- \text{Log IC}_{50} = + 0,23947 (\pm 0,21294) [\text{Log P}] - 13,25671 (\pm 13,38782) [\text{H. B. donor}] + 7,11669 (\pm 11,28870) [q(O_7)] + 0,21850 (\pm 0,10287) [L_x] - 0,40870 (\pm 0,18881) \\ [\rightarrow B4(A)] - 0,23590 (\pm 3,697423) \\ (n = 019; r = 0,959; s = 0,143; F = 29,564; Q^2 = 0,867; \text{SPRESS} = 0,184) \quad (81)$$

Nota: Os modelos selecionados acima com uma cor diferenciada (77 e 81) foram considerados por nós como sendo os modelos com quatro e cinco variáveis descritoras que apresentaram o maior grau de previsibilidade.

Como mencionado anteriormente, o quarto modelo das tabelas 37 e 40 não foi selecionado entre os três melhores contendo cinco variáveis descritoras, pois, notamos que após realizarmos a RLM deste modelo o seu grau de previsibilidade foi menor do que o dos três modelos escolhidos (79, 80 e 81).

5.3.6 Análise dos modelos selecionados

Em QSAR, para evitar a ocorrência de correlação por coincidência, seria interessante estabelecermos uma equação apresentando cinco ou seis compostos para cada variável descritora incluída no modelo (Gaudio e Zandonade, 2001). Nesta pesquisa, o número de derivados de Chalcona estudados é 19, portanto, a equação deveria apresentar no máximo quatro variáveis descritoras. No entanto, Kubinyi ressalta que, para conjuntos de dados com poucos compostos, pode-se violar a regra aplicada ao QSAR clássico desde que a avaliação do modelo seja justificada com o acréscimo de mais variáveis explicativas (Kubinyi, 1993).

De acordo com a literatura (Kubinyi, 2003) a qualidade de um modelo pode ser avaliada em função dos valores **r**, **s**, **F**, **Q²** e **SPRESS** obtidos durante o processo de construção do modelo. O valor **r** serve como medida relativa da qualidade do modelo, enquanto o valor **s** serve como medida absoluta da qualidade do modelo. O valor **F** é um parâmetro indicativo da significância estatística do modelo. Os valores **Q²** e **SPRESS** representam medidas do grau de previsibilidade do modelo.

A) Comparação dos modelos com três descritores

Comparando os modelos estimados com três variáveis explicativas (73, 74 e 75), chegamos à conclusão que o modelo 2 (74) é o que apresenta maior grau de previsibilidade (maior **Q²** e menor **SPRESS**). O modelo 2 (74) apresenta as seguintes variáveis descritoras:

$q(O_7)$ – carga parcial no átomo O_7 ;

L_x – comprimento molecular;

$\exists B4(A)$ – parâmetro de Verloop $B4$ do substituinte A (maior raio do substituinte A).

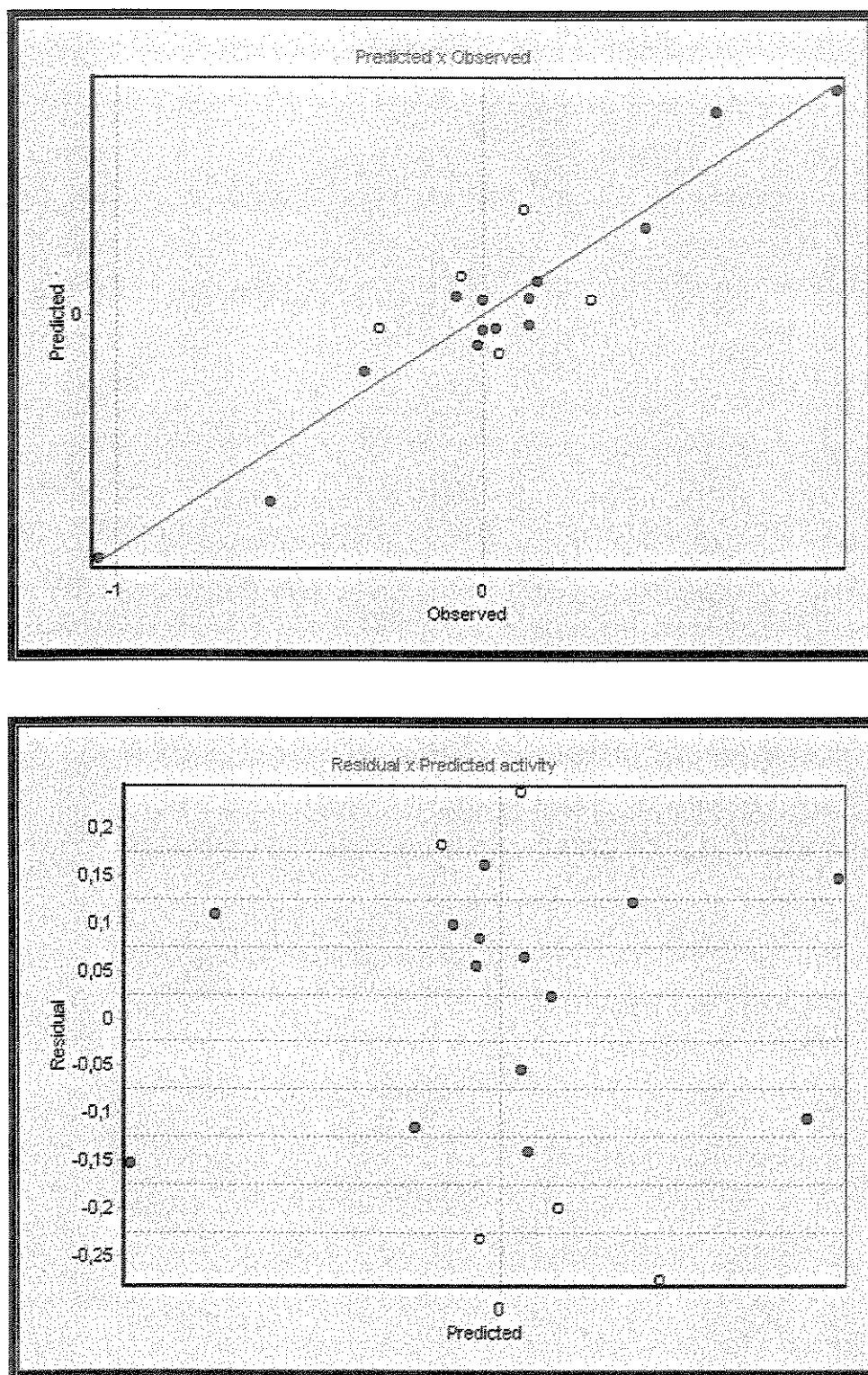
A análise do grau de ajuste de um modelo pode ser feita por intermédio dos diagramas da atividade observada ($-\text{Log IC}_{50 \text{ Obs}}$) em função da atividade prevista ($-\text{Log IC}_{50 \text{ Prev}}$) e também através do diagrama dos resíduos em função da atividade prevista. Compostos que apresentam elevados resíduos (*outliers*) em um determinado modelo devem ser excluídos do modelo para que o ajuste do modelo seja melhorado.

No pacote BuildQSAR, a análise gráfica da regressão pode ser mostrada através de pontos em um gráfico que representam os derivados de Chalcona e podem ser diferenciados pelas cores (azul, amarelo e vermelho) de modo que: azul – representa os compostos que melhor se ajustam ao modelo; vermelho – representa os compostos *outliers* (elevado resíduo) e amarelo – representa os compostos que estão na faixa dos limites de confiança da regressão (95%).

O modelo **2 (74)** com três descritores, é capaz de explicar 87% da variabilidade dos valores observados de atividade biológica, pois apresenta $r = 0,933$ possuindo então um $r^2 = 0,87$ (coeficiente de explicação). A escolha deste modelo indica que a atividade biológica será justificada em função da carga parcial no átomo de oxigênio da carbonila (O_7), em função comprimento da molécula (L_x), e em função do maior raio do substituinte A ($\exists B4$).

Plotamos então os diagramas para analisar se este modelo apresenta *outliers* e para verificar os compostos que se ajustam melhor a este modelo (figura 22).

Figura 22 - Análise gráfica do modelo 2 (74) com 3 variáveis descritoras obtida no programa BuildQSAR: atividade observada x atividade prevista (sup.); resíduos dos valores previstos (inf.).



B) Comparação dos modelos com quatro descritores

Ao comparar os modelos obtidos com quatro variáveis explicativas (76, 77 e 78), percebemos que os modelos 1(76) e 2 (77) são equivalentes, apresentando valores muito próximos de r , s e F . Analisando os valores do coeficiente de correlação (r), desvio padrão (s) e o valor do teste de Fischer (F), o modelo 1 (76) apresenta ligeira vantagem sobre o modelo 2 (77). Como a análise do valor do coeficiente de correlação é um indicativo do grau de ajuste do modelo, podemos então afirmar que o modelo 1 (76) é capaz de explicar 91,5% da variabilidade dos valores observados de atividade biológica, enquanto o modelo 2 (77) é capaz de justificar 90,8% da variabilidade dos valores, ou seja, os modelos 1 (76) e 2 (77) apresentam excelente nível de ajuste.

O grau de ajuste dos modelos 1 (76) e 2 (77) foi também analisado através dos gráficos (figuras 23 e 24) da atividade observada em função da atividade prevista e pelo gráfico dos resíduos em função da atividade prevista.

- O modelo 1 (76) apresenta as seguintes variáveis descritoras:
MR – refratividade molar;
 $q(O_7)$ – carga parcial no átomo O_7 ;
 L_x – comprimento molecular;
↗ B4 (A) – Verloop B4 do substituinte A.
- O modelo 2 (77) apresenta as seguintes variáveis descritoras:
Log P – Log do coeficiente de partição octanol/água (Ghose e Crippen);
H. B. donor – contribuição doadora de ligação de hidrogênio;
 L_x – comprimento molecular;
↗ B4 (A) – Verloop B4 do substituinte A

Observando os valores de Q^2 e SPRESS dos modelos 1 (76) e 2 (77), comparamos o grau de previsibilidade destes modelos. O modelo 2 (77) apresenta maior coeficiente de previsão (Q^2) quando comparado ao modelo 1(76) e apresenta menor desvio da validação cruzada (SPRESS). Percebemos então que escolher um modelo entre muitos que são equivalentes, é uma tarefa nada fácil. Se levarmos em consideração o grau de ajuste do modelo, ficaríamos com o modelo 1(76), mas se fossemos analisar o grau de previsibilidade do modelo,

escolheríamos o modelo 2 (77). Deste modo, preferimos então escolher o modelo 2 (77), por ser o modelo que prevê melhor.

Figura 23 – Análise gráfica do modelo 1 (76) com 4 variáveis descritoras obtida no programa BuildQSAR: atividade observada x atividade prevista (sup.); resíduos dos valores previstos (inf.).

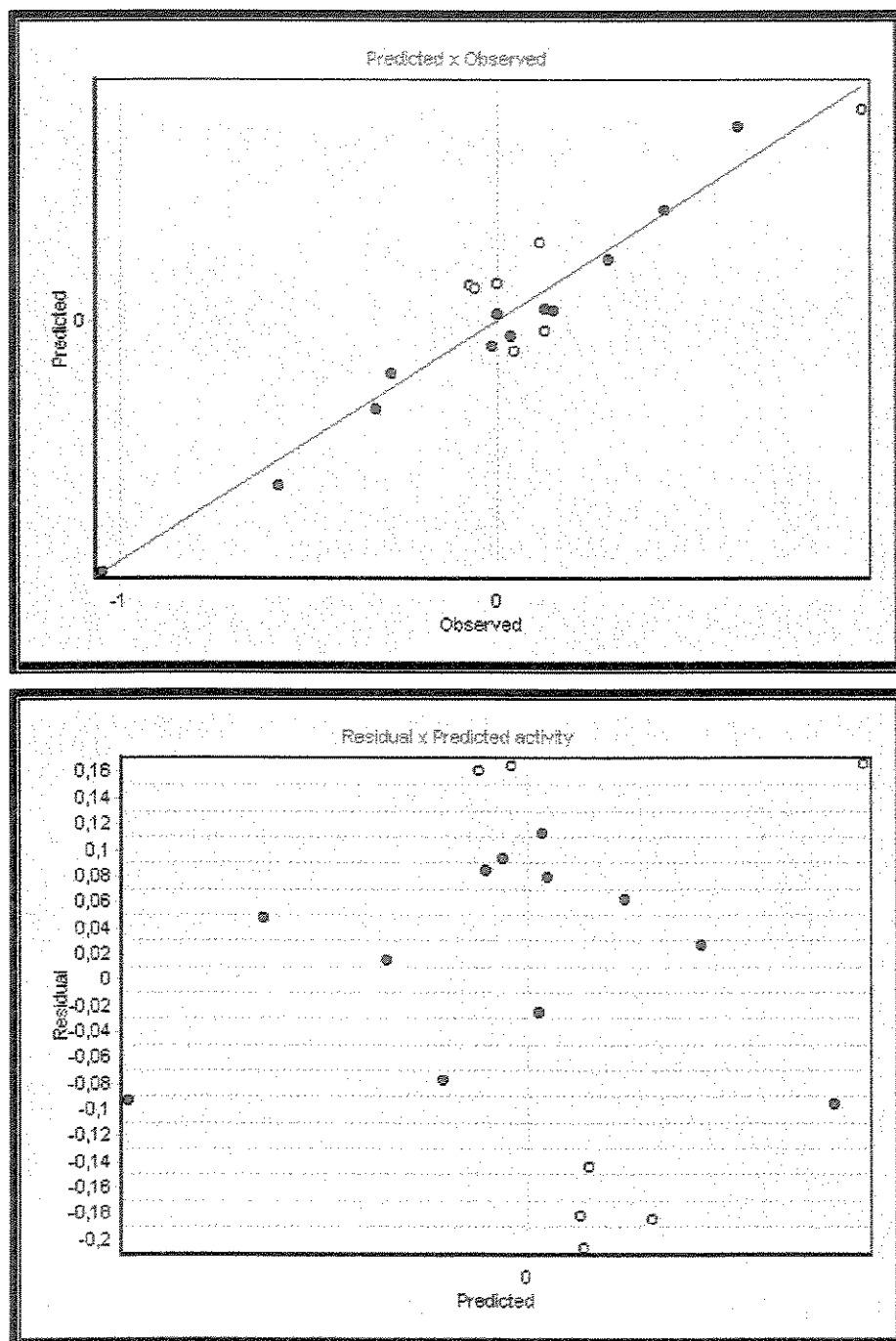
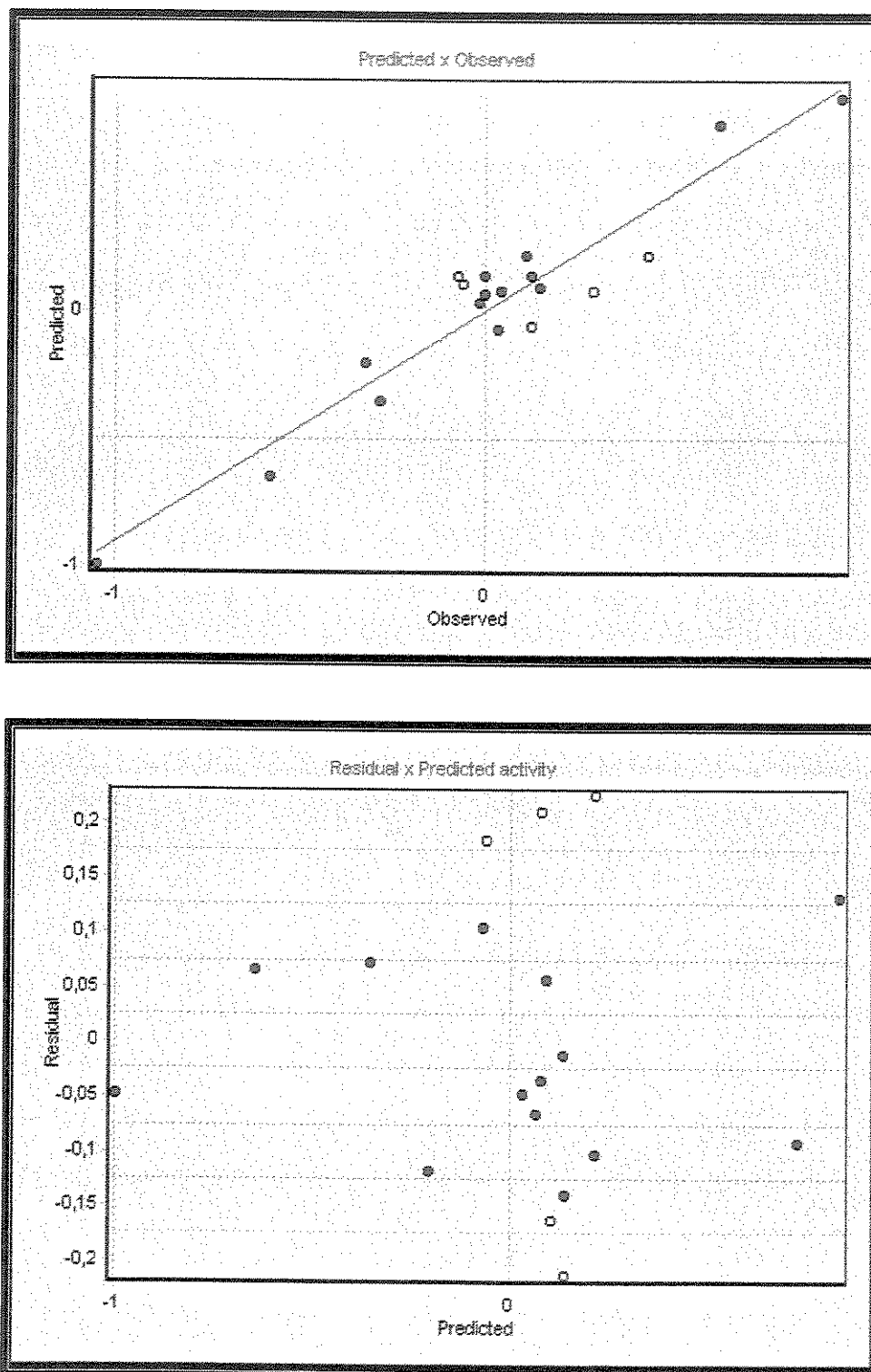


Figura 24 – Análise gráfica do modelo 2 (77) com 4 variáveis descritoras obtida no programa BuildQSAR: atividade observada x atividade prevista (sup.); resíduos dos valores previstos (inf.).



C) Comparação dos modelos com cinco descritores

Verificando os modelos com cinco variáveis descritoras (79, 80 e 81), concluímos que o modelo 3 (81) é o modelo que apresenta maior grau de previsibilidade. Este modelo apresenta as seguintes variáveis descritoras:

Log P – Log do coeficiente de partição octanol/água (Ghose e Crippen);

H. B. donor – contribuição doadora de ligação de hidrogênio;

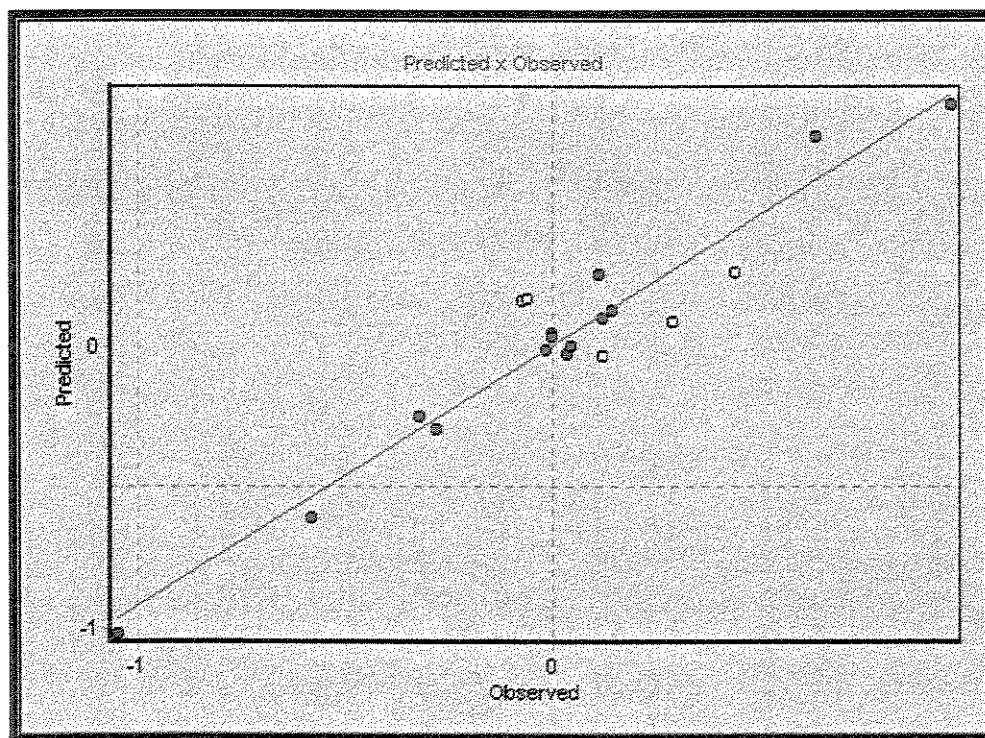
q (O₇) – carga parcial no átomo O₇;

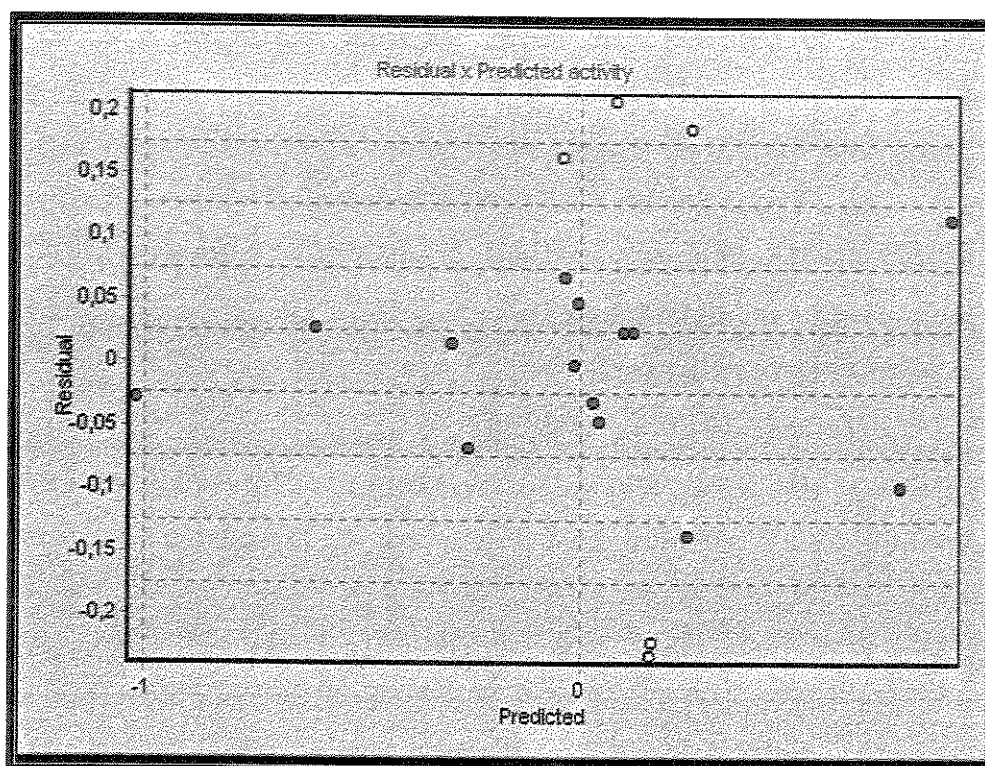
Lx – comprimento molecular;

↻ B4 (A) – Verloop B4 do substituinte A

Os gráficos, atividade observada x atividade prevista e resíduo x atividade prevista também foram plotados para o modelo 3 (81) com cinco descritores (figura 25).

Figura 25 – Análise gráfica do modelo 3 (81) com 5 variáveis descritoras obtida no programa BuildQSAR: atividade observada x atividade prevista (sup.); resíduos dos valores previstos (inf.).





5.3.7 Modelo multidimensional linear proposto

Levando em consideração os parâmetros estatísticos clássicos para validação de modelos em QSAR, devemos levar em conta alguns procedimentos indicativos da qualidade de um modelo. Parâmetros como o número de variáveis escolhidas para o modelo (critério de Topliss) (Topliss & Costello, 1972), o grau de ajuste do modelo, grau de significância e o seu grau de previsibilidade, não podem ser desconsiderados. A validação de um modelo consiste em verificar se o modelo adapta-se convenientemente aos dados observados de atividade biológica. Escolhemos como estimativa de modelo linear, o modelo 2 (77 ou 82) com quatro variáveis independentes.

Modelo proposto:

$$\begin{aligned}
 -\text{Log IC}_{50} = & + 0,29576 (\pm 0,19766) [\text{Log P}] - 15,78087 (\pm 13,06190) [\text{H. B.donor}] + \\
 & + 0,21999 (\pm 0,10515) [\text{Lx}] - 0,37751 (\pm 0,18630) [\text{B4 (A)}] - 2,49014 (\pm 0,961884) \\
 (n = 019; r = 0,953; s = 0,147; F = 34,459; Q^2 = 0,852; \text{SPRESS} = 0,187); \\
 p < 0,0001
 \end{aligned}
 \tag{82}$$

Tabela 44 – Matriz de correlação entre as variáveis participantes do modelo linear proposto; Critério adotado: limite do coeficiente de correlação = 0,4.

	Log P	H. B. donor	Lx	↳ B4 (A)
Log P	1	0,005	0,251	0,044
H. B. donor		1	0,082	0,183
Lx			1	0
↳ B4 (A)				1

A) Avaliação do grau de ajuste e significância do modelo

Para verificarmos o grau de ajuste e a significância do modelo escolhido, observamos os seguintes parâmetros: coeficiente de correlação (r), o desvio padrão e os diagramas da atividade observada x atividade prevista e resíduos x atividade prevista. Nosso modelo apresenta elevada capacidade para explicar os valores observados de atividade biológica, pois possui elevado coeficiente de determinação ou explicação (r^2), o que pode ser confirmado também pelo baixo desvio padrão. O modelo proposto apresenta $r^2 = 0,908$, ou seja, é capaz de explicar 90,8% da variabilidade dos dados observados.

Observando sua equação percebemos o elevado coeficiente de correlação ($r = 0,953$), baixo desvio padrão ($s = 0,147$) e para finalizar a análise do grau de ajuste, verificamos os diagramas da atividade observada x atividade prevista e dos resíduos x atividade prevista, e percebemos uma distribuição normal dos compostos em torno de zero (figura 24). O nosso modelo de regressão, não apresenta compostos com elevado resíduo, ou seja, não possui *outliers*, sendo que entre os 19 compostos pesquisados, 14 se ajustam muito bem ao modelo (pontos em azul) e apenas 5 se encontram na região do limite de confiança (pontos em amarelo).

Com intuito em comprovar o grau de ajuste do nosso modelo realizamos no programa Analyse-it (AME) a análise de variância (ANOVA) da regressão (tabela 45). Observando a tabela da análise de variância do modelo proposto, vimos que a estimativa da variância (s^2) ou quadrado do desvio possui um valor muito baixo sendo igual a 0,02172, o mesmo acontecendo com o valor de p , cujo valor é menor que 0,0001. O valor de p é um dos parâmetros utilizados na avaliação do

grau de significância do modelo. Regressões que apresentam um p-valor menor que 0,05 são consideradas estatisticamente significantes em relação ao valor de r^2 , indicando um erro menor que 5%. Nosso modelo apresentou um p-valor menor que 0,0001, ou seja, o nível de significância do modelo é maior que 99,99%.

Observando a tabela 45, notamos que o modelo foi ajustado em um $r^2 = 0,88$, ou seja, em um $r = 0,938$, verificando portanto que não ocorre falta de ajuste (*lack of fit*) do modelo.

Tabela 45 – Análise de variância (ANOVA) da regressão para o modelo proposto: SS - soma dos quadrados dos desvios, DF – graus de liberdade, MS – média dos quadrados dos desvios.

Test	RLM
Fit	- Log IC ₅₀ = f [Log P, H. B. donnor, Lx, ↗ B4 (A)]
n	19
r^2	0.91
r^2 ajustado	0.88
s	0.1474

Termo	Coefficiente	s	p	Intervalo de confiança (95%)
Intercept	-2.4901	0.4514	<0.0001	-3.4582 to -1.5220
Log P	0.2958	0.0928	0.0066	0.0968 to 0.4947
H.B. donnor	-15.7809	6.1295	0.0220	-28.9273 to -2.6345
Lx	0.2200	0.0493	0.0005	0.1142 to 0.3258
↗ B4 (A)	-0.3775	0.0874	0.0007	-0.5650 to -0.1900

ANOVA

Fonte de variação	SS	DF	MS	F	p
Regressão	2.994831	4	0.748708	34.46	<0.0001
Resíduo	0.304188	14	0.021728		
Total	3.299019	18			

O modelo proposto apresenta elevado valor F e um baixíssimo p -valor. O valor F , juntamente com o p -valor é analisado para avaliar a grau de significância do modelo proposto. O valor F da regressão é o resultado do quociente entre a média quadrática da regressão pela média quadrática dos resíduos. Calculando F , temos:

$$F = 0,748708 \div 0,021728 = 34,459;$$

$p < 0,0001$, ou seja, o modelo possui significância estatística.

B) Avaliação do grau de previsibilidade do modelo

O grau de previsibilidade do modelo é verificado por intermédio da validação cruzada (*cross validation*), executando o cálculo do quadrado do coeficiente de correlação da validação cruzada Q^2 (coeficiente de previsão) e do desvio padrão da validação cruzada (**SPRESS**). Um modelo com elevado grau de previsibilidade para derivados não incluídos no mesmo apresentará um Q^2 elevado e um mínimo **SPRESS**, assim, quanto mais próximo de 1 for Q^2 e mais próximo de zero estiver o **SPRESS**, maior será a previsibilidade do modelo.

Comparando todos os modelos propostos neste trabalho, ficamos entre o modelo 2 (77) com quatro descritores e o modelo 3 (81) com cinco descritores. O modelo 3 (81) (5 descritores) apresenta Q^2 maior do que a nossa proposta de modelo, sendo então um modelo que prevê um pouco melhor do que o nosso modelo, mas como apresenta cinco variáveis independentes, levamos em consideração o critério de Topliss, por termos apenas 19 amostras (derivados de Chalcona). É claro que o modelo 3 (81) não escolhido, nos auxilia na compreensão dos mecanismos envolvidos nas relações entre estrutura química e atividade biológica, pois neste modelo temos a presença de mais uma propriedade descritora, que corresponde à carga parcial no átomo de oxigênio da carbonila [$q(O_7)$].

Mesmo sabendo que o modelo 3 (81) apresenta um grau de previsibilidade maior, a diferença entre este e o nosso modelo em relação ao Q^2 e ao **SPRESS** é muito pequena. Preferimos então levar em consideração o critério de Topliss (número de variáveis independentes) e o princípio da parcimônia.

6.0 Discussão

Partindo do princípio que aproximadamente um milhão de mortes são relatadas anualmente por causa da malária, que a resistência do parasita frente a drogas antimaláricas está aumentando cada vez mais, e levando em consideração o fato de drogas que antigamente eram eficientes e hoje não possuem a mesma eficácia contra os parasitas, torna-se urgente a descoberta de novos protótipos (*lead compound*) e o desenvolvimento racional de fármacos com potente atividade antimalarial, resultando o mínimo de efeitos adversos no organismo humano.

Mesmo sabendo que o número de cepas resistentes a cloroquina está aumentando vertiginosamente, a cloroquina ainda é uma droga utilizada no combate à malária. Como a ação da cloroquina ocorre no interior do vacúolo digestivo do parasita inibindo a conversão do grupo heme em hemozoína por mecanismo de complexação do fármaco com o grupo heme, é importante então evidenciarmos o desenvolvimento de compostos que apresentem uma estrutura química semelhante a cloroquina visando provocar um mecanismo de inibição tal qual semelhante bioquimicamente, ou então obter drogas que apresentem um mecanismo de inibição das enzimas do parasita envolvidas na degradação da hemoglobina. Atualmente, inúmeros compostos estão sendo sintetizados e testados contra o *Plasmodium falciparum* com intuito em verificar a atividade antimalarial. Entre eles, vários derivados de Chalcona foram sintetizados e testados *in vitro* contra cepas resistentes do *Plasmodium falciparum*, sendo que alguns derivados se mostraram ativos contra o parasita (Li *et al*, 1995).

Hoje, o foco das atenções na pesquisa de antimaláricos está voltado para a descoberta de protótipos que sejam inibidores de proteases malarias. Como várias categorias de proteases já foram identificadas (Rosenthal *et al*, 1987) por determinação seqüencial de aminoácidos através da biologia molecular e de técnicas da bioquímica, a estrutura destas proteases tornou-se alvo fácil para agentes antimalariais que tenham ação inibitória enzimática.

Uma protease cisteínica malarial conhecida e identificada é a falcipain. As proteases malarias cisteínicas são enzimas do parasita que participam na degradação da hemoglobina (hemoglobinas) no interior do vacúolo digestivo do

parasita como fonte de aminoácidos, sendo considerado então como um dos alvos promissores para a descoberta de novos agentes quimioterápicos (Rosenthal *et al*, 2002). Já foram identificados resíduos de fenilalanina, serina, asparagina e glutamina em subsítio S₂ e S₃ e resíduo de histidina no subsítio S₁ da falcipain. Comparando a estrutura química desta protease com a da papaína verificou-se uma grande semelhança estrutural no que diz respeito aos resíduos de aminoácidos presente nos sítios de ação (<http://www.cmpharm.ucsf.edu/cohen/>).

Os derivados de Chalcona apresentam uma suposta atividade inibitória de protease cisteína malarial (falcipain) (Li *et al*, 1995), portanto a nossa biomacromolécula alvo é uma enzima do *Plasmodium falciparum* que possui em seus subsítios ativos os resíduos de histidina, glutamato, serina, cisteína, asparagina e treonina. Visto que, existe possibilidade de pesquisa teórica que justifique os mecanismos envolvidos na complexação ligante-biomacromolécula, temos então na pesquisa QSAR uma clara evidência para esclarecimento e elucidação dos processos relacionados à atividade biológica de compostos, correlacionando então a estrutura química destas substâncias inibidoras com a estrutura química da molécula receptora.

Como visto na metodologia e nos resultados deste trabalho, identificamos nove modelos e ao final da pesquisa QSAR, propusemos um modelo multidimensional linear levando em consideração a proposta de Hansch-Fujita (Hansch & Fujita, 1964). Nossa pesquisa de QSAR clássico fundamentou-se na escolha de um modelo que tentasse justificar a atividade biológica dos derivados de Chalcona. Levando em consideração o fato de nossos compostos serem considerados supostos inibidores enzimáticos, preferimos então enveredar na escolha de um modelo que caracterizasse as variações de energia livre de Gibbs durante os processos de complexação fármaco-biomacromolécula receptora (modelo extratermodinâmico).

Na nossa proposta de modelo, aparecem quatro propriedades explicativas: Log P, Hydrogen bond donor, comprimento molecular e parâmetro estereoquímico de Verloop B4 para o substituinte A. Analisando os demais modelos selecionados

pelas técnicas estatísticas utilizadas, percebemos que outras propriedades moleculares também são importantes no estudo das Relações Estrutura-Atividade.

As propriedades moleculares, carga parcial atômica no átomo O₇ [q(O₇)], no átomo C₄ [q(C₄)], no átomo C₃ [q(C₃)], refratividade molar (MR), E_{LUMO} e Hydrogen bond acceptor são propriedades que apareceram nos melhores modelos. Estas propriedades ajudam na elucidação de possíveis interações ligante-biomacromolécula justificadas através de mecanismos de natureza eletrônica ou então por aspectos estereoquímicos.

As propriedades de natureza eletrônica estão relacionadas com a distribuição dos elétrons na molécula que por sua vez pode ser compreendida analisando as cargas parciais atômicas, a eletronegatividade dos elementos na molécula e os aspectos energéticos relacionados com os orbitais.

As propriedades de caráter estereoquímico estão relacionadas com as características topológicas da possível região ativa dos receptores e os ajustes geométricos envolvidos na interação droga-receptor.

Enquanto as propriedades de natureza eletrônica quântica justificam as interações do tipo dipolo-dipolo, ligações de hidrogênio, ligações iônicas e mecanismos de transferência de cargas, as propriedades de caráter estereoquímico estão relacionadas com as características dimensionais e os ajustes geométricos envolvidos no processo de interação que ocorrem na cavidade dos receptores.

Em nossa pesquisa QSAR devemos destacar como propriedade hidrofóbica o Log P determinado pelo método de Ghose e Crippen; como propriedade de caráter eletrônico quântico, destacaremos Hydrogen bond donor, Hydrogen bond acceptor, cargas parciais nos átomos C₄, C₃, O₇ e E_{LUMO}; em termos estereoquímicos daremos destaque ao comprimento molecular e ao parâmetro de Verloop B₄ (substituinte A); e como parâmetro de polarizabilidade notamos a presença marcante da refratividade molar determinada pelo método de Ghose e Crippen.

O sucesso dos trabalhos na área de QSAR depende de uma série de fatores envolvidos durante a pesquisa, seja nos critérios adotados na etapa de

determinação da geometria molecular para as estruturas químicas a serem analisadas, ou então, nas etapas de análise estatística que objetivarão selecionar variáveis independentes, selecionar os melhores modelos e validar os modelos.

Com relação à etapa minimização energética e otimização da geometria das estruturas envolvidas, adotamos como abordagem metodológica um processo de diminuição da energia das moléculas em etapas sucessivas, começando com a minimização energética MM2, depois otimizando a geometria com semi-empírico PM3 e finalmente utilizando um método simplex de minimização energética com opcional de constante dielétrica para meio aquoso. Tínhamos a finalidade em obter conformações para cada derivado estudado com a menor energia possível, ou seja, moléculas com conformação estável, independente se a molécula estava ou não em sua conformação mais estável, até mesmo porque não sabemos se a estrutura responsável pela atividade biológica é a correspondente ao mínimo global.

Pós-determinado a geometria molecular para todos os derivados de Chalcona, sabíamos que a conformação resultante obtida para cada derivado era estável, portanto tínhamos em mão uma conformação molecular correspondente a um mínimo local. Temos a informar que o procedimento foi adotado para todos os derivados, ou seja, todos os compostos passaram pela mesma abordagem metodológica e o mais importante a ser destacado reside no fato de que todas as estruturas obtidas, com exceção da Chalcona 20, resultaram em uma energia final (E4) menor do que o estabelecido como padrão energético de referência - manual do programa Molecular Modeling Pro (a energia deverá ser menor que 25 Kcal/mol).

Com a geometria de cada derivado já estabelecida, executamos cálculo de 69 descritores com finalidade em selecionar por procedimento estatístico RLM as propriedades que melhor explicam a atividade biológica dos compostos estudados. Ao passarmos para a etapa estatística, que é considerada como sendo a etapa “carro-chefe” da pesquisa QSAR, tomamos extremo cuidado a cada passo realizado, objetivando não excluir alguma variável que talvez fosse relevante na construção do modelo. Inicialmente tínhamos 69 propriedades calculadas, que

através de regressão linear simples e análise dos coeficientes de correlação, selecionamos 14 variáveis e que por intermédio de algoritmo genético e busca sistemática, conseguimos obter as dez melhores combinações de variáveis com 3, 4 e 5 descritores. Após executar RLM para as combinações acima, selecionamos as três melhores combinações de descritores obtendo nove modelos lineares, onde finalmente conseguimos propor uma estimativa de modelo linear segundo Hansch-Fujita (Hansch & Fujita, 1964).

Notamos que a adoção de um procedimento seletivo de variáveis em etapas, como comentado anteriormente na metodologia, forneceu-nos combinações muito semelhantes, seja por algoritmo genético ou por busca sistemática.

Podemos crer que, a sistemática adotada por nós na etapa estatística foi “afunilando” cada vez mais o conjunto de variáveis, utilizando como parâmetro o fato das propriedades moleculares não poderem ser altamente correlacionadas entre si e assim acabamos realizando uma “seleção natural” de variáveis independentes obtendo um conjunto de descritores com um baixo coeficiente de correlação entre si. Isto pode ser evidenciado analisando a tabela matriz de correlação (tabela 34) das 14 variáveis selecionadas à RLM e através da matriz de correlação das quatro variáveis descritoras (tabela 44) que participam do modelo proposto.

Uma vez estabelecida uma equação que correlaciona a estrutura química das substâncias pesquisadas com os respectivos valores de atividade biológica, é de extrema importância explorar o máximo de informação contida de modo direto ou indireto nas propriedades moleculares que foram selecionadas como parâmetros explicativos. As propriedades moleculares que participam da equação devem justificar de alguma forma a atividade biológica de uma série de compostos análogos. É claro que, quanto maior for o número de variáveis independentes na equação, mais informação nós teremos para explorar e tentar elucidar os mecanismos envolvidos nas interações fármaco-receptor. Verificando a nossa proposta de modelo (equação 77) e observando os sinais dos coeficientes de regressão de cada propriedade escolhida, podemos afirmar que:

- O parâmetro hidrofóbico **Log P** contribui positivamente para a atividade biológica, ou seja, quanto maior for o Log P da molécula, maior será a atividade do composto;
- O parâmetro eletrônico **H. B. donor** contribui negativamente para a atividade biológica, ou seja, quanto menor for H. B. donor da molécula, maior será a atividade do derivado;
- O parâmetro estereoquímico **Lx** contribui positivamente para a atividade biológica, ou seja, quanto maior for o Lx da molécula, será a atividade do composto, e
- O parâmetro estereoquímico **Verloop B4 (A)** contribui negativamente para a atividade biológica, ou seja, quanto menor for Verloop B4 (A), maior será a atividade do derivado.

Não devemos esquecer que a atividade biológica das substâncias químicas está intrinsecamente associada a mecanismos de interação entre o ligante e a biomacromolécula. A biomacromolécula é a estrutura alvo, ou seja, é o receptor que acomodará os fármacos em regiões específicas denominadas de centros ativos. A acomodação do fármaco nestas regiões envolve variações de energia livre de Gibbs que possivelmente poderá estar relacionada com o mecanismo de ação farmacológico da droga, elucidando assim os aspectos bioquímicos relacionados à estrutura dos fármacos (ligantes) e dos receptores (biomacromoléculas).

Partindo do princípio que já existe na literatura pesquisa a respeito de uma suposta inibição enzimática dos derivados de Chalcona, devemos então pensar nas possíveis interações ocorridas nos subsítios ativos das proteases malarias. Pensando deste modo, o derivado de Chalcona deverá apresentar um maior comprimento molecular possível, desde que o substituinte A possua uma largura menor para que a acomodação fármaco-receptor seja caracterizada com uma maior evidenciação em termos destas possíveis interações.

Em pesquisas futuras, será de grande interesse construirmos modelos de QSAR para os nossos derivados através de métodos quimiométricos multivariados: método PCR (*Principal Component Regression*) e PLS (*Partial Least Squares*). Estes métodos quimiométricos não são prejudicados pela

presença de correlação elevada entre variáveis. Poderíamos também, através destes métodos utilizar um pré-processamento do tipo autoescalamento onde a regressão é realizada fornecendo o mesmo peso para todas as variáveis. A análise de como cada variável está contribuindo no modelo poderá ser caracterizada pela construção do diagrama que mostrará a importância das variáveis independentes (coeficiente de correlação em função das variáveis).

Analizando o modelo proposto, o parâmetro hidrofóbico Log P varia de modo direto com a atividade biológica (-Log IC₅₀), ou seja, estruturas químicas dos derivados de Chalcona que apresentam elevado Log do coeficiente de partição octanol/água são mais lipofílicas e assim atravessam as membranas biológicas com maior facilidade. Substâncias com características estruturais tendendo a lipofilicidade terão suas moléculas se distribuindo e concentrando em regiões mais hidrofóbicas do organismo. A lipofilicidade das substâncias é uma propriedade responsável pelas interações hidrofóbicas entre o fármaco-receptor ou fármaco – meio e também pela solubilidade das substâncias em certos meios.

Analizando a estrutura química dos derivados de Chalcona, podemos afirmar que as regiões responsáveis pelas interações hidrofóbicas correspondem às porções que apresentam anel aromático substituídos com espécies químicas que apresentam baixa eletronegatividade para que não ocorra grande diferença de eletronegatividade entre elementos, e assim, resultando regiões na molécula com caráter hidrofóbico (figura 26). Percebe-se que regiões da molécula que apresentam elementos como oxigênio, flúor e nitrogênio apresentam elevada hidrofilicidade.

Figura 26 – Caráter lipofílico-hidrofílico da Chalcona 07, executado no programa MMP: azul escuro – elevada lipofilicidade, azul claro – baixa lipofilicidade, branco – lipofilicidade intermediária, vermelho - elevada hidrofiliicidade.



Nota: Analisando a figura, observamos a marcante hidrofiliicidade no substituinte A onde encontra - se o nitrogênio e na região da carbonila, enquanto regiões dos anéis aromáticos dos substituintes A e B apresentam elevada lipofilicidade.

Observando o modelo proposto, verificamos que a propriedade molecular **H. B. donor** relaciona de modo inverso com a atividade biológica, ou seja, quanto menor for a contribuição da molécula em fornecer ligação de hidrogênio, maior será a atividade biológica da espécie química. Hydrogen bond donor é uma propriedade de natureza eletrônica relacionada com a distribuição dos elétrons sobre a molécula.

Esta propriedade é determinada em função da somatória das cargas atômicas parciais maiores que + 0,03. Podemos então afirmar que, quanto maior

for a densidade eletrônica em certas regiões da molécula, maior será a contribuição em fornecimento de par eletrônico, ou seja, esta região da molécula terá caráter nucleófilo, sendo uma região sujeita a sofrer ataque de espécies que possuem afinidade por elétrons (ataque eletrófilo).

Esta propriedade está intimamente associada às cargas parciais atômicas, portanto regiões da molécula que apresentam elementos com elevada eletronegatividade, como é o caso do oxigênio presente no grupo carbonila e de alguns átomos substituídos nos anéis aromáticos (oxigênio, nitrogênio, cloro e flúor). Regiões da molécula que apresentam estes elementos possuem maior densidade eletrônica e um caráter nucleofílico pronunciado.

Não podemos esquecer que as Chalconas são estruturas com grupo funcional cetona α - β insaturado (enonas), com ocorrência de um evidente mecanismo de ressonância, contribuindo assim, para com a formação de três híbridos de ressonância (TLV). A deslocalização do par de elétrons da ligação π da carbonila em direção ao oxigênio resulta as formas canônicas dos derivados de Chalcona, onde o oxigênio da carbonila (O_7) (figuras 27 e 28) adquire carga parcial negativa acentuada e os carbonos C_1 e C_4 (carbonila) (figura 21) adquirem carga parcial positiva.

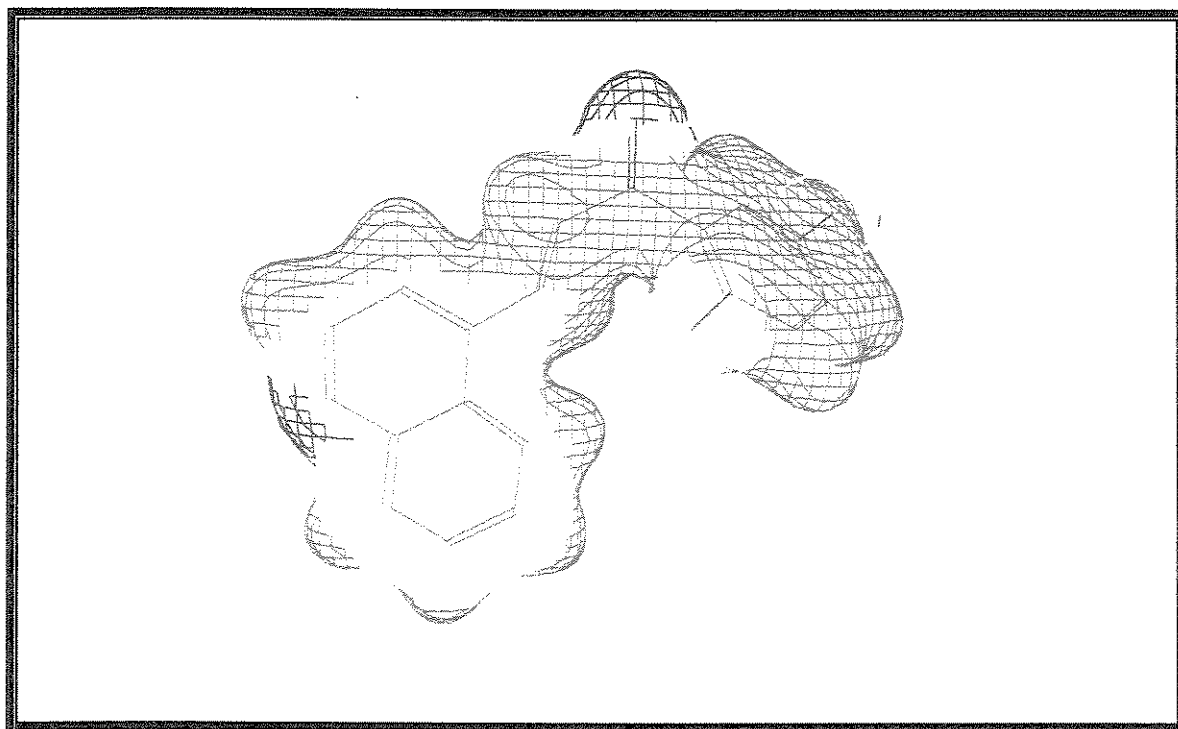
Deste modo, o efeito ressonante ocorrido em enonas justifica a densidade eletrônica acentuada no oxigênio da carbonila e as cargas positivas nos carbonos C_1 e C_4 . É fácil entender que as enonas estão sujeitas à ocorrência de ataques de agentes nucleófilos nas posições C_1 e C_4 , onde o carbono C_1 está sujeito ao ataque de nucleófilo mole e o carbono C_4 está sujeito a ataque de nucleófilo duro.

Outro detalhe importante que deve ser mencionado reside no fato do meio interno do vacúolo digestivo do *Plasmodium falciparum* apresentar caráter ligeiramente ácido (pH = 5,0-5,5). Enonas em meio ácido sofrem protonação no oxigênio da carbonila intensificando o caráter eletrófilo nos carbonos C_1 e C_4 , favorecendo assim a adição nucleofílica nestes carbonos (Costa *et al*, 2003). O importante é saber se o ataque do nucleófilo ocorrerá em carbono C_1 ou carbono C_4 . Ataques nucleofílicos em carbono C_1 ocorrem por intermédio de interações orbitais entre orbital LUMO do carbono C_1 com orbital HOMO do agente

nucleófilo, pelo fato do carbono C₁ apresentar elevado coeficiente LUMO. A adição nucleofílica em carbono C₁ é denominada de adição de Michael e ocorre com a formação de ligação química covalente de caráter irreversível (Morrison & Boyd, 1996).

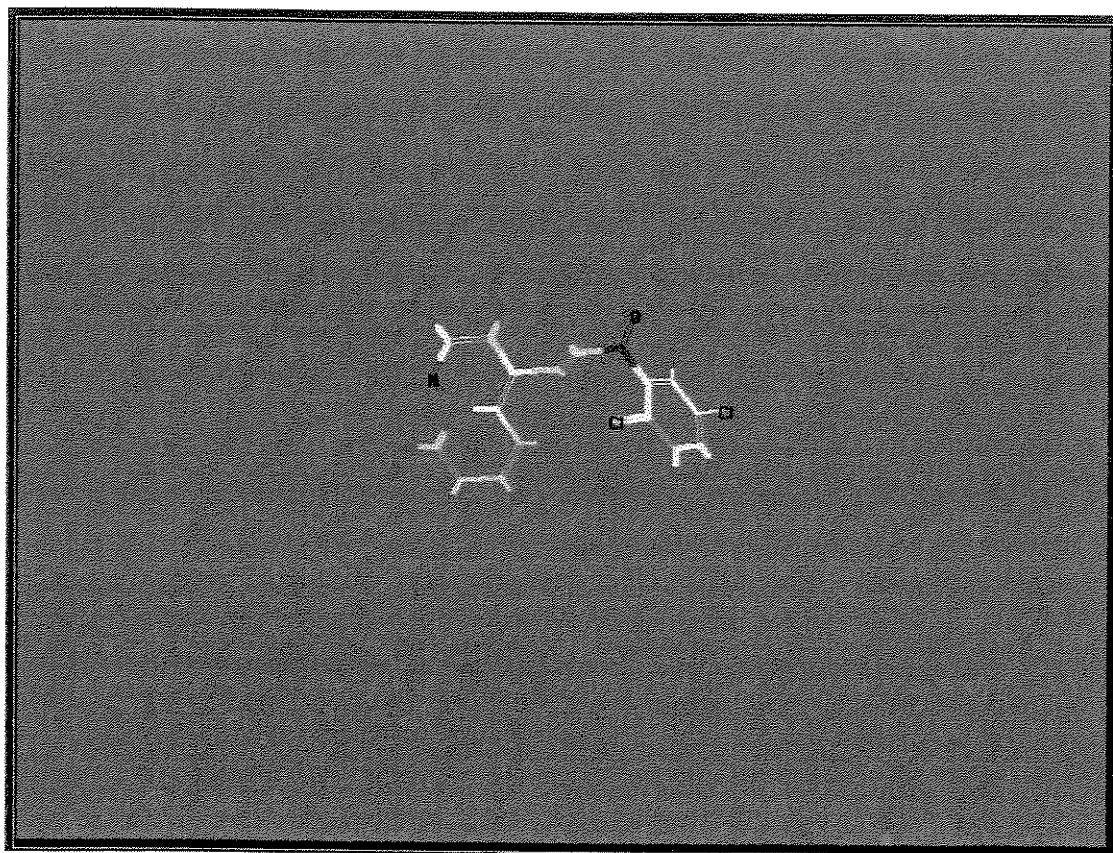
Ataques nucleófilos em C₄ ocorrem devido à intensa atração eletrostática entre carbono C₄ e uma porção negativa do agente nucleófilo. Assim, a adição aos carbonos é realizada por mecanismo de regioseletividade, onde o carbono C₁ está sujeito ao ataque de nucleófilo mole, pois a interação orbitalar predomina sob a interação eletrostática, enquanto o carbono C₄ estará sujeito ao ataque de nucleófilo duro, pois a interação eletrostática neste caso predomina (Costa *et al*, 2003).

Figura 27 – Densidade eletrônica superficial obtida via cargas parciais no programa ChemSite Pro pelo método Combinação Linear de Orbitais Atômicos (LCAO) para a Chalcona 07: verde – levemente positivo, amarelo – levemente negativo, vermelho – intenso negativo.



Nota: Analisando a figura, observamos a elevada densidade eletrônica no oxigênio da carbonila (O₇) e no nitrogênio do substituinte A, ambos em vermelho.

Figura 28 – Cargas atômicas parciais obtidas pelo programa MMP para a Chalcona 07: azul escuro – muito positivo, verde – levemente positivo, branco – neutro, amarelo – levemente negativo, vermelho – muito negativo.



Nota: Observando a figura, notamos que o carbono C₄ da carbonila em azul apresenta elevada carga parcial positiva, estando sujeito à adição do tipo nucleofílica via interação eletrostática.

A presença de grupos elétron-atraentes em substituinte A favorecerá a adição nucleofílica em carbono C₁, ou seja, a adição de Michael será favorecida em função da ocorrência de uma disponível cadeia lateral de aminoácidos com caráter nucleófilo presente na região ativa da protease cisteína.

Sabe-se que os subsítios ativos das proteases cisteínicas do *Plasmodium falciparum* apresentam o aminoácido histidina e que um provável mecanismo de interação entre os derivados de Chalcona e este aminoácido localizado em subsítio ativo poderá estar ocorrendo durante a complexação droga-biomacromolécula (Joachimik *et al*, 2001).

Devemos lembrar que a cloroquina e alguns derivados de Chalcona apresentam na região do substituinte A, o agrupamento quinolínico em suas estruturas químicas e quando foram testados contra cepas resistentes do *Plasmodium falciparum* apresentaram uma atividade considerável (Li *et al*, 1995). O agrupamento quinolínico é uma estrutura heterocíclica que contém nitrogênio e, portanto, em meio ácido (vacúolo digestivo do parasita) o nitrogênio pode também sofrer protonação, facilitando assim interações com os resíduos de aminoácidos presentes na região ativa das proteases.

Observando a densidade eletrônica (figura 27) de alguns derivados de Chalcona, notamos que tanto o oxigênio da carbonila quanto o nitrogênio em grupos quinolínicos apresentam elevada densidade eletrônica, facilitando as interações com agentes eletrofílicos dos centros ativos das proteases.

Tivemos a curiosidade de observar a matriz de correlação entre as propriedades moleculares eletrônicas: H. B. donor, H. B. acceptor, E_{HOMO} , E_{LUMO} , $q(C_1)$, $q(C_3)$, $q(C_4)$ e $q(O_7)$ e $d(C_2-C_4)$. Percebemos que a variável H. B. donor possui baixo coeficiente de correlação para com todas as propriedades mencionadas (tabela 46), com exceção à carga parcial no carbono C_3 , que a carga parcial do oxigênio da carbonila não está correlacionada com H. B. donor e possui boa correlação com a carga parcial do carbono C_4 e que a carga parcial do carbono C_3 possui elevada correlação com a carga parcial do carbono C_1 .

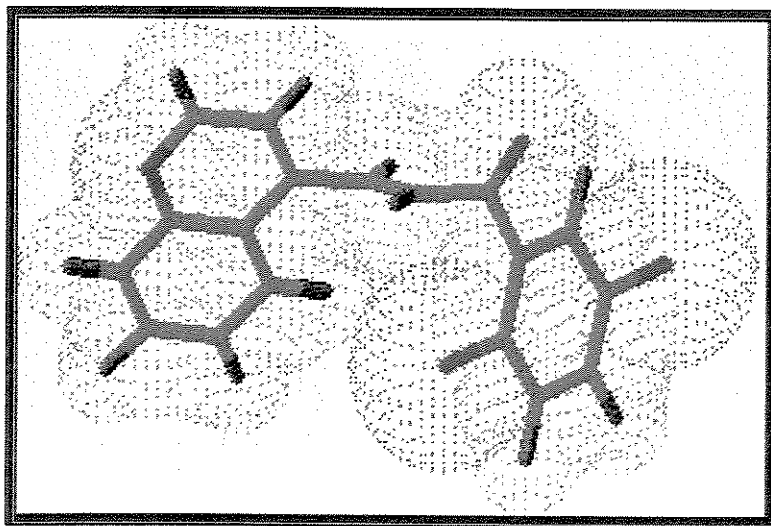
Tabela 46 - Matriz de correlação de alguns parâmetros eletrônicos; Critério adotado: limite do coeficiente de correlação = 0,4.

	E_{HOMO}	E_{LUMO}	H. B. acceptor	H. B. donor	$q(C_1)$	$q(C_3)$	$q(C_4)$	$q(O_7)$	$d(C_2-C_4)$
E_{HOMO}	1	0,486	0,399	0,006	0,029	0,012	0,095	0,093	0,21
E_{LUMO}		1	0,498	0,018	0,088	0,001	0,064	0,048	0,341
H. B. acceptor			1	0,015	0,072	0,043	0,163	0,216	0,141
H. B. donor				1	0,204	0,411	0,241	0,058	0,059
$q(C_1)$					1	0,572	0,013	0,142	0,005
$q(C_3)$						1	0,052	0,027	0,072
$q(C_4)$							1	0,469	0,016
$q(O_7)$								1	0,004
$d(C_2-C_4)$									1

O comprimento molecular (L_x) é uma propriedade estereoquímica, melhor dizendo, propriedade relacionada à dimensão molecular. Analisando a equação proposta como modelo linear, o comprimento molecular é uma grandeza que varia diretamente com a atividade biológica, concluindo então que quanto maior for a extensão da molécula em torno do eixo X (L_x), maior será a sua atividade biológica, determinando uma maior ação inibitória em proteases malarias. Isto nos leva a crer que o provável subsítio ativo do receptor ou a região de reconhecimento molecular apresenta um sulco com grande espaçamento para que a estrutura química seja complementar, e por consequência, caracterizando um “encaixe” mais adequado e eficaz em termos de interações químicas e variações de energia livre.

Deste modo, a molécula ativa deverá possuir uma maior linearidade, apresentando maior superfície de contato para com os aminoácidos da região ativa. Isto pode também ser explicado analisando as características estruturais da porção α - β insaturada dos derivados de Chalcona, que apresenta uma conformação mais estendida (figura 29), possuindo assim uma maior liberdade de movimento no sulco da região ativa, o que justifica o maior ajuste com a protease cisteínica do parasita. Como todos os derivados apresentam a porção α - β insaturada, o que deverá diferencia-los estruturalmente será o formato espacial dos substituintes A e B.

Figura 29: Conformação da Chalcona 07.



Analisando a estrutura química acima em sua porção α - β insaturada, nota-se uma disposição plana justificada pela hibridação dos carbonos sp^2 , indicando que a conjugação da porção alifática fornece uma maior rigidez para a estrutura química, que finalmente resulta uma conformação mais extendida. É claro que a extensão dos substituintes A e B também influencia no comprimento global da molécula. Não podemos desconsiderar o tipo de substituinte, pois o comprimento molecular depende também da geometria molecular, que por sua vez depende dos elementos químicos envolvidos.

Verloop B4 [$\geq$ B4] é uma propriedade molecular topológica relacionada com os substituintes. Analisando o modelo proposto, Verloop B4 do substituinte A varia de modo inverso com a atividade biológica, portanto, quanto menor o valor de $\geq$ B4(A) da molécula, maior será a atividade biológica do composto.

Como Verloop B4 representa o maior raio do substituinte perpendicular ao eixo L1, então o maior raio do substituinte A para um derivado de Chalcona deverá ser tanto quanto menor para que a atividade biológica do composto seja maior. Sendo L1 a extensão do substituinte ao longo do eixo a que este está conectado, então B4 representa a maior “largura” do substituinte. Este dado nos fornece uma importante característica dimensional da porção mais larga do substituinte A, devendo esta ser menor possível para que o substituinte possa acomodar-se na cavidade do subsítio ativo. O subsítio ativo da protease cisteínica que acomoda o substituinte A da molécula provavelmente não possui um grande diâmetro.

Observando as estruturas químicas dos derivados de Chalcona e correlacionando-as com o modelo proposto e com a atividade biológica observada, é possível descrever a influência das características estruturais para as cetonas α - β insaturadas na atividade biológica destes compostos:

- Agrupamento quinolínico como substituinte A favorece a atividade biológica dos compostos;
- Agrupamento metóxi substituído no anel B favorece a atividade biológica dos compostos;
- Substituição cloro próxima ao nitrogênio do agrupamento quinolínico em substituinte A, diminui a atividade biológica dos compostos;

- Substituinte clorofenil em A favorece a atividade biológica dos compostos.

Aparentemente, podemos especular que as propriedades encontradas no modelo proposto apresentam importância relevante e estão correlacionadas com a atividade biológica. No entanto, não podemos desconsiderar as propriedades moleculares, cargas parciais no átomo de oxigênio da carbonila [$q(O_7)$], no carbono C_3 [$q(C_3)$], a refratividade molar, H. B. acceptor e E_{LUMO} , que certamente contribuirão na pesquisa de compostos análogos com uma atividade biológica melhorada.

7.0 Conclusões

Construir um modelo matemático que seja capaz de explicar um fenômeno observado e que também proporcione previsões a respeito do possível mecanismo de ação de substâncias, não é uma tarefa nada fácil. A ação biológica de compostos é um processo de tamanha complexidade, que envolve diversas propriedades relacionadas com a estrutura química dos compostos e da biomacromolécula. Em nossa pesquisa o fenômeno biológico observado refere-se à inibição *in vitro* da parasitemia provocada por cepas cloroquina resistentes de *Plasmodium falciparum* diante uma série de compostos análogos derivados de Chalcona. A série de derivados sintetizados se mostraram ativos, diante comparação com a atividade biológica e o índice de resistência da cloroquina.

Partindo do princípio que durante as interações fármaco-receptor ocorrem manifestações de variação na energia livre de Gibbs, preferimos então escolher como estimativa de modelo uma proposta clássica de QSAR. Como proposta de modelo escolhemos o modelo extratermodinâmico de Hansch- Fujita, que baseia-se numa estimativa de modelo multidimensional linear que correlaciona a atividade biológica dos compostos da série com propriedades moleculares e estruturais. Escolher parâmetros para participar da equação é algo que envolve certa astúcia por parte do pesquisador, visto que as propriedades além de serem selecionadas por técnicas estatísticas devem ser capazes de estimar um provável mecanismo de ação dos compostos estudados.

Não podemos descartar, que o nosso alvo receptor são proteases cisteínicas, enzimas necessárias ao metabolismo do parasita, portanto uma boa estimativa de modelo facilitaria a elucidação de prováveis regiões farmacofóricas dos compostos pesquisados e das características estereoquímicas dos subsítios ativos da biomacromolécula receptora.

Para que uma equação seja promovida a modelo quantitativo é preciso ao final da pesquisa QSAR, selecionar propriedades moleculares explicativas (independentes) através de métodos estatísticos. Além disso, o modelo deverá ser validado estatisticamente. A validação estatística é um procedimento necessário para avaliar a medida da qualidade do modelo e deve ser concretizada

analisando o grau de ajuste, o grau de significância estatística e o grau de previsibilidade do modelo.

Após executarmos, o cálculo de vasta gama de parâmetros (69 variáveis), selecionamos por intermédio de técnicas estatísticas quatorze propriedades de importância relevante. Notamos que as combinações selecionadas por algoritmo genético e por busca sistemática são equivalentes.

Assim, mesmo que tenhamos escolhido como proposta de modelo uma equação que apresenta quatro variáveis descritoras [Log P; H. B. donor; Lx; $\Rightarrow$ B4 (A)], não podemos em hipótese alguma desconsiderar as demais propriedades que não participam da equação modelo.

O modelo proposto possui um bom nível de ajuste, pois apresenta elevado coeficiente de correlação ($r = 0,953$), baixo desvio padrão ($s = 1,147$), sendo também capaz de explicar 90,8% da variabilidade dos dados observados de atividade ($r^2 = 0,908$). Com a concretização da análise de variância (ANOVA) da regressão, constatamos que a estimativa de variância ($s^2 = 0,02172$) é baixa. O modelo apresenta significância estatística, visto que possui elevado valor Fischer ($F = 34,459$) e um baixíssimo p-valor ($p < 0,0001$).

É de extrema importância considerarmos que o modelo também possui elevado grau de previsibilidade. Isto é percebido pelo valor considerável do coeficiente de previsão ($Q^2 = 0,852$) e do baixo desvio padrão da previsão (SPRESS = 0,187).

Um dado de extrema importância e que não deve ser esquecido reside na capacidade dos derivados pesquisados estarem sujeitos aos mecanismos de ressonância, pois as Chalconas são cetonas α - β insaturadas, apresentando então conjugação em sua estrutura química, e estando sujeitas a serem boas aceptoras de Michael, ou seja, participarem como nucleófilos em processos de adição de Michael, formando um aduto de Michael. Esta nucleofilicidade pode ser explicada pelo efeito ressonante que ocorre com as enonas, justificando assim a carga parcial positiva do carbono C₁ [$q(C_1)$], e do carbono C₄ [$q(C_4)$], a elevada densidade eletrônica no oxigênio da carbonila e a sua respectiva carga parcial negativa [$q(O_7)$].

Notamos também que a refratividade molar aparece na maioria dos modelos, sendo considerada também uma variável importante na previsibilidade do mecanismo de ação, assim como o comprimento molecular e o parâmetro estérico de Verloop B4 para o substituinte A. Concluimos que a molécula deverá apresentar uma maior extensão em termos de comprimento e uma pequena largura para o substituinte A. Deste modo, o substituinte A poderá ter um grande comprimento, mas devendo apresentar um menor volume possível. Provavelmente, o subsítio do receptor que interage com o substituinte A dos derivados deverá apresentar um menor espaçamento tridimensional no local ativo, onde substituintes A, mais largos, serão impedidos de interagirem por uma questão estereoquímica (impedimento estérico espacial).

Presenciamos de forma marcante a participação da carga parcial no átomo de oxigênio da carbonila na maioria dos modelos selecionados. Acreditamos que o fato da estrutura possuir insaturação próxima a este átomo, contribua de certa forma com uma densidade eletrônica elevada sobre o átomo de oxigênio. Não devemos descartar a possibilidade do carbono C₄, que está próximo ao oxigênio da carbonila, servir como aceptor de nucleófilo, visto que possui carga parcial positiva com maior intensidade do que o Carbono C₁, estando sujeito assim a interações eletrostáticas com o receptor.

08 Referências Bibliográficas

- Abbas, Abul K.; Lichtman, Andrew H.; Pober, Jordan S. *"Imunologia celular e molecular"*, Tradução: Fernando Gomes do Nascimento e Aurelizia Maria Lemos Xavier. Rio de Janeiro: Livraria e Editora REVINTER Ltda, 1995.
- ACCVIP - Australian Computational Chemistry via the Internet Project. Computational Chemistry module index: Molecular Mechanics Optimisation. Disponível em <http://www.chem.swin.edu.au/modules/mod6/mechoptim.html>. Acesso em 02/05/2003.
- Allinger, N. L. "Calculation of molecular structure and energy by force-field methods". *Adv. Phys. Org. Chem.*, v. 13, p. 1-82, 1976.
- Allinger, N. L. "Conformational Analysis 130. MM2. A Hydrocarbon Force Field Utilizing V1 and V2 Torsional Terms." *J. Am. Chem. Soc.*, v. 99, p.8127-8134, 1977.
- Allinger, N. L. MM2/MMP2 with the 1977 force field submitted to QCPE (Quantum Chemistry Program Exchange, QCPE Program No. 395). 1980.
- Allinger, N. L.; Kok, R. A.; Imam, M. R. "Hydrogen Bonding in MM2." *J. Comp. Chem.*, v. 9, n. 06, p. 591-595, 1988.
- Amaral, Antonia T. & Montanari, Carlos Alberto. "Química Medicinal: 25 anos de planejamento racional de fármacos". *Quím. Nova*, v. 25, n. 1, p. 39-44, 2002.
- Analyse-it Software, Ltd. Versão 1.68, 1997-2000.
- Arango, Héctor Gustavo. *"Bioestatística teórica e computacional"*. Rio de Janeiro: Guanabara Koogan, 2001.
- AREAS, E. P. G. *et al.* "Molecular dynamics simulations at a membrane/water interface". *J. Phys. Chem.*, v. 99, n. 40, p. 14885-14892, 1995.
- AREAS, E. P. G. *et al.* "Molecular dynamics at a cytoplasm/membrane interface of a signal sequence from an *E.coli* maltoporin". *Braz. J. Med. Biol. Res.*, v. 27, n. 2, p. 527-533, 1994.
- Atkins, Peter W. *"Físico-química"*, Tradução: Horacio Macedo. 6ª ed, Rio de Janeiro: LTC, 1999. 2v.
- Atkins, Peter & Jones, Loretta. *"Princípios de Química: questionando a vida moderna e o meio ambiente"*. Porto Alegre: Bookman, 2001.

- Banerjee, Ritu et al. "Four plasmepsins are active in the *Plasmodium falciparum* food vacuole, including a protease with an active-site histidine". *PNAS*, v. 99, n. 2, p. 990-995, 2002.
- Barreiro, Eliezer J. & Fraga, Carlos Alberto Manssour. "Química medicinal: as bases moleculares da ação dos fármacos". Porto Alegre: ARTMED, 2001.
- Barret, A. J. "An introduction to the proteinases". "In *Proteinase Inhibitors*", ed. A. J. Barret & G. Salvesen, p. 3-22. Amsterdam: Elsevier Science, 1986.
- Barton, Allan F. M. "Handbook of solubility Parameters and Other Parameters". CRC Press, 1983.
- Bernard, F. & Schrevel, J. "Purification of a *Plasmodium berghei* neutral endopeptidase and its localization in merozoite". *Mol. Biochem. Parasitol.*, v. 26, p. 167-174, 1987.
- Bondi, A. "Van der Waals volumes and radii". *J Phys. Chem.*, v.68, n. 03, p.441-451, 1964.
- Braun-Breton, C.; Rosenberry, T. L.; Pereira da Silva, L. "Induction of the proteolytic activity of a membrane protein in *Plasmodium falciparum* by phosphatidyl inositol-specific phospholipase C". *Nature*, v. 332, p. 457-459, 1988.
- Braun-Breton, C. et al. "*Plasmodium chabaudi* p68 serine protease activity required for merozoite entry into mouse erythrocytes". *Proc. Natl. Acad. Sci. USA* In press, 1992.
- Bunge, Annik Vivier. "Introdução à Química Quântica". São Paulo: Ed. Edgard Blücher LTDA, 1977.
- Burket, U. & Allinger, N. L. "Molecular Mechanics"; American Chemical Society Monograph, Washington, cap. 1 e 2, 1982.
- Carvalho, Ivone; Pupo, Mônica T.; Borges, Áurea D. I.; Bernardes, Lillian S. C. "Introdução a modelagem de fármacos no curso experimental de química farmacêutica". *Quím. Nova*, v. 26, n. 3, p. 428-438, 2003.
- CDC – U. S. DEPARTMENT OF HEALTH AND HUMAN SERVICES. Centers of Disease Control and Prevention. Disponível em: <http://www.cdc.gov>. Acesso em 19/09/2002.
- ChemSW® Inc.; Molecular Modeling Pro™4.0, Computacional Chemistry Program; Fairfield, 2001.

- ChemSW® Inc.; ChemSite: Interactive 3D Molecular Modeling 5.0; Fairfield, 2001
- Chen, Ming *et al.* "Inhibition of fumarate reductase en *Leishmania major* and *L. donovani* by Chalcones". *Antimicrob. Agents Chemother.*, v. 45, n. 07, p. 2023-2029, 2001.
- Choi, C. Y. H. *et al.* "Spectroscopic characterization of the heme-binding sites in *Plasmodium falciparum* histidine-rich protein 2". *Biochemistry*, v. 38, p. 16916-16924, 1999.
- Cimerman, Benjamin & Cimerman, Sérgio. "*Parasitologia Humana e seus fundamentos gerais*". São Paulo: Ed. Atheneu, 1999.
- Clark, T. "*A Handbook Computacional Chemistry – A Pratica Guide to Chemical Structure and Energy Calculations*". EUA: John Wiley & Sons, 1985.
- Cohen, N. C. "*Guidebook on Molecular Modeling in Drug Design*". San Diego, California: Academic Press, 1996.
- CONSTITUIÇÃO FEDERAL DO BRASIL DE 1988. Promulgada em 05 de outubro de 1988. Disponível em:
<http://www.senado.gov.br/bdtextual/const88/Con1988br.pdf> acesso em 15/10/2003.
- Costa, Maria Cristina Andreazza. "*Relações entre estrutura química de neolignanas e a sua atividade anti-leishmaniose, envolvendo cálculo de solvatação*". 1998. 92 f. Tese (Doutorado em Química) – Instituto de Química, Departamento de Físico-química, Unicamp, Campinas –SP.
- Costa, Paulo; Pilli, Ronaldo; Pinheiro, Sérgio; Vasconcellos, Mário. "*Substâncias Carboniladas e derivados*". Porto Alegre: editora BOOKMAN, 2003.
- Debrant, A. & Delplace, P. "Leupeptin alters the proteolytic processing of P126. The major parasitophorous vacuole antigen of *Plasmodium falciparum*". *Mol. Biochem. Parasitol.*, v. 33, p. 151-158, 1989.
- Egan, T. J. *et al.* "Characterisation of synthetic β -haematin and effects of the antimalarial drugs quinidine, halofantrine, desbutylhalofantrine and mefloquine on its formation". *J. Inorg. Biochem.*, v. 73, p. 101-107, 1999.
- Eggleston, Kathleen K.; Duffin, Kevin L.; Goldberg, Daniel E. "Identification and characterization of Falcilysin, a metallopeptidase involved in hemoglobin catabolism within the malaria parasite *Plasmodium falciparum*". *J. Biol. Chem.*, v. 274, n. 45, p. 32411-32417, 1999.

- Espírito Santo, Larissa Lima do. "Aplicação de métodos semiempíricos ao estudo da estrutura eletrônica de compostos bioativos". 2001. 56 f. Tese (Doutorado em Física) – Instituto de Física Gleb Wataghin, Unicamp, Campinas – SP.
- Ferreira, Marcelo Urbano. "Diversidade Antigênica na Proteína Principal da Superfície de Merozoítos (MSP-1) de *Plasmodium falciparum*: Implicações para o Desenvolvimento de Vacinas Antimaláricas". *Revista Médica Virtual*, 1999.
- Ferreira, Márcia M. C. "Multivariate QSAR". *J. Braz. Chem. Soc.*, v. 13, n. 06, p. 742-753, 2002.
- Ferreira, Márcia Miguel Castro; Montanari, Carlos Alberto; Gaudio, Anderson Coser. "Seleção de variáveis em QSAR". *Quím. Nova*, v. 25, n. 3, p. 439-448, 2002.
- Florence, Alexander Taylor & Attwood, D. "Princípios físico-químicos em farmácia", Tradutores: Zuleika Rothschild (org.), Adolfo Max Rothschild *et al.* São Paulo: Editora da Universidade de São Paulo, 2003.
- Fonseca, Jairo Simon da; Martins, Gilberto de Andrade; Toledo, Geraldo Luciano. "Estatística aplicada". 2ª ed, São Paulo: Atlas, 1985.
- Fukui, K.; Yonezawa, T.; Nagata, C.; Shingu, H. "A molecular orbital theory of reactivity in aromatic hydrocarbons". *J. Chem. Phys.*, v.22, n. 08, p. 1433-1442, 1954.
- FUNASA - Ministério da Saúde: Governo do Brasil. Fundação Nacional de Saúde. Disponível em: <http://www.funasa.gov.br>. Acesso em 24 agosto 2003.
- FUNASA - Ministério da Saúde: Governo do Brasil. Fundação Nacional de Saúde: Manual de terapêutica da malária. Disponível em: http://www.funasa.gov.br/pub/pdfs/manu_therapeutica_malaria.pdf Acesso em 15/10/2003.
- Gaudio, Anderson Coser & Zandonade, Eliana. "Proposição, Validação e análise dos Modelos que Correlacionam Estrutura Química e Atividade Biológica". *Quím. Nova*, v. 24, n. 5, p. 658-671, 2001.
- Gerhauser, Clarissa *et al.* "Cancer Chemopreventive Activity of Xanthohumol, a Natural Product Derived from Hop1". *Molecular Cancer Therapeutics*, v. 01, p. 959-969, 2002.
- Ginsburg, H.; Ward, S. A.; Bray, P. G. "An integrated model of chloroquine action". *Parasitology Today*, v. 15, p. 357, 1999.

- Grellier, P. *et al.* "Purification and identification of a neutral endopeptidase in *Plasmodium falciparum* schizonts and merozoites". *Parasitol. Res.*, v. 75, p. 455-460, 1989.
- Hansch, C. & Fujita, T. " ρ - σ - π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure". *J. Am. Chem. Soc.*, v. 86, p. 1616-1626, 1964.
- Hansch, C. & Leo, A. "Substituent constants for correlation analysis in chemistry and biology". *Wiley and Sons.*, 1979.
- Hanspal, Manjit; Dua, Meenakshi, Takakuwa, Yuichi; chishti, Athar H.; Mizuno, Akiko. "*Plasmodium falciparum* cysteine protease falcipain-2 cleaves erythrocyte membrane skeletal proteins at late stages of parasite development". *Blood*, v. 100, n. 3, p. 1048-1054, 2002.
- Harvey, Richard A.; Champe, Pamela C.; Mycek, Mary J. "*Farmacologia ilustrada*", Tradução: Ricardo M. Oliveira-Filho et al. 2ª ed, Porto Alegre: Artmed Editora, 1998.
- Hehre, W. J.; Yu, J.; Klunzinger, P. E.; Lou, L. "*A brief guide to molecular mechanics and quantum chemical calculations*". EUA: Wavefunction, Inc., 1998.
- Hempelmann, E. & Egan, T. J. "Pigment biocrystallization in *Plasmodium falciparum*". *Tr. Parasitology*, v. 18, p. 11, 2002.
- Herencia, Felipe; Ferrandiz, M. Luiza; Ubeda, Amalia; Dominguez, José N.; Charris, Jaime E.; Lobo, Gricela M.; Alcaraz, M. José. "Synthesis and anti-inflammatory activity of chalcone derivatives". *Elsevier Science*, p. 1169-1174, 1998.
- Horvath, Ari L. "*Molecular design: chemical structure generation from the properties of pure organic compounds*". Amsterdam; London; New York; Tokyo: Elsevier, 1992.
- Humphreys, Michelle J. *et al.* "The aspartic proteinase from the rodent parasite *Plasmodium berghei* as a potential model for plasmepsins from the human malaria parasite, *Plasmodium falciparum*". *FEBS Letters*, v. 463, p. 43-48, 1999.
- Hypercube®, Inc.; Hyperchem™ 6.0, Molecular Modeling System, 2000.
- INTERFARMA - Associação da indústria farmacêutica de pesquisa.
<http://www.interfarma.org.br> . Acesso em 14/08/2003.
- Joachimiak, Macin P. *et al.* "The Impact of Whole Genome Sequence Data on Drug Discovery – A Malaria Case Study". *Mol. Med.*, v. 07, n. 10, p. 698-710, 2001.

- Joback, K. G. & Reid, R. C. "Estimation of pure-component properties from group contributions". *Chem. Eng. Commun.*, v. 57, p. 233-243, 1987.
- Kalant, Harold & Roschlau, Walter H. E. "*Princípios de Farmacologia Médica*". 5ª ed. Rio de Janeiro: ed. Guanabara Koogan, 1991.
- Katz, Michael; Despommier, Dickson D.; Gwadz, Robert W. "*Parasitic diseases*". 2ª ed, Nova York: Springer-Verlag New York Inc, 1988.
- Kayser, Oliver & Kiderlen, Albrecht F. "*In vitro* Leishmanicidal activity of naturally occurring chalcones". *InterScience*, v. 15, p. 148-152, 2001.
- Kayser, Oliver; Kiderlen, Albrecht F.; Croft, Simon L. "Natural Products as potential antiparasitic drugs". Disponível em: <http://userpage.fu-berlin.de/~Kayser/antiparasiticsfromnature.pdf> . Acesso em 02/03/2003.
- Klemba, M. & Goldberg, D. E. "Biological roles of proteases in parasitic protozoa". *Annu. Rev. Biochem.* V 71, p. 275-305, 2002.
- Klopman, G.; Wang, S.; Balthasar, D. M. "Estimation of aqueous Solubility of Organic Molecules by the Group Contribution Approach. Application to the Study of Biodegradation". *J. Chem. Inf. Comput. Sci.*, v. 32, p. 474-482, 1992.
- Korolkovas, Andrejus & Burckhalter, Joseph H. "*Química farmacêutica*". Rio de Janeiro: Guanabara Koogan, 1988.
- Krogsgaard-Larsen, Povl; Liljefors, Tommy; Madsen, Ulf. "*A Textbook of Drug Design and Development*". 2ª ed. The Netherlands: Harwood Academic Publishers, 1996.
- Kubiny, Hugo. "*QSAR: Hansch analysis and related approaches*". Weinheim; New York; Basel; Cambridge; Tokyo: VHC, 1993.
- Kubiny, Hugo. "Strategies and Recent Technologies in the Drug Discovery". *Pharmazie*, v.50, p. 647 – 662, 1995.
- Kubiny, Hugo. "Lecture of the drug design course". Disponível em: <http://home.t-online.de/home/kubiny/lectures.html>. Acesso em 20/02/2003.
- Leitão, Andrei; Montanari, Carlos A.; Donnici, Glaudio I. "Sobre o uso de métodos quimiométricos em química combinatória". *Quím. Nova*, v. 23, n. 02, p. 178-184, 2000.
- Leventhal, Ruth & Cheadle, Russel. "*Parasitologia médica: texto e atlas*". 4ª ed. São Paulo: editorial Premier, 1997.
- Levine, Ira N. "*Quantum Chemistry*". 4ª ed, New Jersey: Prentice Hall, Inc., 1991.

- Li, R. *et al.* "In Vitro antimalarial activity of Chalcones and their derivatives". *J. Med. Chem.*, v. 38, p. 5031-5037, 1995.
- Liu, Mei; Wilairat, Prapon; Go, Mei-Lin. "Antimalarial alkoxylated and hydroxylated chalcones: structure-activity relationship analysis". *J. Med. Chem.*, v. 44, n. 25, p. 4443-4452, 2001.
- Lorenzi, Therezinha F. "*Manual de Hematologia*". 2ª ed, Rio de Janeiro: MEDSI Editora Médica e Científica Ltda, 1999.
- Lüllmann, Heinz & Mohr, Klaus. "*Farmacologia – Texto e Atlas*". 4ª ed, Porto Alegre: Artmed, 2004.
- Lunardi, Fabiane *et al.* "Trypanocidal and Leishmanicidal properties of substitution-containing chalcones". *Antimicrob. Agents Chemother.*, v. 47, n. 04, p. 1449-1451, 2003.
- Lyman, W. J.; Reed, W. F.; Rosenthal, D. H.; Hill, McGraw. "*Handbook of chemical property estimation*". 1982.
- Mann, S. "Crystallochemical strategies in biomineralization". In "*Biomineralization, Chemical and Biochemical Perspectives*" (Mann, S *et al.*, eds), p. 35-62, Weinheim: VHC, 1989.
- Marzzoco, Anita & Torres, Bayardo Baptista. "*Bioquímica Básica*". 2ª ed, Rio de Janeiro: Guanabara Koogan, 1999.
- Mayer, R. *et al.* "Peptide derivatives specific for a *Plasmodium falciparum* proteinase inhibit the human erythrocyte invasion by merozoites". *J. Med. Chem.*, v. 34, p. 3029-3035, 1991.
- McGowan. "Tenside Surfactants". 27:229-230, 1990.
- Mckerrow, James H; Sun, E.; Rosenthal, Philip J.; Bouvier, Jacques. "The proteases and Pathogenicity of Parasitic Protozoa". *Ann. Rev. Microb.*, v. 47, p. 821-853, 1993.
- Mcquarrie, Donald A. & Simon, John D. "*Physical chemistry: a molecular approach*". Sausalito, California: University Science Books, 1997.
- Meyers, Drew. "*Surfaces, interfaces and colloids: Principle and applications*". 2ª ed, Wiley-VCH, Myers, 1999, cap. 11.
- Morgon, Nelson Henrique. "Computação em química teórica: informações técnicas". *Quím. Nova*, v. 24, n. 5, p. 676-682, 2001.

- Morrison, R. & Boyd, R. "Química Orgânica". 13ª ed, Lisboa: Fundação Calouste Gulbenkian, 1996.
- Mundin, Kleber C. "Modelagem Molecular aplicada a sólidos e biomoléculas". In *Anais – IV Escola do CBPF*, Rio de Janeiro, 15 a 26 de julho de 2002, Ligia MCS Rodrigues *et al*, editores, CBPF/ Ao Livro Técnico, Rio de Janeiro, 2003.
- Murray, Robert K.; Granner, Daryl K.; Mayes, Peter A.; Rodwell, Victor W., "Harper: Bioquímica". 9ª ed. São Paulo, ed. Atheneu, 2002.
- Myrdal, P. B.; Krzyzaniak, J. F.; Yalkowsky, S. A. "Modified Trouton's rule for predicting the entropy of boiling". *Ind. Eng. Chem. Res.*, v. 35, p. 1788-1792, 1996.
- Neves, Pereira David. "Parasitologia Humana". 10ª ed. São Paulo: ed. Atheneu, 2000.
- NIH - Center for molecular modeling. Disponível em: http://cmm.info.nih.gov/intro_simulation/course_for_html.html. Acesso em 23/05/2003.
- Norgwyn Montgomery Software Inc. – Geometry Optimiztion: Molecular Mechanics Minimizers. Disponível em: <http://www.norgwyn.com/mmegeom.html> . Acesso em 20/12/2003.
- NorGwyn montgomery SoftwareTM Inc. 1992-2000. Molecular Modeling ProTM. Manual, 4ª ed., Quinn, James A.
- Okamoto, André Kimura. "Estudo Teórico das Relações Estrutura-Atividade Biológica sobre a Inibição por Cumarinas da Mutagênese Induzida pela 2-amino-3-metilimidazo [4,5 – f] quinolina em *Salmonella typhimurium* TA98". 2002. 51 f.. Dissertação (Mestrado em Química) – Instituto de Química, Departamento de Físico-química, Unicamp, Campinas – SP.
- Oliveira, D. B.; Gaudio, A. C. BuildQSAR: A new computer propgram for QSAR analisys. *Quantitative Structure.-Active Relationship*, v.19, n. 06, 2001.
- OMS – Organização Mundial de Saúde. Disponível em: <http://www.who.int/en/> , acesso em 15/10/2003.
- Pagola, S. *et al*. "The structure of malaria pigment (β -haematin)". *Nature*, v. 404, p. 307-310, 2000.
- Patrick, Graham L. "An Introduction to Medicinal Chemistry". 2ª ed. New York: Oxford University Press, 2001.
- Pilar, Frank L. "Elementary Quantum Chemistry". 2ª ed. McGraw-Hill, 1990.

- Pople, J. & Beveridge, D. *"Aproximate Molecular Orbital Theory"*. Mc Graw-Hill, 1970.
- Quinn, James A. "A new method for the calculation of hydrophilic surface area (HSA) from structure". Norgwyn Montgomery Software Inc. Disponível em: <http://www.norgwyn.com/hsa.zip> . Acesso em 23/05/2003.
- Rang, H. P. & Dale, M. M. *"Farmacologia"*. 2ª ed. Rio de Janeiro: ed. Guanabara Koogan, 1993.
- Ratner, Mark A. & Schatz, George C. *"Introduction to Quantum Mechanics in Chemistry"*. New Jersey: Prentice Hall, 2001.
- RBM – Roll Back Malaria. Disponível em: <http://www.rbm.who.int/> acesso em 15/10/2003.
- Rey, Luís. *"Parasitologia: parasitos e doenças parasitárias do homem nas Américas e na África"*. 2ª ed. Rio de Janeiro: ed. Guanabara Koogan, 1991.
- Rogers, Donalds W. *"Computational chemistry using the PC"*. 2ª ed, New York: VHC Publishers, Inc, 1994.
- Roitt, Ivan; Brostoff, Jonathan; Male, David. *"Imunologia"*. 5ª ed. São Paulo: ed. Manole, 1999.
- Rosenthal, Philip J. *et al.* "Identification of three stage-specific proteinases of *Plasmodium falciparum*". *J. Exp. Med.*, v. 166, p. 816-821, 1987.
- Rosenthal, Philip J. "Proteases of malaria parasites: new targets for chemotherapy". *Emerg. Infect. Dis.*, v. 4, n. 1, p. 49-57, 1998.
- Rosenthal, Philip J.; Sijwali, Puran S.; Shenai, Bhaskar R. "Cysteine proteases of malaria parasites: targets for chemotherapy". *Curr. Pharm. Des.*, v. 8, n. 18, p. 1659-1672, 2002.
- Rubin, Emanuel & Farber, John L. *"Patologia"*. Rio de Janeiro: Interlivros, 1990.
- Sabnis, Yogesh A.; desai, Prashant V.; Rosenthal, Philip j.; Avery, Mitchell A. "Probing the structure of falcipain-3, a cysteine protease from *Plasmodium falciparum*: Comparative protein modeling and docking studies". *Protein Sci.*, v. 12, p. 501-509, 2003.
- Segurado, Aluisio Augusto Cotrim; Di santi, Silvia Maria; Shiroma, Mario. "In vivo and in vitro *Plasmodium falciparum* resistance to chloroquine, amodiaquine and quinine in the brazilian Amazon". *Rev. Inst. Med. Trop. S. Paulo*, v. 39, n. 02, p. 85-90, 1997.

- Shenai, Bhaskar R.; Semenov, A. V.; Rosenthal, Philip. J. "Stage-specific antimalarial activity of cysteine protease inhibitors". *Biol. Chem.*, v. 383, n. 5, p. 843-847, 2002.
- Silva, Penildon. "*Farmacologia*". 4ª ed, Rio de Janeiro: Guanabara Koogan, 1994.
- Silverman, Richard B. "*The organic chemistry of drug design and drug action*". San Diego: Academic Press, 1992.
- Sinden, R. E. "Excystation by sporozoites of malaria parasites". *Nature*, v. 252, p. 314, 1974.
- Sullivan, David J. *et al.* "On the molecular mechanism of chloroquine's antimalarial action". *Proc. Natl. Acad. Sci. U. S. A.* v. 93, p. 11865-11870, 1996.
- Sullivan, David J. "Theories on malaria pigment formation and quinoline action". *Int. J. Parasitol.*, v. 32, n. 13, p. 1645-1653, 2002.
- Tamta, Hemlata & Mukhopadhyay, Anup K. "Biochemical targets for malaria chemotherapy". *CRIPS*, v. 04, n. 01, 2003.
- Thomas, Gareth. "*Química Medicinal: uma introdução*". Rio de Janeiro: Guanabara Koogan, 2003.
- Topliss, J. G. & Costello, R. J. "Chance Correlations in Structure-Activity Studies using Multiple Regression Analysis". *J. Med. Chem.*, v.15, p. 1066-1068, 1972.
- Trager, W. & Jensen, J. B. "Human malaria parasites in continuous culture". *Science*, v. 193, p. 673-675, 1976.
- Unger, S. H. & Hansch, C. "On model building in structure-activity relationships. A reexamination of adrenergic blocking activity of β -halo- β -aralkylamines". *J. Med. Chem.*, v. 16, n. 07, p. 745-749, 1973.
- Veber, D. F.; Johnson, S. R.; Cheng, H-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. "Bioavailability". *J. Med. Chem.*, v. 45, p. 2615-2623, 2002.
- Verloop, A.; Hoogenstraaten, W.; Tipker, J. "In drug Design". *Ed. J. Ariens.*, p. 165-207, 1976.
- Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K. "Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure-Activity Relationships. 4. Additional Parameters for Hydrophobic and Dispersive Interactions and Their Application for an Automated Superposition of Certain Naturally Occurring Nucleoside Antibiotics". *J. Chem. Inf. Comput. Sci.*, v. 29, p. 163-172, 1989.

Voet, Donald; Voet, Judith G.; Pratt, Charlotte W. *"Fundamento da Bioquímica"*. Tradução: Arthur Germano Fett Neto *et al.* Porto Alegre: Artmed, 2002.

Wang, Renxiao; Fu, Ying; Lai, Luhua. "A new atom-additive method for calculating partition coefficients". *J. Chem. Inf. Comput. Sci.*, v. 37, n. 03, p. 615-621, 1997.

Wernsdorfer, W. H & MCgregor, L. *"Malaria: Principles and Practice of Malariology"*. Edinburgh: Churchill Livingstone, 1988.

Wiser, Mark F. Tulane University. "Mechanisms of Drug Action and Resistance (Focus on Antimalarials)", 2003. Disponível em:

<http://www.tulane.edu/%7Ewiser/>, acesso em 11/06/2003.

Wiser, Mark F. Tulane University. "Biochemistry of *Plasmodium*", 1999. Disponível em: <http://www.tulane.edu/%7Ewiser/>, acesso em 11/06/2003.

<http://www.bioinformatics.leeds.ac.uk/br>. Acesso em 24/11/2003.

<http://www.caos.kun.nl/> Acesso em 24/11/2003.

<http://www.cdc.gov>. Acesso em 24/11/2003.

<http://www.chemkeys.com/bra/md/mdhtm> Acesso em 24/11/2003.

<http://www.cmpharm.ucsf.edu/cohen/>. Acesso em 24/11/2003.

<http://www.hanzeplein.nl/tropenzorg/malaria.html>. Acesso em 24/11/2003.

<http://www.norgwyn.com/hsa.zip>. Acesso em 24/11/2003.

<http://www.traveldocor.co.uk/malaria.htm> . Acesso em 24/11/2003.