

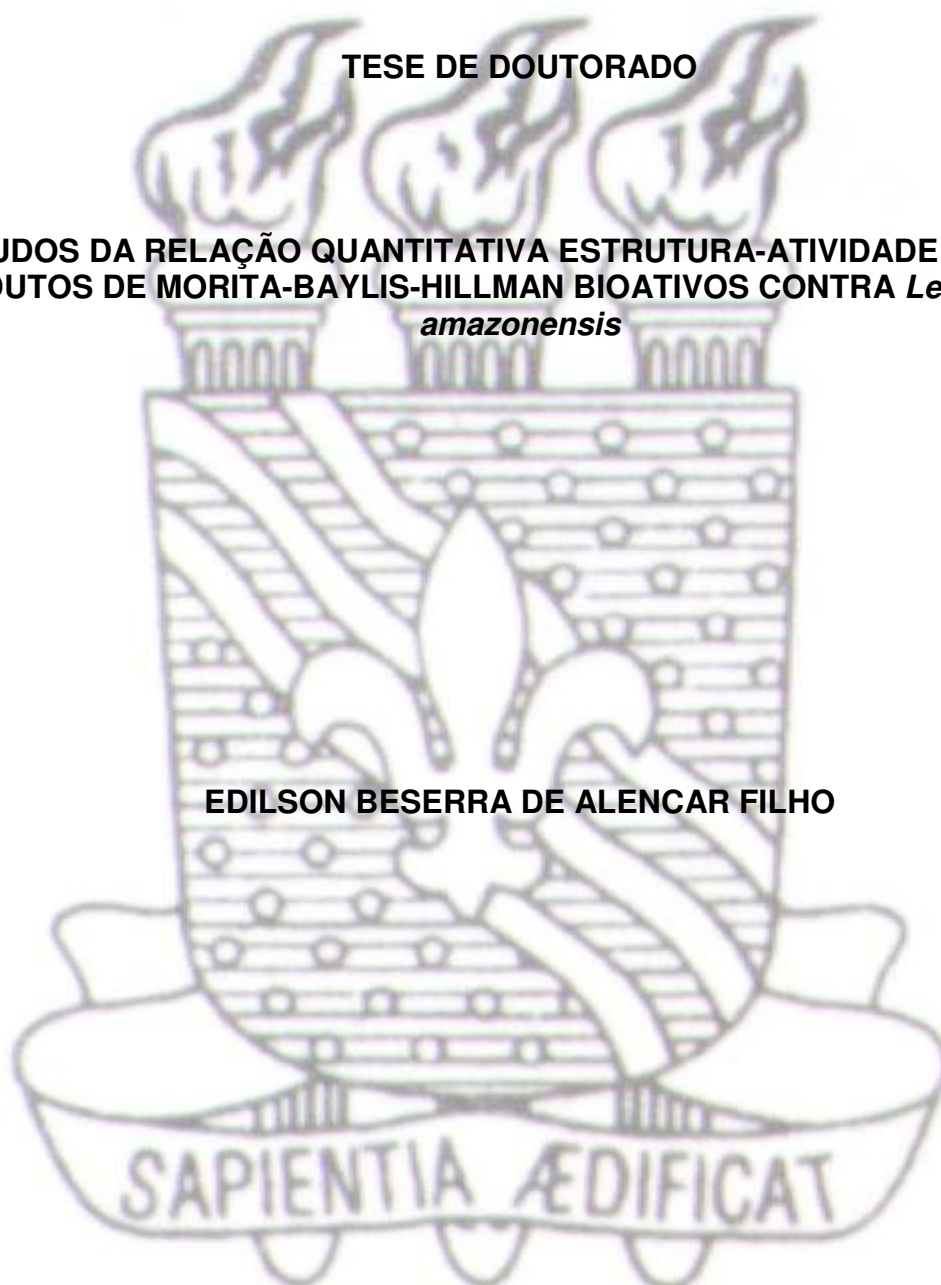


UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

TESE DE DOUTORADO

ESTUDOS DA RELAÇÃO QUANTITATIVA ESTRUTURA-ATIVIDADE (QSAR) DE
ADUTOS DE MORITA-BAYLIS-HILLMAN BIOATIVOS CONTRA *Leishmania
amazonensis*

EDILSON BESERRA DE ALENCAR FILHO



João Pessoa – PB – Brasil
Dezembro /2012



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

TESE DE DOUTORADO

ESTUDOS DA RELAÇÃO QUANTITATIVA ESTRUTURA-ATIVIDADE (QSAR) DE
ADUTOS DE MORITA-BAYLIS-HILLMAN BIOATIVOS CONTRA *Leishmania
amazonensis*

Edilson Beserra de Alencar Filho

Tese apresentada ao Programa de Pós-Graduação em Química do Centro de Ciências Exatas e da Natureza da Universidade Federal da Paraíba, como requisito para obtenção do Título de Doutor em Ciências, área de concentração Química Orgânica.

Orientador: Mário Luiz Araújo de Almeida Vasconcellos
2º Orientador: Karen Cacilda Weber

João Pessoa – PB – Brasil
Dezembro /2012


A368e Alencar Filho, Edilson Beserra de.
Estudos da relação quantitativa estrutura-atividade (QSAR)
de adutos de Morita-Baylis-Hillman bioativos contra Leishmania
amazonensis / Edilson Beserra de Alencar Filho.-- João
Pessoa, 2012.
155f. : il.
Orientadores: Mário Luiz Araújo de Almeida Vasconcellos,
Karen Cacilda Weber
Tese (Doutorado) – UFPB/CCEN
1. Química Orgânica. 2. Adutos de Morita-Baylis-Hillman.
3. Leishmania amazonensis. 4. DFT/PCM. 5. QSAR.

UFPB/BC


CDU: 547(043)

**Estudos da Relação Quantitativa Estrutura –
Atividade (QSAR) de Adutos de Morita-Baylis-
Hillman Bioativos Contra *Leishmania amazonensis*.**

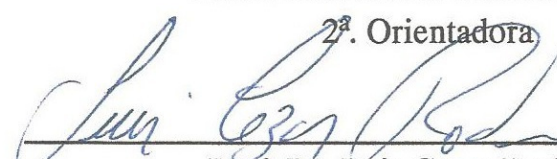
Tese de Doutorado de Edilson Beserra de Alencar Filho aprovada
pela banca examinadora em 14 de dezembro de 2012:



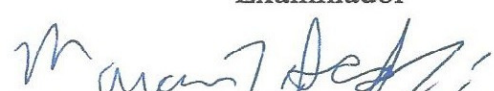
Prof. Dr. Mário Luiz Araújo de Almeida Vasconcellos
Orientador/Presidente




Profa. Dra. Karen Cacilda Weber
2ª. Orientadora



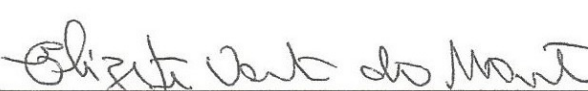
Prof. Dr. Luis Cezar Rodrigues
Examinador



Prof. Dr. Marcus Tullius Scotti
Examinador



Profa. Dra. Luciana Scotti
Examinadora



Profa. Dra. Elizete Ventura do Monte
Examinadora

DEDICATÓRIA

À Seu Edilson (*in memoriam*) e Dona Aurení.

“No fim tu hás de ver que as coisas mais leves são as únicas que o vento não conseguiu levar: um estribilho antigo; um carinho no momento preciso; o folhear de um livro de poemas; o cheiro que tinha um dia o próprio vento.”

Mário Quintana

AGRADECIMENTOS

Ao conjunto de forças criadoras e regentes do Universo a quem chamamos Deus e seus mensageiros, pelo dom da existência consciente e oportunidade de aprendizado;

À minha esposa Bruna e nosso filhote Ike, por este lar feliz onde que escolhemos partilhar nossas vidas, base para a constituição de qualquer empreendimento como foram as árduas horas dedicadas a este trabalho;

Ao meu Orientador Prof. Dr. amigo Mário Vasconcellos, pela formação oferecida, discussões científicas, motivação constante, amizade sincera e confiança depositada; A Profa. Dra. Karen Weber pela disponibilidade em contribuir na construção deste trabalho, ao vivo ou *on line*; ao Prof. Dr. Gerd Bruno pelas dicas “Quânticas” e computacionais; Aos Profs. Elizete Ventura e Wallace Fragoso pelo zelo no enriquecimento deste trabalho.

Aos meus familiares pela torcida e compreensão da ausência, causada pelos voos que resolvi alçar na vida: minha Mãe Aureni pelo apoio incondicional; meu Pai Edilson (*in memoriam*) por imprimir em mim o desejo de fazer o que é correto; meus irmãos e cunhados (Edini, Iarley, Gerúsia, Jussara, César e Arnaldo) que me deram a oportunidade de chegar até aqui; aos sobrinhos pela alegria, carinho, respeito, confiança no Tio coruja;

A minha sogra Áurea pela estadia e amizade nos deslocamentos à João Pessoa, em extensão aos meus cunhados Diego e Daniel; Ao Sogro Edwin e Esposa Cleuma pela torcida e amizade.

Aos amigos do LASOM, Juninho, Fábio, Saulo, Wagner Visconde, Ticiano, Suevy Kid, Natália, Saragina, Hérmersson, Yen, Ramon, pelos inesquecíveis momentos de descontração, amizade e exercício intelectual;

Aos grandes amigos de Pombal, Petrolina e João Pessoa, os quais não são necessários nominar, pela torcida e amizade;

Ao Prof. Dr. Julianeli Tolentino, grande incentivador para que eu aceitasse e persistisse neste desafio, quando ainda amadurecia a ideia de construí-lo, bem como ao Prof. Dr. Jackson Guedes pelo apoio, incentivo e cobertura imprescindíveis para concretização desta Tese;

A todos os colegas e amigos do Colegiado de Ciências Farmacêuticas da UNIVASF pelo agradável ambiente de trabalho e torcida;

RESUMO

Título: Estudos da Relação Quantitativa Estrutura-Atividade (QSAR) de Adutos de Morita-Baylis-Hillman Bioativos contra *Leishmania amazonensis*.

Os Adutos de Morita-Baylis-Hillman (AMBH) compreendem uma classe de moléculas investigadas por nosso grupo de pesquisas nos aspectos sintéticos, teóricos e de bioatividade. Neste trabalho, apresentamos modelos de Relação Quantitativa entre a Estrutura Química e a Atividade Leishmanicida (QSAR) envolvendo 32 AMBH aromáticos. Deste modo, inicialmente foram investigadas as conformações mais estáveis de cada composto através de métodos teóricos, as quais foram utilizadas na construção dos modelos. Foram obtidas curvas de energia potencial utilizando o método semi-empírico AM1, considerando graus de liberdade rotacionais (ligações sigma). A partir destas curvas, a conformação de menor energia para cada molécula foi selecionada e otimizada ao nível B3LYP/6-31+G(d), considerando os efeitos do solvente aquoso usando PCM ("Polarizable Continuum Model"). Dados espectroscópicos de Ressonância Magnética Nuclear de prótons corroboraram o estudo conformacional. Ligações de Hidrogênio Intramoleculares (LHI) se mostraram presentes na maioria das moléculas estudadas, conforme caracterização estrutural e cálculos QTAIM ("Quantum Theory Atoms in Molecules"). Curiosamente, os compostos que apresentaram Ligações de Hidrogênio envolvendo o grupo nitro (NO₂) e a hidroxila (OH) possuem melhores valores de atividade biológica (menor IC₅₀). Uma explicação está baseada no mecanismo de ação redox de nitrocompostos. Observação das cargas NBO ("Natural Bond Orbitals") e análise dos orbitais de fronteira LUKO ("Lowest Unoccupied Kohn-Sham Orbitals") ao nível do grupo *orto*-nitro estão de acordo com estas análises. Considerando os cálculos quânticos, bem como observações estruturais, quatro descritores foram selecionados *a priori* e submetidos a um estudo QSAR ("Quantitative Structure-Activity Relationships") utilizando modelagem PLS ("Partial Least Squares") e MLR ("Multiple Linear Regression"). Uma segunda abordagem QSAR foi realizada a partir de outro conjunto de descritores obtidos através da plataforma *online* E-DRAGON, os quais foram submetidos ao método de seleção de variáveis OPS ("Ordered Predictor Selection"). Os parâmetros de qualidade obtidos para os modelos indicam que ambos são robustos e preditivos. **Palavras chave:** Adutos de Morita-Baylis-Hillman, *Leishmania amazonensis*, DFT/PCM, QSAR.

ABSTRACT

Title: Quantitative Structure-Activity Relationship (QSAR) Studies of Morita-Baylis-Hillman Adducts bioactive against *Leishmania amazonensis*.

The Morita-Baylis-Hillman Adducts (MBHA) is a class of molecules studied by our research group on synthetic, theoretical and bioactivity aspects. In this work, we present Quantitative Structure-Activity Relationship (QSAR) models involving 32 aromatic MBHA. Initially, the most stable conformations of thirty-two MBHA were investigated by theoretical methods, which were used to construct models. For this study, were obtained potential energy curves using AM1 semi-empirical method, considering rotational degrees of freedom (sigma bonds). From these curves, the less energy conformation to each molecule was selected and optimized at B3LYP/6-31+G(d) level, considering solvent effects through Polarizable Continuum Model (PCM). Proton Nuclear Magnetic Resonance data are in agreement with the conformational study. Intramolecular Hydrogen Bonds (IHB) are presents in the most of the studied compounds, according to structural characterization and QTAIM calculations. Curiously, compounds that showed hydrogen bonds involving the nitro and hydroxyl groups have the best values of biological activity (IC_{50}). An explanation is based on redox mechanism of action of nitrocompounds. NBO (Natural Bond Orbital) charges and LUKO (Lowest Unoccupied Kohn-Sham Orbitals) analysis at the *ortho*-nitro group are in agreement with these analyses. Considering quantum calculations and structural observations, four descriptors were selected *a priori* and submitted to a QSAR study using PLS (Partial Least Squares) and MLR (Multiple Linear Regression) modeling. A second QSAR approach was made from the another set of descriptors obtained through the online platform E-Dragon, which were submitted to a variable selection method. The quality parameters obtained for models indicate that both are robust and predictive. **Keywords:** Morita-Baylis-Hillman Adducts, *Leishmania amazonensis*, DFT/PCM, QSAR.

LISTA DE FIGURAS

Figura 2.1: Representação da construção de Componentes Principais.....	38
Figura 2.2: Representação do formaldeído solvatado por moléculas de água e seus dipolos.....	55
Figura 2.3: Representação da influência do dipolo molecular polarizando o dielétrico (cinza) que, por sua vez, produz um campo de reação atuando de volta no soluto..	56
Figura 2.4: Cavidade do soluto definida por esferas centradas nos átomos.....	57
Figura 2.5: Mapa de relevo para região da morfina.....	59
Figura 2.6: Representação QTAIM da molécula BF ₃	60
Figura 2.7: Gráfico molecular obtido por QTAIM para as bases nitrogenadas Guanina e Citosina, interagindo por ligações de hidrogênio.....	61
Figura 4.1: Ângulos diedrais investigados para cada estrutura base dos AMBH estudados.....	66
Figura 4.2: Gráficos obtido após o procedimento RPES para 1a.....	67
Figura 4.3: Parâmetros utilizados para o OPS.....	71
Figura 5.1: Conformações mais estáveis dos AMBH estudados neste trabalho. Em parêntesis os valores de IC ₅₀ em µM para cada molécula.....	74
Figura 5.2: Gráficos Moleculares e Pontos Críticos de Ligação para alguns AMBH.	78
Figura 5.3: Gráficos moleculares para 7d, 23l e 24l após obtenção de geometrias em PCM-B3LYP/6-311++(d,p).....	83
Figura 5.4: Posição relativa do hidrogênio carbinólico em relação ao anel aromático (ângulos diedrais selecionados).....	86
Figura 5.5: Localização dos LUKO's para os AMBH mais ativos da série.....	90
Figura 5.6: Resultados da validação cruzada "leave-n-out".....	95
Figura 5.7: Valores Previstos <i>versus</i> Experimentais de pIC ₅₀ (triângulos vermelhos correspondem as amostras da validação externa).....	96
Figura 5.8: Valores de R ² e Q ² _{loo} para a aleatorização-Y (o triângulo vermelho corresponde aos valores para o modelo original).....	98

Figura 5.9: Resultados da validação cruzada “leave-n-out” para o modelo OPS-PLS.....	105
Figura 5.10: Resultados do procedimento de aleatorização-Y (o triângulo vermelho corresponde aos valores para o modelo original).....	106
Figura 5.11: Representação de par de átomos com volumes de Van der Waals relativamente maiores para 3b, a uma distância topológica de 8 ligações.....	109
Figura 5.12: Estruturas dos AMBH propostas a partir dos estudos QSAR.....	112
Figura 5.13: Geometrias tridimensionais otimizadas dos AMBH propostos.....	112

LISTA DE TABELAS

Tabela 5.1: Parâmetros estruturais para Ligações de Hidrogênio Intramoleculares.	77
Tabela 5.2: Parâmetros topológicos (densidade eletrônica $\rho(r)$ e Laplaciano $\nabla^2 \rho(r)$ ao nível do BCP) para as Ligações de Hidrogênio Intramoleculares.	80
Tabela 5.3: Geometrias otimizadas em PCM-B3LYP/6-311++G(d,p) para investigação de outras possibilidades de LHI.....	81
Tabela 5.4: Parâmetros estruturais e topológicos QTAIM (densidade eletrônica $\rho(r)$ e Laplaciano $\nabla^2 \rho(r)$ ao nível do BCP) para as interações OH...N(py) de 7d, 23l e 24l.	83
Tabela 5.5: Dados ^1H NMR para os AMBH (deslocamento químico do hidrogênio carbinólico).....	84
Tabela 5.6: Cargas NBO no nitrogênio (qN) e oxigênio (qO—considerando o aceptor na LHI) do grupo nitro para os AMBH's.....	89
Tabela 5.7: Valores de atividade transformados e descritores usados no QSAR.	93
Tabela 5.8: Parâmetros estatísticos dos modelos PLS.	94
Tabela 5.9: Resultados obtidos com o procedimento <i>bootstrapping</i> para diferentes modelos PLS.....	97
Tabela 5.10: Principais parâmetros de validação para diferentes modelos MLR, obtidos por escolha aleatória do conjunto teste.	99
Tabela 5.11: Descritores selecionados pelo método OPS.	101
Tabela 5.12: Valores obtidos com procedimento aleatorização-Y para modelo com 12 descritores (a última linha corresponde aos valores obtidos com o modelo original)	102
Tabela 5.13: Valores obtidos com procedimento aleatorização-Y para modelo com 16 descritores (a última linha corresponde aos valores obtidos com o modelo original)	103
Tabela 5.14: Parâmetros de validação do modelo OPS-PLS.....	105
Tabela 5.15: Resultados obtidos com o procedimento <i>bootstrapping</i> para diferentes modelos PLS.....	107
Tabela 5.16: Descritores <i>a priori</i> para as moléculas propostas e valores de atividade previstos.....	113
Tabela 5.17: Descritores selecionados pelo OPS para as moléculas propostas e valores de atividade previstos.	114

LISTA DE ESQUEMAS

Esquema 2.1: Definição e cálculo da constante de Hammett σ	31
Esquema 2.2: Resumo da Técnica MLR aplicada a QSAR.....	36
Esquema 2.3: Representação da Técnica PCR e estimativa dos coeficientes de regressão	39
Esquema 2.4: Resumo da Técnica OPS.....	47
Esquema 2.5: Procedimento autoconsistente para as equações de Kohn-Sham.....	53
Esquema 5.1: Mecanismo reacional da biorredução de nitrocompostos	88

LISTA DE ABREVIATURAS E ACRÔNIMOS

- AM1** – “Austin Model One”
- AMBH** – Adutos de Morita-Baylis-Hillman
- B3** – Combinação dos três parâmetros de Becke
- B3LYP** – Funcional Híbrido (três parâmetros de Becke com Lee, Yang e Parr)
- B88** – Funcional de Troca de Becke
- BCP** – “Bond Critical Point”
- BLYP** – Funcional Híbrido (Becke com Lee, Yang e Parr)
- BP** – “Bond Path”
- DABCO** – 1,4 Diazabicyclo [2,2,2] octano
- DFT** – “Density Functional Theory”
- DNDi** – “Drugs for Negligent Diseases initiative”
- ELUKO** – Energia do Orbital LUKO
- GGA** – “Generalized Gradient Analysis”
- GRE** – Grupo Retirador de Elétron
- HCA** – “Hierarchical Cluster Analysis”
- HOKO** – “Highest Occupied Konh-Shan Orbital”
- HOMO** - “Highest Occupied Molecular Orbital”
- IUPAC** – “International Union of Pure and Applied Chemistry”
- KNN** – “K Nearest Neighbor”
- LCAO** – “Linear Combination of Atomic Orbitals”
- LH** – Ligação de Hidrogênio
- LHI** – Ligação de Hidrogênio Intramolecular
- LogP** – Logaritmo do Coeficiente de Partição octanol/água
- LTA** – Leishmaniose Tegumentar Americana
- LUKO** - “Lowest Unoccupied Konh-Shan Orbital”
- LUMO** - “Lowest Unoccupied Molecular Orbital”
- LV** – Leishmaniose Visceral
- LYP** – Funcional de Correlação de Lee, Yang e Parr
- MLR** – “Multiple Linear Regression”
- NAO** – “Natural Atomic Orbital”

NBO – “Natural Bond Orbital”

NHO – “Natural Hybrid Orbital”

OMS – Organização Mundial da Saúde

PC – “Principal Component”

PCA – “Principal Component Analysis”

PCM – “Polarized Continuum Model”

PCR – “Principal Component Regression”

PLS – “Partial Least Squares”

PRESS – “Predictor Error Sum of Squares for Cross Validation”

QSAR – “Quantitative Structure Activity Relationship”

QTAIM – “Quantum Theory Atoms in Molecules”

REA – Relação Estrutura-Atividade

RMBH – Reação de Morita-Baylis-Hillman

RMN¹H – Ressonância Magnética Nuclear de Prótons

RMSECV - “Root Mean Square Error for Cross Validation”

RPESS – “Relaxed Potential Energy Surface Scan”

SCF – “Self Consistent Field”

SCRf – “Self Consistent Reaction Field”

SIMCA – “Soft Independent Modeling of Class analysis”

SLogP – Logaritmo do Coeficiente de Partição fase orgânica/fase aquosa com contribuição atômica

SUMÁRIO

1. INTRODUÇÃO	17
1.1 Adutos de Morita-Baylis-Hillman e Doenças Negligenciadas	18
1.2 Leishmanioses	24
1.2.1 Fármacos utilizados na Leishmaniose.....	27
2. FUNDAMENTAÇÃO TEÓRICA	30
2.1 Relação Quantitativa Estrutura-Atividade (QSAR)	30
2.1.1 Quimiometria aplicada aos estudos QSAR.....	35
2.3.1.2 Seleção de variáveis pelo método OPS.....	46
2.2 Métodos Computacionais em Química Quântica	48
2.2.1 Teoria do Funcional da Densidade.....	49
2.3 Modelo do Contínuo Polarizável: Inclusão dos efeitos do solvente	54
2.4 Ligações de Hidrogênio e Teoria Quântica de Átomos em Moléculas ...58	
3. OBJETIVOS	64
3.1 Geral	64
3.2 Específicos	64
4. PROCEDIMENTOS COMPUTACIONAIS	66
4.1 Estudo conformacional e obtenção dos descritores escolhidos <i>a priori</i>	66
4.2 Construções de modelos QSAR com descritores escolhidos <i>a priori</i> ...69	
4.3 Construções de modelos QSAR a partir de descritores moleculares diversos	70
4.3.1 Procedimento de Seleção de Variáveis.....	71
4.4 Proposição e avaliação de novos AMBH aromáticos frente aos modelos construídos	72
5. RESULTADOS E DISCUSSÕES	74
5.1 Conformações de mínimo obtidas para os Adutos de Morita-Baylis-Hillman	74

5.2 Relações dos resultados teóricos com dados espectroscópicos experimentais de RMN¹H	84
5.3 Estabelecimento de Relações Estrutura-Atividade (SAR)	86
5.4 Estudos de QSAR dos Adutos de Morita-Baylis-Hillman	92
5.4.1 Construção de um modelo baseado em descritores escolhidos <i>a priori</i>	92
5.4.1.1 Regressão PLS.....	94
5.4.1.2 Regressão Linear Múltipla (MLR)	98
5.4.2 Construção de um modelo baseado em descritores de naturezas diversas	100
5.5 Proposição e previsão da atividade de novos AMBH baseados nos modelos QSAR construídos	111
6. CONCLUSÕES E PERSPECTIVAS	116
6.1 Conclusões.....	116
6.2 Perspectivas.....	117
7. REFERÊNCIAS BIBLIOGRÁFICAS	119
ANEXOS	Erro! Indicador não definido.
ANEXO A - Coordenadas atômicas após otimização completa em DFT/PCM - B3LYP/6-31+G(d)	132
ANEXO B - Trabalhos decorrentes da Tese	154

INTRODUÇÃO

1. INTRODUÇÃO

A noção de que compostos químicos podem interagir com os sistemas orgânicos, promovendo a cura de enfermidades, é muito antiga. A partir do surgimento da Química Moderna por volta do Século XVI, em paralelo aos avanços na microbiologia, fisiologia e patologia, foram estabelecidas as bases para o que se convencionou chamar *Química Medicinal*. Segundo a IUPAC, esta compreende:

Disciplina baseada em Química, envolvendo aspectos das ciências biológicas, médicas e farmacêuticas. Está interessada na invenção, descoberta, desenho, identificação e preparação de compostos biologicamente ativos, o estudo do seu metabolismo, a interpretação do seu modo de ação a nível molecular e a construção de relações estrutura-atividade.

Iniciando com a exploração dos Produtos Naturais e impulsionada pela Revolução Industrial, a Química Medicinal moderna se desenvolveu passando pelos avanços na química sintética, surgimento da biotecnologia e simulação computacional, até chegar na sua concepção atual, a qual não substituiu, mas agregou ao longo dos anos essas diversas estratégias (BARREIRO e FRAGA, 2005).

De modo específico, a partir da década de 1960 com o desenvolvimento da computação moderna, surgiu a área de estudos e pesquisas chamada Química Computacional, a qual viabilizou o estudo de sistemas químicos a partir de métodos teóricos.

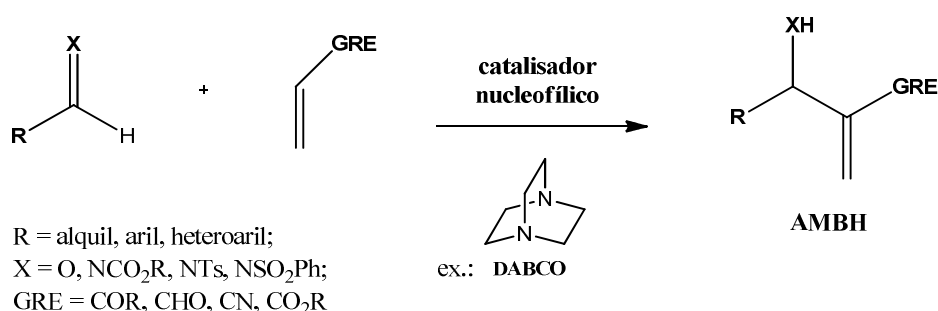
Na perspectiva da descoberta de fármacos, os princípios da Química Computacional podem ser aplicados nos estágios iniciais da descoberta de protótipos bem como no direcionamento de moléculas utilizadas nos estudos de fase clínica. Deste modo, as ferramentas da Química Computacional permitem a concepção de estruturas tridimensionais dos fármacos (estudos conformacionais), estabelecimento de padrões de interação fármaco-receptor (ancoragem molecular), bem como estabelecimento das relações entre propriedades físico-químicas de uma série de moléculas com suas atividades biológicas, QSAR (do inglês “Quantitative

Structure – Activity Relationship”), poupando tempo, esforços e custos no desenvolvimento de moléculas bioativas.

Deste modo, inserido no contexto da Química Medicinal, este trabalho surgiu como uma tentativa de entender a nível molecular a atividade apresentada por uma série congênere de trinta e dois Adutos de Morita-Baylis-Hillman bioativos contra uma das espécies de parasitas causadores da Leishmaniose (*Leishmania amazonensis*). Para este fim, foram criados modelos QSAR que permitissem o estabelecimento de relações entre a estrutura química e a atividade biológica dos compostos estudados, visando à orientação de futuras sínteses através da previsão da tendência de moléculas potencialmente mais ativas. A leishmaniose (considerada uma “doença negligenciada” pela indústria farmacêutica) é endêmica no Brasil e possui um alto número de casos na Região Nordeste. Este fator representou uma motivação adicional para o desenvolvimento desta Tese.

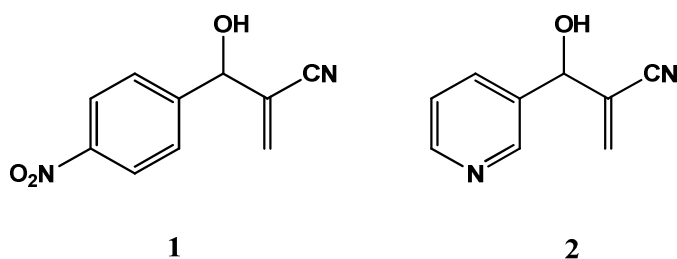
1.1 Adutos de Morita-Baylis-Hillman e Doenças Negligenciadas

A reação de Morita-Baylis-Hillman (RMBH) (MORITA, SUZUKI, HIROSE, 1968; BAYLIS e HILLMAN, 1972) representa uma estratégia de formação de ligações C-C envolvendo um carbono eletrofílico sp^2 (ex. aldeídos, cetonas ou iminas) com a posição α de um alceno ou alcino conectado a um Grupo Retirador de Elétrons (GRE), utilizando fosfinas ou aminas terciárias como catalisadores nucleofílicos, sendo o DABCO mais frequentemente utilizado (Esquema 1.1). O produto desta reação é chamado Aduto de Morita- Baylis-Hillman (AMBH), classe de compostos que foi historicamente utilizada apenas como intermediária na síntese de produtos naturais e outros projetos puramente sintéticos (COELHO e ALMEIDA, 2000).

Esquema 1.1: Reação de Morita-Baylis-Hillman.

A RMBH tem atraído à atenção dos químicos sintéticos por permitir a obtenção de um produto polifuncionalizado, em uma reação “one pot” e que pode ser realizada utilizando solventes não tóxicos ou na ausência de solventes, sendo então um protocolo relativamente barato, eficiente e ecológico para a obtenção de fármacos (BASAVIAH, REDDY e BADSARA, 2010).

Em 1999, os Adutos de Morita-Baylis-Hillman que até então estavam envolvidos apenas em projetos puramente sintéticos, passaram a ser investigados como candidatos a fármacos. Isto se deu a partir do trabalho pioneiro de Kundu e colaboradores (KUNDU *et al.*, 1999), os quais determinaram a atividade de AMBH aromáticos (Figura 1.1) contra o *Plasmodium falciparum* (protozoário causador da malária). Assim esta classe de moléculas foi inserida em um contexto mais amplo: o combate às enfermidades conhecidas como *Doenças Negligenciadas*.

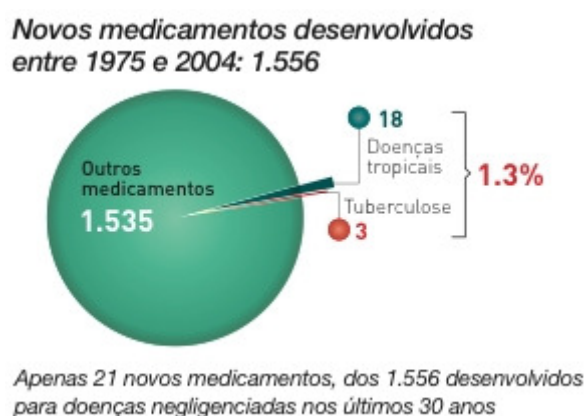
Figura 1.1: AMBH mais ativos contra *P. falciparum* avaliados por Kundu e colaboradores.

Ao longo da última década, este novo paradigma de aplicação dos AMBH tem sido estudado por nosso grupo de pesquisas. Este tem concentrado seus trabalhos na síntese e avaliação biológica de conhecidos e inéditos AMBH,

abordando aspectos de novas metodologias, condições sintéticas, bem como aspectos da Química Computacional.

Dentre as doenças classificadas como negligenciadas encontram-se a Malária, Doença de Chagas, Esquistossomose e Leishmanioses (CAMARGO, 2008), as quais geralmente acometem nichos populacionais alheios ao consumo de produtos oriundos das grandes indústrias farmacêuticas. Deste modo, têm sido historicamente esquecidas quanto aos aspectos de pesquisa, desenvolvimento, inovação e produção, dependendo em grande parte de esforços ainda incipientes dos órgãos governamentais e instituições internacionais sem fins lucrativos como a DNDi (do inglês, “Drugs for Neglected Diseases initiative”). Segundo Chirac (CHIRAC e TORREELE, 2006), dos 1.556 novos medicamentos registrados entre 1975 e 2004, apenas 21 (1,3%) foram destinados para este tipo de doenças (Figura 1.2), também conhecidas como doenças tropicais endêmicas.

Figura 1.2: Perfil quantitativo dos medicamentos desenvolvidos ao longo de 30 anos em comparação aos destinados às doenças tropicais negligenciadas.



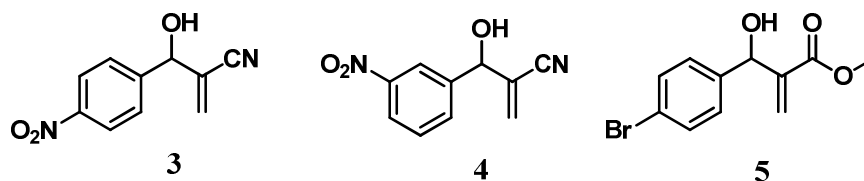
(adaptado de CHIRAC e TORREELE, 2006)

Avaliando os últimos 30 anos, a evolução científica no desenvolvimento de novos procedimentos e moléculas utilizadas na terapêutica é notória, principalmente nos países considerados ricos. Contudo, as doenças tropicais fatais continuam a acometer uma parcela considerável das populações das regiões envolvidas, havendo pouco desenvolvimento no sentido do tratamento destas enfermidades. Em adição, é ainda menor o investimento nas chamadas **Doenças Extremamente Negligenciadas** como a doença do sono, leishmaniose e a doença de Chagas, às

quais são aplicados apenas 5% do financiamento mundial para as Doenças Negligenciadas, o qual já é incipiente (MORAN *et al.*, 2009).

Neste contexto, visando contribuir para a resolução desta problemática, nosso grupo publicou em 2006 um trabalho pioneiro (VASCONCELLOS *et al.*, 2006) o qual descreve a atividade moluscicida de alguns AMBH aromáticos contra o caramujo *Biomphalaria glabrata* (Say), hospedeiro intermediário no ciclo da esquistossomose. Subsequentemente, em 2007, alguns AMBH aromáticos foram apresentados também de forma pioneira como uma classe de compostos muito ativa contra a *Leishmania amazonensis*, causadora das formas cutâneas e mucocutâneas da Leishmaniose, sendo os compostos **3**, **4** e **5** (Figura 1.3) considerados os mais ativos da série (DE SOUZA *et al.*, 2007).

Figura 1.3: AMBH mais ativos da série contra *Leishmania amazonensis*



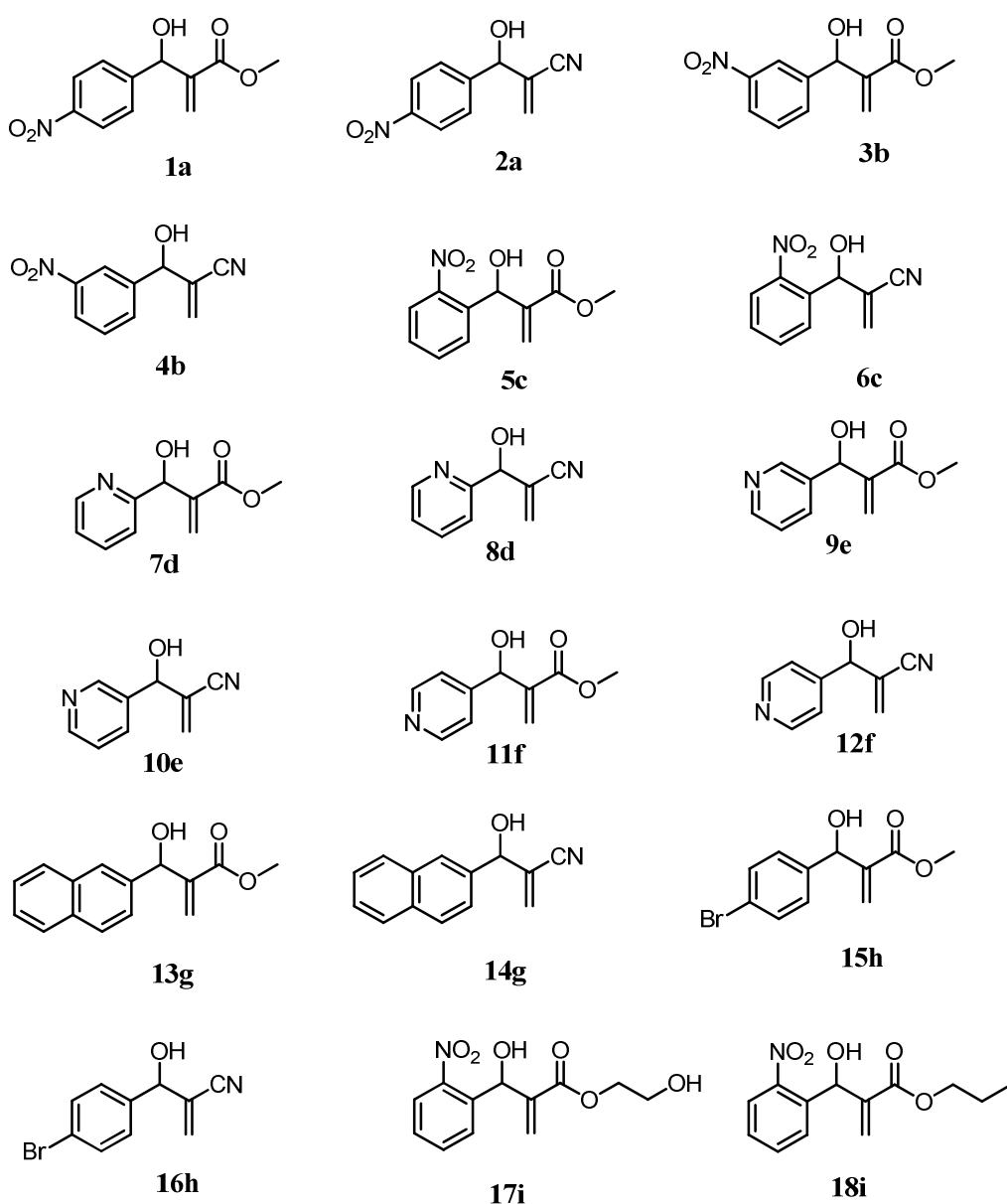
(DE SOUZA *et al.*, 2007)

Em 2009 foram apresentadas as avaliações biológicas de AMBH aromáticos contra *Leishmania chagasi*, causadora da forma visceral da leishmaniose (BARBOSA *et al.*, 2009). No Ano de 2010, um trabalho do nosso grupo apresentou o AMBH 3-hidroxi-2-metileno-3-(4-nitrofenil)propanonitrila como um composto altamente ativo contra *Trypanosoma cruzi*, parasita causador da Doença de Chagas (SANDES *et al.*, 2010). Ainda no ano 2010, nosso grupo apresentou novas condições experimentais para o melhoramento da síntese de dezesseis AMBH, seguido da determinação de suas atividades biológicas contra *L. amazonensis* e *L. chagasi* (JUNIOR *et al.*, 2010).

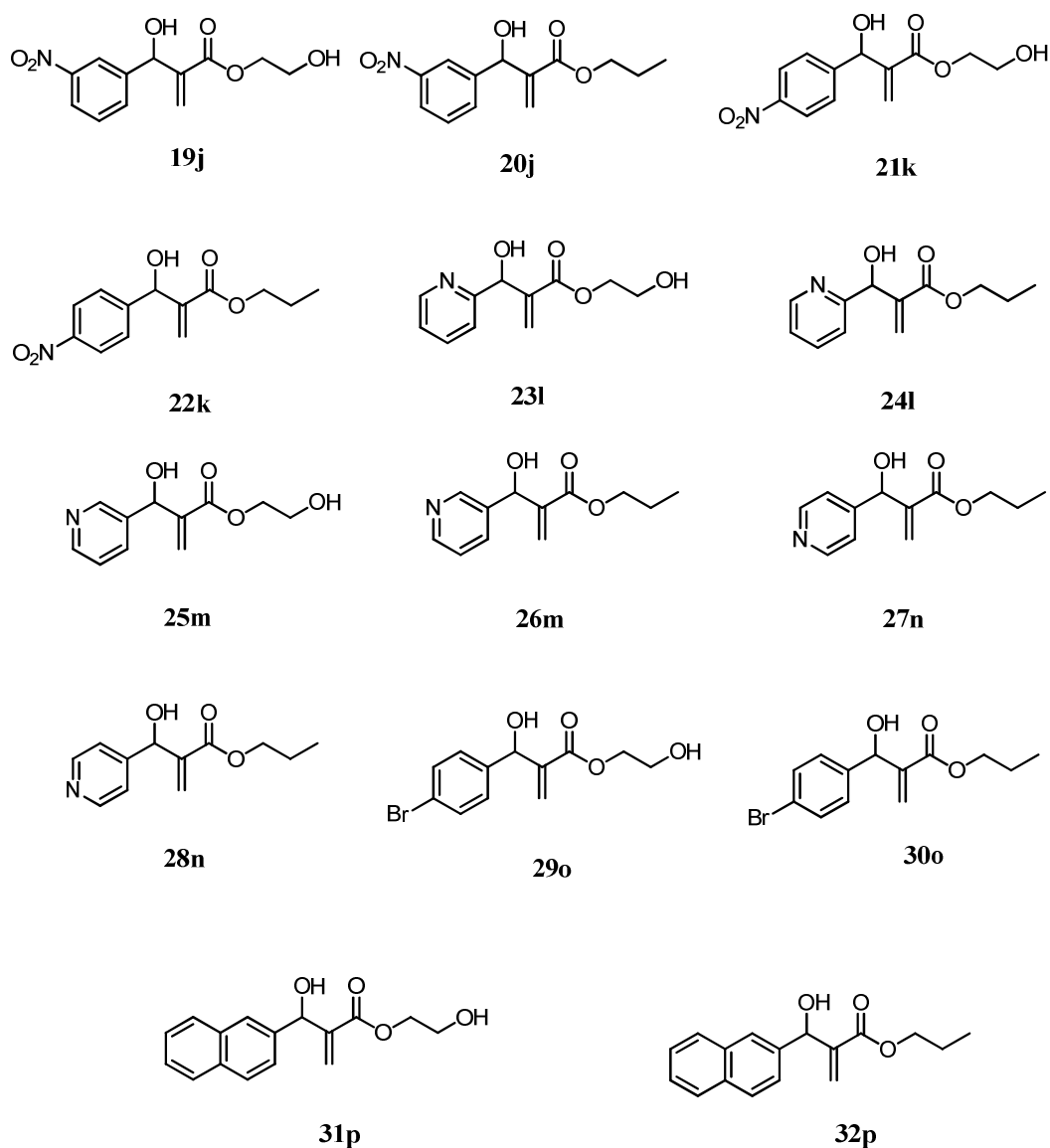
No ano de 2011, nosso grupo publicou a síntese de uma nova série de Adutos de Morita-Baylis-Hillman baseados na estratégia em Química Medicinal conhecida como Bioisosterismo. Esta consiste na troca de um determinado grupo substituinte por outros grupos, dos quais se espera preservar ou aumentar a atividade biológica. O trabalho foi realizado utilizando aldeídos aromáticos e o

acrilato de propila ou acrilato de hidroxietila (SILVA *et al.*, 2011), os quais forneceram 16 novos compostos avaliados contra formas promastigotas de *Leishmania amazonensis*. A Figura 1.4 mostra a estrutura química dos 32 AMBH sintetizados e bioavaliados, apresentados nos dois últimos trabalhos acima citados (JUNIOR *et al.*, 2010 e SILVA *et al.*, 2011). Nos dois trabalhos a avaliação foi feita frente às formas promastigotas do parasita *Leishmania amazonensis*, utilizando o método Eliza/MTT.

Figura 1.4: Trinta e dois AMBH sintetizados e bioavaliados contra espécies de Leishmania.



Continuação da Figura 1.4



(JUNIOR *et al.*, 2010) e (SILVA *et al.*, 2011)

O histórico apresentado demonstra que os Adutos de Morita-Baylis-Hillman correspondem a uma classe promissora no combate às Doenças Negligenciadas, em especial contra a Leishmaniose. Em adição, o conjunto de informações geradas a partir da síntese e avaliação biológica dos 32 compostos da Figura 1.4, contra *Leishmania amazonensis*, enseja a construção de modelos que relacionem a atividade destes compostos com suas estruturas químicas, passo importante no desenvolvimento de uma busca racional por novos fármacos.

1.2 Leishmanioses

A Leishmaniose é uma doença causada por parasitas do gênero *Leishmania*, do qual 15 espécies são responsáveis por infecções em humanos (NGURE *et al.*, 2009; WHO, 2012).

A doença é transmitida através da picada de mosquitos infectados com o parasita. No Brasil estes mosquitos são predominantemente do gênero *Phlebotomus*. (NGURE *et al.*, 2009).

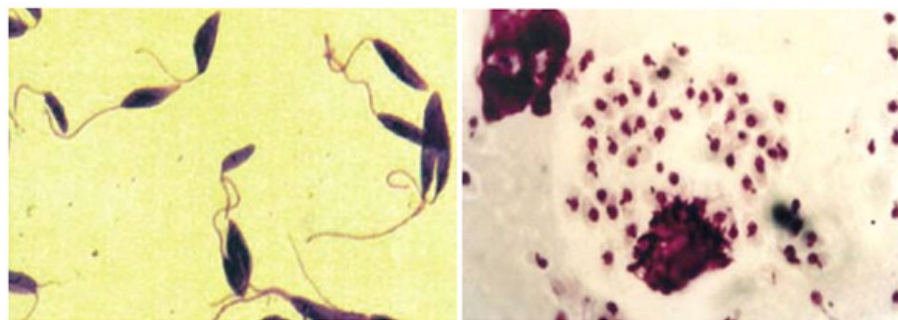
Estima-se que 350 milhões de pessoas vivam sob risco de infecção por este parasita (DESJEUX, 2004). Em geral, a doença acomete as populações mais pobres do globo, haja vista que nos países onde a doença é prevalente estima-se que 80% de suas populações ganhem cerca de US\$ 2 ao dia (DAVIES *et al.*, 2003).

A incidência global da Leishmaniose tem permanecido relativamente estável ao longo dos anos, causando aproximadamente 40.000 óbitos anuais, um número significativo entre as doenças contagiosas. Anualmente, são oficialmente relatados cerca de 280.000 novos casos desta enfermidade em 98 países (ALVAR *et al.*, 2012).

A infecção pelo parasita apresenta as mais diversas manifestações clínicas, a depender da espécie e do estado imunológico do indivíduo acometido. Basicamente, essas manifestações são de dois tipos principais: Leishmaniose Tegumentar Americana (LTA), que acomete pele e mucosas, bem como a Leishmaniose Visceral (LV). A LV é caracterizada por infecção visceral disseminada, sendo a forma mais letal (REITHINGER *et al.*, 2007).

As diversas espécies da *Leishmania spp.* são microorganismos unicelulares que podem apresentar-se em dois tipos distintos (SVS, 2007): as formas promastigota e amastigota. A primeira é uma forma flagelada, encontrada no tubo digestivo do inseto vetor, sendo infectante (Figura 1.5a). A segunda caracteriza-se por ser aflagelada, encontrada nos fagolisossomos dos macrófagos, no organismo dos hospedeiros vertebrados (Figura 1.5b).

Figura 1.5: Formas promastigotas (a) e amastigotas (b) do parasita.

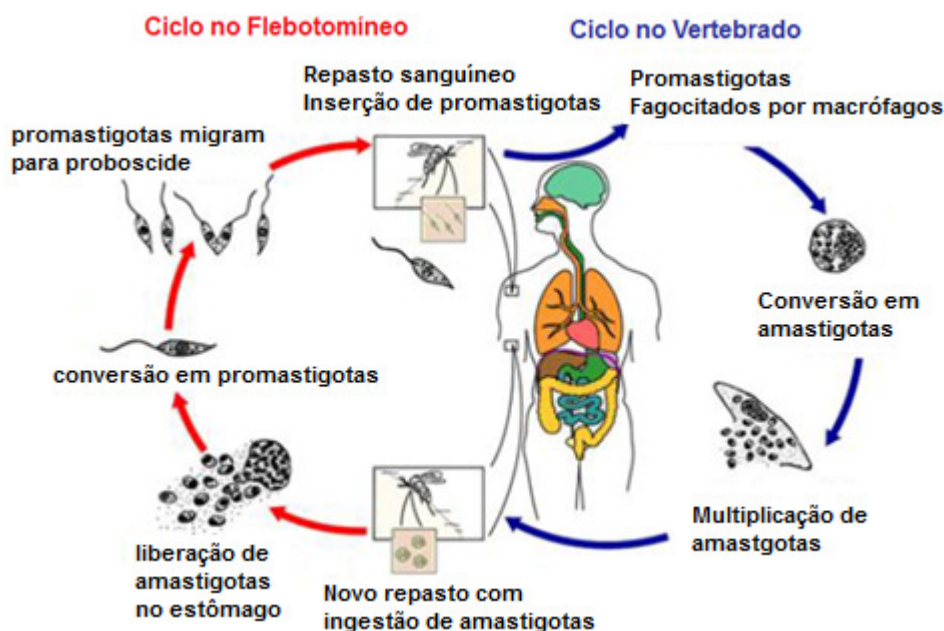


(a) (b)

(adaptado de SVS, 2007)

O ciclo de transmissão da Leishmaniose varia de acordo com a região geográfica, envolvendo uma diversidade de espécies de parasito, vetores, reservatórios e hospedeiros. Contudo, existem vários pontos em comum. Em geral, o ciclo se inicia no momento da picada pelo mosquito, quando as formas promastigotas do parasito são liberadas no sangue do hospedeiro juntamente com a saliva da fêmea do mosquito flebótomo, popularmente conhecido como mosquito palha ou birigui. Caninos e roedores podem participar do ciclo da leishmaniose (ciclo zoonótico), assim como o ser humano (ciclo antropomórfico). Na Figura 1.6 podemos observar todas as fases do ciclo.

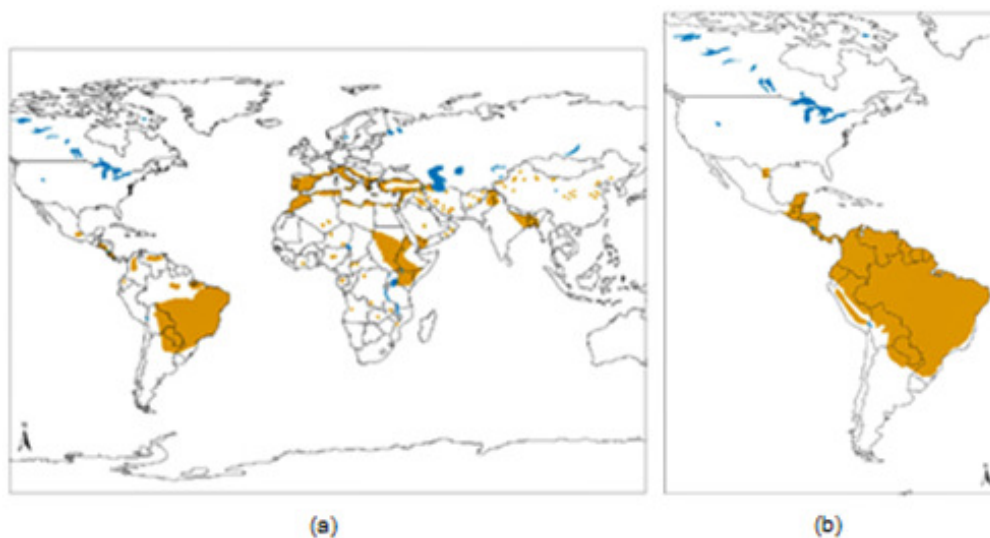
Figura 1.6: Ilustração do ciclo de vida da Leishmania spp.



(adaptado de BARBOSA, 2011)

A Leishmaniose é considerada pela Organização Mundial da Saúde (OMS) como uma das seis mais importantes doenças infecciosas, haja vista seu alto coeficiente de detecção (número de casos novos registrados no decorrer do ano) e capacidade de produzir deformidades (SVS, 2007). A Figura 1.7 apresenta a distribuição de casos desta enfermidade pelo mundo.

Figura 1.7: (a) Em amarelo, distribuição geográfica da Leishmaniose Visceral pelo Mundo; (b) Em amarelo, distribuição geográfica da Leishmaniose cutânea e mucocutânea no continente americano.



(adaptado de WHO/NTD/IDM HIV/AIDS, 2010).

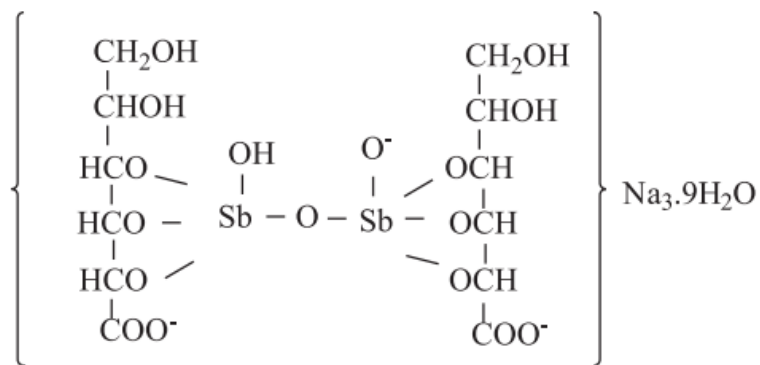
Como podemos perceber da Figura 1.7, a Leishmaniose é uma doença presente em nosso País. Ocorre em ambos os sexos e todas as faixas etárias. As Regiões Norte e Nordeste são as mais afetadas quanto ao número de casos anualmente registrados, ocorrendo respectivamente uma média de 743 e 1850 casos/ano, entre 2005 e 2009 da LV (BRASIL/MS/SVS, 2012). Quanto à Leishmaniose Tegumentar Americana, o número de casos apresentam médias anuais de 9.494 e 6.530 casos/ano entre 2005 e 2008 para as Regiões Norte e Nordeste (BRASIL/MS/SVS, 2012). Estas duas regiões historicamente concentram o maior número de casos da doença anualmente registrados. O prejuízo social desta infecção, assim como em outras parasitoses, pode ser compreendido a partir da diminuição da qualidade de vida destas populações e diminuição da força de trabalho nas áreas atingidas.

1.2.1 Fármacos utilizados na Leishmaniose

A terapêutica da Leishmaniose é realizada basicamente com o uso dos chamados antimoniais pentavalentes, os quais são considerados caros, tóxicos e de baixa eficácia no tratamento da leishmaniose.

O uso medicinal de compostos de antimônio para as mais diversas doenças remonta a um período anterior à Era Cristã. Contudo, somente em 1912, Gaspar de Oliveira Vianna (VIANNA, 1912) observou que um composto desta natureza (o tártaro emético) era eficaz no tratamento da Leishmaniose Tegumentar Americana. Em 1936, Schmidt introduziu na terapia médica o Estibogluconato de sódio (Figura 1.8), conhecido comercialmente como Solustibosan® (Bayer) ou Pentostam® (GSK) (RATH *et al.*, 2003), o qual ainda é utilizado na atualidade.

Figura 1.8: Estibogluconato de sódio.



(adaptado de RATH.*et al.*, 2003)

O tratamento das diferentes formas desta parasitose depende de sua manifestação clínica e da espécie de parasita envolvido (PISCOPO e AZZOPARDI, 2012). Outros fármacos também utilizados nesta terapêutica são o antimoniato de meglumina (Glucantime®), anfotericina B e a miltefosina (PISCOPO e AZZOPARDI, 2012) (Figura 1.9).

FUNDAMENTAÇÃO TEÓRICA

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Relação Quantitativa Estrutura-Atividade (QSAR)

Considerando o planejamento racional de fármacos, uma estratégia importante são os estudos da Relação Quantitativa entre a Estrutura Química e a Atividade Biológica, QSAR (do inglês, “Quantitative Structure-Activity Relationship”). Nestes estudos considera-se que existe uma relação entre as atividades biológicas de uma série congênere (geralmente) e valores de determinadas propriedades moleculares (constitucionais, eletrônicas, etc.), chamadas *descritores físico-químicos*. (TAVARES, 2004).

Nos últimos 50 anos, tem-se observado um crescente interesse pelos estudos de modelagem QSAR aplicados à toxicologia, desenho de fármacos, química ambiental, etc. O crescimento pode ser atribuído em grande parte graças ao desenvolvimento de novas metodologias e técnicas computacionais para tratamento de sistemas multivariados e geração dos descritores físico-químicos de forma rápida e robusta, particularmente devido à Química Computacional.

Apesar de os estudos QSAR terem ganhado forma a partir da década de 60, no século XIX já se tinha a ideia de que características físico-químicas das substâncias estavam diretamente relacionadas à sua atividade.

Em 1863, na França, o pesquisador Cros percebeu uma relação existente entre a toxicidade de alcoóis alifáticos e sua solubilidade em água. Na mesma época, Crum-Brown e Fraser (CRUM-BROWN e FRASER, 1868) formularam a ideia de que a ação fisiológica de uma substância em um sistema biológico (Φ) seria uma função de sua constituição química (C):

$$\Phi = f(C) \quad (1)$$

No início do século XX, Overton e Meyer (MEYER, 1899; OVERTON, 1901) independentemente sugeriram que o efeito narcótico de substâncias orgânicas estava relacionado ao seu coeficiente de partição “ P ” entre as fases orgânica e aquosa. Este coeficiente de partição pode ser definido como a razão entre as

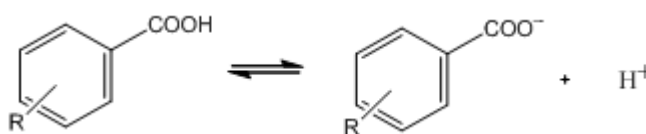
concentrações de equilíbrio de uma substância particionada em duas fases, uma representando a fase orgânica (ou oleosa) e outra representando a fase aquosa:

$$P = \frac{[org.]}{[aq.]} \quad (2)$$

O logaritmo dos valores de P para as mais diversas substâncias, ou $LogP$, é a medida de lipofilicidade geralmente utilizada no estudo das propriedades de fármacos (NOGUEIRA, 2009). Quanto maior o seu valor, maior a tendência lipofílica de uma molécula. Tendo em vista seu significado, este parâmetro passou a ser relacionado principalmente às propriedades farmacocinéticas de transporte dos fármacos pelo organismo, os quais devem atravessar barreiras lipofílicas das membranas celulares para exercerem seus efeitos.

Seguindo a evolução histórica, outros parâmetros relacionando a estrutura química com as propriedades de uma substância foram criados. A partir dos trabalhos de Hammett (HAMMETT, 1935) surge o parâmetro que ficou conhecido como constante do substituinte σ (sigma), que denotava a contribuição de um determinado substituinte do anel aromático ao grau de ionização do ácido benzóico substituído (Esquema 2.1).

Esquema 2.1: Definição e cálculo da constante de Hammett σ .



$$K_R = \frac{[RPhCOO^-][H^+]}{[RPhCOOH]}$$

$$\sigma = \text{Log}K_R - \text{Log}K_H$$

Observando a equação acima, vemos que substituintes com maior tendência a retirar elétrons do anel (por efeito de ressonância ou indutivo) promoveriam maiores valores do parâmetro σ .

Anos depois, Taft desenvolve outro descritor denominado parâmetro estérico, “ES” (TAFT, 1952), o qual representava a mudança na constante de velocidade de hidrólise de ésteres α -substituídos conforme o impedimento estérico causado por um substituinte X. Este parâmetro foi definido de uma forma semelhante a σ , considerando como modelo a hidrólise do acetato de etila.

Em 1962, com base nos trabalhos de Hammett, Hansch e colaboradores publicaram um estudo de relação estrutura-atividade envolvendo substâncias reguladoras de crescimento de plantas e sua relação com a constante de Hammett e um parâmetro de lipofilicidade (π) (HANSCH, 1962). Este parâmetro foi definido com base no $LogP$ e representa uma medida da lipofilicidade relativa de um composto com um dado substituinte R em relação a outro sem este substituinte, contendo hidrogênio H:

$$\pi = LogP_R - LogP_H \quad (3)$$

Este estudo é considerado um dos marcos iniciais do desenvolvimento da moderna metodologia QSAR.

Em outro trabalho que também marcou o início da era QSAR (HANSCH e FUJITA, 1964), Hansch estabelece a relação entre a atividade tóxica de uma série de compostos e alguns parâmetros, utilizando a modelagem por Regressão Linear Múltipla (descrita posteriormente neste trabalho):

$$Log \frac{1}{C_{50}} = a\pi + b\sigma + cES + d \quad (4)$$

onde C_{50} (que compõe a *variável dependente*) é a concentração mínima da substância capaz de gerar efeitos tóxicos em 50% dos indivíduos, π é a conhecida constante de lipofilicidade, σ a constante do substituinte de Hammett e ES o parâmetro de Taft. As três últimas são conhecidas como *variáveis independentes* do modelo.

Na mesma época, Free e Wilson desenvolveram outra abordagem QSAR, a qual é fundamentalmente diferente das abordagens expostas até o presente, pois

não relaciona a atividade biológica a qualquer propriedade físico-química do sistema (FREE e WILSON, 1964). Nela a atividade biológica de um conjunto de análogos pode ser descrita pela contribuição dos substituintes ou de elementos estruturais característicos e em determinadas posições, considerando a estrutura básica da série. Assim, a presença de um grupo (por exemplo o cloro) em uma posição específica de um análogo considerado é expressa como o número 1 (um). Outro análogo que não possua o substituinte na posição considerada, terá o valor desse descritor igual a 0 (zero). Em resumo, um modelo baseado na abordagem Free-Wilson teria a forma:

$$\text{Log } \frac{1}{c} = \sum a_j X_{ij} \quad (5)$$

Onde a_j define a contribuição (coeficiente da regressão) do descritor X_{ij} , o qual por sua vez assume os valores 0 (zero) ou 1 (um) para um determinado grupo em uma dada posição. Um valor positivo do coeficiente de regressão indica que, se o grupo está presente (valor 1), a atividade biológica é aumentada. O valor negativo tem o sentido inverso.

Após estes trabalhos pioneiros, outros estudos foram desenvolvidos na área QSAR, incluindo diversos outros parâmetros (descritores) na modelagem dos sistemas. O desenvolvimento de um modelo QSAR envolve três componentes principais: 1) Um conjunto de dados experimentais de atividades biológicas para uma série congênere de compostos; 2) Dados referentes à estrutura molecular ou propriedades (descritores, variáveis ou preditores) para este grupo de moléculas; 3) Métodos estatísticos para encontrar a relação entre esses dois dados. Os dados biológicos podem ser obtidos da literatura científica ou gerados especificamente para um dado estudo QSAR (GRAMATICA, 2011).

Os descritores utilizados em modelos conhecidos como QSAR clássico incluem parâmetros empíricos (ex. LogP), não empíricos como índices de fragmentos e descritores constitucionais (os quais dependem de conectividades e tipos de átomos na estrutura) bem como químico-quânticos (cargas, valores de energia). Caso haja algum dado experimental ou especulativo sobre o mecanismo de ação, podem-se definir os descritores que conduzirão a um bom modelo, sendo esta técnica baseada na *seleção de descritores a priori*. Por outro lado, ao trabalhar

com compostos bioativos para os quais não é conhecido o mecanismo de ação, ou igualmente existam apenas especulações, pode-se realizar um procedimento que consiste inicialmente na determinação de centenas (ou milhares) de descritores, que representem as mais diversas características moleculares. Estes descritores devem ser então submetidos a um procedimento conhecido como *seleção de variáveis*, o qual selecionará os descritores importantes para um modelo, ou seja, que expliquem a variabilidade apresentada pela variável dependente.

Dentre os programas desenvolvidos para o cálculo de descritores moleculares diversos, um dos mais citados na literatura é o programa DRAGON (TALETE, 2007; GOYAL *et al.*, 2010; BOROTA *et al.*, 2011; GUPTA *et al.*, 2010).

Recentemente, os desenvolvedores deste programa disponibilizaram um recurso *on-line* para o cálculo dos descritores chamado E-DRAGON (TETKO *et al.*, 2005). Utilizando o E-DRAGON, inicialmente carrega-se o arquivo com a molécula previamente construída por um software de modelagem molecular, no formato .sdf. Em seguida podem-se obter diversos descritores moleculares, os quais estão subdivididos em 20 blocos, segundo o tipo. Estes vão de descritores constitucionais como massa molar, número de átomos, número de ligações duplas, triplas, etc.; parâmetros topológicos como a soma das áreas superficiais dos átomos polares; até propriedades moleculares como estimativa do LogP. O conjunto dos descritores possíveis de serem calculados, sua designação e referências podem ser consultados diretamente na página do programa (VCCLAB, 2012). Ao todo podem ser obtidos valores para cerca de 1.600 diferentes descritores.

Uma desvantagem do uso de variáveis de natureza diversa em QSAR é que, não raro, o modelo gerado mesmo após a seleção das variáveis mais importantes (as quais explicam a variabilidade das atividades observadas) pode ser de difícil interpretação qualitativa (entendimento do mecanismo de ação dos compostos), mas podem ser usados para projetar e avaliar moléculas com melhores atividades.

Considerando os métodos matemáticos de construção dos modelos QSAR, pode-se utilizar a Regressão Linear Múltipla, MLR (do inglês, "Multiple Linear Regression"), Regressão por Componentes Principais, PCR (do inglês, "Principal Component Regression") ou a Regressão por Mínimos Quadrados Parciais, PLS (do inglês, "Partial Least Squares"). Estas técnicas serão descritas posteriormente.

Ao longo dos anos, a execução dos estudos QSAR tem incorporado novas características e metodologias para modelar a relação entre a atividade biológica e a estrutura química de um composto. Contudo, independente do tipo de abordagem QSAR, é interessante que um modelo construído tenha relação com o sentido físico do sistema estudado; que seja robusto (capaz de satisfazer procedimentos de validação interna, os quais serão discutidos posteriormente), preditivo e assim permita principalmente incrementar a atividade biológica da classe através da inserção ou modificação de grupos substituintes, orientando a síntese de moléculas potencialmente mais promissoras.

2.1.1 Quimiometria aplicada aos estudos QSAR

A construção de modelos QSAR é dependente de técnicas específicas, as quais estão inseridas campo de estudos que faz interface entre a química e a estatística, denominado *Quimiometria*. Suas ferramentas permitem o processamento e análise de sistemas químicos descritos por muitas variáveis. Nestes sistemas, não raro, surge a necessidade de organizar e compreender uma infinidade de dados gerados por instrumentos cada vez mais modernos, como por exemplo, os dados de absorvância de um espectro de uma amostra (FERREIRA *et al.*, 1999). Uma definição mais formal de Quimiometria pode ser descrita como:

Uma disciplina química que emprega métodos matemáticos e estatísticos para planejar ou selecionar experimentos de forma otimizada e para fornecer o máximo de informação química com a análise dos dados obtidos. (VANDEGINSTE apud FERREIRA *et al.*, 1999)

Outra definição relativamente simples, mas bastante geral foi fornecida por Beebe: *“O processo completo no qual dados são transformados em informação que pode ser usada para tomada de decisão”* (BEEBE, 1998).

Os métodos quimiométricos servem às mais diversas finalidades, podendo ser utilizados para reconhecimento de padrões, como a Análise de Componentes Principais, PCA (do inglês, “Principal Component Analysis”) e a Análise Hierárquica de Agrupamentos, HCA (do inglês, “Hierarchical Cluster Analysis”), as quais se destinam a determinar semelhanças entre amostras descritas por um conjunto de

variáveis (descritores) apropriados (BEEBE, 1998). Outros métodos quimiométricos (SIMCA, KNN) fornecem *modelos de classificação*, os quais são construídos pelo agrupamento de amostras semelhantes e permitem classificar amostras desconhecidas utilizando o modelo construído (BEEBE, 1998). Existem ainda os métodos multivariados que se propõem a construir modelos matemáticos, os quais permitem prever o valor numérico de uma propriedade (ou atividade biológica) de uma substância, relacionando esta propriedade (variável dependente) com descritores apropriados das amostras (variável independente) (BEEBE, 1998). Estes últimos, os quais são chamados *modelos de calibração*, fornecem a base do QSAR e serão discutidos a seguir.

Um dos primeiros modelos de calibração utilizados para análise de dados químicos é a Regressão Linear Múltipla, MLR. Em se tratando de estudos QSAR, a equação de Hansch representa um dos usos pioneiros desta metodologia. O esquema abaixo mostra resumidamente a construção de uma Equação atribuída a Hansch, onde $\log 1/C$ representa a atividade biológica (variável dependente) sendo C a concentração efetiva ou a concentração inibitória mínima para uma dada atividade. b_x são os coeficientes da regressão que são estimados pela MLR. “e” representa o erro ou resíduo entre o valor medido e o valor estimado pelo modelo.

Esquema 2.2: Resumo da Técnica MLR aplicada a QSAR.

$$\log \frac{1}{C} = b_0 + b_1 \sigma + b_2 \log P$$

$$\begin{bmatrix} \log \frac{1}{C} \end{bmatrix} = \begin{bmatrix} 1 & - & - \\ \vdots & \vdots & \vdots \\ 1 & - & - \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} + \begin{bmatrix} \mathbf{e} \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} b_0 + \begin{bmatrix} - \\ \vdots \\ - \end{bmatrix} b_1 + \begin{bmatrix} - \\ \vdots \\ - \end{bmatrix} b_2 + \begin{bmatrix} \mathbf{e} \end{bmatrix}$$

$\mathbf{y} = \mathbf{Xb} + \mathbf{e} = b_0 + b_1 \mathbf{x}_1 + b_2 \mathbf{x}_2 + \mathbf{e} \quad \Rightarrow \quad \hat{\mathbf{b}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}.$

(adaptado de FERREIRA, 2002).

Contudo, com a técnica MLR dois pontos devem ser considerados: Primeiro, é geralmente aceito que a razão compostos/descriptores deve ser maior que cinco (KIRALJ e FERREIRA, 2009); segundo, existe uma considerável chance de aparecerem altas correlações (colinearidade) entre os descriptores, e estas são desvantagens de um modelo MLR. Pode-se tentar resolver estes problemas de duas formas: aplicando um método de seleção de variáveis antes da MLR ou então substituindo esta técnica por calibração multivariada baseada em componentes principais, como em PCR e PLS. Deve-se observar, no entanto, que embora estas últimas técnicas consigam trabalhar com grande número de descriptores altamente correlacionados, existem várias situações onde melhores predições são obtidas quando um subconjunto menor de variáveis é inicialmente selecionado (TEÓFILO, MARTINS e FERREIRA, 2009).

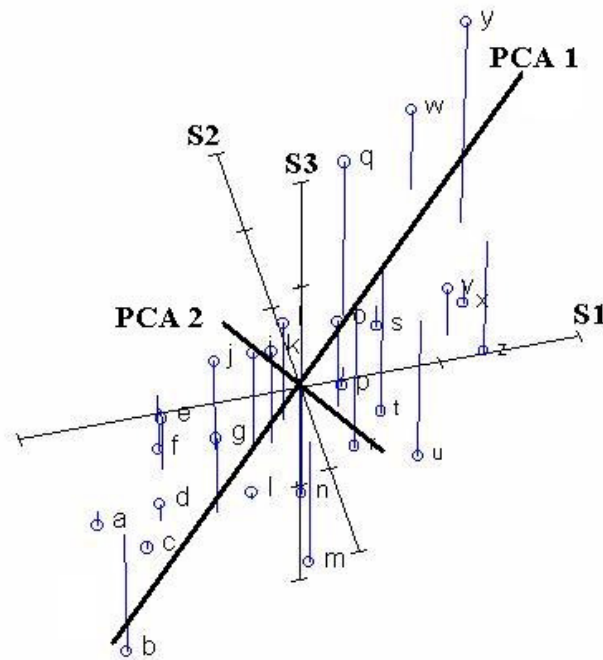
No que concerne aos métodos de regressão PCR e PLS, inicialmente é necessário conhecer os fundamentos da técnica PCA para melhor entendê-los. Esta é uma das metodologias quimiométricas mais utilizadas para o tratamento de dados multivariados, geralmente quando se necessita realizar um reconhecimento de padrões por agrupamento de amostras com propriedades semelhantes. A técnica consiste na substituição das variáveis originais por um novo conjunto, chamado *componentes principais* (PC), sobre os quais os dados são projetados. As PC's representam eixos cartesianos que guardam a maior parte da variabilidade dos dados em um espaço dimensional bem menor. Na prática, o processo de análise PCA parte da matriz dos descriptores, a partir da qual as amostras podem ser projetadas em um espaço multidimensional. A primeira Componente Principal é traçada de forma a incorporar a maior variabilidade dos dados originais projetados, gerando um único eixo e um novo conjunto de coordenadas. A segunda componente principal é traçada sempre perpendicular à primeira e assim sucessivamente. Deste modo, a matriz original de dados $X(I \times J)$, onde "I" é o número de amostras e "J" o número de variáveis sofre a seguinte transformação:

$$X = TL^T \quad (6)$$

O símbolo T denota a *matriz de escores*, representando as posições das amostras no novo conjunto de coordenadas onde as PC's são os eixos e L é a

matriz dos pesos, os quais correspondem à contribuição de cada variável para construção de uma PC. A Figura 2.1 abaixo ilustra o processo de construção de PC's, onde os eixos das variáveis originais correspondem à “Sn”, em três dimensões, reduzidas para duas dimensões:

Figura 2.1: Representação da construção de Componentes Principais.



(Adaptado de: <<http://ordination.okstate.edu/PCA.htm>>)

A Regressão por Componentes Principais PCR utiliza como passo inicial uma PCA, decompondo o conjunto de dados originais em componentes principais. Após este procedimento, realiza um passo de Regressão Linear Múltipla com os dados transformados, gerando um modelo matemático (FERREIRA, 2002).

Considerando que as componentes principais são ortogonais umas às outras e assim não correlacionadas, elas são excelentes regressores para a variável dependente “y” utilizando MLR. A matriz de escores T obtida com as componentes principais é utilizada como variável independente e então se estimam os coeficientes de regressão do modelo, representados por “a” no esquema 2.4. A estimativa dos coeficientes de regressão “ a_x ” de cada *escore* é realizada então pela técnica de mínimos quadrados, como pode ser visualizado no esquema abaixo:

Esquema 2.3: Representação da Técnica PCR e estimativa dos coeficientes de regressão

$$\mathbf{y} = \mathbf{T}\mathbf{a} + \mathbf{e} = a_0 + a_1\mathbf{t}_1 + \dots + a_A\mathbf{t}_A + \mathbf{e} \Rightarrow \hat{\mathbf{a}} = (\mathbf{T}^T\mathbf{T})^{-1} \mathbf{T}^T\mathbf{y}$$

(adaptado de FERREIRA, 2002)

Contudo, a interpretação da PCR só se torna prática se “ a ” puder ser convertida nos coeficientes originais “ b ”:

$$\mathbf{X}\hat{\mathbf{b}} = \mathbf{T}\hat{\mathbf{a}} \quad (7)$$

Como $\mathbf{X} = \mathbf{T}\mathbf{L}^T$, pela regra da pré-multiplicação (isola-se o termo \mathbf{T} multiplicando ambos os lados por $(\mathbf{L}^T)^{-1}$) e sendo \mathbf{L} é uma matriz ortogonal (inversa igual à transposta), tem-se que $\mathbf{X}\mathbf{L} = \mathbf{T}$. Substituindo na equação (7), fica claro que uma possível solução para $\hat{\mathbf{b}}$ é $\mathbf{L}\hat{\mathbf{a}}$ (FERREIRA, 2002). Assim podem-se encontrar valores dos coeficientes de regressão para as variáveis originais.

De forma similar a PCR, a Regressão por Mínimos Quadrados Parciais (“Partial Least Squares”, PLS) tem a capacidade de avaliar dados altamente correlacionados bem como muitas variáveis X (WOLD *et al.*, 2001). A diferença é que, em PLS, o vetor da variável dependente (no caso do QSAR, atividades biológicas) é incluído no processo de decomposição por componentes principais. Deste modo, obtêm-se dois grupos de matrizes de score e pesos e dois blocos de equações:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} = \sum_{a=1}^A \mathbf{t}_a \mathbf{p}_a^t + \mathbf{E} \quad (8)$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Z}^T + \mathbf{F} = \sum_{a=1}^A \mathbf{u}_a \mathbf{z}_a^t + \mathbf{F} \quad (9)$$

Onde T e U representam as matrizes dos escores (t_a e u_a os respectivos vetores coluna), P e Z representam os pesos e E e F a matriz de resíduos. Em PCR, as Componentes Principais são construídas de forma a incorporar a máxima variabilidade dos dados originais, seguindo o esquema da PCA. Em PLS, as Componentes Principais são construídas de forma a obter a melhor relação linear das projeções (escores) das variáveis independentes (que agora passam a se chamar *variáveis latentes*) com os fatores da variável dependente. Deste modo, o modelo final consiste em relacionar a matriz de escores do bloco X (T) com a matriz de escores do bloco Y (U):

$$U = DT + G \quad (10)$$

e assim ao combinar as equações (9) e (10) tem-se a relação:

$$Y = DTZ^T + H \quad (11)$$

Em um trabalho sobre análise multivariada e QSAR, Ferreira (2002) descreve um algoritmo que nos ajuda a entender a regressão PLS, o qual consiste nos seguintes passos:

- 1- Para o cálculo da primeira variável latente, inicia-se com um vetor de escores de y, u. Para uma **variável dependente única** (único y), caso de modelos QSAR, $u = y$. Assim calcula-se o primeiro *fator peso* w (que ainda não é o “peso do escore”):

$$X = yw_1^T + E_x \quad \rightarrow \quad \hat{w}_1 = (y^T y)^{-1} X^T y \quad (12)$$

- 2- Calculam-se os escores (variável latente) de X, ou seja, t, como combinação linear das variáveis originais x e fatores peso como coeficientes:

$$t_1 = X\hat{w}_1 \quad (13)$$

3- Com os escores, calculam-se os pesos (“loadings”) p :

$$\hat{p}_1 = (t_1^T t_1)^{-1} X^T t_1 \quad (14)$$

4- Segue-se a regressão de y no vetor t :

$$y = t_1 q_1 + f \quad (15)$$

Onde o q , o qual representa o primeiro coeficiente de regressão em relação aos escores de X é dado por:

$$\hat{q}_1 = (t_1^T t_1)^{-1} t_1^T y \quad (16)$$

5- Atualiza-se X e y , onde $X = E$ e $y = f$. Assim $E = X - t_1 p_1^T$ e $f = y - t_1 q_1$.

Este processo se repete até o número de fatores desejados. O vetor de regressão (coeficientes de regressão) para as variáveis originais podem ser dados por:

$$\hat{b} = W(P^T W)^{-1} q^t \quad (17)$$

É importante considerar que, independente do método de calibração usado em QSAR, inicialmente deve-se retirar do conjunto de amostras cerca de 20-30% do total, os quais irão compor o que é chamado *conjunto teste*. Como o nome denota, estas amostras serão usadas para testar o modelo. O restante das amostras compõe o *conjunto treinamento* (estas amostras são usadas para construir o modelo). A escolha do conjunto teste deve obedecer a dois critérios principais: incorporar uma diversidade estrutural similar ao conjunto treinamento bem como fornecer previsões externas como interpolações, e não extrapolações (KIRALJ e FERREIRA, 2009).

Ao utilizar técnicas baseadas em componentes principais (PCR ou PLS) deve-se observar que, em geral, os dados da matriz de descritores não têm uma distribuição ótima dos seus valores, possuindo muitas vezes diferentes variâncias e unidades de medida. Deste modo, um conjunto de dados utilizado em análise quimiométrica geralmente requer algum pré-tratamento. Em QSAR o pré-tratamento utilizado é o *autoescalamento*, que corresponde a centrar os valores da variável na média e dividir pelo respectivo desvio padrão (FERREIRA, 2002).

A construção de modelos e a determinação dos parâmetros de validação interna (discutidos posteriormente) são realizadas a partir do conjunto treinamento. A validação interna permite avaliar a robustez do modelo e a capacidade dos descritores explicarem a variabilidade apresentada pela variável dependente. Ao melhor modelo então é aplicado o procedimento de validação externa, avaliando a habilidade preditiva do modelo frente ao conjunto teste.

2.1.1.1 Avaliação dos parâmetros de qualidade do modelo: Procedimentos de validação

A validação de um modelo QSAR requer parâmetros que avaliem a qualidade preditiva e de robustez dos modelos construídos a partir do conjunto treinamento, os quais estão inseridos no contexto da *validação interna*. Não obstante, o modelo deve ser também avaliado utilizando amostras não inseridas na sua construção. Esta etapa, a qual utiliza o já definido conjunto teste, corresponde ao procedimento conhecido como *validação externa*.

Um dos principais métodos de validação interna de modelos QSAR é a *Validação Cruzada*, CV (do inglês, “cross-validation”), utilizando a técnica *leave-one-out*. Este procedimento consiste em retirar uma das amostras (compostos) do conjunto treinamento e construir um modelo de regressão para as amostras que sobram. Com este modelo, realiza-se a previsão do valor da variável dependente (atividade biológica) para a amostra excluída e compara-se com o seu valor verdadeiro. Isto acontece até que todas as amostras tenham sido submetidas ao teste. A partir dos valores de atividade previstos e experimentais podem-se calcular diversos parâmetros de qualidade.

Dentre estes parâmetros está a Soma dos Quadrados dos Erros Residuais de Predição, PRESS (do inglês, “Predicted Residual Error Sum of Squares”), o qual pode ser utilizado na construção de parâmetros derivados. Sua forma é expressa a seguir pela equação 18.

$$PRESS = \sum (y_i^{prev} - y_i^{exp})^2 \quad (18)$$

onde y_i^{prev} é a atividade prevista para a amostra “i” com o modelo construído sem esta amostra e y_i^{exp} corresponde ao valor experimental (verdadeiro) da atividade. É importante destacar que os valores de y_i^{prev} podem ser obtidos a partir do procedimento de validação cruzada (chamado PRESScv) ou então a partir de um modelo construído com todas as amostras do conjunto treinamento (chamado PRESScal, onde o “cal” denota calibração). Quanto menor o seu valor, melhor o modelo.

Outro parâmetro de qualidade, calculado a partir do PRESS, é o RMSEP (do inglês, “Root Mean Square Error of Prediction”), utilizado como estimativa do grau de previsibilidade de um modelo PLS:

$$RMSEP = \sqrt{\frac{PRESS}{n}} \quad (19)$$

onde “n” representa o número de compostos utilizados na construção do modelo. Possui também as versões RMSEcv e RMSEcal

A qualidade estatística dos modelos PLS pode ainda ser atestada através dos parâmetros Coeficiente de Determinação do Modelo, R^2 (obtido na calibração para todo conjunto treinamento) e Coeficiente de Determinação da Validação Cruzada, Q_{loo}^2 :

$$R^2 = 1 - \frac{\sum (y_i^{exp} - y_i^{cal})^2}{\sum (y_i^{exp} - y_i^{med})^2} \quad (20)$$

$$Q_{loo}^2 = 1 - \frac{\sum (y_i^{exp} - y_i^{prev})^2}{\sum (y_i^{exp} - y_i^{med})^2} \quad (21)$$

onde y_i^{cal} corresponde ao valor da variável dependente calculado em um modelo gerado por todas as amostras treinamento, y_i^{prev} aqui denota apenas os valores obtidos com *leave-one-out* e y_i^{med} corresponde à média dos valores de atividade experimentais. Na construção de modelos QSAR, os valores mínimos aceitáveis são $Q_{loo}^2 > 0,5$ e $R^2 > 0,6$ (KIRALJ e FERREIRA, 2009).

Ainda no contexto dos procedimentos de validação interna dos modelos, é prática corrente em QSAR a realização do procedimento *Leave-n-out*. Como o próprio nome sugere, a diferença para o *Leave-one-out* é que mais de uma amostra (“n” amostras”) é retirada; em seguida constrói-se um modelo sem esse bloco de variáveis excluídas e, posteriormente, realiza-se a previsão dos valores da atividade biológica para as amostras do bloco inicialmente retirado. O Q_{lno}^2 é então calculado de forma idêntica ao Q_{loo}^2 . O número de blocos excluídos é dado pela razão M/n , onde M é o número de amostras do conjunto treinamento. Assim, supondo um conjunto treinamento de 50 amostras em um procedimento *Leave-n-out* cujo $n = 2$, o número de blocos de duas amostras retiradas (e por tanto de modelos construídos) será 25. Em geral, recomenda-se utilizar cerca de 30% das amostras do conjunto treinamento como valor de n (KIRALJ e FERREIRA, 2009).

Uma técnica para realização do *Leave-n-out* está implementada no programa *QSAR modeling* (TEÓFILO, MARTINS e FERREIRA, 2009). Nesta técnica se considera inicialmente um valor máximo de “n” (em geral 30% das amostras) e calculam-se valores de Q_{lno}^2 começando com $n = 2$ até o valor máximo estipulado. Assim, para o conjunto de 50 amostras exemplificado, $n_{máx} = 15$. Neste caso teríamos então 14 valores de Q_{lno}^2 . Em adição, para garantir maior confiança na avaliação, recomenda-se realizar dez escolhas aleatórias de M/n blocos para cada n , calculando-se a média dos dez valores Q_{lno}^2 para cada bloco (no nosso exemplo teríamos então 14 médias). O valor de cada média do coeficiente de determinação do *Leave-n-out* deve estar próximo ao Q_{loo}^2 para garantir a estabilidade do modelo (robustez). Em adição, os valores de Q_{lno}^2 médios não devem apresentar desvio maior que 0,1 (KIRALJ e FERREIRA, 2009).

Os parâmetros de qualidade são aplicados a quaisquer tipos de regressão, incluindo PLS e PCR. Conforme relatado anteriormente, nestes casos, as variáveis originais são substituídas por componentes principais, conhecidas como *variáveis latentes*. Assim, um requerimento adicional é definir o número de variáveis latente ótimo para a construção do modelo e isto pode ser feito com base na avaliação de um dos parâmetros acima expostos, em geral o RMSEcv. Contudo, recomenda-se analisar ainda os coeficientes de determinação obtidos, para melhor avaliação.

Na etapa de *validação externa*, utiliza-se o melhor modelo construído para as amostras do conjunto treinamento, o qual passou por todos os testes de validação interna. Com este modelo, realiza-se a previsão dos valores da variável dependente para os compostos do *conjunto teste*, os quais **não foram utilizados na construção do modelo**, e calcula-se o valor de Q_{ext}^2 utilizando equação semelhante à 21. Quanto mais próximo de 1 for o valor de Q_{ext}^2 melhor a capacidade preditiva do modelo proposto.

Outra ferramenta de validação útil em modelos QSAR é a *aleatorização-Y*. Este procedimento consiste em desenvolver modelos paralelos, mantendo os valores dos descritores e permutando os valores das variáveis dependentes entre as amostras. Se realmente existir uma correlação verdadeira, os valores de Q_{loo}^2 e R^2 do modelo original devem ser bem maiores do que os obtidos com os modelos paralelos. Este procedimento descarta a hipótese de correlação ao acaso. Na prática, os modelos com valores de Y permutados não devem possuir R^2 e Q_{loo}^2 maiores que 0,4 (KIRALJ e FERREIRA, 2009).

Na literatura QSAR também se encontra um procedimento de validação interna conhecido como *bootstrapping* (KIRALJ e FERREIRA, 2009). Este também avalia a robustez da relação entre o conjunto de dados. Consiste em dividir aleatoriamente o total de amostras em vários conjuntos treinamento e teste diferentes, sendo então calculados os parâmetros R^2 e Q_{loo}^2 para cada conjunto treinamento. Assim, espera-se que os valores destes parâmetros satisfaçam o mínimo aceitável ($Q_{loo}^2 > 0,5$ e $R^2 > 0,6$), bem como oscilem em torno do valor real (para o modelo original). O objetivo do *bootstrapping* é perturbar o conjunto treinamento, não importando para este procedimento dados estatísticos referentes ao conjunto teste. Diferentes números de divisões (do inglês “splits”) tem sido

relatadas na literatura, variando de dez à centenas (KIRALJ e FERREIRA, 2009). A escolha dos conjuntos, diferentemente da construção do modelo original, não requer necessariamente uma representatividade estrutural dos compostos do conjunto teste em relação aos do conjunto treinamento, sendo feita de forma completamente aleatória. Além disso, diferentemente das validações LOO e LNO (onde uma amostra é excluída apenas uma vez), uma amostra pode ser excluída uma vez, ou várias vezes, ou nunca (KIRALJ e FERREIRA, 2009).

2.3.1.2 Seleção de variáveis pelo método OPS

Dentre os diversos métodos de seleção de variáveis, recentemente foi desenvolvido o denominado método de Seleção dos Preditores Ordenados, OPS (do inglês, “Ordered Predictors Selection”), o qual está implementado no programa QSAR modelling (TEÓFILO, MARTINS e FERREIRA, 2009). Este método baseia-se no uso de um vetor informativo inicial, o qual indica as variáveis independentes (descritores) que possuem melhor relação com as variáveis dependentes. Este vetor informativo pode ser composto pelos coeficientes de correlação da variável dependente com cada variável independente (correlograma). Pode ser também composto pelos próprios coeficientes de regressão, partindo da premissa que baixos coeficientes de regressão denotam pequena importância da variável para o modelo. Ou então pode ser composto por uma avaliação conjunta de diversos vetores informativos.

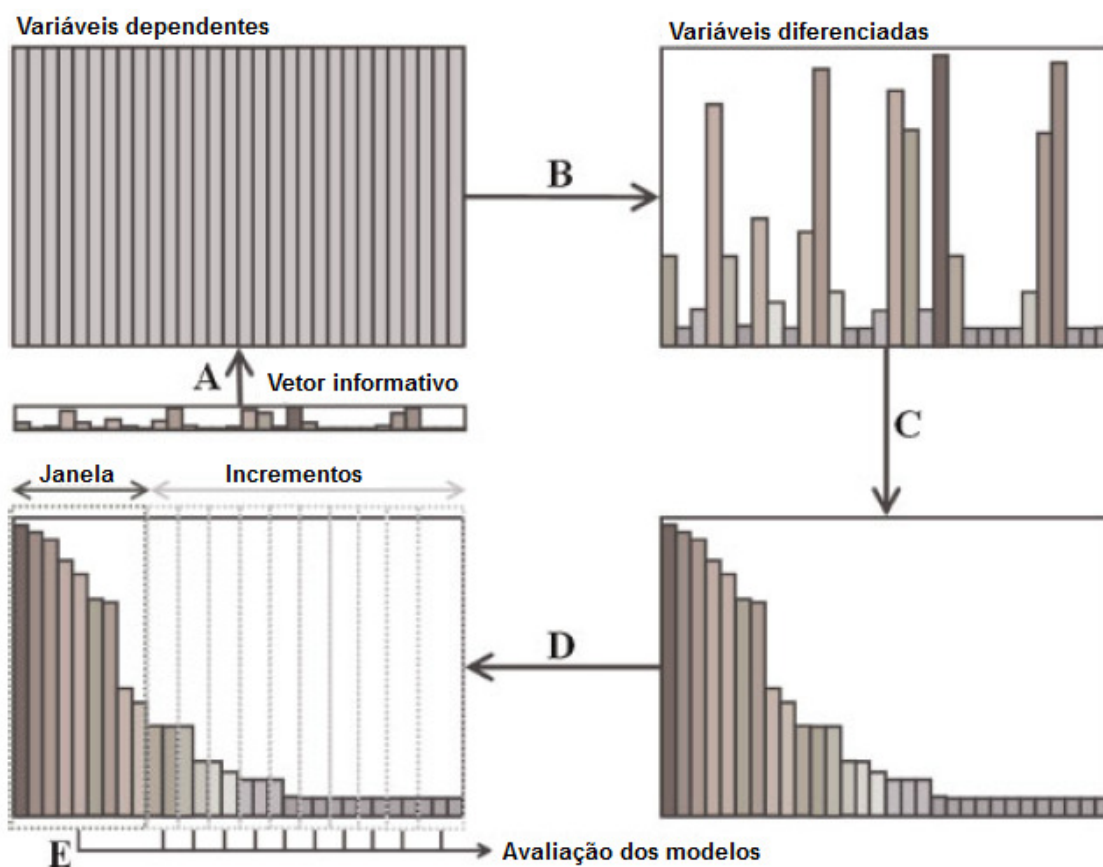
Em um segundo passo do procedimento OPS, a matriz de variáveis independentes X tem suas colunas rearranjadas, de forma a posicioná-las em ordem decrescente do valor absoluto do vetor informativo obtido na etapa anterior (O esquema 2.4 a seguir ilustra e resume os passos do OPS).

No passo “D” do esquema, modelos de regressão multivariada são construídos e avaliados através de validação cruzada. Assim, um subconjunto inicial de variáveis (“janela”) é selecionado para construir e avaliar o *primeiro modelo*. O tamanho desta janela pode ser definido previamente no programa QSAR modelling. Em geral, definem-se janelas maiores para um número maior de variáveis iniciais. Em sequência, a janela é expandida em um número fixo de variáveis (“incremento”), sendo um novo modelo construído e avaliado. Incrementos vão sendo

subsequentemente adicionados até que todas (ou alguma porcentagem) as variáveis sejam incluídas. Os parâmetros da validação cruzada obtidos com os modelos construídos são armazenados e podem ser comparados.

No passo “E” os subconjuntos (janela e incrementos) são comparados usando os parâmetros calculados. Deste modo, o modelo com melhores parâmetros de qualidade deve conter as variáveis com melhor capacidade de previsão, sendo assim as variáveis selecionadas (TEÓFILO, MARTINS e FERREIRA, 2009).

Esquema 2.4: Resumo da Técnica OPS.



Adaptado de (TEÓFILO, MARTINS e FERREIRA, 2009)

2.2 Métodos Computacionais em Química Quântica

As equações da Mecânica Quântica podem ser aplicadas aos sistemas moleculares, descrevendo suas distribuições eletrônicas e permitindo a obtenção de valores de propriedades associadas à estrutura eletrônica. Estas propriedades, por sua vez, podem ser aplicadas aos estudos das relações entre a estrutura química e a atividade biológica de uma série de compostos.

A descrição quântica de uma molécula pode ser feita a partir da resolução da Equação de Schrödinger (SCHRÖDINGER, 1926) a qual pode ser descrita de forma simplificada como:

$$\hat{H}\Psi(\vec{r}, \vec{R}) = E\Psi(\vec{r}, \vec{R}) \quad (22)$$

onde \hat{H} representa o operador *Hamiltoniano* e $\Psi(\vec{r}, \vec{R})$ representa a função de onda molecular, que por sua vez depende das coordenadas dos elétrons e núcleos. Assim, a aplicação do Hamiltoniano à função de onda fornece a energia E do sistema.

O operador Hamiltoniano é dado por:

$$H = T_e + T_n + V_{en} + V_{ee} + V_{nn} \quad (23)$$

onde o primeiro termo corresponde ao operador energia cinética dos elétrons, o segundo referente a energia cinética dos núcleos, o terceiro ao potencial de interação elétron-núcleo, o quarto o potencial elétron-elétron e o quinto ao potencial internuclear.

Em um sistema com muitas partículas, nenhuma destas se move de forma independente das outras, o que torna impossível a descrição exata destes sistemas. À relação de interdependência entre os movimentos eletrônicos e nucleares é dado o nome *correlação* (CRAMER, 2004). Uma das aproximações que se faz à equação (23) é desacopla-los e então calcular a energia eletrônica para coordenadas nucleares fixas. Esta é conhecida como *Aproximação de Born-Oppenheimer* e, a

partir dela, podemos reescrever a equação de Schrödinger *eletrônica* da seguinte forma:

$$(H_{el} + V_N)\Psi_{el}(q_i; q_k) = E_{el}\Psi_{el}(q_i; q_k) \quad (24)$$

Onde H_{el} inclui o primeiro, terceiro e quarto termos da equação 23, V_N representa a energia de repulsão núcleo-núcleo (a qual é constante para um dado conjunto de coordenadas nucleares fixas), q_i representa coordenadas eletrônicas e q_k parâmetros referentes aos núcleos. A partir da aproximação de Born-Oppenheimer se obtém as geometrias moleculares e superfícies de energia potencial. Deste modo a próxima etapa seria resolver a equação de Schrödinger eletrônica para cada configuração nuclear.

Na busca da resolução da Equação de Schrödinger eletrônica, o método *de Hartree* (o qual foi seguido pelas contribuições de Fock e Roothan) representa a base dos métodos conhecidos como *ab initio* (do latim “primeiros princípios”). Estes métodos são assim denominados por descreverem sistemas atômicos e moleculares sem a utilização de parâmetros empíricos. Já os métodos chamados *Semi-empíricos* buscam soluções para a equação de Schrödinger assumindo algumas aproximações integrais bem como o uso de parâmetros empíricos no desenvolvimento dos cálculos.

A principal limitação do Método Hartree-Fock é que este considera as interações intereletrônicas através de um campo médio efetivo. A falta da Correlação Eletrônica explícita pode ocasionar erros no cálculo de energias. A magnitude desse erro associado pode ser importante no tratamento de alguns sistemas, por exemplo, o estudo de curvas de dissociação teóricas, estimativas de pKa teóricos, dentre outros.

2.2.1 Teoria do Funcional da Densidade

Durante o desenvolvimento dos métodos quânticos para tratamento de sistemas químicos, uma alternativa ao uso da função de onda foi considerar a densidade eletrônica. A partir desta ideia, construiu-se o que hoje se conhece como Teoria do Funcional da Densidade, DFT (do inglês, “Density Functional Theory”).

Esta tem sido muito utilizada na descrição teórica de sistemas químicos, por fornecer bons resultados a um custo computacional relativamente baixo.

O uso da densidade eletrônica como variável básica na química teórica remonta ao início do século XX, quando Drude considerou que os elétrons nos metais poderiam ser descritos como um gás homogêneo, desenvolvendo assim sua teoria da condução elétrica e térmica.

Contudo, o uso da densidade eletrônica $\rho(r)$ para a descrição de sistemas atômicos e moleculares foi legitimada a partir dos trabalhos pioneiros de Hohenberg e Kohn (HOHENBERG e KOHN, 1964). Estes forneceram dois teoremas que representam a base do método DFT. O primeiro demonstra que a energia é um funcional da densidade eletrônica:

$$E = E[\rho] \quad (25)$$

O segundo teorema, de forma semelhante ao uso da função de onda molecular, estabelece que o funcional da densidade também obedeça ao *princípio variacional*, ou seja, a energia estimada pela densidade eletrônica aproximada é maior ou igual à energia determinada por um funcional exato.

Cada um dos termos que compõem a energia total podem também ser descritos como funcionais da densidade. Considerando ainda a aproximação de Born-Oppenheimer:

$$E[\rho] = T_e[\rho] + V_{ne}[\rho] + V_{ee}[\rho] \quad (26)$$

onde o primeiro termo após a igualdade corresponde à energia cinética dos elétrons, o segundo o potencial núcleo-elétron (potencial externo) e o terceiro o termo potencial elétron-elétron. Do mesmo modo que o Hamiltoniano eletrônico, o potencial externo relacionado aos elétrons devido às cargas dos núcleos é dado por:

$$\hat{V}_{ne} = \sum_i^N \sum_A^M -\frac{Z_A}{|R_A - r_i|} = \sum_i^N v(r_i) \quad (27)$$

onde Z_A representa a carga nuclear, R_A as coordenadas do núcleo A , r_i as coordenadas do elétron i . Deste modo, a equação 26 pode ser escrita considerando o termo referente ao potencial externo como um funcional da densidade eletrônica (MORGON e COUTINHO, 2007), cuja forma encontra-se na equação abaixo. Assim temos que:

$$E_v[\rho] = \int v(r)\rho(r)dr + F_{HK}[\rho] \quad (28)$$

onde os funcionais da energia cinética dos elétrons e do potencial elétron-elétron são incluídos em um termo único $F_{HK}[\rho]$. Este termo, por ser independente do potencial externo podendo ser aplicado a todos os sistemas eletrônicos, é considerado um funcional universal da densidade. Assim, conhecendo a forma exata deste termo, teríamos a equação exata para a densidade eletrônica do estado fundamental e poderíamos aplicar a qualquer sistema. Contudo, sua forma exata é desconhecida e as tentativas de aproximação é que forneceram as diferentes abordagens DFT.

A DFT ganhou aplicação prática a partir das equações de Kohn e Sham (KOHN e SHAM, 1965). Eles reescreveram a equação para determinação da energia total tornando explícita a repulsão elétron-elétron em termos da interação de Coulomb e definindo uma nova função universal $G[\rho]$:

$$E_v[\rho] = G[\rho] + \frac{1}{2} \iint \frac{\rho(r_1)\rho(r_2)}{|r_1-r_2|} dr_1 dr_2 + \int \rho(r)v(r)dr \quad (29)$$

onde,

$$G[\rho] = T_s[\rho] + E_{xc}[\rho] \quad (30)$$

Na qual $T_s[\rho]$ representa o funcional da energia cinética de um sistema hipotético de elétrons não interagentes. Adicionalmente, assumiu-se que este sistema possui a mesma densidade eletrônica de um sistema de elétrons interagentes. Assim a energia cinética é calculada considerando o modelo de partículas independentes. A parte residual da energia cinética não contemplada com o modelo de elétrons não interagentes estaria inserida no termo $E_{xc}[\rho]$, o qual

ainda inclui os termos troca e correlação. O funcional energia cinética passa então a ser calculado com o mesmo formalismo que o método HF, utilizando o conceito de partículas independentes.

De acordo com Kohn e Sham é possível escrever um Hamiltoniano contendo o termo cinético e um potencial efetivo $v_{ef}(r)$ (KOHN e SHAM, 1965):

$$H^{KS} = -\frac{1}{2}\nabla^2 + v_{ef}(r) \quad (31)$$

Para a obtenção das *funções de Kohn-Sham*, utiliza-se a mesma aproximação empregada no método HF. Deste modo, a função de onda de Kohn-Sham é aproximada por um produto antissimetrizado de funções de um elétron, representado pelo determinante de Slater (SLATER, 1929). Assim temos a seguinte expressão para um elétron:

$$\left(-\frac{1}{2}\nabla^2 + v_{ef}(r)\right)\psi_i^{KS} = \varepsilon_i\psi_i^{KS} \quad (32)$$

A parte da energia cinética é calculada através de um procedimento autoconsistente:

$$T_s[\rho] = \sum_i^N \langle \psi_i^{KS} | -\frac{1}{2}\nabla^2 | \psi_i^{KS} \rangle \quad (33)$$

O potencial efetivo $v_{ef}(r)$ é dado pela expressão:

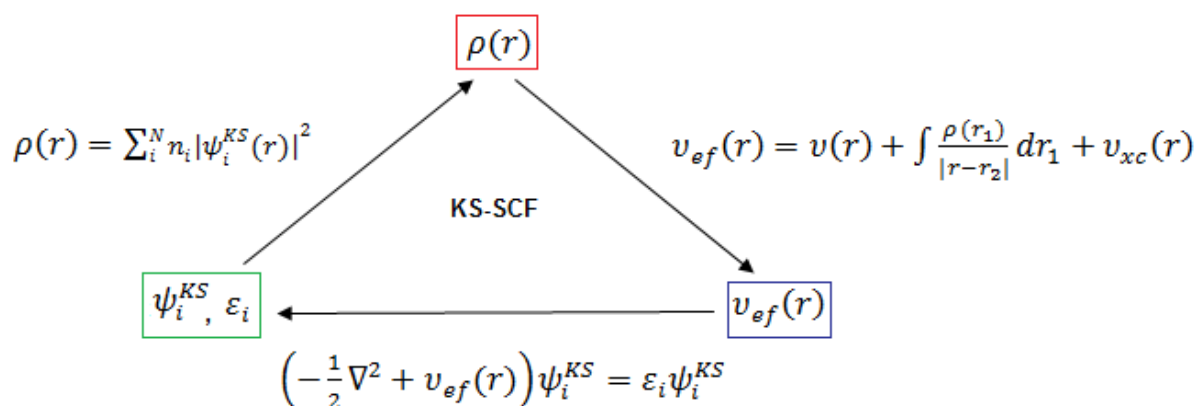
$$v_{ef}(r) = v(r) + \int \frac{\rho(r_1)}{|r-r_1|} dr_1 + v_{xc}(r) \quad (34)$$

Cuja densidade eletrônica a ser utilizada é definida a partir das funções de onda de Kohn-Sham encontradas pela resolução do funcional da energia cinética. O potencial de troca e correlação é definido como a derivada da energia de troca e correlação:

$$v_{xc}(r) = \frac{\partial E_{xc}[\rho]}{\partial \rho(r)} \tag{35}$$

Assim, a resolução das equações de Kohn-Sham é realizada por um procedimento autoconsistente, resumido no esquema 2.5:

Esquema 2.5: Procedimento autoconsistente para as equações de Kohn-Sham.



(Adaptado de MORGON e COUTINHO, 2007)

O potencial de troca e correlação é o termo que guarda a principal problemática da DFT. A tentativa de se obter uma representação para o mesmo é o que diferencia os vários métodos propostos. Inicialmente, cabe destacar que estes métodos separam o funcional de troca e correlação em dois termos, o primeiro correspondente a parcela de troca e o segundo de correlação.

A aproximação da densidade local LDA (do inglês “Local Density Aproximation”) faz parte da primeira geração de aproximações para o termo troca e correlação. Baseia-se na teoria de um gás de elétrons homogêneo onde a energia E_{xc} pode ser calculada para um determinado ponto no espaço conhecendo-se apenas a densidade eletrônica, sem levar em consideração o gradiente. Este modelo traz sérios problemas quando aplicado a sistemas reais, pois não considera situações onde a densidade não é uniforme. Deste modo, outros funcionais foram desenvolvidos a partir de modelos como a Aproximação do Gradiente Generalizado GGA (“Generalized Gradient Aproximation”) e os Métodos Híbridos.

A GGA introduz o gradiente da densidade eletrônica na sua construção. Os funcionais mais populares nesta aproximação correspondem ao funcional de troca de Perdew (PERDEW e WANG, 1986) e de Becke (“B” ou “B88”) (BECKE, 1988). Como exemplos dos funcionais de correlação, temos o de Lee, Yang e Parr “LYP” (LEE, YANG e PARR, 1988).

Os funcionais híbridos, por sua vez, contém termos de troca e correlação GGA incluindo uma contribuição do método Hartree-Fock. Dentre os funcionais híbridos mais utilizados encontra-se o B3LYP, o qual incorpora o termo de troca desenvolvido por Becke, o termo correlação de Lee, Yang e Parr e ainda um funcional de correlação desenvolvido por Vosko, Wilk e Nusair (VOSKO, WILK e NUSAIR, 1980). Ainda possui três parâmetros empíricos (a, b, c), que podem ser ajustáveis.

$$E_{XC}^{B3LYP} = (1 - a)E_X^{LDA} + aE_X^{HF} + b\Delta E_X^{B88} + (1 - c)E_C^{VWN} + cE_C^{LYP} \quad (36)$$

2.3 Modelo do Contínuo Polarizável: Inclusão dos efeitos do solvente

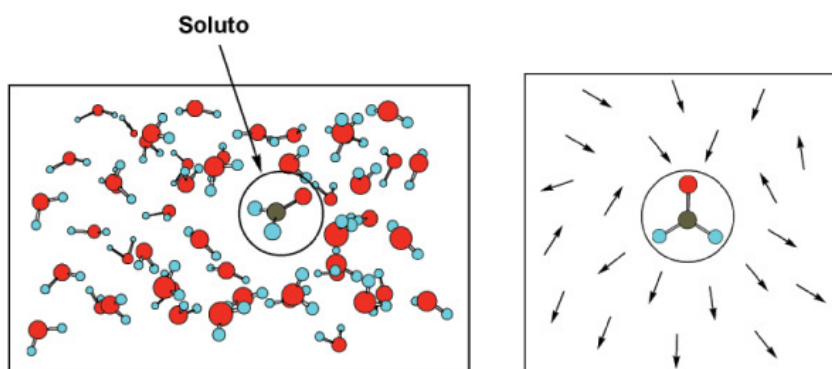
Devido à sua estrutura química e propriedades peculiares, a água é considerada um solvente universal. De fato, grande parte da química ocorre no ambiente aquoso, incluindo os processos que se passam no meio biológico, predominantemente composto por água. Em adição, a presença do solvente pode alterar características intrínsecas de uma molécula como Ligações de Hidrogênio Intramoleculares, as quais tendem a ser enfraquecidas ao serem submetidas a influência do solvente aquoso. Deste modo, nesta Tese consideramos que a simulação do ambiente aquoso poderia ser interessante na obtenção das geometrias otimizadas dos Adutos de Morita-Baylis-Hillman.

Considerando os cálculos em química teórica, existem vários modelos que se propõem a incluir os efeitos do solvente aquoso. Um desses é baseado em um modelo onde o solvente é representado como um contínuo dielétrico envolvendo o soluto, chamado de Modelo de Solvatação Contínua, ou implícito. Trata-se de um método bastante popular em química teórica devido à sua facilidade de utilização,

disponibilidade e seu relativo sucesso, mesmo possuindo algumas limitações. Estas se dão basicamente pelo fato de estes métodos não incluírem explicitamente as interações entre o solvente e os grupos funcionais do sistema molecular de interesse. Contudo, os métodos contínuos podem representar uma poderosa ferramenta no entendimento do comportamento molecular sujeito às influências de um solvente, cuja contribuição eletrostática é usualmente a mais importante (PLIEGO, 2006).

A obtenção de um contínuo dielétrico é realizada a partir da idealização que moléculas do solvente podem ser substituídas por dipolos pontuais ao redor da molécula solvatada. Cada molécula do solvente possuiria então um momento de dipolo associado (Figura 2.2). A contribuição eletrostática para a energia de solvatação, utilizada para medir a interação soluto-solvente, surgiria da interação do dipolo do solvente com a distribuição de carga do soluto (BEN-NAIM e MARCUS, 1984). Assim, a molécula estaria solvatada por vários dipolos.

Figura 2.2: Representação do formaldeído solvatado por moléculas de água e seus dipolos.



(adaptado de PLIEGO, 2006)

Um vetor de polarização \vec{P} pode ser definido a partir de um elemento de volume ΔV e do vetor momento dipolar representado por uma média $\langle \vec{\mu}_i \rangle$:

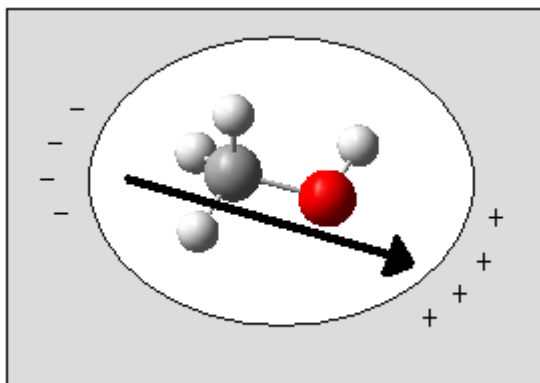
$$\vec{P} = \frac{1}{\Delta V} \sum_{i=1}^n \langle \vec{\mu}_i \rangle \quad (37)$$

O vetor de polarização, deste modo, fornece a ideia da substituição dos dipolos discretos por um campo vetorial contínuo.

Nos modelos contínuos, o solvente é tratado como um material dielétrico, caracterizado principalmente pela sua *constante dielétrica*. Esta representa a resistência de um material à passagem da corrente elétrica, propriedade bastante explorada em capacitores. Assim, quanto maior a constante dielétrica de um material significa que este tem uma maior tendência à polarização.

Nos trabalhos de Onsager (ONSAGER, 1936) foi introduzida a idéia de *campo de reação*. Neste contexto, um dielétrico é polarizado pelo campo elétrico exercido por um dipolo molecular μ , o qual é gerado pela **distribuição de cargas de uma molécula** inserida em uma cavidade na região do dielétrico (Figura 2.3).

Figura 2.3: Representação da influência do dipolo molecular polarizando o dielétrico (cinza) que, por sua vez, produz um campo de reação atuando de volta no soluto.



Considerando que esta cavidade esférica tem um raio a , a distribuição de cargas no dielétrico de constante dielétrica ϵ , a qual resulta dessa polarização, dá origem ao campo de reação:

$$R = \frac{2(\epsilon - 1) \mu}{2\epsilon + 1 a^3} \quad (38)$$

Este campo é paralelo ao momento dipolo e assim atuará de volta no soluto aumentando sua distribuição de cargas.

Uma característica favorável ao uso de modelos contínuos é que estes permitem um tratamento quântico das interações soluto-solvente (MORGON e COUTINHO, 2007). Incorporando o campo de reação, por exemplo, à aproximação de Hartree-Fock, tem-se:

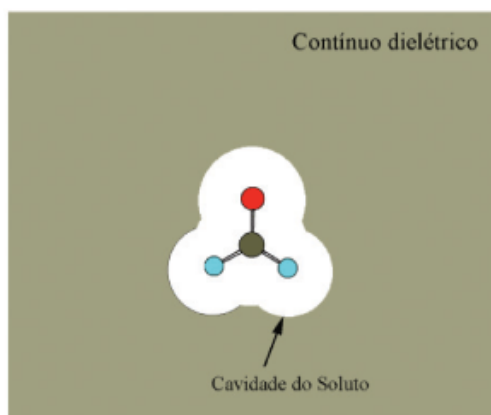
$$\left\{ F_i - \left[\frac{2(\varepsilon - 1)}{2\varepsilon + 1} \right] \frac{1}{a^3} \langle \psi | \mu | \psi \rangle \right\} \chi_i = e_i \chi_i \quad (39)$$

Em que χ_i corresponde aos orbitais moleculares. O termo adicional ao operador de Fock depende do momento dipolo do soluto, que por sua vez depende da função de onda total ψ . Assim tem-se um ciclo adicional autoconsistente, o qual é chamado comumente *Campo de Reação Autoconsistente*, SCRF (do inglês “Self Consistent Reaction Field”).

Um problema que pode ser relacionado aos métodos contínuos é a escolha da forma e tamanho da cavidade para inserção do soluto. Observando a equação 32 vemos que o campo de reação depende do inverso do cubo do raio, sendo assim influenciado diretamente por este. Uma das maneiras para contornar esta situação é fazer com que a cavidade seja formada pela união de esferas centradas nos átomos do soluto. Um dos métodos contínuos mais populares que usam este modelo de cavidade é o Modelo do Contínuo Polarizável, PCM (do inglês “Polarizable Continuum Model”), o qual é oriundo dos trabalhos de Tomasi e colaboradores (TOMASI et al., 1981). Neste modelo, utilizam-se os raios de Van der Waals de cada átomo acrescidos de 20%, formando assim um conjunto de esferas sobrepostas que definem a cavidade (Figura 2.4).

Este modelo tem sido usado, sobretudo, em análises que não requerem interações específicas soluto-solvente, como é o caso de otimização de geometria de moléculas neutras e estudos conformacionais.

Figura 2.4: Cavidade do soluto definida por esferas centradas nos átomos.



(Adaptado de PLIEGO, 2006).

2.4 Ligações de Hidrogênio e Teoria Quântica de Átomos em Moléculas

As Ligações de Hidrogênio (LH) são interações mais fracas que as interações covalentes comuns e ocorrem nas fases sólida, líquida ou gasosa. São compreendidas como sistemas do tipo $XH\cdots YZ$, onde X e Y são átomos de eletronegatividade maior que o hidrogênio (ARUNAN *et al.*, 2011). O grupo X-H é chamado “doador de próton” ou “receptor de elétrons”, enquanto Y é chamado “doador de elétrons” ou “receptor de próton”. Assim, o átomo eletronegativo X atrai elétron da nuvem eletrônica do átomo de hidrogênio, que fica com carga parcialmente positiva, atraindo o par de elétrons do átomo Y. Estas interações, em geral, são consideradas como não-covalentes e descritas basicamente por potenciais do tipo eletrostático. Estão inseridas no grupo das interações não ligantes como as interações de London, forças de Van der Waals e interações iônicas perdendo, em força, apenas para estas últimas. A literatura científica relata ainda a caracterização de ligações de hidrogênio do tipo *não clássicas*, envolvendo outras espécies doadoras de prótons, como C-H, bem como centros de carga π como receptores (TAKAHASHI, SAITO, e KOHNO, 2003).

Na década de 1930, Linus Pauling publicou um trabalho sobre a natureza da ligação química (PAULING, 1939), representando as primeiras bases consistentes para compreender esta interação. Em seu livro “Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry”, afirma que se a distância entre dois átomos interagentes for menor, ou pelo menos igual, à soma dos seus raios de Van der Waals, os átomos estão quimicamente ligados (PAULING, 1939; ROWLAND e TAYLOR, 1996). Este critério de caracterização de uma ligação tem sido amplamente respeitado ao longo do tempo, sendo inclusive utilizado nos estudos sobre ligação de hidrogênio inter ou intramoleculares (OLIVEIRA, ARAÚJO, RAMOS, 2010).

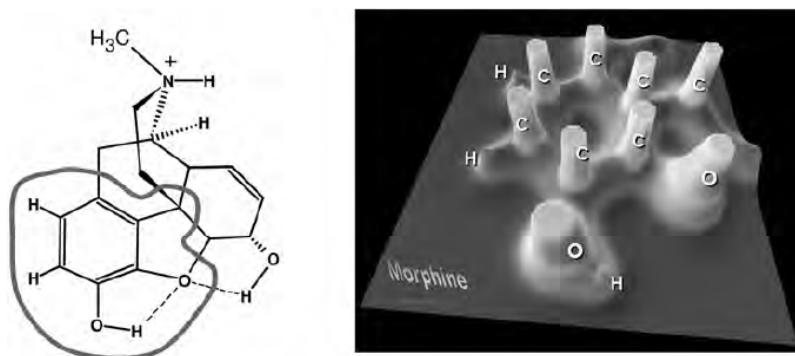
Em 1990, com a publicação do livro “Atoms in Molecules: A Quantum Theory”, Richard Bader estabeleceu o marco de uma nova compreensão da estrutura eletrônica para os sistemas químicos (BADER, 1990). Como sugere o título do livro, esta teoria ficou conhecida como Teoria Quântica de Átomos em Moléculas (“Quantum Theory of Atoms in Molecules”, QTAIM), a qual tem se mostrado eficiente na caracterização de Ligações de Hidrogênio intra e intermoleculares, sendo

inclusive um dos critérios recomendados pela IUPAC para este fim (HUANG *et al.*, 2010; ARUNAN *et al.*, 2011).

A QTAIM considera que a densidade eletrônica possui um papel fundamental na descrição e interpretação das observações experimentais em química. Apesar desta ideia já se encontrar inserida em outras concepções, como na DFT, a essência da QTAIM reside na descrição da topologia molecular, determinada pelo vetor gradiente da densidade eletrônica $\nabla(\rho)$ em cada ponto ocupado por esta.

A topologia da densidade eletrônica é dominada pelas forças atrativas nucleares e, como consequência, os máximos da densidade eletrônica situam-se nas regiões dos núcleos. Seguindo este modelo, um átomo pode ser visualizado como uma região do espaço limitada pelo equilíbrio de forças que os núcleos vizinhos exercem sobre a densidade eletrônica. A Figura 2.5 representa o mapa de relevo da densidade eletrônica para uma região da molécula da morfina.

Figura 2.5: Mapa de relevo para região da morfina.



(adaptado de MATTA e BOYD, 2007)

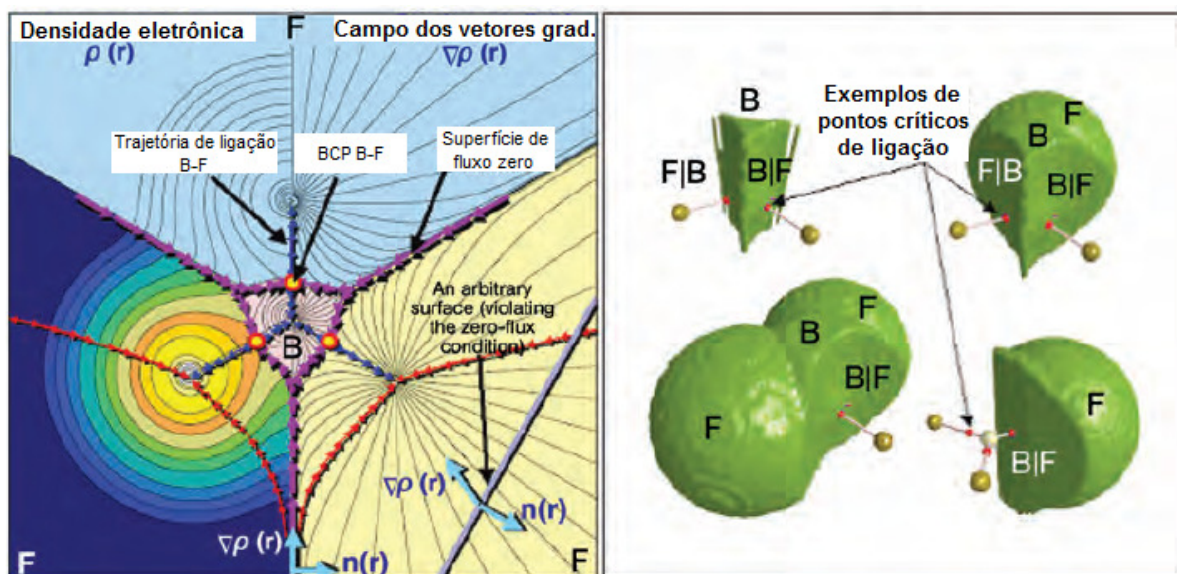
Deste modo, um campo para a densidade eletrônica, que contemple sua noção topológica, deve ser traçado partindo-se dos vetores gradiente $\nabla(\rho)$ para a densidade, os quais definem suas trajetórias ou linhas de contorno. O gradiente de uma função escalar, tal como $\rho(r)$, em um ponto no espaço, é um vetor que aponta na direção em que a densidade aumenta, possuindo uma magnitude proporcional ao aumento nesta direção. Estas regiões para onde a densidade eletrônica converge são chamadas *atratores* e correspondem aos núcleos atômicos.

A ideia de topologia incorpora um particionamento da estrutura molecular em **regiões mononucleares separadas**, identificadas como “átomos nas moléculas”. A superfície que contorna cada região está no fluxo zero do campo dos vetores

gradiente da densidade eletrônica, coincidindo nesta descrição com a fronteira dos átomos.

Além da definição de regiões delimitadas pelo campo de vetores gradiente e respectivos atratores, os quais fornecem os limites dos átomos, a QTAIM descreve a formação das ligações químicas. Entre dois átomos vizinhos pode ser traçada uma linha ligando seus núcleos, a qual representa um máximo local na densidade eletrônica e é chamada *Trajétória de Ligação* (“Bond Path”, BP). O *Ponto Crítico de Ligação* (“Bond Critical Point”, BCP), representa o ponto onde a trajetória de ligação intercepta a superfície de fluxo zero, a qual separa os dois átomos ligados (Figura 2.6).

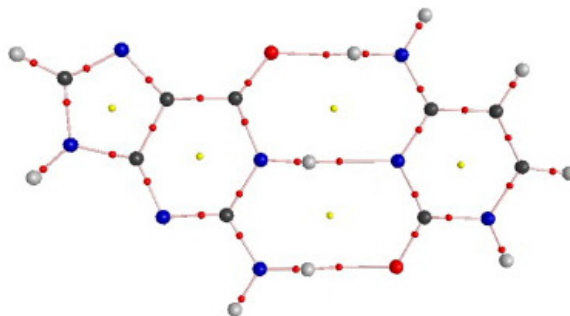
Figura 2.6: Representação QTAIM da molécula BF_3 .



(adaptado de MATTA e BOYD, 2007)

Deste modo, uma coleção de trajetórias de ligação definidas ligando núcleos atômicos em uma geometria de equilíbrio, com seus pontos críticos associados, é conhecida como *gráfico molecular* (Figura 2.7).

Figura 2.7: Gráfico molecular obtido por QTAIM para as bases nitrogenadas Guanina e Citosina, interagindo por ligações de hidrogênio.



(adaptado EBRAHIMI et. al, 2009)

Pontos críticos na topologia da densidade eletrônica são locais onde a densidade eletrônica é um extremo, máximo ou mínimo. Deste modo, a caracterização do Ponto Crítico de Ligação é feita através do Laplaciano da densidade eletrônica $\nabla^2\rho(r)$, cuja expressão é dada pela derivada do gradiente:

$$\nabla^2\rho(r) = \frac{\delta^2\rho}{\delta^2x} + \frac{\delta^2\rho}{\delta^2y} + \frac{\delta^2\rho}{\delta^2z} \quad (40)$$

A expressão para o laplaciano fornece informações importantes sobre a característica da ligação química, onde o sinal do laplaciano pode descrever a concentração de carga eletrônica correspondente à ligação. Segundo o Teorema Virial para a densidade eletrônica (BADER, 1972), podemos relacionar o Laplaciano às parcelas cinética (T) e potencial (V) da densidade eletrônica:

$$\nabla^2\rho(r) = 2T(r) + V(r) \quad (41)$$

Deste modo, considerando a contribuição potencial como um valor sempre negativo e a energia cinética positiva, valores $\nabla^2\rho(r) < 0$ representariam sobreposição da energia potencial, significando concentração de carga entre os núcleos. Esta é uma característica geral das ligações covalentes e insaturadas. Já valores $\nabla^2\rho(r) > 0$ representam preponderância do termo energia cinética e,

portanto, depleção de densidade eletrônica ao nível do BCP, característica de ligações de hidrogênio intra ou intermoleculares.

Além de $\nabla^2\rho(r)$, pode ser observado que o valor da densidade eletrônica ao nível do BCP para a ligação de hidrogênio é relativamente menor que em uma ligação covalente comum. Para esta última apresenta valores decimais, enquanto que para a LH apresenta valores centesimais.

OBJETIVOS

3. OBJETIVOS

3.1 Geral

Construir modelos quantitativos da relação entre a estrutura química e a atividade Leishmanicida 32 Adutos de Morita-Baylis-Hillman (QSAR), os quais poderão ser utilizados para otimizar a síntese de compostos potencialmente mais bioativos.

3.2 Específicos

- Buscar o mínimo conformacional de cada AMBH leishmanicida a ser usado no QSAR, utilizando o método RPES (do inglês, “Relaxed Potential Energy Surface Scan”), seguido de otimização completa em DFT;
- Elucidar características estruturais que possam estar relacionadas às atividades apresentadas pela classe de compostos estudada, como ligações de hidrogênio intramoleculares, presença de grupos funcionais importantes, e outras;
- Caracterizar ligações de hidrogênio intramoleculares através de parâmetro estrutural (distância doador-aceptor) e cálculos QTAIM;
- Comparar os resultados com dados espectroscópicos experimentais previamente relatados por nosso grupo de pesquisas;
- Construir modelos de Relação Quantitativa Estrutura-Atividade a partir das estruturas modeladas;
- Utilizar os modelos QSAR obtidos para proposição de novas moléculas potencialmente mais promissoras, as quais poderão ser sintetizadas e avaliadas por nosso grupo de pesquisas.

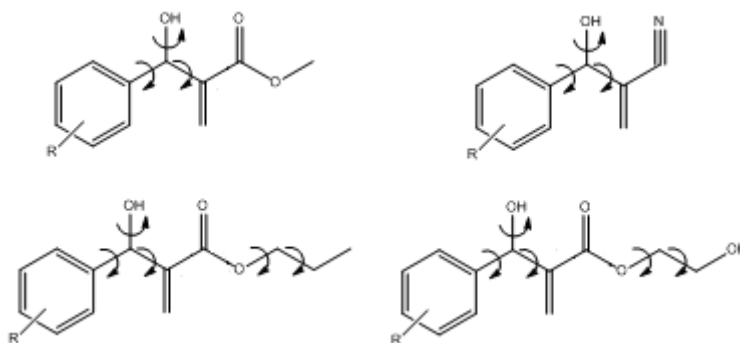
PROCEDIMENTOS COMPUTACIONAIS

4. PROCEDIMENTOS COMPUTACIONAIS

4.1 Estudo conformacional e obtenção dos descritores escolhidos *a priori*

Para obtenção das conformações de mínimo a serem utilizadas nas modelagens QSAR, curvas de energia potencial foram calculadas para cada molécula através de um procedimento implementado no software Gaussian 09W® (FRISCH *et al.*, 2009). Neste procedimento, denominado “Scan” Relaxado na Superfície de Energia Potencial, RPES (‘‘Relaxed Potential Energy Surface Scan’’), inicialmente selecionam-se graus de liberdade rotacionais (ligações simples) considerados importantes. Estes graus de liberdade são representados pelo ângulo diedral envolvendo quatro átomos (dois do eixo de ligação e dois imediatamente vizinhos). Deste modo, durante a efetivação do RPES, o ângulo diedral foi variado em passos de 10 graus até 360 onde, a cada passo, a porção restante da molécula teve sua geometria otimizada usando método semi-empírico AM1 (‘‘Austin Model One’’) (DEWAR *et al.*, 1985) (Figura 4.1).

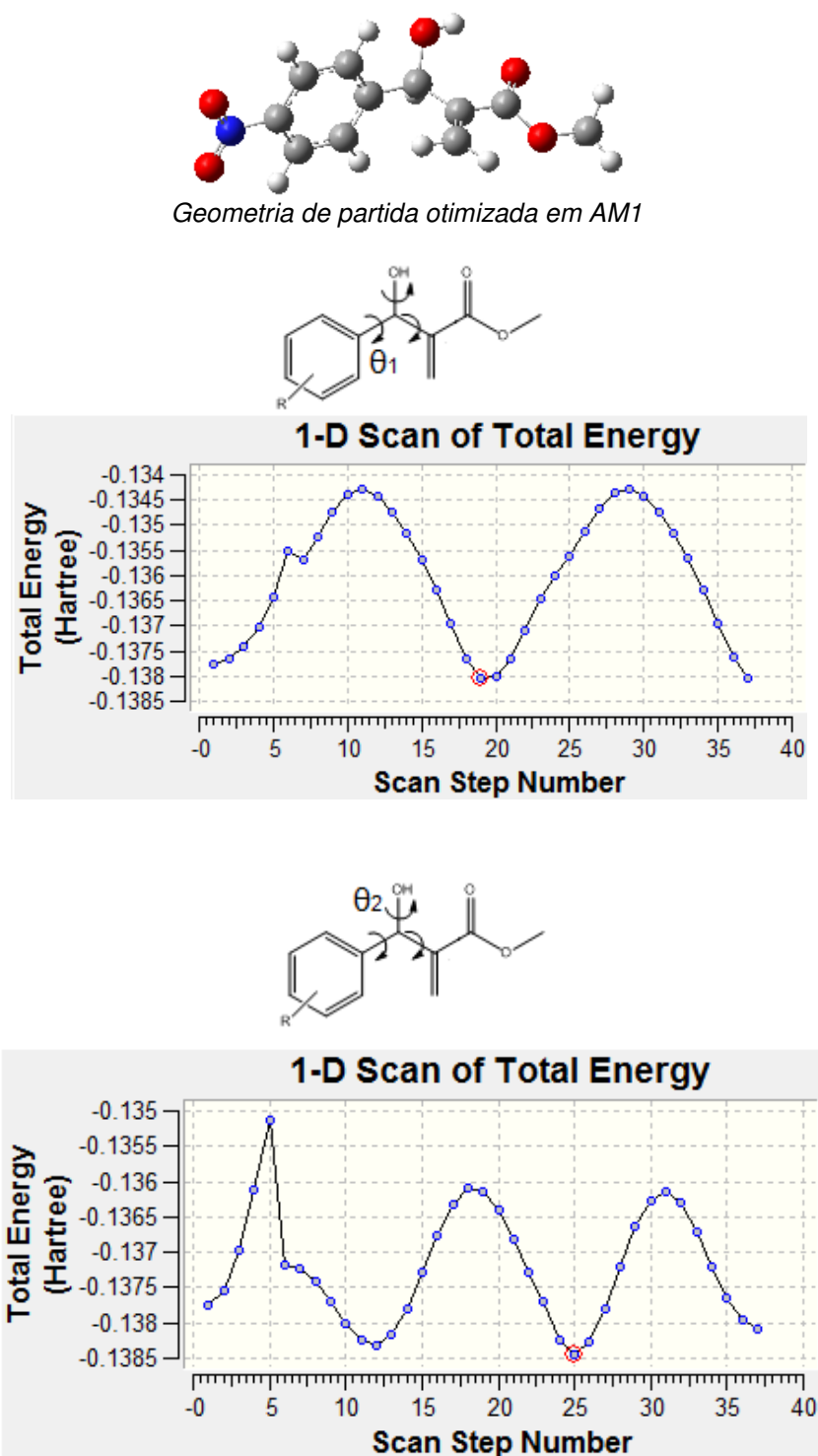
Figura 4.1: Ângulos diedrais investigados para cada estrutura base dos AMBH estudados.

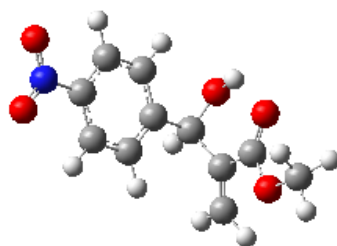
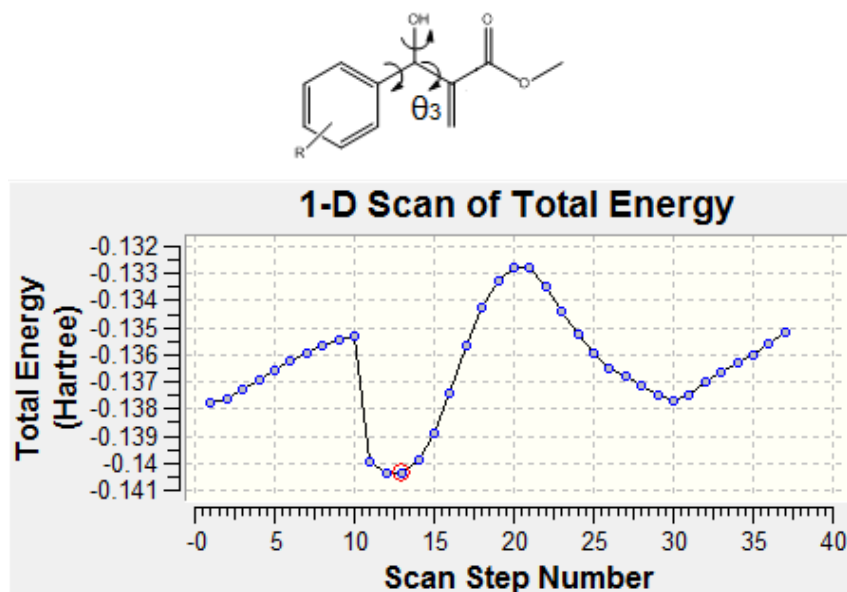


Antes da realização do RPES, a geometria de cada AMBH foi construída com auxílio do programa GaussView5® (DENINGTON, 2009). Nesta etapa buscou-se contemplar duas características consideradas importantes: Ligações de Hidrogênio Intramoleculares OH[⋯]OC bem como a tendência planar da porção éster α,β insaturada. Estas geometrias foram otimizadas usando o método semi-empírico AM1 (‘‘Austin Model One’’) (DEWAR *et al.*, 1985). Em seguida o procedimento RPES fora efetivado. Ao final deste procedimento para cada ligação selecionada, foi visualizada sua respectiva curva através do programa GaussView5®

(DENINGTON, 2009). Para demonstrar a realização deste procedimento, a Figura 4.2 apresenta os gráficos obtidos para cada grau de liberdade considerado na molécula **1a**. Estes gráficos permitiram a observação das conformações de maior e menor energia obtidas.

Figura 4.2: Gráficos obtido após o procedimento RPES para **1a**.





Mínimo conformacional mais estável após RPESS (encontrado no scan em θ_3)

A partir das curvas RPESS para cada molécula, foi selecionado o mínimo conformacional de menor energia. Este foi então submetido a cálculo de otimização de geometria e determinação de frequências vibracionais no nível DFT-B3LYP/6-31+G(d), utilizando método PCM para simulação da influência do ambiente aquoso. Valores de cargas NBO (“Natural Bond Orbital”) e energias dos orbitais LUKO (“Lowest Unoccupied Kohn-Sham Orbitals”) foram obtidos destes cálculos. A visualização das estruturas otimizadas, bem como da localização dos orbitais LUKO foram obtidas através do programa GaussView5® (DENINGTON, 2009).

Os cálculos QTAIM, utilizados especificamente para caracterização de ligações de hidrogênio, foram realizados com auxílio dos programas MORPHY 98 (POPELIER, 1998) e AIM 2000. A partir de cada geometria otimizada em DFT foi realizado um cálculo “single point” B3LYP/6-31+G(d) para obtenção do arquivo .wfn através do programa Gaussian 09W®. Em seguida este arquivo foi submetido ao cálculo QTAIM para geração da topologia da densidade eletrônica, determinação

dos Pontos Críticos de Ligação e parâmetros topológicos associados, no programa MORPHY 98. Os gráficos moleculares foram obtidos no programa AIM 2000.

Cálculos PCM-B3LYP/6-311++G(d,p), realizados para investigar outras possibilidades de formação de Ligações de Hidrogênio Intramoleculares, foram igualmente realizados no programa Gaussian 09W®.

Os valores de SLogP para cada molécula foram obtidos a partir do programa MedChem Designer™ 1.0 (SIMULATION PLUS).

4.2 Construções de modelos QSAR com descritores escolhidos *a priori*

O primeiro modelo QSAR, baseado na técnica PLS, foi concebido utilizando descritores escolhidos *a priori*. Estes descritores foram evidenciados como sendo importantes para a variabilidade apresentada nos valores de atividade biológica. Deste modo, o modelo foi construído com auxílio do programa QSAR Modeling (TEÓFILO, MARTINS e FERREIRA, 2009). Antes das análises quimiométricas, os descritores moleculares foram submetidos a um pré-processamento utilizando o *autoescalamento*, a fim de fornecer aos descritores a mesma importância nas análises. A atividade biológica também sofreu um pré-processamento, sendo convertida em $-\text{LogIC}_{50}$ (pIC_{50}) a partir dos valores originais em M (molar). O cálculo do “negativo” do logaritmo da variável dependente teve como objetivo reduzir o desvio padrão dos dados originais e tornar o valor numérico positivo. Esta transformação também trouxe a conveniência de maiores valores numéricos passarem a corresponder a melhores atividades.

A divisão dos compostos nos conjuntos treinamento e teste se deu considerando 25 compostos para o primeiro e 7 para o segundo, obedecendo um critério de 20-25% para o conjunto teste. Os compostos deste conjunto foram selecionados de forma que representassem diferentes valores de atividade (como interpolações) e padrões estruturais diversos, de forma a incorporar a maior variabilidade possível do conjunto de dados original. Deste modo, foram escolhidos os compostos **5c**, **9e**, **13g**, **20j**, **22k**, **25m**, e **28n** (Figura 1.4).

O número ótimo de variáveis latentes e, conseqüentemente, o melhor modelo, foi escolhido inicialmente com base no menor valor de RMSEcv, seguido da análise dos demais parâmetros de validação interna R^2 , Q_{loo}^2 e Q_{lno}^2 .

O procedimento de validação cruzada “leave-n-out” foi realizado com “n” assumindo valores de 2-8, segundo metodologia disponível no programa QSAR Modeling (discutido na Fundamentação Teórica). Foram realizadas 10 repetições para cada n no cálculo do Q_{ino}^2 .

Para a validação externa e determinação do Q_{ext}^2 foi utilizado o programa Microsoft Excel 2007®.

O procedimento *bootstrapping* foi realizado considerando vinte divisões (“splits”) aleatórias. Os parâmetros de qualidade avaliados nesta etapa foram R^2 e Q_{loo}^2 . O valor de Q_{ext}^2 também foi determinado para comparação com o primeiro modelo obtido.

O segundo modelo QSAR para os descritores *a priori*, obtido através da técnica de MLR, foi construído com o programa Build QSAR (DE OLIVEIRA e GAUDIO, 2000). A definição dos conjuntos treinamento e teste, bem como dos vinte conjuntos do *bootstrapping* foram realizadas de forma idêntica ao modelo PLS. Os parâmetros utilizados para comparação com o modelo PLS foram R^2 , Q_{loo}^2 e Q_{ext}^2 .

4.3 Construções de modelos QSAR a partir de descritores moleculares diversos

Através da plataforma *online* E-DRAGON (TETKO *et al.*, 2005) foram calculados centenas de descritores de natureza diversa, objetivando a busca de outro modelo QSAR. Estes descritores correspondem a parâmetros constitucionais como massa molar, número de átomos, número de ligações duplas, triplas, etc.; parâmetros topológicos como a soma das áreas superficiais dos átomos polares; parâmetros empíricos estimados como LogP. Deste modo, os arquivos de saída dos cálculos PCM/DFT foram inicialmente convertidos para a extensão “sdf”, utilizando o programa OpenBabel. Em seguida, foram carregados na página do E-DRAGON para obtenção dos descritores. As matrizes de descritores obtidas para cada molécula foram salvas no formato “txt” e em seguida utilizadas para a composição da matriz geral dos descritores, a qual foi submetida à seleção de variáveis.

4.3.1 Procedimento de Seleção de Variáveis

O conjunto de centenas de variáveis obtidas com o E-DRAGON foi submetido a uma seleção inicial utilizando o procedimento OPS, implementado no programa QSAR modeling (TEÓFILO, MARTINS e FERREIRA, 2009).

Deste modo, foram definidos o conjunto treinamento e teste utilizando a mesma divisão inicial dos modelos anteriores. Os seguintes parâmetros foram escolhidos na execução do OPS (Figura 4.3) o qual foi aplicado ao conjunto treinamento:

Figura 4.3: Parâmetros utilizados para o OPS

The screenshot shows a dialog box titled "OPS avançado" with the following parameters and options:

- Número de variáveis latentes máximo para o OPS: 10
- Número de variáveis latentes para o modelo: 8
- Número de amostras a retirar na validação cruzada: 1
- Janela: 10 Incremento: 5 Porcentagem de variáveis: 30
- Número de modelos a guardar em cada passo do OPS: 10
- Vetores** (radio buttons):
 - Correlograma
 - Correlograma e vetor de regressão
 - Correlograma e produto
 - Todos os vetores
- Critério para classificar modelos** (radio buttons):
 - RMSECV
 - Rcv
 - Q2
 - Spress
- Executar button

Após esta etapa, foi visualizada uma tabela com valores de Spress para cada modelo (construído com um determinado conjunto de descritores) em ordem decrescente. Spress representa o “desvio padrão das previsões” para a validação *leave-one-out*, calculada a partir de PRESS e graus de liberdade do modelo, conforme demonstrado na seguinte expressão:

$$SPRESS = \sqrt{\frac{PRESS_{CV}}{I - NVL - 1}} \quad (42)$$

A variável I corresponde ao número de amostras e NVL representa o número de variáveis latentes utilizadas no modelo.

Deste modo, foi escolhido o conjunto de variáveis que apresentaram o menor valor *Spress*, o qual correspondeu efetivamente a 12 descritores.

Os descritores selecionados pelo método OPS foram então submetidos à regressão PLS utilizando o programa QSAR Modeling. Foram avaliados parâmetros de qualidade de forma similar aos modelos anteriormente obtidos. Os resultados destes procedimentos e discussões pertinentes constam no capítulo Resultados e Discussões.

4.4 Proposição e avaliação de novos AMBH aromáticos frente aos modelos construídos

Tendo em vista a finalidade prática dos modelos QSAR, foram propostas quatro estruturas de AMBH aromáticos dos quais se esperam possuir bons valores de atividade biológica. As estruturas foram pensadas com base nas características evidenciadas pelo modelo gerado com descritores escolhidos *a priori*. As mesmas serão apresentadas no respectivo tópico em Resultados e Discussões.

Inicialmente, as estruturas propostas foram construídas com auxílio do software GaussView5®. Nesta etapa, buscou-se conservar as características conformacionais de estruturas semelhantes do conjunto inicial das 32 moléculas modeladas. Em seguida, estas geometrias foram submetidas a um cálculo de otimização em PCM-B3LYP/6-31+(d), seguido por cálculo de frequências vibracionais no mesmo nível.

A partir dos cálculos quânticos foram obtidos os valores *ELUKO* para cada AMBH. O valor de *SLogP* foi estimado usando o programa MedChem Designer™ 1.0 (SIMULATION PLUS). Os índices de fragmento também foram determinados. Em seguida estes descritores foram aplicados ao modelo correspondente para obtenção dos valores de atividade previstos.

Em adição, as moléculas propostas foram submetidas a uma análise utilizando o modelo construído com a matriz de descritores E-DRAGON. Neste caso, as estruturas devidamente otimizadas em DFT-PCM foram submetidas ao mesmo procedimento descrito no item 4.3.

RESULTADOS E DISCUSSÕES

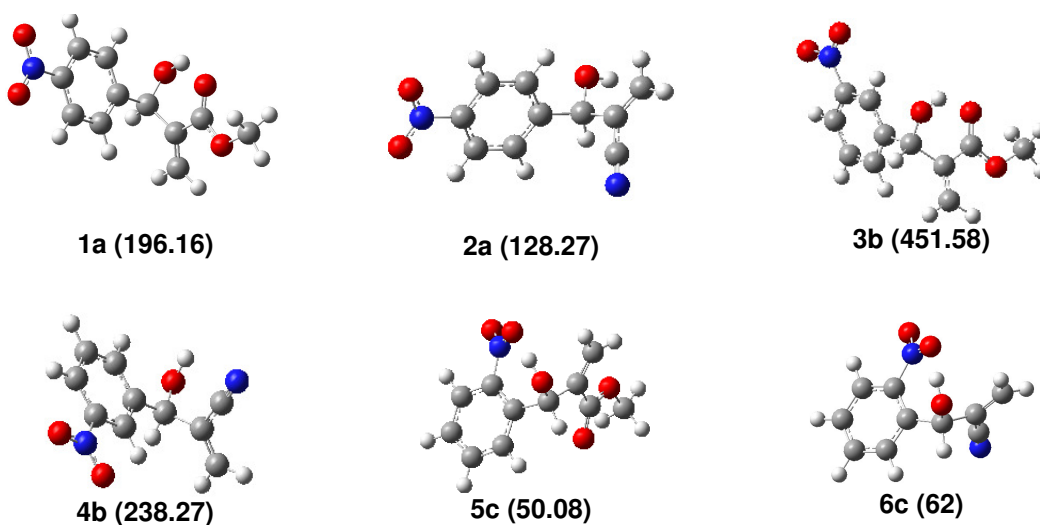
5. RESULTADOS E DISCUSSÕES

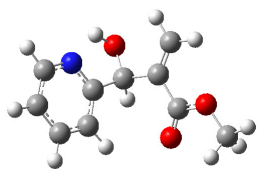
5.1 Conformações de mínimo obtidas para os Adutos de Morita-Baylis-Hillman

Como previamente exposto nos objetivos deste trabalho, inicialmente foi realizado um estudo conformacional dos AMBH através da Técnica RPES, a qual se encontra descrita nos Procedimentos Computacionais. Este estudo permitiu a obtenção dos mínimos conformacionais mais estáveis para cada estrutura, visando utiliza-las como base para os estudos QSAR. Em adição, estes cálculos permitiram a observação de algumas características estruturais e eletrônicas, as quais foram relacionadas à atividade leishmanicida dos compostos estudados.

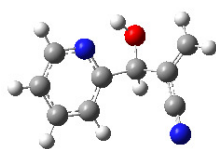
A Figura 5.1 apresenta as conformações mais estáveis encontradas após os cálculos de “scan”, seguidos de otimização em PCM/DFT (geometrias em coordenadas cartesianas no Anexo 8.1). O valor de atividade biológica para cada molécula, previamente determinado por nosso grupo de pesquisas, está indicado abaixo de cada estrutura. Corresponde à concentração mínima que inibe 50% do crescimento dos parasitas (IC_{50} em μM), por meio do método Elisa/MTT (do inglês “Enzyme-Linked Immunosorbent Assay”/ “3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide”) (JUNIOR *et al.*, 2010 e SILVA, *et al.*, 2011).

Figura 5.1: Conformações mais estáveis dos AMBH estudados neste trabalho. Em parêntesis os valores de IC_{50} em μM para cada molécula.

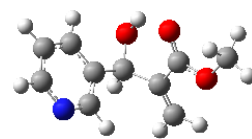




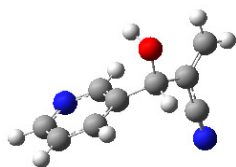
7d (731.37)



8d (302.8)



9e (448.5)



10e (391.77)



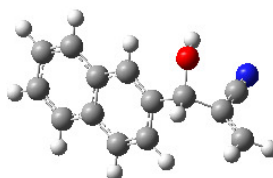
11f (488.98)



12f (250.95)



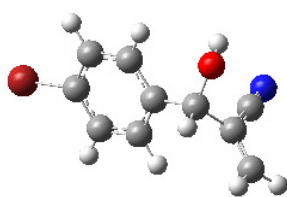
13g (83.5)



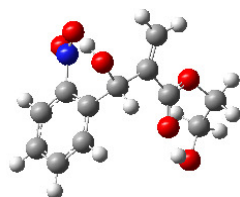
14g (32.88)



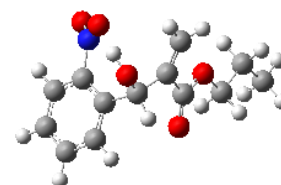
15h (85.54)



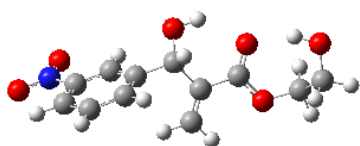
16h (40.23)



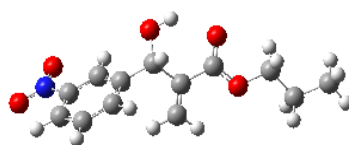
17i (20.52)



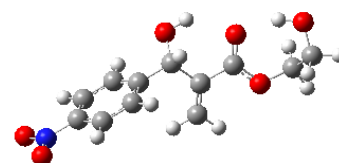
18i (28.38)



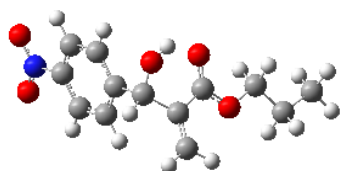
19j (64.57)



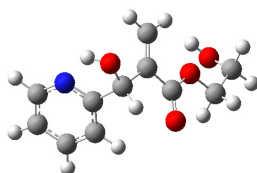
20j (72.23)



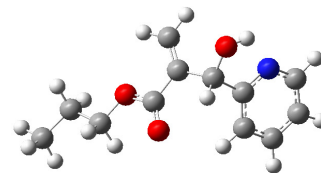
21k (53.03)



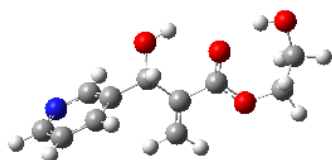
22k (52.04)



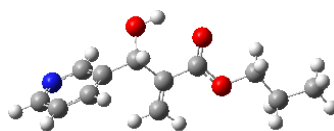
23l (266.32)



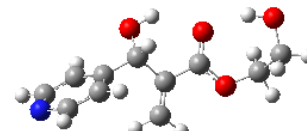
24l (194.07)



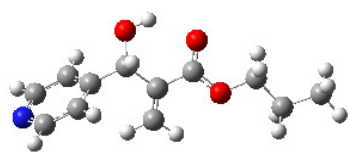
25m (323.77)



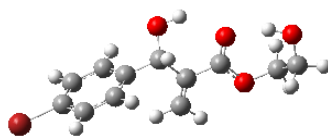
26m (210.81)



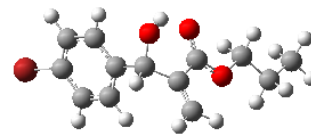
27n (381.65)



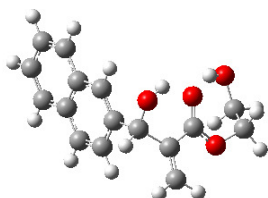
28n (148.6)



29o (56.73)



30o (74.7)



31p (52.54)



32p (54.58)

Observando as conformações de mínimo obtidas, pode-se perceber que muitas destas possuem indicativo de Ligações de Hidrogênio Intramoleculares (LHI) influenciando sua geometria. Este indicativo corresponde ao posicionamento dos doadores/receptores de prótons em cada estrutura. Nos AMBH **1a**, **3b**, **9e**, **11f**, **13g**, **15h**, **19j**, **20j**, **21k**, **22k**, **25m**, **26m**, **27n**, **28n**, **29o**, **30o**, **31p**, **32p**, existe um forte indicativo de interação entre a hidroxila benzílica e o oxigênio da carbonila. Por outro lado, considerando as moléculas **17i**, **19j**, **21k**, **25m**, **27n**, **29o**, **31p** e **32p**, também existe considerável indicativo de LHI entre o oxigênio da carbonila e a hidroxila da cadeia alquílica. Nas moléculas **7d**, **8d**, **23l** e **24l** existe um indicativo de LHI entre a hidroxila benzílica e o nitrogênio da piridina ($\text{OH}\cdots\text{N}(\text{py})$). Os compostos contendo grupo NO_2 em posição *orto* do anel aromático possuem uma particularidade estrutural, onde a conformação mais estável dos “scans”, submetida à otimização completa em DFT/PCM, apresenta um forte indicativo de uma LHI de sete membros, envolvendo as espécies $\text{OH}\cdots\text{O}_2\text{N}$ (**5c**, **6c**, **17i**, **18i**).

Recentemente, a IUPAC (do inglês “International Union of Pure and Applied Chemistry”) publicou um relatório técnico propondo uma definição da ligação de hidrogênio em termos de alguns critérios, os quais podem ser usados como evidência da formação desse tipo de interação (ARUNAN *et al.*, 2011). Deste modo, recomenda-se dentre outros avaliar a distância entre as espécies $\text{XH}\cdots\text{Y}$ (onde X e Y são átomos eletronegativos), bem como observar o aparecimento de pontos críticos de ligação (BCP) entre essas espécies, utilizando a Teoria QTAIM. Considerando o primeiro critério, uma ligação de hidrogênio será caracterizada quando a distância $\text{H}\cdots\text{Y}$ for menor que a soma dos raios de Van der Waals destas espécies. Em adição,

quanto menor for a distância, a interação é considerada mais forte. Na literatura científica, para os casos $\text{OH}\cdots\text{OC}$ e $\text{OH}\cdots\text{N}$, os valores estabelecidos são 2,61 e 2,64 Å respectivamente (BONDI, 1964; ROWLAND e TAYLOR, 1996). Pode ser observado na Tabela 5.1 que todas as interações propostas estão incluídas neste parâmetro de caracterização estrutural.

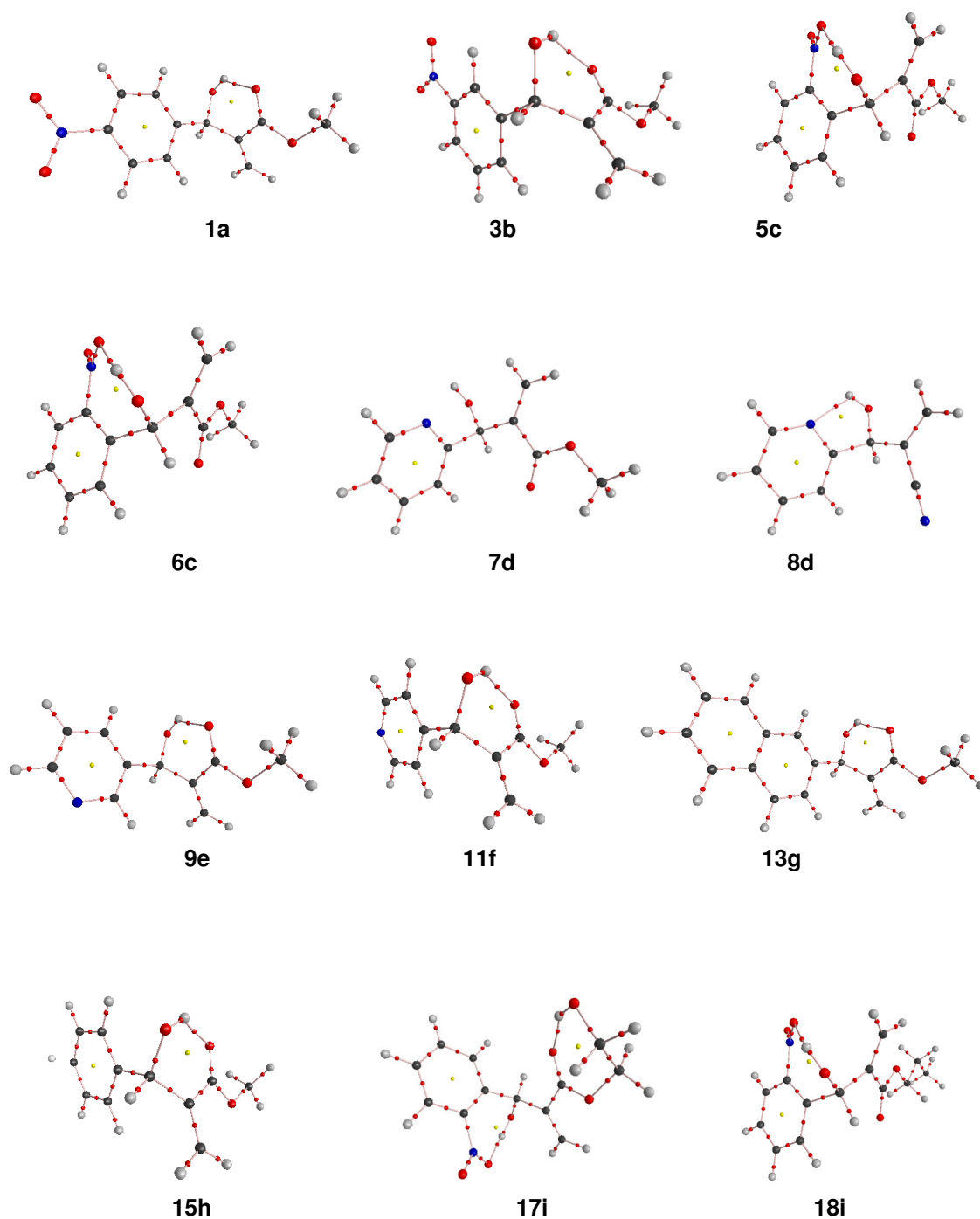
Tabela 5.1: Parâmetros estruturais para Ligações de Hidrogênio Intramoleculares.

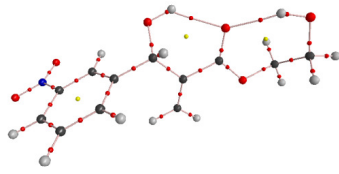
AMBH X \cdots HO	Distância (Angstrom)	AMBH X \cdots HO	Distância (Angstrom)
1a	2,02	21k (C=O \cdots HO)	1,89
3b	2,01	22k	2,01
5c	1,92	23l	2,48
6c	1,90	24l	2,46
7d	2,46	25m(BzOH \cdots O=C)	2,05
8d (N \cdots HO)	2,03	25m (C=O \cdots HO)	1,90
9e	2,02	26m	2,00
11f	2,02	27n (BzOH \cdots O=C)	2,05
13g	2,01	27n (C=O \cdots HO)	1,89
15h	2,02	28n	2,02
17i (NO ₂ \cdots HO)	1,92	29o (BzOH \cdots O=C)	2,01
17i (BzOH \cdots O=C)	1,88	29o (C=O \cdots HO)	1,89
18i (NO ₂ \cdots HO)	1,93	30o	2,02
19j (BzOH \cdots O=C)	2,04	31p (BzOH \cdots O=C)	2,03
19j (C=O \cdots HO)	1,89	31p (C=O \cdots HO)	1,90
21k (BzOH \cdots O=C)	2,05	32p	2,00

Considerando a caracterização de ligações de hidrogênio por QTAIM, além da presença do BCP entre as interações $\text{XH}\cdots\text{Y}$, o valor dos parâmetros topológicos densidade eletrônica $\rho(r)$ do Laplaciano da densidade eletrônica $\nabla\rho(r)$ ao nível do BCP podem ser avaliados, conforme abordado na Fundamentação Teórica. A Figura 5.2 a seguir apresenta os gráficos moleculares obtidos para os AMBH. Podemos perceber que as interações de hidrogênio evidenciadas nos cálculos DFT foram

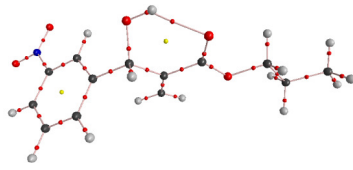
confirmadas através da QTAIM, exceto para as moléculas **7d**, **23l** e **24l**. Adicionalmente, a Tabela 5.2 apresenta os valores dos parâmetros topológicos ao nível dos BCP's, onde pode ser observado que os valores de $\nabla^2\rho(r)$ estão coerentes com a interpretação do Teorema do Virial para a caracterização de uma ligação de hidrogênio.

Figura 5.2: Gráficos Moleculares e Pontos Críticos de Ligação para alguns AMBH

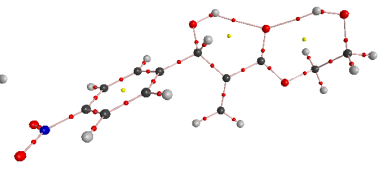




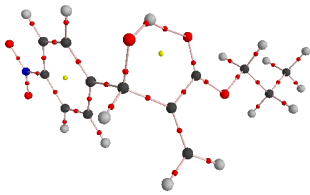
19j



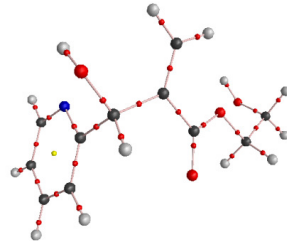
20j



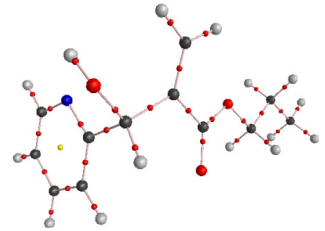
21k



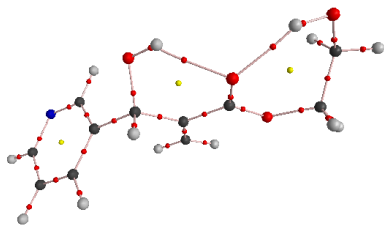
22k



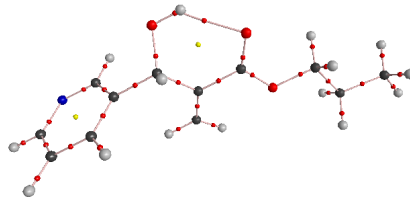
23l



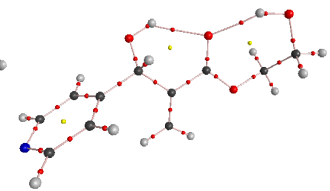
24l



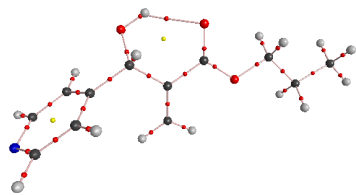
25m



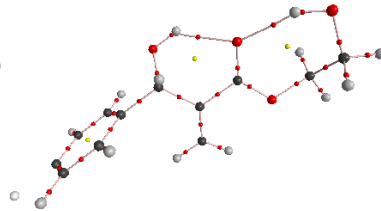
26m



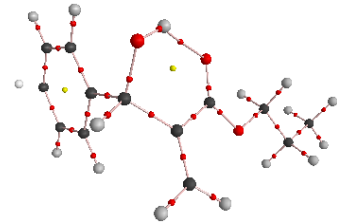
27n



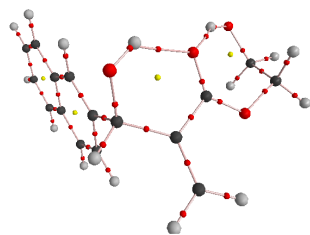
28n



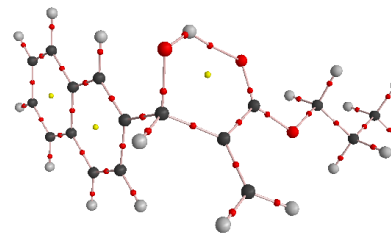
29o



30o



31p



32p

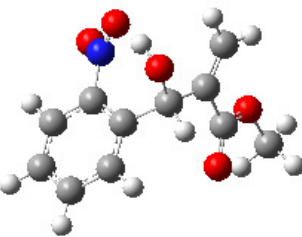

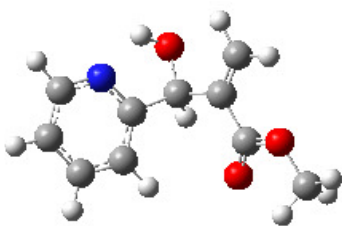

Tabela 5.2: Parâmetros topológicos (densidade eletrônica $\rho(r)$ e Laplaciano $\nabla^2 \rho(r)$ ao nível do BCP) para as Ligações de Hidrogênio Intramoleculares.

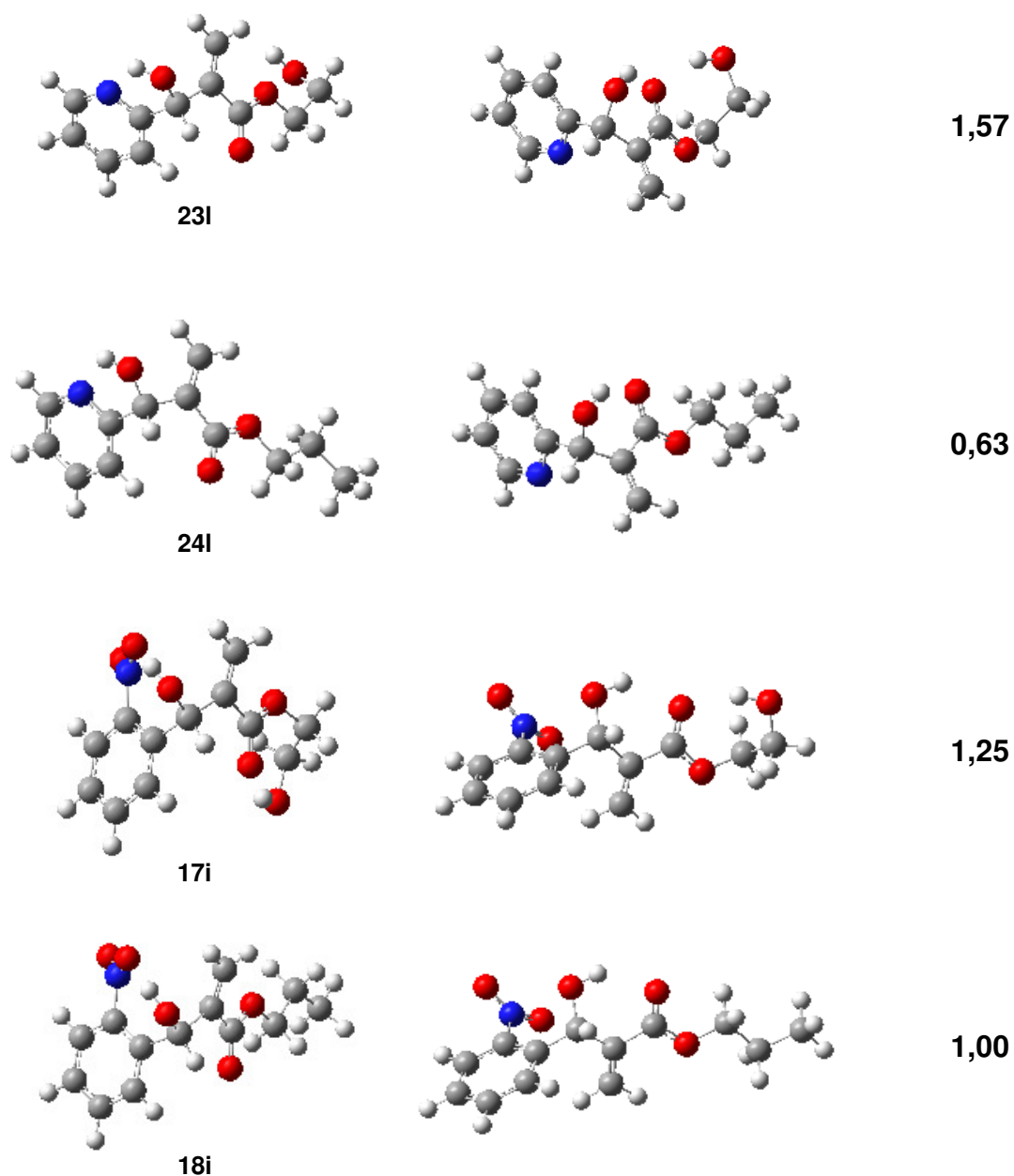
MBHA	$\rho(r)$	$\nabla^2 \rho(r)$
1a	0,02381	0,07883
3b	0,02406	0,08014
5c	0,02778	0,09418
6c	0,02941	0,09981
7d	–	–
8d (N \cdots HO)	0,02599	0,09143
9e	0,02403	0,07953
11f	0,02361	0,07832
13g	0,08017	0,02432
15h	0,07829	0,02369
17i (NO $_2\cdots$ HO)	0,02786	0,09454
17i (ArOH \cdots O=C)	0,03039	0,09993
18i (NO $_2\cdots$ HO)	0,02775	0,09411
19j (ArOH \cdots O=C)	0,02279	0,07686
19j (C=O \cdots HO)	0,02946	0,09723
21k (ArOH \cdots O=C)	0,02261	0,07626
21k (C=O \cdots HO)	0,02949	0,09742
22k	0,08047	0,08047
23l	–	–
24l	–	–
25m (ArOH \cdots O=C)	0,02214	0,07550
25m (C=O \cdots HO)	0,02883	0,09557
26m	0,02501	0,08304
27n (ArOH \cdots O=C)	0,02244	0,07561
27n (C=O \cdots HO)	0,02970	0,09811
28n	0,02406	0,08000
29o (ArOH \cdots O=C)	0,02373	0,07976
29o (C=O \cdots HO)	0,02970	0,09805
30o	0,02398	0,07905
31p (ArOH \cdots O=C)	0,02346	0,07807
31p (C=O \cdots HO)	0,02944	0,09708
32p	0,02471	0,08126

Em algumas moléculas, foi investigada a possibilidade de formação de outras LHI controlando suas geometrias. Assim, as estruturas da figura 5.1 (chamadas de “consideradas”) foram comparadas com outras geometrias subjetivamente construídas (“investigadas”), nas quais foram contempladas outras possibilidades de LHI. De forma a obter uma melhor descrição das LHI e dirimir dúvidas quanto a efetividade de cada tipo de interação na estabilização das estruturas, ambos os tipos de geometrias foram completamente otimizadas em PCM-B3LYP/6-311++(d,p). Estes cálculos tiveram ainda o propósito de avaliar a interação $\text{OH}\cdots\text{N}(\text{py})$ de **7d**, **23I** e **24I**, a qual não fora confirmada pelos cálculos QTAIM usando uma base menor.

A Tabela 5.3 apresenta as conformações consideradas e investigadas, após a mudança do conjunto de base. Nesta tabela constam ainda os valores de energia eletrônica relativa entre as conformações. Como se pode perceber, todas as conformações consideradas neste trabalho representam os mínimos mais estáveis, mesmo em relação a outras possibilidades de formação de LHI.

Tabela 5.3: Geometrias otimizadas em PCM-B3LYP/6-311++G(d,p) para investigação de outras possibilidades de LHI.

Conformação considerada (Cc)	Conformação investigada (Ci)	Energia Relativa (Ci – Cc) (Kcal/mol)
 5c		3,83
 7d		2,57



Com a mudança no conjunto de base para as conformações consideradas (conformações da primeira coluna), observamos que a geometria molecular não teve alterações importantes em relação à Figura 5.1 (o perfil de posicionamento dos átomos e tipos de interações permanece o mesmo), exceto pelo melhor direcionamento e menor distância entre doador e receptor de próton da interação $\text{OH}\cdots\text{N}(\text{py})$, correspondente aos compostos **7d**, **23i** e **24i**. Para estes compostos, a Tabela 5.4 mostra que as distâncias são consideravelmente menores em relação

aos cálculos com menor conjunto de base, evidenciando de forma mais efetiva esta importante interação de hidrogênio de cinco membros.

Deste modo, cálculos QTAIM adicionais foram realizados para as três conformações acima apresentadas, obtidas a partir dos cálculos com maior conjunto de base. Os gráficos moleculares (Figura 5.3), bem como os parâmetros topológicos obtidos para estas conformações (Tabela 5.4) mostram que os cálculos QTAIM confirmaram as evidências das LHI's $\text{OH}\cdots\text{N}(\text{py})$ após a mudança no conjunto de base, não havendo mudanças importantes quanto o perfil das conformações. Curiosamente, houve o aparecimento de BCP envolvendo o oxigênio da carbonila e um hidrogênio aromático em uma interação não clássica. A caracterização da mesma não foi realizada por não trazer informação relevante aos objetivos do presente estudo.

Figura 5.3: Gráficos moleculares para **7d**, **23I** e **24I** após obtenção de geometrias em PCM-B3LYP/6-311++(d,p).

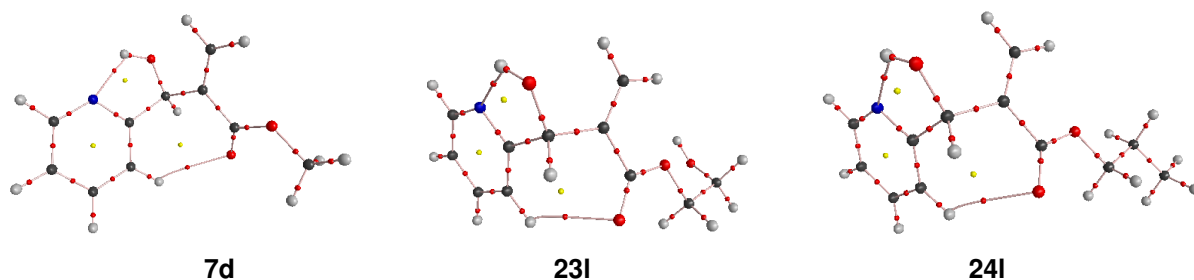


Tabela 5.4: Parâmetros topológicos QTAIM e estruturais (densidade eletrônica $\rho(r)$, Laplaciano $\nabla^2 \rho(r)$ ao nível do BCP, distância anterior-**atual**) para as interações $\text{OH}\cdots\text{N}(\text{py})$ de **7d**, **23I** e **24I**.

MBHA	$\rho(r)$	$\nabla^2 \rho(r)$	$\text{X}\cdots\text{HO}$ (Å)
7d	0,03125	0,10652	2,46 - 1,95
23I	0,03164	0,10728	2,48 - 1,95
24I	0,03170	0,10741	2,46 - 1,95

5.2 Relações dos resultados teóricos com dados espectroscópicos experimentais de RMN¹H

Dados espectroscópicos experimentais de RMN¹H obtidos durante a síntese dos 32 adutos (JUNIOR *et al.*, 2010 e SILVA, *et al.*, 2011) foram analisados de modo a buscar relações com as estruturas modeladas neste trabalho. Em geral, o deslocamento químico correspondente ao hidrogênio carbinólico **HCR₂OH** se apresenta em campo alto quando ocorre interação intramolecular envolvendo a hidroxila benzílica, conforme análises a seguir. Este fenômeno pode ser explicado devido ao maior fluxo de elétrons na direção da LHI, que tornaria este hidrogênio mais desprotegido. A tabela abaixo apresenta os respectivos valores de deslocamento químico (Tabela 5.5).

Tabela 5.5: Dados ¹H NMR para os AMBH (deslocamento químico do hidrogênio carbinólico).

MBHA	δ (ppm)	MBHA	δ (ppm)	MBHA	δ (ppm)	MBHA	δ (ppm)
1a	5.63	9e	5.60	17i	5.57	25m	5.52
2a	5.42	10e	5.33	18i	5.73	26m	5.55
3b	5.60	11f	5.51	19j	5.65	27n	5.56
4b	5.43	12f	5.31	20j	5.62	28n	5.51
5c	5.69	13g	5.75	21k	5.66	29o	5.55
6c	5.98	14g	5.46	22k	5.62	30o	5.52
7d	5.61	15h	5.51	23l	5.57	31p	5.78
8d	5.28	16h	5.27	24l	5.58	32p	5.75

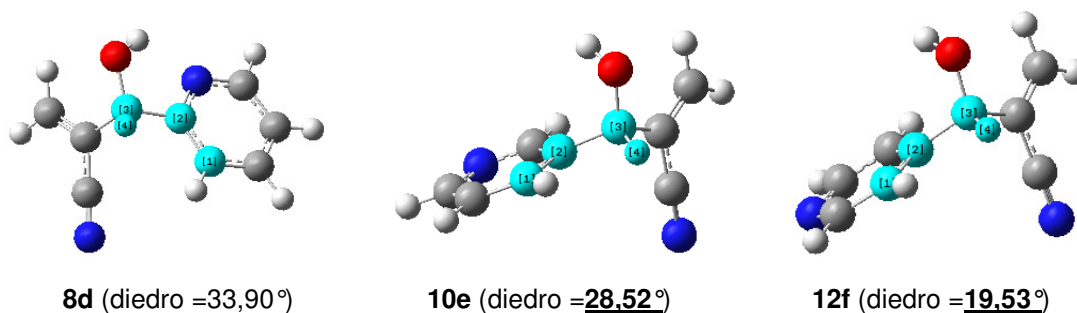
De modo a suprimir qualquer explicação destes fenômenos espectroscópicos por outros efeitos, a análise destes dados foi realizada estritamente entre os regioisômeros. O valor de deslocamento químico no composto **6c** é 0,56 e 0,55 ppm mais alto que o observado nos compostos **4b** e **2a** respectivamente, onde o composto **6c** possui uma LHI envolvendo o grupo nitro em posição *orto* e aqueles compostos não o possuem (Figura 5.1). Similarmente, mas menos expressiva, existe uma diferença entre o deslocamento químico para os regioisômeros **1a**, **3b** e **5c**. Entretanto, **1a** e **3b** podem formar uma interação OH...O=C alternativa (Figura 5.1), a qual explica a menor diferença de deslocamento

químico em relação ao composto **5c**, pois nos dois primeiros existe um efeito adicional de desblindagem do hidrogênio carbinólico.

Os valores de deslocamentos químicos apresentados seguem a mesma análise para os conjuntos de regioisômeros **17i**, **19j**, **21k** e **18i**, **20j**, **22k**, com uma observação: O composto **17i** (*orto*-NO₂) está em campo mais baixo que **19j** e **21k**. Esta aparente controvérsia pode ser explicada se observarmos que a espécie aceptora (grupo carbonila) de **19j** e **21k** interage simultaneamente com dois grupos hidroxila, o que pode causar um maior fluxo de elétrons para a LHI, tornando o hidrogênio carbinólico mais desprotegido em **19j** e **21k**, comparando com **17i**.

Em **outro contexto**, observando os regioisômeros contendo a porção piridina, efeitos de campo mostram sua importância nas análises dos deslocamentos químicos em relação às estruturas modeladas por métodos teóricos. Considerando os regioisômeros **7d**, **9e**, e **11f**, todos contendo o grupo éster metílico, o valor de deslocamento químico relativamente alto de **7d** pode ser explicado devido ao efeito de campo (indutivo) exercido pelo nitrogênio em proximidade ao hidrogênio carbinólico bem como pela interação OH[⋯]N(py), observada nos cálculos de otimização iniciais e de importância melhor evidenciada nos cálculos considerando maior conjunto de base. Fazendo uma comparação entre **8d**, **10e** e **12f** (contendo a porção nitrila), embora havendo a formação de interação OH[⋯]N(py) em **8d**, os compostos **10e** e **12f** mostram valores de deslocamento químico ligeiramente mais altos. Entretanto, observando as conformações destes últimos compostos, o hidrogênio carbinólico mantém-se relativamente mais próximo ao plano dos carbonos do anel, por tanto fora da zona de proteção (“cone de blindagem”) do sistema aromático (Figura 5.4). Nos compostos **23l**, **25m**, e **27n** o efeito indutivo em **23l** somado a interação entre a piridina e a hidroxila explicam seu deslocamento químico ligeiramente maior. Esta mesma explicação pode ser utilizada para **24l**, **26m**, e **28n**.

Figura 5.4: Posição relativa do hidrogênio carbinólico em relação ao anel aromático



As moléculas **13g-16h** e **29o-32p** não foram submetidas a estas análises por não apresentarem relação isomérica, possuindo estruturas consideravelmente diferentes, o que poderia ensejar a concorrência de outros efeitos. No entanto, se observamos com devido cuidado, os deslocamentos químicos relativos entre os subconjuntos duplos **13g-14g**, **15h-16h**, **29o-30o** e **31p-32p** seguem os padrões das discussões precedentes.

5.3 Estabelecimento de Relações Estrutura-Atividade (SAR)

No que diz respeito aos aspectos preliminares de Relação Estrutura-Atividade, SAR (do inglês “Structure-Activity Relationship”), algumas observações importantes podem ser feitas quanto posicionamento do grupo nitro. É conhecido que este grupo está diretamente envolvido na ação de fármacos antiparasitários nitroaromáticos, através da formação de espécies radiculares (JUNIOR *et al.*, 2010 e SILVA, *et al.*, 2011). Observa-se na Figura 5.1 que os compostos contendo o grupo NO₂ em posição *orto* estão entre os mais ativos (menor valor de IC₅₀), inclusive em comparação aos seus regioisômeros. Deste modo, baseado nos cálculos teóricos e geometrias fornecidas, podemos fornecer uma explicação para os dados experimentais.

Objetivando retirar a influência de outros grupos funcionais nas análises, apenas os subconjuntos de regioisômeros nitroaromáticos foram utilizados para comparação. Desta forma, considerando os compostos **3b** e **5c**, observamos que existe uma considerável melhora na atividade biológica quando o grupo muda respectivamente de *m*-NO₂ para *o*-NO₂. Curiosamente o composto **5c** tem um indicativo de LHI de sete membros. A mesma análise pode ser realizada

comparando os compostos **4b** e **6c**. Em adição, os compostos **17i** e **18i** possuem também uma LHI envolvendo o grupo *o*-NO₂ e são as duas moléculas mais ativas das 32 estudadas. Assim destaca-se que a formação de uma LHI de sete membros parece influenciar na potência dos compostos contendo grupo nitro.

Deve-se notar que a influência na atividade se dá a partir do grupo nitro, e não da hidroxila envolvida na LHI. Isto pode ser visualizado tendo como exemplo as moléculas **9e** e **11f**. Ambas indicam uma LHI do tipo OH \cdots O=C, contudo suas atividades biológicas possuem valores relativamente ruins (maior IC₅₀).

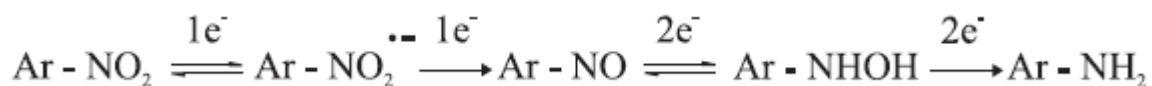
Adicionalmente, poder-se-ia explicar as diferentes atividades apresentadas devido a efeitos de ressonância distintos (tendo em vista as diferentes posições do nitro no anel), e não devido à formação da LHI. Porém essa possibilidade será descartada a seguir.

A posição do grupo nitro no anel deve induzir efeitos eletrônicos de ressonância distintos comparando as posições *orto* e *meta*. Na posição *para* o grupo teria mesma influência da *orto*. Podemos observar que a atividade diminui quando passamos de **1a** para **3b**, o que poderia ser condizente com a explicação de um possível efeito eletrônico provocado pela posição do grupo. No entanto, comparando os compostos **1a** e **5c**, os quais deveriam possuir efeitos eletrônicos similares (e portanto atividades similares caso fosse esta a explicação), observamos que a atividade biológica tem significativa melhora em **5c**, o qual possui LHI envolvendo o grupo nitro. Estes dados indicam que os efeitos eletrônicos indutivos e de ressonância no anel podem ser desconsiderados como influenciadores na atividade. Este é um dado adicional que corrobora a importância da interação OH \cdots O₂N na bioatividade dos adutos estudados.

Em relação à atividade antiparasitária/antibiótica de nitrocompostos já estabelecidos e usados na terapêutica, esta propriedade resulta basicamente de mudanças na estabilidade química dos mesmos, intermediada por interações entre o nitrocomposto e o seu alvo na biofase (PAULA, SERRANO e TAVARES, 2009). Em nível molecular estas mudanças ocorrem devido à facilidade de redução do grupo nitro, característica resultante de seu caráter fortemente acceptor de elétrons, bem como pelo efeito de ressonância entre o nitrogênio e os dois átomos de oxigênio de sua estrutura. A maioria dos compostos nitrados apresenta um mecanismo de biorredução enzimática, o que resulta na formação de radicais livres com toxicidade

preferencial para células bacterianas e parasitárias (PAULA, SERRANO e TAVARES, 2009). Em relação ao processo de redução do grupo nitro, o primeiro passo é a captação de um elétron pelo nitrogênio, formando uma espécie radicalar:

Esquema 5.1: Mecanismo reacional da biorredução de nitrocompostos



(adaptado de PAULA, SERRANO e TAVARES, 2009)

Deste modo, segue-se com a hipótese de que a formação da LHI envolvendo o grupo nitro pode ser responsável pelo aumento no seu potencial de redução, refletindo assim no maior valor de atividade biológica apresentado pelos compostos com esta possibilidade.

Uma análise das cargas atômicas, utilizando o modelo NBO (“Natural Bond Orbitals”), foi realizada de forma a corroborar a hipótese da LHI. Na análise de cargas NBO, as cargas são calculadas pela soma das ocupações dos NAO’s (“Natural Atomic Orbitals”), $\theta_k^{(A)}$. Estes correspondem a funções de onda ortonormais de um elétron, localizadas, que entram na construção dos Orbitais Naturais Híbridos (“Natural Hybrid Orbitals”, NHO’s), h_A , sendo cada um destes uma combinação linear de NAO’s em um dado centro (CRAMER, 2004):

$$h_A = \sum a_k \theta_k^{(A)} \quad (43)$$

Os NHO’s, por sua vez, formam os Orbitais Naturais de Ligação (“Natural Bond Orbitals”, NBO’s). Por tanto NAO’s e NHO’s são a base para a construção de NBO’s. Estes são orbitais de poucos centros, localizados, que descrevem as ligações. Mais precisamente, NBO’s são um conjunto ortonormal de orbitais localizados de máxima ocupação com $N/2$ membros (onde N representa o número de elétrons). Eles representam um modelo que mais aproxima a densidade total de N elétrons da descrição eletrônica proposta por Lewis, a qual denota conjuntos de pares eletrônicos na estrutura molecular.

Conforme colocado nos procedimentos, as cargas foram geradas a partir dos cálculos de otimização de geometria em DFT. Estas se mostraram de acordo com a hipótese da LHI, onde os compostos com a interação $\text{OH} \cdots \text{O}_2\text{N}$ apresentam o nitrogênio com carga mais positiva, o que ocasionaria sua maior avidez por elétrons (Tabela 5.6). É possível observar ainda que a carga do oxigênio tende a se tornar mais negativa com a formação da LHI, sendo este um fator adicional que corrobora a hipótese desta ligação intramolecular. A interação com o hidrogênio na LHI promoveria fluxo de carga proveniente do nitrogênio na direção do oxigênio, causando uma distribuição da carga ainda mais desigual, de forma a atender a esta interação.

Tabela 5.6: Cargas NBO no nitrogênio (qN) e oxigênio (qO—considerando o aceptor na LHI) do grupo nitro para os AMBH's.

MBHA	qN (e)	qO (e)
1a	0,501	-0,412
2a	0,501	-0,412
3b	0.502	-0.409
4b	0.504	-0.407
5c*	0.509	-0.417
6c*	0.511	-0.424
17i*	0.509	-0.417
18i*	0.509	-0.417
19j	0.502	-0.409
20j	0.505	-0.407
21k	0,501	-0,410
22k	0,501	-0,412

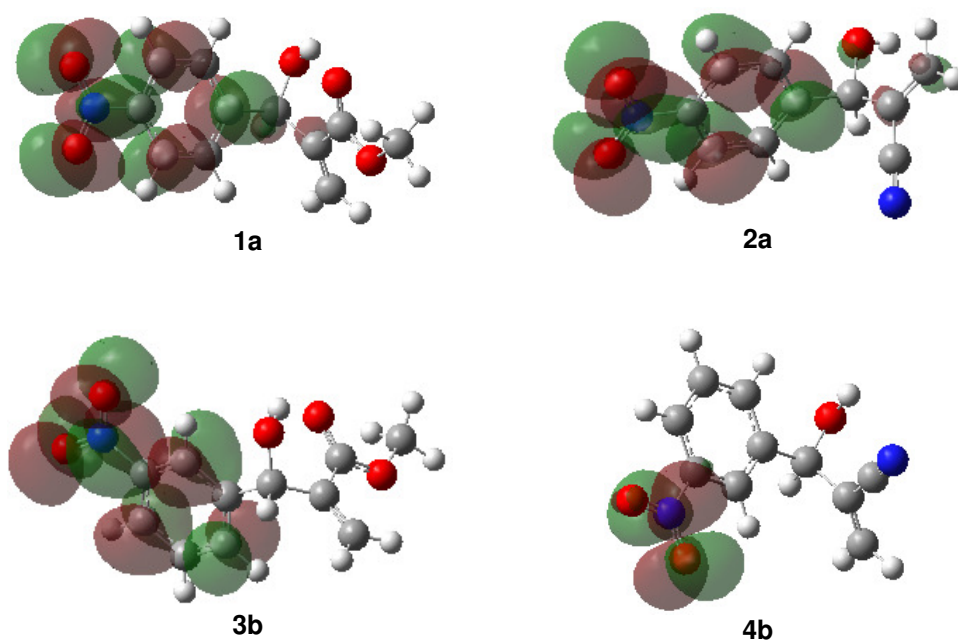
*Compostos com LHI envolvendo o grupo nitro.

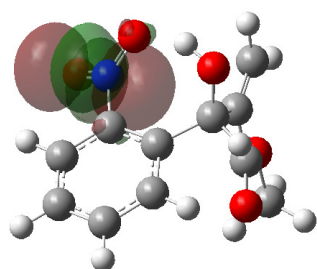
Para a tabela anterior foi aplicado o teste de hipótese de Student (Teste-t), visando observar a significância dos valores de carga obtidos nas quatro moléculas com formação de LHI. Considerou-se como hipótese nula $\bar{x} \leq \mu$, onde \bar{x} corresponde a média dos quatro valores de cargas em módulo e μ as médias das cargas considerando os demais compostos. Os valores de t foram $t = 12,5$ e $t = 6,14$

para **qN** e **qO** respectivamente. Para três graus de liberdade, a tabela t-student indica o valor de corte 5,54, considerando 99% de confiança. Assim rejeitamos a hipótese nula e consideramos significantes as diferenças entre os valores de carga observados.

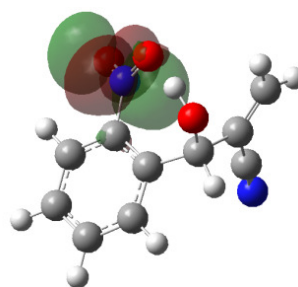
Em adição aos dados expostos, concordando com o processo de biorredução do grupo nitro, esperamos que uma análise dos Orbitais de Fronteira esteja coerente com os relatos feitos até o momento. Os cálculos realizados com o método DFT nos permitem avaliar os chamados *Orbitais de Kohn-Sham*, os quais não constituem os orbitais moleculares clássicos, mas são construídos para dar viabilidade aos cálculos DFT. Deste modo, esperamos que o Orbital de Kohn-Sham desocupado de mais baixa energia (“Lowest Unoccupied Kohn-Sham Orbital”, LUKO) esteja localizado na região do grupo nitro, haja vista que este “orbital virtual” representa possibilidade de alocar elétrons. A Figura 5.5 ilustra a localização dos LUKO's para os compostos contendo o grupo nitro, obtidos com auxílio do programa GaussView 5®. Como pode ser observado, o LUKO concentra-se na região que compreende aquele grupo, corroborando os resultados expostos até o momento.

Figura 5.5: Localização dos LUKO's para os AMBH mais ativos da série.

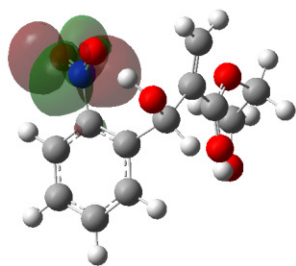




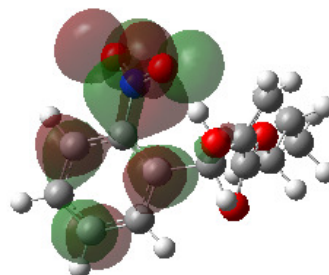
5c



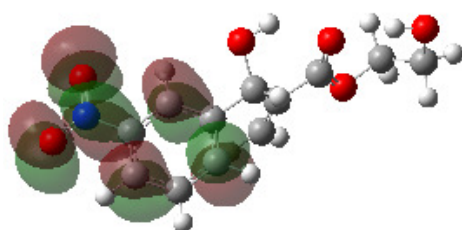
6c



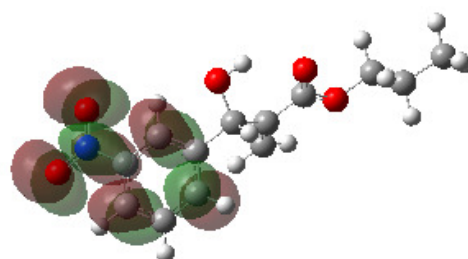
17i



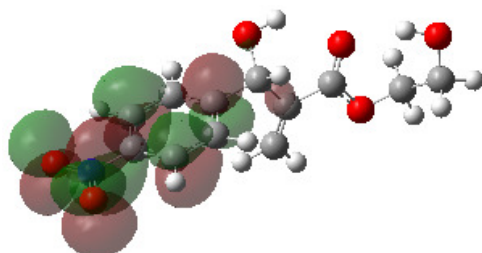
18i



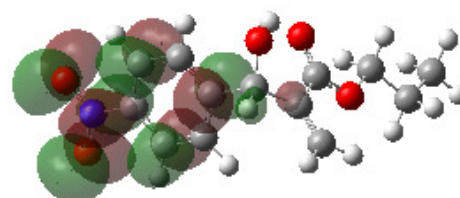
19j



20j



21k



22k

Adicionalmente, pode-se observar na Tabela 5.7 que os valores de energia LUKO são mais baixos para os compostos mais ativos (**5c**, **6c**, **17i**, **18i**) em comparação aos regioisômeros *meta* e *para* menos ativos (**1a**, **2a**, **3b**, **4b**, **19j**, **20j**, **21k**, **22k**). As atividades biológicas na tabela estão apresentadas como pIC_{50} ($-\text{Log}IC_{50}$).

Segundo a Teoria dos Orbitais de Fronteira (FUKUI, YONEZAWA e SHINGU, 1952) a transferência de elétrons ocorre com maior facilidade quanto menor a diferença de energia entre o HOMO (no caso HOKO) do doador e LUMO (LUKO) do acceptor. Assim de uma maneira geral, quanto menor o valor de LUKO

maior a facilidade em receber elétrons. Estas observações estão de acordo com o maior potencial de redução dos grupos nitro que realizam LHI, promovendo uma melhor atividade biológica.

Ainda relacionado às características moleculares importantes para a atividade leishmanicida dos AMBH, outras observações do ponto de vista estrutural merecem destaque. Neste contexto, pode-se observar que compostos contendo grupos lipofílicos como *naftil* ou bromo tendem a possuir valores de bioatividade mais elevados, mesmo considerando ausência do grupo nitro. A medida de lipossolubilidade de uma substância geralmente é avaliada com base no valor do seu LogP o qual, como anteriormente discutido, representa o logaritmo da relação entre a concentração da substância na fase orgânica sobre a fase aquosa. Após o uso pioneiro do LogP por Hansch (estimado teoricamente através da contribuição de cada substituinte pela adição da respectiva constante de lipofilicidade π), outras aproximações foram propostas para a estimativa do seu valor. Uma delas é o parâmetro SLogP, o qual representa um incremento em relação ao LogP por ser um modelo que leva em conta a contribuição dos átomos que estão inseridos nos grupos funcionais (WILDMAN e CRIPPEN, 1999). Outro dado estrutural é que, em geral, homólogos superiores (moléculas contendo grupo metileno adicional na cadeia lateral éster) tendem a assumir melhores valores de atividade em relação aos homólogos inferiores.

5.4 Estudos de QSAR dos Adutos de Morita-Baylis-Hillman

5.4.1 Construção de um modelo baseado em descritores escolhidos *a priori*

Considerando as discussões e resultados apresentados até o momento, alguns descritores eletrônicos e estruturais foram submetidos a construção de um modelo QSAR para os Adutos de Morita-Baylis Hillman leishmanicidas. Os descritores, selecionados *a priori*, foram: Energia do LUKO (*ELUKO*), valor de SLogP, descritores baseados em fragmentos para presença/ausência de NO₂ em posição *orto* (*oNO₂*) e presença/ausência de um metileno adicional na série homóloga (*Me*). Para os dois últimos descritores foi considerado que 1 indica

presença do grupo e 0 ausência do grupo, semelhante a análise Free-Wilson. Na Tabela 5.7 podem ser observados os valores dos descritores considerados.

Tabela 5.7: Valores de atividade transformados e descritores usados no QSAR.

MBHA	pIC₅₀	ELUKO (eV)	oNO₂	Me	SLogP
1a	3,707	-0,11407	0	0	1,681
2a	3,891	-0,11596	0	0	1,437
3b	3,345	-0,11421	0	0	1,608
4b	3,622	-0,11551	0	0	1,372
5c	4,300	-0,11789	1	0	1,594
6c	4,207	-0,12144	1	0	1,361
7d	3,135	-0,06253	0	0	0,459
8d	3,518	-0,06597	0	0	0,389
9e	3,348	-0,07123	0	0	0,392
10e	3,406	-0,06442	0	0	0,352
11f	3,310	-0,07145	0	0	0,456
12f	3,600	-0,06690	0	0	0,387
13g	4,078	-0,07012	0	0	2,734
14g	4,483	-0,07069	0	0	2,607
15h	4,067	-0,16707	0	0	2,443
16h	4,395	-0,17164	0	0	2,294
17i	4,687	-0,11844	1	1	0,946
18i	4,546	-0,11777	1	1	2,429
19j	4,189	-0,11492	0	1	0,952
20j	4,141	-0,11462	0	1	2,445
21k	4,275	-0,11525	0	1	1,019
22k	4,283	-0,11401	0	1	2,523
23l	3,574	-0,06110	0	1	-0,128
24l	3,712	-0,06084	0	1	1,292
25m	3,489	-0,07585	0	1	-0,145
26m	3,676	-0,06854	0	1	1,305
27n	3,418	-0,07605	0	1	-0,197
28n	3,827	-0,06870	0	1	1,248
29o	4,246	-0,07526	0	1	1,738

30o	4,126	-0,06851	0	1	3,315
31p	4,279	-0,07596	0	1	2,033
32p	4,262	-0,06925	0	1	3,553

5.4.1.1 Regressão PLS

Para efetivação da modelagem QSAR, foi utilizado inicialmente o método dos Mínimos Quadrados Parciais (PLS). Conforme descrito nos Procedimentos Computacionais, foram considerados vinte e cinco compostos no conjunto treinamento e sete compostos no conjunto teste (**5c, 9e, 13g, 20j, 22k, 25m, 28n**). A Tabela 5.8 apresenta os parâmetros de qualidade obtidos para o melhor modelo encontrado, construído com uma Variável Latente do PLS. Os parâmetros são: RMSECV – Raiz Quadrada do Erro Quadrático Médio para a Validação Cruzada (“Root Mean Square Error for Cross Validation”); R^2 – Coeficiente de determinação para a calibração; Q^2_{loo} - Coeficiente de determinação para a validação cruzada “leave-one-out”; Q^2_{lno} - Coeficiente de determinação para a validação cruzada “leave-n-out”; Q^2_{ext} - Coeficiente de determinação para a validação externa.

Tabela 5.8: Parâmetros estatísticos dos modelos PLS.

nº LV	RMSECV	R^2	Q^2_{loo}	$Q^2_{lno}^*$	Q^2_{ext}
1	0.2697	0.709	0.611	0.618	0.920
2	0.2797	0.718	0.581	-	-
3	0.2918	0.718	0.544	-	-
4	0.2917	0.718	0.545	-	-

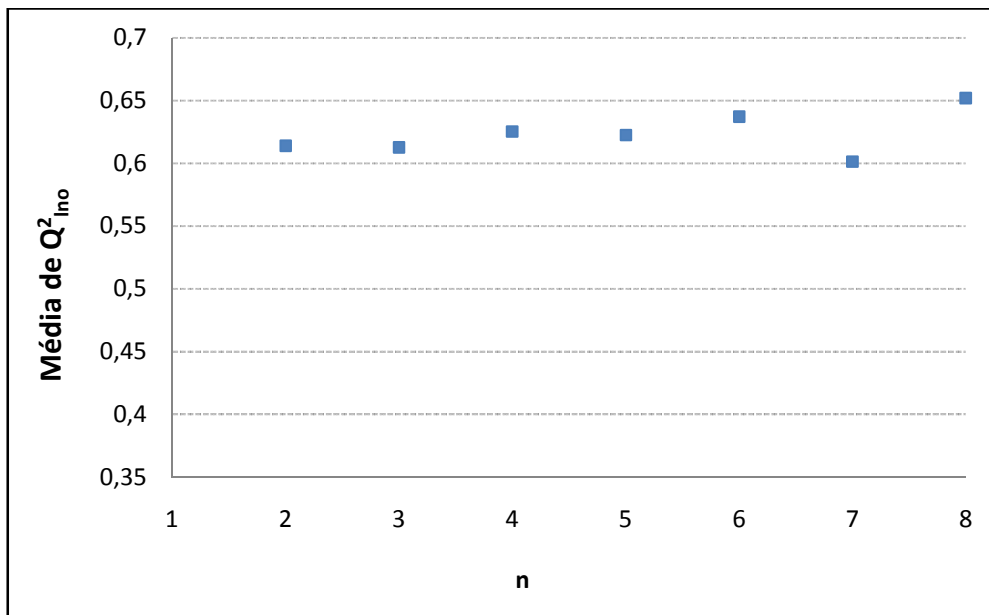
*media para “n” = 8

Os demais procedimentos de validação foram empregados para assegurar que um modelo PLS robusto e confiável fora obtido.

Para o procedimento de validação cruzada “leave-n-out” (com “n” assumindo valores de 2-8), existe a expectativa de que a média dos valores de Q^2_{lno} para cada “n” deva ser próxima de Q^2_{loo} e que não haja um desvio maior que 0,1 entre os

valores. A Figura 5.6 apresenta os resultados obtidos para a validação “leave-n-out”, indicando robustez do modelo PLS selecionado.

Figura 5.6: Resultados da validação cruzada “leave-n-out”.



O vetor de regressão do modelo, considerando as variáveis originais, está expresso na equação abaixo:

$$pIC_{50} = -4.266(ELUKO) + 0.481(oNO_2) + 0.249(Me) + 0.204(SLogP) + 3.04 \quad (44)$$

Podemos ainda visualizar o modelo considerando as variáveis autoescaladas, de forma a avaliar a real contribuição relativa de cada descritor para o modelo:

$$pIC_{50} = -0.315(ELUKO) + 0.361(oNO_2) + 0.288(Me) + 0.459(SLogP) + 3.04 \quad (45)$$

As equações nos indicam que a atividade leishmanicida dos Adutos de Morita-Baylis-Hillman aromáticos tende a aumentar quando as moléculas apresentam menores valores da energia do orbital LUKO, possuem o grupo nitro em posição *orto* ao anel, possuem um metileno a mais na cadeia lateral bem como

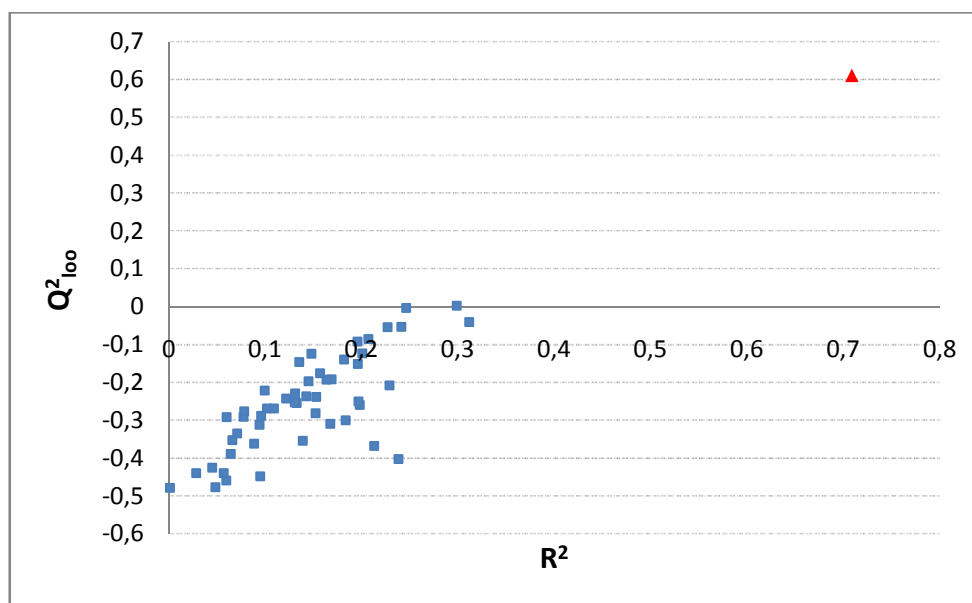
Tabela 5.9: Resultados obtidos com o procedimento *bootstrapping* para diferentes modelos PLS.

Conjunto teste	Q^2_{loo}	R^2	Q^2_{ext}
5, 9, 13, 20, 22, 25, 28*	0,61	0,72	0,92
3, 5, 9, 14, 20, 21, 25	0,77	0,83	0,47
2, 5, 7, 14, 18, 26,32	0,65	0,75	0,68
3, 7, 9, 14, 17, 27, 31	0,76	0,80	0,59
1, 3, 6, 11, 19, 23, 31	0,71	0,77	0,50
32, 10, 15, 20, 2, 28, 6	0,64	0,72	0,73
4, 5, 10, 16, 18, 22, 27	0,57	0,68	0,88
3, 8, 9, 11, 17, 21, 24	0,67	0,72	0,62
1, 3, 5, 7, 9, 11, 13	0,64	0,72	0,62
2, 4, 6, 8, 10, 12, 14	0,72	0,79	0,41
20, 22, 24, 26, 28, 30, 32	0,65	0,72	0,82
2, 7, 9, 15, 22, 29, 31	0,62	0,70	0,64
7, 10, 13, 16, 19, 22, 25	0,62	0,69	0,85
1, 3, 5, 12, 27, 19, 32	0,71	0,77	0,55
4, 6, 9, 17, 25, 31, 32	0,63	0,68	0,81
7, 12, 14, 19, 22, 27, 30	0,71	0,77	0,63
1, 5, 10, 15, 20, 25, 30	0,66	0,73	0,70
2, 8, 12, 17, 23, 29, 31	0,69	0,73	0,56
3, 8, 12, 17, 22, 25, 32	0,67	0,72	0,65
3, 5, 7, 12, 17, 27, 29	0,70	0,75	0,67
3, 11, 15, 18, 21, 23, 24	0,69	0,76	0,53

*Modelo original.

Ainda no contexto dos parâmetros de validação do modelo, podemos analisar os resultados do procedimento Aleatorização-Y (“Y-randomization”). Conforme apresentado na Fundamentação Teórica, esta técnica tem a finalidade de testar a possibilidade de correlação ao acaso entre a variável dependente e os descritores selecionados. Os resultados deste procedimento podem ser visualizados na Figura 5.8, para 50 aleatorizações. É possível verificar que apenas valores baixos de R^2 e Q^2_{loo} foram encontrados nos modelos permutados, o que indica um nível de correlação ao acaso tolerável.

Figura 5.8: Valores de R^2 e Q^2_{100} para a aleatorização-Y (o triângulo vermelho corresponde aos valores para o modelo original).



5.4.1.2 Regressão Linear Múltipla (MLR)

A técnica de MLR também foi realizada de forma comparativa ao modelo PLS, tendo em vista o pequeno número de descritores e ausência de correlação entre os mesmos.

Do mesmo modo que no modelo anterior foi definido um conjunto teste inicial com as moléculas **5c**, **9e**, **13g**, **20j**, **22k**, **25m**, **28n**, sendo o conjunto treinamento composto pelo restante. Em adição, foram escolhidos aleatoriamente vinte conjuntos teste para verificar seus parâmetros de validação frente ao conjunto inicialmente escolhido, bem como a efetividade do procedimento MLR na construção do modelo (*bootstrapping*). Pode-se observar na primeira linha de resultados da Tabela 5.10 que os principais parâmetros de validação para o conjunto treinamento/teste inicialmente escolhido foram aceitáveis, mas não representam melhora em relação ao modelo PLS considerado. Em adição, os dados obtidos a partir da escolha aleatória do conjunto teste (e, em consequência, do conjunto treinamento) forneceram modelos igualmente inferiores. A equação de regressão pode ser visualizada abaixo:

$$pIC_{50} = -3.8467(ELUKO) + 0.4301(oNO2) + 0.3111(Me) + 0.2094(SLOGP) + 3.054 \quad (46)$$

Tabela 5.10: Principais parâmetros de validação para diferentes modelos MLR, obtidos por escolha aleatória do conjunto teste.

Conjunto teste	Q^2_{loo}	R^2	Q^2_{ext}
5, 9, 13, 20, 22, 25, 28	0,54	0,72	0,89
3, 5, 9, 14, 20, 21, 25	0,73	0,83	0,49
2, 5, 7, 14, 18, 26,32	0,65	0,79	0,56
3, 7, 9, 14, 17, 27, 31	0,73	0,82	0,60
1, 3, 6, 11, 19, 23, 31	0,64	0,78	0,48
32, 10, 15, 20, 2, 28, 6	0,58	0,74	0,74
4, 5, 10, 16, 18, 22, 27	0,53	0,70	0,82
3, 8, 9, 11, 17, 21, 24	0,63	0,73	0,61
1, 3, 5, 7, 9, 11, 13	0,55	0,74	0,44
2, 4, 6, 8, 10, 12, 14	0,71	0,82	0,11
20, 22, 24, 26, 28, 30, 32	0,60	0,76	0,08
2, 7, 9, 15, 22, 29, 31	0,57	0,73	0,79
7, 10, 13, 16, 19, 22, 25	0,55	0,72	0,75
1, 3, 5, 12, 27, 19, 32	0,62	0,78	0,55
4, 6, 9, 17, 25, 31, 32	0,54	0,70	0,81
7, 12, 14, 19, 22, 27, 30	0,68	0,80	0,57
1, 5, 10, 15, 20, 25, 30	0,58	0,74	0,69
2, 8, 12, 17, 23, 29, 31	0,63	0,76	0,61
3, 8, 12, 17, 22, 25, 32	0,61	0,76	0,66
3, 5, 7, 12, 17, 27, 29	0,63	0,76	0,66
3, 11, 15, 18, 21, 23, 24	0,64	0,78	0,58

5.4.2 Construção de um modelo baseado em descritores de naturezas diversas

Uma das estratégias utilizadas em QSAR, principalmente quando pouco (ou nada) se conhece sobre o mecanismo de ação da classe de compostos, consiste em determinar grande número de descritores moleculares (de centenas a milhares) e construir um modelo baseado em variáveis selecionadas a partir desse grande conjunto de descritores, com a finalidade prática de orientar a síntese química e estabelecer possíveis explicações para atividade a nível molecular. Contudo, apesar de métodos como PLS e PCR permitirem trabalhar com um conjunto de variáveis bem maiores que o número de amostras, os descritores podem guardar informação redundante, bem como dificultar a interpretação do modo de ação dos compostos, a qual ficaria dependente de um grande número de informações, sendo assim inviável sua racionalização. Deste modo, um bom modelo pode ser construído baseado apenas nos descritores que são importantes para explicar a variabilidade da atividade apresentada e que não sejam (ou sejam pouco) correlacionados.

Um método de seleção de variáveis é aquele que se propõe a reduzir o número de variáveis que podem ser relacionadas às respostas do sistema (atividades biológicas). Este pode ser aplicado em uma etapa preliminar ao PLS ou PCR, bem como para viabilizar o uso da técnica MLR, conforme já exposto na Fundamentação Teórica.

Considerando o exposto, em adição ao modelo construído com descritores escolhidos *a priori*, foram calculados cerca de 1.600 descritores que representam diversas características moleculares utilizando a plataforma *on-line* E-DRAGON (TETKO *et al.*, 2005). Assim, nesta etapa buscou-se avaliar outras perspectivas de construção de um modelo QSAR.

A definição dos conjuntos treinamento e teste iniciais se deu de forma idêntica aos modelos anteriores (**5c, 9e, 13g, 20j, 22k, 25m, 28n** para conjunto teste e o restante para o conjunto treinamento). Os descritores determinados através do E-DRAGON foram então submetidos à seleção de variáveis através do procedimento OPS, conforme descrito nos Procedimentos Computacionais.

Após esta etapa, os 12 descritores selecionados podem ser visualizados na Tabela 5.11. São apresentados os valores para os 32 compostos (apesar de a seleção ter sido feita considerando os 25 compostos do conjunto treinamento).

Tabela 5.11: Descritores selecionados pelo método OPS.

MBHA	MATS8v	BEHm6	MATS 6v	Eig1m	H4u	HTp	Mor14 u	TIC1	nSK	BEL v7	R6e+	H- 049
1a	-0,263	2,617	-0,318	52,221	1,288	6,624	0,355	101,096	17	0,834	0,037	0
2a	-0,81	2,462	-0,212	40,497	0,903	5,776	0,339	78,532	15	0,588	0,067	0
3b	1,447	2,447	-0,065	50,118	1,165	6,608	0,175	101,096	17	0,824	0,057	0
4b	1,425	2,421	0,23	38,69	0,974	5,821	0,539	78,532	15	0,706	0,059	0
5c	-0,535	2,482	-0,323	48,176	1,039	6,804	0,309	101,096	17	0,834	0,075	0
6c	0	2,443	-0,471	37,034	0,803	6,017	0,638	78,532	15	0,602	0,102	0
7d	0,383	2,395	-0,472	36,206	0,737	5,937	0,113	87,832	14	0,695	0,047	1
8d	0	2,244	-0,207	26,544	0,415	5,33	0,323	66,174	12	0,26	0,074	1
9e	0,383	2,398	0,305	36,406	0,878	6,054	0,172	88,587	14	0,75	0,045	2
10e	0	2,262	0,397	26,715	0,668	5,136	0,662	66,929	12	0,272	0,073	2
11f	-0,541	2,391	-0,587	36,638	0,786	6,031	0,251	88,587	14	0,674	0,042	2
12f	0	2,232	-0,724	26,918	0,506	5,163	0,634	66,929	12	0,257	0,065	2
13g	-0,4	2,622	-0,28	54,95	1,378	8,442	0,507	101,819	18	0,834	0,034	0
14g	-0,28	2,482	-0,257	42,865	1,212	7,492	0,841	78,201	16	0,753	0,052	0
15h	-2,384	2,617	-0,49	40,812	0,845	8,543	0,427	90,702	15	0,699	0,041	0
16h	-3,17	2,471	-0,892	30,463	0,669	7,456	0,665	68,729	13	0,26	0,06	0
17i	-0,372	2,57	-0,396	63,842	1,169	7,888	0,708	116,881	19	0,878	0,066	0
18i	-0,048	2,599	-0,226	64,491	1,479	8,313	0,836	123,464	19	0,879	0,066	0
19j	0,704	2,569	-0,202	66,004	1,07	7,041	0,911	116,881	19	0,824	0,056	0
20j	1,185	2,599	-0,014	66,648	1,338	7,585	0,929	123,464	19	0,83	0,06	0
21k	-0,27	2,617	-0,394	68,324	1,229	7,11	0,783	116,881	19	0,974	0,055	0
22k	0,031	2,617	-0,227	68,962	1,694	8,153	0,842	123,464	19	0,996	0,035	0
23l	0,298	2,547	-0,525	50,082	0,822	6,967	0,443	103,007	16	0,721	0,043	1
24l	0,143	2,562	-0,342	50,699	1,139	7,444	0,422	109,315	16	0,829	0,044	1
25m	-0,357	2,548	-0,058	50,304	0,887	6,418	0,774	103,762	16	0,75	0,057	2
26m	0,143	2,559	0,233	50,92	1,171	6,968	0,799	110,07	16	0,829	0,037	2
27n	-0,357	2,546	-0,603	50,557	0,753	6,47	0,808	103,762	16	0,721	0,057	2
28n	0,143	2,579	-0,415	51,172	1,03	6,912	0,881	110,07	16	0,829	0,035	2
29o	-1,078	2,617	-0,454	55,446	0,85	8,606	0,864	106,087	17	0,721	0,056	0
30o	-1,1	2,617	-0,41	56,073	1,208	10,155	0,937	112,49	17	0,829	0,039	0
31p	-0,375	2,765	-0,342	71,548	1,512	9,628	0,889	118,327	20	0,971	0,036	0
32p	-0,216	2,787	-0,257	72,2	1,836	10,113	0,995	125,24	20	1,008	0,031	0

Os descritores da tabela 5.11 correspondentes aos 25 compostos do conjunto treinamento foram submetidos a uma modelagem PLS. Esta forneceu bons parâmetros de validação interna com valores de $Q^2_{loo} = 0,76$, $R^2 = 0,84$ e Q^2_{Ino} médio = 0,75 ($n = 8$) considerando duas variáveis latentes. Contudo o modelo falhou no procedimento de aleatorização-Y, fornecendo valores de Q^2_{loo} para os modelos permutados acima de 0,4 (Tabela 5.12).

Tabela 5.12: Valores obtidos com procedimento aleatorização-Y para modelo com 12 descritores (a última linha corresponde aos valores obtidos com o modelo original)

R^2_{loo}	Q^2
0,308851	-0,02241
0,368384	-0,67492
0,221339	-0,39924
0,264561	-0,70926
0,152069	-0,7864
0,22265	-0,66953
0,394959	-0,27016
0,280108	-0,50382
0,337637	-0,12652
0,237214	-0,66863
0,131146	-0,60721
0,223427	-0,29687
0,432894	0,068098
0,396921	-0,33438
0,140708	-0,44972
0,224194	-0,44809
0,12275	-0,91316
0,228002	-0,61593
0,227723	-0,79233
0,309931	-0,33578
0,270045	-0,75493
0,156088	-0,37601
0,436075	0,095982
0,156468	-1,04482
0,172333	-0,49456
0,239362	-0,95094
0,100046	-0,42204
0,24899	-0,30802
0,198369	-0,80295
0,260752	-0,58464
0,304817	-0,06975
0,321614	-0,22673
0,156075	-1,26517
0,337092	-0,0839
0,33572	-0,5684
0,263246	-0,19502
0,316576	-0,2165
0,169146	-0,60667
0,283024	-0,4886
0,097818	-0,80834
0,203849	-0,71616

0,400638	-0,08181
0,226635	-0,40733
0,465769	-0,09418
0,208117	-0,69019
0,215028	-0,40535
0,195372	-0,32252
0,185277	-0,49573
0,19636	-0,27234
0,301266	-0,24499
0,83869	0,75893

Em seguida, na tentativa de encontrar um modelo capaz de passar nos testes de aleatorização, foram incluídos na matriz anterior os quatro descritores escolhidos *a priori* ELUKO, oNO₂, Me e SLogP, totalizando 16 descritores. Os valores obtidos com um modelo PLS, considerando duas variáveis latentes, foram $Q^2_{loo} = 0,76$, $R^2 = 0,85$ e Q^2_{lno} médio = 0,73 (n = 8). No entanto, de forma similar a tentativa anterior, o modelo não foi satisfatório no teste de aleatorização-Y (Tabela 5.13).

Tabela 5.13: Valores obtidos com procedimento aleatorização-Y para modelo com 16 descritores (a última linha corresponde aos valores obtidos com o modelo original)

R^2_{loo}	Q^2
0,423924	-0,18233
0,260934	-0,68165
0,291109	-0,35618
0,269163	-0,4182
0,359565	-0,93584
0,393908	-0,07172
0,298341	-0,41278
0,221819	-0,44189
0,339583	-0,4035
0,299806	-0,26744
0,15175	-0,92462
0,631391	0,209153
0,286798	-0,27388
0,444529	-0,30013
0,28349	-0,54116
0,228013	-0,86657
0,31146	-0,54007
0,185931	-0,57865
0,344132	-0,25787
0,360888	-0,382
0,250132	-0,8384
0,31859	-0,26248
0,177886	-0,99613
0,228995	-1,02761
0,426976	0,003656
0,381495	-0,11206
0,318737	-0,14009
0,228295	-1,21107

0,26342	-0,31805
0,493507	-0,25311
0,311175	-0,41459
0,3763	-0,22019
0,313724	-0,50783
0,252459	-0,57487
0,274124	-0,24116
0,55437	0,225457
0,346453	-0,36617
0,251795	-0,46973
0,309959	-0,83953
0,274953	-0,17532
0,251824	-0,49404
0,178338	-1,24892
0,315271	-0,20132
0,292762	-0,32726
0,378792	-0,31776
0,364891	-0,25442
0,206961	-0,42341
0,318028	-0,29691
0,349044	-0,43748
0,370132	-0,47893
0,851871	0,762644

Tendo em vista a tendência promissora dos descritores em tela (exceto na aleatorização-Y), procedeu-se uma nova seleção de variáveis utilizando a matriz de 16 descritores descrita no parágrafo anterior. Após o algoritmo OPS, foram fornecidos subconjuntos de descritores, os quais se encontravam classificados em ordem crescente do parâmetro de qualidade Spres. Com estes dados, uma busca sistemática foi realizada, submetendo cada subconjunto de variáveis (começando com o primeiro) à avaliação dos parâmetros de validação interna. O modelo de número 17, o qual corresponde às variáveis **MATS8v, H4u, Mor14u, H-049 e oNO2**, apresentou sucesso em todas as avaliações, com melhor desempenho considerando 2 variáveis latentes. Os coeficientes de regressão obtidos foram então utilizados para calcular os valores de atividade das moléculas do conjunto teste objetivando a obtenção do Q^2_{ext} . A tabela 5.14 apresenta os valores obtidos indicando um modelo QSAR robusto e preditivo.

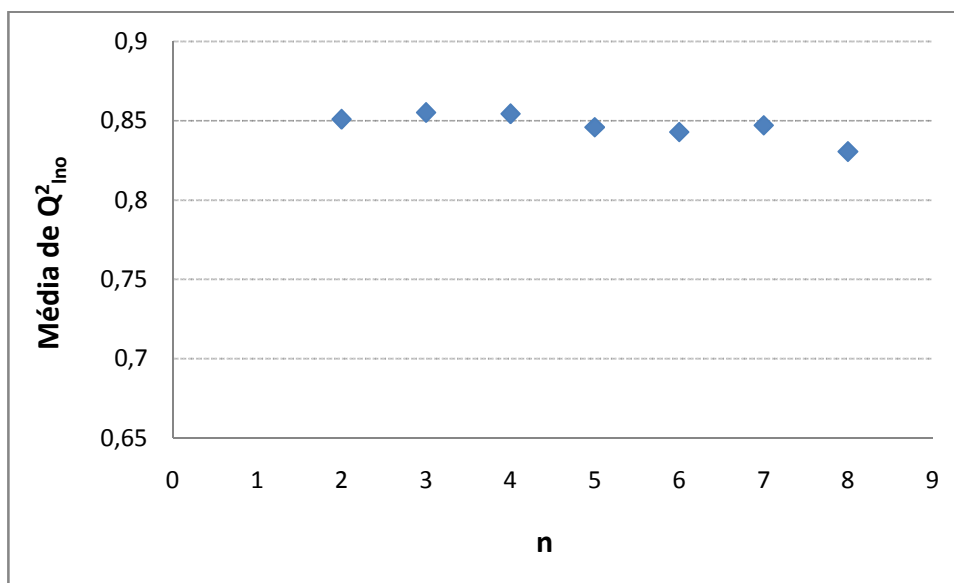
Tabela 5.14: Parâmetros de validação do modelo OPS-PLS.

n° VL	RMSECV	R ²	Q ² _{loo}	Q ² _{Ino} *	Q ² _{ext}
1	0.18042	0.872	0.826	-	-
2	0.16345	0.906	0.857	0.830	0.849
3	0.16592	0.911	0.853	-	-
4	0.1648	0.912	0.855	-	-
5	0.16487	0.912	0.855	-	-

*media para "n" = 8

A Figura 5.9 apresenta o resultado da validação cruzada *leave-n-out*, considerando as médias de dez determinações para cada "n" (de 2 à 8). Como pode ser observado, o perfil do gráfico satisfaz o critério de aceitação deste parâmetro de avaliação (valor próximo ao Q²_{loo} desvio menor que 0,1).

Figura 5.9: Resultados da validação cruzada "leave-n-out" para o modelo OPS-PLS.



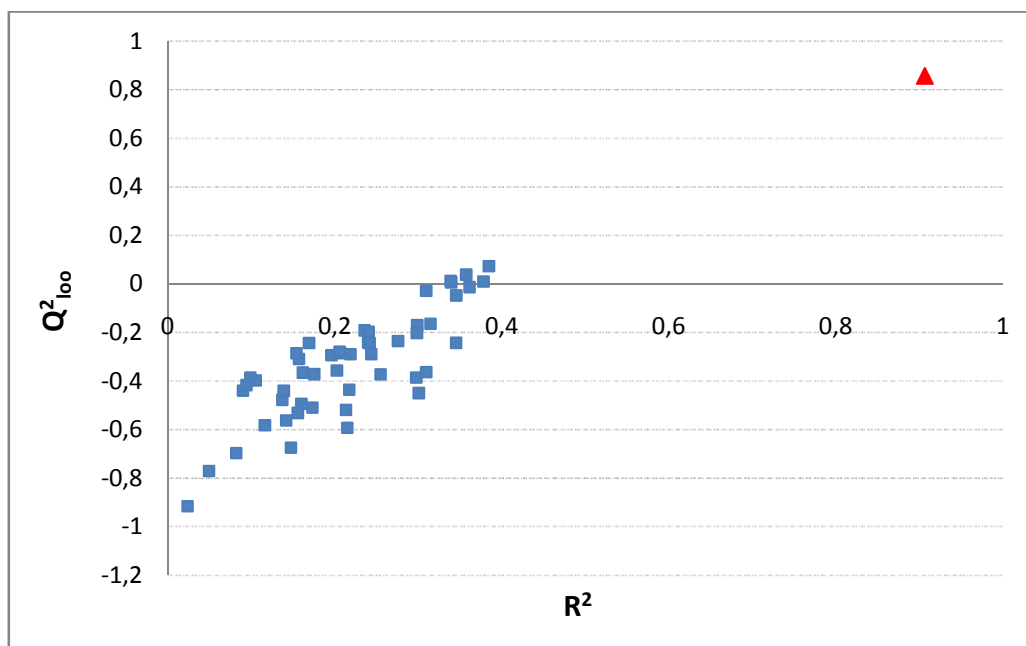
Os vetores de regressão, considerando as variáveis originais bem como as variáveis autoescaladas podem ser visualizados nas equações 47 e 48, respectivamente.

$$pIC_{50} = -0.207(H049) + 0.699(Mor14u) - 0.158(MATS8v) + 0.410(oNO_2) + 0.192(H4u) + 3.3087 \quad (47)$$

$$pIC_{50} = -0.384(H049) + 0.409(Mor14u) - 0.348(MATS8v) + 0.308(oNO_2) + 0.144(H4u) + 3.3087 \quad (48)$$

Os resultados do procedimento de randomização-Y para o modelo OPS-PLS podem ser visualizados na Figura 5.10. Estes mostram que o modelo apresenta tolerável correlação ao acaso.

Figura 5.10: Resultados do procedimento de aleatorização-Y (o triângulo vermelho corresponde aos valores para o modelo original)



Na tabela 5.15 pode ser visualizado o resultado do procedimento *bootstrapping*. Todos os *splits* forneceram modelos com bons parâmetros de validação interna e externa.

Tabela 5.15: Resultados obtidos com o procedimento *bootstrapping* para diferentes modelos PLS.

Conjunto teste	Q^2_{loo}	R^2	Q^2_{ext}
5, 9, 13, 20, 22, 25, 28*	0,857	0,906	0,849
3, 5, 9, 14, 20, 21, 25	0,869	0,916	0,844
2, 5, 7, 14, 18, 26,32	0,828	0,898	0,893
3, 7, 9, 14, 17, 27, 31	0,855	0,906	0,857
1, 3, 6, 11, 19, 23, 31	0,863	0,898	0,905
32, 10, 15, 20, 2, 28, 6	0,864	0,904	0,816
4, 5, 10, 16, 18, 22, 27	0,823	0,890	0,906
3, 8, 9, 11, 17, 21, 24	0,828	0,880	0,909
1, 3, 5, 7, 9, 11, 13	0,830	0,884	0,727
2, 4, 6, 8, 10, 12, 14	0,866	0,914	0,797
20, 22, 24, 26, 28, 30, 32	0,880	0,914	0,588
2, 7, 9, 15, 22, 29, 31	0,824	0,882	0,948
7, 10, 13, 16, 19, 22, 25	0,838	0,889	0,900
1, 3, 5, 12, 27, 19, 32	0,886	0,922	0,732
4, 6, 9, 17, 25, 31, 32	0,856	0,899	0,877
7, 12, 14, 19, 22, 27, 30	0,897	0,933	0,801
1, 5, 10, 15, 20, 25, 30	0,883	0,919	0,702
2, 8, 12, 17, 23, 29, 31	0,843	0,892	0,885
3, 8, 12, 17, 22, 25, 32	0,860	0,902	0,835
3, 5, 7, 12, 17, 27, 29	0,854	0,898	0,858
3, 11, 15, 18, 21, 23, 24	0,816	0,878	0,938

*Modelo original.

Diante dos parâmetros de qualidade apresentados, podemos atestar que as variáveis selecionadas com o método OPS fornecerem um bom modelo QSAR para os Adutos de Morita-Baylis-Hillman. Não obstante, em QSAR faz-se necessário buscar alguma relação entre os descritores moleculares utilizados no modelo e o mecanismo de ação da respectiva classe de compostos.

Descritores como os utilizados nos últimos modelos obtidos, os quais representam características constitucionais e/ou topológicas das mais diversas, podem não possuir uma associação simples com o mecanismo de ação da classe

estudada, mesmo que este seja plenamente conhecido (GONZALES, *apud* ABREU, 2009).

Até o presente, não existem dados experimentais que permitam afirmar conclusivamente o mecanismo de ação leishmanicida dos AMBH, ou o tipo de alvo possivelmente envolvido. O que existe são as especulações já apresentadas em artigos anteriores do nosso grupo de pesquisas e neste trabalho, durante o estudo conformacional e seleção de descritores *a priori*.

Deste modo, a construção do modelo OPS-PLS não tem a pretensão de fornecer explicações conclusivas com relação à atividade dos AMBH. Ela representa (junto com a modelagem *a priori*) um marco inicial neste sentido, bem como um guia para a orientação de futuras sínteses e obtenção de moléculas potencialmente mais promissoras (o que se faz na indústria farmacêutica), poupando custos e tempo nos esforços do nosso grupo de pesquisas e outros interessados neste tema.

Com base nestas considerações, conhecer o significado dos descritores selecionados no procedimento OPS (os quais se mostraram importantes para explicar a atividade dos compostos) poderá ser útil na ocasião em que novos dados experimentais surjam sobre o mecanismo da ação leishmanicida dos AMBH. Estes serão elencados nos parágrafos a seguir.

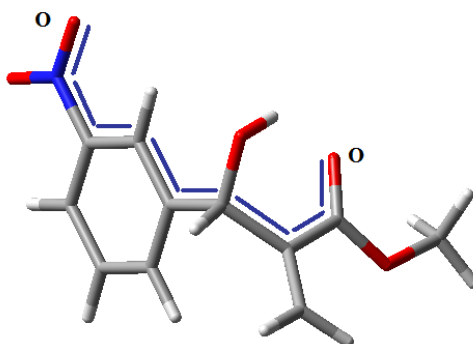
MATS8v: “Moran autocorrelation of lag 8 weighted by van der Waals volume”. Está inserido no grupo dos descritores de autocorrelação 2D, os quais exprimem como os valores de certas propriedades de átomos, separados por intervalos determinados pela distância topológica (“lag”), estão correlacionadas. Para este descritor especificamente, o “v” indica que a propriedade corresponde ao volume de Van der Waals, a uma distância topológica de 8 ligações (MORAN *apud* ABREU, 2009). A forma de cálculo deste descritor é expressa a seguir:

$$I(d) = \frac{\frac{1}{A} \sum_{i=1}^A \sum_{j=1}^A \delta_{ij} \cdot (w_i - \bar{w}) \cdot (w_j - \bar{w})}{\frac{1}{A} \sum_{i=1}^A (w_i - \bar{w})^2} \quad (48)$$

Onde w_i é uma propriedade atômica (como o volume de Van der Waals), \bar{w} é a média do seu valor na molécula, A é o número de átomos e d corresponde à distância topológica (“lag”) entre os átomos. δ_{ij} corresponde ao delta de Kronecker, o qual é 1 se $d_{ij} = d$ e zero caso não seja.

O valor negativo do coeficiente de regressão para esta variável indica que dois átomos a uma distância topológica de 8 ligações, ambos com correlação positiva em relação a média dos volumes de Van der Waals promovem diminuição do valor de pIC_{50} . A figura 5.11 ilustra o significado do descritor. Nesta figura temos dois átomos com volumes de Van der Waals maiores, em relação à média, por tanto uma correlação positiva.

Figura 5.11: Representação de par de átomos com volumes de Van der Waals relativamente maiores para **3b**, a uma distância topológica de 8 ligações.



H-049: “H attached to $C^3(sp^3)/C^2(sp^2)/C^3(sp^2)/C^3(sp)$ ”. Representa o número de átomos de hidrogênio ligados aos carbonos com respectivas hibridações, nos quais o número sobrescrito indica o número de oxidação formal do carbono (soma das ordens de ligação convencionais aos átomos eletronegativos O, N, S, P, Se e halogênios) (PLANCHE *et al.*, 2009). Pode-se claramente observar que os carbonos do anel piridínico, ligados ao nitrogênio, correspondem a átomos $C^2(sp^2)$. Estes são os únicos tipos de carbono, relacionados ao descritor, presentes nos AMBH. Apenas os doze AMBH que contêm o anel piridínico apresentam algum valor para este descritor (Tabela 5.10). Seu coeficiente de regressão no modelo indica que este descritor contribui negativamente para o valor de pIC_{50} . De fato, todas as moléculas contendo a porção piridina são, em geral, menos ativas. Dois aspectos principais podem ser observados neste caso: O primeiro é que a presença deste grupo torna a molécula menos lipofílica (efeito considerado no primeiro modelo QSAR) podendo trazer prejuízo à atividade pelo fator de transporte através das membranas biológicas; O segundo é que a presença do nitrogênio no anel piridínico pode prejudicar o estabelecimento da interação comumente atribuída a anéis aromáticos no sítio de ligação, ou seja, do tipo lipofílico e de preenchimento de “bolsões” hidrofóbicos no ambiente de interação com o alvo.

Mor14u: Pertence ao conjunto de descritores 3D-MoRSE (do inglês, “3D Molecule Representation of Structures based on Electron diffraction”). Estes descritores são calculados utilizando a mesma função de dispersão (espalhamento) utilizada na técnica de difração de elétrons (aplicada para obtenção de curvas de dispersão teóricas). Assim podem fornecer informações sobre as coordenadas tridimensionais dos átomos (STRAND LIFE SCIENCES, 2007). Estes descritores estão diretamente associados à forma e volume ocupado pela molécula, assim maiores valores desse descritor poderiam representar a ocupação da totalidade do sítio de interação no alvo biológico (FREITAS, PAZ e CASTILHO, 2009). O vetor de regressão do modelo PLS indica que maiores valores deste descritor favorecem a atividade.

H4u: “H autocorrelation of lag 4 / unweighted”. Representa um descritor da classe GETAWAY (GEometry, Topology, and Atom-Weights Assembly). Estes descritores buscam combinar a geometria molecular tridimensional (fornecida pela matriz de influência molecular H) com informação química, usando diferentes esquemas de ponderação atômica (massa, polarizabilidade, eletronegatividade). A matriz de influência molecular H é calculada a partir da matriz de coordenadas M (composta por 3 colunas representando as coordenadas x, y, z de todos os átomos) (STRAND LIFE SCIENCES, 2007).

$$H = M(M^T \cdot M)^{-1}M^T \quad (49)$$

Os elementos da diagonal h_{ii} representam a influência (do inglês “leverage”) de cada átomo na determinação da forma da molécula. Os elementos fora da diagonal, h_{ij} , representam o grau de acessibilidade do átomo j para interações com o átomo i.

A forma geral para calcular este descritor, considerando um “lag” k e uma função de ponderação w, seria:

$$Hkw = \sum_{i=1}^n \sum_{j=1}^n (h_{ij} \cdot w_i \cdot w_j) \cdot \delta(k, d_{ij}, h_{ij}) \quad (50)$$

onde d_{ij} corresponde a distância topológica entre os átomos i e j . Assim, $\delta(k, d_{ij} \cdot h_{ij}) = 1$, se $k = d_{ij}$ e $h_{ij} > 0$. Para o descritor **H4u**, $w_i = w_j = 1$, desde que é um descritor não ponderado (“unweighted”). Assim, o seu valor se resume ao somatório das acessibilidades entre os átomos i e j a uma distância topológica de quatro ligações. Pode-se observar a partir da tabela 5.10 que compostos contendo a porção piridina apresentam menores valores deste descritor. De fato, o nitrogênio inserido na sequência de átomos do anel não fornece possibilidade de interações tridimensionais com outros átomos a uma distância de quatro ligações. Ao contrário, moléculas contendo “ramificações” e grupos como o nitro, fornecem possibilidades adicionais de acessibilidade interatômica. Com base no vetor de regressão, vemos que maiores valores deste descritor contribuem para aumentar o valor de pIC_{50} .

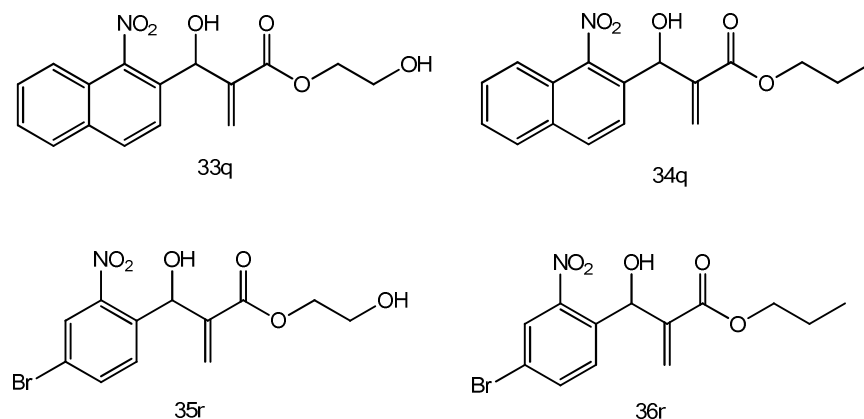
O descritor **oNO2** corresponde ao já conhecido índice de fragmento para presença/ausência do grupo nitro em posição *orto*. Seu aparecimento após a etapa do OPS para os 16 descritores reforça mais uma vez a importância da interação de hidrogênio envolvendo o grupo nitro (indicada pelo estudo conformacional) na atividade dos AMBH leishmanicidas.

5.5 Proposição e previsão da atividade de novos AMBH baseados nos modelos QSAR construídos

Uma das propostas mais interessantes de um modelo QSAR é a possibilidade de fornecer informações sobre a tendência de bioatividade para uma molécula ainda não sintetizada e/ou testada. Isto representa um importante passo durante o planejamento racional de novos fármacos por direcionar a pesquisa de uma série congênere, poupando custos e tempo neste processo.

Deste modo, baseado nos estudos QSAR da presente Tese, propusemos a estrutura de quatro AMBH que devem possuir melhores valores de atividade biológica. As moléculas reúnem características evidenciadas como importantes para a atividade dos AMBH aromáticos, baseadas no estudo considerando descritores *a priori* (Figura 5.12). Não obstante, a previsão de suas atividades deve ser confirmada através do modelo construído a partir da seleção de variáveis.

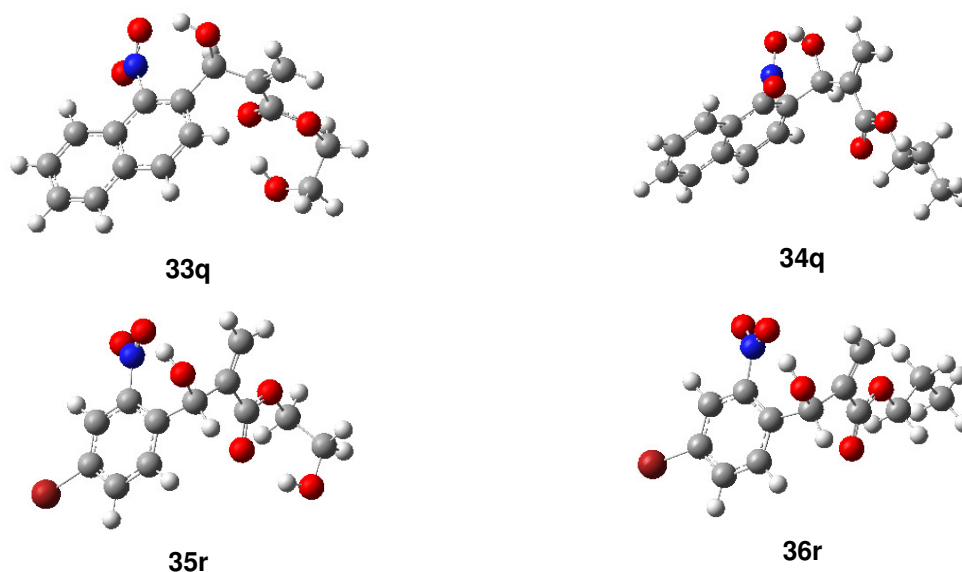
Figura 5.12: Estruturas dos AMBH propostas a partir dos estudos QSAR



Quanto à viabilidade sintética, os aldeídos necessários às correspondentes reações de Morita-Baylis-Hillman são comercialmente disponíveis, conforme busca prévia no banco de dados da empresa Sigma-Aldrich (<http://www.sigmaaldrich.com>). Os acrilatos para a efetivação das reações podem ser facilmente obtidos conforme experiência do nosso grupo de pesquisas nesta área.

As geometrias otimizadas dos AMBH propostos podem ser visualizadas na Figura 5.13.

Figura 5.13: Geometrias tridimensionais otimizadas dos AMBH propostos



Os descritores foram determinados (Tabela 5.16) e aplicados no primeiro modelo QSAR (equação 44), sendo então obtidos os valores de atividade previstos (pIC_{50prev}) para as moléculas propostas (Tabela 5.16). A tabela mostra que os

valores de atividade para os quatro compostos estão na mesma faixa ou são maiores que o mais ativo dos AMBH estudados (**17i**).

Tabela 5.16: Descritores *a priori* para as moléculas propostas e valores de atividade previstos.

MBHA	pIC _{50prev}	ELUKO (eV)	oNO ₂	Me	SLogP
33q	4,671	-0,1157	1	1	1,977
34q	4,994	-0,11739	1	1	3,526
35r	4,649	-0,12279	1	1	1,719
36r	4,957	-0,12228	1	1	3,243
17i	4,688*	-	-	-	-

*pIC₅₀ experimental

Um destaque especial merece ser feito para as moléculas **34q** e **36r**, as quais além de possuírem valor 1 para os dois índices de fragmentos, reúnem outras duas características essenciais que podem explicar seu maior valor de atividade em relação a todas as outras 34 estruturas deste trabalho: os efeitos sinérgicos dos valores de *ELUKO* e *SLogP*. Em primeiro lugar, a presença do bromo em posição *para* (**36r**) tende a retirar elétrons do anel aromático, o que como efeito geral poderia diminuir a densidade eletrônica ao nível do grupo nitro, efeito que pode ser interpretado através do abaixamento da energia do LUKO. Além disso, este grupo volumoso confere uma característica mais lipofílica ao composto onde está inserido, efeito bastante conhecido em química medicinal (BARREIRO e FRAGA, 2008), o que pode ser responsável pelo seu maior *SLogP*. Em adição, apesar de apresentar valor de *ELUKO* relativamente maior, o efeito lipofílico do grupo naftil em **34q**, aumentando seu *SLogP*, tende a melhorar o seu valor de atividade biológica. De fato, esta molécula apresenta o maior valor de pIC_{50prev}, o qual é semelhante a **36r**. Estas observações encorajam ainda a investigação de outros AMBH contendo outros grupos halogênio, inclusive em outras posições do anel aromático, bem como outros substituintes para as cadeias laterais da porção éster.

Em um contexto adicional, de forma a avaliar a potencial atividade das moléculas propostas sob outros aspectos, investigando a relação de coerência entre

os modelos QSAR construídos, os parâmetros **MATS8v**, **H4u**, **Mor14u**, **H-049** e **oNO2** foram também determinados (Tabela 5.17). Estes foram em seguida submetidos à equação 46 para a estimativa dos valores de **pIC_{50prev}** (Tabela 5.17).

Tabela 5.17: Descritores selecionados pelo OPS para as moléculas propostas e valores de atividade previstos.

MBHA	pIC _{50prev} ^a	pIC _{50prev}	H-049	Mor14u	MATS8v	oNO2	H4u
33q	4,671	4,563	0	0.652	-0.424	1	1.673
34q	4,994	4,851	0	1.032	-0.276	1	1.91
35r	4,649	4,361	0	0.454	-0.828	1	1.01
36r	4,957	4,583	0	0.674	-0.713	1	1.457
17i	4,688 ^b	4,688 ^b	-	-	-	-	-

^aA partir do modelo com descritores escolhidos *a priori*.

^bpIC₅₀ experimental

Podemos observar na terceira coluna da Tabela 5.17 que os valores de **pIC₅₀** previstos para as moléculas propostas situam-se na faixa dos compostos mais ativos, de forma semelhante aos resultados com o primeiro modelo. Em adição, o composto **34q** também se caracteriza como o mais ativo do novo conjunto.

Os resultados apresentados possuem considerável relevância por representarem coerentemente as expectativas geradas a partir da construção deste trabalho, fornecendo subsídios racionais para a síntese e avaliação de novos compostos com características estruturais otimizadas, potencialmente ainda mais ativos no combate à Leishmaniose.

CONCLUSÕES E PERSPECTIVAS

6. CONCLUSÕES E PERSPECTIVAS

6.1 Conclusões

A realização deste trabalho permitiu elucidar alguns aspectos estruturais dos Adutos de Morita-Baylis-Hillman estudados por nosso grupo de pesquisas, bem como relacionar alguns descritores moleculares à ação leishmanicida desta classe. Com base nos resultados apresentados, podemos elencar as seguintes conclusões:

- Ligações de Hidrogênio Intramoleculares podem controlar as conformações de Adutos de Morita-Baylis-Hillman estudados por nosso grupo de pesquisas quando há possibilidade de ocorrência, observação realizada após os cálculos DFT/PCM, avaliada por parâmetros estruturais (distância doador-aceitador) e baseados em QTAIM;
- A ligação de hidrogênio $\text{NO}_2 \cdots \text{HO}$, caracterizada por métodos teóricos e suportada por dados espectroscópicos experimentais, mostrou ser um importante fator para o aumento da atividade leishmanicida dos AMBH's estudados, já que todos os dados indicaram haver aumento da eletrofilicidade do grupo nitro e assim do seu potencial de biorredução;
- As observações eletrônicas e estruturais *a priori* conduziram à obtenção de um modelo QSAR robusto e preditivo para a classe de compostos estudada;
- A partir do vetor de regressão do modelo PLS obtido pode-se concluir que, em geral, baixos valores da energia LUKO associados a valores maiores de SLogP bem como valores unitários (presença) dos descritores (*Me*) e (*oNO₂*) podem ser responsáveis por um incremento no valor da atividade leishmanicida de AMBH's aromáticos.
- O cálculo de descritores de naturezas diversas, através da plataforma E-DRAGON, seguido do método de seleção de variáveis OPS, conduziu a uma segunda proposta de modelo QSAR que se mostrou robusto e preditivo para a atividade leishmanicida da classe de compostos estudada;

6.2 Perspectivas

O presente trabalho abre outras possibilidades de investigações a serem realizadas. Assim, como perspectivas deste trabalho, podemos elencar:

- A proposição de várias outras estruturas congêneres de Adutos de Morita-Baylis-Hillman aromáticos (outros grupos halogênio, regioisômeros, porções na cadeia lateral), o cálculo dos descritores evidenciados neste trabalho e o uso destes nos modelos propostos para previsão de sua atividade leishmanicida;
- A síntese das estruturas orientadas pelos resultados obtidos na etapa anterior, de forma a avaliar experimentalmente o comportamento dos compostos que se apresentem como potencialmente mais bioativos contra o parasita *Leishmania*;

REFERÊNCIAS
BIBLIOGRÁFICAS

7. REFERÊNCIAS BIBLIOGRÁFICAS

ABREU, R. M.; FERREIRA, I. C.; QUEIROZ, M. J.; QSAR model for predicting radical scavenging activity of di(hetero)arylamines derivatives of benzo[b]thiophenes. **European Journal of Medicinal Chemistry**. 44, 1952–1958, 2009.

ALVAR, J.; VELEZ, I. D.; BERN, C.; HERRERO, M.; DESJEUX, P.; CANO, J.; JANNIN, J.; DEN BOER, M.; Leishmaniasis Worldwide and Global Estimates of Its Incidence. **PLoS ONE**, 7, e35671, 2012.

ARUNAN, E.; DESIRAJU, G. R.; KLEIN, R. A.; SADLEJ, J.; SCHEINER, S.; ALKORTA, I.; CLARY, D. C.; CRABTREE, R. H.; DANNENBERG, J. J.; HOBZA, P.; KJAERGAARD, H. G.; LEGON, A. C.; MENNUCCI, B.; NESBITT, D. J. Defining the hydrogen bond: an account. **IUPAC recommendations 2011**, 83, 1619–1633, 2011.

BADER, R. F. W.; Atoms in Molecules. A Quantum Theory. **Clarendon Press: Oxford**, 1990.

BADER, R. F. W.; BEDDALL, P. M.; Virial Field Relationship for Molecular Charge Distributions and the Spatial Partitioning of Molecular Properties **Journal of Chemical Physics**, 56, 3320–3329, 1972.

BARBOSA, T. P. Síntese de novos Adutos aromáticos de Morita-Baylis-Hillman: hibridação molecular como ferramenta para obtenção de drogas leishmanicidas. **Tese de doutorado**, João Pessoa, Programa de Pós-Graduação em Química, UFPB, 2010.

BARBOSA, T. P., JUNIOR, C. G. L., SILVA, F. P. L., LOPES, H. M., FIGUEIREDO, L. R. F., SOUSA, S. C. O., BATISTA, G. N., SILVA, T. G., SILVA, T. M. S., OLIVEIRA, M. R., VASCONCELLOS, M. L. A. A., Improved synthesis of seven aromatic Baylis-Hillman adducts (BHA): Evaluation against *Artemia salina* Leach. and

Leishmania chagasi, **European Journal of Medicinal Chemistry**, 44, 1726–1730, 2009.

BARREIRO, Eliezer J.; FRAGA, Carlos Alberto Manssour. A questão da inovação em fármacos no Brasil: proposta de criação do programa nacional de fármacos (Pronfar). **Química Nova**, 28, S53–S63, 2005.

BARREIRO, E. J.; FRAGA, C. A. M.. Química Medicinal: As bases moleculares de ação dos fármacos. 2ª Edição. **Artmed**, 2008.

BASAVAIHAH, D.; REDDY, B. S.; AND BADSARA, S. S.; Recent contributions from the Baylis-Hillman reaction to organic chemistry. **Chemical Review**, 110, 5447–5674, 2010.

BAYLIS, A. B.; HILLMAN, M. E. D., **German Patent** 2155113, 1972.

BECKE, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. **Physics Review A**, 38, 3098–3100, 1988.

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

BEN-NAIM, A.; MARCUS, Y. J. Solvation thermodynamics of nonionic solutes. **Chemical Physics** 81, 2016–2027, 1984.

BONDI, A. Van der Waals Volume and Radii. **Journal of Physical Chemistry**, 68, 441–452, 1964.

BOROTA, A.; MRACEC, M.; GRUIA, A.; RAD-CURPAN, R.; OSTOPOVICI-HALIP, L.; MRACEC, M.. A QSAR study using MTD method and Dragon descriptors for a series of selective ligands of α_2C adrenoceptor. **European Journal of Medicinal Chemistry**, 46, 877–884, 2011.

BRASIL. **Ministério da saúde. Secretaria de Vigilância em Saúde.** Disponível em: <http://portal.saude.gov.br/portal/arquivos/pdf/2_lv_casos_14_10_10.pdf>. Acessado em: 16 de abril de 2012.

CAMARGO, E. P. Doenças tropicais, **Estudos Avançados**, 22, 95–110, 2008. Disponível em: <<http://www.scielo.br/pdf/ea/v22n64/a07v2264.pdf>> Acessado em: 16 de abril de 2012.

CCDC - **The Cambridge Crystallographic Data Centre**, <<http://beta-www.ccdc.cam.ac.uk/pages/Home.aspx>> Acessado em 10 de agosto de 2012

CDC - **Centers for Diseases Control & Prevention Center for Global Health**, <<http://www.dpd.cdc.gov/DPDx/html/Leishmaniasis.htm>> Acessado em: 15 de abril de 2012.

CHIRAC, P.; TORREELE, E. **Lancet**, 1560-1561, 2006. Disponível em: <<http://www.dndi.org.br/pt/doencas-negligenciadas/panorama.html>> Acessado em: 05 de abril de 2012.

COELHO, F.; ALMEIDA, W. P., Reação de Baylis-Hillman: Uma estratégia para a preparação de intermediários multifuncionalizados para síntese orgânica, **Química Nova**, 23, 98–101, 2000.

CRAMER, C. J. **Essentials of Computational Chemistry: Theories and Models**; 2nd ed. **John Wiley & Sons: Chichester**, U.K., 2004.

CRUM-BROWN, A.; FRASER, T.R. **Transactions of the Royal Society of Edinburgh**, 25, 151, 1868.

DAVIES C. R.; KAYE P.; CROFT S.L.; SUNDAR, S. Leishmaniasis: new approaches to disease control. **BMJ**, 326, 377–382, 2003.

DENNINGTON, R. KEITH, T. MILLAM, J. GaussView, Version 5, **Semichem Inc.**, Shawnee Mission KS, 2009.

DE OLIVEIRA, D. B.; GAUDIO, A. C.; BuildQSAR: A New Computer Program for QSAR Analysis. **Quantitative Structure - Activity Relationship**, 19, 599–601, 2000.

DE SOUZA, R. O. M. A., PEREIRA, V. L. P., MUZITANO, M. F., FALCÃO, C. A. B., ROSSI-BERGMANN, B., FILHO, E. B. A., VASCONCELLOS, M. L. A. A., High selective leishmanicidal activity of 3-hydroxy-2-methylene-3-(4-bromophenyl) 103 propanenitrile and analogous compounds, **European Journal of Medicinal Chemistry**, 42, 99–102, 2007.

DESJEUX P. Leishmaniasis: current situation and new perspectives. **Comparative Immunology Microbiology and Infect Diseases**, 27, 305–318, 2004.

DEWAR, M. J. S.; ZOEBISCH, E. G.; HEALY, E. F.; STEWART, J. J. P. The development and use of quantum mechanical molecular models. 76. AMI: a new general purpose quantum mechanical molecular model **Journal of American Chemical Society**, 107, 3902–3909, 1985.

EBRAHIMI, A.; HABIBI KHORASSANI, S.M.; DELARAMI, H. Estimation of individual binding energies in some dimers involving multiple hydrogen bonds using topological properties of electron charge density. **Chemical Physics**, 365, 18–23, 2009.

FERREIRA, M. M. C.; ANTUNES, A. M.; MELGO, M. S.; VOLPE, P. L. O. Quimiometria I: calibração multivariada, um tutorial. **Química Nova**, 22(5), 724–731, 1999.

FERREIRA, M. M. C.; Multivariate QSAR. **Journal of Brazilian Chemical Society**, 13, 742–753, 2002.

FILHO, E. B. A.; MORAES, I. A.; WEBER, K. C.; ROCHA, G. B.; VASCONCELLOS, M. L. A. A. DFT/PCM, QTAIM, ¹H NMR Conformational Studies and QSAR Modeling

of Thirty-two anti-Leishmania amazonensis Morita-Baylis-Hillman Adducts. **Journal of Molecular Structure**, 1022, 72–80, 2012.

FREE, S. M. Jr.; WILSON, J. W. **Journal of Medicinal Chemistry** 7, 395–399, 1964.

FREITAS, H. F.; PAZ, O. S.; CASTILHO, M. S. Estudos de QSAR 3D para um conjunto de inibidores de butirilcolinesterase humana. **Quimica Nova**, 32, 2114–2121, 2009.

FUKUI, K., YONEZAWA, T. AND SHINGU, H. **The Journal of Chemical Physics**, 20, 722, 1952.

Gaussian 09, Revision A.1, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox, **Gaussian, Inc.**, Wallingford CT, 2009.

GOYAL, R. K.; DUREJA, H.; SINGH, G.; MADAN, A. K.; Models for anti-inflammatory activity of 8-substituted-4-anilino-6- aminoquinoline-3-carbonitriles. **Medicinal Chemistry Research**. 21, 1044–1055, 2012.

GRAMATICA, P. **A short history of QSAR evolution**. Disponível em: <http://www.qsarworld.com/Temp_Fileupload/Shorthistoryofqsar.pdf.> Acessado em 01/12/2011.

GUPTA, L; PATEL, A; KARTHIKEYAN, C; TRIVEDI, P. QSAR studies on dihydro-alkoxy-benzyloxypyrimidines (DABOs) derivatives, a new series of potent, broad-spectrum non-nucleoside reverse transcriptase inhibitors. **Journal of Current Pharmaceutical Research**, 01, 19–25, 2010.

HAMMETT, L. P.; Some relations between reaction rates and equilibrium constants. **Chemal Reviews**, 17, 125–136, 1935.

HAMMLET, L. P. Physical Organic Chemistry, 2nd ed., **McGraw-Hill**, New York, 1970.

HANSCH, C.; FUJITA, T. Rho sigma pi analysis: A method for the correlation of biological activity and chemical structure. **Journal of American Chemical Society** 1964, 86, 161–162.

HANSCH, C.; MALONEY, P. P.; FUJITA, T.; MUIR, R. M. Correlation of biological activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. **Nature**, 194, 178–180, 1962.

HOHENBERG, P.; KOHN, W. Inhomogeneous Electron Gas **Physcal Review**, 136, B864–B871, 1964.

HUANG, Z.; YU, L.; DAI, Y.; WANG, H. Hydrogen bonding interactions in cysteine–urea complexes: Theoretical studies of structures, properties and topologies. **Journal of Molecular Structure: THEOCHEM**, 960, 98–105, 2010.

JUNIOR, C. G. L., DE ASSIS, P. A. C., SILVA, F. P. L., SOUSA, S. C. O., DE ANDRADE, N. G., BARBOSA, T. P., NERÍS, P. L. N., SEGUNDO, L. V. G., ANJOS, I. C., CARVALHO, G. A. U., ROCHA, G. B., OLIVEIRA, M. R., VASCONCELLOS, M.

L. A. A., Efficient synthesis of 16 aromatic Morita–Baylis–Hillman adducts: Biological evaluation on *Leishmania amazonensis* and *Leishmania chagasi*, **Bioorganic Chemistry**, 38, 279–284, 2010.

KOHN, W.; SHAM, L. J. Self-consistent equations including exchange and correlation effects. **Physical Review** 140, A1133–A1138, 1965.

KIRALJ, R. and FERREIRA, M. M. C. Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. **Journal of the Brazilian Chemical Society** 20, 770–787, 2009.

KUNBINIY, H. QSAR and 3D QSAR in drug design. **Drug Discovery Today**. Vol. 2, No. 11, 1997.

KUNDU, M. K., SUNDAR, N., KUMAR, S. K., BHAT, S. V., BISWAS, S. V. N., Antimalarial activity of 3-hydroxyalkyl-2-methylene-propionic acid derivatives, **Bioorganic & Medicinal Chemistry Letters**, 9, 731–736, 1999.

PAULING, L. The Nature of the Chemical Bond, third ed., **Cornell University**, USA, 1960.

LEE, C. YANG, W. PARR, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. **Physical Review B**, 37, 785–789, 1988.

MATTA, C. F.; BOYD, R. J. (Eds.), The Quantum Theory of Atoms in Molecules: From SolidState to DNA and Drug Design, **Wiley-VCH, Weinheim** 2007.

MEYER, HH. Zur Theorie der Alkoholnarkose. **Archives of Experimental Pathological Pharmacology** 42, 109–118, 1899.

MORAN, M., GUZMAN, J., ROPARS, A. L., JAMESON, N., Neglected Disease Research and Development: How Much Are We Really Spending? **Plos Medicine**, 6

(2), 2009. Disponível em: <http://www.dndi.org.br/Portugues/doencas_negligenciadas.aspx>. Acessado em: 05 de abril de 2012.

MORGON, N. H.; COUTINHO, K.; Métodos de Química Teórica e Modelagem Molecular. **Editora livraria da física**, 2007.

MORITA, K.; SUZUKI, Z.; HIROSE, H., A Tertiary Phosphine-catalyzed Reaction of Acrylic Compounds with Aldehydes, **Bulletin of the Chemical Society of Japan**, 41, 2815, 1968.

MUCHMORE, S. W. EDMUNDS, J. J. STEWART, K. D. and HAJDUK, P. J. Cheminformatic Tools for Medicinal Chemists. **Journal of Medicinal Chemistry**, 53, 4830–4841, 2010.

NAGY, P. I.; TEJADA, F. R.; KIM, Y. W.; MESSER JR., W. S. Theoretical studies of the solvent effect on conformational equilibria and atomic charges for isolated and hydrogen-bonded dimethoxy thiadiazoles. **Computational and Theoretical Chemistry**, 966, 194–206, 2011.

NGURE, P. K; KIMUTAI, A; ANG', Z. W. N.; RUKUNGA, G; TONUI, W. K. A review of Leishmaniasis in Eastern Africa. **Journal of Nanjing Medical University**, 23, 79–86, 2009.

NOGUEIRA, L. J.; MONTANARI, C. A.; DONNICI, C. L., Histórico da evolução da química medicinal e a importância da lipofilia: de Hipócrates e Galeno a Paracelsus e as contribuições de Overton e de Hansch. **Revista Virtual Química**, 1, 227–240, 2009.

OLIVEIRA, B. G.; ARAUJO, R. C. M. U.; RAMOS, M. N. A topologia molecular QTAIM e a descrição mecânico-quântica de ligações de Hidrogênio e ligações de di-hidrogênio. **Química Nova**, 33, 1155–1162, 2010.

ONSAGER, L. Electric Moments of Molecules in Liquids. **Journal of American Chemical Society**, 58, 1486–1493, 1936

OVERTON, C.E. Studien Uber die Narkose, **Fischer**, Jena, Germany, 1901.

PELIER, P. L. A. MORPHY, **UMIST**, England, EU, 1998.

PARR, R. G. YANG, W. Density Functional Theory, **Oxford University Press**, 1989.

PAULA, F. R.; SERRANO, S. H. P.; TAVARES, L. C. Aspectos mecanísticos da bioatividade e toxicidade de nitrocompostos. **Quimica Nova** 32, 1013–1020, 2009.

PAULING, L.; Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry. **Cornell University Press**, Nova York, 1939.

PISCOPO, T. V.; AZZOPARDI, C. M. Leishmaniasis, Review. **Postgrad Medical Journal**, 82, 649–657, 2006

PLANCHE, A. S.; SCOTTI, M. T.; LÓPEZ, A. G.; EMERENCIANO, V. P.; PÉREZ, E. M.; URIARTE, E. Design of novel antituberculosis compounds using graph-theoretical and substructural approaches. **Molecular Divers**, 13, 445–458, 2009.

PLIEGO JR., J. R.. Modelos contínuos do solvente: fundamentos. **Quimica Nova**. 29, 535–542, 2006.

RATH, S.; TRIVELIN, L. A.; IMBRUNITO, T. R.; TOMAZELA, D. M.; JESÚS, M. N.; MARZAL, P. C.; ANDRADE JUNIOR, H. F.; TEMPONE, A. G. Antimoniais empregados no tratamento da leishmaniose: estado da arte. **Química Nova**, 26, 550–555, 2003.

REITHINGER, R.; DUJARDIN J-C; LOUZIR, H.; PIRMEZ, C.; ALEXANDER, B.; BROOKE, R. S. Cutaneous leishmaniasis. **The Lancet**, 7, 581–596, 2007.

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

ROWLAND, R. S.; TAYLOR, R. Intermolecular Nonbonded Contact Distances in Organic Crystal Structures: Comparison with Distances Expected from van der Waals Radii. **Journal of Physical Chemistry**. 100, 7384–7391, 1996.

SANDES, J. M., BORGES, A. R., JUNIOR, C. G. L, SILVA, F. P. L., CARVALHO, G. A. U., ROCHA, G. B., VASCONCELLOS, M. L. A. A., FIGUEIREDO, R. C. B. Q., 3-Hydroxy-2-methylene-3-(4-nitrophenylpropanenitrile): A new highly active compound against epimastigote and trypomastigote form of *Trypanosoma cruzi*, **Bioorganic Chemistry**, 38, 190–195, 2010.

SCHRÖDINGER, E. An ondulatory theory of the mechanics of atoms and molecules. **The Physical Review**. 28, 1049–1070, 1926.

SILVA, F.P.L.; DE ASSIS, P.A.C.; JUNIOR, C. G. L.; DE ANDRADE, N. G.; DA CUNHA, S. M. D.; OLIVEIRA, M. R. e VASCONCELLOS, M. L. A. A. Synthesis, evaluation against *Leishmania amazonensis* and cytotoxicity assays in macrophages of sixteen new congeners Morita–Baylis–Hillman adducts **European Journal of Medicinal Chemistry** 46, 4295–4301, 2011.

Simulations Plus, Inc. (www-simulations-plus.com), **Lancaster**, CA, USA.

SLATER, J.C. The theory of complex spectra, **Physical Review**, 34, 1293–1323, 1929.

STRAND LIFE SCIENCE P. LTDA, 2007. Disponível em <http://www.strandls.com/sarchitect/documents/manual_html/descctheory.html> Acessado em 20 de agosto de 2012.

SVS - Secretaria de Vigilância em Saúde – Ministério da Saúde. **Manual de Vigilância da Leishmaniose Tegumentar Americana**. 2ª Edição. Série A. Normas e Manuais Técnicos. Brasília: SVS / MS, 2007.

TAFT, R. W. Polar and Steric Substituent Constants for Aliphatic and o-Benzoate Groups from Rates of Esterification and Hydrolysis of Esters **Journal of the American Chemical Society**, 74, 3120–3128, 1952.

TALETE SRL, **DRAGON for Windows (Software for Molecular Descriptor Calculations) Version 5.5** – 2007 – <http://www.taletе.mi.it/>

TAKAHASHI, O.; SAITO, K.; KOHNO, Y. The conformation of 1-alkyl-2-phenylpropan-1-ols studied by ab initio MO calculations. Relevance of the CH/ π and OH/ π hydrogen bonds. **Bulletin of the Chemical Society of Japan**. 76, 2167–2173, 2003.

TAVARES, L. C.; Qsar: The Hansch's approach. **Química Nova**. 27, 631–639, 2004.

TEÓFILO R. F., MARTINS J. P., FERREIRA M. M. C., Sorting variables by using informative vectors as a strategy for feature selection in multivariate regression, **Journal of Chemometrics**, 23, 32–48, 2009.

TEÓFILO, R. F.; FERREIRA, M. M. C.. Building a regression vector for variable selection. **Em: 11th International Conference on Chemometrics in Analytical Chemistry, Montpellier. Book of Abstracts**, 1, 15–20, 2008.

TETKO, I. V.; GASTEIGER, J.; TODESCHINI, R.; MAURI, A.; LIVINGSTONE, D.; ERTL, P.; PALYULIN, V. A.; RADCHENKO, E. V.; ZEFIROV, N. S.; MAKARENKO, A. S.; TANCHUK, V. Y.; PROKOPENKO, V. V. Virtual computational chemistry laboratory - design and description, **Journal of Computer-Aided Molecular Design** 19, 453–463, 2005.

VASCONCELLOS, M. L. A. A., SILVA, T. M. S., CÂMARA, C. A., MARTINS, R. M., LACERDA, K. M., LOPES, H. M., PEREIRA, V. L. P., DE SOUZA, R. O. M. A., CRESPO, L. T. C., Baylis-Hillman adducts with molluscicidal activity against *Biomphalaria glabrata*, **Pest Management Science**, 62, 288–292, 2006.

VCCLAB – Virtual Computational Chemistry Laboratory. **Dragon Descriptors**. Disponível em: < http://www.vcclab.org/lab/indexhlp/dragon_descr.html > acessado em 13 de agosto de 2012

VIANNA, G. O.; **Anais do 7º Congresso Brasileiro de Medicina e Cirurgia** 4, 426, 1912.

VOSKO S.J.; WILK L.; NUSAIR, M. Accurate spin-dependent electron liquid correlation energies for local spin-density calculations - a critical analysis. **Canadian Journal of Physics**, 58, 1200–1211, 1980.

WHO - World Health Organization, **Division of Control of Tropical Diseases**, <<http://www.who.int/topics/leishmaniasis/en/>> Acessado em: 15 de abril de 2012.

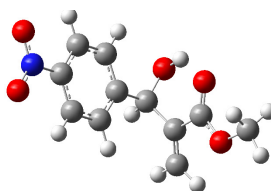
WHO/NTD/IDM HIV/AIDS, Tuberculosis and Malaria (HTM), **World Health Organization**, Disponível em <http://www.who.int/leishmaniasis/leishmaniasis_maps/en/>. 2010 Acessado em: 15 de abril de 2012.

WILDMAN, S. A.; CRIPPEN, G. M. Prediction of physicochemical properties by atomic contributions. **Journal of Chemical Information and Computer Science** 39, 868–873, 1999.

WOLD, S; SJÖSTRÖM, M; ERIKSSON, L. PLS-regression: a basic tool of chemometrics **Chemometric and Intelligent Laboratory Systems** 58: 109–130, 2001.

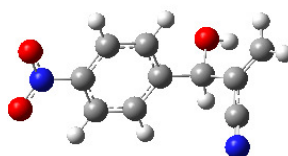
ANEXOS

ANEXO A - Coordenadas atômicas após otimização completa em DFT/PCM - B3LYP/6-31+G(d)



1a

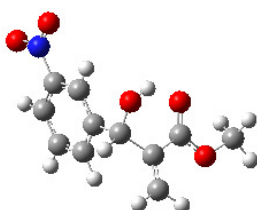
C	0.79311700	-1.04603600	1.10214600
C	0.29765800	0.12349500	1.68525100
C	0.88294200	1.36804100	1.44493500
C	1.99165100	1.43692700	0.60389500
C	2.51810400	0.28036600	0.01161800
C	1.90272600	-0.95679500	0.26763300
C	3.70992900	0.36041700	-0.95155400
C	4.76731300	-0.70139000	-0.66351700
O	4.28125500	1.66507200	-1.00610500
C	5.55004300	-0.50779500	0.59252700
C	5.01192700	-1.71974200	-1.49947200
N	-0.86589600	0.04157000	2.56620400
O	-1.36951400	-1.06854600	2.77300700
O	-1.29881600	1.08549400	3.06765800
O	5.52346200	0.54145000	1.23178000
O	6.28541900	-1.56368700	0.94979300
C	7.11198700	-1.41690800	2.12913200
H	0.31898100	-1.99964400	1.29971700
H	0.47753600	2.26071600	1.90559200
H	2.44912400	2.39880000	0.40235900
H	2.29190600	-1.86276300	-0.18694800
H	3.32568100	0.17463600	-1.96045700
H	4.77149900	1.78882100	-0.16955700
H	5.76975500	-2.46560800	-1.28395600
H	4.45600300	-1.82918400	-2.42696200
H	7.62722400	-2.37055800	2.23385600
H	6.48649700	-1.21742000	3.00179200
H	7.82708700	-0.60382900	1.98642900



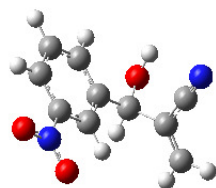
2a

C	0.00100600	-0.59520300	0.70840200
C	0.22633900	0.23882200	1.80471800
C	1.45368000	0.87764400	2.01195700
C	2.47831100	0.66570300	1.09466900
C	2.27869500	-0.16429800	-0.02013400
C	1.03795200	-0.78807700	-0.20418800
C	3.39999000	-0.38189800	-1.02782100
C	4.55918100	-1.18155200	-0.40670400
O	3.83653100	0.88996600	-1.48133500

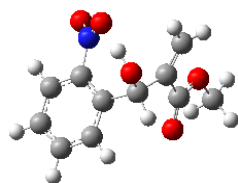
C	4.26005800	-2.54031300	-0.05042900
C	5.79197100	-0.69218100	-0.20261800
N	-0.85475800	0.45192700	2.76725100
O	-1.93337000	-0.12154000	2.57882300
O	-0.64661100	1.19753100	3.73090300
N	3.98240400	-3.63894100	0.21959400
H	-0.96215800	-1.07796600	0.56651200
H	1.60245400	1.52179600	2.87437500
H	3.43840800	1.15546100	1.24882400
H	0.87395900	-1.42916300	-1.07036300
H	3.00122900	-0.97205900	-1.86766300
H	4.31094500	0.77677500	-2.34243800
H	6.58094800	-1.29846400	0.23477900
H	6.02363400	0.33313000	-0.47631300

**3b**

C	2.50512900	0.12462500	-0.19798200
C	2.53831700	1.20025400	-1.08680700
C	1.42465300	2.03673700	-1.13919800
C	0.31567400	1.78992600	-0.32484400
C	0.29374300	0.70074100	0.56138900
C	1.40783700	-0.13698500	0.62588200
C	-0.90874800	0.47217100	1.48965200
C	-2.24234900	0.58827000	0.75885200
C	-2.55775500	-0.51656100	-0.19417100
C	-3.10021800	1.59288000	0.98454000
O	-0.81136200	-0.75488400	2.20917600
O	-3.59228800	-0.26718400	-1.00118700
O	-1.92268700	-1.56839000	-0.22019000
C	-3.98489800	-1.32511400	-1.90782300
N	3.66812100	-0.76517600	-0.12174300
O	3.62802200	-1.72069900	0.66055400
O	4.64089600	-0.52151800	-0.84405500
H	3.40596700	1.37256800	-1.71135900
H	1.41979800	2.88392800	-1.81806400
H	-0.54127500	2.45494300	-0.38321100
H	1.42714500	-0.97607000	1.30859300
H	-0.88729800	1.25387600	2.25606900
H	-4.04766800	1.66366400	0.46063900
H	-2.87251000	2.37237300	1.70716300
H	-0.99987800	-1.46831100	1.56787300
H	-4.83588000	-0.92545700	-2.45735500
H	-3.16239800	-1.56069100	-2.58640500
H	-4.27181300	-2.21485100	-1.34307500

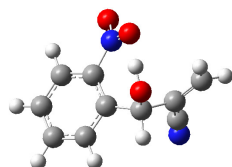
**4b**

C	-0.02515200	-0.02712400	0.05819500
C	-0.01523800	-0.03723900	1.45567000
C	1.21981200	-0.01880500	2.09837100
C	2.40736800	0.00595800	1.35710800
C	2.38264800	0.00637400	-0.04434700
C	1.14639200	-0.00427500	-0.70007900
C	3.67523800	0.07135600	-0.86449100
C	3.71334100	-0.96428000	-1.99079700
O	4.85010300	0.00809100	-0.06637100
C	3.55142200	-2.34330200	-1.61675900
C	3.92665700	-0.64294600	-3.27734100
N	-1.31357300	-0.04107100	-0.63904200
O	-1.31674500	-0.04463200	-1.87473000
O	-2.34729100	-0.04948800	0.03774200
N	3.42094000	-3.44996800	-1.27816000
H	-0.94463300	-0.05754600	2.01787500
H	1.26099500	-0.02160900	3.18674300
H	3.36055400	0.03256500	1.88045700
H	1.08310200	0.01375300	-1.78366100
H	3.72515000	1.05611900	-1.34897700
H	4.93255000	-0.86806300	0.38827400
H	3.97744900	-1.39772100	-4.05857800
H	4.05032300	0.39756300	-3.57091300

**5c**

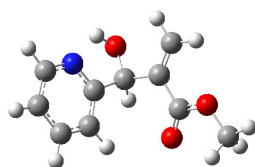
C	0.00823400	-0.14388300	0.06338500
C	0.02734200	-0.06273000	1.45196500
C	1.25059900	0.07615100	2.11176500
C	2.43958200	0.12490600	1.38173300
C	2.46542600	0.05971300	-0.01812000
C	1.20969200	-0.06571300	-0.65068300
C	3.84445000	0.09444900	-0.69652200
C	4.14126400	-1.12806400	-1.56926900
O	4.08435300	1.33011300	-1.35797200
C	4.15804300	-2.43417100	-0.84406600
C	4.49014700	-1.05049600	-2.86007800
N	1.06598900	-0.09287800	-2.10945400
O	1.83796400	0.58983300	-2.79940400
O	0.15949600	-0.76991100	-2.59483100
O	4.20368800	-2.52439900	0.37529700
O	4.13171800	-3.50099300	-1.65980800
C	4.21037800	-4.79928100	-1.03004800
H	-0.92548600	-0.24217500	-0.47776700
H	-0.90349900	-0.10341000	2.00865600

H	1.28373000	0.13902300	3.19543300
H	3.38372800	0.20168900	1.91242700
H	4.57006100	0.07369500	0.12077600
H	3.36783000	1.45254300	-2.00848800
H	4.74212300	-1.93863600	-3.42917400
H	4.54920600	-0.09564900	-3.37116900
H	4.17781600	-5.51590000	-1.84986300
H	3.36134400	-4.94419200	-0.35809700
H	5.14599100	-4.89280900	-0.47366200



6c

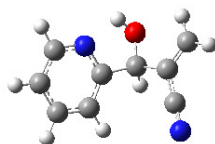
C	2.36992100	0.62304600	0.43971500
C	3.17995800	-0.48937100	0.23117900
C	2.63723300	-1.62836500	-0.36498200
C	1.29335100	-1.64643100	-0.74665300
C	0.44706300	-0.54399800	-0.56799000
C	1.03150800	0.59365100	0.03408100
C	-1.01467500	-0.71904000	-1.00617400
C	-2.02546500	-0.55675300	0.14354800
C	-1.76132100	-1.32963300	1.32577900
O	-1.35895100	0.04613700	-2.14428800
C	-3.14071500	0.18499000	0.07091100
N	-1.52513500	-1.97212600	2.26826900
N	0.28394400	1.83427800	0.26346200
O	-0.64056200	2.12958000	-0.50930700
O	0.61752400	2.55194000	1.20567300
H	2.76577500	1.52268100	0.89553200
H	4.22278500	-0.46134500	0.53029700
H	3.25361100	-2.50644600	-0.53254900
H	0.88381400	-2.54868400	-1.19136200
H	-1.09899800	-1.75858600	-1.33781900
H	-1.17289000	0.98016400	-1.92942100
H	-3.84531700	0.23343400	0.89468300
H	-3.36566600	0.75388600	-0.82458000



7d

C	0.07354100	0.02609700	-0.26867600
C	-0.04251300	-0.33118300	1.07587100
C	1.12944700	-0.51278000	1.81561300
C	2.46357100	-0.02975000	-0.00081900
C	1.34881100	0.18341100	-0.81712600
C	3.88970500	0.16262000	-0.52063500
C	4.65445900	-1.15900000	-0.60066200
O	4.57708600	1.13482800	0.26079300
C	4.19766000	-2.09619800	-1.66878600

C	5.70347400	-1.45146600	0.17953200
H	-1.01519300	-0.46362900	1.54759300
H	1.08266700	-0.78573400	2.87102500
H	1.47970800	0.46700900	-1.85985300
H	3.83151500	0.58411600	-1.53057200
H	4.44163600	0.94205400	1.21825400
H	6.23355100	-2.39439400	0.08855900
H	6.06570800	-0.74124500	0.91763400
H	-0.81371300	0.18510900	-0.88155800
N	2.35700300	-0.37193800	1.29380300
O	3.42748700	-1.77300900	-2.56424000
O	4.71920200	-3.33105700	-1.56570900
C	4.36575900	-4.27098400	-2.60397500
H	4.69820600	-3.90110200	-3.57752200
H	4.88737200	-5.19216600	-2.34377800
H	3.28448700	-4.43049300	-2.61713400

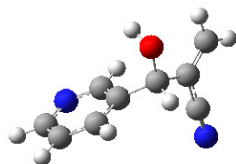
**8d**

C	0.05520700	-0.29957100	0.10361700
C	0.06923300	0.07054500	1.44993000
C	1.28881500	0.43149800	2.02575800
C	2.42677900	0.05848500	0.05705100
C	1.25474000	-0.30349500	-0.61122600
C	3.80116300	0.06848800	-0.62209900
C	4.44276100	-1.33354000	-0.58849900
O	4.64508400	1.01510400	-0.00386200
C	3.79107700	-2.35777600	-1.35527600
C	5.56294800	-1.62152400	0.09158200
N	3.24439500	-3.16387900	-1.99446600
H	-0.84132300	0.09046200	2.04037600
H	1.34469200	0.74098900	3.06670700
H	1.27638100	-0.57235400	-1.66296400
H	3.68127100	0.35971800	-1.67216600
H	4.30522100	1.11715600	0.91061300
H	5.98938700	-2.61930500	0.09171200
H	6.07044700	-0.84489000	0.65338000
H	-0.87506200	-0.57309100	-0.38599600
N	2.44422600	0.42500700	1.34565300

**9e**

N	2.58590000	-1.23699000	1.22643000
C	3.39474300	-1.28078100	0.15540500
C	3.21436600	-0.47190800	-0.96718500
C	2.14404700	0.42411500	-0.98373100
C	1.28255300	0.48079300	0.11589700

C	1.55965300	-0.37353900	1.19318500
C	0.11503700	1.47354900	0.17970800
C	-1.17329500	0.83122200	0.68349700
C	-1.82389000	-0.13552700	-0.24944300
C	-1.71537600	1.14052200	1.86953200
O	-0.08974200	2.14617900	-1.06282500
O	-2.79325200	-0.86808100	0.30626000
O	-1.49979800	-0.23577400	-1.43066600
C	-3.50344200	-1.78081800	-0.56399300
H	4.21860500	-1.98981700	0.20206700
H	3.89799900	-0.54135300	-1.80821900
H	1.97976500	1.07397400	-1.83713300
H	0.92319700	-0.36427400	2.07574500
H	0.38363100	2.26476600	0.88735800
H	-2.63257800	0.67699600	2.21736500
H	-1.24916500	1.87235500	2.52426100
H	-0.48262700	1.48930100	-1.67077000
H	-4.24792500	-2.25709800	0.07226700
H	-3.98325000	-1.22843700	-1.37488600
H	-2.81299400	-2.52184500	-0.97245300

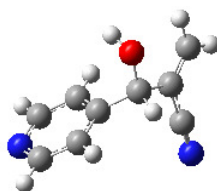


10e

N	0.30537600	0.07061800	0.69062500
C	1.36358600	-0.00710300	-0.13220500
C	1.88874700	1.09616400	-0.81116000
C	1.28667800	2.33931800	-0.62509300
C	0.18086100	2.44368600	0.22887800
C	-0.26607900	1.27473000	0.85790600
C	-0.44790100	3.80537000	0.49445000
C	-1.96652200	3.74834300	0.70647500
O	0.20070300	4.48877800	1.56097100
C	-2.60501000	4.25338100	1.77331300
C	-2.72819400	3.16417500	-0.36325600
N	-3.32259900	2.69760600	-1.24981300
H	1.80784200	-0.99583800	-0.25540000
H	2.74713400	0.97746300	-1.47104200
H	1.66661300	3.22381000	-1.13878200
H	-1.12423900	1.29865800	1.53124700
H	-0.27481700	4.43533400	-0.39066400
H	0.35091500	3.87230600	2.32001800
H	-3.68761300	4.21334900	1.86096700
H	-2.04199000	4.72309300	2.57458000

**11f**

C	0.39340000	0.26750600	0.83118100
N	1.43386400	0.01186300	0.02088000
C	1.77923800	0.98186400	-0.84006800
C	1.12579200	2.21329000	-0.92851400
C	0.04843200	2.47568900	-0.07694400
C	-0.32123600	1.46573000	0.82104600
C	-0.73765900	3.78968800	-0.15923200
C	-1.03871900	4.37835300	1.21488500
O	-0.11153500	4.74546500	-1.01312300
C	-2.27973000	4.43431100	1.71732600
C	0.13545300	4.93091500	1.95224800
O	1.22426700	5.11965300	1.41471800
O	-0.10436100	5.21121100	3.23618700
C	0.98044600	5.80288600	3.98991000
H	0.11801400	-0.52437000	1.52481200
H	2.62132500	0.76571600	-1.49465400
H	1.45265700	2.95135300	-1.65266500
H	-1.15061800	1.60005500	1.50921900
H	-1.69756800	3.56678600	-0.63721900
H	0.67679600	5.06913300	-0.53464700
H	-2.48143900	4.85631900	2.69632500
H	-3.12942500	4.05754300	1.15383700
H	1.83523000	5.12362900	4.01159000
H	1.27060200	6.75631700	3.54296400
H	0.58064000	5.95195600	4.99184400

**12f**

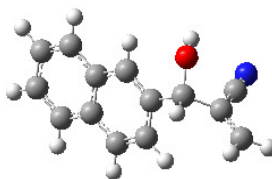
C	-2.18049300	0.44488000	1.35265500
N	-3.15236900	0.45448100	0.42698900
C	-2.81651500	0.07345100	-0.81570100
C	-1.52922800	-0.32689300	-1.17863100
C	-0.52119300	-0.33554400	-0.20897900
C	-0.86264400	0.06684700	1.08758700
C	0.87822600	-0.83087800	-0.56065000
C	1.99060100	0.03527400	0.04407200
C	1.96252900	1.42676100	-0.31460700
O	1.04212600	-2.20447100	-0.22256300
C	2.97455400	-0.43213100	0.82725100
N	1.91568400	2.54671400	-0.63093500
H	-2.46586300	0.75975900	2.35402500
H	-3.61295500	0.08986100	-1.55648700
H	-1.32218000	-0.62554600	-2.20259200

H	-0.12561200	0.09631800	1.88536500
H	0.98940300	-0.79765400	-1.64969300
H	0.71262600	-2.36033100	0.67932800
H	3.74982600	0.21959700	1.21654800
H	3.01952500	-1.48520100	1.08364900



13g

C	0.00752100	0.08203800	0.02479300
C	0.00996900	-0.01352700	1.44077400
C	1.20300500	-0.08656800	2.13069900
C	2.44542900	-0.06685800	1.43785900
C	2.44440600	0.03157400	0.00793500
C	1.19724700	0.10384800	-0.67436000
C	3.69786500	-0.14495800	2.10939300
C	4.88156300	-0.12045400	1.40585400
C	4.88941800	-0.01651000	-0.01541000
C	3.68671400	0.05161300	-0.68783700
C	6.23058200	-0.03556400	-0.75916800
C	7.20340700	1.01433600	-0.22699400
O	6.08213100	0.06884300	-2.17499800
C	8.31659600	0.68981600	0.44539700
C	6.86720100	2.43469400	-0.54016800
O	6.04264200	2.74969700	-1.39507900
O	7.55077700	3.33119700	0.17850800
C	7.31766800	4.72691500	-0.12414500
H	-0.93817800	0.13829600	-0.50800900
H	-0.93354900	-0.02959400	1.98005500
H	1.20803200	-0.16030200	3.21601400
H	1.19629900	0.17702900	-1.75968400
H	3.70977700	-0.22532000	3.19411400
H	5.82446900	-0.18129000	1.94326200
H	3.67368700	0.11261700	-1.77207000
H	6.69208100	-1.01519400	-0.59373800
H	5.83369700	0.99613100	-2.35806700
H	9.00006900	1.44467300	0.82004200
H	8.57274600	-0.34958100	0.63477500
H	7.97270400	5.27631500	0.55049900
H	7.57408700	4.93276900	-1.16570900
H	6.27174100	4.98075300	0.06113300

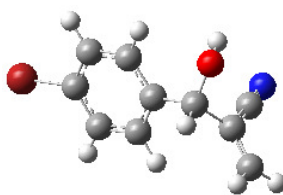
**14g**

C	-0.02075200	-0.06733000	-0.03193300
C	-0.04820700	-0.05527800	1.38666800
C	1.12947900	-0.02140800	2.10536400
C	2.38591000	0.00031900	1.43894100
C	2.41456600	-0.01336600	0.00651200
C	1.18332300	-0.04642800	-0.70604800
C	3.62337500	0.03971300	2.14128200
C	4.82217100	0.05600800	1.46496100
C	4.85880400	0.03639100	0.04020000
C	3.67154900	0.00912100	-0.66270600
C	6.21142500	0.09048700	-0.66906400
C	7.20198400	-0.94153400	-0.12357900
O	6.11366200	-0.00954300	-2.08830700
C	8.38412300	-0.61817200	0.42469800
C	6.82125800	-2.32266600	-0.25170500
N	6.48141400	-3.42857700	-0.38585700
H	-0.95479200	-0.09356200	-0.58697100
H	-1.00293400	-0.07208300	1.90568800
H	1.11136800	-0.01122800	3.19284100
H	1.20559400	-0.05615800	-1.79346500
H	3.61146100	0.05613100	3.22866400
H	5.75317500	0.08528300	2.02503200
H	3.67406600	0.02138700	-1.74886000
H	6.66245700	1.07317900	-0.49691900
H	5.67915700	-0.84619900	-2.32975300
H	9.07447400	-1.37157100	0.79094200
H	8.68132100	0.42227700	0.52078800

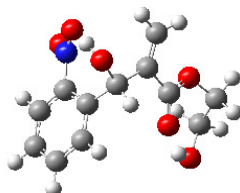
**15h**

C	2.29572600	-0.08088400	0.05926300
C	1.51118400	0.16037600	1.18660400
C	1.80046500	-0.80294300	-1.02406400
C	0.49075000	-1.28937100	-0.97453100
C	-0.32234700	-1.05614600	0.14014600
C	0.20427400	-0.32974800	1.21813700
C	-1.74154600	-1.63338700	0.22084300
C	-2.76246600	-0.60545100	0.69711800
O	-2.15406100	-2.24289000	-1.00293100
C	-3.39141600	-0.70536800	1.87627500
Br	4.08792400	0.58463800	0.00055300
C	-3.05890300	0.50259600	-0.25813400
O	-2.73709300	0.45945100	-1.44322700

O	-3.71324900	1.53434500	0.28377600
C	-4.08539700	2.61271700	-0.60680000
H	1.90667600	0.72073000	2.02752400
H	2.42017200	-0.98826200	-1.89529800
H	0.10516100	-1.86114800	-1.81185300
H	-0.40473000	-0.13951100	2.09770800
H	-1.73234500	-2.44795100	0.95285300
H	-2.32554200	-1.51357200	-1.63053800
H	-4.11744600	0.03244600	2.20156100
H	-3.19412300	-1.53775600	2.54698400
H	-4.61002200	3.33090000	0.02171300
H	-3.19130600	3.06234600	-1.04388000
H	-4.74057000	2.23784000	-1.39621700

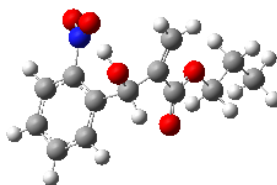
**16h**

C	1.73314100	0.06151600	-0.10025500
C	0.92835300	-0.92643200	-0.66701700
C	1.19629600	1.27566400	0.32127800
C	-0.17536200	1.50033000	0.17337700
C	-1.00809500	0.52193700	-0.38261400
C	-0.43981200	-0.68887700	-0.80511700
C	-2.50010100	0.78696300	-0.58758600
C	-3.37180700	-0.43110600	-0.27996200
O	-2.96426500	1.93218400	0.12680900
C	-4.18445600	-1.00634000	-1.18057000
Br	3.60857600	-0.25020600	0.09039800
C	-3.31625100	-0.93645200	1.06570200
N	-3.25289500	-1.31029100	2.16687800
H	1.35567700	-1.86753300	-0.99682100
H	1.83091200	2.04098600	0.75556800
H	-0.58886300	2.45299200	0.48756900
H	-1.06233200	-1.46156900	-1.24762100
H	-2.67434900	1.03791600	-1.63888800
H	-2.80530500	1.81422000	1.07982600
H	-4.80753800	-1.85961900	-0.93176700
H	-4.23539900	-0.62883000	-2.19783600

**17i**

C	-0.01279100	0.00146900	0.08505300
C	0.02155000	0.02255500	1.47549700
C	1.25589700	0.05240200	2.12841100
C	2.44037600	0.05238700	1.38931500
C	2.45169900	0.04435400	-0.01210000

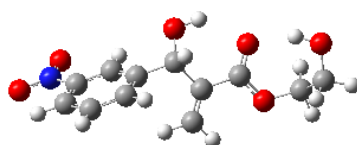
C	1.18587900	0.02974700	-0.63732200
C	3.82536400	0.01535700	-0.70148200
C	4.03269600	-1.18651900	-1.62803600
C	3.95783300	-2.51986400	-0.96570800
O	4.14493600	1.25710600	-1.31457600
C	4.38540300	-1.07926200	-2.91606400
O	3.83715300	-3.54208600	-1.81981300
O	4.00797100	-2.65555500	0.25509300
C	3.97421300	-4.91092000	-1.32599000
C	2.77850500	-5.41264500	-0.52322200
O	2.83001300	-5.10451200	0.86844000
N	1.03067000	0.07335400	-2.09439400
O	0.07009300	-0.50923600	-2.59786200
O	1.85062700	0.71928900	-2.76422400
H	-0.95484900	-0.01174500	-0.45031000
H	-0.90590800	0.02045800	2.03921600
H	1.30086700	0.06791300	3.21333900
H	3.39103800	0.04414500	1.91387100
H	4.55371800	-0.08921300	0.10689400
H	3.43404300	1.45467900	-1.95287800
H	4.57401300	-1.95776600	-3.52322900
H	4.51063600	-0.10959600	-3.38544200
H	4.07215500	-5.48826600	-2.24677300
H	4.90037400	-4.98543600	-0.74808200
H	1.85180700	-5.03371300	-0.97934000
H	2.76338200	-6.50517100	-0.59467100
H	3.15656100	-4.18576200	0.94504500



18i

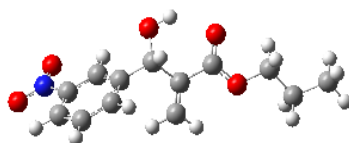
C	-1.79600300	1.53331300	0.29513700
C	-0.90131800	0.92408200	-0.57862400
C	-1.37705600	0.37769400	-1.77285600
C	-2.73786100	0.43764100	-2.07845200
C	-3.67183800	1.05299300	-1.23365100
C	-3.15236200	1.60655300	-0.04351800
C	-5.14351600	1.02439900	-1.67663100
C	-6.08560500	0.35247500	-0.67400900
C	-5.79073300	-1.08597400	-0.39325800
O	-5.59667300	2.30245500	-2.10579000
C	-7.17736300	0.94486600	-0.17325200
O	-6.42665700	-1.56316100	0.68703700
O	-5.04869900	-1.76344700	-1.09291100
C	-6.24896700	-2.97441900	0.99292200
C	-7.05208500	-3.28214900	2.24656500
C	-6.92303500	-4.75729000	2.64746500
N	-3.99251200	2.32962300	0.91598400
O	-4.93724900	3.01045600	0.48943400
O	-3.70762500	2.25931000	2.11173700
H	-1.45595500	1.97562300	1.22417300
H	0.15427500	0.88260400	-0.32953900
H	-0.69244100	-0.10237000	-2.46584100
H	-3.09470900	-0.01951100	-2.99636400

H	-5.16249400	0.40787000	-2.57908300
H	-5.46603800	2.91878900	-1.36108000
H	-7.84803400	0.41574300	0.49488500
H	-7.43390200	1.96563600	-0.43533500
H	-6.59358700	-3.55995900	0.13434600
H	-5.18055900	-3.16451500	1.13611300
H	-6.70179900	-2.63923100	3.06396300
H	-8.10572500	-3.03260900	2.06844200
H	-5.87989500	-5.02460500	2.85682800
H	-7.28791100	-5.42053100	1.85348100
H	-7.50836600	-4.96315200	3.55018300

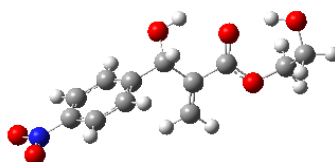


19j

C	-3.51351600	0.02538700	0.00243100
C	-3.89981200	1.36036800	0.14137800
C	-2.92890300	2.34495000	-0.03222600
C	-1.61127500	1.99001200	-0.33762500
C	-1.23688300	0.64492000	-0.46719100
C	-2.20269500	-0.34952700	-0.29736100
C	0.19973200	0.27987100	-0.78384400
C	1.07768700	0.08956700	0.45771300
C	2.50562300	-0.20979700	0.14563400
O	0.19317500	-0.91213000	-1.58170600
C	0.65197500	0.16249100	1.72698800
O	3.32116700	-0.17675800	1.19706200
O	2.88565700	-0.45952100	-1.00212500
C	4.70054900	-0.64690500	1.06063800
C	5.61499700	0.30302500	0.29455500
O	5.61946700	0.10715100	-1.11789400
N	-4.52214100	-1.02494400	0.17684400
O	-4.17335500	-2.20269300	0.04509600
O	-5.67865600	-0.68651600	0.44982000
H	-4.92696600	1.61418700	0.37225200
H	-3.20102700	3.39148300	0.06247800
H	-0.86697700	2.76962100	-0.47872000
H	-1.94432500	-1.39494300	-0.40375700
H	0.63558000	1.09926900	-1.37286800
H	1.12312000	-1.07479200	-1.82730300
H	1.33139300	0.00938900	2.55846500
H	-0.38377300	0.37786800	1.96770700
H	4.68685000	-1.64116300	0.60459600
H	5.02718300	-0.72491200	2.09855300
H	6.64075800	0.12122300	0.63131900
H	5.35676000	1.34111300	0.55105400
H	4.69188600	-0.03617800	-1.39024300

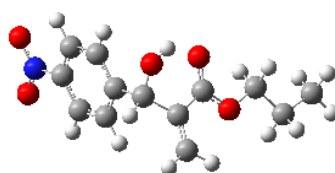
**20j**

C	3.60688300	-0.06781800	-0.21509200
C	4.02970200	1.22942300	-0.51346400
C	3.14015400	2.27741400	-0.28398100
C	1.86591500	2.02094700	0.23189700
C	1.45299500	0.71221900	0.52036000
C	2.33725300	-0.34488500	0.29492800
C	0.06305600	0.45264400	1.06628800
C	-0.99986600	0.24173300	-0.01743000
C	-2.37634900	0.02058200	0.52595900
O	0.12724300	-0.68593900	1.93607500
C	-0.76734400	0.23583600	-1.33728700
O	-3.34082600	0.03794700	-0.39543100
O	-2.59134000	-0.16180000	1.72333200
C	-4.70480600	-0.21185100	0.05346400
C	-5.60760200	-0.17049700	-1.16857900
C	-7.07286100	-0.42216000	-0.79026100
N	4.53044700	-1.18347700	-0.44453900
O	4.15173200	-2.32657600	-0.16863100
O	5.64917500	-0.93259200	-0.90536200
H	5.02325200	1.40696200	-0.90605600
H	3.44319900	3.29719700	-0.49968800
H	1.18651800	2.84994700	0.41340800
H	2.04837800	-1.36324200	0.51954900
H	-0.23828700	1.33124200	1.65504900
H	-0.77287900	-0.78633600	2.30111100
H	-1.57060400	0.07875700	-2.04914900
H	0.22821200	0.38818500	-1.74113400
H	-4.97078200	0.55761700	0.78477200
H	-4.73025000	-1.18768100	0.54824900
H	-5.27088200	-0.92633400	-1.88921300
H	-5.50812000	0.80774400	-1.65553200
H	-7.44041300	0.33684500	-0.08879100
H	-7.20098300	-1.40531700	-0.32090500
H	-7.71014100	-0.39109400	-1.68057400

**21k**

C	3.25588200	0.86393700	0.72484300
C	3.82011200	-0.10035100	-0.11332200
C	3.05824900	-0.79920600	-1.05291700
C	1.69817400	-0.51763800	-1.14790500
C	1.10224400	0.44218200	-0.31599300
C	1.89270700	1.12920300	0.61711600
C	-0.38294200	0.71884400	-0.43294300
C	-1.24976200	-0.19317700	0.44285800
C	-2.71517500	0.02559500	0.26845200

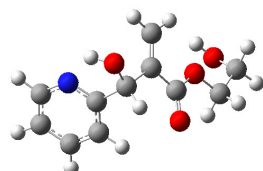
O	-0.61289800	2.09601100	-0.10713400
C	-0.78764900	-1.10463800	1.31053600
O	-3.48553600	-0.89186100	0.84818400
O	-3.16104700	0.98260000	-0.37138700
C	-4.93298100	-0.68428700	0.91877600
C	-5.65947700	-0.87143300	-0.40916000
O	-5.73733300	0.30547900	-1.21052900
N	5.25230100	-0.38321000	-0.00832600
O	5.73857200	-1.23669900	-0.75833300
O	5.91472900	0.24275200	0.82633800
H	3.87196800	1.39181000	1.44244300
H	3.52252900	-1.53728200	-1.69530000
H	1.09658500	-1.04923100	-1.88006900
H	1.44026500	1.87956200	1.25567400
H	-0.67814500	0.54762300	-1.47801400
H	-1.56180200	2.25342500	-0.26910500
H	-1.46394000	-1.71061400	1.90373200
H	0.27567000	-1.26924500	1.45021300
H	-5.24689000	-1.45073400	1.62869700
H	-5.12306700	0.30679300	1.34097800
H	-5.19359800	-1.69972900	-0.96318300
H	-6.69236800	-1.15754400	-0.18633500
H	-4.86149700	0.73719000	-1.17060700



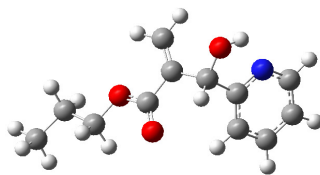
22k

C	-3.00365000	0.26507600	-1.05619400
C	-3.25341900	-0.54376500	0.05405900
C	-2.45207100	-0.48891700	1.19776500
C	-1.37622500	0.39312900	1.21686800
C	-1.09478600	1.21522700	0.11238900
C	-1.92322700	1.14397200	-1.01641200
C	0.06782800	2.21487800	0.17423400
C	1.35388000	1.58497200	0.70064000
C	2.01879900	0.61201100	-0.21733900
O	0.28281900	2.87987900	-1.06797700
C	1.88106000	1.91148400	1.88847200
O	2.99049200	-0.10372500	0.34979600
O	1.69569700	0.49904800	-1.39866000
C	3.72625600	-1.02973100	-0.50325200
C	4.78459200	-1.70641000	0.35219800
C	5.61171900	-2.70477600	-0.46773500
N	-4.38336100	-1.47043700	0.02091400
O	-4.58833300	-2.18267200	1.01064100
O	-5.08762000	-1.50390300	-0.99476000
H	-3.64181800	0.20670100	-1.92955800
H	-2.66807000	-1.12356500	2.04852200
H	-0.74953200	0.43726200	2.10238500
H	-1.72467400	1.78336900	-1.86892900
H	-0.22017800	3.00544500	0.87552500
H	0.69937400	2.22507700	-1.66269500
H	2.79838500	1.45820000	2.24926600
H	1.40362600	2.64672700	2.53117600

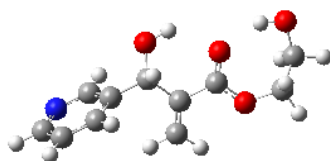
H	3.01450800	-1.74937900	-0.91937500
H	4.16753500	-0.45904000	-1.32629100
H	5.44093100	-0.94057600	0.78427400
H	4.29483000	-2.22177300	1.18800300
H	6.13463200	-2.20776500	-1.29410000
H	4.97889400	-3.49267800	-0.89432900
H	6.36612400	-3.18685200	0.16325000

**23I**

C	0.68268000	-0.61261700	0.11902500
C	1.49856300	-1.74316200	0.02039100
C	2.83589300	-1.62909200	0.40475400
C	3.30086600	-0.39565400	0.86730000
C	2.41183700	0.68194100	0.92174600
N	1.12363400	0.57717200	0.55424500
C	2.83944500	2.04863400	1.45754800
C	2.81156200	3.13071800	0.37752300
C	3.80658800	2.95970200	-0.72149900
O	2.05397200	2.40013600	2.59361600
C	2.01673500	4.20745900	0.43261200
O	3.55468700	3.71475000	-1.80507900
O	4.77093700	2.20925200	-0.64134100
C	4.51674200	3.65381200	-2.88996200
C	4.02693600	4.56816700	-3.99659900
O	2.84722700	4.09139900	-4.64400200
H	-0.36954300	-0.66175000	-0.16563300
H	1.09001100	-2.68424700	-0.34520900
H	3.50665900	-2.48652100	0.34981900
H	4.33560400	-0.26765100	1.18064300
H	3.86843600	1.96438600	1.82617600
H	1.09926300	2.25928100	2.38286500
H	2.03640500	4.96689300	-0.34286200
H	1.33975400	4.35884400	1.26861300
H	5.49386000	3.98469500	-2.51860800
H	4.59437500	2.61837300	-3.23682800
H	3.87530300	5.58510800	-3.60781400
H	4.80507500	4.61004500	-4.76697200
H	2.07057600	4.23766000	-4.05506300

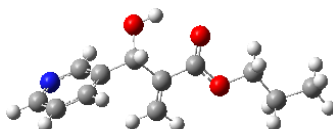
**24l**

C	0.94952800	-1.03004600	0.02841100
C	1.95968300	-1.94987000	-0.26659100
C	3.27518100	-1.61257100	0.05698000
C	3.52678900	-0.37581100	0.65594000
C	2.45166600	0.48228800	0.90560900
N	1.18401800	0.16179800	0.59732300
C	2.64333300	1.83845700	1.58860200
C	2.38529400	3.00214600	0.63155300
C	3.40678000	3.18186100	-0.44362700
O	1.84325100	1.91944900	2.76383400
C	1.36023600	3.85383800	0.76632300
O	3.02557700	4.01904100	-1.42003100
O	4.49821300	2.62370800	-0.42729900
C	3.99129700	4.29460400	-2.47348100
C	3.35621800	5.27927100	-3.44195400
C	4.31216300	5.62988200	-4.58988100
H	-0.09269500	-1.25741400	-0.20159400
H	1.71475700	-2.90311700	-0.73308000
H	4.09388000	-2.30231900	-0.14873600
H	4.53841100	-0.07831500	0.92599400
H	3.68128900	1.90347100	1.93520700
H	0.92529800	1.62801400	2.55411400
H	1.21389900	4.67790500	0.07492700
H	0.65211400	3.75440100	1.58425200
H	4.89819200	4.70270500	-2.01300600
H	4.24606300	3.34808000	-2.96354900
H	2.43153600	4.84378100	-3.84351300
H	3.07465700	6.18973600	-2.89644000
H	4.59113300	4.73776900	-5.16491300
H	5.23482700	6.09103200	-4.21522300
H	3.84156300	6.33838800	-5.28069700

**25m**

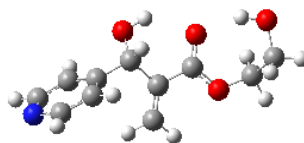
N	-4.11773800	0.41420100	-1.24841300
C	-4.85416900	-0.22164000	-0.32277800
C	-4.34267600	-0.63836900	0.90756700
C	-3.00055000	-0.38376100	1.19104800
C	-2.21268100	0.26979200	0.23863000
C	-2.82739000	0.64717700	-0.96295400
C	-0.75020500	0.55854100	0.50030600
C	0.20341500	-0.51019200	-0.04525000
C	1.64356200	-0.23678500	0.23497300
O	-0.44171800	1.84537300	-0.05945900

C	-0.17083600	-1.61286300	-0.70993800
O	2.47637800	-1.20352200	-0.14534900
O	2.01852600	0.80907300	0.77285300
C	3.89711300	-1.11193500	0.19222300
C	4.66650500	-0.05950000	-0.59991500
O	4.67067700	1.23868200	-0.00918000
H	-5.89676700	-0.40062000	-0.57693300
H	-4.98233800	-1.14208400	1.62592100
H	-2.57166100	-0.68755300	2.14327200
H	-2.25347700	1.16803400	-1.72468500
H	-0.60237000	0.60198200	1.58894500
H	0.47542700	2.04136900	0.20712900
H	0.55879400	-2.33420100	-1.06156500
H	-1.21478300	-1.82298700	-0.91737100
H	4.26243900	-2.10929600	-0.05613700
H	3.98981700	-0.94613200	1.26951400
H	4.28072200	-0.02662000	-1.62954100
H	5.71355500	-0.37580400	-0.64563800
H	3.76138400	1.41486900	0.30158000



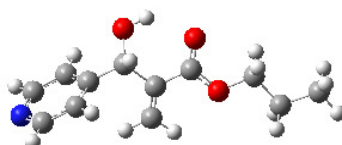
26m

N	-4.29939300	-0.24228500	-1.24104000
C	-4.91286400	-0.83364900	-0.20286500
C	-4.34415500	-0.92824100	1.06849800
C	-3.07587100	-0.38337900	1.27259300
C	-2.41483600	0.22987200	0.20415800
C	-3.07825400	0.27210900	-1.02983100
C	-1.03479600	0.82657800	0.37566800
C	0.10778500	-0.10883100	-0.03448800
C	1.46973600	0.48441200	0.14493600
O	-0.98095900	2.05082800	-0.37496700
C	-0.04654000	-1.35351800	-0.50717400
O	2.47374800	-0.37738100	-0.02915200
O	1.64258200	1.66860100	0.42996100
C	3.83080100	0.13938700	0.09398800
C	4.78917600	-1.01244000	-0.16049600
C	6.25080100	-0.55858100	-0.05538600
H	-5.90122800	-1.24434900	-0.39821600
H	-4.88508400	-1.40968200	1.87757100
H	-2.60672600	-0.43195500	2.25261200
H	-2.60457100	0.75000100	-1.88323600
H	-0.89624100	1.06003900	1.44164800
H	-0.09627600	2.42549400	-0.19965800
H	0.80929100	-1.96522000	-0.77171200
H	-1.02939900	-1.79349500	-0.63991300
H	3.94865300	0.55545500	1.09936600
H	3.95780500	0.94630200	-0.63453800
H	4.59586600	-1.42640000	-1.15814200
H	4.59126900	-1.81147300	0.56504500
H	6.47989900	0.22431400	-0.78878700
H	6.92770200	-1.39935000	-0.24153300
H	6.47592800	-0.16238300	0.94242400



27n

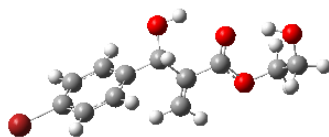
C	-4.37428300	0.21324300	-0.77391800
N	-4.89704400	-0.47247700	0.25519500
C	-4.06773300	-0.77428600	1.26768600
C	-2.71969000	-0.41432300	1.29547500
C	-2.18160700	0.29561100	0.21653800
C	-3.03851300	0.61529000	-0.84139800
C	-0.71986900	0.69187000	0.20987700
C	0.20819400	-0.39018000	-0.35283900
C	1.65576400	-0.03873200	-0.27470600
O	-0.58336000	1.90884400	-0.53631000
C	-0.19019900	-1.55014700	-0.89408700
O	2.48548100	-1.04787700	-0.52981900
O	2.03714400	1.09954400	0.01326900
C	3.91875200	-0.79006600	-0.67617900
C	4.63963300	-0.48462000	0.63306100
O	4.63942600	0.89345900	0.99906400
H	-5.05694800	0.45459400	-1.58616400
H	-4.50300700	-1.32475600	2.09907900
H	-2.10512500	-0.68350500	2.15066300
H	-2.67863400	1.17377500	-1.69863000
H	-0.41543000	0.87710000	1.25001200
H	0.35256300	2.16978700	-0.45166200
H	0.52650400	-2.26602700	-1.28196800
H	-1.24056900	-1.81322300	-0.96408500
H	4.05565900	0.01558500	-1.40351300
H	4.28552300	-1.72781000	-1.09591100
H	5.68996000	-0.76722200	0.50839500
H	4.21679500	-1.10892900	1.43411900
H	3.73979800	1.23426700	0.82660200



28n

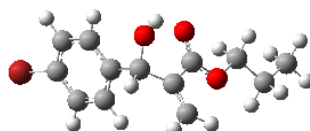
C	-4.42968200	-0.08831900	-1.05463800
N	-4.96732000	-0.91378300	-0.14223200
C	-4.23133100	-1.17195000	0.95112800
C	-2.96371500	-0.63233600	1.17399000
C	-2.40781600	0.22199200	0.21546500
C	-3.16874100	0.49776000	-0.92521300
C	-1.02899900	0.81391500	0.41947600
C	0.10785700	-0.07896100	-0.08924000
C	1.47026000	0.48540700	0.16258400
O	-0.98638800	2.09660500	-0.22107100
C	-0.05081000	-1.25132300	-0.71876800
O	2.47090900	-0.36364900	-0.07909500
O	1.64442600	1.63662700	0.55902000

C	3.82969300	0.12995500	0.10552000
C	4.78314600	-1.00326800	-0.23630300
C	6.24646500	-0.57422200	-0.06926800
H	-5.03723800	0.11620500	-1.93391200
H	-4.67867000	-1.83694800	1.68692000
H	-2.42456500	-0.87642500	2.08566400
H	-2.79388300	1.16224800	-1.69606700
H	-0.87772400	0.95281200	1.49994400
H	-0.10475000	2.46130100	-0.01423800
H	0.80358300	-1.82908900	-1.05468200
H	-1.03508200	-1.66408100	-0.91445200
H	3.93723700	0.45570400	1.14478400
H	3.97143300	0.99636200	-0.54813900
H	4.60210300	-1.32540400	-1.26948300
H	4.56742600	-1.86150700	0.41248900
H	6.91955300	-1.40146100	-0.31902000
H	6.45865900	-0.27027400	0.96315800
H	6.49361900	0.26846800	-0.72657400

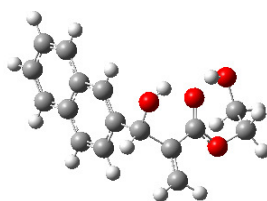


29o

C	-2.75509400	0.50028500	-1.05005100
C	-3.37702200	0.03378700	0.10797100
C	-2.66131300	-0.16214500	1.28793900
C	-1.29238600	0.11734500	1.30095100
C	-0.64140200	0.57835000	0.15068600
C	-1.38496600	0.76867800	-1.02131400
C	0.84790900	0.85026700	0.18098500
C	1.70187000	-0.33228600	-0.29137200
C	3.17158000	-0.08396500	-0.23727700
O	1.11104400	2.01219800	-0.62200400
C	1.22211800	-1.50748400	-0.72388200
O	3.92477500	-1.16170000	-0.44943000
O	3.64022200	1.03484500	-0.00592600
C	5.37003600	-1.01416300	-0.62412900
C	6.13133800	-0.70217900	0.66015600
O	6.23439500	0.68761800	0.96199800
Br	-5.25210300	-0.33910600	0.07571400
H	-3.32494300	0.65341300	-1.96068400
H	-3.15734500	-0.51813200	2.18490300
H	-0.73165400	-0.02750900	2.22135500
H	-0.89371600	1.13439800	-1.91737400
H	1.13510600	1.06668400	1.22010600
H	2.06569600	2.19025400	-0.52891300
H	1.88540900	-2.30364000	-1.04425700
H	0.15573800	-1.70238200	-0.76608700
H	5.66291500	-1.99398500	-1.00398100
H	5.55303500	-0.25470000	-1.39001100
H	5.67805000	-1.25772600	1.49453300
H	7.15696800	-1.06442300	0.53579600
H	5.35885100	1.08383000	0.78355400

**30o**

C	-2.57688800	0.58689100	-1.05983500
C	-2.87859700	-0.19527600	0.05254600
C	-2.06732300	-0.18220100	1.18665700
C	-0.93055600	0.62824400	1.19535800
C	-0.59958200	1.42302100	0.08811500
C	-1.43676800	1.39550300	-1.03287200
C	0.62365600	2.34721900	0.14134400
C	1.86712600	1.63772100	0.66705500
C	2.44764100	0.59639700	-0.23312000
O	0.88243200	2.98334600	-1.11073200
C	2.43313000	1.95013900	1.84085900
O	3.33433000	-0.20782700	0.35701800
O	2.13638100	0.50359800	-1.41911200
C	3.98727800	-1.21073400	-0.47524300
C	4.96130600	-1.97876900	0.40315900
C	5.69097600	-3.06927700	-0.39161000
Br	-4.43856500	-1.30140600	0.02513400
H	-3.21625300	0.57238400	-1.93643300
H	-2.31208200	-0.79194500	2.05028100
H	-0.29990400	0.63468300	2.08019800
H	-1.20485200	2.01326100	-1.89392700
H	0.39776400	3.16588600	0.83287500
H	1.24306700	2.29211300	-1.70033900
H	3.31808500	1.43751100	2.20345200
H	2.02022900	2.73462800	2.47001800
H	3.21422000	-1.86177900	-0.89501400
H	4.49441300	-0.69638100	-1.29757600
H	5.68769000	-1.27737300	0.83265900
H	4.41216500	-2.42824000	1.24007400
H	6.26292200	-2.64169600	-1.22424600
H	4.98641300	-3.80041300	-0.80678000
H	6.39235800	-3.60962100	0.25321800

**31p**

C	4.39781200	-8.40596700	-4.53624100
C	3.44901500	-8.88215600	-3.57347900
C	2.35322100	-8.04092100	-3.23089800
C	2.20931900	-6.79667000	-3.80342700
C	3.15223700	-6.31296600	-4.75630000
C	4.21687200	-7.11554000	-5.11062700
C	2.92907100	-4.93704400	-5.39497700
C	2.74749800	-3.83725900	-4.35076400
C	3.95911100	-3.47197500	-3.56460200
O	3.93984200	-4.59052000	-6.34117200
C	1.58177900	-3.20038900	-4.16843700
O	3.71056100	-2.76014700	-2.46675800

O	5.09412500	-3.80464200	-3.91717300
C	4.81971100	-2.15402700	-1.72891500
C	5.66090000	-3.14560400	-0.93141300
O	6.74002900	-3.72484100	-1.66158300
C	5.49435300	-9.24134200	-4.89083600
C	5.64653200	-10.48778800	-4.31814700
C	4.70724200	-10.95749400	-3.36364200
C	3.63248700	-10.17141300	-3.00064100
H	1.62351300	-8.39513100	-2.50610100
H	1.36297700	-6.17432400	-3.52403100
H	4.93515200	-6.77160600	-5.84903700
H	2.00196700	-4.98237900	-5.97645100
H	4.74491000	-4.38514700	-5.82795300
H	1.46517000	-2.42069900	-3.42307000
H	0.71068400	-3.44673600	-4.77010400
H	5.43576600	-1.58401500	-2.43060900
H	4.30685300	-1.46888100	-1.05249100
H	6.11490200	-2.60183400	-0.09665400
H	5.00449800	-3.92105200	-0.50955700
H	6.39400100	-3.95874800	-2.54518500
H	6.21412900	-8.88049800	-5.62229500
H	6.48869500	-11.11585600	-4.59687900
H	4.83791000	-11.94096100	-2.91974800
H	2.91027900	-10.52898600	-2.26995700



32p

C	-2.91336900	-0.11223300	-0.46848300
C	-3.08694400	-0.56077200	0.88166800
C	-2.17415500	-0.09749300	1.87066200
C	-1.14539900	0.75670200	1.54068500
C	-0.96316400	1.20369400	0.19989800
C	-1.84071400	0.77329800	-0.77362200
C	0.16599200	2.19349800	-0.11240200
C	1.52408400	1.70184300	0.38289700
C	2.11672800	0.55219200	-0.36444600
O	0.22003700	2.55637200	-1.49206500
C	2.17292300	2.28623600	1.39966700
O	3.13869800	-0.03841100	0.25933300
O	1.70227100	0.19550600	-1.46588800
C	3.81433500	-1.12499200	-0.43836000
C	4.94525200	-1.61248400	0.45239400
C	5.71608300	-2.76689600	-0.20072500
C	-3.82384600	-0.56467700	-1.46472800
C	-4.85653500	-1.42117100	-1.14106700
C	-5.02673700	-1.86539400	0.19596400
C	-4.16073100	-1.44370800	1.18437600
H	-2.29979400	-0.42806100	2.89937300
H	-0.46206000	1.09556700	2.31518100
H	-1.72688800	1.11885600	-1.79686300
H	-0.05146400	3.12695800	0.41830000
H	0.58429000	1.78233700	-1.96508900
H	3.13702900	1.92792800	1.74510700
H	1.75036300	3.14653600	1.91263900

H	3.08137200	-1.91312900	-0.63665300
H	4.18216100	-0.74183000	-1.39537800
H	5.62553800	-0.77664300	0.65886000
H	4.52951100	-1.93608400	1.41493600
H	6.52754100	-3.10487100	0.45284000
H	5.06179600	-3.62602200	-0.39341600
H	6.16104300	-2.46088200	-1.15553000
H	-3.69266800	-0.22346700	-2.48927100
H	-5.54517200	-1.75924000	-1.91106000
H	-5.84398500	-2.53965800	0.43847800
H	-4.28948700	-1.78188000	2.21026000

ANEXO B - Trabalhos decorrentes da Tese

Journal of Molecular Structure 1022 (2012) 72–80



Contents lists available at SciVerse ScienceDirect

Journal of Molecular Structure

journal homepage: www.elsevier.com/locate/molstruc

DFT/PCM, QTAIM, ¹H NMR conformational studies and QSAR modeling of thirty-two anti-*Leishmania amazonensis* Morita–Baylis–Hillman AdductsEdilson B.A. Filho^{a,b}, Ingrid A. Moraes^a, Karen C. Weber^b, Gerd B. Rocha^b, Mário L.A.A. Vasconcelos^{b,*}

^a Colegiado de Ciências Farmacêuticas, Universidade Federal do Vale do São Francisco, Petrolina, Pernambuco, Brazil
^b Departamento de Química, Universidade Federal da Paraíba, Campus I, João Pessoa, PB 58059-900, Brazil

HIGHLIGHTS

- ▶ Potential energy curves and DFT/PCM-B3LYP/6-31+(d) geometry optimization.
- ▶ Intramolecular Hydrogen Bonds supported by Quantum Theory Atoms in Molecules.
- ▶ Experimental spectroscopy data related to theoretical results.
- ▶ Quantitative Structure-Activity Relationship using Partial Least Squares (PLS).

ARTICLE INFO

Article history:

Received 9 February 2012
 Received in revised form 17 April 2012
 Accepted 18 April 2012
 Available online 3 May 2012

Keywords:

Morita–Baylis–Hillman Adducts
Leishmania amazonensis
 DFT/PCM
 QTAIM
 QSAR

ABSTRACT

Morita–Baylis–Hillman Adducts (MBHA) has been recently synthesized and bio-evaluated by our research group against *Leishmania amazonensis*, parasite that causes cutaneous and mucocutaneous leishmaniasis. We present here a theoretical conformational study of thirty-two leishmanicidal MBHA by B3LYP/6-31+g(d) calculations with Polarized Continuum Model (PCM) to simulate water influence. Intramolecular Hydrogen Bonds (IHBs) indicated to control the most conformational preferences of MBHA. Quantum Theory Atoms in Molecules (QTAIM) calculations were able to characterize these interactions at Bond Critical Point level. Compounds presenting an unusual seven member IHB between NO₂ group and hydroxyl moiety, supported by experimental spectroscopic data, showed a considerable improvement of biological activity (lower IC₅₀ values). These results are in accordance to redox NO₂ mechanism of action. Based on structural observations, some molecular descriptors were calculated and submitted to Quantitative Structure–Activity Relationship (QSAR) studies through the PLS Regression Method. These studies provided a model with good validation parameters values ($R^2 = 0.71$, $Q^2 = 0.61$ and $Q^2_{ext} = 0.92$).

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

The Morita–Baylis–Hillman Reaction (MBHR) is a relatively recent form to the C–C bond formation [1,2]. This reaction occurs between an sp² electrophilic carbon (e.g. aldehydes, ketones or imines) and the α position of an alkene (or alkyne) connected to an electron-attractor group (EAG), under tertiary phosphines or amines as nucleophilic catalysis, being DABCO a widely used catalyst [3] (Scheme 1). This reaction generates compounds called Morita–Baylis–Hillman Adducts (MBHA) and have been widely used as the starting material in the natural and unnatural products synthesis [4]. The MBHA themselves are today receiving great attention of the synthetic organic chemists which work into academic and industries laboratories on searching on the more

ecological, efficient and cheap synthetic protocols to new drug discovery [4].

In our continuing search for bioactive substances [5–9] and with our efforts towards reactivity of MBH reaction study [10–15] our research group first described the molluscicidal activities of some aromatic MBHA against *Biomphalaria glabrata* (Say) snails, intermediate schistosomiasis host [16]. In sequence, simple aromatic MBHA were presented by the first time as very active class of compounds against the *Leishmania amazonensis*, that causes cutaneous and mucocutaneous infections, MBHA 2, 4 and 15 were the most bioactives, Fig. 1 [17] and following that, the biological evaluation of aromatic MBHA against *Leishmania chagasi* (that cause visceral infections) parasites [18]. Leishmaniasis is one of the neglected diseases caused by different species of the protozoan parasite *Leishmania* and a major public health problem in many developing countries where 350 million people live at risk of infection. Fifteen species of the genus *Leishmania* can infect the man. Main species in the Old World are: *L. major*, *L. tropica*, *L. donovani*

* Corresponding author. Tel.: +55 83 9920 4745.
 E-mail address: mlav@quimica.ufpb.br (M.L.A.A. Vasconcelos).



XVI Simpósio Brasileiro de Química Teórica – SBQT 2011
Ouro Preto – MG, 20-23 Novembro de 2011
30 Anos SBQT

Conformational preferences and QTAIM study of bioactive Morita-Baylis-Hillman Adducts

Ingrid Araújo de Moraes¹(G), Edilson Beserra de Alencar Filho¹(PG), Claudio

Gabriel Lima Júnior²(PG), Mário Luiz Araújo de Almeida Vasconcellos²(PQ)

¹Universidade Federal do Vale do São Francisco, Campus Petrolina-Centro,

Petrolina-PE ²Universidade Federal da Paraíba, Campus I, João Pessoa-PB

Key-words: Conformational study, DFT, PCM, QTAIM.

In this work we presented a conformational study of bioactive Morita-Baylis-Hillman Adducts, as well as Quantum Theory Atoms in Molecules (QTAIM) calculations supporting their Structure-Activity Relationship. Relaxed Potential Energy Surface Scan (Gaussian 09W[®]) were performed using AM1 method, considering important rotational degrees of freedom (sigma bond) to each molecule. Conformational minimum were then submitted to B3LYP/6-31+g(d) calculations with Polarized Continuum Model to simulate water influence. Optimized structures were submitted to QTAIM calculations, using the Morphy98[®] and AIM2000 package. We observed that the conformations are governed by Intramolecular Hydrogen Bonds (IHB), when it is possible. QTAIM parameters such electronic density and laplaciane at bond Critical Point (BCP) level were able to characterize these interactions. Interestingly, compounds where NO₂ group makes IHB with hydroxyl (examples in Fig.1) show a considerable improvement of parasiticidal activity (lower IC₅₀ value), in accordance to redox mechanism of action of NO₂.

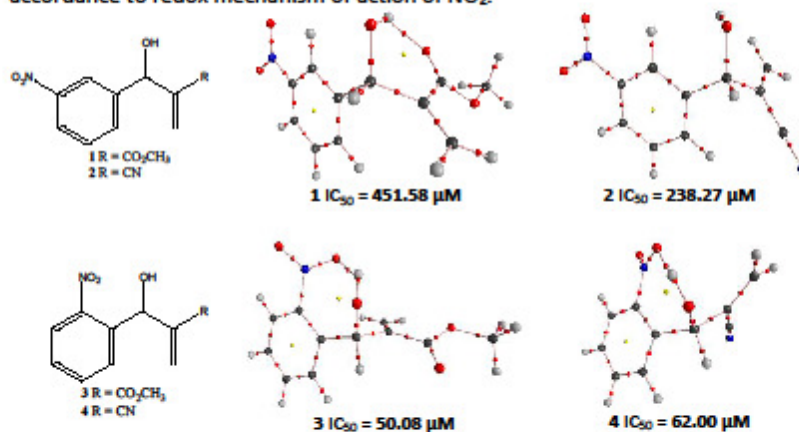


Fig. 1: Conformational preferences, bond critical points and activity of some adducts.

Support: UNIVASF, CNPq, LASOM/UFPB.