

UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE QUÍMICA

**RELAÇÕES ENTRE A ESTRUTURA QUÍMICA DE
NEOLIGNANAS E A SUA ATIVIDADE
ANTI-LEISHMANIOSE, ENVOLVENDO CÁLCULOS DE
SOLVATAÇÃO.**

Maria Cristina Andreazza Costa

TESE DE DOUTORADO

ORIENTADOR: Prof. Dr. Yuji Takahata

Campinas, 13 de novembro de 1998

RELAÇÕES ENTRE A ESTRUTURA QUÍMICA DE
NEOLIGNANAS E A SUA ATIVIDADE
ANTI-LEISHMANIOSE, ENVOLVENDO CÁLCULOS DE
SOLVATAÇÃO.

TESE DE DOUTORADO

MARIA CRISTINA ANDREAZZA COSTA

BANCA EXAMINADORA

Presidente: Prof. Dr. Yuji Takahata

Prof. Dr. Luiz Carlos Gomide Freitas (DQ-UFSCar)

Prof. Dr. Paolo Roberto Livotto (DFQ-UFRGS)

Prof. Dr. Lauro Euclides Soares Barata (IQ-UNICAMP)

Prof. Dr. Yoshiyuki Hase (IQ-UNICAMP)

Relações Entre A Estrutura Química De Neolignanas E A Sua Atividade
Anti-Leishmaniose, Envolvendo Cálculos De Solvatação.

Tese de doutorado

Maria Cristina Andreazza Costa

Orientador: Prof. Dr. Yuji Takahata

Palavras-chave: neolignana, leishmaniose, SAR, análise conformacional.

Instituto de Química, Universidade Estadual de Campinas, 1998.

RESUMO

Pretendemos dar continuidade à análise SAR (Structure-Activity Relationships) de neolignanas testadas em leishmaniose visceral, iniciada por nós em meados de 1991. Uma série de 20 neolignanas, sendo 11 inativas e 9 ativas em *Leishmania donovani*, foi submetida a cálculos de otimização de geometria e de parâmetros físico-químicos, a fim de realizarmos SAR, usando-se métodos de Reconhecimento de Padrões (RP).

Com o objetivo de obtermos maiores informações sobre a relação entre a estrutura das neolignanas e a sua atividade biológica (AB), uma série de novos parâmetros físico-químicos (FQ) foi incluída na análise. Observou-se significativa melhora nos resultados, com os quais foi possível estabelecer uma ordem relativa de AB para os referidos compostos.

A seguir, os métodos RP foram comparados à Regressão Linear Múltipla (RLM), com o intuito de investigarmos a aplicabilidade de RP em SAR. Os resultados obtidos nos confirmaram a capacidade de RP em separar os compostos com maior AB daqueles com menor AB.

A análise conformacional das neolignanas foi aprimorada, através de cálculos de solvatação das mesmas, em água. Obtivemos, com isto, indicações sobre a conformação ativa dos compostos. Comparações entre os mínimos locais encontrados para as neolignanas e a geometria de outros compostos ativos em leishmaniose, bem como estudos sobre a geometria da arginina (possível receptor de drogas ativas em leishmaniose), confirmaram a nossa suposição sobre a conformação ativa.

Structure-Activity Relationships Of Neolignans activities in leishmaniasis, Involving
Solvation Calculations.

Maria Cristina Andreazza Costa

Prof. Dr. Yuji Takahata

Keywords: Neolignans, leishmaniasis, SAR, conformational analysis.

Instituto de Química, Universidade Estadual de Campinas, 1998.

ABSTRACT

We continue to work on SAR analysis (Structure-Activity Relationships) of neolignans tested against visceral leishmaniasis; we initiated these studies in the middle of 1991. Conformational analysis of a series of 20 neolignans, 11 of which are inactivities and 9 are activities against leishmaniasis, was first performed. Then, physico-chemical (PC) parameters were calculated in order to perform a SAR, using Pattern Recognition methods (PR).

In order to obtain more informations about the relationships between the neolignan structures and their biological activity (BA) against leishmaniasis, a series of new PC parameters were included in the analysis. A significant improvement was obtained. It was possible to establish a relative order of BA to the referred compounds.

In the next step, the PR methods were compared with Multiple Linear Regression (MLR). The purpose of the comparison was to investigate the applicability of PR in SAR. The results confirmed that PR are capable of separating the compounds with higher BA of those with lower BA.

Conformational analysis of neolignans was improved by taking account of the effects of solvent water on the compounds. We obtained some indication about the active conformation of the compounds. Comparisons between the local minima found to neolignans and the geometry of the other compounds that are activities in leishmaniasis, as well as studies about the geometry of arginine (a possible receptor of drugs activities against leishmaniasis), have confirmed our supposition about active conformation.

CURRÍCULO

Maria Cristina Andreazza Costa

Nascimento: 11/05/1966 - Itu-SP-Brasil

1986-1990: Graduação: Bacharelado em Química - Unicamp

1987: Auxiliar didático na disciplina de química analítica experimental I

1990: Auxiliar didático na disciplina de química analítica experimental II

1992: Auxiliar didático na disciplina de química orgânica experimental para o curso de engenharia química.

1991-1993: Mestrado em físico-química - Unicamp

1994-1998: Doutorado em físico-química - Unicamp

1996: Estagiário de capacitação docente na disciplina de química analítica experimental II

1998-: Professora no nível de segundo grau, na Escola Técnica de comércio de Valinhos

PARTICIPAÇÃO EM CONGRESSOS

(com apresentação de trabalhos)

14^a. Reunião Anual da Sociedade Brasileira de Química-1991

15^a. Reunião Anual da Sociedade Brasileira de Química-1992

16^a. Reunião Anual da Sociedade Brasileira de Química-1993

17^a. Reunião Anual da Sociedade Brasileira de Química-1994

18^a. Reunião Anual da Sociedade Brasileira de Química-1995

20^a. Reunião Anual da Sociedade Brasileira de Química-1997

VII Simpósio Brasileiro de Química Teórica - 1993

VIII Simpósio Brasileiro de Química Teórica - 1995

IX Simpósio Brasileiro de Química Teórica - 1997

TRABALHOS PUBLICADOS:

COSTA, M.C.A.; TAKAHATA, Y.; Analysis of the photoelectron spectrum of the ethenol with semiempirical ham/3 method. *An. Acad. Bras. Ciênc* **65** (1993) 15-18.

COSTA, M.C.A.; BARATA, L.E.S.; TAKAHATA, Y.; SAR analysis of synthetic neolignans and related compounds which are antileishmaniasis active compounds using pattern recognition methods. *J. Mol. Str.(Theochem)* **340** (1995) 185-192.

COSTA, M.C.A., GAUDIO, A. C.; TAKAHATA,Y.; “A comparative study of principal component and linear multiple regression analysis in QSAT applied to 1,4-dihydropyridine calcium channel antagonists (nifedipine analogues)” *J. Mol. Struct. (Theochem)* **394** (1997) 291-300.

COSTA, M.C.A; TAKAHATA,Y.; Conformational analysis of synthetic neolignans which are active against leishmaniasis, using molecular mechanics method-mm2. *J.Comp.Chem.* **18(5)**-(1997), 712-721.

COSTA, M.C.A., BARATA, L.S.E., TAKAHATA,Y.; Conformation of neolignans that bind to the arginine residue in adenosine-kinase from *Leishmania donovani*. Trabalho aceito para publicação, pelo *J. Mol. Struct. (Theochem)*.

AGRADECIMENTOS

- Quero agradecer especialmente ao prof. Dr. Yuji Takahata, pela dedicação e amizade com que orientou este trabalho e também pelo interesse demonstrado quanto ao meu futuro e formação.
- Agradeço ao prof. Dr. Yoshiyuki Hase, por ter me orientado durante o período em que o Dr. Y. Takahata esteve no exterior.
- À prof^a. Dr^a. Marcia M. C. Ferreira e ao prof. Dr. Roy E. Bruns, agradeço a concessão dos programas de Reconhecimento de Padrões.
- Agradeço ao prof. Dr. Luiz Carlos Gomide Freitas, pela permissão de uso do programa Diadorim e pelos dados de entrada cedidos.
- Ao prof. Dr. Lauro E.S. Barata, agradeço pela possibilidade de participarmos do projeto sobre leishmaniose, que deu início a este trabalho.
- Aos colegas Aline T. Bruni, Fábio Garbujo, José Roberto dos Santos Politi e Vânia E. Barlete, sou grata pelas discussões sobre os parâmetros utilizados no programa Diadorim.
- Aos colegas de sala, Anderson C. Gaudio, Edilson G. Borges, Ignez Marone, Lucicleide R. Cirino e Rosana Vendrame, agradeço pelo agradável ambiente de trabalho e pela amizade.
- Aos colegas Anselmo Elcana, Jair V. Cirino, Júlio Trevas e Paulo Guagdanini, pelas proveitosas discussões e amizade.
- Ao meu marido, Cleber Vinícius Ursini, sou grata pelo demonstrado interesse com que ouvia cada uma das novas conclusões obtidas. Agradeço também por todas as vezes que me fez companhia, enquanto eu finalizava este trabalho.
- Finalmente, agradeço ao CNPq e à FAPESP pelo apoio financeiro, e ao CENAPAD (Centro Nacional de Computação de Alto Desempenho), pelo ambiente computacional utilizado.

ÍNDICE

RESUMO.....	iii
ABSTRACT.....	iv
CURRÍCULO.....	v
TRABALHOS PUBLICADOS.....	vi
AGRADECIMENTOS.....	vii
ÍNDICE.....	1
ÍNDICE DE FIGURAS.....	5
ÍNDICE DE TABELAS.....	8
SÍMBOLOS E ABREVIACÕES.....	11
 CAPÍTULO 1 - INTRODUÇÃO.....	 14
1.1 - A LEISHMANIOSE E AS NEOLIGNANAS.....	14
Subgênero <i>Viannia</i>	15
Subgênero <i>Leishmania</i>	16
Vetores.....	17
Medicamentos.....	18
1.2 - ENSAIOS BIOLÓGICOS.....	21
1.3 - COMPOSTOS ESTUDADOS.....	22
1.4 - OBJETIVOS.....	27
 CAPÍTULO 2 – METODOLOGIA.....	 28
2.1 – ANÁLISE CONFORMACIONAL E A SOLVATAÇÃO EM ÁGUA.....	28
2.1.1-O Método De Monte Carlo.....	30
O Potencial De Interação Intermolecular.....	34
2.1.2- O Método de Mecânica Molecular.....	35
2.2 - RELAÇÕES ENTRE A ESTRUTURA QUÍMICA E A ATIVIDADE BIOLÓGICA.....	 37

2.2.1 Os Métodos de Reconhecimento de Padrões (RP)	38
Análise de Componentes Principais (ACP).....	41
SIMCA.....	46
Agrupamento hierárquico.....	47
KNN.....	52
O Peso de Fisher.....	53
2.2.2 - Métodos de sobreposição molecular, com auxílio computacional ("Computer-Assisted Molecular Fitting - CAMF").....	53
2.3 - Os Métodos Semi-empíricos AM1 e MNDO-PM3.....	55
CAPÍTULO 3 - PARÂMETROS FÍSICO-QUÍMICOS.....	59
3.1 - COEFICIENTE DE PARTIÇÃO.....	59
3.2 – VOLUME MOLECULAR.....	62
3.3 – ÂNGULOS DIEDROS.....	63
3.4 – CARGAS ATÔMICAS.....	63
3.5 – ENERGIAS DOS ORBITAIS DE FRONTEIRA.....	67
3.5.1 - Índices De Fronteira.....	70
Densidade do elétron de fronteira.....	70
Densidade do orbital de fronteira.....	70
Densidade do radical de fronteira.....	70
3.6 – MOMENTOS DIPOLARES.....	71
CAPÍTULO 4 – CÁLCULOS.....	73
4.1 – NEOLIGNANAS ISOLADAS: ANÁLISE CONFORMACIONAL, PARÂMETROS FÍSICO-QUÍMICOS E RELAÇÕES ESTRUTURA-ATIVIDADE.....	73
4.1.1 - Análise Conformacional.....	73
4.1.2 - Parâmetros físico-químicos para as neolignanas isoladas.....	74
4.1.3 - Relações estrutura-atividade para as neolignanas isoladas.....	75
4.2 - ANÁLISE CONFORMACIONAL DE OUTROS COMPOSTOS ATIVOS EM LEISHMANIOSE (AMINO-CETONAS) E DE UM POSSÍVEL RECEPTOR (ARGININA).....	75
4.3 – NEOLIGNANAS SOLVATADAS: ANÁLISE CONFORMACIONAL, CÁLCULO DE PARÂMETROS FÍSICO-QUÍMICOS E RELAÇÕES ESTRUTURA-ATIVIDADE..	76

4.3.1 - Solvatação.....	76
4.3.2 - Relações estrutura-atividade.....	78
CAPÍTULO 5 – INCLUSÃO DE NOVOS PARÂMETROS FÍSICO-QUÍMICOS PARA A CONFORMAÇÃO MAIS PROVÁVEL (E), ENCONTRADA PARA OS COMPOSTOS ISOLADOS.....	79
5.1 – INTRODUÇÃO.....	79
5.2 – CÁLCULOS.....	80
5.3 – RESULTADOS E DISCUSSÃO.....	81
5.4 – CONCLUSÕES.....	106
CAPÍTULO 6 – RLM x ACP.....	107
6.1 – INTRODUÇÃO.....	108
6.2 – METODOLOGIA.....	114
6.3 – RESULTADOS E DISCUSSÃO.....	115
6.4 – CONCLUSÕES.....	125
CAPÍTULO 7 – ANÁLISE CONFORMACIONAL PARA AS NEOLIGNANAS SOLVATADAS EM ÁGUA.....	126
7.1 – INTRODUÇÃO.....	126
7.2 – RESULTADOS E DISCUSSÃO.....	132
7.3 - CONCLUSÕES.....	137
CAPÍTULO 8 – PROCURA DA CONFORMAÇÃO ATIVA DAS NEOLIGNANAS, ATRAVÉS DA COMPARAÇÃO COM OUTROS COMPOSTOS ATIVOS EM LEISHMANIOSE E COM UM POSSÍVEL SÍTIO RECEPTOR.....	138
8.1 - INTRODUÇÃO.....	138
8.2 - RESULTADOS E DISCUSSÃO.....	140
8.3 - CONCLUSÕES.....	151
CAPÍTULO 9 - SUGESTÃO DE NOVOS COMPOSTOS PARA SÍNTESE.....	152
9.1 - PARTE 1: Conjunto teste, proposto com base na análise conformacional para os compostos isolados.....	152
9.1.1 - INTRODUÇÃO.....	152
9.1.2 - CÁLCULOS.....	154
9.1.3 - RESULTADOS E DISCUSSÃO.....	154

9.2 - PARTE 2: Conjunto Teste Proposto Após A Solvatação Dos Compostos.....	159
9.2.1 – INTRODUÇÃO.....	159
Compostos 5 e 10.....	167
Composto 6.....	167
Neolignanas ativas que apresentam o grupo metil no carbono 13 e =O no carbono 12.....	167
Composto 7.....	168
9.3 - SUGESTÃO DE NOVOS COMPOSTOS.....	170
CONCLUSÕES GERAIS.....	172
REFERÊNCIAS.....	174

ÍNDICE DE FIGURAS

Figura 1 - Mosquito-Palha, vetor da leishmaniose.....	17
Figura 2 - Medicamentos usados atualmente contra a leishmaniose; a) antimoniato de meglumine e b) estibogluconato de sódio	19
Figura 3 - Os compostos a (Surinamensina) e b (Virolina) foram isolados do extrato bruto de folhas de <i>Virola Surinamensis</i> e mostraram atividade contra <i>Schistosoma mansoni</i>	21
Figura 4 - a) Estrutura básica das neolignananas, sendo R1 a R8 os substituintes mostrados na tabela 2 e X, oxigênio ou enxofre; t1 e t2 são os eixos torsionais referentes aos ângulos diedros 14-13-12-6 e 18-17-13-12. b) Estrutura básica das aminocetonas estudadas, sendo R11, R22 e R24 os substituintes mostrados na tabela 1.....	24
Figura 5 - Representação dos N objetos num espaço M dimensional, onde M é o número de medidas feitas sobre esses N objetos.....	39
Figura 6 - Representação dos dados y num espaço tridimensional.....	42
Figura 7 - Representação da primeira componente principal num espaço tridimensional.....	43
Figura 8 - Representação das duas primeiras componentes principais num espaço tridimensional.....	44
Figura 9 - Representação de modelos com números diferentes de componentes principais, para cada categoria.....	47
Figura 10 - Representação da distância entre pontos dos grupos A e B, calculada pela conexão simples.....	48
Figura 11 - Representação do dendograma obtido pelo método do agrupamento hierárquico, usando-se a conexão simples.....	51
Figura 12 - Distância entre pontos dos grupos A e B, calculada pela conexão completa.....	51
Figura 13 - Distância entre dois pontos dos grupos A e B, calculada pela conexão por meio de centróides.....	52
Figura 14 - Representação do método KNN para classificação de objetos de um conjunto teste. Os objetos “x”, na figura, são compostos ativos; “.” são os compostos ativos e “*” o composto teste.....	52
Figura 15 - Interação entre o HOMO de uma molécula com o HOMO de outra.....	68

Figura 16 - Interação entre um orbital ocupado de uma molécula e um desocupado de outra.....	68
Figura 17 - Interação entre o HOMO de uma molécula com o LUMO de outra	69
Figura 18 - Resultados da análise RP obtidos para as 21 variáveis da tabela 4. O agrupamento hierárquico é mostrado na figura 13a, a variância das três primeiras componentes principais na figura 13b. Os gráficos dos “scores” e “loadings” das duas primeiras componentes principais são mostrados nas figuras 13c e 13d.....	82
Figura 19a - Agrupamento hierárquico para as neolignanas, utilizando as seguintes variáveis: logP, volume e cargas nas posições 9 (C ₉) , 10 (C ₁₀) e 26 (C ₂₆)	86
Figura 19b - Análise de componentes principais para as neolignanas, utilizando as variáveis: logP, volume e cargas sobre os átomos 9, 10 e 26.....	87
Figura 19c - Gráfico dos “loadings” das duas primeiras componentes principais, obtido para as neolignanas. As variáveis 40, 49, 50, 51 e 52 são: volume, logP e cargas sobre os átomos 9, 10 e 26 (figura 4a), respectivamente.....	88
Figura 20 - Resultados obtidos para as 49 variáveis da tabela 6.....	98
Figura 21a - Dendograma para as neolignanas na conformação E, usando as seguintes variáveis: volume, logP e cargas sobre os átomos 2, 3, 20, 21 e 22.....	102
Figura 21b - ACP para as neolignanas na conformação E, usando as mesmas variáveis descritas na figura 16a.....	103
Figura 21c - Gráfico dos “loadings” das duas primeiras componentes principais para as neolignanas na conformação E; as variáveis 40, 49, 50, 51, 52, 53 e 54 são: volume, logP, cargas sobre os átomos 2, 3, 20, 21 e 22.....	104
Figura 22 - Estrutura básica dos derivados DHP.....	108
Figura 23 - Classificação do conjunto (I) de compostos para-derivados, nos grupos de alta e baixa atividade. Todos os compostos foram classificados corretamente.....	116
Figura 24 - Classificação do conjunto (II) de compostos (orto-derivados), nos grupos de alta e baixa atividade. Os compostos 10 e 17 foram mal classificados.....	118

Figura 25 - Classificação do conjunto (III) de compostos orto e para nos grupos de alta e baixa atividade. Os compostos 13 e 29 foram mal classificados.....	120
Figura 26 - Classificação do conjunto (IV) de compostos (poli e mono derivados) nos grupos de alta e baixa atividade. Os compostos 24, 26, 27, 28, 30, 31, 40 e 42 foram classificados incorretamente.....	122
Figura 27 - Classificação do conjunto (V) nos dois grupos (alta e baixa atividade); os compostos 1, 4 e 11 foram mal classificados.....	124
Figura 28 - Mapa conformacional para a molécula 15.....	128
Figura 29 - Oito mínimos locais encontrados na análise conformacional realizada para as neolignanas isoladas.....	129
Figura 30 - Representação esquemática da adenosina-monofosfato (AMP), adenosina-difosfato (ADP) e adenosina-trifosfato (ATP).....	139
Figura 31 - a) Geometria da arginina obtida de dados de cristalografia; b) Geometria da arginina após rotação de 120° do eixo “c”, em relação à figura 26a. c) Esquema representativo da adenosina. A geometria da adenina foi obtida da literatura [71].....	141
Figura 32 - Interação entre a neolignana 12D e a arginina na conformação B.....	147
Figura 33 - Interação entre uma aminocetona e a arginina na conformação B.....	148
Figura 34 - Interação da neolignana 12D com a arginina, produzida pelo programa PowerFit. a) A arginina está na conformação obtida de dados de cristalografia; b) A arginina está na conformação obtida após a rotação de 120° sobre o eixo “c” (figura 4b).....	149
Figura 35 - Gráfico dos “scores” das duas primeiras componentes principais para os conjuntos treinamento e teste das neolignanas. Os compostos teste 21, 24, 25, 26, 27, 31 e 32 foram classificados como ativos.....	155
Figura 36 - Gráfico dos “loadings” das duas primeiras componentes principais para os conjuntos treinamento e teste das neolignanas. As variáveis 26, 27, 37, 40 e 53 são respectivamente, as cargas sobre os átomos 2, 3, 21, volume e logP.....	156
Figura 37 - Agrupamento hierárquico para os conjuntos treinamento e teste das neolignanas. Os compostos 21, 24, 25, 26, 27, 31 e 32 foram classificados como ativos.....	157
Figura 38 - Neolignanas na conformação D	160

ÍNDICE DE TABELAS

Tabela 1 - Aminocetonas ativas em leishmaniose; ED_{50} é uma medida de atividade biológica que indica a concentração molar, correspondente a 50% da dose efetiva, sendo que quanto maior o valor da dose efetiva, menor a atividade biológica.....	25
Tabela 2 - Os compostos numerados entre 1 e 20 são neolignanas e foram ensaiados em leishmaniose, sendo onze inativos (-) e nove ativos (+). As posições dos substituintes , R e X, são mostradas na figura 4a.....	26
Tabela 3 - Comparação entre as cargas de Mülliken e derivadas do potencial eletrostático (PE), calculadas para as neolignanas ativas (compostos 12 a 20 - tabela 2) Os átomos C18, C19,...etc, são mostrados na figura 4a.....	65
Tabela 4a - Variáveis utilizadas na análise RP das neolignanas testadas em leishmaniose visceral. η , $1/\eta$ e χ são a dureza molecular, moleza molecular e a eletronegatividade de Mülliken; t_1 e t_2 são os ângulos diedros mostrados ns figura 4a.....	83
Tabela 4b - Parâmetros FQ calculados para as 20 neolignanas do conjunto treinamento, na conformação E. ΔH_f é o calor de formação e μ , μ_h e μ_p são os momentos dipolares total, híbrido e pontual, respectivamente; C_8 , C_9 , C_{10} , C_{16} , C_{17} , C_{24} , C_{26} , e C_{28} são as cargas sobre os átomos 8, 9, 10, 16,...28 (figura 4a).	84
Tabela 5 - Valores de atividade anti-leishmaniose para as neolignanas testadas “in vitro”; %I é a porcentagem de inibição.....	89
Tabela 6 - Inclusão de novos parâmetros FQ na análise RP das neolignanas testadas em leishmaniose, para a conformação E. As variáveis 1, 2 e 40 a 49 já haviam sido incluídas na análise da figura 14 e estão dispostas também na tabela 4. As cargas C_8 , C_9 , C_{10} , C_{26} e C_{28} da tabela 4 foram substituídas pelas cargas correspondentes, dos anéis aromáticos: C_2 , C_3 , C_4 , C_{19} , C_{21} e C_{23} . def_8 , def_9 ,...,etc, são as densidades eletrônicas de fronteira sobre os átomos 8, 9, ...etc. dof_8 , dof_9 , etc, são as dendidades de orbital de fronteira sobre os respectivos átomos e drf_8 , drf_9 , etc, são as densidades radicais de fronteira sobre os átomos 8, 9, etc.....	91

Tabela 7 - Resultados da análise SIMCA para o modelo de 4 componentes principais para a categoria 1 e 3 componentes principais para a categoria 2.....	99
Tabela 8 - Análise SIMCA para o modelo de 3 componentes principais para as duas categorias.....	100
Tabela 9 - Análise SIMCA para o modelo de 3 componentes principais para a categoria 1 (inativas) e 2 componentes principais para a categoria 2 (ativas).....	101
Tabela 10 - Dados biológicos expressos como $\log(IC_{50})$, polarizabilidade molecular (α), largura mínima de Verloop (B_1) e comprimento (L) dos substituintes, e barreira rotacional (E_{rot}) para o conjunto de derivados DHP.....	110
Tabela 11 -Carga atômica (C), densidades eletrônica (def) e orbital (dof) de fronteira, e dureza molecular (η) para os 45 derivados DHP calculados com AM1.....	112
Tabela 12 - Matriz de correlação para o conjunto de variáveis usadas na ACP e na RLM para o conjunto (I) de compostos.....	117
Tabela 13 - Matriz de correlação para o conjunto de variáveis usadas na ACP e na RLM para o conjunto (II) de compostos.....	119
Tabela 14 - Matriz de correlação para conjunto de variáveis usadas na ACP e na RLM para o conjunto (III) de compostos.....	121
Tabela 15 - Matriz de correlação para conjunto de variáveis usadas na ACP e na RLM para o conjunto (IV) de compostos.....	123
Tabela 16 - Ordem decrescente de estabilidade para as neolignananas inativas.....	130
Tabela 17 - Ordem decrescente de estabilidade e probabilidade para as neolignananas ativas. Não calculamos os valores de probabilidade para o composto 16, pois devido ao grande número de átomos presentes, não foi possível fazermos uma análise conformacional muito detalhada.....	131
Tabela 18 - A energia estérica (E_{est}) foi calculada com o método MM2 para os compostos isolados e a energia de interação soluto-solvente (E_{st-sv}) foi calculada com o método de Monte Carlo, para os compostos solvatados em água. Os valores mais baixos de energia (E_{est} e E_{st-sv}) foram sublinhados para comparação.....	132

Tabela 19 -A conformação A da arginina foi obtida de dados de cristalografia e a conformação B foi obtida após rotação de 120° do eixo “c”, mostrado nas figuras 4a e 4b. R _E é a barreira rotacional calculada a partir da conformação A.....	142
Tabela 20 - d ₁ é a distância entre os átomos C3 e C6 para a arginina (figura 26), C6 e C18 para as neolignanas (figura 4a) e C6 e C16 para as aminocetonas (figura 4b). Os ângulos torsionais a e b são mostrados nas figuras 4 e 26.	144
Tabela 21 - Conjunto teste, planejado com base na semelhança estrutural das neolignanas ativas em leishmaniose. R1, R2, etc, são substituintes nas posições 1, 2, etc, mostradas na figura 4a do capítulo1; a posição X também é mostrada na figura 4a.....	153

SÍMBOLOS E ABREVIações

AB - Atividade biológica

ACP - Análise de componentes principais

A_e - Afinidade eletrônica

ADP - Adenosina-difosfato

AMP - Adenosina-monofosfato

AM1 - “Austin Model 1”

ATP - Adenosina-trifosfato

B₁ - Largura mínima de Verloop

CAMF - “Computer Assisted Molecular Fitting”

C_n - Carga atômica

Correl - Subprograma do programa Arthur/Unicamp: realiza a correlação entre as variáveis.

CP - Componente Principal

def - densidade eletrônica de fronteira

DHP - 1,4-diidropiridina

dof - densidade do orbital de fronteira

drf - densidade do radical de fronteira

ED₅₀ - Concentração molar correspondente a 50% da dose efetiva

E_{est.} - Energia estérica

E_{rot} - Energia rotacional

E_s - Constante desenvolvida por Taft para parâmetros estéricos

E_{st-sv} - Energia de interação soluto-solvente

FQ - Parâmetros físico-químicos

f - Constante de hidrofobicidade do fragmento

Hier - Subprograma do programa Arthur/Unicamp, o qual realiza o agrupamento hierárquico

HOMO - Orbital molecular mais alto ocupado

I - Potencial de ionização

IC₅₀ - Concentração molar da droga, capaz de inibir determinada função do organismo.

Karlov - Subprograma do programa Arthur/Unicamp, o qual realiza a análise de componentes principais

KNN - Subprograma do programa Arthur/Unicamp, o qual realiza a análise baseada nos k vizinhos mais próximos.

L - Largura (do substituinte) de Verloop

L. donovani - leishmania donovani

LUMO - Orbital molecular mais baixo desocupado

MM2 - Método de mecânica molecular-versão 2

MNDO - “Modified Neglec of Diatomic Overlap”

MNDO-PM3 - “Modified Neglect of Diatomic Overlap - Parametric Method 3”

P - Coeficiente de partição entre um solvente orgânico e outro aquoso

QSAR - Relações quantitativas entre atividade biológica e estrutura química

RLM - Regressão Linear Múltipla

RP - Método de Reconhecimento de Padrões

SAR - Relações entre estrutura química e atividade biológica

SIMCA – “Self Independent Modeling of Class Analogy”

t1 e t2 - eixos torsionais correspondentes aos ângulos diedros 14-13-12-6 e 18-17-13-12, respectivamente, para as neolignanas (figura 4a).

V_{vdw} - Volume de van der Waals

Weight - Subprograma do programa Arthur/Unicamp; realiza o cálculo do peso de Fisher para as variáveis.

α - Polarizabilidade

χ - Eletronegatividade de Mülliken

ΔH_f - Calor de formação

η - Dureza molecular

$1/\eta$ - Moleza molecular

μ - Momento dipolar total

μ_p - Momento dipolar pontual

μ_c - momento dipolar de carga = momento dipolar pontual

μ_H - Momento dipolar híbrido

π - Constante de hidrofobicidade relativa à substituição do hidrogênio

σ - Constante de Hammett, para descrever os efeitos eletrônicos de substituinte no anel benzênico

σ^* - Constante de Taft, para descrever os efeitos eletrônicos de substituintes em estruturas alifáticas.

CAPÍTULO 1

INTRODUÇÃO

1.1 - A LEISHMANIOSE E AS NEOLIGNANAS

A leishmaniose é uma doença tropical causada por um protozoário, o *Leishmania*. Atinge cerca de 400 mil pessoas por ano, não só nos países mais pobres, mas também nos desenvolvidos. A leishmaniose apresenta características clínicas e epidemiológicas diversas, em cada área geográfica, sendo reunidas em quatro grupos [2]:

- 1 - Formas que produzem apenas lesões cutâneas - leishmaniose cutânea.
- 2 - Formas que se complicam com o aparecimento de lesões destrutivas nas mucosas do nariz, boca e faringe - designadas leishmaniose mucocutânea.
- 3 - Formas viscerais, atingindo o baço, fígado, medula óssea e tecidos linfóides - leishmaniose visceral ou calazar.
- 4 - Formas disseminadas cutâneas, que se apresentam em pacientes que haviam sido tratados de leishmaniose visceral- é a leishmaniose cutânea difusa.

A leishmaniose cutânea é doença benigna, a mucocutânea pode causar lesões mutilantes na face, ao passo que a forma visceral é acompanhada de elevada mortalidade, quando não tratada. Os parasitos isolados de todos os casos não são idênticos entre si, existindo vários tipos de *Leishmania*, para cada forma de leishmaniose.

A classificação e a nomenclatura das leishmânias oferecem grandes dificuldades, pois apesar de todas as espécies serem morfologicamente parecidas, causam doenças com características clínicas muito diferentes e peculiares, que não se pode considerar todos como um mesmo agente patogênico. A identificação dos diferentes tipos de *Leishmania* é baseada no quadro clínico que produzem e também nas características geográficas e epidemiológicas. Esses critérios, entretanto, não são muito aceitáveis e até agora não se chegou a uma classificação com aceitação universal.

Atualmente, aceita-se como razoável classificar as leishmânias que atacam o homem, em complexos fenotípicos, agrupados em dois subgêneros: *Viannia* e *Leishmania*.

O subgênero *Viannia* compreende o complexo "*Leishmania brasiliensis*", enquanto o subgênero *Leishmania* compreende os complexos "*Leishmania mexicana*" e *Leishmania donovani*".

Subgênero *Viannia*

Complexo "*Leishmania brasiliensis*"

Os parasitos têm tamanho relativamente pequeno, em torno de 2,3 µm; não provocam a leishmaniose visceral, causando apenas a forma cutânea. As lesões no homem podem ser simples ou múltiplas, os parasitos não são abundantes nas lesões e crescem pobremente em meios de cultura.

Os membros desse complexo são: 1) *Leishmania* (*Viannia*) *brasiliensis*, que apresenta ampla distribuição no Brasil, Venezuela, Guiana Francesa, América Central e nas áreas florestais dos Andes. Forma úlceras cutâneas persistentes, acompanhadas de lesões nasofaringianas destrutivas e desfigurantes. A doença é conhecida por leishmaniose mucocutânea, úlcera de Bauru e ferida brava. 2) *Leishmania* (*Viannia*) *guyanensis*, aparece na Venezuela, Guiana, Suriname e Brasil. Causa ulcerações simples ou múltiplas, mas não ataca as mucosas. 3) *Leishmania* (*Viannia*) *panamensis*, é encontrada no Panamá, produz lesões únicas ou pouco numerosas e se propaga às mucosas. 4) *Leishmania* (*Viannia*) *peruviana*, que distingue-se das demais leishmanioses americanas por encontrar-se longe das áreas florestais, em vales muito elevados e secos dos Andes, na Bolívia, Peru, Equador, Colômbia e Venezuela.

Subgênero *Leishmania*

Complexo "*Leishmania mexicana*"

Os parasitos medem cerca de 3,2 μm e crescem rapidamente nos meios de cultura e no hamster, produzindo neste animal lesões ricas em parasitos. No homem, as lesões localizam-se na pele e são benignas. Os animais reservatórios são roedores e marsupiais que apresentam lesões muito discretas, principalmente na cauda.

Os membros desse complexo são: 1) *Leishmania* (*Leishmania*) *mexicana*, que encontra-se no México, Guatemala e Belise, causando úlceras benignas na pele e com tendência para a cura espontânea; 2) *Leishmania* (*Leishmania*) *pifanoi*, a qual tem sido isolada de alguns casos de leishmaniose cutânea difusa, na Venezuela; 3) *Leishmania* (*Leishmania*) *amazonensis*, peculiar à Bacia Amazônica, sendo encontrada nas regiões Nordeste, Centro-Oeste e Sudeste do Brasil, sendo que raramente a infecção atinge o homem e, quando o faz, produz lesões cutâneas únicas; 4) *Leishmania* (*Leishmania*) *venezuelensis*, que causa lesões nodulares e, 5) *Leishmania* (*Leishmania*) *garnhami*, que causam lesões únicas ou múltiplas, curando-se naturalmente em seis meses.

Complexo "*Leishmania donovani*"

Os parasitos são de pequeno tamanho (2,1 μm de diâmetro) e mostram forte tendência de invadir as vísceras, atingindo o baço, fígado, medula óssea e órgãos linfóides. Os membros deste complexo são as três espécies seguintes: 1) *Leishmania* (*Leishmania*) *donovani*, que encontra-se em países asiáticos e na África Oriental. Infecta indivíduos adultos e é transmitida por flebotomíneos. A doença chama-se leishmaniose visceral ou calazar indiano. 2) *Leishmania* (*Leishmania*) *infantum*, que infecta principalmente crianças em áreas do Mediterrâneo, África Oriental, Próximo Oriente, Ásia Soviética e Norte da China. A doença é o calazar infantil. 3) *Leishmania* (*Leishmania*) *chagasi*, que ataca tanto crianças como adultos e ocorre nas Américas.

Vetores

Os insetos vetores das leishmanioses são pequenos, cor de palha ou castanho-claros, facilmente reconhecíveis pela atitude que adotam quando pousados, pois as asas permanecem entreabertas e ligeiramente levantadas, em vez de se cruzarem sobre o dorso (figura 1). Por isso, são conhecidos no Nordeste brasileiro por "cangalha", orelha-de-veado e "asa-dura"; no sul, por "mosquito-palha" e "birigüi"; na Amazônia, é "tatuíra". Pertencem à família *Psychodidae* e subfamília *Phlebotominae*.



Figura 1 - "Mosquito-palha", vetor da leishmaniose.

A infecção dos insetos dá-se quando eles se alimentam, sugando sangue de um paciente ou animal infectado.

A forma de leishmaniose abordada neste trabalho é a visceral, causada por *Leishmania donovani*. No Brasil, a leishmaniose visceral tem sido detectada em todos os estados costeiros, do Pará ao Paraná e no interior de Minas Gerais, Goiás e Mato Grosso do Sul, mas o maior número está no Nordeste, sobretudo no Ceará [3]. Em meados de 1994,

um surto de leishmaniose atingiu Campinas-SP, tendo sido registrados 33 casos. Também foram confirmados 22 casos em Indaiatuba-SP e um número menor em outras cidades dos arredores.

A leishmaniose visceral ou calazar, durante muito tempo foi confundida com outras endemias, devido à sua semelhança com outras moléstias tropicais de caráter febril e, também, devido ao fato de não chamar a atenção por lesões facilmente visíveis. O calazar caracteriza-se por produzir febre irregular e anemia na sua fase terminal. No local da picada dos insetos vetores pode formar-se pequeno nódulo endurecido (de alguns milímetros) de cor pálida ou levemente pigmentada. Esta lesão inicial localiza-se de preferência no rosto e desaparece antes que surja o quadro sintomático típico.

Na maioria dos casos estudados, a leishmaniose visceral ou começa de modo gradual, imperceptivelmente, com perda de apetite, palidez e ligeiro aumento do baço, com elevação da temperatura, depois; ou então, de forma abrupta, com febre alta (39,5-40 °C), contínua ou não, lembrando infecções intestinais, febre tifóide ou malária. O aumento do volume do baço ocorre com relativa rapidez. Com o progredir da doença, acentua-se a anemia e há marcada tendência às hemorragias. A evolução da doença pode ser rápida, levando o paciente à morte dentro de algumas semanas ou alguns meses [2].

A prevenção da doença depende em grande parte de evitar-se o contato com o vetor. Os medicamentos nem sempre são efetivos e requerem repetidos e extensos períodos de tratamento.

Medicamentos

A primeira linha de medicamentos para a leishmaniose visceral é constituída pelos antimoniais pentavalentes. Duas drogas são atualmente usadas: o antimoniato de meglumine e o estibogluconato de sódio (figura 2). São produtos quimicamente similares e com mesmo grau de toxicidade. Com ambos os produtos, manifestam-se efeitos colaterais como anorexia, náuseas, vômitos, mal-estar, dor de cabeça e, às vezes até lesão renal [2].

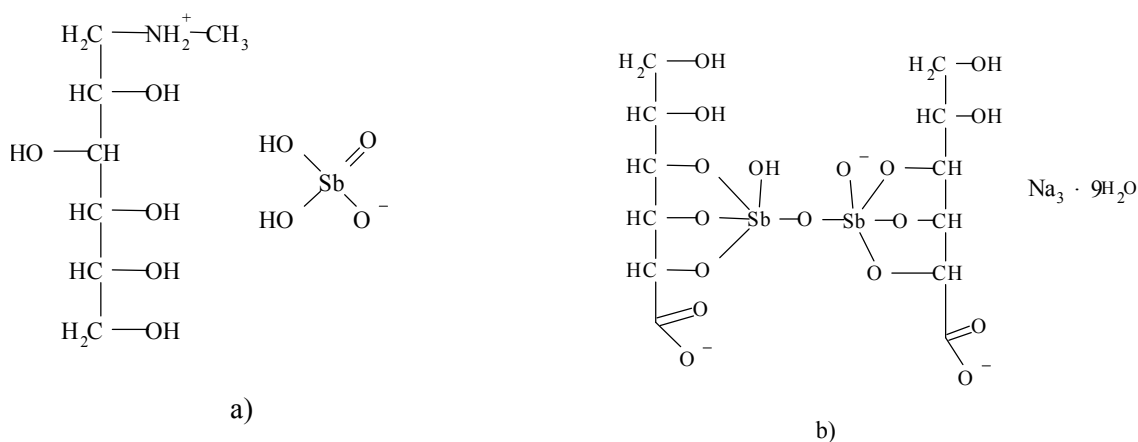


Figura 2 - Medicamentos usados atualmente contra a leishmaniose; a) antimoniato de meglumine e b) estibogluconato de sódio.

Pode ocorrer ausência de resposta ao tratamento e, nesse caso, deve-se passar à segunda linha de medicamentos, que são as Diamidinas Aromáticas. Estas são mais tóxicas do que os primeiros. São eliminados do sangue lentamente, sendo excretadas pela urina durante várias semanas. Os efeitos tóxicos imediatos são: hipotensão, dor de cabeça, fraqueza, dificuldade de respiração, sudorese, sensação de formigamento, vômitos. Nos tratamentos prolongados podem produzir-se lesões hepáticas, pancreáticas, renais e do sistema nervoso.

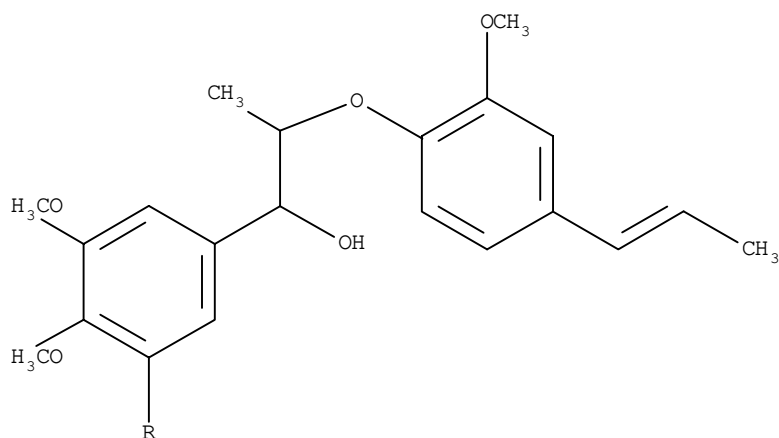
A anfotericina B é também um produto de eliminação lenta e os efeitos colaterais mais comuns, são: anorexia, náuseas, febre, calafrios, anemia. Pode haver produção local de tromboflebite. Por estes motivos, tem-se procurado desenvolver medicamentos menos nocivos ao organismo humano, à base de compostos orgânicos.

Calcula-se que o Brasil disponha de aproximadamente 55 mil espécies vegetais (Barata, L.E.S - comunicação pessoal) e provavelmente 18% delas deve conter propriedades terapêuticas. Entretanto, o país chega a importar matéria-prima para 90% dos remédios sintéticos consumidos pela população, gastando com isto, 2 a 3 bilhões de dólares por ano [3]. Esses números mostram a necessidade de se explorar a nossa flora, no sentido de se investigar as espécies com propriedades medicinais e seus princípios-ativos, com a finalidade de produzir-se novos medicamentos, mais baratos, talvez mais eficientes do que

aqueles usados atualmente e com a possibilidade de se encontrar espécies ativas para doenças , às quais a cura não é ainda conhecida.

Ultimamente, a química e atividade biológica de um certo grupo de compostos, denominados Neolignanas, têm despertado grande interesse. Neolignanas são dímeros formados pelo acoplamento oxidativo de alil e propenil fenóis; esses compostos são encontrados em plantas da família *Myristicaceae*, sendo o gênero *Viola* o mais representativo nas Américas [4].

Em 1970, foram realizados os primeiros estudos pelo grupo do professor Dr. L.E.S. Barata (IQ - UNICAMP), das folhas de *Viola surinamensis*, uma árvore da família *Myristicaceae*, abundante na Amazônia. O extrato bruto dessas folhas apresentou atividade nos ensaios de proteção contra a penetração de cercárias do *Schistosoma mansoni* em camundongos. O fracionamento desse extrato levou ao isolamento de uma mistura de duas neolignanas denominadas Surinamensina (a) e Virolina (b), mostradas na figura 3 [4].

a) R = -OCH₃ (Surinamensina)

b) R = H (Virolina)

Figura 3 - Os compostos a (Surinamensina) e b (Virolina) foram isolados do extrato bruto de folhas de *Virola surinamensis* e mostraram proteção contra a penetração de cercárias de *Schistosoma mansoni*, pela pele.

Com a finalidade de investigar a atividade biológica de neolignanas, o mesmo grupo citado acima sintetizou uma série de compostos análogos àqueles naturais da figura 3, e os submeteu à avaliações nos ensaios antifúngico, antibacteriano, anti-leishmaniose, anti-esquistossomose, anti-PAF (Fator de Agregação Plaquetária) e anti-câncer, tendo mostrado especial interesse em se investigar os fatores responsáveis pela atividade anti-leishmaniose e assim, sintetizar compostos potencialmente mais ativos.

Estudos anteriores realizados com vinte neolignanas (tabela 2) testadas em leishmaniose visceral nos deram informações sobre as conformações mais prováveis [5], mas as relações estrutura-atividade obtidas não foram satisfatórias; assim, continuamos esse trabalho a fim de melhorarmos os resultados obtidos [1, 6]. Para isso, traçamos um roteiro de estudo seguindo os objetivos descritos no final deste capítulo.

1.2 - ENSAIOS BIOLÓGICOS

As neolignanas da tabela 2 foram testadas contra *Leishmania donovani* na London School of Hygiene and Tropical Medicine - University of London, sob coordenação do Dr. Ralph Neal inicialmente e depois, do Dr. Simon Croft. Os ensaios *in vivo* foram realizados

apenas para alguns dos compostos citados, enquanto que os testes *in vitro* foram realizados para todos; sendo assim, utilizaremos os resultados dos testes *in vitro*.

Nos ensaios *in vitro* foram utilizados macrófagos peritoneais de hamsters, infectados com amastigotas (células do leishmania); as drogas foram adicionadas ao meio de cultura e após 7 dias de exposição à droga, os macrófagos foram fixados e tingidos com Giensa. A proporção de macrófagos livres de amastigotas foi determinada microscopicamente. As concentrações iniciais de cada composto testado foram de 300 e 100 μM , e onde a eficácia foi observada, a concentração foi reduzida (3 diluições) nos experimentos subsequentes, até que uma atividade baixa ou nula fosse mostrada. Os resultados foram analisados utilizando regressão linear e teste T para a significância estatística [7].

Neste trabalho, foram analisados não só os compostos ativos, mas também os inativos, pois os métodos que utilizamos (Reconhecimento de Padrões) requerem os dois tipos de compostos.

1.3 - COMPOSTOS ESTUDADOS

As neolignanas estudadas neste trabalho estão dispostas na figura 4a e na tabela 2. No planejamento de fármacos usando química teórica, a etapa mais importante do trabalho é a análise conformacional dos compostos, pois devemos procurar entre os mínimos locais, aquele com maior probabilidade de ligar-se ao receptor; esse mínimo é denominado “conformação ativa” e assim, todas as propriedades físico-químicas são calculadas para os compostos nessa conformação.

Em moléculas flexíveis como as neolignanas, que possuem 4 eixos de rotação livre no seu esqueleto, vários mínimos locais são encontrados.

Como veremos adiante, no capítulo 2, sobre a metodologia, várias considerações foram feitas na tentativa de encontrarmos a conformação ativa mais provável. Além disso, procuramos na literatura outras séries de compostos ativos em leishmaniose, a fim de compararmos a sua geometria com a das neolignanas.

Entre todos os compostos com atividade em leishmaniose que encontramos na literatura [8 - 17], uma série de aminocetonas apresentou alguma semelhança com as

neolignanas, pois ambas as classes de compostos apresentam dois anéis fenílicos com uma carbonila entre eles. As aminocetonas são mostradas na tabela 1 e figura 4b.

Sabe-se que muitas aminocetonas são capazes de inibir uma enzima importante (2,3-oxidosqualeno lanosterol-ciclase) utilizada na biossíntese de esteróis de fungos. As aminocetonas estudadas aqui mostraram também, atividade anti-leishmaniose [8]. As membranas celulares, tanto de fungos como do *Leishmania*, contém ergosterol, em contraste com as membranas celulares de mamíferos que possuem colesterol; a diferença (embora ainda desconhecida) entre a enzima "chave" na síntese de esteróis do *Leishmania* e dos mamíferos, a torna um alvo à pesquisa de novos compostos anti-leishmaniose.

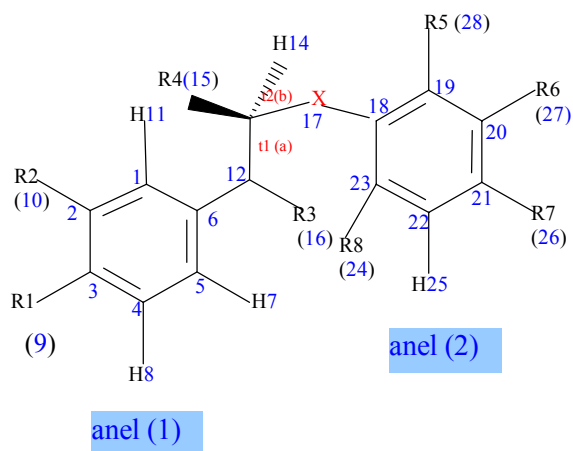
Paralelamente à pesquisa das aminocetonas, encontramos outros estudos que revelam um possível sítio ativo às drogas anti-leishmaniose.

Protozoários parasitas são agentes causadores de uma variedade de doenças fatais, como é o caso do *Leishmania*. Foi observado [18] que esses protozoários não possuem a habilidade de sintetizar purinas, mas captam do hospedeiro, usando enzimas específicas no processo, sendo a adenosina-kinase uma delas. Pesquisadores de vários laboratórios concentraram sua atenção nessas enzimas e foi revelado que o aminoácido arginina está presente apenas na adenosina-kinase do *Leishmania donovani*, não estando presente à mesma enzima de hamsters, um hospedeiro natural do *Leishmania donovani*.

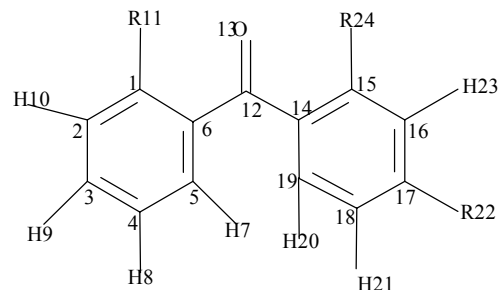
Ghosh, M. e Datta, A.K., observaram que reagentes específicos da arginina - fenilglioxal (PGO), butano-2,3-diona e ciclohexano-1,2-diona - inativam a enzima irreversivelmente. Em contraste, a adenosina-kinase de hamster é insensível a estes reagentes [18].

Como veremos no capítulo 8, a arginina possui geometria favorável à interação com um dos mínimos locais das neolignanas. Essa conformação, por sua vez, é semelhante ao mínimo global das aminocetonas. Isso sugere que a arginina também poderia ser o sítio ativo das aminocetonas. Assim, a arginina poderia ser a “diferença” citada entre a enzima “chave” na síntese de esteróis do *Leishmania* e de mamíferos. Entretanto, isto é apenas uma suposição baseada na semelhança geométrica dos compostos em estudo.

Na tabela 2, os compostos numerados de 1 a 20 são neolignanas pertencentes ao conjunto treinamento, isto é, são aqueles já ensaiados em leishmaniose [7], sendo inativos os 11 primeiros e ativos os restantes.



a) NEOLIGNANAS



b) AMINOCETONAS

Figura 4 - a) Estrutura básica das neolignanas, sendo R1 a R8 os substituintes mostrados na tabela 2 e X, oxigênio ou enxofre; t1 e t2 são os eixos de torção referentes aos ângulos diedros 14-13-12-6 e 18-17-13-12; os dois anéis fenílicos das neolignanas foram identificados por anel (1) e anel (2), para facilitar a discussão no decorrer do trabalho. b) Estrutura básica das amino-cetonas estudadas, sendo R11, R22 e R24 os substituintes mostrados na tabela 1.

Obs: A numeração para as figuras 4 a e 4 b, embora não seja convencional, será utilizada neste trabalho para facilitar a discussão.

Tabela 1 - Aminocetonas ativas em leishmaniose; ED₅₀ é uma medida de atividade biológica que indica a concentração molar, correspondente a 50% da dose efetiva, sendo que quanto maior o valor da dose efetiva, menor a atividade biológica.

COMPOSTO	R11	R22	R24	ATIVIDADE (ED ₅₀)
1	H	(CH ₂) ₇ -NH ₂	H	0,6
2	Cl	(CH ₂) ₇ -NH ₂	H	0,9
3	H	(CH ₂) ₈ -NH ₂	H	0,7
4	H	-CH ₂ -O-(CH ₂) ₆ -NH ₂	CH ₃	3,5

Tabela 2 - Os compostos numerados entre 1 e 20 são neolignanas e foram ensaiados em leishmaniose [7], sendo onze inativos (-) e nove ativos (+). As posições dos substituintes, R e X, são mostrados na figura 4a.

COM- POSTO	R ₁	R ₂	R ₃	R ₄	R ₅	R ₆	R ₇	R ₈	X	ATIV IDAD E
1	H	H	=O	H	H	H	H	H	O	-
2	H	H	=O	H	H	H	Cl	H	O	-
3	H	H	=O	H	H	H	-OCH ₃	H	O	-
4	H	H	=O	H	-OCH ₃	H	H	H	O	-
5	-OCH ₃	-OCH ₃	=O	CH ₃	H	H	H	H	O	-
6	-OCH ₃	-OCH ₃	=O	CH ₃	-OCH ₃	H	H	H	O	-
7	-OCH ₃	-OCH ₃	=O	CH ₃	-OCH ₃	H	(CH) ₂ -CH ₃	H	O	-
8	H	H	=O	H	-NO ₂	H	H	H	O	-
9	H	H	=O	H	H	-CH ₃	H	H	O	-
10	-OCH ₃	-OCH ₃	=O	CH ₃	H	H	H	H	S	-
11	H	H	=O	H	H	H	H	H	S	-
12	-OCH ₃	-OCH ₃	=O	H	H	H	-OCH ₃	H	O	+
13	-OCH ₃	-OCH ₃	=O	H	-OCH ₃	H	H	-OCH ₃	O	+
14	-OCH ₃	-OCH ₃	=O	H	-OCH ₃	H	(CH) ₂ -CH ₃	H	O	+
15	-OH	-OCH ₃	H	CH ₃	-OCH ₃	H	(CH) ₂ -CH ₃	-OCH ₃	O	+
16	-OCH ₃	-OCH ₃	H	CH ₃	-OCH ₃	H	(CH) ₂ -CH ₃	-OCH ₃	O	+
17	H	H	=O	CH ₃	H	H	Cl	H	S	+
18	-OCH ₃	-OCH ₃	=O	CH ₃	H	H	-CH ₃	H	S	+
19	-OCH ₃	-OCH ₃	=O	CH ₃	H	H	Cl	H	O	+
20	-OCH ₃	-OCH ₃	=O	CH ₃	H	H	Cl	H	S	+

1.4 - OBJETIVOS

Os objetivos deste trabalho são mostrados a seguir:

- Procura de descritores (ou parâmetros) físico-químicos que expliquem a atividade anti-leishmaniose das neolignanas estudadas, através da inclusão de novos parâmetros em adição àqueles já calculados em trabalhos anteriores [1, 6].
- Verificação da aplicabilidade dos métodos de Reconhecimento de Padrões (RP) em SAR (Relações entre estrutura química e atividade biológica), utilizados neste trabalho, através de estudos comparativos entre os mesmos e Regressão Linear Múltipla (RLM) que é um método estatístico comumente usado em estudos de QSAR (Relações Quantitativas entre Estrutura-química e Atividade biológica).
- Verificação do efeito da solvatação em água, sobre as conformações encontradas anteriormente [5] para as neolignanas isoladas.
- Comparação entre as conformações encontradas para as neolignanas e as conformações de outros compostos ativos em leishmaniose, com a finalidade de obtermos alguma indicação sobre a conformação ativa das neolignanas;
- Comparação entre as conformações das neolignanas e a geometria do possível receptor;
- Cálculo dos parâmetros físico-químicos, para a conformação supostamente ativa;
- Estabelecimento de relações entre a atividade anti-leishmaniose e a estrutura química das neolignanas;
- Sugestão de novos compostos para síntese;
- Proposta de um roteiro de análise SAR, usando-se apenas química teórica e computacional, para séries de compostos que apresentem semelhanças àqueles que estamos estudando, com relação à diversidade de substituintes e à falta de valores de AB.

CAPÍTULO 2

METODOLOGIA

2.1 - ANÁLISE CONFORMACIONAL E A SOLVATAÇÃO EM ÁGUA.

A primeira etapa no planejamento de fármacos, usando química computacional, é a determinação da geometria na qual serão calculados os parâmetros físico-químicos para correlacionar com a atividade biológica. Se a molécula em questão for rígida, esta geometria procurada poderá ser o seu mínimo global; entretanto, para moléculas flexíveis, com vários eixos de rotação livre, mais de uma conformação de mínima energia será produzida numa análise bem detalhada.

Geralmente, numa molécula com dois eixos de rotação livre por exemplo, um dos eixos é mantido fixo e o outro girado de 0 a 360 graus, a cada 15 ou 10 graus; em seguida, o primeiro eixo que estava fixo é girado de 10 ou 15 graus e fixado nessa posição enquanto o outro é girado novamente de 0 a 360 graus. Todos os valores de energia produzidos são anotados e então podemos verificar as posições de mínima energia.

Nem sempre o mínimo global é a conformação adequada para correlação com a atividade biológica, pois o planejamento de fármacos envolve a procura da conformação ativa, isto é, aquela que melhor se ajusta ao receptor.

Em estudos anteriores, uma metodologia envolvendo termodinâmica estatística foi aplicada ao nosso conjunto de compostos, após termos obtido os respectivos mínimos locais, a fim de procurarmos entre eles, aquele com maior probabilidade de ser a conformação ativa. Essa metodologia considera também que os fármacos são moléculas flexíveis, capazes de alterar a sua conformação no momento de se ligar ao sítio receptor. Esse estudo envolveu cálculos de otimização de geometria com o método de mecânica molecular (MM2), usando-se constante dielétrica para vácuo e para meio biológico, com finalidades comparativas [5, 6]. Notamos, com esse estudo, a influência que a constante dielétrica exerce sobre o valor da energia total calculada para cada conformação; essa influência altera também a estabilidade relativa entre os mínimos locais.

Uma maneira mais realista de se considerar o meio em que o fármaco atua, seria através da solvatação dos compostos em água, um importante constituinte dos sistemas biológicos. Normalmente, cálculos de otimização de geometria de fármacos são realizados na temperatura biológica [19].

Um sistema formado por um composto de interesse biológico, solvatado por moléculas de água, por exemplo, pode ser estudado pelo cálculo de propriedades termodinâmicas de equilíbrio, se conhecermos a função de partição.

A termodinâmica estatística permite quantificar as propriedades macroscópicas, a partir das características moleculares dos sistemas, uma vez que a termodinâmica reflete o comportamento médio do sistema. Assim, é possível calcular as propriedades termodinâmicas do sistema, a partir das interações intermoleculares. Para representar o estado líquido, deve-se considerar a desordem do sistema, existindo assim inúmeras configurações possíveis, isto é, o sistema líquido não pode ser definido por uma única estrutura. Dessa forma, as equações de mecânica estatística não possuem solução analítica e neste caso é usado o cálculo numérico. A simulação computacional pode ser definida como uma determinação numérica computacional de propriedades termodinâmicas, energéticas, estruturais e dinâmicas de um modelo matemático do arranjo molecular correspondente [20]. Essas simulações podem ser realizadas por meio probabilístico, usando o método de Monte Carlo, ou por meio determinístico, integrando as equações de movimento de Newton, com o método de dinâmica molecular.

O método de Monte Carlo [21] envolve o uso de números randômicos no cálculo de uma sequência de estados, cuja evolução é determinada por eventos aleatórios. Por exemplo, supondo-se um sistema contendo N moléculas que interagem segundo uma superfície de potencial conhecida e que estão submetidas a condições externas controladas (temperatura, pressão, etc), configurações espaciais são geradas aleatoriamente e aceitas ou não, de acordo com a distribuição de Boltzmann, de forma que as integrais de interesse são aproximadas de um somatório; valores das observáveis mecânicas de interesse são calculados para cada uma destas configurações. No final deste processo, os valores médios destas observáveis são obtidos. Com este procedimento, pode-se calcular a energia interna, entalpia, volume, etc, do sistema investigado [22].

2.1.1 - O Método de Monte Carlo [21]

O Método de Monte Carlo é um método de simulação computacional para sistemas com muitos graus de liberdade, podendo ser utilizado para calcular valores de integrais em espaços multidimensionais. O seu nome origina-se do fato de gerar “números randômicos” similarmente aos jogos de roleta.

O valor esperado (ou valor médio) de uma função $g(x)$ qualquer, no intervalo entre $[a, b]$ é dado por:

$$\langle g \rangle = \int_a^b g(x) P(x) dx \quad \text{eq. 2.1}$$

onde $P(x)$ é a função probabilidade.

O método de Monte Carlo aproxima $\langle g \rangle$ de um somatório (eq.2.2), no qual são considerados apenas os pontos “ x ” que contribuem predominantemente para o valor da integral:

$$\langle g \rangle = \frac{1}{N} \sum_{i=1}^N g(x_i) \quad \text{eq.2.2}$$

Esses pontos “ x_i ” são gerados através de números randômicos e selecionados de acordo com a probabilidade $P(x)$.

Nos sistemas moleculares, propriedades macroscópicas são representadas pelos valores médios dos estados microscópicos do sistema e o valor médio $\langle A \rangle$ de uma propriedade termodinâmica, em mecânica estatística, pode ser calculado conhecendo-se a função de partição no *ensemble* considerado.

Ensemble é uma terminologia utilizada em mecânica estatística para representar uma coleção imaginária de estados com parâmetros termodinâmicos comuns.

Inicialmente, o sistema molecular é descrito em termos do Hamiltoniano e o *ensemble* apropriado é escolhido. O valor esperado da propriedade termodinâmica A poderia ser calculado então, pela equação 2.3:

$$\langle A \rangle = \frac{\int_{\Omega} A(x) e^{(-H(x)/kT)} dx}{1/(N! h^{3N}) \int_{\Omega} e^{(-H(x)/kT)} dx} \quad \text{eq. 2.3}$$

onde Ω é o espaço de fase, um espaço multidimensional ao qual estão associados as posições e momentos de cada microestado; a variável x está representando uma configuração do sistema. O termo $e^{(-H(x)/kT)}$, no numerador é a função distribuição no *ensemble* considerado, sendo k a constante de Boltzmann e $H(x)$ o Hamiltoniano para o *ensemble* escolhido. O denominador da equação é a função de partição, sendo N o número de moléculas. A equação 2.3 pode ser escrita então, de forma análoga à equação 2.2:

$$\langle A \rangle = \int_{\Omega} A(x) P(x) dx$$

onde a probabilidade $P(x)$ é descrita da seguinte maneira:

$$P(x) = \frac{e^{(-H(x)/kT)}}{1/(N! h^{3N}) \int_{\Omega} e^{(-H(x)/kT)} dx} \quad \text{eq. 2.4}$$

Para descrever o Hamiltoniano na equação 2.3, é necessário que se conheça o momento, posição e a energia potencial das partículas constituintes do sistema, conforme a equação 2.5:

$$H(p, q) = K(p) + U(q) \quad \text{eq. 2.5}$$

onde q é o conjunto de coordenadas e p os momenta das moléculas, sendo que $K(p)$ é a energia cinética.

As interações moleculares e as variações na energia interna das moléculas são responsáveis pelo termo potencial, $U(q)$, que será discutido no item “Potencial de interações intermoleculares”, mais a frente.

Como o espaço de fase de um sistema químico é multidimensional, o cálculo das integrais na equação 2.3 não pode ser efetuado analiticamente ou por métodos numéricos convencionais. Utiliza-se então, o método de Monte Carlo que aproxima o valor esperado de A , na equação 2.3, pelo somatório:

$$\langle A \rangle = \frac{1}{N} \sum_i^N A(x_i) \quad \text{eq. 2.6}$$

No somatório acima, os estados x_i são seleccionados de acordo com a probabilidade apresentada na equação 2.4.

Metrópolis e colaboradores desenvolveram um algoritmo que permite gerar estados distribuídos de acordo com $P(x)$, a partir de uma configuração arbitrária; o algoritmo é apresentado a seguir:

- 1 - Inicialmente, escolhe-se uma configuração x_0 ;
- 2 - Calcula-se a energia total $E(x_0)$ desta configuração;
- 3 - Uma nova configuração x_i é gerada;
- 4 - Calcula-se a energia total $E(x_i)$ da nova configuração;
- 5 - Calcula-se a diferença de energia entre as duas configurações:
 $\Delta E = E_i - E_0$
- 6 - Se $\Delta E < 0$, a nova configuração é aceita; caso contrário, faz-se nova checagem;
- 7 - Selecciona-se um número randômico R , no intervalo $[0,1]$;
- 8 - Se $\exp(-\Delta E/kT) > R$, a nova configuração é aceita; faz-se $x_0 = x_i$ e volta-se à etapa 3.
- 9 - Se $\exp(-\Delta E/kT) < R$, a nova configuração é rejeitada e a configuração anterior x_0 é contada novamente. Volta-se à etapa 3.

A etapa 3 do algoritmo apresentado depende do *ensemble* considerado. No *ensemble* NVT, por exemplo, (número de partículas, volume e temperatura fixos), a nova configuração seria gerada através da rotação e/ou translação aleatória de uma molécula escolhida ao acaso. Já no *ensemble* isotérmico e isobárico, NpT, o volume é uma variável e a nova configuração pode ser gerada variando-se também o volume, de uma quantidade aleatória ΔV .

À medida que a sequência de configurações é gerada, os valores de $A(x_i)$ são calculados e o valor médio $\langle A \rangle$ é obtido no final, pela equação 2.6. Em geral, a

convergência para os valores médios é obtida com aproximadamente 3×10^6 vezes configurações do sistema [21, 22].

Através do método de Monte Carlo com o algoritmo de Metrópolis apresentado, resolveu-se o problema devido ao imenso número de configurações possíveis. Entretanto, existe ainda o problema do grande número de moléculas.

Na prática, o número N de moléculas que constituem o sistema deve ser da ordem de algumas centenas, para que o cálculo seja possível. Entretanto, uma configuração significa um conjunto de $3N$ coordenadas, para N sendo da ordem do número de Avogadro, 10^{23} . Para contornar esse problema, Metrópolis e colaboradores introduziram as “condições de contorno periódicas” [21], que consistem em tomar um cela com paredes fictícias contendo de 100 a 10000 partículas e replicar esta cela infinitas vezes. Assim, uma caixa cúbica com algumas moléculas no seu interior é circundada por réplicas que formam uma rede infinita. Quando as moléculas movem-se na caixa central, as suas imagens repetem os mesmos movimentos nas caixas réplicas e como as paredes são fictícias, com a saída de uma molécula da caixa original, uma das moléculas-imagem atravessa a lateral oposta, na direção do interior desta caixa. Isto torna constante o número de moléculas na caixa de referência. Como as moléculas interagem também com as imagens das réplicas vizinhas, deve-se estabelecer um raio de corte para limitar o número de interações.

O Potencial de Interação Intermolecular

O potencial de interação entre dois corpos, para um sistema de N moléculas, pode ser decomposto na soma de energias de interação E_{AB} entre moléculas a e b. Esse potencial pode ser escrito como na equação 2.7, se considerarmos interação entre moléculas rígidas. Para moléculas com graus de liberdade internos, como rotação, estiramento de ligações, etc., seria necessário acrescentar outros termos de energia potencial.

$$E_{ab} = \sum_i^{n_a} \sum_j^{n_b} [(A_{ij} / r_{ij}^{12}) - (B_{ij} / r_{ij}^6) + (q_i q_j / r_{ij})] \quad \text{eq. 2.7}$$

Na equação 2.7, o somatório é efetuado sobre n_a sítios da molécula a e n_b sítios da molécula b. r_{ij} é a distância entre o sítio i da molécula a e o sítio j da molécula b. Para um dado sítio k, os parâmetros A_{kk} e B_{kk} são escritos como: $A_{kk} = 4\epsilon_k \sigma_k^{12}$ e $B_{kk} = 4\epsilon_k \sigma_k^6$, onde ϵ_k e σ_k são os parâmetros de Lennard-Jones para este sítio. Parâmetros para uma interação diagonal (ij) podem ser obtidos via regra de cruzamento, onde: $A_{ij} = (A_{ii} A_{jj})^{1/2}$ e $B_{ij} = (B_{ii} B_{jj})^{1/2}$. Os sítios são distribuídos sobre a estrutura molecular de forma a assegurar a definição correta da simetria da superfície de potencial. As cargas pontuais, q_k são utilizadas para representar sistemas iônicos e possíveis assimetrias na densidade de carga.

Em geral, os parâmetros necessários para definir a equação 2.7 são calculados através de métodos de química quântica e/ou ajustes para reproduzir dados experimentais [23], sendo que os parâmetros de Lennard-Jones são transferíveis entre moléculas [24].

2.1.2 - O método de Mecânica Molecular

Neste trabalho, o método de mecânica molecular foi usado para obtermos as conformações de mínima energia das neolignanas, as quais foram utilizadas posteriormente, nos cálculos de solvatação. Embora esta etapa tenha sido realizada e apresentada em trabalhos anteriores [5, 6], achamos conveniente descrever brevemente a metodologia, apenas para dar uma noção ao leitor sobre o que se trata a mecânica molecular. Detalhes sobre os parâmetros utilizados podem ser encontrados nas referências 5 e 6.

Apenas o conjunto teste de neolignanas, apresentado na tabela 21 do capítulo 9 (página 153), teve a análise conformacional realizada totalmente neste trabalho, também com o método de mecânica molecular - MM2, seguindo os critérios e parâmetros já estabelecidos para o conjunto treinamento, analisado nas referências 5 e 6.

No método de mecânica molecular, um conjunto de funções potenciais são usados para descrever as interações intramoleculares do sistema em estudo. Esse conjunto de funções é denominado “Campo de força” e contém parâmetros ajustáveis que são otimizados para obter o melhor conjunto de propriedades das moléculas, como geometrias, calores de formação, energias conformacionais ou outras propriedades. Neste método, os movimentos dos núcleos são estudados e os elétrons são considerados implicitamente, como um potencial efetivo; a energia da molécula no estado eletrônico fundamental é função da posição nuclear [25].

Os cálculos de campo de força têm suas origens na espectroscopia vibracional, sendo que em mecânica molecular o campo de força apresenta um potencial harmônico, do tipo mostrado na equação 2.8 :

$$V_{pot} = \frac{1}{2} \sum_i f_{xi} (x_i - x_{0i})^2 \quad \text{eq. 2.8}$$

onde f_{xi} é a constante de força x para a coordenada i .

Os termos (funções potenciais) incluídos num campo de força, usualmente são: estiramento de ligação (V_b), deformação angular (V_θ), torsão (V_{tor}), interações entre átomos não ligados que incluem as interações de van der Waals (vdW) e de dipolos (V_{dip}), e

os termos cruzados: interação torção/deformação angular (V_{tor} / θ) e estiramento/deformação angular (V_b / θ).

Assim, exemplificando com a função de estiramento de ligação:

$$V_b = \frac{1}{2} \sum k_b (r - r_0)^2 \quad \text{eq. 2.9}$$

onde k_b é a constante para o estiramento de ligação e r_0 é o comprimento de ligação de referência; o somatório é feito sobre todas as ligações; k_b e r_0 são parâmetros ajustáveis no campo de força.

A soma de todos os termos do campo de força é chamada energia estérica da molécula e para que o campo de força reproduza bem os dados experimentais, é necessário otimizar esses parâmetros ou introduzir outros. Algumas técnicas experimentais usadas para determinar estruturas são: difração de raios-x para cristais e difração de elétrons e espectroscopia de microondas para gases.

Em mecânica molecular, parâmetros e constantes de força podem ser transferidos de uma molécula para outra; assim é comum otimizarmos os parâmetros para um conjunto de moléculas simples, usando dados experimentais para comparação e posteriormente, usar os mesmos parâmetros para moléculas mais complexas e semelhantes às primeiras.

O campo de força desenvolvido em mecânica molecular, denominado MMI, surgiu em 1973 [26] e em 1977, surgiu a segunda geração de mecânica molecular - MM2 [27], o qual aperfeiçoou consideravelmente o campo de força de 1973 (MMI), devido principalmente à introdução do termo torsional V_2 (inicialmente pensou-se que apenas o potencial para o termo V_3 era suficiente para descrever todos os tipos de ângulos de torção, para um determinado sistema em estudo).

O campo de força de 1977 do MM2 apresentava falhas na reprodução de barreiras de rotação internas nas ligações simples C-C. Além disso, o MM2 também possui a tendência de superestimar as distâncias H...H. Alguns parâmetros torsionais são muito pequenos, enquanto os parâmetros de van der Waals para as interações H...H são muito grandes. Assim, MM2 foi reparametrizado, surgindo um novo campo de força: o MM2' [28]. Parâmetros torsionais e potenciais de átomos não ligados foram reajustados.

Em 1987, um novo campo de força de MM2 surgiu com novos ajustes no campo de força de 1977 [29]. As ligações hidrogênio são consideradas apenas qualitativamente nas

versões anteriores do MM2, sendo que as ligações hidrogênio experimentais são mais curtas e fortes do que aquelas calculadas pelas versões iniciais do MM2. Esta falha é corrigida pelo MM2 (87), através da redução dos raios de van der Waals dos hidrogênios envolvidos nas ligações hidrogênio e pelo incremento na atração de van der Waals entre o hidrogênio e vários átomos eletronegativos.

Utilizamos o campo de força do MM2 (87) para as moléculas estudadas neste trabalho, sendo que incluímos alguns parâmetros do campo de força MM2'. Os parâmetros foram obtidos pelo Prof. Dr. Y. Hase (IQ-Unicamp) através de carta ao autor do método, tendo-nos fornecido, posteriormente. Alguns parâmetros torsionais foram obtidos por nós, na referência 6.

2.2 - RELAÇÕES ENTRE A ESTRUTURA QUÍMICA E A ATIVIDADE BIOLÓGICA.

A relação entre a estrutura de uma série de compostos análogos e a sua atividade biológica é realizada a partir de dados sobre a estrutura química desses compostos e sua atividade; esses dados, que são parâmetros físico-químicos devem representar a hidrofobicidade dos compostos, o efeito estérico e também o efeito eletrônico produzido na molécula, por determinado substituinte [30]. Tais parâmetros podem ser obtidos experimentalmente ou teoricamente, sendo que às vezes os dois tipos são incluídos na análise. Neste trabalho, freqüentemente usaremos os termos "variáveis" e "parâmetros" como sinônimos, ambos significando parâmetros físico-químicos.

Se os dados de atividade biológica (AB) forem numéricos, como por exemplo, valores de ED_{50} (concentração molar correspondente a 50% da dose efetiva) ou de IC_{50} (concentração molar da droga capaz de inibir, em 50%, determinada função no organismo), a relação entre essa atividade e a estrutura química dos compostos é quantitativa e denomina-se "Quantitative Structure-Activity Relationships - QSAR". Por outro lado, se valores numéricos sobre a atividade não são disponíveis, mas apenas temos conhecimento se o composto apresenta atividade ou não, a relação obtida não é quantitativa e denomina-se SAR. No primeiro caso, a relação estrutura-atividade é obtida pelo método estatístico de

Regressão Linear Múltipla (RLM); este é um método estatístico clássico, usado freqüentemente em QSAR.

A série de neolignananas estudadas neste trabalho não dispõe de valores para a atividade anti-leishmaniose e, portanto, não foi possível utilizarmos o método de regressão na análise. Os métodos de Reconhecimento de Padrões (RP) foram aplicados neste caso.

2.2.1 - Os métodos de Reconhecimento de Padrões (RP)

Os métodos RP [31] permitem agrupar objetos de uma mesma categoria, de acordo com a similaridade entre eles. Neste trabalho, os objetos são moléculas pertencentes a duas categorias diferentes: ativa e inativa. Os objetos com categorias conhecidas pertencem ao conjunto chamado treinamento, do qual informações são extraídas para tornar possível a classificação de novos objetos (conjunto teste). Assim, este método possibilita classificar novas moléculas em ativas ou inativas, desde que se conheça os parâmetros físico-químicos adequados, que estejam relacionados com a atividade biológica, isto é, esses parâmetros devem ser capazes de separar os compostos nas duas categorias.

Inicialmente, muitos parâmetros podem ser incluídos, tanto quanto se possa imaginar e obter, uma vez que em RP o número de variáveis pode exceder o número de objetos e que uma das principais características desses métodos é reduzir o número de variáveis, como veremos mais à frente, no item Análise de Componentes Principais.

Apesar do grande número de variáveis ser possível em RP, nem sempre todas elas estão relacionadas com a AB, produzindo resultados ruins. Estas variáveis devem ser identificadas e eliminadas da análise.

Os métodos RP podem ser divididos em dois grupos, sendo que um deles baseia-se na similaridade entre os compostos de determinada categoria e o outro, baseia-se na análise de componentes principais. Para os dois grupos de métodos de RP, inicialmente é construída uma matriz com as variáveis obtidas para cada molécula, como na figura 5, em que os objetos são considerados pontos no espaço M dimensional, onde M é o número de medidas feitas sobre esses objetos.

		Objetos					
		1	2	...	k	...	N
Variáveis	1	y_{11}	y_{12}	...	y_{1k}	...	y_{1N}
	2	y_{21}	y_{22}	...	y_{2k}	...	y_{2N}
	...						
	3	y_{31}	y_{32}	...	y_{3k}	...	y_{3N}
	...						
	i	y_{i1}	y_{i2}	...	y_{ik}	...	y_{iN}
...							
M	y_{M1}	y_{M2}	...	y_{Mk}	...	y_{MN}	
		categoria 1			categoria Q		conjunto teste
		conjunto treinamento					

Figura 5 - Representação dos N objetos num espaço M dimensional, onde M é o número de medidas feitas sobre esses N objetos.

Nos métodos de RP baseados na similaridade entre os compostos, inicialmente é calculada a distância entre os objetos. Neste trabalho foi utilizada a distância euclidiana. A distância euclidiana entre dois pontos i e j de coordenadas Y_{ki} e Y_{kj} é dada por:

$$d_{ij} = \left[\sum_{k=1}^M (Y_{ki} - Y_{kj})^2 \right]^{(1/2)} \quad \text{eq.2.10}$$

onde a soma é feita sobre as M medidas e d_{ij} pode ser relacionado com uma medida de similaridade. Considera-se que quanto menor a distância entre estes, maior a similaridade. Uma medida de similaridade é definida como:

$$s_{ij} = 1 - d_{ij} / (d_{ij})_{\max} \quad \text{eq.2.11}$$

onde $(d_{ij})_{\max}$ é a maior distância interponto no espaço M dimensional. Para objetos idênticos, $s_{ij} = 1$, enquanto que $s_{ij} = 0$ descreve dois pontos que são separados pela maior distância interponto para o conjunto sob estudo.

Antes de iniciar a análise com RP, é necessário escalar os valores obtidos para as propriedades físico-químicas, pois estas possuem unidades diferentes. O escalamento consiste numa transformação dos dados (propriedades físico-químicas), para que eles possam ser comparados numa mesma escala, apesar de possuírem unidades diferentes. Neste trabalho, as propriedades físico-químicas (variáveis) foram transformados tal que a média (eq.10) tornou-se zero e a variância obteve valor 1,0 (eq.11); esse tipo de transformação é chamado de "auto-escalamento" [32].

$$\bar{Y}_i = \sum_{k=1}^N Y_{ik} / N \quad \text{eq.2.12}$$

onde $\bar{Y}_i$ é a média da variável i calculada para os N objetos e Y_{ik} é a medida da variável i para o objeto k .

A variância para a i -ésima variável é:

$$S_i^2 = \sum_{k=1}^N (Y_{ik} - \bar{Y}_i)^2 / (N - 1) \quad \text{eq.2.13}$$

As propriedades calculadas para os objetos, Y_{ik} , foram transformados em Y'_{ik} , pela seguinte equação:

$$Y'_{ik} = (Y_{ik} - \bar{Y}_i) / S_i \quad \text{eq.2.14}$$

tal que:

$$\overline{Y}_i = \sum_{k=1}^N Y_{ik}' / N = 0 \quad \text{eq.2.15}$$

e

$$S_i^2 = \sum_{k=1}^N (Y_{ik}' - \overline{Y}_i')^2 / (N - 1) = 1 \quad \text{eq.2.16}$$

A seguir, serão descritos os métodos de Reconhecimento de Padrões utilizados:

Análise de componentes principais (ACP) [33]

Inicialmente, o conjunto de pontos representando os objetos (ou compostos), localizados no espaço pelos seus valores das propriedades físico-químicas calculadas, é representado por seu ponto central, $\overline{Y}'$ definido por:

$$\overline{Y}' = (\overline{Y}_1 \overline{Y}_2 \dots \overline{Y}_k \dots \overline{Y}_N) \quad \text{eq.2.17}$$

onde $\overline{Y}_k$ é a média da propriedade físico-química k para o conjunto.

A matriz de dados Y , da figura 5 é então representada pela equação 2.18:

$$Y = 1\overline{Y}' + E^{(0)} \quad \text{eq.2.18}$$

onde E é a matriz contendo os desvios das médias.

A figura 6 mostra a representação dos dados Y num espaço tridimensional e a representação do ponto central, $\overline{Y}'$

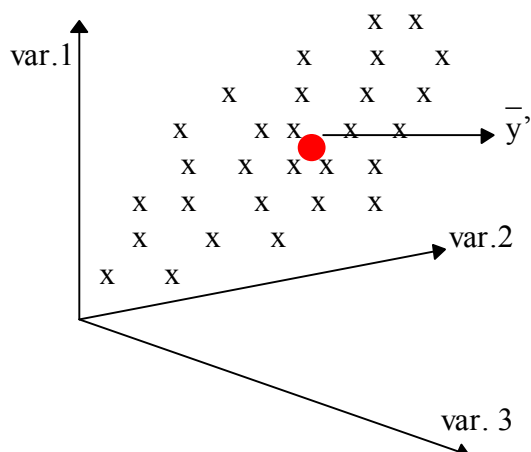


Figura 6 - Representação dos dados Y num espaço tridimensional, e do ponto central, $\bar{Y}$

Agora, uma reta é ajustada aos M pontos no espaço N, de modo que os desvios são os menores possíveis. Os coeficientes de direção desta reta são chamados "Loadings", sendo um para cada variável k, representados pelo vetor P_1 .

Quando cada ponto é projetado nesta linha, obtemos os "Scores" t_{i1} , isto é, a coordenada do ponto i ao longo do eixo P_1 . O valor $t_{i1} P_1$ é subtraído de $E^{(0)}$ para obter os novos desvios $E^{(1)}$:

$$E^{(0)} - t_{i1} P_1 = Y - 1\bar{Y} - t_{i1} P_1 = E^{(1)} \quad \text{-----} \rightarrow Y = 1\bar{Y} + t_{i1} P_1 + E^{(1)}. \quad \text{Eq.2.19}$$

onde $|E^{(1)}| < |E^{(0)}|$

A equação 2.19 descreve o modelo de 1 componente principal que é representada por uma linha no espaço N dimensional e uma equação linear com uma variável, a componente principal (figura 7).

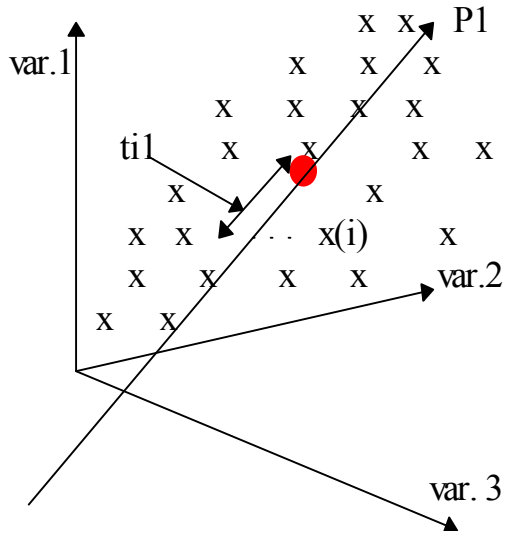


Figura 7- Representação da primeira componente principal, num espaço tridimensional.

Essa linha contém o máximo de variância dos dados, ou seja, o máximo de informação estatística.

Em seguida, subtrai-se o modelo de uma componente principal de Y , o que implica na remoção da direção P_1 :

$$Y - 1. \bar{Y} - t_1 P_1 = E^{(1)} \quad \text{eq. 2.20}$$

Então, ajusta-se uma outra reta aos pontos (perpendicular à primeira componente principal), de modo que os novos desvios são os menores possíveis. Esta reta representa a segunda componente principal:

$$Y = 1. \bar{Y} + t_1 P_1 + t_2 P_2 + E^{(2)} \quad \text{eq.2.21}$$

O modelo de duas componentes principais é representado por um plano no espaço N dimensional e uma equação linear com duas variáveis (as duas componentes principais).

A figura 8 mostra a representação de duas componentes principais.

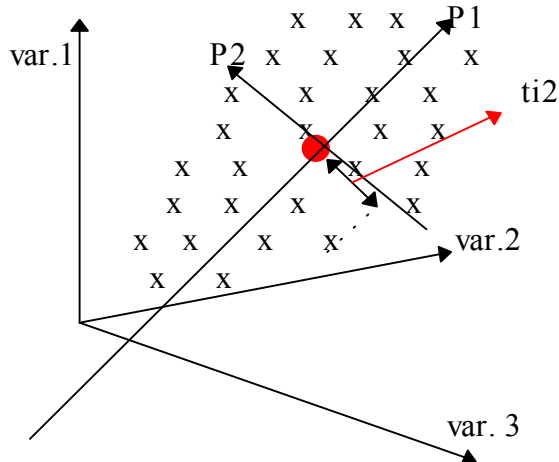


Figura 8- Representação das duas primeiras componentes principais num espaço tridimensional.

Entre todas as retas perpendiculares à linha da primeira componente principal, a linha da segunda componente principal contém a maior variância possível. Este procedimento é repetido N vezes e as últimas componentes principais terão variâncias próximas ou iguais a zero, podendo ser ignoradas.

Interpretação matemática da análise de componentes principais:

Como vimos, o cálculo das componentes principais deve ser realizado de maneira a diminuir a dimensionalidade do problema, sem perda das informações dos dados originais; isto é possível, quando usamos como ponto de partida para este cálculo, a matriz de covariância “C”, que é a matriz de correlação para os dados autoescalados, X:

$$C = X^T X \quad \text{eq. 2.22}$$

onde X^T é a matriz transposta de X.

A correlação entre dois vetores de comprimento unitário quaisquer, é o cosseno do ângulo entre eles e os ângulos entre os vetores podem ser então calculados, a partir de C:

$$\theta_{ik} = \cos^{-1} C_{ik} \quad \text{eq. 2.23}$$

Num problema tridimensional, pode acontecer do ângulo entre duas variáveis ser igual a soma dos ângulos entre as outras duas; por exemplo, $\theta_{13} = \theta_{12} + \theta_{23}$. Neste caso, os três vetores podem ser representados num plano, isto é, a dimensionalidade pôde ser diminuída de três para duas. Nem sempre isto é possível, principalmente quando a dimensionalidade do problema é maior do que três; uma aproximação algébrica é necessária.

Todos os eixos envolvidos no conjunto de dados, devem sofrer rotação, de tal forma que o primeiro novo eixo corresponda à direção de maior variância dos dados e cada eixo sucessivo apresente o máximo de variância residual. Sem alterar a estrutura dos dados, queremos encontrar os eixos ortogonais que representem as direções de máxima variância. Isto é possível quando a matriz de correlação dos dados transformados, é uma matriz diagonal.

Assim, queremos encontrar uma matriz de transformação que, quando aplicada a X, produz um novo conjunto de coordenadas Y, para o qual:

$$Y^T Y = R^T X^T X R = R^T C R = \Lambda, \quad \text{eq.2.24}$$

onde Λ é uma matriz diagonal. Este é um problema de autovetores, no qual desejamos encontrar os vetores em R que, quando aplicados ao sistema, são convertidos em múltiplos de si mesmos:

$$C R = \lambda R, \quad \text{eq. 2.25}$$

onde λ é uma variável escalar, cujas soluções são os elementos diagonais de Λ . Os autovalores λ são as variâncias associadas com cada novo eixo. Resolvendo-se a equação 2.25, os autovetores associados são encontrados e a matriz R obtida.

Agora que temos os autovetores em R, podemos aplicar a rotação sobre os eixos originais e obter as coordenadas dos dados no novo sistema, Y:

$$Y = X R \quad \text{eq. 2.26}$$

As coordenadas são agora os autovetores, os quais são combinações lineares das variáveis originais; a equação acima também pode ser escrita da seguinte forma:

$$X = Y R^T \quad \text{eq. 2.27}$$

As coordenadas Y são os “scores” ($t_1, t_2, \dots$ na equação 2.21) e os elementos dos autovetores (R) são os “loadings” ($p_1, p_2, \dots$ na equação 2.21). Os coeficientes dos autovetores indicam a magnitude da contribuição de uma variável para compor o autovetor.

SIMCA (Self Independent Modeling of Class Analogy) [33]

O método SIMCA constrói um modelo de componentes principais para cada categoria do conjunto de treinamento; esses modelos podem ter números diferentes de componentes principais, conforme a figura 9. O número de componentes principais pode ser determinado pelo método de "Cross Validation".

Uma vez definido o modelo que descreve cada categoria do conjunto de treinamento, o SIMCA pode classificar objetos de categorias desconhecidas (conjunto teste), permitindo ainda, a identificação de amostras que não pertençam às categorias definidas. A classificação do conjunto teste é feita pelo ajuste desses objetos, por regressão múltipla, aos parâmetros dos modelos calculados usando o conjunto de treinamento.

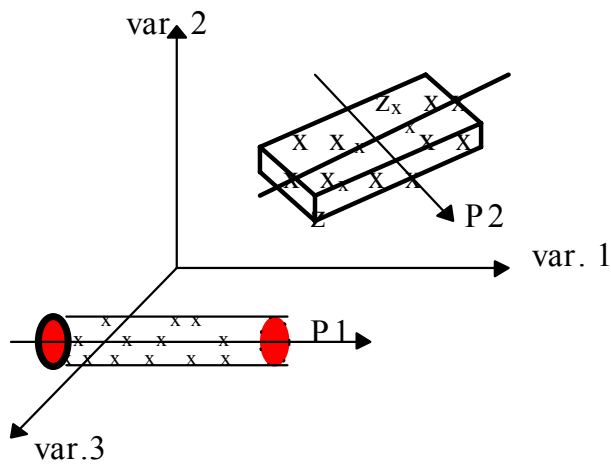


Figura 9 - Representação de modelos com números diferentes de componentes principais, para cada categoria.

Agrupamento hierárquico (HIER) [34]

Neste tipo de análise, os compostos similares são agrupados da seguinte forma: uma matriz de similaridade é obtida e percorrida para o maior valor; os dois pontos produzindo este valor são unidos. Daí em diante, os dois pontos são considerados como único (formando um centro de gravidade) para calcular uma nova e menor matriz de similaridade. Este processo continua até que todos os pontos estejam contidos em um agrupamento.

Existem vários métodos de agrupamento hierárquico e todos são conceitualmente iguais, apresentando apenas variações matemáticas nas definições dos grupos.

As seguintes etapas são seguidas:

- 1) Calcula-se a matriz de distâncias para todos os pontos; no nosso trabalho, foi calculada a distância euclidiana, mostrada anteriormente (eq.2.10).
- 2) Determina-se a menor distância entre os pontos e calcula-se a similaridade entre eles (eq.9); o resultado da similaridade é registrado no dendograma.
- 3) Os dois pontos são então, substituídos por um grupo e a nova matriz de distâncias é calculada. Desta forma, a nova matriz possui uma dimensão a menos do que a matriz inicial.
- 4) As etapas 2 e 3 são repetidas, até que obtenha-se apenas um grupo.

Alguns métodos conhecidos para realizar o agrupamento dos pontos, são: “single link” ou conexão simles, “complete link” ou conexão completa e “centroid” ou conexão por meio de centróides.

O mais conservador deles é o método de **conexão simples**, em que a distância entre o grupo (ou cluster) que acaba de ser formado (por exemplo, grupo AB) e um determinado grupo formado anteriormente (por exemplo, grupo C), é dada pela **menor distância entre os pontos dos dois grupos**:

$$d_{AB-C} = \min(d_{AC}, d_{BC}). \quad \text{eq. 2.28}$$

Podemos também representar a distância calculada pela conexão simples, pela seguinte figura 10:

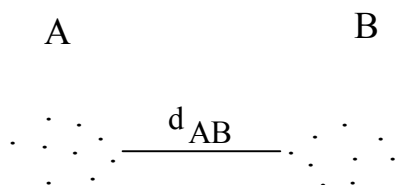


Figura 10 - Representação da distância entre pontos dos grupos A e B, calculada pela conexão simples.

O método de conexão simples foi usado neste trabalho; sendo assim, vamos demonstrar com um exemplo, como o dendograma é obtido:

Supondo que tenhamos seis pontos (moléculas, no nosso caso), de acordo com a metodologia geral para o agrupamento hierárquico, inicialmente, será calculada a matriz de distâncias euclidianas, D_{ij} :

D_{ij}	1	2	3	4	5	6
1	0.0	0.5	0.6	1.55	1.78	2.40
2		0.0	0.62	1.65	1.50	2.30
3			0.0	1.02	1.00	1.58
4				0.0	0.82	1.02
5					0.0	0.68
6						0.0

A menor distância entre pontos é $d_{12} = 0.50$; então, os pontos 1 e 2 serão agrupados, formando o “cluster” A. A similaridade entre eles é calculada, sendo que a maior distância é $d_{16} = 2,40$:

$$S_{ij} = 1 - d_{ij}/(d_{ij})_{\max} = 1 - (0.50/2.40) = 0.792$$

Agora, a nova matriz de distâncias é calculada da seguinte maneira:

Os pontos 1 e 2 são substituídos pelo grupo A; a distância entre todos os outros pontos, permanece a mesma, mas apenas as distâncias entre o grupo A formado e os pontos 3, 4, 5 e 6 devem ser calculadas. É aqui que devemos escolher o método para calcular. No caso da

conexão simples, a distância entre o grupo A (formado pelos pontos 1 e 2) e o ponto 3, por exemplo, será igual à distância entre os pontos 1 e 3 (0,60), a qual é menor do que a distância entre 2 e 3 (0,62).

A nova matriz seria então:

D _{ij}	A	3	4	5	6
A	0.0	0.6	1.55	1.50	2.30
3		0.0	1.02	1.00	1.58
4			0.0	0.82	1.02
5				0.0	0.68
6					0.0

A menor distância é entre os pontos A e 3, e a nova similaridade é calculada e registrada no dendograma: $S_{ij} = 1 - (0,60/2,40) = 0,750$

Agora, o grupo A e o ponto 3 são agrupados, sendo substituídos pelo grupo B e as distâncias entre B e 4, B e 5, B e 6 são calculadas como acima:

D _{ij}	B	4	5	6
B	0,0	1,02	1,00	1,58
4		0,0	0,82	1,02
5			0,0	0,68
6				0,0

A nova menor distância está entre os pontos 5 e 6 = 0,68. A similaridade entre eles é: $S_{56} = 1 - (0,68/2,40) = 0,716$. Os pontos 5 e 6 são unidos, formando o grupo C. A nova matriz é então calculada:

D _{ij}	B	C	4
B	0,0	1,00	1,02
C		0,0	0,82
4			0,0

Os pontos 4 e C são agrupados, formando o grupo D; a similaridade entre eles é: $S_{4C} = 1 - (0,82/2,40) = 0,655$.

D_{ij}	B	D
B	0,0	1,00
D		0,0

O grupo B e D formam um único grupo, incluindo todos os pontos. A similaridade entre B e D é : $1 - 1,00/2,40 = 0,583$.

O dendograma é mostrado a seguir, na figura 11:

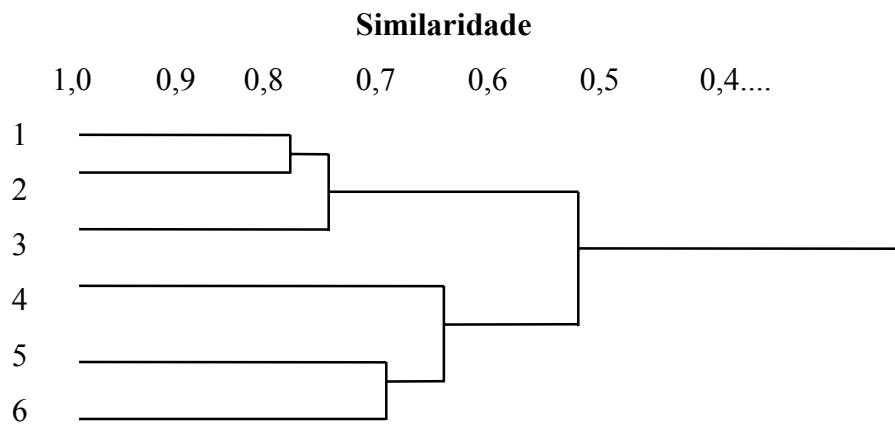


Figura 11 - Representação do dendograma obtido pelo método do agrupamento hierárquico, usando-se a conexão simples.

No caso da conexão completa, a distância entre os grupos é dada pela maior distância entre os pontos de cada grupo (figura 12), ao inverso da conexão simples.

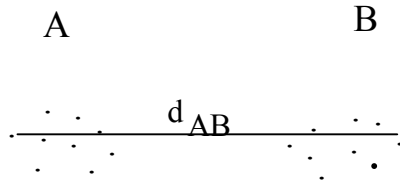


Figura 12 - Distância entre pontos dos grupos A e B, calculada pelo método da Conexão completa.

Na conexão por meio de centróides, é usada a distância de um ponto ao centro de gravidade dos pontos de um outro grupo (figura 13):

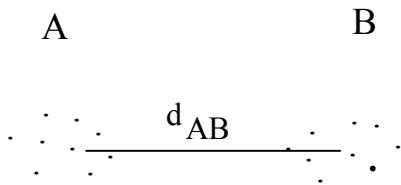


Figura 13 - Distância entre dois pontos dos grupos A e B, calculada pelo método da conexão por meio de centróides.

K Vizinhos Mais Próximos (KNN) [34]

O método do vizinho mais próximo classifica um novo objeto de acordo com a sua distância dos objetos do conjunto treinamento. Os K vizinhos mais próximos no conjunto treinamento são encontrados e o novo objeto (do conjunto teste) é classificado como pertencente à categoria da maioria destes K vizinhos. Este método, assim como o agrupamento hierárquico, também usa a matriz de distâncias euclidianas entre os objetos:

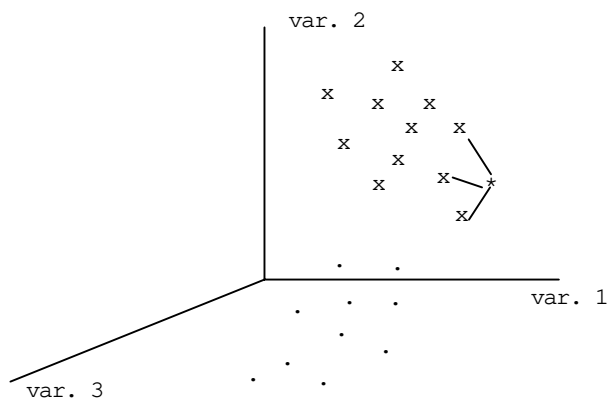


Figura 14 - Representação do método KNN para classificação de objetos de um conjunto teste. Os objetos “x” , na figura, são compostos ativos, “.” os inativos e “*” o composto teste.

Supondo que as moléculas inativas sejam representadas por pontos “.” e as ativas por “x”; o composto “*” seria classificado como ativo pelos 3 vizinhos mais próximos, conforme a figura 14.

O Peso de Fisher [34]

O peso de Fisher estima quantitativamente a utilidade de uma característica para separar duas categorias, conforme a equação 2.29. Utilizamos o peso de Fisher na escolha dos parâmetros FQ para correlacionar com a atividade biológica.

$$w_i^{(q,r)} = \frac{[\bar{x}_i^{(q)} - \bar{x}_i^{(r)}]^2}{[S_i^{(q)}]^2 + [S_i^{(r)}]^2} \quad \text{eq.2.29}$$

Onde: **q e r** são as categorias, “**i**” é a variável e $[S_i^{(q)}]^2$ é a variância da variável “**i**” na categoria q; $x_i^{(q)}$ é o valor da variável “**i**” para a categoria q.

Se $[S_i^{(q)}]^2 + [S_i^{(r)}]^2$ for pequeno e se $[\bar{x}_i^{(q)} - \bar{x}_i^{(r)}]^2$ for grande, o peso da variável “**i**”, $w_i^{(q,r)}$ será grande, isto é, se o desvio da média das medidas de uma variável para as suas categorias for pequeno e se a diferença entre os valores médios da variável “**i**” para as duas categorias for grande, então, o peso será grande.

2.2.2 - Métodos de sobreposição molecular com auxílio computacional (Computer-Assisted Molecular Fitting - CAMF)

Esses métodos foram desenvolvidos para auxiliar análises de relações estrutura-atividade (SAR), através de comparações estruturais de moléculas. Tradicionalmente, tais comparações eram feitas com base em figuras 2D e com modelos moleculares plásticos. Nos últimos dez anos, o avanço no campo de computação gráfica e mecânica molecular, permitiu o desenvolvimento de procedimentos automatizados para comparação de estruturas - os métodos CAMF: “Computer-Assisted Molecular Fitting”.

Nos métodos de sobreposição molecular em 3D, o ajuste é feito pela minimização das distâncias entre pares atômicos pré-selecionados, usando o método dos mínimos quadrados. As seleções dos pares atômicos são baseadas em intuição química ou algum conhecimento prévio sobre o sistema em estudo. Estes métodos são eficientes para sobreposição de moléculas que apresentam partes em comum nas estruturas principais, onde a seleção dos pares atômicos é intuitiva e simples. A desvantagem é a aplicação desses métodos para estruturas que não apresentem muita semelhança, o que dificulta a seleção dos pares atômicos.

Os métodos mais avançados de sobreposição molecular realizam a otimização do emparelhamento de propriedades moleculares intrínsecas entre as duas moléculas, o que inclui análise da forma molecular (“Molecular shape analysis”), distância geométrica e o método do emparelhamento do contorno do potencial eletrostático [35]. Entre esses métodos, o alinhamento estérico e eletrostático (SEAL) é de particular interesse, por ter sido desenvolvido com base em rigorosas considerações teóricas. O emparelhamento das propriedades estéricas e eletrostáticas das moléculas é usado para dirigir o processo de sobreposição. Esse método surgiu em 1992 e desde então, tem sido adaptado e melhorado por muitos grupos de pesquisa na indústria farmacêutica.

A tecnologia PowerFit introduziu o emparelhamento pelo tipo atômico ao SEAL. Isto permitiu discriminar átomos com características estéricas e eletrostáticas muito similares, como átomos de carbono aromático e alifático.

Além do emparelhamento pelo tipo atômico, o PowerFit permite ao usuário especificar o alinhamento entre átomos, pela menor distância. Isto dá ao usuário, a oportunidade de usar a intuição química e conhecimento para influenciar no processo de sobreposição. O PowerFit permite também, realizar a sobreposição com flexibilidade conformacional. Isto é feito através da construção de um grande número de conformações para cada molécula; em seguida, é realizada a sobreposição dessas conformações.

Assim, o potencial de sobreposição molecular é constituído de:

- “Overlap” estérico,
- sobreposição eletrostática,
- sobreposição pelo tipo atômico,
- contração das distâncias e
- energia conformacional (somente para sobreposição flexível).

A superfície de potencial da sobreposição molecular tem muitos mínimos locais. Para encontrar a melhor sobreposição, um método eficiente para a pesquisa global da superfície é utilizado. Este método é chamado “Monte Carlo Simulated Annealing”.

O emparelhamento manual, no nosso caso, foi feito com base nos resultados obtidos anteriormente, através da análise SAR, dos cálculos de solvatação e da comparação com os outros compostos ativos em leishmaniose, conforme está descrito nos capítulos 5, 7 e 8. Esses estudos prévios nos deram condições de realizar o emparelhamento com maior confiança.

No cálculo do potencial de sobreposição molecular, optamos por não utilizar a flexibilidade conformacional, uma vez que tínhamos fortes indicações sobre a conformação ativa. Sendo assim, o nosso potencial consistiu apenas dos quatro primeiros termos: Overlap estérico, sobreposição eletrostática, sobreposição pelo tipo atômico e contração das distâncias.

2.3 – OS MÉTODOS SEMI-EMPÍRICOS AM1 E MNDO-PM3

Os métodos semi-empíricos AM1 (Austin Model 1) e MNDO-PM3 (Modified Neglect of Diatomic Overlap-Parametric Method 3), consistem na segunda e terceira versões do MNDO, respectivamente. O MNDO-PM3 usa o mesmo formalismo do AM1, tendo sido apenas reparametrizado[36]. A diferença entre as equações do MNDO e AM1 está apenas no termo de repulsão entre os caroços (“core”) dos átomos A e B, conforme veremos a seguir. Sendo assim, o entendimento dos métodos AM1 e MNDO-PM3 (ou simplesmente PM3), utilizados neste trabalho, depende basicamente do conhecimento do MNDO.

O método MNDO [37] parte de algumas aproximações feitas pelo NDDO (Neglect of Diatomic Differential Overlap), no qual apenas as integrais de sobreposição entre orbitais atômicos centrados sobre átomos diferentes são anuladas. Além disso, apenas os elétrons de valência são tratados explicitamente, assumindo-se que estes movem-se num campo formado pelos elétrons das camadas internas e pelo núcleo; esta aproximação é chamada aproximação do caroço (“core approximation”);

Os orbitais moleculares OM da camada de valência (ψ_i) são representados pela combinação linear de um conjunto mínimo de bases de orbitais atômicos OA (ϕ_i) da camada de valência (aproximação LCAO):

$$\psi_i = \sum_v C_{vi} \phi_v \quad \text{eq. 2.30}$$

Os coeficientes C_{vi} são encontrados pela resolução das equações de Roothaan. As equações de Roothaan originais, possuem a seguinte forma:

$$\sum_v (F_{\mu v} - E_i S_{\mu v}) C_{vi} = 0 \quad \text{eq. 2.31}$$

Onde E_i é o autovalor do orbital molecular ψ_i , $F_{\mu v}$ são os elementos da matriz de Fock (os quais serão descritos abaixo, para a aproximação NDDO) e $S_{\mu v}$ são os elementos da matriz de recobrimento para os orbitais ϕ_μ e ϕ_v . Na aproximação NDDO, $S_{\mu v}$ é substituída pelo delta de kronecker, $\delta_{\mu v}$, que possui valor zero para $\mu \neq v$ e valor 1 (um) para $\mu = v$. Portanto, a equação de Roothaan assume a seguinte forma:

$$\sum_v (F_{\mu v} - E_i \delta_{\mu v}) C_{vi} = 0 \quad \text{eq. 2.32}$$

Os elementos da matriz de Fock, para o método NNDO são mostrados a seguir:

$$F_{\mu\mu} = U_{\mu\mu} + \sum_B V_{\mu\mu,B} + \sum_{\nu}^A P_{\mu\nu}[(\mu\mu, \nu\nu) - 1/2(\mu\nu, \mu\nu)] + \sum_B \sum_{\lambda, \sigma}^B P_{\lambda, \sigma}(\mu\mu, \lambda\sigma) \quad \text{eq. 2.33}$$

$$F_{\mu\nu} = \sum_B V_{\mu\nu,B} + 1/2 P_{\mu\nu}[3(\mu\nu, \mu\nu) - (\mu\mu, \nu\nu)] + \sum_B \sum_{\lambda, \sigma}^B P_{\lambda, \sigma}(\mu\nu, \lambda\sigma) \quad \text{eq.2.34}$$

$$F_{\mu\lambda} = \beta_{\mu\lambda} - 1/2 \sum_{\nu}^A \sum_{\sigma}^B P_{\nu\sigma}(\mu\nu, \lambda\sigma) \quad \text{eq.2.35}$$

Onde os OA ϕ_{μ} e ϕ_{ν} estão centrados no átomo A e os OA ϕ_{λ} e ϕ_{σ} estão centrados no átomo B e o termo $P_{\mu\nu}$ são os elementos da matriz de densidade.

O termo $U_{\mu\mu}$ representa a soma da energia cinética de um elétron no OA ϕ_{μ} no átomo A e a sua energia potencial devido à atração pelo caroço do próprio átomo A.

O termo $V_{\mu\nu, B}$ representa a atração entre um elétron na distribuição eletrônica $\phi_{\mu}\phi_{\nu}$ do átomo A, e o caroço do átomo B.

A notação $(\mu\mu, \nu\nu)$ representa a integral de repulsão entre dois elétrons no mesmo átomo: $(\mu\mu, \nu\nu) = \int dr_1 dr_2 \phi_{\mu}^*(1)\phi_{\mu}(1) \frac{1}{r_{12}} \phi_{\nu}^*(2)\phi_{\nu}(2); \quad \text{eq.2.36}$

$(\mu\nu, \mu\nu)$ é a integral de troca e $(\mu\nu, \lambda\sigma)$ é a integral de repulsão de dois centros.

$\beta_{\mu\lambda}$ são as integrais de ressonância de um elétron e dois centros, isto é, a atração de um elétron pelos caroços A e B.

Na aproximação MNDO, as integrais da matriz de Fock não são calculadas; ao invés disso, os termos são determinados experimentalmente ou de expressões semi-empíricas, as quais contém parâmetros que são ajustados de acordo com dados experimentais.

A energia total da molécula (E_t) é a soma da energia eletrônica E_{el} e das repulsões entre os caroços dos átomos A e B (E_{AB}):

$$E_t = E_{el} + \sum_{A < B} \sum E_{AB}^{caroço} \quad \text{eq.2.37}$$

A energia eletrônica é obtida pelos vários termos da matriz de Fock e E_{AB} , no MNDO é dada pela equação:

$$E_{AB}^{caroço} = Z_A Z_B (S^A S^A, S^B S^B) + f(R_{AB}) \quad \text{eq.2.38}$$

Onde Z_A e Z_B são as cargas sobre os caroços A e B e $S^A S^A$ é o caroço do átomo A, representado pela distribuição de carga num orbital do tipo S e a função $f(R_{AB})$ inclui um termo exponencial de forma $\exp(-\alpha R_{AB})$, com α ajustável, tal que a repulsão entre átomos neutros se anule no limite $R_{AB} \rightarrow \infty$.

O método MNDO apresentou alguns erros, como a tendência em superestimar repulsões. Dewar e colaboradores publicaram em 1985 [38], uma versão melhorada do MNDO, chamada AM1. Como já dissemos no início deste tópico, a diferença entre MNDO e AM1 está no termo de repulsão entre os caroços dos átomos A e B, que em AM1 é dado por:

$$\begin{aligned}
 E_{AB}^{\text{caroço}} = & Z_A Z_B (S^A S^A, S^B S^B) + f(R_{AB}) + \\
 & + Z_A Z_B (S^A S^A, S^B S^B) \left\{ \sum_i K_{Ai} \exp[L_{Ai} (R_{AB} - M_{Ai})^2] + \right. \\
 & \left. + \sum_j K_{Bj} \exp[L_{Bj} (R_{AB} - M_{Bj})^2] \right\}
 \end{aligned}
 \tag{eq. 2.39}$$

Onde apenas o terceiro termo refere-se à modificação feita pelo AM1. Os parâmetros M, K e L são dados na referência 57, sendo que K e M são ajustáveis.

CAPÍTULO 3

PARÂMETROS FÍSICO-QUÍMICOS

3.1 - COEFICIENTE DE PARTIÇÃO: P

Por volta de 1890, **Richet** [39] mostrou que a toxicidade numa série de compostos simples como álcool, éteres e cetonas é inversamente proporcional à solubilidade em água; a primeira relação entre lipofilicidade e toxicidade foi, então, proposta por Richet.

A lipofilicidade de uma droga é definida pelo **coeficiente de partição P**, o qual é a razão entre as concentrações de equilíbrio da droga, em uma fase orgânica C_{org} e na fase aquosa, C_{aq} , segundo a equação 3.1:

$$P = C_{org} / C_{aq} \quad \text{eq.3.1}$$

Octanol/água é o sistema mais escolhido para medidas de coeficiente de partição em estudos QSAR, devido a sua similaridade a sistemas biológicos: as membranas biológicas apresentam, como o octanol, cadeias alquil (hidrofóbicas) e grupos polares. Além disso, o octanol possui algumas vantagens práticas sobre os outros solventes orgânicos, como o fato de grande variedade de compostos ser solúvel em octanol, o que não ocorre com hidrocarbonetos aromáticos e alifáticos e do octanol apresentar baixa pressão de vapor à temperatura ambiente. Além dessas vantagens, os coeficientes de partição octanol/água são encontrados na literatura para um grande número de drogas [40].

Simultaneamente a Richet, **Overton** explorou a permeabilidade de células animais e de plantas, frente a uma variedade de compostos orgânicos; os estudos de Overton consistem, em grande parte, em explicar a potência de substâncias narcóticas através da correlação com os coeficientes de partição óleo/água e óleo/gás [39].

Entre 1937-1940, **Hammett** desenvolveu seu sistema de constantes σ que descreve os **efeitos eletrônicos** de substituintes do anel benzênico. **Taft** deu sequência aos estudos de Hammett, desenvolvendo um conjunto de valores, σ^* , para a descrição do efeito eletrônico causado por substituintes em estruturas alifáticas. Alguns anos antes, **Taft** havia

obtido os parâmetros E_s para descrever o **efeito estérico**. Com as informações numéricas sobre efeitos eletrônicos e estéricos e, tendo em mãos os princípios básicos da lipofilicidade, **Hansch** e colaboradores iniciaram sua contribuição em 1962, sendo que as mais importantes foram o desenvolvimento da constante de **hidrofobicidade do substituinte**, π (equação 3.2) e o efeito dos substituintes nas propriedades biológicas de uma molécula, pela alteração de algumas das propriedades físico-químicas descritas pelos parâmetros, $\log P$, σ e E_s . O parâmetro $\pi(X)$, na eq. 3.2 é definido como a contribuição do substituinte X na lipofilicidade de uma estrutura R, quando um hidrogênio H é trocado pelo substituinte X [39]:

$$\log P (R-X) = \log P (R-H) + \pi(X) \quad \text{eq.3.2}$$

Hansch propôs então a equação 3.3:

$$\log 1/C = a \log P + b E_s + c \sigma + d \quad \text{eq.3.3}$$

onde $1/C$ representa a potência ou atividade biológica.

De acordo com a equação 3.2, fica óbvio que π_H deve ser zero. Devido à falta de medidas de $\log P$ para compostos de baixo peso molecular, por muitos anos pensou-se que π_H apresentasse valor zero, mas **Davis** (1973) foi o primeiro a sugerir que isso não é verdade e logo depois, **Nys e Rekker** (1973) publicaram os resultados de uma análise de regressão para um grande conjunto de valores de $\log P$. Por esta análise, eles estabeleceram a hidrofobicidade de H, CH, CH₂ e CH₃ e de substituintes em átomos de carbono saturado. Eles definiram então, um novo parâmetro, **f**, o qual é a constante de **hidrofobicidade do fragmento** [33, 35]

$$\log P = \sum_i a_i f_i \quad \text{eq.3.4}$$

Enquanto os valores π são contribuições à lipofilicidade relativas à substituição do hidrogênio, os valores **f** são contribuições absolutas dos substituintes à lipofilicidade total;

a_i indica a frequência com que um dado fragmento ocorre na estrutura. Os parâmetros f podem ser aplicados para compostos aromáticos ou alifáticos. π é mais conveniente quando aplicado a compostos em que um único grupo é trocado. Para trocas mais complexas, o sistema f é mais simples.

A aplicação correta dos valores f de Rekker, requer a consideração de certas correções [40]:

1) Efeito de proximidade: se dois grupos eletronegativos como -OH, -O-, -COOH ou NH₂ são separados por somente um ou dois carbonos saturados, a lipofilicidade total do composto será maior do que a predita; fatores de correção de 0,861 e 0,574 respectivamente, deverão ser adicionados.

2) Hidrogênio ligado a um grupo eletronegativo: um átomo de hidrogênio ligado a -COOH, -COOR, -COR, -CONH₂, etc, aumenta o valor de f , de 0,462 ao invés do valor normal de f_H que seria 0,175.

3) Sistemas conjugados: Se dois anéis aromáticos são conjugados (ex: bifenil) ou "cross-conjugated" (ex: benzofenona), um fator de correção de 0,28 deve ser adicionado.

4) Sistemas aromáticos condensados como naftaleno requerem um fator de correção de 0,31 para cada par de átomos de carbono comum a dois anéis aromáticos.

Para completar o sistema f , foi adicionado um fator de correção na equação 3.5 que opera como um múltiplo (kn) de uma constante C_M [39].

$$\log P = \sum f + Kn.C_M \quad \text{eq.3.5}$$

Essa constante é necessária para corrigir valores de $\log P$ calculados pelo sistema f de Rekker, em relação ao $\log P$ experimental.

Uma versão revisada do sistema de Rekker apresenta regras para a aplicação do fator de correção; esse valor é agora fixado em 0,219 [42].

Leo e colaboradores também desenvolveram uma metodologia baseada no somatório de fragmentos, para calcular $\log P$, mas esta possui um fundamento teórico diferente do método de Rekker; Leo e colaboradores mediram valores de $\log P$ para várias

substâncias de baixo peso molecular incluindo hidrogênio, metano e etano. Com estas informações e outros valores de logP já conhecidos, eles calcularam seus próprios valores f para muitos substituintes. O cálculo de logP é feito pela combinação destas constantes f com outros fatores de correção, como correção para ramificação e ligação [42].

Atualmente, existem vários programas comerciais para o cálculo do logP, através do método do somatório de fragmentos; podemos citar CLOGP e PCMODELS que são baseados no método de Hansch/Leo e Σf , PROLOGP-cdr, Σf -SYBYL e SANALOGP-Er, baseados no método de Rekker [43].

Neste trabalho, usamos valores das constantes f de Rekker para obtenção dos valores de logP, no qual procuramos considerar a correção para o efeito de proximidade.

3.2 - VOLUME MOLECULAR

Craig observou que o coeficiente de partição está correlacionado com o parâmetro estérico (Es) e também com o volume molecular; o volume molecular permite avaliar o efeito estérico causado pela troca de substituintes, de maneira mais acurada do que os parâmetros de Verloop [41] fariam, já que estes estão relacionados com o tamanho do substituinte apenas, enquanto o volume avalia modificações de tamanho ocorridas na molécula toda, devido a possíveis alterações conformacionais que o substituinte pode causar.

O algoritmo utilizado para estimar o volume foi proposto por Higo & Go [44] e o procedimento do cálculo é explicado abreviadamente, a seguir.

Considera-se, inicialmente, que a molécula está contida numa caixa, com arestas de comprimento múltiplo de um valor "a". A seguir, cada aresta é dividida em unidades de "a", de forma que a caixa transforme-se num conjunto de cubos de aresta com valor "a". Cada cubo é então, classificado em três tipos: interno, externo ou superficial à molécula. Se o cubo for externo, será desprezado. Se for interno, passará a fazer parte do volume molecular. Se for superficial, será subdividido em novos cubos de aresta com valor "a". Os novos cubos serão classificados e processados da mesma forma que os primeiros, e o procedimento repetir-se-á até o nível estabelecido pelo operador. Finalmente, determinada fração do volume será computada como pertencente à molécula.

3.3 - ÂNGULOS DIEDROS

Outros parâmetros estéricos utilizados foram alguns ângulos diedros referentes à estrutura básica das neolignanas; estes ângulos variam com a troca dos substituintes, para um mesmo mínimo local; por exemplo, para o mínimo local D, o ângulo diedro referente aos átomos 14-13-12-6 varia em torno de 10 graus quando um substituinte mais volumoso é introduzido nas posições meta ou para do anel (2) mostrado na figura 4a do capítulo 1. Observe, por exemplo, as moléculas inativas 1, 2, 3, 4 e 9, que só diferem entre si pelos substituintes no anel (2) - tabela 2 do capítulo 1; as moléculas 1, 2 e 4, que não são substituídas nas posições meta (R_6) ou para (R_7) ou que possuem substituintes pouco volumosos, apresentam o ângulo diedro 14-13-12-6, para a conformação D, em torno de 50 graus, enquanto as moléculas 3 (substituída em para (R_7) por um $-OCH_3$) e 9 (substituída em meta (R_6) por um $-CH_3$), apresentam um aumento nesse ângulo de 10 graus. Os valores do ângulo diedro discutido, para a conformação D, podem ser verificados na figura 38 do capítulo 9.

3.4 CARGAS ATÔMICAS

As **propriedades eletrônicas** de maior interesse pertencem a três tipos [41]: o primeiro, refere-se à influência que o substituinte produz na reatividade da porção invariante da molécula; a constante σ ou cálculos de orbitais moleculares são usados neste caso; o segundo tipo é o pK_a , necessário para o cálculo da fração ionizada em um determinado pH; finalmente, as energias dos orbitais moleculares fornecem informações sobre as propriedades de oxidação-redução ou sobre a reatividade dos compostos, ao longo de uma série.

As cargas atômicas são utilizadas em estudos de correlações entre a estrutura e atividade biológica [45], sendo um parâmetro do primeiro tipo apresentado, pois mostra a influência que o substituinte produz na estrutura básica da molécula.

Cargas atômicas não são valores esperados da função de onda, não podendo assim serem calculadas diretamente a partir da resolução dessa função, mas podem ser obtidas se

conhecemos a densidade eletrônica. As cargas atômicas derivadas do potencial eletrostático vêm sendo utilizadas com maior frequência ultimamente, em relação às cargas de Mülliken, por ter o primeiro método, superado alguns problemas apresentados pela análise populacional de Mülliken, como o da divisão da população eletrônica de ligação, igualmente entre os dois átomos que formam essa ligação [46].

As cargas derivadas do potencial eletrostático calculadas neste trabalho (usando a base 6-31G**, com a opção CHELPG do programa GAUSSIAN) estão representando adequadamente o efeito dos substituintes nos anéis aromáticos. Por exemplo, o substituinte -OCH₃ é retirador de elétrons pelo efeito de indução, colocando carga + no carbono em que está ligado e é doador de elétrons pelo efeito de ressonância, colocando cargas negativas nas posições orto e para do anel [47]. Observamos que as cargas derivadas do potencial eletrostático mostram melhor esse efeito, do que as de Mülliken. A tabela 3 apresenta esses valores para os compostos ativos.

Na tabela 3, o composto 12 possui um substituinte -OCH₃ na posição 21 (veja a tabela 2 e figura 4a no capítulo 1). A carga de Mülliken na posição 21, marcada com um asterisco, é +0,06, enquanto a carga derivada do potencial eletrostático é +0,39. O substituinte -OCH₃ na posição 21 deve doar elétrons pelo efeito de ressonância, como dissemos, resultando cargas negativas nas posições orto e para. As posições 20 e 22, marcadas com 2 asteriscos, estão em orto à posição 21. As cargas de Mülliken nessas posições são: -0,14 e -0,09, enquanto as derivadas do potencial eletrostático são: -0,21 e -0,26. Verificamos assim, que as cargas derivadas do potencial eletrostático são mais representativas do que as de Mülliken. O mesmo acontece com a molécula 13, na tabela 3, que possui -OCH₃ nas posições 19 e 23 (assinaladas com asteriscos). Observe as posições orto também dessa molécula, assinaladas com 2 asteriscos. O mesmo ocorre para as demais moléculas; as posições assinaladas com um asterisco devem possuir carga positiva e, com dois asteriscos, cargas negativas. Em alguns casos, as cargas de Mülliken apresentam sinal invertido; veja as moléculas 17, 19 e 20, por exemplo.

Tabela 3 - Comparação entre as cargas de Mülliken e derivadas do potencial eletrostático (PE), calculadas para as neolignanas ativas (compostos 12 a 20 - tabela 2). Os átomos C18, C19,...etc, são mostrados na figura 4a. As posições assinaladas com um asterisco (*) devem apresentar carga positiva, devido ao efeito do substituinte, enquanto aquelas com dois asteriscos (**) devem apresentar carga negativa.

COMPOSTO	ÁTOMO	CARGAS DE MÜLLIKEN (ev)	PE
12	C18	+0,06	+0,37
	C19	-0,10	-0,24
	C20 **	-0,14	-0,21
	C21 *	+0,06	+0,39
	C22 **	-0,09	-0,26
	C23	-0,14	-0,20
13	C18	+0,05	+0,10
	C19 *	+0,09	+0,33
	C20 **	-0,18	-0,26
	C21	-0,07	-0,12
	C22 **	-0,18	-0,28
	C23 *	+0,06	+0,25
14	C18	+0,06	+0,17
	C19 *	+0,07	+0,35
	C20 **	-0,14	-0,38
	C21	-0,02	+0,27
	C22	-0,10	-0,31
	C23	-0,15	-0,15

Continuação da tabela 3.

15	C18	+0,06	+0,04
	C19 *	+0,08	+0,38
	C20 **	-0,17	-0,39
	C21	0,00	+0,24
	C22 **	-0,18	-0,39
	C23 *	+0,06	+0,30
16	C18	+0,06	+0,01
	C19 *	+0,08	+0,37
	C20 **	-0,17	-0,39
	C21	0,00	+0,25
	C22 **	-0,18	-0,40
	C23 *	+0,06	+0,29
17	C18	-0,22	+0,30
	C19	-0,03	-0,13
	C20	-0,12	-0,13
	C21 *	-0,09	+0,09
	C22	-0,12	-0,05
	C23	-0,05	-0,29
18	C18	-0,22	+0,19
	C19	-0,04	-0,03
	C20	-0,12	-0,30
	C21	-0,04	+0,23
	C22	-0,12	-0,26
	C23	-0,05	-0,13
19	C18	+0,10	+0,51
	C19	-0,13	-0,29
	C20	-0,06	-0,02
	C21 *	-0,15	+0,05
	C22	-0,07	-0,03
	C23	-0,18	-0,34

Continuação da tabela 3.

20	C18	-0,22	+0,25
	C19	-0,03	-0,10
	C20	-0,12	-0,15
	C21 *	-0,09	+0,11
	C22	-0,12	-0,09
	C23	-0,05	-0,20

3.5 - ENERGIAS DOS ORBITAIS DE FRONTEIRA

As energias do HOMO e do LUMO são usadas em QSAR para estimar propriedades de oxidação-redução [48] pois, essas energias correspondem às energias necessárias para remover um elétron da molécula e para adicionar, respectivamente. Assim, a energia do HOMO é proporcional ao potencial de ionização e a do LUMO, à afinidade eletrônica da molécula.

Os orbitais HOMO (orbital molecular mais alto ocupado) e LUMO (orbital molecular mais baixo desocupado) foram chamados orbitais de fronteira [49]; eles são importantes numa reação química, pois é a interação desses dois orbitais que leva a um maior abaixamento de energia, devido à proximidade entre eles e porque, além disso, as combinações entre orbitais ocupados e desocupados também produzem um importante efeito de abaixamento de energia.

Por exemplo, quando duas moléculas se aproximam, numa reação, seus orbitais interagem formando novos orbitais moleculares ligantes e anti-ligantes. A formação dos orbitais ligantes, normalmente é exotérmica, E_1 enquanto a formação dos orbitais antiligantes é endotérmica, E_2 , conforme a figura 15.

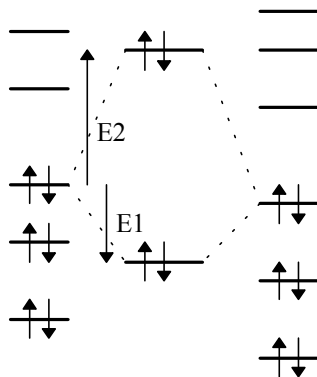


Figura 15 - Interação do HOMO de uma molécula com o HOMO de outra.

No caso das combinações de orbitais ocupados com desocupados, não haverá elétrons para ocupar o orbital antiligante, de energia mais alta (figura 16) [50].

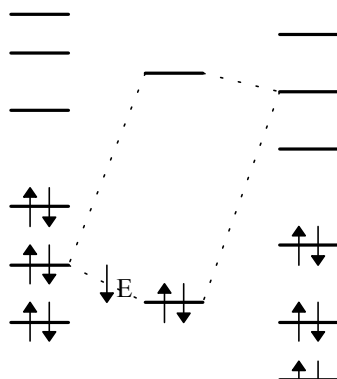


Figura 16 - Interação de um orbital ocupado de uma molécula com um desocupado de outra.

Entretanto, é a interação do HOMO de uma molécula com o LUMO de outra que leva ao maior abaixamento de energia, devido à proximidade entre os dois orbitais, pois, o abaixamento de energia é maior para interação entre orbitais com energias similares do que para interação entre orbitais com energias muito diferentes (figura 17) [50].

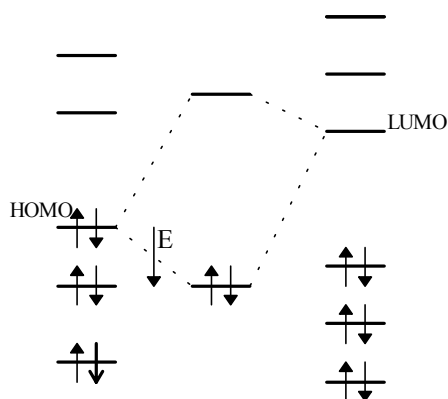


Figura 17 - Interação do HOMO de uma molécula com o LUMO de outra.

A dureza e moleza moleculares também são parâmetros usados em QSAR e correspondem, segundo a teoria de orbital molecular, à facilidade com que os elétrons são rearranjados após uma perturbação. A molécula dura é aquela que possui grande separação entre os orbitais de fronteira e a distribuição eletrônica resiste ao rearranjo de uma perturbação; na molécula "mole", a separação dos orbitais de fronteira é pequena e a distribuição dos elétrons é facilmente rearranjada.

A dureza molecular é expressa pela diferença entre a energia de ionização do átomo neutro e a energia do seu ânion [51]:

$$\eta = \frac{1}{2}(I - A_e) \quad \text{eq.3.6}$$

Onde I é o potencial de ionização e A_e é a afinidade eletrônica da molécula. Este valor é a metade da separação entre os dois orbitais de fronteira. A moleza é expressa pelo inverso da dureza:

$$\frac{1}{\eta} = \frac{2}{\{(I - A_e)\}} \quad \text{eq.3.7}$$

3.5.1 - Índices de Fronteira [52]

Além das energias dos orbitais de fronteira e da dureza e moleza moleculares, outros índices de fronteira utilizados para correlacionar com a atividade biológica, são a densidade do elétron de fronteira (def), a densidade do orbital de fronteira (dof) e a densidade do radical de fronteira (drf).

Densidade do elétron de fronteira:

A densidade do elétron de fronteira é uma medida da reatividade ao ataque eletrofílico na posição r, de acordo com a equação 3.8:

$$def = 2\sum (C_r^{HOMO})^2 \quad \text{eq. 3.8}$$

onde C_r^{HOMO} são os coeficientes de expansão obtidos de cálculos de orbitais moleculares.

Densidade do orbital de fronteira:

É expressa pela equação 3.9, a qual é uma medida da reatividade ao ataque nucleofílico na posição r.

$$def = 2\sum (C_r^{LUMO})^2 \quad \text{eq.3.9}$$

Densidade do radical de fronteira:

É uma medida da reatividade do átomo r frente a um radical. É expressa pela equação 3.10:

$$def = 2\sum (C_r^{HOMO})^2 + 2\sum (C_r^{LUMO})^2 \quad \text{eq.3.10}$$

3.6 - MOMENTOS DIPOLARES

Momento de dipolo, μ , é um vetor, o qual é a medida da magnitude da carga deslocada quando átomos de eletronegatividades diferentes são interligados, e é definido por:

$$\mu = e d_{vec} \quad \text{eq.3.11}$$

onde d_{vec} é o vetor distância que separa as cargas positivas e negativas de magnitude e.

A direção do momento de dipolo de uma molécula é baseada nas eletronegatividades relativas dos átomos componentes e o valor é obtido pelo vetor resultante dos momentos de dipolo de cada ligação presente na molécula.

Um dipolo, numa molécula, pode afetar um átomo não ligado diretamente, através de interações eletrostáticas por um efeito indutivo. Por exemplo, devido à diferença de eletronegatividade entre o C e o Cl, o Cl agirá como um retirador de elétrons, induzindo uma carga positiva no carbono adjacente. Este, por sua vez, tentará "acomodar" sua carga positiva induzida sobre todos os átomos a ele ligados, criando um efeito em todas as ligações sucessivas na molécula:



A formação de dipolos numa molécula, produzida por substituintes de diferentes eletronegatividades, altera propriedades moleculares como a acidez e a basicidade do composto, de maneira que o momento de dipolo pode dar alguma indicação sobre a reatividade do mesmo [53]; assim, momentos dipolares são usados também para correlacionar estrutura química com atividade biológica.[54, 55].

No MOPAC 6.0, o momento dipolar total é dado pela soma do momento de carga pontual (μ_p) (discutido acima) e do momento dipolar de hibridização (μ_h):

$$\mu = \mu_p + \mu_h \quad \text{eq. 3.12}$$

O momento de dipolo de carga pontual ocorre devido à distribuição das cargas pontuais, isto é, das cargas atômicas totais centradas nos respectivos núcleos e é dado pelo somatório do produto entre cada carga atômica (C_n) e o seu vetor posição (R_n):

$$\mu_p = -\sum_n R_n C_n \quad \text{eq.3.13}$$

O momento dipolar de hibridização ocorre no nível atômico, devido à hibridização de orbitais. Por exemplo, os orbitais s e p possuem o centro da carga eletrônica coincidente com o núcleo atômico, devido à simetria desses orbitais, mas se houver hibridização entre os orbitais s e p, o centro da carga eletrônica do orbital híbrido não ficará sobre o núcleo, resultando um momento de dipolo no interior do átomo.

Na molécula de água, por exemplo, teríamos as seguintes contribuições para o momento dipolar total: **o momento dipolar pontual**, devido às cargas positivas sobre os hidrogênios e negativa sobre o oxigênio (o qual pode ser decomposto nas direções x, y e z

do vetor resultante) e o **momento dipolar híbrido**, devido à hibridização dos orbitais s e p do oxigênio.

O momento dipolar total calculado pelo pacote computacional MOPAC 6.0 é a soma dessas contribuições e, assim como as parcelas x, y e z do momento dipolar pontual, estas também podem estar melhor correlacionadas com a atividade biológica, do que o momento dipolar total.

3.7 – ELETRONEGATIVIDADE DE MÜLLIKEN (χ)

A eletronegatividade pode ser entendida como o “poder” de um átomo atrair elétrons para si, numa molécula, segundo Pauling.

Mülliken propôs que, dadas duas espécies “a” e “b”, a energia necessária para “a” retirar um elétron de “b” é $I_b - A_a$, onde I e A são o potencial de ionização e a afinidade eletrônica, respectivamente. Da mesma forma, a energia necessária para “b” retirar um elétron de “a” é $I_a - A_b$. Quando a e b apresentam a mesma eletronegatividade,

$$I_a - A_b = I_b - A_a \rightarrow I_a + A_a = I_b + A_b,$$

o que sugere que podemos definir a eletronegatividade da seguinte forma: $\chi = \frac{1}{2} (I + A)$, onde $\frac{1}{2}$ é um fator arbitrário.

Usamos esse conceito, para estimar a eletronegatividade de moléculas, onde I e A são o potencial de ionização e a afinidade eletrônica da molécula: I_m e A_m .

A quantidade $\chi_m = \frac{1}{2} (I_m + A_m)$ daria a capacidade de uma molécula M1 (ex: fármaco) atrair elétrons de uma outra molécula M2 (ex: receptor), quando M1 e M2 se aproximam.

CAPÍTULO 4

CÁLCULOS

4.1 - NEOLIGNANAS ISOLADAS: ANÁLISE CONFORMACIONAL, PARÂMETROS FÍSICO-QUÍMICOS E RELAÇÕES ESTRUTURA-ATIVIDADE.

4.1.1 - Análise conformacional

A análise conformacional para os 20 compostos da tabela 2 e para as 12 neolignanas do conjunto teste (tabela 21) foi realizada inicialmente para os compostos isolados, com o programa MM2 (campo de força de 1987), acrescido de alguns parâmetros obtidos com a versão MM2' [28]. Os compostos numerados de 1 a 25 já haviam sido analisados, sob esse aspecto, em trabalhos anteriores [5,6], nos quais foram encontrados um máximo de 9 mínimos locais para cada molécula; esses mínimos foram denominados de A a I, de acordo com os valores dos ângulos diedros t1 e t2 encontrados (figura 4a). Assim, o mínimo A apresenta t1 e t2 próximos a 60 graus, para todas as moléculas; B apresenta t1 próximo a 150 graus e t2 próximo a 60 graus, etc. Em algumas moléculas, algumas dessas conformações são enantiômeros de outras, de forma a apresentar um número menor do que 9 para o total de mínimos encontrados. Outras, apesar de não apresentarem enantiômeros, também não possuem todas as 9 conformações. Para as moléculas ativas, por exemplo, apesar de algumas apresentarem os 9 mínimos, outras apresentam um número menor, de forma que apenas 4 conformações são comuns a todas elas: C, D, E e F; a figura 29 (capítulo 7) mostra todas as conformações encontradas para as neolignanas.

Empregando um método estatístico baseado na análise populacional de Boltzmann e com a ajuda de mapas conformacionais, pudemos encontrar os mínimos locais mais prováveis para cada molécula e estes foram utilizados para a realização de cálculos de parâmetros físico-químicos. Esses parâmetros foram empregados para realizar o nosso primeiro estudo de relações estrutura-atividade [1, 6].

Os compostos numerados entre 26 e 32 foram analisados posteriormente e, por serem análogos aos 25 primeiros, empregamos a mesma metodologia.

4.1.2 - Parâmetros físico-químicos para as neolignanas isoladas.

Propriedades hidrofóbicas, eletrônicas e estéricas foram calculadas inicialmente, para as 20 neolignanas do conjunto treinamento e para as 12 do conjunto teste, na conformação E, que apresentou alta probabilidade e é comum a todas as moléculas ativas, o que nos levou a pensar que esta pudesse ser a conformação ativa [5,6].

O **log P**, foi calculado usando-se o método de Rekker, pelo somatório dos parâmetros de hidrofobicidade dos substituintes. O somatório dos fragmentos foi realizado, utilizando valores obtidos na referência 56, aos quais somamos o fator de correção de 0,574 devido ao efeito de proximidade presente em todas as neolignanas, entre os grupos =O e -O- (tabela 2 e figura 4a do capítulo 1).

O **volume de van der Waals (V_{vdw})** foi calculado com o programa surf [57], baseado no algoritmo de Higo e Go [44]. Os **ângulos torsionais t1 e t2** (figura 4a), foram extraídos da análise conformacional realizada com o método MM2. Finalmente, parâmetros eletrônicos como **dureza e moleza** molecular, energias do **HOMO e LUMO**, **momento dipolar**, e **eletronegatividade de Mülliken** foram calculados com o método **semi-empírico MNDO-PM3** [36], usando-se o programa MOPAC 6.0, uma vez que estes são parâmetros originados pelo cálculo de orbitais moleculares e, portanto, não são obtidos de cálculos de mecânica molecular. Nesta etapa do trabalho, foram calculadas as **cargas de Mülliken**, também com o método MNDO-PM3.

Para realizar os cálculos dos parâmetros FQ com o método MNDO-PM3, entramos com os dados de geometria obtidos com a análise conformacional previamente realizada com o método MM2. Assim, cada coordenada foi seguida do inteiro 0, indicando que as mesmas não deveriam ser otimizadas e apenas o cálculo de orbitais moleculares foi realizado para as geometrias pré-estabelecidas.

4.1.3 - Relações estrutura-atividade para as neolignanas isoladas.

Nesta etapa, utilizamos os métodos de Reconhecimento de Padrões, através dos programas Arthur/Unicamp [58] e Pirouette (fornecido pela prof^a Dr^a Marcia M.C. Ferreira, do instituto de química da Unicamp) [59]. O Pirouette realiza a análise de componentes principais e agrupamento hierárquico, como o Arthur/Unicamp, mas é um programa de melhor visualização dos resultados e de mais fácil manipulação dos dados. O Arthur/Unicamp, entretanto, realiza cálculos adicionais, como o Peso de Fisher, KNN e SIMCA, com os quais trabalhamos nesta tese.

4.2 ANÁLISE CONFORMACIONAL DE OUTROS COMPOSTOS ATIVOS EM LEISHMANIOSE (AMINO-CETONAS) E DE UM POSSÍVEL RECEPTOR (ARGININA).

A conformação da benzofenona, que é uma parte comum às aminocetonas estudadas, foi determinada por P. Jin e T. A. Wildeman [60]. Cálculos realizados com AM1 [38] mostraram que a conformação de mínima energia da benzofenona apresenta um ângulo de 33 graus entre os dois anéis fenílicos, concordando com dados de cristalografia [60]. Com base nesses resultados, nós realizamos a análise conformacional das aminocetonas usando o método AM1 também, através do pacote computacional MOPAC 6.0. Foi usada a palavra-chave PRECISE, como um critério de convergência do cálculo SCF; assim, o cálculo é interrompido quando a diferença na energia eletrônica entre duas iterações consecutivas for menor do que 10^{-6} kcal/mol.

Para os substituintes nos anéis fenílicos das aminocetonas, foi realizada a procura conformacional, utilizando “coordenada de reação”. No MOPAC, cada coordenada interna é seguida por um inteiro, indicando a ação a ser realizada sobre aquela coordenada. Assim, o inteiro 1 significa que a coordenada deve ser otimizada, 0 que não deve ser otimizada e -1 indica que deve ser aplicada a coordenada de reação, isto é, aquela coordenada seguida de -1 deve ser otimizada a cada passo estabelecido na última linha dos dados de entrada. No nosso caso, otimizamos as coordenadas referentes aos substituintes, a cada 20°.

A conformação mais estável das aminocetonas foi utilizada para comparação com os mínimos locais encontrados para as neolignanas, a fim de nos auxiliar na busca da conformação ativa.

Dados de cristalografia foram encontrados na literatura para a geometria dos átomos pesados da arginina [61]. As posições dos átomos de hidrogênio foram determinadas através da otimização da geometria da arginina, usando o método semi-empírico AM1, para uniformizar com o método utilizado na otimização da geometria das aminocetonas.

4.3 - NEOLIGNANAS SOLVATADAS: ANÁLISE CONFORMACIONAL, CÁLCULO DE PARÂMETROS FÍSICO-QUÍMICOS E RELAÇÕES ESTRUTURA-ATIVIDADE.

4.3.1 - Solvatação

Estes cálculos foram necessários, para aprimorarmos a nossa análise conformacional e assim, podermos estabelecer novas relações estrutura-atividade com maior eficiência. Entretanto, esta etapa não seria possível sem antes termos obtido os mínimos locais para os compostos isolados, como veremos a seguir.

Nos cálculos de solvatação, calculamos as energias de interação soluto-solvente, utilizando o método de Monte Carlo, através do programa computacional Diadorim [62]. Estes cálculos foram efetuados para todas as conformações encontradas (com o método de mecânica molecular - MM2) para as neolignanas 1, 3, 4, 5, 6, 7, 9, 12, 13, 14, 15 e 16 do conjunto treinamento da tabela 2; estes compostos foram escolhidos por termos disponíveis os parâmetros de Lennard-Jones; como a análise destas doze moléculas foi suficiente para obtermos indicações sobre a conformação ativa, achamos desnecessário continuar com estes cálculos para as oito restantes. O solvente escolhido foi a água, à temperatura biológica (36,5 graus centígrados) e pressão = 1 atm. As cargas derivadas do potencial eletrostático, incluídas nos dados de entrada do programa Diadorim, foram calculadas com o programa Gaussian, usando a opção CHELPG e base 6-31G**, conforme indicações da literatura [24].

Achamos importante, neste ponto, esclarecer que a análise conformacional para as neolignanas solvatadas deu origem a um **novo estudo** de relações estrutura-atividade

(capítulo 9). Portanto, não houve utilização de parâmetros calculados com métodos semi-empíricos e *ab initio* numa mesma análise; ambos tipos de parâmetros foram utilizados, porém em análises diversas.

Inicialmente, uma caixa cúbica contendo 800 moléculas de água nos foi fornecida pelo prof. Dr. Luiz C. G. Freitas da UFSCar; no centro dessa caixa, abrimos um “buraco”, no qual 1 molécula do soluto foi introduzida. O soluto, no nosso caso é uma neolignana, cuja conformação deve ser um dos mínimos locais previamente encontrados numa análise conformacional realizada para os compostos isolados. Devido ao tempo computacional necessário, não seria possível realizar a procura dos mínimos locais para os compostos solvatados. Após a introdução do mínimo local escolhido, a caixa de água contendo o soluto foi equilibrada. Para equilibrar a caixa, são geradas configurações pelo movimento aleatório de rotação e de translação das moléculas. Os valores máximos permitidos para estes movimentos são determinados de forma a obter-se a razão de aproximadamente 45% entre as configurações aceitas e geradas. São realizados também, movimentos de volume da caixa e estes obedecem a mesma proporção de 45%. A energia da configuração é calculada através do potencial de Lennard-Jones e de Coulomb. No início, a energia do sistema varia rapidamente e, no decorrer da simulação, esta variação reduz-se a pequenas oscilações ao redor de um valor médio. Neste ponto, dizemos que o sistema atingiu o equilíbrio. Cerca de 6×10^6 configurações foram necessárias para obter-se o equilíbrio para cada uma das conformações das neolignanas. Após esta fase, iniciou-se o acúmulo final dos valores calculados em cada configuração, para a obtenção de médias estatísticas das observáveis termodinâmicas citadas no capítulo 2, item 2.1. Nesta etapa, cerca de 3×10^6 configurações foram geradas e os valores de interação soluto-solvente puderam ser obtidos para cada conformação das neolignanas estudadas.

4.3.2 - Relações estrutura-atividade.

Após a solvatação, procuramos definir a sobreposição entre as neolignanas e o receptor.

A energia total de interação entre as neolignanas na conformação ativa (D) e o sítio receptor foi obtida com o programa computacional PowerFit [35]. Esse programa é utilizado para realizar a comparação entre estruturas, muito útil em estudos SAR. Tradicionalmente, tais comparações são realizadas com modelos plásticos 3D ou com base em figuras em 2D de moléculas. Esses métodos são denominados “Computer-Assisted Molecular Fitting (CAMF)”.

Inicialmente, utilizamos modelos moleculares plásticos e nos baseamos nos resultados já obtidos com a nossa análise prévia de SAR, para supormos uma possível maneira de interação entre as neolignanas e a arginina (capítulo 8). Em seguida, a nossa suposição pôde ser confirmada pelos resultados obtidos com o PowerFit.

O programa citado nos permite realizar o ajuste automático entre as moléculas, ou manual, utilizando intuição química. No nosso caso, baseando-nos nos resultados da análise das relações entre a estrutura química das neolignanas e a atividade anti-leishmaniose, foi possível fazer o ajuste manual entre as neolignanas e a arginina (receptor); as comparações entre distâncias atômicas das neolignanas e da arginina também nos deram indicações de como o ajuste entre as duas moléculas deveria ocorrer. Sendo assim, ajustamos manualmente as duas moléculas, considerando a conformação D para as neolignanas; para a arginina, consideramos inicialmente a conformação encontrada na literatura e depois, uma conformação suposta por nós no capítulo 8. Calculamos então, as energias de interação entre as neolignanas e a arginina para os dois casos, usando a opção “Atom Based Rigid Fit”.

CAPÍTULO 5

INCLUSÃO DE NOVOS PARÂMETROS FÍSICO-QUÍMICOS PARA A CONFORMAÇÃO MAIS PROVÁVEL (E), ENCONTRADA PARA OS COMPOSTOS ISOLADOS.

5.1 INTRODUÇÃO

No trabalho citado na referência 6, as seguintes propriedades FQ foram calculadas para as vinte neolignanas da tabela 2: energias do HOMO e LUMO, dureza e moleza moleculares, eletronegatividade de Mülliken, momento dipolar, volume de van der Waals, ângulos diedros t1 e t2 e calor de formação. Todos estes parâmetros foram utilizados na análise com RP.

De todos os métodos RP aplicados, apenas o SIMCA apresentou um bom resultado, com 90% de acerto na classificação dos compostos. Os métodos "Agrupamento Hierárquico" e ACP não foram capazes de agrupar os compostos nas respectivas categorias; não mostraremos as tabelas e figuras correspondentes, pois não fazem parte deste trabalho e podem ser observados na referência 6. O método KNN apresentou 60% de classificação correta para 5 vizinhos mais próximos, sendo que esse método foi aplicado posteriormente, não comparecendo na referência 6, como os demais.

Com o objetivo de melhorar os resultados obtidos na referência 6 e também de entendermos melhor a natureza da relação entre as variáveis físico-químicas e a atividade anti-leishmaniose das neolignanas, novos parâmetros físico-químicos foram calculados e incluídos na análise com os métodos RP.

5.2 - CÁLCULOS

Os cálculos deste capítulo estão descritos no capítulo 4, ítem 4.1.2. A conformação escolhida foi a E, de acordo com os resultados obtidos nos trabalhos anteriores, nas referências 5 e 6.

Resumidamente, a conformação E foi escolhida porque dentre as nove conformações encontradas no vácuo (A, B, C...I), a E apresentou alta probabilidade em todas as neolignanas ativas.

A probabilidade de cada mínimo local foi calculada considerando a população relativa entre os mínimos locais, através da função de partição molecular (Z):

$$Z = \sum_t e^{-E(t)/kT} \quad \text{eq. 5.1}$$

onde E é a energia calculada para determinada conformação, k é a constante de Boltzmann, T é a temperatura e t é o ângulo que define a conformação da molécula.

Numa superfície de potencial, verificamos regiões de máxima e de mínima energia; as regiões de mínima energia formam poços ou vales, onde um determinado número de conformações está presente. À primeira vista, podemos pensar que um poço mais profundo (de energia mais baixa) apresenta maior probabilidade do que outro de energia um pouco mais alta; entretanto, se considerarmos a largura dos poços, um número maior de conformações estará presente num poço menos profundo, porém mais largo, do que num poço mais profundo e mais estreito. Esse número de conformações foi levado em consideração nos nossos cálculos. O limite de cada poço foi definido conforme a referência 19.

Cada mínimo local foi associado com a probabilidade:

$$Z(t) = e^{-E(t)/kT} \quad \text{eq.5.2}$$

Foi obtido então, o somatório das funções de probabilidade sobre todos os mínimos, de forma a se obter a função Z (equação 5.1). Em seguida, cada função de probabilidade (associada a cada mínimo) foi normalizada:

$$Z(t) = Z(t)/Z \quad \text{eq.5.3}$$

5.3 RESULTADOS E DISCUSSÃO

As tabelas 4a e 4b mostram os valores para as 21 variáveis consideradas inicialmente: energias do HOMO e LUMO, dureza e moleza moleculares (η e $1/\eta$), eletronegatividade de Müliken (χ), volume, ângulos diedros t_1 e t_2 , calor de formação, momento de dipolo total, híbrido e pontual (μ_t , μ_h e μ_p), cargas sobre os átomos 8, 9, 10, 16, 17, 24, 26 e 28 (figura 4a) e $\log P$.

Os gráficos dos scores das duas primeiras componentes principais e dos loadings, usando as 21 variáveis das tabelas 4a e 4b, são mostrados na figura 18c e 18d; as figuras 18a e b mostram o agrupamento hierárquico e a variância para as três primeiras componentes principais, respectivamente. As duas primeiras componentes principais acumulam apenas 57% da variância total (figura 18b) e por isso, o gráfico dos scores (18c) dessas componentes principais não mostra separação entre as duas categorias. O dendograma (figura 18a) não mostra agrupamento dos compostos ativos e dos inativos; observa-se uma mistura das duas categorias.

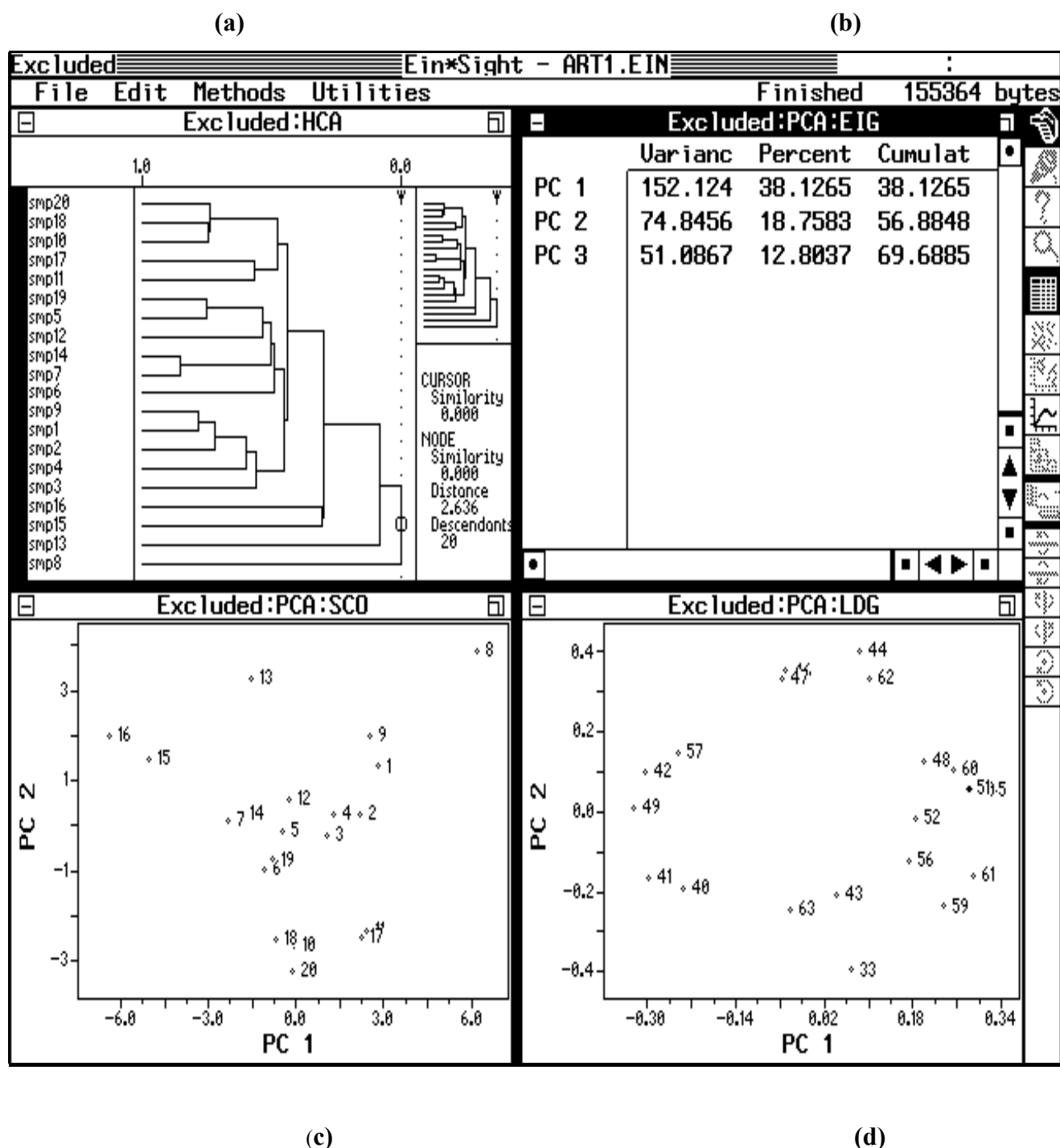


Figura 18 - Resultados da análise RP obtidos para as 21 variáveis da tabela 4. O agrupamento hierárquico é mostrado na figura 18a, a variância das três primeiras componentes principais na figura 18b. Os gráficos dos "scores" e "loadings" das duas primeiras componentes principais são mostrados nas figuras 18c e 18d.

Tabela 4a- Variáveis utilizadas na análise inicial de RP para neolignanas testadas em leishmaniose visceral. η , $1/\eta$ e χ , são a dureza molecular, moleza molecular e a eletronegatividade de Mülliken; t1 e t2 são os ângulos diedros mostrados na figura 4a.

Composto	HOMO	LUMO	η	$1/\eta$	χ	VOLUM E	t1	t2
1	-9,261	-0,573	0,230	4,344	4,917	192,200	175,630	177,610
2	-9,141	-0,635	0,235	4,253	4,888	210,899	176,780	181,250
3	-8,807	-0,555	0,242	4,126	4,681	223,000	168,220	185,060
4	-8,830	-0,482	0,239	4,174	4,656	219,600	174,259	185,751
5	-8,933	-0,452	0,236	4,241	4,692	259,299	177,820	172,210
6	-8,874	-0,389	0,236	4,242	4,631	283,100	154,929	173,090
7	-8,876	-0,437	0,237	4,220	4,656	325,700	178,250	161,560
8	-9,891	-1,155	0,229	4,368	5,523	216,399	178,580	181,000
9	-9,362	-0,518	0,226	4,422	4,940	214,899	180,030	178,230
10	-8,911	-0,632	0,242	4,139	4,772	271,299	168,259	172,880
11	-9,142	-0,685	0,237	4,228	4,913	211,100	158,009	161,800
12	-9,126	-0,557	0,233	4,284	4,841	269,000	177,380	178,970
13	-9,062	-0,369	-0,230	4,347	4,716	292,100	175,130	179,700
14	-8,879	-0,464	0,238	4,208	4,671	312,000	177,429	167,179
15	-8,678	-0,132	0,234	4,273	4,405	333,299	179,320	162,340
16	-8,603	-0,114	0,236	4,245	4,359	376,600	178,730	165,660
17	-9,190	-0,798	0,238	4,196	4,994	240,000	165,700	172,820
18	-8,899	-0,608	0,241	4,141	4,753	287,500	167,820	173230
19	-8,901	-0,513	0,236	4,234	4,747	296,700	177,570	172,060
20	-8,952	-0,796	0,245	4,078	4,874	286,600	160,250	172,800

Tabela 4b - Parâmetros FQ calculados para as 20 neolignanas do conjunto treinamento, na conformação E. ΔH_f é o calor de formação e μ_t , μ_h e μ_p são os momentos dipolares total, híbrido e pontual, respectivamente; C_8 , C_9 , C_{10} , C_{16} , C_{17} , C_{24} , C_{26} e C_{28} são as cargas sobre os átomos 8, 9, 10, 16, ...28 (figura 4a).

Composto	ΔH_f	μ_t	C8	C9	C10	C16	C17	C24
1	-11,420	2,813	0,110	0,105	0,110	-0,315	-0,176	0,110
2	-17,740	2,211	0,110	0,105	0,109	-0,303	-0,172	0,113
3	-46,070	2,933	0,109	0,105	0,108	-0,303	-0,174	0,110
4	-45,300	3,383	0,109	0,104	0,108	-0,306	-0,164	0,112
5	-83,920	2,240	0,118	-0,172	-0,173	-0,308	-0,180	0,108
6	-117,88	2,971	0,117	-0,172	-0,174	-0,308	-0,168	0,111
7	-101,88	3,671	0,120	-0,183	-0,184	-0,311	-0,173	0,117
8	-12,750	5,215	0,109	0,104	0,107	-0,309	-0,160	0,125
9	-18,840	3,056	0,109	0,105	0,108	-0,308	-0,195	0,113
10	-43,850	2,536	0,117	-0,173	-0,174	-0,309	0,073	0,111
11	27,470	2,664	0,109	0,105	0,107	-0,306	0,088	0,120
12	-109,150	1,717	0,121	-0,183	-0,183	-0,308	-0,192	0,114
13	-134,56	4,380	0,120	-0,184	-0,184	-0,313	-0,181	-0,212
14	-99,280	3,616	0,120	-0,183	-0,183	-0,309	-0,174	0,116
15	-115,75	3,819	0,119	-0,243	-0,221	0,063	-0,186	-0,208
16	-134,10	4,580	-0,209	-0,189	-0,184	0,066	-0,186	-0,208
17	20,100	2,067	0,110	0,105	0,108	-0,304	0,076	0,117
18	-52,830	2,729	0,118	-0,173	-0,175	-0,309	0,079	0,119
19	-90,120	1,260	0,118	-0,171	-0,173	-0,307	-0,178	0,112
20	-49,830	2,052	0,118	-0,173	-0,174	-0,308	0,075	0,114

Continuação da tabela 4b

COMPOSTO	C26	C28	LOGP	μ_P	μ_H
1	0,107	0,112	3,730	2,628	0,468
2	0,057	0,117	4,500	2,195	0,520
3	-0,184	0,112	3,810	2,230	0,348
4	0,107	-0,174	3,810	2,484	0,664
5	0,108	0,114	4,210	1,730	0,509
6	0,107	-0,172	4,290	1,910	0,695
7	-0,108	-0,183	5470	2,558	0,598
8	0,120	1,304	3,480	4,608	0,889
9	0,108	0,118	4,268	2,522	0,441
10	0,106	0,123	4,808	1,986	1,687
11	0,105	0,118	4,328	2,713	1,735
12	-0,201	0,117	3,868	1,910	0,208
13	0,108	-0,182	3,857	2,720	0,954
14	-0,108	-0,183	5,052	2,714	0,592
15	-0,105	-0,186	5,563	2812	1,103
16	-0,105	-0,186	6,232	2,774	1,238
17	0,066	0,126	5,509	2,001	1,297
18	-0,068	0,123	5,346	2,214	1,670
19	0,057	0,117	4,976	1,563	0,540
20	0,066	0,126	5,574	1,803	1,321

Com o objetivo de melhorar os resultados obtidos nas figuras 18a, b, c e d, calculamos o peso de Fisher [34] para essas variáveis, que estima a importância relativa de cada uma, na separação dos compostos; o melhor resultado foi obtido com as 5 mais pesadas: volume, logP e cargas sobre os átomos 9, 10 e 26. Os resultados são mostrados nas figuras 19 a, b e c.

A figura 19a mostra o agrupamento da maioria das moléculas ativas (20, 19, 13, 18, 15, 14, 16 e 12) em A e o agrupamento das inativas 11, 9, 2, 8, 4, 1 e 3 em B. Algumas moléculas inativas (6, 5, 10 e 7) foram agrupadas com as ativas em A e a molécula ativa 17 foi agrupada com a maioria das inativas em B.

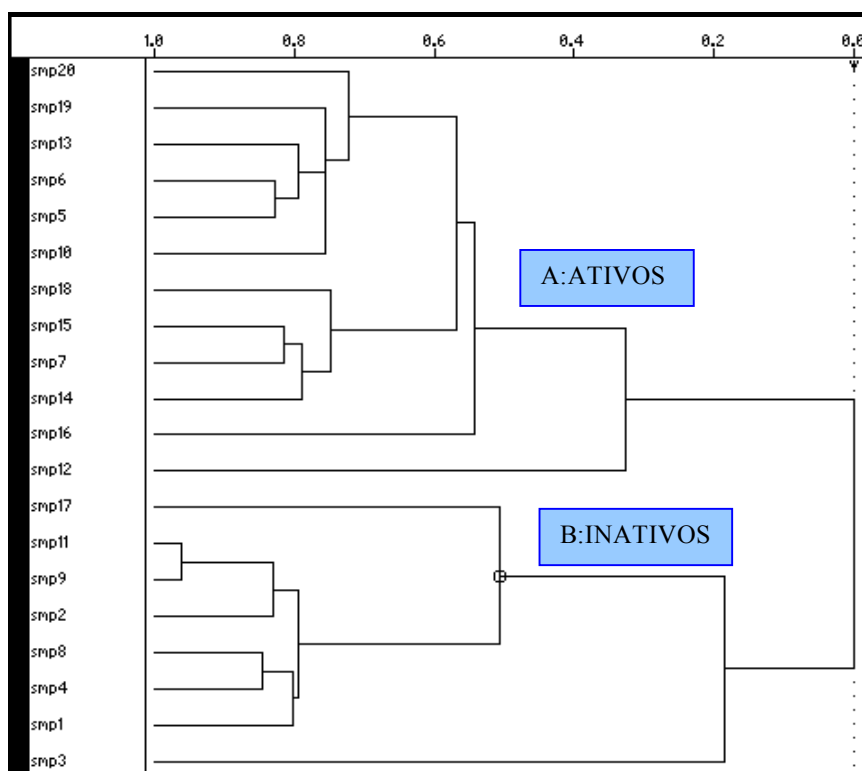


Figura 19 a - Agrupamento hierárquico obtido para as neolignanas, usando as seguintes variáveis: logP, Volume e cargas nas posições 9 (C9), 10 (C10) e 26 (C26).

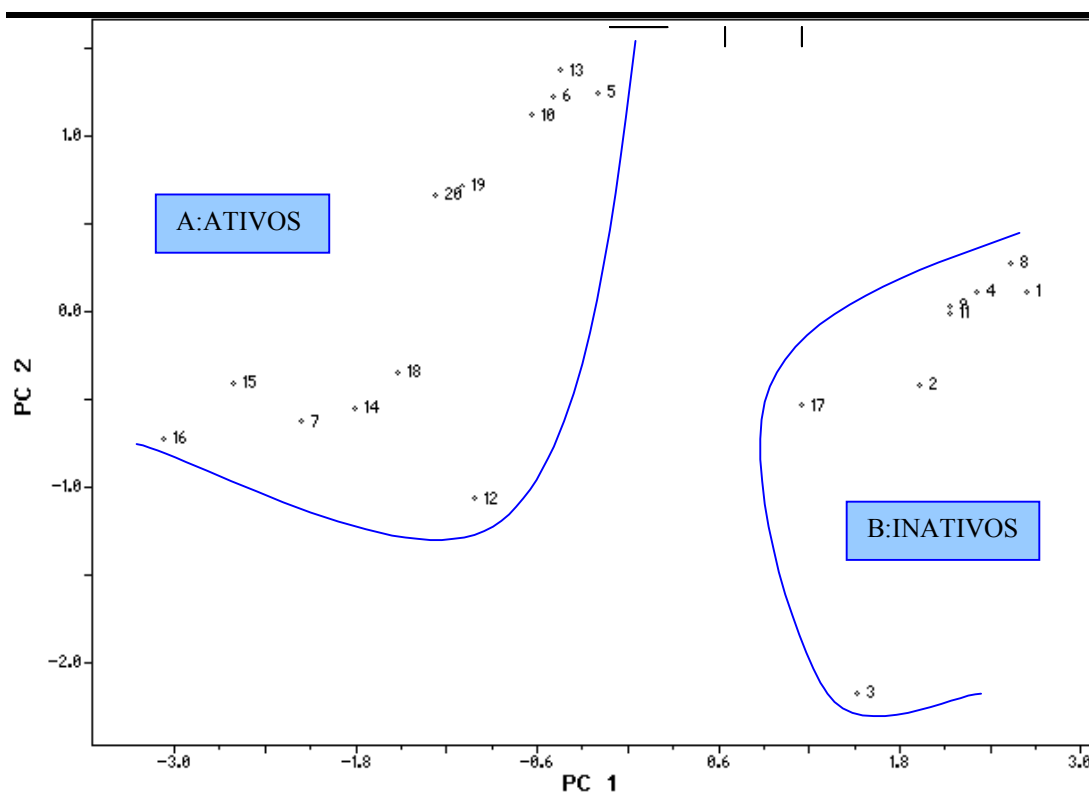


Figura 19b - Análise de componentes principais obtida para as neolignanas, com as variáveis logP, volume e cargas sobre os átomos 9, 10 e 26.

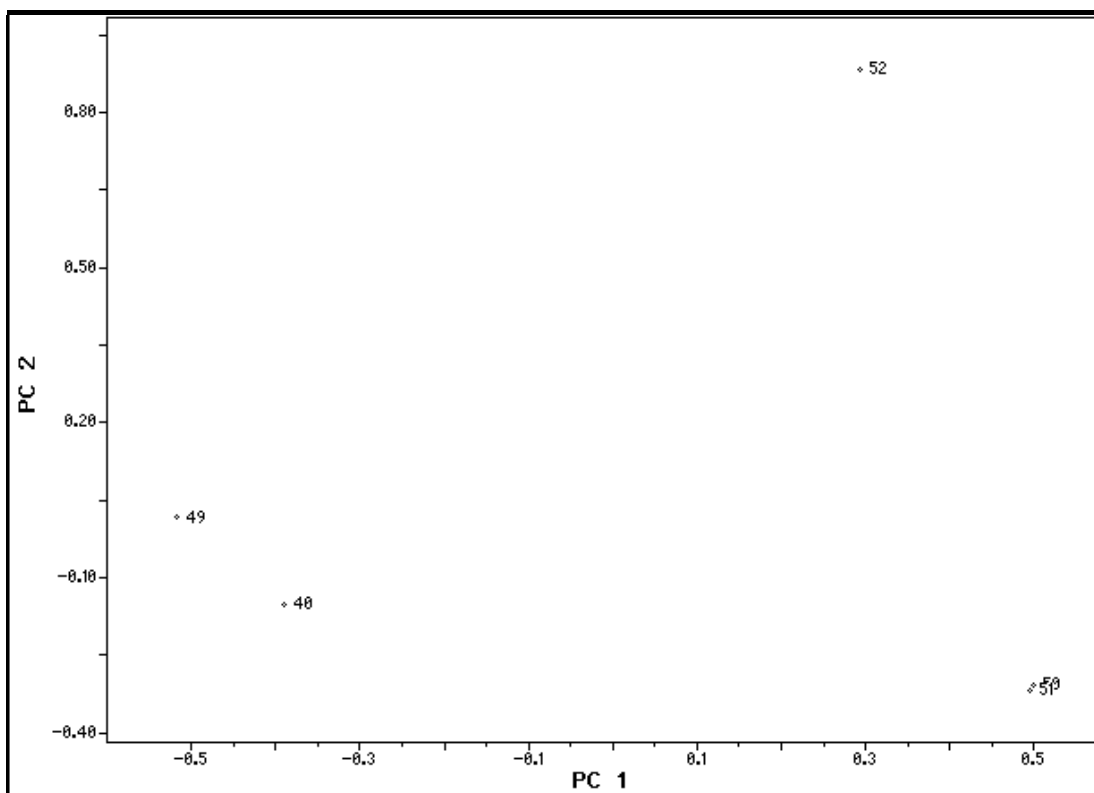


Figura 19 c - Gráfico dos “loadings” das duas primeiras componentes principais, obtido para as neolignanas. As variáveis 40, 49, 50, 51 e 52 são: volume, logP e cargas sobre os átomos 9, 10 e 26 (figura 4a), respectivamente.

As duas primeiras componentes principais acumulam, agora, 85,9% da variância e o gráfico dos scores das duas primeiras componentes principais (figura 19 b) mostra que as moléculas ativas 12, 13, 14, 15, 16, 18, 19, 20, e as inativas 5, 6, 7 e 10 - grupo A - estão separadas das inativas 1, 2, 3, 4, 8, 9, e da ativa 17 - grupo B - (essas moléculas estão dispostas na tabela 2 e figura 4a), pela primeira componente principal (CP1), isto é, um dos conjuntos está disposto à direita do gráfico e o outro à esquerda, na figura 19b. Este resultado é parecido com o agrupamento hierárquico mostrado na figura 19a. Como o auxílio do gráfico dos loadings (19c) para as duas primeiras componentes principais, podemos explicar os agrupamentos obtidos no gráfico dos scores (19b), da seguinte maneira: As variáveis mais importantes, na primeira componente principal são o volume (variável 40), logP (var. 49) e as cargas sobre os átomos 9 e 10 (variáveis 50 e 51); isto

pode ser visto na figura 19c, onde as variáveis 40 e 49 possuem valores negativos altos no eixo da primeira componente principal e as variáveis 50 e 51 possuem valores positivos também altos no mesmo eixo.

A variável 52 (carga sobre o átomo 26) possui valor positivo alto no eixo da segunda componente principal e está separando as moléculas inativas 5, 6 e 10 e as ativas 13, 19 e 20 das demais, como pode ser observado na figura 19b, em relação à segunda componente principal.

A tabela 5 mostra os valores de atividade anti-leishmaniose para alguns compostos (temos valores de AB apenas para estes) [7]. O composto 18 é o mais ativo entre esses e o composto 13 possui AB baixa, sendo o composto 20 o menos ativo entre eles. Esses resultados concordam, de maneira geral, com os obtidos nas figuras 19a e 19b.

Tabela 5 - valores de atividade anti-leishmaniose, para algumas neolignanas testadas "in vitro"; %I a porcentagem de inibição.

NEOLIGNANA	% I "in vitro"
20	0,8
13	14,0
17	63,5
18	96,5

Os compostos inativos 5, 6 e 10 possuem estrutura química parecida com a maioria dos ativos (veja a tabela 2); esses compostos foram agrupados, na nossa análise, com o composto pouco ativo, 13. Assim, de acordo com os nossos resultados e com os valores de AB "in vitro", podemos sugerir que esses compostos inativos seriam "quase ativos". O composto mais ativo, 18, foi separado da maioria das inativas e agrupado com outras moléculas ativas, pela primeira componente principal. Os Compostos ativos que foram agrupados com 18 seriam assim, os mais ativos. Entretanto, o composto 17 com atividade relativamente alta foi agrupado com a maioria das inativas e, o composto inativo 7, com as ativas, ficando próximo do composto 18. Apesar da classificação incorreta dos compostos 7 e 17, este modelo separou a maioria das moléculas em ativas e inativas e colocou numa

posição intermediária do gráfico, algumas inativas de estrutura parecida com a molécula pouco ativa, 13.

Na presente análise foram consideradas cargas sobre os heteroátomos substituídos nos anéis aromáticos e, algumas delas foram selecionadas pelo peso de Fisher, estando entre as cinco variáveis mais importantes. Apesar deste modelo mostrar-se capaz de separar os compostos, preferimos substituir essas cargas pelas cargas dos carbonos em que os heteroátomos estão ligados, de forma a considerar não apenas o tipo de átomo substituído, mas a influência que estes exercem sobre outras posições dos anéis. Se usarmos as cargas dos heteroátomos, todas as moléculas que contêm o grupo -OH em determinada posição do anel, por exemplo, serão agrupadas e outras, substituídas por um outro grupo, como o NH₂ que também é ativador do anel para substituição eletrofílica, formariam um grupo diferente, apesar de ambos grupos exercerem a mesma influência sobre os anéis aromáticos. Se ao invés disso considerarmos as cargas dos carbonos em que estes estão ligados, os dois casos formariam apenas um grupo, tornando a análise mais realista.

Além da substituição das cargas citadas, outros parâmetros físico-químicos foram incluídos: densidade eletrônica (def), orbital (dof) e radical (drf) de fronteira, sobre os átomos 8, 9, 10, 16, 17, 24, 26 e 28; as cargas foram calculadas sobre os átomos 1, 2, 3, 4, 5, 6, 12, 13, 17, 18, 19, 20, 21, 22 e 23. Esses valores para as 20 moléculas são mostrados na tabela 6.

Tabela 6 - Inclusão de novos parâmetros físico-químicos na análise RP das neolignanas testadas em leishmaniose, para a conformação E. As variáveis 1, 2 e 40 a 49 já haviam sido incluídas na análise da figura 19 e estão dispostas também na tabela 4. As cargas C8, C9, C10, C24, C26 e C28 da tabela 4 foram substituídas pelas cargas correspondentes, dos anéis aromáticos: C2, C3, C4, C19, C21 e C23 (continua nas próximas 6 páginas). def 8, def 9,...,etc, são as densidades eletrônicas de fronteira sobre os átomos 8, 9, ...etc. dof 8, dof9, etc, são as densidades orbitais de fronteira sobre os respectivos átomos, e drf 8, drf9, etc, são as densidades radicais de fronteira sobre os átomos 8, 9, etc.

Composto	def8	def9	def 10	def16	def17	def24	def26
	1	2	3	4	5	6	7
1	0,00000	0,00000	0,00000	0,00104	0,91800	0,00003	0,00000
2	0,00000	0,00000	0,00000	0,00167	0,61300	0,00000	1,15600
3	0,00000	0,00000	0,00000	0,00757	0,61700	0,00005	0,55200
4	0,00000	0,00000	0,00000	0,00278	0,67800	0,00000	0,00000
5	0,00001	0,24100	0,28900	0,07090	0,00028	0,00000	0,00000
6	0,00001	0,23000	0,30900	0,05800	0,01960	0,00002	0,00000
7	0,00000	0,00016	0,00009	0,00122	0,41400	0,00013	0,36400
8	0,00000	0,00000	0,00000	0,00510	0,00172	0,00000	0,00000
9	0,00000	0,00000	0,00000	0,00190	0,76800	0,00000	0,00000
10	0,00001	0,25600	0,34200	0,05800	0,00470	0,00000	0,00000
11	0,00000	0,00006	0,00178	0,00520	2,92000	0,00000	0,00000
12	0,00000	0,00046	0,00023	0,00098	0,59100	0,00000	0,64900
13	0,00000	0,21600	0,00147	0,06200	0,00960	0,00209	0,00000
14	0,00000	0,00015	0,00005	0,00027	0,45100	0,00000	0,38400
15	0,00000	0,00280	0,00245	0,00001	0,46200	0,09800	0,11300
16	0,07700	0,31200	0,24300	0,00459	0,01700	0,00203	0,00250
17	0,00001	0,00004	0,00003	0,00486	3,46500	0,00000	0,14300
18	0,00001	0,26100	0,35800	0,06200	0,00368	0,00000	0,00000
19	0,00001	0,24300	0,29200	0,07000	0,00035	0,00000	0,00000
20	0,00001	0,26300	0,35700	0,06500	0,00203	0,00000	0,00000

Tabela 6 - Continuação

Composto	def28	dof8	dof9	dof10	dof16	dof17	dof24
	8	9	10	11	12	13	14
1	0,00000	0,00000	0,00000	0,00000	0,33500	0,03250	0,00007
2	0,00000	0,00000	0,00000	0,00001	0,36600	0,00440	0,00007
3	0,00000	0,00000	0,00000	0,00000	0,28000	0,03980	0,00007
4	0,71500	0,00000	0,00000	0,00000	0,33400	0,04000	0,00008
5	0,00000	0,00002	0,04600	0,00157	0,40200	0,04700	0,00006
6	0,01640	0,00001	0,04500	0,00123	0,3355	0,04500	0,00007
7	0,26600	0,00001	0,03300	0,00147	0,35000	0,04000	0,00006
8	0,00000	0,00000	0,00000	0,00000	0,00007	0,05200	0,05200
9	0,00000	0,00001	0,00000	0,00001	0,36300	0,04100	0,00007
10	0,00000	0,00001	0,01666	0,00028	0,24800	0,61700	0,00035
11	0,00004	0,00000	0,00000	0,00000	0,21600	0,41800	0,00020
12	0,00000	0,00001	0,03300	0,00138	0,37900	0,04300	0,00006
13	0,00329	0,00001	0,03400	0,00140	0,33000	0,04400	0,00147
14	0,27600	0,00001	0,03300	0,00148	0,34800	0,04200	0,00007
15	0,14300	0,00000	0,00000	0,00000	0,00000	0,05600	0,00753
16	0,00465	0,00173	0,00000	0,00000	0,00001	0,06100	0,00477
17	0,00000	0,00001	0,00000	0,00000	0,17900	0,45800	0,00027
18	0,00000	0,00001	0,01600	0,00032	0,25100	0,60400	0,00033
19	0,00000	0,00002	0,04500	0,00145	0,40700	0,05200	0,00006
20	0,00000	0,00001	0,00600	0,00010	0,15100	0,49900	0,00026

Tabela 6 - continuação

Composto	dof26	dof28	drf8	drf9	drf10	drf16	drf17
	15	16	17	18	19	20	21
1	0,00000	0,00001	0,00000	0,00000	0,00000	0,16800	0,47500
2	0,00006	0,00001	0,00001	0,00000	0,00000	0,18400	0,30900
3	0,00022	0,00001	0,00000	0,00000	0,00000	0,14400	0,32800
4	0,00000	0,00009	0,00000	0,00000	0,00000	0,16800	0,35900
5	0,00001	0,00001	0,00001	0,14400	0,14500	0,23600	0,02400
6	0,00001	0,00025	0,00001	0,13800	0,15500	0,19700	0,03200
7	0,00009	0,00022	0,00001	0,01700	0,00078	0,17600	0,22700
8	0,00000	0,47600	0,00000	0,00000	0,00000	0,00259	0,02700
9	0,00001	0,00002	0,00001	0,00000	0,00000	0,18200	0,40500
10	0,00002	0,00002	0,00001	0,13600	0,17100	0,15300	0,31100
11	0,00002	0,00020	0,00004	0,00003	0,00103	0,11100	1,66900
12	0,00005	0,00001	0,00001	0,01700	0,00081	0,19000	0,31700
13	0,00000	0,00050	0,00001	0,01700	0,00077	0,19600	0,02700
14	0,00050	0,00024	0,00001	0,01700	0,00077	0,17400	0,24700
15	0,22100	0,00854	0,00000	0,00138	0,00123	0,00001	0,25900
16	0,24400	0,00834	0,03900	0,15600	0,12200	0,00230	0,03900
17	0,04100	0,00014	0,00001	0,00002	0,00002	0,09200	1,96200
18	0,00027	0,00019	0,00001	0,13900	0,17900	0,15700	0,30400
19	0,00086	0,00001	0,00001	0,14400	0,14700	0,23900	0,02600
20	0,04860	0,00013	0,00001	0,13500	0,17900	0,10800	0,25100

Tabela 6 - Continuação

Composto	drf24	drf26	drf28	C1	C2	C3	C4
	22	23	24	25	26	27	28
1	0,00005	0,00000	0,00001	-0,0340	-0,1321	-0,0549	-0,1328
2	0,00004	0,57800	0,00000	-0,0498	-0,1231	-0,0667	-0,1233
3	0,00006	0,27600	0,00001	-0,0499	-0,1237	-0,0677	-0,1240
4	0,00004	0,00000	0,35800	-0,0461	-0,1243	-0,0685	-0,1251
5	0,00003	0,00000	0,00000	-0,1151	0,05230	0,1035	-0,1898
6	0,00004	0,00000	0,00832	-0,1142	0,05070	0,10270	-0,1905
7	0,00010	0,18200	0,13300	-0,0972	0,04680	0,09770	-0,1725
8	0,02600	0,00000	0,23800	-0,0414	-0,1204	-0,0642	-0,1233
9	0,00003	0,00000	0,00001	-0,0504	-0,1239	-0,0684	-0,1243
10	0,00017	0,00001	0,00001	-0,1154	0,04990	0,10340	-0,1895
11	0,00010	0,00001	0,00012	-0,0547	-0,1236	-0,0686	-0,1230
12	0,00003	0,32500	0,00001	-0,0993	0,04680	0,09910	-0,1716
13	0,00178	0,00000	0,00190	-0,0961	0,04590	0,09710	-0,1735
14	0,00004	0,19200	0,13800	-0,0972	0,04660	0,09850	-0,1725
15	0,05300	0,16700	0,07600	-0,1103	0,00610	0,06600	-0,1076
16	0,00340	0,12300	0,00650	-0,1554	0,08370	0,04780	0,50900
17	0,00014	0,09200	0,00011	-0,0523	-0,1228	-0,0678	-0,1230
18	0,00016	0,00013	0,00010	-0,1147	0,04980	0,10330	-0,1896
19	0,00003	0,00047	0,00000	-0,1156	0,05260	0,10450	-0,1897
20	0,00013	0,02400	0,00006	-0,1147	0,05020	0,10430	-0,1893

Tabela 6 - continuação

Composto	C5	C6	C12	C13	C17	C18	C19
	29	30	31	32	33	34	35
1	-0,0187	-0,1777	0,32500	0,00040	-0,1764	0,10090	-0,1438
2	-0,0358	-0,1640	0,32560	0,00190	-0,1720	0,10390	-0,1347
3	-0,0370	-0,1636	0,32510	0,00340	-0,1738	0,06130	-0,1007
4	-0,0370	-0,1618	0,3272	0,00000	-0,1639	0,06640	0,07420
5	-0,0334	-0,1631	0,33170	0,03600	-0,1801	0,10000	-0,1378
6	-0,0336	-0,1621	0,33320	0,03330	-0,1680	0,06360	0,07810
7	-0,0353	-0,1579	0,33020	0,03450	-0,1732	0,06320	0,06700
8	-0,0324	-0,1588	0,33340	0,00390	-0,1599	0,18320	-0,4272
9	-0,0363	-0,1606	0,3254	0,00840	-0,1946	0,09040	-0,1308
10	-0,0317	-0,1608	0,34570	-0,1955	0,07250	-0,2147	-0,0416
11	-0,0367	-0,1572	0,34190	-0,2448	0,08830	-0,2269	-0,0479
12	-0,0340	-0,1613	0,32850	0,00820	-0,1924	0,05600	-0,0968
13	-0,0361	-0,1575	0,32800	0,00940	-0,1808	0,04810	0,08870
14	-0,0339	-0,1597	0,33040	0,00340	-0,1739	0,05910	0,06800
15	-0,0982	-0,0794	-0,0556	0,07840	-0,1864	0,05960	0,08220
16	-0,1827	-0,0471	-0,0563	0,08030	-0,1863	0,05870	0,08230
17	-0,0362	-0,1573	0,34030	-0,1956	0,07630	-0,2165	-0,0322
18	-0,0317	-0,1602	0,34590	-0,1944	0,07060	-0,2221	-0,0375
19	-0,0327	-0,1642	0,33170	0,03490	-0,1780	0,10290	-0,1300
20	-0,0309	-0,1616	0,34580	-0,1944	0,07490	-0,2162	-0,0322

Tabela 6 - continuação

Composto	C20	C21	C22	C23	logP	HOMO	LUMO
	36	37	38	39	40	41	42
1	-0,0645	-0,1393	-0,0634	-0,1905	3,7300	-9,2610	-0,5730
2	-0,0602	-0,1523	-0,0594	-0,1848	4,5000	-9,1410	-0,6350
3	-0,1548	0,07240	-0,01089	-0,1503	3,81000	-8,8070	-0,5550
4	-0,1643	-0,0990	-0,1025	-0,1627	3,81000	-8,8300	-0,4820
5	-0,0676	-0,1345	-0,0686	-0,1866	4,21000	-8,9330	-0,4520
6	-0,1683	-0,0943	-0,1084	-0,1572	4,2900	-8,8740	-0,3890
7	-0,1432	-0,0259	-0,0977	-0,1510	5,47000	-8,8760	-0,4370
8	0,01320	-0,1549	-0,0097	-0,2028	3,48000	-9,8910	-1,1550
9	-0,0432	-0,1311	-0,0677	-0,1760	4,26800	-9,3620	-0,5180
10	-0,1177	-0,0716	-0,1175	-0,0584	4,80800	-8,9110	-0,6320
11	-0,1207	-0,0689	-0,1189	-0,0446	4,32800	-9,1420	-0,6850
12	-0,1393	0,06300	-0,0983	-0,1401	3,86800	-9,1260	-0,5570
13	-0,1768	-0,0695	-0,1810	0,06390	3,85700	-9,0620	-0,3690
14	-0,1439	-0,0233	-0,0961	-0,1484	5,05200	-8,8790	-0,4640
15	-0,1704	-0,0022	-0,1764	0,0636	5,56300	-8,6780	-0,1320
16	-0,1697	-0,0025	-0,1759	0,06320	6,2320	-8,6030	-0,1140
17	-0,1171	-0,0906	-0,1163	-0,0476	5,50900	-9,1900	-0,7980
18	-0,1205	-0,0421	-0,1222	-0,0525	5,34600	-8,8990	-0,6080
19	-0,0643	-0,1482	-0,0654	-0,1788	4,97600	-8,9810	-0,5130
20	-0,1164	-0,0908	-0,1162	-0,0489	5,57400	-8,9520	-0,7960

Tabela 6 - continuação

Composto	η	$1/\eta$	χ	μ	t1	t2	V
	43	44	45	46	47	48	49
1	0,23000	4,34400	4,91700	2,81300	175,630	177,610	192,200
2	0,23500	4,25300	4,88800	2,21100	176,780	181,250	210,899
3	0,24200	4,12600	4,68100	2,93300	168,220	185,060	223,000
4	0,23900	4,17400	4,65600	3,38300	174,259	185,751	219,600
5	0,23600	4,24100	4,69200	2,24000	177,820	172,210	259,299
6	0,23600	4,24200	4,63100	2,97100	154,929	173,090	283,100
7	0,23700	4,22000	4,65600	3,67100	178,250	161,560	325,700
8	0,22900	4,36800	5,52300	5,21500	178,580	181,000	216,399
9	0,22600	4,42200	4,94000	3,05600	180,030	178,230	214,899
10	0,24200	4,13900	4,77200	2,53600	168,259	172,880	271,299
11	0,23700	4,22800	4,91300	2,66400	158,009	161,800	211,100
12	0,23300	4,2840	4,84100	1,71700	177,380	178,970	269,000
13	-0,2300	4,34700	4,71600	4,38000	175,130	179,700	292,100
14	0,23800	4,20800	4,67100	3,61600	177,429	167,179	312,00
15	0,23400	4,27800	4,40500	3,81900	179,320	162,340	333,299
16	0,23600	4,24500	4,35900	4,58000	178,730	165,660	376,600
17	0,23800	4,19600	4,99400	2,06700	165,700	172,820	240,00
18	0,24100	4,14500	4,5300	2,72900	167,820	173,230	287,500
19	0,23600	4,23400	4,74700	1,26000	177,570	172,060	296,700
20	0,24500	4,07800	4,87400	2,05200	168,250	172,800	286,600

Os resultados obtidos considerando todas essas variáveis, incluindo as variáveis 1, 2 e 40 a 49 da tabela 6 (usadas na análise anterior), não separaram os compostos nas duas categorias, usando-se apenas duas componentes principais, uma vez que a variância acumulada para as duas primeiras componentes foi de apenas 47,6%. A figura 20a, b, c e d mostram o dendograma, a variância e os gráficos dos scores e loadings dessas duas componentes.

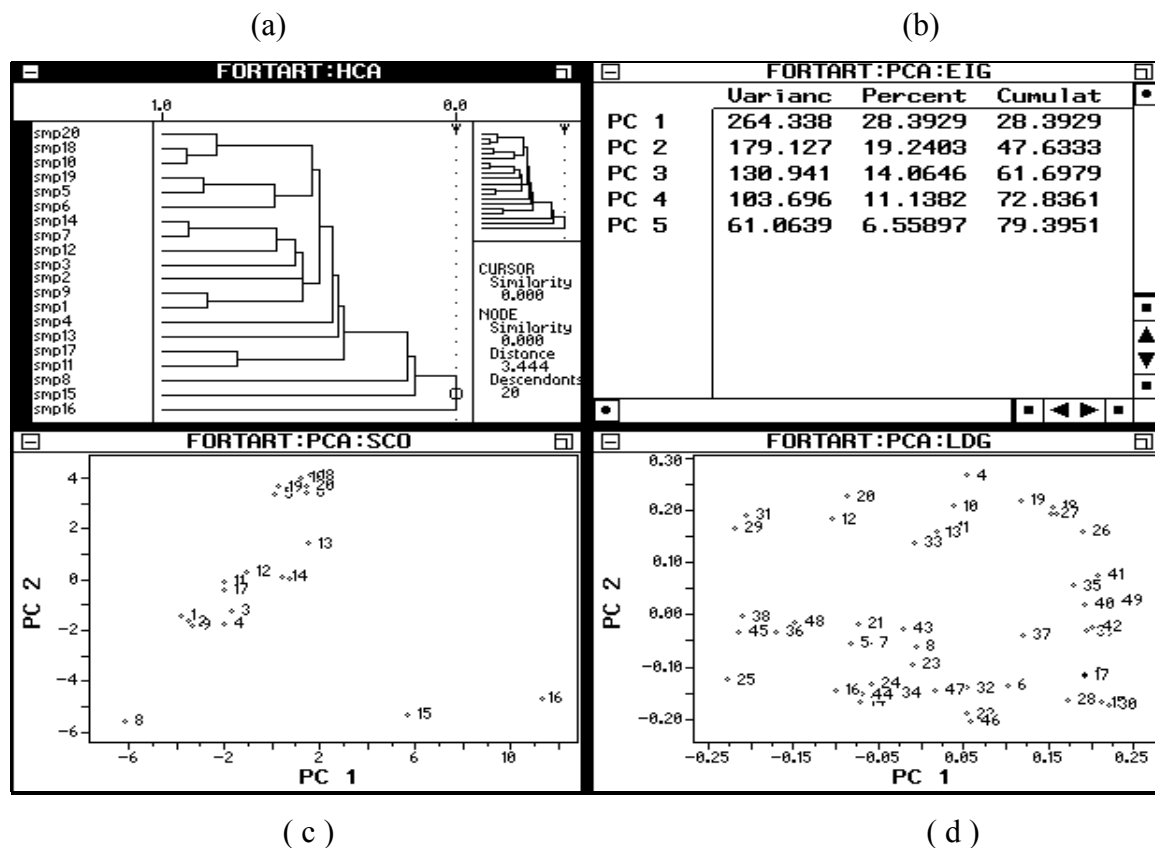


Figura 20 - Resultados obtidos para as 49 variáveis da tabela 6.

Nesta etapa, o método SIMCA, que constrói um modelo para cada categoria, usando números variáveis de componentes principais para cada uma, foi útil para nos ajudar a definir as variáveis relacionadas com as moléculas 5, 6, 7, 10 e 17 que geralmente formam um grupo separado ou são classificadas incorretamente, conforme foi mostrado nos resultados das figuras 19a e b.

Como já vimos, as componentes principais são construídas através da combinação linear das variáveis originais, sendo que numa mesma CP, as variáveis apresentam

coeficientes diferentes umas das outras e, conseqüentemente, importâncias diferentes; sendo assim, podemos definir quais variáveis são mais importantes na separação dos compostos. Por exemplo, se numa análise com 10 variáveis precisamos de 3 CP, mas apenas seis dessas variáveis apresentarem coeficientes significativos nas 3 CP, então, as 4 variáveis restantes não são importantes.

Podemos ainda, verificar quais variáveis (entre as mais importantes) estão relacionadas com determinado composto. Por exemplo, no nosso caso, as moléculas 5, 6, 7, 10 e 17 estão sendo classificadas incorretamente; por isso, desejamos saber quais variáveis são responsáveis por isso. O método SIMCA constrói um modelo de componentes principais para cada categoria, permitindo que analisemos cada categoria separadamente. Aplicamos então, o método SIMCA ao nosso conjunto treinamento e verificamos que o melhor resultado foi com o modelo de 4 componentes principais para a categoria 1 (inativos) e 3 componentes para a categoria 2 (ativos); 90% dos compostos foram classificados corretamente, sendo que agora, apenas as moléculas 7 e 19 foram mal classificadas (tabela 7).

Na análise ACP, apenas as duas primeiras CP são utilizadas; então este resultado já mostra que algumas variáveis presentes na terceira e quarta componentes principais são importantes, pois agora os compostos 5, 6, 10 e 17 foram classificados corretamente.

Tabela 7 - Resultado da análise SIMCA para o modelo de 4 CP para a categoria 1 e 3CP para a categoria 2.

CATEGORIA	NÚMERO DE COMPONENTES PRINCIPAIS	Nº TOTAL DE COMPOSTOS	COMPOSTOS CLASSIFICADOS INCORRETAMENTE
INATIVA (1)	4	11	7
ATIVA (2)	3	9	19

PORCENTAGEM DE CLASSIFICAÇÃO CORRETA: 90%

Vamos verificar o que ocorre com as moléculas inativas, quando diminuimos o número de componentes principais da categoria 1, para 3 (modelo de 3 CP para as duas categorias). Os compostos 5, 7 e 11 são classificados incorretamente (tabela 8); isto indica que as variáveis mais importantes na quarta componente principal da categoria 1, na análise anterior, são importantes na classificação dos compostos 5 e 11 que haviam sido bem

agrupados no modelo de 4 e 3 CP. Essas variáveis são: def 26, def24, def28, drf26 e carga sobre o átomo 20; elas estão relacionadas com o **anel (2)**. Como o composto 5 é parecido com os ativos devido aos substituintes -OCH₃ nas posições R1 e R2 do anel 1 (tabela 2), entendemos que para agrupá-lo às inativas, precisamos utilizar parâmetros comuns às outras inativas, ou seja, aqueles do anel 2.

Tabela 8 - Análise SIMCA para o modelo de 3 componentes principais para as duas categorias.

CATEGORIA	NÚMERO DE COMPONENTES PRINCIPAIS	Nº TOTAL DE COMPOSTOS	COMPOSTOS CLASSIFICADOS INCORRETAMENTE
INATIVA (1)	3	11	compostos: 5, 7, 11
ATIVA (2)	3	9	

PORCENTAGEM DE CLASSIFICAÇÃO CORRETA: 85%

Ao diminuirmos 1 componente principal também na categoria 2 (modelo de 3 e 2 CP - tabela 9), todas as inativas são classificadas corretamente, mas 3 ativas (14, 17 e 19) agrupam-se com as inativas. A terceira CP da categoria 2 possui as seguintes variáveis mais importantes: def9, def10, def16, def24, drf9 e drf10, sendo que a maioria está relacionada com o anel (1). A retirada dos parâmetros relacionados com o anel (1), aproximou os compostos 5, 7 e 11 das inativas, anteriormente agrupado com as ativas, pois, esses compostos assemelham-se com as ativas devido aos substituintes no anel (1), principalmente; por outro lado, esses mesmos parâmetros do anel (1) são necessários para agrupar os compostos 14 e 19 com as outras ativas, que agora foram classificadas incorretamente.

Tabela 9 - Análise SIMCA para o modelo de 3 componentes principais para a categoria 1 (inativas) e 2 componentes principais para a categoria 2 (ativas).

Categoria	Número De Componentes Principais	Nº Total De Compostos	Compostos Classificados Incorretamente
INATIVA (1)	3	11	
ATIVA (2)	2	9	Compostos: 14, 17, 19

PORCENTAGEM DE CLASSIFICAÇÃO CORRETA: 85%

Com esses resultados, concluímos que a eliminação de variáveis relacionadas com o anel (1), apesar de levar à classificação correta dos compostos inativos 5, 7 e 11, não é adequada porque provoca a classificação incorreta dos compostos ativos. Assim, se os compostos referidos 5, 6, 7 e 10 se parecem com os ativos por causa do anel (1), mas não podemos eliminar essas variáveis, talvez a inclusão de variáveis relacionadas com o anel (2) ajude na classificação correta desses compostos.

Fazendo uso das observações descritas acima, selecionamos, além do logP, volume e das cargas sobre os átomos 2, 3 e 21, mais duas cargas relacionadas com o anel 2, C20 e C22. Os resultados estão dispostos nas figuras 21a, b e c.

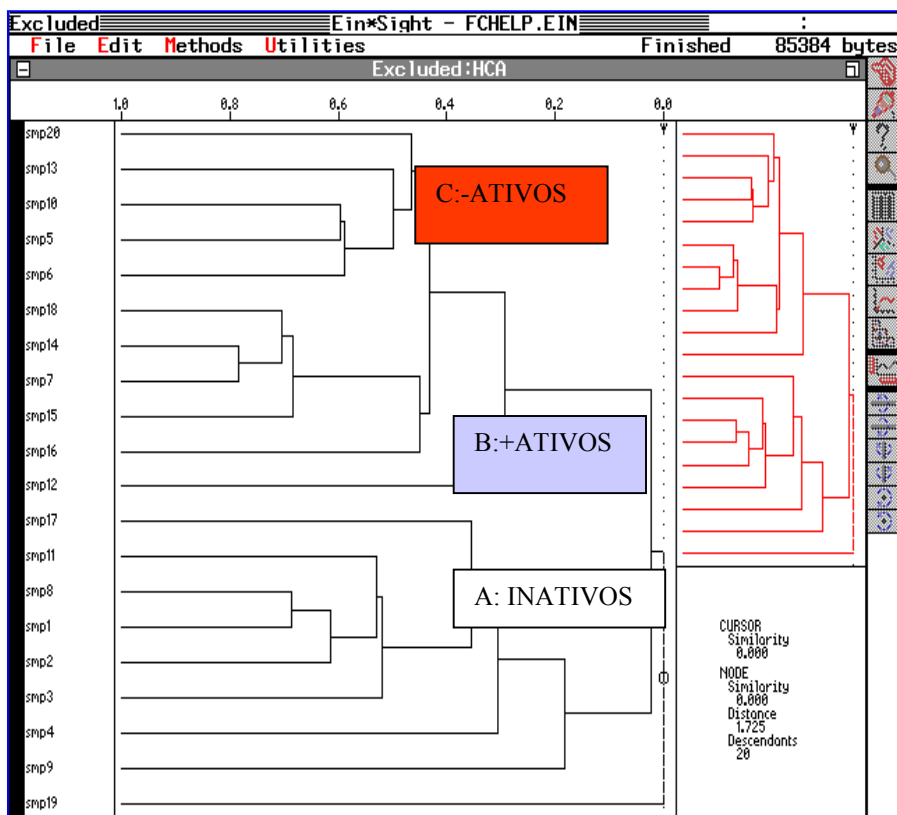


Figura 21 a- Dendrograma obtido para as neolignanas na conformação E, usando as seguintes variáveis: volume, logP, cargas sobre os átomos 2, 3, 20, 21 e 22.

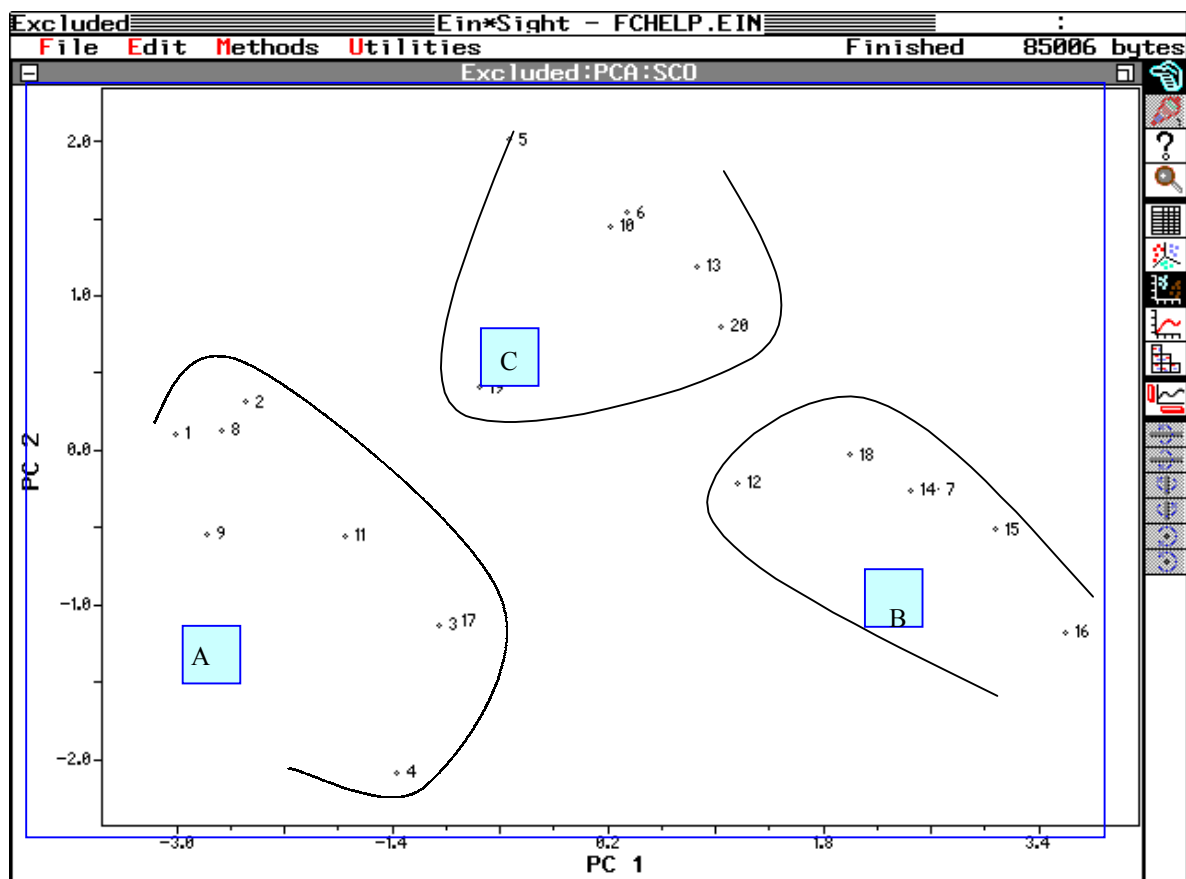


Figura 21 b - Análise de componentes principais para as neolignanas na conformação E, usando as variáveis descritas na figura 21 a.

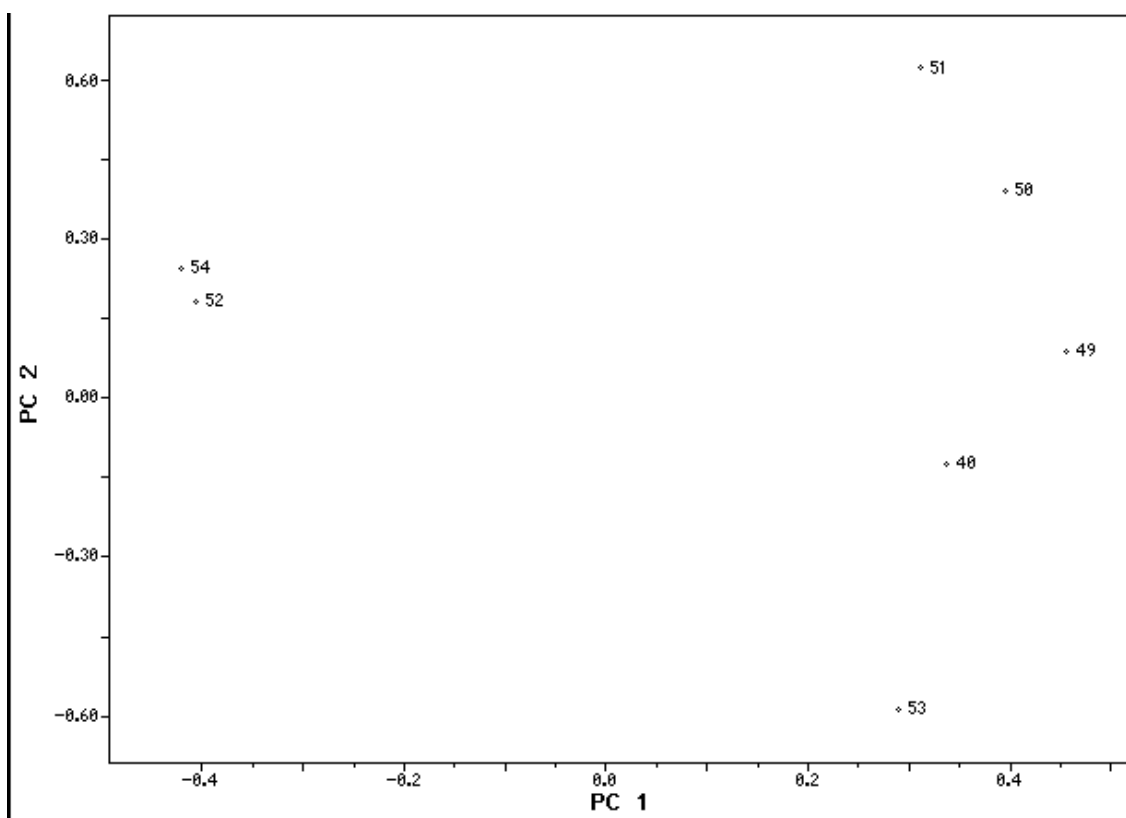


Figura 21 c - Gráfico dos "loadings" das duas primeiras componentes principais para as neolignanas na conformação E; as variáveis 40, 49, 50, 51, 52, 53 e 54 são: volume, logP, cargas sobre os átomos 2, 3, 20, 21 e 22.

Nesta análise, as figuras 21 a e b mostram que o composto 20 também foi agrupado com 5, 6 e 10, além do composto 13, conforme a análise anterior; este resultado concorda com os dados de AB, uma vez que o composto 20 e 13 possuem valores baixos e próximos de AB (0,8 e 14% de inibição a 30 μ M). Estes compostos estão no grupo C da figura 21b e são separados pela segunda CP, na qual as variáveis mais importantes são as cargas sobre os átomos 3 e 21 (variáveis 51 e 53). Todas as outras são importantes na primeira componente principal, que separa a maioria dos compostos ativos (grupo B) dos inativos (grupo A).

Segundo esta análise, a seguinte ordem de AB foi estabelecida:

18, 12, 14, 15 e 16 -----> mais ativos (grupo B)

13, 20, 19 -----> menos ativos (grupo C)

5, 6, 10-----> “quase ativos” (grupo C)

1, 2, 3, 4, 8, 9 e 11 -----> inativos (grupo A)

7 e 17 -----> classificados incorretamente.

5.4 - CONCLUSÕES

Com a presente análise, realizada com o conjunto treinamento das neolignanas na conformação E, concluímos que as variáveis mais importantes na classificação dos compostos, são aquelas relacionadas com o efeito estérico e hidrofobicidade das moléculas, como volume e logP e também os parâmetros eletrônicos dos átomos dos anéis aromáticos, como as cargas. Esses resultados nos dão duas importantes indicações sobre a atividade das neolignanas: 1) a interação com o receptor deve ocorrer também através dos anéis aromáticos, e não necessariamente pelos heteroátomos, como havíamos pensado inicialmente. 2) O volume da molécula e o coeficiente de partição octanol-água, sendo fatores importantes nesta análise, indicam que no geral, a atividade anti-leishmaniose das neolignanas está relacionada com o aumento de volume e de hidrofobicidade, pois as neolignanas mais ativas apresentam valores de volume e logP maiores.

Embora os resultados tenham melhorado significativamente em relação àqueles da referência 1, antes de prosseguirmos o trabalho, decidimos verificar o comportamento dos métodos de RP aplicados em SAR, frente a outro método estatístico comumente usado em QSAR - a análise de regressão linear múltipla (RLM). Essa comparação foi feita a fim de investigarmos o motivo de alguns compostos continuarem sendo classificados incorretamente com RP.

CAPÍTULO 6

RLM x ACP

A análise de componentes principais (ACP) é um método de Reconhecimento de Padrões bastante popular e tem sido utilizado com frequência em estudos de relações estrutura-atividade (SAR) [63]. Os métodos de RP são particularmente importantes em SAR, quando valores numéricos sobre a atividade biológica não são disponíveis. Nesses casos, é necessário usar um método capaz de classificar os compostos em ativos ou inativos, o que pode ser feito com o uso de ACP.

A análise de regressão linear múltipla (RLM) é um método estatístico muito usado em estudos de relações quantitativas entre estrutura química e atividade biológica (QSAR). Ele relaciona propriedades físico-químicas com valores de atividade biológica de uma série de compostos.

Neste trabalho, escolhemos os métodos de Reconhecimento de Padrões, por não termos os valores de atividade biológica, como já explicamos anteriormente. No capítulo 5, discutimos os resultados obtidos para as neolignanas (tabela 2), usando inicialmente, cerca de 50 parâmetros físico-químicos. Comparamos com resultados obtidos num trabalho anterior, onde apenas 10 variáveis haviam sido utilizadas, e verificamos que a inclusão de parâmetros hidrofóbicos e estéricos foi importante para a melhora dos resultados. A molécula 17 foi classificada incorretamente por apresentar estrutura bastante parecida com as inativas. Como foram utilizados parâmetros dos três tipos citados e usados com frequência em QSAR/SAR, pensamos que a classificação incorreta de alguns compostos pudesse ser uma limitação do método estatístico aplicado.

Com o objetivo de verificarmos a confiabilidade da análise de componentes principais em SAR, aplicamos este método a uma série de compostos à qual já havia sido realizada uma análise de QSAR, com o método de regressão linear múltipla. Esses compostos não apresentam nenhuma relação com as neolignanas ou com a leishmaniose; foram usados apenas como objeto de estudo dos métodos estatísticos citados. Por isso, este capítulo será descrito como um trabalho separado, contendo os seguintes tópicos:

Introdução, Cálculos, Resultados e Discussão. Este trabalho foi também publicado na referência 64.

6.1 INTRODUÇÃO:

Os derivados da 1,4-diidropiridina (DHP) - figura 22 - são conhecidos por antagonistas de cálcio. Eles têm sido usados no tratamento de doenças cardiovasculares como angina, alguns tipos de arritmias cardíacas, hipertensão e outras. Essas drogas agem diretamente nos canais de cálcio localizados na membrana celular, bloqueando o fluxo de íons cálcio (Ca^{2+}) do meio extracelular para o citoplasma celular [65]. Nos últimos dez anos, os derivados DHP têm sido objeto de um número de estudos QSAR onde parâmetros clássicos e derivados de química quântica, têm sido usados [66,67].

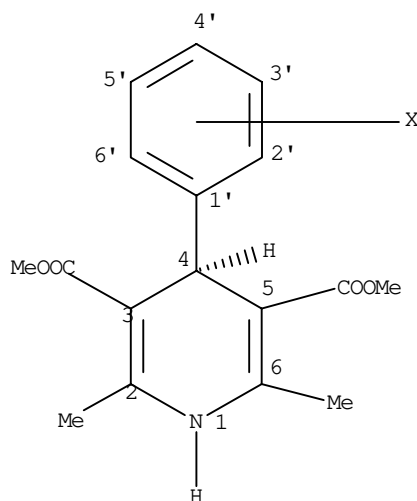


Figura 22 - Estrutura básica dos derivados da DHP

Gaudio e colaboradores realizaram estudos de QSAR em uma série de 45 derivados DHP, através da análise RLM e usaram parâmetros clássicos e químico-quânticos [68]. Os parâmetros calculados usados por Gaudio e colaboradores foram: carga atômica líquida (C_n), momento de dipolo total (μ), de carga (μ_c) e de hibridização (μ_H), dureza (η), polarizabilidade molecular (α), índices de fronteira como densidade eletrônica de fronteira (def), densidade orbital de fronteira (dof) e densidade radical de fronteira (drf), energia da barreira rotacional do anel fenílico (E_{rot}) e o volume de van der Waals (V_{vdw}).

Parâmetros clássicos como constante hidrofóbica de Hansch (π), constante eletrônica de Hammet (σ) e parâmetros de Verloop (B_1 e L) foram também usados. Os parâmetros químico-quânticos foram calculados usando o método AM1 [66]. As tabelas 10 e 11 mostram esses parâmetros. Baseando-se nas equações RLM obtidas, eles estimaram a natureza das interações droga-receptor.

Cada equação RLM, tal como Eq. (6.1), é uma combinação linear dos parâmetros físico-químicos ($p_1, p_2, \dots$):

$$\log (1/C) = c_0 + c_1 \cdot p_1 + c_2 \cdot p_2 + c_3 \cdot p_3 \dots \quad \text{eq. 6.1}$$

onde C é a concentração molar da droga, necessária para produzir uma determinada resposta biológica e $c_0, c_1, c_2, \dots$ são constantes determinadas pelo método RLM. A interpretação de uma equação QSAR é baseada no significado físico dos parâmetros ($p_1, p_2, \dots$), bem como na magnitude e sinal das constantes ($c_1, c_2, \dots$).

Para comparar os dois métodos, aplicamos a análise de componentes principais a exatamente o mesmo conjunto de 45 derivados de DHP, estudado por RLM. O principal objetivo foi verificar se o método ACP classificaria os compostos em "mais ativos" e "menos ativos", uma vez que neste caso todos os derivados apresentavam atividade biológica. Apesar dos dois métodos estatísticos serem completamente diferentes, ambos podem ser usados para relacionar atividade biológica com parâmetros físico-químicos.

Tabela 10 - Dados biológicos expressos como $\log (1/IC_{50})$, polarizabilidade molecular (α), largura mínima de Verloop (B_1) e comprimento (L) dos substituintes, e barreira rotacional (E_{rot}) para o conjunto de derivados DHP.

Nº	COMPOSTO	$\log 1/IC_{50}$	α	B_1	L	E_{rot}
1	3'-Br	8,89	26,53	1,95	3,83	1,66
2	2'-CF ₃	8,82	26,43	1,98	3,30	9,73
3	2'-Cl	8,66	25,75	1,80	3,52	4,73
4	3'-NO ₂	8,40	27,45	1,70	3,44	2,05
5	2'-CH=CH ₂	8,35	28,00	1,60	4,29	6,36
6	2'- NO ₂	8,29	27,09	1,70	3,44	6,54
7	2'-Me	8,22	26,24	1,52	3,00	6,67
8	2'-Et	8,19	27,49	1,52	4,11	7,41
9	2'- Br	8,12	26,10	1,95	3,83	5,60
10	2'-CN	7,80	26,80	1,60	4,23	5,82
11	3'- Cl	7,80	26,06	1,80	3,52	1,58
12	3'-F	7,68	25,27	1,35	2,65	1,97
13	H	7,68	24,98	1,00	2,06	1,48
14	3'-CN	7,46	27,23	1,60	4,23	1,71
15	3'-I	7,38	26,99	2,15	4,23	1,65
16	2'-F	7,37	25,16	1,35	2,65	4,80
17	2'-I	7,33	26,51	1,98	4,23	6,20
18	2'-OMe	7,24	27,12	1,52	3,98	3,88
19	3'- CF ₃	7,13	26,62	1,35	3,30	1,80
20	3'-Me	6,96	26,34	1,35	3,00	1,49
21	2'-OEt	6,96	28,47	1,50	4,92	5,27
22	3'-OMe	6,72	27,34	1,35	3,98	1,61
23	3'-NMe ₂	6,05	29,25	1,50	3,53	1,67
24	3'-OH	6,00	25,72	1,35	2,74	1,41
25	3'-NH ₂	5,70	26,39	1,50	2,93	1,54

26	3'-OAc	5,22	29,06	1,35	4,87	1,51
27	3'-OCOPh	5,20	36,18	1,70	8,15	1,57
28	2'-NH ₂	4,40	26,24	1,50	2,93	5,82
29	4'-F	6,89	25,36	1,35	2,65	1,52
30	4'-Br	5,40	26,70	1,95	3,83	1,43
31	4'-I	4,64	27,16	2,15	4,23	1,40
32	4'-NO ₂	5,50	27,68	1,70	3,44	1,40
33	4'-NMe ₂	4,00	29,50	1,50	3,53	1,55
34	4'-CN	5,46	27,45	2,06	4,23	1,41
35	4'-Cl	5,09	26,19	1,80	3,52	1,47
36	2',6'-Cl ₂	8,72	26,65			9,24
37	F ₅	8,36	26,71			8,52
38	2'F,6'-Cl	8,12	26,04			9,44
39	2',3'-Cl ₂	7,72	26,87			4,99
40	2'-Cl,5'-NO ₂	7,52	28,39			5,25
41	3',5'-Cl ₂	7,03	27,13			1,63
42	2'-OH,5'-NO ₂	7,00	28,26			4,37
43	2',5'-Me ₂	7,00	27,63			6,66
44	2',4'-Cl ₂	6,40	27,04			4,71
45	2',4',5'-(OMe) ₃	3,00	31,98			3,84

Tabela 11 - Carga atômica (C), densidades eletrônica (def) e orbital (dof) de fronteira, e dureza molecular (η) para os 45 derivados DHP calculados com AM1.

Composto	C_3	C_6	def_4	dof_5	dof	η
1	-0,1683	-0,1075	0,00041	0,00347	0,00930	4,173
2	-0,0955	-0,1275	0,00013	0,04050	0,06001	4,199
3	-0,1323	-0,1111	0,00023	0,01681	0,03488	4,179
4	-0,1338	-0,0799	0,00028	0,0897	0,02145	4,151
5	-0,1270	-0,1192	0,00020	0,02520	0,16782	4,167
6	-0,0869	-0,1296	0,00020	0,14475	0,14228	4,147
7	-0,1393	-0,1150	0,00145	0,00772	0,01810	4,160
8	-0,1394	-0,1154	0,00036	0,19349	0,21433	4,161
9	-0,1132	-0,1185	0,00028	0,00315	0,01132	4,184
10	-0,0986	-0,1258	0,00027	0,00308	0,01104	4,178
11	-0,0647	-0,1182	0,00025	0,01375	0,02800	4,169
12	0,0841	-0,1352	0,00020	0,00987	0,02562	4,171
13	-0,1343	-0,1195	0,00015	0,03618	0,05323	4,164
14	-0,0228	-0,0961	0,00032	0,00443	0,01173	4,173
15	-0,2659	-0,1012	0,00028	0,01418	0,02820	4,169
16	-0,1743	-0,0952	0,00092	0,00411	0,01020	4,171
17	-0,1079	-0,1207	0,00030	0,00233	0,00805	4,198
18	-0,2109	-0,0879	0,00008	0,04871	0,20501	4,177
19	-0,1658	-0,0902	0,00019	0,00446	0,01043	4,177
20	-0,0756	-0,1223	0,00015	0,00411	0,01118	4,162
21	-0,2132	-0,0877	0,00020	0,00657	0,01875	4,177
22	0,0719	-0,1517	0,00024	0,01593	0,03053	4,163
23	0,0442	-0,1457	0,00025	0,00357	0,01220	3,972
24	0,0735	-0,1551	0,00012	0,00566	0,01268	4,161
25	0,0460	-0,1597	0,00010	0,00789	0,01632	3,979
26	0,0544	-0,1273	0,00009	0,02339	0,2173	4,166
27	0,554	-0,1252	0,00022	0,00345	0,01169	4,168
28	-0,2026	-0,0825	0,00035	0,00269	0,00952	3,903
29	-0,1689	-0,0962	0,00025	0,00328	0,1162	4,170

30	-0,1073	-0,1213	0,00028	0,00372	0,01019	4,167
31	-0,1013	-0,1237	0,01180	0,00417	0,01170	4,166
32	-0,0706	-0,1339	0,00007	0,01824	0,03209	4,121
33	-0,1553	-0,0970	0,20599	0,00156	0,00583	3,951
34	-0,0949	-0,1258	0,04055	0,00326	0,00924	4,167
35	-0,1290	-0,1112	0,27441	0,00268	0,00825	4,170
36	-0,1346	-0,0358	0,00014	0,15594	0,04693	4,134
37	0,0205	0,1041	0,00006	0,01948	0,01969	4,077
38	-0,1767	-0,0214	0,00026	0,00567	0,01092	4,131
39	-0,0674	-0,1110	0,00045	0,00337	0,01036	4,180
40	-0,1450	-0,0343	0,00049	0,00070	0,00745	3,980
41	-0,0578	-0,1148	0,00006	0,00509	0,01200	4,171
42	-0,2353	-0,0127	0,00018	0,00207	0,00739	4,027
43	-0,1357	-0,1136	0,06596	0,00150	0,0893	4,160
44	-0,1293	-0,1036	0,27320	0,00194	0,00782	4,177
45	-0,2580	-0,1327	0,26425	0,00217	0,01189	3,916

6.2 METODOLOGIA

Gaudio e colaboradores realizaram a análise de regressão linear sobre quatro grupos de compostos extraídos do conjunto de 45 derivados da DHP; o grupo I consistiu dos para-monoderivados, o grupo II dos orto-monoderivados, III dos orto e para-monoderivados e o IV, de todos os 45 compostos. Nós aplicamos ACP a exatamente os mesmos subconjuntos, em adição a um novo conjunto (V) que consistiu de alguns compostos selecionados. Com isso, pudemos comparar os resultados dos dois métodos: ACP vs. RLM. Usamos o pacote de programas computacionais denominado ARTHUR/UNICAMP, já citado [58]. Antes de aplicarmos ACP aos conjuntos de DHPs, foi necessário escalar os dados, conforme descrito no capítulo 2, item 2.2. Então, aplicamos ACP a cada conjunto de variáveis, inicialmente usando todas as variáveis, depois as dez mais importantes (segundo o Peso de Fisher calculado [34]), etc, até encontrarmos o melhor conjunto de variáveis para separar os compostos nas duas categorias: alta ou baixa atividade. O peso de Fisher foi calculado com o programa WEIGHT; os coeficientes de correlação entre as variáveis usadas em RLM e ACP foram calculados com o programa CORREL; o gráfico dos "scores" das duas primeiras componentes principais foi obtido a partir dos dados obtidos com o programa KARLOV, que realiza a análise de componentes principais; os programas WEIGHT, CORREL e KARLOV, são subprogramas do programa ARTHUR/UNICAMP.

6.3 RESULTADOS E DISCUSSÃO:

A primeira série estudada foi o conjunto (I) que consiste dos oito compostos para mono-substituídos; são os compostos 13 e 29-35 (tabela 10). Os valores de $\log (1/IC_{50})$ variam entre 7,55 (composto 13) e 4,00 (composto 33). Nós escolhemos os compostos 13 e 29 para formarem o conjunto de alta atividade e os seis remanescentes foram escolhidos para o conjunto de baixa atividade. A separação entre os grupos de alta e baixa atividade é superior a 1,39 unidades na escala de $\log (1/IC_{50})$. Assim, temos dois grupos claramente definidos. A questão é: "O método ACP é capaz de separar o grupo de alta atividade daquele de baixa atividade?" A figura 23 mostra os resultados da análise de componentes principais para o conjunto (I). Os compostos de alta atividade e de baixa atividade formam grupos distintamente separados na figura. As duas primeiras componentes principais (CP), que são obtidas pela combinação linear das variáveis (ou parâmetros físico-químicos) originais (conforme explicado no capítulo 2, item 2.2), estão descritas nas equações 6.2a e 6.2b em termos das variáveis mais importantes:

$$CP1 = -0,471 \alpha - 0,602 B_1 - 0,645L \quad (6.2a)$$

$$CP2 = +0,859 \alpha - 0,479 B_1 - 0,181L \quad (6.2b)$$

onde α é a polarizabilidade molecular, B_1 e L são a largura e comprimento mínimo de Verloop, respectivamente. A variância em CP1 é 78% e em CP2, 22%. A equação 6.2a mostra que, em valores absolutos, o coeficiente de L é o mais alto e, em seguida o de B_1 . Na equação 6.2b, α é a variável mais importante. Os dois grupos de compostos (alta e baixa atividade) estão melhor separados pela primeira CP, significando que L e B_1 são as principais variáveis usadas na classificação dos oito compostos nos grupos de alta e baixa atividade.

Os dois derivados DHPs para-substituídos (H e F) exibem alta atividade, enquanto os seis para-substituídos (NO_2 , CN, Br, Cl, I e NMe_2) exibem baixa atividade. Mahmoudian e Richards [67], trabalharam com um conjunto de cinco para-substituídos da DHP (H, F, Me, NO_2 e Cl); eles aplicaram RLM na análise e encontraram boa correlação entre a atividade biológica e as variáveis B_1 e L . Gaudio e colaboradores aplicaram RLM

aos oito para-substituídos do conjunto (I). Eles encontraram que o volume de van der Waals (V_{vdw}) possui alta correlação com a atividade biológica, mas nenhum dos parâmetros L , B_1 ou α apresentaram correlação significativa com a atividade biológica, quando RLM foi aplicada ao conjunto (1). Entretanto, existe boa correlação entre as variáveis L e B_1 e o volume, conforme a tabela 12.

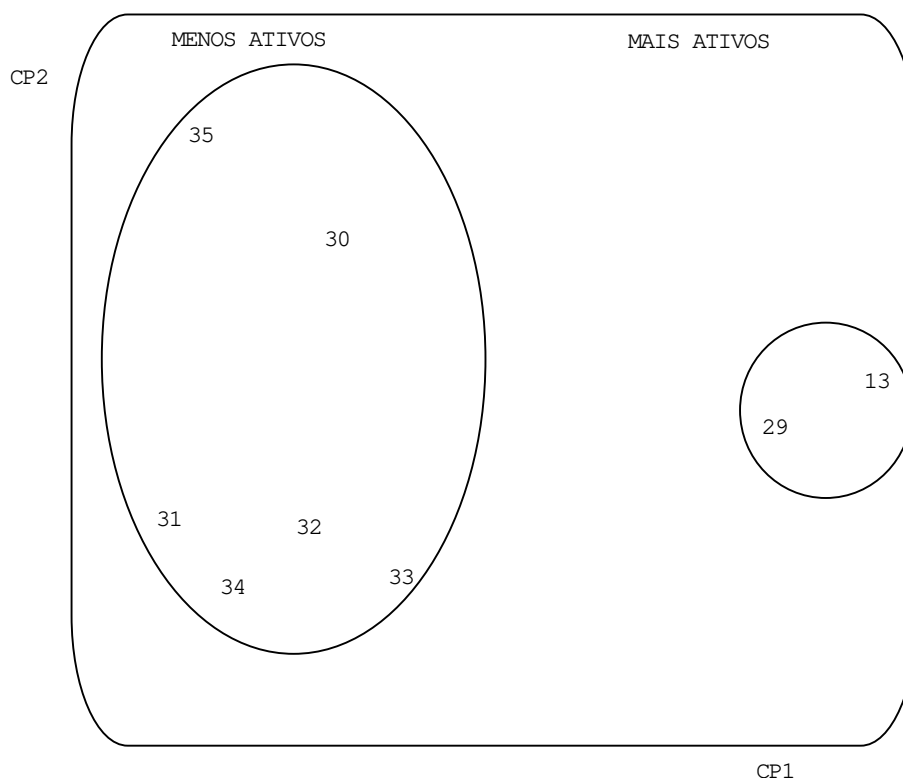


Figura 23 - Classificação do conjunto (I) de compostos para-derivados, nos grupos de alta e baixa atividade. Todos os compostos foram classificados corretamente.

Tabela 12- Matriz de correlação para o conjunto de variáveis usadas na ACP e na RLM para o conjunto (I) de compostos.

RLM	ACP		
	α	B ₁	L
V _{vdw}	0,903	0,442	0,622

As quatro variáveis da tabela 12 são estéricas. Portanto, podemos concluir que a atividade biológica das DHPs está relacionada com fatores estéricos.

A seguir, aplicamos ACP ao conjunto (II), dos orto-monoderivados. Gaudio e colaboradores usaram 13 orto derivados na análise RLM. Nós aplicamos ACP ao mesmo conjunto. Inicialmente, os 13 derivados foram divididos em dois conjuntos, de acordo com os valores da atividade biológica, de forma que os compostos 2, 3, 5, 6, 7, 8 e 9 (tabela 10) foram agrupados como de alta atividade e os seis derivados remanescentes (10, 13, 16, 17, 18 e 21) foram escolhidos como de baixa atividade. A figura 24 mostra os resultados da ACP para os 13 derivados. Onze compostos foram bem classificados, enquanto os compostos 10 e 17 foram classificados incorretamente. As duas componentes principais obtidas estão descritas nas equações 6.3 e 6.4, em termos das variáveis mais importantes:

$$CP1 = 0,64 C_{6'} - 0,64 C_{3'} - 0,43 E_{rot} \quad (6.3)$$

$$CP2 = 0,32 C_{6'} - 0,29 C_{3'} - 0,90 E_{rot} \quad (6.4)$$

onde, $C_{6'}$ e $C_{3'}$ são as cargas atômicas nas posições 6'e 3' do anel fenílico e E_{rot} é a energia da barreira rotacional do anel fenílico em torno do anel piridínico. $C_{6'}$ e $C_{3'}$ são parâmetros eletrônicos, enquanto E_{rot} é estérico.

A razão pela qual esses dois compostos (10 e 17) foram classificados incorretamente é desconhecida; entretanto, podemos estimar a razão. Os valores da atividade biológica dos 13 derivados da DHP variam entre 9,00 e 7,00. O critério para

divisão do conjunto (II) em duas categorias foi o seguinte: Os derivados com atividade biológica maior do que 8,00 foram agrupados como de alta atividade e aqueles que apresentaram valores menores do que 8,00 foram agrupados como de baixa atividade. Contudo, a diferença entre o composto 9 que pertence ao grupo de alta atividade e o composto 10 (de baixa atividade) é apenas 0,32 unidades na escala $\log (1/IC_{50})$. A diferença entre as duas categorias é muito pequena, não havendo uma clara distinção entre os valores de $\log (1/IC_{50})$. Provavelmente, isto causou a dificuldade na separação das duas categorias pelo método ACP.

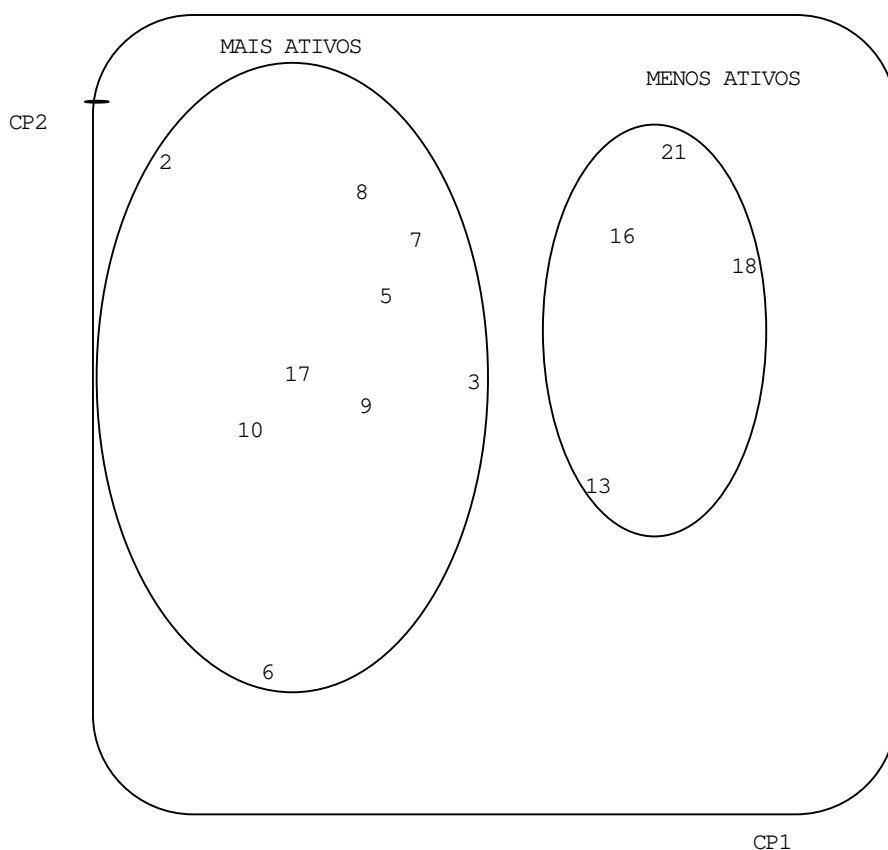


Figura 24 - Classificação do conjunto (II) de compostos (orto-derivados), nos grupos de alta e baixa atividade. Os compostos 10 e 17 foram mal classificados.

O método RLM selecionou as seguintes variáveis: π , C_3' e μ_H , duas das quais são diferentes daquelas usadas na ACP (Eq. 6.3 e 6.4). Somente a variável C_3' foi usada nos dois métodos. A tabela 11 mostra a matriz de correlação para os dois conjuntos de

variáveis. Com exceção de C_3' , a correlação entre as variáveis é baixa. Os dois métodos estatísticos selecionaram conjuntos diferentes de variáveis.

Tabela 13 - Matriz de correlação para o conjunto de variáveis usadas nas análises ACP e RLM para o conjunto (II) de compostos.

RLM	ACP		
	E_{rot}	C_6'	C_3'
π	0,317	0,424	0,467
C_3'	0,564	0,981	1,000
μ_H	0,038	0,268	0,322

O conjunto (III) consiste dos derivados orto e para. A maioria dos orto derivados apresentam valores de $\log(1/IC_{50})$ maiores do que 7,00. O composto 28 é o único orto-derivado com valor mais baixo do que 4,00. Por outro lado, a maioria dos para derivados apresentam valores de $\log(1/IC_{50})$ mais baixos do que 5,50. O composto 29 é o único para derivado com valor de 6,89. Os derivados com valor de $\log(1/IC_{50})$ igual ou maior do que 6,89 foram agrupados como de alta atividade e aqueles com atividade igual ou menor do que 5,50 foram agrupados como os de baixa atividade. Os resultados foram plotados na figura 25. A separação entre os dois grupos é clara, mas os compostos 13 e 29 foram classificados incorretamente. Eles deveriam pertencer ao grupo de alta atividade. As duas primeiras componentes principais obtidas foram:

$$CP1 = -0,417 \eta - 0,650 \text{dof}_2' - 0,636 E_{rot} \quad (6.5)$$

$$CP2 = 0,907 \eta - 0,247 \text{dof}_2' - 0,341 E_{rot} \quad (6.6)$$

A variância na CP1 é 49,0 % e na CP2 é 30,1% . Estas equações mostram que dof_2' , E_{rot} e η são variáveis importantes na classificação; dof_2' e η são parâmetros eletrônicos, e E_{rot} é estérico.

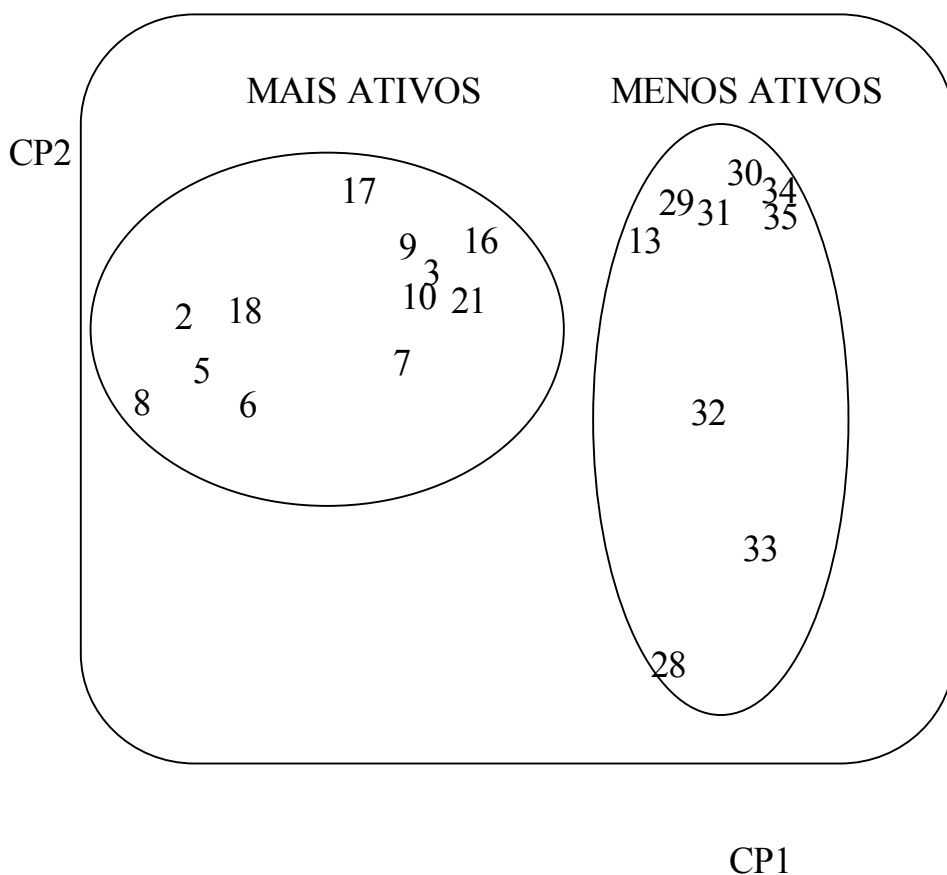


Figura25 - Classificação do conjunto (III) nos grupos de alta e baixa atividade. Os compostos 13 e 29 foram mal classificados.

A tabela 14 é a matriz de correlação para as três variáveis usadas em ACP e para as quatro usadas na análise RLM para o mesmo conjunto de compostos. Os coeficientes de correlação na matriz são todos baixos (menores do que 0,50), indicando que a correlação entre as variáveis dos dois grupos é baixa.

Tabela 14 - Matriz de correlação para o conjunto de variáveis usadas em ACP e em RLM, para o conjunto (III) de compostos.

RLM	ACP		
	η	dof _{2'}	E _{rot}
dof _{3'}	0,12	0,47	0,16
C _{3'}	0,33	0,12	0,03
μ_H	0,16	0,40	0,13
B ₁	0,22	0,25	0,06

O conjunto (IV) consiste de todos os 45 compostos. Os compostos cujos valores de $\log (1/IC_{50})$ são iguais ou maiores do que 6,72 foram agrupados como de alta atividade; estes são os compostos 1-22, 29 e 36-43. Os compostos com valores de atividade igual ou menor do que 6,40 foram agrupados como de baixa atividade. Esses são os compostos 23-28, 30-35, 44 e 45. As duas primeiras componentes principais, CP1 e CP2 geradas, estão nas equações 6.7 e 6.8.

$$CP1 = 0,67 \eta - 0,71 \text{dof}_{4'} + 0,24 E_{\text{rot}} \quad (6.7)$$

$$CP2 = -0,33 \eta - 0,01 \text{dof}_{4'} + 0,94 E_{\text{rot}} \quad (6.8)$$

A variância na CP1 é 43% e na CP2 é 33%. O índice de fronteira $\text{dof}_{4'}$ e a dureza η são significantes na equação 6.7, enquanto a energia rotacional é mais importante na equação 6.8. A figura 26 mostra a separação do conjunto (IV), de acordo com ACP. A maioria dos 45 compostos está agrupada na região central ao longo do eixo CP1. A separação entre os dois grupos não é muito clara. Entretanto, podemos dividir o espaço aproximadamente nas regiões de alta e baixa atividade, como na figura 26. Os seis compostos de baixa atividade, 24, 26, 27, 28, 30 e 31 e os dois de atividade alta, 40 e 42 foram classificados incorretamente. Assim, um total de oito compostos foram classificados

incorretamente, o que corresponde a 18% do total de 45 compostos. Esta porcentagem alta de classificação incorreta deve-se à pequena diferença entre os valores de atividade do grupo de alta e de baixa atividade; essa diferença é de apenas 0,5 unidades de $\log(1/IC_{50})$.

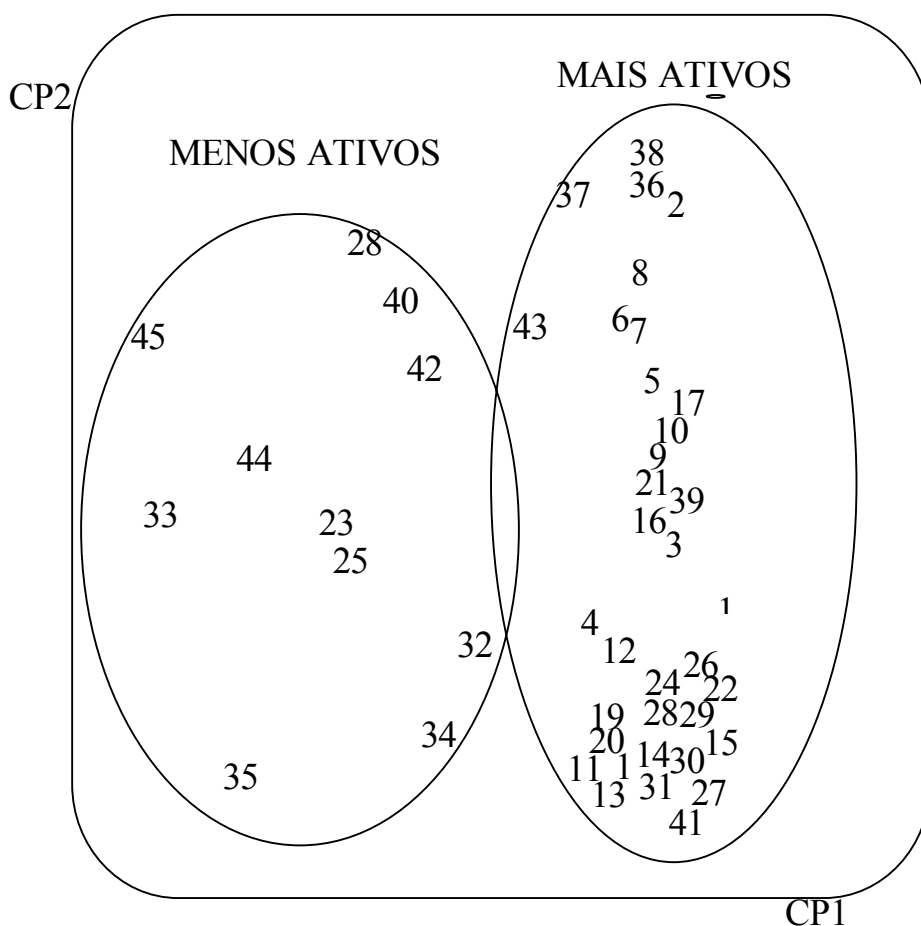


Figura 26 - Classificação do conjunto (IV) de compostos (poli e mono derivados) nos grupos de alta e baixa atividade. Os compostos 24, 26, 27, 28, 30, 31, 40 e 42 foram classificados incorretamente.

A tabela 15 mostra a matriz de correlação para as variáveis usadas em ACP e em RLM.

Tabela 15 - Matriz de correlação para o conjunto de variáveis usadas em ACP e em RLM para o conjunto (IV) de compostos.

RLM	ACP		
	E_{rot}	η	def_4
π	0,30	0,46	0,10
σ_m	0,31	0,31	0,03
def_5	0,07	0,49	0,62
V_{vdw}	0,16	0,21	0,28
B_1	0,49	0,38	0,23
E_{rot}	1,0	0,10	0,10

Os coeficientes de correlação são baixos, com exceção da E_{rot} que foi a única variável usada nas duas análises.

Com o objetivo de verificar se é possível melhorar a separação dos compostos pelo uso de ACP, nós construímos dois conjuntos bem separados, selecionando 27 compostos entre os 45. Chamamos este conjunto de (V). O grupo de atividade alta consiste de 14 compostos: 1-11 e 36-38, com valores de $\log(1/IC_{50})$ entre 8,89 e 7,80. O grupo de atividade baixa consiste de 13 compostos: 23-28, 30-35 e 45 que apresenta valores de $\log(1/IC_{50})$ entre 6,05 e 3,0. O resultado da ACP é mostrado na figura 27. Os grupos de alta e baixa atividade foram separados satisfatoriamente. Apenas três compostos (1, 4 e 11) foram classificados incorretamente. Agora, esse erro representa apenas 11% ao invés de 18%. A comparação entre as figuras 27 e 26 mostra que a separação entre os grupos de alta e baixa atividade foi significativamente melhorada para o conjunto (V), na figura 27.

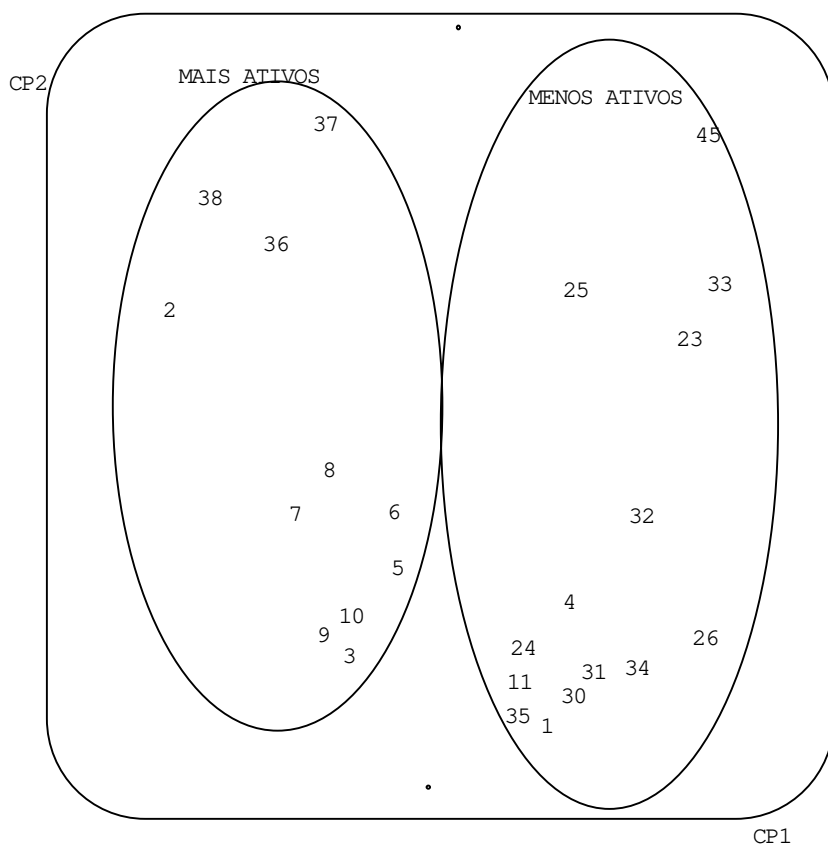


Figura 27- Classificação do conjunto (V) nos dois grupos (alta e baixa atividade); os compostos 1, 4 e 11 foram mal classificados.

As equações 6.9 e 6.10 mostram as duas componentes principais para o conjunto (V).

$$CP1 = -0,33 \eta - 0,68 \text{ dof}_{5'} - 0,66 E_{\text{rot}} \quad \text{eq. 6.9}$$

$$CP2 = 0,94 \eta - 0,15 \text{ dof}_{5'} - 0,31 E_{\text{rot}} \quad \text{eq. 6.10}$$

A variância na CP1 é 52% e na CP2 é 31%.

6.4 - CONCLUSÕES:

O método de ACP mostrou-se capaz de classificar os derivados DHP em dois grupos: alta e baixa atividade, embora em alguns casos tenha ocorrido a classificação incorreta de alguns compostos. ACP foi aplicada a 4 subconjuntos idênticos aos estudados por RLM, tendo apresentado bons resultados.

A diferença na escolha das variáveis pelos dois métodos estatísticos, pode ser devido ao diferente propósito de RLM e ACP. Embora ambos correlacionem atividade biológica com parâmetros físico-químicos, RLM é adequada para o estudo de relações quantitativas entre atividade biológica e estrutura-química (QSAR), enquanto ACP é adequado para a classificação de uma determinada série de compostos em categorias, o que pode ser chamado de relações (qualitativas) entre estrutura química e atividade biológica (SAR).

A conclusão mais importante para o nosso objetivo, que é verificar a aplicabilidade dos métodos de RP em SAR, é a de que ACP foi capaz de separar os compostos nas duas categorias e que é comum ocorrer alguma porcentagem de erro na classificação, o que não prejudica a análise de forma geral se pudermos entender os motivos que estão levando a esse erro. Neste caso, que tínhamos valores de atividade biológica, verificamos erros na classificação dos compostos com valores muito próximos de atividade, quando estes foram forçosamente separados em mais ativos e menos ativos. No caso das neolignananas, no capítulo 5, apesar de não termos os valores de AB, construímos uma tabela indicando uma ordem relativa de atividade, baseando-nos nos resultados da análise ACP.

Analogamente aos resultados obtidos com a análise ACP aplicada às DHP, concluímos que as neolignananas em posição intermediária no gráfico dos “scores” podem realmente formar um terceiro conjunto, nem muito ativo, nem muito inativo, como havíamos pensado. Assim, a classificação dessas neolignananas pode estar correta e as atividades delas devem apresentar valores próximos.

CAPÍTULO 7

ANÁLISE CONFORMACIONAL PARA AS NEOLIGNANAS SOLVATADAS EM ÁGUA.

7.1 - INTRODUÇÃO

No capítulo sobre a metodologia (capítulo 2, item 2.1), abordamos a análise conformacional de fármacos; estes, normalmente são moléculas flexíveis com uma ou mais ligações simples sobre as quais, pode ocorrer rotação livre. Técnicas experimentais como cristalografia podem determinar somente a conformação mais estável; por outro lado, cálculos de otimização de geometria podem produzir várias conformações de mínima energia (mínimos locais) e, considerando que o fármaco pode ajustar-se ao receptor no momento da interação com o mesmo, qualquer um desses mínimos locais poderá ser a conformação ativa.

Em trabalhos anteriores [5,6], mapas conformacionais foram produzidos para uma melhor visualização das regiões de mínima energia encontradas; um exemplo desses mapas é mostrado na figura 28 para a neolignana ativa, 15. Os mínimos locais são denominados A, B, C, ...H, de acordo com os ângulos torsionais t_1 e t_2 da figura 4a. Para todas as neolignanas estudadas, o mínimo local A corresponde a t_1 e t_2 próximos a 60° ; o mínimo local B corresponde a t_1 próximo a 150° e t_2 próximo a 60° . Apenas para as moléculas 15 e 16 foi encontrado um mínimo a mais, denominado I, devido à estrutura básica diferente das mesmas, pois estas são as únicas neolignanas estudadas que não apresentam uma carbonila no esqueleto.

Os mapas conformacionais foram obtidos, plotando-se os valores de energia *versus* os ângulos torsionais t_1 e t_2 e correspondem às superfícies de energia potencial para cada molécula, sendo constituídos de “vales” e “montanhas”. Os “vales” são justamente as conformações de mínima energia, ou simplesmente mínimos locais.

Uma análise rápida dos mapas poderia nos levar a pensar que os mínimos locais correspondentes aos “vales” mais profundos apresentariam maior probabilidade de ser a conformação ativa, uma vez que apresentando energia mais baixa, são mais estáveis.

Entretanto, existem mínimos locais correspondentes a “vales” mais rasos, porém mais largos, que são mais prováveis do que aqueles profundos e estreitos; isso ocorre, porque os “vales” rasos e largos podem apresentar um número maior de conformações possíveis, do que aqueles profundos e estreitos. Para decidirmos quais mínimos locais seriam mais prováveis, utilizamos uma metodologia baseada na distribuição de Boltzmann, a qual foi brevemente explicada no capítulo 5, item 5.2.

Na figura 29 estão mostradas todas as conformações encontradas para a maioria das neolignanas; essa análise conformacional foi realizada para as neolignanas isoladas. Os resultados obtidos nas referências 5 e 6, usando-se constante dielétrica para vácuo e meio biológico, nos mostraram que de maneira geral, as conformações que apresentam os anéis fenílicos mais próximos entre si (B e G) ou numa posição intermediária (C, E, F, H) são favorecidas tanto no vácuo quanto no meio biológico. Entretanto, em meio biológico houve um acréscimo das conformações em que os anéis estão mais próximos. As conformações A e D são aquelas que apresentam os anéis fenílicos mais afastados.

Figura 28 - Mapa conformacional produzido para a molécula 15.

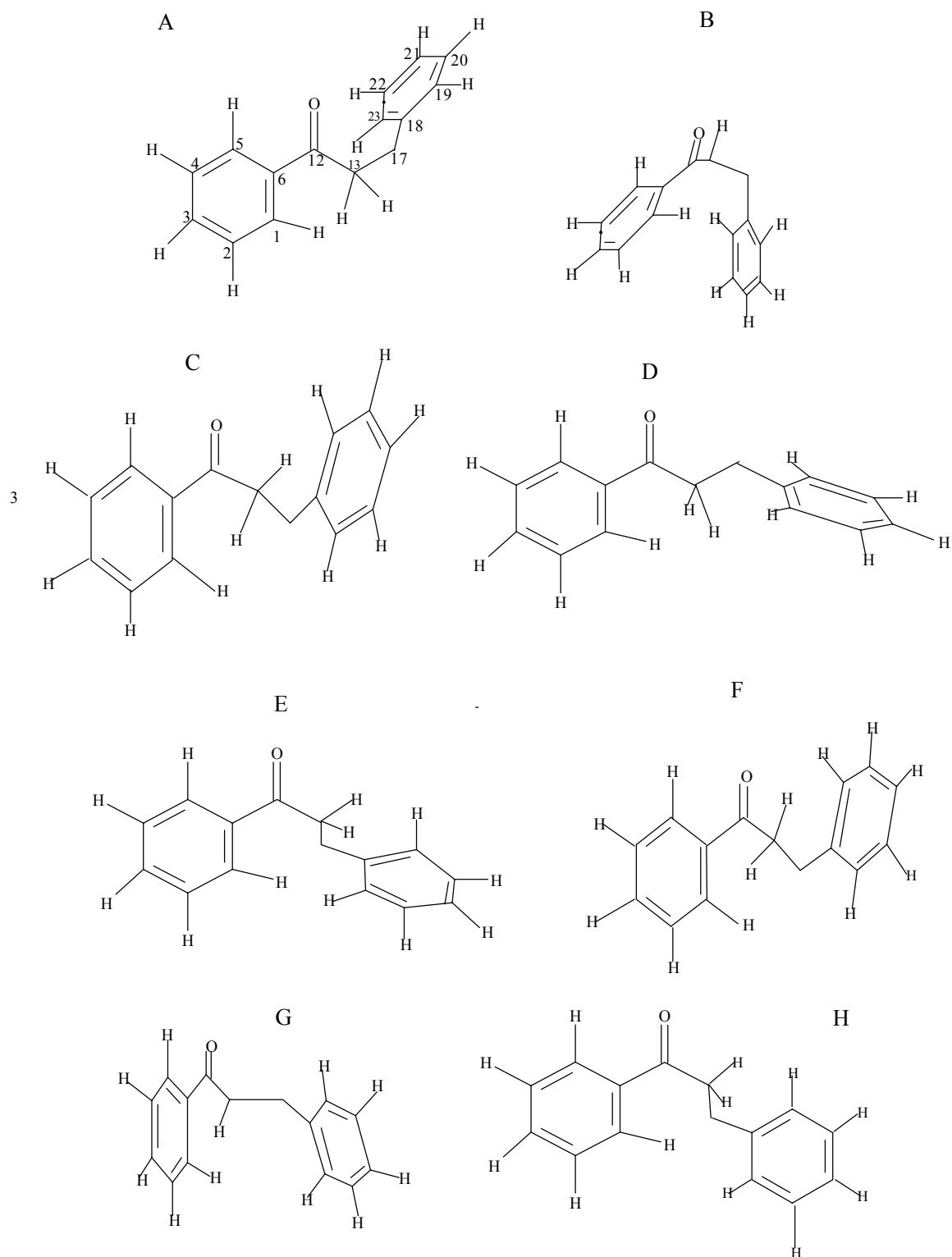


Figura 29- Oito mínimos locais encontrados na análise conformacional realizada para as neolignanas isoladas.

As conformações encontradas para as moléculas ativas e inativas são mostradas nas tabelas 16 e 17, em ordem decrescente de estabilidade para as inativas e de probabilidade para as ativas. A estabilidade foi determinada pelo cálculo de energia realizado com o método de mecânica molecular, MM2, nas referências 5 e 6; a probabilidade foi calculada nos mesmos trabalhos, usando a metodologia baseada na distribuição de Boltzmann, já citada. Não haveria sentido em calcularmos a probabilidade para as conformações presentes nas moléculas inativas, uma vez que trata-se da conformação ativa mais provável; por isso, foram produzidos mapas conformacionais apenas para as moléculas ativas e também apenas para estas, calculamos as probabilidades.

Tabela 16 - Ordem decrescente de estabilidade para as neolignanas inativas.

MOLÉCULA	ORDEM DECRESCENTE DE ESTABILIDADE, NO VÁCUO
1	$E > C > A > B > D$
2	$E > C > A > B > D$
3	$E > C > D > A > B$
4	$F > H > C > B > E > G > A > D$
5	$E > C > B > H > F = A > D > G$
6	$F = C > B > E > H > A > D > G$
7	$C > A > E > D > H > F$
8	$H > F > C > A = E$
9	$E = H > D > C > A$
10	$B > E > F > A > C > D > H > G$
11	$B > C > E > A > D$

Tabela 17- Ordem decrescente de estabilidade e probabilidade para as moléculas ativas. Não calculamos os valores de probabilidade para o composto 16, pois devido ao grande número de átomos presentes, não foi possível fazermos uma análise conformacional muito detalhada.

Molécula	Mínimo local	Estabilidade	Probabilidade
12	A, C, D, E	E>D>C>A	E>D>C>A
13	A, C, D, E	E>D>A>C	E>D>A>C
14	A, C, D, E, F, H	H>E>F>C>D>A	H>C=E>F>D>A
15	A, B, C, D, E, F, G, H, I	B>H>E>C>D>A>F>G>I	B>>E>H>C>A=D>F=G=I
16	A, B, C, D, E, F, G, H, I	B>E>H>C>D>A>F>G>I	-----
17	B, C, D, E, F, G	B>F>E>D>C>G	B>F>E>D>C>G
18	B, C, D, E, F, G, H	B>E>F>D>C=G>H	B>F>E>D=C=G=H
19	A, B, C, D, E, F, G, H	B>E>C>A>F>H>D>G	B>C>E>A>F>H=D>G
20	B, C, D, E, F, G, H	B>E>F>D>C=G>H	B>E>F>C=D=G=H

Na procura da conformação ativa, consideramos duas possibilidades: 1) Se existir algum mínimo local presente apenas nas moléculas inativas e ausente nas ativas, este poderá ser descartado, pois a conformação ativa deve ser uma daquelas pertencentes às moléculas ativas; 2) Se existir algum mínimo local que não está presente em todas as ativas, este também será descartado, pois a conformação ativa, sendo a mesma para todas as neolignanas, deve pertencer a todas que apresentarem atividade.

Verificamos nas tabelas 16 e 17 que não há nenhuma conformação ausente em todas as ativas. Entretanto, podemos ver que apenas alguns mínimos locais estão presentes em todas as ativas; estes são as conformações C, D, E e F (nas moléculas XII e XIII, a conformação F é um enantiômero da conformação C e, por este motivo, a conformação F não aparece na tabela 17 para essas moléculas).

Nos trabalhos realizados anteriormente [5, 6], concluímos que uma dessas deve ser a conformação ativa. A conformação E, além de estar presente em todas as ativas, apresenta alta probabilidade em todas elas; assim, havíamos escolhido esse mínimo local para realizar os cálculos dos parâmetros físico-químicos e das relações estrutura-atividade, como discutimos no capítulo 5.

Continuando essa análise conformacional no presente trabalho, realizamos cálculos de interação soluto-solvente para as neolignanas solvatadas por moléculas de água, com o objetivo de aprimorarmos a simulação do meio biológico e assim, de obtermos mais informações sobre a conformação ativa. Esses cálculos estão descritos no capítulo 4 (item 4.3).

7.2 - RESULTADOS E DISCUSSÃO

A solvatação em água foi realizada para algumas neolignanas ativas e inativas, em todas as conformações encontradas nos trabalhos anteriores [5,6]: A, B, ...I. Na tabela 18 estão dispostos os resultados de energia de interação soluto-solvente (neolignana-água, neste caso) e também os valores de energia estérica calculados com MM2, nas referências 5 e 6, para comparação.

Tabela 18 - A energia estérica (E_{est}) foi calculada com o método MM2 para os compostos isolados e a energia de interação soluto-solvente $E_{\text{st-sv}}$ foi calculada com o método de Monte Carlo, para os compostos solvatados em água. Os valores mais baixos de energia (E_{est} e $E_{\text{st-sv}}$) foram sublinhados para comparação.

MOLÉCULA	CONFORMAÇÃO	E_{est} (kcal)	$E_{\text{st-sv}}$ (kcal)
1	A	3,3	-38,6
	B	4,0	-39,6
	C	3,0	-37,1
	D	4,7	<u>-43,0</u>
	E	<u>2,9</u>	-35,1
3	A	11,0	<u>-53,2</u>
	B	11,4	-38,9
	C	9,9	-39,4
	D	10,8	-45,7
	E	<u>8,3</u>	-43,2
4	A	7,6	<u>-48,0</u>
	B	7,0	-43,8
	C	7,0	-36,2
	D	8,5	-40,9
	E	7,0	-36,9
	F	<u>6,5</u>	-40,2
	G	7,6	-42,0
	H	7,0	-35,1
5	A	15,1	-54,3
	B	<u>14,7</u>	<u>-105,3</u>
	C	15,1	-96,2
	D	16,7	-88,6
	E	14,9	-83,0
	F	15,9	-91,8
	G	17,5	-102,1

	H	16,5	-75,8
6	A	19,4	-59,3
	B	<u>17,4</u>	-57,9
	C	17,6	-48,5
	D	20,6	<u>-60,5</u>
	E	18,9	-57,6
	F	18,5	-50,9
	G	21,2	-49,9
	H	21,0	-56,2
7	A	20,6	<u>-66,8</u>
	C	<u>20,0</u>	-53,9
	D	24,6	-59,5
	E	20,1	-64,6
	F	35,6	-58,6
	H	25,1	-62,2
9	A	4,7	-44,1
	C	4,6	-32,8
	D	4,4	<u>-44,7</u>
	E	<u>4,0</u>	-38,8
	H	4,1	-39,5
12	A	26,2	-58,1
	C	25,9	-54,0
	D	25,7	<u>-61,1</u>
	E	<u>25,3</u>	-56,7
13	A	32,2	-63,5
	C	32,4	-57,3
	D	31,9	<u>-67,0</u>
	E	<u>31,5</u>	-65,2
14	A	18,9	-60,2
	C	18,4	-56,6

	D	18,6	<u>-69,8</u>
	E	17,8	-59,0
	F	18,3	-56,0
	H	<u>17,7</u>	-57,1
15	A	20,6	<u>-125,7</u>
	B	<u>17,9</u>	-51,0
	C	19,8	-57,4
	D	20,3	-54,9
	E	19,6	-57,5
	F	25,1	-50,1
	G	27,0	-52,8
	H	19,4	-50,2
	I	27,5	-54,2
16	A	35,9	<u>-59,9</u>
	B	<u>32,4</u>	-55,7
	C	35,1	-53,7
	D	35,7	-55,8
	E	34,0	-54,8
	F	40,2	-57,2
	G	41,1	-52,2
	H	34,7	-59,8
	I	41,7	-59,8

Na tabela 18, vemos que as conformações A e D foram favorecidas para a maioria das doze moléculas estudadas, sendo que apenas uma (inativa) apresentou o mínimo B como mais estável. Comparando com os resultados obtidos com MM2, para as moléculas isoladas (tabelas 16 e 17), observamos que os cálculos para moléculas isoladas, neste caso, tendem a estabilizar as conformações mais "dobradas" (B, C, E, F e H) e para moléculas solvatadas, são estabilizadas as conformações mais "estendidas" (A e D).

A estabilização das conformações "estendidas" pela solvatação em água, nos indica que as interações com o solvente, neste caso, ocorrem com maior eficiência do que nas conformações "dobradas". As moléculas estudadas contêm heteroátomos como oxigênio ou cloro substituídos nos anéis fenílicos, que podem formar pontes de hidrogênio com a água ou outro tipo de interação eletrostática; sendo assim, as conformações "estendidas" parecem "facilitar" a formação dessas interações.

Em estudos anteriores, verificamos que as únicas conformações presentes em todas as moléculas ativas são: C, D, E e F. Assim, uma delas deve ser a conformação ativa, isto é, que interage com o receptor. Apesar de não termos realizado estes cálculos para as 20 neolignanas do conjunto treinamento, os mesmos já nos indicam que a conformação D deve ser usada no prosseguimento dos cálculos, em preferência à E - como havíamos concluído da análise para os compostos isolados [5,6] - pois, as conformações A e D foram estabilizadas na maioria dos casos, mas apenas D está presente em todas as ativas.

7.3 - CONCLUSÕES

No capítulo 5, concluímos que o volume molecular é um parâmetro importante na classificação dos compostos , e que a atividade anti-leishmaniose das neolignanas parece estar relacionada com o **aumento de volume molecular**; essa conclusão é concordante com os resultados obtidos após a solvatação, pois as conformações com maior probabilidade de serem ativas são justamente aquelas em que os anéis fenílicos estão mais afastados entre si (A e D).

A escolha da conformação D parece razoável, pois esta, além de ter sido uma das mais estabilizadas em água, está presente em todas as neolignanas ativas e apresenta volume maior do que as outras, em que os anéis estão mais próximos.

CAPÍTULO 8

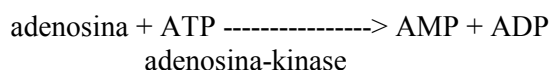
PROCURA DA CONFORMAÇÃO ATIVA DAS NEOLIGNANAS, ATRAVÉS DA COMPARAÇÃO COM OUTROS COMPOSTOS ATIVOS EM LEISHMANIOSE E COM UM POSSÍVEL SÍTIO RECEPTOR.

8.1 - INTRODUÇÃO:

No capítulo anterior, vimos que as conformações A e D foram mais estabilizadas pela água do que as demais e trabalhos anteriores selecionaram apenas quatro mínimos locais entre os oito encontrados para as neolignanas. Esses quatro são: C, D, E e F. Assim, até agora a conformação D parece ser a conformação ativa mais provável.

Como já vimos no capítulo 1 (item 1.3), a arginina é um possível sítio ativo do leishmania, porque está presente apenas na adenosina-kinase desse protozoário, isto é, a adenosina-kinase dos animais hospedeiros do leishmania não contém este resíduo de arginina. Assim, pode-se desenvolver uma droga capaz de interagir com a arginina, a fim de inibir a função enzimática apenas da adenosina-kinase do leishmania, sem alterar a mesma enzima do hospedeiro.

As kinases são enzimas específicas que participam dos processos de fosforilação; no caso da adenosina-kinase, o substrato da enzima é a adenosina. Isso significa que a enzima vai interagir com a molécula da adenosina durante a fosforilação:



onde, AMP, ADP e ATP são a adenosina-monofosfato, adenosina-difosfato e adenosina-trifosfato, respectivamente, conforme a figura 30.

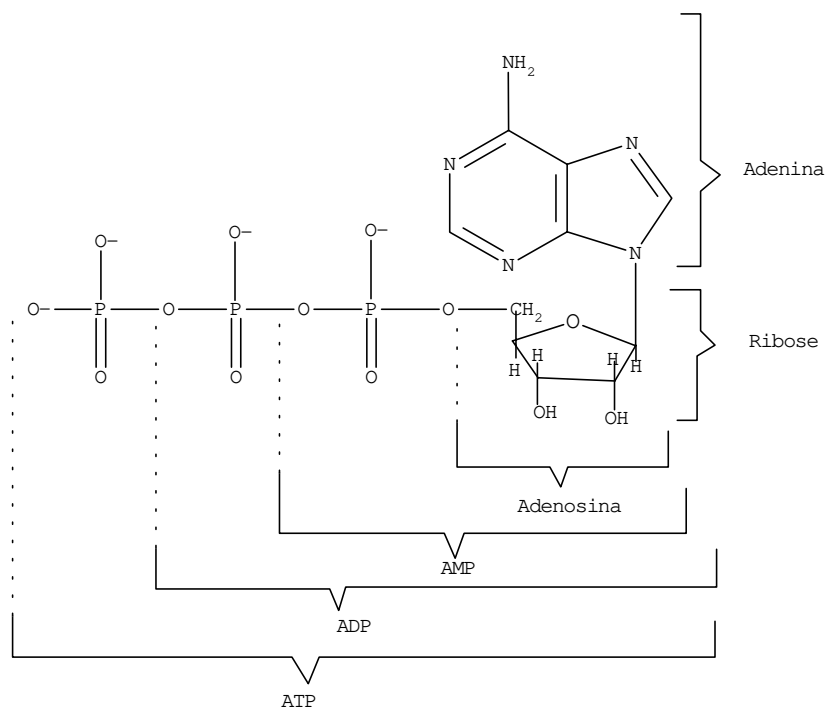


Figura 30 - Representação esquemática da adenosina-monofosfato (AMP), adenosina-difosfato (ADP) e adenosina-trifosfato (ATP).

A figura 30 mostra que a adenosina é formada pela base purínica adenina e por um anel de ribose, que, segundo a literatura, fica num plano perpendicular ao plano da adenina [69].

Uma das maneiras pela qual as enzimas interagem com o seu substrato é formando o complexo chave-fechadura, no qual ambos possuem geometrias complementares; a função da enzima é encaixar-se ao substrato, a fim de colocá-lo na posição correta para interagir com outros compostos [69]. No caso da inibição competitiva, a droga inibidora da função enzimática deve também possuir geometria complementar à respectiva enzima e, portanto, deve ser de estrutura similar à do substrato.

Se a arginina for realmente o sítio ativo do leishmania, as neolignananas ativas devem possuir uma geometria que permita ocorrer interações com a arginina, a ponto de inativá-la.

O objetivo desta etapa do trabalho é investigar qual das possíveis conformações ativas das neolignanas seria mais parecida com a geometria da arginina. Isso nos daria uma indicação a mais sobre a conformação ativa das neolignanas.

Os cálculos desta etapa do trabalho foram discutidos no capítulo 4, item 4.2.

8.2 - RESULTADOS E DISCUSSÃO:

Não foi encontrada na literatura, a conformação da arginina presente na adenosina-kinase do leishmania, mais adequada para a continuidade deste trabalho; deste modo, optou-se por utilizar dados de cristalografia para a arginina, com a seguinte fórmula mínima: arginina.2H₂O [61].

A geometria da arginina obtida de dados de cristalografia, é mostrada na figura 31a; a figura 31c mostra o esquema da adenosina. A geometria da porção adenina da adenosina foi obtida da literatura [69]. Como já foi mencionado, sabe-se que a adenina está num plano perpendicular ao da ribose [69].

Figura 31 - a) Geometria da arginina obtida de dados de cristalografia [61]; b) Geometria da arginina após rotação de 120° do eixo “c”, em relação à figura 31a. c) Esquema representativo da adenosina. A geometria da adenina foi obtida da literatura [69].

Comparando a geometria do grupo guanidínico na arginina com a geometria da adenina, observamos que os átomos N9, C6, N10 e N8 da arginina podem superpor-se aos átomos C6, C5, N7 e C4 da adenina, respectivamente, através de ligações hidrogênio entre os grupos -NH da arginina e os átomos de nitrogênio da adenina.

Se o eixo “c” da arginina (figura 31a) for girado de 120 graus (figura 31b), uma outra ligação hidrogênio pode ocorrer entre o H15 da arginina e o O17 da adenosina.

A tabela 19 mostra os valores do calor de formação calculados para as duas geometrias da arginina e a energia da barreira rotacional. A geometria obtida de dados de cristalografia foi denominada A e aquela obtida após rotação de 120 graus, foi chamada de B. A barreira rotacional (E_R) foi calculada partindo-se da conformação A. Esses cálculos foram realizados com o método semi-empírico AM1. Os valores obtidos indicam que a arginina pode estar na conformação B, pois essa conformação é mais estável do que A e a barreira rotacional encontrada é baixa (3,09 kcal), permitindo a rotação sobre o eixo “c”.

Como a conformação A é observada nos dados de cristalografia, esta deve ser mais estável do que B, para a arginina na forma de cristal. A discrepância entre a conformação observada e calculada, deve-se provavelmente, ao fato de que a conformação calculada é obtida para a molécula isolada e a observada, para a molécula no cristal. A diferença entre os valores pode também ter ocorrido devido ao método semi-empírico utilizado. Em qualquer um desses casos, entretanto, podemos assumir que B é quase tão estável quanto A, uma vez que a barreira rotacional entre as duas conformações é baixa: 3,09 kcal/mol. Essas considerações sugerem que a arginina pode adquirir tanto a conformação A quanto a B, quando interage com o substrato. Contudo, se estiver na conformação B, poderá interagir com o substrato, pelos seus grupos amino e guanidínico.

Tabela 19 - A conformação A da arginina foi obtida de dados de cristalografia e a conformação B foi obtida após rotação de 120° do eixo “c”, mostrado nas figuras 4a e 4b. R_E é a barreira rotacional calculada a partir da conformação A.

CONFORMAÇÃO	ΔH_f (kcal/mol)	R_E (kcal/mol)
A	-56,77	3,09
B	-58,52	

Os elétrons π dos dois anéis aromáticos das neolignanas podem fazer interações eletrostáticas com os referidos grupos da arginina. Para isso, a posição dos anéis deve coincidir com a dos grupos guanidínico e amino da arginina. A distância entre os dois grupos na arginina é de 5,0 Å, considerando os átomos C3 e C6 na figura 31a. A

conformação D da neolignana 12, por exemplo, é a única que apresenta uma distância próxima a esse valor, considerando os átomos C6 e C18, na figura 4a.

Entre outros compostos ativos em leishmnirose, encontramos uma série de amino cetonas. Estas moléculas são semelhantes às neolignanas, porque ambas as séries apresentam dois anéis fenílicos com uma carbonila entre eles. A estrutura básica das aminocetonas está mostrada na figura 4b e tabela 1 (páginas 25 e 26). A geometria dessas moléculas foi calculada conforme está explicado no capítulo 4.

A conformação D das neolignanas é aproximadamente planar, como as amino-cetonas. Esta conformação produz a posição adequada dos anéis para interagir com a arginina; as outras possuem os ângulos torsionais a ou b (tabela 20) próximos a 60 graus, o que causa nas neolignanas uma conformação "dobrada", encurtando a distância entre os anéis e impossibilitando a interação das neolignanas nas conformações C, E e F com a arginina. Esses resultados e para os outros compostos ativos são mostrados na tabela 20. Esta tabela contém também, os ângulos torsionais a e b para todas as neolignanas, arginina e para as amino-cetonas.

Tabela 20 - d_l é a distância entre os átomos c_3 e c_6 para a arginina (fig.31), c_6 e c_{18} para as neolignanas (fig. 4a) e c_6 e c_{16} para as aminocetonas (fig.4b). Os ângulos torsionais a e b são mostrados nas figuras 4 e 31.

COMPOSTO	CONFORMAÇÃO	d_l (Å)	ÂNGULOS TORSIONAIS	
			a	b
arginina	A ou B	5,0	162	175
neolignana 12	C	3,6	46	63
	D	5,0	162	175
	E	4,4	-60	180
	F	3,6	-46	-63
neolignana 13	C	3,4	24	64
	D	5,1	-175	179
	E	4,5	-66	180
	F	3,4	-24	-64
neolignana 14	C	3,2	26	61
	D	5,0	-176	179
	E	4,4	-63	167
	F	3,7	-49	-67
neolignana 15	C	3,9	65	74
	D	5,0	-177	161
	E	4,1	-60	162
	F	3,4	-42	-65

neolignana 16	C	3,9	60	76
	D	4,6	-155	142
	E	4,1	-61	179
	F	3,3	-39	-65
neolignana 17	C	3,4	49	45
	D	5,3	-139	-177
	E	4,8	-74	173
	F	3,8	-72	-50
neolignana 18	C	3,4	56	39
	D	5,3	-140	-177
	E	4,8	-72	173
	F	4,0	-68	-63
neolignana 19	C	3,4	52	51
	D	4,9	-160	177
	E	4,3	-69	171
	F	3,7	-57	-63
neolignana 20	C	3,4	58	38
	D			
	E			
	F	4,0	170	-63
todas aminocetonas		4.9	-150	-180

Com os resultados e conclusões obtidos até agora, sugerimos que: 1) as neolignanas devem interagir com o receptor através dos seus anéis aromáticos, na conformação D; 2) a arginina interage pelos seus grupos guanidínico e amino, nas conformações A ou B (figura 31) .

Com a ajuda de modelos moleculares, sobreposemos a neolignana ativa 12 D e a arginina na conformação B; a sobreposição das duas moléculas é mostrado na figura 32. Nesse caso, podemos pensar em interações eletrostáticas entre os grupos -NH da arginina e os elétrons π do anel 1 da neolignana, ou entre os grupos -NH da arginina e o oxigênio 16 da neolignana e ainda, poderiam ocorrer interações de dipolos entre o oxigênio da carbonila da arginina e a carga positiva no carbono 21 da neolignana, localizado no anel 2; o -NH do grupo amino também poderia interagir com alguma ligação dupla do anel 2.

A figura 33 mostra o mesmo tipo de sobreposição com uma amino-cetona; neste caso, a sobreposição também parece ser possível e as interações ocorreriam principalmente entre os grupos -NH guanidínicos da arginina e os elétrons π do anel 1 ou o oxigênio da carbonila das amino-cetonas; o grupo amino da arginina interagiria com os elétrons π do anel 2.

cliv

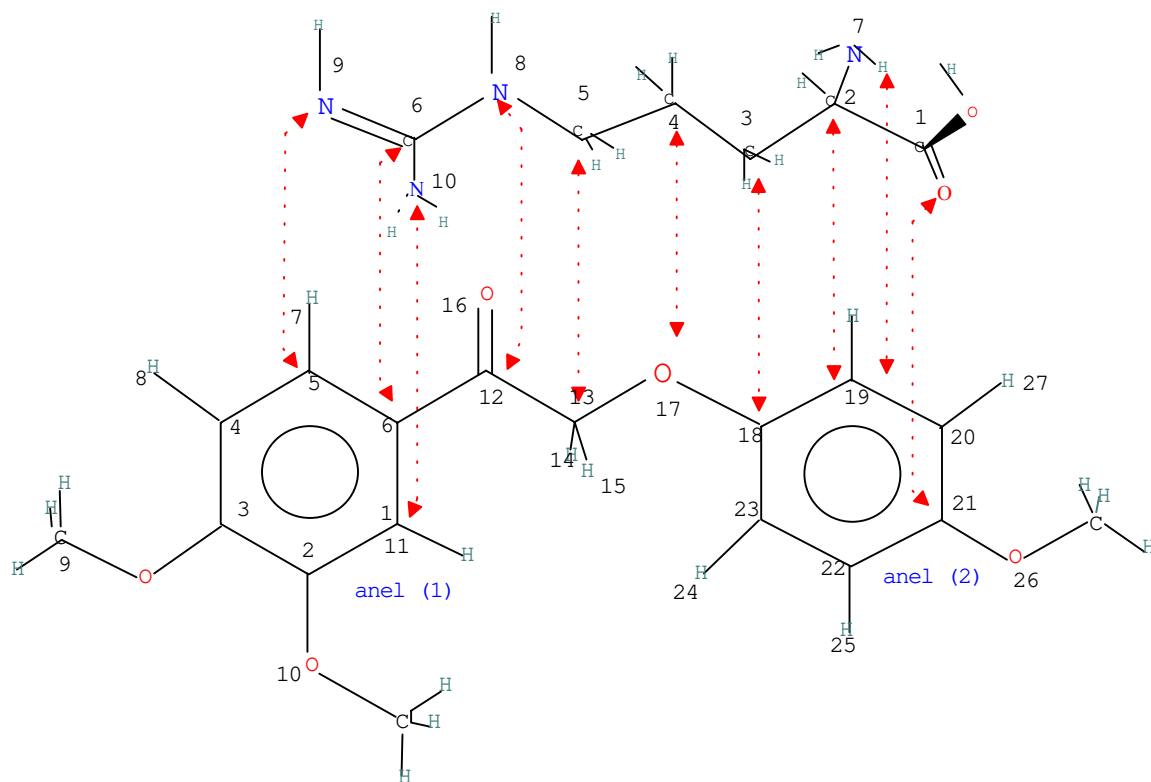


Figura 32 - Interação entre a neolignana 12 D e a arginina na conformação B.

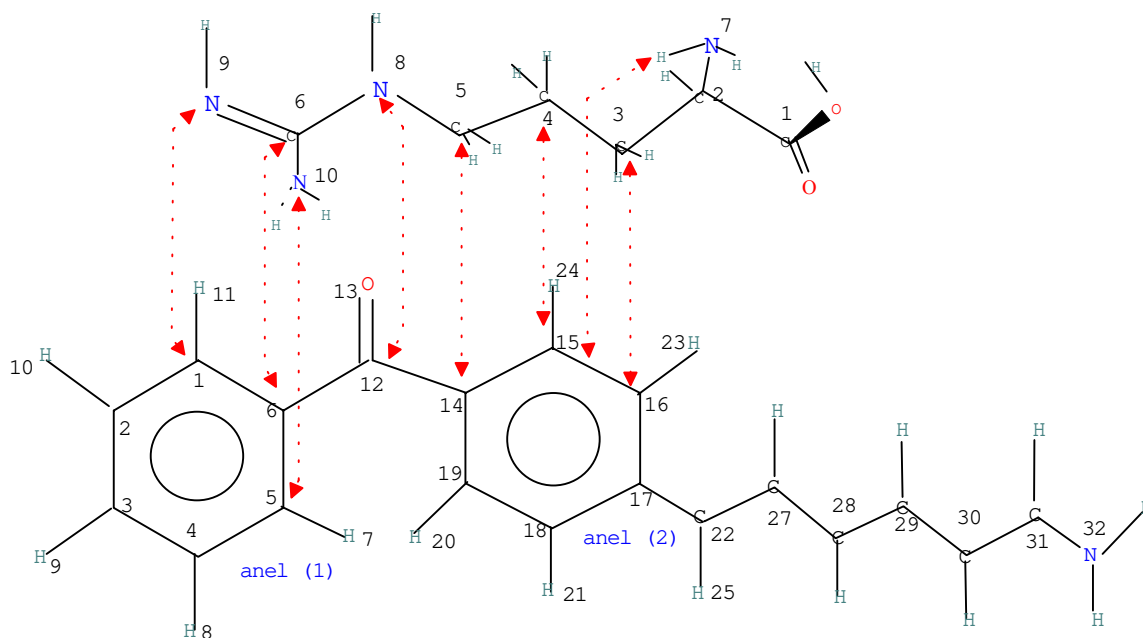


Figura 33 - Interação entre uma amino-cetona e a arginina na conformação B.

Para confirmar essas sugestões, usamos o programa computacional PowerFit para realizar a interação entre a neolignana 12 D e a arginina. A metodologia sobre o PowerFit foi descrita no capítulo 2.

A figura 34 mostra a interação da arginina nas suas duas possíveis conformações, A e B, com a conformação D da neolignana 12 (12 D). A interação da arginina na conformação A com a neolignana 12 D apresenta -900,74 kcal de energia, enquanto a interação da conformação B da arginina com a mesma neolignana 12 D, apresenta -982,02 kcal.

Figura 34 - Interação da neolignana 12D com a arginina, produzida pelo programa PowerFit*. a) A arginina está na conformação obtida de dados de cristalografia; b) A arginina está na conformação obtida após rotação de 120° sobre o eixo “c”(fig. 4b).

* Os valores de energia para a figura 34 foram obtidos com uma versão demonstrativa do PowerFit; por isso, estes valores podem estar incorretos. Estamos interessados apenas numa análise qualitativa destes resultados.

Esses resultados mostram que a interação arginina-12D é mais estável para a conformação B da arginina, conforme havíamos proposto inicialmente.

A sobreposição entre as duas moléculas não se deu exatamente como sugerimos inicialmente, pois na figura 32, mostramos uma possível interação entre a arginina e a neolignana 12D, em que os carbonos 12, 13 e 17 da neolignana ficariam sobrepostos aos átomos 8, 5 e 4, respectivamente, da arginina. Conforme os resultados obtidos com o PowerFit, isto não ocorre, mas a interação entre os anéis fenílicos da neolignana e os grupos guanidínico e amino da arginina (figura 34b), concorda com a nossa suposição, na figura 32; as duas figuras mostram a interação com os anéis fenílicos.

8.3 - CONCLUSÕES

Com estes resultados, obtivemos mais uma vez a indicação de que a conformação ativa mais provável das neolignanas é a D; esta foi selecionada inicialmente, juntamente com outras 3 (C, D, E, F), entre as oito possíveis na análise conformacional para os compostos isolados (capítulo 5); em seguida, foi uma das mais estabilizadas (A e D) na solvatação em água (capítulo 7), e, finalmente, é a conformação que melhor se “encaixa” à arginina e também é a mais parecida com as aminocetonas que também apresentam atividade anti-leishmaniose.

Todos os resultados, a partir do capítulo 5, nos dão fortes indicações de termos encontrado a conformação ativa das neolignanas, bem como o modo de interação com o receptor. Sabemos ainda que existe uma tendência de aumento de AB anti-leishmaniose, conforme aumenta o volume e o logP das neolignanas. Agora temos em mãos, resultados e conclusões suficientes para propormos alguns compostos para síntese, com um bom nível de confiança.

CAPÍTULO 9

SUGESTÃO DE NOVOS COMPOSTOS PARA SÍNTESE

Este capítulo será dividido em duas partes, pois a sugestão de compostos para síntese foi realizada em duas etapas.

O primeiro conjunto de compostos foi planejado após obtermos os resultados do capítulo 5, em que toda a análise foi realizada para os compostos isolados. Em seguida, os resultados dos capítulos seguintes (em que as neolignanas do conjunto treinamento foram solvatadas e comparadas com outros compostos ativos em leishmaniose) nos levaram a procurar uma nova metodologia para se propor os compostos para síntese; assim, um novo conjunto teste foi sugerido.

9.1 - PARTE 1: Conjunto teste, proposto com base na análise conformacional para os compostos isolados.

9.1.1 - INTRODUÇÃO:

No capítulo 5, todos os resultados foram obtidos segundo a análise conformacional realizada para os compostos isolados, o que nos levou a conclusões diferentes daquelas após a solvatação em água. Conforme a análise dos compostos isolados, a conformação com maior probabilidade de ser a ativa é aquela em que os anéis fenílicos estão mais próximos entre si (conformação E), conforme foi explicado no capítulo 5, item 5.2. Os métodos de Reconhecimento de Padrões (RP) aplicados ao conjunto treinamento na conformação E, inicialmente, levaram aos resultados apresentados nas figuras 19 a, b e c do capítulo 5. Os compostos do nosso primeiro conjunto teste foram então planejados, baseando-nos na semelhança estrutural com os compostos ativos do conjunto treinamento e em seguida, a nova análise de RP incluindo os novos compostos foi realizada, a fim de

verificarmos se os compostos teste seriam classificados como ativos ou inativos. O conjunto teste é mostrado a seguir, na tabela 21; a numeração começa a partir de 21, pois estes compostos foram analisados juntamente com os 20 compostos do conjunto treinamento, dispostos na tabela 2 do capítulo 1.

Tabela 21 - Conjunto teste, planejado com base na semelhança estrutural das neolignananas ativas em leishmaniose. R1, R2, etc, são os substituintes nas posições 3, 2, etc, mostradas na figura 4.a do capítulo 1; a posição X também é mostrada na figura 4.a.

Com- postos	R1	R2	R3	R4	R5	R6	R7	R8	X
21	-OCH ₃	-OCH ₃	H	CH ₃	-OCH ₃	H	(CH) ₂ -CH ₃	-OCH ₃	O
22	H	H	=O	H	Cl	H	H	H	O
23	H	H	=O	H	H	H	Cl	H	S
24	-OCH ₃	-OCH ₃	=O	CH ₃	-OCH ₃	H	H	-OCH ₃	O
25	-OCH ₃	-OCH ₃	=O	H	H	H	-CH ₃	H	S
26	-OCH ₃	-OCH ₃	=O	H	H	H	-CH ₃	H	O
27	-OCH ₃	-OCH ₃	=O	H	H	H	(CH) ₂ -CH ₃	H	O
28	H	H	=O	CH ₃	H	H	Cl	H	O
29	H	H	=O	CH ₃	H	H	H	H	O
30	H	H	=O	CH ₃	H	H	H	H	S
31	-OCH ₃	-OCH ₃	=O	CH ₃	Cl	H	H	H	O
32	-OCH ₃	-OCH ₃	=O	H	Cl	H	H	H	O

9.1.2 - CÁLCULOS:

A análise conformacional para este conjunto teste de compostos foi realizada através do método de mecânica molecular - MM2, como para as neolignanas do conjunto treinamento, conforme foi discutido no item 4.1.1 do capítulo 4. Os parâmetros físico-químicos foram calculados na conformação E (conformação ativa encontrada na análise dos compostos isolados), usando os mesmos métodos do conjunto treinamento (item 4.1.2 - capítulo 4); apenas os cinco parâmetros mais importantes encontrados para o conjunto treinamento, foram calculados para os compostos teste: o logP, com o método de Rekker [39], o volume, com o programa Surf [57], e as cargas sobre os carbonos 2, 3 e 21, com o método semi-empírico MNDO-PM3 [36]. Em seguida, o conjunto teste foi analisado através dos seguintes métodos RP: Knn, Hier e ACP, contidos nos programas Arthur/Unicamp [58] e Pirouette [59] e os compostos classificados como ativos pelos três métodos citados, foram sugeridos para síntese.

9.1.3 - RESULTADOS E DISCUSSÃO:

O resultado da análise de componentes principais é mostrado nas figuras 35 e 36. Na figura 35, o gráfico dos “scores” das duas primeiras componentes principais mostra a separação dos compostos do conjunto treinamento em ativos e inativos, como na figura 19b do capítulo 5, mas agora, além do conjunto treinamento o gráfico mostra também a classificação do conjunto teste. As pequenas diferenças entre este gráfico e o da figura 19b, devem-se ao fato de que anteriormente utilizamos as cargas sobre os heteroátomos 9, 10 e 26 e agora, sobre os carbonos em que estão ligados (2, 3 e 21), pois, conforme já discutimos, estas representam melhor a influência do substituinte no anel aromático, do que as cargas sobre os heteroátomos.

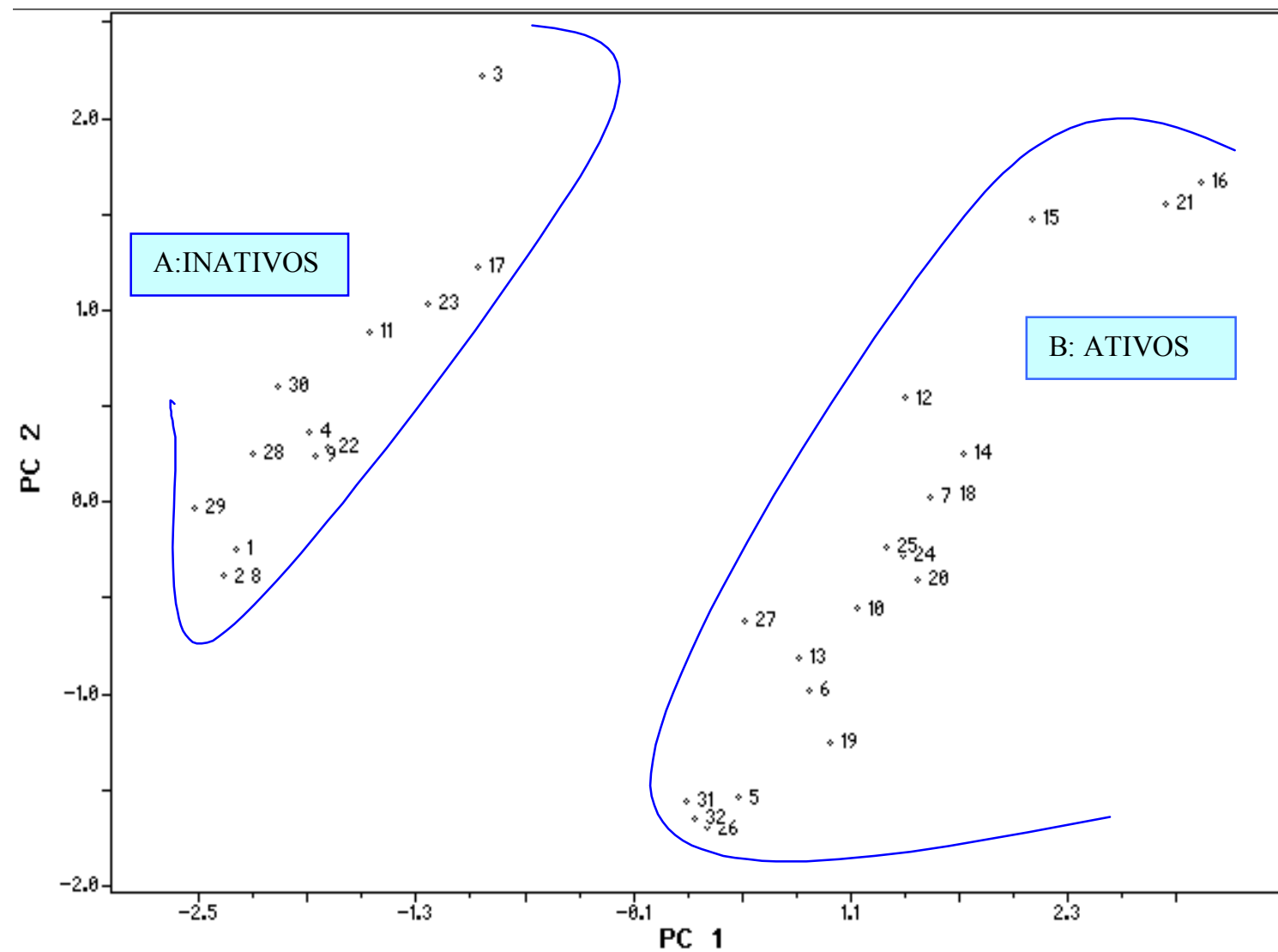


Figura 35 - Gráfico dos “scores” das duas primeiras componentes principais para os conjuntos treinamento e teste das neolignanas. Os compostos teste **21, 24, 25, 26, 27, 31 e 32** foram classificados como ativos.

O gráfico dos “loadings” mostra que as cargas sobre os átomos 2, 3 e 21 (variáveis 26, 27 e 37) são muito importantes na primeira componente principal para a separação dos compostos, enquanto o logP (variável 53) e o volume (variável 49), contribuem menos. No caso da segunda componente principal, as variáveis 26, 27 e 37 continuam sendo as mais importantes; em seguida, o logP (variável 53) também contribue para a separação na segunda componente principal e, finalmente o volume (variável 49) parece não ser muito significativo na separação dos compostos pela segunda componente principal. apresenta peso muito pequeno na segunda componente principal.

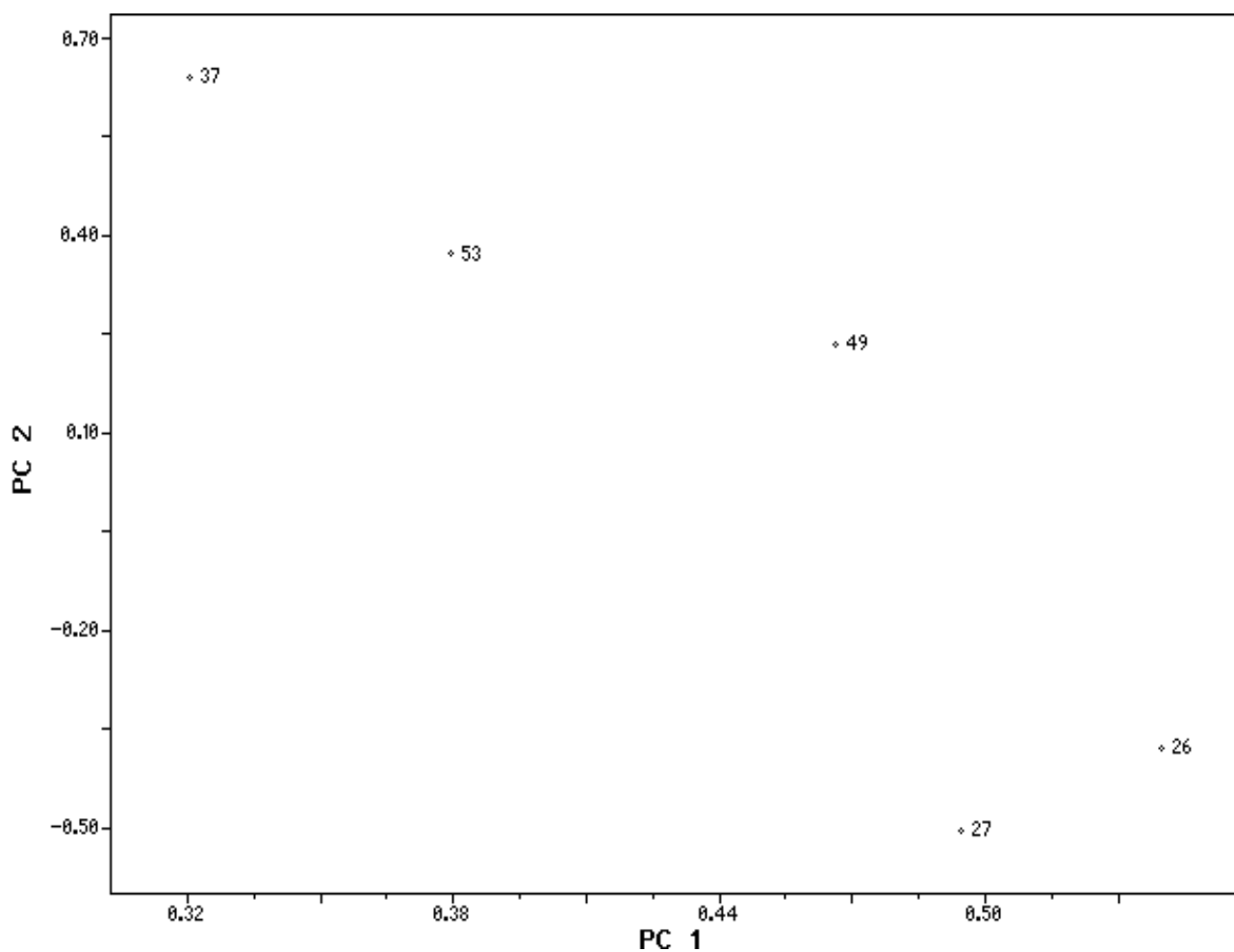


Figura 36 - Gráfico dos “loadings” das duas primeiras componentes principais para os conjuntos de treinamento e teste das neolignanas. As variáveis 26, 27, 37, 49 e 53 são respectivamente, as cargas sobre os átomos 2, 3, 21, volume e logP.

A figura 37 mostra o agrupamento hierárquico, o qual classificou como ativos os compostos **21, 24, 25, 26, 27, 31 e 32** (categoria B, assinalada no dendograma), assim como a análise ACP.

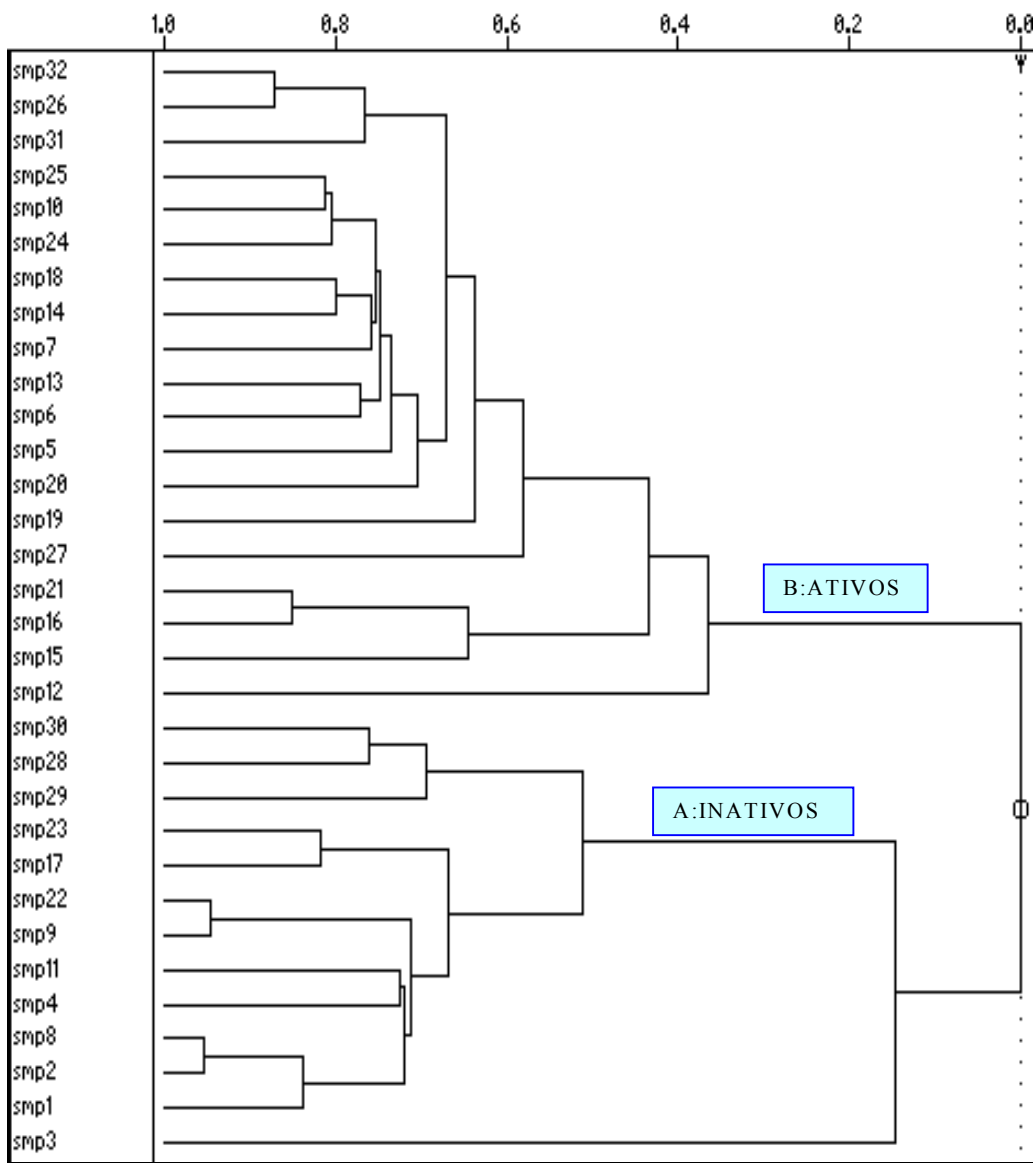


Figura 37 - Agrupamento hierárquico para os conjuntos treinamento e teste das neolignanais. Os compostos **21, 24, 25, 26, 27, 31 e 32** foram classificados como ativos. Os compostos estão indicados à esquerda da figura, por smp1, smp2, etc.

Os resultados obtidos com os dois métodos apresentados (ACP e Hier) são concordantes, porém, pelo método dos k vizinhos mais próximos, apenas os compostos **21**, **24**, **25**, **26** e **27** foram classificados como ativos. Assim, estes foram inicialmente sugeridos para síntese.

Entre os compostos sugeridos, apenas o **24** foi testado em *leishmania donovani*, mas o resultado foi negativo.

No decorrer do capítulo 5, concluímos que algumas cargas incluídas em posições específicas do anel (2) (cargas 20, 21 e 22) foram importantes na separação das duas categorias de compostos (figuras 21a, b e c), o que poderia alterar também os resultados para o conjunto teste. Nos capítulos seguintes, entretanto, obtivemos indicações de que a conformação ativa seria outra, diferente daquela estudada no capítulo 5 e só após a posse desta nova informação é que propusemos o novo conjunto teste.

A solvatação do compostos (capítulo7) nos mostrou que a conformação ativa deveria ser uma daquelas em que os anéis estão mais estendidos (A ou D) e não a E, como havíamos pensado. A comparação das neolignanas com outros compostos ativos em leishmaniose e com o sítio ativo do receptor confirmou que a conformação D deve ser a conformação ativa. Finalmente, o estudo da interação do sítio ativo com as neolignanas confirmaram que a interação com o receptor deve ocorrer pelos anéis fenílicos; esse estudo nos mostrou também com maior precisão os locais de interação com o anel (2), que seria próximo aos carbonos 19 e 21.

Ao mesmo tempo que obtivemos estas informações sobre a conformação ativa e sobre a interação com o receptor, comparamos as cargas derivadas do potencial eletrostático com aquelas de Mülliken utilizadas anteriormente no capítulo 5 e concluímos que deveríamos prosseguir o trabalho utilizando aquelas derivadas do potencial eletrostático, conforme foi explicado no capítulo 3. Apesar de termos discutido este assunto no capítulo 3 que refere-se aos parâmetros físico-químicos utilizados, foi após a solvatação dos compostos, apresentada no capítulo 7 (que exigia cargas derivadas do potencial eletrostático), que comparamos as cargas calculadas pelos dois métodos e optamos por **não** continuar com as cargas de Mülliken.

9.2 - PARTE 2: Conjunto Teste Proposto Após A Solvatação Dos Compostos.

9.2.1 - INTRODUÇÃO:

Os métodos de Reconhecimento de Padrões foram muito úteis e eficazes na parte inicial deste trabalho; foi possível, por exemplo, definir as variáveis relacionadas com a atividade biológica e propor o conjunto teste apresentado na parte 1 deste capítulo. Entretanto, os compostos **5, 6, 7, 10 e 17** apresentam a tendência de serem classificados incorretamente (como foi discutido anteriormente no capítulo 5), dificultando a sugestão de compostos para síntese.

Considerando este fato e lembrando que agora dispomos de maiores informações sobre as relações estrutura-atividade das neolignanas, optamos por não utilizar novamente os métodos RP na sugestão de novos compostos. Ao invés disso, tentamos encontrar respostas para a inatividade dos compostos **5, 6, 7 e 10** na estrutura conformacional destes e na interação dos mesmos com o receptor (arginina).

A figura 38 mostra todas as neolignanas na conformação D. Inicialmente, vamos observar o anel (2) de todas as neolignanas ativas - composto **12 a 20** - e daquelas inativas que apresentam o grupo metóxi no anel (1) - compostos **5, 6, 7 e 10**.

Figura 38 - Neolignanas na conformação D.

Figura 38 - Continuação

Compostos 5 e 10:

Os compostos inativos **5 e 10** não apresentam nenhum substituinte no anel (2), ao contrário de todas as ativas. Como vimos nos resultados do capítulo 8, a interação da arginina com o anel (2), ocorre próxima às posições 19 e 21, para o composto **12** do exemplo (figuras 32 e 34), ou seja, na parte superior do anel (2) (carbono 19) e na posição “para” (carbono 21) ao átomo da ponte (oxigênio no caso do composto **12**).

Se observarmos os compostos ativos, na figura 38, todos apresentam um substituinte pelo menos em uma dessas posições. Esses substituintes, em todos os casos, conferem uma carga positiva ao carbono em que está ligado (tabela 3 - veja as cargas derivadas do potencial eletrostático para os carbonos substituídos 19 e/ou 21 para os compostos ativos); sendo assim, concluímos que um **composto ativo deve ter um substituinte desse tipo (geralmente -OCH₃) nas posições 19 e/ou 21**; explicamos com isso, a **inatividade dos compostos 5 e 10**.

Composto 6:

Observando agora o composto **6** na conformação D (figura 38), verificamos que apesar de estar substituído por um **-OCH₃** no carbono **19**, **este localiza-se em posição contrária** ao carbono 19 dos compostos ativos **12, 13, 14, 15 e 16**, por exemplo (figura 38). Assim, o substituinte fica na parte inferior do anel (2), onde não ocorreria interação com a arginina (figuras 32 e 34b). **Talvez por esse motivo, o composto 6 seja inativo.**

Neolignanas ativas que apresentam o grupo metil no carbono 13 e =O no carbono 12:

A inversão da posição do carbono 19 observada para o composto **6**, devido à rotação do anel (2), ocorre também para os compostos **ativos 17, 18, 19 e 20**; porém, estes apresentam o carbono 21 substituído, por onde também a arginina pode interagir.

Observando mais cuidadosamente essas neolignanas, verificamos que todas aquelas que apresentam =O no carbono 12 e, ao mesmo tempo, o grupo metil (-CH₃) no carbono 13, apresentam essa inversão de posição do carbono 19, ou os anéis (1) e (2) ficam perpendiculares entre si, impedindo a interação com a arginina. Nestes casos, a interação só poderia ocorrer pelo carbono 21 (observe os compostos **6, 7, 17, 18, 19 e 20**). Em contrapartida, aqueles que possuem hidrogênio ao invés de -CH₃ no carbono 13, não apresentam essa inversão (compostos **12, 13, 14**). Assim, chegamos a mais uma conclusão sobre a atividade das neolignanas: **Se o composto apresentar o grupo -CH₃ no carbono 13, o anel 2 deverá estar substituído no carbono 21.**

No caso dos compostos **15 e 16**, que também apresentam -CH₃ na posição 13, este problema não ocorre, pois não possuindo =O no esqueleto, possuem estrutura diferente das demais e o substituinte poderá estar nos carbonos 19 e/ou 21, como nas outras neolignanas ativas.

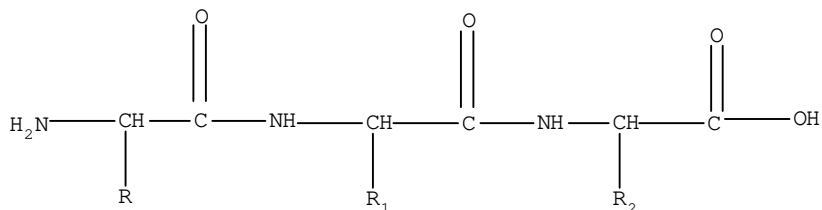
Composto 7:

O composto **7** parece estar condizente com as conclusões sobre a atividade e no entanto, é inativo. Este apresenta o grupo metil no carbono 13 e a inversão no carbono 19 ocorre; contudo, ele possui um substituinte na posição 21. Observe que de todas as neolignanas na conformação D que apresentam o grupo metil no carbono 13, esta é a única em que os anéis estão planares (figura 38), devido talvez à interação entre a carbonila e o oxigênio do grupo metóxi do carbono 19. Se observarmos a figura 34, veremos que um grupo metóxi nessa posição impedirá a aproximação da arginina, pois irá coincidir com um grupo $-\text{CH}_2$ da arginina. Assim, isso explicaria a inatividade do composto **7** e uma nova ressalva deve ser incluída ao sugerirmos novos compostos: **Se o composto apresentar o grupo metil no carbono 13 e $=\text{O}$ no carbono 12, deverá estar substituído no carbono 21; entretanto, se além disso o composto estiver substituído também na posição 19, esse substituinte não deverá levar à planaridade entre os anéis fenílicos (1) e (2).**

Até agora, procuramos justificar a inatividade das neolignanas que apresentam o grupo metóxi no anel (1) (neolignanas **5, 6, 7 e 10**), como a maioria das ativas; agora, vamos tentar explicar a inatividade daquelas que não apresentam os grupos metóxi no anel (1) - compostos **1, 2, 3, 4, 8, 9 e 11**.

É muito provável que o resíduo de arginina da adenosina-kinase do leishmania, esteja localizado em posição intermediária, não sendo portanto, N-terminal ou C-terminal.

Nas proteínas, os aminoácidos estão ligados uns aos outros por ligações amida, como no exemplo abaixo (para um conjunto de três aminoácidos), onde R, R_1 e R_2 são as cadeias laterais de cada aminoácido.



No caso da arginina, a cadeia lateral contém o grupo guanidínico. Se a arginina tivesse C-terminal, apresentaria o grupo carboxila como na figura 34a ou 34b, o qual interagiria com a neolignana. A maior probabilidade é que a arginina esteja ligada a outros aminoácidos e assim, o grupo -OH da carboxila seria substituído pelo -NH do aminoácido seguinte. Então, ao invés da interação ocorrer pelo -OH da arginina, ocorreria pelo -NH do outro aminoácido. Isso tudo não altera em nada as nossas conclusões, pois de qualquer maneira a interação com o anel (2) pode ocorrer.

As proteínas são reativas e apresentam um comportamento específico, devido aos grupos ativos livres das cadeias laterais, como os grupos guanidínicos das argininas [47]. A interação do anel (2) com o grupo amino da arginina ou amido (devido à ligação com outro aminoácido), é importante para a atividade das neolignanas, talvez pela sobreposição que deve ocorrer entre a droga e o receptor, mas a **especificidade** da atividade está principalmente relacionada com a interação entre o anel (1) e o grupo guanidínico. **Sendo assim, todos os compostos ativos devem conter os substituintes metóxi no anel (1), que permitem tal interação. Os compostos 1, 2, 3, 4, 8, 9 e 11 não apresentam tal substituinte no anel (1), o que explica a inatividade dos mesmos, em leishmaniose.**

9.3 - SUGESTÃO DE NOVOS COMPOSTOS.

Com base nas conclusões citadas no item 9.2, uma nova neolignana deve ser planejada da maneira descrita a seguir, considerando as diferentes estruturas básicas:

- 1 - Todas devem conter os grupos “metóxi” no anel (1);
- 2 - Se a estrutura básica não tiver o grupo metil no carbono 13 (nesse caso os anéis fenílicos são planares entre si), deverá ter um substituinte que confira uma carga positiva ao carbono em que estiver ligado, nas posições 19 e/ou 21 (por exemplo, um metóxi);
- 3 - Se a estrutura básica tiver o grupo metil no carbono 13 (neste caso, os anéis fenílicos não são planares entre si), então, o substituinte deve necessariamente estar no carbono 21, pois a posição 19 não é alcançada pela arginina (ex: compostos 18 e 19).
- 4 - Se a estrutura básica tiver o grupo metil no carbono 13 e também um substituinte do tipo -OCH₃ no carbono 19, o composto ficará planar novamente, mas haverá impedimento estérico impossibilitando a aproximação da arginina. Nesse caso, tal composto deve ser inativo (ex: composto 7).

Voltando ao conjunto teste da tabela 21, vamos analisar cada um dos compostos planejados.

Os **compostos 22, 23, 28, 29 e 30** seriam **inativos** por não apresentarem a condição 1, apresentada acima. O **composto 31**, não sendo planar (-CH₃ no carbono 13 (R4)), deveria estar substituído na posição 21 (R7), ao invés da 19 (R5). O **composto 32**, entretanto, poderia ser **ativo**, contrariando o resultado apresentado na parte 1 deste capítulo. O composto **32** obedece à condição 1 e é do tipo descrito no item 2, ou seja, não contém o metil no carbono 13. Assim, o substituinte no carbono 19 (R5) permite a interação com a arginina.

Agora, vamos aos compostos teste classificados como ativos, anteriormente; são aqueles de número **21, 24, 25, 26 e 27**. Todos obedecem à condição 1; os compostos **25, 26 e 27** são do tipo descrito no item 2; o **composto 25 deve ser ativo**, pois apesar de possuir

enxofre na ponte, o que elimina a planaridade entre os anéis, ele está substituído no carbono 21. Os compostos **26 e 27 também podem ser ativos**, pois estão substituídos numa das posições descritas no item 2 (C19 ou C21). **O composto 24 deve ser inativo**; ele é do tipo apresentado no item 4: possui metil no carbono 13 e o grupo metóxi na posição 19 (R5), que o torna planar, ao mesmo tempo que impossibilita a aproximação da arginina devido ao impedimento estérico causado pelo substituinte no carbono 19, na parte inferior do anel (2). O composto **21** pode ser ativo, pois apresenta as condições 1 (substituinte no anel (1)) e 3 (substituinte no carbono 21 (R7), uma vez que não é planar).

Com esta nova análise, concluímos que o conjunto teste apresentado anteriormente sofre pequenas modificações: os compostos **22, 23, 28, 29, 30 e 31** continuam sendo inativos e os compostos **21, 25, 26 e 27** continuam sendo ativos; apenas os compostos **24 e 32** apresentariam atividades invertidas: o composto **24** passa a ser inativo e o **32**, ativo. Sendo assim, sugerimos para síntese os seguintes compostos: **21, 25, 26, 27 e 32**.

CONCLUSÕES GERAIS

Quando iniciamos este trabalho, o nosso principal objetivo era sugerir novos compostos para síntese; para isto, traçamos um roteiro baseando-nos nos resultados obtidos anteriormente, a fim de elucidarmos o motivo de alguns resultados inesperados e de obtermos maiores esclarecimentos sobre o modo de interação das neolignanas com o receptor do leishmania.

No tópico “objetivos”, descrevemos as diversas etapas que deveríamos seguir; a primeira delas, foi **a inclusão de novos parâmetros físico-químicos**, pois acreditávamos que aqueles usados até então, não estariam explicando adequadamente a atividade anti-leishmaniose. **A inclusão das cargas sobre os carbonos dos anéis fenílicos foi essencial para a separação dos compostos nas duas categorias**, mas foi também nessa etapa que observamos a classificação incorreta dos compostos 5, 6, 7, 10 e 17 (figura 19b, p.87). Os valores da atividade anti-leishmaniose “in vitro”, de algumas neolignanas, nos mostraram que as neolignanas 5, 6 e 10 poderiam pertencer a uma terceira categoria, pois estas formaram um grupo com o composto 13 de atividade baixa (tabela 5 – p.89 e figura 19b, p.87).

A análise realizada com o método SIMCA nos ajudou a definir as variáveis relacionadas com as neolignanas 5, 6, 7, 10 e 17 que formam um grupo separado ou que são agrupadas incorretamente. Como o SIMCA constrói um modelo para cada categoria, a inclusão ou retirada de uma componente principal para determinada categoria, nos mostrou quais variáveis poderiam ser incluídas para melhorar a separação. **Concluimos que deveríamos incluir mais variáveis relacionadas com o anel fenílico 2**, a fim de aproximarmos os compostos inativos 5, 6, 7 e 10 da categoria inativa. Normalmente, o método SIMCA não é usado com esta finalidade, mas apenas para classificação dos compostos em categorias. Sendo assim, **concluimos também que o SIMCA pode ser muito útil na escolha de variáveis**.

A inclusão de outras cargas sobre carbonos do anel 2, aproximou o composto 20 que também possui atividade baixa, dos compostos 5, 6, 10 e 13 agrupados anteriormente (figura 21b, p.103); com isso, pudemos estabelecer uma ordem relativa de atividade anti-leishmaniose, para as neolignanas estudadas, na página 105.

Um outro objetivo nosso, foi verificar se a aplicação de ACP em SAR pode nos levar a resultados confiáveis, uma vez que pelo menos dois dos nossos compostos (7 e 17) continuaram sendo mal classificados. Para isso, aplicamos ACP a vários conjuntos de derivados de DHP (capítulo 6), aos quais possuíamos valores de AB e que já haviam sido estudados com RLM. Pudemos verificar que em alguns dos conjuntos estudados, compostos com valores de atividade biológica muito próximos também foram classificados incorretamente, **em concordância com as nossas conclusões sobre as neolignanas**. Concluímos então, que a **aplicação de ACP em SAR é confiável** e que os **compostos “mal classificados”** possivelmente apresentam **valores de AB próximos** à categoria em que foram agrupados.

No final do capítulo 6, já tínhamos condições de propor novos compostos para síntese, baseando-nos na ordem relativa de atividade proposta por nós (p.105) e considerando que os compostos aparentemente mal classificados, apresentariam AB próximas àquelas da categoria em que estivessem agrupados. No entanto, **foi a solvatação das neolignanas em água, juntamente com a comparação entre os mínimos locais destas e de outros compostos ativos em leishmaniose, que nos proporcionaram talvez, os resultados e conclusões mais importantes sobre a interação das neolignanas com o receptor.**

Concluímos, após os resultados dos capítulos 7 e 8, que a conformação ativa das neolignanas é uma daquelas em que os anéis fenílicos estão mais afastados (**conformação D**), a qual é a única capaz de se sobrepor e interagir com o possível receptor. Reunindo esta conclusão com a de que a interação se dá provavelmente pelos anéis fenílicos (capítulo 5), passamos a investigar mais atentamente, a posição dos substituintes nos anéis fenílicos de cada uma das neolignanas, na conformação D (capítulo 9); verificamos que, algumas das neolignanas inativas com estrutura semelhante às ativas, apresentavam rotação no anel 2 de tal forma a impedir a interação dos seus substituintes com a arginina, tornando-as inativas. Com isso, estabelecemos novo método para propor compostos para síntese, conforme está descrito no final do capítulo 9.

Os resultados obtidos a partir do capítulo 7, tornaram desnecessário o cálculo de parâmetros FQ na conformação D e a aplicação de RP, e por isto, esta etapa dos objetivos foi eliminada.

REFERÊNCIAS

- [1] Costa, M.C.A., Barata, L.E.S. e Takahata, Y., *J. Mol. Struct. (Theochem)*, **340**, 185 (1995).
- [2] Rey, L. "Parasitologia" 2nd ed. Ed. Guanabara Koogan S.A., Rio de Janeiro, RJ, cap. 19 (1991).
- [3] -, *Globo Ciência*, **13**, 32, (1992).
- [4] Barata, L.E.S, 2. Brazilianisch-Deutsches Symposium fur Naturstoffchemie (1991).
- [5] Costa, M.C.A. e Takahata, Y., *J. Comput. Chem.* **18**, 712, (1996).
- [6] Costa, M.C.A., Dissertação de Mestrado, Universidade Estadual de Campinas, Campinas-SP, Brasil, (1993).
- [7] Paine, A., Barata, L. E. S., Neal, R. A., Croft, S. L., Phillipson, J.D., Santos, L. S. e Ferri, P. H., Anti-leishmanial activity of neolignans from *Virola* species and synthetic analogues related to neolignans – manuscrito não publicado (1998).
- [8] Gebre-Hiwot, A. And Frommel, D., *J. Antimicrob. Chemother.*, **32**, 837, (1993).
- [9] Singh, K., Hasan, A., Pratap, R., Guru, P.Y. and Bhakuni, D.S. *J. Indian Chem. Soc.* **66**, 686, (1989)
- [10] Deo, K., Avasthi, K., Pratap, R., Bhakuni, D.S. and Joshi, M.N. *Indian J. Chem.* **5**, 459 (1990).
- [11] Tandon, J.S., Srivastava, V. and Guru, P.Y. *J. Nat. Prod.*, **54** 1102, (1991)
- [12] Adinolfi, L. E. and Bonventre, P. F., *Infect. Immun.* **58** (9) 3151 (1990)
- [13] Mukhopadhyay, R. and Madhubala, R. *Biochem. Pharmacology*, **47**, 253, (1994).
- [14] Steck, E. A. and Kinnamon, K.E. *Experimental Parasitology*, **52**, 404, (1981).
- [15] Dodd, R.H., Quannes, C., Robert-Gero, M. and Potier, P3. *J. Med. Chem.* **32**, 1272, (1989)
- [16] Lawrence, F. and Robert-Gero, M. *J. Euk, Microbiol.*, **40**, 581, (1993).
- [17] Hiwot, A. G and Frommel, D., *J. Antimicrob. Chemother.* **32**, 837, (1993).
- [18] Ghosh, M. And Datta, A. K., *Biochem. J.* **298**, 295, (1994).
- [19] Richards, W.G. "Quantum Pharmacology", Butter & Tanner Ltd., Great Britain, (1977).

- [20] Beveridge, D.L. e DiCapua, F.M. - *Annu. Rev. Biophys. Chem.* **18**, 431, (1989).
- [21] Allen, M.P.; Tildesley, D., "Computes Simulation of Liquids", Clarendon Press, Oxford (1987).
- [22] Freitas, L.C.G., Botelho, L. F., *Química Nova*, **17**, 489, (1994).
- [23] J. Chem. Soc. Faraday Trans., (1993), 90, (12), número especial sobre superfícies de potencial e energias em caminhos de reação.
- [24] Jorgensen, W. L. and Nguyen, T.B., *J. Comp. Chem.*, **14**, 195-205 (1993).
- [25] Burket, U. and Allinger, N.L. "Molecular Mechanics". ACS Monograph, Washington D.C., cap. 1 e 2 (1982).
- [26] Allinger, N.L., *Adv. Phys. Org. Chem.* **13**, 1 (1976)
- [27] Allinger, N. L., *J. Am. Chem. Soc.*, vol. 99, **25**, 8127 (1977)
- [28] Jaime, C. And Osawa, E., *Tetrahedron*, **39**, 17, 2769 (1983).
- [29] Allinger, N. L., et al., *J. Comp. Chem.* Vol. 9, **6**, 591, (1988)
- [30] Martin, Y. C. "Quantitative Drug Design", Marcel Dekker, Inc., New York, N.Y. (1978).
- [31] Kowalski, B.R., "Chemometrics. Mathematics and Statistics in Chemistry", D. Reidel Publ. Comp., Dordrecht, (1984).
- [32] Kowalski, B. R. and Bender, C.F., *J. Am. Chem. Soc.*, **9**, 5632 (1972)
- [33] Wold, S., Esbensen, K. and Geladi, P., *Chem. Intell. Lab. Syst.*, **2**, 37 (1987).
- [34] Sharaf, M.A; Ilman, D.L., Kowalski, B.R., "Chemometrics", Wiley (1986)
- [35] PowerFit v.1.0 (User's Guide), Microsimulations, 478 Green Mountain Road, Mahwah, NJ 07430, (1996).
- [36] Stewart, J.J.P., *J. Comp. Chem.*, **10**, 2, 221 (1989).
- [37] Dewar, M.J.S. and Thiel, W., *J. Am. Chem. Soc.*, **99:15**, 4899 (1977)
- [38] Dewar, M.J.S., et al., *J. Am. Chem. Soc.*, **107**, 3902 (1985)
- [39] Rekker, R. F., *Quant. Struct. - Act. Relat.* **11**, 195 (1992)
- [40] Kubinyi, H., *Progr. Drug. Res.*, **32**, 102, (1979).
- [41] Martin, Y. C., "Quantitative Drug Design", Marcel Dekker, Inc., New York, N.Y., pp. 68-70, 87 (1978).
- [42] Mannhold, R., Rekker, R. F., Sonntag, C., Laak, A. M. T., Dross, K. and Polymeropoulos, E. E., *J. Pharm. Sci.*, **84**, 1410 (1995).

- [43] van de Waterbeend, H. and Mannhold, R., *Quant. Struct.-Act. Relat.* **15**, 410, (1996).
- [44] Higo, J. and Go, N., *J. Comp. Chem.*, **10**, 3, 376 (1989)
- [45] Kireev, D. B., Fetisov, V. I., Zefirov, N. S., *J. Mol., Struct. (Theochem)*, **304**, 143, (1994).
- [46] Guadagnini, P. H., Souza, A. A. e Bruns, R. E., *Química Nova*, **19** 148 (1995)
- [47] Allinger, N. L. et al., "Química Orgânica", 2nd ed. , Guanabara Dois, Rio de Janeiro, RJ, pp. 325, 655, (1978).
- [48] Martin, Y.C. et al. *J. Med. Chem.*, **16**, 1089 (1973).
- [49] Fukui, K. et al., *J. Chem. Phys.*, **20**, 722 (1952).
- [50] Fleming, I., "Frontier Orbitals and Organic Chemical Reactions", London, John Wiley & Sons, Ltd., (1978).
- [51] Shriver, D.F., et al. "Inorganic Chemistry", Oxford University Press (1990).
- [52] Chin, C. and Song, P., "Reactivity Indices for biomolecules", Lubbock, Texas Tech Press (1981).
- [53] Hirsch, J.A., "Concepts in Theoretical Organic Chemistry", Allynand Bacon, Inc., Boston, USA, pp. 88,89, (1975).
- [54] Lien, E.J. et al. *J. Pharm. Sic.*, **71**,6, 641 (1982)
- [55] Garrett, E.R., et al., *J. Med. Chem.*, **9**, 203 (1966).
- [56] Nys, G.G. and Rekker, R.F., *Eur. J. Med. Chem. Chim. Ther.*, **9**, 361 (1974).
- [57] Gaudio, A. C. and Takahata, Y. *J.Comput. Chem.*, **16**, 277 (1992).
- [58] Scarminio, I.S. and Bruns, R.E., *Trends Anal. Chem.* **8**, 326 (1989)
- [59] Pirouette, Multivariate Data Analysis for IBM PC Systems, version 2.0, Infometrix, Seattle, WA. (1996).
- [60] Pauline, J and Wildman, T.A., *J.Phys.Chem.*, **95**, 20 (1991).
- [61] Ramachandran, G.N., Mazumdar, S.K., Venkatesan, K. And Lakshminarayanan, A. *V. J.Mol. Biol.* **15**, 232-242 (1966).
- [62] Programa Diadorim, escrito em linguagem Fortran por Luiz Carlos Gomide Freitas, Departamento de Química, UFSCar, 1992.
- [63] Niemiin, G.J., in W. Darcher and J. Devilers (Eds.), "Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology", Dluwer Academic Publishers, Dordrecht, p. 153, (1990)

- [64] Costa, M.C.A., Gaudio, A. C., and Takahata, Y., *J.Mol. Struct. (Theochem)*, **394**, 291,(1997).
- [65] Janis, R. A.and Silver, P.*J. Adv. Drug. Res.*, **16**, 309, (1987).
- [66] Berntsson, P. And Wold, S, *Quant.Struct.-Act. Relat.* , **5**, 45, (1986).
- [67] Mahmodian, M. And Richards, W. G., *J. Pharm. Pharmacol.*, **38** , 272, (1986).
- [68] Gaudio, A. C., Korolkovas,A. and Takahata, Y., *J. Pharm. Sci*, **83** , 1410 (1994).
- [69] Lehninger, A. L. "Bioquímica", Edgard Blucher ed., vol. 1, São Paulo, p. 150-161 (1976).