

UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Química

Departamento de Físico-Química

**Aplicações de métodos quimiométricos no estudo da relação
estrutura-atividade de: *a)* Congêneros de bifenis policlorados
b) Derivados de 4-Aminometiltioxantenona e
5-Aminometilbenzotiopiranoindazol**

Lucicleide Ribeiro Cirino

Orientadora Profa. Dra. Márcia M. C. Ferreira

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Química do Instituto de Química da Universidade Estadual de Campinas como parte dos requisitos para obtenção do título de Doutor em Físico-Química.

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

Cirino, Lucicleide Ribeiro

C496a Aplicações de métodos quimiométricos no estudo da relação estrutura-atividade de *a)* congêneros de bifenis policlorados; *b)* derivados de 4-aminometiltioxantenona e 5-aminometilbenzotilpiranoindazol / Lucicleide Ribeiro Cirino. — Campinas, SP: [s, n], 2001.

Orientadora: Márcia M. C. Ferreira.

Tese (doutorado) — Universidade Estadual de Campinas.
Instituto de Química.

1. Quimiometria. 2. QSAR. 3. PCBs. 4. Drogas anti-tumor.
I. Ferreira, Márcia M. C.. II. Universidade Estadual de Campinas.
Instituto de Química. III. Título.

Agradeço

a Deus por ter me proporcionado chegar até aqui, sei que cada ser tem uma missão, mesmo que não esteja clara ou compreensível aos olhos humanos.

ao Jair por todo apoio, incentivo e amor.

à Profa. Dra. Márcia M. C. Ferreira pela oportunidade de realizar este trabalho, por sua orientação, pela amizade e ótima convivência.

à Denise Mills e ao Prof. Dr. Subhash Basak pela utilização do programa POLLY.

Ao Dr. Rudolf Kiralj pelas sugestões no tratamento das drogas.

Aos colegas e amigos Aline, Cris, Cristiano, Edilton, Edilson, Fabiana, Hermes, Marlon, Maximiliano, Sayuri e Shim.

a todo o pessoal do CENAPAD-SP.

à Bel, André e Celi da coordenadoria de pós-graduação.

ao CNPq pelo auxílio financeiro concedido.

Às pessoas que mais amo:

Jair

meus pais

meu irmão

CURRICULUM VITAE

1. DADOS PESSOAIS

Nome: Lucicleide Ribeiro Cirino

Nascimento: 17/03/1972, São João de Meriti/RJ - Brasil

2. FORMAÇÃO ACADÊMICA/TITULAÇÃO

1997-2001 Doutorado em Química.

Universidade Estadual de Campinas, UNICAMP, São Paulo, Brasil.

Título: Estudo da relação estrutura atividade de: a) bifenis policlorados.

b) derivados de 4-aminometiltioxantenona e 5-aminometilbenzotilpiranoindazol.

Orientadora: Profa. Dra. Márcia Miguel Castro Ferreira.

Bolsista do Conselho Nacional de Desenvolvimento Científico e Tecnológico/DF.

1995-1997 Mestrado em Física.

Universidade Estadual de Campinas, UNICAMP, São Paulo, Brasil.

Título: Fotoluminescência resolvida no tempo em $a - Si_{1-x}C_x : H$.

Orientador: Prof. Dr. Leandro Russovski Tessler.

Bolsista do Conselho Nacional de Desenvolvimento Científico e Tecnológico/DF.

1991-1994 Graduação em Física.

Universidade Federal do Rio de Janeiro, UFRJ, Rio de Janeiro, Brasil.

3. PRODUÇÃO CIENTÍFICA

3.1 Trabalhos completos em eventos

- i. CIRINO, Lucicleide Ribeiro, CIRINO, José Jair Vianna, FERREIRA, Márcia Miguel Castro. *Activity/cytotoxicity qsar of 4-aminomethylthioxanthenone and 5-aminomethylbenzothiopyranoindazole derivatives*. In: 13TH EUROPEAN SYMPOSIUM ON QSAR, 2000, Düsseldorf. Rational Approaches To Drug Design - Abstract Book. 2000. v.1. p.63.
- ii. CIRINO, Lucicleide Ribeiro, CIRINO, José Jair Vianna, FERREIRA, Márcia Miguel Castro. *PCB's: estudo da relação entre coeficiente de partição octanol-água e propriedades físico-químicas*. In: X SIMPÓSIO BRASILEIRO DE QUÍMICA TEÓRICA, 1999, Caxambú. Livro do Resumos do X SBQT. 1999.

3.2 Trabalhos resumidos em eventos

- i. CIRINO, Lucicleide Ribeiro, RIBEIRO, Fabiana A L, FERREIRA, Márcia Miguel Castro. *Correlação quantitativa estrutura-propriedades de hidrocarbonetos policíclicos aromáticos*. In: 23ª REUNIÃO ANUAL DA SBQ, 2000, Poços de Caldas. Livro de Resumos Md-057. 2000. v.2. p.57.
- ii. NAGATA, Noemi, CIRINO, Lucicleide Ribeiro, FERREIRA, Márcia Miguel Castro. *Utilização de qsar e calibração multivariada para avaliar a toxicidade aguda em vibrio fisheri. ii. compostos narcóticos*. In: X SIMPÓSIO BRASILEIRO DE QUÍMICA TEÓRICA, 1999, Caxambú. Livro do Resumos do X SBQT. 1999. v.1. p.136-136.
- iii. CIRINO, Lucicleide Ribeiro, MARÍNEZ, Luz Elena Bolívar. *Flutuação de potencial em a – Si_{1-x}C_x : H*. In: IX SIMPÓSIO BRASILEIRO DE QUÍMICA TEÓRICA, 1997, Caxambú. Livro de Resumos do IX SBQT. 1997.
- iv. CIRINO, Lucicleide Ribeiro, TESSLER, Leandro Russovski. *Dcaimento da fotoluminescência em a – Si_{1-x}C_x : H*. In: XIX ENCONTRO NACIONAL DE FÍSICA DA MATÉRIA CONDENSADA, 1996, Águas de Lindóia. Livro de Resumos do XIX Encontro Nacional de Física da Matéria Condensada. 1996.
- v. ALMEIDA, Danilo Paiva, CARLOS, Antônio, CIRINO, Lucicleide Ribeiro. *Electron impact multiple ionization cross section of neon*. In: VII SIMPÓSIO BRASILEIRO DE QUÍMICA TEÓRICA, 1993, Caxambú. Livro do Resumos da VII SBQT. 1993.
- vi. ALMEIDA, Danilo Paiva, CARLOS, Antônio, CIRINO, Lucicleide Ribeiro. *Simulação de monte-carlo para um espectrômetro de massa por tempo de voo*. In: XVI REUNIÃO DE TRABALHOS SOBRE FÍSICA NUCLEAR NO BRASIL, 1993, Serra Negra. 1993.

3.3 Artigos completos publicados em periódicos

- i. TESSLER, Leandro Russovski, CIRINO, Lucicleide Ribeiro. *Time resolved photoluminescence of a-Si_{1-x}C_x : H*. Mrs Symp. Proc. Amorphous Silicon Technology, San Francisco, v.420, p.783-786, 1996.

4. ATIVIDADE ACADÊMICA LIGADA AO ENSINO

4.1 Auxiliar Didático

- i. Cálculo Integral e Diferencial II e III — Segundo semestre de 1992.
- ii. Mecânica Clássica — Segundo semestre de 1995 e primeiro semestre de 1996.
- iii. Física II — Segundo semestre de 1996.

Conteúdo

Lista de Tabelas	xiii
Lista de Figuras	xvii
Resumo	xxi
Abstract	xxiii
1 Introdução	1
1.1 Bifenis Policlorados	3
1.2 Derivados de 4-aminometiltioxantenona e 5-aminometilbenzotilpiranoindazol	5
2 Metodologia	7
2.1 Análise Conformacional	8
2.2 Métodos Semi-Empíricos	8
2.3 Parâmetros Utilizados	11
2.3.1 Parâmetros Físicos–Químicos	12
2.3.2 Parâmetros Topológicos	16
2.3.3 Parâmetros Estéricos	20
2.4 Métodos Quimiométricos	20
2.4.1 Pré-Processamentos	20
2.4.2 Análise por Componentes Principais	21
2.4.3 Regressão Linear Múltipla	23
2.4.4 Regressão por Componentes Principais	24
2.4.5 Regressão por Mínimos Quadrados Parciais	25

2.4.6	Validação	26
2.5	Principais Programas Utilizados	28
3	Bifenis Policlorados	31
3.1	Coefficiente de Partição octanol-água	34
3.2	Solubilidade	40
3.3	Coefficiente de Partição carbono orgânico	47
3.4	Constante da lei de Henry	52
3.5	Coefficiente de Atividade Aquosa	57
3.6	Ponto de Fusão	63
3.7	Toxicidade IC ₅₀	69
3.8	Toxicidade EC ₅₀	73
3.9	Propriedades Eletrônicas	77
3.10	Conclusão	90
4	Derivados de 4M e 5M	91
4.1	Intercalação DNA-droga	94
4.2	Propriedades Eletrônicas	98
4.3	Análise por Componentes Principais	103
4.4	Atividade EC ₅₀	105
4.5	Conclusão	109
A	Métodos Ab Initio	111
A.1	Conjuntos de Base	112
A.2	Conjuntos de Base 6-31G, 6-311G e MC-311G	112
B	Nomenclatura dos PCBs.	115
C	Tabela de Dados	119

Lista de Tabelas

1.1	Classificação com o número de congêneros de bifenis em cada grupo homólogo.	4
3.1	Comparação dos valores obtidos por cálculos AM1 para os ângulo entre os planos dos anéis de alguns PCBs, com resultados de cálculos ab initio (STO-3G) da literatura.	33
3.2	Porcentagem da variância acumulada pelo modelo para $\log K_{OW}$	38
3.3	<i>Loadings</i> das variáveis no modelo para $\log K_{OW}$	38
3.4	Vetor de regressão para $\log K_{OW}$. $^2\chi$ é o índice de conectividade de caminho de ordem 2, e $^6\chi_{ch}$ o índice de conectividade de cadeia de ordem 6.	38
3.5	Valor de $\log K_{OW}$ previsto e experimental para algumas moléculas.	38
3.6	Valor Previsto e Experimental de $\log K_{OW}$	39
3.7	Porcentagem da variância acumulada pelo modelo para $\log S$	43
3.8	<i>Loadings</i> das variáveis acumulada pelo modelo para $\log S$	43
3.9	Vetor de regressão para $\log S$. $^6\chi^b$ é o índice de complexidade de vértice, e P_9 o número de caminho de comprimento 9.	43
3.10	Comparação do valor de $\log S$ previsto e experimental para validação. . . .	43
3.11	Valor Previsto e Experimental de $\log S$	44
3.12	Momento de dipolo para as moléculas selecionadas com o mesmo número de átomos de cloro.	45
3.13	Momento de dipolo para as moléculas selecionadas com número de átomos de cloro diferentes.	46
3.14	Porcentagem da variância acumulada pelo modelo para $\log K_{OC}$	49
3.15	<i>Loadings</i> das variáveis no modelo para $\log K_{OC}$	49
3.16	Vetor de regressão para $\log K_{OC}$. H^D índice de complexidade de distância, e W o índice de Winer.	49

3.17	Comparação do valor de $\log K_{OC}$ previsto e experimental para validação. . .	49
3.18	Valor Previsto e Experimental de $\log K_{OC}$	50
3.19	Porcentagem da variância acumulada pelo modelo para $\log H$	55
3.20	<i>Loadings</i> das variáveis no modelo para $\log H$	55
3.21	Vetor de regressão para $\log H$	55
3.22	Comparação do valor de $\log H$ previsto e experimental para validação. . . .	55
3.23	Valor Previsto e Experimental de $\log H$ ($\text{Pa } m^3/\text{mol}$).	56
3.24	Porcentagem da variância acumulada pelo modelo para $-\log Y_W$	59
3.25	<i>Loadings</i> das variáveis no modelo par $-\log Y_W$	59
3.26	Vetor de regressão para $-\log Y_W$	59
3.27	Comparação do valor de $-\log Y_W$ previsto e experimental para validação. . .	59
3.28	Valor Previsto e Experimental de $-\log Y_W$	60
3.29	Porcentagem da variância acumulada pelo modelo para mp	65
3.30	<i>Loadings</i> das variáveis no modelo pra mp	65
3.31	Vetor de regressão para mp	65
3.32	Comparação entre o valor de mp previsto e experimental para validação. . .	65
3.33	Valor Previsto e Experimental de mp	66
3.34	Porcentagem da variância acumulada pelo modelo para $\log IC_{50}$	71
3.35	<i>Loadings</i> das variáveis no modelo para $\log IC_{50}$	71
3.36	Vetor de regressão para $\log IC_{50}$	71
3.37	Comparação entre o valor de $\log IC_{50}$ previsto e experimental para validação.	71
3.38	Valor Previsto e Experimental de $\log IC_{50}$	72
3.39	Porcentagem da variância acumulada pelo modelo para $\log EC_{50}$	75
3.40	<i>Loadings</i> das variáveis no modelo para $\log EC_{50}$	75
3.41	Vetor de regressão para $\log EC_{50}$. ${}^5\chi^v$ o índice de conectividade de valência de ordem 5, e ${}^6\chi_{pc}^v$ o índice de conectividade de cluster de valência de ordem 6.	75
3.42	Comparação entre o valor de $\log EC_{50}$ previsto e experimental para validação.	75
3.43	Valor Previsto e Experimental de $\log EC_{50}$	76
4.1	Tabela com as moléculas estudadas. * moléculas consideradas inativas. . . .	93
4.2	Porcentagem da variância acumulada pelo modelo para $\log EC_{50}$	107
4.3	<i>Loadings</i> das variáveis no modelo para $\log EC_{50}$	107

4.4	Vetor de regressão para $\log EC_{50}$. Bac é o índice de Balaban e Ss é a soma de índices eletrotopológicos.	107
4.5	Valor Previsto e Experimental de $\log EC_{50}$	108
B.1	Estrutura das moléculas de PCBs.	116
C.1	Variáveis utilizadas nos modelos das moléculas de PCB.	119
C.2	Variáveis utilizadas nos modelos das moléculas de PCB.	124
C.3	Propriedades previstas para as moléculas de PCB.	132
C.4	Variáveis utilizadas nos modelos das moléculas de 4M e 5M. Valores de $\log EC_{50}$ previstos.	141

Lista de Figuras

1.1	Esquema das moléculas de bifenis policlorados.	4
2.1	Molécula de 2-metilpentano e sua representação de grafo.	16
2.2	Pré-processamento de dados. Os dados de cada variável são representados por uma barra de variância e seu centro. A) Dados sem pré-processamento. B) Resultado após centrar na média. C) Resultado depois de escalar pela variância. D) Dados autoescalados.	22
2.3	Representação de uma componente principal para o caso de duas variáveis $\mathbf{x}_1$ e $\mathbf{x}_2$. A) Os <i>loadings</i> (p_1 e p_2) são os cosenos do vetor direção. B) Os <i>scores</i> (t_1 e t_2) são as projeções das amostras na PC.	23
3.1	Barreira de rotação para molécula de bifenil.	32
3.2	Gráfico de $\log K_{OW}$ versus descritores, mostrando a correlação.	37
3.3	Gráfico dos valores previstos e experimentais de $\log K_{OW}$	37
3.4	Gráfico do resíduo de <i>Student</i> versus a <i>leverage</i> obtido pelo modelo escolhido.	37
3.5	Gráfico de $\log S$ versus descritores, mostrando a correlação.	42
3.6	Gráfico do valor previsto e o experimental de $\log S$	42
3.7	Moléculas de PCBs.	46
3.8	Gráfico de $\log K_{OC}$ versus descritores, mostrando a correlação.	48
3.9	a) Gráfico de $\log K_{oc}$ previsto versus $\log K_{oc}$ experimental.	48
3.10	Gráfico de $\log H$ versus descritores, mostrando a correlação.	54
3.11	Gráfico de $\log H$ experimental versus $\log H$ previsto obtido pelo modelo.	54
3.12	Gráfico de $-\log Y_W$ versus descritores, mostrando a correlação.	58
3.13	Gráfico dos valores de $-\log Y_W$ experimental versus os valores previstos pelo modelo.	58
3.14	Gráfico <i>mp</i> versus descritores, mostrando a correlação.	64

3.15 a) Gráfico de <i>mp</i> versus o volume. b) Gráfico da variação do volume para PCBs com o mesmo número de átomos de cloro.	64
3.16 Gráfico dos valores <i>mp</i> experimental versus os valores previstos pelo modelo	66
3.17 Gráfico de log IC ₅₀ versus descritores, mostrando a correlação.	70
3.18 Gráfico de log IC ₅₀ previsto pelo modelo versus o experimental.	72
3.19 Gráfico de log EC ₅₀ versus descritores, mostrando a correlação.	74
3.20 a) Gráfico de log EC ₅₀ previsto e experimental.	74
3.21 Propriedades eletrônicas da molécula 2-CB.	79
3.22 Propriedades eletrônicas da molécula 3-CB.	80
3.23 Propriedades eletrônicas da molécula 2,6-CB.	81
3.24 Propriedades eletrônicas da molécula 3,4-CB.	82
3.25 Propriedades eletrônicas da molécula 2,3,6-CB.	83
3.26 Propriedades eletrônicas da molécula 3,3',4-CB.	84
3.27 Propriedades eletrônicas da molécula 2,2',3,3'-CB.	85
3.28 Propriedades eletrônicas da molécula 3,3',4,5,5'-CB.	86
3.29 Propriedades eletrônicas da molécula 2,2',4,4',6,6'-CB.	87
3.30 Propriedades eletrônicas da molécula 2,2',3,4,4',5,6,6'-CB.	88
3.31 Propriedades eletrônicas da molécula 2,2',3,3',4,4',5,5',6,6'-CB.	89
4.1 Esquema das drogas estudadas: a) moléculas de 1 até 44 da tabela 4.1; b) moléculas de 45 até 55.	92
4.2 Geometria da molécula de hycantone, obtida por cristalografia.	92
4.3 Estrutura cristalográfica de 5' – d(CpGpTpApCpG) – 3' complexado com <i>N, N – bis</i> (aminopropil) – 2 – anthraquinonesulfonamida.	95
4.4 Estrutura cristalográfica de d(CGGCCG) complexado com adriamicina.	95
4.5 Distâncias entre alguns átomos da molécula de sulfonamida intercalada e o DNA.	96
4.6 Proposta da intercalação da molécula 1 com o DNA.	97
4.7 Orbitais moleculares de HOMO e LUMO, e superfície se potencial eletrostático das molécula 2 (ativa) e 34 (inativa).	99
4.8 Orbitais moleculares de HOMO e LUMO, e superfície se potencial eletrostático das molécula 4 (ativa) e 35 (inativa).	100
4.9 Orbitais moleculares de HOMO e LUMO, e superfície se potencial eletrostático das molécula 8 (ativa) e 36 (inativa).	101

4.10	Orbitais moleculares de HOMO e LUMO, e superfície se potencial eletro- tático de duas moléculas inativas (11 e 14).	102
4.11	Gráfico dos <i>scores</i> mostrando a separação entre drogas ativas e inativas. b) Gráfico dos <i>loadings</i> na primeira e terceira PCs.	104
4.12	Gráfico de log EC ₅₀ versus descritores.	106
4.13	Gráfico de log EC ₅₀ previsto e experimental.	106
B.1	Esquema das moléculas de bifenis policlorados.	115

Resumo

Métodos quimiométricos foram utilizados com a finalidade de construir modelos preditivos de propriedades e atividade biológicas de várias moléculas. Descritores eletrônicos, estéricos e topológicos foram calculados e utilizados na construção dos modelos. Os cálculos desses descritores foram feitos com a utilização do método semi-empírico AM1, após a análise conformacional das moléculas.

Na primeira parte do trabalho apresentamos os resultados obtidos no estudo de bifenis policlorados (PCBs), que são moléculas de interesse ambiental. A análise conformacional feita na molécula de bifenil, utilizando o método *ab initio* em nível HF/6-31G*, mostra que o ângulo entre os anéis está em torno de 45°. Esta molécula foi utilizada como base na construção das 209 moléculas de PCBs. As propriedades estudadas foram o coeficiente de partição octanol-água, coeficiente de partição carbono-orgânico, solubilidade, constante da lei de Henry, coeficiente de atividade aquosa e ponto de fusão. As atividades biológicas estudadas foram as toxicidades EC₅₀ e IC₅₀. Foi possível obter modelos multivariados para todas as propriedades e atividades biológicas estudadas. Os resultados obtidos para as toxicidades indicaram que quanto mais lipofílica mais tóxica será a molécula de PCB.

Na segunda parte deste trabalho, estudamos um grupo de 55 moléculas derivadas de 4-aminometiltioxantenona e 5-aminometilbenzotioipiranoindazol. Algumas moléculas deste grupo apresentam atividade anti-tumor, enquanto outras são consideradas inativas. A atividade estudada foi EC₅₀. As propriedades determinantes no comportamento ativo ou inativo das drogas foram a área e a eletronegatividade. Os orbitais de HOMO, LUMO e superfície de potencial eletrostático não apresentam diferenças significativas. Uma análise da estrutura destas moléculas, juntamente com os resultados obtidos pelos métodos quimiométricos mostra que o substituinte *R*⁷ deve ficar o mais afastado possível da estrutura básica da molécula para que ela seja ativa.

Abstract

Chemometrical methods were used to build models to predict biological activities and physical properties of a big set of molecules. Eletronic, steric, and topological descriptors were used to build those models. The AM1 semiempirical method was used to calculate them, after the conformational analysis investigation.

At first, are presented the results obtained in the study of the PCBs (Polychlorinated Biphenis) which are molecules of enviromental interest. The conformational analysis done on the biphenyl molecule, at the HF-631G* level theory, showed that the dihedral angle between the two rings is about 45°. This molecule was used to buil the 209 PCBs derivatives. The studied properties were: octanol/water partition coefficient, organic carbon partition coefficient, solubility, Henry's Law constant, aqueous activity coefficient and melting point. The biological activities studied were the EC₅₀ and IC₅₀. It was possible to obtain multivariate models for all physical properties and biological activities studied. The obtained results on toxicity indicated that the more lipophilic the PCB molecule is, more toxic it will be.

The last part of this work consider the study of 55 derivatives of 4-aminomethylthioxanthenone and 5-aminomethylbenzothiopyranoindazol. Some of these molecules have antitumor activity while others are considered inactive. The evaluated activity was the EC₅₀. The behaviour of the studied drugs as active or inactive were determined by its the electronegativity and surface area. The HOMO and LUMO molecular orbitals and also electrostatic potential show no significant differences. Structural investigations of these molecules, together with the results obtained by means of chemometrical methods, showed that the substituent named R^7 should be as farther as possible from the molecule's basic structure to assure its activity.

Capítulo 1

Introdução

O estudo das relações entre estrutura e atividade, e relações entre estrutura e propriedade é de grande importância na bioquímica, química medicinal e ambiental, tanto para estudar e propor novos compostos com características de interesse, como no estudo de propriedades e características de compostos já existentes.

A modelagem molecular é um processo iterativo usado para estudar as relações propostas entre as características químicas de uma molécula, ou série de moléculas, à sua atividade biológica. Com um detalhamento dos processos biológicos responsáveis pela atividade, a hipótese geralmente é refinada pelo estudo da similaridade estrutural e das diferenças entre moléculas ativas e inativas.

Experimentalmente as possibilidades combinatórias para esta estratégia em um sistema simples pode ser numericamente explosiva o que exigiria altos investimentos financeiros além de não haver garantia de se obter resultados positivos. A fim de reduzir os custos e aumentar a eficiência da pesquisa, é indispensável a utilização de novas ferramentas de trabalho. Uma alternativa é analisar quantitativamente variações de atividades biológicas com descritores moleculares, que são as características moleculares relacionadas com a estrutura, os quais podem ser obtidos para cada composto. As análises de QSAR (Quantitative Structure Activity Relationship) e QSPR (Quantitative Structure Properties Relationship)¹⁻³ têm sido utilizadas com sucesso.

Os modelos de QSAR são matemáticos e correlacionam a estrutura molecular à atividade biológica (ex. toxicidade). Esses modelos se baseiam na aceitação de que a estrutura da molécula, isto é, sua geometria e propriedades eletrônicas, contém as características responsáveis pelas suas propriedades físicas, químicas e biológicas. Ou seja, propriedades

moleculares são determinadas pela estrutura química. Assumindo que a toxicidade de uma substância é determinada por suas propriedades, então, há uma relação estrutura, propriedade e atividade. A formulação de uma relação estrutura-atividade envolve três componentes. O primeiro são os dados da atividade para um grupo de compostos. Segundo, os dados de propriedades (descritores físico-químicos) para estes compostos. Terceiro, estes dados podem ser relacionados por meio de uma função matemática.

$$Atividade = f(descritores) \quad (1.1)$$

Compostos são normalmente considerados como estruturas bidimensionais. Entretanto, as atividades biológicas são manifestações de estruturas de moléculas tridimensionais, assim como as propriedades moleculares. Então é importante: (1) determinar com acurácia as variações na estrutura que são consistentes com a produção de uma atividade específica, (2) definir os caminhos em que as alterações na estrutura, e conseqüentemente nas propriedades de um composto, influenciam na atividade.

Uma das mais eficientes ferramentas para alcançar estes objetivos é a utilização da quimiometria, área da química que utiliza métodos matemáticos e estatísticos, permitindo a obtenção do máximo de informações de problemas complexos, com múltiplas variáveis (neste caso, os descritores).

Existem muitos exemplos da aplicação de métodos quimiométricos no estudo de QSAR, como o trabalho divulgado pelo *The Sandoz Institute for Medical Research* no desenvolvimento de novos agentes analgésicos,⁴ o estudo de N²-fenilguanina como inibitor da herpes,⁵ o estudo dos efeitos da toxicidade de benzil-álcool.⁶ Também podemos citar o estudo dos mecanismos de ação da toxicidade para *Vibrio fischeri*,⁷ o estudo para prever a toxicidade de hidrocarbonetos halogenados,⁸ o estudo de mutagenicidade de um grupo de moléculas de hidrocarbonetos clorados utilizando 11 descritores estruturais, entre os quais, as cargas calculadas e a eletronegatividade.⁹ Podemos citar também alguns trabalhos do nosso grupo,¹⁰⁻¹² e o trabalho sobre naftoquinonas,¹³ desenvolvido junto com o grupo de química-quântica da USP-São Carlos.

Neste trabalho, o objetivo foi obter relações entre estrutura, propriedades e atividades, utilizando-se métodos quimiométricos. Os sistemas objetos de estudo desta tese são: um sistema composto por moléculas de interesse ambiental, e um segundo sistema composto por uma série de moléculas de interesse medicinal.

O presente trabalho foi desenvolvido segundo a metodologia abordada no Capítulo

2, onde se descreve resumidamente os métodos quimiométricos e mecânico-quânticos utilizados. No Capítulo 3, são estudadas as relações quantitativas obtidas para os PCBs. As relações quantitativas obtidas para 4-aminometiltioxantenona e 5-aminometilbenzotilpiraindoindazol são exploradas no Capítulo 4.

1.1 Bifenis Policlorados

Na primeira parte desta tese, estudamos um grupo de compostos chamados bifenis policlorados, aos quais nos referiremos como PCBs. Atualmente há mais de 60.000 compostos químicos em produção. Destes, cerca de 500 são de interesse ambiental, por causa da sua presença em quantidade detectável em vários componentes do meio ambiente, sua toxicidade, sua tendência à bioacumulação e sua persistência. Todas estas características despertam grande interesse no estudo destes compostos, com a finalidade de se obter informações detalhadas sobre o comportamento dos mesmos no meio ambiente.¹⁴ Exemplos de alguns grupos desses compostos são: bifenis policlorados (PCBs), fenóis, clorobenzenos, dioxinas policloradas, hidrocarbonetos aromáticos policíclicos e dibenzofuranos.

Os PCBs constituem uma família de compostos formados pela adição de átomos de cloro ao bifenil, que consiste de dois anéis benzênicos ligados por uma ligação simples carbono-carbono (fig. 1.1). Os PCBs possuem fórmula geral $C_{12}H_nCl_y$, com $n = 0, \dots, 10$ e $y = 10 - n$, de maneira que existem 209 congêneros de PCBs, que são organizados em subcategorias, os homólogos, formados por diferentes congêneros que tenham o mesmo número de átomos de cloro¹⁵ (tab.1). O número de identificação de cada congênero (de 1-209) foi criado por Ballschmiter e Zell¹⁶ e mais tarde adotado pela IUPAC. Outra maneira de se referir a cada composto é indicando a posição do átomo de cloro no anel, de acordo com a figura 1.1, seguida pelo termo CB (ex. 2, 2', 3-CB). Na tabela B.1 (apêndice B), está o número de identificação de cada congênero, juntamente com a posição do átomo de cloro no anel.

Os PCBs são largamente utilizados na indústria como fluidos de transferência de calor, fluidos hidráulicos, diluentes orgânicos, fluidos dielétricos em transformadores elétricos e capacitores. A versatilidade industrial usual dos PCBs está diretamente relacionada às suas propriedades físicas que incluem resistência à oxidação e redução, excelente propriedade de isolante elétrico, estabilidade térmica e não inflamabilidade.¹⁷

Infelizmente estas propriedades físicas, juntamente com seu amplo uso e descarte impróprio, tem resultado na contaminação de cada componente do ecossistema global por

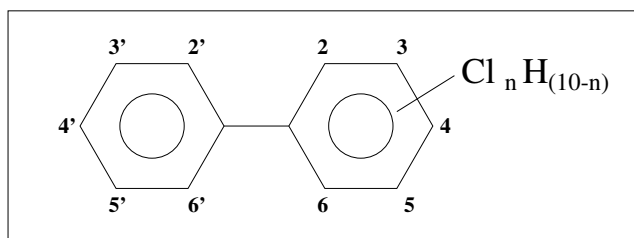


Figura 1.1: Esquema das moléculas de bifenis policlorados.

Tabela 1.1: Classificação com o número de congêneros de bifenis em cada grupo homólogo.

Homólogo	Número de substituintes (Cl)	Número de congêneros
Monoclorobifenil	1	3
Diclorobifenil	2	12
Triclorobifenil	3	24
Tetraclorobifenil	4	42
Pentaclorobifenil	5	46
Hexaclorobifenil	6	42
Heptaclorobifenil	7	24
Octaclorobifenil	8	12
Nonaclorobifenil	9	3
Decaclorobifenil	10	1

PCBs. Nos seres humanos os problemas de saúde que podem ocorrer dependem da dose de exposição. Alguns problemas são: irritação na pele e olhos, depressão, perda de memória, fadiga, perda de sono, perda de apetite e peso, náusea e vômito. Os PCBs são considerados cancerígenos, além de afetarem o sistema reprodutivo, e causarem danos irreversíveis em filhos de mulheres que sofreram algum tipo de exposição.¹⁸

Os PCBs são encontrados no ar, água e solo, sendo considerados virtualmente insolúveis em água, são absorvidos por sedimentos em rios, lagos e oceanos. A natureza lipofílica e a persistência dos PCBs contribuem para seu alto potencial de bioacumulação nos níveis mais altos da cadeia alimentar. Isto é, resíduos de PCBs são rotineiramente detectados em peixes e tecido adiposo humano.¹⁹

A dinâmica dos PCBs no meio ambiente é determinada por suas propriedades químicas (hidrofobicidade, persistência) e pelos processos biológicos (estrutura da rede alimentar, eficiência de assimilação).

Aceita-se cada vez mais o fato de que o único método rigoroso de resolver os proble-

mas causados pelos PCBs no meio ambiente é obter as propriedades de todos os congêneros, e tratar cada congênero separadamente. Desta forma, estudar a dinâmica dos PCBs no meio ambiente exige a medida de vários parâmetros. Infelizmente poucos congêneros são sintetizados em quantidade suficiente para que seja possível a determinação de suas propriedades,²⁰ além das dificuldades experimentais e do alto custo envolvido.

1.2 Derivados de 4-aminometiltioxantenona e 5-aminometilbenzotilpiranoindazol

Câncer (ou neoplasia, ou tumor maligno) é uma classe de doenças caracterizadas pelo crescimento descontrolado de células aberrantes. O câncer mata pela invasão destrutiva de órgãos normais por estas células, o que se dá por extensão direta ou por disseminação à distância, pelo sangue, linfa ou superfície serosa. O comportamento anormal das células cancerosas é geralmente espelhado por mutações genéticas, expressões de características oncológicas, ou secreção anormal de hormônios ou enzimas. Todos os cânceres invadem ou se metastatizam, mas cada tipo específico tem características clínicas e biológicas, que devem ser estudadas para um adequado diagnóstico, tratamento e acompanhamento.

Vários agentes podem causar ou contribuir diretamente para uma seqüência de eventos que levem a um meio do câncer se desenvolver. O caminho final comum dos cânceres é alguma alteração genética, que converte uma célula bem constituída, participante do corpo como um todo, numa outra, “renegada”, destrutiva, que não responde mais aos comandos de uma comunidade de células. Promotores (oncogenes) e supressores têm um papel central em muitos casos. Substâncias químicas como o benzeno, nitrosaminas, agentes físicos, como radiação gama e ultravioleta, e agentes biológicos, como alguns tipos de vírus, contribuem para a carcinogênese em algumas circunstâncias. O agente carcinogênico mais importante para a população em geral é o tabaco, pois ele causa ou contribui para o desenvolvimento de aproximadamente um terço de todos os cânceres, principalmente em pulmão, esôfago, bexiga, cabeça e pescoço.

Entre as modalidades terapêuticas utilizadas está a quimioterapia, considerada o primeiro tratamento sistemático para o câncer. Na maioria da vezes consiste em uma associação de drogas, pouco eficazes se utilizadas sozinhas, pois nos tumores existem subpopulações de células com sensibilidade diferente às drogas antineoplásicas. Os mecanismos de ação das drogas são diferentes, mas sempre acabam em lesão de DNA celular. A

toxicidade contra células normais é a causa dos efeitos colaterais (náuseas, vômitos). A quimioterapia pode ser usada como tratamento principal (leucemias, linfomas, câncer de testículo), mas normalmente é adjuvante, após tratamento cirúrgico ou radioterápico.

Pesquisas sobre câncer procuram obter drogas com eficácia cada vez maior. Entender a relação da eficácia com a estrutura possibilitaria o desenvolvimento de novas drogas, mais eficientes. Mas como já foi mencionado, sintetizar e testar estas drogas nem sempre alcança o objetivo desejado, além de requerer altos gastos.

Na segunda parte do nosso trabalho, estudamos um grupo de drogas com atividade anti-tumor, que são os derivados de 4-aminometiltioxantenona (4M)* e 5-aminometilbenzotiopiranoindazol (5M).

Apesar do uso de cálculos teóricos em estudos de QSAR para moléculas biológicas parecer inconsistente, visto que os cálculos são feitos para molécula isolada e no vácuo, e os processos ocorrem em ambiente biológico, alguns pontos devem ser considerados. Apesar de ser possível realizar cálculos simulando um meio dielétrico, que estaria mais próximo do real, ou cálculos da molécula solvatada, isto envolve um alto custo computacional. Por outro lado, existe na literatura aplicações de cálculos teóricos no estudo de moléculas biológicas, com bons resultados, o que mostra ser viável esta metodologia, desde que se respeite suas limitações. Por fim, a partir dos resultados obtidos é possível ter uma idéia do comportamento da droga, mas não se tem a ambição de explicar os processos envolvidos, os quais consistem em mecanismos complexos.

**Esta nomenclatura não é oficial, e só será utilizada por nós por uma questão de praticidade.*

Capítulo 2

Metodologia

Neste capítulo descrevemos as principais etapas do nosso trabalho, assim como os métodos utilizados.

Inicialmente foi feita uma análise conformacional, para todas as moléculas*, utilizando o método semi-empírico AM1.²¹ Em seguida foram calculadas as propriedades moleculares. Os resultados dos cálculos forneceram os valores das energias do HOMO (Highest Occupied Molecular Orbital) e LUMO (Lowest Unoccupied Molecular Orbital), o calor de formação (ΔH), momento de dipolo total e suas componentes, a polarizabilidade e as cargas eletrostáticas. Estas e outras propriedades são descritas neste capítulo.

Outro importante conjunto de parâmetros é obtido através da topologia da estrutura molecular. Parâmetros químicos foram desenvolvidos via teoria de graphos.²² Em termos gerais estes parâmetros quantificam a estrutura molecular descrevendo características como disposição dos átomos na molécula, forma e ramificação. Estes parâmetros, conhecidos como índices topológicos, são usados com sucesso na quantificação de moléculas similares, no estudo de toxicidade e propriedades biomédicas.^{23,24}

Após termos obtido todos os dados (parâmetros) acima mencionados, o passo seguinte consistiu no tratamento destes dados, correlacionando-os com a atividade/propriedade de interesse (eq. 1.1). Com esta finalidade utilizamos os métodos de quimiometria PCA (Principal Component Analysis)²⁵ e PLS (Partial Least Squares).^{25,26}

*Os detalhes sobre a análise conformacional para os PCBs e os derivados de 4M e 5M são encontrados nos capítulos 3 e 4 respectivamente.

2.1 Análise Conformacional

As propriedades de uma molécula dependem criticamente de sua estrutura tridimensional, ou conformacional. A análise conformacional consiste no estudo das conformações da molécula e a influência nas propriedades. Portanto, é de suma importância antes de obter os parâmetros de uma determinada molécula, certificar se que a geometria utilizada é a mais adequada. A pesquisa conformacional tem como objetivo identificar a conformação preferida da molécula, pois esta é a conformação que determina seu comportamento. Isto é, localizar os pontos mínimos na superfície de energia potencial. A análise conformacional torna-se necessária pois os pacotes computacionais de otimização de geometria, que utilizam como parâmetro a geometria que oferece uma menor energia potencial, podem durante o processo de otimização encontrar um mínimo local, e não conseguir ultrapassar algumas barreiras de energia, ficando confinados neste ponto.

Entre os métodos de pesquisa conformacional está a pesquisa sistemática. A pesquisa sistemática explora a conformação espacial mudando regularmente a geometria. Em geral a mudança ocorre em algum ângulo diedro. Um problema da pesquisa sistemática é o número de conformações que são geradas $\#C$ (eq. 2.1 - em que θ é o incremento no ângulo i). Este número pode ser explosivo, por exemplo, numa pesquisa em um composto com 4 ângulos diedros, em que é feita uma análise com intervalos de 30° , o número de estruturas geradas é 20736.

$$\#C = \prod_{i=1}^N \frac{360}{\theta_i} \quad (2.1)$$

No nosso estudo, foi utilizada a pesquisa sistemática para obter a conformação mais estável.

2.2 Métodos Semi-Empíricos

A evolução da teoria dos orbitais moleculares resultou em dois métodos principais, *ab initio*[†] e semi-empíricos. Destes, o *ab initio* não necessita nenhum parâmetro determinado experimentalmente, mas estes métodos são muito lentos e sua aplicação rotineira, num nível qualquer de precisão, a sistemas com algumas dezenas de átomos não é praticável.

[†] No apêndice A apresentamos uma breve descrição dos principais conjuntos de base.

Em contraste, os métodos semi-empíricos empregam parâmetros determinados experimentalmente e sua precisão é limitada à dos dados experimentais usados na determinação dos parâmetros. Entretanto os métodos semi-empíricos são rápidos o suficiente para aplicações rotineiras em sistemas grandes. Em consequência, alguns métodos semi-empíricos se tornaram populares, em especial o MNDO (Modified Neglect of Diatomic Overlap)²⁷ do qual derivam o AM1 (Austin Model 1)²¹ e o MNDO-PM3 (Modified Neglect of Diatomic Overlap - Parametric Method 3).²⁸

O bom desempenho dos métodos semi-empíricos não é devido ao rigor do tratamento teórico mas ao fato de que seus parâmetros ajustáveis são otimizados para reproduzir propriedades químicas importantes. Uma comparação entre métodos ab initio e semi-empíricos mostrou que para o cálculo de calores de formação em sistemas para os quais os métodos semi-empíricos foram parametrizados, a precisão destes métodos é comparável à dos métodos ab initio com funções de base extensas.²⁹

No método MNDO e no método AM1 o calor de formação é descrito por

$$\Delta H_f = E_{\text{Elect}} + E_{\text{Nuc}} - \sum_A E_{el}^A + \sum_A \Delta H_f^A \quad (2.2)$$

em que ΔH_f^A é o calor de formação experimental do átomo A . E_{el}^A é a energia eletrônica de cada átomo

$$E_{el}^A = f(P_{ii}, U_{ii}, \langle ii|jj \rangle, \langle ij|ij \rangle) \quad (2.3)$$

em que P_{ii} é a população, no estado fundamental, do orbital atômico i . U_{ii} são as energias de 1 elétron dos orbitais atômicos do íon resultante da remoção de todos os elétrons de valência (representados pelos parâmetros U_{ss} e U_{pp}). $\langle ii|jj \rangle$ são as integrais de 2 elétrons e 1 centro que são representadas pelos parâmetros $\langle ss|ss \rangle = G_{ss}$, $\langle ss|pp \rangle = G_{sp}$, $\langle pp|pp \rangle = G_{pp}$, $\langle pp|p'p' \rangle = G_{p2}$ e $\langle sp|sp \rangle = H_{sp}$. Aqui p e p' são orbitais atômicos p em um único centro mas diferindo na forma de sua função angular.

E_{Nuc} é a energia de repulsão nuclear dada por

$$E_{\text{Nuc}} = \sum_{i < j} E_N(i, j) \quad (2.4)$$

com

$$E_N(i, j) = Z_A Z_B \langle AA|BB \rangle \left(1 + e^{(-\alpha_A R_{ij})} + e^{(-\alpha_B R_{ij})} \right) \quad (2.5)$$

$$+ \frac{Z_i Z_j}{R_{ij}} \left(\sum_k a_{kA}^{-bk_e A (R_{ij} - c_{kA})^2} \right) + \left(\sum_k a_{kB}^{-bk_e B (R_{ij} - c_{kB})^2} \right) \quad (2.6)$$

em que o átomo i é do tipo A , o átomo j é do tipo B , $\langle AA|BB \rangle$ é uma integral de 2 centros do tipo $\langle ss|ss \rangle$, Z_i é o número de elétrons de valência no átomo i e R_{ij} é a distância interatômica.

Somente o primeiro termo desta expressão está presente no método MNDO. O segundo termo foi introduzido no AM1 para corrigir as excessivas repulsões nucleares de longa distância. As quantidades α_A , a_{ik} , b_{ik} e c_{ik} são parâmetros otimizáveis.

A energia eletrônica total é dada por

$$E_{\text{Elect}} = \frac{1}{2} \mathbb{P} (\mathbb{H} + \mathbb{F}) \quad (2.7)$$

com $\mathbb{P}$, $\mathbb{H}$ e $\mathbb{F}$ sendo as matrizes densidade, de 1 elétron e de Fock, respectivamente. A matriz de Fock é composta de termos de 1 e de 2 elétrons. Os elementos da diagonal desta última são definidos por

$$F_{\mu\mu} = U_{\mu\mu} + \sum_B V_{\mu\mu,B} + \sum_\nu P_{\nu\nu} [\langle \mu\mu|\nu\nu \rangle - \langle \mu\nu|\mu\nu \rangle] + \sum_B \sum_\lambda \sum_\sigma P_{\lambda\sigma} \langle \mu\mu|\lambda\sigma \rangle \quad (2.8)$$

em que $V_{\mu\mu,B}$ é o potencial sobre o orbital atômico ϕ_μ devido a todos os núcleos no sistema, exceto aquele do orbital ϕ_μ . Os elementos da matriz de Fock fora da diagonal são dados por

$$F_{\mu\nu} = \frac{S_{\mu\nu}(\beta_\mu + \beta_\nu)}{2} - \sum_{\lambda\sigma} P_{\lambda\sigma} \langle \mu\lambda|\nu\sigma \rangle \quad (2.9)$$

em que β_μ e β_ν são parâmetros atômicos do tipo β_s ou β_p . O recobrimento $S_{\mu\nu}$ ocorre sobre dois orbitais de Slater do tipo

$$\phi_\mu = N r^{n-1} e^{-\xi r} Y_l^m(\theta, \phi) \quad (2.10)$$

em que ξ é o expoente de um orbital atômico do tipo ξ_s ou ξ_p . Integrais do tipo $\langle \mu\lambda|\nu\sigma \rangle$

são funções dos expoentes de Slater e das integrais de 2 elétrons e 1 centro. O mais simples dos termos de 2 centros é a integral $\langle ss|s's' \rangle$ que é a repulsão entre dois elétrons em um orbital s . Esta integral tem a forma

$$\langle ss|s's' \rangle = \frac{27,21}{\sqrt{\frac{1}{R} + \frac{1}{4} \left(\frac{27,21}{G_{ss_A}} + \frac{27,21}{G_{ss_B}} \right)^2}} \quad (2.11)$$

Todas as outras integrais de 2 elétrons e 2 centros envolvem os expoentes de Slater.³⁰

Doze variáveis são identificadas por elemento dentro do método MNDO, as quais podem ser consideradas como parâmetros. Num primeiro esforço somente sete delas foram otimizadas às quais são designados os símbolos U_{ss} , U_{pp} , β_s , β_p , ξ_s , ξ_p e α . Às cinco variáveis restantes (G_{ss} , G_{sp} , G_{pp} , G_{p2} e H_{sp}), foram atribuídos valores determinados de espectros atômicos.³¹ O método AM1 introduz um segundo conjunto de parâmetros na forma de Gaussianas de interação caroço-caroço para corrigir as excessivas repulsões de longa distância na expressão original de repulsão do método MNDO. Assinalando-se mais de quatro destas Gaussianas por átomo, cerca de doze parâmetros são introduzidos. Cada elemento no método AM1 é descrito por 10 a 19 parâmetros.

A base teórica do método AM1 é similar à do método MNDO, a única diferença sendo a presença no AM1 de Gaussianas radiais que modificam a repulsão caroço-caroço. O método PM3 é uma reparametrização do método MNDO e por esta razão é designado como MNDO-PM3 (MNDO - Parametric Method 3). Para cada elemento, neste método, há um conjunto de 18 parâmetros com exceção do hidrogênio que possui 11 parâmetros. Estes parâmetros são os seguintes: U_{ss} , U_{pp} , β_s , β_p , ξ_s , ξ_p , α , G_{ss} , G_{sp} , G_{pp} , G_{p2} , H_{sp} , a_1 , b_1 , c_1 , a_2 , b_2 e c_2 . Para o hidrogênio os sete parâmetros envolvendo os orbitais p não aparecem.

Os métodos MNDO e AM1 são parametrizados para os elementos H, B, C, Si, N, O, S, F, Cl, Br, I, Hg e Zn e o método MNDO-PM3 é parametrizado para os elementos H, C, Si, Ge, Sn, Pb, N, P, As, Bi, O, S, Se, Te, F, Cl, Br, I, Al, Be, Mg, Zn, Cd e Hg.

2.3 Parâmetros Utilizados

A construção de modelos quantitativos que expliquem a atividade biológica e propriedades de uma determinada série de compostos requer a utilização de propriedades moleculares. As propriedades moleculares utilizadas neste trabalho são descritas nas seções seguintes.

2.3.1 Parâmetros Físicos–Químicos

Calor de formação

O calor de formação molecular no estado padrão é obtido considerando-se a energia que é liberada quando os átomos ionizados e seus elétrons de valência combinam-se para formar a molécula (E_{lib}), a energia necessária para ionizar os elétrons de valência dos átomos da molécula (E_{io}), e a entalpia padrão de atomização ($\Delta H_f^0(A)$).

A energia E_{lib} é igual a soma da energia eletrônica e da energia de repulsão nuclear:

$$E_{\text{lib}} = E_{\text{eletr}} + \sum_{A=1}^M \sum_{B=A+1}^M \frac{Z_A Z_B}{R_{AB}} \quad (2.12)$$

No método AM1 a energia eletrônica é:

$$E_{\text{eletr}} = \frac{1}{2} \sum_{\mu} \sum_{\nu} P_{\mu\nu} (H_{\mu\nu} + F_{\mu\nu}) \quad (2.13)$$

Na equação (2.13) $P_{\mu\nu}$ e $F_{\mu\nu}$ são os elementos da matriz densidade ($\mathbb{P}$) e da matriz de Fock ($\mathbb{F}$), e $H_{\mu\nu}$ são os elementos da matriz de integrais de 1 elétron ($\mathbb{H}$).

A equação para o calor de formação molecular, no estado padrão, (ΔH_f^0) pode então ser escrita como:

$$\Delta H_f^0 = \frac{1}{2} \sum_{\mu} \sum_{\nu} P_{\mu\nu} (H_{\mu\nu} + F_{\mu\nu}) + \sum_{A=1}^M \sum_{B=A+1}^M \frac{Z_A Z_B}{R_{AB}} + \sum_A E_{\text{io}} + \sum_A \Delta H_f^0(A) \quad (2.14)$$

Orbitais de Fronteira

As energias de HOMO (E_{homo}) e LUMO (E_{lumo}) são descritores muito utilizados. Em diferentes reações químicas estes orbitais podem ser fundamentais: a formação do estado de transição se deve à interação entre os orbitais de fronteira das espécies reagentes.

A diferença de energia entre os orbitais HOMO e LUMO (conhecido como *gap*), é um importante índice de estabilidade química. Um *gap* de valor alto representa uma molécula com alta estabilidade, ou seja, baixa reatividade em reações químicas.

Potencial de ionização e afinidade eletrônica

O potencial de ionização (PI) caracteriza o comportamento nucleofílico da molécula, e a afinidade eletrônica (AE) caracteriza o comportamento eletrofílico. Estes parâmetros estão relacionados com as energias de HOMO e LUMO, através do teorema de Koopmans,³² que resumidamente diz que:

$$PI = -E_{HOMO} \quad (2.15)$$

$$AE = -E_{LUMO} \quad (2.16)$$

em que E significa a energia do orbital molecular HOMO ou LUMO. Não podemos deixar de mencionar um ponto discutível do teorema de Koopmans, que é o fato do orbital no estado ionizado ser assumido como sendo o mesmo do estado não ionizado, negligenciando o fato que os orbitais no estado ionizado são diferentes dos orbitais no estado não ionizado.

Dureza e moleza molecular

São medidas da intensidade da atração entre os núcleos atômicos e a nuvem eletrônica molecular. Uma molécula é considerada mole, quando a separação dos orbitais de fronteira (HOMO-LUMO) é pequena, e a distribuição de elétrons é facilmente rearranjada por uma perturbação. A molécula dura é aquela que possui grande separação entre os orbitais de fronteira, e portanto tende a ter alta estabilidade, ou baixa reatividade em reações químicas. A dureza molecular é expressa como a metade da separação entre os dois orbitais de fronteira:³³

$$\chi = \frac{(E_{LUMO} - E_{HOMO})}{2} \quad (2.17)$$

A moleza é expressa pelo inverso da dureza.

Eletronegatividade

Foi inicialmente definida por Pauling como "*a capacidade de um átomo em atrair elétrons de outro átomo*". Mulliken³⁴ definiu a eletronegatividade de um átomo A como a média do seu potencial de ionização PI e sua afinidade eletrônica AE:

$$\eta_A = \frac{PI_A + AE_A}{2} \quad (2.18)$$

Momento de dipolo

Representa a distribuição de cargas em uma molécula, sendo a medida da magnitude da carga deslocada quando átomos de eletronegatividades diferentes são interligados. O momento de dipolo de uma molécula, pode oferecer indicação sobre a reatividade da mesma.³⁵ O momento de dipolo é representado por um vetor μ , e tem contribuição tanto eletrônica como nuclear. A contribuição nuclear pode ser calculada usando a fórmula para um sistema de cargas discretas:

$$\mu_{nuclear} = \sum_{A=1}^M Z_A \mathbf{R}_A \quad (2.19)$$

em que M é o número de átomos da molécula, Z e $\mathbf{R}$ são respectivamente a carga e a posição de cada átomo. A origem da contribuição eletrônica de uma distribuição contínua da nuvem eletrônica molecular, pode ser calculada usando-se o operador momento dipolar $\mathbf{r}$, na função de onda molecular, integrando-se sobre todo o espaço ($d\tau$).

$$\mu_{elctr} = \int d\tau \psi_0 \left(\sum_{i=1}^N -\mathbf{r}_i \right) \psi_0 \quad (2.20)$$

substituindo o operador momento dipolar $\mathbf{r}$, pela soma das contribuições de cada elétron, podemos reescrever a equação 2.20 em função de matriz densidade $\mathbb{P}$:

$$\mu_{elctr} = \sum_{\mu=1}^K \sum_{\nu=1}^K \mathbb{P}_{\mu\nu} \int d\tau \phi_{\mu}(-\mathbf{r}) \phi_{\nu} \quad (2.21)$$

Polarizabilidade

Polarizabilidade é a medida da suceptibilidade da molécula ser polarizada. A presença de um campo elétrico externo F , pode induzir o momento de dipolo μ , e distribuir as cargas de uma molécula ou átomo. Este processo é chamado polarização:

$$\mu = \mu_o + \alpha F + \frac{1}{2}\beta F^2 + \frac{1}{6}\gamma F^3 \quad (2.22)$$

em que μ_o é o momento de dipolo permanente, α é a polarizabilidade, β a primeira hiperpolarizabilidade e γ a segunda hiperpolarizabilidade. Alinhando-se o campo elétrico externo a um dos eixos, eixo x por exemplo ($F=(F_x,0,0)$), os valores do momento dipolar perturbado (μ_x) e das polarizabilidades ao longo deste eixo podem ser obtidos,

$$\mu_x = \mu_x^0 + \alpha_{xx}F_x + \frac{1}{2}\beta_{xxx}F_x^2 + \frac{1}{6}\gamma_{xxx}F_x^3 \quad (2.23)$$

estimando-se μ_x para quatro intensidades de campo elétrico diferentes, produz-se quatro equações e quatro incógnitas. Resolvendo-se o sistema para α_{xx} :

$$\alpha_{xx}F_x^2 = \frac{2}{3}[\mu_{x(F_x)} - \mu_{x(-F_x)}] - \frac{1}{12}[\mu_{x(2F_x)} - \mu_{x(-2F_x)}] \quad (2.24)$$

Usando o mesmo procedimento é possível obter os valores de α_{yy} e α_{zz} . A polarizabilidade média é calculada através da média dos seus componentes:

$$\alpha = \frac{1}{3}(\alpha_{xx} + \alpha_{yy} + \alpha_{zz}) \quad (2.25)$$

Cargas

As cargas atômicas são geradas a partir do potencial eletrostático calculado em vários pontos ao redor da molécula. A idéia é ajustar as cargas para se obter o valor do potencial eletrostático que foi calculado a partir da função de onda.³⁶ O potencial eletrostático molecular, ε_p , representa a energia de interação de uma unidade de carga positiva em algum ponto do espaço, $\mathbf{r}_p$, com o núcleo e os elétrons da molécula:

$$\varepsilon_p = \sum_A \frac{Z_A}{R_{AP}} - \sum_\mu \sum_\nu P_{\mu\nu} \int \frac{\phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r})}{r_p} d\mathbf{r} \quad (2.26)$$

em que Z_A são os números atômicos, $P_{\mu\nu}$ são os elementos da matriz de densidade e R_{AP} e $\mathbf{r}_p$ são as distâncias entre as cargas e os núcleos e os elétrons respectivamente. Desta forma a primeira somatória é sobre os centros atômicos e a segunda somatória é sobre as funções de base.

Como primeiro passo para se obter as cargas, define-se os pontos ao redor da molécula. É importante que os pontos escolhidos fiquem fora do raio de van der Waals da molécula, e a distância máxima não seja maior que $2.8\text{\AA}$. Em seguida calcula-se o potencial segundo a equação (2.26) para os pontos escolhidos. Por fim, utilizando-se a equação (2.27), é feito um ajuste das cargas atômicas (Q_A) pelo método dos mínimos quadrados para reproduzir o potencial eletrostático,

$$\varepsilon_P^{approx} = \sum_A \frac{Q_A}{R_{AP}} \quad (2.27)$$

2.3.2 Parâmetros Topológicos

Para a representação de grafos nas moléculas, os átomos são representados como vértices e as ligações como traços³⁷ (fig(2.1)). A seguir são descritos os índices topológicos que foram utilizados neste trabalho.

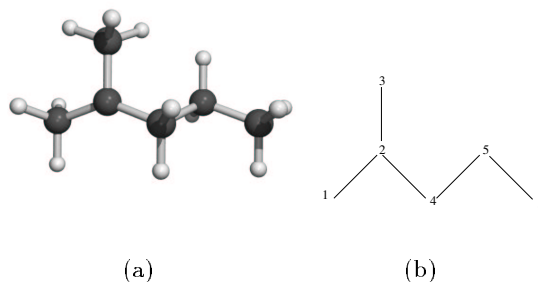


Figura 2.1: Molécula de 2-metilpentano e sua representação de grafo.

Número de Wiener

O número de Wiener foi introduzido em 1947,³⁸ e foi o primeiro índice de grafo introduzido na química. Para definir o número de Wiener, consideremos uma representação de grafo de uma molécula, e $v_1, v_2, \dots, v_n$ os vértices. A distância entre os vértices v_i e v_j é o número de traços (ligações) considerando-se o menor caminho que conecte v_i e v_j :

$$W = \frac{1}{2} \sum_{i,j} (D_{i,j}) \quad (2.28)$$

em que $D_{i,j}$ é um elemento da matriz distância $\mathbf{D}$. Para a molécula da figura (2.1) a matriz distância é:

$$\mathbf{D} = \begin{bmatrix} 0 & 1 & 2 & 2 & 3 & 4 \\ 1 & 0 & 1 & 1 & 2 & 3 \\ 2 & 1 & 0 & 2 & 3 & 4 \\ 2 & 1 & 2 & 0 & 1 & 2 \\ 3 & 2 & 3 & 1 & 0 & 1 \\ 4 & 3 & 4 & 2 & 1 & 0 \end{bmatrix} \quad (2.29)$$

Índices de Conectividade

O conceito de índice de conectividade molecular foi introduzido por Randić.³⁹ A idéia é quantificar a estrutura molecular de forma que fragmentos e subestruturas sejam incorporados como índices numéricos. Primeiro calcula-se, para cada átomo i , o descritor de valência (δ^v) e o descritor de elétrons sigma δ :

$$\delta_i^v = \frac{Z^v - h}{Z - Z^v - 1} \quad (2.30)$$

$$\delta_i = \sigma - h \quad (2.31)$$

em que σ é o número de elétrons sigma, Z^v é o número de elétrons de valência, Z é o número atômico e h o número de átomos de H ligados ao átomo i .

Os índices de conectividade molecular são simbolizados por ${}^m\chi_t$. As subestruturas de uma molécula e a ordem do índice de conectividade são definidas pela fragmentação da representação de grafo:

- a** – Átomos (ordem zero, $m = 0$);
 - b** – Uma ligação (primeira ordem, $m = 1$);
 - c** – Fragmentos de duas ligações consecutivas (segunda ordem, $m = 2$);
 - d** – Três ligações contínuas (terceira ordem de caminho, $m = 3$, $t = P$);
 - e** – Fragmentos incluindo *cluster*, três átomos ligados a um átomo central ($m = 3$, $t = C$);
-

f – Fragmentos incluindo caminho/*cluster*, se refere à quatro ligações adjacentes, com 3 ligações em um átomo central, equivalente à estrutura do isopentano ($m = 4$, $t = \text{PC}$);

g – Fragmento de cadeia, com anéis de 3, 4, 5, ... átomos ($m = 3, 4, 5, \dots$, $t = \text{CH}$);

O índice de conectividade de valência ${}^m\chi_t$ é calculado multiplicando-se os valores de δ de cada átomo da subestrutura, a raiz quadrada deste produto é somada com todos os outros produtos de subestruturas de ordem m e tipo t :

$${}^m\chi_t = \sum (\Pi_{k=1}^{m+1} \delta_i)^{-0.5} \quad (2.32)$$

O índice de conectividade de valência é obtido usando-se o valor de δ^v no lugar de δ .

Número de Balaban

O índice de Balaban^{40,41} é conhecido também como a média da soma das distâncias, e é baseado na matriz **D** (eq.2.29). Cada átomo é descrito por seu valor numérico, que é igual à soma dos elementos da linha correspondente na matriz distância. Para cada par de vértices adjacentes (i,j) é calculado o produto da soma das distâncias ($d_i d_j$), e então é obtido a média da soma das distâncias (J).

$$J = \left(\frac{M}{m+1} \right) \sum_{i,j} (d_i d_j)^{-0.5} \quad (2.33)$$

$$m = M - n + 1 \quad (2.34)$$

$$d_i = \sum_{j=1}^n (D_{ij}) \quad (2.35)$$

em que m é o número *cyclomatic*, para uma representação de grafo com n vértices e M adjacências se for uma representação acíclica (eq. 2.34), caso seja uma representação de grafo cíclica, m é o número de anéis no grafo. d_i é a soma das distâncias.

Índice de Harary

O índice de Harary (H_D) é obtido a partir da matrix de distância $\mathbf{D}$, e recebeu este nome em homenagem a Frank Harary.⁴² H_D é definido como:

$$H_D = \frac{1}{2} \sum_{ij} \frac{1}{D_{ij}} \quad (2.36)$$

Índice Eletrotopológico

Os índices eletrotopológicos são valores numéricos calculados para cada átomo na molécula, que contém informações sobre o ambiente topológico dos átomos e a interação eletrônica devido a outros átomos da molécula.^{43, 44}

Para se obter o valor do estado eletrotopológico de um átomo (S_i), é necessário primeiro se obter o valor do estado intrínseco I_i :

$$I_i = \frac{\left[\left(\frac{2}{N} \right)^2 \delta_i^y + 1 \right]}{\delta_i} \quad (2.37)$$

em que δ_i^y e δ_i foram definidos nas equações 2.30 e 2.31 respectivamente, e N é o número de *grupos* terminais. No exemplo da fig. 2.1 $N=1$ (átomo 1 ou 3), em uma molécula de 4-clorofenol $N=2$, e os grupos terminais são OH e Cl.

A perturbação no átomo i , originária da presença de todos os outros átomos j , é uma função da diferença entre os valores intrínsecos, e dependente da distância:

$$\Delta I_i = \sum \frac{I_i - I_j}{r_{ij}^2} \quad (2.38)$$

em que r_{ij} é o número de traços (ligações) entre os átomos i e j . O estado eletrotopológico é dado pela soma dos estados intrínsecos e das perturbações.

$$S_i = I_i + \Delta I_i \quad (2.39)$$

2.3.3 Parâmetros Estéricos

Os parâmetros estéricos moleculares utilizados foram o volume molecular (V), a área da superfície molecular (S) e a ovalidade molecular (Ova).^{45,46} A área da superfície molecular e o volume molecular foram calculados utilizando-se o raio de van der Waals de cada átomo da molécula para calcular as esferas, e considerando a sobreposição destas esferas. A ovalidade é definida como uma razão entre a área e o volume:

$$Ova = \frac{1}{3 \cdot 3\pi^{1/3}} \frac{S}{V^{2/3}} \quad (2.40)$$

2.4 Métodos Quimiométricos

Atualmente o desenvolvimento computacional proporciona um número muito grande de descritores (parâmetros) moleculares, com economia de tempo, tornando necessário o uso de técnicas estatísticas para se obter a relação da equação 1, fornecendo o máximo de informações químicas, com a análise dos dados obtidos.

Nesta seção descrevemos resumidamente os principais métodos quimiométricos incluindo os utilizados no tratamento dos nossos dados.

Para iniciarmos o tratamento, os dados obtidos para cada amostra, são organizados em uma matriz $\mathbf{X}(n,m)$ (matriz de variáveis independentes) com n linhas e m colunas, em que cada linha representa uma amostra, e cada coluna corresponde a um descritor específico. O outro conjunto de dados é constituído das variáveis dependentes (atividades ou propriedade de interesse) e organizado no vetor $\mathbf{y}(n,1)$ (variável dependente), em que o total de elementos deste vetor é igual ao número de amostras n . O processo geral consiste de duas etapas: Modelagem, que estabelece uma relação matemática entre $\mathbf{X}$ e $\mathbf{y}$, e validação que otimiza a relação no sentido de uma melhor descrição da atividade. Concluída cada uma destas etapas, o modelo está apto para a previsão de outras moléculas.

2.4.1 Pré-Processamentos

O pré-processamento consiste em preparar os dados com a finalidade de remover os efeitos dominantes que não sejam de interesse primário.

Essencialmente há três possibilidades de preparar os dados:

- 1 – Primeiramente quando não é preciso escalar; isto ocorre quando todas as variáveis da matriz $\mathbf{X}$ foram medidas na mesma unidade, como por exemplo dados de espectroscopia.
- 2 – Segundo, escalar pela variância; isto é necessário quando as variáveis da matriz $\mathbf{X}$ são medidas em diferentes unidades (ex: %, km, L). Para escalar pela variância, basta dividir todos os componentes de uma determinada variável pelo seu desvio padrão, desta forma a variância de todas as variáveis passa ser unitária:

$$x_{ij}(sv) = \frac{x_{ij}}{s_j}, \quad \begin{matrix} i = 1, \dots, n \\ j = 1, \dots, m \end{matrix} \quad (2.41)$$

$$s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \quad (2.42)$$

em que $x_{ij}(sv)$ são os elementos de cada variável escalados pela variância, e s_j^2 é a variância desta determinada variável.

- 3 – Terceiro, centrar na média, em que os valores dos dados de cada variável é subtraído pela média desta variável.

$$\bar{x}_{ij}(cm) = x_{ij} - \bar{x}_j \quad (2.43)$$

$$\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij} \quad (2.44)$$

em que $\bar{x}_j$ é a média de uma determinada variável.

Na figura 2.2²⁶ é mostrada uma ilustração de dados escalados, centrados na média e autoescalados. O mais usual é utilizar dados autoescalados, que são dados centrados na média e escalados pela variância. Desta forma todos os dados deste trabalho foram autoescalados.

2.4.2 Análise por Componentes Principais

Um número de variáveis muito grande pode dificultar a análise do problema em questão. Com o objetivo de identificar quais variáveis são relevantes, a análise de componentes

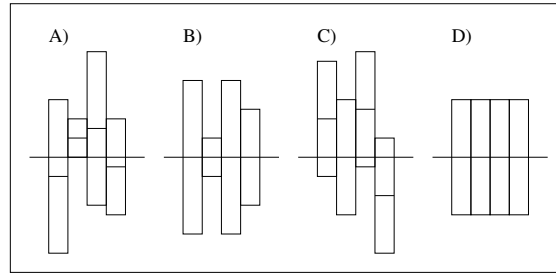


Figura 2.2: Pré-processamento de dados. Os dados de cada variável são representados por uma barra de variância e seu centro. A) Dados sem pré-processamento. B) Resultado após centrar na média. C) Resultado depois de escalar pela variância. D) Dados autoescalados.

principais é utilizada.

Decomposição em componentes principais é uma forma de definir novas variáveis: combinações lineares das variáveis iniciais em um novo sistema de eixos ortogonais. Os novos eixos são definidos de modo a descreverem o espalhamento entre os pontos usando o menor número possível de eixos. Desta forma a natureza multivariada dos dados pode ser visualizada em poucas dimensões. Estes novos eixos são denominados componentes principais (PC), variáveis latentes ou autovetores.

Matematicamente o processo consiste na decomposição da matriz $\mathbf{X}$ em valores singulares (SVD).⁴⁷

$$\mathbf{X} = \mathbf{TP}^T \quad (2.45)$$

em que $\mathbf{P}$ é a matriz dos *loadings*, e $\mathbf{T}$ a matriz dos *scores*. Os *loadings* representam a contribuição de cada variável para a PC, e os *scores* são as novas coordenadas das amostras. Para ilustrar os *loadings* e os *scores*, mostramos a figura 2.3.²⁶

Um algoritmo utilizado para calcular as PCs é o algoritmo de NIPALS²⁶ (Nonlinear iterative partial least squares). Ele calcula $\mathbf{t}_1$ e $\mathbf{p}'_1$ da matriz $\mathbf{X}$. O produto $\mathbf{t}_1\mathbf{p}'_1$ é subtraído de $\mathbf{X}$, resultando no resíduo $\mathbf{E}_1$. Este resíduo pode ser usado para calcular $\mathbf{t}_2$ e $\mathbf{p}'_2$.

Os passos do algoritmo de NIPALS estão a seguir:

1. Especifica-se um vetor $\mathbf{x}_j$ da matriz $\mathbf{X}$ e faz: $\mathbf{t}_h = \mathbf{x}_j$.
2. Calcula-se $\mathbf{p}'_h : \mathbf{t}'_h \mathbf{X} / \mathbf{t}'_h \mathbf{t}_h$.

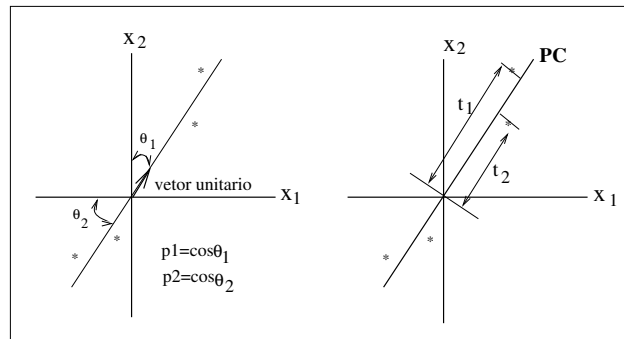


Figura 2.3: Representação de uma componente principal para o caso de duas variáveis x_1 e x_2 . A) Os *loadings* (p_1 e p_2) são os cosenos do vetor direção. B) Os *scores* (t_1 e t_2) são as projeções das amostras na PC.

3. Normaliza $\mathbf{p}'_h$: $\mathbf{p}'_{h_{new}} = \mathbf{p}'_{h_{old}} / \|\mathbf{p}'_{h_{old}}\|$
4. Calcula-se $\mathbf{t}_h$: $\mathbf{t}_h = \mathbf{X}\mathbf{p}_h / \mathbf{p}'_h\mathbf{p}_h$.
5. Compara-se o $\mathbf{t}_h$ usado no passo 2 com o $\mathbf{t}_h$ obtido pelo passo 4. Se forem iguais, significa que o processo iterativo convergiu, então pare. Se forem diferentes, volte ao passo 2.

2.4.3 Regressão Linear Múltipla

Assumindo que exista uma relação linear entre as variáveis da matriz $\mathbf{X}$ com n amostras e m variáveis independentes, e o vetor $\mathbf{y}$, podendo ser representada matematicamente como:

$$y = b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_mx_m + e \quad (2.46)$$

$$y = \sum_{j=1}^m b_jx_j + e \quad (2.47)$$

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e} \quad (2.48)$$

onde, b_j são os coeficientes de regressão e e os erros ou resíduos. As equações 2.46 e 2.47 descrevem uma dependência multilinear para uma amostra, enquanto para n amostras a notação fica como na equação 2.48, tendo $\mathbf{y}$ e $\mathbf{b}$ como vetores colunas ($n \times 1$) e ($m \times 1$) respectivamente, e $\mathbf{X}$ a matriz de variáveis independentes ($n \times m$). Fazer regressão linear

múltipla (MLR), significa determinar os vetores $\mathbf{b}$ e $\mathbf{e}$. Considerando-se a dimensão na matriz $\mathbf{X}$, três casos podem ocorrer:

- 1- $m > n$. Há mais variáveis do que amostras, desta forma há um número infinito de soluções para $\mathbf{b}$, o que não é o desejado.
- 2- $m = n$. O número de amostras e variáveis são iguais, neste caso há uma única solução para $\mathbf{b}$, e conseqüentemente $\mathbf{e}$ é um vetor de zeros:

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b} = 0 \quad (2.49)$$

Entretando, na prática nem sempre se encontra esta situação.

- 3- $m < n$. Neste caso o número de amostras é maior que o número de variáveis. Não há uma solução exata para $\mathbf{b}$, mas pode-se obter uma solução minimizando-se o vetor resíduo $\mathbf{e}$:

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b} \quad (2.50)$$

O mais popular método quando se tem a situação do ítem **3**, é chamado mínimos quadrados, cuja solução é:

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (2.51)$$

Observando a equação 2.51 nos deparamos com o mais frequente problema em MLR, a inversa de $\mathbf{X}'\mathbf{X}$ pode não existir. Este problema também é conhecido como colinearidade, determinante zero ou singularidade.

2.4.4 Regressão por Componentes Principais

Em regressão por componentes principais (PCR), a equação 2.48, passa a ser escrita em função da matriz *scores* $\mathbf{T}$ de $\mathbf{X}$:

$$\mathbf{y} = \mathbf{T}\mathbf{b} + \mathbf{e} \quad (2.52)$$

com solução:

$$\mathbf{b} = (\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'\mathbf{b} \quad (2.53)$$

Com o uso dos *scores* o problema de colinearidade, que existia em MLR, fica resolvido, por garantir uma inversibilidade de matriz na solução de $\mathbf{b}$. Outra vantagem em PCR está na possibilidade de se eliminar as componentes principais menos importantes. Entretanto PCR oferece alguns riscos quando se descarta alguma PC que contém informação relevante, ou quando a PC usada na regressão representa ruído.

2.4.5 Regressão por Mínimos Quadrados Parciais

O método de regressão por mínimos quadrados parciais (PLS) assim como PCR utiliza a análise por componentes principais. Entretanto, o método PLS leva em consideração além das informações contidas na matriz $\mathbf{X}$, as informações provenientes do vetor $\mathbf{y}$. A relação entre $\mathbf{X}$ e $\mathbf{y}$ é obtida utilizando-se o *score* $\mathbf{t}$, da matriz $\mathbf{X}$, e o *score* $\mathbf{u}$, de $\mathbf{y}$. Os *scores* de cada matriz calculados separadamente e aplicados na equação $b_h = \mathbf{u}'_h \mathbf{t}_h / \mathbf{t}'_h \mathbf{t}_h$ não resulta no melhor modelo possível, pois as PCs calculadas para cada bloco de dados apresentam uma fraca relação entre si. Desta forma, um algoritmo iterativo pode oferecer uma melhor relação entre os dados, e conseqüentemente um melhor modelo. A seguir é apresentado um algoritmo para PLS.²⁶

1. Faça $\mathbf{u}_{start} = \mathbf{y}$.

No bloco $\mathbf{X}$:

2. $\mathbf{w}' = \mathbf{u}'\mathbf{X} / \mathbf{u}'\mathbf{u}$
3. $\mathbf{w}'_{new} = \mathbf{w}'_{old} / \|\mathbf{w}'_{old}\|$ (normalização).
4. $\mathbf{t} = \mathbf{X}\mathbf{w} / \mathbf{w}'\mathbf{w}$.

Calcule os *loadings* de $\mathbf{X}$ e reescale os *scores*.

5. $\mathbf{p}' = \mathbf{t}'\mathbf{X} / \mathbf{t}'\mathbf{t}$
6. $\mathbf{p}'_{new} = \mathbf{p}'_{old} / \|\mathbf{p}'_{old}\|$ (normalização)
7. $\mathbf{t}_{new} = \mathbf{t}_{old} / \|\mathbf{p}'_{old}\|$

$$8. \mathbf{w}'_{new} = \mathbf{w}'_{old} / \|\mathbf{p}'_{old}\|$$

Encontre o coeficiente de regressão:

$$9. b = \mathbf{u}'\mathbf{t}/\mathbf{t}'\mathbf{t}$$

Calcule os resíduos:

$$10. \mathbf{E}_h = \mathbf{E}_{h-1} - \mathbf{t}_h \mathbf{p}'_h; \mathbf{X} = \mathbf{E}_0 \text{ (para uma componente } h).$$

$$11. \mathbf{F}_h = \mathbf{F}_{h-1} - b_h \mathbf{t}_h \mathbf{q}'_h; \mathbf{y} = \mathbf{F}_0 \text{ (para uma componente } h).$$

Volte ao passo 1 para encontrar a próxima componente.

O número de PCs que será utilizado é um ponto importante na validação do modelo (obtido pelo método PLS ou PCR).⁴⁸ Apesar de se poder calcular um número de componentes principais iguais ao posto da matriz $\mathbf{X}$, nem sempre todas são utilizadas. A principal razão está no fato de algumas das PCs com menor contribuição não descreverem informações sistemáticas mas sim informações aleatórias. Neste trabalho o método utilizado para se determinar o número de PCs que terá o modelo foi a validação cruzada. Neste método remove-se uma amostra i do conjunto e constrói-se o modelo, o qual é usado para prever a atividade ou propriedade da amostra removida ($\hat{y}_i$). Isso deve ser feito para todas as amostras. Após calcular os erros de previsão ($\hat{y}_i - y_i$), calcula-se a soma do quadrado do erro de previsão ($PRESS$):

$$PRESS = \sum (\hat{y}_i - y_i)^2 \quad (2.54)$$

Este processo é repetido para modelos com 1, 2, 3 e quantas PC forem necessárias. O número de PC mais adequado é o que corresponde ao menor valor de $PRESS$.

2.4.6 Validação

A precisão do modelo pode ser verificada a partir de algumas propriedades capazes de medir o grau de ajuste do modelo com o conjunto de dados em questão. Entre estas propriedades está o $PRESS$ que já mencionamos e algumas outras que descrevemos a seguir.

Coefficiente de correlação definido como a raiz quadrada da razão entre a soma dos quadrados dos valores previstos em relação à média pela soma dos quadrados dos valores observados em relação à média:

$$r^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} \quad (2.55)$$

em que $\bar{y}$ é a média de $\mathbf{y}$. Quanto mais próximo a soma dos quadrados dos valores previstos em relação à média for da soma dos quadrados dos valores observados em relação à média, melhor o ajuste do modelo linear ao conjunto de dados. Ou seja, modelos bem ajustados possuem r^2 próximo à unidade.

Coefficiente de correlação das previsões obtido da razão de *PRESS* pela soma dos quadrados dos valores previstos em relação à média:

$$Q^2 = 1 - \frac{PRESS}{\sum(\hat{y}_i - \bar{y})^2} \quad (2.56)$$

Como o segundo termo da equação (2.56) deve ter o menor valor possível, Q^2 deverá ser próximo da unidade para que a regressão tenha boa capacidade preditiva.

Desvio padrão das previsões é calculado de acordo com a seguinte equação:

$$S_{PRESS} = \sqrt{\frac{PRESS}{n - k - 1}} \quad (2.57)$$

em que n é o número de amostras, e k é o número de PCs. Assim como *PRESS*, S_{PRESS} deverá apresentar o menor valor possível.

Quadrado do desvio padrão da regressão definido como:

$$s^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{n - k - 1} \quad (2.58)$$

Ou seja, quanto menor o valor de s melhor a regressão.

Teste de Fischer definido como:

$$F = \frac{\sum(\hat{y}_i - \bar{y})^2}{(k - 1)s^2} \quad (2.59)$$

Como $\sum(\hat{y}_i - \bar{y})^2$ deve ser o maior possível, e s^2 o menor possível, uma regressão com boa qualidade de ajuste equivale a um valor de F na ordem de dezenas ou cen-

tenas. Os valores de F devem ser comparados com $F_{(k,n-k-1)}$, disponível em tabelas encontradas em livros de estatística.⁴⁹ Em geral o valor de F deve ser cinco vezes maior que o valor mínimo aceitável (encontrado nas tabelas). Este valor encontrado nas tabelas não costuma ser apresentado, mas sim o nível de confiabilidade tomado como referência.

“**Leverage**” que pode ser interpretada geometricamente como a distância da amostra ao centróide do conjunto de dados. É utilizada na detecção de amostras anômalas, conhecidas como *outliers*. Usualmente o limite máximo sugerido é:

$$h = \frac{3k}{n} \quad (2.60)$$

Resíduo de Student que verifica a influência de cada amostra no modelo. Amostras mal modeladas têm resíduos altos. A análise do gráfico dos resíduos de *Student* versus a *leverage* para cada amostra, permite determinar os *outliers*. Amostra com resíduo e *leverage* altos devem sempre ser excluídas.

2.5 Principais Programas Utilizados

1. DVIPS 5.76⁵⁰ - Converte arquivo DVI para PostScript.
 2. GAUSSIAN 94⁵¹ - Calcula orbitais moleculares.
 3. GEPOL93 8.0⁵² - Calcula a área e o volume molecular.
 4. GIMP 1.0⁵³ - Faz tratamento de imagem.
 5. GNUPLOT 3.5⁵⁴ - Faz gráficos.
 6. GV 3.5.8⁵⁵ - Visualiza arquivos PostScript.
 7. L^AT_EX⁵⁶ - Editor de texto.
 8. MOLDEN 3.2⁵⁷ - Visualiza moléculas.
 9. MATLAB 5.0⁵⁸ - Resolve numericamente problemas lineares e não lineares.
 10. MOPAC 6.0⁵⁹ - Calcula orbitais moleculares.
-

11. POLLY⁶² - Calcula índices topológicos.
 12. PLS_TOOLBOX 1.5⁶³ - Rotinas do Matlab utilizadas para aplicações quimiométricas.
 13. SPARTAN 5.0⁶⁵ - Calcula orbitais moleculares.
 14. WHIM-3D⁶⁶ - Calcula índices topológicos e eletrotopológicos.
 15. XDVI⁶⁷ - Visualiza arquivos DVI.
 16. XFIG 3.1⁶⁸ - Desenha figuras.
 17. XV 3.10a⁶⁹ - Visualiza imagens.
-

Capítulo 3

Bifenis Policlorados

Inicialmente foi feito um estudo para se obter a geometria da molécula de bifenil, a qual foi utilizada como esqueleto na montagem das moléculas de PCBs.

Existem muitos estudos experimentais^{70,71} e teóricos^{72,73} sobre a geometria e a barreira de rotação da molécula de bifenil. O ângulo entre os planos dos anéis depende fortemente do estado em que o bifenil se encontra, e é sensível ao meio ambiente (gasoso, líquido ou sólido). Acima de 40K, os dados experimentais apontam para uma geometria planar no estado sólido. Por outro lado em solução, em cristal abaixo de 40K, e em fase gasosa, são encontrados diferentes ângulos entre os planos dos anéis,⁷⁴ mas sempre apontando para uma estrutura não planar. São apontados dois fatores responsáveis por esta duplicidade de estruturas: o efeito da delocalização eletrônica, que favorece uma conformação coplanar, e uma repulsão entre os átomos de hidrogênio nas posições orto dos anéis, que é mínima na conformação perpendicular.⁷⁵

A determinação experimental em fase gasosa estima a altura da barreira rotacional para a molécula na conformação coplanar igual a 6.0 ± 2.1 KJ/mol, e para a molécula na conformação perpendicular igual a 6.5 ± 2.0 KJ/mol. O ângulo obtido utilizando-se difração de elétrons é $44.4 \pm 1.2^\circ$.⁷⁶ Estudos anteriores usando mecânica molecular e métodos semiempíricos não obtiveram uma conclusão definitiva.⁷⁴ O primeiro estudo utilizando métodos ab initio foi feito no nível Hartree-Fock,⁷⁷ depois muitos outros foram realizados para se determinar a conformação mais estável, empregando-se bases STO-3G. Os valores obtidos variam entre $38.6^\circ - 43.8^\circ$.^{74,78}

Para obtermos a conformação de menor energia, construímos uma molécula de bifenil e variamos o ângulo entre os planos dos anéis em intervalos de 9° , de 0° até 90° e calculamos

a energia para cada conformação utilizando o método *ab initio* empregando-se a base HF/6-31G*. O resultado é apresentado na figura 3.1. A conformação com menor energia obtida apresenta um ângulo de 45° entre os anéis, esta conformação foi então otimizada. Este resultado está de acordo com resultados encontrados na literatura,^{79,80} em que o valor do ângulo encontrado varia de 44.74° a 46.26° para cálculos realizados com a mesma base (HF/6-31G*) utilizada por nós.

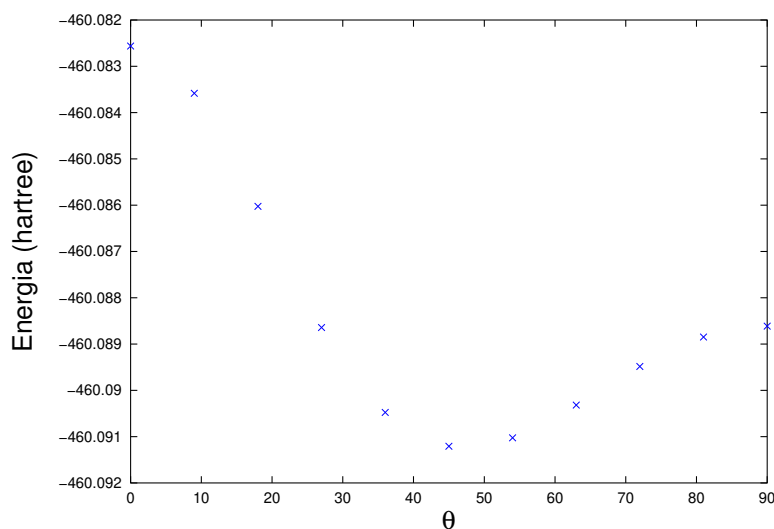


Figura 3.1: Barreira de rotação para molécula de bifenil.

A partir da molécula de bifenil construímos todas as 209 moléculas de PCBs utilizando o programa MOLDEN.⁵⁷ Após termos todas as moléculas, fizemos uma análise conformacional de cada molécula utilizando o método semi-empírico AM1. Esta análise foi feita variando-se o ângulo entre os planos dos anéis, em intervalos de 30° , entre 0° e 180° . Os resultados foram comparados com resultados encontrados na literatura,⁸¹ onde os cálculos foram realizados utilizando-se método *ab initio* com a base STO-3G (tab. 3.1).

Em seguida as propriedades moleculares foram calculadas utilizando-se o programa Spartan 5.⁶⁵ O método semi-empírico escolhido foi o AM1 (Austin Model 1).²¹ Apesar de se encontrar na literatura que o melhor método de parametrização para átomos de cloro

Tabela 3.1: Comparação dos valores obtidos por cálculos AM1 para os ângulo entre os planos dos anéis de alguns PCBs, com resultados de cálculos ab initio (STO-3G) da literatura.

PCB	θ_{AM1}	$\theta_{lit.}$
3-	40.44	41.6
3,4-	40.53	41.4
2,6-	91.90	90.0
3,3',4-	40.90	41.7
2,3,6-	89.52	90.0
2,2',3,3'-	107.9	104.9
3,3',4,5,5'-	40.0	41.7
2,2',4,4',6,6'-	90.0	90.0

seja o semi-empírico MNDO-PM3 (Modified Neglect of Diatomic Overlap - Parametric Method Number 3),²⁸ como tínhamos algumas moléculas onde o número de ligações C – H era maior que o número de ligações C – Cl, resolvemos comparar as energias de formação obtidas pelos dois métodos, para todas as moléculas. Observamos que para as moléculas cujo o número de ligações C – H predominava, as energias obtidas pelo método AM1 eram menores, e para as demais moléculas as energias de formação obtidas pelos dois métodos eram iguais.

Os índices topológicos foram calculados pelo programa Polly, o qual gerou um total de 98 índices para as 209 moléculas. Os métodos quimiométricos utilizados foram: PCA e PLS. Foram tratados por esses métodos as propriedades experimentais extraídas da literatura. Estes dados experimentais iniciais possibilitaram a elaboração de modelos de previsão para outras moléculas (QSPR). Isto ocorre porque, apesar de ser possível a síntese de todas as moléculas, poucos são os PCBs sintetizados em quantidade suficiente para análise experimental. Desta forma, há grandes dificuldades em se obter muitas das propriedades de vários PCBs, e a literatura é escassa em relação a estes dados, daí a importância da construção de modelos e previsão das mesmas.

As propriedades experimentais estudadas foram: coeficiente de partição octanol-água, coeficiente de partição carbono-orgânico, constante da lei de Henry, solubilidade em água, coeficiente de atividade aquosa (todos os dados utilizados foram obtidos à 25 °C) e ponto de fusão. Com o valor das propriedades para todos os PCBs, iniciou-se o estudo de QSAR. As atividades biológicas estudadas foram as toxicidades IC₅₀ e EC₅₀.

3.1 Coeficiente de Partição octanol-água

O coeficiente de partição octanol-água (K_{OW}) mede a lipofilicidade. K_{OW} é definido como a razão da concentração de equilíbrio de uma substância dissolvida em um sistema de duas fases, formado por dois solventes imiscíveis, água e octanol. K_{OW} é considerado um parâmetro importante em estudos de QSAR, pois dá uma descrição da interação do composto com o meio, o que é muito importante, principalmente quando se usa variáveis que descrevem a molécula isolada. K_{OW} é considerado um parâmetro chave no estudo do comportamento ambiental de PCBs. A determinação de K_{OW} tem sido recomendada pelo programa de evolução de compostos perigosos (OECD).⁸² Nos últimos anos K_{OW} vem sendo muito utilizado em várias áreas, com muitas publicações divulgando a correlação de K_{OW} com outras propriedades físicas, químicas e biológicas dos compostos e o sucesso de seu uso na estimativa de fatores de bioconcentração, coeficiente de partição carbono-orgânico, toxicidade e solubilidade.^{83,84}

Para PCBs o estudo de K_{OW} é limitado porque existem valores inconsistentes para alguns congêneros, enquanto para muitos os congêneros o valor de K_{OW} não foi determinado. Entre as técnicas experimentais utilizadas para medir K_{OW} se destacam: “*shake-flask*”, que é o método clássico de determinação de K_{OW} , técnica geradora de coluna (GC) e medidas de cromatografia líquida de fase reversa de alta eficiência e de camada fina (RP-HPLC e RP-TLC). Todas estas técnicas apresentam suas limitações, como dificuldade analítica na determinação de K_{OW} em situações de equilíbrio extremamente pouco aquosa⁸⁵ ou a absorção e volatilização do soluto (GC).⁸⁶

Para se obter uma análise com dados mais consistente, os valores de K_{OW} escolhidos foram os obtidos pela mesma técnica experimental. Nós utilizamos medidas de K_{OW} obtidas por “*thin-layer chromatography - retention time*” (TLC-RT).⁸⁷

Inicialmente selecionamos os descritores que apresentavam melhor correlação linear com $\log K_{OW}$: volume, calor de formação (ΔH), eletronegatividade (η), energia de LUMO (E_{lumo}), ovalidade, índice de conectividade de cadeia de ordem 6 (${}^6\chi_{ch}$) e índice de Winer (W) (fig.3.2). Analisando os valores de *loadings* destas variáveis, foi feita uma nova seleção dos descritores que tinham maior contribuição nas primeiras PCs. A regressão foi feita para 4 modelos, cada um com uma combinação de descritores. Todos os modelos apresentaram bons coeficientes de regressão. O modelo escolhido foi o que conseguiu descrever todas as moléculas com o menor resíduo.

A modelagem foi feita utilizando-se 30 amostras, pois procuramos medidas feitas

com a mesma técnica experimental. O modelo com 2 componentes principais, teve o coeficiente de regressão igual a 0.997. Na tabela 3.2 estão as variâncias de cada componente principal (PC), e na tabela 3.3 estão as contribuições de cada variável para as componentes principais (*loadings*). A primeira PC acumula 92.50% da variância total, e é o resultado da contribuição de todas as variáveis. A segunda PC acumula uma variância de 6.96%, e apresenta contribuição significativa da variável E_{lumo} , seguida pelo volume. A terceira PC contribui com apenas 0.42%. O valor de PRESS é menor quando utiliza-se duas PCs, o que indica que a terceira PC pode estar representando os ruídos das variáveis, principalmente o volume e o ΔH . Na tabela 3.4 estão os descritores utilizados na modelagem com coeficientes do vetor de regressão e os dados estatísticos obtidos. Os dados estatísticos mostrados são: coeficiente de correlação (r), quadrado do desvio padrão da regressão (s^2), coeficiente de correlação das previsões (Q^2), desvio padrão das previsões (S_{PRESS}), o teste de Fischer (F), e o número de amostras utilizadas na modelagem (n). Estes dados estatísticos foram utilizados em todos os modelos apresentados neste trabalho. Na figura 3.3 e na tabela 3.6 são mostrados os valores previstos por validação cruzada para todas as moléculas que participaram do modelo versus os valores experimentais. Observa-se que os erros estão em torno de 2%.

Durante a modelagem, foi verificado o resíduo de *Student* e a *leverage* para todas as amostras (fig.3.4). Observamos que não há *outliers*, pois todas as amostras apresentaram o valores de *leverage* dentro da faixa considerada normal, neste caso menor que 0.2. Os valores de resíduo de *Student* para as amostras, estão dentro da faixa prevista (-2.5 a 2.5). Este procedimento foi seguido para todas as análises apresentadas nesta tese. Amostras com alto valor de *leverage* ou resíduo de *Student*, receberam um atenção especial, pois poderiam ser *outliers*. As consideradas *outliers* ficaram de fora do processo de modelagem. Para validar o modelo, foi previsto o valor de $\log K_{OW}$ para três amostras que não fizeram parte da modelagem (validação externa). O maior erro obtido foi de 0.2%, como pode ser verificado na tabela 3.5. Uma vez obtida a função matemática (eq. 1.1) que relaciona os descritores com a propriedade em questão, esta relação (modelo) é utilizada para previsão. Os resultados previstos para as 209 moléculas de PCBs estão na tabela C.3, no apêndice C.

Analisando a tabela 3.4 e a figura 3.2 observamos que $\log K_{OW}$ aumenta com o volume, que nos PCBs está correlacionado com o aumento do número de átomos de cloro, e com a posição destes átomos, visto que átomos de cloro na posição orto, tendem aumentar

o ângulo entre os anéis de benzeno. Este resultado está de acordo com dados da literatura, que afirmam que a contribuição para a lipofilicidade com o aumento do número de átomos de cloro é maior quando estes estão na posição orto.⁸⁸ O sinal negativo em ΔH indica que quanto menor seu valor, maior o valor de K_{OW} . A E_{lumo} contribui negativamente, e diminui com o aumento do número de átomos de cloro. A E_{lumo} está relacionada à afinidade eletrônica (eq. 2.16), então, moléculas com maior afinidade eletrônica contribuem para um aumento em K_{OW} .

Como K_{OW} depende da concentração de PCBs em octanol e em água, um aumento em K_{OW} significa um aumento da concentração de PCBs em octanol e/ou uma diminuição da concentração em água. Alguns autores sugerem que podem surgir pontes de hidrogênio, sendo que o PCB se comportaria como eletroceptor (devido sua relação com E_{lumo}).⁸⁹

Na literatura existem estudos para prever $\log K_{OW}$ com bons resultados. Podemos citar o trabalho de Hansen⁹⁰ que correlaciona $\log K_{OW}$ com a área, e obtém um coeficiente de correlação $r = 0.997$. Podemos citar também o trabalho de Gramatica,⁹¹ em que dois índices topológicos foram utilizados. Entretanto, nenhum destes trabalhos possibilita uma interpretação da influência de outras características moleculares no comportamento lipofílico da molécula.

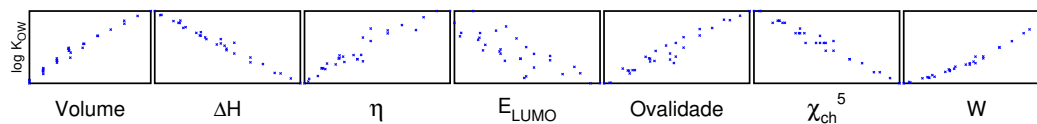


Figura 3.2: Gráfico de $\log K_{OW}$ versus descritores, mostrando a correlação.

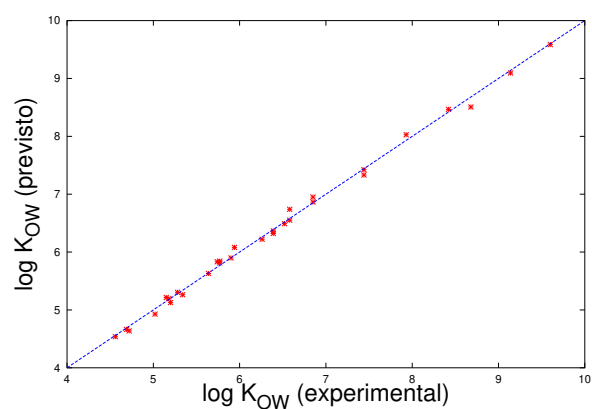


Figura 3.3: Gráfico dos valores previstos e experimentais de $\log K_{OW}$.

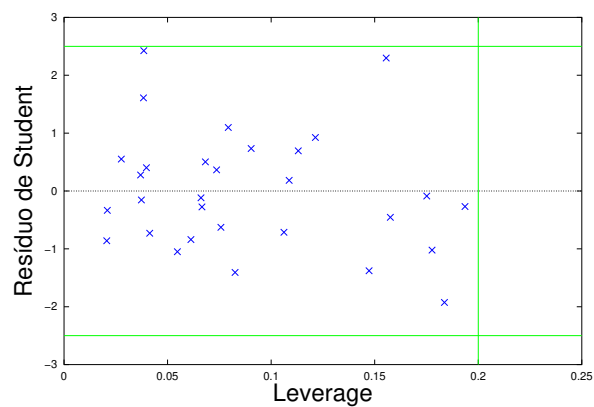


Figura 3.4: Gráfico do resíduo de *Student* versus a *leverage* obtido pelo modelo escolhido.

Tabela 3.2: Porcentagem da variância acumulada pelo modelo para $\log K_{OW}$.

# PC	Esta PC	Total
1	92.4954	92.4954
2	6.9621	99.4575
3	0.4172	99.8747

Tabela 3.3: *Loadings* das variáveis no modelo para $\log K_{OW}$

Descritor	PC1	PC2	PC3
Volume	0.4635	0.4946	0.7226
ΔH	-0.4649	0.0576	0.5087
E_{lumo}	-0.3725	0.8986	-0.3480
${}^2\chi$	0.4658	0.1127	-0.1829
${}^6\chi_{ch}$	-0.4619	-0.1727	0.3094

Tabela 3.4: Vetor de regressão para $\log K_{OW}$. ${}^2\chi$ é o índice de conectividade de caminho de ordem 2, e ${}^6\chi_{ch}$ o índice de conectividade de cadeia de ordem 6.

Descritor	Vetor de regressão
Volume	0.3661
ΔH	-0.1266
E_{lumo}	-0.1447
${}^2\chi$	0.1776
${}^6\chi_{ch}$	-0.2149

$$r = 0.997 \quad s^2 = 0.006 \quad Q^2 = 0.997$$

$$S_{PRESS} = 0.08 \quad F = 8.3 \times 10^3 \quad n = 30$$

Tabela 3.5: Valor de $\log K_{OW}$ previsto e experimental para algumas moléculas.

Congênero	Valor experimental	Valor previsto
2,4'-	5.09	5.17
2,3,4,4'-	6.24	6.34
3,3',4,4',5,5'-	7.55	7.66

Tabela 3.6: Valor Previsto e Experimental de $\log K_{OW}$.

PCB	$\log K_{OW}$	
	Previsto	Experimental
1	4.54	4.56
2	4.64	4.72
3	4.66	4.69
4	4.93	5.02
5	5.13	5.20
7	5.22	5.15
9	5.19	5.18
11	5.26	5.34
15	5.31	5.28
18	5.63	5.64
28	5.83	5.74
29	5.84	5.77
31	5.81	5.77
37	5.90	5.90
52	6.22	6.26
54	6.08	5.94
61	6.32	6.39
70	6.37	6.39
77	6.49	6.52
80	6.55	6.58
87	6.74	6.58
101	6.86	6.85
116	6.95	6.85
128	7.33	7.44
153	7.42	7.44
185	8.03	7.93
194	8.51	8.68
202	8.47	8.42
206	9.09	9.14
209	9.59	9.60

3.2 Solubilidade

Solubilidade (S) em água, assim como pressão de vapor, é uma propriedade de saturação, isto é, é a medida da capacidade máxima que uma fase (água) tem para dissolver um composto (PCB). A taxa de solubilidade é um fator físico-químico que influencia na taxa de biodegradação. O método convencional para preparar soluções saturadas consiste em colocar um excesso de soluto na água e obter o equilíbrio misturando lentamente, prevenindo a formação de emulsão e suspensão. O método de medir a concentração na solução saturada depende da solubilidade do soluto e suas propriedades químicas.

Como já foi mencionado, os PCBs são lipofílicos, ou seja, apresentam baixa solubilidade em água, conseqüentemente apresentam concentração muito pequena em água, originando dificuldades na medida. A solubilidade dos PCBs diminui com o aumento do número de átomos de cloro, sabe-se que em geral a substituição de um átomo de hidrogênio por um átomo de cloro, causa uma queda na solubilidade.

Assim como K_{OW} , as dificuldades experimentais resultam em uma discrepância entre os valores divulgados na literatura. Os dados de solubilidade utilizados neste trabalho, foram medidos por “*generator column-GC/ECD*”,⁸⁷ considerado um método preciso.

Os descritores com maior correlação com $\log S$ são mostrados na figura 3.5. São eles a área superficial molecular, o calor de formação (ΔH), a eletronegatividade (η), a energia de LUMO (E_{lumo}), o índice de complexidade de vértice (H^v), o número de caminho de comprimento de ordem 9 (P_9) e o índice de Winer (W). Foram construídos 3 modelos com estes descritores, o modelo escolhido foi aquele que apresentou o maior índice de correlação e que descreveu um maior número de amostras com o menor valor de *PRESS*.

A primeira PC acumula 77.02% da variância total dos dados (tab. 3.7), e é o resultado da composição de todas variáveis com coeficientes de módulos próximos (tab. 3.8). A segunda PC acumula uma variância total de 19.63% (tab. 3.7) e tem uma contribuição significativa de H^v (tab.3.8). A terceira PC representa apenas 2.55% da variância total, e o valor de *PRESS* aumenta, o que indica a representação de ruído.

O modelo PLS com 2 componentes principais foi feito utilizando-se 31 moléculas e 5 descritores com um coeficiente de correlação $r = 0.965$. Os descritores utilizados na modelagem foram: área, ΔH , E_{lumo} , H^v e P_9 . Os valores previstos e experimentais de $\log S$ são mostrados na figura 3.6 e na tabela 3.11. Nota-se que os erros estão abaixo de 20%, exceto para 8 compostos, os quais estão em negrito na tabela. Na tabela 3.9 estão os descritores utilizados na modelagem com os coeficientes do vetor de regressão, e os dados

estatísticos do modelo. O valor previsto para duas amostras que não participaram do modelo estão na tabela 3.10.

Os resultados (tab. 3.9) mostram que $\log S$ recebe uma contriuição negativa da área e dos índices topológicos, os quais podem estar completando informações que não foram fornecidas pela área. O sinal negativo nos índices topológicos indica que um aumento na ramificação da molécula influencia negativamente a solubilidade. Nas moléculas de PCBs este aumento na ramificação é uma consequência do aumento do número de átomos de cloro. A E_{lumo} contribui positivamente para $\log S$. O aumento na E_{lumo} significa uma diminuição na afinidade eletrônica, ou seja no caráter eletroceptor da molécula. O sinal positivo em ΔH significa que quanto maior ΔH maior a solubilidade. De alguma forma moléculas com menor caráter eletroceptor interagem mais com a água.

Alguns trabalhos utilizam o volume molar,⁹² a área superficial,⁹³ e o índice de conectividade (χ)⁹⁴ para prever $\log S$. Mas observando o valor de $\log S$ para as três amostras que foram deixadas de fora para serem utilizadas na validação do modelo (tab. 3.10), podemos observar que $\log S$ é dez vezes menor para a terceira amostra em relação à primeira amostra. Entretanto, observando as tabelas C.1 e C.2 (apêndice C) verificamos que nenhum dos descritores que participaram da modelagem tem o mesmo comportamento. Este fato contribui para a idéia que a solubilidade não é resultado da variação em um único tipo de descritor, mas da combinação de um grupo de descritores.

A partir dos modelos obtidos para $\log K_{OW}$ e $\log S$, fizemos a previsão destas propriedades para todas as moléculas (tabela C.3). Estes valores obtidos foram incluídos na matriz **X**, que foi utilizada para a modelagem das outras propriedades.

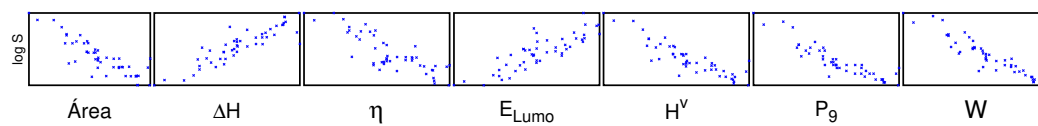


Figura 3.5: Gráfico de $\log S$ versus descritores, mostrando a correlação.

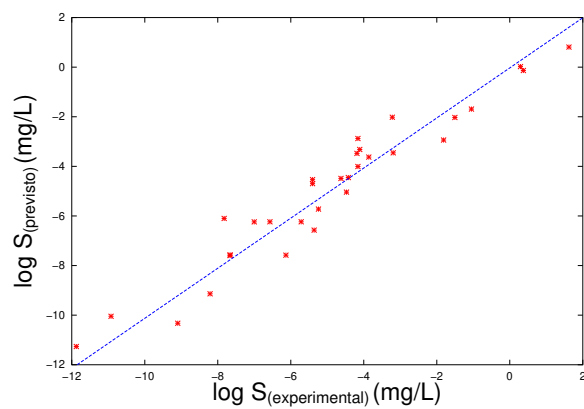


Figura 3.6: Gráfico do valor previsto e o experimental de $\log S$.

Tabela 3.7: Porcentagem da variância acumulada pelo modelo para $\log S$.

# PC	Esta PC	Total
1	77.02	77.02
2	19.63	96.65
3	2.55	99.21

Tabela 3.8: *Loadings* das variáveis acumulada pelo modelo para $\log S$.

Descritor	PC1	PC2
Área	-0.4962	-0.1955
ΔH	0.4998	0.1360
E_{lumo}	0.4746	-0.2337
H^v	-0.1716	0.9339
P_9	-0.4993	-0.1285

Tabela 3.9: Vetor de regressão para $\log S$.⁶ χ^b é o índice de complexidade de vértice, e P_9 o número de caminho de comprimento 9.

Descritor	Vetor de regressão
Área	-0.2682
ΔH	0.2687
E_{lumo}	0.1939
H^v	-0.0063
P_9	-0.2633

$$r = 0.965 \quad s^2 = 0.719 \quad Q^2 = 0.931$$

$$S_{PRESS} = 0.848 \quad F = 375.24 \quad n = 31$$

Tabela 3.10: Comparação do valor de $\log S$ previsto e experimental para validação.

Congênero	Valor experimental	Valor previsto
2,4',5'-	-2.408	-2.604
2,2',4,6,6'-	-4.131	-4.571

Tabela 3.11: Valor Previsto e Experimental de $\log S$.

PCB	$\log S$ (mg/L)	
	Previsto	Experimental
1	0.811	1.621
3	0.021	0.293
10	-0.137	0.372
11	-1.689	-1.050
15	-2.020	-3.219
28	-2.94	-1.814
30	-2.029	-1.505
37	-3.479	-4.186
40	-2.879	-4.160
49	-3.318	-4.110
52	-3.456	-3.194
61	-3.626	-3.863
70	-4.008	-4.160
83	-4.697	-5.404
86	-4.488	-4.625
87	-4.537	-5.404
88	-4.453	-4.423
101	-5.038	-4.474
128	-6.100	-7.824
129	-6.236	-5.714
134	-6.239	-7.002
136	-5.721	-5.240
153	-6.571	-5.360
155	-6.236	-6.571
171	-7.582	-6.133
185	-7.596	-7.663
187	-7.567	-7.663
194	-9.138	-8.210
206	-10.327	-9.097
208	-10.046	-10.925
209	-11.264	-11.876

Para compararmos a característica polar de algumas moléculas de PCBs, e verificarmos como esta característica influencia a solubilidade e o coeficiente de partição octanol-água, analisamos dois grupos de moléculas. No primeiro grupo, selecionamos três moléculas com o mesmo número de átomos de cloro (tab. 3.12). A posição dos átomos de cloro no anel determina a polaridade das moléculas. Para o nosso exemplo, observamos que mesmo se os dois átomos de cloro estiverem em um anel, e o terceiro átomo de cloro no outro anel em uma posição fixa (4'), a polaridade é influenciada pela distância entre os dois átomos de cloro que estão no mesmo anel. Quando estes dois átomos de cloro estão mais próximos (PCB 22 (fig.3.7a)) o momento de dipolo é maior do que quando os átomos de cloro estão mais afastados (PCBs 31 (fig. 3.7b) e 39 (fig. 3.7c)). A diminuição no momento de dipolo é acompanhado de uma diminuição no valor da solubilidade, e por um aumento no coeficiente de partição octanol-água. Este comportamento pode indicar que moléculas mais polares podem interagir mais com a água, e por isto são mais solúveis.

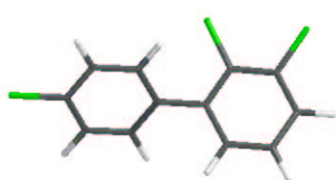
Tabela 3.12: Momento de dipolo para as moléculas selecionadas com o mesmo número de átomos de cloro.

Estrutura	PCB	μ (eV)	$\log S$ (mg/L)	$\log K_{OW}$
2 3 4'	22	2.274	-2.297	5.75
2 4' 5	31	1.089	-2.604	5.81
3 4' 5	39	0.188	-3.256	5.91

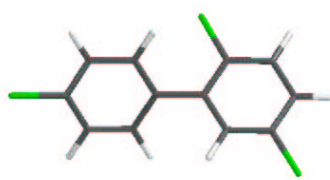
No segundo grupo de moléculas (tab. 3.13), observamos que mesmo quando a polaridade aumenta devido à presença de um átomo de cloro na posição 4' (PCB 104 (fig. 3.7d)), a solubilidade diminui e o coeficiente de partição octanol-água aumenta. Outro fato observado é que moléculas apolares (PCBs 55 (fig. 3.7e) e 155 (fig. 3.7f)) apresentam diferentes valores de solubilidade e de coeficiente de partição octanol-água. Nestas moléculas ocorre uma diminuição no valor da solubilidade e um aumento no coeficiente de partição octanol-água, quando se aumenta o número de átomos de cloro. O aumento no número de átomos de cloro é um fator determinante no tamanho da molécula, e na sua interação com as moléculas de água. Ou seja, o tamanho molecular é um fator mais determinante no processo de interação das moléculas de PCBs com as moléculas de água do que a polaridade, quando se compara moléculas que não tenham o mesmo número de átomos de cloro.

Tabela 3.13: Momento de dipolo para as moléculas selecionadas com número de átomos de cloro diferentes.

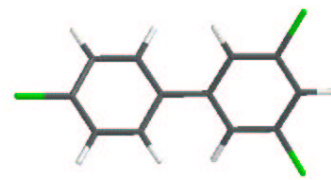
Estrutura	PCB	μ	$\log S$	$\log K_{OW}$
2 2' 6 6'	54	0	-2.704	6.08
2 2' 4' 6 6'	104	1.398	-4.571	6.80
2 2' 4 4' 6 6'	155	0	-6.236	7.39



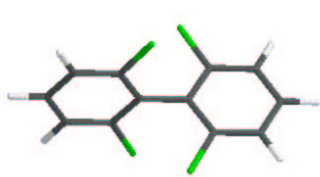
(a) 2 3 4'-CB



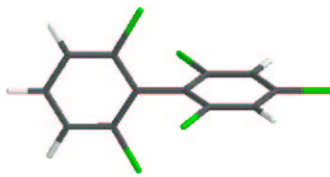
(b) 2 4' 5-CB



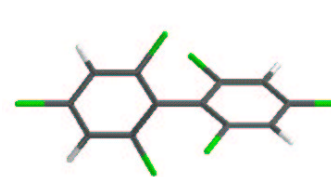
(c) 3 4' 5-CB



(d) 2 2' 6 6'-CB



(e) 2 2' 4' 6 6'-CB



(f) 2 2' 4 4' 6 6'

Figura 3.7: Moléculas de PCBs.

3.3 Coeficiente de Partição carbono orgânico

O conteúdo de carbono orgânico é um fator determinante da adsorção no solo, tornando o coeficiente de partição carbono-orgânico (K_{OC}) o mais usado parâmetro para indicar a mobilidade do composto no solo.⁹⁵ Assim como K_{OW} , K_{OC} é um coeficiente de partição entre a água e um meio orgânico, e devido a baixa solubilidade dos PCBs em água, ocorrem muitos problemas práticos na sua determinação.

Para compostos altamente hidrofóbicos como PCBs, a partição solo-água pode ser descrita como:

$$K_d = \frac{C_s}{C_w} \quad (3.1)$$

em que C_s e C_w são as concentrações do composto (PCB) no solo e na água respectivamente. Entretanto, K_d se refere à partição do composto em um solo específico. No caso dos PCBs a partição é controlada principalmente pela matéria orgânica do solo, então K_d pode ser normalizado pela fração de carbono orgânico (f_{oc}).⁹⁰

$$K_{OC} = \frac{K_d}{f_{oc}} \quad (3.2)$$

Os dados experimentais utilizados na modelagem foram obtidos no livro de Mackay.⁸⁷

Os descritores com maior correlação com $\log K_{OC}$ (eletronegatividade (η), ovalidade, índice de Winer (W), índice de complexidade de distância (H^D), $\log K_{OW}$, e $\log S$) são mostrados na figura 3.8. O modelo utilizou 62 amostras e 3 componentes principais.

A primeira PC representa 95.81% da variância total, como é mostrado na tabela 3.14, e a segunda e terceira representam 3.16% e 0.70% respectivamente. As três componentes principais foram utilizadas, visto que o valor de $PRESS$ diminui, e 99.68% da variância pode ser representada. Na tabela 3.15 podemos observar que todas as variáveis contribuem igualmente em módulo para a primeira PC. A segunda PC recebe uma contribuição significativa da eletronegatividade (η). Enquanto a terceira PC recebe uma contribuição significativa de $\log S$. Os valores previstos pelo modelo para $\log K_{OC}$ estão na figura 3.9 e na tabela 3.18, juntamente com os valores experimentais. Observamos que o erro obtido é muito baixo (em torno de 2%). Isto é devido ao fato de K_{OC} ser uma medida relacionada com a lipofilicidade dos compostos, e os PCBs serem altamente lipofílicos.

Analisando a tabela 3.16, observamos que o comportamento de $\log K_{OC}$ está coeren-

te com os sinais obtidos para $\log S$ e $\log K_{OW}$. Quanto menor for a solubilidade maior será o valor de K_{OC} , e quanto maior for K_{OW} maior será K_{OC} . Portanto, o comportamento de K_{OC} é similar ao comportamento de K_{OW} , ou seja, preferência dos PCBs pelo meio orgânico e não aquoso.

Na tabela 3.16 estão os descritores utilizados na modelagem e os coeficientes do vetor de regressão. Três amostras não fizeram parte da construção do modelo, e foram utilizadas na validação (tabela 3.17). O maior erro obtido foi de aproximadamente 1%. Os valores de $\log K_{OC}$ previstos para todos os PCBs estão na tabela C.3.

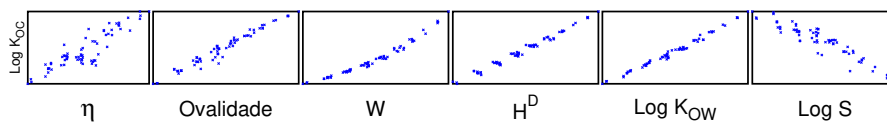


Figura 3.8: Gráfico de $\log K_{OC}$ versus descritores, mostrando a correlação.

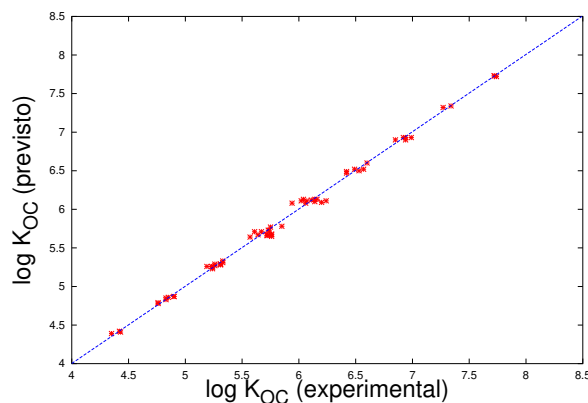


Figura 3.9: a) Gráfico de $\log K_{oc}$ previsto versus $\log K_{oc}$ experimental.

Tabela 3.14: Porcentagem da variância acumulada pelo modelo para $\log K_{OC}$.

# PC	Esta PC	Total
1	95.81	95.81
2	3.16	98.97
3	0.70	99.68

Tabela 3.15: *Loadings* das variáveis no modelo para $\log K_{OC}$.

Descritor	PC1	PC2	PC3
η	0.4267	-0.8891	0.2841
W	0.4526	0.2101	-0.3151
H^D	0.4481	0.3671	0.6878
$\log S$	-0.4524	-0.0450	0.5871
$\log K_{OW}$	0.4556	0.1694	-0.0466

Tabela 3.16: Vetor de regressão para $\log K_{OC}$. H^D índice de complexidade de distância, e W o índice de Winer.

Descritor	Vetor de regressão
η	0.0531
W	0.2019
H^D	0.4605
$\log S$	-0.1402
$\log K_{OW}$	0.2498

$$\begin{array}{lll}
 r = 0.999 & s^2 = 0.006 & Q^2 = 0.995 \\
 S_{PRESS} = 0.077 & F = 6.87 \times 10^3 & n = 62
 \end{array}$$

Tabela 3.17: Comparação do valor de $\log K_{OC}$ previsto e experimental para validação.

Congênero	Valor experimental	Valor previsto
2,4-	4.83	4.84
2,2',4,5'-	5.71	5.67
2,2',3,3',4,4',5,5',6,6'-	8.09	8.12

Tabela 3.18: Valor Previsto e Experimental de $\log K_{OC}$.

PCB	$\log K_{OC}$	
	Previsto	Experimental
1	4.39	4.35
2	4.42	4.42
3	4.41	4.43
4	4.78	4.76
8	4.83	4.83
9	4.85	4.83
10	4.79	4.76
11	4.87	4.90
12	4.86	4.85
14	4.87	4.90
18	5.25	5.24
20	5.26	5.23
21	5.26	5.19
26	5.29	5.31
28	5.29	5.31
29	5.29	5.26
30	5.23	5.24
31	5.28	5.31
33	5.27	5.26
35	5.33	5.33
37	5.31	5.33
40	5.64	5.57
44	5.67	5.64
47	5.66	5.72
50	5.67	5.76
51	5.65	5.76
53	5.68	5.76
54	5.66	5.72
60	5.71	5.67
61	5.71	5.61
66	5.73	5.74
70	5.73	5.73
75	5.68	5.72
<i>continua na próxima página.</i>		

<i>continuação da página anterior.</i>		
PCB	log K_{OC}	
	Previsto	Experimental
77	5.77	5.75
80	5.78	5.85
83	6.13	6.04
86	6.11	6.02
87	6.11	6.07
88	6.12	6.11
95	6.13	6.16
99	6.13	6.14
100	6.11	6.24
101	6.10	6.14
104	6.09	6.20
110	6.08	6.06
116	6.08	5.94
128	6.47	6.42
129	6.49	6.42
134	6.52	6.49
136	6.50	6.53
153	6.52	6.57
169	6.60	6.60
170	6.90	6.85
171	6.90	6.94
180	6.93	6.92
185	6.93	6.94
187	6.93	6.99
194	7.32	7.27
202	7.34	7.34
206	7.73	7.72
207	7.72	7.74
208	7.73	7.72

3.4 Constante da lei de Henry

A constante da lei de Henry (H) é essencialmente um coeficiente de partição ar-água que pode ser determinado pela medida da concentração de soluto em ambas as fases. A partir desta definição, H oferece informações de como e onde um composto tende a se acumular no equilíbrio. Substância com baixo valor de H tende a se acumular em fase aquosa, enquanto que substância com alto valor de H se acumula mais na fase gasosa. Como o ar e a água predominam em um modelo de ecossistema dividido em subsistemas, e como a água age como ligação entre todos estes subsistemas, conhecer H é muito importante no estudo do comportamento da substância no ecossistema.

Existem modelos teóricos para se determinar H , utilizando dados de potencial químico do soluto na fase aquosa e gasosa.⁹⁶ O valor de H pode também ser obtido dos dados de solubilidade (S) e pressão de vapor⁹⁷ (p_s):

$$H = \frac{p_s}{S} \quad (3.3)$$

$$\log H = \log p_s - \log S \quad (3.4)$$

Shiu *et al*²⁰ discute métodos de determinar, estimar e expressar este coeficiente. Neste trabalho, os dados utilizados foram obtidos no livro de Mackay.⁸⁷ As variáveis que apresentaram melhor correlação com $\log H$, são apresentadas na figura 3.10. São elas o volume, a ovalidade, o índice de complexidade de vértice (H^v), o índice de conectividade de caminho de ordem 6 (χ^6), e o $\log$ da solubilidade ($\log S$). O melhor modelo utilizou 5 descritores, para o grupo de 23 amostras. Outras duas amostras não foram usadas na construção do modelo, e foram utilizadas na validação.

A tabela 3.19 mostra que a primeira PC representa 83.68% da variância total, e é o resultado da contribuição de todas as variáveis, como pode ser observado na tabela 3.20. A segunda PC que representa 15.55% da variância. A terceira PC é acompanhada por um aumento no valor de PRESS. Desta maneira o modelo foi escolhido com duas PCs. Na tabela 3.21 estão os descritores utilizados no modelo, com os coeficientes do vetor de regressão e os dados estatísticos do modelo.

Como foi mencionado, H está correlacionado com o comportamento do composto em meio aquoso, e portanto com a solubilidade. O sinal positivo (tab. 3.21) para o coeficiente da solubilidade indica que conforme aumenta a solubilidade, aumenta-se o valor de H ,

o que aparentemente contradiz a equação 3.4. Este comportamento pode ser explicado analisando-se esta mesma equação, ou seja, se o valor de H diminui apesar da solubilidade diminuir também, então a pressão de vapor está diminuindo de maneira mais significativa do que a solubilidade. Podemos concluir que conforme se aumenta o número de átomos de cloro, a tendência do composto se acumular em meio aquoso é maior do que em meio gasoso.

Os valores de $\log H$ experimental e previsto pelo modelo estão na figura 3.11 e na tabela 3.23. O erro para a maioria das moléculas é menor do que 20%, exceto para dois compostos, os quais estão em negrito na tabela. A previsão para as duas amostras que não participaram da modelagem está na tabela 3.22.

Shiu²⁰ relaciona H com a pressão de vapor e a solubilidade, mas Nirmalakhanden⁹⁸ discute as dificuldades para se obter estes valores quando os compostos são hidrofóbicos e tem baixo valor de pressão de vapor, além dos erros que estão associados com estas medidas. Gramatica⁹¹ obtém um modelo para H utilizando descritores topológicos, e o número de átomos de cloro. Enquanto que o modelo obtido por nós, apesar de apresentar um coeficiente de correlação menor do que o obtido por outros autores, oferece uma idéia de como não só o número, mas a posição dos átomos de cloro na molécula podem influenciar sua interação com o meio, visto que a ovalidade é uma consequência tanto do número de átomos de cloro, como da posição destes na molécula.

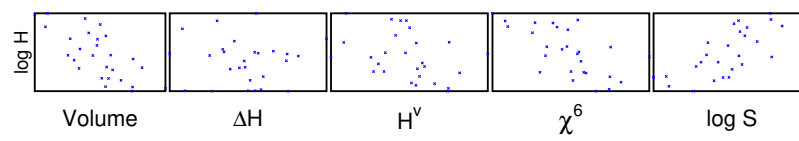


Figura 3.10: Gráfico de log H versus descritores, mostrando a correlação.

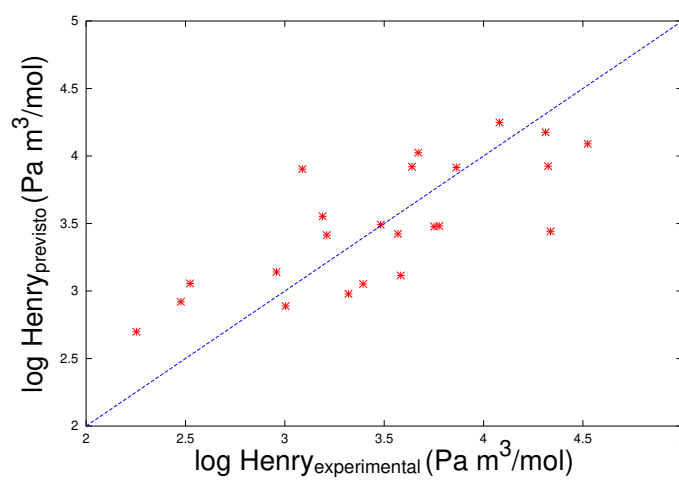


Figura 3.11: Gráfico de log H experimental versus log H previsto obtido pelo modelo.

Tabela 3.19: Porcentagem da variância acumulada pelo modelo para log H.

# PC	Esta PC	Total
1	83.68	83.68
2	15.55	99.24
3	0.65	99.89

Tabela 3.20: *Loadings* das variáveis no modelo para log H.

Descritor	PC1	PC2
Volume	-0.4562	0.4018
Ovalidade	-0.4437	0.4467
H^v	-0.3873	-0.6910
χ^6	-0.4721	0.2896
log S	0.4712	-0.2787

Tabela 3.21: Vetor de regressão para log H .

Descritor	Vetor de regressão
Volume	-0.0383
Ovalidade	0.1178
H^v	-0.7617
χ^6	-0.0447
log S	0.0596

$$r = 0.751 \quad s^2 = 0.180 \quad Q^2 = 0.563$$

$$S_{PRESS} = 0.424 \quad F = 25.81 \quad n = 23$$

Tabela 3.22: Comparação do valor de log H previsto e experimental para validação.

Congênero	Valor experimental	Valor previsto
2,3',5-	3.250	3.847
2,2',3,4'-	3.520	3.538

Tabela 3.23: Valor Previsto e Experimental de $\log H$ (Pa m^3 /mol).

PCB	$\log H$ (Pa m^3 /mol)	
	Previsto	Experimental
1	4.175	4.312
2	3.925	4.325
3	3.477	3.751
4	4.248	4.080
9	4.024	3.671
15	2.889	3.004
18	4.090	4.524
29	3.553	3.190
33	3.482	3.777
40	3.903	3.088
47	3.141	2.958
52	3.915	3.863
77	2.699	2.253
87	3.414	3.211
101	3.423	3.569
128	2.920	2.477
155	3.442	4.337
171	3.052	3.394
202	3.920	3.640
206	2.979	3.320
207	3.115	3.583
208	3.492	3.482
209	3.056	2.523

3.5 Coeficiente de Atividade Aquosa

A forma geral do potencial químico de um soluto é:

$$\mu_A(s) = \mu_A^*(l) + RT \ln Y_W \quad (3.5)$$

em que Y_W é a atividade do soluto A .

Os dados experimentais utilizados foram obtidos no artigo de Myrdal.⁹⁹ Os descritores que apresentaram melhor correlação com $-\log Y_W$ (fig.3.12) foram o volume molecular, o calor de formação (ΔH), a energia de LUMO (E_{lumo}), a ovalidade, o índice de Winer (W) e o $\log S$. O modelo foi obtido com 80 amostras e 2 componentes principais.

Na tabela 3.24 observamos que a primeira PC representa 91.97% da variância total, e que a segunda PC representa 7.79%, e a terceira PC representa apenas 0.14% desta variância. Ocorre também que o menor valor de *PRESS* corresponde à duas PCs. A primeira PC é o resultado da combinação de todas as variáveis tendo o módulo dos coeficientes com valores próximos (tab. 3.25). A segunda PC tem como maior contribuição a variável E_{lumo} seguida pelo volume. As contribuições das variáveis E_{lumo} e $\log S$ apresentaram sinais negativos enquanto que o volume e o W apresentaram sinais positivos (tab. 3.26). O sinal negativo de E_{lumo} mostra que quanto maior o caráter eletroaceptor da molécula menor será Y_W (os valores utilizados foram $-\log Y_W$), e compostos menos solúveis em água apresentam uma menor atividade aquosa.

Na tabela 3.26 estão os descritores utilizados na modelagem e os coeficientes do vetor de regressão juntamente com os dados estatísticos desta modelagem. As amostras que foram utilizadas para validação tiveram seus valores de $-\log Y_W$ previstos, os quais se encontram na tabela 3.27. Os erros obtidos estão abaixo de 10%. Os valores previstos versus os experimentais estão na figura 3.13 e na tabela 3.28. O maior erro obtido foi de 10%. Os valores previstos para todos os PCBs estão na tabela C.3. Nesta tabela é possível observar que houve extrapolação para a amostra 209, em que o valor previsto é 5% maior que o maior valor utilizado na modelagem. Como os modelos obtidos neste trabalho são modelos aplicáveis apenas a uma determinada região, a extrapolação não é recomendada. Mas considerando o quanto se extrapolou, podemos aceitar o valor obtido.

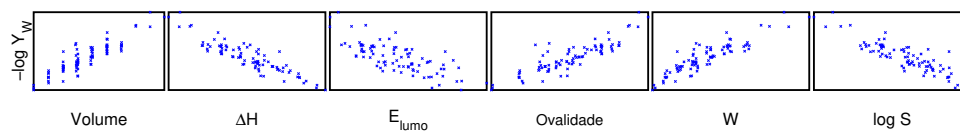


Figura 3.12: Gráfico de $-\log Y_W$ versus descritores, mostrando a correlação.

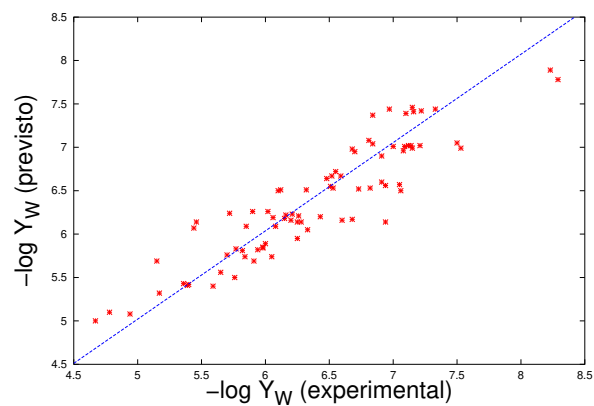


Figura 3.13: Gráfico dos valores de $-\log Y_W$ experimental versus os valores previstos pelo modelo.

Tabela 3.24: Porcentagem da variância acumulada pelo modelo para $-\log Y_W$.

# PC	Esta PC	Total
1	91.97	91.97
2	7.79	99.76
3	0.14	99.89

Tabela 3.25: *Loadings* das variáveis no modelo par $-\log Y_W$

Descritor	PC1	PC2
Volume	0.5088	0.3788
E_{lumo}	-0.4535	0.8838
W	0.5145	0.2680
$\log S$	-0.5203	-0.0602

Tabela 3.26: Vetor de regressão para $-\log Y_W$.

Descritor	Vetor de regressão
Volume	0.3381
E_{lumo}	-0.0094
W	0.3007
$\log S$	-0.2692

$$r = 0.910 \quad s^2 = 0.091 \quad Q^2 = 0.831$$

$$S_{PRESS} = 0.302 \quad F = 378.85 \quad n = 80$$

Tabela 3.27: Comparação do valor de $-\log Y_W$ previsto e experimental para validação.

Congênero	Valor experimental	Valor previsto
2,3,4'-	5.82	5.81
2,2',4,6,6'-	6.73	6.52
2,2',3,4,5,5'-	7.09	7.01

Tabela 3.28: Valor Previsto e Experimental de $-\log Y_W$.

PCB	$-\log Y_W$	
	Previsto	Experimental
1	5.00	4.67
2	5.08	4.94
3	5.10	4.78
6	5.40	5.59
7	5.43	5.36
8	5.42	5.40
9	5.41	5.39
10	5.32	5.17
11	5.50	5.76
15	5.56	5.65
16	5.69	5.91
18	5.74	5.84
19	5.69	5.15
22	5.81	5.82
24	5.74	6.05
26	5.84	5.98
28	5.89	6.00
29	5.83	5.77
30	5.76	5.70
31	5.85	5.98
33	5.82	5.94
37	5.95	6.25
40	6.09	6.08
42	6.14	6.25
44	6.14	6.28
46	6.07	5.44
47	6.20	6.43
48	6.17	6.68
50	6.14	5.46
52	6.16	6.20
53	6.09	5.85
54	6.05	6.33
56	6.22	6.16
<i>continua na próxima página.</i>		

<i>continuação da página anterior.</i>		
PCB	-log Y_W	
	Previsto	Experimental
60	6.24	5.72
61	6.16	6.60
63	6.21	6.26
66	6.26	5.90
69	6.14	6.94
70	6.23	6.21
74	6.26	6.02
75	6.18	6.15
76	6.19	6.06
82	6.51	6.12
86	6.51	6.32
87	6.53	6.53
88	6.50	7.06
95	6.50	6.10
97	6.55	6.51
100	6.56	6.94
101	6.60	6.91
104	6.52	6.73
110	6.57	7.05
114	6.67	6.59
116	6.53	6.82
118	6.72	6.55
122	6.64	6.48
124	6.67	6.52
128	6.99	7.53
129	6.99	7.15
134	6.96	7.08
135	6.95	6.70
136	6.90	6.91
137	7.02	7.12
138	7.02	7.14
139	7.01	7.00
141	7.01	7.09
151	6.98	6.68
<i>continua na próxima página.</i>		

<i>continuação da página anterior.</i>		
PCB	-log Y_W	
	Previsto	Experimental
153	7.05	7.50
156	7.08	6.81
158	7.04	6.84
163	7.02	7.21
170	7.44	6.97
171	7.42	7.22
174	7.37	6.84
180	7.46	7.15
183	7.44	7.33
185	7.39	7.10
187	7.41	7.16
194	7.89	8.23
202	7.78	8.29

3.6 Ponto de Fusão

É a temperatura na qual sob uma pressão específica, líquido e sólido coexistem em equilíbrio. Os dados experimentais utilizados na modelagem foram obtidos no livro de Mackay.¹⁴ Foram utilizadas 70 amostras. Os descritores que apresentaram melhor correlação linear com o ponto de fusão (mp) foram: volume molecular, calor de formação (ΔH), energia de lumo (E_{lumo}), ovalidade, índice de caminho de ligação de ordem 4 (χ_b^4) e índice de caminho de ordem 6 (χ^6). O gráfico destes descritores com mp está na figura 3.14. Observamos nesta figura que há grupos (linhas verticais) no gráfico do volume, então expandimos o gráfico $mp \times$ volume (fig. 3.15). As linhas verticais representam os grupos de homólogos, são oito grupos de homólogos pois o PCB com maior número de átomo de cloro é o 202 (2,2',3,3',5,5',6,6'). Este comportamento era esperado visto que o volume do átomo de hidrogênio é bem menor do que o volume do átomo de cloro. Portanto, a substituição de um átomo de hidrogênio por um átomo de cloro causa um aumento significativo no volume molecular. Entretanto, isto não significa que PCBs pertencentes a um mesmo grupo homólogo tenham o mesmo volume. Na figura 3.15b é possível observar a variação do volume para PCBs de um mesmo grupo homólogo. Esta figura apresenta dois grupos, sendo que as moléculas de menor volume tem dois átomos de cloro nas posições 2 e 2'.

A tabela com a relação da variância acumulada por cada componente principal (tab. 3.29) mostra que as duas primeiras PCs acumulam 97.36% dos dados, enquanto que a terceira PC acumula apenas 2.31%. O menor valor de *PRESS* é obtido para duas PCs. Estes fatos indicam que a variância na terceira PC deve estar associada ao ruído contido nas variáveis.

Na tabela 3.26 estão os descritores utilizados na modelagem (ΔH , E_{lumo} , ovalidade e χ^6), e os coeficientes do vetor de regressão juntamente com os dados estatísticos desta modelagem. Três amostras não fizeram parte da construção do modelo, e foram utilizadas na validação (tabela 3.32). O maior erro obtido foi de aproximadamente 5%.

Todas as variáveis contribuem igualmente em módulo para a primeira PC. A segunda PC tem contribuição significativa da E_{lumo} (tab. 3.30). O sinal negativo em ΔH (tab. 3.31) indica que quanto menor ΔH , maior o ponto de fusão. O mesmo ocorre com a variável E_{lumo} , quanto menor seu valor maior o ponto de fusão. A dependência com a ovalidade mostra a dependência de mp com o tamanho e a simetria molecular. Esta relação do ponto de fusão com a forma molecular já é conhecida.¹⁰⁰

Os valores previstos pelo modelo, e os valores experimentais estão na figura 3.16 e na

tabela 3.33. Observa-se que os erros estão abaixo de 25%, exceto para os compostos que estão em negrito na tabela. Os valores previstos para as amostras que não participaram da modelagem estão na tabela 3.32. O modelo obtido apresenta resultado comparável com o trabalho de Gramatica,⁹¹ em que o coeficiente de correlação é $r = 0.852$, com a vantagem que nosso modelo recebe contribuição de diferentes tipos de descritores. Observamos na tabela C.3 que para as últimas moléculas (204-209) ocorreram extrapolações. A extrapolação para a molécula 204 é de 3%, chegando à 30% para a molécula 209. Esta extrapolação pode ter ocorrido porque os valores realmente sejam maiores, visto que as moléculas apresentam mais átomos de cloro (amostras 205-209). Mas como comentamos anteriormente, o modelo é um modelo aplicável a uma região determinada, e não é recomendado utilizar valores extrapolados.

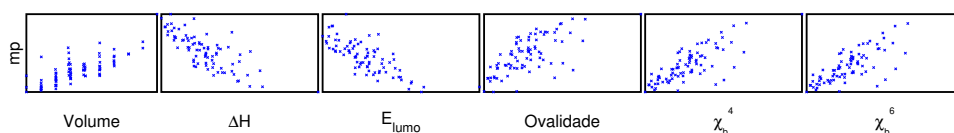


Figura 3.14: Gráfico *mp* versus descritores, mostrando a correlação.

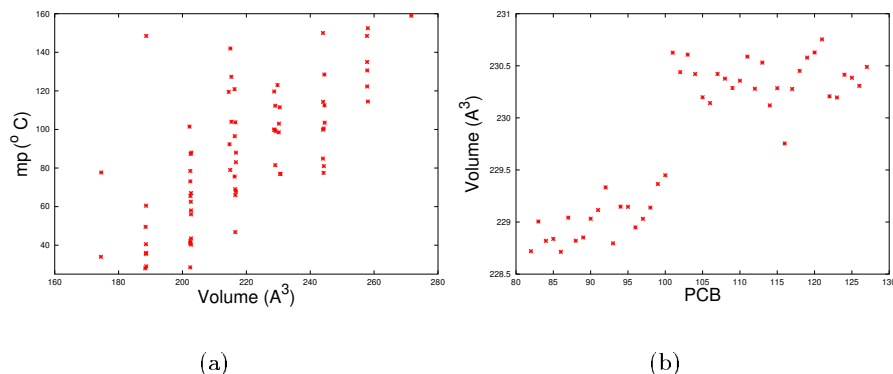


Figura 3.15: a) Gráfico de *mp* versus o volume. b) Gráfico da variação do volume para PCBs com o mesmo número de átomos de cloro.

Tabela 3.29: Porcentagem da variância acumulada pelo modelo para mp .

# PC	Esta PC	Total
1	87.23	87.23
2	10.12	97.36
3	2.31	99.66

Tabela 3.30: *Loadings* das variáveis no modelo pra mp

Descritor	PC1	PC2	PC3
ΔH	-0.5262	0.2112	0.2811
E_{lumo}	-0.4473	-0.8554	0.2492
Ovalidade	0.5128	-0.4183	-0.3483
χ^6	0.5100	-0.2208	0.8588

Tabela 3.31: Vetor de regressão para mp .

Descritor	Vetor de regressão
ΔH	-0.2013
E_{lumo}	-0.4056
Ovalidade	0.0778
χ^6	0.2152

$$r = 0.852 \quad s^2 = 3.21 \quad Q^2 = 0.723$$

$$S_{PRESS} = 20.28 \quad F = 173.95 \quad n = 70$$

Tabela 3.32: Comparação entre o valor de mp previsto e experimental para validação.

Congênero	Valor experimental	Valor previsto
2,3',4'-	65.5	69.0
2,2',3,4,5-	99.9	100.6
2,2',3,3',4,5,5',6,6'-	182.0	189.5

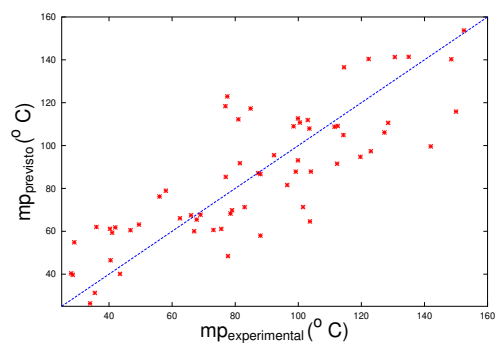


Figura 3.16: Gráfico dos valores mp experimental versus os valores previstos pelo modelo

Tabela 3.33: Valor Previsto e Experimental de mp .

PCB	$mp^{\circ}C$	
	Previsto	Experimental
1	26.3	34.0
2	40.2	16.5
3	48.4	77.7
5	40.3	28.0
7	55.7	24.3
8	46.5	40.5
9	39.4	23.0
10	31.3	35.5
11	54.9	29.0
12	63.1	49.5
14	62.0	36.0
16	39.7	28.5
18	40.1	43.5
20	61.8	42.0
21	71.3	101.5
22	60.6	73.1
23	59.4	41.0
<i>continua na próxima página.</i>		

<i>continuação da página anterior.</i>		
PCB	<i>mp° C</i>	
	Previsto	Experimental
26	61.2	40.3
28	76.2	56.0
29	68.3	78.5
30	66.1	62.5
31	60.0	67.0
34	78.9	58.0
37	87.2	87.3
39	86.7	88.0
42	67.8	69.0
44	60.5	46.8
45	61.1	75.6
47	71.3	83.0
48	67.5	66.0
49	65.5	67.8
52	58.0	88.0
53	64.6	103.7
56	81.6	96.5
60	99.6	142.0
61	95.5	92.3
65	69.8	79.0
66	106.1	127.3
70	87.9	104.0
80	121.0	164.0
82	94.7	119.7
87	91.5	112.3
88	93.1	99.9
95	87.8	99.2
97	91.8	81.5
101	85.4	77.0
105	111.9	103.0
114	109.0	98.5
116	97.4	123.0
118	108.8	111.5
120	118.4	76.9
<i>continua na próxima página.</i>		

<i>continuação da página anterior.</i>		
PCB	<i>mp° C</i>	
	Previsto	Experimental
128	115.9	150.0
129	117.3	84.9
133	110.6	128.5
134	112.7	99.9
136	104.9	114.3
137	122.9	77.5
138	112.3	81.0
151	110.7	100.5
153	107.9	103.5
155	109.1	112.5
166	114.0	162.5
170	141.4	135.0
171	140.4	122.3
174	141.3	130.6
180	136.5	114.5
182	153.7	152.5
185	140.3	148.5
194	161.4	159.0
202	160.8	161.0

O estudo apresentado até este momento, dá uma idéia do comportamento dos PCBs em meio aquoso, no solo e no ar. Todas estas propriedades são importantes pois dão uma idéia da mobilidade dos PCBs no meio ambiente.

Cerca de um bilhão de dólares são gastos em pesquisas para se determinar as reações tóxicas causadas por PCBs. Existem duas fontes que originam dados de toxicidades: *a*) experimentos em animais, que são deliberadamente expostos; *b*) estudos em seres humanos que foram acidentalmente expostos. Os dados em animais possibilitam expor um número necessário de animais à taxas de concentrações maiores. Mas o sistema imunológico dos animais difere do sistema humano, e os resultados não são diretamente aplicados à população humana. Entretanto, possibilitam uma idéia do perigo que cada congênero de PCB oferece.¹⁰¹

3.7 Toxicidade IC₅₀

Concentração necessária para imobilizar 50% dos indivíduos testados. Os experimentos foram feitos em células cerebrais de ratos. Os dados foram retirados do artigo de Kodavanti.¹⁰² O modelo foi obtido com 12 amostras e 2 componentes principais. As variáveis que apresentaram uma melhor correlação com log IC₅₀ são o calor de formação (ΔH), a energia de homo (E_{homo}), a solubilidade (log S) e o índice de caminho de ligação de ordem 6 (χ_b^6). Estes descritores estão na figura 3.17.

A tabela 3.34 mostra que a primeira PC acumula um total de 84.78% da variância, enquanto que a segunda PC acumula 14.90%, totalizando 99.68% da variância total. O modelo utilizou estas duas PCs. Analisando a tabela 3.35 observamos que a primeira PC tem contribuição de todas as variáveis, enquanto que a principal contribuição da segunda PC vem da E_{homo} . A comparação entre os valores experimentais e os valores previstos para log IC₅₀ estão na figura 3.18 e na tabela 3.38. Nota-se que o maior erro obtido foi de 7%.

Na tabela 3.36 estão os coeficientes do vetor de regressão dos descritores utilizados na modelagem, juntamente com os dados estatísticos. A variável ΔH contribui positivamente com a concentração IC₅₀. Um menor valor de ΔH se refere à estabilidade molecular, quanto mais estável for esta molécula, menor a concentração de IC₅₀ e maior a toxicidade. As variáveis E_{homo} e log S contribuem positivamente com a concentração IC₅₀. A E_{homo} está associada ao caráter doador de elétrons da molécula. O log S indica que a interação dos PCBs com as células testadas dependem do quão solúvel é o PCB. Quanto mais solúvel, maior a concentração necessária para imobilizar 50% dos indivíduos testados, ou seja menos tóxico o PCB. Isto é uma evidência de que a interação está ocorrendo em um meio hidrofóbico, o que era esperado, visto que as células do sistema nervoso são ricas em

lipídios.

As amostras reservadas para validação tiveram seus valores de $\log IC_{50}$ previstos e estão na tabela 3.37, onde se verifica que o maior erro foi de 8%. A previsão para todas as moléculas de PCBs está na tabela C.3. A toxicidade aumenta com aumento do número de átomos de cloro (quanto menor o valor de $\log IC_{50}$, maior a toxicidade). Por exemplo, a concentração prevista para o PCB 83 (2,2',3,3',5-CB) foi $\log IC_{50}=1.953$, enquanto que para o PCB 129 (2,2',3,3',4,5-CB), o qual apresenta um átomo de cloro a mais, o valor previsto foi $\log IC_{50}=1.668$, ou seja, o PCB 129 com maior número de átomos de cloro é mais tóxico do que o PCB 83. Para um mesmo grupo de homólogos, a presença do átomo de cloro na posição orto, em geral aumenta a toxicidade. Por exemplo, os PCBs 4 (2,2'-CB) e 10 (2,6-CB), são os mais tóxicos entre os diclorofenis. Estes PCBs apresentam os átomos de cloro na posição orto, sendo que para o PCB 10, que tem os átomos de cloro no mesmo anel ($\log IC_{50}=2.015$) a toxicidade é maior do que a toxicidade do PCB 4 ($\log IC_{50}=2.064$), que tem os átomos de cloro um em cada anel. Não podemos deixar de mencionar as extrapolações, que ocorreram basicamente com as moléculas com 1, 7, 8 e 9 átomos de cloro, e ficaram em torno de 5%, com exceção da amostra 209 que teve uma extrapolação de 13%. Isto ocorreu porque a modelagem não teve nenhuma molécula pertencente a estes grupos de homólogos. Entretanto, estes valores mesmos que extrapolados, servem como referência para o fato das moléculas serem as mais tóxicas (moléculas com 7, 8 e 9 átomos de cloro).

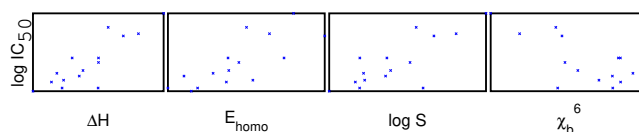


Figura 3.17: Gráfico de $\log IC_{50}$ versus descritores, mostrando a correlação.

Tabela 3.34: Porcentagem da variância acumulada pelo modelo para log IC₅₀.

# PC	Esta PC	Total
1	84.78	84.78
2	14.90	99.68

Tabela 3.35: *Loadings* das variáveis no modelo para log IC₅₀.

Descritor	PC1	PC2
ΔH	0.5327	-0.2455
E_{homo}	0.3923	0.8958
χ_6^b	-0.5361	0.1857
$\log S$	0.5243	-0.3207

Tabela 3.36: Vetor de regressão para log IC₅₀.

Descritor	Vetor de regressão
ΔH	0.2161
E_{homo}	0.3265
χ_6^b	-0.2275
$\log S$	0.2136

$$r = 0.884 \quad s^2 = 0.007 \quad Q^2 = 0.782$$

$$S_{PRESS} = 0.086 \quad F = 32.28 \quad n = 12$$

Tabela 3.37: Comparação entre o valor de log IC₅₀ previsto e experimental para validação.

Congênero	Valor experimental	Valor previsto
2,4,4'-	1.932	2.039
2,2',5,5'-	1.589	1.875
2,2',4,4',5,5'-	1.887	1.660

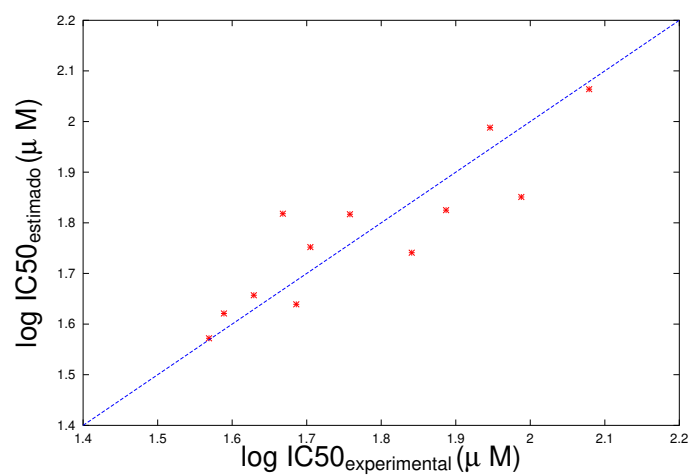


Figura 3.18: Gráfico de $\log IC_{50}$ previsto pelo modelo versus o experimental.

Tabela 3.38: Valor Previsto e Experimental de $\log IC_{50}$.

PCB	$\log IC_{50}(\mu M)$	
	Previsto	Experimental
4	2.064	2.079
19	1.988	1.946
47	1.817	1.758
50	1.851	1.988
104	1.752	1.705
105	1.818	1.668
118	1.825	1.887
128	1.621	1.589
133	1.657	1.629
136	1.741	1.841
156	1.639	1.686
180	1.572	1.569

3.8 Toxicidade EC₅₀

O valor EC₅₀, representa a quantidade de PCB necessária para se observar uma reação neurotoxicológica em 50% dos indivíduos testados. EC₅₀ foi obtido com estudos em ratos e camundongos. Os valores experimentais para as 29 amostras utilizadas neste modelo, foram obtidos do review de Giesy.¹⁸ As variáveis que apresentaram uma melhor correlação com o log EC₅₀ foram a dureza, o log K_{OW} , o índice de conectividade de cluster de ordem 6 (${}^6\chi_{pc}$), o índice de conectividade de valência de ordem 5 (χ_v^5) e o índice de conectividade de cluster de valência de ordem 6 (${}^6\chi_{pc}^v$). Estes descritores estão no gráfico da figura 3.19. Os descritores utilizados na modelagem foram: dureza, χ_v^5 , ${}^6\chi_{pc}^v$ e log K_{OW}

Na tabela 3.39 observamos que a primeira PC acumula uma variância de 70.81% e a segunda PC acumula 26.43% da variância total, totalizando 97.24%. A terceira PC representa apenas 1.70%, e o valor de *PRESS* é muito próximo do valor obtido com duas PCs. Todas as variáveis contribuem para a primeira PC com coeficientes de módulo próximos. A segunda PC é resultado da composição principalmente da dureza, seguida pelo log K_{OW} e por χ_v^5 (tab. 3.40). O valor previsto de log EC₅₀ para cada amostra está na tabela 3.43, juntamente com o valor experimental, estes valores também estão na figura 3.18. Os erros obtidos ficaram abaixo de 10%. Na tabela 3.42 estão os valores previsto e experimental de três amostras separadas para validação. O maior erro obtido para estas amostras foi de 17%.

Na tabela 3.41 estão os descritores utilizados na modelagem com coeficientes do vetor de regressão e os dados estatísticos. A dureza molecular se refere ao *gap* molecular, quanto menor o valor da dureza (menor o *gap*), mais mole é a molécula e maior sua reatividade. Então, quanto mais mole for a molécula maior o valor de EC₅₀, ou seja menos tóxica é a molécula. Este fato era esperado, pois quanto mais reativa a molécula, maior a possibilidade dela se decompor. O sinal de K_{OW} negativo, indica que quanto mais lipofílica for a molécula de PCB, maior sua toxicidade, isto é, conforme aumenta o valor de K_{OW} , a concentração de PCB necessária para que ocorra o efeito neurotoxicológico diminui. Isto pode ser explicado pelo fato que já foi mencionado, que as células do sistema nervoso são ricas em lipídios, com o aumento da lipofilicidade do PCB, aumenta a possibilidade de interação entre os PCBs e estas células.

Observando as previsões para log EC₅₀ para os 209 PCBs na tabela C.3, verificamos claramente que não existe correlação direta entre o volume ou a forma molecular, com a toxicidade como ocorreu nos caso anteriores. Por outro lado, há uma tendência positiva

entre índices de conectividade de valência e a toxicidade do PCB. Estes índices são calculados utilizando-se a equação 2.32, que contém o número de elétrons de valência do átomo em questão. Então, a distribuição das cargas na estrutura, neste caso, é mais importante que a forma molecular. Na tabela C.3 também observamos que duas extrapolações ocorrem, para as moléculas 77 e 81. Estas extrapolações são de 3% e 4%, esta faixa de extrapolação pode ser aceita, pois está dentro da faixa de erro do modelo.

O comportamento da toxicidade EC_{50} é similar ao da toxicidade IC_{50} , como pode ser verificado na tabela C.3. Em um mesmo grupo de homólogos a toxicidade é maior quando há átomos de cloro na posição orto. Os mais tóxicos em cada grupo de homólogos são: PCB 17 (2,2',4-CB) no grupo de triclorobifenis, os PCBs 40 (2,2',3,3'-CB), 42 (2,2',3,4'-CB), 53 (2,2',5,6'-CB) e 54 (2,2',6,6'-CB) no grupo de tetraclorobifenis, os PCBs 104 (2,2',4,6,6'-CB) e 125 (2',3,4,5,6'-CB) no grupo de pentaclorobifenis, o PCB 135 (2,2',4,4',6,6'-CB) no grupo de hexaclorobifenil, o PCB 184 (2,2',3,4,4',6,6'-CB) no grupo de heptaclorobifenil, os PCBs 197 (2,2',3,3',4,4',6,6'-CB) e 202 (2,2',3,3',5,5',6,6'-CB) no grupo de octaclorobifenis, e o PCB 207 (2,2',3,3',4,4',5,6,6'-CB) no grupo de nonaclorobifenis.

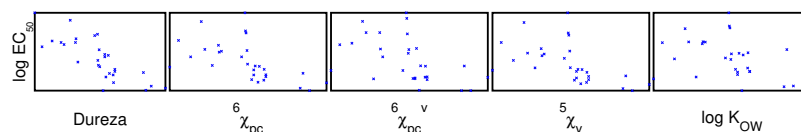


Figura 3.19: Gráfico de $\log EC_{50}$ versus descritores, mostrando a correlação.

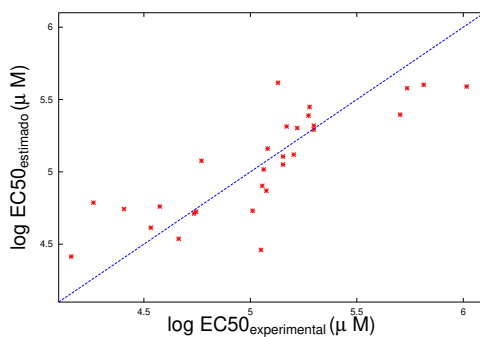


Figura 3.20: a) Gráfico de $\log EC_{50}$ previsto e experimental.

Tabela 3.39: Porcentagem da variância acumulada pelo modelo para log EC₅₀.

# PC	Esta PC	Total
1	70.81	70.81
2	26.43	97.24
3	1.70	98.93

Tabela 3.40: *Loadings* das variáveis no modelo para log EC₅₀.

Descritor	PC1	PC2	PC3
Dureza	-0.4324	-0.6654	-0.0136
${}^5\chi^v$	-0.5252	0.4041	-0.7425
${}^6\chi_{pc}^v$	-0.5622	0.2531	0.6683
log K_{OW}	-0.4702	0.5743	0.0427

Tabela 3.41: Vetor de regressão para log EC₅₀. ${}^5\chi^v$ o índice de conectividade de valência de ordem 5, e ${}^6\chi_{pc}^v$ o índice de conectividade de cluster de valência de ordem 6.

Descritor	Vetor de regressão
Dureza	-0.6862
${}^5\chi^v$	-0.1026
${}^6\chi_{pc}^v$	-0.0555
log K_{OW}	-0.0694

$$r = 0.822 \quad s^2 = 0.068 \quad Q^2 = 0.676$$

$$S_{PRESS} = 0.262 \quad F = 54.12 \quad n = 29$$

Tabela 3.42: Comparação entre o valor de log EC₅₀ previsto e experimental para validação.

Congênero	Valor experimental	Valor previsto
2,2',5,5'-	4.454	4.702
2,2',3,3',4,4',6-	4.898	4.844
2,2',3,4,4',5,6-	5.914	4.885

Tabela 3.43: Valor Previsto e Experimental de $\log EC_{50}$.

PCB	$\log EC_{50}(\mu M)$	
	Previsto	Experimental
1	5.119	5.204
2	5.395	5.704
3	5.602	5.814
4	4.414	4.159
5	5.051	5.153
6	5.106	5.153
7	5.319	5.298
8	5.292	5.298
10	4.537	4.663
11	5.389	5.273
12	5.616	5.130
13	5.590	6.016
18	4.743	4.407
24	4.870	5.075
26	5.161	5.081
28	5.449	5.278
30	4.731	5.011
31	5.314	5.170
33	5.303	5.220
39	5.578	5.737
44	4.712	4.736
47	4.724	4.745
49	4.760	4.575
50	4.787	4.263
75	5.077	4.771
100	5.016	5.063
103	4.904	5.056
104	4.614	4.533
155	4.461	5.050

3.9 Propriedades Eletrônicas

Como para muitas propriedades estudadas, o caráter eletroceptor (E_{lumo}) e eletrodoador da molécula (E_{homo}) estiveram presentes, assim como outras propriedades eletrônicas tais como dureza molecular, eletronegatividade e dipolo. Aachamos conveniente uma análise do potencial eletrostático, dos orbitais moleculares de HOMO e LUMO e da densidade de elétrons, o que foi feito para algumas moléculas, com diferente número de átomos de cloro em posições diferentes.

A densidade de elétrons, $\rho(\mathbf{r})$, para uma molécula é definida em termos da soma do quadrado de todos os orbitais moleculares ocupados,

$$\rho(\mathbf{r}) = \sum_{\mu} \sum_{\nu} P_{\mu\nu} \phi_{\mu}(\mathbf{r}) \phi_{\nu}(\mathbf{r}) \quad (3.6)$$

em que, como foi mencionado no capítulo 2, $P_{\mu\nu}$ são os elementos da matriz densidade de um elétron e ϕ os orbitais atômicos. A densidade eletrônica pode ser visualizada utilizando-se curvas de contorno, onde cada curva representa uma igual densidade de elétrons. Enquanto a densidade de elétrons reflete somente a distribuição eletrônica, o potencial eletrostático tem contribuição tanto do núcleo como dos elétrons. O potencial eletrostático é obtido segundo a equação 2.26. Como as forças eletrostáticas são as principais responsáveis pela interação de longo alcance entre as moléculas, o potencial eletrostático pode ser de grande ajuda no estudo da interação entre moléculas. Os orbitais moleculares, $\psi_i(\mathbf{r})$, são expressos como combinação dos orbitais atômicos $\phi_{\mu}(\mathbf{r})$,

$$\psi_i(\mathbf{r}) = \sum_{\mu} c_{\mu i} \phi_{\mu}(\mathbf{r}) \quad (3.7)$$

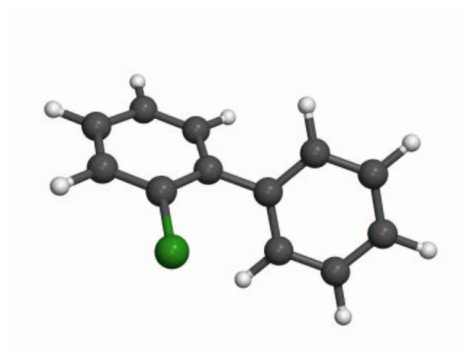
em que $c_{\mu i}$ são os coeficientes do orbital.

As figuras seguintes apresentam um esquema da molécula, os orbitais moleculares de LUMO e HOMO, a densidade de elétrons ($0.02 \text{ e}\text{\AA}^3$) e a superfície de potencial eletrostático. A representação da densidade de elétrons, indica que quanto mais próximas as linhas maior a densidade eletrônica naquele ponto. O potencial eletrostático é representado

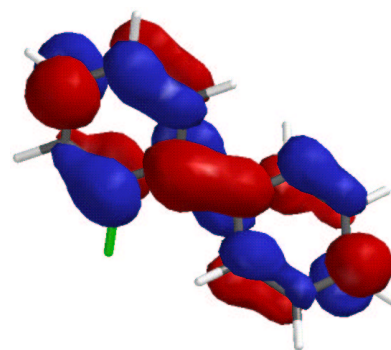
por um mapa de cores, onde a cor azul indica onde o potencial é positivo, e as regiões em vermelho onde ele é negativo.

Nas figuras de 3.21-3.31 observamos que os orbitais de LUMO se concentram principalmente nos átomos dos anéis, com exceção das moléculas 10 e 24 (figuras 3.23 e 3.25 respectivamente), em que o orbital de LUMO se concentra apenas em um anel. Nestas moléculas os átomos de cloro estão apenas em um dos anéis e nas posições orto. Outra exceção ocorre na molécula 204 (fig. 3.30): nesta molécula o orbital de LUMO se concentra no anel com cinco átomos de cloros. Para estas moléculas o comportamento do HOMO também difere das outras moléculas. Os orbitais de HOMO se concentram nos átomos dos anéis e nos átomos de cloro, sendo que quando existe átomo de cloro na posição orto e para, não existe densidade de HOMO no átomo de cloro da posição meta, caso exista.

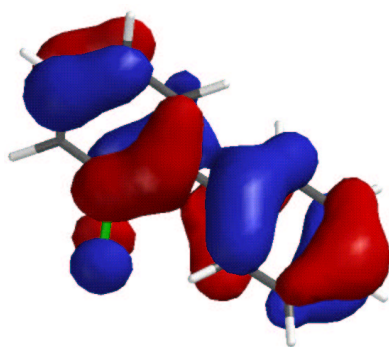
A densidade de elétrons se concentra nos átomos de carbono (figs. 3.21 e 3.22) quando o número de átomos de cloro é pequeno. Quando o número de átomos de cloro aumenta, observa-se que a concentração de elétrons, passa a ocorrer não apenas no anel, mas nos átomos de cloro também. O anel aromático é conhecidamente eletroceptor (com alta afinidade eletrônica). Os átomos de cloro também tem alta afinidade eletrônica. Na molécula 2-CB (fig. 3.21) observamos no anel que não contém átomo de cloro, que a alta afinidade eletrônica do anel aromático é bastante visível, os átomos de hidrogênios estão azuis e o anel está amarelo. No outro anel, o átomo de cloro e o anel estão amarelos. Este comportamento é também observado na molécula 3-CB (fig. 3.22). A molécula 3,3',4-CB (fig. 3.26) apresenta uma densidade eletrônica maior na região onde estão os dois átomos de cloro vizinhos. Na molécula 2,2',3,3'-CB (fig. 3.27), a densidade de elétrons é maior nos átomos de cloro e no anel. Isto ocorre para todos os exemplos mostrados. Nos átomos de cloro ainda ocorre regiões em que o potencial é positivo, indicando a alta afinidade eletrônica dos mesmos.



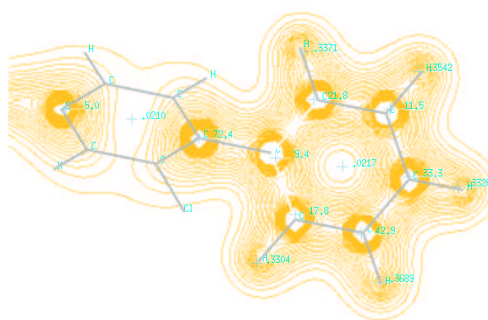
(a) Molécula



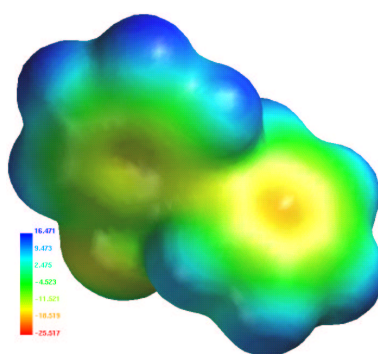
(b) LUMO



(c) HOMO



(d) Densidade de elétrons



(e) Potencial eletrostático

Figura 3.21: Propriedades eletrônicas da molécula 2-CB.

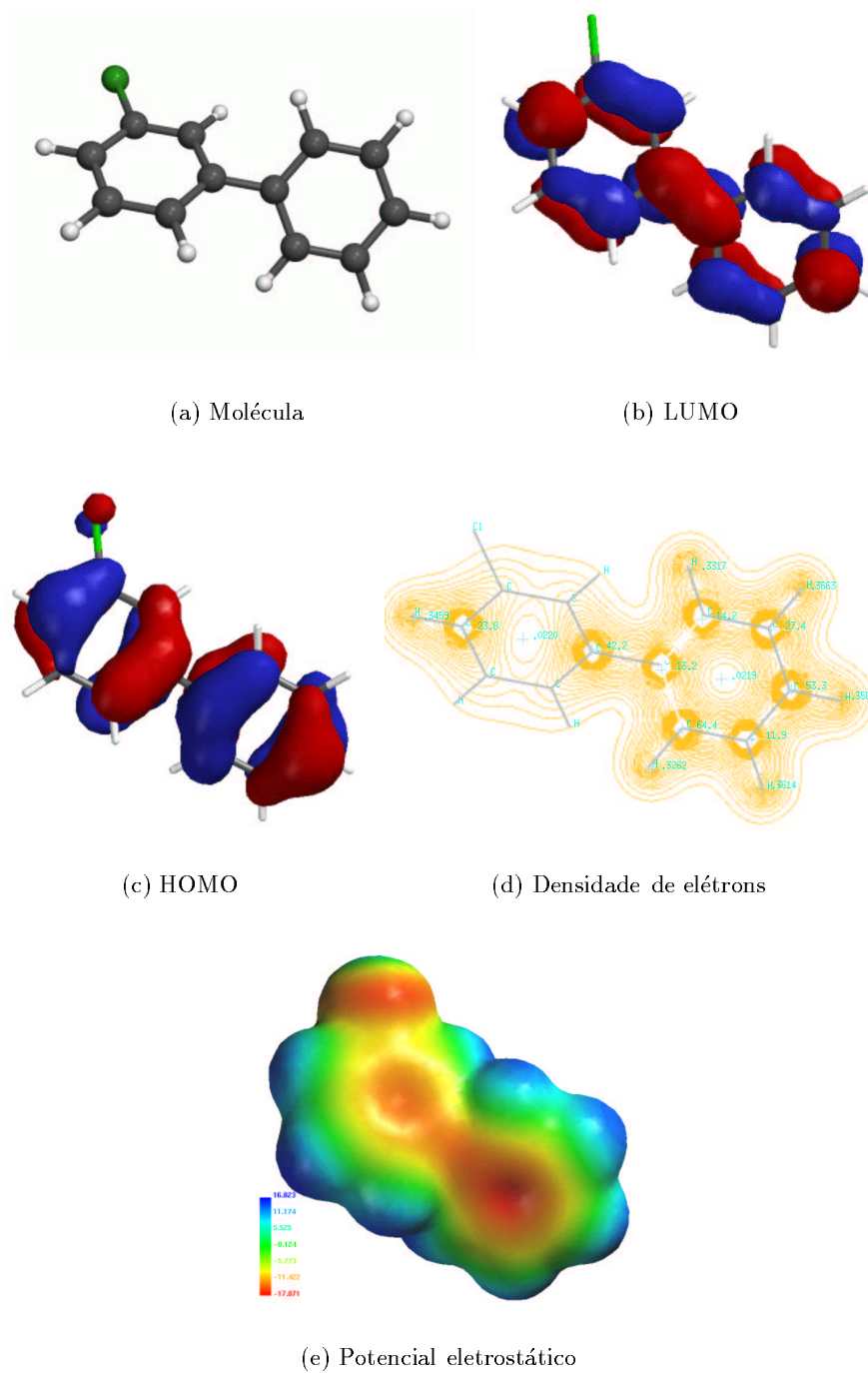


Figura 3.22: Propriedades eletrônicas da molécula 3-CB.

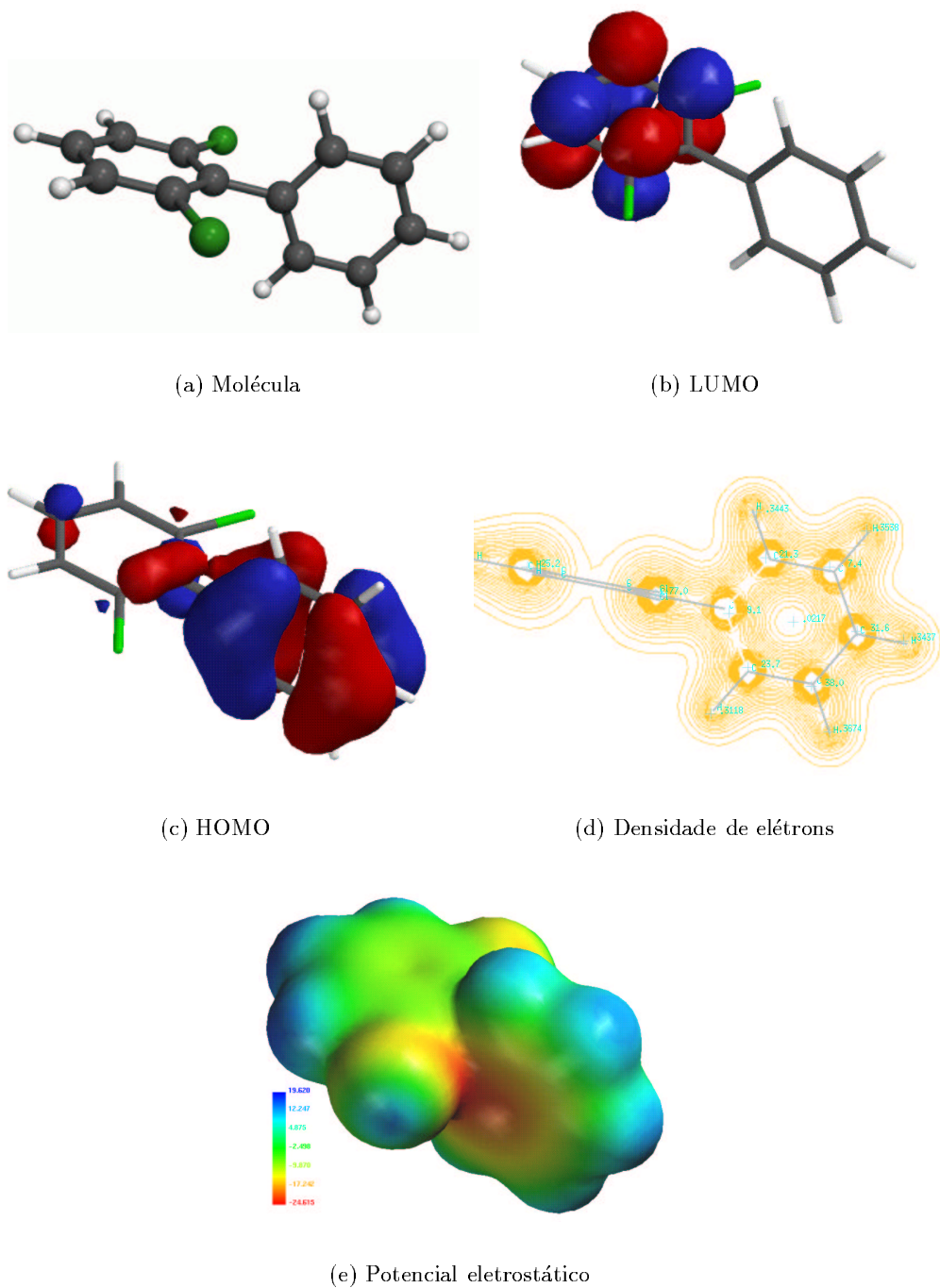


Figura 3.23: Propriedades eletrônicas da molécula 2,6-CB.

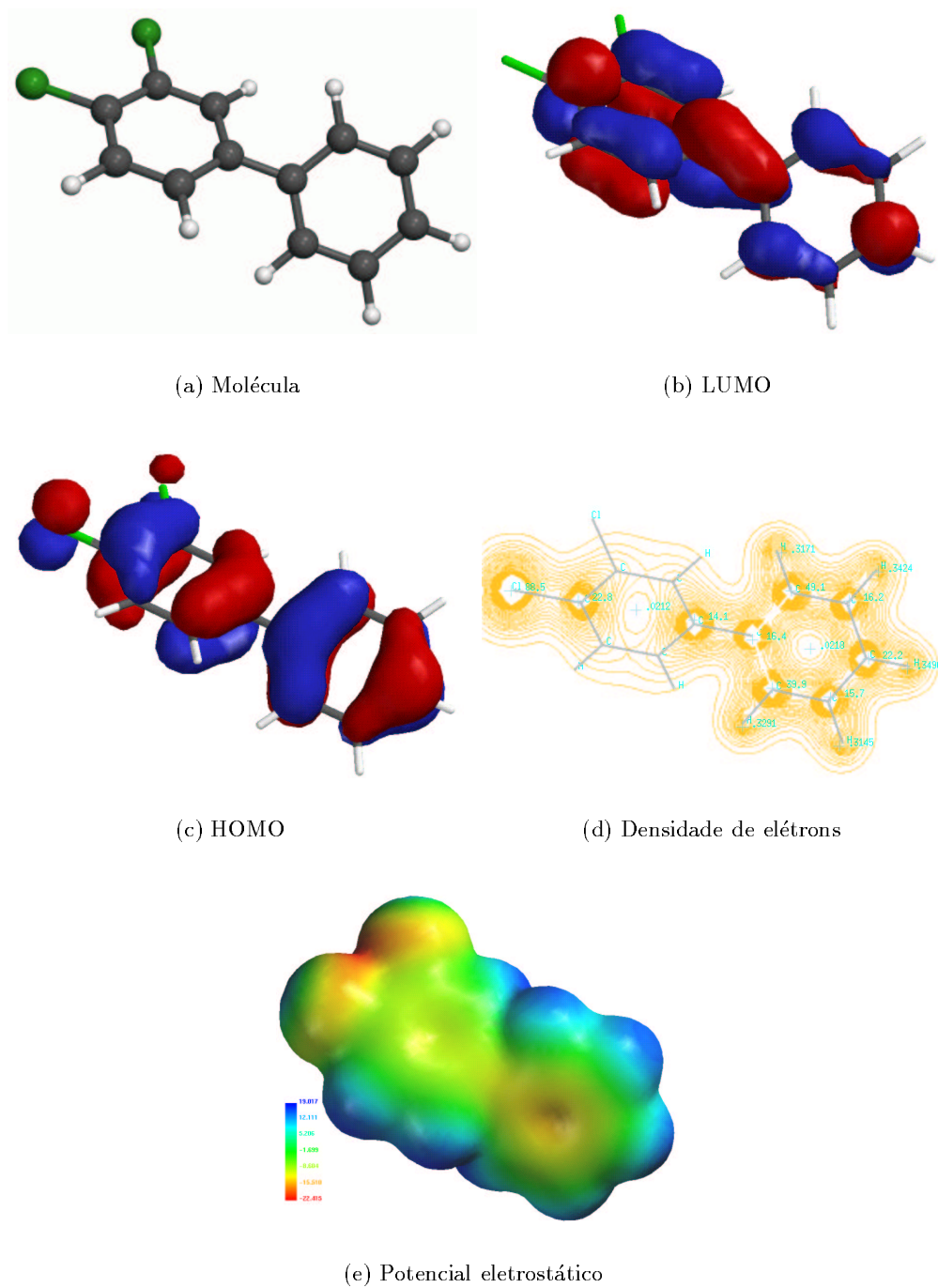


Figura 3.24: Propriedades eletrônicas da molécula 3,4-CB.

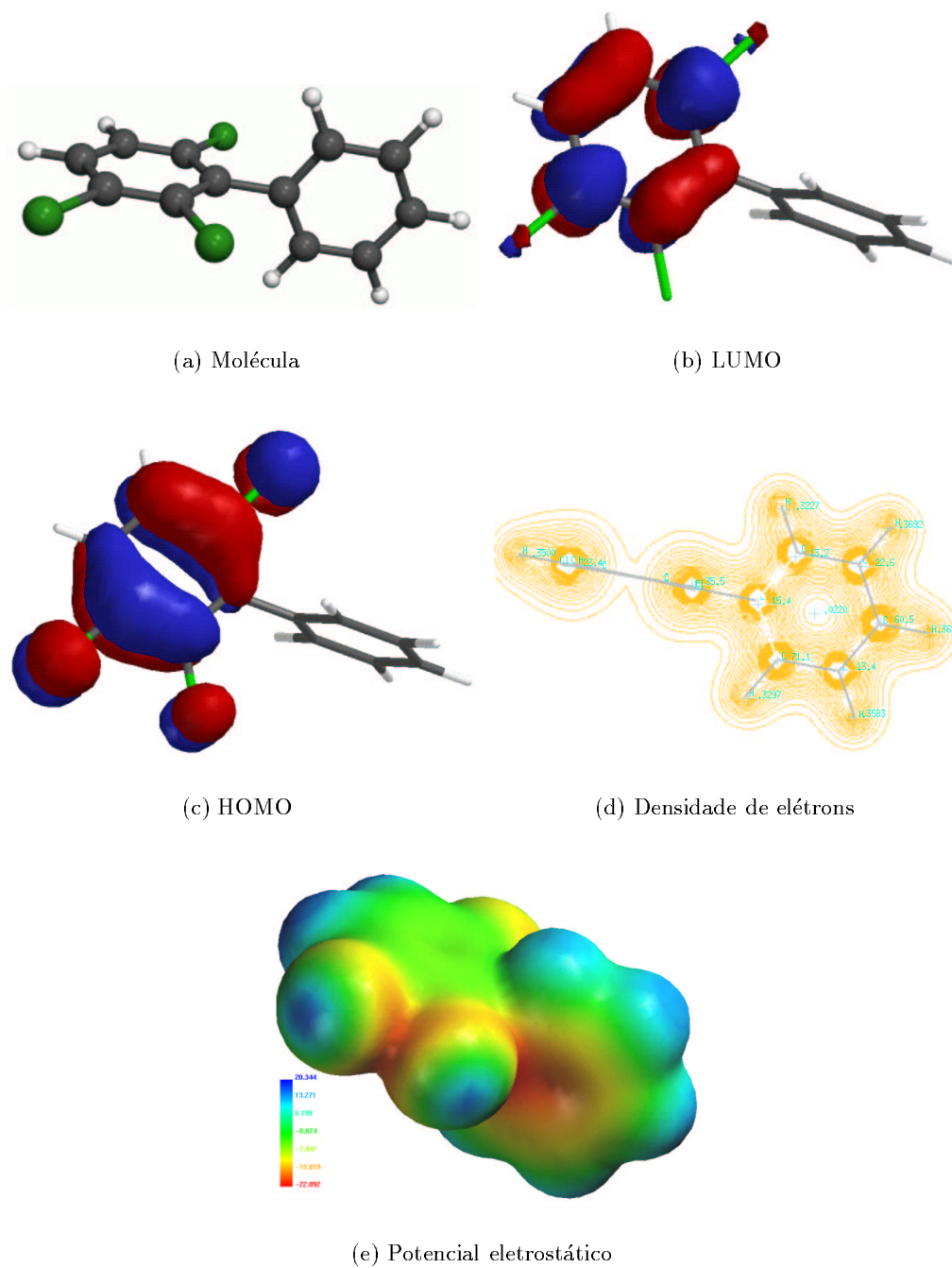


Figura 3.25: Propriedades eletrônicas da molécula 2,3,6-CB.

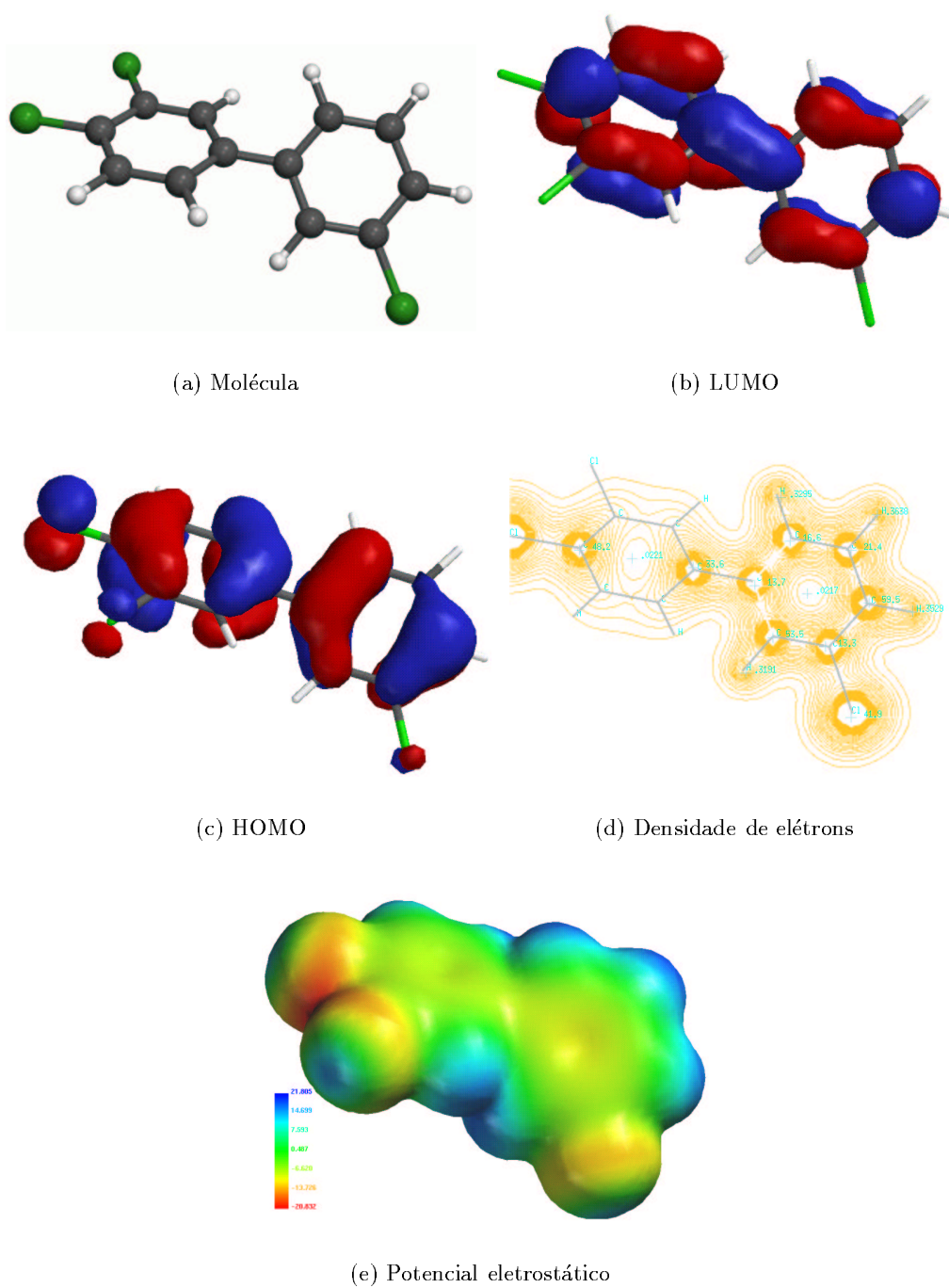


Figura 3.26: Propriedades eletrônicas da molécula 3,3',4-CB.

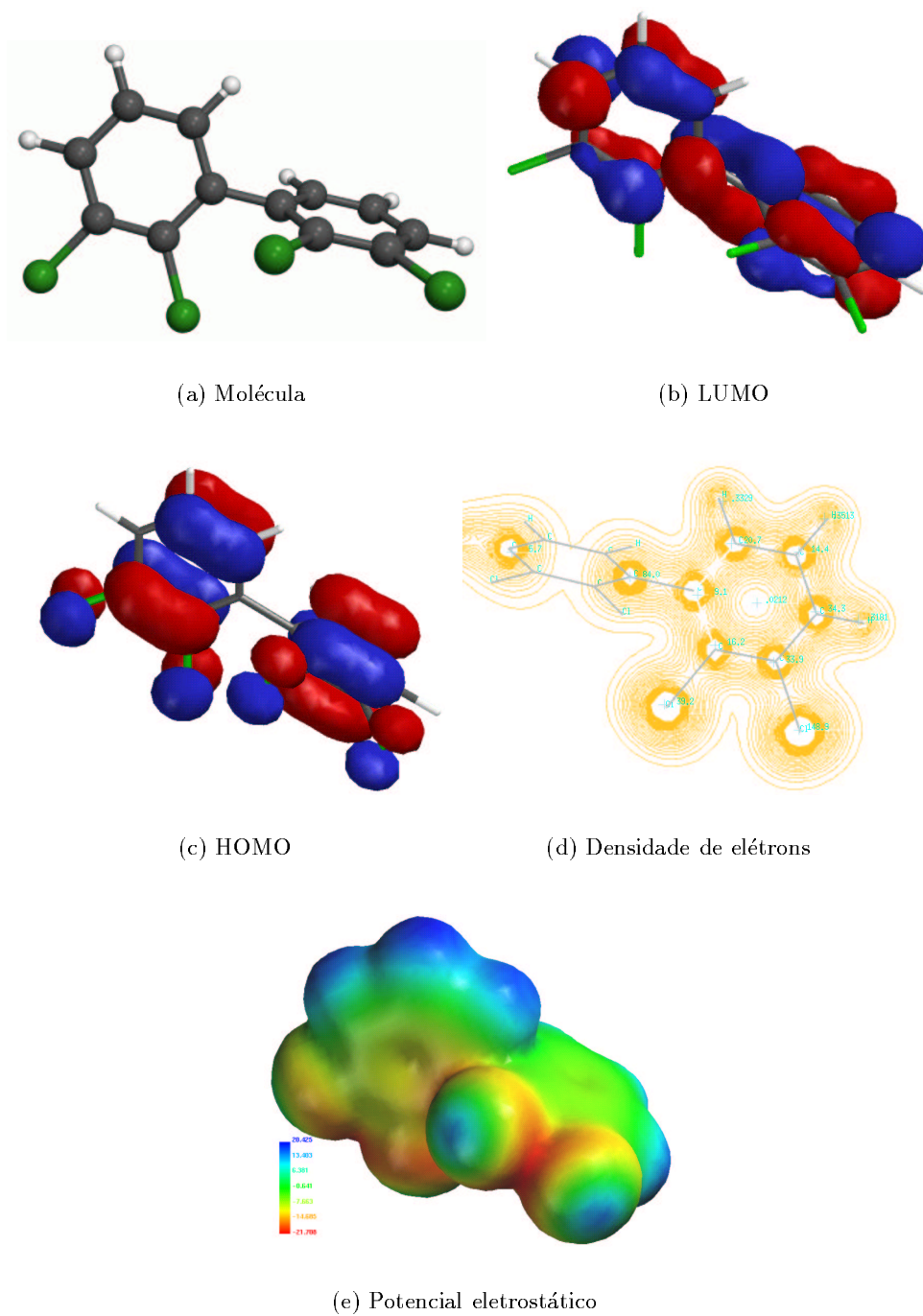
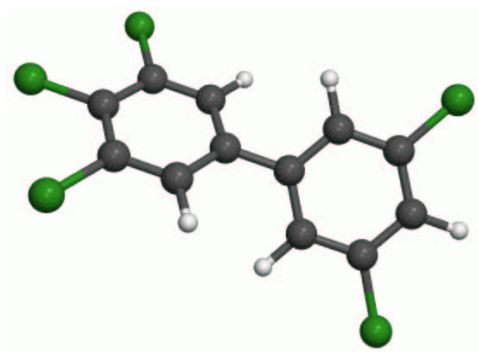
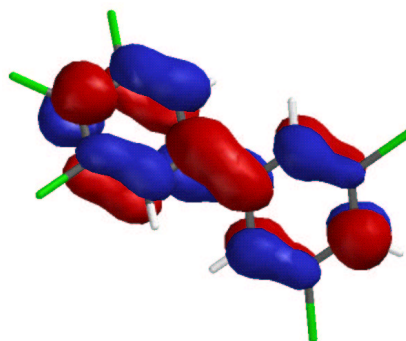


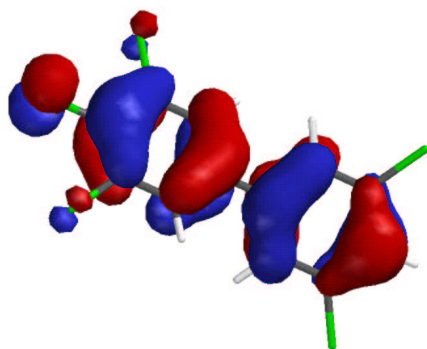
Figura 3.27: Propriedades eletrônicas da molécula 2,2',3,3'-CB.



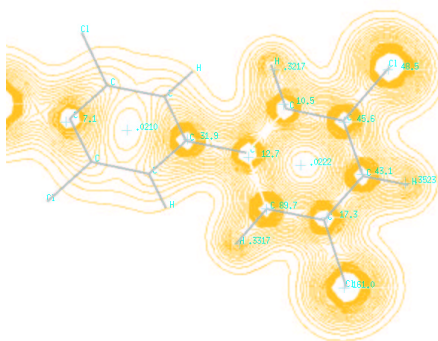
(a) Molécula



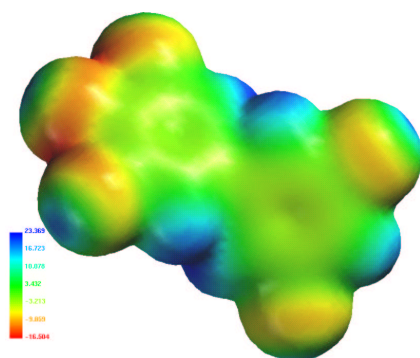
(b) LUMO



(c) HOMO

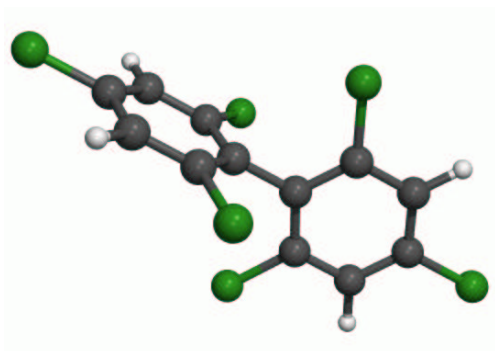


(d) Densidade de elétrons

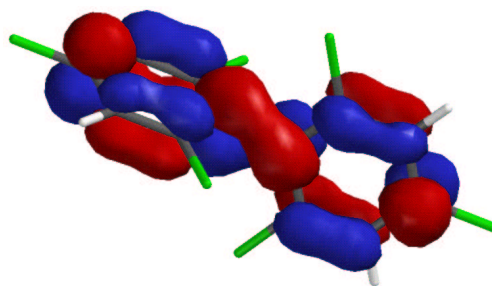


(e) Potencial eletrostático

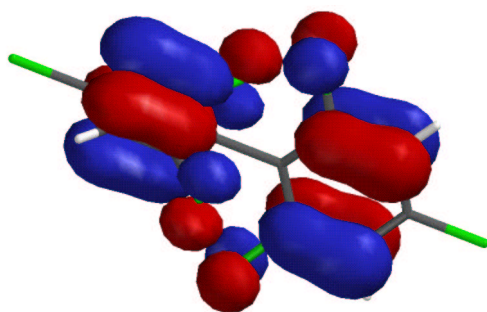
Figura 3.28: Propriedades eletrônicas da molécula 3,3',4,5,5'-CB.



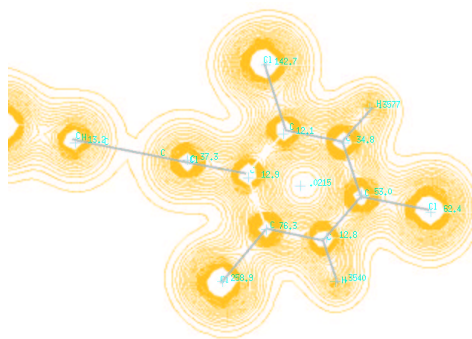
(a) Molécula



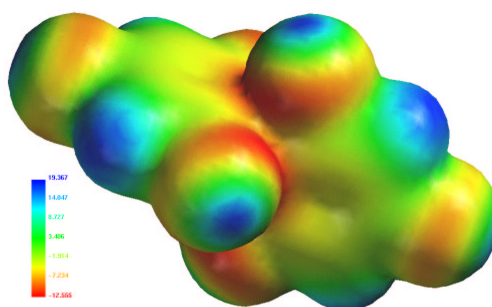
(b) LUMO



(c) HOMO



(d) Densidade de elétrons



(e) Potencial eletrostático

Figura 3.29: Propriedades eletrônicas da molécula 2,2',4,4',6,6'-CB.

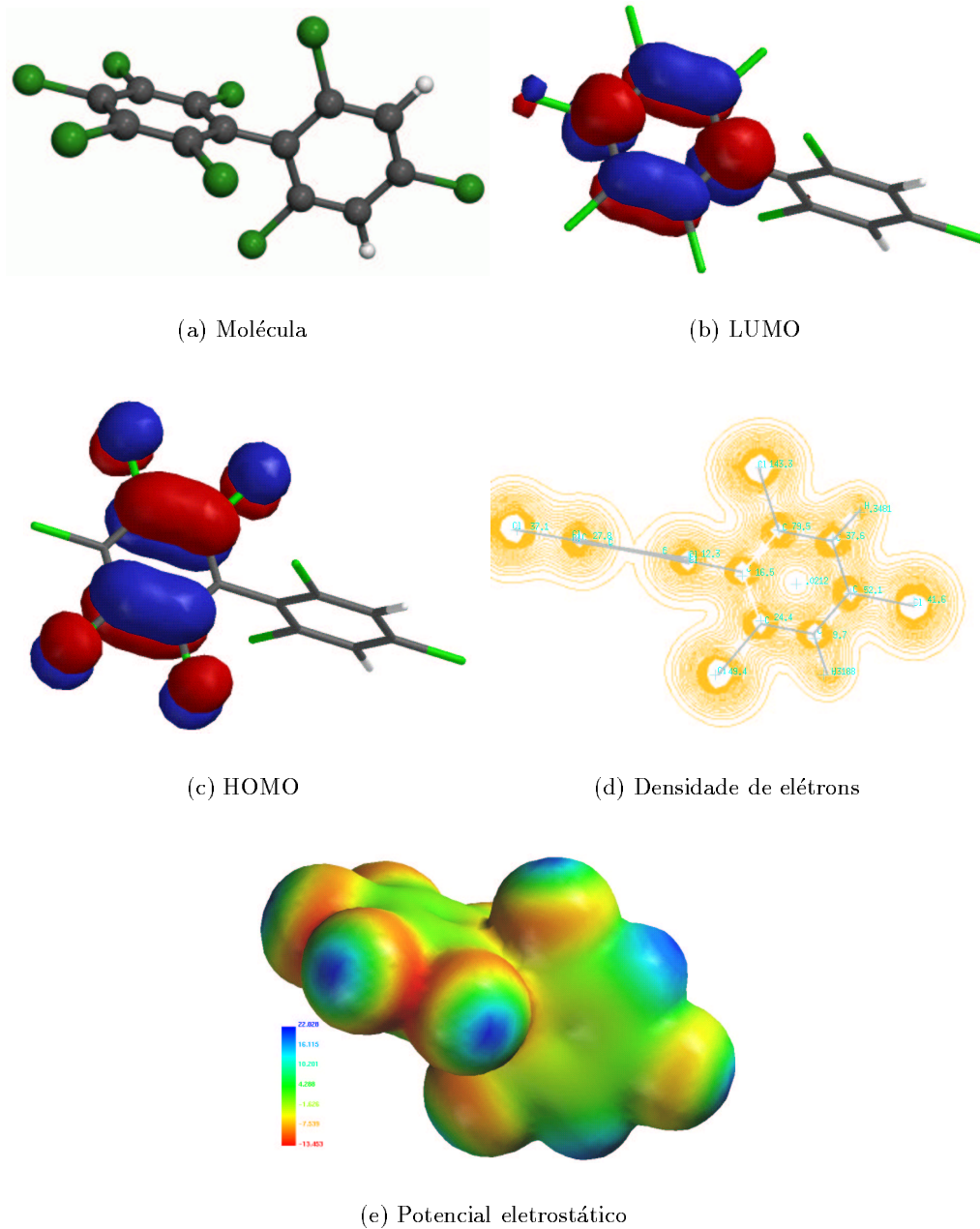
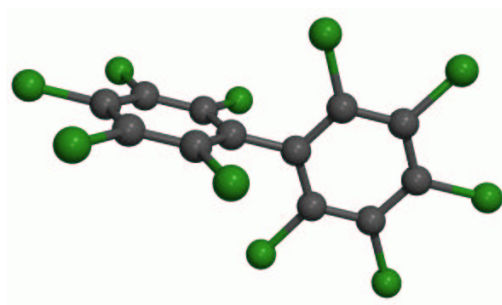
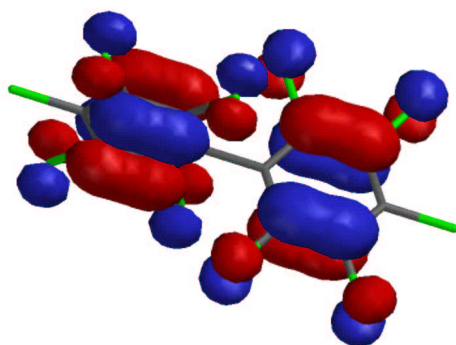
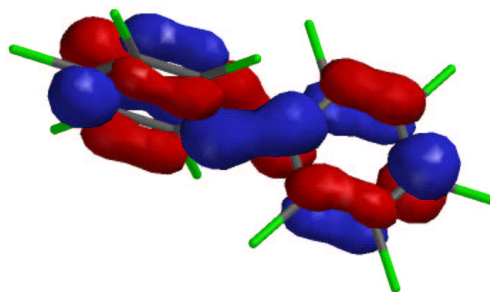


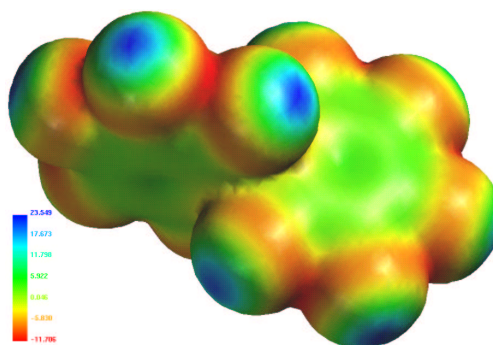
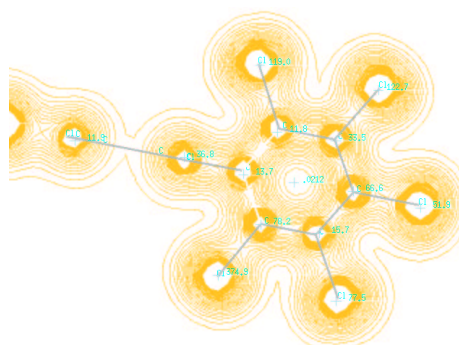
Figura 3.30: Propriedades eletrônicas da molécula 2,2',3,4,4',5,6,6'-CB.



(b) LUMO



(d) Densidade de elétrons



(e) Potencial eletrostático

Figura 3.31: Propriedades eletrônicas da molécula 2,2',3,3',4,4',5,5',6,6'-CB.

3.10 Conclusão

O objetivo inicial deste trabalho foi atingido, ou seja, obter relações matemáticas das propriedades e atividades propostas, e fazer previsões das mesmas. Estas previsões são importantes, uma vez que para muitas moléculas estes valores não são encontrados na literatura, e estudos sobre a dinâmica dos PCBs no meio ambiente são extremamente importantes e necessários.

Os resultados obtidos para IC_{50} e EC_{50} são coerentes uma vez que contém informações que podem ser correlacionadas com o meio onde o PCB interage. E as propriedades eletrônicas que foram utilizadas na modelagem de ambos (energia de homo e dureza) contém aproximadamente a mesma informação.

Os índices topológicos foram importantes na obtenção dos modelos.

Observamos que nenhuma propriedade apresenta dependência com um único tipo de descritor (eletrônico, topológico ou termodinâmico). Os modelos obtidos como uma combinação de classes diferentes de descritores mostraram sua potencialidade visto que nos possibilita ver quais parâmetros influenciam e de que maneira, pois são considerados simultaneamente os efeitos estéricos, eletrônicos e topológicos. Além de permitir uma melhor previsão destas propriedades e atividades para as demais moléculas.

Apesar de não termos indícios claros sobre a formação de pontes de hidrogênio entre os PCBs e a água, e os PCBs e o octanol, esta hipótese deve ser investigada. Sugerimos para um trabalho futuro um estudo utilizando simulação computacional com método de monte-carlo, o qual permite simular a estruturação de moléculas de água ou octanol em relação ao PCB. Este estudo permitirá uma análise sobre as possíveis pontes de hidrogênio, e como os átomos das moléculas estão dispostos em relação ao meio. Nossa sugestão é que estes cálculos sejam feitos para congêneros com diferentes números de cloro, e com algumas posições diferentes.

As medidas correlacionadas com a interação dos PCBs com a água foram as que apresentaram maiores erros. Isto pode estar correlacionado com a dificuldade nas medidas experimentais, uma vez que os PCBs são lipofílicos.

Capítulo 4

Derivados de 4M e 5M

Como foi mencionado na introdução, o objetivo das drogas citotóxicas é causar alguma lesão ao DNA da célula maligna, direta ou indiretamente. A lesão direta, depende do tipo de interação droga-DNA e da reatividade química intrínseca da droga. A lesão indireta, geralmente é resultado de interações entre a droga e *agentes* envolvidos no metabolismo do DNA, como a topoisomerase II. Os principais danos sobre o DNA que podem ocorrer em processos citotóxicos são: interações entre pares de base, ligações covalentes dos compostos como um todo em sítios de intercalação ou sítios externos, geração de quebra de cadeia através da interação com a topoisomerase II.¹⁰³

Muitas drogas agem de maneira complexa, por exemplo, os derivados de elipticina, moléculas muito estudadas e utilizadas no tratamento de câncer humano, podem se ligar ao DNA através de intercalação, geração de espécies oxidantes e interferência na atividade catalítica da topoisomerase II.^{104,105} Mas, a contribuição de cada mecanismo para a citotoxicidade das drogas não está claro.

O objetivo deste trabalho é contribuir com informações que ajudem no entendimento da citotoxicidade dos derivados de 4-aminometiltioxantenona (4M) e 5-aminometilbenzotiopiranoindazol (5M), e obter modelos que possam prever esta citotoxicidade para moléculas similares.

As moléculas estudadas foram sintetizadas e testadas por Perni et al.¹⁰⁶ Na tabela 4.1 está a lista com os dois grupos de moléculas. O esquema destas moléculas se encontra na figura 4.1. As moléculas foram construídas utilizando-se o programa Molden,⁵⁷ a partir de dados cristalográficos de uma molécula de hycantone (fig. 4.2),¹⁰⁷ que é a molécula 1 na tabela 4.1. Como existem muitos diedros para se fazer a análise conformacional, utilizando-

se esta estrutura cristalográfica, a análise conformacional foi feita apenas nos diedros dos substituintes R^4 . As estruturas foram pré-otimizadas com o programa Mopac 6.⁵⁹ A seguir foi feita uma análise conformacional sistemática, utilizando o método semi-empírico AM1, do programa Spartan 5.⁶⁵ A partir da conformação de menor energia, foram calculadas as propriedades utilizadas na análise. Os descritores calculados foram os mesmos utilizados no estudo dos PCBs, entretanto os índices topológicos foram os calculados pelo programa WHIM-3D.⁶⁶ Os métodos quimiométricos utilizados foram: PCA e PLS.

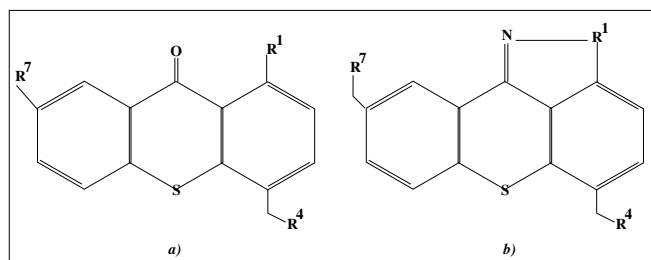


Figura 4.1: Esquema das drogas estudadas: a) moléculas de 1 até 44 da tabela 4.1; b) moléculas de 45 até 55.

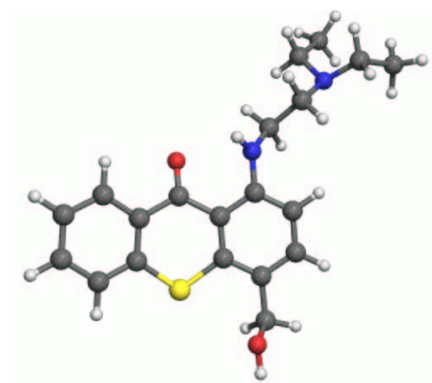


Figura 4.2: Geometria da molécula de hycantone, obtida por cristalografia.

Tabela 4.1: Tabela com as moléculas estudadas. * moléculas consideradas inativas.

Comp.	R ¹	R ⁴	R ¹
1	H	OH	NHCH ₂ CH ₂ NEt ₂
2	H	NHCH ₃	NHCH ₂ CH ₂ NEt ₂
3	H	NHCHO	NHCH ₂ CH ₂ NEt ₂
4	H	NH ₂	NHCH ₂ CH ₂ NEt ₂
5	H	N(Et)CHO	NHCH ₂ CH ₂ NEt ₂
6	H	NHEt	NHCH ₂ CH ₂ NEt ₂
7*	H	NHC ₆ H ₅	NHCH ₂ CH ₂ NEt ₂
8	H	NHSO ₂ CH ₃	NHCH ₂ CH ₂ NEt ₂
9	H	NHSO ₂ Et	NHCH ₂ CH ₂ NEt ₂
10	H	NHSO ₂ C ₆ H ₅	NHCH ₂ CH ₂ NEt ₂
11*	H	NHSO ₂ - 4 - CH ₃ - C ₆ H ₅	NHCH ₂ CH ₂ NEt ₂
12*	H	NHSO ₂ - 4 - Cl - C ₆ H ₄	NHCH ₂ CH ₂ NEt ₂
13	H	NHSO ₂ - 3,4 - Cl - C ₆ H ₅	NHCH ₂ CH ₂ NEt ₂
14*	H	NHSO ₂ - 4 - F - C ₆ H ₅	NHCH ₂ CH ₂ NEt ₂
15	H	NHSO ₂ - 2 - F - C ₆ H ₅	NHCH ₂ CH ₂ NEt ₂
16	H	NHSO ₂ - 4 - CH ₃ O - C ₆ H ₄	NHCH ₂ CH ₂ NEt ₂
17	H	N(CH ₃)SO ₂ CH ₃	NHCH ₂ CH ₂ NEt ₂
18	H	N(Et)SO ₂ CH ₃	NHCH ₂ CH ₂ NEt ₂
19	H	N(nPr)SO ₂ CH ₃	NHCH ₂ CH ₂ NEt ₂
20	H	N(CH ₃)SO ₂ Et	NHCH ₂ CH ₂ NEt ₂
21	H	N(CH ₃)SO ₂ C ₆ H ₅	NHCH ₂ CH ₂ NEt ₂
22	H	NHCOCH ₃	NHCH ₂ CH ₂ NEt ₂
23	H	NHCOCF ₃	NHCH ₂ CH ₂ NEt ₂
24	H	NHCO ₂ C ₆ H ₅	NHCH ₂ CH ₂ NEt ₂
25	H	NHCO ₂ CH ₃	NHCH ₂ CH ₂ NEt ₂
26	H	NHPO(OEt) ₂	NHCH ₂ CH ₂ NEt ₂
27*	H	NEt ₂	NHCH ₂ CH ₂ NEt ₂
28	H	NHCH ₃	NHCH ₂ CH ₂ CH ₂ NEt ₂
29	H	NHCH ₃	NHCH ₂ CH ₂ NMe ₂
30	H	NHCH ₃	NHCH ₂ CH ₂ CH ₂ NMe ₂
31	H	NHSO ₂ CH ₃	NHCH ₂ CH ₂ NMe ₂
32	H	NHSO ₂ CH ₃	NHCH ₂ CH ₂ CH ₂ NMe ₂
33*	H	NHCH ₃	NHCH ₂ CH ₂ CH ₂ NEt ₂
34*	Br	NHCH ₃	NHCH ₂ CH ₂ NEt ₂
35*	Br	NH ₂	NHCH ₂ CH ₂ NEt ₂
36*	Br	NHSO ₂ CH ₃	NHCH ₂ CH ₂ NEt ₂
37	OCH ₃	NHCH ₃	NHCH ₂ CH ₂ NEt ₂
38	OCH ₃	NHCHO	NHCH ₂ CH ₂ NEt ₂
39	OCH ₃	NHSO ₂ CH ₃	NHCH ₂ CH ₂ NEt ₂
40	OCH ₃	NHCO ₂ CH ₃	NHCH ₂ CH ₂ NEt ₂
41	OCH ₃	NHSO ₂ C ₆ H ₅	NHCH ₂ CH ₂ NEt ₂
42	OH	NHCH ₃	NHCH ₂ CH ₂ NEt ₂
43	OH	NHSO ₂ CH ₃	NHCH ₂ CH ₂ NEt ₂
44	OH	NHCO ₂ CH ₃	NHCH ₂ CH ₂ NEt ₂
45	H	NHCH ₃	NHCH ₂ CH ₂ NEt ₂
46	H	NHCHO	NHCH ₂ CH ₂ NEt ₂
47	H	NH ₂	NHCH ₂ CH ₂ NEt ₂
48	H	NHSO ₂ CH ₃	NHCH ₂ CH ₂ NEt ₂
49	H	NHCO ₂ CH ₃	NHCH ₂ CH ₂ NEt ₂
50	OCH ₃	NHCHO	NHCH ₂ CH ₂ NEt ₂
51	OCH ₃	NH ₂	NHCH ₂ CH ₂ NEt ₂
52	OCH ₃	NHSO ₂ CH ₃	NHCH ₂ CH ₂ NEt ₂
53	OCH ₃	NHCO ₂ CH ₃	NHCH ₂ CH ₂ NEt ₂
54	OH	NHSO ₂ CH ₃	NHCH ₂ CH ₂ NEt ₂
55	OH	NHCO ₂ CH ₃	NHCH ₂ CH ₂ NEt ₂

4.1 Intercalação DNA-droga

Como não há nenhuma estrutura experimental nem calculada de complexos de DNA com tioxantenona ou benzotilpiranoindazol, nós apresentamos um modelo de intercalação de um tioxantenona em um fragmento de DNA. Na figura 4.3 está a estrutura cristalográfica de 5' – d(CpGpTpApCpG) – 3' complexado com duas moléculas de *N,N* – bis(aminopropil) – 2 – antraquinonesulfonamida¹⁰⁸ e também a estrutura de d(CGGCCG) complexado com duas moléculas de adriamicina,¹⁰⁹ mostrando a intercalação destas drogas com o DNA. Para uma visão mais clara, também mostramos as moléculas sem o DNA, mantendo a distância e posição das mesmas, e o DNA sem as moléculas. A adriamicina contém em sua estrutura, três anéis fundidos com um grupo substituinte flexível. Nossas drogas têm uma estrutura básica que lembra a da adriamicina e da sulfonamida. Portanto, estes complexos foram utilizados como base para a intercalação da molécula 1 do nosso trabalho (fig. 4.2) no DNA. Manualmente substituímos a molécula de sulfonamida pela molécula 1, procurando uma posição semelhante.

Nota-se que as bases do DNA estão em posição aproximadamente paralela à droga. A molécula da droga deve ter uma parte plana que possibilite a interação com as bases do DNA, e também uma parte flexível que permita uma maior interação com o DNA. A estrutura básica não pode ser muito pequena, pois são necessárias interações da droga com as bases do DNA.

Na figura 4.5 são mostradas as distâncias entre os átomos da droga e os átomos do DNA. Observamos que a menor distância (2.59Å) ocorre entre um átomo de oxigênio da molécula e um hidrogênio do DNA, e a maioria das distâncias variam de 3.0 - 4.0Å. Esta faixa de distância ocorre entre as bases do DNA e a estrutura básica da droga, assim como com o seu substituinte. Estas distâncias possibilitam uma forte interação $\pi - \pi$ entre as bases aromáticas do DNA e os anéis da droga, e interações de van der Waals entre a droga e o DNA.

Na figura 4.6 intercalamos a molécula 1 com o DNA. Procuramos deixar o substituinte R¹ mais para fora e os anéis ficaram no centro entre as bases de DNA.

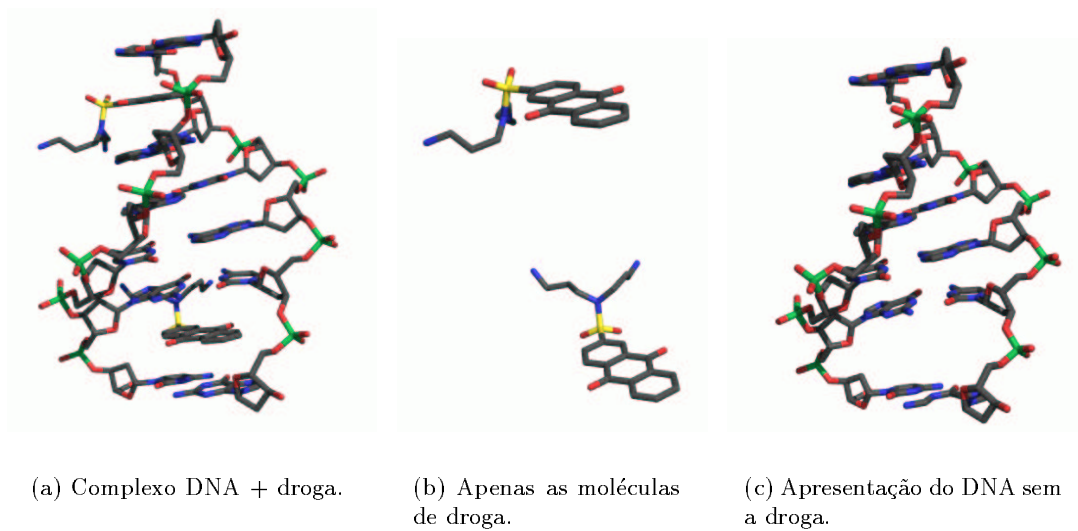


Figura 4.3: Estrutura cristalográfica de 5' – d(CpGpTpApCpG) – 3' complexado com *N, N* – bis(aminopropil) – 2 – anthraquinonesulfonamida.

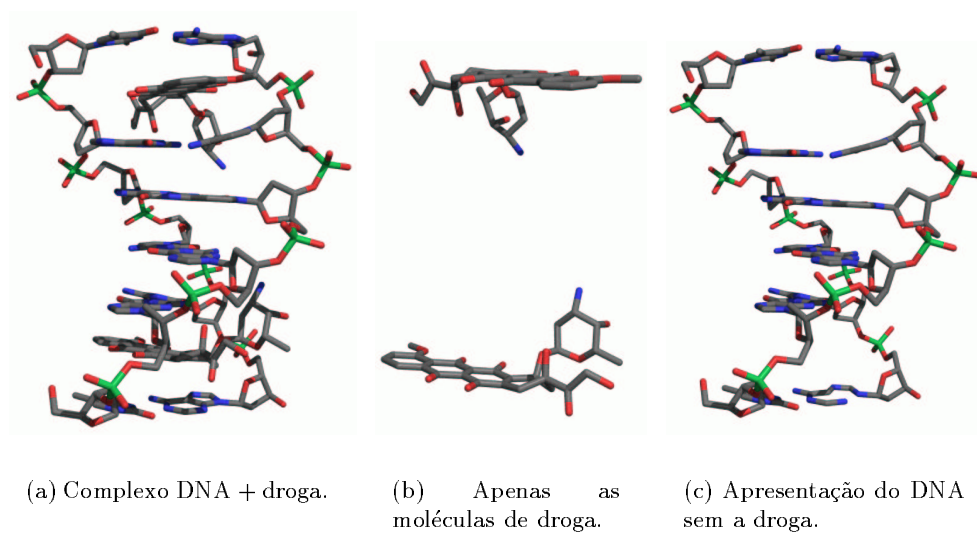


Figura 4.4: Estrutura cristalográfica de d(CGGCCG) complexado com adriamicina.

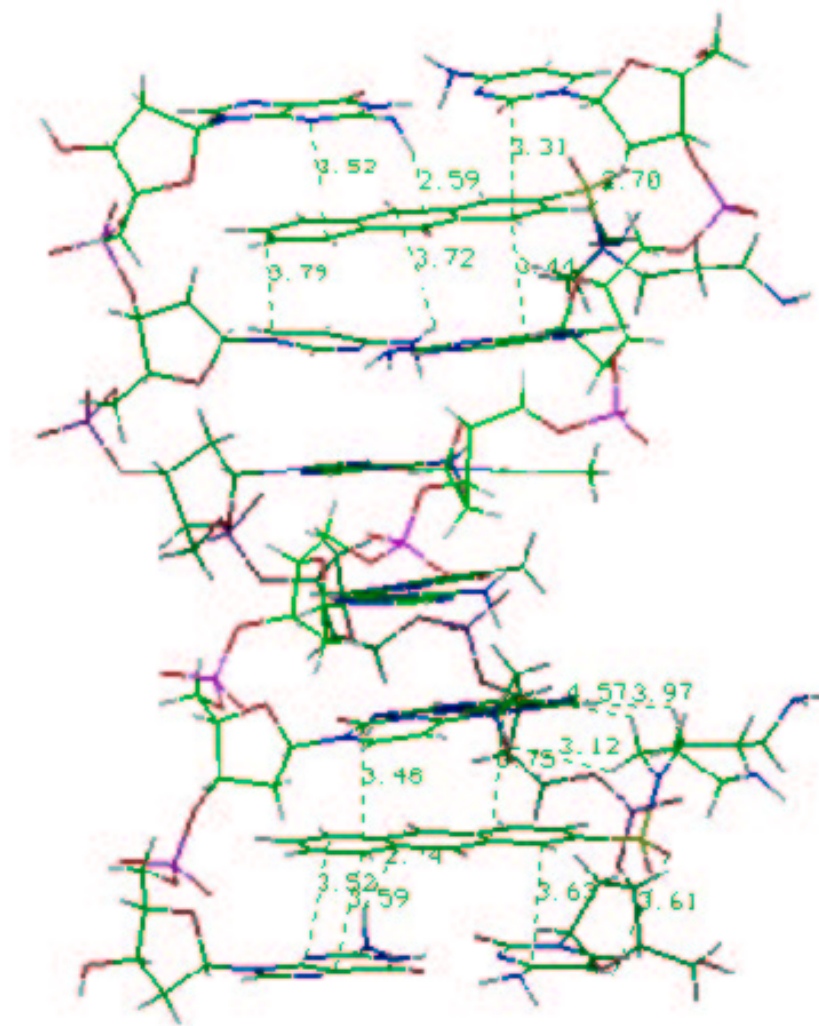


Figura 4.5: Distâncias entre alguns átomos da molécula de sulfonamida intercalada e o DNA.

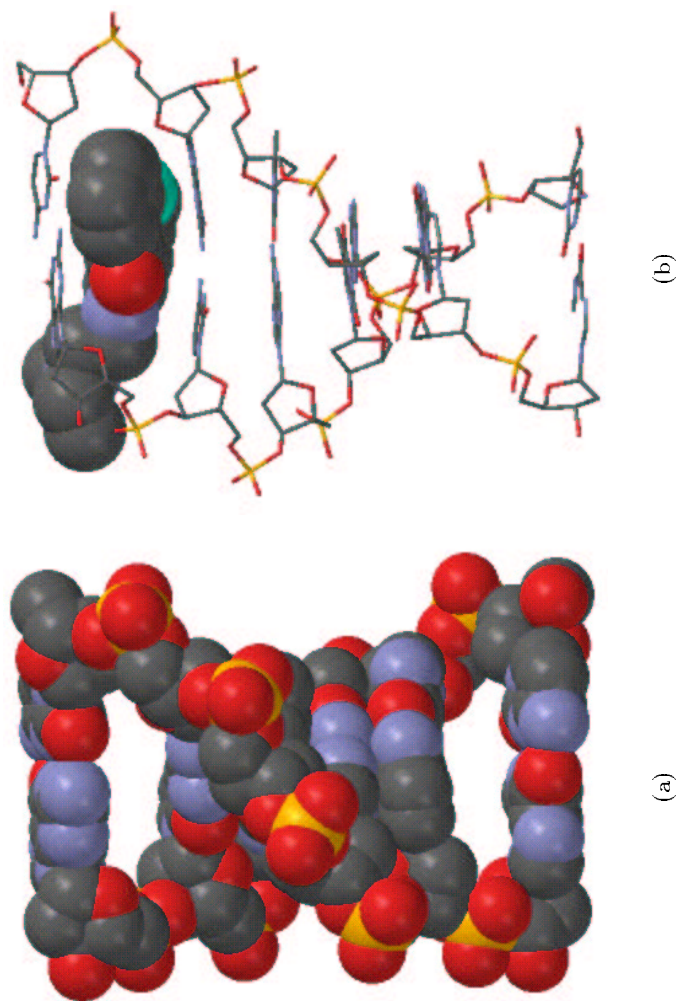


Figura 4.6: Proposta da intercalação da molécula 1 com o DNA.

4.2 Propriedades Eletrônicas

O estudo do potencial eletrostático tem proporcionado bons resultados no estudo de moléculas orgânicas,^{110,111} e como é conhecido para outras drogas, a interação droga-DNA é fortemente dependente do campo eletrostático gerado no processo.¹¹² A seguir apresentamos os orbitais de HOMO e LUMO, e a superfície de potencial eletrostático de algumas moléculas consideradas ativas e inativas, segundo a classificação de Perni, e que tenham estruturas bem similares.

Observamos que quase não existe diferença entre os orbitais de HOMO e LUMO das moléculas ativas e inativas das figuras 4.7, 4.8 e 4.9. Entretanto pode-se observar que o átomo de bromo na posição R^7 , aumenta o volume nesta posição em 6 vezes, considerando-se que o raio de van der Waals do hidrogênio é ≈ 1.0 e do bromo ≈ 1.85 . Apesar de ser um fator importante, o aumento do volume nesta posição não deve ser aceito como responsável pelo fato da molécula ser inativa, visto que outras moléculas com substituintes na posição R^7 foram consideradas ativas, como é o caso das moléculas 37, 38, 39, 40, 41, 50, 51, 52 e 53. Estas moléculas apresentam o substituinte R^7 (OCH_3) com volume maior que o volume do bromo.

Nas figuras 4.7, 4.8 e 4.9 observa-se que a diferença entre as superfícies de potencial eletrostático destas moléculas é mínima, mas pode-se observar a afinidade eletrônica do átomo de bromo, visto que o potencial eletrostático na região onde ele se encontra está menos positivo.

Na figura 4.10 apresentamos os orbitais de LUMO e HOMO e a superfície de potencial eletrostático para duas drogas inativas. Os orbitais de LUMO se localizam no meio do substituinte (R^4). Para estas moléculas o fato de serem inativas pode estar associado ao impedimento estérico, visto que a posição do substituinte R^4 está mais próximo da estrutura básica da molécula, o que poderia impedir a penetração da mesma entre as bases do DNA. Estes fatos indicam que é necessário que a base da molécula não tenha seu volume muito aumentado por substituintes na posição R^7 , ou que os substituintes da posição R^4 não fiquem muito próximos da estrutura básica.

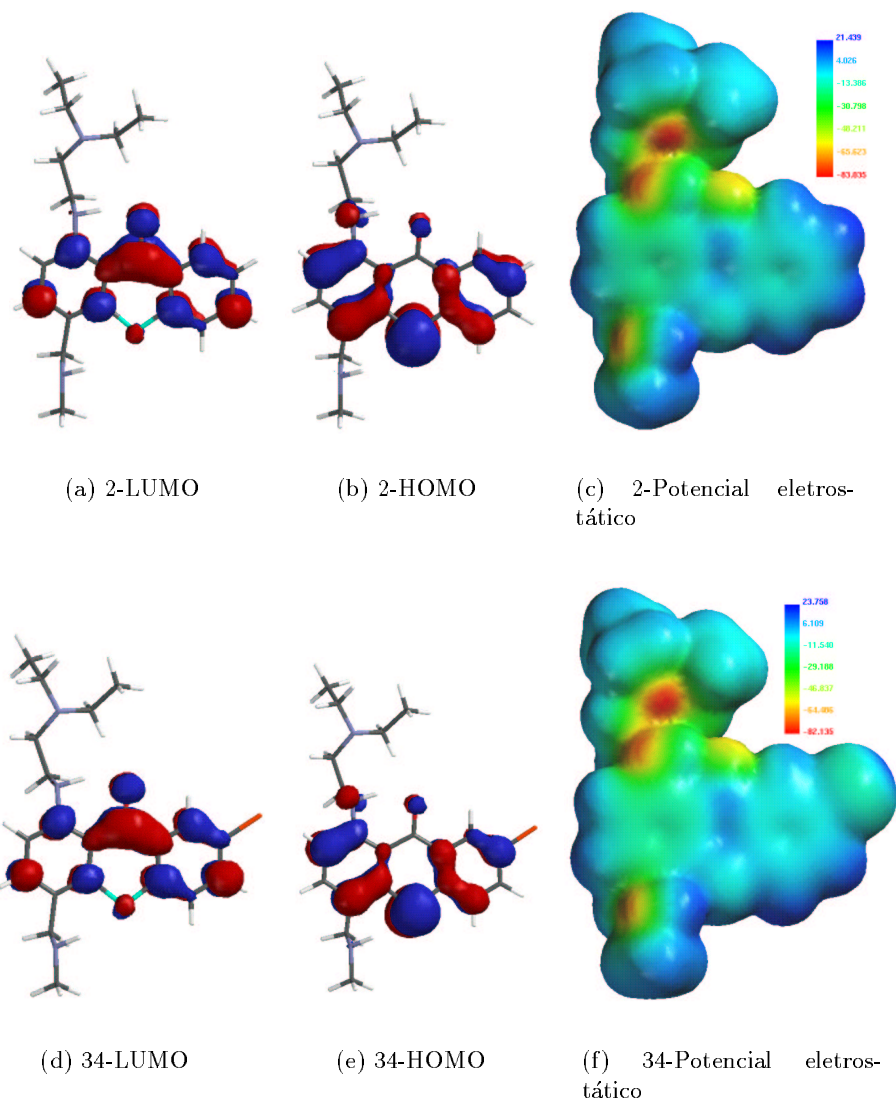


Figura 4.7: Orbitais moleculares de HOMO e LUMO, e superfície se potencial eletrostático das molécula 2 (ativa) e 34 (inativa).

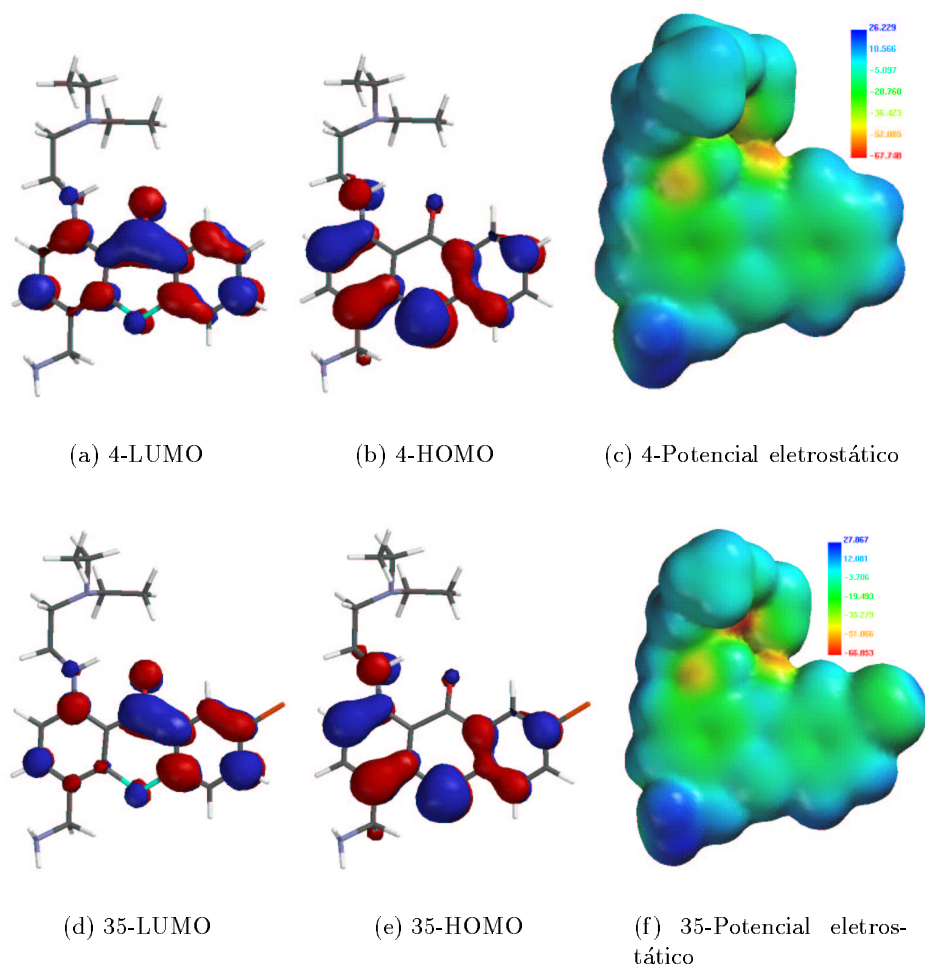


Figura 4.8: Orbitais moleculares de HOMO e LUMO, e superfície se potencial eletrostático das molécula 4 (ativa) e 35 (inativa).

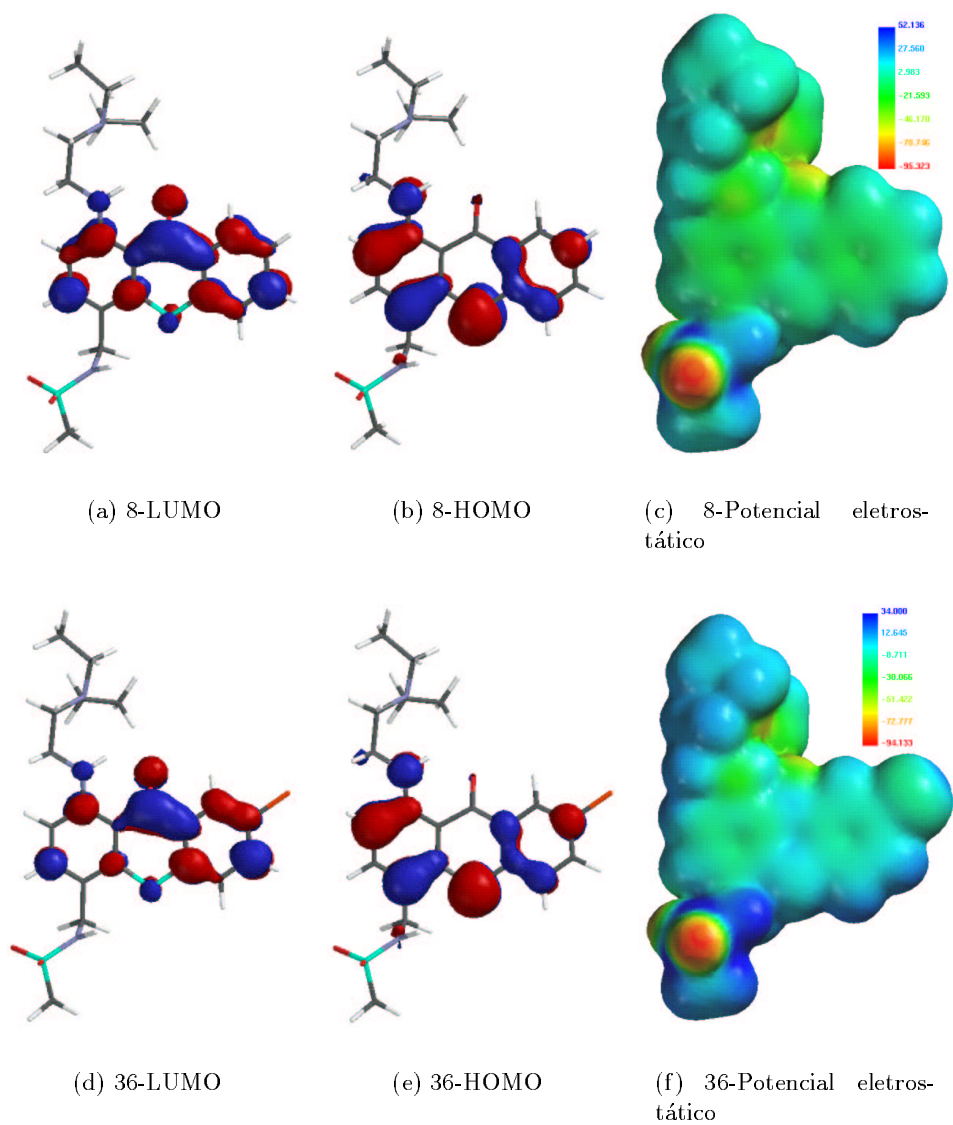


Figura 4.9: Orbitais moleculares de HOMO e LUMO, e superfície de potencial eletrostático das molécula 8 (ativa) e 36 (inativa).

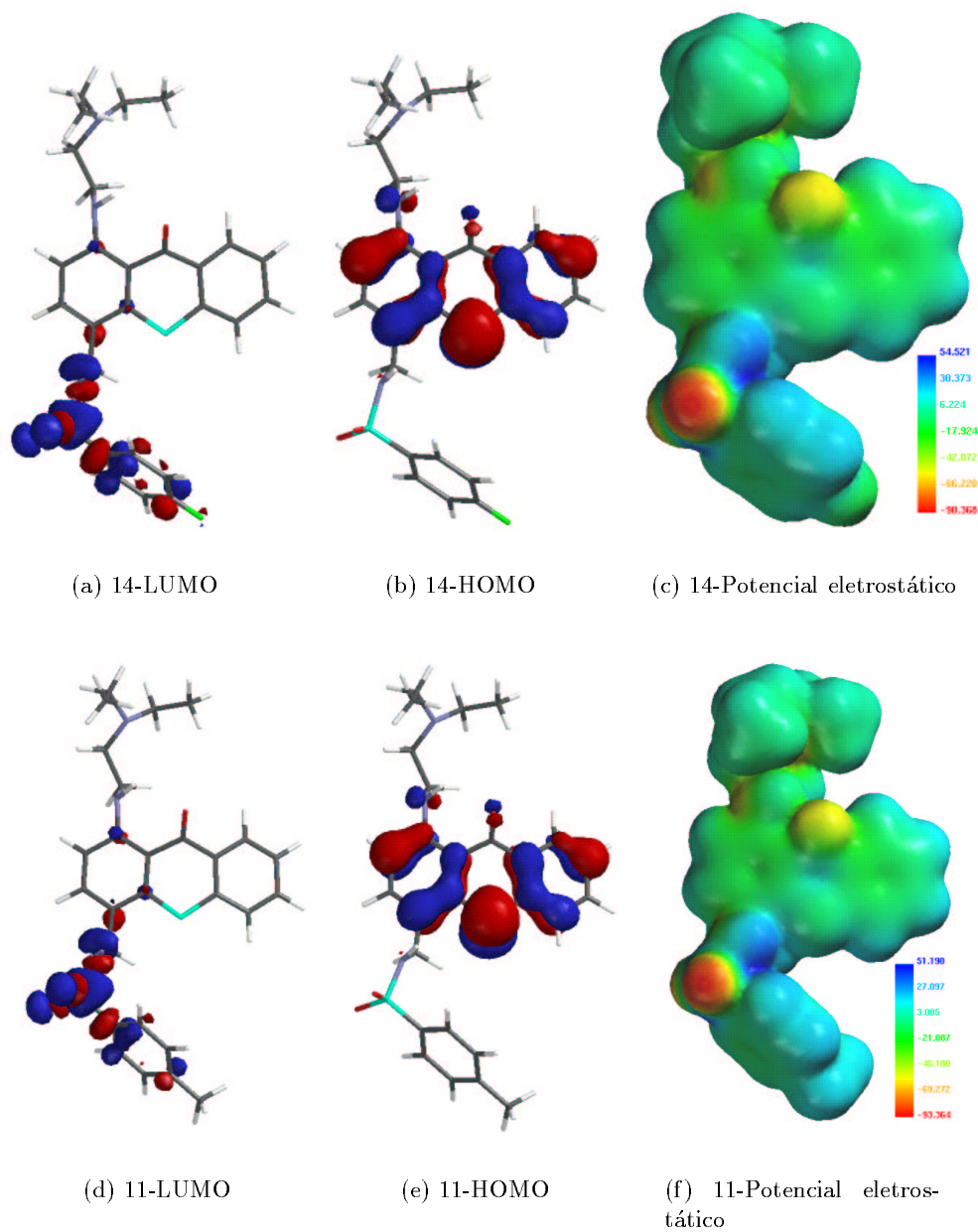


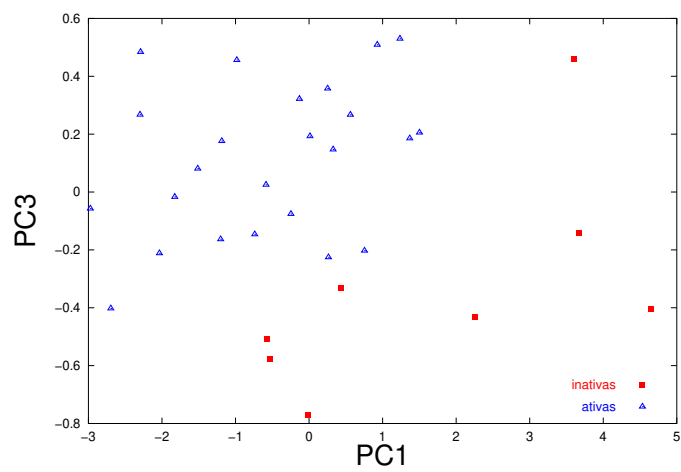
Figura 4.10: Orbitais moleculares de HOMO e LUMO, e superfície se potencial eletrostático de duas moléculas inativas (11 e 14).

4.3 Análise por Componentes Principais

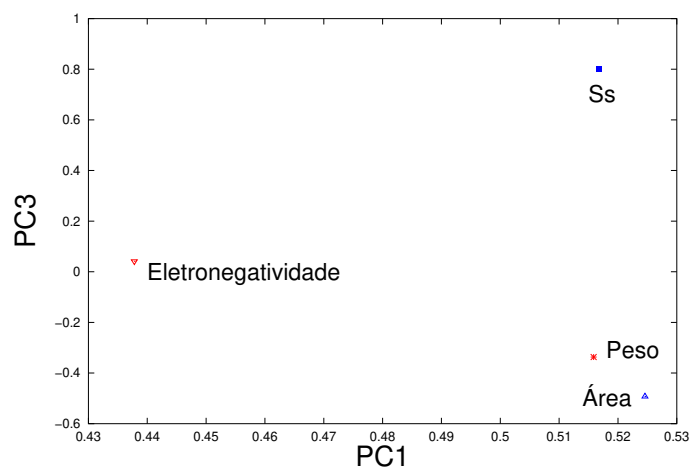
Análise por componentes principais foi utilizada, a fim de determinar quais propriedades moleculares (variáveis) poderiam estar influenciando no comportamento ativo ou inativo das drogas, e como seria esta influência. As variáveis que contribuíram para uma melhor separação foram: eletronegatividade, área, peso e a soma dos índices eletrotopológicos.

A primeira e a terceira PCs foram capazes de separar as drogas ativas das drogas inativas. Na figura 4.11a observa-se esta separação. A primeira PC acumula uma variância de 82.93%, e a terceira PC acumula uma variância de 3.04%, totalizando 85.97% da variância total. Analisando o gráfico dos *loadings* (fig. 4.11b) juntamente com o gráfico dos *scores* (fig. 4.11a), podemos notar que valores muito baixo de eletronegatividade, são relacionados com a maior atividade da droga. A área e o peso contribuem para o comportamento inativo das drogas. Moléculas com área muito grande tendem a ser inativas. Como estamos trabalhando com um grupo de moléculas que apresentam a mesma estrutura básica, um aumento na área destas moléculas está relacionado com o aumento no tamanho dos substituintes, o que poderia causar algum impedimento estérico.

Comparando esta análise com a estrutura das moléculas inativas (tab.4.1), podemos concluir que no caso das drogas 11, 12 e 13, o tamanho da molécula pode estar sendo determinante para o seu comportamento inativo. E para as drogas 33, 34, 35 e 36, a alta eletronegatividade do átomo de bromo poderia ser a responsável pelo comportamento inativo destas drogas.



(a)



(b)

Figura 4.11: Gráfico dos *scores* mostrando a separação entre drogas ativas e inativas. b) Gráfico dos *loadings* na primeira e terceira PCs.

4.4 Atividade EC₅₀

O valor de EC₅₀ é a concentração do agente testado que causou uma redução de 50% na fluorescência do complexo *calf thymus DNA/ethidium bromide*. A molécula de brometo de ethidium se intercala entre os nucleotídeos, e o complexo apresenta fluorescência quando exposto à luz ultravioleta. O complexo do brometo de DNA/ethidium absorve fortemente a luz UV em 300 nm e emite uma luz alaranjada em 590nm. A diminuição na fluorescência está associada com uma diminuição no número de DNA, ou seja, de células malignas.

A tabela 4.5 contém os valores experimentais dos 23 compostos.¹⁰⁶ Os descritores que apresentaram uma melhor correlação com log EC₅₀ foram: eletronegatividade (η), calor de formação (ΔH), área da superfície molecular, energia de LUMO (E_{lumo}), índice de Balaban (BAC) e a soma dos índices eletrotopológicos (Ss). O gráfico destes descritores está na figura 4.12.

Observamos na tabela 4.2, que a primeira PC acumula uma variância de 77,79% e junto com a segunda PC, acumulam uma variância total de 96.97%. O valor de *PRESS* diminui da primeira para segunda PC, e aumenta na terceira PC. Então o modelo foi feito com 2 PCs principais. O modelo feito com 23 amostras e 2 componentes principais, apresentou um coeficiente de correlação igual a 0.830.

A primeira PC é o resultado da contribuição de todas as variáveis com coeficientes próximos. A segunda PC tem contribuições significativas da eletronegatividade (η) e de Ss, onde Ss é a soma dos índices eletrotopológicos. Na tabela 4.4 estão os coeficientes do vetor de regressão para log EC₅₀. O sinal positivo indica que um aumento na η e em Ss resulta em um aumento de log EC₅₀. Desta forma molécula muito eletronegativa e com alto valor de Ss precisa de maior concentração para fornecer o mesmo efeito tóxico que molécula com valor de eletronegatividade e Ss menor. Isto explica o fato dos compostos contendo átomos de bromo serem inativos. O mesmo ocorre com a área dos compostos analisados, onde quanto maior a área, maior a concentração necessária para obter o mesmo efeito. Isto poderia estar correlacionado com o espaço disponível entre as bases do DNA.

Os valores previstos de log EC₅₀ estão na figura 4.13 e na tabela 4.5. Os erros são altos para a maioria das moléculas, mas a tendência qualitativa é facilmente observada, com exceção das moléculas em negrito. Estes erros podem estar relacionados com a técnica experimental utilizada para se obter os valores de EC₅₀. Os resultados estatísticos deste modelo estão na tabela 4.4. Os valores previstos para todas as drogas (ativas) estão na tabela C.4. Também está nesta tabela o valor de log EC₅₀ previsto para uma droga

inativa (14). Este valor é alto mostrando que o modelo conseguiu prever que esta droga é inativa. Observamos que os compostos 13, 15 e 16 tiveram os valores previstos para $\log EC_{50}$ muito alto, indicando que estes compostos são inativos. Estes compostos são os que apresentam maior área, ou seja, este resultado está de acordo com a análise por componentes principais feita anteriormente. Observando a tabela 4.4, verificamos que ocorreu extrapolações para algumas moléculas. Algumas moléculas extrapolam 10%, mas outras tem valores extrapolados de até mais de 100%, como é o caso da amostra número 13. Desta maneira, não aconselhamos a utilização dos valores extrapolados pelo modelo, mesmo assim podemos concluir que os compostos 29, 37 45 47 e 51 devem ser bastante ativos (que apresentam menor valor de $\log EC_{50}$). Estas moléculas têm em comum os grupos $NHCH_3$ ou NH_2 na posição R^4 .

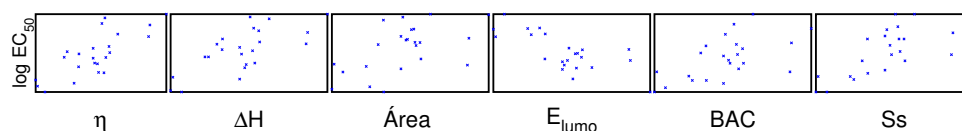


Figura 4.12: Gráfico de $\log EC_{50}$ versus descritores.

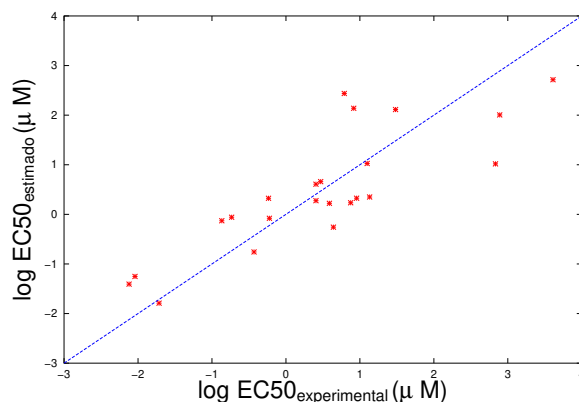


Figura 4.13: Gráfico de $\log EC_{50}$ previsto e experimental.

Tabela 4.2: Porcentagem da variância acumulada pelo modelo para log EC₅₀.

# PC	Esta PC	Total
1	77.79	77.79
2	19.18	96.97
3	1.83	98.80

Tabela 4.3: *Loadings* das variáveis no modelo para log EC₅₀.

Descritor	PC1	PC2	PC3
η	0.4154	0.7745	-0.0767
Área	0.5330	-0.3148	-0.7297
Bac	0.5502	-0.1811	0.1810
Ss	0.4905	-0.5180	0.6549

Tabela 4.4: Vetor de regressão para log EC₅₀. Bac é o índice de Balaban e Ss é a soma de índices eletrotopológicos.

Descritor	Vetor de regressão
η	0.6614
Área	0.5940
Bac	0.4247
Ss	0.4935

$$r = 0.830 \quad s^2 = 0.73 \quad Q^2 = 0.689$$

$$S_{PRESS} = 0.0362 \quad F = 44.30 \quad n = 23$$

Tabela 4.5: Valor Previsto e Experimental de $\log EC_{50}$.

Molécula	$\log EC_{50} (\mu M)$	
	Previsto	Experimental
1	-0.078	-0.223
2	-0.129	-0.868
3	1.017	2.833
4	-0.058	-0.734
8	2.005	2.890
24	2.139	0.916
37	-0.758	-0.431
38	1.026	1.099
39	2.113	1.482
40	2.715	3.611
42	0.351	1.131
43	2.437	0.788
45	-1.791	-1.715
46	0.277	0.405
47	-1.406	-2.120
48	0.223	0.588
49	0.327	0.956
50	-0.258	0.642
51	-1.251	-2.040
52	0.660	0.470
53	0.324	-0.236
54	0.234	0.875
55	0.608	0.405

4.5 Conclusão

Neste estudo foi possível obter a correlação quantitativa da atividade EC_{50} para uma série de moléculas derivadas de 4-amonometiltioxantenona e 5- aminometilbenzotilpiranoindazol, permitindo a previsão da atividade de moléculas similares sem que seja necessário sintetizá-las.

Os resultados obtidos utilizando-se análise por componentes principais indicam que as variáveis que estão determinando o comportamento ativo e inativo das drogas são a eletronegatividade, a área, o peso e a soma dos índices eletrotopológicos. O resultado da análise conclui que quanto mais eletronegativa a droga, menor sua atividade. E quanto maior a área, menos ativa a molécula.

Os resultados obtidos para a atividade EC_{50} estão de acordo com os resultados obtidos na PCA, indicando que um aumento na área está associado a uma diminuição na atividade. Este aumento na área é devido principalmente aos substituintes R^4 . Acreditamos que moléculas muito grandes não conseguem penetrar entre as bases do DNA.

Através de estudo gráfico, observamos que é necessário que a estrutura básica das drogas seja “plana” para que esta possa penetrar entre as bases de DNA. A atividade está associada a interações $\pi - \pi$ entre a estrutura básica da droga e as bases de DNA, e interação dos substituintes com o DNA, mas estes substituintes não podem se localizar na molécula de forma que ocasionem um efeito estérico.

Como em todo estudo de modelagem, é importante não apenas a análise quantitativa, a qual permite obter os fatores que podem potencializar a ação da droga e seus derivados, sempre que possível é necessário analisar como ocorre a interação da droga. Mesmo que este passo seja muito complexo, e esteja longe de uma situação ideal.

Apêndice A

Métodos Ab Initio

O termo “ab initio” se refere a uma classe de procedimentos computacionais que não usa parâmetros empíricos em sua formulação exceto pelas constantes físicas fundamentais e , m , h e c . Segundo este modelo todas as propriedades desejadas são calculadas do início. Para isso, definem-se conjuntos de funções de base para a expansão dos orbitais moleculares. As equações de Hartree-Fock-Roothaan (A.1) são então resolvidas exatamente sem nenhuma aproximação, na matriz de Fock,

$$F_{\mu\nu} = H_{\mu\nu} + \sum_{\lambda\sigma} P_{\lambda\sigma} \left[(\chi_{\mu}\chi_{\nu}|\chi_{\lambda}\chi_{\sigma}) - \frac{1}{2} (\chi_{\mu}\chi_{\lambda}|\chi_{\nu}\chi_{\sigma}) \right] \quad (\text{A.1})$$

em que

$H_{\mu\nu}$ - Hamiltoniano do caroço

$P_{\lambda\sigma}$ - Matriz densidade

χ_{μ} - Spin orbitais

Na prática, isso significa que se forem usados métodos numéricos para calcular as integrais a precisão desses métodos deve ser, de no mínimo, 8 dígitos. Em geral os ciclos de auto-consistência (SCF – Self Consistent Field) devem ser repetidos até que a variação da energia total seja menor ou igual a 10^{-5} ua ($\approx 0,26$ kJ/mol).

Se o objetivo for ir além do esquema Hartree-Fock (HF), no qual não são considerados os efeitos da correlação eletrônica, deve-se aplicar metodologias para compensar este efeito. Entre os métodos usuais de tratamento da correlação eletrônica podemos citar o CI (Configuration Interaction), MCSCF (Multiconfiguration SCF), MBPT (Many Body

Perturbation Theory) e MPPT (Møller-Plesset Perturbation Theory). A descrição dessas metodologias pode ser encontrada na referência (114).

A.1 Conjuntos de Base

Muitos métodos ab initio funcionam com conjuntos de base LCAO (Linear Combination of Atomic Orbitals). Para átomos ou moléculas diatômicas usa-se, geralmente, um conjunto de base que consiste de orbitais do tipo Slater (STO - Slater Type Orbitals),

$$\chi_{nlm}^{SL} = N r^{n-1} e^{-\zeta_n r} Y_{lm}(\vartheta, \varphi) \quad (\text{A.2})$$

em que

N - constante de normalização

Y - esféricas harmônicas

χ_{nlm}^{SL} - funções de Slater

ζ - expoente de Slater

Para moléculas com 3 átomos ou mais átomos aparecem integrais com 3 e 4 centros na matriz de Fock A.1. A avaliação dessas integrais com orbitais STO se torna muito trabalhosa e para moléculas grandes sua resolução é impraticável. Nesse contexto foram introduzidos os orbitais Gaussianos (GTO - Gaussian Type Orbitals),

$$\chi_{uvw}^G = N x^u y^v z^w e^{-\zeta_n r^2} \quad (\text{A.3})$$

A natureza Gaussiana da parte radial torna o cálculo das integrais de múltiplos centros mais fácil.¹¹⁵

A.2 Conjuntos de Base 6-31G, 6-311G e MC-311G

Esses conjuntos de base são denominados “split valence”, o que significa que há duas funções para descrever separadamente a parte eletrônica interna e o nível de valência do átomo.

A base 6-31G foi desenvolvida por Pople e colaboradores¹¹⁶ e é um dos conjuntos de base mais usados em cálculos ab initio. Nesse conjunto, são usadas 3 funções gaussianas para descrever a estrutura eletrônica dos átomos. A primeira delas é uma contração de

6 gaussianas primitivas e representa a estrutura eletrônica interna. A segunda função é obtida pela contração de 3 gaussianas primitivas e a terceira é 1 função gaussiana não contraída. A segunda e terceira funções representam a camada de valência do átomo.¹¹⁴ Por essa razão esse conjunto é denominado 6-31. As funções contraídas são obtidas por combinação linear das gaussianas primitivas. Esse conjunto de base é adequado para estudos de geometria de moléculas.¹¹⁵

O conjunto de base 6-311G foi desenvolvido por Pople e colaboradores¹¹⁷ e sua principal característica é a otimização dos expoentes e coeficientes das gaussianas em nível MP2 para os átomos do primeiro período da tabela periódica para uma melhor descrição de suas camadas de valência. Nessa base a estrutura eletrônica interna é representada por uma contração de 6 gaussianas primitivas, e a camada de valência é desdobrada em três funções.

A base MC-311G¹¹⁸ estende essa característica para os elementos do segundo período.

Apêndice B

Nomenclatura dos PCBs.

Na tabela B.1 está a nomenclatura de cada congênero, juntamente com a posição do átomo de cloro no anel.

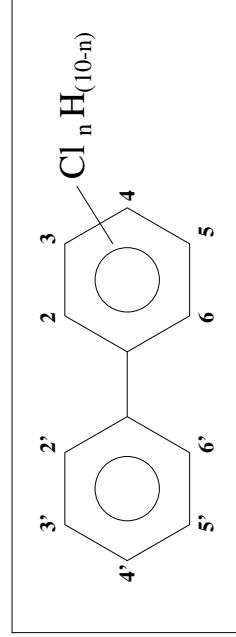


Figura B.1: Esquema das moléculas de biphenil policlorados.

Tabela B.1: Estrutura das moléculas de PCBs.

N.º	Estrutura	N.º	Estrutura	N.º	Estrutura	N.º	Estrutura
Monoclorobifenil				Tetraclorobifenis			
1	2	52,2',5,5'	107	2,3,3',4',5	161	2,3,3',4,5',6	
2	3	53,2',5,6'	108	2,3,3',4,5'	162	2,3,3',4',5,5'	
3	4	54,2',6,6'	109	2,3,3',4,6	163	2,3,3',4',5,6	
Diclorobifenis		55,3,3',4	110	2,3,3',4',6	164	2,3,3',4',5',6	
		56,3,3',4'	111	2,3,3',5,5'	165	2,3,3',5,5',6	
2,2'		57,3,3',5	112	2,3,3',5,6	166	2,3,4,4',5,6	
2,3		58,3,3',5'	113	2,3,3',5',6	167	2,3',4,4',5,5'	
2,3'		59,3,3',6	114	2,3,4,4',5	168	2,3',4,4',5',6	
2,4		60,3,4,4'	115	2,3,4,4',6	169	3,3',4,4',5,5'	
2,4'		61,3,4,5	116	2,3,4,5,6			
9	2,5	62,3,4,6	117	2,3,4',5,6	Heptaclorobifenis		
2,6		63,3,4',5	118	2,3',4,4',5	170	2,2',3,3',4,4',5	
1,3,3'		64,3,4',6	119	2,3,4,4',6	171	2,2',3,3',4,4',6	
1,3,4		65,3,5,6	120	2,3',4,5,5'	172	2,2',3,3',4,5,5'	
1,3,4'		66,3,4,4'	121	2,3',4,5',6	173	2,2',3,3',4,5,6	
1,3,5		67,3,4,5	122	2',3,3',4,5	174	2,2',3,3',4,5,6'	
15	4,4'	68,3',4,5'	123	2',3,4,4',5	175	2,2',3,3',4,5',6	
Triclorobifenis		69,3',4,6	124	2',3,4,5,5'	176	2,2',3,3',4,6,6'	
		70,3',4',5	125	2',3,4,5,6'	177	2,2',3,3',4',5,6	
2,2',3		71,3,4',6	126	3,3',4,4',5	178	2,2',3,3',5,5',6	
1,2,2',4		72,3',5,5'	127	3,3',4,5,5'	179	2,2',3,3',5,6,6'	
1,2,2',5		73,3',5',6			180	2,2',3,4,4',5,5'	
1,2,2',6		74,4,4',5	Hexaclorobifenis		181	2,2',3,4,4',5,6	
2,3,3'		75,4,4',6	128	2,2',3,3',4,4'	182	2,2',3,4,4',5,6'	

continua na próxima página

<i>continuação da página anterior.</i>					
<i>N.º</i>	<i>Estrutura</i>	<i>N.º</i>	<i>Estrutura</i>	<i>N.º</i>	<i>Estrutura</i>
22,3,4	76 2',3,4,5	129 2,2',3,3',4,5	183 2,2',3,4,4',5',6		
22,3,4'	77,3',4,4'	130 2,2',3,3',4,5'	184 2,2',3,4,4',6,6'		
22,3,5	78,3',4,5	131 2,2',3,3',4,6	185 2,2',3,4,5,5',6		
22,3,6	79,3',4,5'	132 2,2',3,3',4,6'	186 2,2',3,4,5,6,6'		
22,3',4	80,5,5'	133 2,2',3,3',5,5'	187 2,2',3,4',5,5',6		
22,3',5	81,4,4',5	134 2,2',3,3',5,6	188 2,2',3,4',5,6,6'		
22,3',6		135 2,2',3,3',5,6'	189 2,3,3',4,4',5,5'		
22,4,4'	Pentaclorobifenis	136 2,2',3,3',6,6'	190 2,3,3',4,4',5,6		
22,4,5	82,2',3,3',4	137 2,2',3,4,4',5	191 2,3,3',4,4',5',6		
22,4,6	83,2',3,3',5	138 2,2',3,4,4',5'	192 2,3,3',4,5,5',6		
32,4',5	84,2',3,3',6	139 2,2',3,4,4',6	193 2,3,3',4',5,5',6		
32,4',6	85,2',3,4,4'	140 2,2',3,4,4',6'			
33 2',3,4	86,3,4,5	141 2,2',3,4,5,5'	Octaclorobifenis		
34 2',3,5	87,2',3,4,5'	142 2,2',3,4,5,6	194 2,2',3,3',4,4',5,5'		
33,3',4	88,2',3,4,6	143 2,2',3,4,5,6'	195 2,2',3,3',4,4',5,6		
33,3',5	89,2',3,4,6'	144 2,2',3,4,5',6	196 2,2',3,3',4,4',5',6		
33,4,4'	90,3,4',5	145 2,2',3,4,6,6'	197 2,2',3,3',4,4',6,6'		
38,4,5	91,2',3,4',6	146 2,2',3,4',5,5'	198 2,2',3,3',4,5,5',6		
39,4',5	92,2',3,5,5'	147 2,2',3,4',5,6	199 2,2',3,3',4',5,5',6		
	93,2',3,5,6	148 2,2',3,4,5,6'	200 2,2',3,3',4,5,6,6'		
Tetraclorobifenis	94,2',3,5,6'	149 2,2',3,4',5',6	201 2,2',3,3',4,5',6,6'		
22,3,3'	95,2',3,5',6	150 2,2',3,4',6,6'	202 2,2',3,3',5,5',6,6'		
42,2',3,4	96,3,6,6'	151 2,2',3,5,5',6	203 2,2',3,4,4',5,5',6		
42,2',3,4'	97,2',3',4,5	152 2,2',3,5,6,6	204 2,2',3,4,4',5,6,6'		
42,2',3,5	98,2',3,4,6	153 2,2',4,4',5,5'	205 2,3,3',4,4',5,5',6'		
42,2,3,5'	99,2',4,4',5	154 2,2',4,4,5,6'			
<i>continua na próxima página</i>					

<i>continuação da página anterior.</i>					
N°. Estrutura	N°. Estrutura	N°. Estrutura	N°. Estrutura	N°. Estrutura	N°. Estrutura
43 ,2',3,6	102',4,4',6	155	2,2',4,4',6,6'	Nonaclorobifenis	
42 ,3,6'	102',2',4,5,5'	156	2,3,3',4,4',5	206	2,2',3,3',4,4',5,5',6
42 ,2',4,4'	102',2',4,5,6'	157	2,3,3',4,4',5'	207	2,2',3,3',4,4',5,6,6'
48 ,2',4,5	103',2',4,5',6	158	2,3,3',4,4',6	208	2,2',3,3',4,5,5',6,6'
49 ,2',4,5'	102',2',4,6,6'	159	2,3,3',4,5,5'		
50 ,4,6	103',3,3',4,4'	160	2,3,4',4,6,6'	Decaclorobifenis	
52 ,2',4,6'	103',3',4,5			209	2,2',3,3',4,4',5,5',6,6'

Apêndice C

Tabela de Dados

Tabela C.1: Variáveis utilizadas nos modelos das moléculas de PCB.

PCB	Area ($\text{\AA}^2$)	Volume ($\text{\AA}^3$)	Ovalidade	$\Delta H(\text{Kcal/mol})$	μ (Debye)	η (eV)	Dureza (ev)	E_{lumo} (eV)	E_{homo} (eV)
1	207.888	174.473377	42.692.226	4.645	4.533	-0.113	-9.178		
2	208.192	174.552.378	40.543496	4.695	4.426	-0.27	-9.121		
3	208.168	174.587.8	40.402.566	4.664	4.342	-0.322	-9.006		
4	225.242	188.586.416	37.112.611	4.726	4.796	0.071	-9.522		
5	222.99	188.349403	37.37	2.135	4.811	4.539	-0.273	-9.35	
6	224.398	188.582.411	35.748	2.356	4.809	4.523	-0.286	-9.331	
7	224.46	188.568.411	36.06453	4.811	4.433	-0.378	-9.245		
8	224.407	188.577	1.411	35.537	2.243	4.771	4.445	-0.326	-9.217
9	224.385	188.566.411	35.976	0.437	4.816	4.47	-0.346	-9.285	
10	224.978	188.522.415	37.758	0.856	4.86	4.751	-0.109	-9.61	
11	224.631	188.65612	33.651	2.171	4.865	4.411	-0.454	-9.276	
12	223.231	188.464404	34.958	2.444	4.818	4.314	-0.504	-9.132	
13	224.612	188.652.412	33.476.331	4.83	4.329	-0.501	-9.16		
14	224.553	188.64412	34.081.711	4.878	4.409	-0.469	-9.288		
15	224.63	188.646.412	33.33	0.001	4.803	4.26	-0.543	-9.063	
16	240.025	202.441.44	13793	2.482	4.851	4.721	-0.13	-9.573	
17	24623	202.65948	30.527	2.128	4.918	4.677	-0.241	-9.595	
18	24634	202.66648	30.473417	4.857	4.621	-0.235	-9.478		
19	240.209	202.473441	32.036.238	4.836	4.692	-0.143	-9.528		
20	239.552	202.463437	30.537	2.769	4.96	4.54	-0.42	-9.5	
21	237.747	202.19427	32.314	2.56	4.946	4.42	-0.526	-9.366	
22	239.509	202.45636	30.319	2.274	4.914	4.454	-0.46	-9.368	
23	239.333	202.412.486	13163517	4.977	4.468	-0.509	-9.446		

continua na próxima página

continuação da página anterior.										
PCB	Area (Å ²)	Volume (Å ³)	Ovalidade	ΔH(Kcal/mol)	μ (Debye)	η (eV)	Dureza (ev)	E _{LUMO} (eV)	E _{HOMO} (eV)	
24	239.735	202.373438	32.651	0.936	4.982	4.567	-0.414	-9.549		
25	240.917	202.666.444	29.236.879	4.957	4.431	-0.527	-9.388			
26	240.888	202.609444	29.142974	4.955	4.462	-0.493	-9.418			
27	24402	202.62647	30.83	2.09	4.893	4.666	-0.227	-9.558		
28	240.916	202.664444	29.024103	4.921	4.364	-0.557	-9.285			
29	239.351	202.456.435	30.872.686	4.957	4.362	-0.594	-9.319			
30	24333	202.586.447	13869	0.589	5.08	4.655	-0.425	-9.734		
31	240.832	202.663443	28.941.089	4.925	4.399	-0.526	-9.324			
32	24414	202.621.447	30.637	2.366	4.877	4.649	-0.228	-9.526		
33	239.506	202.487	1.436	30.195	3.135	4.91	4.42	-0.49	-9.331	
34	240.815	202.66643	29.324	2.344	4.378	4.515	-0.463	-9.492		
35	235.7	203.34417	30.389	2.524	4.983	4.151	-0.831	-9.134		
36	24044	202.742.444	27.304409	5.036	4.401	-0.635	-9.437			
37	239.657	202.552.437	27.988.237	4.942	4.238	-0.705	-9.18			
38	234.058	202.608	32.217	2.823	4.972	4.14	-0.832	-9.112		
39	24004	202.731.444	27.134	0.188	4.993	4.316	-0.676	-9.309		
40	254.824	216.30662	26.562	2.838	4.941	4.701	-0.24	-9.642		
41	254.779	216.28662	26.768	2.994	5.05	4.624	-0.427	-9.674		
42	256.42	216.519471	25.301	2.322	5.008	4.659	-0.349	-9.667		
43	256.383	216.507	1.47	25.662	2.125	5.007	4.566	-0.441	-9.573	
44	256.427	216.523471	25.259	2.047	4.946	4.612	-0.333	-9.558		
45	254.961	216.312.463	26.992.531	4.987	4.564	-0.423	-9.55			
46	255.003	216.33463	26.769	2.419	4.905	4.661	-0.244	-9.566		
47	258.016	216.737	1.479	24.04633	5.027	4.636	-0.391	-9.662		
48	256.647	216.552.472	25.404	2.347	5.026	4.529	-0.497	-9.556		
49	258.022	216.74479	23.996.237	4.973	4.599	-0.374	-9.572			
50	256.551	216.532.471	25.949509	5.043	4.585	-0.458	-9.628			
51	256.595	216.5671	25.48	2.163	4.94	4.678	-0.261	-9.618		
52	258.029	216.743479	23.952	0.057	4.96	4.603	-0.357	-9.563		
53	256.601	216.35671	25.452.258	4.868	4.611	-0.257	-9.478			
54	255.182	216.36664	26.955	0	4.882	4.714	-0.169	-9.596		
55	254.348	216.307	1.46	25.571	2.671	5.083	4.434	-0.649	-9.517	
56	254.61	216.364461	25.067	3.074	5.04	4.437	-0.603	-9.477		
57	253.183	215.291.457	24.81.301	4.936	4.3	-0.636	-9.236			
58	252.942	215.281.456	24.722.698	5.003	4.379	-0.625	-9.382			
59	254.36	215.33464	26.306.279	4.844	4.387	-0.456	-9.231			
60	12821	214.981.448	25.566.294	4.918	4.243	-0.676	-9.161			
61	249.85	214.71441	26.943747	4.914	4.24	-0.674	-9.154			
62	252.589	215.014455	27.61528	4.887	4.334	-0.552	-9.221			
63	253.204	215.299458	24.718	0.576	4.93	4.274	-0.657	-9.204		
64	254.36	215.333464	26.71233	4.848	4.384	-0.464	-9.232			
65	252.399	214.934454	27.611	0.698	4.848	4.268	-0.579	-9.116		
66	253.108	215.297	1.457	24.598.499	4.917	4.226	-0.691	-9.143		
67	255.766	215.523471	25.064045	4.885	4.329	-0.556	-9.214			
68	254.884	215.612.466	23.766	0.859	4.996	4.31	-0.686	-9.305		
69	256.22	215.658.473	25.469	0.929	4.967	4.501	-0.466	-9.468		
70	253.196	215.331.457	24.544391	4.913	4.246	-0.668	-9.159			

continua na próxima página

continuação da página anterior.

PCB	Area (Å ²)	Volume (Å ³)	Ovalidade	ΔH (Kcal/mol)	μ (Debye)	η (eV)	Dureza (ev)	E_{lumo} (eV)	E_{homo} (eV)
71	254.513	215.371.465	26.173	2.213	4.792	4.5	-0.293	-9.292	
72	254.971	215.63466	23.754	0.914	4.958	4.306	-0.652	-9.264	
73	255.42	215.66669	25.49781	4.926	4.494	-0.433	-9.42		
74	253.244	215.333458	24.683	0.094	4.911	4.195	-0.717	-9.106	
75	256.214	215.652.473	25.387	0.58	4.958	4.491	-0.467	-9.449	
76	12367	215.013448	25.667	2.396	4.923	4.266	-0.657	-9.188	
77	247.811	214.168.432	23.321.51	4.963	3.992	-0.972	-8.955		
78	247.769	214.13431	23.467	1.584	5.012	4.066	-0.946	-9.079	
79	249.525	214.46344	22.489	0.767	5.037	4.077	-0.96	-9.114	
80	12252	214.762.449	2.682	0.019	5.156	4.209	-0.946	-9.365	
81	247.77	214.143431	23.363	0.896	4.97	3.996	-0.974	-8.966	
82	267.161	228.72.477	2.708	2.094	4.912	4.448	-0.464	-9.36	
83	268.863	229.00585	20.81463	4.869	4.378	-0.491	-9.247		
84	267.478	228.82.478	22.34613	4.845	4.385	-0.46	-9.23		
85	265.573	228.838.468	2.183715	4.971	4.308	-0.662	-9.279		
86	267.213	228.714477	2.875	2.115	4.902	4.326	-0.575	-9.228	
87	268.993	229.042.486	4.891	4.4	-0.491	-9.292			
88	267.559	228.821.479	22.53958	4.907	4.328	-0.579	-9.235		
89	267.618	228.852.479	22.294	2.207	4.888	4.466	-0.422	-9.354	
90	268.997	229.032.486	20.74589	4.892	4.396	-0.497	-9.288		
91	269.273	229.11687	2.394231	4.857	4.383	-0.474	-9.24		
92	270.697	229.333494	19.924	0.817	4.877	4.378	-0.5	-9.255	
93	267.418	228.796.478	22.547	1.288	4.856	4.269	-0.587	-9.126	
94	269.365	229.148.487	2.467	1.559	4.821	4.382	-0.439	-9.202	
95	269.349	229.147	1.487	2.382	0.695	4.851	4.38	-0.471	-9.231
96	267.954	228.949481	22.961	0.893	4.803	4.383	-0.42	-9.186	
97	268.991	229.031.486	20.848.717	4.891	4.35	-0.541	-9.241		
98	269.345	229.139487	2.508.491	4.912	4.42	-0.492	-9.332		
99	270.848	229.36695	19.891	0.863	4.904	4.342	-0.562	-9.247	
100	271.18	229.449496	20.541	0.832	4.973	4.471	-0.502	-9.444	
101	273.034	230.626.501	18.962.216	5.117	4.521	-0.596	-9.639		
102	271.612	230.439494	20.408	2.465	5.036	4.525	-0.511	-9.562	
103	272.937	230.607	1.501	19.505	0.167	5.067	4.501	-0.566	-9.569
104	271.522	230.421.494	20.93398	5.094	4.612	-0.481	-9.706		
105	269.329	230.198.483	20.171	2.439	5.153	4.35	-0.803	-9.503	
106	269.11	230.142.482	20.945	2.103	5.215	4.376	-0.839	-9.591	
107	270.948	230.422.491	19.05577	5.175	4.392	-0.783	-9.566		
108	270.678	230.377	1.489	19.332.78	5.223	4.433	-0.79	-9.655	
109	270.873	230.287	1.491	2.321.446	5.226	4.482	-0.744	-9.709	
110	271.188	230.357	1.492	20.386	2.525	5.118	4.525	-0.593	-9.643
111	272.186	230.588.497	18.24	0.93	5.235	4.457	-0.778	-9.692	
112	270.871	230.28.491	2.263157	5.202	4.456	-0.746	-9.659		
113	272.514	230.531.499	19.583881	5.156	4.556	-0.6	-9.712		
114	268.988	230.12.481	20.727	1.08	5.175	4.309	-0.867	-9.484	
115	270.886	230.28691	2.117	0.857	5.219	4.471	-0.748	-9.69	
116	267.265	229.734473	24.91975	5.255	4.447	-0.808	-9.703		
117	270.883	230.277	1.491	2.053	0.735	5.212	4.458	-0.754	-9.669

continua na próxima página

<i>continuação da página anterior.</i>										
PCB	Area (Å ²)	Volume (Å ³)	Ovalidade	ΔH(Kcal/mol)	μ (Debye)	η (eV)	Dureza (ev)	E _{lumo} (eV)	E _{homo} (eV)	
118	270.952	230.451.491	18.748.063	5.164	4.299	-0.865	-9.464			
119	272.782	230.5195	19.412.858	5.143	4.513	-0.63	-9.656			
120	272.243	230.628.497	17.933	0.074	5.217	4.36	-0.857	-9.577		
121	274.108	230.794507	18.612.062	5.198	4.555	-0.642	-9.753			
122	269.32	230.207	1.483	20.248	2.937	5.18	4.414	-0.765	-9.594	
123	269.29	230.196.483	20.104663	5.151	4.341	-0.81	-9.491			
124	270.698	230.41689	18.908	2.275	5.175	4.362	-0.813	-9.536		
125	271.205	230.386.492	20.5	3.519	5.095	4.629	-0.466	-9.725		
126	269.509	230.308.483	17.994173	5.2	4.203	-0.997	-9.403			
127	270.872	230.48949	17.191.036	5.269	4.282	-0.977	-9.541			
128	284.319	243.97506	16.715	2.509	5.236	4.608	-0.628	-9.843		
129	284.295	243.968.506	17.039	2.743	5.243	4.498	-0.744	-9.741		
130	285.915	244.196.513	15.645	2.023	5.173	4.546	-0.628	-9.719		
131	284.464	243.989506	17.315	2.268	5.229	4.478	-0.751	-9.707		
132	284.501	244.006.507	16.872	2.628	5.143	4.556	-0.588	-9.699		
133	287.52	244.417	1.521	14.58316	5.171	4.546	-0.625	-9.717		
134	284.478	243.9507	17.279	2.021	5.191	4.455	-0.736	-9.646		
135	286.101	244.226.514	15.817	1.71	5.115	4.52	-0.595	-9.636		
136	284.682	244.03507	17.043307	5.056	4.544	-0.512	-9.6			
137	285.882	244.183513	15.78729	5.256	4.499	-0.756	-9.755			
138	286.179	244.243515	15.377	1.807	5.188	4.52	-0.668	-9.709		
139	286.06	244.209514	16.037	1.318	5.263	4.499	-0.764	-9.761		
140	286.085	244.22514	15.877	2.134	5.226	4.586	-0.641	-9.812		
141	285.885	244.184513	15.753942	5.223	4.472	-0.751	-9.696			
142	282.461	243.67	1.497	19.342	2.417	5.256	4.443	-0.812	-9.699	
143	284.469	244.002.506	17.169	3.123	5.168	4.499	-0.669	-9.666		
144	286.058	244.208.514	16.024125	5.193	4.432	-0.761	-9.625			
145	284.642	244.022.507	17.422	2.169	5.176	4.499	-0.678	-9.675		
146	287.777	244.462.522	14.308	0.929	5.195	4.517	-0.678	-9.713		
147	286.061	244.204514	15.998.065	5.207	4.457	-0.75	-9.664			
148	287.688	244.447	1.522	14.818.052	5.153	4.499	-0.654	-9.652		
149	286.358	244.27	1.515	15.528.857	5.117	4.518	-0.598	-9.635		
150	286.272	244.234515	16.013087	5.103	4.53	-0.573	-9.634			
151	286.068	244.206.514	15.99	0.348	5.181	4.438	-0.743	-9.618		
152	284.653	244.021.507	17.404586	5.111	4.46	-0.651	-9.57			
153	288.038	244.507	1.523	14.041	0.013	5.202	4.508	-0.695	-9.71	
154	287.949	244.504523	14.532.06	5.15	4.499	-0.651	-9.649			
155	287.844	244.46322	15	0	5.247	4.663	-0.583	-9.91		
156	285.722	244.151.512	15.668.267	5.296	4.48	-0.816	-9.775			
157	284.11	244.046.504	15.42.992	5.274	4.342	-0.933	-9.616			
158	285.901	244.17513	15.941.581	5.271	4.443	-0.829	-9.714			
159	285.493	244.219511	14.809	0.922	5.338	4.38	-0.959	-9.718		
160	283.655	243.817	1.503	18.256.44	5.327	4.432	-0.896	-9.759		
161	287.23	244.3552	15.157	0.918	5.322	4.484	-0.838	-9.805		
162	287.361	244.452	14.416.659	5.235	4.556	-0.679	-9.791			
163	285.907	244.177	1.513	15.887	1.683	5.268	4.444	-0.823	-9.712	
164	285.952	244.20513	15.649	2.801	5.219	4.555	-0.665	-9.774		

continua na próxima página

continuação da página anterior.

PCB	Area (Å ²)	Volume (Å ³)	Ovalidade	ΔH (Kcal/mol)	μ (Debye)	η (eV)	Dureza (ev)	E_{lumo} (eV)	E_{homo} (eV)
165	287.233	244.35152	15.109		5.281	4.453	-0.828	-9.734	
166	283.666	243.814503	18.042		0.232	5.326	4.423	-0.903	-9.749
167	285.835	244.304512	14.02805		5.278	4.294	-0.984	-9.573	
168	287.542	244.427	1.521	14.676	2.097	5.262	4.548	-0.714	-9.81
169	284.308	244.15505	13.327		0.003	5.322	4.192	-1.13	-9.513
170	299.036	257.802527	12.207		2.006	5.319	4.494	-0.825	-9.812
171	299.205	257.824527	12.435		2.018	5.327	4.494	-0.832	-9.821
172	300.63	258.018534	1.157		1.486	5.306	4.471	-0.835	-9.778
173	297.251	257.521519	14.281		2.349	5.326	4.424	-0.901	-9.75
174	299.214	257.82927	12.346		2.474	5.242	4.489	-0.754	-9.731
175	300.806	258.044535	1.394167		5.274	4.431	-0.843	-9.705	
176	299.401	257.858528	12.56584		5.224	4.46	-0.763	-9.684	
177	299.212	257.81927	12.404		2.02	5.263	4.454	-0.809	-9.718
178	300.815	258.041535	1.369		0.949	5.254	4.44	-0.814	-9.694
179	299.405	257.851528	12.534177		5.178	4.454	-0.724	-9.632	
180	300.891	258.064535	10.879		0.967	5.301	4.47	-0.831	-9.771
181	298.842	257.73926	13.0026		5.349	4.437	-0.913	-9.786	
182	300.796	258.048535	1.351798		5.258	4.493	-0.764	-9.751	
183	3064	258.08936	1.102	0.91	5.273	4.431	-0.842	-9.705	
184	300.973	258.072535	1.547		0.998	5.267	4.494	-0.774	-9.761
185	298.841	257.738526	13.001295		5.294	4.382	-0.912	-9.677	
186	3093	259.926531	14.371		2.524	5.262	4.431	-0.832	-9.693
187	3069	258.08536	1.073	0.858	5.257	4.441	-0.816	-9.698	
188	300.989	258.073535	1.533	0.176	5.196	4.458	-0.738	-9.653	
189	298.799	257.86525	10.949	0.859	5.385	4.298	-1.087	-9.683	
190	298.683	257.707	1.525	12.9208	5.37	4.397	-0.973	-9.767	
191	300.657	258.021534	1.264677		5.384	4.48	-0.905	-9.864	
192	300.011	257.883531	12.157		0.099	5.414	4.43	-0.984	-9.845
193	300.667	258.021534	1.218893		5.339	4.452	-0.887	-9.791	
194	313.74	271.621547	7.762072		5.382	4.484	-0.898	-9.866	
195	31.987	271.351539	9.448778		5.409	4.428	-0.981	-9.837	
196	313.919	271.647	1.548	7.96892	5.354	4.449	-0.905	-9.803	
197	314.095	271.67	1.548	8.138137	5.32	4.479	-0.84	-9.799	
198	313.588	271.572546	8.424		0.992	5.373	4.38	-0.993	-9.753
199	312.175	271.3854	9.568932		5.321	4.405	-0.916	-9.726	
200	314.11	271.668548	8.132		0.901	5.272	4.43	-0.842	-9.702
201	313.93	271.68548	7.94809		5.319	4.449	-0.87	-9.768	
202	314.124	271.667	1.549	8.129	0.016	5.239	4.441	-0.797	-9.68
203	313.847	271.617	1.547	8.126	0.111	5.373	4.382	-0.991	-9.755
204	313.759	271.602547	8.1565		5.351	4.424	-0.926	-9.775	
205	313.44	271.5546	8.302		0.931	5.48	4.43	-1.049	-9.91
206	326.703	285.173559	5.03		0.885	5.451	-1.054	-9.849	
207	326.88	285.19456	5.184		0.962	5.405	-0.992	-9.818	
208	326.902	285.19956	5.184		0.939	5.37	-0.995	-9.746	
209	339.667	298.72872	2.268		0.019	5.451	-1.053	-9.849	

Tabela C.2: Variáveis utilizadas nos modelos das moléculas de PCB.

PCB	χ^2	χ^6	χ_b^6	χ^v	χ_{ch}^6	χ_p^{vc}	W_9	H^D	H^v	
324	1.373	0.403	0.99	0.185	1.138	240	14	3.42	2.56	
43	1.441	0.427	0.985	0.185	0.891	246	16	3.409	2.608	
418	1.509	0.451	0.985	0.185	0.791	252	16	3.395	2.696	
4851	1.551	0.48	1.219	0.167	1.824	287	18	3.544	2.553	
5829	1.566	0.49	1.123	0.17	1.765	291	18	3.537	2.592	
5957	1.697	0.54	1.133	0.167	1.433	294	18	3.534	2.594	
5957	1.726	0.555	1.113	0.17	1.507	298	18	3.526	2.667	
5946	1.662	0.521	1.146	0.167	1.353	301	18	3.521	2.684	
957	1.57	0.492	1.319	0.17	1.658	293	18	3.537	2.594	
5861	1.628	0.519	1.122	0.17	1.945	286	16	3.546	2.528	
11	6.064	1.683	0.527	1.141	0.167	1.206	301	22	3.524	2.644
12926	1.689	0.535	1.125	0.17	1.317	303	20	3.517	2.711	
13	6.052	1.776	0.558	1.14	0.167	1.105	308	20	3.511	2.725
14	6.076	1.726	0.555	1.116	0.17	1.328	298	20	3.529	2.609
15	6.04	1.84	0.581	1.14	0.167	1.005	315	24	3.499	2.815
16	6.357	1.794	0.594	1.342	0.151	2.423	344	22	3.651	2.587
17	6.485	1.865	0.621	1.344	0.151	2.194	352	22	3.64	2.665
18	6.485	1.809	0.6	1.536	0.151	2.325	346	22	3.651	2.587
19	6.388	1.789	0.591	1.42	0.151	2.742	338	22	3.66	2.53
20	6.463	1.873	0.622	1.267	0.151	2.06	352	24	3.641	2.63
21	6.335	1.855	0.624	1.233	0.158	2.467	352	22	3.639	2.684
22	6.451	1.857	0.609	1.279	0.151	1.98	360	22	3.629	2.714
23	6.475	1.784	0.599	1.416	0.158	2.461	348	22	3.648	2.6
24	6.366	1.75	0.586	1.42	0.158	2.802	342	20	3.655	2.566

continua na próxima página.

continua na próxima página.

continuação da página anterior.										
PCB	χ^2	χ^6	χ_b^6	${}^5\chi^v$	${}^6\chi^{ch}$	${}^6\chi_p^{uc}$	W	P_9	H^D	H^u
51	7.022	2.065	0.721	1.546	0.136	3.113	410	26	3.748	2.646
52	7.119	2.046	0.714	1.854	0.136	2.827	412	28	3.75	2.625
53	7.022	2.108	0.741	1.726	0.136	3.225	403	26	3.758	2.565
54	6.926	2.01	0.696	1.691	0.136	3.757	394	28	3.767	2.514
55	6.969	2.165	0.756	1.377	0.139	2.762	421	28	3.735	2.723
56	6.959	2.085	0.72	1.408	0.136	2.487	424	28	3.731	2.738
57	7.109	2.071	0.725	1.561	0.139	2.756	416	30	3.744	2.641
58	7.109	2.201	0.776	1.386	0.136	2.478	417	30	3.742	2.638
59	7	2.118	0.748	1.552	0.139	3.079	409	26	3.751	2.606
60	6.957	2.147	0.743	1.39	0.139	2.681	430	30	3.724	2.807
61	6.852	2.006	0.711	1.503	0.147	3.533	412	26	3.745	2.684
62	6.884	2.023	0.726	1.493	0.147	3.685	407	24	3.75	2.648
63	7.097	2.08	0.719	1.572	0.139	2.675	425	26	3.733	2.72
64	6.988	1.999	0.693	1.577	0.139	3.018	418	24	3.739	2.69
65	6.884	1.896	0.671	1.671	0.147	3.898	402	24	3.758	2.58
66	7.087	2.263	0.79	1.398	0.136	2.228	433	30	3.723	2.811
67	7.099	2.146	0.752	1.56	0.139	2.587	423	28	3.735	2.727
68	7.237	2.403	0.852	1.376	0.136	2.218	426	26	3.733	2.706
69	7.14	2.351	0.847	1.352	0.139	2.708	416	24	3.742	2.66
70	7.087	2.086	0.721	1.604	0.136	2.38	426	28	3.732	2.74
71	6.991	2.185	0.773	1.396	0.136	2.649	417	24	3.74	2.68
72	7.237	2.198	0.777	1.582	0.136	2.37	419	30	3.743	2.639
73	7.14	2.42	0.876	1.361	0.136	2.618	410	24	3.751	2.574
74	7.087	2.133	0.739	1.572	0.139	2.507	432	30	3.725	2.811
75	7.128	2.209	0.785	1.377	0.139	2.648	425	28	3.731	2.75
76	6.971	2.192	0.778	1.379	0.139	2.639	419	26	3.738	2.697

continuação da página anterior.

continuação da página anterior.										
PCB	χ^2	χ^6	χ_b^6	${}^5\chi^v$	${}^6\chi^{ch}$	${}^6\chi_p^{vc}$	W	P_9	H^D	H^u
103	7.668	2.411	0.891	1.825	0.124	3.715	480	30	3.843	2.665
104	7.571	2.275	0.835	1.79	0.124	4.248	470	32	3.851	2.618
105	7.464	2.377	0.854	1.518	0.124	3.189	502	36	3.819	2.832
106	7.486	2.295	0.838	1.648	0.129	3.829	488	34	3.834	2.725
107	7.605	2.286	0.823	1.702	0.124	3.183	496	34	3.828	2.748
108	7.614	2.497	0.911	1.497	0.124	3.179	494	34	3.829	2.732
109	7.518	2.395	0.889	1.625	0.129	3.962	482	30	3.839	2.69
110	7.496	2.292	0.835	1.694	0.124	3.507	488	30	3.834	2.718
111	7.754	2.378	0.873	1.681	0.124	3.174	488	38	3.838	2.652
112	7.518	2.247	0.829	1.804	0.129	4.175	476	32	3.847	2.622
113	7.646	2.507	0.933	1.66	0.124	3.478	480	32	3.844	2.618
114	7.474	2.302	0.831	1.66	0.129	3.748	498	34	3.824	2.805
115	7.506	2.273	0.833	1.65	0.129	3.901	492	32	3.829	2.774
116	7.271	2.12	0.793	1.724	0.139	5.235	470	28	3.849	2.653
117	7.506	2.149	0.779	1.828	0.129	4.113	486	28	3.836	2.702
118	7.595	2.359	0.849	1.701	0.124	3.013	504	36	3.82	2.836
119	7.636	2.522	0.933	1.493	0.124	3.136	496	32	3.826	2.775
120	7.745	2.474	0.905	1.68	0.124	3.004	496	34	3.83	2.736
121	7.786	2.761	1.037	1.459	0.124	3.106	488	28	3.836	2.671
122	7.476	2.353	0.856	1.513	0.124	3.268	492	34	3.829	2.736
123	7.464	2.377	0.854	1.518	0.124	3.189	502	36	3.819	2.832
124	7.605	2.35	0.856	1.709	0.124	3.16	494	34	3.83	2.738
125	7.508	2.535	0.945	1.489	0.124	3.41	484	28	3.837	2.678
126	7.573	2.389	0.849	1.538	0.124	2.856	510	40	3.813	2.847
127	7.723	2.38	0.857	1.53	0.124	2.869	502	40	3.823	2.751
128	7.873	2.536	0.951	1.688	0.111	4.43	569	40	3.916	2.822
continua na próxima página.										

continuação da página anterior.										
PCB	χ^2	χ^6	χ_b^6	${}^5\chi^v$	${}^6\chi^{ch}$	${}^6\chi_p^{vc}$	W	P_9	H^D	H^{su}
158.217	2.541	0.975	1.89	0.111	4.74	555	40	3.93	2.722	
156	7.982	2.511	0.936	1.789	0.113	4.255	577	42	3.912	2.835
157	7.982	2.649	0.99	1.624	0.111	3.969	578	42	3.91	2.832
158.014	2.569	0.975	1.767	0.113	4.39	570	38	3.917	2.803	
158.132	2.605	0.986	1.768	0.113	4.247	568	42	3.921	2.738	
160	7.905	2.474	0.951	1.857	0.12	5.512	552	36	3.932	2.697
168.163	2.787	1.074	1.733	0.113	4.36	561	36	3.927	2.703	
168.122	2.533	0.953	1.807	0.111	3.964	571	42	3.919	2.75	
168.014	2.424	0.916	1.945	0.113	4.603	563	36	3.924	2.733	
168.014	2.625	1.002	1.788	0.111	4.269	562	36	3.925	2.72	
168.163	2.619	1.008	1.912	0.113	4.575	554	40	3.934	2.636	
166	7.893	2.373	0.901	1.881	0.12	5.451	563	36	3.922	2.777
168.112	2.627	0.985	1.807	0.111	3.794	580	42	3.911	2.835	
168.154	2.875	1.106	1.587	0.111	3.898	571	36	3.917	2.776	
168.09	2.571	0.947	1.655	0.111	3.657	587	48	3.905	2.848	
180.91	2.737	1.065	1.947	0.101	5.479	650	46	4.002	2.825	
178.423	2.706	1.062	0.008	0.101	5.736	642	44	4.007	2.798	
178.531	2.721	1.02	1.18	0.101	5.454	642	46	4.01	2.741	
178.304	2.62	1.042	1.17	0.105	6.556	626	42	4.019	2.703	
178.423	2.723	1.072	1.81	0.101	5.883	632	42	4.016	2.713	
178.563	2.81	1.112	1.66	0.101	5.691	634	42	4.016	2.709	
180.54	2.653	1.052	0.321	0.101	6.256	624	42	4.022	2.685	
178.423	2.646	1.042	1.75	0.101	5.921	634	42	4.014	2.724	
178.563	2.728	1.082	3.33	0.101	5.877	626	46	4.023	2.637	
178.454	2.671	1.086	1.78	0.101	6.413	616	44	4.029	2.609	
180.521	2.727	1.082	1.128	0.101	5.312	652	46	4.003	2.827	
continua na próxima página.										

<i>continuação da página anterior.</i>									
PCB	log K_{ow}	log S (mg/L)	log K_{oc}	log H (Pa m ³ /mol)	-log Y_w	mp (°C)	log IC_{50} (μM)	log EC_{50} (μM)	
98	6.76	-4.358	6.3.701	6.5	101.1	1.913	5.123		
99	6.82	-4.954	6.13.034	6.59	94.7	1.94	5.315		
100	6.8	-4.776	6.13.31	6.56	93.2	1.855	5.016		
101	6.86	-5.038	6.3.423	6.6	85.4	1.766	4.852		
102	6.8	-4.556	6.09.691	6.54	89.9	1.811	4.837		
103	6.86	-4.789	6.13.716	6.56	98.4	1.79	4.904		
104	6.8	-4.571	6.09.954	6.52	84.9	1.752	4.614		
105	6.95	-5.614	6.12.848	6.68	111.9	1.818	5.366		
106	6.99	-5.511	6.13.391	6.65	808	1.788	5.248		
107	6.97	-5.569	6.13.285	6.67	98.7	1.793	5.228		
108	6.98	-5.557	6.13.354	6.66	121.4	1.734	5.177		
109	6.95	-5.029	6.09.577	6.59	909	1.733	4.989		
110	6.84	-4.742	6.08.446	6.57	91.2	1.775	4.89		
111	7	-5.96	6.13.767	6.69	107.1	1.715	5.08		
112	6.95	-5.209	6.13.925	6.59	95.2	1.766	4.999		
113	6.88	-5.024	6.3.944	6.58	111.5	1.714	4.828		
114	7	-5.584	6.13.986	6.67	109	1.835	5.41		
115	6.94	-5.18	6.09.159	6.62	96.4	1.751	5.014		
116	6.95	-4.682	6.08.76	6.53	97.4	1.781	4.992		
117	6.94	-4.937	6.09.529	6.59	83.4	1.776	4.995		
118	7.01	-5.921	6.13.835	6.72	808	1.825	5.458		
119	6.89	-5.075	6.09.153	6.62	111.9	1.737	4.983		
120	7.04	-5.874	6.13.341	6.7	18.4	1.759	5.32		
121	6.93	-4.946	6.09.675	6.6	134	1.668	4.895		
122	6.93	-5.395	6.13.335	6.64	110.9	1.781	5.207		
123	6.95	-5.636	6.12.848	6.68	12.4	1.823	5.387		
									<i>continua na próxima página.</i>

<i>continuação da página anterior.</i>									
PCB	log K_{ow}	log S (mg/L)	log K_{oc}	log H (Pa m ³ /mol)	-log Y_w	mp (°C)	log IC ₅₀ (μM)	log EC ₅₀ (μM)	
176	7.94	-7.282	6.93.627	7.35	134.2	1.63	4.778		
177	7.96	-7.396	6.93.429	7.38	134.1	1.616	4.844		
178	7.99	-7.838	6.93.867	7.41	141.3	1.604	4.847		
179	7.92	-7.36	6.93	4.008	7.34	13.9	1.649	4.752	
180	7.99	-7.873	6.93.911	7.46	136.5	1.572	4.859		
181	8.03	-7.592	6.93.174	7.41	18.1	1.585	4.885		
182	7.96	-7.41	6.93.179	7.4	13.7	1.572	4.815		
183	8	-7.759	6.93.059	7.44	136.9	1.604	4.917		
184	7.98	-7.562	6.93.313	7.4	136.2	1.582	4.755		
185	8.03	-7.596	6.93.536	7.39	140.3	1.632	4.967		
186	8.04	-7.233	6.93.803	7.36	13.4	1.634	4.836		
187	7.99	-7.567	6.93.437	7.41	130.3	1.614	4.848		
188	7.96	-7.204	6.93.712	7.35	19.3	1.631	4.798		
189	8.13	-8.709	6.92.834	7.56	154.8	1.598	5.331		
190	8.06	-7.871	6.92.993	7.46	140.3	1.59	5.031		
191	8.04	-7.885	6.93.002	7.47	19.1	1.514	4.9		
192	8.1	-7.998	6.93.468	7.45	19.9	1.528	4.962		
193	8.02	-7.857	6.93.357	7.45	12.3	1.562	4.913		
194	8.51	-9.138	7.32.881	7.89	161.4	1.47	4.757		
195	8.54	-8.899	7.33.018	7.84	367	1.499	4.844		
196	8.52	-8.88	7.33.027	7.85	966	1.495	4.811		
197	8.49	-8.716	7.33.166	7.81	157.3	1.509	4.684		
198	8.58	-9.053	7.33.454	7.84	171.3	1.519	4.925		
199	8.51	-8.613	7.33.581	7.78	164.1	1.55	4.804		
200	8.49	-8.588	7.33.549	7.79	160.6	1.551	4.759		
201	8.5	-8.801	7.33.4	7.82	161.3	1.518	4.764		
									<i>continua na próxima página.</i>

<i>continuação da página anterior.</i>									
PCB	log K_{OW}	log S (mg/L)	log K_{OC}	log H (Pa m ³ /mol)	-log Y_w	mp (°C)	log IC_{50} (μ M)	log EC_{50} (μ M)	
202	8.47	-8.786	7.33.92	7.78	160.8	1.555	4.691		
203	8.58	-9.076	7.33.026	7.87	160.1	1.526	4.924		
204	8.55	-8.754	7.33.277	7.81	167.3	1.515	4.813		
205	8.61	-9.325	7.33.96	7.91	13.9	1.445	4.902		
206	9.09	-10.327	7.73.978.3	191.4	1.419	4.823			
207	9.07	-10.036	7.73.115.25	187.3	1.439	4.735			
208	9.07	-10.046	7.73.492.23	18.5	1.469	4.781			
209	9.59	-11.264	8.13.05669	22.7	1.366	4.663			

Tabela C.4: Variáveis utilizadas nos modelos das moléculas de 4M e 5M. Valores de $\log \text{EC}_{50}$ previstos.

Molécula	Área ($\text{\AA}^2$)	Peso	η (eV)	BAC	Ss	log EC ₅₀ (μM)
1	388.584	356.488	4.217	30	54	-0.475
2	420.971	369.531	4.202	33	53	-0.183
3	422.598	383.514	4.302	36	60	1.3
4	395.275	355.504	4.228	30	52	-0.399
5	459.054	411.568	4.401	52	63	2.627
6	439.316	383.558	4.176	36	54	-0.105
7	477.438	431.602	4.322	18	62	-
8	456.537	433.593	4.358	56	67	2.324
9	476.718	447.62	4.313	59	69	2.326
10	514.149	495.664	4.324	34	77	4.84
11	567.919	509.691	4.791	45	78	-
12	532.038	530.109	4.459	45	81	-
13	547.011	564.554	4.595	58	85	7.882
14	522.426	513.654	4.462	45	85	6.596
15	521.075	513.654	4.405	45	85	6
16	545.782	525.69	4.246	50	82	4.473
17	469.801	447.62	4.293	69	69	1.596
18	490.633	461.647	4.317	76	70	2.056
19	513.095	475.674	4.288	79	72	2.261
20	493.162	461.647	4.345	72	70	2.536
21	529.963	509.691	4.367	45	79	5.305
22	440.414	397.541	4.253	45	61	0.902
23	461.716	451.511	4.522	72	84	4.812
24	502.125	459.612	4.289	25	71	4.109
25	451.182	413.54	4.393	48	65	2.713
26	536.726	491.591	4.325	74	72	3.296
27	475.203	411.612	4.163	52	57	-
28	437.964	383.558	4.24	34	54	0.585
29	380.606	341.477	4.137	27	50	-1.641
30	398.788	355.504	4.342	28	51	0.801
31	421.101	405.539	4.332	50	64	1.34
32	436.894	419.566	4.457	51	66	3.025
33	437.269	383.558	4.231	42	54	-

continua na próxima página.

<i>continuação da página anterior.</i>						
Molécula	Área ($\text{\AA}^2$)	Peso	η (eV)	BAC	Ss	log EC ₅₀ (μM)
34	439.623	448.432	4.412	42	55	-
35	415.464	434.405	4.402	39	54	-
36	476.69	512.494	4.499	69	70	-
37	448.549	399.557	4.075	49	58	-1.102
38	451.127	413.54	4.254	52	65	1.173
39	486.838	463.619	4.313	76	73	2.19
40	486.231	443.566	4.411	66	70	3.289
41	545.718	525.69	4.271	50	82	4.722
42	430.327	385.53	4.224	42	58	0.279
43	467.391	449.592	4.366	69	73	2.595
44	465.189	429.539	4.406	59	70	3.097
45	413.791	366.531	4.033	25	48	-2.094
46	417.687	380.514	4.245	28	55	0.546
47	392.121	352.504	4.091	22	48	-1.836
48	453.97	430.593	4.17	44	63	0.554
49	447.281	410.54	4.2	38	61	0.792
50	455.051	410.54	4.128	42	60	-0.007
51	423.586	382.53	4.06	36	53	-1.645
52	483.182	460.619	4.173	62	68	0.866
53	479.388	440.566	4.155	54	65	0.681
54	467.979	446.592	4.126	55	69	0.455
55	458.805	426.539	4.197	47	66	1.047

Bibliografia

- [1] Franke, R. *Theoretical Drug Design Methods*. Elsevier, (1984).
- [2] Benigni, R. and Richard, M. *METHODS: A Companion to Methods in Enzymology* **14**, 264–276 (1998).
- [3] Schultz, T. W. and Seward, J. R. *The Sci. Total Environment* **249**, 73–84 (2000).
- [4] Walpolex, C. S. J., Wigglerworth, R., Bevan, S., Campbell, E. A., Dray, A., James, I. F., Masdin, K. J., Perkins, M. N., and Winter, J. *J. Med. Chem.* **36**, 2381 (1993).
- [5] Gaudio, A., Richards, W. G., and Takahata, Y. *J. Mol. Graphics and Modelling* **18**, 33–41 (2000).
- [6] Kapur, S., Shusterman, A., Verma, R. P., Hansch, C., and Selassie, C. D. *Chemosphere* **41**, 1643–1649 (2000).
- [7] Cronin, M. and Schultz, T. *Ecotox. and Environmental Safety* **39**, 65–69 (1998).
- [8] Trohalaki, S., Gifford, E., and Pachter, R. *Computers and Chemistry* **24**, 421–427 (2000).
- [9] Dunn, W., Wold, S., Edlund, U., Helberg, S., and Gasteiger, J. *Quant. Struct. Act. Relat.* **3**, 131 (1984).
- [10] Cirino, L. R., Cirino, J. J. V., and Ferreira, M. M. C. *13th European Symp. on Quantitative Structure-Activity Relationships, Abstract Book* , 63 (2000).
- [11] Cirino, L. R., Ribeiro, F. A. L., and Ferreira, M. M. C. *23^a Reunião Anual da SBQ - Livro de Resumos MD-057* **2**, 57 (2000).
- [12] Nagata, N., Cirino, L. R., and Ferreira, M. M. C. *X Simpósio Brasileiro de Química Teórica - Livro de Resumos* , 63 (1999).
- [13] Subramanian, S., Ferreira, M. M. C., and Trisc, M. *Structural Chemistry* **9**, 47 (1998).
- [14] Mackay, D., Shiu, W. Y., and Ma, K. C. *Illustrated Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemical*, volume 1,2,3. Lewis Publishers, (1993).

- [15] Rachdawong, P. and Christensen, E. R. *Environ. Sci. Technol.* **31**(9), 2686 (1997).
- [16] Ballschmiter, K. and Zell, M. *Fresenius Z. Anal. Chem.* **20**, 302 (1980).
- [17] Hutzinger, O., Safe, S., and Zitko, V. *The Chemistry of PCBs*. CRC Press, (1974).
- [18] Giesy, J. P. and Kannan, K. *Crit. Rev. in Toxicol.* **28**(6), 511 (1998).
- [19] Mullin, M. D., Pochini, C. M., McCrindle, S., Romkes, M., Safe, S. H., and Safe, L. M. *Environ. Sci. Technol.* **18**(6), 468 (1984).
- [20] Shiu, W. Y. and Mackay, D. *J. Phys. Chem. Ref. Data* **15**(2), 911 (1986).
- [21] Dewar, M. J. S., Zoebish, E. G., Healy, E. F., and Stewart, J. J. P. *J. Am. Chem. Soc.* **107**, 3902 (1985).
- [22] Balaban, A. T. *Chemical Application of Graph Theory*. Academic, (1967).
- [23] Basak, S. C. *Med. Sci. Res.* **15**, 605 (1987).
- [24] Raychaudhury, C., Ray, S. K., Ghosh, J. J., Roy, A. B., and Basak, S. C. *J. Comput. Chem.* **5**, 581 (1984).
- [25] Beebe, K. R., Pell, P. J., and Seasholtz, M. B. *Chemometrics: A practical Guide*. Willey, (1998).
- [26] Geladi, P. and Kowalski, B. R. *Anal. Chem. Acta* **1**, 181 (1986).
- [27] Dewar, M. J. S. and Thiel, W. *J. Am. Chem. Soc.* **99**, 4899 (1977).
- [28] Stewart, J. J. P. *J. Comput. Chem.* **10**, 209, 221 (1989).
- [29] Dewar, M. J. S. and Storch, D. M. *J. Am. Chem. Soc.* **107**, 3898 (1985).
- [30] Dewar, M. J. S. and Thiel, W. *Theor. Chim. Acta (Berl.)* **46**, 89 (1977).
- [31] Oleari, L., DiSipio, L., and DeMichelis, G. *Mol. Phys.* **10**, 97 (1966).
- [32] Koopmans, T. A. *Physica* **1**, 104 (1933).
- [33] Shriver, D. F. *Inorganic Chemistry*. Oxford University Press, (1990).
- [34] Mulliken, R. S. *J. Chem. Phys.* **2**, 782 (1934).
- [35] Hirsch, J. A. *Concepts in Theoretical Organic Chemistry*. Allynand Bacon, Inc., (1975).
- [36] Chirlian, L. E. and Francl, M. M. *J. Comput. Chem.* **8**, 894 (1987).
- [37] Kier, L. B. and Hall, L. H. *Molecular Connectivity in Structure-Activity Analysis*. John Wiley & Sons, (1986).
- [38] Wiener, H. *J. Am. Chem. Soc.* **69**, 2636 (1947).

- [39] Randić, M. *J. Am. Chem. Soc.* **97**, 6609 (1975).
- [40] Balaban, A. T. *Chem. Phys. Lett.* **89**, 399 (1982).
- [41] Balaban, A. T. *Pure Appl. Chem.* **55**, 199 (1983).
- [42] Plavšić, D., Nikolić, S., Trinajstić, N., and Mihalić, Z. *J. Math. Chem* **12**, 235 (1993).
- [43] Kier, L. B. and Hall, L. H. *J. Math. Chem* **7**, 229 (1991).
- [44] Hall, L. H., Mohnen, B. K., and Kier, L. B. *J. Chem. Inf. Comput. Sci.* **31**, 76 (1991).
- [45] Huang, M., Doerge, D., Bodor, N., Pop, E., and Brewster, M. E. *Int. J. Quant. Chemistry* (22), 171 (1995).
- [46] Bodor, N. S. and Huang, M. *Int. J. Quant. Chemistry* (61), 127 (1997).
- [47] Golub, G. H. and van Loan, C. F. *Matrix Computation*. The John Hopkins University Press, (1989).
- [48] Ferreira, M. M. C., Antunes, A. M., Melgo, M. S., and Volpe, P. L. O. *Química Nova* **5**(22), 724 (1999).
- [49] Draper, N. R. and Smith, H. *Applied Regression Analysis*. John Wiley & Sons, (1981).
- [50] *DVIPS* – Radical Eye Software. (1997).
- [51] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Gill, P. M. W., Johnson, B. G., Robb, M. A., J. R. Cheeseman, T. K., Petersson, G. A., Montgomery, J. A., Raghavachari, K., Al-Laham, M. A., Zakrzewski, V. G., Ortiz, J. V., Foresman, J. B., Cioslowski, J., Stefanov, B. B., Nanayakkara, A., Challacombe, M., Peng, C. Y., Ayala, P. Y., Chen, W., Wong, M. W., Andres, J. L., Replogle, E. S., Gomperts, R., Martin, R. L., Fox, D. J., Binkley, J. S., Defrees, D. J., Baker, J., Stewart, J. J. P., Head-Gordon, M., Gonzalez, C., and Pople, J. A. *Gaussian 94, Revision D.2*. Gaussian, Inc, (1995).
- [52] Pascual-Ahuir, J. L., Silla, E., and Tunon, I. *GEPOL93 (GEometria POLihedro)*. Departamento de Química Física, Facultad de Química, Universidad de Valencia, Espanha, (1993).
- [53] Kimball, S. and Mattis, P. *THE GIMP* Version 1.0.
- [54] Williams, T. and Kelley, C. *GNUPLOT — An Interactive Plotting Program* Version 3.5. (1995).
- [55] Plass, J. *GV PostScript and PDF previewer* Version 3.5.8, www.thep.physik.uni-mainz.de.
- [56] Lamport, L. *L^AT_EX A Document Preparation System*. Addison-Wesley, (1995).
- [57] Schaftenaar, G. *Molden Version 3.2*. (1996).

- [58] C.Moler, Little, J., Kleiman, S., and Bangert, S. *MATLAB* Version 5.0. The MathWorks, Inc., (1997).
- [59] Stewart, J. J. P. *J. Comput.-Aided Mol. Des.* **4**(1), 1–105 (1990).
- [60] Eaton, J. W. *OCTAVE – A hight-level interactive language for numerical computation* Version 2.0.1. (1996).
- [61] Infometrix. *Multivariate Data Analysis for IBM PC Systems* Version 2.1.
- [62] *POLLY*. Natural Resources Research Insitute, University of Minnesota, Duluth.
- [63] Wise, B. M. and Gallagher, N. B. *PLS_Toolbox* Version 1.5. Eigenvector Technologies, (1995).
- [64] Anger, S., Bayer, D., Dailey, C. C. C., Dilger, A., Enzmann, S. D. A., Farmer, D., Wegner, T., Young, C., et al. *POV-Ray - Persistence of Vision Ray Tracer* Version 3.0. (1997).
- [65] Dehre, W. J., Burke, L. D., and Shusterman, A. J. *Spartan User's Guide*. Wavefunction, Inc., (1993).
- [66] Todeschini, R. and Gramatica, P. *Quant. Struct.-Act. Relat.* **16**, 113–119 (1997).
- [67] *DVI Previewer for the X Window System*. Version 22-08.
- [68] Sutanthavibul, S. *XFIG Facility for Interactive Generation of figures under X11*.
- [69] Bradley, J. *XV Interactive Image Display for the X Window System*. Version 3.10a, (1994).
- [70] Kurland, R. J. and Wise, W. B. *J. Am. Chem. Soc.* **86**, 1877 (1964).
- [71] Bastiansen, O. and Samdal, S. *J. Mol. Struct. (Theochem)* **128**, 115 (1985).
- [72] Brédas, J. L., Street, G. B., Thérman, B., and André, J. M. *J. Chem. Phys.* **83**, 1323 (1985).
- [73] Cioslowski, J. and Mixon, S. T. *J. Am. Chem. Soc.* **114**, 4382 (1992).
- [74] Häfelinger, G. and Regelman, C. *J. Comput. Chem.* **6**, 368 (1985).
- [75] Rubio, M., Merchán, M., and Ortí, E. *Theor. Chim. Acta* **91**, 17 (1995).
- [76] Almennigen, A., Bastiansen, O., Fernholt, L., Cyvin, B. C., Cyvin, S. J., and Samdal, S. *J. Mol. Struct. (Theochem)* **128**, 59 (1985).
- [77] Almlöf, J. *Chem. Phys* **6**, 135 (1974).
- [78] McKinney, J. D., Gottschalk, K. E., and Pedersen, L. *J. Mol. Struct. (Theochem)* **104**, 445 (1983).
- [79] Häfelinger, G. and Regelman, C. *J. Comput. Chem.* **8**, 1057 (1987).

- [80] Tsuzuki, S. and Tanabe, K. *J. Phys. Chem.* **95**, 139 (1991).
- [81] Mizukami, Y. *J. Mol. Struct. (Theochem)* **11**, 488 (1999).
- [82] Watari, H., Tanaka, M., and Suzuki, N. *Anal. Chem.* **54**, 702 (1982).
- [83] Dearden, J. C. *Environ. Sci. Technol.* **61**, 203 (1985).
- [84] Mallholt, H. and Peters, R. H. *Environ. Sci. Technol.* **22**, 1479 (1988).
- [85] Rapaport, R. A. and Eisenreich, S. J. *Environ. Sci. Technol.* **18**, 163 (1984).
- [86] Miller, M. M., Ghodbane, S., Wasik, S. P., Tewari, Y. B., and Martire, D. E. *J. Chem. Eng. Data* **29**, 184 (1984).
- [87] Mackay, D., Shiu, W. Y., and Ma, K. C. *Illustrated Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemical*, volume 3. Lewis Publishers, (1993).
- [88] Ghose, A. K., Pritchett, A., and Crippen, M. *J. Comput. Chem.* **9**, 80 (1988).
- [89] Cramer, C. J., Famini, G. R., and Lowrey, A. H. *Acc. Chem. Res.* **26**, 599 (1993).
- [90] Hansen, B. G., Paya-Perez, A. B., Rahman, M., and Larsen, B. R. *Chemosphere* **39**, 2209 (1999).
- [91] Gramatica, P., Navas, N., and Todeschini, R. *Chemometrics* (40), 53 (1998).
- [92] Pearlman, R. S., Yalkowsky, S. H., and Banerjee, S. *J. Phys. Chem. Ref. Data* (13), 555 (1984).
- [93] Mackay, D., Mascarenhas, R., Shiu, W. Y., Valvani, S. C., and Yalkowsky, S. *Chemosphere* **9**, 274 (1980).
- [94] Nirmalakhanden, N. N. and Speece, R. E. *Environ. Sci. Technol.* **26**(6), 708 (1989).
- [95] Karickhoff, S., Brown, D., and Scott, T. *Water Res.* **13**, 241 (1979).
- [96] Nirmalakhandan, N. N. and Speece, R. E. *Environ. Sci. Technol.* **22**(11), 1349 (1988).
- [97] L. P. Burkhard, A. W. A. and Armstrong, D. E. *Environ. Sci. Technol.* **19**(6), 500 (1985).
- [98] Nirmalakhanden, N. N. and Speece, R. E. *Environ. Sci. Technol.* **22**(11), 1349 (1988).
- [99] Myrdal, P., Ward, G. H., Dannenfelser, R. M., Mishra, D., and Yalkowsky, S. H. *Chemosphere* **24**, 1047 (1992).
- [100] Dearden, J. C. *Chemom. Intell. Lab. Syst.* **24**, 77 (1994).
- [101] Coglian, J. *PCBs: Cancer Dose-Response Assessment and Application to Environmental Mixtures*. National Center for Environmental Assessment Office of Research and Development—U.S. Environmental Protection Agency, (1996).

- [102] Kodavanti, P. R. S., Ward, T. R., McKinney, J. D., and Tilson, H. *Arch. Toxicol.* **70**, 150 (1996).
- [103] Albert, A. *Selective Toxicity*. Methuen, (1951).
- [104] Goodwin, S., Smith, A. F., and Horning, E. C. *J. Am. Chem. Soc.* **81**, 1903 (1959).
- [105] Sizum, P., Auclair, C., Lescot, E., Paoletti, C., Perly, B., and Fermandjiani, S. *Biopolymers* **27**, 1096 (1988).
- [106] Perni, R. B., Wentland, M. P., Huang, J. I., Powles, R. G., Aldous, S., Klingbeil, K. M., Peverly, A. D., Robison, R. G., Corbett, T. H., Jones, J. L., Mattes, K. C., Rake, J. B., and Coughlin, S. A. *J. Med. Chem* **41**, 3645–3654 (1998).
- [107] Wei, C. H. and Einstein, J. R. *Acta Cryst. Sec. B* **34**, 205 (1978).
- [108] Obtido na base de dados, código NDB: DDF079.
- [109] Leonard, G. A., Hambley, T. W., McAuley-Hecht, K., Brown, T., and Hunter, W. N. *Acta Cryst. Sec. D* **49**, 458 (1993).
- [110] Ferenczy, G. G., Reynolds, C. A., and Richards, W. G. *J. Comput. Chem.* **11**, 159 (1990).
- [111] Luque, F. J., Illas, F., and Orozco, M. *J. Comput. Chem.* **11**, 416 (1990).
- [112] Courseille, C., Busetta, B., and Hospital, H. *Acta Cryst. Sec. B* **30**, 2628 (1974).
- [113] Atkins, P. W. *Physical Chemistry*. Oxford University Press, fourth edition, (1990).
- [114] Levine, I. N. *Quantum Chemistry*. Prentice Hall, fourth edition, (1991).
- [115] Szabó, G. N., Surján, P. R., and Ángyán, J. G. *Applied Quantum Chemistry*. Reidel, first edition, (1987).
- [116] Hehre, W. J., Ditchfield, R., and Pople, J. A. *J. Chem. Phys.* **56**(5), 2257 (1972).
- [117] Krishnan, R., Binkley, J. S., and Pople, J. A. *J. Chem. Phys.* **72**(1), 650 (1980).
- [118] MacLean, A. D. and Chandler, G. S. *J. Chem. Phys.* **72**, 5639 (1980).