

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS FÍSICAS E MATEMÁTICAS
CURSO DE PÓS-GRADUAÇÃO EM QUÍMICA**

TESE

**ÍNDICE SEMI-EMPÍRICO ELETROTOPOLÓGICO: UM NOVO
DESCRITOR MOLECULAR USADO EM ESTUDOS DE
CORRELAÇÃO QUANTITATIVA ENTRE ESTRUTURA E
RETENÇÃO CROMATOGRÁFICA PARA COMPOSTOS DE
INTERESSE**

ÉRICA SILVA SOUZA

**FLORIANÓPOLIS
2009**

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO DE CIÊNCIAS FÍSICAS E MATEMÁTICAS
CURSO DE PÓS-GRADUAÇÃO EM QUÍMICA**

TESE

**ÍNDICE SEMI-EMPÍRICO ELETROTOPOLÓGICO: UM NOVO
DESCRITOR MOLECULAR USADO EM ESTUDOS DE
CORRELAÇÃO QUANTITATIVA ENTRE ESTRUTURA E
RETENÇÃO CROMATOGRÁFICA PARA COMPOSTOS DE
INTERESSE**

Tese apresentada ao Curso de Pós-Graduação em Química, do Departamento de Química da Universidade Federal de Santa Catarina, como requisito para obtenção do grau de Doutor(a) em Química.

Área de Concentração: Química Analítica

Orientadora: Prof^a. Dra. Vilma Edite Fonseca Heinzen

Co-orientador: Prof. Dr. Rosendo Augusto Yunes

Doutoranda: Érica Silva Souza

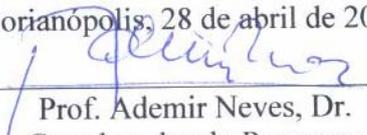
**FLORIANÓPOLIS
2009**

Érica Silva Souza

**ÍNDICE SEMI-EMPÍRICO ELETROTOPOLÓGICO: UM NOVO
DESCRITOR MOLECULAR USADO EM ESTUDOS DE CORRELAÇÃO
QUANTITATIVA ENTRE ESTRUTURA E RETENÇÃO
CROMATOGRÁFICA PARA COMPOSTOS DE INTERESSE**

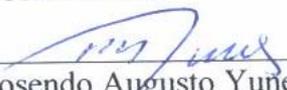
Esta tese foi julgada e aprovada para a obtenção do título de Doutora em Química -
Área Química Analítica - no Programa de Pós-Graduação em Química da
Universidade Federal de Santa Catarina.

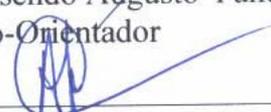
Florianópolis, 28 de abril de 2009.

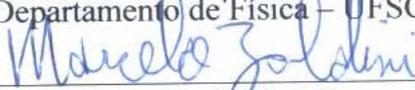

Prof. Ademir Neves, Dr.
Coordenador do Programa

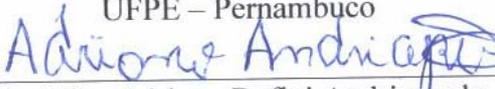
BANCA EXAMINADORA

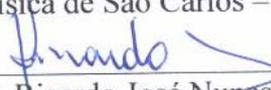

Prof. Dra. Vilma Edite Fonseca Heinzen
Orientadora

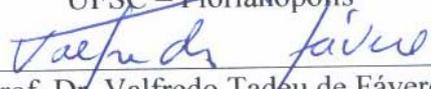

Prof. Dr. Rosendo Augusto Yunes
Co-Orientador

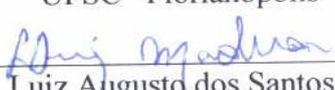

Prof. Dr. Carlos Alberto Kuhnen
Departamento de Física - UFSC


Prof. Dr. Marcelo Zaldini Hernandez
UFPE - Pernambuco


Prof. Dr. Adriano Defini Andricopulo
Instituto de Física de São Carlos - USP


Prof. Dr. Ricardo José Nunes
UFSC - Florianópolis


Prof. Dr. Valfredo Tadeu de Fávère
UFSC - Florianópolis


Prof. Dr. Luiz Augusto dos Santos Madureira
UFSC - Florianópolis

Às pessoas que acreditam em um sonho que tenham coragem, paciência e sabedoria.

AGRADECIMENTOS

Aos orientadores, Profa. Dra. Vilma Edite Fonseca Heinzen, Prof. Dr. Rosendo Augusto Yunes e Prof. Dr. Carlos Alberto Kuhnen, pela orientação imprescindível, pelo exemplo de profissionais dignos e apaixonados pela ciência e, principalmente, por acreditarem no sucesso deste trabalho e na minha capacidade de enfrentar obstáculos.

Aos meus familiares, Fátima, Nila (*in memoriam*) e Jocemar, que apesar da distância, sempre me incentivaram em todos os momentos da minha vida e desejaram a minha felicidade.

A todos os meus amigos, alguns de longa data, outros nem tanto, pela companhia e amizade nas horas de lazer, para alguns pela cumplicidade quase que fraternal diária e para outros pela motivação e vivacidade sendo exemplos de perseverança.

Aos colegas de laboratório Anna Célia, Berenice e Ledilege pela colaboração na realização deste trabalho.

Aos Professores do Departamento de Química da Universidade Federal de Santa Catarina que contribuíram para a construção de todo um aprendizado, o qual aguçou a minha criatividade e possibilitou o meu avanço tanto acadêmico quanto pessoal.

Aos funcionários, em particular a Graça e o Jadir, pela presteza, profissionalismo e gentileza a mim dispensada em mais de uma década durante todo período desde a Graduação até o Doutorado.

Ao Departamento de Química-UFSC.

Ao Programa de Pós-Graduação em Química da UFSC.

Ao CNPq pelo suporte financeiro.

“Enquando houver alguém que acredite em uma idéia, a idéia viverá.”
José Ortega y Gasset

Sumário

	Página
Resumo	ix
Abstract	xi
Lista de Figuras	xiii
Lista de Tabelas	xiv
Lista de Abreviaturas	xv
1. Introdução.....	1
2. Revisão de literatura.....	3
2.1 Correlação quantitativa entre estrutura molecular e propriedade – QSPR / QSAR / QSRR.....	3
2.1.1 Modelo de correlação entre estrutura-propriedade.....	4
2.2. Propriedades moleculares.....	7
2.2.1. Retenção cromatográfica.....	8
2.2.2. Interações intermoleculares.....	9
2.3. Descritores moleculares.....	11
2.3.1. Descritores topológicos.....	12
2.3.1.1 Teoria dos grafos moleculares.....	15
2.3.2. Descritores químico-quânticos.....	16
2.3.3.1 Mecânica quântica e o princípio variacional.....	17
2.3.3.2. Método variacional linear.....	21
2.4. Índice semi-empírico topológico (I_{ET}).....	24
2.5. Classes de compostos estudadas.....	27
2.5.1. Alcanos e alcenos.....	27
2.5.2. Aldeídos e cetonas.....	29
2.5.3. Ésteres.....	30
3. Objetivos.....	31
4. Metodologia.....	32
4.1. Índices de retenção cromatográficos.....	32
4.1.1. Alcanos e alcenos.....	32
4.1.2. Aldeídos e cetonas.....	32
4.1.3. Ésteres alifáticos.....	32
4.2. Cálculo do índice semi-empírico eletrotológico (I_{SET}).....	32
4.2.1. Cálculo da carga atômica líquida.....	33

4.2.2. Alcanos e alcenos.....	34
4.2.3. Aldeídos e cetonas.....	35
4.2.4. Ésteres alifáticos.....	38
4.3. Análise de regressão e validação dos modelos de QSPR / QSRR.....	41
4.3.1. Validação cruzada.....	42
4.3.2. Validação externa dos modelos.....	43
5. Resultados e discussões.....	44
5.1.1. Aplicação do I_{SET} para estimar a retenção cromatográfica de alcanos e alcenos.....	44
5.2. Aplicação do I_{SET} para estimar a retenção cromatográfica de aldeídos e cetonas.....	49
5.2.1. Aplicação do I_{SET} para estimar a retenção cromatográfica de ésteres.....	55
6. Conclusões.....	65
6.1. Recomendações para trabalhos futuros.....	68
7. Referências bibliográficas.....	69
8. Anexos.....	90

Resumo

Estudos quantitativos entre estrutura e retenção cromatográfica vêm sendo desenvolvidos nas últimas três décadas com base em vários tipos de descritores topológicos (como índices de conectividade) e descritores quânticos. Em meados da década de 90 foi proposto o índice semi-empírico topológico (I_{ET}), utilizado inicialmente para alcanos e alcenos sendo posteriormente estendido para outros compostos orgânicos. O cálculo do I_{ET} é efetivado por meio de valores tabelados atribuídos aos carbonos primários, secundários, terciários e quaternários da molécula. Neste trabalho, um novo índice é proposto, baseado na observação de que a retenção cromatográfica de alcanos e alcenos depende fortemente da carga atômica líquida dos átomos de carbono dessas moléculas, o que determina uma boa correlação entre as cargas atômicas líquidas e os valores atribuídos a cada átomo de carbono no cálculo do I_{ET} . Esta correlação permite a definição de um índice, denominado semi-empírico eletrotológico (I_{SET}) que depende dos valores de carga atômica líquida, calculadas por meio de métodos semi-empíricos de química quântica. Inicialmente as geometrias moleculares são obtidas através dos cálculos de Mecânica Molecular (MM+) e posteriormente otimizadas utilizando o método semi-empírico AM1 com o algoritmo de Polak-Ribiere e técnicas de minimização de gradiente nos quais os limites de convergência são 0,0001 e gradiente RMS 0,0001 kcal (A.mol)⁻¹. A carga atômica líquida é obtida pela análise de população de Mulliken. Este novo índice é aplicado inicialmente para os hidrocarbonetos alifáticos (alcanos e alcenos) e estendido para aldeídos, cetonas e ésteres. Para estes compostos orgânicos a definição de I_{SET} é generalizada de maneira a incluir o momento de dipolo molecular e as cargas líquidas nos átomos pertencentes aos respectivos grupos funcionais, as quais são explicitadas via definição de um momento de dipolo local equivalente.

A correlação linear simples entre os índices de retenção cromatográficos e os valores dos índices semi-empíricos eletrotológicos para alcanos e alcenos é de boa qualidade e estabilidade interna de acordo com os parâmetros estatísticos: $r=0,9990$; $s=10,74$; $N=179$. A validade externa do modelo é testada subdividindo o grupo em três subgrupos e sendo cada um deles predito usando os outros dois subgrupos como conjunto de treinamento. Os resultados obtidos para os grupos de treinamento ($r^2=0,9980$) e predição ($r^2=0,9980$), indicam que o modelo proposto tem alta estabilidade estatística e validade.

Os modelos de QSRR para aldeídos e cetonas usando o I_{SET} são também de boa qualidade como mostrado pelos seguintes parâmetros estatísticos: $r = 0,9994$; $s = 10,31$; $N = 15$ para aldeídos e $r = 0,9997$; $s = 11,72$; $N = 40$ para cetonas. Este modelo também mostra uma boa estabilidade externa com $r^2 = 0,9978$ para os subgrupos de treinamento e com $r^2 = 0,9977$ para os subgrupos de predição.

Modelos de QSRR utilizando I_{SET} também são obtidos para descrever a retenção cromatográfica de 100 ésteres alifáticos em diferentes fases estacionárias (SE-30, OV-7, DC-710, OV-225 e XE-60) com parâmetros estatísticos variando entre: $r=0,9990$ a $0,9959$; e $s=7,65$ a $16,02$. Regressões lineares simples entre o índice de retenção e I_{SET} são de boa qualidade estatística e grande estabilidade interna apresentando ótima capacidade de predição para um grupo externo, especialmente para fases estacionárias de baixa polaridade, mostrando que interações moleculares específicas ocorrem em fases estacionárias mais polares. Um modelo combinado para a predição do índice de retenção usando o índice proposto (I_{SET}) e a polaridade de McReynolds das fases estacionárias é de boa qualidade estatística conforme observado pelos parâmetros de mérito da regressão entre os valores de retenção experimentais versus os calculados: $r = 0,9978$; $s = 12,94$; $N = 500$.

Este novo índice, desenvolvido para alcanos e alcenos, e generalizado para cetonas, aldeídos e ésteres (moléculas polares) permite obter índices de retenção com alta estabilidade e validade estatística por regressão linear simples. Este índice usa um descritor que codifica tanto a distribuição de carga quanto detalhes estruturais, relacionados ao tamanho, ramificação e presença de heteroátomos. Abrindo assim, novas possibilidades para estudos cromatográficos e de outras propriedades moleculares de diferentes funções orgânicas.

PALAVRAS-CHAVE: Retenção Cromatográfica; Descritores Topológicos; QSPR; QSRR; Índice Semi-empírico Eletrotológico; Alcanos; Alcenos; Aldeídos; Cetonas; Ésteres.

Abstract

Quantitative studies between molecular structure and chromatographic retention has been developed in the last three decades based on many types of topological descriptors (as connectivity indices) as well as quantum descriptors. In the middle of the 90th decade a semi-empirical topological index (I_{ET}) was proposed and initially applied for alkanes and alkenes; then it was extended for other classes of organic compounds. The semi-empirical topological index is calculated by means of tabulated values attributed to the primary, secondary, tertiary and quaternary carbon atoms of the molecule. In this work, a new index is proposed based on the fact that the chromatographic retention of alkanes and alkenes strongly depends on the net charge of the carbon atoms of these molecules, which implies a good correlation between the net atomic charge and the values attributed to each carbon atom in the I_{ET} calculation. This correlation allows the definition an index, the semi-empirical electrotopological index (I_{SET}), that depends on the net atomic charges which are calculated through semi-empirical quantum chemical methods. The molecular geometries are first obtained through molecular mechanics (MM+) calculations and subsequently optimized with AM1 method using the Polak-Ribierre algorithm and minimization gradient techniques where the convergence limits are 0,0001 and RMS gradient of 0,0001 kcal (A.mol)⁻¹. The net atomic charge is obtained using the Mulliken population analysis. This new index is initially applied for aliphatic hydrocarbons (alkanes and alkenes) and extended for aldehydes, ketones and esters. For these organic compounds the definition of I_{SET} is generalized to include the molecular dipole moment and the net charges of the atoms belonging to the respective functional group which appears through a definition of an equivalent local dipole moment.

A simple linear regression between the chromatographic retention indices and the values of semi-empirical electrotopological indices for alkanes and alkenes is of good quality and internal stability according the statistical parameters: $r=0.9990$; $s=10.74$; $N=179$. The external validity of the model is verified dividing the group into three subgroups with each group being predicted using the two subgroups as training set. The results obtained for the training ($r^2=0.9980$) and prediction ($r^2=0.9980$) sets indicate that the proposed model has high statistical stability and validity.

The QSRR models for aldehydes and ketones using the I_{SET} are of good stability too, as shown by the following statistical parameters: $r = 0.9994$; $s = 10.31$; N

= 15 for aldehydes and $r = 0.9997$; $s = 11.72$; $N = 42$ for ketones. This model also shows also a good external stability with $r^2 = 0.9978$ for the training subgroups and with $r^2 = 0.9977$ for the prediction subgroups.

QSRR models by means of I_{SET} are also obtained to describe the chromatographic retention of 100 aliphatic esters in different stationary phases (SE-30, OV-7, DC-710, OV-225 and XE-60) with statistical parameters between: $r = 0.9990$ to 0.9959 ; and $s = 7.65$ to 16.02 . Single linear regressions between retention index and the I_{SET} are of good statistical quality and of great internal stability. This reveals optimal predictive capacity for an external group especially for stationary phases of low polarity showing that specific molecular interactions occur in stationary phases with high polarity. A combined model to predict the retention index using the proposed index (I_{SET}) and the McReynolds polarity of the stationary phases is of good statistical quality as observed from the obtained merit parameters of the single linear regression between the experimental retention values versus calculated indices: $r = 0.9978$; $s = 12.94$; $N = 500$.

This new index, developed for alkanes and alkenes and generalized for ketones, aldehydes and esters (polar molecules) allows obtaining retention indices with high stability and statistical validity by single linear regression. This index uses a descriptor that codifies both the charge distribution and structural details related to the size, branches and presence of heteroatoms. Thus, the present method opens new possibilities for studies of chromatography and other molecular properties of various organic functions.

KEYWORDS: Chromatographic retention; Topological Descriptors; QSPR; QSRR; Semi-empirical Electrotopological Index; Alkanes; Alkenes; Aldehydes; Ketones; Esters.

Lista de Figuras

	Página
Figura 1: Diagrama de fluxo das etapas envolvidas num modelo de QSPR.....	6
Figura 2: A molécula do 3-metilpentano como fórmula estrutural (a) e grafo molecular enumerado (b) onde foram omitidos os átomos de hidrogênio.....	15
Figura 3: Regressão linear simples entre os índices de retenção experimentais (usando esqualano como fase estacionária) e número de átomos de carbono para alcanos lineares e ramificados.....	25
Figura 4: Regressão linear simples entre os valores de SET_i e as cargas atômicas líquidas (δ_i).....	35
Figura 5: Índices de retenção experimentais (IR_{Exp}) versus índices de retenção calculados (IR_{Calc}) para os dados de 179 alcanos e alcenos.....	48
Figura 6: Índices de retenção experimentais (IR_{Exp}) versus I_{SET} para 55 aldeídos e cetonas.....	53
Figura 7: Índices de retenção experimentais (IR_{Exp}) versus índice de retenção calculados (IR_{Calc}) para 40 cetonas e 15 aldeídos.....	53
Figura 8: Valores residuais versus índices retenção calculados (IR_{Calc}) para 40 cetonas e 15 aldeídos.....	54
Figura 9: Regressão linear simples entre os índices de retenção (IR_{Exp}) e os índices semi-empírico eletrotopológicos (I_{SET}) para ésteres nas fases estacionárias SE-30, OV-7, DC-710, XE-60 e OV-225.....	60
Figura 10: Correlação entre o intercepto (b_0) e a polaridade de McReynolds das fases estacionárias (Mp) dos modelos QSRR obtidos usando I_{SET}	62
Figura 11: (a) Índices de retenção experimentais (IR_{Exp}) versus calculados (IR_{Calc}) para 500 dados. (b) Gráfico dos valores de resíduos versus os índices de retenção calculados (IR_{Calc}) dos ésteres em cinco fases estacionária com diferentes polaridades.....	63

Lista de Tabelas

	Página
Tabela 1: Lista das características desejadas para os índices topológicos proposta por Randic citada no artigo publicado por Mihalic e Trinajstic.....	14
Tabela 2: Valores de C_i e log de C_i para os átomos de carbono primário, secundário, terciário e quaternário em alcanos.....	25
Tabela 3: Os valores de carga atômica líquida (δ_i) e SET_i para cada átomo de carbono das moléculas <i>cis-penten-2-eno</i> e <i>trans-penten-2-eno</i>	44
Tabela 4: Valores dos índices de retenção experimentais, IR_{Exp} , e dos índices semi-empírico eletrotológico, I_{SET} , para diferentes alcanos e alcenos.....	46
Tabela 5: Verificação da validade estatística do modelo de QSRR obtido para alcanos e alcenos.....	49
Tabela 6: Os valores de carga atômica líquida (δ_i) e SET_i para cada átomo de carbono e oxigênio das moléculas 3-hexanona e hexanal.....	50
Tabela 7: Valores experimentais do índice de retenção (IR_{Exp}) (HP-1 50C°) e I_{SET} para diferentes aldeídos e cetonas.....	52
Tabela 8: Verificação da validação estatística do modelo (N=55).....	54
Tabela 9: Valores de cargas atômicas líquidas e SET_i para cada átomo de carbono e oxigênio das moléculas de formiato de propila e isobutirato de isopropila.....	55
Tabela 10: Valores dos índices de retenção de cinco fases estacionárias (SE-30, OV-7, DC-710, XE-60 e OV-225) e dos índices semi-empíricos eletrotológicos (I_{SET}).....	58
Tabela 11: Resultados estatísticos para os modelos de QSRR e as constantes de McReynolds de cinco fases estacionárias.....	60
Tabela 12: Índices de retenção experimentais e predito nas fases estacionárias SE-30 e OV-7 para o grupo externo (ésteres).....	64

Lista de Abreviaturas

α : polarizabilidade molecular

a_i : coeficientes da expansão

AM1: modelo Austin Model 1

b_0 : intercepto

b_1 : inclinação

CG: cromatografia gasosa

CG-EM: cromatografia gasosa acoplada à espectrômetria de massa

CGL: cromatografia de partição gás-líquido

C_{Ac} : átomo de carbono no lado ácido do grupo funcional COO

c_i e \hat{c}_i : valores experimentais e previstos, respectivamente

\bar{c}_i : valor médio de c_i

C_i : valor atribuído a cada átomo de carbono e/ou grupo funcional i da molécula

ce : correção estérea

C_j : valor atribuído a cada átomo de carbono e/ou grupo funcional adjacente a i

CoMFA: Análise Comparativa do Campo Molecular do inglês *Comparative Molecular Field Analysis*

d : comprimento da ligação

δ_A : carga líquida do átomo A

δT : solubilidade total

ε : permissividade relativa do meio

ε_0 : permissividade do vácuo ($8,854 \cdot 10^{-12} F / m$ (ou $C^2 / N \cdot m^2$))

E_o : energia mínima do sistema

f: teste de hipótese nula (F -test)

Φ : função de onda tentativa (proposta)

φ_i : função de onda conhecida

h : constante de Planck

H : operador Hamiltoniano

H_{ij} : integral de ressonância

HCA: análise hierárquica de agrupamentos

I : energia de ionização

I.D.: diâmetro interno

I_{ET} : Índice Semi-empírico Topológico

I_{SET} : Índice Semi-empírico Eletrotológico

IR_A : Índice de retenção de Kováts de um composto A

IR_{Exp} : Índice de retenção experimental

IR_{Calc} : Índice de retenção calculado

J: Índice de Balaban

k_B : constante de Boltzmann ($1,38 \cdot 10^{-23} J / K$)

KNN: K-ésimo vizinho mais próximo (do inglês *K-nearest neighbors algorithm*)

log P: coeficiente de partição octanol/água

m_e : massa do elétron

m_k : massa do núcleo

mK : Índice Kappa

MLR: Regressão Linear Múltipla (do inglês *Multiple Linear Regression*)

MNDO: Negligência Intermediária Modificada do Recobrimento Diferencial (do inglês *Modified Neglected Differential Overlap*)

Mp: constantes de McReynolds

MTI: Índice de Schultz

M_w : massa molecular

OEI: índice par/ímpar (do inglês *odd-even index*)

PCA: Análise de Componentes Principais (do inglês *Principal component analysis*)

PCR: Regressão por Componentes Principais (do inglês *Principal Component Regression*)

PEI: Índice de Efeito da Polarizabilidade (do inglês *polarizability effect index*)

PLS: Regressão por Mínimos Quadrados Parciais (do inglês *Partial Least Square Regression*)

PM3 e PM5: Métodos de Parametrizações 3 e 5 (do inglês *Parameterized Model number 3 and 5*)

Q_A : carga calculada no átomo ligado A

QSAR: Correlação Quantitativa entre Estrutura e Atividade (do inglês *Quantitative Structure-Activity Relationships*)

QSPR: Correlação Quantitativa entre Estrutura e Propriedade (do inglês *Quantitative Structure-Property Relationships*)

QSRR: Correlação Quantitativa entre Estrutura e Retenção Cromatográfica (do inglês *Quantitative Structure-Retention Relationships*)

r_i : distância entre os centros interagentes

r_{ab} : distância entre as partículas a e b

r : coeficiente de correlação

r^2 : coeficiente de determinação

r^2_{CV} : coeficiente de validação cruzada

RLM: Regressão Linear Múltipla (do inglês *Multiple Linear Regression*)

RMS: Raiz Quadrática Média dos Valores (do inglês *Root Mean Square*)

s_p : coordenada de spin da partícula

s : desvio padrão

sce : segunda correção estérea

S_{ij} : integral de sobreposição (*overlap*)

SET_i e SET_j : valores dos fragmentos originados das cargas atômicas líquidas de cada átomo de carbono i e os j carbonos adjacentes a i

SVij: índice do efeito estéreo (do inglês *steric effect index*)

S_W : área superficial acessível ao solvente

T : temperatura em Kelvin (K)

θ : ângulo entre ligações

T_{Bp} : temperatura de ebulição

$t_{R(A)}$: tempo de retenção ajustado do composto A

$t_{R(N)}$: tempo de retenção ajustado do composto padrão homólogo com um número N de carbonos

$t_{R(N+1)}$: tempo de retenção ajustado do composto padrão homólogo com um número $(N + 1)$ de carbonos

μ : momento de dipolo

μ_F : momento de dipolo equivalente local

v : volume molar

V_i : volume de integração

V : energia potencial

∇^2 : operador Laplaciano

W: índice de Wiener

Ψ : função de onda

$|\Psi|^2$: densidade de probabilidade

Z : número atômico

Z_A : carga eletrônica do átomo isolado A

1. Introdução

Os estudos de QSPR representam um importante papel na química moderna e possibilitam prognosticar uma propriedade de interesse num determinado composto, antes mesmo dele ser sintetizado, e conseqüentemente vem atraindo os químicos durante várias gerações. Com o desenvolvimento de modelos obtidos através de QSPR, utilizando descritores moleculares, tal possibilidade passou a tornar-se uma realidade, e tem sido empregada com sucesso para uma ampla e variada faixa de aplicações¹⁻⁷.

Uma vez estabelecida uma correlação entre estrutura molecular e propriedade, compostos que apresentam uma propriedade de interesse, incluindo aqueles ainda não sintetizados, podem ser rapidamente prognosticados. Assim, torna-se possível selecionar compostos mais promissores para sintetizar e testar em laboratório. Os métodos QSPR evitam que recursos financeiros sejam desperdiçados em longos e laboriosos testes, acelerando o desenvolvimento da síntese de novas moléculas que podem ser utilizadas para os mais diversos fins, justificando o contínuo desenvolvimento de estudos sobre este tema. Isto é especialmente importante na área de procura de novos fármacos mais potentes e seletivos para o tratamento de diversas doenças que afetam o homem¹⁻².

Os trabalhos iniciais acerca de correlação quantitativa entre estrutura molecular e retenção cromatográfica (QSRR) foram realizados por Martin (*apud* Kaliszan, 1987) em 1949³ onde foi verificada a existência de uma correlação entre os parâmetros cromatográficos e o número de átomos de carbono para séries homólogas. Os trabalhos realizados por Heinzen e colaboradores (1992 e 1993)⁸⁻⁹ iniciaram com o emprego dos índices de conectividade molecular desenvolvidos em 1977 por Kier e Hall (1976, 1977 e 2001)¹⁰⁻¹², até a posterior formulação do índice denominado I_{ET} aplicado para alcanos e alcenos lineares e ramificados etc (1999)¹³.

O I_{ET} se fundamenta no fato de que a retenção cromatográfica se deve a interação de cada átomo da molécula com a fase estacionária. Esta interação, no entanto, depende da presença de outros átomos vizinhos na molécula que produzem um efeito estéreo entre si e com a fase estacionária. Desta forma, tanto os fatores eletrostáticos bem como os estéreos estão envolvidos neste fenômeno, e são representados por valores atribuídos para cada fragmento da molécula constituindo uma tabela a qual deve ser utilizada para o cálculo do I_{ET} . O índice semi-empírico topológico foi ampliado para compostos cíclicos, alicíclicos e aromáticos representando uma importante etapa na direção de estudos futuros de correlação quantitativa entre estrutura e atividade (QSAR), já que muitos dos compostos

biologicamente ativos apresentam em suas estruturas moleculares anéis aromáticos¹⁴⁻¹⁸. Esta mesma metodologia foi aplicada, com sucesso, para o prognóstico de outras propriedades físico-químicas e biológicas de interesse (ponto de ebulição, coeficiente de partição, solubilidade, etc.), baseada no sucesso já obtido e reportado previamente¹⁹⁻²⁷. Uma limitação deste índice é o fato de que, para cada função orgânica, existem tabelas específicas, as quais se tornam mais complexas à medida que se têm diferentes átomos na constituição da molécula.

Na primeira parte desta tese, será apresentada uma revisão da literatura abordando os principais conceitos inerentes ao tema, assim como diversos artigos, com o objetivo de contribuir para que o leitor avalie e compreenda a importância do presente estudo para a química e áreas afins. Em seguida, serão descritos os objetivos gerais e específicos a serem alcançados com o desenvolvimento deste trabalho de tese. Posteriormente, a metodologia utilizada no desenvolvimento de um novo descritor molecular, o índice semi-empírico eletrotológico (I_{SET}), será detalhada para alcanos e alcenos, aldeídos e cetonas, e ésteres. Os resultados e as discussões serão apresentados aplicando o I_{SET} para prognosticar a retenção cromatográfica das funções anteriores. Por fim, apresentaremos as conclusões deste trabalho de doutorado e recomendações para trabalhos futuros.

2. Revisão de literatura

2.1. Correlação quantitativa entre estrutura molecular e propriedade – QSPR / QSAR / QSRR

A sigla QSPR (do inglês *Quantitative Structure-Property Relationships*) pode ser considerada como a mais abrangente para designar as diferentes conexões existentes entre a estrutura e as propriedades moleculares. Sendo assim, para correlações quantitativas entre estrutura química e atividade biológica empregamos QSAR (do inglês *Quantitative Structure-Activity Relationships*), ou ainda, para correlações entre estrutura química e retenção cromatográfica, QSRR (do inglês *Quantitative Structure-Retention Relationships*)²⁸.

A percepção das correlações quantitativas existentes entre estrutura química quer nos aspectos estéreos, eletrônicos ou topológicos e propriedades é conhecida há muito tempo na Química. Isto se baseia no fato de que a menor variação da estrutura de uma molécula deve conter as informações responsáveis pelas mudanças nas suas propriedades físicas, químicas ou biológicas. Presume-se então, que um comportamento uniforme pode ser gerado para todos os compostos dentro de uma mesma classe. É intuitivo esperar que compostos com propriedades físico-químicas semelhantes, também tenham estruturas semelhantes²⁹⁻³⁶.

As relações entre estrutura e propriedade são estabelecidas através de modelos matemáticos e alguns aspectos teóricos e práticos devem ser observados para a proposição, validação e análise dos mesmos³⁷⁻³⁸. Para a construção de tais modelos matemáticos, a estrutura química deve ser codificada para que possa representar as características mais relevantes de um composto no que se refere a uma dada propriedade e simultaneamente estabelecer uma boa correlação com os dados experimentais^{30-31,39}. Ademais, há outras questões importantes referentes ao acesso, manipulação e quantidade de informações disponíveis acerca da estrutura molecular. No entanto, mais importante do que o conhecimento é a necessidade de uma organização racional desses dados^{1-2,40-41}.

No campo de QSAR podem ser considerados dois diferentes métodos: o método de Hansch-Fujita (1964)⁴² onde os modelos tentam caracterizar e isolar as contribuições de cada um dos fatores físico-químicos responsáveis pela atividade biológica/atuação farmacológica, fornecendo uma explicação para o efeito de cada

parâmetro no modelo^{3, 23,30,36,41-44}. O outro método emprega uma extensa variedade de descritores moleculares que são calculados a partir da estrutura da molécula⁴¹. Neste método encontram-se aqueles conhecidos como de reconhecimento de padrões, método que revela ou apresenta um padrão dentro de um conjunto de dados⁴³⁻⁴⁶. O método DARC-PELCO⁴⁷⁻⁵² é um exemplo, o qual tem origem em uma molécula hipotética, denominada de hiperestrutura, onde estão incluídas todas as características estruturais da molécula sob investigação. Este método foi empregado por Chrétien e Dubois (1977 e 1981)⁵⁰⁻⁵¹ para predição de índices de retenção de Kováts de alcenos obtendo bons resultados.

Correlações significativas são obtidas a partir da habilidade do descritor utilizado em quantificar exatamente as características moleculares microscópicas relacionadas às propriedades macroscópicas⁵³⁻⁵⁴. Vários pesquisadores vêm realizando estudos para correlacionar e prever índices de retenção obtendo informação estrutural das moléculas de diversos tipos de compostos, tais como: odoríferos, olefinas, drogas, bifenilas policloradas, usando parâmetros físico-químicos ou descritores moleculares^{1-2,55-59}. Duas importantes revisões de estudos de QSRR, as quais são de grande ajuda em estudos nessa área de conhecimento, foram publicadas recentemente por Kaliszan (2007)¹ e Héberger (2007)².

2.1.1. Modelo de correlação entre estrutura-propriedade

O modelo de QSPR representa a relação entre dois conjuntos numéricos via uma expressão algébrica, onde um deles indica a propriedade e o outro, a estrutura da molécula. Conforme Hall e Kier (2001), tal modelo deveria extrair o paralelismo existente entre a variação da estrutura e a propriedade⁵⁹.

Prognosticar propriedades físicas, químicas e biológicas, bem como obter alguma informação sobre o comportamento molecular é o objetivo fundamental dos estudos de QSPR/QSAR nos quais são utilizados descritores moleculares simples^{1-2,57-60}. Esta relação quantitativa obtida através do modelo é capaz de prognosticar variáveis denominadas contínuas^{a*} ou classificar qualitativamente variáveis denominadas discretas ou não contínuas^{b*} 39,61-63.

^{a*} por exemplo: ponto de ebulição; toxicidade DL_{50} ; concentrações de odor passíveis de serem detectadas; atividade antiviral; coeficientes de partição octanol/água; refração molar; polaridade; acidez e basicidade do hidrogênio; polarizabilidade; retenção cromatográfica, etc.

^{b*} por exemplo: doce/amargo, tóxico/não tóxico, carcinogênico/não carcinogênico.

Um dos problemas cruciais na construção de modelos de correlação é a forma de representação da estrutura molecular para QSPR, ou seja, a tradução da estrutura em uma quantidade numérica para reter o máximo de informação estrutural possível^{1-2,64}.

Além disso, é crucial selecionar o descritor mais informativo entre um grande número de descritores correlacionados. Uma grande quantidade de artigos interessantes tem sido publicada nos últimos anos¹⁻²; onde novas tendências podem ser observadas neste campo. Como os descritores são altamente correlacionados e numerosos, selecionar um modelo apropriado não é uma tarefa fácil e muitos laboratórios utilizam modelos de QSPR para demonstrar a utilidade e as vantagens de técnicas quimiométricas desenvolvidas recentemente. Do mesmo modo, muitas vezes, modelos de QSPR demonstram a aplicabilidade dos descritores topológicos^{2,65}.

Cada descritor molecular pode ser calculado a partir da estrutura ou por meio de simulação computacional e associa um número simples a uma substância química ordenando um grupo de compostos de acordo com um valor numérico de uma propriedade específica. De acordo com critérios estatísticos previamente estabelecidos, o modelo de QSPR é válido se a variável independente (descritor molecular) estiver relacionada matematicamente (linearmente ou não) com a magnitude da propriedade em estudo (variável dependente)¹².

Estudos de QSRR necessitam de parâmetros de retenção cromatográfica quantitativamente comparáveis de uma série suficientemente grande de analitos e de um conjunto de seus descritores estruturais. Através da utilização de técnicas quimiométricas, os parâmetros da retenção são caracterizados em termos de vários descritores. Se os modelos de QSRR obtidos são estatística e fisicamente significantes, então eles podem ser aplicados para¹:

- (A) identificar os descritores moleculares mais úteis;
- (B) prever retenção de um novo analito e identificar analitos desconhecidos;
- (C) adquirir uma visão do mecanismo molecular de separação de um determinado sistema cromatográfico;
- (D) comparar quantitativamente as propriedades individuais de separação em diferentes fases estacionárias;

(E) avaliar outras propriedades que não são organolépticas dos analitos, como constantes de lipofilicidade e de dissociação;

(F) estimar bioatividade relativa dentro de um conjunto de compostos potencialmente ativos, bem como outras propriedades.

Para a construção do modelo que tem como objetivo a previsão de propriedades contínuas, vários métodos multivariados considerados clássicos podem ser utilizados. Dentro destes métodos existem: MLR (regressão linear múltipla), PLS (regressão por mínimos quadrados parciais), PCR (regressão por componentes principais) e outros^{45-46,66-67}. Já para modelos qualitativos, destinados à classificação de propriedades discretas, são empregados métodos de reconhecimento de padrões como: HCA (Análise Hierárquica de Agrupamentos), PCA (Análise de Componentes Principais) ou ainda a técnica conhecida como KNN (K-ésimo vizinho mais próximo)^{45-46,64,68-73}.

Uma das diversas maneiras com as quais podemos ilustrar as etapas envolvidas no desenvolvimento de um modelo de QSPR é via um diagrama de fluxo como apresentado por Mihalic e Trinajstic(1992)³⁰ (Figura 1).

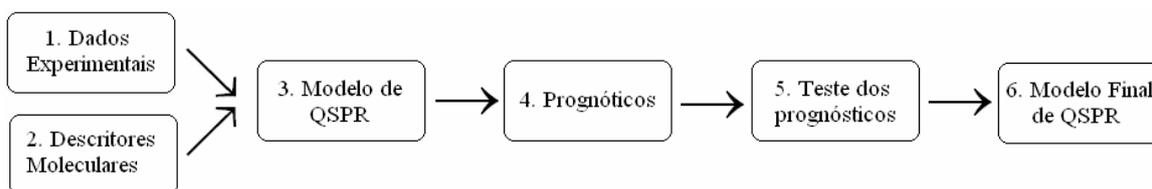


Figura 1: Diagrama de fluxo das etapas envolvidas num modelo de QSPR.

1. Dados Experimentais: são obtidos experimentalmente de forma confiável, sendo este grupo chamado de “grupo de dados”.

2. Descritores Moleculares: podem ser eletrônicos, geométricos ou topológicos e devem ser selecionados e computados adequadamente.

3. Modelo de QSPR: nesta etapa os dois conjuntos numéricos (Dados Experimentais e Descritores Moleculares), são correlacionados por meio de métodos matemáticos resultando numa expressão algébrica aceitável. O modelo obtido deve ser avaliado

estatisticamente pelo coeficiente de correlação (r), pelo desvio padrão (s) e analisado por sua estabilidade.

4. Prognósticos: os valores das propriedades moleculares de interesse são previstas para moléculas que não fazem parte do grupo de treinamento usando o modelo de QSPR obtido inicialmente.

5. Teste dos Prognósticos: realizado através de determinações experimentais, onde é possível confirmar os prognósticos ou não obtidos anteriormente. O modelo deve ser refeito se estes testes não apresentarem resultados satisfatórios, produzindo assim outro modelo até a obtenção de um adequado.

6. Modelo Final de QSPR: se os testes realizados confirmarem os prognósticos, o modelo de QSPR é então aceito na sua forma final com seus dados estatísticos.

Por conseguinte, o modelo final de QSPR construído por meio das etapas pode ser empregado no prognóstico de propriedades de compostos ou ainda na pesquisa de novos compostos que apresentem propriedades consideradas relevantes.

2.2. Propriedades moleculares

Os estudos de QSPR também podem fornecer informações importantes para o entendimento das características estruturais das moléculas relacionadas à propriedades físicas, químicas, biológicas e termodinâmicas. A habilidade para modelar teoricamente tais propriedades parece uma proposta atraente em função da falta de informações experimentais para um grande número de compostos. Muitas são as propriedades de uma molécula que podem ser previstas por meio de uma correlação com sua estrutura molecular^{28,74-75}. Em princípio, toda e qualquer propriedade está codificada na sua estrutura química e QSPRs geram eficientes maneiras de decodificar estas estruturas. Um exemplo dos benefícios de estudos em QSPR é a previsão de uma propriedade. Uma excelente revisão publicada por Katritzky *et al.* (2000) aborda estudos de QSPR onde as propriedades se destacam por sua importância tecnológica³³. Entre as diversas propriedades citadas encontramos: ponto de ebulição, temperatura crítica, densidade,

coeficiente de partição octanol-água, solubilidade aquosa de líquidos e sólidos, índices de retenção cromatográfica, e concentração crítica micelar.

A retenção cromatográfica encontra-se entre as inúmeras propriedades moleculares e tem atraído nossa atenção em estudos recentes de QSRR sendo o tema central desta tese. Nosso objetivo é identificar e quantificar diversos compostos, assim como também auxiliar na elucidação do comportamento cromatográfico dos mesmos.

2.2.1. Retenção cromatográfica

Em 1952, A. J. P. Martin e A. T. James desenvolveram a cromatografia gasosa moderna⁷⁶, a qual se tornou uma das técnicas analíticas mais importante e extensivamente utilizada. A importância desta técnica analítica é evidenciada pelos inúmeros artigos de revisão encontrados na literatura com aplicação nas mais variadas áreas⁷⁷⁻⁷⁸.

Em comemoração aos 50 anos da cromatografia gasosa, Brinkman e Jansen (2002)⁷⁹ publicaram um editorial onde relataram que James e Martin publicaram em 1952 o primeiro artigo sobre cromatografia gasosa, ou mais corretamente, sobre cromatografia de partição gás-líquido (CGL). Entretanto, em 1941 o conceito de CGL já havia sido introduzido, quando Martin e Synge publicaram um dos mais importantes artigos sobre cromatografia líquido-líquido, sendo por isso agraciados com o prêmio Nobel de Química em 1952. Neste artigo os autores escreveram (*apud* Brinckman e Jansen, 2002)⁷⁹:

A fase móvel não necessita ser um líquido, mas pode ser um vapor [...] Separações muito refinadas de substâncias voláteis deveriam ser, portanto, possíveis numa coluna onde o gás permanente flui sobre um gel impregnado com solvente não volátil onde as substâncias a serem separadas obedecem à lei de Raoult.

Por aproximadamente uma década esta idéia permaneceu adormecida, tornando-se realidade somente em 1950, onde ficou evidente a necessidade de uma padronização na apresentação dos dados de retenção cromatográfica. Para suprir essa necessidade, Kováts em 1958, propôs um sistema⁸⁰⁻⁸¹ que proporcionava uma apresentação uniforme e precisa dos dados de retenção, criando assim o denominado índice de retenção de Kováts sendo o mais utilizado na prática da cromatografia gasosa contemporânea, estando seus valores disponíveis, em laboratórios privados, publicações e em bibliotecas disponíveis comercialmente como a Biblioteca Sadtler⁸². Esse índice é utilizado no lugar dos tempos de retenção porque relaciona o tempo de retenção de um soluto com os tempos de retenção

dos alcanos lineares por meio de uma escala logarítmica, usados até então como padrões. O índice de retenção de Kováts de um composto A (IR_A) sob condições isotérmicas pode ser calculado pela Equação 1:

$$IR_A = 100N + 100 \frac{\log t_{R(A)} - \log t_{R(N)}}{\log t_{R(N+1)} - \log t_{R(N)}} \quad (1)$$

onde, $t_{R(A)}$ é o tempo de retenção ajustado do composto A ;

$t_{R(N)}$ é o tempo de retenção ajustado do composto padrão homólogo com um número N de carbonos;

$t_{R(N+1)}$ é o tempo de retenção ajustado do composto padrão homólogo com um número $N + 1$ de carbonos.

O tempo de retenção ajustado é obtido pela diferença entre o tempo de retenção medido e o tempo morto da coluna, sendo que este é o tempo necessário para uma substância muito pouco retida (o ar ou gás inerte) eluir e é dependente da velocidade de fluxo do sistema.

A polaridade da fase estacionária, a temperatura da coluna e a estrutura molecular dos analitos contribuem para o índice de retenção de um composto. Assim, o valor do índice de retenção de Kováts indica, efetivamente, a posição de um composto desconhecido em relação aos n -alcanos, sendo que todos se encontram em condições experimentais idênticas.

Algumas questões teóricas e práticas para os sistemas de índices de retenção foram abordadas por Tarjan *et al.*⁸³⁻⁸⁴ em artigos de revisão comemorativos ao 25° e 30° aniversário dos índices de retenção a partir do trabalho inicialmente desenvolvido por Kováts.

De acordo com uma revisão publicada por Eiceman, Hill e Gardea-Torredy em 1998, a retenção cromatográfica passou por significativos avanços, abordando principalmente o conceito de índice de retenção unificado e a classificação de fases estacionárias empregando o coeficiente de Kováts⁸⁵.

2.2.2. Interações intermoleculares

Interações intermoleculares entre as moléculas do soluto e da fase estacionária, conhecidas como forças de: orientação (dipolo-dipolo), indução (dipolo-dipolo induzido ou

forças de Debye), dispersão (ou forças de London), ligação de hidrogênio e complexação doador-receptor de elétrons; são responsáveis por um processo físico bastante complexo que é a retenção cromatográfica^{1,86-91}. Estas interações são naturalmente dependentes do ambiente topológico, da geometria e/ou da estrutura eletrônica destas moléculas⁹².

Estas interações intermoleculares, segundo Kaliszan (2007)¹, podem ser divididas em duas classes, sendo uma delas as chamadas forças de van der Waals, que englobam as forças de indução, orientação e dispersão, e consideradas como não específicas e de natureza “mais física”. Na segunda classe temos as forças de ligações de hidrogênio e de complexação doador-receptor de elétrons, que apresentam um caráter “mais químico”.

A seguir detalhamos as interações intermoleculares de maior interesse para fundamentar o desenvolvimento deste trabalho¹:

a) Forças de Orientação (dipolo-dipolo): decorrentes da existência de um momento de dipolo permanente, característico de moléculas polares. Estas interações decrescem com o aumento da temperatura, sendo que em casos extremos de temperaturas muito elevadas podem ser consideradas como desprezíveis. A expressão que melhor quantifica a energia de interação dipolo-dipolo é apresentada abaixo:

$$E_{d-d} = -W^2 \frac{2}{3k_B T} \frac{\mu_1^2 \mu_2^2}{\epsilon r^6}. \quad (2)$$

Onde: $W = 1/(4\pi\epsilon_0)$; $\epsilon_0 = 8,854.10^{-12} F/m$ (ou $C^2/N.m^2$) é a permissividade do vácuo; μ é o momento de dipolo; $k_B = 1,38.10^{-23} J/K$ é a constante de Boltzmann; T é a temperatura em Kelvin; ϵ é a permissividade relativa do meio; r_i é a distância entre os centros interagentes.

b) Forças de Indução (dipolo-dipolo induzido): resultantes da indução de um dipolo elétrico em moléculas não polares pela interação destas com o momento de dipolo permanente de uma molécula polar vicinal. Em moléculas polares o momento de dipolo induzido é relativamente pequeno quando comparado ao permanente. Estas interações são consideradas fracas e não dependem da temperatura. A expressão abaixo quantifica esta interação:

$$E_{d-di} = -W^2 \frac{\alpha_1 \mu_2^2 + \alpha_2 \mu_1^2}{\epsilon r^6}, \quad (3)$$

onde α é a polarizabilidade molecular.

c) Forças de Dispersão (dipolo instantâneo-dipolo induzido): oriundas da formação de dipolos instantâneos em todas as moléculas. Em moléculas apolares a flutuação contínua

da densidade eletrônica acarreta, em certo instante, um pequeno momento de dipolo que tende a polarizar moléculas vizinhas. São também conhecidas com forças de dispersão de London, por terem sido reconhecidas pela primeira vez pelo físico polonês Fritz London, sendo proporcionais a área superficial ou o volume⁷⁶. As forças de dispersão de London predominam entre os três tipos de interações de van der Waals (indução, orientação e dispersão), sendo consideradas fracas e também independentes da temperatura. Esta interação pode ser expressa segundo:

$$E_d = -W^2 \frac{3}{2} \frac{I_1 I_2 \alpha_1 \alpha_2}{I_1 + I_2 \epsilon r^6}, \quad (4)$$

sendo I é a energia de ionização.

d) Ligações de Hidrogênio: as polarizações moleculares originam as interações que ocorrem entre o átomo de hidrogênio e dois ou mais átomos eletronegativos, como O, N e F, formando uma ligação entre eles.

2.3. Descritores moleculares

Foram desenvolvidos muitos termos matemáticos denominados de descritores moleculares para a descrição da estrutura química na forma de um número ou um conjunto de números^{74,93-94}. Segundo a metodologia de Hansch e Leo (1964)⁴², as QSARs eram desenvolvidas tradicionalmente pela caracterização de propriedades físico-químicas dos compostos⁹⁵. Por volta de 1950, Wiener⁹⁶ desenvolveu o índice estrutural de natureza topológica e posteriormente Randic introduziu o índice de ramificação molecular. Mas em 1975, Kier e Hall (2001)⁵⁹ sugeriram o uso de descritores para caracterizar topologicamente as moléculas, desenvolvendo assim, o índice de conectividade molecular de primeira ordem, $^1\chi$, o qual foi correlacionado linearmente com o ponto de ebulição de alcanos de cadeia curta. Quase duas décadas depois, de acordo com Katritzky e Gordeeva (1947)⁹⁶, os descritores moleculares podem ser obtidos empiricamente ou não, sendo classificados como: topológicos, eletrônicos, geométricos ou ainda descritores combinados^{12,54,98-105}.

Os descritores empíricos correspondem às propriedades físicas, ao passo que os descritores não empíricos abrangem os índices topológicos e químico-quânticos¹². São definidos como índices topológicos quando tem sua origem restrita na conectividade e na composição da estrutura. No entanto, quando são derivados da geometria molecular tridimensional são chamados de geométricos e aqueles obtidos a partir da distribuição de

carga parcial são denominados eletrônicos. Os descritores combinados associam simultaneamente a estrutura eletrônica e a conectividade dos átomos na molécula, ou a geometria molecular e a estrutura eletrônica^{12,54,104-105}. Descritores físicos são aqueles obtidos diretamente de medidas experimentais segundo Anker e Jurs (1990)⁹². Ademais, vários parâmetros físico-químicos (hidrofóbicos, estéreo ou eletrônicos), obtidos a partir de medidas experimentais ou calculados por métodos químico-quânticos, podem ser utilizados para descrever a estrutura molecular de um composto¹⁰⁶.

Atualmente, um grande número de novos índices está disponível na literatura, incluindo programas/pacotes computacionais que efetuam os cálculos diretamente. Entre eles o programa ADAPT de Jurs (1996)¹⁰⁷ e o programa SPARC desenvolvido pela Agência de Proteção Ambiental dos Estados Unidos. Existem também os programas CODESSA (1998)⁶⁶ e DRAGON (2000)¹⁰⁸, que combinam uma grande variedade de descritores clássicos não empíricos junto com novos descritores químico-quânticos e combinados. Além disso, esses programas também realizam o tratamento estatístico dos dados para o desenvolvimento de QSPR⁴³.

O fato de existir atualmente um crescente interesse da comunidade científica pelos descritores moleculares, constatado pelo grande número de descritores propostos, como salientado por Todeschini e Consonni (2000)¹⁰⁸. São muitos os descritores definidos e facilmente calculados através de programas computacionais. Com a imensa quantidade de dados disponíveis, faz-se necessário o uso de modelos de regressão, retropropagação de erro para selecionar os melhores descritores para a construção do modelo de predição de dados³⁷.

O maior desafio está em selecionar o descritor molecular mais apropriado a ser empregado no modelo de QSPR/QSAR a fim de obter uma correlação significativa sob o ponto de vista estatístico, e de fácil interpretação estrutural^{1-2,28,45}. Apesar dos descritores moleculares serem classificados de várias maneiras, os topológicos e químico-quânticos serão detalhados na seqüência, em virtude do descritor aqui proposto estar fundamentado neles.

2.3.1. Descritores topológicos

Os descritores numéricos que caracterizam a estrutura química (o tamanho e a forma) de uma molécula por um único número¹⁴ são denominados índices topológicos. O grafo molecular é o primeiro passo da análise topológica e representa a conexão entre os átomos de uma molécula^{31,66}.

Quando um número simples, como o índice topológico, é utilizado para representar o grafo molecular resulta em uma considerável perda de informação, pois a estrutura tridimensional da molécula passa a ser descrita por um índice topológico que é uma quantidade unidimensional^{12,31,104-105}.

Estrada (2002) dividiu a história dos índices topológicos em três períodos distintos¹⁰⁹:

-Primeiro período: iniciou-se com Harry Wiener (1947) que primeiro definiu o índice estrutural de natureza topológica⁹⁷ (W , índice de Wiener). A continuação ocorreu através do estudo de Hosoya (1971)^{86,110} que introduziu o termo topológico para os descritores moleculares baseados no grafo molecular. Neste período poucos índices topológicos eram desenvolvidos e praticamente todas as aplicações se restringiam às propriedades físicas dos alcanos.

- Segundo período: iniciou-se em 1975 quando Milan Randić introduziu seu índice de ramificação molecular¹¹¹, χ . Após esta publicação uma grande proliferação destes índices foi observada na literatura e também as aplicações dos índices topológicos em diversas áreas da química^{12,31,35,112}. Os índices de conectividade molecular, desenvolvidos e aplicados por Kier e Hall nos anos 70 e 80, podem ser considerados como os representantes mais importantes deste período^{11-12,59}. Outros índices foram desenvolvidos neste período, entre eles: J , índice de Balaban (1982)^{29,113}; $^m K$, índice de Kier da forma molecular - Índice Kappa (1985)¹¹⁴; MTI , índice de Schultz (1989)³¹.

- Terceiro período: iniciou-se nos anos 90, sendo caracterizado por uma introdução mais seletiva destes descritores, pelo emprego de determinadas regras para a definição de tais índices^{30,93} e pela expansão das aplicações em casos mais complexos como na química medicinal. O número e a qualidade das aplicações dos índices topológicos na química medicinal tiveram um acentuado crescimento a partir da metade da década de 90^{113,115-117}.

O excelente desempenho desses índices, quando comparado às outras classes de descritores estruturais, pode ser atribuído:

- (i) a facilidade em calcular esses índices a partir do grafo molecular;
- (ii) as boas correlações obtidas com várias propriedades/atividades;

(iii) a fácil interpretação dos resultados, porque freqüentemente fornecem uma maneira simples de medir a ramificação molecular, forma, tamanho, semelhanças e diferenças moleculares^{38,118}.

Segundo Randic (1992)¹¹⁹, os melhores descritores são aqueles que apresentam simplicidade estrutural, porque ajudam na interpretação e explicação estrutural relacionada à propriedade. A tabela 1 enumera as características desejadas para os índices topológicos proposta por Randic.

Tabela 1: Lista das características desejadas para os índices topológicos proposta por Randic citada no artigo publicado por Mihalic e Trinajstic (1992)³⁰.

NÚMERO	CARACTERÍSTICAS
1	Interpretação estrutural direta
2	Boa correlação com pelo menos uma propriedade molecular
3	Boa discriminação de isômeros
4	Bem definido
5	Generalizável
6	Independente linearmente
7	Simplicidade
8	Não baseado em propriedades físico-químicas
9	Não relacionado superficialmente com outros índices
10	Eficiência de interpretação
11	Baseado em conceitos estruturais familiares
12	Correta dependência de tamanho
13	Mudança gradual com a gradual variação da estrutura

A discriminação de isômeros é considerada como um “ponto fraco” de muitos dos índices topológicos utilizados, podendo ser citado como exceção o índice de Balaban (1977 e 1990)¹²⁰⁻¹²¹. Porém, apesar do baixo potencial discriminatório de muitos índices, sua aplicação como descritores nos modelos estrutura-propriedade/atividade tem resultado em correlações de sucesso.

Apesar dos inúmeros índices topológicos disponíveis na literatura, pesquisas nesta área continuam em pleno desenvolvimento com o objetivo de ampliar cada vez mais sua aplicação, criar índices novos e mais abrangentes ou ainda aperfeiçoar aqueles já existentes.

A transformação de uma molécula num índice topológico que expresse as suas características, diferindo de molécula para molécula, é possível empregando a teoria dos

grafos³⁰. Será apresentada a seguir uma breve introdução da aplicação da teoria dos grafos na química para estudos de correlações estrutura-propriedade.

2.3.1.1. Teoria dos grafos moleculares

Grafos moleculares são construídos a partir da constituição das moléculas usando diferentes convenções. Sendo que os vértices, ou pontos correspondem aos átomos, e as linhas, ou arestas, representam as ligações covalentes entre os átomos^{30,67}, onde os átomos de hidrogênio e suas ligações são omitidos representando-se apenas o esqueleto molecular.^{30,39,44} Características geométricas como comprimentos de ligação, ângulos de ligação ou configurações tridimensionais, não são levados em consideração¹¹⁸. Portanto, é um conceito associado mais à topologia molecular do que à geometria, e trata da análise de todas as conseqüências da conectividade em um sistema químico³⁸⁻³⁹. A teoria dos grafos auxilia as áreas de: síntese, polímeros, cinética química, química sintética, termoquímica, equilíbrio químico, etc.³⁹ e vem sendo aplicada nos estudos de correlação estrutura-atividade e estrutura-propriedade com sucesso.

A fórmula estrutural do composto 3-metilpentano e o correspondente diagrama do grafo molecular é mostrado na Figura 2.

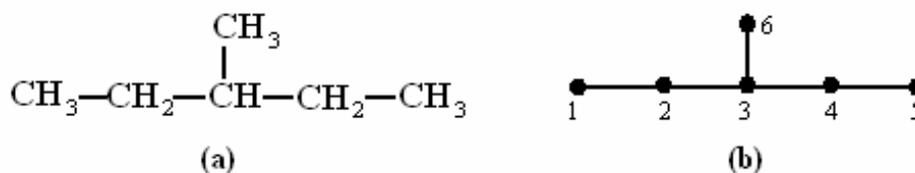


Figura 2: A molécula do 3-metilpentano como fórmula estrutural (a) e grafo molecular enumerado (b) onde foram omitidos os átomos de hidrogênio.

Por meio de um grafo molecular, como no exemplo acima, a estrutura química de um composto pode ser expressa por meio de diferentes matrizes, polinômios, espectros, momentos espectrais, seqüência da contagem das distâncias, caminhos e percursos, originando assim diferentes índices topológicos.

2.3.2. Descritores químico-quânticos

Recentemente, os progressos no desenvolvimento de processadores de alto desempenho, bem como de algoritmos eficientes, vem contribuindo de forma decisiva para a implementação de rotinas de cálculos quanto-mecânicos moleculares rápidos e precisos.

Assim, cálculos de química quântica têm servido como fonte de novos descritores moleculares, os quais podem em princípio, expressar todas as propriedades geométricas e eletrônicas de sistemas moleculares e suas interações.

A química quântica fornece uma descrição mais precisa e detalhada dos efeitos eletrônicos do que nos métodos empíricos. Métodos químico-quânticos podem ser aplicados em QSPR pela obtenção direta dos descritores a partir da função de onda molecular. Ademais, descritores eletrônicos obtidos a partir da onda molecular podem também ser separados para átomos ou grupo de átomos, permitindo a descrição de várias regiões moleculares separadamente. Métodos baseados em campos de força moleculares clássicos e métodos de química-quântica são capazes de minimizar a energia potencial da estrutura molecular. Ambas as abordagens podem ser utilizadas para obtenção de propriedades de termodinâmicas (calor de formação) e eletrônicas (momento de dipolo), porém somente cálculos químico-quânticos são capazes de estimar cargas atômicas, orbitais moleculares, energias, e vários outros descritores eletrônicos úteis em estudos de QSPR/QSAR⁵⁴. Em princípio, a química-quântica é capaz de fornecer uma descrição quantitativa precisa de estruturas moleculares e suas propriedades químicas.

Segundo Katritzky e Fara (2005)¹²² os descritores moleculares podem ser classificados de acordo com o método de cálculo. Uma revisão de aplicações de descritores químico-quânticos em estudos de QSAR/QSPR aplicados para propriedades químicas, físicas, bioquímicas e farmacológicas de compostos foi publicada por Karelson *et al.*(1996)⁵⁴. Conforme estes autores, os descritores químico-quânticos se dividem em:

(i) Descritores Geométricos: são tridimensionais e caracterizam a molécula como um todo (tamanho e forma) permitindo a diferenciação de heteroátomos^{58,123} entre moléculas que são topologicamente muito similares⁹². Alguns exemplos são⁹⁷: volume molecular; área superficial, entre outros.

(ii) Descritores Eletrônicos: refletem principalmente alguma característica da distribuição de carga parcial e podem elucidar alguma propriedade físico-química^{12,92,97,123}.

Alguns parâmetros da estrutura eletrônica molecular utilizados como descritores eletrônicos são: cargas atômicas; momento de dipolo; energia dos orbitais HOMO e LUMO; polarizabilidade molecular, etc.

O cálculo desses descritores químico-quânticos tem por base o surgimento da mecânica quântica e o uso do princípio variacional na solução da equação de Schrödinger aplicada em sistemas moleculares. Em virtude da importância desses descritores para o desenvolvimento deste trabalho, apresentamos a seguir os princípios básicos da mecânica quântica e o princípio variacional.

2.3.3.1. Mecânica quântica e o princípio variacional

Os cálculos empregados para a obtenção dos descritores moleculares quânticos utilizados em QSRR têm por base o método variacional linear em mecânica quântica, cujos princípios básicos serão detalhados a seguir após uma breve introdução histórica.

Ao final do século XIX na Química o conceito de molécula proposto por Stanislao Cannizzaro já era largamente aceito, assim como Dmitri Mendeliev chegara à tabela periódica dos elementos mesmo ainda não tendo uma interpretação adequada para este comportamento periódico na natureza; o problema da estrutura do benzeno fora esclarecido por Friedrich August Kekulé; e Svante August Arrhenius havia elucidado problemas fundamentais de reações químicas¹²⁴.

Já na Física, a Mecânica de Isaac Newton estava bem estabelecida com os trabalhos de Leonhard Euler, Joseph-Louis de Lagrange, William Rowan Hamilton entre outros, e era aplicada com sucesso no movimento planetário. Os trabalhos de Benjamin Thompson, conhecido como Conde de Rumford e James Prescott Joule resultaram na descoberta da equivalência entre trabalho e calor. As pesquisas de Sadi Carnot levaram ao conceito de entropia e a formulação da 2ª lei da termodinâmica, as quais foram continuadas por Josiah Gibbs que completou o desenvolvimento da termodinâmica. Nos campos da óptica e eletromagnetismo James Clerk Maxwell na década de 1860 unificava os campos elétricos e magnéticos, e Heinrich Rudolf Hertz em 1887 verificou pela primeira vez de forma experimental a propagação das ondas eletromagnéticas previstas por Maxwell. Esta era a situação em fins do século XIX e à sua estrutura conceitual pertinente nos referimos como sendo a Física Clássica¹²⁴.

Porém a Física Clássica apresentava algumas falhas no que se refere à elucidação da radiação do corpo negro, do efeito fotoelétrico e da estabilidade atômica (existência das linhas espectrais). A busca para o entendimento destes fenômenos originou o

desenvolvimento de uma teoria que iria revolucionar a descrição física de sistemas microscópicos. Esta teoria conjugava conceitos da mecânica clássica com conceitos novos relacionados à energia e propriedades intrínsecas das partículas. Em 1900 Max Planck postula a quantização da energia solucionando a problema da radiação do corpo negro. Em 1905 Albert Einstein demonstra que aos quanta de energia, anteriormente postulados por Planck, estariam associados à partículas (fótons), da mesma maneira que a matéria é constituída por átomos, explicando assim o efeito fotoelétrico. A estabilidade atômica é resolvida por Niels Bohr postulando que os elétrons poderiam ocupar apenas certas órbitas e a passagem de uma órbita para outra, dar-se-ia com absorção ou emissão de um fóton. Durante a década de 1920 com os trabalhos de Louis de Broglie (dualidade onda-partícula), Erwin Schrödinger (equação da onda), Max Born (interpretação probabilística da função de onda) e Werner Heisenberg (princípio da incerteza) estruturaram-se definitivamente os alicerces da Mecânica Quântica¹²⁵.

Na Mecânica Quântica a descrição mais apropriada do comportamento de sistemas microscópicos (elétrons, núcleos, átomos, moléculas) é dada ou está contida em uma função matemática de suas coordenadas e do tempo, denominada função de onda (Ψ). Max Born interpretou $|\Psi|^2$ como a densidade de probabilidade de um elétron ser encontrado em um volume de contorno em um instante t . Em um problema unidimensional, a probabilidade de que a partícula seja encontrada no intervalo dx em torno do ponto x_1 é dada por $|\Psi(x_1)|^2 dx$. Tridimensionalmente, podemos interpretar $|\Psi(x, y, z, t)|^2 dv$ como a probabilidade de encontrar a partícula no volume dv no instante t . Rigorosamente falando, Ψ também é dependente da coordenada de spin da partícula (s_p). Para uma partícula confinada em um volume de integração, V_i , sabemos que a soma das probabilidades de encontrá-la em algum lugar, dentro deste volume, deve ser a unidade, condição que expressamos por meio de:

$$\int_{V_i} |\Psi|^2 dv = 1, \quad (5)$$

que é denominada condição de normalização da função de onda. A interpretação física de $|\Psi|^2$, estabelece que funções de onda aceitáveis devem ser unívocas, finitas em todo espaço, contínuas e normalizáveis. Em particular, para sistemas moleculares, Ψ eletrônica deve ser anti-simétrica com respeito à troca de coordenadas de dois elétrons quaisquer, de modo a satisfazer o *princípio de exclusão de Pauli*, que proíbe dois elétrons de ocuparem o mesmo estado quântico¹²⁴.

Um dos postulados fundamentais da mecânica quântica estabelece que a evolução temporal de qualquer sistema físico é descrito pela *equação de Schrödinger*

$$H\Psi = i\hbar \frac{\partial \Psi}{\partial t}, \quad (6)$$

onde $\hbar = \frac{h}{2\pi}$ (i é a unidade imaginária e h é a constante de Planck). Sendo H o *operador Hamiltoniano* do sistema,

$$H = -\frac{\hbar^2}{2m} \nabla^2 + V(x, y, z, t), \quad (7)$$

$$\text{no qual } \nabla^2 = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \right) \text{ (operador Laplaciano),} \quad (8)$$

onde o primeiro termo de H é identificado como sendo a energia cinética ($p^2/2m$) e o segundo como a energia potencial. Quando a energia potencial (V) não depende explicitamente do tempo, $V = V(x, y, z)$, em geral podemos separar Ψ em coordenadas espaciais e temporal, ou seja,

$$\Psi(x, y, z, t) = \psi(x, y, z) f(t). \quad (9)$$

Substituindo a equação acima na equação de Schrödinger obtemos,

$$i\hbar \psi(x, y, z) \frac{df}{dt} = H\psi(x, y, z) f(t) \quad (10)$$

e dividindo-a por $\Psi = \psi f$, temos:

$$\frac{i\hbar}{f} \frac{df}{dt} = \frac{H\psi}{\psi} = E, \quad (11)$$

onde E é uma constante, identificada como sendo a energia total do sistema. A equação anterior fornece duas equações, uma para $f(t)$ e outra para ψ . Para $f(t)$ temos,

$$\frac{df}{f} = -\frac{iE}{\hbar} dt, \text{ integrando-a: } f = e^{-\frac{iEt}{\hbar}}. \quad (12)$$

Para a parte espacial (ψ) obtemos,

$$H\psi = E\psi, \quad (13)$$

que é a equação de Schrödinger para estados estacionários¹²⁴. Assim denominados, pois a densidade de probabilidade para tais estados permanece constante com o tempo,

$$|\Psi|^2 = \Psi^* \Psi = f^* f \psi^* \psi = \psi^* \psi e^{+\frac{iEt}{\hbar}} e^{-\frac{iEt}{\hbar}} = \psi^* \psi = |\psi(x, y, z)|^2. \quad (14)$$

Cada sistema molecular apresenta vários níveis de energia de modo que a equação para estados estacionários é escrita de forma mais correta como $H\psi_n = E_n\psi_n$. Uma função

ψ_n que satisfaz a equação de Schrödinger para estado estacionário é dita uma autofunção do operador H . A constante E_n é dita um autovalor associado à autofunção ψ_n . Em geral, a equação de Schrödinger tem muitas autofunções E_n aceitáveis para uma dada molécula, onde cada uma é caracterizada por um autovalor E_n diferente, menos para estados quânticos degenerados, onde para várias funções ψ_n temos um mesmo autovalor associado E_n ¹²⁶.

O operador Hamiltoniano para sistemas moleculares possui cinco contribuições para a energia total: a energia cinética dos elétrons e núcleos, a atração dos elétrons aos núcleos, as repulsões intereletrônicas e internucleares conforme a equação a seguir,

$$H = -\sum_i \frac{\hbar^2}{2m_e} \nabla_i^2 - \sum_k \frac{\hbar^2}{2m_k} \nabla_k^2 - \sum_i \sum_k \frac{e^2 Z_k}{r_{ik}} + \sum_{i<j} \frac{e^2}{r_{ij}} + \sum_{k<l} \frac{e^2 Z_k Z_l}{r_{kl}}, \quad (15)$$

onde i e j referem-se aos elétrons, k e l referem-se aos núcleos, m_e é a massa do elétron, m_k é a massa do núcleo k , e Z é o número atômico e r_{ab} é a distância entre as partículas a e b . Assim, ψ_n é uma função de $3n$ coordenadas (x , y e z , por exemplo) onde n é o número total de partículas (núcleos e elétrons)^{124,126}.

Obter funções de onda exatas para os sistemas moleculares é extremamente difícil, pois os termos que compõem o Hamiltoniano levam em conta atração e repulsão aos pares, implicando um movimento interdependente entre partículas, denominado *correlação*. Simplificamos o problema recorrendo à aproximação de *Born-Oppenheimer*, que considera os núcleos fixos no espaço já que a razão entre as massas eletrônica e nuclear é $(m_e / m_k) \ll 1$, e o movimento dos elétrons pode ser considerado como independente do movimento dos núcleos¹²⁴. Assim separamos o movimento nuclear do eletrônico e resolvemos a equação de Schrödinger para obter os estados eletrônicos moleculares. A equação de Schrödinger para a parte nuclear fornece os espectros rotacionais e vibracionais moleculares.

A energia total de um sistema molecular descrito pela função de onda Ψ normalizada é dada por:

$$E_0 = \int_{V_i} \Psi^* H \Psi dv. \quad (16)$$

Entretanto, na prática torna-se impossível a solução exata da equação de Schrödinger. Um dos teoremas fundamentais aplicadas à mecânica quântica é o princípio

variacional de Rayleigh-Ritz¹²⁴, o qual estabelece que não pode existir um autovalor de H mais baixo que a energia mínima do sistema (E_0), ou seja,

$$E = \int_{V_i} \Phi^* H \Phi dv \geq E_0, \quad (17)$$

onde Φ são funções normalizadas que representam estimativas da função Ψ , que é a solução exata da equação de Schrödinger para o sistema. Quanto mais a função Φ escolhida for semelhante a Ψ (função de onda real ou verdadeira), menor será a diferença entre E e E_0 .

Sabendo-se que não existe solução analítica para moléculas, faz-se obrigatório o uso de algum método que permita obter funções de onda que descrevam adequadamente o sistema molecular. Uma forma prática de resolver este problema é o emprego do método variacional de Rayleigh-Ritz¹²⁴. No cálculo variacional não linear a função de onda tentativa (ou seja, a função de onda proposta) é tomada como dependente de um ou mais parâmetros $\Psi = \Psi(\alpha, \beta, \gamma)$ e logo $E = E(\alpha, \beta, \gamma)$ e podemos minimizar E com relação aos parâmetros α, β, γ de forma a obter a melhor energia com a função de onda adotada. Outra maneira de se obter um valor aproximado para a energia de um sistema qualquer é considerarmos a função de onda que descreve o estado quântico do sistema como uma combinação linear de funções de onda conhecidas. Isto quer dizer que a função de onda tentativa, Φ , depende linearmente dos parâmetros variacionais. Aqui não é necessário que as funções Φ sejam normalizadas.

2.3.3.2. Método variacional linear

No método variacional linear a função estimada Φ é escrita como combinação linear de orbitais atômicos denominados de *conjunto base*, isto é,

$$\Phi = \sum_{i=1}^N a_i \varphi_i = a_1 \varphi_1 + a_2 \varphi_2 + \dots + a_N \varphi_N \quad (18)$$

onde o conjunto de N funções φ_i são funções de onda conhecidas (não necessariamente normalizadas), e os a_i são os coeficientes da expansão. As funções de onda atômicas φ_i podem ser, por exemplo, orbitais hidrogenóides $1s, 2s, 2p, 3s, 3p, 3d$ etc. A forma e o número dessas funções são escolhidas de modo a cobrir da melhor maneira possível a região da ligação química^{124,126}. Podemos ilustrar o método variacional considerando a função estimada Φ construída a partir de dois orbitais,

$$\Phi = a_1\varphi_1 + a_2\varphi_2. \quad (19)$$

A energia do sistema é calculada usando

$$E = \frac{\int \Phi^* H \Phi dv}{\int \Phi^* \Phi dv}, \quad (20)$$

substituindo Φ vem

$$E = \frac{|a_1|^2 \int \varphi_1^* H \varphi_1 dv + (a_1^* a_2 + a_1 a_2^*) \int \varphi_1^* H \varphi_2 dv + |a_2|^2 \int \varphi_2^* H \varphi_2 dv}{|a_1|^2 \int |\varphi_1|^2 dv + (a_1^* a_2 + a_1 a_2^*) \int \varphi_1^* \varphi_2 dv + |a_2|^2 \int |\varphi_2|^2 dv}. \quad (21)$$

Assumindo por simplicidade que as φ_i são reais e normalizadas, e que os coeficientes a_i sejam reais, reescrevemos a equação acima como:

$$E = \frac{a_1^2 H_{11} + 2a_1 a_2 H_{12} + a_2^2 H_{22}}{a_1^2 + 2a_1 a_2 S_{12} + a_2^2}, \quad (22)$$

onde as notações H_{ij} e S_{ij} representam as integrais no numerador e denominador, respectivamente. O elemento de matriz, H_{ij} é chamado *integral de ressonância* e S_{ij} é chamado *integral de sobreposição (overlap)*. A integral de sobreposição indica o recobrimento dos orbitais atômicos centrados em diferentes sítios. As integrais de ressonâncias H_{ii} podem ser interpretadas como correspondendo à energia de um único elétron ocupando o orbital atômico φ_i ^{124,126}.

Aplicando a condição de minimização $\frac{\partial E}{\partial a_1} = 0$ e $\frac{\partial E}{\partial a_2} = 0$ na equação anterior,

obtemos o seguinte sistema de equações:

$$(H_{11} - E)a_1 + (H_{12} - ES_{12})a_2 = 0 \quad (23)$$

$$(H_{12} - ES_{12})a_1 + (H_{22} - E)a_2 = 0$$

Sabemos da Álgebra Linear que um conjunto de N equações em N incógnitas tem uma solução não-trivial ($a_1 \neq 0$ e $a_2 \neq 0$) se e somente se o determinante formado pelos coeficientes das incógnitas (as várias quantidades $H_{ij} - ES_{ij}$) for igual a zero, ou seja,

$$\begin{vmatrix} H_{11} - E & H_{12} - ES_{12} \\ H_{12} - ES_{12} & H_{22} - E \end{vmatrix} = 0, \quad (24)$$

que é denominado *determinante secular* para o conjunto de equações (24) (N=2). Quando a função de onda estimada é uma combinação de N funções φ_i conhecidas não normalizadas (Equação 12), o determinante secular tem a forma,

$$\begin{vmatrix} H_{11} - ES_{11} & H_{12} - ES_{12} & \cdots & H_{1N} - ES_{1N} \\ H_{21} - ES_{21} & H_{22} - ES_{22} & \cdots & H_{2N} - ES_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ H_{N1} - ES_{N1} & H_{N2} - ES_{N2} & \cdots & H_{NN} - ES_{NN} \end{vmatrix} = 0. \quad (25)$$

Em geral haverá N raízes E , isto é, haverá N energias E_i (algumas das quais podem ser iguais uma a outra, caso em que as raízes são ditas *degeneradas*) onde cada valor de E_i dará origem a um conjunto diferente de coeficientes a_i , obtidos resolvendo-se o conjunto de Equação (23). Estes coeficientes definirão uma função de onda estimada Φ que descreve o estado quântico do sistema. O menor valor de E_i é a energia mínima obtida com a função estimada, sendo os demais valores de energias associados aos estados excitados^{124,126}.

O método variacional linear é a base na qual se assentam os modernos métodos computacionais empregados no cálculo de propriedades moleculares e mecanismos de reação. Existe uma grande variedade de implementações deste método com diversos níveis de refinamento. Podemos classificar em dois os grupos os métodos baseados em mecânica quântica: *os métodos ab initio* e *os métodos semi-empíricos*^{126,127}.

Nos primeiros, efetua-se o cálculo de cada interação elétron-elétron, elétron-núcleo e núcleo-núcleo para a obtenção da energia total do sistema, isto é, todos os elétrons são levados em conta com o mesmo rigor matemático^{124,126}.

Nos *métodos semi-empíricos*, apenas os elétrons de valência entram no cálculo da energia total. Ao passo que os demais elétrons (elétrons de cerne) têm seus valores de energia de interação já previamente calculados e parametrizados com base em dados experimentais, por exemplo: potenciais de ionização, momento de dipolo, calores de formação etc. A relativa imprecisão destes métodos é compensada com uma redução considerável no tempo computacional, o que permite tratar sistemas moleculares com grande número de átomos, o que atualmente ainda é inviável usando os *métodos ab initio*^{124,126}.

Uma grande variedade de parametrizações resultou em diversos *métodos semi-empíricos*, sendo os mais empregados MNDO, AM1, PM3 e PM5¹²⁸. O método MNDO é ainda utilizado apesar de apresentar alguns problemas como: o impedimento estéreo tendendo a ser fortemente desfavorecido, resultando em calores de formação excessivamente altos e pequenos anéis são previstos como estáveis demais. Com o objetivo de sanar tais problemas, Michael J. S. Dewar e colaboradores (1985) desenvolveram o modelo Austin Model 1 (AM1) através da modificação da forma funcional do modelo MNDO (cujo erro primário era os comprimentos de ligação) via alteração do termo de repulsão nuclear no Hamiltoniano. Inicialmente o AM1 foi parametrizado para os elementos C, H, O e N, sendo que atualmente inclui parametrizações para B, F, Al, Si, P, S, Cl, Zn, Ge, Br, I e Hg. Em 1989 James J. P Stewart desenvolveu o PM3 a partir da otimização simultânea dos parâmetros para boa parte dos elementos químicos acima, mas limitando a expansão da função de onda molecular a duas funções gaussianas por átomos ao invés de quatro no termo de repulsão nuclear. Uma nova versão do PM3 (PM5) foi desenvolvida por Stewart e colaboradores (2002) apresentando um conjunto de parâmetros melhorados e que tem sido empregada em cálculos de energias de reação em química orgânica¹²⁸.

Contudo, os métodos PM3 e PM5 não necessariamente superam os métodos MNDO e AM1 demandando maior tempo computacional o que em muitos casos significa que se pode adotar o método AM1, pois os resultados alcançados terão a precisão requerida.

2.4. Índice semi-empírico topológico (I_{ET})

Para codificar a informação encontrada na estrutura molecular, tais como número de átomos, suas conectividades, número de insaturações (caso existam) e a presença de heteroátomos, descritores topológicos derivados da teoria do grafo molecular têm sido propostos e utilizados nas últimas décadas¹⁻². Porém, umas das limitações desses índices, era a incapacidade de diferenciar estruturas isoméricas *cis*- e *trans*- de alcenos¹³. Ademais, o princípio da aditividade do índice de retenção proposto por Kováts não é observado para alcanos ramificados, uma vez que a retenção de átomos terciários ou quaternários diminui devido ao efeito estéreo provocado pelos átomos ou grupos vizinhos. Isto pode ser observado através do desvio da correlação linear entre o índice de retenção cromatográfico

e o número de átomos de carbono para alcanos ramificados comparando com os n-alcanos, os quais apresentam uma relação linear conforme mostrado na Figura 3.

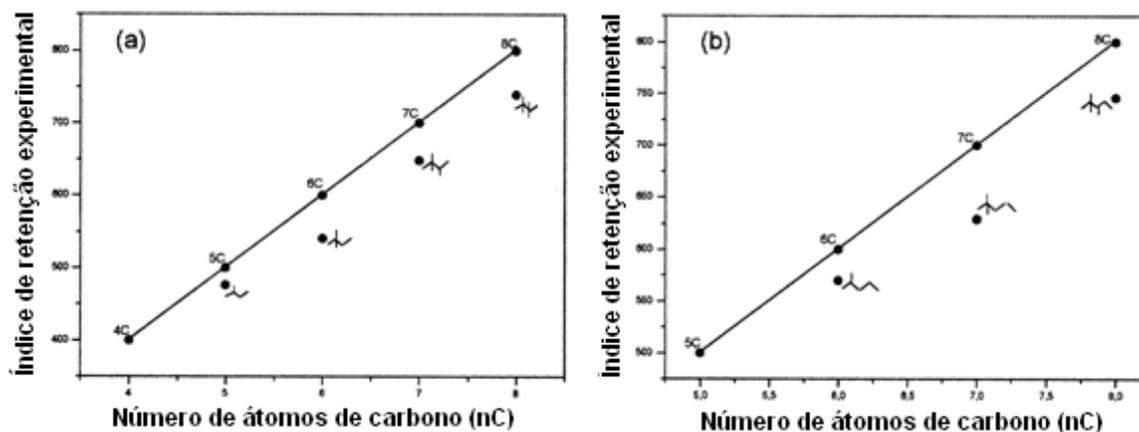


Figura 3: Regressão linear simples entre os índices de retenção experimentais (usando esqualano como fase estacionária¹³) e número de átomos de carbono para alcanos lineares e ramificados

Portanto, é evidente a partir da Figura 3, que os átomos de carbono secundário, terciário e quaternário possuem um valor menor do que 100 unidades de índice de retenção. Sendo assim, Heinzen e colaboradores (1999)¹³, baseados no comportamento de retenção cromatográfica e através de deduções teóricas, atribuíram valores numéricos aproximados de: 100 para o átomo de carbono do grupo metila de acordo com Kováts, 90 para os átomos de carbono secundários, 80 para os átomos de carbono terciários e 70 para os quaternários, conforme a Tabela 2, desenvolvendo o Índice semi-empírico topológico (I_{ET}).

Tabela 2: Valores de C_i e \log de C_i para os átomos de carbono primário, secundário, terciário e quaternário em alcanos.

Fragmento	Valor	C_i	$\log C_i$
-CH ₃	100	1,000	0,0000
-CH ₂ -	90	0,900	-0,04576
-CH<	80	0,800	-0,09691
>C<	70	0,700	-0,1549

C_i : Valores normalizados dos átomos de carbono

As moléculas descritas pelo I_{ET} são primeiramente representadas como grafos moleculares de acordo com a teoria do grafo químico, e posteriormente são atribuídos os respectivos valores para os fragmentos (CH₃, CH₂, CH e C). Além do mais, a determinação do grau ou do valor do efeito estereo presente em cada átomo de carbono depende também do tamanho do substituinte e não somente do seu grau de substituição. Como as tendências

dos desvios de linearidade são logarítmicas, foi também necessário adicionar o valor do logaritmo do átomo(s) de carbono adjacente(s) de acordo com a sua característica como carbono primário, secundário, terciário, ou quaternário. Assim o índice proposto, I_{ET} , é calculado para cada molécula segundo a equação abaixo¹³:

$$I_{ET} = \sum(C_i + \log C_j) \quad (26)$$

sendo, C_i é o valor atribuído a cada átomo de carbono e/ou grupo funcional i da molécula e C_j é o valor atribuído a cada átomo de carbono e/ou grupo funcional adjacente a i .

A determinação da retenção cromatográfica de alcanos, por este índice, se baseia na suposição de que as forças de dispersão de London são as principais responsáveis pela interação de cada átomo de carbono com a fase estacionária, sendo esta interação reduzida pelo efeito estéreo da vizinhança¹¹⁶. No caso dos alcenos^{13-14,129}, todas as forças de van der Waals estão envolvidas no processo de retenção cromatográfica. Entretanto, a influência dos efeitos conformacionais sobre as forças intermoleculares tornou difícil a previsão destes efeitos sobre a retenção, baseando-se somente em considerações teóricas, como no caso dos alcanos. Desse modo, valores aos átomos de carbono da dupla ligação de alcenos lineares foram calculados através de uma aproximação numérica¹³.

Posteriormente, este índice foi ampliado para estimar a retenção cromatográfica de ésteres, aldeídos e cetonas^{16,24} em fases estacionárias de baixa polaridade. Para este grupo de compostos as forças de van der Waals, principalmente as de orientação e de indução, apresentam uma maior contribuição no comportamento cromatográfico observado.

Os grupos -COO- (éster) e C=O (aldeído e cetona) foram considerados como um único vértice do grafo molecular, sendo atribuídos valores a estes grupos observando os índices de retenção experimentais dos compostos. Devido à importância do efeito estéreo e interações intramoleculares mútuas entre o grupo funcional e os átomos vizinhos das moléculas ramificadas, foram atribuídos valores aos átomos de carbono nas posições α , β , γ diferentes daqueles previamente atribuídos aos alcanos¹³.

Devido aos resultados obtidos, este índice foi aplicado para outras classes de compostos orgânicos, entre elas: álcoois¹⁹, oxo compostos¹⁸, alquil benzenos²⁶, alcanos halogenados²⁷ na obtenção de correlações estrutura-retenção e outras propriedades. Os modelos de QSRR obtidos apresentaram resultados satisfatórios indicando o potencial de previsão do índice semi-empírico topológico empregando a regressão linear simples com

um único descritor molecular. Além disso, este índice mostrou-se simples, original e capaz de diferenciar isômeros *-cis / -trans* de alcenos lineares¹³.

Porém, com o acréscimo de possibilidades de vizinhanças, as tabelas geradas para o cálculo do I_{ET} das moléculas tornaram-se mais extensas. Enquanto para os alcanos havia 4 possibilidades, para os alcenos era acrescentado mais 12 possibilidades de fragmentos devido aos valores específicos que representavam a ligação C=C¹³. Já para as funções cetonas e aldeídos os valores específicos para a ligação C=O e para as ligações vizinhas representavam um acréscimo de mais 9 possibilidades²⁴, e para os ésteres a tabela utilizada contém mais 17 fragmentos¹⁶. Tais acréscimos dificultaram o cálculo manual do I_{ET} de diferentes funções.

Para contornar tais limitações e tornar o índice independente do uso de tabelas específicas, iniciou-se uma busca para correlacionar os valores C_i atribuídos a cada átomo de carbono com uma propriedade independente da retenção cromatográfica. Entre as propriedades testadas estão: volume molecular, área superficial, polarizabilidade, eletronegatividade, momento de dipolo, etc. Todavia, as cargas atômicas líquidas (δ_i) apresentaram uma boa correlação. No capítulo de metodologia, seção 4.2., encontra-se uma explanação detalhada para a construção do I_{SET} .

2.5. Classes de compostos estudadas

A seguir reportaremos a importância de cada um desses grupos para a química e áreas relacionadas.

2.5.1. Alcanos e alcenos

Uma grande variedade de alcanos metil ramificados são sintetizados por insetos, sendo que as diferenças estruturais entre estes compostos são bastante sutis. Tais compostos, em geral, consistem de cadeias principais lineares longas com um número ímpar de átomos de carbono, estendendo-se até 51 átomos. Características relevantes destes compostos são devido as posições restritas das ramificações metila. Além de serem considerados como agentes impermeabilizantes, também são considerados como compostos químicos específicos de atração denominados de feromônios de agregação e/ou sexual.

Nos últimos 30 anos, de acordo com Koutek *et al.* (1998 e 2002), os estudos envolvendo feromônios de insetos têm despertado grande interesse na comunidade científica, principalmente na procura de novas tecnologias no controle das pragas¹³⁰⁻¹³¹. Para identificar os metil-alcanos a cromatografia gasosa (CG) e a cromatografia gasosa acoplada à espectrometria de massa (CG-EM) são os principais métodos usados¹³².

A análise de alcanos polimetil-substituídos relatada por Kissin *et al.* (1986)¹³³ permitiu a previsão das possíveis estruturas destes compostos, facilitando assim, a identificação dos picos no cromatograma através do princípio da aditividade. Modelos de QSRR desenvolvidos por Katritzky *et al.* (2000)¹³² para 178 *mono-, di-, tri- e tetra-*metil alcanos produzidos por insetos possibilitaram a interpretação da correlação entre os índices de retenção cromatográficos com as suas estruturas moleculares. O programa computacional CODESSA foi utilizado para descrever as moléculas e o melhor modelo foi obtido com descritores topológicos embasados em descritores químico-quânticos. Para este mesmo grupo de hidrocarbonetos eluidos em uma coluna capilar não polar de polisiloxano-DB-1, Carlson *et al.* (1998)¹³⁴ descreveram um protocolo de identificação. A análise detalhada da composição de lipídeos da teia da aranha através de cromatografia gasosa, espectrometria de massa e derivatizações químicas pelo método numérico geral, permitiram a Schulz (2001)¹³⁵ estimar os índices de retenção dos alcanos metil ramificados.

Os alcenos são produtos importantes de muitas reações e processos químicos, como craqueamento catalítico e térmico, pirólise, oligomerização, dehidrogenação, isomerização, etc¹³⁷. No ambiente, esta classe de compostos representa contaminantes principalmente nas emissões antropogênicas de alcenos voláteis, como componentes de combustíveis não queimado¹³⁸. Foto-oxidantes também são formados durante as reações de alcenos na atmosfera, sendo que as diferenças nas velocidades de reações representam um papel importante. Por exemplo, a velocidade relativa para a reação dos hexenos acíclicos com ozônio está na faixa de 1:100 e a velocidade de reação aumenta bruscamente com o aumento do número de grupos alquil adjacentes à dupla ligação¹³⁹. Esta é a razão pela qual a indústria petrolífera deve produzir uma gasolina que apresente níveis reduzidos de emissão de gases, através dos escapamentos dos automóveis, com componentes considerados prejudiciais a saúde (alcenos e compostos aromáticos). Portanto, métodos analíticos efetivos são necessários para assegurar que os valores exigidos pelos padrões de qualidade do ar estejam sendo respeitados¹³⁷.

Chrétien e Dubois (1986 e 1977)⁴⁹⁻⁵⁰ utilizaram o método designado de Darc-Pelco para alcenos ramificados e obtiveram bons resultados na previsão dos índices de retenção de Kováts. Khorasheh *et al.* (1989)¹⁴⁰ empregaram o método teórico grafo para monoalquil de alta massa molar e polimetil alcenos com o objetivo de estimar os índices de retenção destes compostos. A correlação entre os valores previstos e experimentais apresentou parâmetros estatísticos satisfatórios. Duvenbeck e Zinn (1993)¹⁴¹ descreveram o uso de modelos de vértices e arestas, desenvolvidos a partir da teoria dos grafos, para correlacionar dados de retenção de alcenos ramificados utilizando índices topológicos como o índice de conectividade molecular a nível atômico e o índice eletrotológico. Hu e Zhang (1993)¹⁴² correlacionaram índices de retenção cromatográficos para um grande número de alcenos (C₅-C₁₀) em diferentes fases estacionárias, a várias temperaturas. Os descritores empregados foram parâmetros de solubilidade total (δT), volume molar (v), e número de átomos de carbono (N) nas moléculas do soluto, obtendo bons resultados de QSRR. Outra publicação sobre QSRR envolvendo alcenos ramificados foi a reportada por Dimov e Osman (1996)¹⁴³, onde os autores desenvolveram um esquema de seleção de descritores que foi aplicado a alcenos. As equações obtidas foram formadas por dois grupos de descritores: básicos e de ajuste. Os descritores básicos foram aqueles que proporcionaram cálculos dos valores dos índices de retenção próximos aos valores experimentais. Já os descritores de ajuste foram aqueles que possibilitaram uma maior aproximação dos valores calculados aos valores experimentais. Posteriormente, os autores estudaram o efeito da conformação destes mesmos compostos na previsão dos índices de retenção¹⁴⁴.

2.5.2. Aldeídos e cetonas

As plantas em geral têm capacidade para sintetizar, acumular e emitir substâncias voláteis que podem atuar como moléculas que conferem odores e aromas devido as suas interações específicas com órgãos olfativos e gustativos. Essas substâncias de baixa massa molecular incluem aldeídos, cetonas, álcoois, ésteres, etc e são comercialmente importantes para as indústrias alimentícias, farmacêuticas e químicas sendo usados como aromatizantes, drogas, pesticidas e matérias-primas. Devido à baixa abundância desses voláteis em plantas, muitos dos produtos naturais estão sendo substituídos por seus análogos sintéticos. Neste sentido, pesquisas recentes nas indústrias químicas vêm sendo desenvolvidas para um uso mais amplo destes compostos como produtos de limpeza

doméstica, higiene pessoal etc, o que vem gerando um grande numero de patentes nesta área¹⁴⁵⁻¹⁴⁶.

2.5.3. Ésteres

Os ésteres são considerados o principal grupo de odores frutados. Sendo assim, esta classe de compostos tem sido o objeto de vários estudos de QSPR/QSAR. Boelens (1983)¹⁴⁰ classificou os fortes odores frutados de 106 ésteres etílicos e, em seguida, por regressão múltipla, correlacionou estas características com a estrutura molecular. Na continuidade deste trabalho, Sell (1988)¹⁴⁷ investigou para uma série de ésteres alifáticos os efeitos do impedimento estérico da função éster e a presença de insaturação na vizinhança da função. Rossiter (1996 e 1996)^{69,148} estudou a correlação quantitativa entre odor/estrutura para 27 ésteres alifáticos utilizando três abordagens diferentes: Regressão, Análise de Componentes Principais (PCA) e Análise Comparativa do Campo Molecular (CoMFA) e concluiu que essas técnicas podem ser usadas para identificação da estrutura/odor e que elas podem ser aplicadas, com certo grau de confiança, para resolver problemas onde essa relação não é tão nítida.

Os principais problemas na aplicação da correlação estrutura/atividade para odores são as seguintes: a limitada compreensão do mecanismo olfativo e as dificuldades associadas com medições precisas de odores. No entanto, esses problemas foram minimizados com os progressos recentes nas ciências biológicas e no desenvolvimento de programas de computador mais sofisticados para a aplicação de QSAR¹⁴⁸

3. Objetivos

Objetivo geral

Propor a construção de um novo descritor molecular baseado na carga atômica líquida calculada através de métodos computacionais químico-quânticos. Bem como, o prognóstico da retenção cromatográfica através de estudos de correlação quantitativa entre estrutura molecular e retenção cromatográfica usando este descritor (o índice semi-empírico eletrotopológico).

Objetivos específicos

- Investigar a correlação existente entre os valores C_i atribuídos a cada átomo de carbono pelo I_{ET} e a carga atômica líquida atribuída para os mesmos átomos de carbono na molécula;
- Verificar a possibilidade da utilização dos valores das cargas atômicas líquidas de cada átomo de carbono na molécula no lugar dos valores tabelados atribuídos a cada átomo de carbono no cálculo do I_{ET} .
- Desenvolver e otimizar este novo índice (semi-empírico eletrotopológico - I_{SET}) para prognosticar a retenção cromatográfica de compostos orgânicos de interesse, inicialmente os alcanos e alcenos;
- Testar a habilidade de previsão dos modelos de QSRR obtidos com o novo índice, através de parâmetros estatísticos clássicos, de testes de validação internos e externos, e do procedimento de validação cruzada (r_{CV}^2);
- Aplicar o Índice Semi-Empírico Eletrotopológico (I_{SET}) para prognosticar a retenção cromatográfica de aldeídos, cetona e ésteres através da correlação quantitativa entre estrutura molecular e retenção cromatográfica (QSRR);
- Testar a habilidade de previsão dos modelos de QSRR obtidos, através de parâmetros estatísticos clássicos, de testes de validação internos e externos, e do procedimento de validação cruzada (r_{CV}^2).

4. Metodologia

4.1. Índices de retenção cromatográficos

Os índices de retenção cromatográficos dos compostos de interesse foram obtidos da literatura.

4.1.1. Alcanos e alcenos

Os índices de retenção de 85 alcanos e 94 alcenos (Tabela 4, página 46) estudados foram determinados a 100 °C por cromatografia gasosa de alta performance com coluna capilar (200m x 0,25mm I.D.) recoberta com esqualano^{50,149-151}.

4.1.2. Aldeídos e cetonas

Os índices de Kováts de 15 aldeídos e 40 cetonas alifáticas estudados foram determinados por cromatografia gasosa com a fase estacionária HP-1 a 50°C conforme mostrado na Tabela 7 na página 52^{18,24,56,72,82,152-154}.

4.1.3. Ésteres alifáticos

Os índices de retenção de 100 ésteres alifáticos estudados foram determinados em isotermas a 150 °C em cinco diferentes fases estacionárias (SE-30, OV-7, DC-710, XE-60 e OV-225)^{64,82,155-156}. Os ésteres e seus índices de retenção cromatográficos estão mostrados na Tabela 10 na página 58. As cinco fases estacionárias apresentam baixa e média polaridade sendo que todas usam polisiloxano, variando apenas o percentual do grupo fenila nos substituintes. As constantes de McReynolds das fases estacionárias¹⁵⁶ estão dispostas na Tabela 11 na página 60.

4.2. Cálculo do índice semi-empírico eletrotopológico (I_{SET})

O desenvolvimento do I_{ET} para alcanos e alcenos lineares foi brevemente descrito no capítulo de revisão bibliografia seção 2.4, onde os valores atribuídos para os átomos de

carbono primário ($\text{CH}_3 = 1,000$), secundário ($\text{CH}_2 = 0,900$), terciário ($\text{CH} = 0,800$) e quaternário ($\text{C} = 0,700$), são calculados por aproximação numérica, baseados no comportamento da retenção cromatográfica experimental das moléculas e através de deduções teóricas¹³. O valor do I_{ET} para cada molécula é calculado segundo a Equação (26) (página 26).

Com a observação da existência de uma boa correlação entre as cargas atômicas líquidas (δ_i) dos átomos de carbono e os valores C_i atribuídos a cada átomo de carbono da molécula, definiu-se o índice semi-empírico eletrotológico substituindo-se os valores de C_i por SET_i e C_j por SET_j , conforme a equação a seguir:

$$I_{SET} = \sum (SET_i + \log SET_j) \cdot \quad (27)$$

Onde SET_i e SET_j são valores originados das cargas atômicas líquidas de cada átomo de carbono i e os j carbonos adjacentes a i .

Os valores de SET_i para os alcanos são obtidos através de uma correlação linear simples obtida a partir da média aritmética dos valores calculados das cargas atômicas líquidas (δ_i) dos átomos de carbono primário ($\text{CH}_3 = -0,210$), secundário ($\text{CH}_2 = -0,158$), terciário ($\text{CH} = -0,101$) e quaternário ($\text{C} = -0,041$) em diferentes moléculas e os valores atribuídos aos átomos de carbono primário ($\text{CH}_3 = 1,000$), secundário ($\text{CH}_2 = 0,900$), terciário ($\text{CH} = 0,800$) e quaternário ($\text{C} = 0,700$). Com esta correlação foi possível obter os valores dos fragmentos (SET_i) a partir da carga atômica, não sendo mais necessário consultar tabelas e posteriormente aplicá-los para outras funções.

4.2.1. Cálculo da carga atômica líquida

As estruturas moleculares foram desenhadas através do pacote computacional Hyperchem¹⁵⁷, onde foi encontrada a conformação de menor energia para cada molécula usando o método de mecânica molecular (MM_+) e posteriormente o método semi-empírico mecânico-quântico (AM1)¹⁵⁷⁻¹⁵⁸ com Algoritmo Polak-Ribiere e gradiente de minimização, onde o limite de convergência foi 0,0001 e o gradiente RMS de $0,0001 \text{ kcalA}^{-1}\text{mol}^{-1}$. A carga atômica líquida dos átomos de carbono foi calculada pela análise populacional de Mulliken¹⁵⁹.

A análise populacional de Mulliken permite atribuir uma carga líquida para cada átomo (δ_A) na molécula. O valor desta carga líquida é a diferença entre a carga eletrônica do átomo isolado (Z_A) e a carga calculada no átomo ligado (Q_A)¹⁵⁸⁻¹⁵⁹, isto é,

$$\delta_A = Z_A - Q_A. \quad (28)$$

O método semi-empírico AM1 tem sido amplamente utilizado em compostos orgânicos^{128,160}, pois além de ser computacionalmente mais rápido, como por exemplo, que o método PM3, fornece cargas líquidas que diferenciam melhor os carbonos CH₃, CH₂, CH e C, sendo a faixa de valores de cargas líquida obtidas com o AM1 mais ampla que a obtida com o PM3.

Apesar do fato de que o cálculo das cargas atômicas ser menos confiável do que outras propriedades moleculares, e de que diferentes métodos semi-empíricos fornecem valores para as cargas com uma pequena diferença numérica, é importante reconhecer que o seu cálculo é fácil e indicam as tendências da distribuição da densidade de carga nas moléculas, além de um maior poder de discriminação entre isômeros. Uma vez que muitas reações químicas ou propriedades físico-químicas dependem fortemente da densidade eletrônica local, cargas atômicas parciais e descritores baseados na carga são atualmente utilizados como índices de reatividade química. Por exemplo, cargas líquidas atômicas podem ser consideradas como índices de reatividade não-direcional¹⁶¹.

4.2.2. Alcanos e alcenos

A correlação linear simples entre os valores da média aritmética das cargas atômicas líquidas (δ_i) e os valores atribuídos aos átomos de carbono primário, secundário, terciário e quaternário podem ser visto na Equação (27) e Figura 4.

$$SET_i = -1,7712\delta_i + 0,6242 \quad (29)$$

$$r = 0,9995; s = 0,0050; N = 4$$

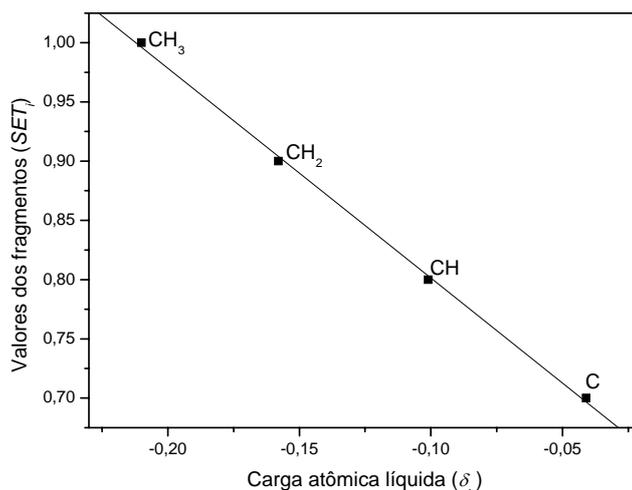


Figura 4: Regressão linear simples entre os valores de SET_i e as cargas atômicas líquidas (δ_i)

O cálculo do índice semi-empírico eletrotológico (I_{SET}) para cada molécula das funções alcanos e alcenos é feito com a Equação (27) com os valores de SET_i obtidos com a Equação (29) sendo as cargas atômicas líquidas obtidas via cálculos AM1 e análise populacional de Mulliken.

4.2.3. Aldeídos e cetonas

No que concerne a retenção cromatográfica de funções orgânicas tais como cetonas e aldeídos, as quais são moléculas polares, deve ser observado que somente as cargas atômicas líquidas não vão responder pelas interações dessas moléculas com a fase estacionária. Assim, devemos levar em conta a existência de um momento dipolar, até então desprezível para as moléculas de alcanos e alcenos. Considerando a fase estacionária recoberta com material de baixa polaridade, observamos que as interações entre a fase estacionária e as moléculas de alcanos e alcenos são eletrostáticas com contribuição de forças dispersivas. Em moléculas polares essas interações aumentam lentamente em relação às interações correspondentes aos hidrocarbonetos. Observa-se que a retenção cromatográfica das moléculas 2-hexanona, 3-hexanona e hexanal fornecem valores experimentais de retenção de 767, 764 e 776, respectivamente. Para o hidrocarboneto correspondente, o heptano, o índice de retenção é 700. Devido à presença do heteroátomo (oxigênio) há um aumento em torno de 10% no índice de retenção. Portanto, as interações

entre moléculas polares e a fase estacionária aumentam lentamente e isto claramente deve-se a redistribuição de carga que ocorre devido à presença do heteroátomo. As forças dispersivas entre estas moléculas e a fase estacionária incluem interações carga-dipolo e dipolo-dipolo induzido, as quais são fracas em relação às forças de interação eletrostáticas. Assim, as distribuições dipolares de carga em tais moléculas levam a um pequeno aumento nas interações do soluto com a fase estacionária em relação ao caso dos hidrocarbonetos onde o momento de dipolo é praticamente zero. A princípio, a partir das informações dadas acima é possível inferir que o índice de retenção possa ser calculado usando a Equação (27) e aplicar a Equação (28) também para heteroátomos, mas incluindo alterações sutis que associam os efeitos das interações dipolares dispersivas. Claramente o maior efeito na distribuição de carga devido a presença de heteroátomos ocorre em suas vizinhanças e o excesso de carga nestes átomos acarretam em interações eletrostáticas mais intensas que as fracas interações dispersivas dipolares. Todos esses fatos podem ser incluídos nos cálculos do índice de retenção por meio de um pequeno acréscimo nos valores de SET_i para heteroátomos e átomos de carbono ligados a eles. Uma maneira de obtermos um pequeno acréscimo nos valores de SET_i é levarmos em conta o momento de dipolo molecular, o qual tem sido usado como um descritor químico-quântico em estudos de QSPR/QSAR^{2,54}. Entretanto, verificou-se que somente a introdução do momento de dipolo não é suficiente para determinar o comportamento cromatográfico destas moléculas. Assim, foi necessário introduzir um momento de dipolo equivalente local associado às cargas líquidas dos átomos pertencentes ao grupo carbonila que contribuem para o aumento do valor da retenção. Isto foi efetivado multiplicando-se os valores de SET_i destes átomos por uma função A_μ a qual deve depender do momento de dipolo da molécula e da carga líquida dos átomos de carbono e oxigênio. Desde que $A_\mu = 1$, quando o momento dipolar é zero ou próximo a zero (como no caso dos alcanos e alcenos) uma primeira tentativa de se alcançar tal função é considerar uma dependência linear com o momento de dipolo molecular, μ , isto é, $A_\mu = 1 + (\mu / \mu_F)$ onde μ_F é uma função local (em unidades de momento de dipolo), isto é, um “momento de dipolo equivalente local”, no sentido de que depende das cargas líquidas nos átomos de carbono e oxigênio. Esta definição de A_μ se aplica somente se $\mu / \mu_F < 1$, desde que A_μ deve refletir apenas um pequeno acréscimo nas interações devido às forças dispersivas dipolares. Entretanto, boas escolhas para a definição de μ_F para cetonas e aldeídos (como veremos abaixo) implicam que a razão μ / μ_F pode ser

maior que a unidade, mostrando claramente que não é possível aplicar a definição acima para A_μ . Por outro lado, considerando-se que $\mu / \mu_F > 1$ então A_μ não pode ser uma função polinomial de μ / μ_F . Assim, A_μ deve ter uma dependência mais fraca com o momento de dipolo que uma dependência linear. Verificou-se que esta dependência tem um comportamento logarítmico, isto é, esta fraca dependência pode ser alcançada via uma função logarítmica, pois a função $f(x) = x$ cresce muito mais rapidamente que a função $f(x) = \log(1+x)$. Levando estes fatos em conta é possível encontrar uma definição de A_μ que difere levemente da unidade e que depende de forma logarítmica com momento de dipolo molecular, como na Equação (29),

$$A_\mu = 1 + \log\left(1 + \frac{\mu}{\mu_F}\right). \quad (30)$$

Onde μ é o momento de dipolo molecular calculado e μ_F é o momento de dipolo equivalente local, o qual depende das cargas dos átomos pertencentes à ligação C=O. Naturalmente, μ_F deve ser relacionado às cargas líquidas dos átomos de oxigênio desde que deve refletir alguma contribuição para as interações eletrostáticas entre estes átomos e a fase estacionária. Neste sentido μ_F pode ser relacionado à diferença das cargas líquidas dos átomos do grupo funcional C=O. Assim, μ_F pode ser definido de diferentes maneiras e algumas definições podem ser usadas em cálculos preliminares. Como esperado, para cetonas a melhor definição foi

$$\mu_F = d |\delta_C - \delta_O| \quad (31)$$

onde d é o comprimento da ligação C=O e $|\delta_C - \delta_O|$ é o valor absoluto da diferença das cargas líquidas destes átomos. Esta definição de μ_F é uma tentativa de se levar em conta a contribuição das interações eletrostáticas provenientes das cargas do oxigênio e carbono ligado a ele. Para os aldeídos o átomo de carbono da ligação C=O é terminal estando ligado a um hidrogênio. Assim, verificou-se a necessidade de se considerar a carga líquida positiva nesta região polar da molécula como sendo a soma das cargas líquidas nos átomos de carbono e hidrogênio. Portanto, para aldeídos a melhor definição para μ_F foi

$$\mu_F = d |\delta_C + \delta_H - \delta_O|. \quad (32)$$

Sendo assim, a Equação (30) mostra que há um aumento nas interações entre as moléculas e a fase estacionária devido a existência de um momento de dipolo, e que esta contribuição pode ser aumentada devido a carga dos heteroátomos se $\mu_F > \mu$, ou podem decrescer se

$\mu_F < \mu$. No caso das cetonas e aldeídos a função local μ_F é numericamente menor que o momento de dipolo mostrando que A_μ possui uma contribuição apreciável devido às cargas líquidas destes átomos. Isto revela a contribuição do oxigênio para as interações eletrostáticas entre o soluto e a fase estacionária.

Desta maneira, para incluir as interações dispersivas dipolares no cálculo do índice de retenção, multiplicamos os valores de SET_i do heteroátomo (oxigênio) e do carbono ligado a ele pela função dipolar A_μ , isto é, neste modelo o I_{SET} é calculado como especificado na Equação (33),

$$I_{SET} = \sum I_{SET_i} = \sum (A_\mu SET_i + \log A_\mu SET_j), \quad (33)$$

onde os valores de SET_i para os átomos de carbono e oxigênios são obtidos usando-se a Equação (29) (página 34) através de cálculos AM1 das cargas líquidas. Na expressão acima para I_{SET} a função dipolar A_μ é tomada como a unidade para o restante dos átomos de carbono das moléculas. A Equação (33) (I_{SET} para os aldeídos e cetonas) reduz à Equação (27) (I_{SET} para alcanos) quando o momento de dipolo é zero ou próximo de zero, como no caso dos alcanos e alcenos, desde que $A_\mu = 1$ para $\mu = 0$.

4.2.4. Ésteres alifáticos

Para ésteres os maiores efeitos na distribuição de carga são devidos à presença dos dois átomos de oxigênio e ocorrem em suas vizinhanças (carbonos adjacentes). O excesso de cargas nestes átomos implica em interações que são mais intensas que as fracas forças dispersivas. Assim de maneira similar ao caso das cetonas e aldeídos o cálculo de I_{SET} para ésteres deve incluir o momento de dipolo molecular e as cargas líquidas nos átomos de oxigênio por meio de um momento de dipolo equivalente local. Portanto aplicamos a Equação (45) onde os SET_i são obtidos via Equação (29) sendo que a função dipolar A_μ é calculada com a Equação (30). Mas agora devemos observar que μ_F pode, ou depender das cargas líquidas dos átomos de oxigênio, ou depender das cargas líquidas dos átomos pertencentes ao grupo $O_a=C-O_b-C$. Fica claro que existem diferentes maneiras de definir o momento de dipolo equivalente local. Ésteres possuem dois átomos de oxigênio e é

possível, a princípio, definirmos duas funções locais, uma dependente das cargas e do comprimento da ligação C=O_a e outra dependente das cargas e do comprimento da ligação C-O_b. Logo fica implícito a necessidade de efetuar alguns cálculos com diferentes definições para μ_F . Após alguns cálculos preliminares verificou-se que para ésteres a diferença de carga $\delta_o - \delta_c$ não fornece resultados razoáveis, como no caso de cetonas e aldeídos, o que pode ser entendido observando que os átomos de oxigênio blindam parcialmente a carga do carbono da carbonila.

Como esperado, para ésteres as melhores definições do momento de dipolo equivalente foram

$$\mu_{F1} = d_1 |\delta_{Oa}| \text{ e } \mu_{F2} = d_2 |\delta_{Ob}| \quad (34 \text{ e } 35)$$

onde d_1 e d_2 são os comprimentos de ligação das ligações C_a=O_a e C_a-O_b sendo $|\delta_{Oa}|$ e $|\delta_{Ob}|$ os valores absolutos das cargas líquidas nos átomos de oxigênio (O_a e O_b). O momento de dipolo equivalente local é então calculado como a soma vetorial de dois dipolos, isto é,

$$\mu_F = (\mu_{F1}^2 + \mu_{F2}^2 + 2\mu_{F1}\mu_{F2} \cos \theta)^{1/2}, \quad (36)$$

onde θ é o ângulo entre as ligações C=O_a e C-O_b. Entretanto para formiatos ocorre uma distribuição de carga específica na região polar destas moléculas e uma melhor definição para o momento local foi alcançada observando que o carbono da ligação C_a=O_a é terminal sendo ligado apenas a um hidrogênio. Assim foi necessário considerar as interações eletrostáticas provenientes das cargas nos átomos de oxigênio, carbono e hidrogênio pertencentes ao grupo HC_aO_aO_bC_{Al} (C_{Al} significa carbono no lado alcoólico do éster). Portanto, dipolos equivalentes foram construídos a partir das cargas líquidas dos grupos HC_aO_a, HC_aO_b e O_bC_{Al}. Os dipolos equivalentes associados a estas cargas líquidas são:

$$\mu_{F1} = d_1 |\delta_H + \delta_{Ca} - \delta_{Oa}|, \quad \mu_{F2} = d_2 |\delta_H + \delta_{Ca} - \delta_{Ob}| \text{ e } \mu_{F3} = d_3 |\delta_{CAl} - \delta_{Ob}| \quad (32, 32 \text{ e } 31)$$

onde d_1 e d_2 são os comprimentos calculados das ligações C_a=O_a e C_a-O_b e d_3 é o comprimento da ligação C_{Al}-O_b. Em uma primeira abordagem, os momentos locais μ_{F2} e μ_{F3} são tomados como colineares e um novo momento local é obtido pela diferença entre μ_{F2} e μ_{F3} , isto é,

$$\mu_{F4} = \mu_{F2} - \mu_{F3} \quad (37)$$

O momento de dipolo equivalente local final é definido como

$$\mu_F = (\mu_{F1}^2 + \mu_{F4}^2 + 2\mu_{F1}\mu_{F4} \cos \theta)^{1/2} \quad (36)$$

onde θ é o ângulo entre as ligações C=O_a e C-O_b. Logo para formiatos a carga do átomo de hidrogênio também deve ser considerada, como ocorre no caso dos aldeídos, pois nestes casos a carga no hidrogênio contribui de forma explícita para a carga positiva da região polar destas moléculas. Desta maneira a presente abordagem para o cálculo de I_{SET} considera importantes características topológicas e eletrônicas ($SET_i(\delta_i)$) em diferentes funções orgânicas, contidas também no momento de dipolo da molécula e nos momentos de dipolos equivalentes locais, como pode ser visto a partir das melhores definições encontradas para μ_F . Para ésteres μ_F é sempre menor que o momento de dipolo molecular. Assim, A_μ tem uma contribuição apreciável dos átomos envolvidos nas definições encontradas acima para μ_F , revelando em particular a contribuição dos átomos de oxigênio para as interações eletrostáticas entre soluto e fase estacionária.

A aplicação da Equação (33) para o cálculo de I_{SET} para ésteres mostra que esta equação pode ser tomada como uma definição geral do índice eletrotológico podendo ser aplicada para diferentes funções orgânicas, as quais são especificadas através de definições apropriadas do momento de dipolo equivalente local μ_F . Aplicações preliminares da Equação (33) em ésteres (com as devidas definições de μ_F acima mencionadas) mostraram que os cálculos superestimaram o índice de retenção para ésteres ramificados e subestimaram os resultados para ésteres metílicos. Estes fatos revelaram a necessidade de considerar outras definições de μ_F para ésteres ramificados e metílicos. Uma solução mais simples foi encontrada levando-se em consideração os efeitos estéreos em ésteres ramificados e metílicos. A simplicidade está em considerar o impedimento estéreo do átomo de carbono C_{Al} do grupo C_aO_aO_bC_{Al} e o átomo de carbono no lado ácido do grupo funcional COO (que denominamos aqui de carbono C_{Ac}). Como é sabido o fator $\log SET_j$ fornece precisamente o efeito estéreo do átomo j . Assim, para incluir uma correção estérea (ce) na Equação (33) para ésteres ramificados foi adicionado o termo

$$ce = n \log SET(C_{Ac}) + n \log SET(C_{Al}), \quad (38)$$

onde n é o número de ramificações do éster. Por outro lado, para ésteres metílicos o átomo de carbono C_{Al} é ligado a três átomos de H e é necessário remover o efeito superestimado do termo $\log A_{\mu} SET_j$ na Equação (33). Para ésteres metílicos isto é facilmente alcançado incluindo uma segunda correção estérea (sce) adicionando à Equação (33) o termo

$$sce = -\log SET(C_{Al}). \quad (39)$$

Resultados amplamente satisfatórios foram obtidos com esta abordagem revelando que neste modelo os complexos efeitos estéreo podem ser incluídos via impedimento estéreo dos dois carbonos ligados aos lados alcoólico e ácido do grupo funcional COO , usando-se as cargas líquidas a partir das quais são calculados os valores de SET_i .

Os cálculos de I_{SET} para alcanos, alcenos, aldeídos e cetonas foram efetivados via planilhas EXCEL. No caso das funções com heteroátomos (aldeídos e cetonas) algumas limitações/dificuldades foram encontradas neste procedimento face ao grande número de variáveis envolvidas nas definições de μ_F . Para ésteres, o grau de complexidade cresce devido à presença de dois heteroátomos, pois aumenta o número de variáveis, entre elas: cargas nos oxigênios, carbonos ligados a eles, assim como os respectivos comprimentos e ângulos de ligação. Isto implica na possibilidade de mais de uma definição de momento dipolar equivalente com caráter vetorial na mesma molécula. Tais dificuldades inviabilizam o uso racional das planilhas de EXCEL para o cálculo de I_{SET} de uma amostragem considerável de moléculas com este grau de complexidade. Sendo assim, foi desenvolvido um código FORTRAN (vide anexo I) para efetuar estes cálculos tendo por base a leitura dos dados de saída (cargas líquidas, momento de dipolo e posições atômicas) do pacote computacional HyperChem.

4.3. Análise de regressão e validação dos modelos de QSPR / QSRR

Todos os modelos de correlações lineares simples entre os índices de retenção experimentais (IR_{Exp}) e o I_{SET} , foram determinados usando o programa computacional *Origin*¹⁶².

Para testar a qualidade de previsão dos modelos obtidos foram utilizados como parâmetros estatísticos o coeficiente de correlação (r), o desvio padrão (s), o coeficiente de determinação (r^2) e o teste de hipótese nula (F -test) (f). Para testar o potencial e a validade de prognóstico dos modelos obtidos foi empregado o coeficiente de validação cruzada (r_{CV}^2) usando “*leave-one-out*” executado pelo programa *Tsar 3.3 for Windows*¹⁶³.

4.3.1. Validação cruzada

A validação cruzada é um método prático e confiável para testar a capacidade preditiva do modelo construído. Segundo a definição de Beebe, Pell e Seasholtz (1998)⁴⁵, a validação cruzada é:

“Um processo empregado para validar modelos onde um grupo de calibração é dividido em subgrupos de calibração e validação. Um modelo é construído com o subgrupo de calibração e é usado para prever o subgrupo de validação. Este processo é repetido utilizando diferentes subgrupos até que cada uma das amostras do conjunto tenha sido incluída em um subgrupo de validação. Os resultados estimados são então empregados para validar o desempenho do modelo.”

O método de validação cruzada é considerado como um teste de validação interno do modelo de correlação e as etapas envolvidas no seu desenvolvimento consistem, resumidamente, do seguinte¹⁶⁴:

- a) remove-se uma (“*leave-one-out*”) ou mais amostras i do conjunto de calibração e constrói-se o modelo como anteriormente;
- b) usa-se o novo modelo para prever os dados removidos \hat{c}_i ;
- c) calcula-se o r_{CV}^2 definido pela equação¹⁰⁸:

$$r_{CV}^2 = 1,0 - \frac{\sum_{i=1}^N (c_i - \hat{c}_i)^2}{\sum_{i=1}^N (c_i - \bar{c}_i)^2}; \quad (40)$$

onde N é o nº de compostos; c_i e \hat{c}_i são os valores experimentais e previstos, respectivamente; \bar{c}_i é o valor médio de c_i . No entanto, gráficos dos valores calculados em função dos experimentais e gráficos dos valores residuais em função dos calculados, da

propriedade de interesse, também podem ser usados como evidência da validade de um modelo de QSPR¹⁶⁷⁻¹⁶⁸.

4.3.2. Validação externa dos modelos

A estabilidade e validade do modelo podem também ser avaliadas através de testes de validação externa^{54,123,167-168}. Neste estudo, esta validação externa foi adicionalmente empregada para alguns grupos de compostos investigados.

Um dos testes consiste em dividir aleatoriamente o conjunto ou grupo original de dados em um subgrupo de treinamento e outro subgrupo de teste externo. Posteriormente, um novo modelo de QSRR é construído com o subgrupo de treinamento, sendo este então empregado para prever a propriedade do subgrupo de teste externo. Um grupo de compostos externos, que não se encontravam no conjunto original da construção do modelo, também pode ser empregado como grupo de teste na validação externa.

Outra sistemática de validação externa empregada foi a de dividir sistematicamente o conjunto original de dados em três subgrupos diferentes, sendo que o primeiro contém os compostos 1, 4, 7 etc. A propriedade de cada subgrupo foi então prevista (grupo de teste) através do modelo obtido utilizando os dois subgrupos restantes como grupo de treinamento³³.

5. Resultados e discussões

Os resultados pertinentes, bem como as discussões, obtidos neste estudo serão apresentados separadamente para a melhor definição de μ_F alcançada para estimar a retenção cromatográfica de cada função.

5.1. Aplicação do I_{SET} para estimar a retenção cromatográfica de alcanos e alcenos

Neste estudo 179 alcanos e alcenos foram considerados, sendo as cargas líquidas calculadas via análise populacional de Mulliken e posteriormente foram obtidos os SET_i a partir da Equação (29) (página 34). Os valores de I_{SET} de cada molécula são obtidos segundo a Equação (27) (página 33).

A seguir, exemplificamos o cálculo do I_{SET} para as moléculas *cis*-penten-2-eno e *trans*-penten-2-eno representadas pelos grafos químicos abaixo. Os valores das cargas atômicas líquidas e os respectivos SET_i são mostrados na Tabela 3.

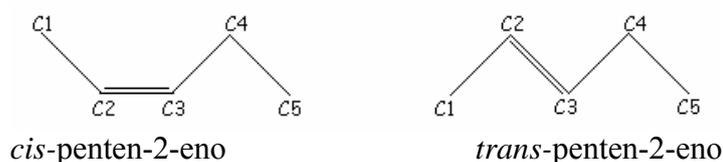


Tabela 3: Os valores de carga atômica líquida (δ_i) e SET_i para cada átomo de carbono das moléculas *cis*-penten-2-eno e *trans*-penten-2-eno.

Átomos de carbono	<i>cis</i> -penten-2-eno		<i>trans</i> -penten-2-eno	
	δ_i	SET_i	δ_i	SET_i
C1	-0,187	0,9555	-0,185	0,9520
C2	-0,173	0,9307	-0,170	0,9254
C3	-0,165	0,9165	-0,167	0,9200
C4	-0,133	0,8598	-0,132	0,8580
C5	-0,209	0,9945	-0,207	0,9910

cis-penten-2-eno

$$I_{SETC1} = 0,9555 + \log 0,9307$$

$$I_{SETC2} = 0,9307 + \log 0,9555 + \log 0,9165$$

$$I_{SETC3} = 0,9165 + \log 0,9307 + \log 0,8598$$

$$I_{SETC4} = 0,8598 + \log 0,9165 + \log 0,9945$$

$$I_{SETC5} = 0,9945 + \log 0,8598$$

$$I_{SET} = 4,3653$$

trans-penten-2-eno

$$I_{SETC1} = 0,9520 + \log 0,9254$$

$$I_{SETC2} = 0,9254 + \log 0,9520 + \log 0,9200$$

$$I_{SETC3} = 0,9200 + \log 0,9254 + \log 0,8580$$

$$I_{SETC4} = 0,8580 + \log 0,9200 + \log 0,9910$$

$$I_{SETC5} = 0,9910 + \log 0,8580$$

$$I_{SET} = 4,3481$$

O cálculo AM1 revela que as estruturas otimizadas dos isômeros *-cis* e *-trans* têm uma ligeira diferença na distribuição de carga. Como pode ser visto acima, os resultados da análise de população de Mulliken mostram a carga atômica líquida de cada átomo de carbono para cada isômero, o que implica que os valores de fragmentos, SET_i , diferem o suficiente para resultar em valores diferentes de I_{SET} ¹⁶⁹.

Em uma etapa preliminar, buscou-se a correlação entre a retenção cromatográfica e o I_{SET} ignorando os efeitos estéreos dos átomos de carbono adjacentes, isto é, calculando o I_{SET} por meio da Equação (27), porém sem o termo $\log SET_j$. A correlação assim obtida está expressa na Equação (41):

$$IR_{Exp} = -40,3887 + 109,5053I_{SET} \quad (41)$$

$$r = 0,9941; s = 26,05; N = 179; f = 14933; r^2 = 0,9887; r_{CV}^2 = 0,9881$$

Quando os efeitos estéreos são considerados, isto é, incluindo o termo $\log SET_j$, os resultados obtidos podem ser observados na Equação (42):

$$IR_{Exp} = -36,8298 + 120,8252I_{SET} \quad (42)$$

$$r = 0,9990; s = 10,74; N = 179; f = 88782; r^2 = 0,9980; r_{CV}^2 = 0,9980.$$

Como pode ser observado, a inclusão dos efeitos estéreos dos átomos de carbono adjacentes produz uma melhora significativa nos valores de r^2 , s e r_{CV}^2 , mostrando que esses efeitos devem ser considerados.

Os valores de I_{SET} calculados para cada um dos 179 compostos entre alcanos e alcenos e os índices de retenção experimentais estão mostrados na Tabela 4¹⁷¹.

Tabela 4: Valores dos índices de retenção experimentais, IR_{Exp} , (esqualano 100 C°) e dos índices semi-empírico eletrotopológico, I_{SET} , para diferentes alcanos e alcenos¹⁶⁹

N°	Composto	IR_{Exp}	I_{SET}	N°	Composto	IR_{Exp}	I_{SET}
1	Butano	400,0	3,6343	46	3,4-dimetilundecano	1247,0	10,7082
2	Pentano	500,0	4,4457	47	4,4-dimetilheptano	828,6	7,2960
3	Hexano	600,0	5,2622	48	4,5-dimetiloctano	943,1	8,2708
4	Heptano	700,0	6,0787	49	2,2,3-trimetilbutano	641,1	5,5324
5	Octano	800,0	6,8952	50	2,2,3-trimetilpentano	738,6	6,3524
6	Nonano	900,0	7,7117	51	2,2,3-trimetilhexano	823,3	7,1627
7	Decano	1000,0	8,5282	52	2,2,3-trimetilheptano	914,4	7,9685
8	Undecano	1100,0	9,3447	53	2,3,3-trimetilhexano	841,7	7,1492
9	Dodecano	1200,0	10,1612	54	2,3,3-trimetilheptano	931,7	7,9570
10	2-metilbutano	475,4	4,3303	55	2,3,4-trimetilhexano	849,7	7,3035
11	2-metilpentano	569,8	5,1448	56	2,3,4-trimetilheptano	933,4	8,1151
12	2-metilhexano	666,8	5,9552	57	2,4,4-trimetilhexano	809,7	7,1714
13	2-metilheptano	765,0	6,7717	58	3,3,4-trimetilheptano	936,6	7,9610
14	2-metiloctano	864,8	7,5882	59	3,3,5-trimetilheptano	907,7	7,9793
15	2-metiononano	963,9	8,4047	60	3,4,4-trimetilheptano	932,2	7,9626
16	2-metilundecano	1164,0	10,0377	61	3,4,5-trimetilheptano	945,0	8,1111
17	3-metilpentano	584,6	5,1311	62	2,2,3,4-tetrametilpentano	821,9	6,9948
18	3-metilhexano	676,5	5,9501	63	2,2,3,4-tetrametilhexano	908,8	7,8065
19	3-metilheptano	772,6	6,7641	64	2,2,4,4-tetrametilpentano	774,6	6,8663
20	3-metiloctano	870,8	7,5806	65	2,2,4,4-tetrametilhexano	888,6	7,6537
21	3-metilnonano	969,6	8,3971	66	2,3,3,5-tetrametilhexano	903,3	7,8266
22	3-metilundecano	1169,6	10,0300	67	2,3,4,5-tetrametilhexano	923,1	7,9745
23	4-metilheptano	767,4	6,7692	68	2,2-dimetil-3-etilpentano	824,4	7,1231
24	4-metiloctano	863,3	7,5831	69	2,2-dimetil-3-etilhexano	902,1	7,9279
25	4-metilnonano	960,0	8,3996	70	2,3-dimetil-4-etilhexano	949,4	8,1185
26	4-metilundecano	1158,6	10,0326	71	2,4-dimetil-3-etilpentano	838,4	7,2839
27	5-metilnonano	957,4	8,3924	72	2,4-dimetil-3-etilhexano	929,8	8,0963
28	6-metilundecano	1151,8	10,0254	73	2,4-dimetil-4-etilhexano	920,7	7,9610
29	3-etilpentano	686,6	5,9193	74	2,4-dimetil-3-	915,1	7,9389
30	3-etilheptano	867,4	7,5522	75	3,4-dietilhexano	945,8	8,2542
31	3-etiloctano	964,0	8,3687	76	2-etil-2-metilheptano	953,0	8,1018
32	4-etilheptano	858,2	7,5573	77	3-etil-2-metilpentano	762,4	6,6057
33	4-etiloctano	951,5	8,3713	78	3-etil-2-metilhexano	844,4	7,4511
34	2,2-dimetilbutano	537,6	4,8756	79	3-etil-2-metilheptano	941,0	8,2374
35	2,3-dimetilbutano	568,1	5,0112	80	3-etil-2,2,4-trimetilpentano	903,9	7,7677
36	2,3-dimetilpentano	672,5	5,8209	81	3-etil-3-metilhexano	856,0	7,2536
37	2,3-dimetilhexano	760,8	6,6353	82	3-etil-4-metilhexano	855,6	7,4159
38	2,3-dimetilheptano	855,5	7,4493	83	3-etil-4-metilheptano	940,5	8,2349
39	2,3-dimetiloctano	952,1	8,2657	84	4-etil-4-metilheptano	937,6	8,0777
40	2,3-dimetilundecano	1251,4	10,7152	85	4-etil-3-metilheptano	935,7	8,2349
41	3,3-dimetilpentano	660,2	5,6641	86	2-metil-1-propeno	383,0	3,4377
42	3,3-dimetilheptano	837,5	7,2970	87	Buten-1-eno	384,0	3,5874
43	3,3-dimetiloctano	932,0	8,1100	88	<i>Trans</i> -buten-2-eno	405,0	3,5771
44	3,4-dimetilhexano	771,6	6,6279	89	<i>Cis</i> -buten-2-eno	417,0	3,5754
45	3,4-dimetilheptano	858,0	7,4423	90	Penten-1-eno	482,0	4,4010

Continuação

Nº	Composto	IR_{Exp}	I_{SET}	Nº	Composto	IR_{Exp}	I_{SET}
91	2-metil-buten-1-eno	488,0	4,2011	135	<i>Cis</i> -nonen-3-eno	887,5	7,5928
92	<i>Trans</i> -penten-2-eno	499,0	4,3481	136	<i>Trans</i> -nonen-2-eno	896,4	7,6094
93	<i>Cis</i> -penten-2-eno	504,0	4,3653	137	<i>Cis</i> -nonen-2-eno	901,9	7,6311
94	<i>Cis</i> -4-metil-penten-2-eno	557,0	4,9483	138	<i>Cis</i> -decen-5-eno	981,6	8,3624
95	2,3-dimetil-buten-1-eno	560,0	4,8500	139	Decen-1-eno	982,5	8,4774
96	<i>Trans</i> -4-metil-penten-2-eno	562,0	4,9806	140	<i>Trans</i> -decen-4-eno	982,5	8,3784
97	2-metil-penten-1-eno	581,0	5,0189	141	<i>Cis</i> -decen-4-eno	982,8	8,3675
98	Hexen-1-eno	583,0	5,2140	142	<i>Trans</i> -decen-5-eno	984,1	8,3810
99	<i>Trans</i> -hexen-3-eno	592,0	5,1192	143	<i>Trans</i> -decen-3-eno	985,8	8,3822
100	<i>Cis</i> -hexen-3-eno	594,6	5,1572	144	<i>Cis</i> -decen-3-eno	985,8	8,4212
101	<i>Ttrans</i> -hexen-2-eno	597,0	5,1608	145	<i>Trans</i> -decen-2-eno	996,7	8,4242
102	<i>Cis</i> -hexen-2-eno	604,0	5,1799	146	<i>Cis</i> -decen-2-eno	1001,7	8,4511
103	<i>Trans</i> -4,4-dimetil-penten-2-eno	615,0	5,5038	147	<i>Cis</i> -undecen-5-eno	1078,2	9,1849
104	3,3-dimetil-penten-1-eno	630,0	5,5372	148	<i>Cis</i> -undecen-4-eno	1080,5	9,1779
105	<i>Cis</i> -4,4-dimetil-penten-2-eno	639,0	5,4580	149	<i>Trans</i> -undecen-4-eno	1081,1	9,1905
106	2,4-dimetil-penten-1-eno	640,0	5,7077	150	<i>Trans</i> -undecen-5-eno	1081,8	9,1984
107	3,4-dimetil-penten-1-eno	641,0	5,7156	151	Undecen-1-eno	1082,4	9,3041
108	2,3-dimetil-penten-1-eno	654,0	5,6610	152	<i>Cis</i> -undecen-3-eno	1085,3	9,2223
109	<i>Cis</i> -4-metil-hexen-2-eno	657,0	5,8067	153	<i>Trans</i> -undecen-3-eno	1085,4	9,2013
110	4-metil-hexen-1-eno	657,0	5,9009	154	<i>Cis</i> -dodecen-6-eno	1175,0	9,9978
111	<i>Trans</i> -4-metil-hexen-2-eno	667,0	5,7763	155	<i>Cis</i> -dodecen-5-eno	1175,6	9,9988
112	2-metil-hexen-1-eno	677,0	5,8303	156	<i>Trans</i> -dodecen-6-eno	1179,6	10,0278
113	Hepten-1-eno	683,0	6,0305	157	<i>Cis</i> -dodecen-4-eno	1179,6	9,9943
114	<i>Trans</i> -hepten-3-eno	688,0	5,9353	158	<i>Trans</i> -dodecen-5-eno	1180,6	10,0089
115	<i>Cis</i> -hepten-3-eno	691,0	5,9717	159	<i>Trans</i> -dodecen-4-eno	1180,6	10,0105
116	<i>Trans</i> -hepten-2-eno	698,0	5,9774	160	Dodecen-1-eno	1183,0	10,1215
117	<i>Cis</i> -hepten-2-eno	704,7	6,0017	161	<i>Trans</i> -dodecen-3-eno	1185,1	10,0083
118	2,4,4-trimetil-penten-1-eno	708,0	6,2349	162	<i>Cis</i> -dodecen-3-eno	1185,1	10,0388
119	2,4,4-trimetil-penten-2-eno	718,0	6,0831	163	<i>Cis</i> -tridecen-6-eno	1271,2	10,8134
120	<i>Cis</i> -2,2-dimetil-hexen-3-eno	720,0	6,2575	164	<i>Cis</i> -tridecen-5-eno	1273,8	10,8144
121	2,3-dimetil-hexen-1-eno	741,0	6,4809	165	<i>Trans</i> -tridecen-6-eno	1277,4	10,8477
122	2,5-dimetil-hexen-2-eno	751,0	6,5114	166	<i>Trans</i> -tridecen-4-eno	1279,9	10,8270
123	<i>Trans</i> -2-metil-hepten-3-eno	757,0	6,5846	167	<i>Trans</i> -tridecen-5-eno	1279,5	10,8305
124	Octen-1-eno	782,0	6,8606	168	Tridecen-1-eno	1283,1	10,9346
125	<i>Trans</i> -octen-4-eno	784,0	6,7515	169	<i>Cis</i> -tridecen-3-eno	1284,4	10,8681
126	<i>Cis</i> -octen-4-eno	784,2	6,7345	170	<i>Trans</i> -tridecen-3-eno	1284,9	10,8377
127	<i>Trans</i> -octen-3-eno	788,2	6,7544	171	<i>Cis</i> -tetradecen-7-eno	1366,7	11,6248
128	<i>Cis</i> -octen-3-eno	789,8	6,7729	172	<i>Cis</i> -tetradecen-6-eno	1368,6	11,6238
129	<i>Trans</i> -octen-2-eno	797,0	6,7938	173	<i>Cis</i> -tetradecen-5-eno	1372,0	11,6283
130	<i>Cis</i> -octen-2-eno	802,0	6,8241	174	<i>Trans</i> -tetradecen-7-eno	1374,5	11,6488
131	Nonen-1-eno	882,5	7,6609	175	<i>Trans</i> -tetradecen-6-eno	1375,7	11,6547
132	<i>Trans</i> -nonen-4-eno	884,2	7,5610	176	<i>Cis</i> -tetradecen-4-eno	1377,7	11,6274
133	<i>Cis</i> -nonen-4-eno	885,4	7,6027	177	<i>Trans</i> -tetradecen-5-eno	1378,4	11,6469
134	<i>Trans</i> -nonen-3-eno	886,4	7,5743	178	<i>Trans</i> -tetradecen-4-eno	1379,3	11,6374
				179	Tetradecen-1-eno	1383,2	11,7545

Como pode ser observado através dos parâmetros estatísticos, este modelo explica mais de 99% da variação nos valores dos índices de retenção experimentais para este conjunto de compostos. Esse é um bom resultado considerando que os dados experimentais são oriundos de diferentes laboratórios.

Os parâmetros estatísticos obtidos com o novo descritor, I_{SET} , são comparáveis aos alcançados anteriormente para 157 alcanos usando os descritores ${}^3W_{CH}$ ($r=0,9862$ e $s=31,0$), ${}^1\chi$ ($r=0,9902$ e $s=26,0$) e I_{ET} ($r=0,9901$ e $s=26,2$). Para 79 alcenos lineares o descritor I_{ET} apresentou $r=0,9999$ e $s=2,35$. Porém neste caso, houve a necessidade de atribuir valores diferentes aos fragmentos para as várias posições da dupla ligação¹³.

A correlação entre os dados de retenção experimentais (IR_{Exp}) e calculados (IR_{Calc}) para todos os compostos são mostrados na Figura 5a. Os valores residuais foram plotados contra os calculados para verificar as suas relações (Figura 5b).

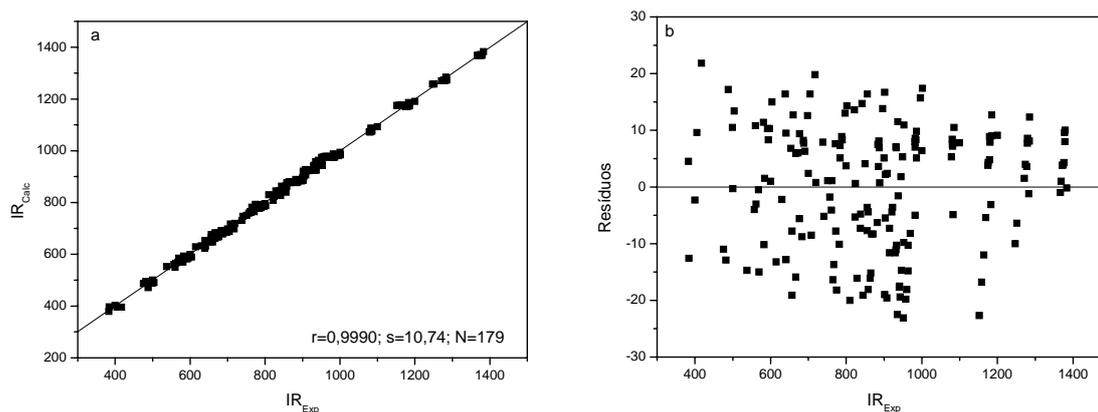


Figura 5: Índices de retenção experimentais (IR_{Exp}) versus índices de retenção calculados (IR_{Calc}) para os dados de 179 alcanos e alcenos¹⁷¹

Como pode ser observado na Figura 5b, os valores dos resíduos, como função dos índices de retenção calculados, distribuem-se aleatoriamente em torno da origem, indicando que o modelo é adequado.

Constata-se que 67,03% dos compostos têm valores residuais menor que 10 unidades ($s = 10,74$) e 97,8% revelam valores residuais inferiores a 20 unidades de retenção ($\approx 2s$). Quatro compostos (6-metilundecano, 4-etiloctano, 4-etil-3-metilheptano e 2-cis-buteno), correspondendo a 2,2% dos dados amostrados têm valores residuais acima de $2s$.

A fim de verificar a validade estatística do modelo, todo o conjunto de dados foi sistematicamente dividido em três subgrupos e cada subconjunto foi predito usando os

outros dois subconjuntos como grupo de treinamento. Os resultados estão apresentados na Tabela 5.

Tabela 5: Verificação da validade estatística do modelo de QSRR obtido para alcanos e alcenos¹⁶⁹

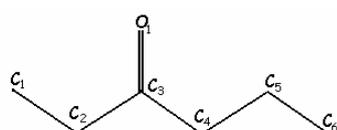
Grupo de treinamento	r^2	Grupo de teste	r^2
1 e 2	0,9982	3	0,9977
1 e 3	0,9980	2	0,9980
2 e 3	0,9983	1	0,9978
Média	0,9982	Média	0,9978

Os resultados da Tabela 5 mostraram que o coeficiente de determinação médio do grupo de treinamento ($r^2 = 0,9980$) e o coeficiente de determinação médio do grupo prognosticado ($r^2 = 0,9980$), indicam que o modelo proposto tem uma elevada estabilidade e validade estatística.

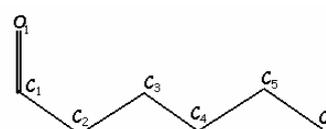
Então, o modelo de QSRR para 179 alcanos e alcenos rapidamente obtido com o I_{SET} usando as cargas atômicas líquidas calculadas a partir do método químico-quântico AM1, possui parâmetros estatísticos de boa qualidade. Assim, a unificação do método químico-quântico com o topológico decodifica as informações mais relevantes contidas na estrutura tridimensional da molécula, sendo capaz de discriminar isômeros geométricos, como alcenos *-cis* e *-trans*.

5.2 Aplicação do I_{SET} para estimar a retenção cromatográfica de aldeídos e cetonas

As moléculas 3-hexanona e hexanal representadas no grafo molecular abaixo foram tomadas como exemplos de cálculo do I_{SET} usando o presente método. Os valores de cargas atômicas líquidas e SET_i são dadas na Tabela 6.



3-hexanona



hexanal $\delta_H = +0.086$

Tabela 6: Os valores de carga atômica líquida (δ_i) e SET_i para cada átomo de carbono e oxigênio das moléculas 3-hexanona e hexanal¹⁷⁰

Átomos	3-hexanona			hexanal		
	δ_i	SET_i	$A_\mu SET_i$	δ_i	SET_i	$A_\mu SET_i$
O ₁	-0,288	1,1346	1,9507	-0,289	1,1363	1,9328
C ₁	-0,206	0,9892	-	+0,183	0,2995	0,5094
C ₂	-0,212	0,9998	-	-0,233	1,0371	-
C ₃	+0,224	0,2268	0,3899	-0,155	0,8988	-
C ₄	-0,212	0,9998	-	-0,158	0,9041	-
C ₅	-0,155	0,8988	-	-0,159	0,9059	-
C ₆	-0,212	0,9998	-	-0,211	0,9980	-

3-hexanona

$$\mu_F = d |\delta_C - \delta_O| \quad (31)$$

$$\mu_F = 1,2342 |0,224 - [-0,288]| = 0,6319$$

hexanal

$$\mu_F = d |\delta_C + \delta_H - \delta_O| \quad (32)$$

$$\mu_F = 1,2314 |0,183 + 0,086 - [-0,289]| = 0,6871$$

$$A_\mu = 1 + \log\left(1 + \frac{\mu}{\mu_F}\right) \quad (30)$$

$$A_\mu = 1 + \log[1 + (2,6790/0,6319)] = 1,7193 \quad \left| \quad A_\mu = 1 + \log[1 + (2,7640/0,6871)] = 1,7009$$

$$I_{SET} = \sum I_{SET_i} = \sum (A_\mu SET_i + \log A_\mu SET_i) \quad (33)$$

3-hexanona

$$I_{SETO1} (=O) = A_\mu SET_{O1} + \log A_\mu SET_{C3} = 1,9507 + \log 0,3899 = 1,5416$$

$$I_{SETC1} (-CH_3) = SET_{C1} + \log SET_{C2} = 0,9892 + \log 0,9998 = 0,9891$$

$$I_{SETC2} (-CH_2-) = SET_{C2} + \log SET_{C1} + \log A_\mu SET_{C3} = 0,9998 + \log 0,9892 + \log 0,3899 = 0,5860$$

$$I_{SETC3} (>C<) = A_\mu SET_{C3} + \log SET_{C2} + \log A_\mu SET_{O1} + \log SET_{C4} = 0,3899 + \log 0,9998 + \log 1,9507 + \log 0,9998 = 0,6799$$

$$I_{SETC4} (-CH_2-) = SET_{C4} + \log A_\mu SET_{C3} + \log SET_{C5} = 0,9998 + \log 0,3899 + \log 0,8988 = 0,5444$$

$$I_{SETC5} (-CH_2-) = SET_{C5} + \log SET_{C4} + \log SET_{C6} = 0,8988 + \log 0,9998 + \log 0,9998 = 0,8986$$

$$I_{SETC6} (-CH_3) = SET_{C6} + \log SET_{C5} = 0,9998 + \log 0,8988 = 0,9535$$

$$I_{SET} = 1,5416 + 0,9891 + 0,5860 + 0,6799 + 0,5444 + 0,8986 + 0,9535 = 6,1931$$

hexanal

$$I_{SET01} (=O) = A_{\mu}SET_{01} + \log A_{\mu}SET_{C1} = 1,9328 + \log 0,5094 = 1,6398$$

$$I_{SETC1} (=CH-) = A_{\mu}SET_{C1} + \log A_{\mu}SET_{01} + \log SET_{C2} = 0,5094 + \log 1,9328 + \log 1,0371 = 0,8114$$

$$I_{SETC2} (-CH_2-) = SET_{C2} + \log A_{\mu}SET_{C1} + \log SET_{C3} = 1,0371 + \log 0,5094 + \log 0,8988 = 0,6978$$

$$I_{SETC3} (-CH_2-) = SET_{C3} + \log SET_{C2} + \log SET_{C4} = 0,8988 + \log 1,0371 + \log 0,9041 = 0,8708$$

$$I_{SETC4} (-CH_2-) = SET_{C4} + \log SET_{C3} + \log SET_{C5} = 0,9041 + \log 0,8988 + \log 0,9059 = 0,8148$$

$$I_{SETC5} (-CH_2-) = SET_{C5} + \log SET_{C4} + \log SET_{C6} = 0,9059 + \log 0,9041 + \log 0,9980 = 0,8612$$

$$I_{SETC6} (-CH_3) = SET_{C6} + \log SET_{C5} = 0,9980 + \log 0,9059 = 0,9550$$

$$I_{SET} = 1,6398 + 0,8114 + 0,6978 + 0,8708 + 0,8148 + 0,8612 + 0,9550 = 6,6508$$

A Tabela 7 mostra os valores calculados de I_{SET} usando a Equação (30) e os índices de retenção experimentais para 15 aldeídos e 40 cetonas.

Tabela 7: Valores experimentais do índice de retenção (IR_{Exp})³⁸ (HP-1 a 50C°) e I_{SET} para diferentes aldeídos e cetonas¹⁷⁰.

Nº	Composto	IR_{Exp}	I_{SET}	Nº	Composto	IR_{Exp}	I_{SET}
1	Propanona	469,7	4,0158	29	2-dodecanona	1369,5	11,0872
2	2-butanona	574,7	4,5952	30	7-tridecanona	1447,6	11,8516
3	3-metil-2-butanona	639,9	5,2258	31	2-tridecanona	1469,7	11,9027
4	2-pentanona	665,4	5,3987	32	2-tetradecano	1570,0	12,7170
5	3-pentanona	675,4	5,3900	33	8-pentadecanona	1646,7	13,4817
6	2,2-dimetil-3-butanona	691,8	5,8039	34	2-pentadecanona	1670,2	13,5319
7	4-metil-2-pentanona	720,1	6,0484	35	2-hexadecanona	1770,5	14,2179
8	2-metil-3-pentanona	731,9	6,0227	36	9-heptadecanona	1845,9	15,1070
9	3-metil-2-pentanona	733,5	6,0746	37	2-heptadecanona	1870,8	15,1706
10	3-hexanona	764,1	6,1931	38	2-octadecanona	1971,0	15,9849
11	2-hexanona	767,0	6,1987	39	10-nonadecanona	2044,9	16,7312
12	5-metil-3-hexanona	815,6	6,8401	40	2-nonadecanona	2071,3	16,8010
13	2-metil-3-hexanona	818,9	6,8444	41	acetaldeído	360,4	3,3967
14	5-metil-2-hexanona	835,4	6,8815	42	Isobutanal	540,3	4,8898
15	4-heptanona	851,8	7,0130	43	Butanal	571,1	5,0052
16	3-heptanona	864,9	6,9897	44	Isovaleraldeído	635,0	5,7104
17	2-heptanona	867,5	7,0080	45	2-metil-butanal	645,3	5,7028
18	2-metil-3-heptanona	917,5	7,6230	46	Valeraldeído	674,4	5,8293
19	5-metil-3-heptanona	921,7	7,6145	47	2-metil-valeraldeído	729,5	6,5245
20	3-octanona	964,8	7,8049	48	2-etil-butanal	742,1	6,5319
21	2-octanona	967,5	7,8306	49	Hexanal	776,5	6,6508
22	5-nonanona	1051,4	8,5885	50	Heptanal	877,2	7,4709
23	2-nonanona	1068,7	8,6458	51	Octanal	977,8	8,2859
24	3-metil-2-nonanona	1122,0	9,2996	52	Nonanal	1085,1	9,1078
25	3-decanona	1164,8	9,4363	53	Decanal	1192,0	9,9216
26	2-decanona	1168,9	9,4583	54	Undecanal	1298,9	10,7395
27	6-undecanona	1249,2	10,2266	55	Dodecanal	1405,6	11,5562
28	2-undecanona	1269,2	10,2706				

As correlações obtidas para 40 cetonas e 15 aldeídos podem ser observadas nas equações (43) e (44), respectivamente:

$$IR_{Exp} = 123,8746I_{SET} - 11,8849 \quad (43)$$

$$r = 0,9997; s = 11,78; N = 40; f = 60364,3; r^2 = 0,9993; r_{CV}^2 = 0,9993$$

e

$$IR_{Exp} = 128,7244I_{SET} - 86,9288 \quad (44)$$

$$r = 0,9994; s = 10,31; N = 15; f = 1264,6; r^2 = 0,9989; r_{CV}^2 = 0,9986$$

Os resultados apresentados indicam que para moléculas polares é necessário a introdução do momento de dipolo e o momento de dipolo equivalente local associado com

as cargas líquidas dos átomos do grupo carbonila de forma a obter uma boa correlação entre a retenção cromatográfica e o descritor desenvolvido.

Comparando-se com os resultados obtidos utilizando o descritor I_{ET} para 31 cetonas e aldeídos ($r=0,9995$ e $s=5,47$)²⁴, observa-se que o descritor proposto é capaz decodificar tão bem a molécula quanto o seu predecessor.

As correlações entre os valores de IR_{Exp} e I_{SET} para todos os compostos são mostradas na Figura 6.

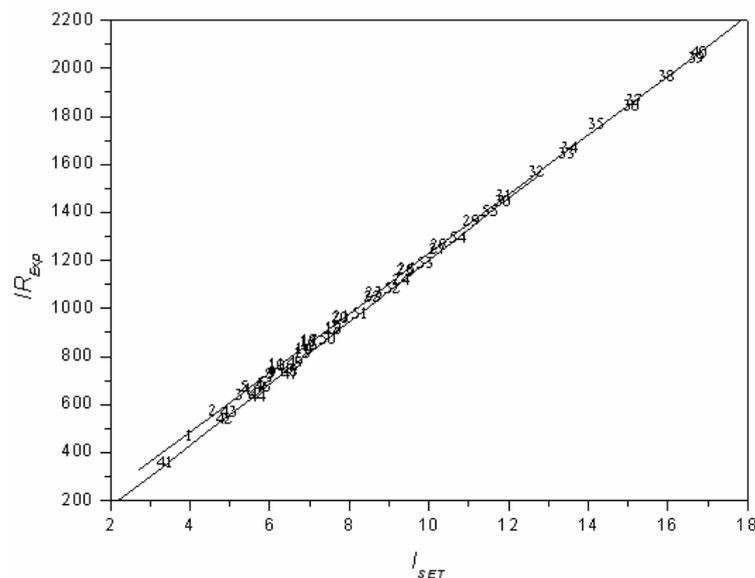


Figura 6: Índices de retenção experimentais (IR_{Exp}) versus I_{SET} para 55 aldeídos e cetonas.

As correlações entre os índices de retenção experimentais (IR_{Exp}) e calculados (IR_{Calc}) para os compostos estudados são mostrados na Figura 7.

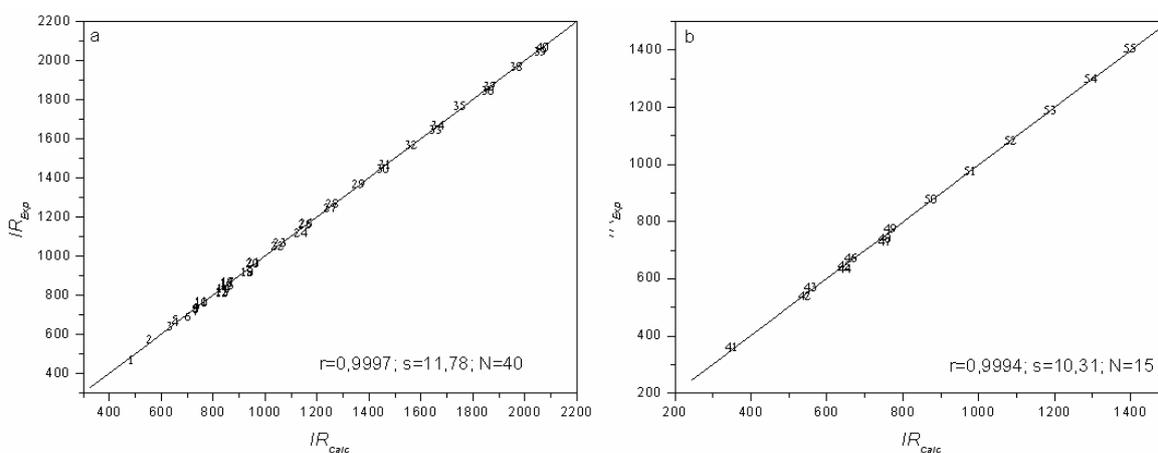


Figura 7: Índices de retenção experimentais (IR_{Exp}) versus índices de retenção calculados (IR_{Calc}) para 40 cetonas^a e 15 aldeídos^b.

Os valores residuais ($IR_{Exp} - IR_{Calc}$) versus os valores calculados para aldeídos e cetonas são mostrados na Figura 8.

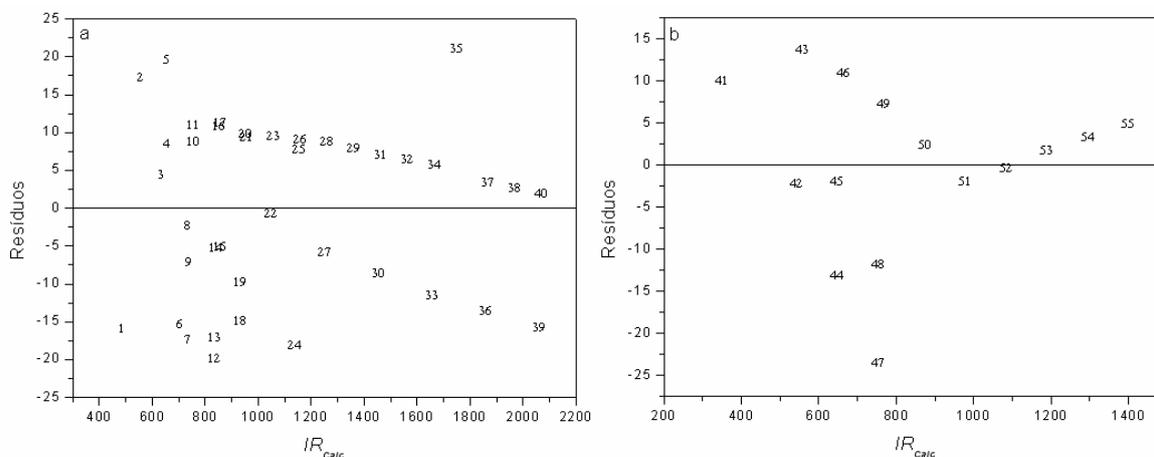


Figura 8: Valores residuais versus índices retenção calculados (IR_{Calc}) para 40 cetonas^a e 15 aldeídos^b

Observa-se na Figura 8b que a maior parte dos compostos apresentam resíduos aceitáveis. Porém uns poucos compostos apresentam resíduos com valores acima do desvio padrão, o que pode ser atribuídos tanto a incertezas experimentais quanto a discrepâncias nos valores obtidos para as cargas atômicas líquidas dos átomos dessas moléculas.

Os resultados da validação estatística estão apresentados na Tabela 8 e foram efetuados da mesma forma que para os alcanos e alcenos.

Tabela 8: Verificação da validação estatística do modelo (N=55)¹⁷⁰

Grupo de Treinamento	r^2	r_{CV}^2	Grupo de predição	r^2	r_{CV}^2
1 e 2	0,9988	0,9985	3	0,9972	0,9969
1 e 3	0,9979	0,9977	2	0,9977	0,9972
2 e 3	0,9984	0,9979	1	0,9971	0,9964
Média	0,9984	0,9980	Média	0,9973	0,9968

Os resultados obtidos usando o descritor I_{SET} são similares aqueles publicados por Ren¹⁵² em modelos de regressão linear múltipla para 33 aldeídos e cetonas usando os índices topológicos Xu e AI , e também similares aos resultados obtidos por Héberger e colaboradores usando os descritores químico-quânticos (S_W e μ)⁷² e propriedades físico-químicas (T_{Bp} , M_W , $\log P$)¹⁵³ para 31 e 35 compostos, respectivamente.

S_W : área superficial acessível ao solvente, T_{Bp} : temperatura de ebulição, M_W : massa molecular, $\log P$: coeficiente de partição octanol/água.

5.3. Aplicação do I_{SET} para estimar a retenção cromatográfica de ésteres

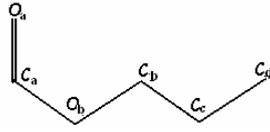
Os cálculos de I_{SET} para alcanos, alcenos, aldeídos e cetonas foram realizados via planilhas EXCEL usando quatro decimais. Devido à presença de dois heteroátomos nos ésteres, o número de variáveis (cargas nos oxigênios, carbonos ligados a eles, assim como os respectivos comprimentos e ângulos de ligação) aumenta, inviabilizando o uso racional das planilhas de EXCEL. Assim, um código FORTRAN foi desenvolvido utilizando dupla precisão para efetuar estes cálculos. Por uma questão de praticidade, os cálculos dos exemplos a seguir foram feitos com um número menor e adequado de casas decimais.

As moléculas de formiato de propila e isobutirato de isopropila representadas no grafo a seguir foram tomadas como exemplo de cálculo do I_{SET} e os respectivos valores de cargas atômicas líquidas e SET_i são mostrados na Tabela 9.

Tabela 9: Valores de cargas atômicas líquidas e SET_i para cada átomo de carbono e oxigênio das moléculas de formiato de propila e isobutirato de isopropila¹⁷¹

Átomos	formiato de propila			isobutirato de isopropila		
	δ_i	SET_i	$A_{,,}SET_i$	δ_i	SET_i	$A_{,,}SET_i$
O _a	-0,351	1,2463	1,8412	-0,359	1,2605	2,1994
O _b	-0,288	1,1346	1,6761	-0,278	1,1168	1,9487
C _a	+0,257	0,1683	0,2486	+0,305	0,0832	0,1452
C _b	-0,015	0,6505	0,9610	+0,036	0,5601	0,9773
C _c	-0,165	0,9165	-	-0,096	0,7941	-
C _d	-0,212	0,9998	-	-0,210	0,9963	-
C _e	-	-	-	-0,210	0,9963	-
C _f	-	-	-	-0,237	1,0441	-
C _g	-	-	-	-0,215	1,0051	-

a) formiato de propila:



$$\mu_{F1} = d_{C_a-O_a} |\delta_{C_a} - \delta_{O_a} + \delta_H| \quad (32)$$

$$\mu_{F1} = 1,229079 |0,257249 - (-0,351472) + 0,172434| = 0,960101$$

$$\mu_{F2} = d_{C_a-O_b} |\delta_{C_a} - \delta_{O_b} + \delta_H| \quad (32)$$

$$\mu_{F2} = 1,359716 |0,257249 - (-0,288289) + 0,172434| = 0,976238$$

$$\mu_{F3} = d_{C_b-O_b} |\delta_{C_b} - \delta_{O_b}| \quad (31)$$

$$\mu_{F3} = 1,438218 |-0,014937 - (-0,288289)| = 0,393139$$

$$\mu_{F4} = \mu_{F2} - \mu_{F3} \quad (37)$$

$$\mu_{F4} = 0,976238 - 0,393139 = 0,583099$$

$$\mu_F = (\mu_{F1}^2 + \mu_{F4}^2 + 2\mu_{F1}\mu_{F4} \cos \theta)^{1/2} \quad (36)$$

$$\mu_F = [(0,960101)^2 + (0,583099)^2 + (2 * 0,960101 * 0,583099 * \cos 119,5)]^{1/2} = 0,842880$$

$$A_\mu = 1 + \log\left(1 + \frac{\mu}{\mu_F}\right) = 1 + \log\left(1 + \frac{1,6870}{0,8428}\right) = 1,4773 \quad (30)$$

$$I_{SET} = \sum I_{SET_i} = \sum (A_\mu SET_i + \log A_\mu SET_j) \quad (33)$$

$$I_{SET_a} = A_\mu SET_{O_a} + \log A_\mu SET_{C_a} = 1,8412 + \log 0,2486 = 1,2368$$

$$I_{SET_{O_b}} = A_\mu SET_{O_b} + \log A_\mu SET_{C_a} + \log A_\mu SET_{C_b} = 1,6761 + \log 0,2486 + \log 0,9610 = 1,0545$$

$$I_{SET_{C_a}} = A_\mu SET_{C_a} + \log A_\mu SET_{O_a} + \log A_\mu SET_{O_b} = 0,2486 + \log 1,8412 + \log 1,6761 = 0,7381$$

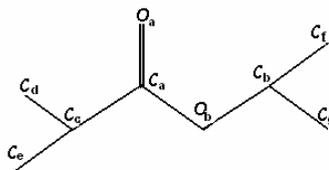
$$I_{SET_{C_b}} = A_\mu SET_{C_b} + \log A_\mu SET_{O_b} + \log SET_{C_c} = 0,9610 + \log 1,6761 + \log 0,9165 = 1,1477$$

$$I_{SET_{C_c}} = SET_{C_c} + \log A_\mu SET_{C_b} + \log SET_{C_d} = 0,9165 + \log 0,9610 + \log 0,9998 = 0,8994$$

$$I_{SET_{C_d}} = SET_{C_d} + \log SET_{C_c} = 0,9998 + \log 0,9165 = 0,9622$$

$$I_{SET} = 1,2368 + 1,0545 + 0,7381 + 1,1477 + 0,8994 + 0,9622 = 6,0387$$

b) isobutirato de isopropila :



$$\mu_{F1} = d_{C_a-O_a} |\delta_{O_a}| = 1,232948 |-0,358939| = 0,442553 \quad (34)$$

$$\mu_{F2} = d_{C_a-O_b} |\delta_{O_b}| = 1,364974 |-0,278490| = 0,380131 \quad (35)$$

$$\mu_F = (\mu_{F1}^2 + \mu_{F2}^2 + 2\mu_{F1}\mu_{F2} \cos \theta)^{1/2} \quad (36)$$

$$= [(0,442553)^2 + (0,380131)^2 + (2 * 0,442553 * 0,380131 * \cos 119,023)]^{1/2} = 0,420853$$

$$A_\mu = 1 + \log\left(1 + \frac{\mu}{\mu_F}\right) = 1 + \log\left(1 + \frac{1,9180}{0,4208}\right) = 1,7449 \quad (30)$$

Sendo esta molécula um éster ramificado com $n = 2$, então devemos usar a correção estérea (ce) conforme a equação abaixo:

$$I_{SET} = \left(\sum I_{SET_i}\right) + sc = \left(\sum A_\mu SET_i + \log A_\mu SET_j\right) + n \log SET_{C_{Ac}} + n \log SET_{C_{Al}} \quad (45)$$

$$I_{SET_{O_a}} = A_\mu SET_{O_a} + \log A_\mu SET_{C_a} = 2,1994 + \log 0,1452 = 1,3626$$

$$I_{SET_{O_b}} = A_\mu SET_{O_b} + \log A_\mu SET_{C_a} + \log A_\mu SET_{C_b} = 1,9487 + \log 0,1452 + \log 0,9773 = 1,1021$$

$$I_{SET_{C_a}} = A_\mu SET_{C_a} + \log A_\mu SET_{O_a} + \log A_\mu SET_{O_b} + \log SET_{C_c}$$

$$= 0,1452 + \log 2,1994 + \log 1,9487 + \log 0,7941 = 0,6788$$

$$I_{SET_{C_b}} = A_\mu SET_{C_b} + \log A_\mu SET_{O_b} + \log SET_{C_f} + \log SET_{C_g}$$

$$= 0,9773 + \log 1,9487 + \log 1,0441 + \log 1,0051 = 1,2896$$

$$I_{SET_{C_c}} = SET_{C_c} + \log A_\mu SET_{C_a} + \log SET_{C_d} + \log SET_{C_e}$$

$$= 0,7941 + \log 0,1452 + \log 0,9963 + \log 0,9963 = -0,0462$$

$$I_{SET_{C_d}} = SET_{C_d} + \log SET_{C_c} = 0,9963 + \log 0,7941 = 0,8975$$

$$I_{SET_{C_e}} = SET_{C_e} + \log SET_{C_c} = 0,9963 + \log 0,7941 = 0,8975$$

$$I_{SET_{C_f}} = SET_{C_f} + \log A_\mu SET_{C_b} = 1,0441 + \log 0,9773 = 1,0353$$

$$I_{SET_{C_g}} = SET_{C_g} + \log A_\mu SET_{C_b} = 1,0051 + \log 0,9773 = 0,9963$$

$$sc = 2 \log SET_{C_b} + 2 \log SET_{C_c} = 2 \log 0,5601 + 2 \log 0,7941 = -0,7020$$

$$I_{SET} = 1,3626 + 1,1021 + 0,6788 + 1,2896 - 0,0462 + 0,8975 + 0,8975 + 1,0353 + 0,9963 - 0,7020$$

$$= 7,5115$$

A Tabela 10 apresenta os valores calculados de I_{SET} e os índices de retenção experimentais para 100 ésteres lineares e ramificados.

Tabela 10: Valores dos índices de retenção de cinco fases estacionárias (SE-30, OV-7, DC-710, XE-60 e OV-225) e dos índices semi-empíricos eletrotológicos (I_{SET})¹⁷¹

Nº	Composto	SE-30	OV-7	DC-710	XE-60	OV-225	I_{SET}
1	Formiato de propila	608	662	703	821	834	6,0387
2	Formiato de butila	709	766	805	923	934	6,8701
3	Formiato de pentila	807	866	903	1021	1031	7,6858
4	Formiato de hexila	906	968	1002	1120	1128	8,5112
5	Formiato de octila	1104	1171	1201	1318	1324	10,1444
6	Acetato de metila	509	559	609	737	759	5,2056
7	Acetato de propila	695	755	796	920	932	6,8215
8	Acetato de butila	794	855	895	1022	1030	7,6480
9	Acetato de pentila	891	958	996	1122	1132	8,4025
10	Acetato de heptila	1084	1151	1181	1299	1305	10,0882
11	Acetato de octila	1182	1252	1280	1397	1401	10,9107
12	Acetato de isopropila	643	686	722	842	840	6,3210
13	Acetato de isopentila	859	922	955	1086	1087	8,1012
14	Propionato de metila	617	667	710	837	851	5,9612
15	Propionato de etila	692	743	783	897	905	6,7012
16	Propionato de propila	789	851	884	1003	1013	7,5108
17	Propionato de pentila	980	1047	1082	1202	1211	9,1446
18	Propionato de hexila	1074	1142	1176	1300	1307	9,9675
19	Propionato de heptila	1173	1242	1270	1388	1392	10,7753
20	Propionato de isopropila	733	785	817	920	928	7,0056
21	Propionato de isobutila	848	906	938	1057	1062	8,0329
22	Propionato de iopentila	948	1014	1042	1156	1169	8,7925
23	Butirato de metila	702	767	803	923	938	6,7703
24	Butirato de etila	778	840	876	986	999	7,5262
25	Butirato de butila	969	1036	1069	1187	1193	9,1349
26	Butirato de pentila	1062	1131	1164	1284	1288	9,9412
27	Butirato de hexila	1156	1225	1258	1380	1388	10,7642
28	Butirato de isopropila	820	878	905	1010	1019	7,8011
29	Butirato de isobutila	933	993	1023	1143	1146	8,7154
30	Butirato de isopentila	1029	1093	1124	1243	1247	9,5886
31	Pentanoato de etila	876	941	975	1085	1105	8,3094
32	Pentanoato de propila	971	1036	1069	1179	1196	9,0964
33	Pentanoato de butila	1063	1130	1163	1279	1289	9,9220
34	Pentanoato de pentila	1155	1222	1257	1375	1382	10,7282
35	Pentanoato de hexila	1247	1316	1349	1470	1476	11,5513
36	Pentanoato de isopropila	915	971	1000	1105	1116	8,5874
37	Pentanoato de isobutila	1027	1087	1117	1235	1240	9,6203
38	Pentanoato de isopentila	1132	1185	1217	1333	1340	10,3762
39	Hexanoato de metila	902	974	1006	1133	1140	8,3867
40	Hexanoato de etila	976	1045	1073	1196	1203	9,1100
41	Hexanoato de butila	1156	1231	1256	1377	1383	10,7431
42	Hexanoato de pentila	1246	1325	1349	1471	1508	11,5493
43	Hexanoato de hexila	1337	1417	1440	1565	1575	12,3724
44	Hexanoato de isopropila	1008	1074	1115	1207	1206	9,4076
45	Hexanoato de isobutila	1119	1189	1211	1332	1333	10,4412
46	Hexanoato de isopentila	1212	1287	1308	1430	1432	11,1961
47	Heptanoato de metila	998	1062	1094	1212	1219	9,1928
48	Heptanoato de etila	1068	1135	1165	1283	1289	9,9322
49	Heptanoato de propila	1155	1224	1253	1370	1375	10,7395

continuação

Nº	Composto	SE-30	OV-7	DC-710	XE-60	OV-225	I_{SET}
50	Heptanoato de isopropila	1105	1172	1202	1319	1325	10,2120
51	Heptanoato de isobutila	1213	1284	1311	1428	1432	11,2465
52	Octanoato de metila	1094	1162	1192	1309	1315	10,0329
53	Octanoato de isopropila	1200	1270	1298	1415	1419	11,0326
54	Octanoato de isobutila	1306	1379	1404	1521	1524	12,0669
55	Decanoato de etila	1354	1428	1452	1569	1571	12,3802
56	Isobutirato de etila	732	793	831	968	935	7,0167
57	Isobutirato de butila	931	987	1019	1130	1136	8,6499
58	Isobutirato de pentila	1024	1082	1115	1225	1230	9,4558
59	Isobutirato de hexila	1117	1177	1208	1323	1323	10,2790
60	Isobutirato de isopropila	780	829	849	953	947	7,5115
61	Isobutirato de isobutila	899	946	975	1096	1086	8,5664
62	Isobutirato de isopentila	994	1045	1075	1196	1188	9,2980
63	Isobutirato de metilpropila	892	954	989	1106	1115	8,3204
64	Isopentanoato de metila	763	823	854	969	982	7,2346
65	Isopentanoato de butila	1021	1081	1111	1224	1231	9,5371
66	Isopentanoato de pentila	1112	1174	1204	1318	1325	10,3430
67	Isopentanoato de hexila	1204	1267	1296	1414	1417	11,1662
68	Isopentanoato de isopropila	874	923	950	1049	1064	8,1431
69	Isopentanoato de isobutila	985	1039	1066	1177	1184	9,2330
70	Isopentanoato de isopentila	1081	1136	1164	1278	1281	9,9907
71	Isohexanoato de metila	875	936	970	1094	1104	8,1146
72	Isohexanoato de etila	943	1004	1036	1158	1163	8,7801
73	Isohexanoato de propila	1035	1094	1128	1249	1253	9,5870
74	Isohexanoato de butila	1125	1189	1220	1340	1345	10,4121
75	Isohexanoato de pentila	1215	1281	1313	1433	1438	11,2181
76	Isohexanoato de isopropila	979	1028	1059	1169	1172	9,0156
77	Isohexanoato de isobutila	1089	1147	1175	1296	1309	10,1081
78	Isohexanoato de isopentila	1181	1244	1272	1393	1394	10,8636
79	2-metilbutirato de metila	762	820	858	976	987	7,2145
80	2-metilbutirato de etila	832	892	928	1045	1055	7,8835
81	2-metilbutirato de propila	924	987	1021	1138	1146	8,6905
82	2-metilbutirato de butila	1016	1082	1113	1231	1238	9,5156
83	2-metilbutirato de pentila	1110	1178	1207	1325	1330	10,3217
84	2-metilbutirato de hexila	1204	1275	1302	1419	1423	11,1447
85	2-metilbutirato de isobutila	980	1044	1077	1194	1202	9,1567
86	2-metilbutirato de isopentila	1072	1139	1169	1287	1293	9,9145
87	2-metilpentanoato de metila	853	908	945	1054	1065	8,0240
88	2-metilpentanoato de etila	917	973	1006	1118	1121	8,6899
89	2-metilpentanoato de propila	1009	1064	1096	1204	1213	9,4969
90	2-metilpentanoato de butila	1097	1157	1189	1296	1303	10,3210
91	2-metilpentanoato de isopropila	952	999	1021	1121	1128	8,8739
92	2-metilpentanoato de isobutila	1064	1116	1144	1254	1257	9,9628
93	2-metilpentanoato de isopentila	1153	1208	1238	1352	1350	10,7184
94	2-etilbutirato de metila	845	899	934	1049	1053	7,8968
95	2-etilbutirato de etila	914	969	1000	1122	1113	8,5635
96	2-etilbutirato de propila	1005	1060	1092	1204	1206	9,3699
97	2-etilbutirato de pentila	1183	1241	1273	1390	1387	11,0014
98	2-etilbutirato de hexila	1273	1332	1364	1482	1478	11,8248
99	2-etilbutirato de isobutila	1060	1112	1139	1253	1249	9,8351
100	2-etilbutirato de isopentila	1149	1202	1232	1352	1343	10,5429

Para verificar a aplicabilidade do I_{SET} na predição do índice de retenção em diferentes fases estacionárias e a influência da polaridade da fase estacionária foram observadas as equações de regressões lineares simples entre os índices de retenção (IR_{Exp}) e os índices semi-empíricos eletrotopológicos (I_{SET}) usando 100 ésteres, os quais são apresentados para cada fase estacionária separadamente, como mostrado na Figura 9 e Tabela 11.

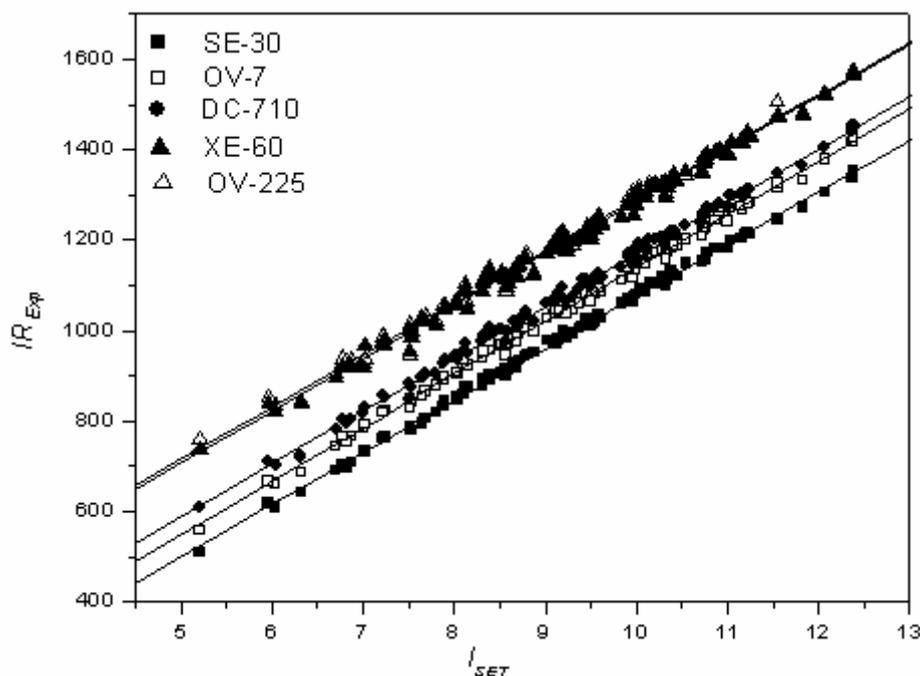


Figura 9: Regressão linear simples entre os índices de retenção (IR_{Exp}) e os índices semi-empírico eletrotopológicos (I_{SET}) para ésteres nas fases estacionárias SE-30, OV-7, DC-710, XE-60 e OV-225.

Tabela 11: Resultados estatísticos para os modelos de QSRR e as constantes de McReynolds de cinco fases estacionárias

Equação N°	Fase estacionária	Mp	b_0	b_1	r	r^2	r^2_{CV}	s	f
46	SE-30	217	-74,7563	115,0361	0,9990	0,9981	0,9980	7,65	51777
47	OV-7	592	-39,5558	117,8831	0,9986	0,9973	0,9972	9,42	35905
48	DC-710	827	11,1427	115,8248	0,9984	0,9968	0,9967	10,01	30690
49	XE-60	1785	128,1568	115,8136	0,9971	0,9942	0,9940	13,51	16852
50	OV-225	1813	141,4299	115,0095	0,9959	0,9918	0,9914	16,02	11814

Mp : constantes de McReynolds¹⁰; b_0 : intercepto; b_1 : inclinação; b_1 médio= **115,9134**; r : coeficiente de correlação; r^2 : coeficiente de determinação; r^2_{CV} : coeficiente de validação cruzada; s : desvio padrão; f : teste de hipótese nula (F-test). Composição das fases estacionárias: SE-30 (100% de metilsiloxano) e OV-7 (20% de fenila e 80% de metilsiloxano).

Os resultados estatísticos para os modelos finais de QSRR mostram que, para todas as fases estacionárias, os coeficientes de correlação são melhores que 0,9959, especialmente para as fases estacionárias SE-30, OV-7 e DC-710 (ficam entre 0,9990 e 0,9984). Os coeficientes de validação cruzada demonstram que os modelos finais são estatisticamente significantes, com valores entre 0,9980 e 0,9914, os quais são muito próximos dos correspondentes valores de r^2 .

Os resultados obtidos por regressão linear simples usando o I_{SET} são melhores que os publicados por Lu *et al.* (2006)¹⁵⁶ e por Liu *et al.* (2007)⁶⁴. Sendo que estes autores usaram modelos de regressão linear múltipla entre IR e índices topológicos para 90 ésteres saturados. Liu *et al.* (2007)⁶⁴ aplicando novos índices topológicos (índice de efeito de polarizabilidade (*PEI*), índice par-ímpar (*OEI*) e índice de efeito estéreo (*SVij*), previamente desenvolvidos por eles⁶⁴, encontraram, para cada fase estacionária, correlações com coeficientes entre 0,9989 e 0,9977 e coeficientes de validação cruzada entre 0,9988 e 0,9975. Por meio da análise estatística eles mostraram que o efeito da polarizabilidade, o tamanho molecular e as ramificações apresentam contribuições dominantes para o IR dos ésteres estudados, mas o efeito estéreo também fornece alguma contribuição. Estes modelos de QSRR foram comparados com outros resultados similares, por Lu *et al.* (2006)¹⁵⁶ onde índices topológicos tais como o índice de Lu e novos índices baseados em distância-tipo de átomo (*distance-based atom-type DAI topological indices*) foram aplicados em métodos de MLR. Os modelos de Liu *et al.* (2007)⁶⁴ produzem melhores correlações com índices de retenção de Kováts do que os índices topológicos previamente introduzidos por Lu *et al.* (2006)¹⁵⁶. Além disso, os parâmetros estatísticos obtidos com a presente abordagem apresentam também uma boa concordância com aqueles obtidos para ésteres saturados usando o I_{ET} , (N=83; $r=0,9992$; $s=8,32$) com SLR¹⁶.

A partir da Tabela 11 e Figura 9 pode-se observar que o coeficiente linear (b_0 – intercepto) aumenta linearmente com a polaridade da fase estacionária, enquanto que o coeficiente angular (b_1 - inclinação) não apresenta variações significativas. O valor de b_0 varia de acordo com a intensidade da interação entre as moléculas do soluto e da fase estacionária, mostrando o grau de influência das propriedades das fases estacionárias no índice de retenção. As constantes de McReynolds são usadas como indicadores para avaliar a polaridade da coluna cromatográfica¹⁷²⁻¹⁷⁵. A correlação entre o coeficiente b_0 e as constantes de McReynolds (*Mp*) sugere que as moléculas do soluto apresentam

diferentes comportamentos de retenção. Isto ocorre face à predominância de diferentes mecanismos de interação destas moléculas com as diferentes fases estacionárias devido ao aumento da polaridade. Este fato pode ser constatado pela Equação (50) e Figura 10:

$$b_o = -108,1958 + 0,1351Mp \quad (50)$$

$$r = 0,9967; s = 9,10; n = 5$$

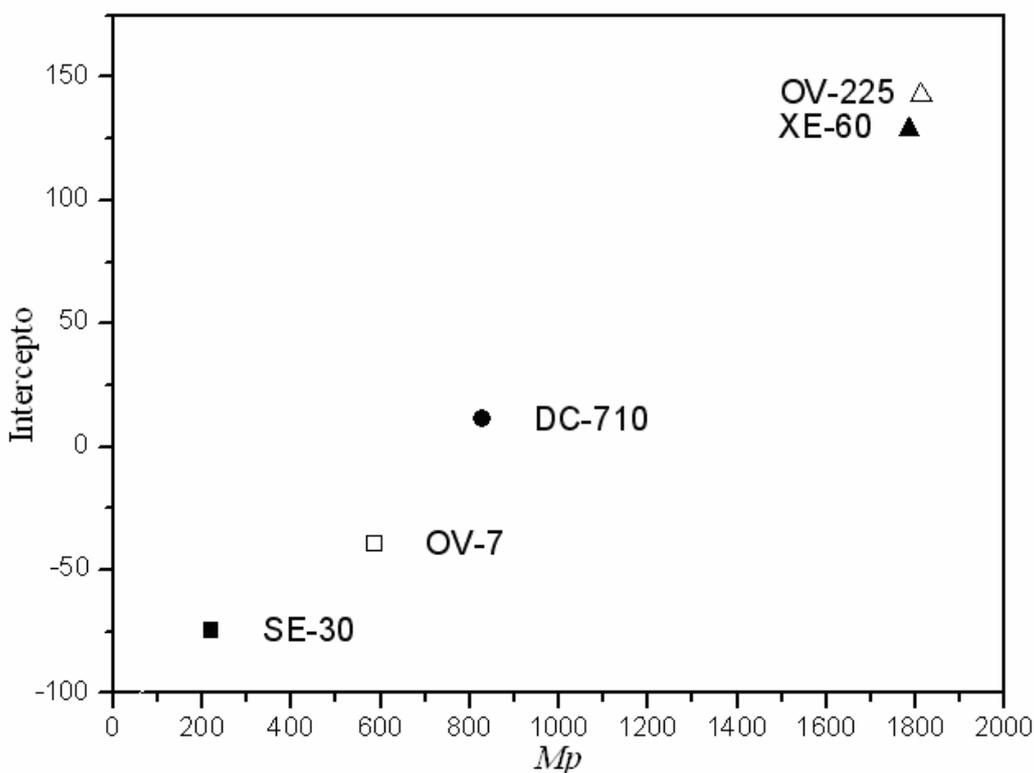


Figura 10: Correlação entre o intercepto (b_o) e a polaridade de McReynolds das fases estacionárias (Mp) dos modelos QSRR obtidos usando I_{SET} .

A boa correlação linear entre b_o e Mp obtida, como pode ser visto a partir dos parâmetros estatísticos, indica a existência de interações intermoleculares similares que contribuem para a retenção cromatográfica destes compostos.

A partir da Equação (50) e do valor médio dos coeficientes angulares (b_1 médio no rodapé da Tabela 11) é possível deduzir a seguinte equação combinada¹⁷¹ ($RI_{Calc} = b_o + b_{1_{ave}} I_{SET}$):

$$RI_{Calc} = (-108,1958 + 0,1351Mp) + 115,9134I_{SET} . \quad (51)$$

Sendo assim, um parâmetro que representa uma propriedade da fase estacionária, a polaridade de McReynolds (Mp), foi adicionada ao modelo. Desta maneira foi construída

uma nova equação que inclui a polaridade das diferentes fases estacionárias. Uma indicação gráfica da qualidade deste modelo combinado pode ser vista na Figura 11, onde se mostra a correlação entre os índices de retenção experimentais (IR_{Exp}) e os calculados (IR_{Calc}), e também o gráfico dos valores residuais versus os índices de retenção calculados (IR_{Calc}) para os compostos estudados.

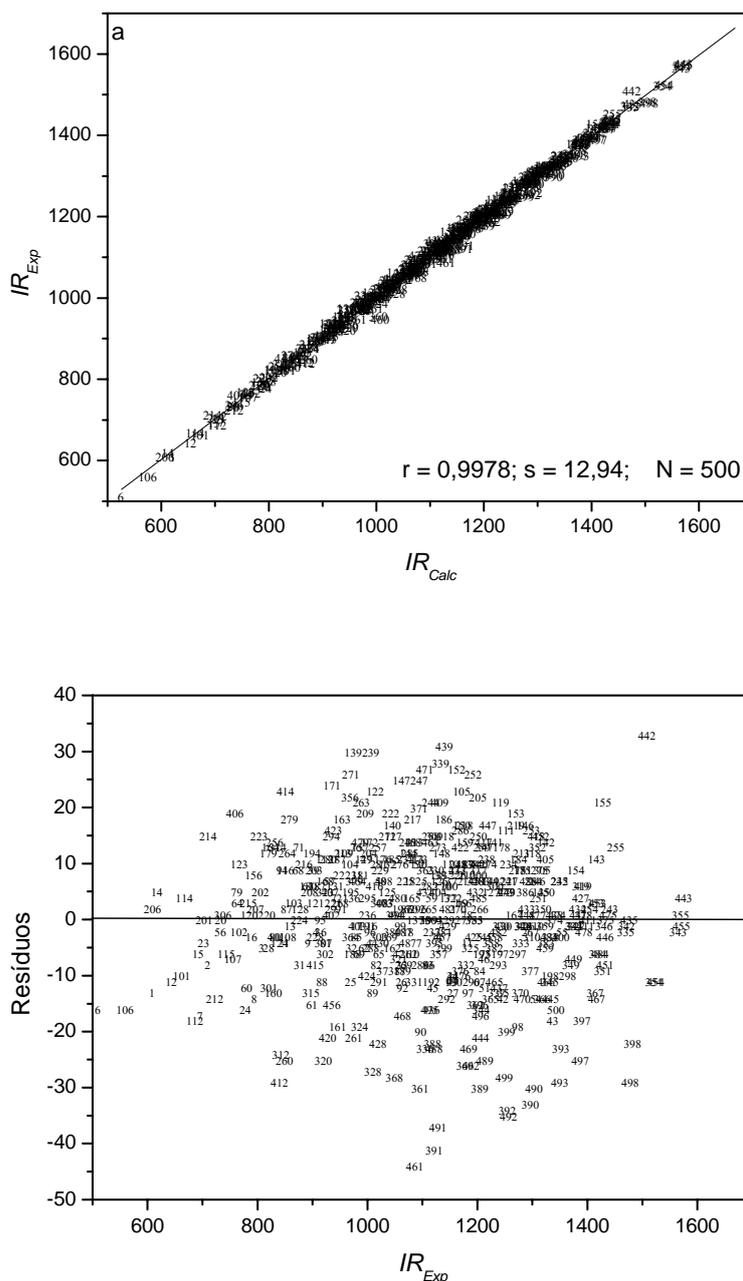


Figura 11: (a) Índices de retenção experimentais (IR_{Exp}) versus calculados (IR_{Calc}) para 500 dados. (b) Gráfico dos valores de resíduos versus os índices de retenção calculados (IR_{Calc}) dos ésteres em cinco fases estacionárias com diferentes polaridades

Para verificar a validação externa dos modelos foram tomados 10 compostos externos, não incluídos nos modelos originais de QSRR, e testados nas fases estacionárias SE-30 e OV-7. As equações (46) e (47) da Tabela 11 e a Equação combinada (51) foram utilizadas para prever a retenção cromatográfica destes compostos. Os valores dos *IR* experimentais e preditos, assim como seus resíduos (ΔIR) são mostrados na Tabela 12.

Tabela 12: Índices de retenção experimentais e preditos nas fases estacionárias SE-30 e OV-7 para o grupo externo¹⁷¹

No.	Composto	SE-30 (Exp.)	SE-30 (eq.46)	ΔIR	SE-30 (eq.51)	ΔIR	OV-7 (Exp.)	OV-7 (eq.47)	ΔIR	OV-7 (eq.51)	ΔIR
1	Formiato de heptila	1005	998	7,3	1002	3,3	1070	1059	10,6	1052	17,6
2	Acetato de hexila	988	993	-4,8	997	-8,8	1056	1054	1,6	1047	8,5
3	Propionato de butila	886	889	-2,9	892	-6,1	949	948	1,1	943	6,2
4	Butirato de propila	875	881	-6,0	884	-9,2	938	940	-1,9	935	3,2
5	Pentanoato de metila	807	798	9,4	800	6,8	871	854	16,6	851	20,2
6	Hexanoato de propila	1064	1066	-2,1	1071	-6,7	1138	1130	8,4	1121	16,6
7	Octanoato de etila	1165	1162	2,8	1167	-2,5	1234	1228	6,0	1218	15,9
8	Isopentanoato de etila	839	834	4,5	837	1,7	893	892	0,8	888	5,0
9	Isohexanoato de hexila	1306	1310	-4,3	1317	-10,7	1374	1380	-5,8	1367	6,6
10	2-etilbutirato de butila	1093	1098	-5,1	1103	-9,9	1150	1162	-12,3	1154	-3,6

Estes resultados indicam que os modelos de QSRR apresentam não somente alta estabilidade interna, como podem ser observados pelos valores de r^2 e r_{CV}^2 , mas também boa habilidade preditiva para o grupo externo. Assim, o novo descritor introduzido (I_{SET}) é capaz de descrever adequadamente essas moléculas, originando um bom modelo para a predição, de forma satisfatória, do índice de retenção cromatográfico destes compostos além de moléculas externas.

6. Conclusão

Este trabalho foi desenvolvido com propósito de prognosticar a retenção cromatográfica por meio de estudos de correlação quantitativa entre estrutura molecular e retenção cromatográfica via uma expressão algébrica, onde um deles indica a propriedade e o outro, o descritor molecular relacionado a estrutura da molécula.

Uma etapa decisiva na construção do modelo está na seleção de um descritor molecular apropriado, a ser empregado para obter uma correlação significativa através da codificação da estrutura química, com a finalidade de reter as informações estruturais necessárias e adequadas sob o ponto de vista estatístico. A retenção cromatográfica é um processo físico-químico bastante complexo, o qual compreende interações intermoleculares entre as moléculas do soluto e da fase estacionária, conhecidas como forças de van der Waals. E, também, é dependente do ambiente topológico, da geometria e/ou da estrutura eletrônica destas moléculas. Assim, é natural esperar que seja possível a definição de um descritor que dependa explicitamente das propriedades moleculares como cargas atômicas, momento de dipolo, polarizabilidade, etc, as quais são responsáveis pelas interações intermoleculares.

A fim de se atingir tais objetivos, um novo descritor molecular denominado I_{SET} foi desenvolvido por meio da correlação entre os valores de SET_i para átomos de carbono primário, secundário, terciário e quartenário, obtidos através da observação do comportamento cromatográfico e deduções teóricas, com a carga atômica líquida calculada pelo método semi-empírico AM1. Usando esta correlação os valores de SET_i foram calculados para todos os tipos de átomos de carbono de alcanos e alcenos. Com isso, as limitações existentes com o descritor precedente (I_{ET}) foram contornadas, tornando o novo índice independente do uso de tabelas específicas, além de atribuir valores para os fragmentos derivados de uma propriedade independente da retenção cromatográfica.

A aplicação do I_{SET} em modelos de QSRR para 179 alcanos e alcenos lineares e ramificados demonstraram boa qualidade com parâmetros estatísticos ($r = 0,9990$; $s = 10,74$). A validade externa do modelo foi testada e os resultados obtidos para os grupos de treinamento ($r^2 = 0,9982$) e predição ($r^2 = 0,9978$), indicam que o modelo proposto tem alta estabilidade estatística e validade. Além disso, o novo índice contém informações sobre a estrutura tridimensional das moléculas, sendo capaz de discriminar isômeros geométricos como *-cis* e *-trans* de alcenos.

Em seguida, o I_{SET} foi aplicado para moléculas polares como cetonas, aldeídos e ésteres, mas foi necessário incluir no índice o momento de dipolo molecular e um momento de dipolo equivalente local, definido a partir das cargas dos heteroátomos e átomos adjacentes.

Para cetonas, o momento de dipolo equivalente local foi representado pelas cargas atômicas líquidas dos átomos de oxigênio e do carbono da carbonila. O modelo obtido é de boa qualidade como mostrado pelos parâmetros estatísticos ($r = 0,9997$; $s = 11,72$; $N = 42$). Ao passo que para o modelo de QSRR obtido para os aldeídos, a melhor definição do momento de dipolo equivalente local incluiu também a carga do hidrogênio conectado ao carbono da ligação C=O. Sendo também de boa qualidade o modelo conforme os parâmetros estatísticos gerados ($r = 0,9994$; $s = 10,31$; $N = 15$). Para verificar a validade estatística do modelo, todo o conjunto de dados foi sistematicamente dividido em três subgrupos e cada subconjunto foi predito usando os outros dois subconjuntos como grupo de treinamento. Estes modelos também mostraram uma boa estabilidade externa com $r^2 = 0,9984$ para os subgrupos de treinamento e com $r^2 = 0,9973$ para os subgrupos de predição.

No entanto, para o desenvolvimento do modelo de QSRR dos ésteres, devido ao excesso de carga pela presença dos dois átomos de oxigênio, também foi necessário incluir o momento de dipolo molecular e as cargas líquidas nos átomos de oxigênio por meio de um momento de dipolo equivalente local. Para tal, duas definições de momento de dipolo equivalente locais foram utilizadas. A correlação obtida para 100 ésteres alifáticos em diferentes fases estacionárias (SE-30, OV-7, DC-710, OV-225 e XE-60) representada pelos parâmetros estatísticos ($r=0,9990$ a $0,9959$; e $s=7,65$ a $16,02$) é de boa qualidade.

A correlação linear entre os coeficientes lineares (b_0) e as constantes de McReynolds (Mp) obtida, indica que existe similaridade entre as fases SE-30, OV-7, DC-710, XE-60 e OV-225, como também aponta a existência de interações intermoleculares similares que contribuem para a retenção cromatográfica destes compostos. Assim, foi possível deduzir uma equação combinada incluindo a polaridade de McReynolds, um parâmetro que representa uma propriedade da fase estacionária. Este modelo combinado apresentou boa qualidade conforme observado pelos parâmetros de mérito da regressão entre os valores de retenção experimentais versus os calculados ($r = 0,9978$; $s = 12,94$; $N = 500$).

Em geral, os resultados alcançados com este novo descritor foram satisfatórios mostrando que foi possível decodificar a estrutura molecular e manter as informações

relevantes para realizar a previsão de retenção cromatográfica. Com exceção de algumas moléculas, cujos resíduos foram um pouco elevados, o que tanto pode ser atribuído a desvios experimentais quanto a incertezas no cálculo semi-empírico das cargas atômicas líquidas.

Apesar de não ser um dos objetivos, a automatização dos cálculos no caso dos ésteres via o desenvolvimento de um código FORTRAN, tornou possível cálculos mais precisos, rápidos e eficientes. Além de possibilitar os cálculos usando várias definições do momento de dipolo equivalente local.

Esta nova abordagem que incluiu também moléculas polares permitiu a obtenção do índice de retenção com alta estabilidade e validade via regressão linear simples usando um único descritor que decodifica a molécula tanto nos aspectos relacionados a distribuição de carga quanto nos detalhes estruturais, relacionados a tamanho, ramificação e presença de heteroátomos. Isto abre novas possibilidades para o estudo de QSRR e QSAR de diferentes funções orgânicas.

6.1 Recomendações para trabalhos futuros

Encontram-se listadas abaixo algumas sugestões para continuidade deste trabalho, visto que os resultados obtidos foram de boa qualidade estatística:

- Expandir o índice Semi-empírico Eletrotopológico para outras classes de compostos, por exemplo: álcoois, alcanos halogenados, aromáticos e alquil benzenos;
- Aplicação deste índice para prever outras propriedades físico-químicas e biológicas de interesse;
- Verificar a possibilidade de inclusão desse índice no programa computacional E-DRAGON.

7. Referências

1. Kaliszan, R. QSRR: Quantitative Structure-(Chromatographic) retention relationships. *Chemical Reviews*, v. 107, no. 7, p. 3212-3246, 2007.
2. Héberger, K. Quantitative structure-(chromatographic) retention relationships. *J. Journal of Chromatography A*, v. 1158, p. 273-305, 2007.
3. Kaliszan, R. *Quantitative Structure-Chromatographic Retention Relationships*. 1. ed. New York: John Wiley & Sons, 1987.
4. Kaliszan, R. *Structure and Retention in Chromatography*. A Chemometric Approach. Harwood Academic, Amsterdam, 1997.
5. Leitão, A. Montanari, C. A.; Donnici, C. L. Sobre o Uso de Métodos Quimiométricos em Química Combinatória. *Química Nova*, v.23, no.2, p.178-184, 2000.
6. Ferreira, M. M. C. Multivariate QSAR. *Journal of the Brazilian Chemical Society*, v. 13, no. 6, 742-753, 2002.
7. Silva, L. R.; Ferreira, M. M. C. Estudo do Coeficiente de Partição Octanol-Água de Bifenilas Policloradas (PCBs) Utilizando Parâmetros Topológicos. *Química Nova*, v. 26, no. 3, p.312-318, 2003.
8. Heinzen, V. E. F.; Yunes, R. A. Relationship between gas chromatographic retention indices and molecular connectivity indices of chlorinated pesticides and structurally related compounds. *Journal of Chromatography A*, v. 598, no.2, p.243-50, 1992.
9. Heinzen, V. E. F.; Yunes R. A. Correlation between gas chromatographic retention indices of linear alkylbenzene isomers and molecular connectivity indices, *Journal of Chromatography A*, v. 654, no.1, p.183-189, 1993.

10. Kier, L.B.; Hall, L.H. *Molecular connectivity in chemistry and drugs research*, Academic Press: New York, 1976.
11. Kier, L. B.; Hall, L. H. Structure-Activity studies on odor molecules using molecular connectivity. *Journal of Theoretical Biology*, v. 67, no. 3, p. 585-595, 1977.
12. Randic, M. The connectivity index 25 years after. *Journal of Molecular Graphics and Modelling*, v.20, no.1, p.19-35, 2001.
13. Heinzen, V. E. F.; Soares, M. F.; Yunes, R. A. Semi-empirical topological method for prediction of the chromatographic retention of *cis*- and *trans*-alkene isomers and alkanes. *Journal of Chromatography A*, v. 849, no.2, p.495-506, 1999.
14. Junkes, B. S.; Amboni, R. D. M. C.; Heinzen, V. E. F.; Yunes, R. A. Use of a semi-empirical topological method to predict the chromatographic retention of branched alkenes, *Chromatographia*, v. 55, no.1-2, p.75-81, 2002.
15. Junkes, B. S.; Amboni, R. D. M. C.; Heinzen, V. E. F.; Yunes, R. A. Quantitative structure-retention relationships (QSRR), using the optimum semi-empirical topological index, for methyl-branched alkanes produced by insects. *Chromatographia*, v. 55, no.11-12, p.707-713, 2002.
16. Amboni, R. D. M. C.; Junkes, B. S.; Heinzen, V. E. F.; Yunes, R. A. Semi-empirical topological method for prediction of the chromatographic retention of esters. *Journal of Molecular Structure: THEOCHEM*, v. 579, p.53-62, 2002.
17. Junkes, B. S.; Amboni, R. D. M. C.; Yunes, R. A.; Heinzen, V. E. F. Semi-empirical topological index: a novel molecular descriptor for quantitative structure-retention relationship studies. *Internet Electronic Journal of Molecular Design.*, v. 2, p.33-49, 2003.
18. Amboni, R. D. M. C.; Junkes, B. S.; Yunes, R. A.; Heinzen, V. E. F. Quantitative structure-property relationship study of chromatographic retention indices and normal

boiling points for oxo compounds using the semi-empirical topological method, *Journal of Molecular Structure: THEOCHEM*, v. 586, p.71-80, 2002.

19. Junkes, B. S.; Amboni, R. D. M. C.; Yunes, R. A.; Heinzen, V. E. F. Prediction of chromatographic retention of saturated alcohols on stationary phases of different polarity applying the novel semi-empirical topological index, *Analytica Chimica Acta*, v. 477, p.29-39, 2003.

20. Soares, M. F. *Aplicação da relação entre estrutura e retenção cromatográfica (QSRR) empregando diferentes descritores para cumarinas, acetofenonas e triterpenos. Estudo de um novo índice para alcanos e alcenos*. 2000, 154f. Tese (Doutorado em Química Analítica) – Programa de Pós-Graduação em Química, UFSC, Florianópolis.

21. Amboni, R. *Estudo de correlação quantitativa entre estrutura e propriedade (QSPR) usando descritores topológicos para compostos carbonílicos alifáticos*. 2001. Tese (Doutorado em Química) – Curso de Pós-Graduação em Química, UFSC, Florianópolis.

22. Junkes, B. S. *Índice semi-empírico topológico: desenvolvimento e aplicação de um novo descritor molecular em estudos de correlação quantitativa estrutura-propriedade (QSPR)*. 2003. Tese (Doutorado em Química) – Curso de Pós-Graduação em Química, UFSC, Florianópolis.

23. Arruda, A. C. S. *Ampliação e aplicação do método semi-empírico topológico (I_{ET}) e modelos QSRR/QSPR/QSAR para compostos alifáticos halogenados e cicloalcanos*. 2008. Tese (Doutorado em Química) – Curso de Pós-Graduação em Química, UFSC, Florianópolis.

24. Heinzen, V. E. F.; Junkes, B. S.; Amboni, R. D. M. C.; Yunes, R. A. Application of the Semi-empirical Topological Index in Quantitative Structure-Chromatographic Retention Relationship (QSRR) Studies of Aliphatic Ketones and Aldehydes on Stationary Phases of Different Polarity. *Journal of the Brazilian Chemical Society*, v. 15, no. 2, p. 183-189, 2004.

25. Heinzen, V. E. F.; Junkes, B. S.; Yunes, R. A.; Arruda, A. C.; Porto, L. C. Semi-empirical topological index: a tool for QSPR/QSAR studies. *Journal of Molecular Modeling*, v. 11, p. 128-134, 2005.
26. Porto, L. C.; Souza, E. S.; Junkes, B. S.; Yunes, R. A.; Heinzen, V. E. F. Semi-Empirical Topological Index: development of QSPR/QSRR and optimization for alkylbenzenes. *Talanta* (Oxford), v. 76, no.2, p. 407-412, 2008.
27. Arruda, A. C. S.; Souza E. S.; Junkes, B. S.; Yunes, R. A.; Heinzen, V. E. F. Semi-empirical topological index to predict properties of halogenated aliphatic compounds. *Journal of Chemometrics*, v. 22, no.3-4, p. 186-194, 2008.
28. Romanelli, G. P.; Cafferata, L. F. R.; Castro, E. A. An improved QSAR study of toxicity of saturated alcohols. *Journal of Molecular Structure: THEOCHEM*, v. 504, p.261-265, 2000.
29. Balaban, A. T. Can topological indices transmit information on properties but not on structures? *Journal of Computer-Aided Molecular Design*, v. 19, no. 9-10, p.651-660, 2005.
30. Mihalic, Z.; Trinajstic, N. A graph-theoretical approach to structure property relationships. *Journal of Chemical Education*, v. 69, no. 9, p. 701-712, 1992.
31. Neves, P. J., Costa, J. B. N., Ndiya, P. M. TOP: Um programa de cálculo de descritores topológicos para uso em correlações entre estrutura e atividade. *Química Nova*, v. 21, n. 6, p. 709-713, 1998.
32. Wang, Z. Y.; Zeng X. L.; Zhaia Z. C. Prediction of supercooled liquid vapor pressures and n-octanol/air partition coefficients for polybrominated diphenyl ethers by means of molecular descriptors from DFT method. *Science of the Total Environment*, v. 389, p. 296-305, 2008.
33. Katritzky, A. R.; Maran, V.; Lobanov, V.C.; Karelson, M. Structurally diverse quantitative structure-property relationship correlation of technologically relevant

physical properties. *Journal of Chemical Information and Computer Sciences*, v. 40, no. 1, p.1-18, 2000.

34. Nelson, S. D.; Seybold, P. G. Molecular structure-property relationships for alkenes, Molecular structure-property relationships for alkenes. *Journal of Molecular Graphics & Modelling*, v. 20, no. 1, 36-53, 2001.

35. Seybold, P. G.; May, M.; Bagal, U. A. Molecular structure-property relationships. *Journal of Chemical Education*, v. 64, no. 7, p. 575-582, 1987.

36. Gaudio, A. C. Modelos clássicos de estudo quantitativo das relações entre estrutura química e atividade biológica. *Química Nova*, v. 19, no. 13, p. 278- 289, 1996.

37. Pompe, M.; Novic, M. Prediction of gas-chromatographic retention indices using topological descriptors. *Journal of Chemical Information and Computer Sciences*, v. 39, no. 1, p. 59- 67, 1999.

38. Gaudio, A. C.; Zandonade, E. Proposição, validação e análise dos modelos que correlacionam estrutura química e atividade biológica. *Química Nova*, v. 24, n. 5, p. 658-671, 2001.

39. Hansen, P. J.; Jurs, P. C. Chemical applications of graph theory. *Journal of Chemical Education*, v. 65, no. 7, p. 574-580, 1988.

40. Hansch, C.; Hoekman, D.; Gao, H. Comparative QSAR: Toward a deeper understanding of chemicobiological interactions. *Chemical Reviews*, v. 96, no. 3, p. 1045-1075, 1996.

41. Schweitzer, R. C.; Morris, J. B. The development of a quantitative structure property relationship (QSPR) for the prediction of dielectric constants using neural networks. *Analytica Chimica Acta*, v. 384, no. 3, p.285-303, 1999.

42. Hansch, C.; Fujita, T. ρ , σ , π Analysis. A method for the correlation of biological activity and chemical structure. *Journal of the American Chemical Society*, v.86, no. 8, p.1616-1626, 1964.
43. Kubinyi, H. *QSAR: Hansch Analysis and Related Approaches*. Mannhold, R., Krosgaard-Larsen, P., Timmerman, H., Eds., vol. 1, VCH: Weinheim, 1993.
44. Tute, M. S. In *Physical Chemical Properties of Drugs*; Yalkowsky, S. H.; Sinkula, A. A.; Valvani, S.C., Eds.; Marcel Dekker: New York, 1980;
45. Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. *Chemometrics: a practical guide*. 1 ed. New York : John Wiley & Sons, 1998.
46. Livingstone, D. *Data Analysis for Chemists. Applications to QSAR and Chemical Product Design*. New York: Oxford University Press Inc. 1995.
47. Duperray, B.; Chastrette, M.; Makabeh. M. C.; Pacheco, H. Comparative analysis of Hansch, Free-Wilson and Darc-Pelco correlations in a bactericide series of halogenophenols. *European Journal of Medicinal Chemistry*, v. 11, no. 4, p.323-336, 1976.
48. Mercier, C.; Dubois, J. E. Comparison of molecular connectivity and Darc-Pelco methods-performance in anti-microbial, halogenated phenol QSARS. *European Journal of Medicinal Chemistry*, v. 14, no. 5, p. 415-423, 1979.
49. Dubois, J. E.; Mercier, C., Panaye, A. Analysis of bactericide activity for aliphatic-alcohols and beta-naphthol populations using Hansch and Darc-Pelco methods-lengthening chain effect. *European Journal of Medicinal Chemistry*, v. 36, p. 135-169, 1986.
50. Chrétien, J. R.; Dubois, J. E. Topological analysis of gas-liquid-chromatographic behavior of alkenes. *Analytical Chemistry*, v. 49, n. 6, p. 747-756, 1977.

51. Chrétien, J. R.; Dubois, J. E.; Hirsch, R. F. Gas-solid chromatography on cation-exchange resins - Darc topological analysis of the behavior of alkenes on nickel ions. *Journal of Chromatography*, v. 207, no. 1, p. 115-130, 1981.
52. Mercier, C.; Troullier, G.; Dubois, J. E. Darc computer-aided-design in anticholinergic research. *Quantitative Structure-Activity Relationships*, v. 9, no. 2, p.88-93, 1990.
53. Katritzky, A. R. ; Tulp, I.; Fara, D. C.; Lauria, A.; Maran, U.; Acree, W. E. A general treatment of solubility. 3. Principal component analysis (PCA) of the solubility of diverse solutes in diverse solvents. *Journal of Chemical Information and Modeling*, v. 45, no. 4, p. 913-923, 2005.
54. Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-Chemical descriptors in QSAR/QSPR studies. *Chemical Reviews*, v. 96, no. 3, p.1027-1043, 1996.
55. Randic, M. The structural origin of chromatographic retention data. *Journal of Chromatography*, v. 161, p. 1-14, 1978.
56. Zhang, X.; Lu, P. Unified equation between Kováts indices on different stationary phases for select types of compounds. *Journal of Chromatography*, v. 731, no. 1-2, p. 187-199, 1996.
57. Chen, S.; Liu, H.; Wang, Z. Study of quantitative structure-retention relationship (QSRR) of gas chromatography for polychlorinated dibenzothiophenes on non-polar columns. *QSAR & Combinatorial Science*, v. 26, n.8, p.889-896, 2007.
58. Ren, Y.; Liu, H.; Yao, X.; Liu, M. An accurate QSRR model for the prediction of the GCxGC-TOFMS retention time of polychlorinated biphenyl (PCB) congeners. *Analytical and Bioanalytical Chemistry*, v. 388, n. 1, p. 165- 172, 2007.
59. Hall, L.; H.; Kier, L. B. Issues in representation of molecular structure. The development of molecular connectivity. *Journal of Molecular Graphics & Modelling*, v. 20, p. 4-18, 2001.

60. Katritzky, A. R.; Petrukhin, R.; Tatham, D.; Interpretation of quantitative structure-property and –activity relationships. *Journal of Chemical Information and Modeling*, v. 41, p. 679-685, 2001.
61. Casañola-Martín, G. M.; Marrero-Ponce, Y.; Khan, M. T. H.; Ather, A.; Sultan, S.; Torrensc F.; Rotondoh R. TOMOCOMD-CARDD descriptors-based virtual screening of tyrosinase inhibitors: Evaluation of different classification model combinations using bond-based linear indices. *Bioorganic & Medicinal Chemistry*. v. 15, p. 1483–1503, 2007.
62. Abraham, M. H.; Sánchez-Moreno, R.; Cometto-Muñiz, J. E.; Cain, W. S. A. Quantitative Structure–Activity Analysis on the Relative Sensitivity of the Olfactory and the Nasal Trigeminal Chemosensory Systems. *Chemical Senses*. v. 32, p. 711–719, 2007.
63. Bertinetto, C.; Duce, C.; Micheli, A.; Solaro, R.; Starita, A.; Tiné, M. R. Prediction of the glass transition temperature of (meth)acrylic polymers containing phenyl groups by recursive neural network. *Polymer*. v. 48, p. 7121-7129, 2007.
64. Liu, F.; Liang, Y.; Cao, C.; Zhou, N. QSPR study of GC retention indices for saturated esters on seven stationary phases based on novel topological indices. *Talanta*. v. 72, p. 1307–1315, 2007.
65. Rücker, C.; Rücker, G.; Meringerjj, M. y-Randomization and Its Variants in QSPR/QSAR. *Journal of Chemical Information and Modeling*, v. 47, p. 2345-2357, 2007.
66. Ivanciuc, O.; Ivanciuc, T.; Balaban, A. T. Quantitative structure-property relationship study of normal boiling for halogen-/oxygen-/sulfur-containing organic compounds using the CODESSA program, *Tetrahedron*, v. 54, p.9129-42, 1998.
67. Ivanciuc, O.; Ivanciuc, T.; Balaban, A. T. Quantitative structure-property relationship for the normal boiling temperature of acyclic carbonyl compounds, *Internet Electronic Journal of Molecular Design*. v. 1, p. 252-68, 2002.

68. Derde, M. P.; Massart, D. L. Supervised pattern recognition: the ideal method? *Analytica Chimica Acta*, v. 191, p. 1-16, 1986.
69. Rossiter, K. A. Quantitative relationships between structure and the fruity odor of esters. *Perfum & Flavor*, v. 21, p. 33-46, 1996.
70. Tominaga, Y. Comparative study of class data analysis with PCA-LDA, SIMCA, PLS, and k-NN. *Chemometrics and Intelligent Laboratory Systems*, v.49, p.105-115, 1999.
71. Héberger, K.; Görgényi, M. Principal component analysis of Kováts indices for carbonyl compounds in capillary gas chromatography. *Journal of Chromatography A*, v. 845, p.21-31, 1999.
72. Körtvélyesi, T.; Görgényi, M.; Héberger, K.; Correlation between retention indices and quantum chemical descriptors of ketones and aldehydes on stationary phases of different polarity. *Analytica Chimica Acta*, v. 28, p. 73-82, 2001.
73. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, v. 58, p.109-130, 2001.
74. Liu, S.; Cai, S.; Cao, C.; Li, Z. Molecular electronegative distance vector (MEDV) related to 15 properties of alkanes. *Journal of Chemical Information and Computer Sciences*, v. 40, no. 6, p. 1337-1348, 2000.
75. Duchowicz, P.; Castro, E. A.; Toporov, A. A. QSPR modeling of normal boiling point of aldehydes, ketones, and esters by means of nearest neighboring codes correlation weighting, *Journal of the Argentina Chemical Society*, v. 90, 2002.
76. James, A. T.; Martin, J. P. Gas-liquid partition chromatography: the separation and micro-estimation of volatile fatty acids from formic acid to dodecanoic acid. *Biochemical Journal*, v. 50, p.679-690, 1952.

77. Eiceman, G. A.; Hill, H. H.; Davani, B.; Gardea-Torres, J. Gas chromatography. *Analytical Chemistry*, v. 68, no. 12, p. R291-R308, 1996.
78. Bartle, K. D.; Myers, P. History of gas chromatography. *Trac – Trends in Analytical Chemistry*, v. 21, no. 9-10, p.547-557, 2002.
79. Brinkman, U. A. T.; Janssen, H. G. 50 years of gas chromatography. *Trac – Trends in Analytical Chemistry*, v. 21, no. 9-10, p.545-546, 2002.
80. Kováts, E. Gas-chromatographische charakterisierung organischer verbindungen. Teil 1: Retentionsindeces aliphatischer halogenide, alkole, aldehyde und ketone. *Helvetica Chimica Acta*, v.41, p.1915-1932, 1958.
81. Kováts, E. sz. Zu Fragen der polarität. die method der linearkombination der wechselwirkungskräfte (LKWW). *Chimia*, v. 22, p. 459-463, 1968.
82. *The Sadtler Standard Gas Chromatography Retention Index Library*, Sadtler Research Laboratories, Philadelphia, PA, 1985.
83. Budahegyi, M. V.; Lombosi, E. R; Lombosi, T. S; Mészáros, S. Y; Nyiredy, Sz.; Tarján, G.; Timar, I.; Takács, J. M. Twenty-fifth anniversary of the retention index system in gas-liquid chromatography. *Journal of Chromatography*, v. 271, p.213-307, 1983.
84. Tarján, G.; Nyiredy, Sz.; Györ, M.; Lombosi, E. R.; Lombosi, T. S.; Budahegyi, M. V.; Mészáros, S. Y.; Takács, J. M. Thirtieth anniversary of the retention index according to kováts in gas-liquid chromatography. *Journal of Chromatography*, v. 472, p.1-91, 1989.
85. Eiceman, G. A.; Hill, H. H.; Davani, B.; Gardea-Torres, J. Gas-chromatography. *Analytical Chemistry*, v. 70, p. 321R-339R, 1998.

86. Hosoya, H. Topological index. A newly proposed quantity characterizing topological nature of structural isomers of saturated hydrocarbons. *Bulletin of the Chemical Society of Japan*, v. 44, no. 9, p.2332-2339, 1971.
87. Peng, T. C. Prediction of retention indices V. Influence of electronic effects and column polarity on retention index. *Journal of Chromatography A*, v. 903, p.117-143, 2000.
88. Scott, R. P. W. The role of molecular interactions in chromatography. *Journal of Chromatography*, v. 122, p.35-63, 1976.
89. Lamparczy, K. H.; Radecki, A. The role of electric interactions in the retention index concept-implications in quantitative structure-retention studies. *Chromatographia*, v. 18, p. 615-618, 1985.
90. Lamparczy, K. H. The role of electric interactions in the retention index concept-universal interaction indexes for GLC, HPLC, and TLC. *Chromatographia*, v. 20, p.283-288, 1985.
91. Ciola, R. *Introdução à cromatografia em fase gasosa*. Edgard Blücher, São Paulo, 1973.
92. Anker, L. S.; Jurs, P. C. Quantitative structure-retention relationship studies of odor-active aliphatic compounds with oxygen-containing functional groups. *Analytical Chemistry*, v. 62, p.2676-2684, 1990.
93. Randic, M. Generalized molecular descriptors. *Journal of Mathematical Chemistry*, v. 7, p.155- 168, 1991.
94. Randic, M.; Zupan, J. On Interpretation of Well-Known Topological Indices. *Journal of Chemical Information and Modeling*, v. 41, p. 550-560, 2001.

95. Pogliani, L. From molecular connectivity indices to semi empirical connectivity terms: recent trends in graph theoretical descriptors. *Chemical Review*, v. 100, p.3827-3858, 2000.
96. Wiener, H. Structural determination of paraffin boiling point. *Journal of the American Chemical Society*, v. 69, p.17-20, 1947.
97. Katritzky, A. R.; Gordeeva, E. V. Traditional topological indices vs electronic, geometrical, and combined molecular descriptors in QSAR/QSPR research. *Journal of Chemical Information and Computer Sciences*, v. 33, p.835-857, 1993.
98. Ivanciuc, O.; Ivanciuc, T.; Cabrol-Bass, D.; Balaban, A. T.; Optimum structural descriptors derived from the Ivanciuc-Balaban Operator. *Internet Electronic Journal of Molecular Design*, v. 1, p. 319-331, 2002.
99. Sutter, J. M.; Petersen, T. A.; Jurs, P. C. Prediction of gas chromatographic retention indices of alkylbenzenes. *Analytica Chimica Acta*, v. 342, n. 2-3, p.112-113, 1997.
100. Payares, P.; Dias, D.; Olivero, J.; Vivas, R.; Gómez, I. Prediction of the gas chromatographic relative retention times of flavonoids from molecular structure. *Journal of Chromatography A*, v. 771, p.213-219, 1997.
101. Olivero, J.; Kannan, K. Quantitative structure-retention relationships of polychlorinated naphthalenes in gas chromatography. *Journal of Chromatography A*, v. 849, p.621-627, 1999.
102. Olivero, J.; Gracia, T.; Payares, P.; Vivas, R.; Diaz, D.; Daza, E.; Geerlings, P. Molecular structure and gas chromatographic retention behavior of the components of ylang-ylang oil. *Journal of Pharmaceutical Sciences*, v. 86, no. 5, p.625-630, 1997.

103. Bonchev, D.; Markel, E.; Dekmezian, A. Topological analysis of long-chain branching patterns in polyolefins, *Journal of Chemical Information and Computer Sciences*, v. 41, p.1274-1285, 2001.
104. Randic, M.; Basak, C. A. New descriptor for structure-property and structure-activity correlations. *Journal of Chemical Information and Computer Sciences*, v. 41, p. 650-656, 2001.
105. Randic, M. On history of the randic index and emerging hostility toward chemical graph theory. *Match-Communications In Mathematical And In Computer Chemistry*, v. 59,p. 105-124, 2008.
106. Basak, S. C.; Nikolic, S.; Trinajstic, N. QSPR modeling: Graph connectivity indices versus line graph connectivity indices. *Journal of Chemical Information and Computer Sciences*, v. 40, p. 927-933, 2000.
107. Wessel, M. D.; Sutter, J. M.; Jurs, P. C. Prediction of reduced ion mobility constants of organic compounds from molecular structure. *Analytical Chemistry*, v. 68, p. 4237-4243, 1996.
108. Todeschini, R.; Consonni, V. *Handbook of molecular descriptors*. VCH, Wiley, 2000.
109. Estrada, E. The structural interpretation of randic index. *Internet Electronic Journal of Molecular Design*, v. 1, p. 360-366, 2002.
110. Hosoya, H. The topological index Z before and after 1971. *Internet Electronic Journal of Molecular Design*, v. 1, p. 428-442, 2002.
111. Randic, M. On characterization of molecular branching, *Journal of the American Chemical Society*, v. 97, p. 6609-6615, 1975.
112. Sabljic, A.; Trinajstic, N. Quantitative structure-activity relationships: the role of topological indices. *Acta Pharm. Jugosl.*, v. 31, p.189-214, 1981.

113. Balaban, A. T. Topological and stereochemical molecular descriptors for databases useful in QSAR, similarity/dissimilarity and drug design. *SAR and QSAR in Environmental Research*, v. 8, p. 1-21, 1998.
114. Kier, L. B.; Hall, L. H. *Molecular connectivity in structure-activity analysis*. Letchworth: Research Studies Press, 1986.
115. Estrada, E. Recent advances on the role of topological indices in drug discovery research. *Current Medicinal Chemistry*, v. 8, no. 13, p. 1573-1588, 2001.
116. Ríos-Santamarina, I.; García-Domenech, R.; Cortijo, J.; Santamaria, P.; Morcillo, E.; Gálvez, J. Natural compounds with bronchodilatador activity selected by molecular topology, *Internet Electronic Journal of Molecular Design*, v. 1, p. 70-79, 2002.
117. Estrada, E. Edge adjacency relationships in molecular graphs containing heteroatoms: A new topological index related to molar volume. *Journal of Chemical Information and Computer Sciences*, v. 35, p. 701-707, 1995.
118. Bonchev, D. Topological order in molecules 1. Molecular branching revisited. *Journal of Molecular Structure: THEOCHEM*, v. 366, p. 137-156, 1995.
119. Randić, M. Chemical structure-what is "she"? *Symposium on Graph Theory in Chemistry*, v. 69, n. 9, p. 713-718, 1992.
120. Bonchev, D.; Trinajstić, N. Information-theory, distance matrix, and molecular branching. *Journal of Chemical Physics*, v. 67, p. 4517-4533, 1977.
121. Müller, W. R.; Szymanski, K.; Knop, J. V.; Trinajstić, N. Molecular topological index. *Journal of Chemical Information and Computer Sciences*, v.30, p.160-3, 1990.

122. Katritzky, A. R.; Fara, D. C. How chemical structure determines physical, chemical and technological properties: An overview illustrating the potential of quantitative structure-property relationships for fuels science. *Energy & Fuels*, v. 19, p. 922-935, 2005.
123. Ren, B. Novel atomic-level-based AI topological descriptors: Applications to QSPR/QSAR modeling. *Journal of Chemical Information and Computer Sciences*, v. 42, p. 858- 868, 2002.
124. McQuarrie, D. A.; Simon, J. D. *Physical Chemical: A Molecular Approach*. Sausalito: University Science Books, c1997.
125. Tegmark, M.; Wheeler, J.A. 100 years of quantum mysteries. *Scientific American*, v. 284, no. 2, p. 68-75, 2001.
126. Lowe, J. P. *Quantum Chemistry*. 2nded. San Diego: Academic Press, 1993.
127. Freitas, L. C. G. Prêmio Nobel de Química em 1998: Walter Kohn e John A. Pople. *Química Nova*, vol. 22, no. 2, 1999
128. Bredow, T.; Jug, K. Theory and range of modern semiempirical molecular orbital methods. *Theoretical Chemistry Accounts*, v. 113, p. 1-14, 2005.
129. Soares, M. F.; Boing, C. R.; Heinzen, V. E. F.; Yunes, R. A. Aplicação da teoria de QSRR na interpretação da retenção cromatográfica de uma série de acetofenonas, *Anais Associedade Brasileira de Química*, v. 49, p. 24-30, 2000.
130. Koutek, B.; Hoskovec, M.; Streinz, L.; Vrkocová, P.; Ruzicka, K. Additivity of vaporization properties in pheromone-like homologous series. *Journal of Chemical Society Perkin Trans. 2*, v. 6, p. 1351-1355, 1998.
131. Navarro, D. M. A. F.; Murta, M. M.; Duarte, A. G.; Lima, I. S.; Nascimento, R. R.; Sant'ana, A. E. G.; Aspectos práticos relacionados ao uso do rincoferol, o feromônio de agregação da broca-do-olho-do-coqueiro *rhyncophorus palmarum* L.

- (Coleóptera: curculionidae) no controle de pragas do coqueiro. Análise de sua eficiência em campo. *Química Nova*, v. 25, p. 32-36, 2002.
132. Katritzky, A. R.; Chen, K.; Maran, U.; Carlson, D. A. QSPR Correlation and predictions of GC retention indexes for methyl-branched hydrocarbons produced by insects. *Analytical Chemistry*, v. 72, p. 101-109, 2000.
133. Kissin, Y. V.; Payne, W. B. Gas chromatographic analysis of polymethyl-substituted alkanes. *Journal of Chromatographic Science*. v. 24, p. 164-169, 1986.
134. Carlson, D. A.; Bernier, U. R.; Sutton, B. D. Elution patterns from capillary GC for methyl-branched alkenes. *Journal Chemical Ecology*, v.24, p. 1845-1865, 1998.
135. Schulz, S. Composition of the silk lipids of the spider *Nephila clavipes*. *Lipids*, v. 36, p. 637-647, 2001.
136. Chastrette, M.; Cretin, D.; Aïdi, C. El. Structure-odor relationships: using neural networks in the estimation of camphoraceous or fruity odors and olfactory thresholds of aliphatic alcohols. *Journal of Chemical Information and Computer Sciences*, v. 36, p. 108-113, 1996.
137. Sojác, L.; Addová, G.; Kubinec, R.; Kraus, A.; Hu, G. Gas chromatographic-mass spectrometric characterization of all acyclic C5-C7 alkenes from fluid catalytic cracked gasoline using polymethyl siloxane and squalene stationary phases. *Journal of Chromatography A*, v. 947, p. 103-117, 2002.
138. Chen, E. N.; Drinkwater, D. E.; McCann, J. M. Compositional analysis of hydrocarbon groups in *Journal of Chromatographic Science*, v. 33, p. 535-559, 1995.
139. Ramnäs, O.; Östermark, U.; Petersson, G. Assessment by gas-chromatography mass-spectrometry of hexenes emitted to air from petrol. *Journal of Chromatography A*, v. 638, p. 65-69, 1993.

140. Khorasheh, F.; Gray, M. R.; Selucky, M. Correlation for Kováts retention index of C₉-C₂₆ monoalkyl and polymethyl alkanes and alkenes *Journal of Chromatograph*, v. 481, p. 1-16, 1989.
141. Duvenbeck, C. H.; Zinn, P. List operations on chemical graphs. 3. Development of vertex and edge models for fitting retention index data. *Journal of Chemical Information and Computer Sciences*, v. 33, p. 211-219, 1993.
142. Hu, Z.; Zhang, H. Prediction of gas chromatographic retention indices of alkenes from the total solubility parameters. *Journal of Chromatography A*, v. 653, p. 275-282, 1993.
143. Dimov, N.; Osman, A. Selection of molecular descriptors used in quantitative structure-gas chromatographic retention relationships. II. Isoalkanes and alkenes. *Analytica Chimica Acta*, v. 323, p. 15-25, 1996.
144. Dimov, N.; Osman, A. Influence of conformation on the accuracy of quantitative-structure retention relationship calculations in gas chromatography. *Journal of Chromatography A*, v. 773, p. 368-371, 1997.
145. Perring, K. D.; Duprey, R. J. H.; Ness, J. N. Perfume composition for, e.g. room fresheners or room deodorants, clothes deodorants, bathroom accessories, or household cleansers, includes ethers, aldehydes or nitriles, esters, alcohols, or ketones Patent Number: **WO2006095200-A1; EP1855761-A1; US2008108542-A1**
146. Koivusalmi, E.; Kilpeläinen, I.; Karhunen, P. Base oil production, used in e.g. lubricating oils for automobiles, involves condensing feedstock comprising aldehyde or ketone of biological origin in presence of aldol condensation catalyst, hydrodeoxygenating, and hydroisomerizing
Patent Number: **WO2007068797-A2; FI200505665-A; EP1966356-A2**
147. Sell, C. S. *Flavors and Fragrances: A World Perspective; Proceedings of the 10th International Congress of Essential Oils, Fragrances and Flavors*; Lawrence,

B. M., Mookherjee, B. D.; Willis, B. J. Eds.; Elsevier Science Publishers B. V.: Amsterdam, The Netherlands, 1988.

148. Rossiter, K. J. Structure-odor relationships. *Chemical Reviews*, v. 96, p. 3201-3240, 1996.

149. Bosnjak, N.; Mihalic, Z.; Trinajstic, N. Application of topological indexes to chromatographic data – calculation of the retention indexes of alkanes. *Journal of Chromatography*, v. 540, p. 430-440, 1991.

150. Dubois, J. E.; Chretien, J.R.; Sojak, L. Topological analysis of the behaviour of linear alkenes up to tetradecenes in gas-liquid chromatography on squalane. *Journal of Chromatography*, v. 194, p. 121-134, 1980.

151. Hilal, S. H.; Carreira, L. A.; Karickhoff, S. W.; Melton, C. M. Estimation of Gas-Liquid-Chromatographic Retention Times from Molecular-Structure. *Journal of Chromatography A*, v. 662, p. 269-280, 1994.

152. Ren, B. Atom-type-based AI topological descriptors for quantitative structure–retention index correlations of aldehydes and ketones. *Chemometrics and Intelligent Laboratory Systems*, v. 66, p. 29-39, 2003.

153. Héberger, K.; Görgényi, M.; Sjöström, M. Partial least squares modeling of retention data of oxo compounds in gas chromatography. *Chromatographia*, v. 51, p. 595-600, 2000.

154. Héberger, K.; Gorgenyi, M.; Kowalska, T. Temperature dependence of Kováts indices in gas chromatography revisited. *Journal of Chromatography A*, v. 973, p. 135-142, 2002.

155. Amboni, R. D. M. C.; Junkes, B. S.; Yunes, R. A.; Heinzen, V. E. F. Quantitative Structure-Odor Relationships of Aliphatic Esters Using Topological Indices, *Journal of Agricultural and Food Chemistry*, v. 48, p. 3517-3521, 2000.

156. Lu, C.; Guo, W.; Yin, C. Quantitative structure-retention relationship study of gas chromatographic retention indices of saturated esters on different stationary phases using novel topological indices. *Analytica Chimica Acta*, v. 561, p. 96-102, 2006.
157. **HyperChem** for Windows, Release 7.01, serial number 12-701-150170036, Hypercube, 2002.
158. Smith, W. B. *Introduction to Theoretical Organic Chemistry and Molecular Modeling*. VCH: New York, 1996, p. 142-143.
159. Mulliken, R. S. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. *Journal of Chemical Physics*, v. 23, p. 1833-1840, 1955.
160. Gough, K.M.; Belohorcova, K.; Kaiser, K. L.E. Quantitative structure-activity relationships (QSARs) of Photobacterium phosphoreum toxicity of nitrobenzene derivatives. *Science of the Total Environment*, v. 142, p. 179-190, 1994.
161. Kikuchi, O. Systematic QSAR Procedures with Quantum Chemical Descriptors *Quantitative Structure-Activity Relationships*, v. 6, p. 179-184, 1987.
162. MicroCal Origin version 5.0, 1997.
163. TsarTM 3.3 for Windows, Oxford Molecular Ltd., 2000.
164. Ferreira, M. M. C.; Antunes, A. M.; Melgo, M. S.; Volpe, P. L. O. Quimiometria I: Calibração multivariada, um tutorial. *Química Nova*, v. 22, p. 724-731, 1999.
165. Ren, B. Novel atom type AI indices for QSPR studies of alcohols. *Journal of Computational & Chemistry*, v. 26, p. 223-235, 2002.
166. Ren, B. Application of novel atom-type AI topological indices in the structure-property correlations. *Journal of Molecular Structure: THEOCHEM*, v. 586, p. 137-148, 2002.

167. Liu, S.; Liu, H.; Xia, Z.; Cao, C.; Li, Z. S. Molecular distance-edge vector (μ) and chromatographic retention index of alkanes. *Journal of the American Chemical Society*, v. 47, p. 455-460, 2000.
168. Wold, S.; Eriksson, L. Validation Tools. In: H. van de Waterbeemd. *Chemometric Methods in Molecular Design*. Weinheim; New York; Basel; Cambridge; Tokyo: VCH, v. 359, p. 309-318, 1995.
169. Souza, E. S.; Junkes, B. S.; Kuhnen, C. A.; Yunes, R. A.; Heinzen, V. E. F. On a new semi-empirical electrotopological index for QSRR models. *Journal of Chemometrics*, v. 22, p. 378-384, 2008.
170. Souza, E. S.; Kuhnen, C. A.; Junkes, B. S.; Yunes, R. A.; Heinzen, V. E. F. Modeling the semi-empirical electrotopological index in QSPR studies for aldehydes and ketones. *Journal of Chemometrics*, 2009. (in press)
171. Souza, E. S.; Kuhnen, C. A.; Junkes, B. S.; Yunes, R. A.; Heinzen, V. E. F.; Quantitative Structure-Retention Relationship Modeling of Esters on Stationary Phases of Different Polarity. *Journal of Molecular Graphics and Modelling*, 2009. (aceito)
172. Rohrschneider, L. Eine methode zur chrakterisierung von gaschromatographischen trennflüssigkeiten. *Journal of Chromatography A*, v. 22, p. 6-22, 1966.
173. McReynolds, W. O. Characterization of some liquid phases. *Journal of Chromatographic Science*, v. 8, p.685-691, 1970.
174. Morales, R. ; Blanco, C. ; Furton, K. G. The gas-liquid chromatographic stationary phase properties of liquid organic salts: Anomalous selectivity variation when employing the Rohrschneider/McReynolds system. *Talanta*, v. 40, no. 10, p. 1541-1549, 1993.

175. Kollie, T. O.; Poole, C. F. Influence of solute size and the non-polar interaction term on the selection of test solutes for the classification of stationary phase selectivity in gas chromatography. *Journal of Chromatography A*, v. 556, no. 1-2, p. 457-484, 1991.