UNIVERSIDADE DE SÃO PAULO

FACULDADE DE CIÊNCIAS FARMACÊUTICAS

Programa de Pós-Graduação em Fármaco e Medicamentos

Área de Insumos Farmacêuticos

Aplicação de metodologias do CADD (*Computer-Aided Drug Design*) a um conjunto de pirrolidina carboxamidas: mapeamento do farmacóforo e planejamento de novos protótipos tuberculostáticos potenciais

Bárbara Athayde Vaz Galvão da Silva

Dissertação para obtenção do grau de

MESTRE

Orientador:

Prof^a. Tit. Elizabeth Igne Ferreira

São Paulo

2012

Bárbara Athayde Vaz Galvão da Silva

Aplicação de Metodologias do CADD (*Computer-Aided Drug Design*) a um conjunto de pirrolidina carboxamidas: mapeamento do farmacóforo e planejamento de novos protótipos tuberculostáticos potenciais

[Versão Corrigida]

Dissertação apresentada à Faculdade de Ciências Farmacêuticas da Universidade de São Paulo para obtenção do grau de MESTRE em Ciências: Fármaco e Medicamentos

Área de concentração: Insumos farmacêuticos

Orientador:

Prof^a. Tit. Elizabeth Igne Ferreira

São Paulo

2012

Bárbara Athayde Vaz Galvão da Silva

Aplicação de Metodologias do CADD (*Computer-Aided Drug Design*) a um conjunto de pirrolidina carboxamidas: mapeamento do farmacóforo e planejamento de novos protótipos tuberculostáticos potenciais

[Versão Corrigida]

Comissão Julgadora da
Dissertação para obtenção do grau de Mestre

Prof^a. Tit. Elizabeth Igne Ferreira
orientador/presidente

Prof. Dr. Osvaldo Andrade Santos-Filho
FIOCRUZ

Prof. Dr. Carlos Maurício Rabello Sant'Anna
UFRRJ

São Paulo, _____ de _____.

*Dedico este trabalho à Dama da
Modelagem Molecular no Brasil,
Orientadora, Pesquisadora,
Professora e Grande Amiga, Dr^a.
Kerly F. M. Pasqualoto, pelo
conhecimento e dedicação
imensuráveis.*

AGRADECIMENTOS

Convivi com pessoas incríveis nestes anos, tive experiências que contribuíram imensamente para meu crescimento acadêmico e profissional, mas, sem dúvida nenhuma, minha maior satisfação no mestrado foi ter sido acompanhada duas ORIENTADORAS extremamente renomadas em suas respectivas áreas.

- ☆ À Prof^a. Tít. Elizabeth Igne Ferreira por me receber como aluna em seu grupo, pela orientação, por disponibilizar todos os meios necessários para que este projeto fosse desenvolvido, foi excepcional trabalhar com pesquisadora tão renomada na área*
- ☆ À Dr^a. Kerly Pasqualoto pela orientação, suporte em todos os momentos e pelas oportunidades que por ela foram criadas para meu crescimento acadêmico*
- ☆ Ao Dr. Roberto Parise Filho, professor que me acompanha desde a graduação, me incentivou a fazer mestrado, pelo intermédio dele conheci minhas orientadoras. Tive a oportunidade de trabalhar com ele novamente, agora como docente da FCF-USP, é o responsável pela minha dedicação e interesse em química farmacêutica.*

AGRADECIMENTOS

- ☆ *As minhas grandes amigas Kely M. Turra e Juliana Pachioni pelo café filosófico de todas as manhãs... É sempre gratificante discussões académicas com grandes amigas, surgiram muitas ideias destes cafés*
- ☆ *Ao Prof. Assoc. Leoberto Costa Tavares pelo conhecimento e experiência compartilhados, assim como aos alunos Salomão, Fanny e Alex, pela amizade e companhia em diversas disciplinas*
- ☆ *Aos alunos e professores da QF, Prof^a. Carlota, Prof^a. Vení, Prof. Gustavo, Ricardo, Marina, Soraya, Vanessa, Jeanine, Natan, Charles, Matheus, Camila, Marcela.*
- ☆ *Ao secretários da pós-graduação David e Elisabeth por todo o apoio*
- ☆ *Ao Prof. Hopfinger pelo suporte e pelas licenças colaborativas dos programas WOLF, 4D-QSAR e MOLSIM*
- ☆ *À CAPES pelo apoio financeiro*

- ☆ *Por fim o agradecimento especial ao meu filho RODRIGO pela renovação diária, não importa as dificuldades do caminho a ser percorrido, ele faz o impossível tornar-se alcançável.*

SUMÁRIO

LISTA DE FIGURAS	viii
LISTA DE TABELAS	xii
LISTA DE QUADROS	xv
LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS	xvi
RESUMO	xviii
ABSTRACT	xix
1. INTRODUÇÃO	20
1.1. Tuberculose	21
1.2. Planejamento racional de novos fármacos e modelagem molecular	27
1.3. Enoil-acp redutase (ENR) – Alvo importante no planejamento de novos agentes tuberculostáticos	29
2. JUSTIFICATIVA E OBJETIVO	41
3. METODOLOGIA	44
3.1. Dados biológicos	45
3.2. Construção e minimização de energia dos modelos moleculares tridimensionais (3D)	51
3.3. Simulações de dinâmica molecular (DM)	52
3.4. Análise QSAR multivariada	54
3.4.1. Cálculo de variáveis independentes ou descritores	56

3.4.2. Seleção de variáveis e construção dos modelos	61
3.4.3. Validação dos modelos QSAR da análise multivariada	65
4. RESULTADOS E DISCUSSÃO	70
4. 1. Simulações de Dinâmica Molecular	71
4.2. Análise QSAR multivariada	84
4.2.1. Seleção de variáveis e construção de modelos QSAR (GFA-PLS) com os dados da Tabela A (descritores eletrônicos calculados com método B3LYP/3-21G*)	86
4.2.2. Seleção de variáveis e construção de modelos QSAR (GFA-PLS) com os dados da Tabela B (descritores eletrônicos calculados com método HF/3-21G*)	97
4.2.3. Seleção de variáveis e construção de modelos QSAR (GFA-PLS) com dados da Tabela C (descritores eletrônicos calculados com método B3LYP/6-31G*)	108
4.2.4. Seleção de variáveis e construção de modelos QSAR (GFA-PLS) com dados da Tabela D (descritores eletrônicos calculados com método HF/6-31G*)	119
4.3. Interpretação do modelo final da análise multivariada	136
4.4. Proposição de novas estruturas	140
5. CONCLUSÕES E PERSPECTIVAS	141
REFERÊNCIAS BIBLIOGRÁFICAS	144
ANEXOS	157

LISTA DE FIGURAS

Figura 1. Número de casos novos de TB no País e respectivas regiões nos anos de 1990, 2000 e 2010 (MINISTÉRIO DA SAÚDE, 2011). _____	22
Figura 2. Número de casos notificados de MDR – TB no mundo no ano de 2010 (WHO, 2011). _____	23
Figura 3. Número de casos notificados de TB em pacientes portadores do HIV no mundo no ano de 2010 (WHO, 2011). _____	23
Figura 4. Número de casos notificados de TB e MDR-TB no Brasil, entre os anos de 2005 a 2009 (WHO, 2011) _____	24
Figura 5. Representação esquemática de componentes estruturais importantes da parede celular do <i>M. tuberculosis</i> (adaptado de SAIER, 2009). _____	30
Figura 6. Classes de micolatos encontrados na parede celular do <i>M. tuberculosis</i> . (Adaptado de PASQUALOTO, 2003; TAKAYAMA et al., 2005). _____	33
Figura 7. Inibidores de ENRs _____	34
Figura 8. Esquema de interação da isoniazida com resíduos de aa do sítio ativo da InhA código 1zid (ROZWARSKI et al., 1998). _____	35
Figura 9. Esquema de interação de um derivado benzodiazoborina com resíduos de aa do sítio ativo da FabI código 1dfg (BALDOCK et al., 1996). _____	36
Figura 10. Representação das interações entre a triclosana e resíduos de aa do sítio ativo da FabI código 1qsg (STEWART, et. al., 1999). _____	37
Figura 11. Estrutura química geral das pirrolidina carboxamidas. _____	38
Figura 12. Representação das interações entre análogo pirrolidina carboxamida e resíduos de aa do sítio ativo da InhA código 2h7i (HE et. al., 2006). _____	39
Figura 13. Dispersão dos dados de atividade biológica dos 46 análogos da classe pirrolidina carboxamidas. Faixa de dispersão delimitada pelo tracejado vermelho. _____	50
Figura 14. Estrutura da unidade fundamental do conjunto de pirrolidina carboxamidas. ____	53

- Figura 15.** Diedros considerados como descritores na análise de QSAR multivariada. θ_1 representado em azul; θ_2 em vermelho, θ_3 em verde, θ_4 em roxo e θ_5 em amarelo . _____ 58
- Figura 16.** Fluxograma da abordagem QSAR multivariada partindo do conjunto inicial de dados (compostos investigados) até o cálculo de variáveis independentes ou descritores. _____ 60
- Figura 17.** Fluxograma referente às etapas de seleção de variáveis e construção de modelos QSAR com o programa WOLF 5.5 (ROGERS, 1994). _____ 64
- Figura 18.** Gráficos referentes à segunda parte da simulação de DM (100 ps a 298 K) dos 46 ligantes. A energia do sistema foi calculada sem considerar as contribuições E_{Hb} e E_{solv} . _____ 73
- Figura 19.** Superfície de van der Waals dos ligantes (A) TBP1, (B) TBP39 e (C) TBP46 (programa Viewer Lite 5.0, Accelrys 2002). _____ 79
- Figura 20.** Segunda parte da simulação de DM (100 ps a 298 K) dos ligantes TBP39, TBP1 e TBP46. E_{tot} sem considerar as contribuições de E_{solv} e E_{Hb} . _____ 79
- Figura 21.** Valores de energia potencial total para os confôrmers selecionados dos PAC. 80
- Figura 22.** Dispersão gráfica do descritor de área de superfície polar (PSA) em relação à atividade biológica. _____ 85
- Figura 23.** Dispersão gráfica do descritor de índice topológico de efeito estérico para o nitrogênio N_1 do núcleo comum às pirrolidina carboxamidas em relação à atividade biológica. _____ 85
- Figura 24.** Dispersão gráfica do descritor de excentricidade para o nitrogênio N_1 do núcleo comum às pirrolidina carboxamidas em relação à atividade biológica. _____ 85
- Figura 25.** Variação dos valores de q^2_{LNO} com a retirada de amostras ($N = 2$ a 23) do modelo 6. _____ 90
- Figura 26.** Visualização gráfica dos valores de q^2_{grand} versus r^2_{grand} para 30 aleatorizações do modelo 6. O dado destacado com retângulo pontilhado vermelho ao modelo real (valores de q^2 e r^2 do modelo 6). _____ 92

Figura 27. Dendrograma obtido com o método HCA para os compostos da tabela A (n=46).	93
Figura 28. Gráfico de dispersão dos valores de resíduo de Student versus leverage para os compostos da Tabela A.	94
Figura 29. Variação dos valores de q^2_{LNO} com a retirada de amostras (N = 2 a 23) do modelo 2.	101
Figura 30. Visualização gráfica dos valores de q^2_{yrand} versus r^2_{yrand} para 30 aleatorizações do modelo 2. O dado destacado com retângulo pontilhado refere-se ao modelo real (valores de q^2 e r^2 do modelo 2).	103
Figura 31. Dendrograma obtido com o método HCA para os compostos da tabela B (n=46).	104
Figura 32. Gráfico de dispersão dos valores de resíduo de Student versus leverage.	105
Figura 33. Variação dos valores de q^2_{LNO} com a retirada de amostras (N = 2 a 22) do modelo 3.	112
Figura 34. Visualização gráfica dos valores de q^2_{yrand} versus r^2_{yrand} para 30 aleatorizações dos valores de y no modelo 3. O dado destacado com retângulo pontilhado amarelo alaranjado se refere ao modelo real (valores de q^2 e r^2 do modelo 3.	114
Figura 35. Dendrograma obtido pelo método HCA para os compostos da tabela de dados C (n=45)	115
Figura 36. Gráfico de dispersão dos valores de resíduo de Student versus leverage.	116
Figura 37. Variação dos valores de q^2_{LNO} com a retirada de amostras (N = 2 a 22) do modelo 7	122
Figura 38. Visualização gráfica dos valores de q^2_{yrand} versus r^2_{yrand} para 30 aleatorizações do modelo 7. O dado destacado com retângulo pontilhado vermelho refere-se ao modelo real (valores de q^2 e r^2 do modelo 7).	125
Figura 39. Dendrograma obtido com o método HCA para os compostos da tabela D de dados (n=45)	126
Figura 40. Gráfico de dispersão dos valores de resíduo de Student vs. leverage.	126

-
- Figura 41.** Variação dos valores de q^2_{LNO} com a retirada de amostras ($N = 2$ a 17) do modelo 1. _____ 130
- Figura 42.** Visualização gráfica dos valores de q^2_{yrand} versus r^2_{yrand} para 30 aleatorizações dos valores de y no modelo 1. O dado destacado com retângulo pontilhado amarelo se refere ao modelo real. _____ 132
- Figura 43.** Visualização gráfica dos valores de pCl_{50} experimental e pCl_{50} previsto pelo modelo 1. _____ 134
- Figura 44.** Representação esquemática para a interpretação do modelo 1 gerado com a tabela D (HF/6-31G*). _____ 136
- Figura 45.** Mapas de potencial eletrostático do ligante mais ativo (TBP39) e menos ativo (TBP14), calculados com método HF/6-31G* (programa GaussView 3.0). _____ 137
- Figura 46.** Esquema das relações observadas e sugeridas entre os dados obtidos com os mapas de potencial eletrostático e as interações entre análogo pirrolidina carboxamida e resíduos de aa do sítio ativo da InhA código 2h7l (HE et. al., 2006). _____ 139

LISTA DE TABELAS

Tabela 1. Medidas estatísticas, número de descritores e número de amostras atípicas dos oito melhores modelos, utilizando fator de ajuste de 0,7 ($n = 46$)	86
Tabela 2. Matriz de correlação-cruzada de resíduos de ajuste entre pares de modelos para os oito melhores modelos	87
Tabela 3. Matriz de correlação-cruzada entre descritores do modelo 6 ($n = 46$)	88
Tabela 4. Valores de q^2_{LNO} e das respectivas variações obtidas na validação cruzada LNO do modelo 6	89
Tabela 5. Valores de q^2_{yrand} e r^2_{yrand} para 30 aleatorizações do y (y-randomization) para o modelo 6	91
Tabela 6. Medidas estatísticas, número de descritores e número de amostras atípicas (outliers) dos oito melhores modelos, utilizando fator de ajuste de 0,01 ($n = 36$)	95
Tabela 7. Matriz de correlação de resíduos de ajuste entre pares de modelos para os oito melhores modelos	96
Tabela 8. Matriz de correlação entre descritores do modelo 2 ($N = 36$)	97
Tabela 9. Medidas estatísticas, número de descritores e número de amostras atípicas dos oito melhores modelos, utilizando fator de ajuste de 0,5 ($n = 46$)	98
Tabela 10. Matriz de correlação-cruzada de resíduos de ajuste entre pares de modelos para os oito melhores modelos	98
Tabela 11. Matriz de correlação-cruzada entre descritores do modelo 2 ($n = 46$)	99
Tabela 12. Valores de q^2_{LNO} e das respectivas variações resultantes na validação cruzada LNO do modelo 2.	100
Tabela 13. Valores de q^2_{yrand} e r^2_{yrand} para 30 aleatorizações do y (y-randomization) no modelo 2	102
Tabela 14. Medidas estatísticas, número de descritores e número de amostras atípicas (outliers) dos oito melhores modelos, utilizando fator de ajuste de 0,01 ($n = 36$)	106

Tabela 15. Matriz de correlação de resíduos de ajuste entre pares de modelos para os oito melhores modelos _____	106
Tabela 16. Matriz de correlação entre descritores do modelo 4 (n = 36) _____	107
Tabela 17. Medidas estatísticas, número de descritores e número de amostras atípicas dos oito melhores modelos, utilizando fator de ajuste de 0,3 (n = 45) _____	108
Tabela 18. Matriz de correlação-cruzada de resíduos de ajuste entre pares de modelos para os oito melhores modelos _____	109
Tabela 19. Matriz de correlação-cruzada entre descritores do modelo 3 (n = 45) _____	110
Tabela 20. Valores de q^2_{LNO} e das respectivas variações resultantes na validação cruzada LNO do modelo 3 _____	111
Tabela 21. Valores de q^2_{yrand} e r^2_{yrand} para 30 aleatorizações dos valores de y para o modelo 3 _____	113
Tabela 22. Medidas estatísticas, número de descritores e número de amostras atípicas (outliers) dos oito melhores modelos, utilizando fator de ajuste de 1,0 (n = 35) _____	117
Tabela 23. Matriz de correlação de resíduos de ajuste entre pares de modelos para os oito melhores modelos _____	117
Tabela 24. Matriz de correlação-cruzada entre descritores do modelo 4 (n = 35) _____	118
Tabela 25. Medidas estatísticas, número de descritores e número de amostras atípicas dos oito melhores modelos, utilizando fator de ajuste de 0,3 (n = 45) _____	119
Tabela 26. Matriz de correlação-cruzada de resíduos de ajuste entre pares de modelos para os oito melhores modelos _____	120
Tabela 27. Matriz de correlação-cruzada entre descritores do modelo 7 (n = 45) _____	121
Tabela 28. Valores de q^2_{LNO} e das respectivas variações resultantes na validação cruzada LNO do modelo 7 _____	122
Tabela 29. Valores de q^2_{yrand} e r^2_{yrand} para 30 aleatorizações do y (y-randomization) no modelo 7 _____	123
Tabela 30. Medidas estatísticas, número de descritores e número de amostras atípicas (outliers) dos oito melhores modelos, utilizando fator de ajuste de 0,5 (n = 35) _____	127

Tabela 31. Matriz de correlação de resíduos de ajuste entre pares de modelos para os oito melhores modelos _____	128
Tabela 32. Matriz de correlação entre descritores do modelo 1 (n = 35) _____	129
Tabela 33. Valores de q^2_{LNO} e das respectivas variações resultantes na validação cruzada LNO do modelo 1 (n = 35) _____	130
Tabela 34. Valores de q^2_{yrand} e r^2_{yrand} para 30 aleatorizações dos valores de y para o modelo 1 _____	131
Tabela 35. Validação externa do conjunto de avaliação utilizando o modelo 1 _____	133
Tabela 36. Previsão dos valores de pCl_{50} para a classe de arilamidas com o modelo 1 _	135
Tabela 37. Previsão dos valores de pCl_{50} com o modelo 1 para os compostos propostos.	137

LISTA DE QUADROS

- Quadro 1.** Esquema de tratamento I para TB pulmonar e extrapulmonar em adultos e adolescentes (PNCT, 2010) _____ 26
- Quadro 2.** Retratamento em recidivantes e retorno após abandono do Esquema I (PNCT, 2010) _____ 26
- Quadro 3.** Conjunto de 26 estruturas da classe química pirrolidina carboxamidas mono ou dissustituídas e respectivos valores de atividade biológica expressos em CI_{50} ($\mu\text{mol/L}$) e pCI_{50} _____ 46
- Quadro 4.** Conjunto de 20 estruturas da classe química pirrolidina carboxamidas com substituintes volumosos e respectivos valores de atividade biológica expressos em CI_{50} ($\mu\text{mol/L}$) e pCI_{50} _____ 47
- Quadro 5.** Valores das contribuições de energia em kcal/mol para os confôrmeros de menor energia selecionados dos PAC dos ligantes do conjunto investigado provenientes das simulações de DM de 200 ps a 298 K _____ 81

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

TB = tuberculose

SIDA = síndrome da imunodeficiência adquirida humana

FAS = *fatty acid synthase* (ácido graxo sintetase)

CADD = *computer-aided drug design* (desenvolvimento de fármacos auxiliado por computador)

acp = *acyl carrier protein* (proteína acil-carreadora)

ENR = enoil redutase

InhA = enoil-acp redutase de *Mycobacterium tuberculosis*

QSAR = *quantitative structure-activity relationships* (relações quantitativas entre estrutura química e atividade biológica)

IR = independente do receptor

OMS = Organização Mundial da Saúde

MDR-TB = *multidrug resistant* (cepas multirresistentes)

XRD-TB = *extensively drug-resistant* (cepas extremamente resistentes)

DOTS = *Directly Observed Treatment, Short-Course* (tratamento de curta duração com observação direta)

PNCT = Programa Nacional de Controle da Tuberculose

IUPAC = *International Union of Pure and Applied Chemistry*

TDM = *trehalose dimycolate* (dimicolato de trealose)

NAD = nicotinamida adenina dinucleotídeo

KatG = enzima catalase peroxidase

FabI = enoil-acp redutase de *Escherichia coli*

PfENR = enoil-acp redutase *Plasmodium falciparum*

HTS = *High Throughput Screening* (ensaios biológicos automatizados em larga escala)

PDB = *Protein Data Bank* (banco de dados de proteínas)

CI₅₀ = concentração inibitória 50%

pCI_{50 prev} = valor da concentração inibitória de 50% prevista

pCI_{50 exp} = valor da concentração inibitória de 50% experimental

CoMFA = *Comparative Molecular Field Analysis* (análise comparativa de campos moleculares)

LAPEN = Laboratório de Planejamento e Síntese de Quimioterápicos Potencialmente Ativos

CLAE = cromatografia líquida de alta eficiência

MM+ = campo de força empírico (*Molecular Mechanics*)

PM3 = método semi-empírico (*Parametrized Model 3*)

DM = dinâmica molecular

ps = unidade de tempo em picosegundos

K = unidade de temperatura em Kelvin

PAC = perfil de amostragem conformacional

RMSD = *root-mean square deviation*

MLR = *multiple linear regression* (regressão linear múltipla)

PLS = *Partial Least Squares* (regressão por quadrados mínimos)

VL = variável latente

CHELPG = *CHarges from Electrostatic Potentials using a Grid based method*

HOMO = *Highest Occupied Molecular Orbital* (energia do orbital molecular ocupado de maior energia)

LUMO = *Lowest Unoccupied Molecular Orbital* (energia do orbital molecular desocupado de menor energia)

PCA = *Principal Component Analysis* (análise por componentes principais)

DFT = *Density Functional Theory* (teoria do funcional de densidade)

HF = *Hartree – Fock*

Carga_C₂ = carga HF/6-31G* no carbono C₂ do núcleo fundamental das pirrolidina carboxamidas

μ = momento de dipolo

α = polarizabilidade

Θ = ângulo de torção ou diedro

RM = refratividade molecular

ClogP = valor calculado do coeficiente de partição *n*-octanol-água

ClogD = valor calculado do coeficiente de distribuição de espécies ionizadas

PSA = *Polar Surface Area* (área de superfície polar)

MSA = *Molecular Surface Area* (área de superfície molecular)

MEP = *Electrostatic Potential Maps* (mapas de potencial eletrostático)

GFA = *Genetic Function Approximation* (aproximação da função genética)

d = *smoothing factor* (fator de ajuste)

LSE = *Least Square Error* (erro de mínimos quadrados)

LOF = *Lack-of-fit* (fator de desajuste de Friedman)

LOO = *Leave-one-out*

LNO = *Leave-N-out*

q² = coeficiente de correlação cruzada

r² = coeficiente de correlação

R = coeficiente de determinação

HCA = *Hierarchical Cluster Analysis* (análise hierárquica de agrupamentos)

DP = desvio-padrão

SILVA, B. A. V. G. **Aplicação de Metodologias do CADD (*Computer-Aided Drug Design*) a um conjunto de pirrolidina carboxamidas: mapeamento do farmacóforo e planejamento de novos protótipos tuberculostáticos potenciais.** 2011. 176p. Dissertação (Mestrado) – Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo, 2011. [Versão corrigida].

RESUMO

A situação da tuberculose (TB) foi alterada de forma significativa pela síndrome de imunodeficiência adquirida (SIDA ou AIDS) e pelo aparecimento de novas cepas do *Mycobacterium tuberculosis* resistentes ao tratamento quimioterápico, que justificariam a pesquisa de novos agentes antimicobacterianos. Alvos interessantes têm emergido para o planejamento racional de novos fármacos contra a TB, particularmente, considerando processos metabólicos específicos que ocorrem durante a biossíntese da parede celular micobacteriana e que envolvem a síntese de ácidos graxos (FAS-II, *fatty acid synthase*). O sistema FAS-II constitui diferença bioquímica importante entre mamíferos e micobactérias. A enzima enoil-acp (*acyl carrier protein*, proteína acil-carregadora) redutase (ENR) é alvo determinante no sistema FAS-II, responsável pela etapa de alongamento dos ácidos micólicos, que são os principais componentes da parede celular do *M. tuberculosis*. O presente projeto tem como objetivo a aplicação de metodologias do planejamento de fármacos auxiliado por computador, CADD (*Computer-Aided Drug Design*), em um conjunto de derivados pirrolidina carboxamidas descritos como inibidores potenciais da ENR do *M. tuberculosis* (InhA) com intuito de mapear o farmacóforo, investigar a orientação dos ligantes no sítio ativo e os tipos de interações que se estabelecem com os resíduos de aminoácidos do sítio de interação. Inicialmente, investigaram-se as relações quantitativas entre estrutura química e atividade biológica (QSAR, *quantitative structure-activity relationships*) com aplicação de abordagem multivariada. O melhor modelo QSAR indicou que propriedades estruturais, termodinâmicas e eletrônicas devem ser consideradas no processo de planejamento e proposição de novos protótipos potencialmente tuberculostáticos.

SILVA, B. A. V. G. **Aplicação de Metodologias do CADD (*Computer-Aided Drug Design*) a um conjunto de pirrolidina carboxamidas: mapeamento do farmacóforo e planejamento de novos protótipos tuberculostáticos potenciais.** 2011. 176p. Dissertação (Mestrado) – Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo, 2011. [Versão corrigida].

ABSTRACT

The incidence of tuberculosis (TB) disease has significantly changed considering the acquired immunodeficiency syndrome (AIDS) co-infection as well as the emergence of new *Mycobacterium tuberculosis* strains resistant to the currently chemotherapy. These facts support the search for novel antimycobacterial agents. Interesting targets have been elucidated and could be used for the rational design of new drugs against TB, primarily those related to specific biochemical metabolic pathways that occur during the cell wall biosynthesis, specially involved in the fatty acid synthase (FAS) system. The FAS-II system is an important biochemical difference between mammals and mycobacteria. The enoyl-ACP reductase (ENR) is the key enzyme in the FAS-II system, responsible for the elongation step of mycolic acids, which are the major components in the *M. tuberculosis* cell wall. This research project aims the application of computer-aided drug design (CADD) methodologies to a set of pyrrolidine carboxamide derivatives, which were previously reported as potential *M. tuberculosis* ENR (InhA) inhibitors, for mapping the pharmacophore, investigating the ligands' orientation at the active site and also the interaction types regarding the amino acid residues in the active site. Initially, the quantitative structure-activity relationships (QSAR) were performed applying a multivariate approach. The best QSAR model indicated the structural, thermodynamic, and electronic properties must be taken into account in the design of novel leads as potential antituberculosis agents.

INTRODUÇÃO

1. INTRODUÇÃO

1.1. Tuberculose

A tuberculose (TB) é doença infecciosa crônica, cujos agentes etiológicos são micobactérias do “complexo tuberculose” (*Mycobacterium bovis*, *Mycobacterium africanum* e, principalmente, *Mycobacterium tuberculosis*). A TB é exemplo clássico de doença causada por parasita intracelular. Na maioria dos casos, a infecção é transmitida de pessoa para pessoa por meio da inalação de gotículas infectadas. A TB pode envolver qualquer órgão sistêmico, mas o pulmão é o órgão mais comum da lesão primária e o principal comprometido (SENSI, GRASSI, 1996).

Desde 1993, a Organização Mundial da Saúde (OMS) declarou a TB como emergência global de saúde pública devido as suas proporções. Cerca de um terço da população mundial está infectada pelo *M. tuberculosis*. Em 1996, houve aproximadamente oito milhões de novos casos da doença, com três milhões de mortes. Cada pessoa doente pode infectar de 10 a 15 pessoas somente em um ano (WHO, 1998, 2002a). Os indivíduos infectados não necessariamente desenvolverão a doença. Esta depende do sistema de defesa específico do hospedeiro, da resposta imunológica e de fatores genéticos (WOLINSKY, 1992; SENSI, GRASSI, 1996).

O Brasil está em 19º lugar dentre os vinte e dois países que concentram 80% dos casos de TB no mundo e somado ao Peru apresentam 50% dos casos registrados nas Américas. Notificaram-se, no Brasil, cerca de oitenta e cinco mil casos de tuberculose por ano, sendo setenta e um mil casos novos, e registraram-se cerca de 4.800 óbitos decorrentes da doença (MINISTÉRIO DA SAÚDE, 2011).

De acordo com os dados do Ministério da Saúde, o número de casos novos no Brasil foi de 74.567 no ano de 1990, 81.181 casos em 2000 e 70.997 casos 2010. A região sudeste é a responsável pela maior variação no número de casos novos, correspondendo a aproximadamente 45% do total de novos casos. Na figura 1 está apresentado o número de casos novos de TB nos anos de 1999, 2000 e 2010 no Brasil e nas respectivas regiões.

Figura 1. Número de casos novos de TB no País e respectivas regiões nos anos de 1990, 2000 e 2010 (MINISTÉRIO DA SAÚDE, 2011).

A OMS estimou que entre 2002 e 2020 cerca de um bilhão a mais de pessoas estarão infectadas, cento e cinquenta milhões estarão doentes e trinta e seis milhões morrerão de TB, se o controle não for fortalecido (WHO, 2002a, 2009a).

A situação mundial de TB foi alterada significativamente pela síndrome de imunodeficiência adquirida (SIDA ou AIDS) e pelo aparecimento de novas cepas do *M. tuberculosis* resistentes ao tratamento quimioterápico, seja por um fármaco isolado ou pela terapia combinada. Neste caso, são denominadas de cepas multirresistentes (*MDR, multidrug-resistant*). Além disso, já são descritos casos de cepas extremamente resistentes (*XDR-TB, extensively drug-resistant*), definidas como cepas que adquirem resistência aos fármacos de primeira escolha e a, no

mínimo, três fármacos de segunda escolha do esquema terapêutico da TB (WHO, 2010).

A figura 2 mostra a notificação mundial de casos de MDR-TB no ano de 2010, as regiões com maior índice de notificação são norte asiático, leste europeu e sul da África.

Figura 2. Número de casos notificados de MDR – TB no mundo no ano de 2010 (WHO, 2011).

Conforme citado anteriormente, a AIDS contribuiu para o reaparecimento de casos de TB, especialmente como infecção oportunista. No ano de 2010 foram notificados casos de co-infecção HIV/TB na América do norte, norte europeu e Oceania como mostra a figura 3.

Figura 3. Número de casos notificados de TB em pacientes portadores do HIV no mundo no ano de 2010 (WHO, 2011).

De acordo com a Organização Mundial da Saúde, nos anos de 2007 e 2008, registraram-se aproximadamente 27.000 casos de MDR-TB no mundo, destes 832 foram nas Américas, o que caracteriza baixa incidência de cepas resistentes nesta região (WHO, 2009b). A figura 4 apresenta a comparação entre os anos de 2005 a 2009 relacionando o número de casos de TB e o número de MDR-TB no Brasil, que é considerado um país que apresenta baixa incidência de cepas multi-resistentes.

Figura 4. Número de casos notificados de TB e MDR-TB no Brasil, entre os anos de 2005 a 2009 (WHO, 2011)

De acordo com dados da OMS (1997, 2002a, 2002d, 2009a), mais de cinquenta milhões de pessoas podem estar infectadas por cepas resistentes. O esquema terapêutico mais longo para TB contribui para a baixa aderência do paciente ao tratamento. Além disso, também favorece o aparecimento de resistência, a prescrição de fármacos inadequados (fármaco isolado ou em combinação) e o sistema falho de distribuição de medicamentos. No caso de infecção por HIV, em que o sistema de defesa do paciente já está comprometido, a resistência pode ser considerada consequência natural (WHO, 2002a). Embora os esquemas terapêuticos preconizados para a TB sejam amplamente aceitos, esforços

devem ser realizados para adaptá-los às diferentes situações socioeconômicas e às várias apresentações da doença.

A estratégia mundial para o controle da TB sugerida pela OMS é o tratamento diretamente supervisionado ou monitorado, denominada DOTS (***D**irectly **O**bserved **T**reatment **S**hort-Course*). Esta estratégia envolve a intervenção governamental e visa aumentar a probabilidade de cura dos pacientes, em função da garantia do tratamento assistido, além de diminuir o aparecimento de resistência ao tratamento combinado, já que o paciente recebe a medicação e é acompanhado durante todo o tratamento por profissionais dos serviços de saúde. No Brasil, o monitoramento do tratamento é realizado em três visitas semanais, durante os dois primeiros meses, e uma visita semanal nos últimos quatro meses de tratamento (MINISTÉRIO DA SAÚDE, 2010).

Considerando as diretrizes da estratégia de DOTS, no País, tem-se o Programa Nacional de Controle da Tuberculose (PNCT), cujo propósito fundamental é promover o controle e diminuir a transmissão da doença. Além disso, este programa tem como meta identificar pelo menos 70% de casos estimados anualmente e promover o tratamento e a cura em, no mínimo, 85% dos casos (MINISTÉRIO DA SAÚDE, 2010).

Os fármacos do tratamento de TB são divididos em primeira (estreptomicina, isoniazida, rifampicina, etambutol e pirazinamida) e segunda escolha (ácido p-aminosalicílico, cicloserina, amicacina, etionamida e capreomicina), de acordo com a eficácia e a toxicidade dos fármacos. O PNCT adota diferentes esquemas de tratamento de acordo com a faixa etária dos pacientes e com a fase da doença. Nos quadros 1 e 2 estão apresentados o esquema básico para o tratamento de TB

pulmonar e extrapulmonar em adultos e adolescentes e o esquema de tratamento em caso de recidivas.

Quadro 1. Esquema de tratamento I para TB pulmonar e extrapulmonar em adultos e adolescentes (PNCT, 2010)

	Duração (meses)	Fármaco	Dose (mg/kg/dia)			
			até 20 kg	21-35 kg	36-45 kg	> de 45 kg
1ª F A S E	2	rifampicina	10	300	450	600
		isoniazida	10	300	300	400
		pirazinamida	35	1000	1500	2000
2ª F A S E	4	rifampicina	10	300	450	600
		isoniazida	10	200	300	400

Quadro 2. Retratamento em recidivantes e retorno após abandono do Esquema I (PNCT, 2010)

	Duração (meses)	Fármaco	Dose (mg/kg/dia)			
			até 20 kg	21-35 kg	36-45 kg	> de 45 kg
1ª F A S E	2	rifampicina	10	300	450	600
		isoniazida	10	200	300	400
		etambutol	25	600	800	1200
		pirazinamida	35	1000	1500	2000
2ª F A S E	4	rifampicina	10	300	450	600
		isoniazida	10	200	300	400
		etambutol	25	600	800	1200

Diante do reaparecimento da TB se faz necessária a pesquisa por novos agentes antimicobacterianos, uma vez que os organismos infectantes tornam-se resistentes aos fármacos convencionalmente disponíveis e alguns destes fármacos apresentam efeitos colaterais graves. Além disso, o tratamento convencional das micobacterioses em pacientes imunodeprimidos é insatisfatório (SENSI, GRASSI, 1996).

O sequenciamento genético de diversos agentes infecciosos permite o acesso a informações relevantes sobre os processos metabólicos envolvidos na doença e disponibilizam novos alvos para o desenvolvimento de novos quimioterápicos

potenciais. Muitos processos metabólicos específicos da micobactéria ocorrem durante a biossíntese dos componentes da parede celular e novos alvos interessantes têm emergido para o planejamento racional de novos agentes antimicobacterianos (BARRY, 1997). Os ácidos micólicos, componentes principais da parede celular do *M. tuberculosis*, são exemplos destes alvos e o estudo de sua estrutura química em conjunto com seu processo biossintético constitui ferramenta importante para o planejamento de novos agentes potenciais contra tuberculose (DUBNAU, 2000; GLICKMAN, 2000; GLICKMAN, 2001; PASQUALOTO, FERREIRA, 2001).

1.2. Planejamento racional de novos fármacos e modelagem molecular

A busca por novos fármacos envolve tanto a pesquisa por entidades químicas novas como estudos sobre a estrutura química de fármacos já existentes, com o propósito de aumentar a potência, diminuir a toxicidade, melhorar o perfil farmacocinético, ou até de utilizar a estrutura química de um fármaco como molde para um novo composto químico com atividade biológica específica. Para isso, procura-se sempre programar as etapas a serem realizadas para aumentar a probabilidade de êxito e diminuir gastos (KOROLKOVAS, 1982).

Apesar dos avanços tecnológicos, a informação principal que permite o planejamento racional de novas moléculas é o conhecimento da etiologia de determinada doença, que se deseja combater, ou, pelo menos, dos processos bioquímicos que são alterados e comprometem a resposta biológica, ao nível molecular (COHEN, 1996; WERMUTH, 2008).

De acordo com os glossários de termos empregados em Química Medicinal (WERMUTH *et al.*, 1998) e no planejamento de fármacos (SANT'ANNA, 2002),

recomendações da IUPAC para 1997, modelagem molecular é a investigação das estruturas e das propriedades moleculares utilizando a química computacional e as técnicas de visualização gráfica a fim de fornecer uma representação tridimensional do sistema em estudo, sob um dado conjunto de circunstâncias ou critérios definidos pelo pesquisador ou usuário.

O desenvolvimento de projetos envolvendo programas de modelagem molecular e suas aplicações na pesquisa farmacêutica constituem campo de estudo denominado planejamento de fármacos auxiliado por computador (CADD, *Computer-Aided Drug Design*), que tem como expectativa a descoberta de moléculas promissoras (mais potentes e seletivas) considerando propriedades tridimensionais específicas. O CADD pode ser aplicado a qualquer molécula ativa cuja interação com um alvo molecular específico (receptor, enzima, canais de íons, proteínas transportadoras, DNA) seja postulada (COHEN, 1996; OHLSTEIN, RUFFOLO, ELLIOT, 2000; WERMUTH, 2008). Assim, a modelagem molecular utiliza duas estratégias principais do CADD: o planejamento direto ou dependente do receptor e o planejamento indireto ou independente do receptor, observando a disponibilidade ou não da estrutura tridimensional da biomacromolécula alvo (COHEN, 1996; OOMS, 2000; WERMUTH, 2008).

A contribuição do CADD na descoberta de novos protótipos possui valor limitado devido ao fato de que os estudos de modelagem molecular permitiriam somente a identificação de novos ligantes provenientes de sistemas biológicos já conhecidos. Do mesmo modo, métodos de relações quantitativas entre estrutura química e atividade biológica (QSAR, *Quantitative Structure-Activity Relationships*), que evidenciam as propriedades físico-químicas indispensáveis ao efeito biológico,

possibilitariam a otimização somente de conjuntos de compostos já descobertos (WERMUTH, 2008).

O CADD não é, entretanto, uma rota direta para a descoberta de novos fármacos, mas fornece informações mais detalhadas e importantes para alcançar este objetivo (OOMS, 2000). A modelagem molecular como ferramenta para o planejamento de fármacos permite, por exemplo, a redução do número de compostos a serem sintetizados a partir do reconhecimento de propriedades estruturais e eletrônicas que resultariam em compostos pouco ativos ou inativos (WERMUTH, 2008). Sabe-se que no processo de discriminação das moléculas de um fármaco no organismo, a biomacromolécula alvo reconhece a micromolécula por meio de suas propriedades físico-químicas (eletrônica, estereoquímica e lipofílica), que influenciam tanto na distribuição na biofase quanto na interação com a biomacromolécula alvo (TAVARES, FERREIRA, 2006). Desta forma, avanços tecnológicos nas áreas de caracterização estrutural de biomacromoléculas, de ciência computacional e de biologia molecular têm contribuído para que o planejamento de novas moléculas se torne factível (OOMS, 2000).

1.3. *Enoil-acp redutase (ENR) – Alvo importante no planejamento de novos agentes tuberculostáticos*

As bactérias Gram-positivas ácido-resistentes dos gêneros *Mycobacterium*, *Corynebacterium* e *Nocardia*, possuem a parede celular de composição diferenciada. Esta parede celular é composta por camadas de peptidoglicanos ligados covalentemente aos arabinogalactanos, que podem estar esterificados com ácidos graxos α -alquil e/ou β -hidroxiácidos de cadeia longa. Estes ácidos graxos, no gênero *Mycobacterium*, são denominados ácidos micólicos e formam uma parede celular

mais espessa que nas demais bactérias Gram-positivas ácido-resistentes (SAIER, 2009). A composição da membrana e parede celular do *M. tuberculosis* está representada na figura 5.

Figura 5. Representação esquemática de componentes estruturais importantes da parede celular do *M. tuberculosis* (adaptado de SAIER, 2009).

O mecanismo biossintético de ácidos graxos, denominado sistema FAS (*Fatty Acid Synthase*), envolve ciclos repetitivos de condensação, ceto-redução, desidratação e enoil-redução. As enzimas que participam deste processo são consideradas alvos ideais para o planejamento de agentes antibacterianos e antimicobacterianos mais seletivos, devido à diferença na organização molecular entre os sistemas FAS de bactérias/micobactérias e de mamíferos, ou seja, diferença no complexo de enzimas envolvidas no sistema FAS (McCARTHY, HARDIE, 1984; BARRY *et. al.*, 1998; HEATH, WHITE, ROCK, 2001; ASSELINEAU *et. al.*, 2002).

Mamíferos, aves e leveduras produzem ácidos graxos via sistema FAS-I, no qual a atividade enzimática reside numa única cadeia polipeptídica de enzimas multifuncionais compondo o sistema de síntese de ácidos graxos tipo I (BARRY *et. al.*, 1998; MARRAKCHI, LANÉELLE, QUÉMARD, 2000). Em contraste, a maioria das

plantas e bactérias utiliza um sistema FAS-II, em que cada atividade enzimática corresponde a um polipeptídio individual (ROCK, CRONAN, 1996; BARRY *et al.*, 1998; MARRAKCHI, LANÉELLE, QUÉMARD, 2000). Já as micobactérias possuem ambos os sistemas de biossíntese de ácidos graxos, FAS-I e FAS-II (BLOCH, 1977; BARRY *et al.*, 1998; MARRAKCHI, LANÉELLE, QUÉMARD, 2000).

No sistema FAS I, o alongamento inicia com duas unidades consecutivas de carbono, utilizando grupos acetil e malonil ligados a uma coenzima como substratos, até que a cadeia alquílica de ácidos graxos intermediários atinja aproximadamente 16 carbonos (ROZWASRKI *et al.*, 1999). Intermediários de 20 e 26 carbonos também são produzidos. O sistema FAS II das micobactérias tem preferência por substratos iniciais de ácidos graxos com 16 carbonos, que garantem o alongamento mais uniforme da cadeia de ácidos graxos (BLOCH, 1997; BARRY *et al.*, 1998; PASQUALOTO, 2003). Os produtos de cadeia mais longa do sistema FAS-II são os precursores dos ácidos micólicos, principais componentes da parede celular micobacteriana (BRENAN, NIKAIDO, 1995; LEE, BRENNAN, BESRA, 1996; BARRY *et al.*, 1998; MARRAKCHI, LANÉELLE, QUÉMARD, 2000).

Takayama e colaboradores (2005) demonstraram que um único gene (Rv2524c) é responsável pela ativação do sistema FAS I multifuncional no *Mycobacterium* e o gene ativo leva à expressão da proteína contendo todas as funções necessárias para a síntese de novos ácidos graxos.

O sistema FAS II envolve ciclos repetitivos de condensação, ceto-redução, desidratação e enoil-redução (HEATH, WHITE, ROCK, 2001, 2002; MARRAKCHI, ZHANG, ROCK, 2002). Na última etapa do sistema FAS II atua a enzima ENR, alvo deste projeto.

A bactéria *Escherichia coli* apresenta o sistema FAS-II melhor caracterizado e inclui as seguintes enzimas-acp (***acyl carrier protein***, proteínas acil-carregadoras): β -cetoacil-acp sintetases (FabB, FabF e FabH), β -cetoacil-acp redutase (FabG), β -hidroxi-acil-acp desidrases ou desidratases (FabA e FabZ) e enoil-acp redutase (FabI, inicialmente denominada EnvM) (ROZWARSKI *et. al.*, 1999).

A enzima ENR catalisa a reação principal, que regula o ciclo de alongamento de ácidos graxos (STEWART *et. al.*, 1999). A reação consiste na redução estereoespecífica, NADH-dependente, de ácidos graxos α,β -insaturados ligados à proteína acil-carregadora (ROZWARSKI *et. al.*, 1999; STEWART *et. al.*, 1999).

Os ácidos micólicos são ácidos graxos β -hidroxilados de cadeia lateral α -alquílica longa, caracterizados por uma porção muito hidrofóbica entre os carbonos C₅₄ a C₆₃. Encontram-se ligados covalentemente a moléculas de arabinogalactanos e peptideoglicanos da parede celular. Estes componentes e um glicolípido livre, denominado dimicolato de trealose (TDM, *trehalose dimycolate*), conferem à micobactéria uma parede celular complexa com espessa camada de lipídeos na porção externa, que determina a patogenicidade do microorganismo (YUAN, BARRY, 1998).

Acredita-se que a presença destes ácidos graxos de cadeia tão longa é essencial à sobrevivência do *Mycobacterium* em ambiente químico agressivo, como o encontrado nos fagolisossomos no interior dos macrófagos (JANIN, 2007).

Três classes de micolatos estão presentes no *M. tuberculosis*: o ácido α -micólico, que é o mais abundante (cerca de 70%); o ácido metóxi-micólico e ácido ceto-micólico (figura 6). Alguns possuem um anel ciclopropano na cadeia longa que contribui para a manutenção da integridade estrutural da parede celular do bacilo. A presença de anel ciclopropano ou de insaturações na cadeia principal permite

configurações *cis* ou *trans* dos micolatos (TAKAYAMA, WANG, BESRA, 2005; PASQUALOTO, 2003; BARRY *et. al.*, 1998).

Figura 6. Classes de micolatos encontrados na parede celular do *M. tuberculosis*. (Adaptado de PASQUALOTO, 2003; TAKAYAMA *et. al.*, 2005).

A ENR do *M. tuberculosis*, InhA, apresenta sequência total de aminoácidos com 28% de homologia em relação à ENR da *E. coli* (FabI) (ROZWASRKI *et. al.*, 1998). Na região do sítio ativo, a semelhança estrutural entre InhA/FabI é de 40% (BALDOCK *et. al.*, 1996), o que indica uma conservação importante dos principais resíduos de aminoácido envolvidos no reconhecimento molecular.

Pelo menos três inibidores de classes diferentes da ENR complexados com seus respectivos alvos enzimáticos foram caracterizados: isoniazida (INH), que inibe a InhA, ENR do *M. tuberculosis* (ROZWARSKI *et. al.*, 1998); diazaborinas (BALDOCK *et. al.*, 1996; BALDOCK *et. al.*, 1998) e triclosana (McMURRY, OETHINGER, LEVY, 1998; STEWART *et. al.*, 1999), compostos que inibem a FabI, ENR da *E. coli*. Estudos relataram que a triclosana também inibe a InhA (PARIKH, *et. al.*, 2000). As estruturas destes inibidores estão apresentadas na figura 7.

Figura 7. Inibidores de ENRs.

A isoniazida (hidrazida do ácido isonicotínico) é utilizada como fármaco de primeira escolha nos esquemas terapêuticos de TB. Age como um bioprecursor que necessita de ativação pela enzima KatG. A forma ativa, que pode ser um ânion ou radical, liga-se covalentemente ao átomo C₄ da nicotinamida do cofator (NAD) complexado à enzima InhA, resultando na formação do aduto acilpiridina:NAD (ROZWARSKI *et al.*, 1998; ROZWARSKI *et al.*, 1999), que funciona como inibidor irreversível. Na figura 8 estão representadas as interações entre a forma ativa da isoniazida ligada covalentemente ao cofator NAD (aduto acilpiridina:NAD) e os resíduos de aminoácidos (aa) do sítio ativo da biomacromolécula alvo (InhA). Mecanismos de resistência descritos à isoniazida envolvem a perda de atividade da enzima catalase e deleções no gene catalase peroxidase (KatG) (BANERJEE, *et al.*, 1994).

Figura 8. Esquema de interação da isoniazida com resíduos de aa do sítio ativo da InhA (arquivo LIGPLOT). A ligação covalente do aduto acilpiridina:NAD está indicada com a seta laranja [Fonte: Banco de dados PDBsum (LASKOWSKI, 2007), código 1zid (ROZWARSKI *et al.*, 1998)].

As diazaborinas (1,2-diidro-1-hidroxi-2-(organossulfonil)areno[d][1,2,3] diazaborinas; fig. 7) constituem uma família experimental de compostos que têm propriedades antibacterianas contra uma variedade de bactérias Gram-negativas e apresentam como característica estrutural importante um anel heterocíclico 1,2-diazina contendo boro como terceiro heteroátomo. O átomo de boro presente nas diazaborinas participa de ligação covalente com o grupo 2-hidroxila da ribose

próxima ao anel nicotinamida do cofator (BALDOCK *et. al.*, 1996; BALDOCK *et. al.*, 1998). Na figura 9 estão representadas interações entre um derivado benzodiazaborina e os resíduos de aa do sítio ativo da FabI, ENR da *E. coli*. Observa-se a ligação covalente entre a hidroxila da ribose próxima ao anel nicotinamida do cofator NAD e o átomo de boro do derivado benzodiazaborina.

Figura 9. Esquema de interação de um derivado benzodiazaborina com resíduos de aa do sítio ativo da FabI (arquivo LIGPLOT). A ligação covalente entre o oxigênio da hidroxila da ribose (em vermelho), próxima à nicotinamida do NAD, e o átomo de boro da benzodiazaborina (em verde) está representada [Fonte: Banco de dados PDBsum (LASKOWSKI, 2007), código 1dfg (BALDOCK *et. al.*, 1996)].

A triclosana (éter 2,4,4'-triclouro-2'-hidroxidifenila; fig. 5), que possui atividade antimicrobiana de amplo espectro, é utilizada como aditivo antibacteriano em produtos de higiene pessoal e em antissépticos hospitalares devido ao seu perfil de

segurança (BHARGAVA, LEONARD, 1996). Foi descrita com inibidor da ENR da *E. coli*, do *M. tuberculosis*, de cepas resistentes de *Staphylococcus aureus*, *Haemophilus influenzae* e *Plasmodium falciparum* (SIVARAMAN *et. al.*, 2004). Diversos análogos da triclosana foram estudados e apresentaram atividade inibitória sobre a FabI e a InhA (PEROZZO *et. al.*, 2002). Recentemente, Morde e colaboradores (2009) verificaram a inibição de análogos da triclosana sobre a ENR do *P. falciparum*, PfENR.

A comparação de estruturas cristalografadas dos inibidores, diazoborina e triclosana, em complexo com a FabI revelou que ambos interagem somente com a porção ribose do cofator NAD. Entretanto, a triclosana, ao invés de ligação covalente, forma ligações hidrogênio com o grupo 2'-hidroxila da ribose próxima ao anel nicotinamida do cofator (STEWART *et. al.*, 1999), conforme representado na figura 10.

Figura 10. Representação das interações entre a triclosana e resíduos de aa do sítio ativo da FabI [Fonte: Banco de dados PDBsum (LASKOWSKI, 2007), código 1qsg (STEWART, *et. al.*, 1999).

O advento da robótica e a miniaturização dos testes *in vitro* em células modificadas contribuíram para o desenvolvimento dos ensaios biológicos automatizados em larga escala (HTS, *High Throughput Screening*) (MACARRON, 2006; ANDRADE *et. al.*, 2008). He e colaboradores (2006) descobriram um conjunto de pirrolidina carboxamidas como nova classe de potentes inibidores da InhA com a aplicação de HTS. A potência do composto líder ou protótipo foi incrementada em mais de 160 vezes e o valor de afinidade quantitativa, expresso em CI_{50} , chegou a $0,14 \mu M$. A estrutura cristalografada da InhA com cinco inibidores está disponível no banco de dados de proteínas PDB (*Protein Data Bank*) (BERMAN *et al.*, 2000). A estrutura química geral das pirrolidina carboxamidas está apresentada na figura 11.

Figura 11. Estrutura química geral das pirrolidina carboxamidas.

Observando-se as cinco estruturas cristalografadas da classe pirrolidina carboxamidas, verifica-se que o resíduo de aa Tyr156 do sítio de interação também está envolvido na interação da izoniazida, triclosana e diazaborinas. Porém, a classe pirrolidina carboxamidas interage de forma semelhante à triclosana, por meio de ligação de hidrogênio na mesma porção ribose do cofator (NAD), conforme figura 12.

Figura 12. Representação das interações entre análogo pirrolidina carboxamida e resíduos de aa do sítio ativo da InhA [Fonte: Banco de dados PDBsum (LASKOWSKI, 2007), código 2h7i (HE *et. al.*, 2006)].

Kumar e Siddiqi (2008) selecionaram o subconjunto das pirrolidina carboxamidas descritas por He e colaboradores (2006) e aplicaram a análise comparativa de campos moleculares (CoMFA, *Comparative Molecular Field Analysis*) para auxiliar na explicação do modo de interação desta classe de inibidores. Os resultados obtidos por Kumar e Siddiqi (2008) podem ser utilizados para direcionar a aplicação de outras metodologias do CADD no planejamento de novos protótipos anti-TB potenciais.

O conhecimento cada vez mais aprofundado acerca da biossíntese de ácidos graxos das bactérias/micobactérias e a caracterização bioquímica e estrutural das enzimas envolvidas no processo de síntese de ácidos graxos têm contribuído para a

identificação de alvos bioquímicos importantes e promissores para o processo de planejamento de novos protótipos potenciais (BROWN, WRIGHT, 2005). A disponibilidade destas informações permite, por exemplo, o desenvolvimento de estudos de CADD visando a obter novos agentes antibacterianos e antimicobacterianos eficientes contra cepas sensíveis e resistentes.

JUSTIFICATIVA E OBJETIVO

2. JUSTIFICATIVA E OBJETIVO

Considerando a incidência de TB mundial, a taxa de mortalidade, o quadro desta doença no Brasil, principalmente na região sudeste do País, além dos índices elevados de resistência do microrganismo aos fármacos convencionalmente utilizados, há necessidade do desenvolvimento de fármacos específicos e de um esquema terapêutico racional.

A aplicação de métodos do CADD e a modelagem molecular contribuem de forma relevante ao processo de planejamento de novos fármacos, reduzindo etapas sintéticas, quantidade de ensaios e de animais de experimentação, além de auxiliar no desenvolvimento de moléculas inéditas, mais seletivas que o fármaco protótipo.

As enzimas que participam do mecanismo biossintético de ácidos graxos (sistema FAS) são consideradas alvos ideais para o planejamento de novos agentes antibacterianos/antimicobacterianos devido à diferença na organização molecular entre os sistemas FAS de bactérias/micobactérias e mamíferos (McCARTHY, HARDIE, 1984; BARRY *et al.*, 1998; HEATH, WHITE, ROCK, 2001; ASSELINEAU *et al.*, 2002).

Desta forma, o presente projeto tem como objetivo a aplicação das metodologias do CADD em uma classe de inibidores da InhA, análogos de pirrolidina carboxamidas, com intuito de mapear o farmacóforo, investigar a orientação destes ligantes no sítio ativo, os tipos de interações que se estabelecem com os resíduos de aminoácidos no sítio de interação, além de verificar quais modificações moleculares seriam interessantes para gerar pontos adicionais de interação no sítio ativo da enzima. Inicialmente a investigação quantitativa que relaciona as propriedades moleculares derivadas da estrutura dos análogos pirrolodina

carboxamidas e a inibição da biomacromolécula alvo, InhA, foi explorada com aplicação de abordagem multivariada. Tais informações permitirão o planejamento e proposição de novos protótipos potencialmente tuberculostáticos, mais potentes e seletivos.

METODOLOGIA

3. METODOLOGIA

O presente projeto foi realizado sob a responsabilidade da Dr^a. Kerly Fernanda Mesquita Pasqualoto, pesquisadora do Laboratório de Bioquímica e Biofísica do Instituto Butantan, e desenvolvido no setor de Modelagem Molecular do Laboratório de Planejamento e Síntese de Quimioterápicos Potencialmente Ativos em Endemias Tropicais - LAPEN, FCF-USP.

A realização deste estudo contou com uma licença colaborativa de programas de modelagem molecular e de QSAR (MOLSIM 3.2, 4D-QSAR 3.0, WOLF 5.5) desenvolvidos no The Chem21 Group, Inc. sob a coordenação do Prof. Dr. Anton J. Hopfinger, Universidade do Novo México, EUA.

3.1. *Dados biológicos*

Selecionaram-se 46 inibidores da InhA, pertencentes à classe das pirrolidina carboxamidas, ensaiados sob o mesmo protocolo farmacológico. A atividade biológica foi avaliada como a concentração de composto que resulta em 50% de inibição na atividade enzimática da InhA à temperatura de 298K (25 °C) (He *et. al.*, 2006). Nos quadros 3 e 4 estão listadas as 46 estruturas do conjunto de pirrolidina carboxamidas investigado e seus respectivos valores de atividade biológica (CI_{50} , μM), além dos valores de atividade linearizados (pCI_{50} ou $\log 1/CI_{50}$ ou $-\log CI_{50}$).

Quadro 3. Conjunto de 26 estruturas da classe química pirrolidina carboxamidas mono ou dissustituídas e respectivos valores de atividade biológica expressos em CI_{50} ($\mu\text{mol/L}$) e pCI_{50}

Composto	R ₁	R ₂	CI ₅₀	pCI ₅₀
TBP 1	H	H	10,66 ± 0,51	4,97
TBP 2	H	2-COOMe	34,88 ± 2,06	4,46
TBP 3	H	3-Br	0,89 ± 0,05	6,05
TBP 4	H	4-Br	28,02 ± 4,29	4,55
TBP 5	H	3-Cl	1,35 ± 0,05	5,87
TBP 6	H	4-I	14,50 ± 0,79	4,97
TBP 7	H	3-Me	16,79 ± 0,54	4,46
TBP 8	H	3-CF ₃	3,51 ± 0,09	6,05
TBP 9	H	3-NO ₂	10,59 ± 0,48	4,98
TBP 10	H	3-CH-(CH ₃) ₂	5,55 ± 0,21	4,78
TBP 11	H	4-OCOMe	73,58 ± 9,97	5,45
TBP 12	2-Cl	4-Cl	56,02 ± 10,23	4,98
TBP 13	2-Cl	5-Cl	56,50 ± 11,09	5,26
TBP 14	2-Me	5-Cl	0,97 ± 0,03	4,13
TBP 15	2-Me	4-Br	37,41 ± 1,76	4,25
TBP 16	2-Me	5-Me	10,05 ± 0,33	4,25
TBP 17	3-Me	5-Me	3,14 ± 0,12	6,01
TBP 18	2-Me	3-Cl	23,12 ± 1,00	4,43
TBP 19	2-Me	4-NO ₂	31,37 ± 1,45	5,00
TBP 20	3-F	5-F	1,49 ± 0,05	5,50
TBP 21	3-Cl	5-Cl	0,39 ± 0,01	4,64
TBP 22	3-Br	5-CF ₃	0,85 ± 0,05	4,50
TBP 23	3-OMe	5-CF ₃	1,30 ± 0,04	5,83
TBP 24	3-CF ₃	5-CF ₃	3,67 ± 0,17	6,41
TBP 25	2-OMe	5-Cl	1,60 ± 0,06	6,07
TBP 26	3-Cl	4-F	14,83 ± 0,98	5,89

Quadro 4. Conjunto de 20 estruturas da classe química pirrolidina carboxamidas com substituintes volumosos e respectivos valores de atividade biológica expressos em CI_{50} ($\mu\text{mol/L}$) e pCI_{50}

Composto	Estrutura Química	CI_{50}	pCI_{50}
TBP 27		$3,39 \pm 0,10$	5,47
TBP 28		$0,75 \pm 0,04$	6,13
TBP 29		$0,41 \pm 0,01$	6,39
TBP 30		$0,39 \pm 0,01$	6,41
TBP 31		$5,18 \pm 0,34$	5,23
TBP 32		$6,41 \pm 0,12$	5,19
TBP 33		$1,39 \pm 0,02$	5,86

Quadro 4 (cont.). Conjunto de 20 estruturas da classe química pirrolidina carboxamidas com substituintes volumosos e respectivos valores de atividade biológica expressos em CI_{50} ($\mu\text{mol/L}$) e pCI_{50}

Composto	Estrutura Química	CI_{50}	pCI_{50}
TBP 34		$2,57 \pm 0,17$	5,59
TBP 35		$5,51 \pm 0,22$	5,26
TBP 36		$4,47 \pm 0,28$	5,35
TBP 37		$0,85 \pm 0,05$	6,07
TBP 38		$0,46 \pm 0,01$	6,34
TBP 39		$0,14 \pm 0,01$	6,85

Quadro 4 (cont.). Conjunto de 20 estruturas da classe química pirrolidina carboxamidas com substituintes volumosos e respectivos valores de atividade biológica expressos em Cl_{50} ($\mu\text{mol/L}$) e pCl_{50}

Composto	Estrutura Química	Cl_{50}	pCl_{50}
TBP 40		$0,62 \pm 0,05$	6,21
TBP 41		$0,36 \pm 0,03$	6,44
TBP 42		$0,32 \pm 0,02$	6,50
TBP 43		$1,29 \pm 0,10$	5,89
TBP 44		$3,94 \pm 0,34$	5,40
TBP 45		$13,55 \pm 0,85$	4,87
TBP 46		$29,23 \pm 2,17$	4,53

A faixa de dispersão dos valores linearizados de atividade biológica variou de 4,13 a 6,85 unidades logarítmicas, conforme o gráfico apresentado na figura 13. De acordo com esta dispersão, neste estudo, consideraram-se como compostos menos ativos aqueles distribuídos entre 4 e 5 unidades logarítmicas, como moderadamente ativos aqueles entre 5 e 6 e como mais ativos aqueles compostos distribuídos acima de 6 unidades logarítmicas.

Figura 13. Dispersão dos dados de atividade biológica dos 46 análogos da classe pirrolidina carboxamidas. Faixa de dispersão delimitada pelo tracejado vermelho.

Outro conjunto de inibidores da InhA, pertencentes à classe arilamidas e composto de 24 estruturas, foi selecionado para verificar a capacidade de previsão do melhor modelo QSAR construído e validado com os conjuntos de treinamento e de avaliação, respectivamente, dos análogos pirrolidina carboxamidas. Embora as arilamidas não compartilhem o mesmo farmacóforo das pirrolidina carboxamidas, parecem apresentar o mesmo mecanismo de interação na InhA (He *et.al*, 2006; He *et. al.*, 2007). Os valores de pCl_{50} , obtidos pelo mesmo protocolo farmacológico, variaram entre 4,41 a 7,05 e a estrutura tridimensional de um dos análogos arilamida em complexo com a enzima InhA está disponível no PDB (código de entrada 2NSD e resolução 1,90Å). A metodologia de construção dos modelos moleculares tridimensionais (3D), minimização de energia e seleção dos confôrmeros para o

cálculo das propriedades moleculares aplicada à classe pirrolidina carboxamida também foi aplicada aos compostos da classe arilamida.

3.2. Construção e minimização de energia dos modelos moleculares tridimensionais (3D)

Há um carbono assimétrico no núcleo comum a todos os análogos pirrolidina carboxamida investigados, que permite configuração *R* ou *S* para cada inibidor. He e colaboradores (2006) resolveram a mistura racêmica por cromatografia líquida de alta eficiência (CLAE) e realizaram o ensaio biológico para cada um dos enantiômeros separadamente e para a mistura racêmica. Ambos os enantiômeros apresentaram atividade inibitória, mas a contribuição da configuração *S* foi maior para este efeito.

Estruturas de cinco complexos InhA-NAD-inibidor, obtidas por difração de raios-X, foram depositadas no PDB (códigos de entrada 2h7i, 2h7l, 2h7m, 2h7n e 2h7p e resolução 1,62Å a 1,90Å) e apresentam os inibidores na configuração *S* no sítio de interação da enzima (He *et al.*, 2006). Estes inibidores fazem parte do conjunto de 46 ligantes investigados neste estudo. Então, os dados experimentais dos complexos InhA-NAD-inibidor supracitados foram considerados no processo de construção dos modelos moleculares 3D dos inibidores ou ligantes isolados, ou seja, as estruturas 3D dos cinco inibidores na enzima-alvo foram utilizadas como referência contemplando a configuração *S* do ligante.

Os processos de reconstrução das estruturas de referência e de construção e otimização de geometria dos modelos moleculares 3D dos compostos do conjunto investigado, sem quaisquer restrições, foram realizados em campo de força empírico MM+ utilizando o módulo “Build” do programa HyperChem 7.51 (Hercube Inc.,

2002). As cargas atômicas parciais foram calculadas com o método semi-empírico PM3 (STEWART, 1989) no programa HyperChem 7.51 (Hypercube Inc., 2002), a fim de melhor parametrizar alguns átomos presentes nos ligantes do conjunto investigado.

O programa MOLSIM 3.2 (DOHERTY, 1994) foi utilizado para minimizar a energia dos modelos moleculares 3D dos compostos antes do desenvolvimento das simulações de dinâmica molecular (DM). Empregaram-se os métodos de minimização de Declive Máximo (*Steepest Descent*) com 500 iterações e o de Gradientes Conjugados (*Conjugate Gradient*) com 20.000 iterações. O critério de convergência de energia adotado no processo de minimização foi de 0,01 kcal/mol. Constante dielétrica de 3,5 foi utilizada porque simula o ambiente dos sistemas biológicos, considerando que a constante dielétrica no interior de enzimas varia de 2 a 5 (ADMET/MI-QSAR, The Chem21 Group, Inc., 2001). A temperatura inicial e de referência foi de 298 K (25 °C), que corresponde à temperatura do ensaio biológico descrito por He e colaboradores (2006).

3.3. Simulações de dinâmica molecular (DM)

As simulações de DM permitem investigar a evolução temporal das configurações de cada molécula de interesse obtendo-se o perfil de amostragem conformacional (PAC). A utilização da DM no estudo comparativo entre os inibidores de InhA possibilita explorar as diversas trajetórias espaciais dos grupos substituintes no decorrer do tempo, na busca pela conformação mais frequente próxima da bioativa (MARTÍNEZ, BORIN, SKAF, 2007).

Desenvolveram-se simulações de DM de 200 ps (200.000 passos, cada passo de 1 fs ou 0,001 ps) a 298 K (temperatura do ensaio biológico) para se obter o PAC de cada composto do conjunto investigado. Arquivos de trajetória foram salvos a cada 20 passos resultando em 10.000 confôrmeros de cada ligante. Massas fictícias (5.000 u.m.a.) foram atribuídas aos átomos do núcleo comum do conjunto investigado, conforme indicado na figura 14, para restringir mudanças mais bruscas de posições atômicas em relação aos ligantes experimentais que estavam no sítio ativo da biomacromolécula alvo, a enzima InhA.

A partir da região de equilíbrio da simulação de DM foi selecionado um confôrmero de energia mais baixa, posteriormente minimizado, que mantivesse a integridade estrutural em relação ao inibidor experimental do complexo 2h7i (inibidor não-substituído). O critério adotado para a avaliação da integridade estrutural foi o valor de desvio quadrático médio de posições atômicas (RMSD, *root-mean square deviation*), que não deveria ultrapassar 0,01 Å em comparação ao dado experimental. A sobreposição da estrutura do ligante experimental e dos confôrmeros selecionados das simulações de DM foi realizada no programa HyperChem 7.51 (Hypercube Inc., 2002).

Figura 14. Estrutura da unidade fundamental do conjunto de pirrolidina carboxamidas. Massas fictícias foram atribuídas durante as simulações de DM aos átomos do núcleo comum destacados em roxo.

A energia potencial total (E_{tot}) do confôrmero selecionado do PAC de cada composto corresponde à soma das seguintes contribuições de energia: de

deformação axial ou estiramento (E_{est}), de deformação angular (E_{ang}), de deformação torsional (E_{tors}), de interações do tipo 1-4 ou Lennard-Jones ($E_{1,4}$), de carga ou eletrostática (E_{el}), de van der Waals (E_{vdW}), de solvatação (E_{solv}) e de ligações de hidrogênio (E_{Hb}). As contribuições de energia E_{solv} e E_{Hb} foram calculadas somente para o confômero de energia mais baixa selecionado do PAC de cada composto. O modelo de camada de hidratação implícita, descrito por Hopfinger (1973), foi utilizado para o cálculo da contribuição de energia E_{solv} dos ligantes do conjunto investigado.

3.4. Análise QSAR multivariada

As pesquisas em QSAR têm como objetivo a construção de modelos matemáticos que relacionem a estrutura química (propriedades físico-químicas, estruturais e conformacionais) e a resposta biológica (variável dependente). Tais relações possibilitam a compreensão das forças de interação envolvidas no mecanismo de ação de fármacos e fundamentam o desenvolvimento de novos compostos exibindo propriedades biológicas desejáveis (FERREIRA, 2002a; FERREIRA, 2002b).

Em QSAR, são duas as principais abordagens utilizadas na construção de modelos matemáticos (FERREIRA, 2002b). A primeira refere-se ao método de regressão linear múltipla (MLR, *Multiple Linear Regression*), que se adapta ao tratamento de dados através do método clássico descrito por Hansch-Fujita (GAUDIO, 1996; HANSCH, LEO, 1995; KUBINYI, 1993; MARTIN, 1978). A segunda abordagem utiliza métodos quimiométricos baseados na análise de componentes

principais (*Principal Components Analysis*, PCA) e consiste em análise multivariada (WOLD, ESBENSEN, GELADI, 1987).

O método de regressão MLR, apesar de sua simplicidade e facilidade de interpretação dos resultados, apresenta a desvantagem de ser extremamente sensível à presença de colinearidade entre as variáveis independentes (descritores). Portanto, a aplicação do método de MLR em um conjunto de variáveis altamente correlacionadas pode gerar coeficientes de regressão instáveis e sem significado (FERREIRA, 2002a; FERREIRA, 2002b). Uma das maneiras de contornar este problema é utilizar previamente um método de seleção de variáveis independentes para otimizar a construção dos modelos QSAR. Assim, a probabilidade de que descritores com a mesma informação apareçam no mesmo modelo QSAR diminui (GAUDIO, ZANDONADE, 2001).

Por outro lado, métodos de regressão que utilizam a análise de componentes principais como fundamento não apresentam a limitação relacionada à existência de colinearidade. Um exemplo de método de regressão que constitui a segunda abordagem de construção de modelos QSAR é a regressão por mínimos quadrados parciais (*Partial Least Square*, PLS) (FERREIRA, 2002a; FERREIRA, 2002b).

Na análise QSAR multivariada por PLS o número de descritores ou variáveis independentes é maior ou igual ao número de compostos do conjunto investigado. O método de regressão PLS agrupa os descritores que apresentam informações semelhantes em uma única variável latente (VL) ou componente PLS e utiliza as VLS disponíveis e não intercorrelacionadas para a construção do modelo QSAR. Então, além de evitar a colinearidade o método de regressão PLS funcionaria também como ferramenta prévia de seleção de variáveis à construção de modelos QSAR (FERREIRA, 2002a).

Análise QSAR multivariada preliminar foi desenvolvida considerando o conjunto total de 46 inibidores da InhA, porém os modelos finais foram construídos com os conjuntos de treinamento de cada tabela de dados e a capacidade de previsão externa foi verificada com o conjunto de avaliação (validação externa).

O confômero selecionado para cada ligante na etapa de simulação de DM foi empregado para gerar os descritores ou variáveis independentes. Os dados biológicos já mencionados no item 3.1 (valores de pCl_{50} - Quadros 3 e 4) constituíram as variáveis dependentes da análise QSAR.

3.4.1. Cálculo de variáveis independentes ou descritores

Descritores ou variáveis independentes de diversas naturezas (eletrônica, estereoquímica, lipofílica, estrutural, termodinâmica) foram calculados considerando os confômeros selecionados dos PAC gerados nas simulações de DM.

Os descritores termodinâmicos foram gerados na etapa de simulação de DM e correspondem às contribuições de energia consideradas no cálculo da energia potencial total (E_{tot}). Então, os descritores termodinâmicos incluíram os valores de energia de E_{tot} , E_{est} , E_{ang} , E_{tors} , E_{el} , $E_{1,4}$, E_{vdW} , E_{Hb} e E_{solv} para o confômero selecionado do PAC de cada ligante do conjunto investigado.

As cargas atômicas parciais de potencial eletrostático (algoritmo CHELPG) (BRENEMAN, WIBERG, 1990) foram calculadas para os confômeros selecionados dos PAC resultantes das simulações de DM utilizando o programa Gaussian 03W (Gaussian Inc., 2003). Empregaram-se os métodos *ab initio* em nível de teoria Hartree-Fock (HF) (FOCK, 1930) e funcional híbrido B3LYP (BECKE; LEE, YANG, PARR) (1988), com os conjuntos de bases 3-21G* (inclui o átomo de iodo, ligante TBP6) e 6-31G* (não considera o átomo do iodo, exclui o ligante TBP6). O funcional

híbrido B3LYP (BECKE, 1988; LEE, YANG, PARR, 1988) combina os níveis de teoria HF e do funcional de densidade (DFT, *Density Functional Theory*) (LEACH, 2001). As cargas de potencial eletrostático (CHELPG) nos átomos do farmacóforo das pirrolidina carboxamidas (figura 14) foram incluídas no conjunto de variáveis independentes consideradas na análise QSAR multivariada. Quatro tabelas diferentes foram geradas por conta dos diferentes métodos utilizados no cálculo dos descritores eletrônicos.

Além das cargas de potencial eletrostático, calcularam-se os seguintes descritores eletrônicos nas condições já mencionadas: energias de orbitais moleculares de fronteira, E_{HOMO} e E_{LUMO} (HOMO, *highest occupied molecular orbital*, orbital molecular ocupado de maior energia, e LUMO, *lowest unoccupied molecular orbital*, orbital molecular desocupado de menor energia); momento de dipolo total (μ_{tot}) e momento de dipolo nos eixos x (μ_x), y (μ_y) e z (μ_z).

Descritores que expressam lipofilicidade/hidrossolubilidade, conectividade (topológicos) e características estereoquímicas foram calculados com os programas HyperChem 7.51 (Hypercube Inc., 2002); Marvin Beans 5.2.1 (ChemAxon Ltd., 1999-2010) e Gaussian 03W (Gaussian, Inc., 2004).

Em relação às características estereoquímicas, os ângulos de torção (diedros) considerados na análise QSAR estão representados na figura 15 e correspondem aos seguintes átomos: grupo substituinte $R_3\text{-N-C=O}$ (diedro 1, θ_1), H-C-C=O (θ_2) e H-C-C(O)-NH (θ_3), C-C(O)-NH-C (θ_4) e C(O)-NH-C-C(Ar) (θ_5), respectivamente. Os valores de θ foram obtidos com o programa HyperChem 7.51 (Hypercube Inc., 2002).

Figura 15. Diedros considerados como descritores na análise de QSAR multivariada. θ_1 representado em azul; θ_2 em vermelho, θ_3 em verde, θ_4 em roxo e θ_5 em amarelo .

Entre os descritores que expressam lipofilicidade, que são importantes na interpretação do perfil farmacocinético, calcularam-se o coeficiente de partição de espécies ionizadas (ClogD) nos valores de pH 1,5; 5,0; 6,5 e 7,4 (Marvin Beans 5.2.1, ChemAxon Ltd., 1999-2010) e o coeficiente de partição na forma neutra (ClogP) (HyperChem 7.51, Hypercube Inc., 2002). O valor de ClogP foi obtido a partir de dados experimentais de partição octanol/água de diversos fragmentos e grupos químicos para o cálculo do logP de uma molécula, segundo os métodos de Ghose, Pritchett e Crippen (1988).

As propriedades topológicas refletem forma e tamanho molecular e procuram descrever através de características bidimensionais propriedades relacionadas à superfície das moléculas (BASAK *et al.*, 2000). As variáveis independentes desta natureza calculadas e utilizadas na análise QSAR multivariada estão relacionadas a seguir com as respectivas definições.

- ☆ Volume de van der Waals – calculado a partir de uma grade virtual para delimitar a superfície da molécula e considera o raio atômico de cada átomo pelo método de Bodor, Gabanvi e Wong (1989) (programa HyperChem 7.51, Hypercube Inc., 2002).
- ☆ Área de superfície polar 2D (PSA, *polar surface area*) – utiliza os átomos polares da molécula no cálculo e pode ser correlacionada com a capacidade de transporte passivo na membrana biológica (programa Marvin Beans 5.2.1,

ChemAxon Ltd., 1999-2010). A área de superfície polar topológica estimada é baseada no método desenvolvido por Ertl, Rhode e Selzer (2000) e os resultados são muito aproximados aos da PSA 3D, mas cerca de 100 vezes mais rápidos.

- ☆ Área de superfície molecular (MSA, *molecular surface area*) – é calculada com o método desenvolvido por Ferrara, Apostolakis e Caflisch (2002), que corresponde ao mesmo método utilizado para o cálculo do volume de van der Waals (programa Marvin Beans 5.2.1, ChemAxon Ltd., 1999-2010).
- ☆ Descritores topológicos baseados no contorno ou bordas da superfície molecular – índice de Platt (PLATT, 1947, 1952) e índice de Randic (RANDIC, 1993) (programa Marvin Beans 5.2.1, ChemAxon Ltd., 1999-2010).
- ☆ Descritores topológicos baseados em distâncias – índice de Balaban (BALABAN, 1982); índice de Harary (PLAVSIC *et. al.*, 1993); índice de Szeged (KHADIKAR *et. al.*, 1995); índice de polaridade de Wiener, índice ou valor de Wiener e a variante índice superior de Wiener (WIERNER, 1947); excentricidade (maior valor na correspondente linha da matriz de distância para cada átomo) (programa Marvin Beans 5.2.1, ChemAxon Ltd., 1999-2010).
- ☆ Índice topológico de efeito estérico (CAO, LIU, 2004) - relacionado ao impedimento estérico (volume específico relativo) de átomos de grupos substituintes presentes no sistema em estudo.

Além das propriedades já mencionadas, foram calculados descritores de caráter misto como refratividade molar (RM) (HyperChem 7.51) e polarizabilidade (α) (HyperChem 7.51). O cálculo de RM de uma determinada molécula considera valores experimentais de refratividade de fragmentos ou grupos químicos

(SUBHASH *et. al.*, 2002; GHOSE, CRIPPEN, 1987). Já o cálculo de α considera a eletronegatividade dos átomos da molécula e diferentes incrementos são atribuídos a diferentes tipos de átomos (MILLER, 1990).

Calcularam-se, então, 65 descritores ou variáveis independentes. Uma vez que as variáveis independentes obtidas constituem natureza diversa e, portanto, apresentam diferentes ordens de grandeza, há necessidade de pré-processar os dados para que possam ser utilizados comparativamente na análise QSAR (FERREIRA, 2002a). O pré-processamento empregado nesta análise QSAR foi o autoescalamento, que consiste em centrar na média e dividir pelo desvio-padrão cada coluna correspondente da tabela de descritores (FERREIRA 2002a).

O fluxograma apresentado na figura 16 contém as etapas iniciais da abordagem multivariada até o cálculo dos descritores (tabelas de variáveis independentes com descritores eletrônicos obtidos por diferentes métodos) que serão submetidos à seleção de variáveis e, posteriormente, construção de modelos QSAR.

Figura 16. Fluxograma da abordagem QSAR multivariada partindo do conjunto inicial de dados (compostos investigados) até o cálculo de variáveis independentes ou descritores.

3.4.2. Seleção de variáveis e construção dos modelos

Nas tabelas geradas no item **3.4.1** as linhas correspondem ao número de compostos do conjunto investigado e as colunas às variáveis independentes, sendo a última coluna de cada tabela referente ao dado biológico (valores de pCl_{50}), variável dependente da análise QSAR.

Neste estudo, antes da etapa de seleção de variáveis e construção de modelos QSAR procedeu-se à inspeção visual dos dados gerados, que pode ser considerada como pré-filtro para a seleção de variáveis (KIRALJ, FERREIRA, 2009). O comportamento ou distribuição das variáveis independentes (eixo x), uma de cada vez, em relação aos valores de pCl_{50} (eixo y) foi visualizado em gráficos. Descritores não dispersos em relação ao dado biológico foram eliminados já nesta etapa. A correlação linear entre as variáveis independentes e a variável dependente também foi calculada e considerada na investigação (FERREIRA, 2002; 2002a). No entanto, descritores com boa dispersão em relação aos valores de pCl_{50} , mas com valores não tão altos de correlação linear (0,3 ou mais), foram ainda mantidos nesta etapa, já que funções do tipo não linear podem ser estabelecidas em análise QSAR utilizando o programa WOLF 5.5 (ROGERS, 1994). Foram selecionados 42 descritores com a utilização destes dois filtros.

Conforme já mencionado, análise QSAR multivariada preliminar foi desenvolvida empregando-se o conjunto total de dados para gerar os modelos, ou seja, os compostos ainda não foram separados em conjunto de treinamento e conjunto de avaliação. Neste momento do estudo, a intenção foi buscar as melhores condições que justificassem a obtenção de modelos matemáticos utilizando todos os ligantes do conjunto investigado.

Os dados das tabelas geradas foram embaralhados no programa MATLAB v.7.3.0 (The MathWorks, Inc., 2011) antes de iniciar o processo de construção e otimização de modelos QSAR no programa WOLF 5.5 (ROGERS, 1994). O embaralhamento foi aplicado para evitar qualquer tipo de tendência em relação à ordem de inserção de dados.

A metodologia de algoritmo genético GFA (*Genetic Function Approximation*) (ROGERS, HOPFINGER, 1994) foi empregada na seleção de variáveis e otimização de modelos, enquanto que o método de regressão PLS foi utilizado na construção de modelos QSAR. Os dois métodos estão implementados no programa WOLF 5.5 (ROGERS, 1994).

O algoritmo GFA utiliza uma população de múltiplos bons modelos e testa somente os modelos finais, que agregam as melhores funções. Modelos melhores são gerados à medida que se desenvolvem as recombinações genéticas, que têm como finalidade reorganizar os termos que conferem melhor perfil aos modelos. Os modelos iniciais são gerados por meio de seleção aleatória de um determinado número de características (descritores) do conjunto de dados investigado e funções básicas considerando estas características são construídas de acordo com as especificações do usuário sobre o tipo de função a ser empregado. A partir daí, os modelos genéticos são gerados com sequências aleatórias daquelas funções básicas. O algoritmo GFA pode gerar modelos que utilizem não somente funções lineares, mas também funções quadráticas (polinômios de segunda ordem), *splines* e outras funções não-lineares (ROGERS, HOPFINGER, 1994).

Oito melhores modelos são selecionados pelo programa WOLF 5.5 e classificados pelo fator de falta de ajuste de Friedman (LOF, *lack-of-fit*), que é a medida de mínimos quadrados penalizada. O fator de ajuste ou *smoothing factor* (d),

que faz parte da definição da medida estatística LOF, é o único parâmetro determinado pelo usuário (eq. 1; ROGERS, 1994) e altera o equilíbrio entre o número de variáveis independentes nos modelos e a redução no valor do erro de mínimos quadrados (LSE, *least squares error*). O valor padrão do fator de ajuste é 1,0 (ROGERS, 1994).

$$\text{LOF} = \frac{\text{LSE}}{\left(1 - \frac{c + dp}{M}\right)^2} \quad (\text{eq. 1})$$

c = número de funções no modelo

d = fator de ajuste

p = número total de características contidas em todas as funções de base

M = número de amostras pertencentes ao conjunto de treinamento

Neste estudo, testaram-se a função linear e polinômios de segunda ordem e variaram-se os valores do fator de ajuste de 1,0 a 0,4. O número de ciclos de operações genéticas (*crossovers*) foi de 50.000 a 70.000 e a probabilidade de mutação de 10% foi testada em cada ciclo de recombinação genética. Avaliaram-se os oito melhores modelos de acordo com os seguintes parâmetros estatísticos (ERIKSSON, *et. al.*, 2003): coeficiente de determinação múltipla (r^2), coeficiente de validação cruzada *leave-one-out* (LOO) (q^2), LSE e LOF. A matriz de correlação-cruzada de descritores também foi calculada para eliminar pares de descritores do modelo selecionado que apresentassem valores de coeficiente de correlação de Pearson (R) maiores que 0,7 (método GFA-PLS). Da mesma forma, calculou-se a matriz de correlação-cruzada de resíduos de ajuste entre pares de modelos para determinar se os oito modelos selecionados apresentavam informações comuns ou distintas. Pares de modelos altamente correlacionados ($R \approx 1$) foram considerados o mesmo modelo, enquanto que os pares de modelos pouco correlacionados ($R < 0,7$) foram considerados diferentes entre si. Os ligantes que apresentaram valor da

diferença entre atividade experimental e calculada (resíduo) maior que dois desvios em relação ao desvio-padrão da média dos resíduos do modelo (limite de confiança de 95%) foram considerados compostos desviantes estatisticamente ou amostras atípicas (*outliers*).

As etapas referentes à obtenção de modelos matemáticos com o programa WOLF 5.5 (ROGERS, 1994) estão apresentadas na figura 17.

Figura 17. Fluxograma referente às etapas de seleção de variáveis e construção de modelos QSAR com o programa WOLF 5.5 (ROGERS, 1994).

Os modelos preliminares obtidos com o conjunto completo de dados foram submetidos aos procedimentos de validação preliminar, descritos no item 3.4.3.

Uma vez determinadas as melhores condições para geração dos modelos, realizou-se a separação do conjunto original de dados (tabelas A, B, C e D) em conjunto de treinamento (construção de modelos) e de avaliação (validação externa) considerando os seguintes critérios: análise hierárquica de agrupamentos (HCA);

valores de resíduos de Student *versus leverage* ou alavancamento (expressa a influência da amostra na construção do modelo) de cada amostra no modelo; gradiente de atividade biológica e diversidade estrutural (KIRALJ, FERREIRA, 2009). Dez compostos foram selecionados para compor o conjunto de avaliação. A quantidade de compostos está de acordo com a faixa de 20 a 25% preconizada (TROPSHA, 2002; KIRALJ, FERREIRA, 2009).

As tabelas de dados para a construção dos modelos QSAR finais constituíram-se de 35 ou 36 amostras (considerando a presença ou ausência do composto com iodo) e, como já mencionado, as condições adotadas para a construção e avaliação dos modelos finais foram àquelas previamente determinadas para o conjunto total de compostos (análise QSAR preliminar). Variaram-se os valores do fator de ajuste de 1,0 a 0,05. O número de ciclos de operações genéticas (crossovers) foi de 50.000 a 70.000 e a probabilidade de mutação de 10% foi testada em cada ciclo de recombinação genética. O método de regressão MLR foi utilizado nesta etapa para construção dos modelos e, por conta disso, o valor aceitável de R na matriz de correlação entre descritores diminuiu de 0,7 para 0,5 (HOPFINGER *et al.*, 1997).

3.4.3. Validação dos modelos QSAR da análise multivariada

Após seleção do melhor modelo de cada tabela de dados, procedeu-se com a etapa de validação no programa Pirouette v.3.11 (Infometrix, Inc., 2003) aplicando-se o procedimento de validação cruzada *leave-N-out* (LNO) (CLARK, 2004; BAUMANN, 2003), para avaliar a robustez dos modelos, e de aleatorização dos valores de **y** (*y-randomization*) (RÜCKER, 2007; WOLD, 1995), que indica a chance

de correlação ou correlação ao acaso entre a variável dependente e os descritores dos modelos gerados.

No procedimento de validação cruzada LNO, subconjuntos com N amostras do conjunto inicial de dados são aleatoriamente excluídos. A validação LNO é desenvolvida para $N = 2, 3$, etc. e os coeficientes de validação cruzada LNO (q^2_{LNO}) são calculados da mesma maneira (fórmula semelhante) que para a validação cruzada LOO. É recomendado que N represente uma fração de amostras significativa, ou seja, que pelo menos 20 a 30% de amostras sejam retiradas do total do conjunto inicial (GRAMATICA, 2007; KIRALJ, FERREIRA, 2009). Recentemente foi demonstrado que a repetição da validação LNO para conjuntos embaralhados com a utilização do valor médio de q^2_{LNO} e o respectivo desvio-padrão para cada N é estatisticamente mais confiável do que desenvolver o procedimento LNO uma só vez (TÉOFILO, MARTINS, FERREIRA, 2009). No final do processo, considera-se bom modelo ou modelo robusto aquele cujos valores de q^2_{LNO} estiverem próximos ao valor de q^2_{LOO} , com pequenas variações em todos os valores de N até alcançar o valor crítico de N . O N crítico é o valor máximo de N em que o valor de q^2_{LNO} ainda é estável e alto (próximo do valor de q^2_{LOO}). Para cada validação LNO, as variações de valores expressam o quanto q^2_{LNO} oscila do valor médio e esta oscilação não deve exceder 0,1. No caso de réplicas de LNO, critério mais rigoroso é considerado, ou seja, o valor de dois desvios em relação ao desvio-padrão médio das repetições para cada valor de N ($N = 2, 3$, etc) não deve ser maior que 0,1, incluindo o valor crítico de N (KIRALJ, FERREIRA, 2009).

Neste estudo, os procedimentos de validação cruzada LNO foram desenvolvidos em triplicata. Para as tabelas de dados preliminares, com 46 e 45 compostos, retiraram-se subconjuntos com até 23 e 22 compostos, respectivamente,

que correspondem a 50 e 49% do total de amostras do conjunto inicial de dados. O critério de avaliação supramencionado e considerado como o mais rigoroso foi adotado, já que a validação cruzada LNO foi realizada com três repetições.

Considerando a validação cruzada LNO para o conjunto de treinamento, nas tabelas compostas por 36 ou 35 amostras foram retiradas subconjuntos de até 18 compostos, que corresponde a 50% do total de amostras.

No procedimento *y-randomization* várias corridas de cálculo são realizadas para gerar modelos QSAR mantendo fixas as colunas correspondentes aos descritores da tabela de dados (matriz **X**) e embaralhando somente os valores da coluna da variável dependente (vetor **y**). Os modelos obtidos nestas condições devem apresentar baixa qualidade estatística (valores baixos de $q^2_{y\text{rand}}$ e $r^2_{y\text{rand}}$) e nenhum significado real (BAUMANN, STIEFL, 2004; ERIKSSON et al., 2003; CLARK, FOK, 2004; WOLD, ERIKSSON, 1995; RÜCKER, RÜCKER, MERINGER, 2007). São duas as dificuldades em relação à aleatorização dos valores de **y**: (1) como avaliar os resultados/modelos de cada corrida de cálculo; e (2) como saber o número ideal de aleatorizações de valores de **y** para o conjunto de dados (KIRALJ, FERREIRA, 2009).

A abordagem simples descrita por Eriksson e Wold (1995) pode ser resumida como um conjunto de decisões que expressam a falta de qualidade dos modelos obtidos no processo de aleatorização do **y** com base na avaliação dos valores de $q^2_{y\text{rand}}$ e $r^2_{y\text{rand}}$ e na relação $r^2_{y\text{rand}} > q^2_{y\text{rand}}$. Desta forma, têm-se as seguintes situações:

- ☆ $q^2_{y\text{rand}} < 0,2$ e $r^2_{y\text{rand}} < 0,2 \Rightarrow$ nenhuma correlação ao acaso;
- ☆ qualquer $q^2_{y\text{rand}}$ e $0,2 < r^2_{y\text{rand}} < 0,3 \Rightarrow$ correlação ao acaso insignificante;
- ☆ qualquer $q^2_{y\text{rand}}$ e $0,3 < r^2_{y\text{rand}} < 0,4 \Rightarrow$ correlação ao acaso tolerável;

☆ qualquer $q^2_{y\text{rand}}$ e $r^2_{y\text{rand}} > 0,4 \Rightarrow$ correlação ao acaso importante.

Assim, a frequência de correlação ao acaso corresponde ao número de aleatorizações que resultaram em modelos com correlações falsas (falsos bons modelos), o que pode ser facilmente visualizado em gráficos que apresentam os valores de $q^2_{y\text{rand}}$ *versus* $r^2_{y\text{rand}}$ e que incluem, também, os valores das medidas estatísticas do modelo real, q^2_{LOO} e r^2_{LOO} (KIRALJ, FERREIRA, 2009).

A abordagem de Eriksson e Wold (1995) consiste de dez aleatorizações de y para qualquer conjunto de dados. O teste é suficientemente sensível porque falsos bons modelos facilmente falham em uma ou mais corrida aleatorizada de cálculo (pelo menos 10%). Muitos pesquisadores sugerem centenas ou milhares de aleatorizações independente do tamanho do conjunto de dados, enquanto que outros argumentam que o número de aleatorizações depende do tamanho do conjunto de dados. Kiralj e Ferreira (2009) demonstraram que 10 e 1.000 aleatorizações podem gerar a mesma informação qualitativa e as medidas estatísticas resultantes destas duas abordagens podem não ser claramente distintas quando as relações entre $q^2_{y\text{rand}}$ e $y^2_{y\text{rand}}$ são inspecionadas. Portanto, é esperado que modelos pobres apresentem desempenho indesejado no procedimento de *y-randomization*, enquanto que bons modelos devem estar livres de quaisquer correlações ao acaso, mesmo para um número pequeno de aleatorizações. Os autores verificaram, também, que 10 a 25 aleatorizações de y são suficientes no processo de validação de modelos QSAR.

Neste estudo os procedimentos de *y-randomization* foram desenvolvidos com 30 aleatorizações para todos os conjuntos de dados. A avaliação dos resultados foi baseada nos critérios descritos por Eriksson e Wold (1995) considerando a avaliação dos valores de $q^2_{y\text{rand}}$ e $r^2_{y\text{rand}}$ e na relação $r^2_{y\text{rand}} > q^2_{y\text{rand}}$. A frequência de correlação

ao acaso foi visualizada em gráficos com os valores de q^2_{yrand} *versus* r^2_{yrand} , incluindo os valores das medidas estatísticas do modelo real, q^2_{LOO} e r^2_{LOO} .

Realizou-se, também, a validação externa para verificar a capacidade de previsão do modelo QSAR final. A separação dos compostos para compor o conjunto de avaliação considerou a representatividade de diversidade estrutural e de gradiente de atividade do conjunto original. A análise hierárquica de agrupamentos (HCA; método completo e distância Euclideana) e a avaliação dos resíduos de Student *versus* leverage (alavancamento, influência das amostras no modelo) para o conjunto original foram desenvolvidas para auxiliar na escolha dos compostos do conjunto de avaliação. Os valores dos descritores selecionados no modelo final obtidos para os compostos do conjunto de avaliação ($n = 10$) foram substituídos na equação para gerar os valores calculados de pCl_{50} . Consideraram-se compostos ou ligantes bem previstos pelo modelo final aqueles que apresentaram o valor da diferença entre pCl_{50} calculada ou prevista e pCl_{50} experimental (resíduo) menor ou igual a um desvio em relação ao desvio-padrão da média dos resíduos do modelo (limite de confiança de 99%).

RESULTADOS E DISCUSSÃO

4. RESULTADOS E DISCUSSÃO

4. 1. Simulações de Dinâmica Molecular

Além de se estabelecer as condições necessárias para o equilíbrio dos sistemas investigados, neste estudo, as simulações de DM foram desenvolvidas para selecionar confôrmeros dos ligantes do conjunto de pirrolidina carboxamidas com características estruturais semelhantes aos dados experimentais, uma vez que os ligantes das estruturas cristalografadas estão no sítio de interação da biomacromolécula alvo (enzima InhA) e, portanto, apresentam orientação e disposição espacial dos grupos funcionais mais próximas à conformação bioativa. Os confôrmeros de energia mais baixa selecionados dos PAC foram minimizados e considerados os modelos moleculares iniciais nos cálculos dos descritores utilizados na análise QSAR multivariada.

Cabe ressaltar, também, que os PAC resultantes das simulações de DM serão empregados em um próximo estudo, em andamento, que consiste na aplicação da metodologia de QSAR-4D independente do receptor (IR) (HOPFINGER *et al.*, 1997), com o propósito de mapear o farmacóforo, explorar a liberdade conformacional e de alinhamento dos derivados pirrolidina carboxamidas. Para esta abordagem, apesar de estarem disponíveis cinco inibidores cristalografadas na enzima InhA, a geometria do receptor foi negligenciada porque o mecanismo de ação ainda não está totalmente elucidado.

O PAC resultante da DM de 200 ps a 298 K para cada ligante (figura 18) foi avaliado em função dos valores de E_{tot} sem considerar as contribuições de E_{solv} de E_{Hb} . Somente os ligantes TBP6 e TBP46 não apresentaram os respectivos PAC

equilibrados nos primeiros 100 ps de simulação, provavelmente devido à natureza do substituinte R_2 (I e Br em posição *para* no anel aromático). Considera-se que a equilibração do sistema acontece quando não há oscilação brusca dos valores de energia (LEACH, 2001). Os gráficos representando a segunda parte da DM (100ps) de cada composto estão representados na figura18.

Figura 18. Gráficos referentes à segunda parte da simulação de DM (100 ps a 298 K) dos 46 ligantes. A energia do sistema foi calculada sem considerar as contribuições E_{Hb} e E_{solv} .

Figura 18 (cont.). Gráficos referentes à segunda parte da simulação de DM (100 ps a 298 K) dos 46 ligantes. A energia do sistema foi calculada sem considerar as contribuições E_{Hb} e E_{solv} .

Figura 18 (cont.). Gráficos referentes à segunda parte da simulação de DM (100 ps a 298 K) dos 46 ligantes. A energia do sistema foi calculada sem considerar as contribuições E_{Hb} e E_{solv} .

Figura 18 (cont.). Gráficos referentes à segunda parte da simulação de DM (100 ps a 298 K) dos 46 ligantes. A energia do sistema foi calculada sem considerar as contribuições E_{Hb} e E_{solv} .

Figura 18 (cont.). Gráficos referentes à segunda parte da simulação de DM (100 ps a 298 K) dos 46 ligantes. A energia do sistema foi calculada sem considerar as contribuições E_{Hb} e E_{solv} .

Figura 18 (cont.). Gráficos referentes à segunda parte da simulação de DM (100 ps a 298 K) dos 46 ligantes. A energia do sistema foi calculada sem considerar as contribuições E_{Hb} e E_{solv} .

A faixa de variação dos valores de E_{tot} (sem considerar E_{solv} e E_{Hb}) para os ligantes TBP1 a TBP30 e ligante TBP33 foi de 50 kcal/mol $\pm$ 7 kcal/mol. Para os ligantes que possuem substituintes volumosos, como TBP31, TBP32, TBP34 a TBP43 (quadro 4) os valores de E_{tot} variaram de 100 $\pm$ 11 kcal/mol. Já os ligantes TBP44, TBP45 e TBP46 apresentaram os menores valores de E_{tot} (25 $\pm$ 2 kcal/mol). Estes compostos possuem dois anéis aromáticos substituídos. No entanto, comparar valores de E_{tot} de diferentes compostos e tentar relacioná-los com o padrão de

substituição dos compostos não é informação útil no processo de planejamento, uma vez que se trata de moléculas com diferentes números de átomos.

Para os compostos TBP1, TBP39 e TBP46, que apresentaram faixas de E_{tot} distintas (figura 20 – gráfico DM), porque possuem número de átomos diferentes, é possível observar como característica estrutural importante os substituintes na posição R_3 , conforme representado na figura 19.

Figura 19. Superfície de van der Waals dos ligantes (A) TBP1, (B) TBP39 e (C) TBP46 (programa Viewer Lite 5.0, Accelrys 2002). Os modelos 3D estão representados em palito. Os átomos de carbono estão em cinza, oxigênios em vermelho, bromo em vermelho escuro, flúor em azul claro (superfície transparente em vermelho), nitrogênios em azul e hidrogênios em branco.

Figura 20. Segunda parte da simulação de DM (100 ps a 298 K) dos ligantes TBP39, TBP1 e TBP46. E_{tot} sem considerar as contribuições de E_{solv} e E_{Hb} .

De acordo com He e colaboradores (2006), a presença de anel aromático em substituição ao anel cicloexil nos ligantes TBP44, TBP45 e TBP46 e de substituintes eletronegativos alteram o estado de ressonância destes compostos, o que reflete de forma desfavorável na atividade.

Como já mencionado, calcularam-se as contribuições intramoleculares de E_{solv} e de E_{Hb} para o confômero de cada ligante selecionado a partir do equilíbrio do PAC. A figura 21 apresenta a relação da energia potencial total (E_{tot}), incluindo as contribuições de E_{solv} e E_{Hb} , para cada confômero selecionado. Os valores de todas as contribuições de energia estão listados no quadro 5.

Figura 21. Valores de energia potencial total para os confômeros selecionados dos PAC.

$E_{\text{tot}} = E_{\text{est}} + E_{\text{ang}} + E_{\text{tors}} + E_{1,4} + E_{\text{vdW}} + E_{\text{el}} + E_{\text{solv}} + E_{\text{Hb}}$. Compostos com valores de E_{tot} menores que 30 kcal/mol estão no círculo de linha verde pontilhada; compostos com valores de E_{tot} entre 30 e 70 kcal/mol estão na faixa tracejada em azul e compostos com valores de E_{tot} maiores que 70 kcal/mol estão no retângulo de linha vermelha pontilhada.

A energia potencial está relacionada à mudança de posição ou de coordenadas atômicas de cada elemento da estrutura de um ligante na busca por arranjos conformacionais mais estáveis. Então, o valor de E_{tot} de cada ligante está associado ao confômero mais frequente, de energia mais baixa, selecionado do PAC, uma vez que as simulações foram processadas com ligantes isolados (estado não ligado), sem a presença da biomacromolécula alvo. A interação com

biomacromolcula alvo est relacionada energia livre do sistema (ΔG , energia de Gibbs), que se refere atividade inibitria do sistema (pCl_{50}).

De acordo com os valores de E_{tot} , considerando todas as contribuies intramoleculares de energia (figura 19), podem-se distinguir trs faixas: valores de E_{tot} menores que 30 kcal/mol (TBP44, TBP45 e TBP46), valores de E_{tot} entre 30 e 70 kcal/mol (31 ligantes) e valores de E_{tot} maiores que 70 kcal/mol (TBP31, TPB32, TBP34 a TPB43).

Quadro 5. Valores das contribuies de energia em kcal/mol para os confrmers de menor energia selecionados dos PAC dos ligantes do conjunto investigado provenientes das simulaes de DM de 200 ps a 298 K

Composto	E_{est}	E_{ang}	E_{tors}	$E_{1,4}$	E_{vdW}	E_{el}	E_{solv}	E_{Hb}	E_{tot}
TBP 1	6,57	16,21	17,13	15,66	-3,33	-2,34	-1,77	-0,44	47,68
TBP 2	6,88	19,72	17,63	17,44	-2,36	-3,27	-0,85	-5,02	50,17
TBP 3	7,80	14,97	19,14	15,87	-4,46	-2,99	-1,55	-0,45	48,34
TBP 4	5,70	13,32	18,48	16,31	-2,67	-3,24	-1,88	-0,44	45,58
TBP 5	7,11	15,82	17,22	14,30	-3,96	-3,03	-2,51	-0,45	44,49
TBP 6	11,01	12,05	19,82	14,09	-3,61	-2,56	-1,81	-0,44	48,55
TBP 7	10,77	15,51	14,6	14,21	-3,45	-1,69	-0,92	-0,45	48,58
TBP 8	6,10	16,35	18,24	21,94	-3,48	-5,53	-3,32	-0,46	49,84
TBP 9	9,21	16,69	18,57	28,51	-1,72	-14,44	-6,74	-0,20	49,88
TBP 10	10,82	17,30	19,73	19,67	-4,60	-3,46	0,20	-0,46	59,20
TBP 11	6,99	19,86	17,10	0,40	-5,25	3,79	-4,63	-1,32	36,94
TBP 12	5,51	14,35	17,74	18,20	-4,48	-3,20	-2,87	-0,50	44,76
TBP 13	5,98	17,45	15,48	16,84	-4,33	-3,10	-3,11	-0,50	44,70
TBP 14	6,40	12,13	11,71	17,15	-3,65	-3,35	-1,83	-0,48	38,08
TBP 15	8,51	16,21	17,05	16,14	-4,34	-2,42	-0,62	-0,45	50,08
TBP 16	14,08	21,57	18,91	17,65	-3,55	-2,71	0,15	-0,49	65,61
TBP 17	8,66	16,25	17,28	14,47	-4,33	-1,41	0,37	-0,44	50,86
TBP 18	6,66	18,61	17,15	16,93	-4,49	-3,28	-0,99	-0,50	50,09
TBP 19	5,43	20,76	16,77	21,36	-2,32	-13,08	-5,95	-0,23	42,74
TBP 20	7,28	16,06	18,01	7,72	-3,44	-0,81	-3,13	-0,48	41,20
TBP 21	7,18	13,43	20,41	13,59	-4,43	-2,81	-3,10	-0,47	43,80

Quadro 5 (cont). Valores das contribuições de energia em kcal/mol para os confôrmeros de menor energia selecionados dos PACs dos ligantes do conjunto investigado provenientes das simulações de DM de 200 ps a 298 K

Composto	E_{est}	E_{ang}	E_{tors}	$E_{1,4}$	E_{vdW}	E_{el}	E_{solv}	E_{Hb}	E_{tot}
TBP 22	7,75	16,21	16,81	23,68	-4,21	-6,82	-3,49	-0,47	49,46
TBP 23	9,76	20,97	20,28	21,02	-3,75	-6,96	-4,30	-0,69	56,33
TBP 24	7,94	18,02	16,56	22,96	-3,31	-8,24	-5,22	-0,48	48,23
TBP 25	8,83	23,05	18,16	17,71	-5,19	-5,42	-2,93	-1,16	53,06
TBP 26	6,07	16,08	17,16	17,74	-3,41	-3,09	-2,88	-0,45	47,22
TBP 27	9,92	23,12	16,25	17,81	-4,14	0,18	-1,82	-0,67	60,65
TBP 28	13,90	18,61	8,84	35,66	-3,51	-8,93	-0,63	0,00	63,94
TBP 29	8,13	22,09	14,97	23,01	-5,32	-1,73	-1,34	-0,44	59,36
TBP 30	12,28	22,97	8,19	24,70	-2,35	-4,95	-1,62	-0,45	58,77
TBP 31	19,41	31,53	24,96	39,55	-1,48	-6,67	0,29	0,00	107,59
TBP 32	16,50	34,05	26,78	38,88	-4,06	-9,41	-1,09	0,00	101,65
TBP 33	12,40	27,20	9,79	18,40	-6,57	-4,29	-0,94	-0,46	55,53
TBP 34	11,73	33,47	13,14	28,24	-5,38	-7,50	-0,43	-0,26	73,01
TBP 35	18,92	34,8	26,23	39,82	-7,45	-9,96	-1,99	0,00	100,37
TBP 36	19,64	27,08	25,10	41,04	-2,47	-8,88	-1,86	0,00	99,65
TBP 37	15,00	29,53	22,06	20,56	2,96	-2,02	-1,36	-0,57	86,16
TBP 38	12,40	38,74	22,13	38,26	1,69	-6,95	0,36	0,00	106,63
TBP 39	19,06	30,74	14,36	34,42	10,91	-6,11	-1,92	-1,58	99,89
TBP 40	19,01	34,74	18,45	43,26	-0,32	-10,56	-0,54	0,00	104,04
TBP 41	17,49	39,80	13,31	31,19	9,78	-5,39	-1,89	-0,53	103,76
TBP 42	17,87	35,24	18,70	45,18	1,75	-10,09	0,31	0,00	108,96
TBP 43	24,85	41,59	12,32	33,77	6,88	-3,78	-1,63	-0,53	113,47
TBP 44	3,69	6,24	9,84	15,94	-3,06	-3,34	-4,09	-0,38	24,84
TBP 45	4,41	6,00	10,97	16,15	-3,50	-3,43	-3,12	-0,39	27,10
TBP 46	4,75	6,33	6,95	16,22	-3,61	-2,54	-3,77	-0,49	23,84

* Valores de contribuições de energia em kcal/mol. E_{est} = contribuição de energia de deformação axial ou estiramento; E_{ang} = contribuição de energia de deformação angular; E_{tors} = contribuição de energia de deformação torsional; $E_{1,4}$ = contribuição de energia de interação do tipo 1-4 ou Lennard-Jones; E_{vdW} = contribuição de energia de van der Waals; E_{el} = contribuição de energia eletrostática; E_{solv} = contribuição de energia de solvatação; E_{Hb} = contribuição de energia de ligação de hidrogênio; E_{tot} = energia potencial total, que corresponde à soma de todas as contribuições de energia.

Considerando, ainda, as três faixas de valores de E_{tot} dos confôrmeros selecionados dos PAC, os maiores valores de E_{tot} foram obtidos para os ligantes que

possuem substituintes mais volumosos ligados diretamente ao farmacóforo das pirrolidina carboxamidas e cujas contribuições de E_{ang} foram também maiores. Os confômeros com valores intermediários de E_{tot} apresentaram valores de contribuição de E_{ang} também intermediários. Os confômeros dos ligantes TBP44, TBP45 e TBP46, que são os únicos que possuem anel aromático em substituição ao cicloexil, apresentaram menores valores tanto de E_{tot} quanto de contribuição de E_{ang} (quadro 5).

A contribuição intramolecular de E_{Hb} para a maioria dos ligantes foi em torno de 0,5 kcal/mol em módulo. No entanto, os ligantes TBP28, TBP31, TBP32, TBP35, TBP36, TBP38, TBP40 e TBP42 não apresentaram contribuição deste parâmetro termodinâmico na E_{tot} ($E_{Hb} = 0$). Este fato está relacionado estruturalmente à presença de um nitrogênio terciário no núcleo comum das pirrolidinas carboxamidas investigadas, que só aparece nos ligantes supracitados.

A manutenção da integridade estrutural do confômero selecionado da simulação de DM de cada ligante em relação ao ligante cristalografado no sítio ativo da InhA (código do PDB 2h7i) foi avaliada por meio da superposição dos modelos moleculares. Os valores de RMSD variaram entre 10^{-7} e $0,1\text{\AA}$, indicando que a integridade estrutural dos modelos moleculares dos ligantes investigados foi mantida em relação ao inibidor experimental, durante as simulações de DM. Estas conformações foram minimizadas e utilizadas para o cálculo de descritores na análise QSAR multivariada. As contribuições de energia provenientes das simulações de DM foram consideradas na análise QSAR como descritores termodinâmicos.

4.2. Análise QSAR multivariada

Na etapa de inspeção visual do comportamento das variáveis independentes em relação ao dado biológico (valores de pCl_{50}), descritores foram descartados considerando os critérios mencionados no item 3.4.2. Os valores de correlação linear entre os descritores calculados e os valores de pCl_{50} variaram de 0,38 a 0,58 em módulo (Tabela E - Anexos). Descritores que apresentaram valores mais baixos de correlação linear, mas com boa distribuição (dispersão, variabilidade) em relação ao dado biológico, não foram eliminados nesta etapa, já que o programa WOLF 5.5 (ROGERS, 1994) permite a utilização de funções não lineares na construção de modelos QSAR.

A seguir, estão apresentados os gráficos de dispersão de descritores que não foram considerados na análise QSAR preliminar devido ao comportamento (dispersão) em relação ao dado biológico. Entre eles estão: PSA ($R = 0,20$), índice topológico de efeito estérico ($R = 0,22$ a $0,58$) e excentricidade ($R = 0,03$ a $0,24$). Estes descritores não apresentaram uma distribuição ou dispersão ampla em relação ao dado biológico e, portanto, não foram considerados adequados como representantes da variabilidade estrutural para o conjunto investigado.

Para os descritores calculados de índice topológico de efeito estérico e de excentricidade, as contribuições aparecem como valores numéricos em cada átomo do grupo fundamental do composto (figura 14). No entanto, apenas um gráfico de cada descritor está apresentado (figuras 22 a 24) porque o comportamento foi semelhante em todos os átomos selecionados.

Figura 22. Dispersão gráfica do descritor de área de superfície polar (PSA) em relação à atividade biológica.

Figura 23. Dispersão gráfica do descritor de índice topológico de efeito estérico para o nitrogênio N₁ do núcleo comum às pirrolidina carboxamidas em relação à atividade biológica.

Figura 24. Dispersão gráfica do descritor de excentricidade para o nitrogênio N₁ do núcleo comum às pirrolidina carboxamidas em relação à atividade biológica.

4.2.1. Seleção de variáveis e construção de modelos QSAR (GFA-PLS) com os dados da Tabela A (descritores eletrônicos calculados com método B3LYP/3-21G*)

A tabela A (Anexos) contém 46 amostras e 42 descritores e foi empregada para gerar modelos QSAR preliminares com a metodologia GFA-PLS, utilizando o programa WOLF 5.5 (ROGERS, 1994). Os melhores modelos foram obtidos com as seguintes condições: fator de ajuste de 0,7; 10% de mutação para cada ciclo de recombinação (*crossover*) e 50.000 ciclos de operações genéticas.

Na Tabela 1 estão apresentadas as medidas estatísticas, o número de descritores e o número de amostras atípicas ou *outliers* dos oito melhores modelos para o conjunto de 46 pirrolidina carboxamidas.

Cabe lembrar que os valores estatísticos mínimos para modelos de regressão aceitáveis em QSAR são $q^2 > 0,5$ e $r^2 > 0,6$ (TROPSHA, GRAMATICA, GOMBAR, 2003; GOLBRAIKH, TROPSHA, 2002) e que uma diferença grande entre os valores de r^2 e q^2 (maior que 0,2-0,3) indica claramente que o modelo está superajustado (*Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship*. OECD, 2007; ERIKSSON *et. al.*, 2003), ou seja, que as variáveis independentes presentes no modelo são intercorrelacionadas e, portanto, não expressam informação única ao modelo.

Tabela 1. Medidas estatísticas, número de descritores e número de amostras atípicas dos oito melhores modelos, utilizando fator de ajuste de 0,7 (n = 46)

Modelo	nº de descritores	r^2	q^2	LOF	LSE	<i>outliers</i>
1	4	0,60	0,54	0,28	0,20	1
2	4	0,60	0,54	0,28	0,20	1
3	4	0,60	0,54	0,28	0,20	1
4	4	0,60	0,54	0,28	0,20	1
5	4	0,60	0,54	0,28	0,20	1
6	4	0,60	0,53	0,28	0,20	0
7	4	0,60	0,53	0,28	0,20	1
8	4	0,60	0,53	0,28	0,21	1

Os valores de r^2 permaneceram em 0,60 e os valores de q^2 variaram entre 0,53 a 0,54. Os valores de LSE variaram entre 0,20 a 0,21 e de LOF permaneceu em 0,28. Apenas o modelo 6 não apresentou amostras atípicas ou *outliers*.

A matriz de correlação-cruzada de resíduos de ajuste entre pares de modelos dos oito melhores modelos foi calculada (tabela 2) e os valores de R foram maiores que 0,7, indicando a presença de modelos equivalentes, ou seja, existe um só modelo (modelo 6).

Tabela 2. Matriz de correlação-cruzada de resíduos de ajuste entre pares de modelos para os oito melhores modelos

Modelo	1	2	3	4	5	6	7	8
1	1,00							
2	1,00	1,00						
3	0,99	0,99	1,00					
4	0,99	0,99	1,00	1,00				
5	0,99	0,99	1,00	1,00	1,00			
6	0,72	0,71	0,72	0,72	0,72	1,00		
7	0,99	0,99	0,99	0,99	0,99	0,72	1,00	
8	0,99	0,99	0,99	0,99	0,99	0,71	0,98	1,00

O modelo 6 (eq.2) apresentou funções lineares e não lineares (quadráticas) e está apresentado a seguir.

$$pCl_{50} = 4,90 + 0,13 (\theta_5 + 0,66)^2 + 0,45 ClogD_{pH6,5} + 0,13 (\theta_1 + 0,28)^2 + 0,20 (\mu_T + 0,40)^2 \quad (eq.2)$$

$$n = 46 \quad q^2 = 0,53 \quad r^2 = 0,60 \quad LOF = 0,28 \quad LSE = 0,20 \quad outliers = 0$$

As propriedades físico-químicas ou descritores do modelo 6 são: o diedro C(O)-N-R³-C (θ_5 , figura 15); a partição calculada de espécies ionizadas e não-ionizadas em pH 6,5 (ClogD_{pH6,5}); o diedro R³-N-C=O (θ_1 , figura 15) e momento de dipolo total (μ_T). Todos os descritores contribuem favoravelmente ao efeito biológico (coeficientes de regressão positivos).

A matriz de correlação-cruzada entre os descritores do melhor modelo (modelo 6) foi calculada (tabela 3) e os valores de coeficiente de correlação de Pearson (R), em módulo, foram menores que 0,7, indicando que cada descritor fornece informação independente ao modelo.

Tabela 3. Matriz de correlação-cruzada entre descritores do modelo 6 ($n = 46$)

	μ_T	θ_1	θ_5	$\text{ClogD}_{\text{pH}6.5}$
μ_T	1,00			
θ_1	-0,13	1,00		
θ_5	0,51	-0,17	1,00	
$\text{ClogD}_{\text{pH}6.5}$	-0,38	-0,09	-0,40	1,00

A análise preliminar com aplicação do método B3LYP/3-21G* para o cálculo dos descritores eletrônicos permitiu a construção de bons modelos QSAR considerando todas as amostras do conjunto de dados, ou seja, com significativa representatividade.

Na etapa de validação preliminar, procedeu-se a análise de validação cruzada LNO, em triplicata e com retirada de até 23 amostras, ou seja, 50% das amostras do conjunto inicial (46 compostos). Os dados estatísticos avaliados estão presentes na tabela 4 e última linha corresponde à validação cruzada LOO ($N = 1$). A figura 25 demonstra o comportamento dos valores médios de q^2_{LNO} à medida que as amostras são retiradas (N).

Tabela 4. Valores de q^2_{LNO} e das respectivas variações obtidas na validação cruzada LNO do modelo 6

LNO	q^2_{LNO} (I)	q^2_{LNO} (II)	q^2_{LNO} (III)	$\langle q^2_{LNO} \rangle$	DP	2DP
2	0,54	0,55	0,54	0,54	0,00	0,01
3	0,55	0,52	0,55	0,54	0,01	0,03
4	0,54	0,54	0,54	0,54	0,00	0,00
5	0,56	0,54	0,56	0,55	0,01	0,02
6	0,53	0,52	0,53	0,53	0,01	0,02
7	0,52	0,50	0,52	0,51	0,01	0,02
8	0,54	0,55	0,54	0,55	0,00	0,01
9	0,54	0,47	0,54	0,52	0,04	0,08
10	0,48	0,51	0,48	0,49	0,02	0,04
11	0,51	0,51	0,51	0,51	0,00	0,00
12	0,53	0,52	0,53	0,53	0,01	0,01
13	0,54	0,50	0,54	0,53	0,02	0,05
14	0,50	0,52	0,50	0,51	0,01	0,01
15	0,53	0,53	0,53	0,53	0,00	0,01
16	0,52	0,51	0,52	0,52	0,00	0,01
17	0,52	0,53	0,52	0,53	0,01	0,01
18	0,49	0,53	0,49	0,51	0,03	0,05
19	0,38	0,54	0,38	0,43	0,09	0,18
20	0,39	0,50	0,39	0,43	0,06	0,13
21	0,46	0,51	0,46	0,48	0,03	0,05
22	0,49	0,52	0,49	0,50	0,02	0,03
23	0,44	0,46	0,44	0,45	0,01	0,03
LOO	0,54	0,54	0,54	0,54	0,00	0,00

q^2_{LNO} (I), (II), (III) = valores de q^2_{LNO} para a primeira, segunda e terceira repetições, respectivamente; $\langle q^2_{LNO} \rangle$ = valor médio de q^2_{LNO} ; DP = desvio-padrão da média; 2DP = valor de dois desvios do valor médio de q^2_{LNO} .

De acordo com a tabela 4, o modelo 6 apresentou oscilação importante dos valores de q^2_{LNO} , ou seja, valores de 2DP maiores que 0,1, quando foram retirados subconjuntos de $N = 19$ amostras (2DP = 0,18 ~ 0,2). Portanto, o valor de N crítico para este conjunto de dados foi $N = 18$ (39 % do total de amostras). Na figura 25 está a visualização dos valores médios de q^2_{LNO} com a retirada dos subconjuntos de amostras ($N = 2$ a 23) para o modelo 6.

Figura 25. Variação dos valores de q^2_{LNO} com a retirada de amostras ($N = 2$ a 23) do modelo 6.

Como já mencionado na seção Metodologia, para que um modelo seja robusto no procedimento de validação cruzada LNO, o valor de dois desvios em relação ao desvio-padrão da média das repetições em cada N não deve ser maior que 0,1 para pelo menos 20 a 30 % das amostras retiradas do conjunto inicial (KIRALJ, FERREIRA, 2009), que neste caso, corresponderia de 9 a 14 compostos.

Procederam-se 30 corridas de aleatorização do $\mathbf{y}$, mantendo-se fixos os descritores originais da tabela de dados que gerou o modelo 6 (matriz $\mathbf{X}$). Os resultados foram avaliados de acordo com os critérios definidos por Eriksson e Wold (1995), que consideram a inspeção dos valores de q^2_{yrand} e r^2_{yrand} e da relação $r^2_{yrand} > q^2_{yrand}$.

Na tabela 5 estão apresentados os valores de r^2_{yrand} e q^2_{yrand} para as 30 aleatorizações de $\mathbf{y}$ obtidas para o modelo 6. Os valores de r^2 e q^2 para os valores reais de $\mathbf{y}$ estão indicados na última linha da tabela (modelo real). Para quaisquer valores de q^2_{yrand} os valores de r^2_{yrand} permaneceram entre 0,2 e 0,3, o que indica que o modelo 6 não apresenta correlação ao acaso significativa.

Tabela 5. Valores de $q^2_{y\text{rand}}$ e $r^2_{y\text{rand}}$ para 30 aleatorizações do y (*y-randomization*) para o modelo 6

y_{rand}	$r^2_{y\text{rand}}$	$q^2_{y\text{rand}}$
y_1	0,1	-0,0
y_2	0,0	-0,3
y_3	0,1	-0,0
y_4	0,1	-0,0
y_5	0,1	0,0
y_6	0,0	-0,0
y_7	0,1	0,0
y_8	0,1	0,0
y_9	0,1	-0,0
y_{10}	0,2	0,1
y_{11}	0,2	0,0
y_{12}	0,0	-0,0
y_{13}	0,1	-0,0
y_{14}	0,2	0,0
y_{15}	0,1	-0,0
y_{16}	0,1	0,0
y_{17}	0,0	-0,2
y_{18}	0,1	0,0
y_{19}	0,0	-0,2
y_{20}	0,0	-0,1
y_{21}	0,2	0,0
y_{22}	0,1	-0,0
y_{23}	0,1	0,0
y_{24}	0,1	0,0
y_{25}	0,1	-0,2
y_{26}	0,0	-0,1
y_{27}	0,1	0,0
y_{28}	0,0	0,1
y_{29}	0,2	-0,0
y_{30}	0,2	0,0
y_{modelo}	0,6	0,5

A frequência de correlação ao acaso pode ser visualizada no gráfico com os valores de $q^2_{y\text{rand}}$ *versus* $r^2_{y\text{rand}}$ (figura 26), que inclui os valores das medidas estatísticas do modelo real, q^2_{LOO} e r^2_{LOO} . Verifica-se que todos os valores de $q^2_{y\text{rand}}$ e $r^2_{y\text{rand}}$ estão distantes dos valores de q^2 e r^2 do modelo real (modelo 6), reforçando que a correlação ao acaso é insignificante (ERIKSSON, WOLD, 1995).

Figura 26. Visualização gráfica dos valores de q^2_{yrand} versus r^2_{yrand} para 30 aleatorizações do modelo 6. O dado destacado com retângulo pontilhado vermelho ao modelo real (valores de q^2 e r^2 do modelo 6).

O modelo 6 se apresentou robusto após o procedimento de validação cruzada LNO e não demonstrou correlação ao acaso significativa de acordo com os resultados da aleatorização dos valores de $\mathbf{y}$. Portanto, as condições utilizadas para gerar o modelo 6 serão mantidas para a construção do modelo final ou auxiliar, com o conjunto de treinamento.

Cabe ressaltar que quando se trabalha com conjuntos não muito grandes de dados em análise QSAR (menos de 50 compostos), o procedimento de construção de modelos com todos os compostos antes da separação em conjunto de treinamento e de avaliação é preconizado. Os conjuntos com maior número de compostos apresentam melhor representatividade estatística. No entanto, quando se utilizam todos os compostos para a construção dos modelos alguns processos de validação, como validação externa e *bootstrapping* (conjunto de avaliação), por exemplo, não podem ser realizados. Então, a separação de dados se faz necessária para que todos os processos de validação atualmente recomendados possam ser desenvolvidos (KIRALJ, FERREIRA, 2009).

Os modelos construídos com todos os compostos geralmente são denominados de modelos reais (maior representatividade) e àqueles construídos

com o conjunto de treinamento (após a separação do conjunto de dados) de modelos auxiliares. O objetivo da construção dos modelos auxiliares é possibilitar a aplicação de validações que não podem ser desenvolvidas com os modelos reais (KIRALJ, FERREIRA, 2009). Neste estudo, esta estratégia de construção de modelos QSAR foi adotada.

O conjunto de avaliação foi selecionado da matriz original de dados (tabela A) considerando os critérios descritos na seção Metodologia. A figura 27 mostra o dendrograma que classifica os compostos por similaridade, considerando a distância Euclideana. Compostos mais semelhantes entre si apresentam valores de índice de similaridade próximos a 1.

Figura 27. Dendrograma obtido com o método HCA para os compostos da tabela A (n=46).

Utilizaram-se também os valores de resíduos de *Student versus leverage* ou alavancamento (expressa a influência da amostra na construção do modelo) de cada amostra no modelo, conforme figura 28. Compostos com valores elevados de *leverage* e resíduo de Student, ou somente de *leverage*, não devem ser retirados do modelo para que não seja alterado o ajuste da regressão, previamente obtido na construção do modelo QSAR preliminar, com o conjunto total (FERREIRA, 2002; KIRALJ, FERREIRA, 2009).

Figura 28. Gráfico de dispersão dos valores de resíduo de Student *versus leverage* para os compostos da Tabela A.

Do conjunto de dados com descritores eletrônicos gerados por B3LYP/3-21G* foram selecionados 10 compostos para o conjunto de avaliação (22%), considerando os critério já mencionados, que foram utilizados na etapa de validação externa do modelo QSAR final (TBP41; TBP29; TBP3; TBP33; TBP27; TBP8; TBP10; TBP7; TBP45 e TBP19).

O conjunto de treinamento foi composto por 36 compostos. Os descritores previamente selecionados na análise QSAR preliminar foram as variáveis independentes consideradas na análise (Anexos A). A metodologia GFA-MLR foi

utilizada na etapa de construção e otimização dos modelos QSAR finais no programa WOLF 5.5 (ROGERS, 1994). Os melhores modelos foram obtidos com as seguintes condições: fator de ajuste de 0,01; 10% de mutação para cada ciclo de recombinação (*crossover*) e 70.000 ciclos de operações genéticas.

Na tabela 6 estão apresentadas as medidas estatísticas, o número de descritores e o número de amostras atípicas ou *outliers* dos oito melhores modelos para o conjunto de treinamento ($n = 36$).

Tabela 6. Medidas estatísticas, número de descritores e número de amostras atípicas (*outliers*) dos oito melhores modelos, utilizando fator de ajuste de 0,01 ($n = 36$)

Modelo	nº descritores	r^2	q^2	LOF	LSE	<i>outliers</i>
1	4	0,58	0,47	0,29	0,23	0
2	4	0,57	0,48	0,29	0,23	0
3	4	0,57	0,48	0,29	0,23	1
4	4	0,57	0,45	0,29	0,23	1
5	4	0,57	0,47	0,29	0,23	0
6	4	0,57	0,48	0,29	0,23	0
7	4	0,57	0,48	0,29	0,23	0
8	3	0,54	0,44	0,29	0,24	0

Os valores de r^2 variaram de 0,54 a 0,58 e os valores de q^2 de 0,44 a 0,48. O valor de LOF foi 0,29 e LSE variou de 0,23 a 0,24. Os modelos 2, 6 e 7 apresentaram as mesmas variáveis e as mesmas medidas estatísticas. Selecionou-se o modelo 2 para prosseguir a análise.

A matriz de correlação de resíduos de ajuste entre pares de modelos para os oito melhores modelos foi calculada e está apresentada na tabela 7. Os valores de coeficiente de correlação foram muito próximos de 1 ($R \approx 1$), indicando que os modelos são equivalentes, ou seja, existe apenas um modelo. Então, o modelo 2 que apresenta valor de q^2 maior que o modelo 1 foi selecionado como o melhor modelo.

Tabela 7. Matriz de correlação de resíduos de ajuste entre pares de modelos para os oito melhores modelos

Modelo	1	2	3	4	5	6	7	8
1	1,00							
2	1,00	1,00						
3	0,92	0,93	1,00					
4	1,00	0,99	0,93	1,00				
5	1,00	1,00	0,92	0,99	1,00			
6	1,00	1,00	0,92	0,99	1,00	1,00		
7	0,99	1,00	0,92	0,99	1,00	1,00	1,00	
8	0,96	0,96	0,88	0,95	0,96	0,96	0,96	1,00

O modelo 2 (eq.3), apresentada a seguir, apresentou funções lineares e não lineares (quadráticas).

$$pCl_{50} = 4,90 + 0,10 (\theta_5 + 0,66)^2 + 0,23 (\mu_T + 0,12)^2 + 0,40 \text{ClogD}_{\text{pH}_{6,5}} + 0,12 (\theta_1 + 0,41)^2$$

(eq.3)

$$n = 36 \quad q^2 = 0,48 \quad r^2 = 0,57 \quad \text{LOF} = 0,29 \quad \text{LSE} = 0,23 \quad \text{outliers} = 0$$

Onde: θ_1 é o diedro $R_3\text{-N-C=O}$; $\text{ClogD}_{\text{pH}_{6,5}}$ corresponde ao valor calculado de $\log D$ (coeficiente de partição de espécies ionizadas) em pH 6,5; μ_T corresponde ao momento de dipolo total e θ_5 é o diedro $\text{C(O)-N-R}_2\text{-C}$.

Todos os descritores do modelo 2 contribuem favoravelmente à resposta biológica (coeficientes positivos). O modelo 2 possui descritores que expressam propriedades moleculares de caráter eletrônico (μ_T), estereoquímico (θ_1 e θ_5) e de partição (ClogD em pH 6,5).

A matriz de correlação entre os descritores do modelo selecionado como o melhor modelo foi calculada e está apresentada na tabela 8. O valor de coeficiente de correlação de Pearson (R) entre os descritores θ_5 e μ_T , em módulo, foi maior que 0,5 e indica que estas variáveis são correlacionados entre si e, portanto, não fornecem informação independente ao modelo (superajuste). Além disso, o modelo

apresentou valor de $q^2 < 0,5$ e valor de $r^2 < 0,6$. Então, este conjunto de dados (tabela A final) foi descartado para posteriores análises de validação. A tabela A não gerou modelo QSAR final satisfatório.

Tabela 8. Matriz de correlação entre descritores do modelo 2 (N = 36)

	μ_T	θ_1	θ_5	ClogD _{6,5}
μ_T	1,00			
θ_1	-0,07	1,00		
θ_5	0,56	-0,16	1,00	
ClogD _{6,5}	-0,44	-0,03	-0,46	1,00

4.2.2. Seleção de variáveis e construção de modelos QSAR (GFA-PLS) com os dados da Tabela B (descritores eletrônicos calculados com método HF/3-21G*)

Os dados da tabela B (Anexos), contendo 46 amostras e 42 descritores, foram empregados para gerar modelos QSAR preliminares com a metodologia GFA-PLS, utilizando o programa WOLF 5.5 (ROGERS, 1994). Os melhores modelos foram obtidos com as seguintes condições: fator de ajuste de 0,3; 10 % de mutação para cada ciclo de recombinação (*crossover*) e 50.000 ciclos de operações genéticas.

Na Tabela 9 estão apresentadas as medidas estatísticas, o número de descritores e o número de amostras atípicas ou *outliers* dos oito melhores modelos para o conjunto de 46 pirrolidina carboxamidas.

Tabela 9. Medidas estatísticas, número de descritores e número de amostras atípicas dos oito melhores modelos, utilizando fator de ajuste de 0,5 ($n = 46$)

Modelo	Nº de descritores	r^2	q^2	LOF	LSE	outliers
1	5	0,62	0,51	0,26	0,19	1
2	5	0,63	0,52	0,26	0,19	0
3	5	0,63	0,52	0,25	0,19	1
4	5	0,62	0,52	0,26	0,19	1
5	5	0,62	0,50	0,26	0,19	1
6	4	0,60	0,50	0,26	0,21	1
7	4	0,60	0,51	0,26	0,20	0
8	5	0,63	0,52	0,26	0,19	1

Os valores de r^2 variaram de 0,60 a 0,63 e os valores de q^2 de 0,50 a 0,52. Os valores de LOF e LSE foram de 0,25 a 0,26 e de 0,19 a 0,21, respectivamente. Apenas os modelos 2 e 7 não apresentaram amostras atípicas, o modelo 2 apresentou maiores valores de q^2 e r^2 e, portanto, foi considerado o melhor modelo.

A matriz de correlação-cruzada de resíduos de ajuste entre pares de modelos dos oito melhores modelos foi calculada e está apresentada na tabela 10. Os valores do coeficiente de correlação foram altos ($R \approx 1$) entre todos os modelos, indicando que os modelos são equivalentes, ou seja, existe apenas um modelo (modelo 2).

Tabela 10. Matriz de correlação-cruzada de resíduos de ajuste entre pares de modelos para os oito melhores modelos

Modelo	1	2	3	4	5	6	7	8
1	1,00							
2	0,83	1,00						
3	0,99	0,84	1,00					
4	0,98	0,84	1,00	1,00				
5	0,98	0,82	0,97	0,97	1,00			
6	0,93	0,78	0,94	0,94	0,89	1,00		
7	0,94	0,78	0,93	0,93	0,90	0,99	1,00	
8	0,99	0,84	1,00	1,00	0,98	0,94	0,94	1,00

O modelo 2 (eq.4) apresentou funções lineares e não lineares (quadráticas) e está apresentada a seguir:

$$pCI_{50} = 4,94 + 0,13 (\theta_1 + 0,42)^2 + 0,49 ClogD_{pH6,5} + 0,20 (\mu_T + 0,39)^2 + 0,10 (\theta_5 + 0,66)^2 - 0,02 (E_{HB} - 0,13)^2$$

(eq.4)

$$n = 46 \quad q^2 = 0,52 \quad r^2 = 0,63 \quad LOF = 0,26 \quad LSE = 0,19 \quad outliers = 0$$

Onde: θ_1 é o diedro $R^3-N-C=O$ (figura 3), $ClogD_{pH6,5}$ corresponde ao valor calculado de $\log D$ (coeficiente de partição de espécies ionizadas) em pH 6,5, μ_T corresponde ao momento de dipolo total, θ_5 é o diedro $C(O)-N-R^2-C$ (figura 3) e E_{HB} corresponde à contribuição da energia de ligação de hidrogênio.

Os descritores do modelo 2 contribuem favoravelmente à resposta biológica (coeficientes positivos), exceto E_{HB} . Expressam propriedades moleculares de caráter eletrônico (μ_T) e estereoquímico (θ_1 e θ_5), de partição de espécies ionizadas e não-ionizadas em função do pH do compartimento no organismo ($ClogD$ em pH 6,5) e de caráter termodinâmico (E_{HB}).

A matriz de correlação-cruzada entre os descritores do modelo selecionado como o melhor modelo foi calculada e está apresentada na tabela 11. Os valores de coeficiente de correlação de Pearson (R), em módulo, foram menores que 0,7, indicando que cada descritor fornece informação independente ao modelo.

Tabela 11. Matriz de correlação-cruzada entre descritores do modelo 2 ($n = 46$)

	μ_T	θ_1	θ_5	$ClogD_{pH6,5}$	E_{HB}
μ_T	1,00				
θ_1	-0,13	1,00			
θ_5	0,51	-0,17	1,00		
$ClogD_{pH6,5}$	-0,38	-0,09	-0,40	1,00	
E_{HB}	-0,10	0,00	0,01	0,10	1,00

Então, a análise QSAR preliminar com os dados da tabela B (método HF/3-21G* para o cálculo dos descritores eletrônicos) permitiu a construção de bons modelos considerando todas as amostras do conjunto de dados, ou seja, com significativa representatividade.

Na etapa de validação preliminar, procedeu-se a análise de validação cruzada LNO, em triplicata e com retirada de até 23 amostras, ou seja, 50% das amostras do conjunto inicial (46 compostos). É recomendado que pelo menos 20 a 30 % das amostras sejam retiradas neste processo (KIRALJ, FERREIRA, 2009). Para a Tabela B, seriam pelo menos 14 compostos. Os dados estatísticos avaliados estão presentes na tabela 12 e a última linha corresponde à validação cruzada LOO ($N = 1$). A figura 30 demonstra o comportamento dos valores médios de q^2_{LNO} à medida que as amostras são retiradas (N).

Tabela 12. Valores de q^2_{LNO} e das respectivas variações resultantes na validação cruzada LNO do modelo 2.

LNO	q^2_{LNO} (I)	q^2_{LNO} (II)	q^2_{LNO} (III)	$\langle q^2_{LNO} \rangle$	DP	2DP
2	0,55	0,57	0,55	0,56	0,01	0,03
3	0,58	0,16	0,58	0,44	0,24	0,48
4	0,51	0,55	0,51	0,52	0,02	0,04
5	0,58	0,59	0,58	0,58	0,01	0,02
6	0,57	0,07	0,57	0,40	0,29	0,58
7	0,55	0,48	0,55	0,53	0,04	0,08
8	0,55	0,32	0,55	0,47	0,13	0,27
9	0,54	0,51	0,54	0,53	0,02	0,04
10	0,55	0,55	0,55	0,55	0,00	0,00
11	0,54	0,12	0,54	0,40	0,24	0,49
12	0,49	0,06	0,49	0,35	0,25	0,50
13	0,53	0,13	0,53	0,40	0,23	0,46
14	0,53	0,56	0,53	0,54	0,01	0,03
15	0,57	0,33	0,57	0,49	0,14	0,28
16	0,53	0,34	0,53	0,47	0,11	0,23
17	0,56	0,34	0,56	0,49	0,12	0,24
18	0,36	0,36	0,36	0,36	0,00	0,00
19	0,35	0,35	0,35	0,35	0,00	0,01
20	0,35	0,39	0,35	0,37	0,02	0,04
21	0,17	0,38	0,17	0,24	0,12	0,24
22	0,21	0,01	0,21	0,15	0,11	0,23
23	0,24	0,01	0,24	0,17	0,13	0,26
LOO	0,57	0,57	0,57	0,57	0,00	0,00

q^2_{LNO} (I), (II), (III) = valores de q^2_{LNO} para a primeira, segunda e terceira repetições, respectivamente; $\langle q^2_{LNO} \rangle$ = valor médio de q^2_{LNO} ; DP = desvio-padrão da média; 2DP = valor de dois desvios do valor médio de q^2_{LNO} .

De acordo com a tabela 12, o modelo 2 apresenta oscilação importante dos valores de q^2_{LNO} considerando bem menos de 30% das amostras retiradas (14 compostos). Valores de 2DP maiores que 0,1 foram observados para $N = 3, 6, 8, 11$ a 13, 15 a 17, 21 a 23. Apesar desta inconsistência dos resultados obtidos para a validação cruzada LNO deste conjunto, somente este procedimento de validação não é suficiente à exclusão do modelo 2. Na figura 29 está a visualização dos valores médios de q^2_{LNO} com a retirada dos subconjuntos de amostras ($N = 2$ a 23) para o modelo 2.

Figura 29. Variação dos valores de q^2_{LNO} com a retirada de amostras ($N = 2$ a 23) do modelo 2.

Procederam-se 30 corridas de aleatorização do $\mathbf{y}$, mantendo-se fixos os descritores originais da tabela de dados que gerou o modelo 2 (matriz $\mathbf{X}$). Como já descrito, os resultados foram avaliados de acordo com os critérios definidos por Eriksson e Wold (1995), que consideram a inspeção dos valores de q^2_{yrand} e r^2_{yrand} e da relação $r^2_{yrand} > q^2_{yrand}$.

Na tabela 13 estão apresentados os valores de r^2_{yrand} e q^2_{yrand} para as 30 aleatorizações de $\mathbf{y}$ obtidas para o modelo 2. Os valores de r^2 e q^2 para os valores reais de $\mathbf{y}$ estão indicados na última linha da tabela (modelo real). Os valores de

$q^2_{y_{rand}}$ foram menores que 0,2 e os valores de $r^2_{y_{rand}}$ não ultrapassaram 0,2, o que indica que para o modelo 2 não há correlação ao acaso.

Tabela 13. Valores de $q^2_{y_{rand}}$ e $r^2_{y_{rand}}$ para 30 aleatorizações do y (*y-randomization*) no modelo 2

y_{rand}	$r^2_{y_{rand}}$	$q^2_{y_{rand}}$
y_1	0,1	0,0
y_2	0,1	-0,0
y_3	0,0	-0,0
y_4	0,1	-0,0
y_5	0,1	0,0
y_6	0,1	0,0
y_7	0,1	-0,0
y_8	0,2	-0,0
y_9	0,1	0,0
y_{10}	0,2	0,1
y_{11}	0,2	0,0
y_{12}	0,1	-0,0
y_{13}	0,1	-0,1
y_{14}	0,2	-0,0
y_{15}	0,1	-0,0
y_{16}	0,1	0,0
y_{17}	0,0	-0,2
y_{18}	0,1	0,0
y_{19}	0,1	0,0
y_{20}	0,1	0,0
y_{21}	0,2	0,0
y_{22}	0,1	0,0
y_{23}	0,2	-0,0
y_{24}	0,2	-0,0
y_{25}	0,1	0,0
y_{26}	0,1	0,0
y_{27}	0,0	-0,0
y_{28}	0,1	0,0
y_{29}	0,1	0,0
y_{30}	0,2	-0,0
y_{modelo}	0,6	0,5

A frequência de correlação ao acaso pode ser visualizada na figura 30, que apresenta o gráfico com os valores de $q^2_{y_{rand}}$ *versus* $r^2_{y_{rand}}$, incluindo os valores das

medidas estatísticas do modelo real, q^2_{LOO} e r^2_{LOO} . Verifica-se que todos os valores de q^2_{yrand} e r^2_{yrand} estão distantes dos valores de q^2 e r^2 do modelo real (modelo 2).

Figura 30. Visualização gráfica dos valores de q^2_{yrand} versus r^2_{yrand} para 30 aleatorizações do modelo 2. O dado destacado com retângulo pontilhado refere-se ao modelo real (valores de q^2 e r^2 do modelo 2).

Apesar do modelo 2 ter apresentado medidas estatísticas além do limite preconizado para o teste de validação cruzada LNO (valores de 2DP acima de 0,1 para vários subconjuntos de M), este modelo não foi excluído do estudo QSAR. No procedimento de validação cruzada LNO as amostras são retiradas aleatoriamente para cada subconjunto de N compostos (neste caso até $N = 23$) e o procedimento LNO é muito sensível à ordem das amostras no conjunto de dados, portanto apenas o resultado insatisfatório deste teste não é considerado suficiente para a exclusão do modelo gerado (KIRALJ, FERREIRA, 2009).

Quanto ao teste de aleatorização dos valores de $\mathbf{y}$, as 30 repetições ou corridas demonstraram que o modelo 2 não apresenta correlação ao acaso, ou seja, nenhum modelo gerado se aproximou aos valores de q^2 e r^2 do modelo real (modelo 2).

O conjunto de avaliação foi selecionado da matriz original de dados (tabela B) considerando os critérios descritos na seção Metodologia. A figura 31 mostra o dendrograma que classifica os compostos por similaridade e na figura 32 está o gráfico de dispersão dos valores de resíduos de Student *versus leverage* ou alavancamento. Neste conjunto de dados (tabela B), também, 10 compostos foram considerados para o conjunto de avaliação (22%), também considerando representatividade estrutural e gradiente de dado biológico (dendrograma) além de valores baixos de *leverage* (influência não significativa do compostos na regressão com todos os compostos), que serão utilizados na etapa de validação externa do modelo QSAR final (TBP41; TBP28; TBP30; TBP23; TBP33; TBP34; TBP27; TBP45; TBP26 e TBP7).

Figura 31. Dendrograma obtido com o método HCA para os compostos da tabela B (n=46).

Figura 32. Gráfico de dispersão dos valores de resíduo de Student *versus* leverage.

O conjunto de treinamento, utilizado na construção dos modelos finais, apresenta 36 compostos. Os descritores previamente selecionados na análise QSAR preliminar foram as variáveis independentes consideradas na análise (Anexos).

A metodologia GFA-MLR foi utilizada na etapa de construção e otimização dos modelos QSAR finais no programa WOLF 5.5 (ROGERS, 1994). Os melhores modelos foram obtidos com as seguintes condições: fator de ajuste de 0,01; 10% de mutação para cada ciclo de recombinação (*crossover*) e 70.000 ciclos de operações genéticas.

Na tabela 14 estão apresentadas as medidas estatísticas, o número de descritores e o número de amostras atípicas ou *outliers* dos oito melhores modelos para o conjunto de treinamento com 36 pirrolidina carboxamidas.

Tabela 14. Medidas estatísticas, número de descritores e número de amostras atípicas (*outliers*) dos oito melhores modelos, utilizando fator de ajuste de 0,01 ($n = 36$)

Modelo	nº de descritores	r^2	q^2	LOF	LSE	<i>outliers</i>
1	5	0,59	0,49	0,30	0,22	0
2	5	0,59	0,49	0,30	0,22	0
3	5	0,59	0,49	0,30	0,22	0
4	5	0,59	0,50	0,30	0,22	0
5	5	0,59	0,49	0,30	0,22	0
6	5	0,59	0,49	0,30	0,22	0
7	5	0,59	0,49	0,30	0,22	0
8	5	0,59	0,47	0,30	0,22	0

Os valores de r^2 foram 0,59 para todos os modelos e os valores de q^2 variaram de 0,47 a 0,50. Os valores de LOF foram de 0,30 e LSE de 0,22. Nenhum modelo apresentou amostras atípicas. O modelo 4 foi selecionado como o melhor modelo por apresentar maior valor de q^2 .

A matriz de correlação de resíduos de ajuste entre pares de modelos dos oito melhores modelos foi calculada (tabela 15) e os valores de R foram próximos de 1, indicando a existência de apenas um modelo.

Tabela 15. Matriz de correlação de resíduos de ajuste entre pares de modelos para os oito melhores modelos

Modelo	1	2	3	4	5	6	7	8
1	1,00							
2	0,99	1,00						
3	1,00	0,99	1,00					
4	1,00	0,99	1,00	1,00				
5	1,00	0,99	1,00	0,99	1,00			
6	1,00	0,99	1,00	0,99	1,00	1,00		
7	1,00	0,99	1,00	1,00	1,00	1,00	1,00	
8	1,00	0,99	1,00	0,99	1,00	1,00	1,00	1,00

O modelo 4 (eq.5), descrita a seguir, apresentou apenas funções não lineares (quadráticas).

$$pCl_{50} = 5,04 + 0,11 (\theta_5 + 0,66)^2 + 0,51 \text{ClogD}_{\text{pH}_{6,5}} + 0,17 (\mu_T + 0,66)^2 + 0,11 (\theta_1 + 0,69)^2 - 0,05 (E_{\text{HB}} + 1,40)^{2No}$$

(eq.5)

$$n = 36 \quad q^2 = 0,50 \quad r^2 = 0,59 \quad \text{LOF} = 0,30 \quad \text{LSE} = 0,22 \quad \text{outliers} = 0$$

Onde: θ_5 é o diedro $C(O)-N-R^3-C$; $ClogD_{pH_{6,5}}$ corresponde à partição calculada de espécies ionizadas e não-ionizadas em pH 6,5; μ_T corresponde ao momento de dipolo total; θ_1 é o diedro $R^3-N-C=O$ e E_{HB} corresponde à contribuição da energia de ligação de hidrogênio.

No modelo 4, exceto o descritor E_{HB} , os demais descritores contribuem de forma favorável ao efeito biológico (coeficientes de regressão positivos). Os descritores presentes no modelo 4 expressam propriedades de caráter eletrônico (μ_T), partição de espécies ionizadas e não-ionizadas em função do pH do compartimento no organismo ($ClogD$ em pH 6,5), propriedades termodinâmicas (E_{HB}) e de caráter estereoquímico (θ_5).

A matriz de correlação entre os descritores do melhor modelo (modelo 4) foi calculada (tabela 16) e as variáveis μ_T e θ_5 apresentaram valor de coeficiente de correlação de Pearson (R), maior que 0,5 (em módulo), indicando que são correlacionadas entre si e, portanto, não fornecem informação independente ao modelo (superajuste). Por este motivo, este conjunto de dados (tabela C final) foi descartado para posteriores análises de validação. A tabela B também não gerou modelo QSAR satisfatório.

Tabela 16. Matriz de correlação entre descritores do modelo 4 ($n = 36$)

	E_{HB}	μ_T	θ_1	θ_5
E_{HB}	1,00			
μ_T	-0,16	1,00		
θ_1	-0,02	-0,14	1,00	
θ_5	0,00	0,52	-0,16	1,00

4.2.3. Seleção de variáveis e construção de modelos QSAR (GFA-PLS) com dados da Tabela C (descritores eletrônicos calculados com método B3LYP/6-31G*)

A tabela de dados C, que contém 45 amostras e os 42 descritores pré-selecionados (Anexos), foi empregada para gerar modelos QSAR preliminares com a metodologia GFA-PLS, utilizando o programa WOLF 5.5 (ROGERS, 1994). Os melhores modelos foram obtidos com as seguintes condições: fator de ajuste de 0,3; 10% de mutação para cada ciclo de recombinação (*crossover*) e 70.000 ciclos de operações genéticas.

Na Tabela 17 estão apresentadas as medidas estatísticas, o número de descritores e o número de amostras atípicas ou *outliers* dos oito melhores modelos para o conjunto de 45 pirrolidina carboxamidas.

Tabela 17. Medidas estatísticas, número de descritores e número de amostras atípicas dos oito melhores modelos, utilizando fator de ajuste de 0,3 (n = 45)

Modelo	Nº de descritores	r^2	q^2	LOF	LSE	<i>outliers</i>
1	5	0,64	0,54	0,25	0,18	3
2	6	0,67	0,49	0,25	0,17	1
3	5	0,64	0,54	0,25	0,18	0
4	5	0,64	0,54	0,25	0,18	0
5	6	0,66	0,53	0,25	0,17	1
6	4	0,62	0,53	0,25	0,19	1
7	4	0,62	0,52	0,25	0,19	1
8	5	0,64	0,53	0,25	0,18	1

Os valores de r^2 variaram de 0,62 a 0,67 e os valores de q^2 de 0,49 a 0,54. O valor de LOF foi de 0,25 para todos os modelos e os valores de LSE variaram de 0,17 a 0,19. Apenas os modelos 3 e 4 não apresentaram amostras atípicas.

A matriz de correlação-cruzada de resíduos de ajuste entre pares de modelos dos oito melhores modelos foi calculada e está demonstrada na tabela 18. O modelo

5 apresentou valor de $R < 0,7$ o que indica que este modelo não é similar aos modelos restantes. Como o modelo 5 possui amostras atípicas, apesar de distinto, foi descartado.

Como supramencionado, os modelos 3 e 4 não apresentaram amostras atípicas. Com a verificação mais detalhada destes dois modelos, percebeu-se que eles também compartilhavam as mesmas medidas estatísticas e descritores, então o modelo 3 foi selecionado como o melhor modelo.

Tabela 18. Matriz de correlação-cruzada de resíduos de ajuste entre pares de modelos para os oito melhores modelos

Modelo	1	2	3	4	5	6	7	8
1	1,00							
2	0,93	1,00						
3	0,78	0,83	1,00					
4	0,93	1,00	0,83	1,00				
5	0,45	0,54	0,49	0,54	1,00			
6	0,89	0,94	0,75	0,94	0,50	1,00		
7	0,89	0,93	0,74	0,93	0,50	1,00	1,00	
8	0,93	0,99	0,82	0,93	0,53	0,93	0,93	1,00

O modelo 3 (eq.6) apresentou funções lineares e não-lineares (modelo misto) e está descrito a seguir.

$$pCl_{50} = 5,59 + 0,08 (Carga_C_2 + 0,56)^2 + 0,22 (\mu_T + 0,13)^2 - 0,11(Carga_N_{10} - 1,25)^2 - 0,16 (E_{EST} - 0,36)^2 + 0,53 ClogD_{pH6,5}$$

(eq.6)

$$n = 45 \quad q^2 = 0,54 \quad r^2 = 0,64 \quad LOF = 0,25 \quad LSE = 0,18 \quad outliers = 0$$

Onde: Carga_C₂ refere-se à carga de potencial eletrostático do carbono 2 no núcleo comum das pirrolidina carboxamidas; μ_T corresponde ao momento de dipolo total; carga_N₁₀ refere-se à carga de potencial eletrostático do nitrogênio 10 no núcleo comum das pirrolidina carboxamidas; E_{EST} refere-se à contribuição de energia de deformação axial ou de estiramento e ClogD_{pH6,5} é o valor de calculado

de logD (coeficiente de partição de espécies ionizadas) em pH 6,5.

O modelo 3 apresentou descritores de natureza eletrônica, termodinâmica e de partição incluindo potencial de ionização. Considerando os valores dos coeficientes de regressão do modelo 3, verifica-se que os descritores carga_C₂, μ_T e ClogD_{pH6,5} favorecem o efeito biológico, enquanto que carga_N₁₀ e E_{EST} desfavorecem o efeito biológico.

A matriz de correlação-cruzada entre os descritores do modelo selecionado como o melhor modelo foi calculada (tabela 19) e os valores de *R*, em módulo, foram menores que 0,7, indicando que cada descritor fornece informação independente ao modelo.

Tabela 19. Matriz de correlação-cruzada entre descritores do modelo 3 (n = 45)

	E _{EST}	μ_T	carga_C ₂	carga_N ₁₀	ClogD _{pH6,5}
E _{EST}	1,00				
μ_T	-0,15	1,00			
carga_C ₂	-0,09	0,12	1,00		
carga_N ₁₀	0,36	-0,03	-0,09	1,00	
ClogD _{pH6,5}	0,66	-0,38	0,12	0,30	1,00

A análise preliminar com o método B3LYP/6-31G* no cálculo dos descritores eletrônicos também permitiu a construção de bons modelos considerando todas as amostras do conjunto de dados, ou seja, com significativa representatividade.

Na etapa de validação preliminar, procedeu-se a análise de validação cruzada LNO, em triplicata e com retirada de até 22 amostras, ou seja, 49% das amostras do conjunto inicial (45 compostos). Os dados estatísticos avaliados estão presentes na tabela 20 e a última linha corresponde à validação cruzada LOO (*N* = 1). Como critério de avaliação dos resultados considerou-se que o valor de dois desvios em relação ao desvio-padrão da média das repetições, para cada valor de *N* (*N* = 2, 3,

etc), não deveria ser maior que 0,1 para pelo menos 20 a 30 % das amostras retiradas do conjunto inicial (pelo menos de 9 a 14 compostos) (KIRALJ, FERREIRA, 2009).

Tabela 20. Valores de q^2_{LNO} e das respectivas variações resultantes na validação cruzada LNO do modelo 3

LNO	q^2_{LNO} (I)	q^2_{LNO} (II)	q^2_{LNO} (III)	$\langle q^2_{LNO} \rangle$	DP	2DP
2	0,57	0,57	0,56	0,57	0,01	0,01
3	0,57	0,55	0,56	0,56	0,01	0,02
4	0,55	0,54	0,55	0,55	0,01	0,01
5	0,58	0,52	0,55	0,55	0,03	0,06
6	0,54	0,51	0,56	0,54	0,03	0,05
7	0,60	0,57	0,57	0,58	0,02	0,03
8	0,52	0,54	0,56	0,54	0,02	0,04
9	0,53	0,49	0,55	0,52	0,03	0,07
10	0,52	0,55	0,47	0,51	0,04	0,08
11	0,55	0,55	0,58	0,56	0,02	0,03
12	0,53	0,43	0,60	0,52	0,08	0,17
13	0,47	0,28	0,59	0,45	0,16	0,32
14	0,57	0,42	0,56	0,51	0,08	0,17
15	0,53	0,44	0,58	0,52	0,07	0,14
16	0,53	0,51	0,55	0,53	0,02	0,04
17	0,52	0,51	0,56	0,53	0,03	0,06
18	0,54	0,58	0,51	0,54	0,04	0,07
19	0,55	0,61	0,49	0,55	0,06	0,12
20	0,58	0,59	0,54	0,57	0,03	0,05
21	0,56	0,38	0,55	0,50	0,10	0,20
22	0,59	0,60	0,58	0,59	0,01	0,02
LOO	0,57	0,57	0,57	0,57	0,00	0,00

q^2_{LNO} (I), (II), (III) = valores de q^2_{LNO} para a primeira, segunda e terceira repetições, respectivamente; $\langle q^2_{LNO} \rangle$ = valor médio de q^2_{LNO} ; DP = desvio-padrão da média; 2DP = valor de dois desvios do valor médio de q^2_{LNO} .

De acordo com a tabela 20, o modelo 3 apresentou oscilação importante dos valores de q^2_{LNO} com a retirada de subconjuntos de 12 amostras (27 %), mais de 20 % das amostras do conjunto inicial. O valor de N crítico para este conjunto foi $N = 11$ (24 % do total de amostras). Valores de 2DP maiores que 0,1 foram observados para $N = 12$ a 15 e 21.

A figura 33 demonstra o comportamento dos valores médios de q^2_{LNO} à medida que as amostras são retiradas (N).

Figura 33. Variação dos valores de q^2_{LNO} com a retirada de amostras ($N = 2$ a 22) do modelo 3.

Procederam-se 30 corridas de aleatorização de valores de $\mathbf{y}$ (*y-randomization*), mantendo-se fixos os descritores originais da tabela de dados C, que gerou o modelo 3 (matriz $\mathbf{X}$). Como já descrito, os resultados foram avaliados de acordo com os critérios definidos por Eriksson e Wold (1995), que consideram a inspeção dos valores de q^2_{yrand} e r^2_{yrand} e da relação $r^2_{yrand} > q^2_{yrand}$.

Na tabela 21 estão apresentados os valores de r^2_{yrand} e q^2_{yrand} para as 30 aleatorizações dos valores de $\mathbf{y}$ obtidas para o modelo 3. Os valores de r^2 e q^2 para os valores reais de $\mathbf{y}$ estão indicados na última linha da tabela (modelo real). Os dados demonstraram que para quaisquer valores de q^2_{yrand} os valores de r^2_{yrand} não ultrapassaram 0,4, o que indica correlação ao acaso tolerável.

Tabela 21. Valores de $q^2_{y_{rand}}$ e $r^2_{y_{rand}}$ para 30 aleatorizações dos valores de y para o modelo 3

y_{rand}	$r^2_{y_{rand}}$	$q^2_{y_{rand}}$
y_1	0,1	-0,0
y_2	0,0	-0,0
y_3	0,1	0,0
y_4	0,0	-0,1
y_5	0,1	-0,1
y_6	0,2	0,0
y_7	0,1	-0,0
y_8	0,1	0,0
y_9	0,3	0,1
y_{10}	0,1	-0,1
y_{11}	0,1	-0,1
y_{12}	0,2	0,0
y_{13}	0,1	-0,0
y_{14}	0,1	0,0
y_{15}	0,1	0,0
y_{16}	0,3	0,2
y_{17}	0,2	0,1
y_{18}	0,1	0,0
y_{19}	0,2	0,0
y_{20}	0,1	0,0
y_{21}	0,1	0,0
y_{22}	0,1	0,0
y_{23}	0,1	0,0
y_{24}	0,1	0,0
y_{25}	0,2	0,0
y_{26}	0,1	-0,1
y_{27}	0,2	0,0
y_{28}	0,1	0,0
y_{29}	0,1	0,0
y_{30}	0,1	0,0
y_{modelo}	0,6	0,5

A frequência de correlação ao acaso pode ser visualizada na figura 34, que apresenta o gráfico com os valores de $q^2_{y_{rand}}$ *versus* $r^2_{y_{rand}}$, incluindo os valores das medidas estatísticas do modelo real, q^2_{LOO} e r^2_{LOO} .

Figura 34. Visualização gráfica dos valores de q^2_{yrand} versus r^2_{yrand} para 30 aleatorizações dos valores de y no modelo 3. O dado destacado com retângulo pontilhado amarelo alaranjado se refere ao modelo real (valores de q^2 e r^2 do modelo 3).

Verifica-se que os valores de q^2_{yrand} e r^2_{yrand} estão em sua maioria ainda distantes dos valores de q^2 e r^2 do modelo real (modelo 6). Para as corridas de aleatorização y_9 e y_{16} , os valores de r^2_{yrand} estiveram entre 0,3 e 0,4 o que indica correlação ao acaso tolerável (ERIKSSON, WOLD, 1995).

O modelo 3 se apresentou robusto para o procedimento de validação cruzada LNO até a retirada de 27 % das amostras do conjunto inicial de dados e demonstrou chance de correlação tolerável para o resultado da aleatorizações dos valores da variável dependente. Portanto, as condições utilizadas para gerar o modelo 3 foram mantidas para a construção do modelo final ou auxiliar, após a separação dos conjuntos de dados em conjunto de treinamento e de avaliação.

O conjunto de avaliação foi selecionado da matriz original de dados (tabela C) considerando os critérios descritos na seção Metodologia. A figura 35 apresenta o dendrograma que classifica os compostos por índice de similaridade e na figura 36 estão os valores de resíduos de Student versus *leverage* ou alavancamento. Estes resultados foram considerados na separação do conjunto de avaliação, conforme já descrito.

Neste conjunto de dados, também, 10 compostos foram considerados para o conjunto de avaliação (22%), que serão utilizados na etapa de validação externa do modelo QSAR final (TBP42; TBP21; TBP28; TBP20; TBP34; TBP24; TBP31; TBP45; TBP15 e TBP11).

Figura 35. Dendrograma obtido pelo método HCA para os compostos da tabela de dados C (n=45)

Figura 36. Gráfico de dispersão dos valores de resíduo de Student *versus* leverage.

O conjunto de treinamento, utilizado na construção dos modelos finais, apresenta 35 compostos (o composto TBP6 foi excluído por conta do conjunto de base utilizado no cálculo dos descritores eletrônicos). Os descritores previamente selecionados na análise QSAR preliminar foram as variáveis independentes consideradas na construção dos modelos finais (Anexos).

A metodologia GFA-MLR foi utilizada na etapa de construção e otimização dos modelos QSAR finais [programa WOLF 5.5 (ROGERS, 1994)]. Os melhores modelos foram obtidos com as seguintes condições: fator de ajuste de 1; 10% de mutação para cada ciclo de recombinação (*crossover*) e 70.000 ciclos de operações genéticas.

Na tabela 22 estão apresentadas as medidas estatísticas, o número de descritores e o número de amostras atípicas ou *outliers* dos oito melhores modelos para o conjunto de treinamento ($n = 35$).

Tabela 22. Medidas estatísticas, número de descritores e número de amostras atípicas (*outliers*) dos oito melhores modelos, utilizando fator de ajuste de 1,0 (n = 35)

Modelo	nº de descritores	r^2	q^2	LOF	LSE	<i>outliers</i>
1	4	0,70	0,49	0,25	0,15	0
2	4	0,70	0,49	0,25	0,15	0
3	4	0,69	0,49	0,25	0,15	0
4	4	0,69	0,51	0,25	0,15	0
5	4	0,69	0,51	0,25	0,15	0
6	5	0,69	0,51	0,25	0,15	0
7	4	0,69	0,51	0,25	0,15	0
8	4	0,74	0,43	0,25	0,13	0

Os valores de r^2 variaram de 0,69 a 0,74 e q^2 entre 0,43 a 0,51. O valor de LOF permaneceu em 0,25 e o valor de LSE variou entre 0,13 a 0,15. Nenhum modelo apresentou amostras atípicas. Os modelos 4 a 7 apresentaram os maiores valores de q^2 (0,51).

A matriz de correlação de resíduos de ajuste entre pares de modelos dos oito melhores modelos foi calculada (tabela 23) e os valores de R foram próximos de 1, indicando que os modelos são equivalentes (existe apenas um modelo).

Tabela 23. Matriz de correlação de resíduos de ajuste entre pares de modelos para os oito melhores modelos

Modelo	1	2	3	4	5	6	7	8
1	1,00							
2	1,00	1,00						
3	1,00	1,00	1,00					
4	1,00	1,00	1,00	1,00				
5	1,00	1,00	1,00	1,00	1,00			
6	1,00	1,00	1,00	1,00	1,00	1,00		
7	0,99	0,99	0,99	0,99	0,99	0,99	1,00	
8	0,99	0,99	0,99	0,99	0,99	0,99	1,00	1,00

O modelo 4 (eq. 7), demonstrado a seguir, apresentou funções lineares e não lineares (quadráticas).

$$pCI_{50} = 5,72 + 0,68 \text{ ClogD}_{\text{pH}_{6,5}} + 0,28 (\mu_T + 0,49)^2 - 0,17 (\text{Carga_N}_{10} - 1,25)^2 - 0,15 (\text{E}_{\text{EST}} - 0,05)^2$$

(eq.7)

$$n = 35 \quad q^2 = 0,51 \quad r^2 = 0,69 \quad \text{LOF} = 0,25 \quad \text{LSE} = 0,15 \quad \text{outliers} = 0$$

Onde: $\text{ClogD}_{\text{pH}_{6,5}}$ corresponde ao valor calculado de logD (coeficiente de partição de espécies ionizadas) em pH 6,5; μ_T corresponde ao momento de dipolo total; Carga_N_{10} corresponde à carga de potencial eletrostático (CHELPG) no N_{10} do núcleo fundamental das pirrolidina carboxamidas e E_{EST} corresponde à contribuição da energia de deformação axial ou de estiramento.

Com exceção dos descritores $\text{ClogD}_{\text{PH}6.5}$ e μ_T , os demais descritores selecionados no modelo 4 contribuem de forma não favorável à resposta biológica (coeficientes negativos).

A matriz de correlação entre os descritores do modelo selecionado como o melhor modelo foi calculada (tabela 24) e os descritores E_{EST} e $\text{ClogD}_{\text{pH}_{6,5}}$ apresentaram valor de coeficiente de correlação de Pearson (R) maior que 0,5 (em módulo), indicando que são correlacionados entre si e, portanto, não fornecem informação independente ao modelo (superajuste). Por este motivo, este conjunto de dados (tabela C final) foi descartado para posteriores análises de validação. A tabela C também não gerou modelo QSAR final satisfatório.

Tabela 24. Matriz de correlação-cruzada entre descritores do modelo 4 ($n = 35$)

	E_{EST}	μ_T	Carga_N_{10}	$\text{ClogD}_{6,5}$
E_{EST}	1,00			
μ_T	-0,21	1,00		
Carga_N_{10}	0,28	-0,13	1,00	
$\text{ClogD}_{6,5}$	0,66	-0,46	0,22	1,00

4.2.4. Seleção de variáveis e construção de modelos QSAR (GFA-PLS) com dados da Tabela D (descritores eletrônicos calculados com método HF/6-31G*)

Os dados pré-selecionados da tabela D (seção Anexos – 42 descritores e 45 amostras) foram empregados para gerar modelos QSAR preliminares com a metodologia GFA-PLS, utilizando o programa WOLF 5.5 (ROGERS, 1994). Neste caso, o conjunto de base 6-31G* restringe moléculas que contenham o elemento iodo. Então, o composto TBP6 foi excluído desta análise preliminar.

Os melhores modelos foram obtidos com as seguintes condições: fator de ajuste de 0,3; 10% de mutação para cada ciclo de recombinação (*crossover*) e 50.000 ciclos de operações genéticas.

Na Tabela 25 estão apresentadas as medidas estatísticas, o número de descritores e o número de amostras atípicas ou *outliers* dos oito melhores modelos para o conjunto de 45 pirrolidina carboxamidas.

Tabela 25. Medidas estatísticas, número de descritores e número de amostras atípicas dos oito melhores modelos, utilizando fator de ajuste de 0,3 (n = 45)

Modelo	N ^o de descritores	r ²	q ²	LOF	LSE	outliers
1	6	0,70	0,58	0,23	0,16	1
2	5	0,70	0,61	0,21	0,16	0
3	6	0,71	0,60	0,22	0,15	1
4	6	0,70	0,57	0,23	0,16	1
5	5	0,70	0,62	0,22	0,16	0
6	6	0,70	0,60	0,23	0,16	1
7	5	0,70	0,62	0,21	0,16	0
8	5	0,70	0,60	0,23	0,16	3

Os valores de r² variaram de 0,62 a 0,71 e os valores de q² de 0,57 a 0,62. Os valores de LOF variaram entre 0,21 a 0,23 e de LSE entre 0,15 a 0,16. Dentre os modelos que não apresentaram amostras atípicas (modelos 2, 5 e 7), o modelo 7 apresentou maior valor de q² e menor valor de LOF.

A matriz de correlação-cruzada de resíduos de ajuste entre pares de modelos dos oito melhores modelos foi calculada (tabela 26) e os valores de R foram próximos de 1, indicando a existência de apenas um modelo.

Tabela 26. Matriz de correlação-cruzada de resíduos de ajuste entre pares de modelos para os oito melhores modelos

Modelo	1	2	3	4	5	6	7	8
1	1,00							
2	0,84	1,00						
3	0,90	0,97	1,00					
4	0,99	0,85	0,90	1,00				
5	0,83	0,99	0,97	0,83	1,00			
6	0,98	0,85	0,91	0,99	0,85	1,00		
7	0,84	1,00	0,97	0,84	1,00	0,85	1,00	
8	0,71	0,84	0,80	0,72	0,82	0,71	0,83	1,00

O modelo 7 (eq.8) apresentou funções lineares e não-lineares (quadráticas) e está descrito a seguir.

$$pCI_{50} = 6,41 - 0,25 (E_{EST} - 1,34)^2 - 0,31 \theta_4 - 0,13 (E_{SOLV} + 1,32)^2 - 0,14 (E_{EL} + 0,51)^2 + 0,19 (Carga_C_2 + 0,61)^2$$

(eq.8)

$$n = 45 \quad q^2 = 0,62 \quad r^2 = 0,70 \quad LOF = 0,21 \quad LSE = 0,16 \quad outliers = 0$$

Onde: E_{EST} corresponde à contribuição da energia de deformação axial ou de estiramento; θ_4 é o diedro C-C(O)-N-R² (figura 15); E_{SOLV} corresponde à contribuição da energia de solvatação; E_{EL} corresponde à contribuição de energia eletrostática e $Carga_C_2$ corresponde à carga de potencial eletrostático (CHELPG) no carbono 2 do núcleo fundamental das pirrolidina carboxamidas (figura 14).

Os descritores selecionados no modelo 7 contribuem de forma não favorável à resposta biológica (coeficientes negativos), exceto o descritor referente à carga no C₂. Expressam propriedades de caráter eletrônico ($Carga_C_2$), estereoquímico (θ_4) e, também, propriedades termodinâmicas (E_{HB} , E_{SOLV} e E_{EL}).

A matriz de correlação-cruzada entre os descritores do modelo selecionado

como o melhor modelo foi calculada (tabela 27) e os valores de coeficiente de correlação de Pearson (R), em módulo, foram menores que 0,7, indicando que cada descritor fornece informação independente ao modelo.

Tabela 27. Matriz de correlação-cruzada entre descritores do modelo 7 ($n = 45$)

	E_{EST}	E_{EL}	E_{SOLV}	carga_ C_2	θ_4
E_{EST}	1,00				
E_{EL}	-0,38	1,00			
E_{SOLV}	0,46	0,16	1,00		
carga_ C_2	-0,15	0,12	-0,06	1,00	
θ_4	0,09	-0,12	-0,09	-0,08	1,00

A análise preliminar com o método HF/6-31G* no cálculo dos descritores eletrônicos permitiu a construção de bons modelos considerando todas as amostras do conjunto de dados, ou seja, com significativa representatividade.

O modelo 7, apresentou as melhores medidas estatísticas (maiores valores de q^2 e r^2 , menores valores de LOF e LSE) em relação aos modelos gerados com as demais tabelas A, B e C. Como já mencionado, de acordo com este método de cálculo de descritores eletrônicos, o composto TBP6 não foi incluído na análise.

Na etapa de validação preliminar, procedeu-se a análise de validação cruzada LNO, em triplicata e com retirada de até 22 amostras, ou seja, 49% das amostras do conjunto inicial (45 compostos). Os dados estatísticos avaliados estão presentes na tabela 28 e a última linha corresponde à validação cruzada LOO ($N = 1$). A figura 37 demonstra o comportamento dos valores médios de q^2_{LNO} à medida que as amostras são retiradas (N).

Tabela 28. Valores de q^2_{LNO} e das respectivas variações resultantes na validação cruzada LNO do modelo 7

LNO	$q^2_{LNO}(I)$	$q^2_{LNO}(II)$	$q^2_{LNO}(III)$	$\langle q^2_{LNO} \rangle$	DP	2DP
2	0,61	0,61	0,61	0,61	0,00	0,01
3	0,61	0,62	0,60	0,61	0,01	0,02
4	0,60	0,60	0,60	0,60	0,00	0,00
5	0,59	0,58	0,62	0,60	0,02	0,04
6	0,59	0,61	0,60	0,60	0,01	0,02
7	0,61	0,65	0,57	0,61	0,04	0,08
8	0,59	0,56	0,59	0,58	0,02	0,03
9	0,61	0,64	0,61	0,62	0,02	0,03
10	0,65	0,55	0,56	0,59	0,05	0,11
11	0,64	0,59	0,58	0,60	0,04	0,07
12	0,58	0,61	0,53	0,57	0,04	0,08
13	0,53	0,51	0,61	0,55	0,05	0,10
14	0,58	0,58	0,61	0,59	0,02	0,04
15	0,54	0,50	0,58	0,54	0,04	0,08
16	0,57	0,54	0,56	0,56	0,01	0,03
17	0,59	0,55	0,51	0,55	0,04	0,08
18	0,62	0,58	0,48	0,56	0,07	0,14
19	0,62	0,59	0,52	0,58	0,05	0,11
20	0,62	0,57	0,57	0,59	0,03	0,06
21	0,58	0,48	0,60	0,55	0,06	0,13
22	0,60	0,59	0,54	0,58	0,03	0,06
LOO	0,62	0,62	0,62	0,62	0,00	0,00

$q^2_{LNO}(I)$, (II) , (III) = valores de q^2_{LNO} para a primeira, segunda e terceira repetições, respectivamente; $\langle q^2_{LNO} \rangle$ = valor médio de q^2_{LNO} ; DP = desvio-padrão da média; 2DP = valor de dois desvios do valor médio de q^2_{LNO} .

Figura 37. Variação dos valores de q^2_{LNO} com a retirada de amostras ($N = 2$ a 22) do modelo 7

De acordo com a tabela 28, o modelo 7 não apresentou oscilação importante de valores de q^2_{LNO} , ou seja, valores de 2DP não foram maiores que 0,1. Portanto, o

valor de N máximo de compostos retirados para este conjunto foi $N = 22$ (49% do total de amostras). O N crítico é o valor máximo de N em que o valor de q^2_{LNO} ainda é estável e alto (próximo do valor de q^2_{LOO}). Então, para o procedimento preliminar de validação cruzada LNO o modelo 7 foi considerado robusto. Cabe ressaltar que os valores de 2DP para $N = 10, 18, 19$ e 21 apresentaram variação na segunda decimal, mas tal variação não é considerada significativa (0,11; 0,14 e 0,13 ~0,1).

Na tabela 29 estão apresentados os valores de r^2_{yrand} e q^2_{yrand} para as 30 aleatorizações dos valores de y obtidas para o modelo 7. Os valores de r^2 e q^2 para os valores reais de y estão indicados na última linha da tabela (modelo real). Os valores de q^2_{yrand} foram menores que 0,2 exceto para y_{25} e os valores de r^2_{yrand} ultrapassaram 0,2 em y_{22} e y_{29} , o que indica que para o modelo 2 não há correlação ao acaso significativa (ERIKSSON, WOLD, 1995) pois para quaisquer valores de q^2_{yrand} , os valores de r^2_{yrand} permaneceram entre 0,2 e 0,3.

Tabela 29. Valores de q^2_{yrand} e r^2_{yrand} para 30 aleatorizações do y (y -randomization) no modelo 7

y_{rand}	r^2_{yrand}	q^2_{yrand}
y_1	0,1	0,0
y_2	0,1	-0,1
y_3	0,2	0,0
y_4	0,1	0,0
y_5	0,0	-0,2
y_6	0,2	0,0
y_7	0,1	-0,1
y_8	0,1	0,0
y_9	0,0	-0,2
y_{10}	0,1	-0,1
y_{11}	0,1	0,0
y_{12}	0,2	0,0
y_{13}	0,1	0,0
y_{14}	0,1	-0,1
y_{15}	0,1	0,0

Tabela 29 (cont.). Valores de $q^2_{y\text{rand}}$ e $r^2_{y\text{rand}}$ para 30 aleatorizações do **y** (*y-randomization*) no modelo 7

y_{rand}	r²_{yrand}	q²_{yrand}
y₁₆	0,1	0,0
y₁₇	0,0	-0,2
y₁₈	0,1	0,0
y₁₉	0,2	0,0
y₂₀	0,1	0,0
y₂₁	0,2	0,0
y₂₂	0,3	0,2
y₂₃	0,1	0,0
y₂₄	0,1	0,0
y₂₅	0,0	-0,3
y₂₆	0,1	0,0
y₂₇	0,2	0,0
y₂₈	0,1	0,0
y₂₉	0,3	0,2
y₃₀	0,1	-0,1
y_{modelo}	0,7	0,6

A frequência de correlação ao acaso pode ser visualizada na figura 38, que apresenta o gráfico com os valores de $q^2_{y\text{rand}}$ *versus* $r^2_{y\text{rand}}$, incluindo os valores das medidas estatísticas do modelo real, q^2_{LOO} e r^2_{LOO} . Verifica-se que todos os valores de $q^2_{y\text{rand}}$ e $r^2_{y\text{rand}}$ estão distantes dos valores de q^2 e r^2 do modelo real (modelo 7), para quaisquer valores de $q^2_{y\text{rand}}$, os valores de $r^2_{y\text{rand}}$ não ultrapassaram 0,3, exceto para **y₂₂** e **y₂₉** ($r^2_{y\text{rand}}$ entre 0,3 e 0,4) indicando correlação ao acaso tolerável, de acordo com os critérios de Eriksson e Wold (1995).

Figura 38. Visualização gráfica dos valores de q^2_{yrand} versus r^2_{yrand} para 30 aleatorizações do modelo 7. O dado destacado com retângulo pontilhado vermelho refere-se ao modelo real (valores de q^2 e r^2 do modelo 7).

O modelo 7 se apresentou robusto após o procedimento de validação cruzada LNO e não demonstrou chance de correlação quando os valores de y foram embaralhados (*y-randomization*), portanto as condições utilizadas para gerar o modelo 7 foram mantidas para a geração do modelo final ou auxiliar com o conjunto de treinamento (separação dos dados iniciais).

O conjunto de avaliação foi selecionado da matriz original de dados (tabela D) considerando os critérios descritos na seção Metodologia. A figura 39 apresenta o dendrograma que classifica os compostos por índice de similaridade e na figura 40 estão os valores de resíduos de Student *versus leverage* ou alavancamento. Neste conjunto de dados, também, 10 compostos foram considerados para o conjunto de avaliação (22%), que serão utilizados na etapa de validação externa do modelo QSAR final (TBP21; TBP40; TBP28; TBP43; TBP33; TBP25; TBP10; TBP1; TBP45 e TBP26).

Figura 39. Dendrograma obtido com o método HCA para os compostos da tabela D de dados (n=45)

Figura 40. Gráfico de dispersão dos valores de resíduo de Student vs. *leverage*.

O conjunto de treinamento, utilizado na construção dos modelos finais, apresenta 35 compostos (o composto TBP6 foi excluído por conta do conjunto de base utilizado no cálculo dos descritores eletrônicos). Os descritores previamente selecionados na análise QSAR preliminar foram as variáveis independentes consideradas na análise (Anexos).

A metodologia GFA-MLR foi utilizada na etapa de construção e otimização dos modelos QSAR finais [programa WOLF 5.5 (ROGERS, 1994)]. Os melhores modelos foram obtidos com as seguintes condições: fator de ajuste de 0,5; 10% de mutação para cada ciclo de recombinação (*crossover*) e 70.000 ciclos de operações genéticas.

Na tabela 30 estão apresentadas as medidas estatísticas, o número de descritores e o número de amostras atípicas ou *outliers* dos oito melhores modelos para o conjunto de 35 pirrolidina-carboxamidas.

Tabela 30. Medidas estatísticas, número de descritores e número de amostras atípicas (*outliers*) dos oito melhores modelos, utilizando fator de ajuste de 0,5 (n = 35)

Modelo	nº de descritores	r^2	q^2	LOF	LSE	<i>outliers</i>
1	5	0,70	0,59	0,27	0,17	0
2	5	0,70	0,59	0,28	0,17	0
3	5	0,69	0,59	0,28	0,17	0
4	5	0,69	0,59	0,28	0,17	0
5	5	0,69	0,59	0,28	0,17	0
6	5	0,69	0,59	0,28	0,17	0
7	5	0,69	0,59	0,28	0,17	0
8	5	0,69	0,59	0,28	0,17	0

Os valores de r^2 variaram de 0,69 a 0,70 e o valor de q^2 foi de 0,59 para os oito modelos. Os valores de LOF variaram de 0,27 a 0,28 e o valor de LSE foi de 0,17 para todos os modelos. Nenhum modelo apresentou amostras atípicas.

A matriz de correlação de resíduos de ajuste entre pares de modelos dos oito melhores modelos foi calculada e está demonstrada na tabela 31. Os valores de R

foram próximos de 1, indicando a existência de apenas um modelo. O modelo 1 foi selecionado como o melhor modelo porque apresentou menor valor de LOF.

Tabela 31. Matriz de correlação de resíduos de ajuste entre pares de modelos para os oito melhores modelos

Modelo	1	2	3	4	5	6	7	8
1	1,00							
2	1,00	1,00						
3	1,00	1,00	1,00					
4	1,00	1,00	1,00	1,00				
5	1,00	1,00	1,00	1,00	1,00			
6	1,00	1,00	1,00	1,00	1,00	1,00		
7	0,99	0,99	0,99	0,99	0,99	0,99	1,00	
8	0,99	0,99	0,99	0,99	0,99	0,99	1,00	1,00

O modelo 1 (eq.9), descrita a seguir, apresentou funções lineares e não-lineares (modelo misto).

$$pCI_{50} = 6,24 + 0,19 (Carga_C_2 + 0,72)^2 - 0,31 \theta_4 - 0,16 (E_{SOLV} + 1,01)^2 - 0,11 (E_{EL} + 0,74)^2 - 0,23 (E_{EST} - 1,34)^2$$

(eq.9)

$$n = 35 \quad q^2 = 0,59 \quad r^2 = 0,70 \quad LOF = 0,27 \quad LSE = 0,17 \quad outliers = 0$$

Onde: Carga_C₂ refere-se à carga de potencial eletrostático do C₂ no núcleo comum das pirrolidina-carboxamidas; θ_4 é o diedro R³-N-C=O; E_{SOLV} refere-se à contribuição de energia de solvatação; E_{EL} refere-se à contribuição de energia de eletrostática e E_{EST} corresponde à contribuição da energia de deformação axial ou de estiramento.

O modelo 1 apresentou descritores de natureza eletrônica, termodinâmica e estereoquímica. Considerando os valores dos coeficientes de regressão do modelo 1, verifica-se que apenas o descritor Carga_C₂ favorece o efeito biológico (coeficiente de regressão positivo).

A matriz de correlação entre os descritores do modelo selecionado como o melhor modelo foi calculada (tabela 32) e todos os valores de *R* (em módulo) foram

menores que 0,5, indicando que cada descritor fornece informação independente ao modelo. O modelo 1 foi utilizado, então, nos processos de validação.

Tabela 32. Matriz de correlação entre descritores do modelo 1 ($n = 35$)

	E_{EST}	E_{EL}	E_{SOLV}	carga_ C_2	θ_4
E_{EST}	1,00				
E_{EL}	-0,37	1,00			
E_{SOLV}	0,46	0,25	1,00		
carga_ C_2	-0,19	0,05	-0,06	1,00	
θ_4	0,10	-0,14	-0,09	-0,10	1,00

Na validação cruzada LNO é recomendado que pelo menos 20 a 30 % das amostras sejam retiradas (KIRALJ, FERREIRA, 2009). Para a tabela D corresponderia à retirada de 7 a 10 compostos. Os dados estatísticos avaliados estão presentes na tabela 33 e a última linha corresponde à validação cruzada LOO ($N = 1$). A figura 34 demonstra o comportamento dos valores médios de q^2_{LNO} à medida que as amostras são retiradas (N).

Como já mencionado na seção Metodologia, quando a validação cruzada LNO é realizada com repetições o critério de avaliação dos resultados é mais rigoroso, ou seja, considera-se que o valor de dois desvios em relação ao desvio-padrão da média das repetições, para cada valor de N ($N = 2, 3$, etc), não deve ser maior que 0,1 para pelo menos 20 a 30 % das amostras retiradas do conjunto inicial (KIRALJ, FERREIRA, 2009).

De acordo com a tabela 33, o modelo 1 apresenta oscilação importante dos valores de q^2_{LNO} a partir de 10 amostras retiradas. O N crítico para este conjunto de dados foi igual a 9 (valores de q^2_{LNO} permaneceram estáveis, semelhantes ao q^2_{LOO}). Valores de 2DP maiores que 0,1 foram observados para $N = 10, 12$ e 17. Os resultados obtidos para a validação cruzada LNO deste conjunto foi considerado satisfatório.

Tabela 33. Valores de q^2_{LNO} e das respectivas variações resultantes na validação cruzada LNO do modelo 1 ($n = 35$)

LNO	q^2_{LNO} (I)	q^2_{LNO} (II)	q^2_{LNO} (III)	$\langle q^2_{LNO} \rangle$	DP	2DP
2	0,59	0,61	0,58	0,59	0,02	0,03
3	0,59	0,59	0,57	0,58	0,01	0,02
4	0,58	0,61	0,56	0,58	0,02	0,05
5	0,58	0,58	0,61	0,59	0,01	0,03
6	0,57	0,61	0,52	0,57	0,04	0,09
7	0,61	0,64	0,60	0,62	0,02	0,04
8	0,60	0,59	0,56	0,58	0,02	0,05
9	0,60	0,60	0,56	0,59	0,02	0,05
10	0,62	0,52	0,51	0,55	0,06	0,13
11	0,46	0,46	0,46	0,46	0,00	0,01
12	0,44	0,58	0,48	0,50	0,07	0,14
13	0,53	0,56	0,60	0,56	0,03	0,06
14	0,57	0,59	0,62	0,59	0,03	0,05
15	0,56	0,52	0,59	0,56	0,03	0,07
16	0,60	0,57	0,54	0,57	0,03	0,05
17	0,66	0,58	0,27	0,51	0,21	0,41
LOO	0,60	0,60	0,60	0,60	0,00	0,00

q^2_{LNO} (I), (II), (III) = valores de q^2_{LNO} para a primeira, segunda e terceira repetições, respectivamente; $\langle q^2_{LNO} \rangle$ = valor médio de q^2_{LNO} ; DP = desvio-padrão da média; 2DP = valor de dois desvios do valor médio de q^2_{LNO} .

Na figura 41 está a visualização dos valores médios de q^2_{LNO} com a retirada dos subconjuntos de amostras para o modelo 1.

Figura 41. Variação dos valores de q^2_{LNO} com a retirada de amostras ($N = 2$ a 17) do modelo 1.

Procederam-se 30 corridas de aleatorização do y , mantendo-se fixos os descritores originais da tabela de dados que gerou o modelo 1 (matriz X). Como já descrito, os resultados foram avaliados de acordo com os critérios definidos por Eriksson e Wold (1995), que consideram a inspeção dos valores de $q^2_{y\text{rand}}$ e $r^2_{y\text{rand}}$ e da relação $r^2_{y\text{rand}} > q^2_{y\text{rand}}$.

Na tabela 34 estão apresentados os valores de $r^2_{y\text{rand}}$ e $q^2_{y\text{rand}}$ para as 30 aleatorizações de y obtidas para o modelo 1. Os valores de r^2 e q^2 para os valores reais de y estão indicados na última linha da tabela (modelo real).

Os valores de $q^2_{y\text{rand}}$ foram menores que 0,2 e os valores de $r^2_{y\text{rand}}$ não ultrapassaram 0,3, o que indica correlação ao acaso insignificante para o modelo 1.

Tabela 34. Valores de $q^2_{y\text{rand}}$ e $r^2_{y\text{rand}}$ para 30 aleatorizações dos valores de y para o modelo 1

y_{rand}	$r^2_{y\text{rand}}$	$q^2_{y\text{rand}}$
y_1	0,0	-0,2
y_2	0,1	-0,1
y_3	0,1	0,0
y_4	0,2	0,0
y_5	0,2	0,0
y_6	0,1	-0,1
y_7	0,2	0,0
y_8	0,1	-0,1
y_9	0,4	0,1
y_{10}	0,1	-0,1
y_{11}	0,1	-0,1
y_{12}	0,0	-0,2
y_{13}	0,2	0,0
y_{14}	0,2	0,0
y_{15}	0,1	0,0
y_{16}	0,2	0,0
y_{17}	0,1	0,0
y_{18}	0,2	0,0
y_{19}	0,0	-0,4
y_{20}	0,1	-0,1
y_{21}	0,2	0,0
y_{22}	0,3	0,0
y_{23}	0,2	0,0
y_{24}	0,1	-0,3

Tabela 34 (Cont.). Valores de $q^2_{y_{rand}}$ e $r^2_{y_{rand}}$ para 30 aleatorizações dos valores de y para o modelo 1

y_{rand}	$r^2_{y_{rand}}$	$q^2_{y_{rand}}$
y_{25}	0,2	0,0
y_{26}	0,1	0,0
y_{27}	0,2	0,0
y_{28}	0,1	0,0
y_{29}	0,2	0,0
y_{30}	0,1	0,0
y_{modelo}	0,6	0,7

A frequência de correlação ao acaso pode ser visualizada na figura 42, que apresenta o gráfico com os valores de $q^2_{y_{rand}}$ versus $r^2_{y_{rand}}$, incluindo os valores das medidas estatísticas do modelo real, q^2_{LOO} e r^2_{LOO} . Verifica-se que todos os valores de $q^2_{y_{rand}}$ e $r^2_{y_{rand}}$ estão distantes dos valores de q^2 e r^2 do modelo real (modelo 1).

Figura 42. Visualização gráfica dos valores de $q^2_{y_{rand}}$ versus $r^2_{y_{rand}}$ para 30 aleatorizações dos valores de y no modelo 1. O dado destacado com retângulo pontilhado amarelo se refere ao modelo real.

O modelo 1 se apresentou robusto para o procedimento de validação cruzada LNO até a retirada de 25,7 % das amostras do conjunto de treinamento e não demonstrou chance de correlação no resultado da aleatorização dos valores da variável dependente.

Na validação externa (tabela 35), os valores dos descritores $Carga_{C_2}$, θ_4 , E_{SOLV} , E_{EL} e E_{EST} calculados para os compostos do conjunto de avaliação foram substituídos no modelo 1 e avaliou-se a diferença do dado biológico experimental com o previsto pelo modelo, dado demonstrado na figura 43. As amostras que apresentaram valor de resíduo menor ou igual ao valor de um desvio-padrão da média do modelo foram consideradas bem previstas pelo modelo. Conforme a tabela 19, o poder de previsão do modelo 1 foi de 70%.

Tabela 35. Validação externa do conjunto de avaliação utilizando o modelo 1

Amostra	pCl ₅₀ exp	pCl ₅₀ prev	resíduos
TBP21	6,41	5,57	0,84
TBP40	6,21	5,82	0,39
TBP28	6,13	6,01	0,12
TBP43	5,89	5,86	0,03
TBP33	5,86	5,51	0,35
TBP25	5,80	5,56	0,24
TBP10	5,26	5,15	0,11
TBP1	4,97	4,99	-0,02
TBP45	4,87	4,69	0,18
TBP26	4,83	4,93	-0,10
Valor do desvio-padrão		(DP)	0,27

Figura 43. Visualização gráfica dos valores de pCl₅₀ experimental e pCl₅₀ previsto pelo modelo 1. Os compostos do conjunto de treinamento estão representados em losangos preenchidos em cinza escuro enquanto que os compostos do conjunto de avaliação estão apresentados em losangos preenchidos em roxo.

Além dos procedimentos de validação, a capacidade de previsão do modelo final para outra classe de compostos também foi testada. Para realizar a previsão de uma classe não análoga às pirrolidina carboxamidas, o modelo selecionado com os coeficientes não autoescalados (eq.10), apresentado a seguir foi utilizado para o cálculo dos valores de pCl₅₀.

$$\begin{aligned} \text{pCl}_{50} = & 6,30 - 0,01 * (\text{E}_{\text{EST}} - 17,49)^2 - 0,01 * (\text{E}_{\text{EL}} + 7,50)^2 - 0,004 * \Theta_4 \\ & + 76,37 * (\text{Carga_C}_2 - 0,72)^2 - 0,06 * (\text{E}_{\text{SOLV}} + 3,77)^2 \end{aligned}$$

(eq.10)

Utilizaram-se 24 arilamidas (Quadros A e B – Anexos) (HE, *et. al.*, 2007) das quais 19 apresentaram valores de resíduo menores que o valor do desvio-padrão médio (tabela 36). Portanto, poder de previsão do modelo 1 foi de 79%, indicando que o modelo foi capaz de prever a atividade de compostos não congêneres ao conjunto de treinamento, mas que apresentam o mesmo modo de interação com a InhA.

Tabela 36. Previsão dos valores de pCl_{50} para a classe de arilamidas com o modelo 1

Amostra	pCl_{50} exp	pCl_{50} prev	resíduos
TBA24	5,69	5,88	-0,19
TBA8	5,73	4,72	1,01
TBA21	6,70	5,84	0,86
TBA15	5,11	9,16	-4,05
TBA2	4,78	4,74	0,04
TBA22	5,98	6,05	-0,06
TBA16	4,85	5,87	-1,02
TBA19	6,40	4,22	2,18
TBA10	5,01	4,41	0,60
TBA7	6,00	4,33	1,67
TBA4	5,51	4,15	1,36
TBA3	5,20	4,38	0,83
TBA11	5,17	4,32	0,85
TBA14	4,50	4,04	0,46
TBA6	4,81	4,50	0,31
TBA12	5,22	4,52	0,70
TBA23	5,72	4,99	0,73
TBA9	4,86	4,79	0,07
TBA5	5,03	4,24	0,79
TBA20	7,05	4,53	2,51
TBA18	5,13	6,06	-0,93
TBA13	4,75	5,09	-0,33
TBA17	5,29	6,50	-1,21
TBA1	4,41	3,72	0,69
Valor do desvio padrão		(DP)	1,30

4.3. Interpretação do modelo final da análise multivariada

O modelo mais promissor na análise QSAR multivariada foi obtido com a tabela D (descritores eletrônicos calculados com o método HF/6-31G*).

A representação da interpretação das relações entre estrutura química e atividade biológica para o modelo 1 (eq.9) podem ser visualizadas na figura 44.

$$pCl_{50} = 6,24 + 0,19 (Carga_C_2 + 0,72)^2 - 0,31 \theta_4 - 0,16 (E_{SOLV} + 1,01)^2 - 0,11 (E_{EL} + 0,74)^2 - 0,23 (E_{EST} - 1,34)^2$$

(eq.9)

Figura 44. Representação esquemática para a interpretação do modelo 1 gerado com a tabela D (HF/6-31G*).

A principal região da molécula responsável pelo modo de interação com o alvo biológico é a carbonila do anel pirrolidínico, que estabelece ligação de hidrogênio (átomo de oxigênio da carbonila) com o cofator NAD da InhA, inibindo assim a enzima. A diferença de potencial eletrostático entre o oxigênio e o C₂ (carbonila) no anel pirrolidínico, no modelo gerado, estaria representada pelo descritor Carga_C₂. Na verdade, o modelo QSAR gerado mascarou a importância da carga do oxigênio. Já o diedro θ_4 seria responsável pela orientação da porção mais volumosa da molécula no sítio ativo, representada pelos substituintes R₁ e R₂.

Dentre as contribuições termodinâmicas do modelo final estão as contribuições de energia de deformação axial ou de estiramento, de solvatação e de carga.

A contribuição negativa de energia de solvatação poderia estar relacionada à capacidade do ligante de estabelecer ligação de hidrogênio com resíduos de aminoácido complementares no sítio ativo do alvo. Na avaliação da estrutura cristalografada do complexo enzima-NAD-ligante TBP1 (2h7i), observaram-se dois pontos importantes de ligação de H no sítio de interação: com a porção nicotinamida do NAD (cofator) e com o resíduo de aa Tyr158 da enzima InhA. Interações similares foram observadas para os complexos 2h7l, 2h7m, 2h7n e 2h7p.

A contribuição de energia eletrostática (de carga) pode ser expressa com a visualização do mapa de potencial eletrostático (MPE) (programa GaussView 3.0). Na figura 45 estão apresentados os MPE's do ligante mais ativo (TBP39) e menos ativo (TBP14) do conjunto de treinamento, respectivamente. Regiões em vermelho intenso indicam maior distribuição de densidade eletrônica, enquanto que regiões em azul expressam regiões de menor distribuição de densidade eletrônica.

Figura 45. Mapas de potencial eletrostático do ligante mais ativo (TBP39) e menos ativo (TBP14), calculados com método HF/6-31G* (programa GaussView 3.0).

Comparando-se os ligantes TPB39 (mais ativo) e TBP14 (menos ativo), observam-se diferenças quanto às características eletrônicas. A carbonila do anel pirrolidínico, ponto de interação de ligação de hidrogênio com o resíduo de aa Tyr156 da InhA e com a porção nicotinamida do NAD, no ligante TBP39 apresenta menor densidade eletrônica (vermelho menos intenso) que no ligante TBP14. Por outro lado, esta porção do ligante TPB39 parece estar menos estericamente impedida.

Verifica-se importante diferença entre os substituintes do anel aromático: a molécula TBP39 apresenta substituinte volumoso que interagiria em um bolso/região auxiliar no sítio ativo da enzima InhA, conforme demonstra figura 46, enquanto que a molécula menos ativa (TBP14) possui substituinte menos volumoso.

Figura 46. Esquema das relações observadas e sugeridas entre os dados obtidos com os mapas de potencial eletrostático e as interações entre análogo pirrolidina carboxamida e resíduos de aa do sítio ativo da InhA [Fonte: Banco de dados PDBsum (LASKOWSKI, 2007), código 2h7l (HE *et. al.*, 2006).

4.4. Proposição de novas estruturas

Com base nas propriedades moleculares quantificadas pelo modelo 1 interpretado no item 4.3., propuseram-se novas estruturas químicas. Tiveram os valores de atividade biológica previstas conforme a tabela 37.

Tabela 37. Previsão dos valores de pCl_{50} com o modelo 1 para os compostos propostos.

Amostra	Fórmula molecular	pCl_{50} prev
PT4	$C_{22}H_{18}O_2N_3$	6,58
PT5	$C_{30}H_{33}O_3N_3$	5,15
PT6	$C_{19}H_{25}O_4N_3$	5,48
PT8	$C_{31}H_{23}O_3N_3$	5,80
PT9	$C_{35}H_{25}O_3N_3$	5,60
PT10	$C_{26}H_{18}O_2N_3$	5,46
TBP39		6,85
TBP14		4,13

Foram propostas seis moléculas que apresentaram atividade prevista intermediária (entre pCl_{50} 5,0 e 6,0). O composto PT4 apresentou atividade prevista considerada alta, pCl_{50} de 6, aproximou-se do composto mais ativo da classe pirrolidina carboxamidas, o composto TBP39 com pCl_{50} de 6,85.

CONCLUSÕES E PERSPECTIVAS

5. CONCLUSÕES E PERSPECTIVAS

No desenvolvimento da análise QSAR multivariada preliminar, descritores diversos foram calculados considerando as conformações selecionadas dos PAC e métodos distintos foram empregados nos cálculos de cargas de potencial eletrostático (Tabela A, B, C e D). Os dados da Tabela D (descritores eletrônicos calculados com método HF/6-31G*) geraram os melhores modelos preliminares.

Nas validações preliminares dos modelos QSAR com todos os compostos, aplicaram-se os procedimentos de validação cruzada LNO e aleatorização dos valores de **y** (*y-randomization*) que indicaram a obtenção de modelos robustos e sem correlação ao acaso. Estes resultados foram bons prognósticos para a obtenção do modelo final.

A tabela D foi a que gerou o melhor modelo QSAR final ($n = 35$; 5 descritores; $q^2 = 0,61$; $r^2 = 0,71$; $LSE = 0,16$; $LOF = 0,32$; *outliers* = 0). Descritores termodinâmicos (contribuições de energia de deformação axial ou de estiramento, eletrostática e de solvatação), eletrônico (carga de potencial eletrostático (CHELPG) no carbono C₂) e estrutural [diedro Θ_4 (C-O-N-R₂)] apresentaram a melhor correlação com a atividade inibitória sobre a enzima InhA. Ainda, o modelo QSAR final permitiu a interpretação química e biológica a partir da observação dos descritores presentes no modelo que serão úteis na proposição de novos tuberculostáticos potenciais.

O modelo selecionado como o melhor modelo apresentou-se robusto e sem chance de correlação de acordo com os procedimentos de validação LNO e aleatorização dos valores de **y**. A capacidade de previsão externa do modelo final foi de 70% (conjunto de avaliação - validação externa). Além disso, o modelo final foi capaz de prever a atividade de 79% dos ligantes de um conjunto de arilamidas, classe não congênere.

Apesar de uma das principais limitações da abordagem multivariada ser o fato de negligenciar a estrutura tridimensional dos compostos no processo de cálculo de alguns descritores, o modelo de QSAR obtido e validado por diferentes métodos estatísticos, apresentou capacidade importante de previsão de estruturas não congêneres. Portanto, este modelo teórico pode ser utilizado como ferramenta para o planejamento virtual de novos protótipos inibidores da enzima InhA.

Como perspectiva, pretende-se finalizar a análise QSAR-4D IR, em fase de desenvolvimento, a fim de explorar mais o farmacóforo e a orientação do conjunto de pirrolidina-carboxamidas no sítio de interação da InhA. Desta forma, a proposição de novos protótipos seria facilitada, pois nesta abordagem os descritores são medidas de ocupação que informam a posição e a natureza do provável grupo substituinte no ligante que auxiliaria ou não a resposta biológica esperada.

REFERÊNCIAS BIBLIOGRÁFICAS

ADMET/MI-QSAR Version 1.0 User's Guide; The Chem21 Group Inc. Chicago, 2001.

ANDRADE, C.H., PASQUALOTO, K.F.M., ZAIM, M.H., FERREIRA, E.I. Abordagem racional no planejamento de novos tuberculostáticos: Inibidores da InhA, enoil-ACP redutase do *M. tuberculosis*. *Rev. Bras. Cienc. Farm.*, v. 44, p. 167-179, 2008.

ASSELINEAU, C., ASSELINEAU, G., LANÉELLE, G., LANÉELLE, M.A. The biosynthesis of mycolic acids by Mycobacteria: current and alternative hypotheses. *Prog. Lipid. Res.*, v. 41, p. 501-523, 2002.

BALABAN, A. T. Distance Connectivity Index. *Chem. Phys. Lett.*, v. 89, p. 399-404, 1982.

BALDOCK, C., de BOER, G.J. de, RAFFERTY, J.B., STUITJE, A.R., RICE, D.W. Mechanism of action of diazaborines. *Biochem. Pharmacol.*, v. 55, p. 1541-1549, 1998.

BALDOCK, C., RAFFERTY, J.B., SEDELNIKOVA, S.E., BAKER, P.J., STUITJE, A. R., SLABAS, A.R., HAWKES, T.R., RICE, D.W. A mechanism of drug action revealed by structural studies of enoyl reductase. *Science*, v. 274, p. 2107-2110, 1996.

BANERJEE, A., DUBNAU, E., QUEMARD, A, BALASUBRAMANIAN, V. *et. al.* InhA, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis*. *Science*. v. 263, p. 227-230, 1994.

BARRY, C.E. III, LEE, R.E., MDLULI, K., SAMPSON, A.E., SCHROEDER, B.G., SLAYDEN, R.A., YUAN, Y. Mycolic acids: structure, biosynthesis and physiological functions. *Prog. Lipid Res.*, v. 37, p. 143-179, 1998.

BARRY, C.E. III. New horizons in the treatment of tuberculosis. *Biochem. Pharmacol.*, v. 54, p. 1165-1172, 1997.

BASAK, S. C., BALABAN, A. T., GRUNWALD, G. D., GUTE, B. D. Topological Indices: their nature and mutual relatedness. *J. Chem. Inf. Comput. Sci.*, V. 40, P. 891-898, 2000.

-
- BECKE, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A.*, v. 38, p. 3098-3100, 1988.
- BERMAN, H.M., WESTBROOK, J., FENG, Z., GILLILAND, G., BHAT, T. N., WEISSIG, H., SHINDYALOV, I. N., BOURNE, P.E. The Protein Data Bank. *Nuc. Acids Res.*, v. 28, P. 235-242, 2000.
- BRENEMAN, C. M., WIBERG, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.*, v. 11, n. 3, p. 361-373, 1990.
- BHARGAVA, H.N., LEONARD, P.A. Triclosan: applications and safety. *Am. J. Infect. Control.*, v. 24, p. 209-218, 1996.
- BHATT, A., MOLLE, V., BESRA, G.S., JACOBS, W.R., KREMER, L. The *Mycobacterium tuberculosis* FAS-II condensing enzymes: their role in mycolic acid biosynthesis, acid-fastness, pathogenesis and in future drug development. *Mol. Microbiol.*, v. 64, n. 6, p. 1442-1454, 2007.
- BLOCH, K. Control mechanisms for fatty acid synthesis in *Mycobacterium smegmatis*. *Adv. Enzymol. Relat. Areas Mol. Biol.*, v. 45, p. 1-84, 1977.
- BODOR, N., GABANYI, Z., WONG, C. K. A new method for the estimation of partition coefficient. *J. Am. Chem. Soc.*, v. 111, p. 3783-3786, 1989.
- BRENAN, P.J., NIKAIIDO, H. The envelope of mycobacteria. *Annu. Rev. Biochem.*, v. 64, p. 29-63, 1995.
- BRENEMAN, C. M., WIBERG, K. B. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.*, v. 11, p. 361-373, 1990.
- BROWN, E.D., WRIGHT, G.D. New targets and screening approaches in antimicrobial drug discovery. *Chem Rev.*, v. 105, p. 759-774, 2005.
- BUNIN, B. A., SIESEL, B., MORALES, G. A., BAJORATH, J. Chemoinformatics theory. In: *Chemoinformatics: Theory, practice and products*. Ed. Springer Holanda, 2007, cap.1.9, p. 31-35.

-
- CAMPBELL, J.W., CRONAN, J.E.Jr. Bacterial fatty acid biosynthesis: targets for antibacterial drug discovery. *Annu. Rev. Microbiol.*, v. 55, p. 305-332, 2001.
- CAO, C., LIU, L. Topological steric effect index and its application. *J. Chem. Inf. Comput. Sci.*, v. 44, n. 2, p. 678 - 687, 2004.
- COHEN, N.C. *Guidebook on molecular modeling in drug design*. San Diego: Academic Press, 1996. 361p.
- DEWAR, M. J. S.; ZOEBISCH, E. G.; HEALLY, E. F.; STEWART, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.*, v.107, p.3902-3909, 1985.
- DOHERTY, D. *MOLSIM: Molecular Mechanics and Dynamics Simulation Software*, version 3.0 – User's Guide. The Chem21 Group Inc., 1780 Wilson Dr., Lake Forest, IL, 60045, EUA, 1994.
- DUARTE, H. A., ROCHA, W. R. Teoria do funcional de densidade. In: MORGON, N. H., COUTINHO, K. *Métodos de química teórica e modelagem molecular*. Ed. Livraria da Física, São Paulo. 2007. Cap. 3, 2007, p. 73-111.
- DUBNAU, E., CHAN, J., RAYNAUD, C., MOHAN, V.P., LANEELLE, M.A., et al. Oxygenated mycolic acids are necessary for virulence of *Mycobacterium tuberculosis* in mice. *Mol. Microbiol.*, v.36, p.360 -637, 2000.
- DYE, C., S. SCHEELE, P. DOLIN, V. PATHANIA, M. C. RAVIGLIONE. Consensus statement. Global burden of tuberculosis: estimated incidence, prevalence, and mortality by country. WHO Global Surveillance and Monitoring Project. *JAMA*, v. 282, p. 677-686, 1999.
- ERIKSSON, L., JAWORSKA, J., WORTH, A. P., CRONIN, M. T. D., McDOWELL, R. M. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environ. Health Perspect.*, v. 111, p. 1361-1375, 2003.
- ERTL, P., ROHDE, B., SELZER, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.*, v. 43, n. 20, p. 3714-3717, 2000.

-
- FERRARA, P., APOSTOLAKIS, J., CAFLISCH, A. Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins*, v. 46, p. 24-33, 2002.
- FERREIRA, M.M.C. Multivariate QSAR. *J. Braz. Chem. Soc.*, v. 13, p. 742-753, 2002a.
- FERREIRA, M. M. C. Seleção de variáveis em QSAR. *Quim. Nova*, v. 25, p. 439-448, 2002b.
- FOCK, V. A. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Z. Phys.* v. 61, p. 126-148, 1930.
- GAUDIO, A. C. Modelos clássicos de estudo quantitativo das relações entre estrutura química e atividade biológica. *Quim. Nova*, v. 19, p. 278-289, 1996.
- GAUDIO, A. C., ZANDONADE, E. Proposição, validação e análise dos modelos que correlacionam estrutura química e atividade biológica. *Quim. Nova.*, v. 24, n. 5, p. 658-671, 2001.
- Gaussian 03, Revision C.02, FRISCH, M. J., TRUCKS, G. W., SCHLEGEL, H. B., SCUSERIA, G. E., ROBB, M. A., CHEESEMAN, J. R., MONTGOMERY, JR., J. A., VREVEN, T., KUDIN, K. N., BURANT, J. C., MILLAM, J. M., IYENGAR, S. S., TOMASI, J., BARONE, V., MENNUCCI, B., COSSI, M., SCALMANI, G., REGA, N., PETERSSON, G. A., NAKATSUJI, H., HADA, M., EHARA, M., TOYOTA, K., FUKUDA, R., HASEGAWA, J., ISHIDA, M., NAKAJIMA, T., HONDA, Y., KITAO, O., NAKAI, H., KLENE, M., LI, X., KNOX, J. E., HRATCHIAN, H. P., CROSS, J. B., BAKKEN, V., ADAMO, C., JARAMILLO, J., GOMPERTS, R., STRATMANN, R. E., YAZYEV, O., AUSTIN, A. J., CAMMI, R.; POMELLI, C., OCHTERSKI, J. W., AYALA, P. Y., MOROKUMA, K., VOTH, G. A., SALVADOR, P., DANNENBERG, J. J., ZAKRZEWSKI, V. G., DAPPRICH, S., DANIELS, A. D., STRAIN, M. C., FARKAS, O., MALICK, D. K., RABUCK, A. D., RAGHAVACHARI, K., FORESMAN, J. B., ORTIZ, J. V., CUI, Q., BABOUL, A. G., CLIFFORD, S., CIOŚŁOWSKI, J., STEFANOV, B. B., LIU, G., LIASHENKO, A., PISKORZ, P., KOMAROMI, I., MARTIN, R. L., FOX, D. J., KEITH, T., AL-LAHAM, M. A., PENG, C. Y., NANAYAKKARA, A., CHALLACOMBE, M., GILL, P. M. W., JOHNSON, B., CHEN, W., WONG, M. W., GONZALEZ, C., POPLE, J. A. Gaussian, Inc., Wallingford CT, 2004.
- GHOSE, A. K., PRITCHETT, A., CRIPPEN, G. M. Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships III: Modeling hydrophobic interactions. *J. Comput. Chem.*, v. 9, p. 80-90, 1988.

-
- GLEN, R. C. A fast empirical method for the calculation of molecular polarizability. *J. Comput. Aid. Mol. Des.*, v. 8, p. 457-466, 1994.
- GLICKMAN, M.S., COX, J.S., JACOBS, W.R., Jr. A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of *Mycobacterium tuberculosis*. *Mol. Cell*, v. 5, p. 717-727, 2000.
- GLICKMAN, M.S., JACOBS, W.R., Jr. Microbial pathogenesis of *Mycobacterium tuberculosis*: dawn of a discipline. *Cell*, v. 104, p.477-485, 2001.
- GOLBRAIKH, A., TROPSHA, A. Beware of q²! *J. Mol. Graphics Mod.*, v. 20, p. 269-276, 2002.
- GRAMATICA, P. principles of QSAR models validation: internal and external. *QSAR Comb. Sci.*, v. 26, p. 694-701, 2007.
- Guidance Document an the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models. n. 69, OECD: Paris, 2007.
<http://www.oecd.org/dataoecd/55/35/38130292.pdf>
- HANSCH, C., LEO, A. *Exploring QSAR: fundamentals and applications in chemistry and biology*. American Chemical Society: Washington, D.C., 1995.
- HE, X., ALIAN, A., STROUD, R., MONTELLANO, P.R.O. Pyrrolidine Caroxamides as a novel class of inhibitors of enoyl acyl carrier protein reductase from *Mycobacterium tuberculosis*. *J. Med. Chem.*, v. 49, p. 6308-6323, 2006.
- HEATH, R.J., WHITE, S.W., ROCK, C.O. Inhibitors of fatty acid synthesis as antimicrobial chemotherapeutics. *Appl. Microbiol. Biotechnol.*, v. 58, p. 695-703, 2002.
- HEATH, R.J., WHITE, S.W., ROCK, C.O. Lipid biosynthesis as a target for antibacterial agents. *Prog. Lipid. Res.*, v. 40, p. 467-497, 2001.
- HEATH, R.J., YU, Y.T., SHAPIRO, M.A., OLSON, E., ROCK, C.O. Broad spectrum antimicrobial biocides target the FabI component of fatty acid synthesis. *J. Biol. Chem.*, v. 273, p. 30316-30320, 1998.

HEATH, R.J., ROCK, C.O. Enoyl-acyl carrier protein reductase (fabI) plays a determinant role in completing cycles of fatty acid elongation in *Escherichia coli*. *J. Biol. Chem.*, v. 270, n. 44, p. 26538-26542, 1995.

HOPFINGER, A.J. 4D-QSAR Software, v. 3.0 – User's Manual. The Chem21 Group Inc., 1780 Wilson Dr., Lake Forest, IL, 60045, EUA, 1999.

HOPFINGER, A.J., WANG, S., TOKARSKI, J.S., JIN, B., ALBUQUERQUE, M., MADHAV, P.J., DURAISWAMI, C. Construction of 3D-QSAR models using the 4D-QSAR analysis formalism. *J. Am. Chem. Soc.*, v.119, p. 10509-10524, 1997.

HyperChem Program Release 7.51 for Windows; Hypercube, Inc., Gainesville, FL 2002.

JANIN, Y.L. Antituberculosis drugs: ten years of research. *Bioorg. Med. Chem.*, v. 15, p. 2479-2513, 2007.

KHADIKAR, P. V., DESHPANDE, N. V., KALE, P. P., DOBRYNIN, A., GUTMAN, I., DOMOTOR, G. The Szeged Index and an Analogy with the Wiener Index. *J. Chem. Inf. Comput. Sci.*, v. 35, n. 3, p. 547-550, 1995.

KIRALJ, R., FERREIRA, M. M. C. Basic validation procedures for regression models in QSAR and QSPR studies: theory and applications. *J. Braz. Chem. Soc.*, v. 20, n. 4, p. 770-787, 2009.

KOROLKOVAS, A. *Química farmacêutica*. Rio de Janeiro: Guanabara Koogan, cap. 2, 1982. p. 223-248.

KUBINYI, H. QSAR and 3D QSAR in drug design Part 1: methodology. *Drug Discov. Today*, v. 2, p. 457-467, 1997.

KUBINYI, H. QSAR and 3D QSAR in drug design. Part 2: Applications and problems. *Drug Discov. Today*, v.2, p. 538-546, 1997.

KUBINYI, H. Methods and principles in medicinal chemistry. In: MANHOLD, R., LARSIN, K. P., TIMMERMAN, H. eds. *QSAR: Hansch analysis and related approaches*. Weinheim: VCH, vol. 1, 1993, 240p.

-
- KUMAR, A., SIDDIQI, M. I. CoMFA based *de novo* design of Pyrrolidine Carboxamides as inhibitors of Enoyl Acyl Carrier Protein Reductase from *Mycobacterium tuberculosis*. *J. Mol. Model.*, v.14, p. 923-935, 2008.
- KUO, M.R., MORBIDONI, H.R., ALLAND, D. Targeting tuberculosis and malaria through inhibition of enoyl reductase: compound activity and structural data. *J. Biol. Chem.*, v. 278, p. 20851-20859, 2003.
- LASKOWSKI, R. A. Enhancing the functional annotation of PDB structures in PDBsum using key figures extracted from the literature. *Bioinformatics*, v. 23, p.1824-1827, 2007.
- LEACH, A.R. *Molecular Modelling: Principles and Applications*. 2 ed. Pearson Education, Londres, 2001, 768p.
- LEE, C., YANG, W., PARR, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B.*, v. 37, p. 785-789, 1988.
- LEE, R.E., BRENNAN, P.J., BESRA, G.S. *Mycobacterium tuberculosis* cell envelope. *Curr. Top. Microbiol. Immunol.*, v. 215, p. 1-27, 1996.
- LEVY, C.W., BALDOCK, C., WALLACE, A.J., SEDELNIKOVA, S., VINER, R.C., CLOUGH, J.M., STUITJE, A.R., SLABAS, A.R., RICE, D.W., RAFFERTY, J.B. A study of the structure-activity relationship for diazaborine inhibition of *Escherichia coli* enoyl-acp reductase. *J. Mol. Biol.*, v. 309, p. 171-180, 2001.
- MACARRON, R. Critical review of the role of HTS in drug discovery. *Drug Discov. Today*, v. 11, p. 277-279, 2006.
- MARRAKCHI, H., ZHANG, Y. M., ROCK, C. O. Mechanistic diversity and regulation of type II fatty acid synthesis. *Biochem. Soc. Trans.*, v. 30, p. 1050-1055, 2002.
- MARRAKCHI, H., LANÉELLE, G., QUÉMARD, A. InhA, a target of the antituberculosis drug isoniazid, is involved in a mycobacterial fatty acid elongation system, FAS II. *Microbiology*, v. 146, pt. 2, p. 289-296, 2000.
- MARTIN, Y. C. *Quantitative drug design: a critical introduction*. Marcel Dekker: New York, 1978, 292p.

MARTÍNEZ, L., BORIN, I.A., SKAF, M.S. Fundamentos de simulação por dinâmica molecular. In: MORGON, N. H., COUTINHO, K., Eds. *Métodos de química teórica e modelagem molecular*, Ed. Livraria da Física, São Paulo. 2007. Cap. 12, p. 413-451.

MATLAB versão 6.3.0 para Windows; The MathWors, Inc., 1984 – 2011.

Marvin Beans versão 5.2.1. para windows; ChemAxon Ltd. 1999-2010.

McCARTHY, A. D., HARDIE, D. G. Fatty acid synthase – an example of protein evolution by gene fusion. *Trens. Biochem. Sci.*, v. 9, n. 2, p. 60-65, 1984.

McMURRY, L.M., OETHINGER, M., LEVY, S.B. Triclosan targets lipid synthesis. *Nature*, v. 394, p. 531-532, 1998.

MILLER, K. J. Additivity methods in molecular polarizability. *J. Am. Chem. Soc.*, v. 112, n. 23, p. 8533–8542, 1990.

MINISTÉRIO DA SAÚDE, Tratamento Diretamente Observado (TOD) da Tuberculose na Atenção Básica. 2011, Brasília – DF. (2011a)

MINISTÉRIO DA SAÚDE, em 03/09/2011, (2011b)
http://portal.saude.gov.br/portal/saude/profissional/visualizar_texto.cfm?idtxt=31109

MINISTÉRIO DA SAÚDE, em 20/06/2010,
http://portal.saude.gov.br/portal/saude/profissional/area.cfm?id_area=1527

MINISTÉRIO DA SAÚDE, *Programa Nacional de Controle da Tuberculose*, em 20/06/2010, (2010b)
http://portal.saude.gov.br/portal/saude/profissional/visualizar_texto.cfm?idtxt=31109.

MINISTÉRIO DA SAÚDE, em 27/04/2009,
http://portal.saude.gov.br/portal/saude/profissional/visualizar_texto.cfm?idtxt=25340.

-
- MORDE, V. A., SHAIKH, M. S., PISSURLINKAR, R. R. S., COUTINHO, E. C. Molecular modeling studies, synthesis, and biological evaluation of Plasmodium falciparum enoyl-acyl carrier reductase (PfENR) inhibitors. *Mol. Divers.*, v. 13, p. 501-517, 2009.
- MORGON, N. H., CUSTÓDIO, R. Teoria do funcional de densidade. *Quím. Nova*, v. 18, n. 1, p. 44-55, 1995.
- OHLSTEIN, E.H., RUFFOLO, R. R., ELLIOTT, J. D. Drug discovery in next millennium. *Annu. Rev. Pharmacol. Toxicol.*, v.40, p. 177-191, 2000.
- OOMS, F. Molecular modeling and computer aided drug design. Examples of their applications in medicinal chemistry. *Curr. Med. Chem.*, v. 7, p. 141-158, 2000.
- PARIKH, S.L., XIAO, G., TONGE, P.J. Inhibition of InhA, the enoyl reductase from *Mycobacterium tuberculosis*, by triclosan and isoniazid. *Biochemistry*, v. 39, p. 7645-7650, 2000.
- PARIKH, S., MOYNIHAN, D.P., XIAO, G., TONGE, P.J. Roles of tyrosine 158 and lysine 165 in the catalytic mechanism of InhA, the enoyl-acyl carrier protein reductase from *Mycobacterium tuberculosis*. *Biochemistry*, v. 38, p. 13623-13634, 1999.
- PASQUALOTO, K.F.M., FERREIRA, E.I, SANTOS-FILHO, O.A., HOPFINGER, A.J. Rational design of new antituberculosis agents: Receptor-independent four-dimensional quantitative structure-activity relationship analysis of a set of isoniazid derivatives. *J. Med. Chem.*, v. 47, p. 3755-3764, 2004.
- PASQUALOTO, K.F.M. *Planejamento racional de tuberculostáticos com base na estrutura dos ácidos micólicos da parede celular do Mycobacterium tuberculosis*. 122p. Faculdade de Ciências Farmacêuticas, USP, São Paulo, 2003. [Tese de doutorado].
- PASQUALOTO, K.F.M., FERREIRA, E.I. An approach for the rational design of new antituberculosis agents. *Curr. Drug Targets*, v. 2, p. 427-437, 2001.
- PEROZZO, R., KUO, M., SIDHU, A. S., VALIYAVEETIL, J. T., BITTMAN, R., JACOBS, Jr, W. R., FIDOCK, D. A., SACCHETTINI, J. C. Structural elucidation of the specificity of the antibacterial agent triclosan for malarial enoyl acyl carrier protein reductase. *J. Biol. Chem.*, v. 277, n.15, p. 13106-13114, 2002.

-
- PLATT, J. R. Influence of neighbor bonds on additive bond properties in paraffins, *J. Chem. Phys.*, v. 15, p. 419-420, 1947.
- PLATT, J. R. Prediction of isomeric differences in paraffin properties. *J. Phys. Chem.*, v. 56, p. 328-336, 1952.
- PLAVSIC, D., NIKOLIC, S., TRINAJSTIC, N., MIHALIC, Z. J. On the Harary Index for the characterization of chemical graphs. *Math. Chem.*, v. 12, p. 235-250, 1993.
- RANDIC, M. Novel molecular descriptor for structure-property studies. *Chem. Phys. Lett.*, v. 211, p. 478-483, 1993.
- ROCK, C.O., CRONAN, J.E. *Escherichia coli* as a model for the regulation of dissociable (type II) fatty acid biosynthesis. *Biochim. Biophys. Acta*, v.1302, p. 1-16, 1996.
- ROGERS, D. *WOLF Genetic Function Approximation*. Reference Manual, version 5.5; Molecular Simulation Inc.: Burlington, MA, 1994.
- ROGERS, D., HOPFINGER, A.J. Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships. *J. Chem. Inf. Comput. Sci.*, v. 34, p. 854-866, 1994.
- ROUJEINIKOVA, A., SEDELNIKOVA, S., de BOER, G.J., STUITJE, A.R., SLABAS, A.R., RAFFERTY, J.B., RICE, D.W. Inhibitor binding studies on enoyl reductase reveal conformational changes related to substrate recognition. *J. Biol. Chem.*, v. 274, p. 30811-30817, 1999.
- ROZWARSKI, D.A., VILCHÈZE, C., SUGANTINO, M., BITTMAN, R., SACCHETTINI, J.C. Crystal structure of the *Mycobacterium tuberculosis* enoyl-acyl reductase, InhA, in complex with NAD⁺ and C16 fatty acyl substrate. *J. Biol. Chem.*, v. 274, p. 15582-15589, 1999.
- ROZWARSKI, D.A., GRANT, G.A., BARTON, D.H.R., Jr. JACOBS, W.R., SACCHETTINI, J.C. Modification of the NADH of the isoniazid target (InhA) from *Mycobacterium tuberculosis*. *Science*, v. 279, p. 98-102, 1998.
- SAIER, Jr, M. H. Cell structure chemical composition: cell membrane prokaryotic. In: SCHAECHTER, M., Ed. *Encyclopedia of microbiology*, 3 ed., Ed. Elsevier, California, 2009. p. 956-971.

-
- SANT'ANNA, C.M. Glossário de termos usados no planejamento de fármacos (recomendações da IUPAC para 1997). *Quim. Nova*, v. 25, p. 505-512, 2002.
- SENSI, P., GRASSI, G.G. Antimycobacterial agents. In: BURGER, A., WOLFF, M.E. Eds. *Burger's medicinal chemistry and drug discovery*. 5. ed. New York: John Wiley & Sons, Inc., 1996. v. 2, chap. 34, p. 575-635.
- SIMAS, A. M., ROCHA, G. B. Métodos semi-empíricos de estrutura eletrônica em química quântica. In: MORGON, N. H., COUTINHO, K. *Métodos de química teórica e modelagem molecular*, São Paulo, Ed. Livraria da Física, Cap. 2, 2007, p. 29-71.
- SIVARAMAN, S., SULLIVAN, T. J., JOHNSON, F., NOVICHENOK, P., CUI, G., SIMMERLING, C., TONGE, P. J. Inhibition of the bacterial enoyl reductase FabI by triclosan: a structure – reactivity analysis of FabI inhibition by triclosan analogues. *J. Med. Chem.*, v. 47, n. 3, p. 509-518, 2004.
- STEWART, J. J. P. Optimization of parameters for semiempirical methods. *J. Comput. Chem.*, v. 10, p. 209-221, 1989.
- STEWART, M.J., PARIKH, S., XIAO, G., TONGE, P.J., KISKER, C. Structural basis and mechanism of enoyl reductase inhibition by triclosan. *J. Mol. Biol.*, v. 290, p. 859-865, 1999.
- SULLIVAN, T. J., TRUGLIO, J. J., BOYNE, M. E., NOVICHENOK, P., ZHANG, X., STRATTON, C., LI, H. J., KAUR, T., AMIN, A., JOHNSON, F., SLAYDEN, R. A., KISKER, C., TONGE, P. J. High affinity inhA inhibitors with activity against drug-resistant strains of *Mycobacterium tuberculosis*. *ACS. Chem. Biol.*, v. 1, p. 43-53, 2006.
- TAKAYAMA, k., WANG, C., BESRA, G.S. Pathway to synthesis and processing of mycolic acids in *Mycobacterium tuberculosis*. *Clin. Microbiol. Rev.*, v. 18, p. 81-101, 2005.
- TAVARES, L. C. QSAR: a abordagem de Hansch. *Quim. Nova*, v. 27, n. 4, p.631-639, 2004.
- TAVARES, L. C.; FERREIRA, E. I. Relações quantitativas estrutura-atividade. Fundamentos e aplicação da análise de Hansch. Ver editores, local de edição e editora *Prática de Química Farmacêutica e Medicinal*, 2002. p. 36-66

TOKARSKI, J. S., HOPFINGER, A. J. Constructing protein models for ligand-receptor binding thermodynamic simulations: an application to a set of peptidomimetic renin inhibitors. *J. Chem. Inf. Comput. Sci.*, v. 37, p. 779-791, 1997.

TROPSHA, A., GRAMATICA, P. GOMBAR, V. K. The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.*, v. 22, p. 69-77, 2003.

VEYRON-CHURLET, R., GUERRINI, O., MOUREY, L., DAFFÉ, M., ZERBIB, D. Protein-protein interactions within the Fatty Acid Synthase-II system of *Mycobacterium tuberculosis* are essential for mycobacterial viability. *Mol. Microbiol.*, v. 54, p. 1161-1172, 2004.

VILCHÈZE, C., MORBIDONI, H.R., WEISBROD, T.R., *et al.* Inactivation of inhA-encoded fatty acid synthase II (FAS II) enoyl-acyl carrier protein reductase induces accumulation of the FAS I end products and cell lyses of *Mycobacterium smegmatis*. *J. Bacteriol.*, v. 182, p. 4059-4067, 2000.

ViewerLite versão 5.0, for windows; Accelrys Inc. 2002.

WERMUTH, C.G. Strategies in the search for new lead compounds or original working hypotheses. In: _____ Ed. *The Practice of medicinal chemistry*. 2 ed. San Diego: Academic Press, 2003. Chap. 5, p. 69-89.

WERMUTH, C.G., GANELLIN, C.R., LINDBERG, P., MITSCHER, L.A. Glossary of terms used in medicinal chemistry. *Annu. Rep. Med. Chem.*, v. 33, p. 385-395, 1998.

WIENER, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.*, v. 69, n. 1, p. 17-20, 1947.

WOLD, S., ESBENSEN, K., GELADI, P. Principal component analysis. *Chemom. Intell. Lab. Syst.*, v. 2, p. 37-52, 1987.

WORLD HEALTH ORGANIZATION. *MDR – TB, 2011*. [https://extranet.who.int/sree/Reports?op=vs&path=/WHO_HQ_Reports/G2/PROD/EXT/MDRTB_Indicators].

WORLD HEALTH ORGANIZATION. *MDR – TB, 2010*. [http://www.who.int/tb/publications/global_report/2009/update/a-8_full.pdf].

WORLD HEALTH ORGANIZATION. *WHO TB burden estimates by country, 2009a*. [http://www.who.int/tb/country/data/download/en/index1.html].

WORLD HEALTH ORGANIZATION. *WHO TB country graphs, 2009b*. [http://www.who.int/tb/country/data/2009/bra.pdf].

WORLD HEALTH ORGANIZATION. *An expanded DOTS framework for effective tuberculosis control*. WHO/CDS/TB/2002.297. 2002c [www.who.int/gtb/dots/2002-297.htm].

WORLD HEALTH ORGANIZATION. *Country Profiles: Brazil*. Annual Report, 1997. [www.who.int/gtb/publications/tbrep_97/countries/brazil.htm].

WORLD HEALTH ORGANIZATION. *Global tuberculosis control: surveillance, planning, financing*. WHO Report 2002. WHO/CDS/TB/2002.295. 2002d [www.who.int/gtb/publications/globrep02/index.html].

WORLD HEALTH ORGANIZATION. Report of a WHO workshop on *Tuberculosis Control and Medical Schools*. Italy. WHO/TB/98.236. 1998 [www.who.int/gtb/publications/tb_control/index.htm].

WORLD HEALTH ORGANIZATION. *Tuberculosis*. Fact Sheet n°104 - Revised in August 2002. 2002a [www.who.int/mediacentre/factsheets/who104/en/index.htm].

WORLD HEALTH ORGANIZATION. *World TB day highlights report*. WHO/CDS/STB/2002.2002b[stoptb.org/world.tb.day/WTBD_2002/Final_Highlights_Report_2002.pdf].

YAUN, Y., BARRY, III, C. E. A common mechanism for the biosynthesis of methoxy and cyclopropyl mycolic acids in *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA*, v. 93, p. 12828-12833, 1996.

ANEXOS

Tabela A. Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método B3LYP/3-21G*) para os 46 compostos aleatorizados

Composto	E _{EST}	E _{ang}	E _{TORS}	E _{1,4}	E _{vdW}	E _{EL}	E _{SOLV}	E _{HBD}	E _{TOT}
TBP45	-1,139	-1,667	-2,233	-0,640	0,415	0,648	-1,021	0,059	-1,490
TBP46	-1,206	-1,703	-1,337	-0,647	0,372	0,396	-0,631	0,199	-1,361
TBP9	-0,267	-0,518	0,357	0,576	-0,326	-2,733	-2,796	0,448	-0,463
TBP39	1,660	1,039	-0,581	1,161	-5,276	-0,366	0,085	-1,395	1,508
TBP4	-0,953	-0,892	0,337	-0,631	0,046	0,450	0,110	0,125	-0,632
TBP5	-0,677	-0,615	0,056	-0,830	0,552	0,509	-0,272	0,115	-0,675
TBP42	1,427	1,538	0,386	2,225	-1,686	-1,497	1,415	0,720	1,866
TBP41	1,353	2,043	-0,815	0,841	-4,833	-0,161	0,101	0,006	1,661
TBP22	-0,552	-0,571	-0,035	0,098	0,650	-0,568	-0,852	0,084	-0,480
TBP7	0,039	-0,649	-0,528	-0,839	0,352	0,890	0,680	0,111	-0,514
TBP3	-0,542	-0,709	0,484	-0,674	0,748	0,521	0,307	0,119	-0,524
TBP26	-0,881	-0,586	0,043	-0,489	0,336	0,492	-0,489	0,116	-0,568
TBP23	-0,159	-0,044	0,739	-0,165	0,470	-0,608	-1,335	-0,211	-0,209
TBP21	-0,664	-0,880	0,767	-0,900	0,736	0,572	-0,619	0,089	-0,703
TBP29	-0,478	0,080	-0,445	0,032	1,085	0,879	0,427	0,126	-0,089
TBP11	-0,701	-0,167	0,030	-2,205	1,058	2,447	-1,533	-1,054	-0,973
TBP2	-0,722	-0,182	0,148	-0,519	-0,075	0,441	0,723	-6,016	-0,452
TBP32	1,159	1,406	2,187	1,602	0,591	-1,304	0,581	0,720	1,578
TBP15	-0,404	-0,571	0,019	-0,648	0,701	0,683	0,859	0,111	-0,455
TBP37	0,866	0,905	1,135	-0,210	-2,160	0,796	0,415	-0,040	0,967
TBP24	-0,515	-0,371	-0,091	0,027	0,297	-0,971	-1,885	0,075	-0,528
TBP8	-0,875	-0,556	0,284	-0,074	0,364	-0,201	-0,753	0,105	-0,464
TBP31	1,729	1,127	1,782	1,668	-0,420	-0,525	1,406	0,720	1,812
TBP28	0,651	-0,305	-1,812	1,283	0,376	-1,167	0,856	0,720	0,091
TBP14	-0,816	-1,024	-1,172	-0,548	0,431	0,418	0,138	0,080	-0,928
TBP34	0,226	1,342	-0,853	0,549	1,109	-0,761	0,972	0,372	0,449
TBP18	-0,765	-0,305	0,041	-0,570	0,760	0,438	0,639	0,047	-0,455
TBP20	-0,644	-0,588	0,233	-1,481	0,348	1,140	-0,642	0,074	-0,805
TBP6	0,085	-1,033	0,636	-0,851	0,415	0,643	0,148	0,126	-0,516
TBP33	0,357	0,647	-1,600	-0,424	1,575	0,151	0,668	0,103	-0,240
TBP10	0,048	-0,451	0,616	-0,299	0,803	0,387	1,352	0,103	-0,095
TBP30	0,334	0,178	-1,956	0,199	-0,079	-0,036	0,263	0,121	-0,112
TBP1	-0,783	-0,571	0,036	-0,695	0,305	0,705	0,172	0,124	-0,550
TBP25	-0,341	0,187	0,266	-0,492	1,034	-0,170	-0,518	-0,832	-0,338
TBP44	-1,346	-1,677	-1,589	-0,668	0,199	0,421	-1,215	0,212	-1,450
TBP35	1,633	1,489	2,065	1,695	1,920	-1,460	0,040	0,720	1,527
TBP19	-1,006	-0,067	-0,044	-0,131	-0,091	-2,347	-2,325	0,409	-0,745
TBP17	-0,374	-0,567	0,070	-0,813	0,697	0,970	1,453	0,134	-0,424
TBP27	-0,128	0,194	-0,160	-0,483	0,623	1,421	0,141	-0,175	-0,038
TBP13	-0,898	-0,434	-0,331	-0,579	0,697	0,489	-0,629	0,046	-0,667
TBP12	-0,990	-0,778	0,172	-0,444	0,756	0,461	-0,482	0,053	-0,665
TBP43	2,793	2,242	-1,036	1,096	-3,696	0,296	0,257	0,006	2,044
TBP38	0,357	1,926	1,151	1,541	-1,662	-0,605	1,443	0,720	1,774
TBP16	0,686	0,023	0,433	-0,498	0,391	0,600	1,319	0,067	0,157
TBP36	1,774	0,633	1,813	1,816	-0,032	-1,153	0,116	0,720	1,499
TBP40	1,650	1,483	0,331	2,035	-0,875	-1,631	0,907	0,720	1,672

Tabela A (cont.). Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método B3LYP/3-21G*) para os 46 compostos aleatorizados

Composto	E _{HOMO}	E _{LUMO}	μ_T	μ_x	μ_y	μ_z	carga_N ₁	carga_C ₂
TBP45	-1,333	0,435	0,777	-0,934	1,560	1,584	1,532	1,228
TBP46	-1,016	0,434	0,215	1,025	1,531	-0,635	0,180	0,126
TBP9	-0,608	-3,594	2,249	0,368	-2,050	-0,792	-0,983	0,580
TBP39	-0,339	-0,110	-1,722	-0,087	-0,546	-0,019	-0,103	-0,675
TBP4	-0,429	0,280	-0,044	-0,804	-0,933	0,459	0,039	-0,439
TBP5	-0,475	0,443	0,758	-0,565	-1,240	-0,840	1,540	-0,277
TBP42	3,227	-0,699	-1,005	-0,550	-0,475	1,120	0,316	-0,403
TBP41	-0,336	-0,543	-0,401	-0,765	-0,554	-0,605	-0,613	0,809
TBP22	-0,552	-0,580	-1,816	0,683	0,868	0,827	-1,502	0,977
TBP7	0,131	1,198	0,723	1,409	1,644	0,319	-0,392	0,320
TBP3	-0,287	0,434	0,878	1,052	1,825	-0,729	-0,739	0,964
TBP26	-0,636	0,145	0,450	0,565	-1,219	1,464	0,268	-0,571
TBP23	-0,247	-0,208	1,706	-1,292	-1,252	-0,699	-0,086	0,517
TBP21	-0,563	-0,073	-0,716	-0,421	1,172	-0,780	-0,823	0,779
TBP29	-0,129	0,771	1,110	1,566	-0,096	2,196	-1,722	1,121
TBP11	-0,316	0,450	-0,397	-0,990	0,067	-0,780	0,673	0,032
TBP2	0,170	-0,793	0,959	-1,274	-0,742	-0,802	0,190	0,110
TBP32	2,177	1,045	-0,651	-0,132	-0,346	-1,263	-0,217	-0,273
TBP15	-0,052	0,506	0,099	-0,885	-0,868	-0,254	0,396	0,282
TBP37	-0,535	-0,122	-1,925	0,308	-0,253	1,165	-0,951	-0,913
TBP24	-1,030	-0,969	-0,778	0,515	-0,867	0,866	-0,042	-0,608
TBP8	-0,880	-0,065	-2,308	-0,288	0,129	-0,349	-0,097	-0,676
TBP31	1,432	1,126	0,018	1,223	0,834	1,739	0,916	-1,474
TBP28	1,554	-1,587	1,466	1,372	2,065	0,273	-1,850	0,550
TBP14	0,306	-0,601	0,993	-1,634	-0,074	-0,553	1,962	-4,787
TBP34	0,195	0,050	1,397	1,782	1,311	-0,821	-0,577	0,532
TBP18	-0,497	0,634	-0,685	-0,992	-0,295	0,218	-0,137	0,356
TBP20	-0,491	0,298	-0,478	-0,623	-0,826	-0,073	-0,547	0,482
TBP6	-0,402	0,252	-0,137	1,095	-0,677	1,247	-0,620	0,505
TBP33	-0,114	-0,004	0,115	-1,235	-0,354	-0,318	0,963	-0,111
TBP10	-0,087	1,130	-0,117	1,252	-0,208	1,580	-0,305	-0,187
TBP30	-0,215	-0,429	0,987	1,282	0,495	-1,882	0,999	-0,354
TBP1	0,101	0,989	0,125	-0,961	-0,704	-0,540	-0,360	0,529
TBP25	-0,135	0,361	-0,121	-1,021	0,024	-0,965	0,881	-0,464
TBP44	-0,780	0,468	0,355	-0,709	-1,140	-0,243	-0,427	0,222
TBP35	0,146	0,818	-0,967	-0,026	0,042	1,972	-0,448	0,070
TBP19	-1,056	-3,098	-0,774	0,833	-0,489	1,335	0,175	-0,326
TBP17	0,029	1,253	1,029	-1,417	-0,790	-0,294	-0,151	1,070
TBP27	-0,374	0,921	-1,125	1,006	0,987	0,602	1,481	-1,086
TBP13	-0,418	-0,143	0,460	1,376	1,507	0,002	-0,139	-0,702
TBP12	-0,311	-0,064	-0,718	-0,910	-0,398	0,072	-0,195	0,315
TBP43	-0,348	-0,580	0,070	1,151	1,236	-0,805	1,188	1,356
TBP38	3,506	-0,923	-1,252	0,607	0,310	-1,051	-1,352	1,200
TBP16	0,360	1,197	0,664	-1,055	-0,456	-1,290	-0,605	-0,186
TBP36	0,071	0,758	-0,106	0,001	1,355	-1,264	-1,043	0,702
TBP40	1,588	-1,214	0,640	-0,900	-1,111	-0,392	3,325	-1,220

Tabela A (cont.). Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método B3LYP/3-21G*) para os 46 compostos aleatorizados

Composto	carga_C ₃	carga_C ₄	carga_C ₅	carga_O ₆	carga_H ₇	carga_C ₈	carga_O ₉	carga_N ₁₀
TBP45	-1,312	-0,844	-0,011	-1,124	1,034	0,776	-0,156	-0,470
TBP46	0,463	-0,321	1,330	-1,037	-0,425	-0,441	0,122	0,145
TBP9	-1,003	-0,274	1,130	-0,321	0,089	0,509	-0,055	-0,711
TBP39	1,272	-1,108	0,820	0,130	0,646	0,168	0,627	0,006
TBP4	0,378	0,538	-1,067	0,387	-0,657	0,630	-0,567	-0,812
TBP5	-0,875	0,928	-0,889	0,458	-0,766	-0,408	0,392	-0,131
TBP42	0,699	1,515	-1,688	-0,239	-1,182	-2,255	2,962	2,179
TBP41	0,273	0,609	-1,035	-0,924	0,663	-0,550	0,658	0,354
TBP22	-0,992	-0,009	0,338	-0,347	0,314	0,109	-0,157	-0,196
TBP7	0,111	0,020	0,076	-0,310	-0,101	-0,017	-0,117	-0,218
TBP3	-0,962	0,016	0,543	-0,472	-0,446	-0,185	0,155	0,082
TBP26	0,378	0,684	-0,735	0,358	-0,714	-0,186	0,012	-0,157
TBP23	-1,491	0,511	-0,352	0,184	-0,978	0,862	-0,586	-1,051
TBP21	-0,269	-0,490	0,298	-0,117	0,841	0,304	-0,188	-0,163
TBP29	-2,512	-0,517	1,712	-0,489	0,303	1,350	-1,714	-1,063
TBP11	0,193	0,454	-0,758	-0,335	0,223	-0,605	1,024	-0,030
TBP2	1,039	-1,316	0,150	-0,078	0,335	2,014	-1,419	-1,936
TBP32	1,785	-1,956	0,676	0,258	1,345	2,160	-1,201	-0,800
TBP15	-0,350	-0,065	-0,095	0,230	0,188	-0,018	-0,054	-0,246
TBP37	2,204	-2,160	1,336	0,046	1,725	1,410	0,044	-0,927
TBP24	0,956	-0,151	-0,044	0,535	-0,367	0,289	0,076	-0,740
TBP8	0,846	0,384	-0,699	0,567	-0,763	0,645	-0,782	-0,961
TBP31	1,936	-1,525	1,085	0,066	0,777	0,106	0,282	0,950
TBP28	-0,858	-0,145	2,533	-0,023	-1,181	-1,743	0,235	3,589
TBP14	0,442	1,749	-0,004	5,978	-0,716	-2,573	3,569	0,344
TBP34	-0,558	0,689	-0,283	-0,770	-1,205	-0,047	-0,997	-0,194
TBP18	-1,125	0,607	0,004	0,249	-0,551	-0,608	-0,179	0,484
TBP20	-0,171	-0,024	-0,079	-0,557	0,255	-0,025	-0,097	-0,066
TBP6	0,159	0,314	-0,080	-0,406	3,054	-0,270	-0,081	-0,076
TBP33	-0,424	0,915	-0,987	0,139	-1,107	-0,224	-0,257	0,070
TBP10	-0,204	-0,135	-0,444	0,214	0,406	1,440	-1,931	-1,389
TBP30	0,143	0,421	-1,280	-0,155	-0,225	0,379	-0,477	-0,380
TBP1	-0,378	0,484	-0,341	-0,367	-0,107	-0,099	0,021	-0,269
TBP25	0,051	1,127	-1,010	0,157	-1,237	-0,594	-0,045	0,244
TBP44	1,237	-0,157	1,022	-0,769	-0,045	-0,688	0,745	0,303
TBP35	1,562	-2,739	1,240	-0,278	2,265	2,441	-1,703	-0,901
TBP19	0,720	0,270	-0,372	0,037	-0,301	-0,752	1,127	0,124
TBP17	-1,107	0,913	-0,456	-0,268	-0,803	-0,568	0,403	0,060
TBP27	-0,349	1,562	-1,338	0,717	-1,524	-0,625	0,484	-0,297
TBP13	0,728	0,112	-0,188	0,494	-0,492	0,011	0,076	-0,345
TBP12	0,312	0,135	-0,422	-0,238	-0,111	0,143	-0,029	-0,426
TBP43	-1,727	1,018	-1,502	-0,288	-0,805	-0,567	0,463	0,463
TBP38	-0,992	-1,475	1,949	-0,658	1,651	-0,024	-0,803	2,207
TBP16	-0,365	1,351	-1,037	-0,010	-1,275	-0,077	-1,076	-0,200
TBP36	0,305	-1,487	1,866	-0,615	0,691	0,086	0,056	1,323
TBP40	-0,171	-0,429	-0,907	-0,008	1,279	-1,683	1,134	2,227

Tabela A (cont.). Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método B3LYP/3-21G*) para os 46 compostos aleatorizados

Composto	Vol. Vw	ClogP	RM	α	Área superf.	θ_1	θ_2	θ_3
TBP45	-1,013	-0,319	-0,805	-0,841	-0,988	-0,775	-0,409	-1,315
TBP46	-1,050	-0,461	-0,816	-0,828	-1,033	-0,342	-0,044	-0,805
TBP9	-0,816	-1,403	-0,766	-0,856	-0,661	-0,426	0,291	0,248
TBP39	1,502	1,768	1,625	1,545	1,786	0,082	-0,155	-1,478
TBP4	-0,767	-0,552	-0,750	-0,752	-0,741	0,263	0,285	0,317
TBP5	-0,887	-0,836	-0,901	-0,845	-0,833	0,189	0,285	0,317
TBP42	1,565	1,819	1,617	1,801	1,128	0,458	0,038	-0,497
TBP41	1,817	2,174	1,871	1,789	2,053	-1,848	-0,575	-2,006
TBP22	-0,382	0,340	-0,430	-0,546	-0,176	0,089	0,258	0,167
TBP7	-0,871	-0,886	-0,888	-0,857	-0,868	-0,729	-0,005	-0,635
TBP3	-0,779	-0,552	-0,750	-0,752	-0,721	0,774	0,250	0,172
TBP26	-0,859	-0,694	-0,889	-0,857	-0,850	0,421	0,244	0,325
TBP23	-0,287	-0,714	-0,492	-0,566	-0,145	-0,133	0,295	0,295
TBP21	-0,632	-0,309	-0,644	-0,589	-0,544	1,267	0,289	0,296
TBP29	0,420	0,755	0,434	0,424	0,807	-1,094	0,700	1,662
TBP11	-0,428	-1,869	-0,562	-0,518	-0,135	-1,743	-0,028	-1,049
TBP2	-0,426	-1,626	-0,541	-0,518	-0,664	0,434	0,281	0,318
TBP32	1,653	0,806	1,534	1,462	1,565	0,702	0,021	0,308
TBP15	-0,486	-0,086	-0,480	-0,508	-0,504	0,377	0,266	0,188
TBP37	0,262	0,178	0,328	0,322	0,187	0,670	-0,536	-1,625
TBP24	-0,304	0,431	-0,518	-0,686	-0,151	0,365	0,265	0,123
TBP8	-0,732	-0,461	-0,838	-0,893	-0,740	0,375	0,330	0,495
TBP31	1,877	1,130	1,793	1,716	1,854	-0,431	-0,324	-1,688
TBP28	0,718	0,927	0,866	1,071	0,169	1,506	-6,379	4,092
TBP14	-0,546	-0,359	-0,631	-0,601	-0,702	2,483	0,332	0,880
TBP34	0,755	0,137	0,834	0,743	0,904	0,096	0,408	0,737
TBP18	-0,619	-0,359	-0,631	-0,601	-0,794	1,039	0,464	0,741
TBP20	-1,057	-1,079	-1,135	-1,125	-1,003	0,560	0,235	0,078
TBP6	-0,607	-0,086	-0,494	-0,434	-0,620	0,807	0,274	0,313
TBP33	0,242	0,380	0,385	0,322	0,141	0,818	0,202	-0,112
TBP10	-0,278	-0,147	-0,398	-0,371	-0,368	0,598	0,264	0,473
TBP30	0,106	0,350	0,188	0,180	0,161	-1,707	0,347	0,434
TBP1	-1,153	-1,352	-1,158	-1,101	-1,240	-0,686	-0,097	-0,866
TBP25	-0,463	-1,089	-0,555	-0,518	-0,398	-1,845	0,242	0,156
TBP44	-1,176	-0,734	-0,967	-0,921	-1,083	-0,046	-0,085	-0,953
TBP35	1,725	0,937	1,546	1,450	1,808	-0,968	0,148	0,555
TBP19	-0,546	-0,927	-0,496	-0,613	-0,463	-1,178	0,237	-0,088
TBP17	-0,573	-0,410	-0,618	-0,615	-0,450	1,293	0,358	0,511
TBP27	0,553	0,188	0,506	0,508	0,601	0,850	0,262	0,060
TBP13	-0,648	-0,309	-0,644	-0,589	-0,645	-0,418	0,290	0,319
TBP12	-0,662	-0,309	-0,644	-0,589	-0,677	0,935	0,283	0,339
TBP43	2,133	2,569	2,117	2,032	2,296	-0,277	0,257	-0,586
TBP38	1,094	0,846	1,265	1,455	0,549	0,924	-0,332	-0,804
TBP16	-0,555	-0,410	-0,618	-0,615	-0,743	-1,034	0,396	0,509
TBP36	1,872	1,181	1,780	1,730	2,037	-2,283	-0,425	-1,465
TBP40	1,308	1,424	1,371	1,557	0,893	-0,410	0,299	0,546

Tabela A (cont.). Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método B3LYP/3-21G*) para os 46 compostos aleatorizados

Composto	ClogD pH1,5	ClogD pH5	ClogD pH6,5	ClogD pH7,4	MSA	Platt	Randic
TBP45	0,030	-0,172	-0,298	-0,343	-1,306	-1,710	-1,352
TBP46	-0,078	-0,313	-0,447	-0,494	-1,383	-1,710	-1,352
TBP9	-3,326	-3,036	-3,314	-3,407	-0,677	-0,654	-0,661
TBP39	1,640	1,920	1,904	1,896	1,342	1,054	1,407
TBP4	0,023	-0,172	-0,298	-0,343	-0,878	-0,805	-0,872
TBP5	-0,093	-0,332	-0,467	-0,514	-0,910	-0,805	-0,872
TBP42	1,473	1,731	1,705	1,694	1,573	1,607	1,585
TBP41	1,864	2,212	2,211	2,208	1,664	1,356	1,696
TBP22	0,661	0,648	0,565	0,535	-0,359	-0,504	-0,592
TBP7	-0,158	-0,417	-0,556	-0,605	-0,757	-0,504	-0,592
TBP3	0,023	-0,172	-0,298	-0,343	-0,883	-0,805	-0,872
TBP26	0,009	-0,200	-0,328	-0,373	-0,850	-0,805	-0,872
TBP23	-0,006	-0,219	-0,348	-0,393	-0,025	-0,152	-0,200
TBP21	0,342	0,233	0,128	0,091	-0,723	-0,805	-0,872
TBP29	0,987	1,072	1,011	0,988	0,412	0,401	0,493
TBP11	-0,818	-1,274	-1,459	-1,522	-0,291	-0,303	-0,267
TBP2	-0,057	-0,285	-0,417	-0,464	-0,365	-0,303	-0,269
TBP32	-1,818	-0,624	0,376	0,555	1,731	1,758	1,678
TBP15	0,400	0,309	0,207	0,172	-0,529	-0,504	-0,592
TBP37	0,480	0,412	0,317	0,283	0,068	0,401	0,322
TBP24	0,741	0,751	0,674	0,645	-0,018	-0,202	-0,309
TBP8	0,103	-0,078	-0,199	-0,242	-0,576	-0,504	-0,592
TBP31	-1,565	-0,888	0,317	0,797	1,981	2,060	1,958
TBP28	0,806	0,836	0,763	0,736	0,612	0,703	0,718
TBP14	-2,101	-2,734	-1,905	-1,109	-0,372	-0,504	-0,592
TBP34	0,697	0,695	0,614	0,585	0,753	0,853	0,797
TBP18	-0,194	-0,464	-0,606	-0,655	-0,592	-0,504	-0,592
TBP20	-0,325	-0,633	-0,785	-0,837	-0,947	-0,805	-0,872
TBP6	0,139	-0,021	-0,140	-0,181	-0,790	-0,805	-0,872
TBP33	0,277	0,148	0,039	0,000	0,156	0,401	0,303
TBP10	0,371	0,271	0,168	0,131	-0,102	0,099	-0,012
TBP30	0,661	0,648	0,565	0,535	0,067	0,099	0,206
TBP1	-0,528	-0,897	-1,062	-1,119	-1,106	-0,805	-0,872
TBP25	-0,209	-0,483	-0,626	-0,675	-0,380	-0,453	-0,480
TBP44	-0,194	-0,464	-0,606	-0,655	-1,427	-1,710	-1,352
TBP35	-1,681	0,054	0,694	0,726	1,859	1,758	1,678
TBP19	-0,673	-1,086	-1,261	-1,321	-0,294	-0,353	-0,378
TBP17	0,212	0,064	-0,050	-0,091	-0,373	-0,202	-0,309
TBP27	0,603	0,572	0,485	0,454	0,559	0,451	0,602
TBP13	-0,129	-0,379	-0,517	-0,564	-0,751	-0,805	-0,872
TBP12	-0,129	-0,379	-0,517	-0,564	-0,763	-0,805	-0,872
TBP43	1,495	1,731	1,705	1,694	2,028	1,658	1,986
TBP38	0,639	0,657	0,574	0,545	0,889	1,306	1,121
TBP16	-0,260	-0,549	-0,695	-0,746	-0,394	-0,202	-0,309
TBP36	-1,478	-0,012	0,912	1,039	1,842	1,758	1,678
TBP40	1,147	1,317	1,269	1,250	1,283	1,306	1,295

Tabela A (cont.). Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método B3LYP/3-21G*) para os 46 compostos aleatorizados

Composto	Balaban	Harary	Hyper Wiener	Szeged	Wiener	Wiener polarity
TBP45	0,667	-0,776	-0,712	-0,718	-0,734	-0,785
TBP46	1,034	-0,959	-0,871	-0,875	-0,900	-0,990
TBP9	0,361	-0,590	-0,564	-0,592	-0,568	-0,581
TBP39	-0,986	1,658	1,464	1,561	1,606	1,466
TBP4	0,973	-0,969	-0,844	-0,856	-0,885	-0,990
TBP5	1,034	-0,959	-0,871	-0,875	-0,900	-0,990
TBP42	-0,864	1,475	1,225	1,481	1,308	1,773
TBP41	-1,170	1,883	1,825	1,753	1,901	1,875
TBP22	1,279	-0,159	-0,276	-0,304	-0,237	-0,171
TBP7	1,034	-0,959	-0,871	-0,875	-0,900	-0,990
TBP3	1,034	-0,959	-0,871	-0,875	-0,900	-0,990
TBP26	0,667	-0,769	-0,717	-0,726	-0,740	-0,683
TBP23	1,095	0,043	-0,108	-0,143	-0,049	0,033
TBP21	0,789	-0,763	-0,747	-0,744	-0,754	-0,785
TBP29	-0,803	0,167	0,394	0,263	0,321	-0,069
TBP11	-0,068	-0,427	-0,340	-0,421	-0,356	-0,581
TBP2	0,299	-0,368	-0,481	-0,512	-0,439	-0,274
TBP32	-1,354	1,591	1,906	1,805	1,785	1,466
TBP15	0,667	-0,769	-0,721	-0,726	-0,740	-0,683
TBP37	-1,598	0,285	0,233	0,190	0,226	-0,069
TBP24	0,789	0,502	0,230	0,183	0,335	0,443
TBP8	0,116	-0,377	-0,409	-0,448	-0,399	-0,376
TBP31	-1,354	1,591	1,906	1,805	1,827	1,466
TBP28	-1,109	0,922	0,157	0,625	0,375	1,261
TBP14	0,850	-0,751	-0,769	-0,763	-0,768	-0,683
TBP34	-1,782	0,777	0,620	0,691	0,632	1,057
TBP18	0,850	-0,746	-0,771	-0,765	-0,771	-0,581
TBP20	0,789	-0,763	-0,747	-0,744	-0,754	-0,785
TBP6	0,973	-0,969	-0,844	-0,856	-0,885	-0,990
TBP33	-1,292	0,369	0,053	0,140	0,120	0,545
TBP10	0,361	-0,590	-0,564	-0,592	-0,568	-0,581
TBP30	-0,374	0,014	0,046	0,039	0,034	-0,069
TBP1	-0,129	-1,145	-0,993	-1,002	-1,043	-1,194
TBP25	0,606	-0,560	-0,639	-0,635	-0,613	-0,478
TBP44	1,034	-0,959	-0,871	-0,875	-0,900	-0,990
TBP35	-1,476	1,815	2,270	2,119	2,104	1,671
TBP19	0,055	-0,390	-0,372	-0,413	-0,381	-0,274
TBP17	0,789	-0,763	-0,747	-0,744	-0,754	-0,785
TBP27	-1,231	0,323	0,802	0,505	0,638	0,033
TBP13	0,850	-0,759	-0,769	-0,763	-0,768	-0,683
TBP12	0,789	-0,759	-0,744	-0,744	-0,754	-0,683
TBP43	-1,415	2,094	2,251	2,237	2,255	1,977
TBP38	-1,660	1,339	0,831	1,088	0,974	1,261
TBP16	0,850	-0,751	-0,769	-0,763	-0,768	-0,683
TBP36	-1,354	1,591	1,906	1,805	1,785	1,466
TBP40	-0,619	1,270	0,882	1,056	1,006	1,671

Tabela B. Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método HF/3-21G*) para os 46 compostos aleatorizados

Composto	E _{EST}	E _{ang}	E _{TORS}	E _{1,4}	E _{vdW}	E _{EL}	E _{SOLV}	E _{HBD}
TBP45	-1,139	-1,667	-2,233	-0,640	-0,286	0,648	-1,021	0,059
TBP46	-1,206	-1,703	-1,337	-0,647	-0,256	0,396	-0,631	0,199
TBP9	-0,267	-0,518	0,357	0,576	0,225	-2,733	-2,796	0,448
TBP39	1,660	1,039	-0,581	1,161	3,637	-0,366	0,085	-1,395
TBP4	-0,953	-0,892	0,337	-0,631	-0,032	0,450	0,110	0,125
TBP5	-0,677	-0,615	0,056	-0,830	-0,381	0,509	-0,272	0,115
TBP42	1,427	1,538	0,386	2,225	1,162	-1,497	1,415	0,720
TBP41	1,353	2,043	-0,815	0,841	3,332	-0,161	0,101	0,006
TBP22	-0,552	-0,571	-0,035	0,098	-0,448	-0,568	-0,852	0,084
TBP7	0,039	-0,649	-0,528	-0,839	-0,243	0,890	0,680	0,111
TBP3	-0,542	-0,709	0,484	-0,674	-0,516	0,521	0,307	0,119
TBP26	-0,881	-0,586	0,043	-0,489	-0,232	0,492	-0,489	0,116
TBP23	-0,159	-0,044	0,739	-0,165	-0,324	-0,608	-1,335	-0,211
TBP21	-0,664	-0,880	0,767	-0,900	-0,508	0,572	-0,619	0,089
TBP29	-0,478	0,080	-0,445	0,032	-0,748	0,879	0,427	0,126
TBP11	-0,701	-0,167	0,030	-2,205	-0,729	2,447	-1,533	-1,054
TBP2	-0,722	-0,182	0,148	-0,519	0,052	0,441	0,723	-6,016
TBP32	1,159	1,406	2,187	1,602	-0,408	-1,304	0,581	0,720
TBP15	-0,404	-0,571	0,019	-0,648	-0,483	0,683	0,859	0,111
TBP37	0,866	0,905	1,135	-0,210	1,489	0,796	0,415	-0,040
TBP24	-0,515	-0,371	-0,091	0,027	-0,205	-0,971	-1,885	0,075
TBP8	-0,875	-0,556	0,284	-0,074	-0,251	-0,201	-0,753	0,105
TBP31	1,729	1,127	1,782	1,668	0,290	-0,525	1,406	0,720
TBP28	0,651	-0,305	-1,812	1,283	-0,259	-1,167	0,856	0,720
TBP14	-0,816	-1,024	-1,172	-0,548	-0,297	0,418	0,138	0,080
TBP34	0,226	1,342	-0,853	0,549	-0,764	-0,761	0,972	0,372
TBP18	-0,765	-0,305	0,041	-0,570	-0,524	0,438	0,639	0,047
TBP20	-0,644	-0,588	0,233	-1,481	-0,240	1,140	-0,642	0,074
TBP6	0,085	-1,033	0,636	-0,851	-0,286	0,643	0,148	0,126
TBP33	0,357	0,647	-1,600	-0,424	-1,086	0,151	0,668	0,103
TBP10	0,048	-0,451	0,616	-0,299	-0,554	0,387	1,352	0,103
TBP30	0,334	0,178	-1,956	0,199	0,054	-0,036	0,263	0,121
TBP1	-0,783	-0,571	0,036	-0,695	-0,210	0,705	0,172	0,124
TBP25	-0,341	0,187	0,266	-0,492	-0,713	-0,170	-0,518	-0,832
TBP44	-1,346	-1,677	-1,589	-0,668	-0,137	0,421	-1,215	0,212
TBP35	1,633	1,489	2,065	1,695	-1,324	-1,460	0,040	0,720
TBP19	-1,006	-0,067	-0,044	-0,131	0,063	-2,347	-2,325	0,409
TBP17	-0,374	-0,567	0,070	-0,813	-0,481	0,970	1,453	0,134
TBP27	-0,128	0,194	-0,160	-0,483	-0,429	1,421	0,141	-0,175
TBP13	-0,898	-0,434	-0,331	-0,579	-0,481	0,489	-0,629	0,046
TBP12	-0,990	-0,778	0,172	-0,444	-0,521	0,461	-0,482	0,053
TBP43	2,793	2,242	-1,036	1,096	2,548	0,296	0,257	0,006
TBP38	0,357	1,926	1,151	1,541	1,146	-0,605	1,443	0,720
TBP16	0,686	0,023	0,433	-0,498	-0,270	0,600	1,319	0,067
TBP36	1,774	0,633	1,813	1,816	0,022	-1,153	0,116	0,720
TBP40	1,650	1,483	0,331	2,035	0,603	-1,631	0,907	0,720

Tabela B (cont.). Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método HF/3-21G*) para os 46 compostos aleatorizados

Composto	E _{TOT}	E _{HOMO}	E _{LUMO}	μ_T	μ_X	μ_Y	μ_Z	carga_N ₁
TBP45	-1,490	-0,426	0,450	1,209	-1,080	1,576	1,797	1,847
TBP46	-1,361	-0,212	0,493	0,324	1,038	1,534	-0,708	0,380
TBP9	-0,463	-0,826	-3,520	2,484	0,402	-2,172	-0,853	-0,779
TBP39	1,508	0,870	-0,237	-1,669	0,003	-0,556	-0,177	0,112
TBP4	-0,632	-0,548	0,339	-0,076	-0,825	-0,931	0,386	0,114
TBP5	-0,675	-0,806	0,592	0,727	-0,582	-1,211	-0,865	1,556
TBP42	1,866	2,708	-0,956	-0,929	-0,545	-0,546	1,135	0,439
TBP41	1,661	0,887	-0,720	-0,392	-0,710	-0,613	-0,592	-0,782
TBP22	-0,480	-0,819	-0,595	-1,917	0,741	0,786	0,742	-1,461
TBP7	-0,514	-0,558	1,246	0,632	1,393	1,630	0,256	-0,178
TBP3	-0,524	-0,578	0,490	0,886	1,071	1,800	-0,755	-0,940
TBP26	-0,568	-0,842	0,112	0,291	0,510	-1,205	1,467	0,270
TBP23	-0,209	-0,740	-0,299	1,410	-1,202	-1,205	-0,635	-0,058
TBP21	-0,703	-0,832	-0,010	-0,714	-0,501	1,195	-0,573	-1,021
TBP29	-0,089	-0,001	0,786	1,121	1,613	0,001	2,179	-1,500
TBP11	-0,973	-0,743	0,769	-0,167	-0,986	0,053	-0,919	0,578
TBP2	-0,452	-0,545	-0,703	1,150	-1,267	-0,850	-0,855	0,253
TBP32	1,578	0,115	0,999	-0,660	-0,017	-0,411	-1,171	-0,160
TBP15	-0,455	-0,591	0,513	-0,061	-0,852	-0,825	-0,239	0,332
TBP37	0,967	1,072	-0,204	-1,772	0,289	-0,312	1,259	-0,650
TBP24	-0,528	-0,987	-0,996	-0,954	0,500	-0,827	0,898	-0,029
TBP8	-0,464	-0,921	-0,062	-2,248	-0,259	0,244	-0,390	-0,102
TBP31	1,812	0,722	1,065	-0,021	1,193	0,860	1,748	-0,473
TBP28	0,091	2,243	-1,831	1,140	1,327	1,962	0,231	-1,815
TBP14	-0,928	-0,575	-0,122	1,122	-1,667	-0,039	-0,595	1,376
TBP34	0,449	0,801	-0,151	1,173	1,682	1,305	-0,779	-0,394
TBP18	-0,455	-0,720	0,690	-0,723	-1,000	-0,228	0,136	-0,316
TBP20	-0,805	-0,845	0,227	-0,724	-0,505	-0,802	-0,081	-0,473
TBP6	-0,516	-0,354	0,358	-0,207	1,105	-0,630	1,320	-0,675
TBP33	-0,240	1,039	-0,200	0,174	-1,265	-0,267	-0,395	0,746
TBP10	-0,095	-0,403	1,160	-0,085	1,269	-0,185	1,634	-0,312
TBP30	-0,112	0,798	-0,552	1,052	1,340	0,426	-1,784	0,933
TBP1	-0,550	-0,542	1,098	0,180	-0,982	-0,683	-0,587	-0,127
TBP25	-0,338	-0,628	0,381	0,136	-1,056	-0,150	-0,971	1,045
TBP44	-1,450	-0,149	0,549	0,390	-0,698	-1,169	-0,319	-0,339
TBP35	1,527	0,135	0,727	-0,881	-0,185	0,030	1,963	-0,278
TBP19	-0,745	-0,948	-2,554	-0,805	0,854	-0,526	1,283	0,314
TBP17	-0,424	-0,007	1,213	0,880	-1,365	-0,756	-0,331	-0,378
TBP27	-0,038	-0,037	0,927	-1,053	0,950	1,114	0,617	1,338
TBP13	-0,667	-0,746	-0,043	0,510	1,419	1,486	-0,080	-0,010
TBP12	-0,665	-0,720	0,056	-0,753	-0,926	-0,334	0,038	-0,289
TBP43	2,044	0,940	-0,743	0,039	1,140	1,237	-0,742	1,259
TBP38	1,774	2,956	-1,242	-1,409	0,595	0,279	-0,889	-1,558
TBP16	0,157	-0,120	1,174	0,713	-1,085	-0,402	-1,261	-0,479
TBP36	1,499	0,342	0,746	-0,097	0,021	1,405	-1,131	-1,008
TBP40	1,672	2,144	-1,420	0,574	-0,895	-1,090	-0,414	3,695

Tabela B (cont.). Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método HF/3-21G*) para os 46 compostos aleatorizados

Composto	carga_C ₂	carga_C ₃	carga_C ₄	carga_C ₅	carga_O ₆	carga_H ₇	carga_C ₈	carga_O ₉
TBP45	1,052	-1,076	-1,032	0,131	-0,326	1,898	0,754	-0,118
TBP46	0,131	0,596	-0,587	1,439	-0,359	0,033	-0,335	0,063
TBP9	0,585	-1,062	-0,273	1,196	-0,263	0,124	0,424	-0,071
TBP39	-0,821	1,415	-1,176	0,737	-0,198	0,315	0,135	0,807
TBP4	-0,504	0,338	0,520	-1,058	-0,157	-0,091	0,746	-0,939
TBP5	-0,280	-0,904	0,988	-0,937	-0,145	-0,133	-0,462	0,331
TBP42	-0,501	0,709	1,533	-1,850	-0,246	-0,226	-2,309	2,929
TBP41	1,000	0,333	0,536	-1,123	-0,343	0,312	-0,674	2,972
TBP22	1,038	-1,002	0,074	0,318	-0,246	0,168	0,133	-0,365
TBP7	0,315	0,078	-0,117	0,118	4,448	0,110	0,024	-0,066
TBP3	1,050	-0,948	0,039	0,635	-0,256	-0,043	-0,117	-0,140
TBP26	-0,592	0,361	0,680	-0,692	-0,164	-0,105	-0,238	-0,079
TBP23	0,514	-1,521	0,577	-0,343	-0,164	-0,212	0,820	-0,403
TBP21	0,897	-0,257	-0,439	0,399	-0,219	0,334	0,330	-0,442
TBP29	1,054	-2,354	-0,479	1,670	-0,275	0,178	1,167	-1,894
TBP11	0,192	0,155	0,408	-0,779	-0,269	0,166	-0,623	1,062
TBP2	0,152	0,879	-1,241	0,182	-0,227	0,191	2,105	-1,359
TBP32	-0,404	1,876	-1,900	0,609	-0,168	0,525	2,279	-1,030
TBP15	0,287	-0,457	0,088	-0,121	-0,166	0,126	-0,043	-0,252
TBP37	-1,046	2,151	-1,780	0,879	-0,176	0,551	1,134	0,457
TBP24	-0,683	0,977	-0,253	0,059	4,768	-6,078	0,468	-0,140
TBP8	-0,672	0,794	0,339	-0,631	-0,138	-0,105	0,710	-0,823
TBP31	-1,638	1,807	-1,522	0,939	-0,193	0,366	0,116	0,504
TBP28	0,669	-1,052	0,186	2,341	-0,222	-0,324	-2,025	0,262
TBP14	-4,334	0,237	1,513	0,290	0,416	-0,177	-2,020	1,948
TBP34	0,550	-0,549	0,739	-0,352	-0,313	-0,242	-0,061	-0,991
TBP18	0,317	-1,037	0,630	0,148	-0,156	-0,083	-0,601	-0,274
TBP20	0,496	-0,150	-0,007	-0,084	-0,279	0,175	-0,030	-0,118
TBP6	0,617	0,084	0,363	-0,056	-0,290	1,018	-0,315	-0,166
TBP33	-0,057	-0,495	0,952	-0,895	-0,178	-0,231	-0,317	-0,153
TBP10	-0,213	-0,177	-0,042	-0,493	-0,171	0,197	1,460	-1,907
TBP30	-0,399	0,174	0,383	-1,198	-0,212	0,012	0,474	-0,602
TBP1	0,598	-0,408	0,387	-0,419	-0,252	0,092	-0,076	0,042
TBP25	-0,540	0,072	1,163	-1,057	-0,186	-0,303	-0,658	-0,058
TBP44	0,292	1,318	-0,368	1,103	-0,312	0,129	-0,688	0,755
TBP35	-0,145	1,748	-2,866	1,339	-0,236	0,834	2,606	-1,493
TBP19	-0,470	0,760	0,108	-0,261	-0,178	0,029	-0,636	0,910
TBP17	1,193	-1,066	0,921	-0,350	-0,236	-0,149	-0,509	0,271
TBP27	-1,071	-0,437	1,613	-1,371	-0,115	-0,351	-0,612	0,457
TBP13	-0,805	0,686	0,179	-0,215	-0,140	-0,059	-0,060	-0,032
TBP12	0,381	0,306	0,165	-0,313	-0,241	0,046	0,182	-0,244
TBP43	1,342	-1,929	1,186	-1,680	-0,215	-0,145	-0,612	0,478
TBP38	1,365	-0,742	-1,634	2,024	-0,299	0,635	-0,061	-0,628
TBP16	-0,257	-0,402	1,350	-1,121	-0,192	-0,264	-0,010	-1,117
TBP36	0,836	0,297	-1,447	1,812	-0,298	0,296	-0,247	0,405
TBP40	-1,493	-0,124	-0,456	-0,970	-0,210	0,459	-1,726	1,252

Tabela B (cont.). Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método HF/3-21 G*) para os 46 compostos aleatorizados

Composto	carga_N ₁₀	Vol. Vw	ClogP	RM	α	Área superf.	θ_1	θ_2
TBP45	-0,418	-1,013	-0,319	-0,805	-0,841	-0,988	-0,775	-0,409
TBP46	0,174	-1,050	-0,461	-0,816	-0,828	-1,033	-0,342	-0,044
TBP9	-0,525	-0,816	-1,403	-0,766	-0,856	-0,661	-0,426	0,291
TBP39	0,041	1,502	1,768	1,625	1,545	1,786	0,082	-0,155
TBP4	-0,854	-0,767	-0,552	-0,750	-0,752	-0,741	0,263	0,285
TBP5	-0,052	-0,887	-0,836	-0,901	-0,845	-0,833	0,189	0,285
TBP42	2,033	1,565	1,819	1,617	1,801	1,128	0,458	0,038
TBP41	0,447	1,817	2,174	1,871	1,789	2,053	-1,848	-0,575
TBP22	-0,213	-0,382	0,340	-0,430	-0,546	-0,176	0,089	0,258
TBP7	-0,190	-0,871	-0,886	-0,888	-0,857	-0,868	-0,729	-0,005
TBP3	0,095	-0,779	-0,552	-0,750	-0,752	-0,721	0,774	0,250
TBP26	-0,118	-0,859	-0,694	-0,889	-0,857	-0,850	0,421	0,244
TBP23	-0,984	-0,287	-0,714	-0,492	-0,566	-0,145	-0,133	0,295
TBP21	-0,239	-0,632	-0,309	-0,644	-0,589	-0,544	1,267	0,289
TBP29	-1,098	0,420	0,755	0,434	0,424	0,807	-1,094	0,700
TBP11	0,020	-0,428	-1,869	-0,562	-0,518	-0,135	-1,743	-0,028
TBP2	-1,982	-0,426	-1,626	-0,541	-0,518	-0,664	0,434	0,281
TBP32	-1,053	1,653	0,806	1,534	1,462	1,565	0,702	0,021
TBP15	-0,188	-0,486	-0,086	-0,480	-0,508	-0,504	0,377	0,266
TBP37	-0,822	0,262	0,178	0,328	0,322	0,187	0,670	-0,536
TBP24	-0,857	-0,304	0,431	-0,518	-0,686	-0,151	0,365	0,265
TBP8	-0,997	-0,732	-0,461	-0,838	-0,893	-0,740	0,375	0,330
TBP31	0,756	1,877	1,130	1,793	1,716	1,854	-0,431	-0,324
TBP28	3,796	0,718	0,927	0,866	1,071	0,169	1,506	-6,379
TBP14	0,264	-0,546	-0,359	-0,631	-0,601	-0,702	2,483	0,332
TBP34	-0,108	0,755	0,137	0,834	0,743	0,904	0,096	0,408
TBP18	0,507	-0,619	-0,359	-0,631	-0,601	-0,794	1,039	0,464
TBP20	-0,052	-1,057	-1,079	-1,135	-1,125	-1,003	0,560	0,235
TBP6	-0,005	-0,607	-0,086	-0,494	-0,434	-0,620	0,807	0,274
TBP33	0,203	0,242	0,380	0,385	0,322	0,141	0,818	0,202
TBP10	-1,356	-0,278	-0,147	-0,398	-0,371	-0,368	0,598	0,264
TBP30	-0,411	0,106	0,350	0,188	0,180	0,161	-1,707	0,347
TBP1	-0,230	-1,153	-1,352	-1,158	-1,101	-1,240	-0,686	-0,097
TBP25	0,360	-0,463	-1,089	-0,555	-0,518	-0,398	-1,845	0,242
TBP44	0,375	-1,176	-0,734	-0,967	-0,921	-1,083	-0,046	-0,085
TBP35	-1,156	1,725	0,937	1,546	1,450	1,808	-0,968	0,148
TBP19	0,122	-0,546	-0,927	-0,496	-0,613	-0,463	-1,178	0,237
TBP17	0,035	-0,573	-0,410	-0,618	-0,615	-0,450	1,293	0,358
TBP27	-0,312	0,553	0,188	0,506	0,508	0,601	0,850	0,262
TBP13	-0,209	-0,648	-0,309	-0,644	-0,589	-0,645	-0,418	0,290
TBP12	-0,403	-0,662	-0,309	-0,644	-0,589	-0,677	0,935	0,283
TBP43	0,469	2,133	2,569	2,117	2,032	2,296	-0,277	0,257
TBP38	2,104	1,094	0,846	1,265	1,455	0,549	0,924	-0,332
TBP16	-0,205	-0,555	-0,410	-0,618	-0,615	-0,743	-1,034	0,396
TBP36	1,196	1,872	1,181	1,780	1,730	2,037	-2,283	-0,425
TBP40	2,043	1,308	1,424	1,371	1,557	0,893	-0,410	0,299

Tabela B (cont.). Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método HF/3-21G*) para os 46 compostos aleatorizados

Composto	Θ_3	ClogD pH1,5	ClogD pH5	ClogD pH6,5	ClogD pH7,4	MSA	Platt
TBP45	-1,315	0,030	-0,172	-0,298	-0,343	-1,306	-1,710
TBP46	-0,805	-0,078	-0,313	-0,447	-0,494	-1,383	-1,710
TBP9	0,248	-3,326	-3,036	-3,314	-3,407	-0,677	-0,654
TBP39	-1,478	1,640	1,920	1,904	1,896	1,342	1,054
TBP4	0,317	0,023	-0,172	-0,298	-0,343	-0,878	-0,805
TBP5	0,317	-0,093	-0,332	-0,467	-0,514	-0,910	-0,805
TBP42	-0,497	1,473	1,731	1,705	1,694	1,573	1,607
TBP41	-2,006	1,864	2,212	2,211	2,208	1,664	1,356
TBP22	0,167	0,661	0,648	0,565	0,535	-0,359	-0,504
TBP7	-0,635	-0,158	-0,417	-0,556	-0,605	-0,757	-0,504
TBP3	0,172	0,023	-0,172	-0,298	-0,343	-0,883	-0,805
TBP26	0,325	0,009	-0,200	-0,328	-0,373	-0,850	-0,805
TBP23	0,295	-0,006	-0,219	-0,348	-0,393	-0,025	-0,152
TBP21	0,296	0,342	0,233	0,128	0,091	-0,723	-0,805
TBP29	1,662	0,987	1,072	1,011	0,988	0,412	0,401
TBP11	-1,049	-0,818	-1,274	-1,459	-1,522	-0,291	-0,303
TBP2	0,318	-0,057	-0,285	-0,417	-0,464	-0,365	-0,303
TBP32	0,308	-1,818	-0,624	0,376	0,555	1,731	1,758
TBP15	0,188	0,400	0,309	0,207	0,172	-0,529	-0,504
TBP37	-1,625	0,480	0,412	0,317	0,283	0,068	0,401
TBP24	0,123	0,741	0,751	0,674	0,645	-0,018	-0,202
TBP8	0,495	0,103	-0,078	-0,199	-0,242	-0,576	-0,504
TBP31	-1,688	-1,565	-0,888	0,317	0,797	1,981	2,060
TBP28	4,092	0,806	0,836	0,763	0,736	0,612	0,703
TBP14	0,880	-2,101	-2,734	-1,905	-1,109	-0,372	-0,504
TBP34	0,737	0,697	0,695	0,614	0,585	0,753	0,853
TBP18	0,741	-0,194	-0,464	-0,606	-0,655	-0,592	-0,504
TBP20	0,078	-0,325	-0,633	-0,785	-0,837	-0,947	-0,805
TBP6	0,313	0,139	-0,021	-0,140	-0,181	-0,790	-0,805
TBP33	-0,112	0,277	0,148	0,039	0,000	0,156	0,401
TBP10	0,473	0,371	0,271	0,168	0,131	-0,102	0,099
TBP30	0,434	0,661	0,648	0,565	0,535	0,067	0,099
TBP1	-0,866	-0,528	-0,897	-1,062	-1,119	-1,106	-0,805
TBP25	0,156	-0,209	-0,483	-0,626	-0,675	-0,380	-0,453
TBP44	-0,953	-0,194	-0,464	-0,606	-0,655	-1,427	-1,710
TBP35	0,555	-1,681	0,054	0,694	0,726	1,859	1,758
TBP19	-0,088	-0,673	-1,086	-1,261	-1,321	-0,294	-0,353
TBP17	0,511	0,212	0,064	-0,050	-0,091	-0,373	-0,202
TBP27	0,060	0,603	0,572	0,485	0,454	0,559	0,451
TBP13	0,319	-0,129	-0,379	-0,517	-0,564	-0,751	-0,805
TBP12	0,339	-0,129	-0,379	-0,517	-0,564	-0,763	-0,805
TBP43	-0,586	1,495	1,731	1,705	1,694	2,028	1,658
TBP38	-0,804	0,639	0,657	0,574	0,545	0,889	1,306
TBP16	0,509	-0,260	-0,549	-0,695	-0,746	-0,394	-0,202
TBP36	-1,465	-1,478	-0,012	0,912	1,039	1,842	1,758
TBP40	0,546	1,147	1,317	1,269	1,250	1,283	1,306

Tabela B (cont.). Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método HF/3-21G*) para os 46 compostos aleatorizados

Composto	Randic	Balaban	Harary	Hyper Wiener	Szeged	Wiener	Wiener polarity
TBP45	-1,352	0,667	-0,776	-0,712	-0,718	-0,734	-0,785
TBP46	-1,352	1,034	-0,959	-0,871	-0,875	-0,900	-0,990
TBP9	-0,661	0,361	-0,590	-0,564	-0,592	-0,568	-0,581
TBP39	1,407	-0,986	1,658	1,464	1,561	1,606	1,466
TBP4	-0,872	0,973	-0,969	-0,844	-0,856	-0,885	-0,990
TBP5	-0,872	1,034	-0,959	-0,871	-0,875	-0,900	-0,990
TBP42	1,585	-0,864	1,475	1,225	1,481	1,308	1,773
TBP41	1,696	-1,170	1,883	1,825	1,753	1,901	1,875
TBP22	-0,592	1,279	-0,159	-0,276	-0,304	-0,237	-0,171
TBP7	-0,592	1,034	-0,959	-0,871	-0,875	-0,900	-0,990
TBP3	-0,872	1,034	-0,959	-0,871	-0,875	-0,900	-0,990
TBP26	-0,872	0,667	-0,769	-0,717	-0,726	-0,740	-0,683
TBP23	-0,200	1,095	0,043	-0,108	-0,143	-0,049	0,033
TBP21	-0,872	0,789	-0,763	-0,747	-0,744	-0,754	-0,785
TBP29	0,493	-0,803	0,167	0,394	0,263	0,321	-0,069
TBP11	-0,267	-0,068	-0,427	-0,340	-0,421	-0,356	-0,581
TBP2	-0,269	0,299	-0,368	-0,481	-0,512	-0,439	-0,274
TBP32	1,678	-1,354	1,591	1,906	1,805	1,785	1,466
TBP15	-0,592	0,667	-0,769	-0,721	-0,726	-0,740	-0,683
TBP37	0,322	-1,598	0,285	0,233	0,190	0,226	-0,069
TBP24	-0,309	0,789	0,502	0,230	0,183	0,335	0,443
TBP8	-0,592	0,116	-0,377	-0,409	-0,448	-0,399	-0,376
TBP31	1,958	-1,354	1,591	1,906	1,805	1,827	1,466
TBP28	0,718	-1,109	0,922	0,157	0,625	0,375	1,261
TBP14	-0,592	0,850	-0,751	-0,769	-0,763	-0,768	-0,683
TBP34	0,797	-1,782	0,777	0,620	0,691	0,632	1,057
TBP18	-0,592	0,850	-0,746	-0,771	-0,765	-0,771	-0,581
TBP20	-0,872	0,789	-0,763	-0,747	-0,744	-0,754	-0,785
TBP6	-0,872	0,973	-0,969	-0,844	-0,856	-0,885	-0,990
TBP33	0,303	-1,292	0,369	0,053	0,140	0,120	0,545
TBP10	-0,012	0,361	-0,590	-0,564	-0,592	-0,568	-0,581
TBP30	0,206	-0,374	0,014	0,046	0,039	0,034	-0,069
TBP1	-0,872	-0,129	-1,145	-0,993	-1,002	-1,043	-1,194
TBP25	-0,480	0,606	-0,560	-0,639	-0,635	-0,613	-0,478
TBP44	-1,352	1,034	-0,959	-0,871	-0,875	-0,900	-0,990
TBP35	1,678	-1,476	1,815	2,270	2,119	2,104	1,671
TBP19	-0,378	0,055	-0,390	-0,372	-0,413	-0,381	-0,274
TBP17	-0,309	0,789	-0,763	-0,747	-0,744	-0,754	-0,785
TBP27	0,602	-1,231	0,323	0,802	0,505	0,638	0,033
TBP13	-0,872	0,850	-0,759	-0,769	-0,763	-0,768	-0,683
TBP12	-0,872	0,789	-0,759	-0,744	-0,744	-0,754	-0,683
TBP43	1,986	-1,415	2,094	2,251	2,237	2,255	1,977
TBP38	1,121	-1,660	1,339	0,831	1,088	0,974	1,261
TBP16	-0,309	0,850	-0,751	-0,769	-0,763	-0,768	-0,683
TBP36	1,678	-1,354	1,591	1,906	1,805	1,785	1,466
TBP40	1,295	-0,619	1,270	0,882	1,056	1,006	1,671

Tabela C. Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método B3LYP/6-31G*) para os 45 compostos aleatorizados

Composto	E _{EST}	E _{ang}	E _{TORS}	E _{1,4}	E _{vdW}	E _{EL}	E _{SOLV}	E _{HBD}
TBP25	-0,335	0,164	0,278	-0,510	-0,712	-0,155	-0,509	-0,820
TBP32	1,148	1,385	2,187	1,578	-0,410	-1,281	0,578	0,715
TBP9	-0,262	-0,542	0,369	0,556	0,216	-2,701	-2,762	0,445
TBP43	2,764	2,221	-1,015	1,074	2,516	0,308	0,258	0,009
TBP22	-0,544	-0,595	-0,021	0,079	-0,450	-0,550	-0,839	0,086
TBP26	-0,869	-0,609	0,057	-0,507	-0,236	0,503	-0,481	0,118
TBP16	0,680	0,000	0,444	-0,516	-0,273	0,610	1,308	0,069
TBP28	0,645	-0,329	-1,786	1,261	-0,263	-1,146	0,850	0,715
TBP31	1,711	1,105	1,784	1,645	0,280	-0,508	1,394	0,715
TBP40	1,634	1,461	0,342	2,011	0,590	-1,606	0,900	0,715
TBP27	-0,124	0,172	-0,145	-0,500	-0,431	1,426	0,143	-0,170
TBP2	-0,713	-0,206	0,161	-0,536	0,045	0,452	0,718	-5,947
TBP30	0,332	0,155	-1,930	0,180	0,048	-0,022	0,263	0,123
TBP23	-0,155	-0,067	0,748	-0,183	-0,327	-0,589	-1,317	-0,206
TBP45	-1,190	-1,728	-1,314	-0,664	-0,260	0,407	-0,621	0,200
TBP17	-0,368	-0,591	0,083	-0,829	-0,482	0,977	1,440	0,136
TBP20	-0,635	-0,612	0,245	-1,495	-0,244	1,147	-0,632	0,075
TBP37	0,858	0,883	1,142	-0,229	1,468	0,805	0,414	-0,037
TBP11	-0,691	-0,190	0,044	-2,217	-0,728	2,446	-1,513	-1,039
TBP34	0,226	1,320	-0,833	0,529	-0,763	-0,742	0,965	0,371
TBP8	-0,863	-0,580	0,296	-0,093	-0,255	-0,186	-0,741	0,106
TBP36	1,756	0,611	1,815	1,792	0,016	-1,131	0,118	0,715
TBP4	-0,941	-0,916	0,349	-0,648	-0,038	0,461	0,112	0,126
TBP39	1,644	1,017	-0,563	1,138	3,594	-0,349	0,088	-1,377
TBP29	-0,471	0,057	-0,428	0,013	-0,747	0,887	0,426	0,127
TBP3	-0,535	-0,733	0,495	-0,691	-0,517	0,531	0,307	0,120
TBP12	-0,978	-0,801	0,185	-0,461	-0,522	0,472	-0,473	0,055
TBP41	1,340	2,023	-0,796	0,820	3,292	-0,146	0,104	0,009
TBP44	-1,330	-1,701	-1,564	-0,684	-0,142	0,433	-1,198	0,212
TBP15	-0,397	-0,595	0,032	-0,665	-0,485	0,692	0,853	0,112
TBP33	0,355	0,624	-1,575	-0,442	-1,081	0,164	0,664	0,104
TBP7	0,040	-0,673	-0,510	-0,855	-0,247	0,898	0,676	0,113
TBP13	-0,887	-0,457	-0,315	-0,596	-0,482	0,500	-0,619	0,049
TBP24	-0,507	-0,394	-0,076	0,008	-0,209	-0,951	-1,861	0,077
TBP10	0,050	-0,474	0,626	-0,316	-0,554	0,399	1,340	0,104
TBP46	-1,125	-1,691	-2,204	-0,657	-0,289	0,658	-1,006	0,062
TBP42	1,413	1,517	0,398	2,200	1,144	-1,473	1,403	0,715
TBP38	0,355	1,905	1,157	1,517	1,128	-0,587	1,431	0,715
TBP5	-0,668	-0,638	0,070	-0,846	-0,383	0,520	-0,266	0,117
TBP21	-0,655	-0,904	0,776	-0,916	-0,509	0,582	-0,609	0,091
TBP19	-0,993	-0,090	-0,030	-0,150	0,056	-2,317	-2,297	0,407
TBP14	-0,805	-1,048	-1,150	-0,565	-0,300	0,430	0,140	0,082
TBP35	1,617	1,468	2,065	1,671	-1,316	-1,436	0,043	0,715
TBP18	-0,755	-0,329	0,055	-0,587	-0,525	0,450	0,635	0,050
TBP1	-0,773	-0,595	0,050	-0,712	-0,214	0,715	0,174	0,125

Tabela C (cont.). Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método B3LYP/6-31G*) para os 45 compostos aleatorizados

Composto	E _{TOT}	E _{HOMO}	E _{LUMO}	μ _T	μ _x	μ _y	μ _z	carga_N ₁
TBP25	-0,346	-0,297	0,340	-0,213	-0,963	-0,144	-0,814	0,852
TBP32	1,554	1,557	1,025	-0,599	-0,141	-0,372	-1,256	0,017
TBP9	-0,471	-0,748	-3,683	2,526	0,516	-2,138	-0,781	-1,178
TBP43	2,016	0,124	-0,455	0,146	1,290	1,165	-0,673	1,436
TBP22	-0,487	-0,611	-0,568	-1,634	0,733	0,916	0,905	-1,518
TBP26	-0,575	-0,667	0,183	0,382	0,688	-1,173	1,663	0,210
TBP16	0,145	0,166	1,129	0,646	-1,013	-0,504	-1,246	-0,789
TBP28	0,079	2,041	-1,417	1,430	1,453	1,919	0,256	-1,842
TBP31	1,786	1,089	1,079	-0,135	1,286	0,703	1,823	1,208
TBP40	1,647	2,094	-1,050	0,595	-0,873	-1,106	-0,316	3,253
TBP27	-0,049	-0,421	0,915	-1,053	1,088	0,957	0,585	1,237
TBP2	-0,459	0,003	-0,800	1,107	-1,241	-0,883	-0,756	-0,005
TBP30	-0,123	-0,192	-0,342	0,964	1,367	0,401	-1,856	0,850
TBP23	-0,218	-0,384	-0,164	1,486	-1,192	-1,213	-0,646	-0,358
TBP45	-1,361	-0,842	0,365	0,214	1,015	1,457	-0,638	0,731
TBP17	-0,432	-0,175	1,168	0,959	-1,358	-0,807	-0,282	-0,445
TBP20	-0,810	-0,552	0,330	-0,420	-0,593	-0,865	-0,071	-0,634
TBP37	0,948	0,313	-0,008	-2,026	0,364	-0,303	1,229	-0,650
TBP11	-0,976	-0,496	0,357	-0,561	-0,886	0,014	-0,707	0,562
TBP34	0,434	0,069	0,076	1,356	1,888	1,218	-0,652	-0,674
TBP8	-0,472	-0,842	-0,012	-2,008	-0,312	-0,024	-0,368	-0,259
TBP36	1,475	-0,163	0,774	-0,292	-0,014	1,153	-1,190	-0,615
TBP4	-0,639	-0,511	0,228	-0,222	-0,678	-0,946	0,537	0,004
TBP39	1,485	-0,011	0,005	-1,624	-0,111	-0,600	0,025	-0,007
TBP29	-0,100	-0,210	0,700	1,167	1,661	-0,141	2,420	-1,973
TBP3	-0,531	-0,400	0,371	0,855	1,086	1,711	-0,697	-0,919
TBP12	-0,671	-0,402	-0,003	-0,658	-0,892	-0,444	0,073	-0,299
TBP41	1,636	0,103	-0,427	-0,492	-0,789	-0,524	-0,423	-0,445
TBP44	-1,450	-0,511	0,518	0,169	-0,618	-1,093	-0,242	0,011
TBP15	-0,463	-0,227	0,420	-0,055	-0,769	-0,894	-0,156	0,159
TBP33	-0,250	0,251	0,058	0,254	-1,263	-0,373	-0,319	0,949
TBP7	-0,522	-0,076	1,137	0,616	1,460	1,509	0,409	-0,114
TBP13	-0,673	-0,479	-0,062	0,393	1,444	1,377	0,059	-0,029
TBP24	-0,535	-0,935	-0,827	-0,746	0,674	-0,857	1,091	-0,200
TBP10	-0,106	-0,276	1,054	-0,068	-0,161	1,575	-0,496	-0,461
TBP46	-1,489	-1,043	0,384	0,649	-0,785	1,472	1,718	1,555
TBP42	1,840	3,029	-0,713	-1,026	-0,490	-0,458	1,368	0,225
TBP38	1,748	3,654	-0,760	-1,429	0,666	0,249	-0,901	-1,264
TBP5	-0,681	-0,549	0,449	0,703	-0,527	-1,223	-0,808	1,279
TBP21	-0,708	-0,609	-0,010	-0,670	-0,426	1,108	-0,643	-0,896
TBP19	-0,750	-1,003	-3,264	-0,661	1,007	-0,506	1,489	0,250
TBP14	-0,932	-0,115	-0,874	1,383	-1,738	-0,184	-0,603	1,772
TBP35	1,504	-0,090	0,818	-0,926	0,059	-0,059	2,152	-0,341
TBP18	-0,462	-0,569	0,594	-0,496	-1,032	-0,353	0,233	-0,422
TBP1	-0,556	-0,087	0,962	0,010	-0,880	-0,714	-0,493	-0,222

Tabela C (cont.). Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método B3LYP/6-31G*) para os 45 compostos aleatorizados

Composto	carga_C ₂	carga_C ₃	carga_C ₄	carga_C ₅	carga_O ₆	carga_H ₇	carga_C ₈	carga_O ₉
TBP25	-0,213	-0,489	1,169	-1,076	-0,018	-1,296	-0,639	-0,162
TBP32	-0,555	2,009	-1,729	0,666	0,581	1,589	2,161	-0,311
TBP9	0,803	-1,141	-0,182	1,395	-0,491	0,231	0,529	-0,145
TBP43	0,934	-1,390	0,725	-1,230	0,395	-0,948	-0,013	-0,193
TBP22	1,079	-0,920	0,011	0,524	-0,197	0,467	0,164	-0,205
TBP26	-0,602	0,296	0,769	-0,766	0,556	-0,758	-0,321	-0,134
TBP16	0,022	-0,751	1,442	-1,130	0,040	-1,536	-0,164	-0,295
TBP28	0,612	-0,364	0,024	2,368	-0,360	-1,067	-2,098	-0,086
TBP31	-1,965	2,335	-1,333	0,658	0,058	1,205	0,319	-0,139
TBP40	-1,456	0,207	-0,392	-0,905	-0,041	1,424	-1,722	0,015
TBP27	-1,152	-0,338	1,232	-1,255	1,077	-1,345	-0,369	-0,136
TBP2	0,230	0,447	-1,049	0,537	-0,088	0,261	2,003	-0,343
TBP30	-0,182	-0,087	0,512	-1,261	-0,344	-0,263	0,399	-0,226
TBP23	0,599	-1,482	0,456	-0,157	0,736	-0,993	0,953	-0,241
TBP45	0,049	0,387	-0,177	1,270	-1,904	-0,322	-0,484	-0,138
TBP17	1,297	-0,993	0,825	-0,287	-0,377	-0,786	-0,476	-0,108
TBP20	0,550	-0,289	-0,053	0,167	-0,608	0,428	0,122	-0,182
TBP37	-1,267	2,614	-1,778	1,095	0,102	1,944	1,403	-0,145
TBP11	0,065	0,329	0,377	-0,810	-0,443	0,403	-0,404	-0,064
TBP34	0,729	-0,618	0,734	-0,262	-1,244	-1,325	-0,074	-0,286
TBP8	-0,552	0,629	0,582	-0,853	0,745	-0,799	0,498	-0,251
TBP36	0,430	1,052	-1,039	1,164	-0,802	0,827	0,107	-0,130
TBP4	-0,437	0,290	0,637	-1,038	0,504	-0,857	0,486	-0,218
TBP39	-0,877	1,538	-0,999	0,665	0,161	0,915	0,464	-0,111
TBP29	1,407	-2,363	-0,386	2,147	-0,767	0,311	1,151	-0,336
TBP3	1,063	-0,792	0,178	0,580	-0,390	-0,464	-0,242	-0,139
TBP12	0,475	0,047	0,301	-0,383	-0,410	-0,123	0,025	-0,166
TBP41	0,997	0,311	-3,104	-1,165	-1,297	0,767	-0,545	-0,071
TBP44	0,201	1,100	-0,056	0,776	-1,369	0,204	-0,590	-0,074
TBP15	0,341	-0,278	-0,138	0,198	0,448	0,329	0,236	-0,216
TBP33	-0,163	-0,392	0,899	-1,067	0,425	-1,158	0,017	-0,248
TBP7	0,228	0,204	0,094	-0,076	-0,345	0,041	0,052	-0,185
TBP13	-0,720	0,366	0,420	-0,263	0,743	-0,658	-0,223	-0,126
TBP24	-0,641	0,618	0,173	0,059	1,042	-0,578	0,084	-0,125
TBP10	0,161	-0,692	0,131	-0,153	0,293	0,407	1,026	-0,347
TBP46	0,978	-0,656	-0,700	0,082	-1,516	1,031	0,770	6,496
TBP42	-0,403	0,013	1,041	-1,889	-0,037	-0,435	-2,543	0,397
TBP38	1,041	-0,607	-1,164	1,772	-0,618	2,186	-0,484	-0,124
TBP5	-0,158	-0,806	0,896	-0,890	0,548	-0,783	-0,372	-0,106
TBP21	0,865	-0,255	-0,270	0,384	-0,063	0,904	0,179	-0,151
TBP19	-0,439	0,556	0,461	-0,517	0,387	-0,374	-0,761	-0,004
TBP14	-4,211	0,140	1,520	-0,092	4,802	-0,736	-2,188	0,026
TBP35	-0,122	1,553	-2,409	1,493	-0,089	2,570	2,437	-0,374
TBP18	0,437	-1,126	0,635	0,317	0,669	-0,744	-0,688	0,247
TBP1	0,523	-0,212	0,714	-0,793	-0,495	-0,097	-0,186	-0,140

Tabela C (cont.). Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método B3LYP/6-31G*) para os 45 compostos aleatorizados

Composto	carga_N ₁₀	Vol. Vw	ClogP	RM	α	Área superf.	θ_1	θ_2
TBP25	0,183	-0,473	-1,079	-0,561	-0,523	-0,409	-1,820	0,246
TBP32	-0,692	1,628	0,795	1,510	1,439	1,540	0,717	0,027
TBP9	-0,705	-0,823	-1,389	-0,770	-0,858	-0,670	-0,407	0,294
TBP43	0,063	2,104	2,538	2,088	2,004	2,267	-0,258	0,260
TBP22	-0,160	-0,393	0,334	-0,437	-0,550	-0,188	0,107	0,261
TBP26	-0,145	-0,867	-0,688	-0,893	-0,859	-0,858	0,437	0,248
TBP16	-0,216	-0,565	-0,407	-0,624	-0,619	-0,752	-1,013	0,398
TBP28	3,558	0,699	0,915	0,848	1,051	0,154	1,518	-6,307
TBP31	0,970	1,851	1,116	1,767	1,691	1,827	-0,411	-0,314
TBP40	2,214	1,285	1,406	1,348	1,534	0,874	-0,391	0,302
TBP27	-0,413	0,536	0,184	0,491	0,494	0,583	0,864	0,265
TBP2	-1,746	-0,437	-1,610	-0,547	-0,523	-0,673	0,450	0,284
TBP30	-0,449	0,092	0,344	0,176	0,168	0,146	-1,682	0,349
TBP23	-1,022	-0,298	-0,708	-0,499	-0,570	-0,157	-0,115	0,298
TBP45	0,183	-1,056	-0,458	-0,820	-0,830	-1,040	-0,323	-0,038
TBP17	-0,071	-0,583	-0,407	-0,624	-0,619	-0,461	1,306	0,361
TBP20	-0,148	-1,063	-1,069	-1,136	-1,124	-1,010	0,576	0,239
TBP37	-0,765	0,247	0,174	0,314	0,309	0,172	0,686	-0,525
TBP11	-0,097	-0,439	-1,850	-0,568	-0,523	-0,148	-1,718	-0,022
TBP34	-0,240	0,736	0,134	0,816	0,727	0,884	0,113	0,410
TBP8	-0,848	-0,741	-0,458	-0,842	-0,895	-0,749	0,391	0,332
TBP36	1,250	1,846	1,166	1,754	1,704	2,009	-2,257	-0,415
TBP4	-0,706	-0,775	-0,548	-0,755	-0,755	-0,750	0,280	0,288
TBP39	-0,130	1,478	1,747	1,600	1,522	1,760	0,100	-0,147
TBP29	-0,940	0,404	0,745	0,420	0,410	0,788	-1,072	0,699
TBP3	0,065	-0,787	-0,548	-0,755	-0,755	-0,730	0,789	0,254
TBP12	-0,348	-0,671	-0,307	-0,649	-0,594	-0,686	0,950	0,286
TBP41	0,275	1,791	2,148	1,844	1,764	2,025	-1,823	-0,563
TBP44	0,194	-1,181	-0,728	-0,970	-0,922	-1,090	-0,028	-0,079
TBP15	-0,401	-0,496	-0,087	-0,487	-0,513	-0,514	0,393	0,270
TBP33	-0,194	0,227	0,374	0,371	0,309	0,126	0,833	0,206
TBP7	-0,327	-0,878	-0,878	-0,892	-0,859	-0,876	-0,709	0,001
TBP13	-0,273	-0,657	-0,307	-0,649	-0,594	-0,654	-0,399	0,293
TBP24	-0,621	-0,316	0,424	-0,525	-0,690	-0,163	0,382	0,269
TBP10	-1,108	-0,289	-0,147	-0,406	-0,377	-0,379	0,614	0,267
TBP46	-0,382	-1,019	-0,317	-0,809	-0,843	-0,995	-0,754	-0,398
TBP42	2,521	1,541	1,797	1,593	1,775	1,107	0,474	0,044
TBP38	2,392	1,073	0,835	1,243	1,432	0,531	0,938	-0,323
TBP5	-0,190	-0,895	-0,828	-0,904	-0,847	-0,841	0,206	0,288
TBP21	-0,103	-0,641	-0,307	-0,649	-0,594	-0,554	1,280	0,292
TBP19	0,101	-0,556	-0,919	-0,503	-0,617	-0,474	-1,156	0,241
TBP14	0,539	-0,556	-0,357	-0,636	-0,605	-0,710	2,492	0,334
TBP35	-0,796	1,700	0,925	1,522	1,427	1,782	-0,947	0,153
TBP18	0,046	-0,628	-0,357	-0,636	-0,605	-0,802	1,053	0,465
TBP1	-0,315	-1,158	-1,339	-1,159	-1,101	-1,245	-0,665	-0,090

Tabela C (cont.). Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método B3LYP/6-31G*) para os 45 compostos aleatorizados

Composto	e_3	ClogD pH1,5	ClogD pH5	ClogD pH6,5	ClogD pH7,4	MSA	Platt
TBP25	0,161	-0,204	-0,478	-0,622	-0,672	-0,396	-0,470
TBP32	0,312	-1,795	-0,618	0,369	0,545	1,707	1,734
TBP9	0,253	-3,287	-3,002	-3,281	-3,375	-0,691	-0,670
TBP43	-0,573	1,481	1,711	1,684	1,672	2,002	1,633
TBP22	0,172	0,657	0,640	0,555	0,525	-0,375	-0,520
TBP26	0,329	0,011	-0,198	-0,328	-0,373	-0,864	-0,820
TBP16	0,511	-0,254	-0,543	-0,691	-0,742	-0,410	-0,219
TBP28	4,058	0,800	0,826	0,752	0,724	0,592	0,682
TBP31	-1,664	-1,545	-0,878	0,310	0,784	1,955	2,034
TBP40	0,547	1,137	1,302	1,252	1,233	1,260	1,283
TBP27	0,066	0,599	0,566	0,477	0,445	0,540	0,432
TBP2	0,322	-0,053	-0,282	-0,416	-0,463	-0,381	-0,319
TBP30	0,436	0,657	0,640	0,555	0,525	0,049	0,081
TBP23	0,299	-0,003	-0,217	-0,347	-0,393	-0,043	-0,169
TBP45	-0,790	-0,075	-0,310	-0,445	-0,492	-1,395	-1,721
TBP17	0,513	0,212	0,063	-0,053	-0,094	-0,389	-0,219
TBP20	0,084	-0,318	-0,627	-0,779	-0,832	-0,961	-0,820
TBP37	-1,602	0,478	0,407	0,310	0,275	0,050	0,382
TBP11	-1,031	-0,806	-1,260	-1,446	-1,510	-0,307	-0,319
TBP34	0,736	0,693	0,687	0,604	0,575	0,732	0,832
TBP8	0,497	0,105	-0,077	-0,200	-0,243	-0,591	-0,520
TBP36	-1,443	-1,458	-0,012	0,899	1,023	1,817	1,734
TBP4	0,320	0,026	-0,170	-0,298	-0,343	-0,892	-0,820
TBP39	-1,457	1,625	1,898	1,880	1,871	1,319	1,033
TBP29	1,653	0,979	1,059	0,997	0,974	0,393	0,382
TBP3	0,177	0,026	-0,170	-0,298	-0,343	-0,896	-0,820
TBP12	0,343	-0,125	-0,375	-0,514	-0,562	-0,777	-0,820
TBP41	-1,979	1,847	2,187	2,184	2,180	1,640	1,333
TBP44	-0,937	-0,189	-0,459	-0,602	-0,652	-1,439	-1,721
TBP15	0,193	0,399	0,305	0,202	0,166	-0,544	-0,520
TBP33	-0,104	0,277	0,146	0,035	-0,004	0,138	0,382
TBP7	-0,622	-0,153	-0,413	-0,553	-0,602	-0,771	-0,520
TBP13	0,323	-0,125	-0,375	-0,514	-0,562	-0,765	-0,820
TBP24	0,128	0,736	0,743	0,663	0,634	-0,035	-0,219
TBP10	0,475	0,370	0,267	0,163	0,126	-0,119	0,081
TBP46	-1,295	0,033	-0,170	-0,298	-0,343	-1,318	-1,721
TBP42	-0,485	1,460	1,711	1,684	1,672	1,549	1,583
TBP38	-0,789	0,635	0,649	0,565	0,535	0,868	1,283
TBP5	0,320	-0,089	-0,329	-0,465	-0,512	-0,924	-0,820
TBP21	0,300	0,341	0,230	0,124	0,086	-0,738	-0,820
TBP19	-0,080	-0,663	-1,074	-1,250	-1,310	-0,310	-0,369
TBP14	0,878	-2,075	-2,704	-1,888	-1,101	-0,388	-0,520
TBP35	0,556	-1,659	0,053	0,683	0,714	1,834	1,734
TBP18	0,741	-0,189	-0,459	-0,602	-0,652	-0,607	-0,520
TBP1	-0,851	-0,519	-0,888	-1,054	-1,111	-1,119	-0,820

Tabela C (cont.). Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método B3LYP/6-31G*) para os 45 compostos aleatorizados

Composto	Randic	Balaban	Harary	Hyper Wiener	Szeged	Wiener	Wiener polarity
TBP25	-0,499	0,627	-0,581	-0,655	-0,652	-0,631	-0,500
TBP32	1,654	-1,331	1,569	1,881	1,781	1,762	1,444
TBP9	-0,679	0,382	-0,611	-0,581	-0,609	-0,587	-0,603
TBP43	1,962	-1,393	2,071	2,226	2,211	2,230	1,956
TBP22	-0,610	1,300	-0,180	-0,294	-0,322	-0,257	-0,193
TBP26	-0,889	0,688	-0,790	-0,734	-0,743	-0,758	-0,705
TBP16	-0,327	0,872	-0,772	-0,786	-0,779	-0,786	-0,705
TBP28	0,697	-1,087	0,900	0,138	0,604	0,354	1,239
TBP31	1,934	-1,331	1,569	1,881	1,781	1,803	1,444
TBP40	1,273	-0,597	1,248	0,860	1,034	0,984	1,649
TBP27	0,581	-1,209	0,301	0,781	0,485	0,617	0,011
TBP2	-0,288	0,321	-0,390	-0,498	-0,529	-0,457	-0,296
TBP30	0,186	-0,352	-0,008	0,027	0,020	0,014	-0,091
TBP23	-0,219	1,117	0,021	-0,126	-0,161	-0,068	0,011
TBP45	-1,368	1,055	-0,980	-0,887	-0,891	-0,918	-1,012
TBP17	-0,327	0,811	-0,784	-0,763	-0,761	-0,772	-0,807
TBP20	-0,889	0,811	-0,784	-0,763	-0,761	-0,772	-0,807
TBP37	0,302	-1,576	0,264	0,213	0,171	0,206	-0,091
TBP11	-0,286	-0,046	-0,449	-0,357	-0,439	-0,375	-0,603
TBP34	0,776	-1,760	0,755	0,599	0,670	0,611	1,035
TBP8	-0,610	0,137	-0,399	-0,427	-0,466	-0,418	-0,398
TBP36	1,654	-1,331	1,569	1,881	1,781	1,762	1,444
TBP4	-0,889	0,994	-0,990	-0,860	-0,872	-0,903	-1,012
TBP39	1,384	-0,964	1,636	1,441	1,538	1,583	1,444
TBP29	0,473	-0,781	0,145	0,374	0,243	0,301	-0,091
TBP3	-0,889	1,055	-0,980	-0,887	-0,891	-0,918	-1,012
TBP12	-0,889	0,811	-0,780	-0,760	-0,761	-0,772	-0,705
TBP41	1,673	-1,148	1,860	1,800	1,729	1,877	1,853
TBP44	-1,368	1,055	-0,980	-0,887	-0,891	-0,918	-1,012
TBP15	-0,610	0,688	-0,790	-0,737	-0,743	-0,758	-0,705
TBP33	0,283	-1,270	0,347	0,034	0,121	0,100	0,523
TBP7	-0,610	1,055	-0,980	-0,887	-0,891	-0,918	-1,012
TBP13	-0,889	0,872	-0,780	-0,786	-0,779	-0,786	-0,705
TBP24	-0,327	0,811	0,481	0,211	0,164	0,314	0,421
TBP10	-0,031	0,382	-0,611	-0,581	-0,609	-0,587	-0,603
TBP46	-1,368	0,688	-0,797	-0,729	-0,735	-0,752	-0,807
TBP42	1,562	-0,842	1,452	1,202	1,458	1,285	1,751
TBP38	1,099	-1,637	1,316	0,810	1,066	0,952	1,239
TBP5	-0,889	1,055	-0,980	-0,887	-0,891	-0,918	-1,012
TBP21	-0,889	0,811	-0,784	-0,763	-0,761	-0,772	-0,807
TBP19	-0,397	0,076	-0,411	-0,389	-0,431	-0,400	-0,296
TBP14	-0,610	0,872	-0,772	-0,786	-0,779	-0,786	-0,705
TBP35	1,654	-1,454	1,792	2,244	2,094	2,080	1,649
TBP18	-0,610	0,872	-0,767	-0,787	-0,782	-0,788	-0,603
TBP1	-0,889	-0,107	-1,166	-1,008	-1,018	-1,060	-1,217

Tabela D. Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método HF/6-31G*) para os 45 compostos aleatorizados

Composto	E _{EST}	E _{ang}	E _{TORS}	E _{1,4}	E _{vdW}	E _{EL}	E _{SOLV}	E _{HBD}	E _{TOT}
TBP25	-0,335	0,164	0,278	-0,510	-0,712	-0,155	-0,509	-0,820	-0,346
TBP32	1,148	1,385	2,187	1,578	-0,410	-1,281	0,578	0,715	1,554
TBP9	-0,262	-0,542	0,369	0,556	0,216	-2,701	-2,762	0,445	-0,471
TBP43	2,764	2,221	-1,015	1,074	2,516	0,308	0,258	0,009	2,016
TBP22	-0,544	-0,595	-0,021	0,079	-0,450	-0,550	-0,839	0,086	-0,487
TBP26	-0,869	-0,609	0,057	-0,507	-0,236	0,503	-0,481	0,118	-0,575
TBP16	0,680	0,000	0,444	-0,516	-0,273	0,610	1,308	0,069	0,145
TBP28	0,645	-0,329	-1,786	1,261	-0,263	-1,146	0,850	0,715	0,079
TBP31	1,711	1,105	1,784	1,645	0,280	-0,508	1,394	0,715	1,786
TBP40	1,634	1,461	0,342	2,011	0,590	-1,606	0,900	0,715	1,647
TBP27	-0,124	0,172	-0,145	-0,500	-0,431	1,426	0,143	-0,170	-0,049
TBP2	-0,713	-0,206	0,161	-0,536	0,045	0,452	0,718	-5,947	-0,459
TBP30	0,332	0,155	-1,930	0,180	0,048	-0,022	0,263	0,123	-0,123
TBP23	-0,155	-0,067	0,748	-0,183	-0,327	-0,589	-1,317	-0,206	-0,218
TBP45	-1,190	-1,728	-1,314	-0,664	-0,260	0,407	-0,621	0,200	-1,361
TBP17	-0,368	-0,591	0,083	-0,829	-0,482	0,977	1,440	0,136	-0,432
TBP20	-0,635	-0,612	0,245	-1,495	-0,244	1,147	-0,632	0,075	-0,810
TBP37	0,858	0,883	1,142	-0,229	1,468	0,805	0,414	-0,037	0,948
TBP11	-0,691	-0,190	0,044	-2,217	-0,728	2,446	-1,513	-1,039	-0,976
TBP34	0,226	1,320	-0,833	0,529	-0,763	-0,742	0,965	0,371	0,434
TBP8	-0,863	-0,580	0,296	-0,093	-0,255	-0,186	-0,741	0,106	-0,472
TBP36	1,756	0,611	1,815	1,792	0,016	-1,131	0,118	0,715	1,475
TBP4	-0,941	-0,916	0,349	-0,648	-0,038	0,461	0,112	0,126	-0,639
TBP39	1,644	1,017	-0,563	1,138	3,594	-0,349	0,088	-1,377	1,485
TBP29	-0,471	0,057	-0,428	0,013	-0,747	0,887	0,426	0,127	-0,100
TBP3	-0,535	-0,733	0,495	-0,691	-0,517	0,531	0,307	0,120	-0,531
TBP12	-0,978	-0,801	0,185	-0,461	-0,522	0,472	-0,473	0,055	-0,671
TBP41	1,340	2,023	-0,796	0,820	3,292	-0,146	0,104	0,009	1,636
TBP44	-1,330	-1,701	-1,564	-0,684	-0,142	0,433	-1,198	0,212	-1,450
TBP15	-0,397	-0,595	0,032	-0,665	-0,485	0,692	0,853	0,112	-0,463
TBP33	0,355	0,624	-1,575	-0,442	-1,081	0,164	0,664	0,104	-0,250
TBP7	0,040	-0,673	-0,510	-0,855	-0,247	0,898	0,676	0,113	-0,522
TBP13	-0,887	-0,457	-0,315	-0,596	-0,482	0,500	-0,619	0,049	-0,673
TBP24	-0,507	-0,394	-0,076	0,008	-0,209	-0,951	-1,861	0,077	-0,535
TBP10	0,050	-0,474	0,626	-0,316	-0,554	0,399	1,340	0,104	-0,106
TBP46	-1,125	-1,691	-2,204	-0,657	-0,289	0,658	-1,006	0,062	-1,489
TBP42	1,413	1,517	0,398	2,200	1,144	-1,473	1,403	0,715	1,840
TBP38	0,355	1,905	1,157	1,517	1,128	-0,587	1,431	0,715	1,748
TBP5	-0,668	-0,638	0,070	-0,846	-0,383	0,520	-0,266	0,117	-0,681
TBP21	-0,655	-0,904	0,776	-0,916	-0,509	0,582	-0,609	0,091	-0,708
TBP19	-0,993	-0,090	-0,030	-0,150	0,056	-2,317	-2,297	0,407	-0,750
TBP14	-0,805	-1,048	-1,150	-0,565	-0,300	0,430	0,140	0,082	-0,932
TBP35	1,617	1,468	2,065	1,671	-1,316	-1,436	0,043	0,715	1,504
TBP18	-0,755	-0,329	0,055	-0,587	-0,525	0,450	0,635	0,050	-0,462
TBP1	-0,773	-0,595	0,050	-0,712	-0,214	0,715	0,174	0,125	-0,556

Tabela D (cont.). Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método HF/6-31G*) para os 45 compostos aleatorizados

Composto	E _{HOMO}	E _{LUMO}	μ_T	μ_X	μ_Y	μ_Z	carga_N ₁	carga_C ₂
TBP25	-0,319	0,366	-0,027	-1,001	-0,214	-0,851	0,984	-0,247
TBP32	-0,049	0,996	-0,649	-0,083	-0,367	-1,209	0,100	-0,764
TBP9	-0,501	-3,537	2,746	0,504	-2,273	-0,858	-0,978	0,811
TBP43	0,205	-0,648	0,114	1,225	1,237	-0,665	1,367	0,917
TBP22	-0,456	-0,518	-1,624	0,753	0,963	0,751	-1,497	1,133
TBP26	-0,464	0,175	0,300	0,621	-1,168	1,561	0,151	-0,666
TBP16	-0,146	1,110	0,680	-1,049	-0,433	-1,246	-0,789	0,020
TBP28	0,567	-1,697	1,175	1,361	1,932	0,168	-1,713	0,698
TBP31	0,124	1,009	-0,168	1,200	0,776	1,733	1,505	-2,168
TBP40	0,540	-1,270	0,525	-0,884	-1,077	-0,360	3,499	-1,796
TBP27	-0,123	0,898	-0,984	0,990	1,126	0,534	0,965	-1,035
TBP2	-0,400	-0,655	1,265	-1,256	-0,936	-0,840	0,025	0,287
TBP30	0,155	-0,494	0,991	1,347	0,413	-1,776	0,593	-0,162
TBP23	-0,401	-0,147	1,269	-1,165	-1,138	-0,641	-0,417	0,627
TBP45	-0,162	0,425	0,292	0,984	1,540	-0,720	0,839	0,103
TBP17	-0,104	1,114	0,825	-1,311	-0,763	-0,353	-0,653	1,413
TBP20	-0,487	0,326	-0,570	-0,554	-0,827	-0,133	-0,520	0,532
TBP37	0,242	-0,131	-1,892	0,320	-0,323	1,209	-0,257	-1,510
TBP11	-0,356	0,697	-0,379	-0,898	0,041	-0,852	0,394	0,253
TBP34	0,161	-0,142	1,138	1,732	1,278	-0,650	-0,490	0,757
TBP8	-0,485	0,067	-1,930	-0,351	0,086	-0,445	-0,281	-0,492
TBP36	0,014	0,753	-0,296	-0,024	1,270	-1,107	-0,507	0,496
TBP4	-0,315	0,287	-0,256	-0,717	-0,922	0,421	0,095	-0,537
TBP39	0,185	-0,145	-1,600	-0,080	-0,581	-0,157	0,214	-1,118
TBP29	-0,101	0,692	1,155	1,630	-0,028	2,263	-1,767	1,327
TBP3	-0,307	0,418	0,889	1,058	1,792	-0,757	-1,130	1,186
TBP12	-0,439	0,098	-0,697	-0,919	-0,355	0,350	-0,340	0,489
TBP41	0,186	-0,629	-0,469	-0,770	-0,551	-0,460	-0,605	1,263
TBP44	-0,171	0,576	0,221	-0,630	-1,115	-0,342	0,111	0,350
TBP15	-0,370	0,424	-0,209	-0,756	-0,837	-0,184	0,012	0,374
TBP33	0,231	-0,163	0,296	-1,291	-0,267	-0,424	0,658	-0,094
TBP7	-0,347	1,170	0,522	1,383	1,573	0,298	0,085	0,159
TBP13	-0,422	0,036	0,431	1,421	1,435	-0,062	0,047	-0,830
TBP24	-0,534	-0,724	-0,847	0,657	-0,804	1,032	-0,191	-0,723
TBP10	-0,224	1,062	-0,044	1,289	-0,268	1,674	-1,023	0,368
TBP46	-0,224	0,404	0,914	-0,878	1,570	1,769	1,804	0,836
TBP42	0,705	-0,821	-0,979	-0,508	-0,481	1,279	0,502	-0,614
TBP38	0,787	-1,106	-1,566	0,608	0,281	-0,807	-1,468	1,099
TBP5	-0,384	0,571	0,688	-0,558	-1,204	-0,857	1,206	-0,128
TBP21	-0,432	0,046	-0,679	-0,504	1,190	-0,528	-1,027	0,943
TBP19	-0,536	-3,081	-0,741	0,923	-0,518	1,352	0,253	-0,591
TBP14	-0,294	-0,266	1,553	-1,789	-0,109	-0,681	1,027	-3,488
TBP35	6,164	0,762	-0,900	-0,066	-0,012	2,026	-0,322	-0,377
TBP18	-0,324	0,638	-0,548	-1,042	-0,255	0,110	-0,576	0,362
TBP1	-0,388	1,056	0,069	-0,922	-0,676	-0,565	0,113	0,534

Tabela D (cont.). Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método HF/6-31G*) para os 45 compostos aleatorizados

Composto	carga_C ₃	carga_C ₄	carga_C ₅	carga_O ₆	carga_H ₇	carga_C ₈	carga_O ₉
TBP25	-0,471	1,295	-1,223	-0,013	-1,037	-0,722	0,012
TBP32	2,014	-1,997	0,660	0,852	1,054	2,217	-0,968
TBP9	-1,201	-0,274	1,493	-0,467	0,054	0,423	0,101
TBP43	-1,644	0,880	-1,456	1,051	-0,718	0,026	-0,266
TBP22	-0,948	0,048	0,468	-0,157	0,153	0,130	-0,389
TBP26	0,259	0,767	-0,717	0,774	-0,565	-0,399	0,181
TBP16	-0,858	1,565	-1,257	0,442	-1,099	-0,111	-1,118
TBP28	-0,568	0,286	2,076	-0,578	-0,996	-2,374	0,704
TBP31	2,104	-1,541	0,421	0,264	0,800	0,321	0,497
TBP40	0,207	-0,458	-0,962	-0,002	0,724	-1,853	1,806
TBP27	-0,497	1,342	-1,264	1,340	-0,989	-0,392	0,285
TBP2	0,273	-1,173	0,604	-0,179	0,063	2,076	-1,519
TBP30	-0,092	0,418	-1,110	-0,235	-0,299	0,469	-0,591
TBP23	-1,522	0,503	-0,150	1,251	-0,799	0,897	-0,539
TBP45	0,497	-0,599	1,447	-2,573	-0,106	-0,341	0,040
TBP17	-0,963	0,862	-0,191	-0,312	-0,654	-0,459	0,412
TBP20	-0,333	-0,085	0,149	-0,611	0,178	0,065	-0,134
TBP37	2,611	-1,794	0,557	0,508	1,127	1,161	0,487
TBP11	0,292	0,223	-0,813	-0,722	0,242	-0,400	0,902
TBP34	-0,657	0,837	-0,376	-1,509	-0,989	-0,134	-0,968
TBP8	0,499	0,571	-0,833	0,852	-0,572	0,503	-0,756
TBP36	1,102	-1,240	1,057	-1,166	0,467	-0,015	0,727
TBP4	0,217	0,670	-1,119	0,730	-0,673	0,516	-0,751
TBP39	1,657	-1,413	0,665	0,165	0,680	0,530	0,534
TBP29	-2,202	-0,403	2,091	-0,833	0,045	1,227	-1,599
TBP3	-0,802	0,110	0,721	-0,467	-0,404	-0,214	0,064
TBP12	0,053	0,324	-0,331	-0,423	-0,255	0,031	-0,228
TBP41	0,445	0,428	-1,375	-1,875	0,538	-0,654	0,963
TBP44	1,092	-0,323	0,791	-1,875	0,174	-0,593	0,798
TBP15	-0,405	-0,094	0,184	0,763	0,082	0,186	-0,586
TBP33	-0,460	0,898	-0,962	0,741	-0,822	-0,030	-0,600
TBP7	0,190	-0,129	-0,045	-0,323	0,057	0,086	-0,082
TBP13	0,292	0,479	-0,323	0,996	-0,572	-0,311	0,191
TBP24	0,562	0,015	0,185	1,417	-0,411	0,216	0,092
TBP10	-0,570	0,011	-0,102	0,065	0,091	1,188	-1,792
TBP46	-0,389	-1,180	0,248	-1,820	4,796	0,780	-0,360
TBP42	0,562	1,017	-1,626	-0,323	-0,537	-2,430	4,184
TBP38	-0,203	-1,571	1,787	-0,645	1,334	-0,573	0,709
TBP5	-0,870	0,990	-0,946	0,752	-0,640	-0,460	0,483
TBP21	-0,264	-0,277	0,427	0,098	0,458	0,030	-0,031
TBP19	0,566	0,269	-0,362	0,807	-0,280	-0,671	1,123
TBP14	0,031	1,348	0,334	2,315	-0,751	-1,579	-0,525
TBP35	1,657	-2,909	1,728	0,120	1,772	2,533	-1,481
TBP18	-1,048	0,684	0,439	1,229	-0,650	-0,704	-0,280
TBP1	-0,219	0,621	-0,989	-0,423	-0,072	-0,194	0,271

Tabela D (cont.). Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método HF/6-31G*) para os 45 compostos aleatorizados

Composto	carga_N ₁₀	Vol. Vw	ClogP	RM	α	Área superf.	Θ_1	Θ_2
TBP25	0,285	-0,473	-1,079	-0,561	-0,523	-0,409	-1,820	0,246
TBP32	-0,913	1,628	0,795	1,510	1,439	1,540	0,717	0,027
TBP9	-0,530	-0,823	-1,389	-0,770	-0,858	-0,670	-0,407	0,294
TBP43	-0,009	2,104	2,538	2,088	2,004	2,267	-0,258	0,260
TBP22	-0,203	-0,393	0,334	-0,437	-0,550	-0,188	0,107	0,261
TBP26	-0,109	-0,867	-0,688	-0,893	-0,859	-0,858	0,437	0,248
TBP16	-0,283	-0,565	-0,407	-0,624	-0,619	-0,752	-1,013	0,398
TBP28	3,799	0,699	0,915	0,848	1,051	0,154	1,518	-6,307
TBP31	0,779	1,851	1,116	1,767	1,691	1,827	-0,411	-0,314
TBP40	2,030	1,285	1,406	1,348	1,534	0,874	-0,391	0,302
TBP27	-0,406	0,536	0,184	0,491	0,494	0,583	0,864	0,265
TBP2	-1,773	-0,437	-1,610	-0,547	-0,523	-0,673	0,450	0,284
TBP30	-0,482	0,092	0,344	0,176	0,168	0,146	-1,682	0,349
TBP23	-0,968	-0,298	-0,708	-0,499	-0,570	-0,157	-0,115	0,298
TBP45	0,183	-1,056	-0,458	-0,820	-0,830	-1,040	-0,323	-0,038
TBP17	-0,093	-0,583	-0,407	-0,624	-0,619	-0,461	1,306	0,361
TBP20	-0,116	-1,063	-1,069	-1,136	-1,124	-1,010	0,576	0,239
TBP37	-0,671	0,247	0,174	0,314	0,309	0,172	0,686	-0,525
TBP11	-0,057	-0,439	-1,850	-0,568	-0,523	-0,148	-1,718	-0,022
TBP34	-0,160	0,736	0,134	0,816	0,727	0,884	0,113	0,410
TBP8	-0,876	-0,741	-0,458	-0,842	-0,895	-0,749	0,391	0,332
TBP36	1,096	1,846	1,166	1,754	1,704	2,009	-2,257	-0,415
TBP4	-0,731	-0,775	-0,548	-0,755	-0,755	-0,750	0,280	0,288
TBP39	-0,144	1,478	1,747	1,600	1,522	1,760	0,100	-0,147
TBP29	-1,000	0,404	0,745	0,420	0,410	0,788	-1,072	0,699
TBP3	0,051	-0,787	-0,548	-0,755	-0,755	-0,730	0,789	0,254
TBP12	-0,331	-0,671	-0,307	-0,649	-0,594	-0,686	0,950	0,286
TBP41	0,327	1,791	2,148	1,844	1,764	2,025	-1,823	-0,563
TBP44	0,264	-1,181	-0,728	-0,970	-0,922	-1,090	-0,028	-0,079
TBP15	-0,350	-0,496	-0,087	-0,487	-0,513	-0,514	0,393	0,270
TBP33	-0,100	0,227	0,374	0,371	0,309	0,126	0,833	0,206
TBP7	-0,321	-0,878	-0,878	-0,892	-0,859	-0,876	-0,709	0,001
TBP13	-0,168	-0,657	-0,307	-0,649	-0,594	-0,654	-0,399	0,293
TBP24	-0,732	-0,316	0,424	-0,525	-0,690	-0,163	0,382	0,269
TBP10	-1,105	-0,289	-0,147	-0,406	-0,377	-0,379	0,614	0,267
TBP46	-0,378	-1,019	-0,317	-0,809	-0,843	-0,995	-0,754	-0,398
TBP42	2,333	1,541	1,797	1,593	1,775	1,107	0,474	0,044
TBP38	2,343	1,073	0,835	1,243	1,432	0,531	0,938	-0,323
TBP5	-0,108	-0,895	-0,828	-0,904	-0,847	-0,841	0,206	0,288
TBP21	-0,158	-0,641	-0,307	-0,649	-0,594	-0,554	1,280	0,292
TBP19	0,106	-0,556	-0,919	-0,503	-0,617	-0,474	-1,156	0,241
TBP14	0,478	-0,556	-0,357	-0,636	-0,605	-0,710	2,492	0,334
TBP35	-1,001	1,700	0,925	1,522	1,427	1,782	-0,947	0,153
TBP18	0,482	-0,628	-0,357	-0,636	-0,605	-0,802	1,053	0,465
TBP1	-0,281	-1,158	-1,339	-1,159	-1,101	-1,245	-0,665	-0,090

Tabela D (cont.). Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método HF/6-31G*) para os 45 compostos aleatorizados

Composto	Θ_3	ClogD pH1,5	ClogD pH5	ClogD pH6,5	ClogD pH7,4	MSA	Platt
TBP25	0,161	-0,204	-0,478	-0,622	-0,672	-0,396	-0,470
TBP32	0,312	-1,795	-0,618	0,369	0,545	1,707	1,734
TBP9	0,253	-3,287	-3,002	-3,281	-3,375	-0,691	-0,670
TBP43	-0,573	1,481	1,711	1,684	1,672	2,002	1,633
TBP22	0,172	0,657	0,640	0,555	0,525	-0,375	-0,520
TBP26	0,329	0,011	-0,198	-0,328	-0,373	-0,864	-0,820
TBP16	0,511	-0,254	-0,543	-0,691	-0,742	-0,410	-0,219
TBP28	4,058	0,800	0,826	0,752	0,724	0,592	0,682
TBP31	-1,664	-1,545	-0,878	0,310	0,784	1,955	2,034
TBP40	0,547	1,137	1,302	1,252	1,233	1,260	1,283
TBP27	0,066	0,599	0,566	0,477	0,445	0,540	0,432
TBP2	0,322	-0,053	-0,282	-0,416	-0,463	-0,381	-0,319
TBP30	0,436	0,657	0,640	0,555	0,525	0,049	0,081
TBP23	0,299	-0,003	-0,217	-0,347	-0,393	-0,043	-0,169
TBP45	-0,790	-0,075	-0,310	-0,445	-0,492	-1,395	-1,721
TBP17	0,513	0,212	0,063	-0,053	-0,094	-0,389	-0,219
TBP20	0,084	-0,318	-0,627	-0,779	-0,832	-0,961	-0,820
TBP37	-1,602	0,478	0,407	0,310	0,275	0,050	0,382
TBP11	-1,031	-0,806	-1,260	-1,446	-1,510	-0,307	-0,319
TBP34	0,736	0,693	0,687	0,604	0,575	0,732	0,832
TBP8	0,497	0,105	-0,077	-0,200	-0,243	-0,591	-0,520
TBP36	-1,443	-1,458	-0,012	0,899	1,023	1,817	1,734
TBP4	0,320	0,026	-0,170	-0,298	-0,343	-0,892	-0,820
TBP39	-1,457	1,625	1,898	1,880	1,871	1,319	1,033
TBP29	1,653	0,979	1,059	0,997	0,974	0,393	0,382
TBP3	0,177	0,026	-0,170	-0,298	-0,343	-0,896	-0,820
TBP12	0,343	-0,125	-0,375	-0,514	-0,562	-0,777	-0,820
TBP41	-1,979	1,847	2,187	2,184	2,180	1,640	1,333
TBP44	-0,937	-0,189	-0,459	-0,602	-0,652	-1,439	-1,721
TBP15	0,193	0,399	0,305	0,202	0,166	-0,544	-0,520
TBP33	-0,104	0,277	0,146	0,035	-0,004	0,138	0,382
TBP7	-0,622	-0,153	-0,413	-0,553	-0,602	-0,771	-0,520
TBP13	0,323	-0,125	-0,375	-0,514	-0,562	-0,765	-0,820
TBP24	0,128	0,736	0,743	0,663	0,634	-0,035	-0,219
TBP10	0,475	0,370	0,267	0,163	0,126	-0,119	0,081
TBP46	-1,295	0,033	-0,170	-0,298	-0,343	-1,318	-1,721
TBP42	-0,485	1,460	1,711	1,684	1,672	1,549	1,583
TBP38	-0,789	0,635	0,649	0,565	0,535	0,868	1,283
TBP5	0,320	-0,089	-0,329	-0,465	-0,512	-0,924	-0,820
TBP21	0,300	0,341	0,230	0,124	0,086	-0,738	-0,820
TBP19	-0,080	-0,663	-1,074	-1,250	-1,310	-0,310	-0,369
TBP14	0,878	-2,075	-2,704	-1,888	-1,101	-0,388	-0,520
TBP35	0,556	-1,659	0,053	0,683	0,714	1,834	1,734
TBP18	0,741	-0,189	-0,459	-0,602	-0,652	-0,607	-0,520
TBP1	-0,851	-0,519	-0,888	-1,054	-1,111	-1,119	-0,820

Tabela D (cont.). Variáveis independentes calculadas e autoescaladas (descritores eletrônicos calculados com método HF/6-31G*) para os 45 compostos aleatorizados

Composto	Randic	Balaban	Harary	Hyper Wiener	Szeged	Wiener	Wiener polarity
TBP25	-0,499	0,627	-0,581	-0,655	-0,652	-0,631	-0,500
TBP32	1,654	-1,331	1,569	1,881	1,781	1,762	1,444
TBP9	-0,679	0,382	-0,611	-0,581	-0,609	-0,587	-0,603
TBP43	1,962	-1,393	2,071	2,226	2,211	2,230	1,956
TBP22	-0,610	1,300	-0,180	-0,294	-0,322	-0,257	-0,193
TBP26	-0,889	0,688	-0,790	-0,734	-0,743	-0,758	-0,705
TBP16	-0,327	0,872	-0,772	-0,786	-0,779	-0,786	-0,705
TBP28	0,697	-1,087	0,900	0,138	0,604	0,354	1,239
TBP31	1,934	-1,331	1,569	1,881	1,781	1,803	1,444
TBP40	1,273	-0,597	1,248	0,860	1,034	0,984	1,649
TBP27	0,581	-1,209	0,301	0,781	0,485	0,617	0,011
TBP2	-0,288	0,321	-0,390	-0,498	-0,529	-0,457	-0,296
TBP30	0,186	-0,352	-0,008	0,027	0,020	0,014	-0,091
TBP23	-0,219	1,117	0,021	-0,126	-0,161	-0,068	0,011
TBP45	-1,368	1,055	-0,980	-0,887	-0,891	-0,918	-1,012
TBP17	-0,327	0,811	-0,784	-0,763	-0,761	-0,772	-0,807
TBP20	-0,889	0,811	-0,784	-0,763	-0,761	-0,772	-0,807
TBP37	0,302	-1,576	0,264	0,213	0,171	0,206	-0,091
TBP11	-0,286	-0,046	-0,449	-0,357	-0,439	-0,375	-0,603
TBP34	0,776	-1,760	0,755	0,599	0,670	0,611	1,035
TBP8	-0,610	0,137	-0,399	-0,427	-0,466	-0,418	-0,398
TBP36	1,654	-1,331	1,569	1,881	1,781	1,762	1,444
TBP4	-0,889	0,994	-0,990	-0,860	-0,872	-0,903	-1,012
TBP39	1,384	-0,964	1,636	1,441	1,538	1,583	1,444
TBP29	0,473	-0,781	0,145	0,374	0,243	0,301	-0,091
TBP3	-0,889	1,055	-0,980	-0,887	-0,891	-0,918	-1,012
TBP12	-0,889	0,811	-0,780	-0,760	-0,761	-0,772	-0,705
TBP41	1,673	-1,148	1,860	1,800	1,729	1,877	1,853
TBP44	-1,368	1,055	-0,980	-0,887	-0,891	-0,918	-1,012
TBP15	-0,610	0,688	-0,790	-0,737	-0,743	-0,758	-0,705
TBP33	0,283	-1,270	0,347	0,034	0,121	0,100	0,523
TBP7	-0,610	1,055	-0,980	-0,887	-0,891	-0,918	-1,012
TBP13	-0,889	0,872	-0,780	-0,786	-0,779	-0,786	-0,705
TBP24	-0,327	0,811	0,481	0,211	0,164	0,314	0,421
TBP10	-0,031	0,382	-0,611	-0,581	-0,609	-0,587	-0,603
TBP46	-1,368	0,688	-0,797	-0,729	-0,735	-0,752	-0,807
TBP42	1,562	-0,842	1,452	1,202	1,458	1,285	1,751
TBP38	1,099	-1,637	1,316	0,810	1,066	0,952	1,239
TBP5	-0,889	1,055	-0,980	-0,887	-0,891	-0,918	-1,012
TBP21	-0,889	0,811	-0,784	-0,763	-0,761	-0,772	-0,807
TBP19	-0,397	0,076	-0,411	-0,389	-0,431	-0,400	-0,296
TBP14	-0,610	0,872	-0,772	-0,786	-0,779	-0,786	-0,705
TBP35	1,654	-1,454	1,792	2,244	2,094	2,080	1,649
TBP18	-0,610	0,872	-0,767	-0,787	-0,782	-0,788	-0,603
TBP1	-0,889	-0,107	-1,166	-1,008	-1,018	-1,060	-1,217

Tabela E. Valores de correlação linear entre as variáveis independentes e a atividade biológica.

	μ_T	μ_X	μ_Y	μ_Z	carga_N ₁	carga_C ₂	carga_C ₃
B3LYP/3-21G*	-0,12	0,09	0,06	-0,13	-0,07	-0,03	-0,16
HF/3-21G*	-0,15	0,10	0,06	-0,11	-0,08	-0,02	-0,14
B3LYP/6-31G*	-0,11	0,11	0,05	-0,09	-0,09	-0,01	-0,11
HF/6-31G*	-0,13	0,10	0,06	-0,10	-0,10	0,00	-0,07
	carga_C ₄	carga_C ₅	carga_O ₆	carga_H ₇	carga_C ₈	carga_O ₉	carga_N ₁₀
B3LYP/3-21G*	0,01	0,10	0,07	0,00	-0,20	0,18	0,35
HF/3-21G*	0,03	0,07	-0,09	-0,05	-0,23	0,25	0,34
B3LYP/6-31G*	-0,14	0,08	0,03	0,13	-0,21	-0,18	0,36
HF/6-31G*	-0,01	0,06	-0,04	-0,06	-0,22	0,24	0,34

E _{EST}	E _{ang}	E _{TORS}	E _{1,4}	E _{vdW}	E _{EL}	E _{SOLV}	E _{HBD}
0,46	0,44	-0,07	0,43	0,44	-0,21	0,26	0,19
E _{TOT}	E _{HOMO}	E _{LUMO}	Vol. Vw	ClogP	RM	α	Área superf.
0,47	0,32	-0,13	0,46	0,57	0,47	0,48	0,45
θ_1	θ_2	θ_3	ClogD pH1,5	ClogD pH5	ClogD pH6,5	ClogD pH7,4	MSA
0,13	-0,16	0,04	0,46	0,54	0,56	0,58	0,43
Platt	Randic	Balaban	Harary	Hyper Wiener	Szeged	Wiener	Wiener polarity
0,43	0,45	-0,38	0,48	0,40	0,44	0,43	0,47

Quadro A. Conjunto de 18 estruturas da classe química arilamidas e respectivos valores de atividade biológica expressos em Cl_{50} ($\mu\text{mol/L}$) e pCl_{50} .

Composto	R ₁	R ₂	n	X	Cl ₅₀	pCl ₅₀
TBA1	H	H	0	N	38,86 ± 1,35	4,41
TBA2	4-CH ₃	H	0	N	16,64 ± 0,49	4,78
TBA3	4-CH ₃	3-CF ₃	0	N	6,26 ± 0,33	5,21
TBA4	4-CH ₃	3-Cl	0	N	3,07 ± 0,48	5,51
TBA4	3-CH ₃	3-Cl	0	N	9,43 ± 0,80	5,03
TBA6	3-CH ₃	4-NO ₂	0	N	15,47 ± 1,52	4,81
TBA7	3,4-Me ₂	3-Cl	0	N	0,99 ± 0,03	6,00
TBA8	3,4-Me ₂	3-CF ₃	0	N	1,85 ± 0,08	5,73
TBA9	2-F	3-Cl	0	N	13,87 ± 1,03	4,86
TBA10	4-F	3-Cl	0	N	9,74 ± 0,62	5,01
TBA11	3-Cl	3-Cl	0	N	6,73 ± 0,27	5,17
TBA12	3,4-Cl ₂	3-Cl	0	N	6,05 ± 0,58	5,22
TBA13	3,4-Cl ₂	H	0	N	17,62 ± 1,24	4,75
TBA14	H	H	1	N	31,50 ± 1,65	4,50
TBA15	3-Cl	H	1	C	7,74 ± 0,25	5,11
TBA16	2-F	H	1	C	14,11 ± 0,42	4,85
TBA17	4-CH ₃	H	1	C	5,16 ± 0,45	5,29
TBA18	3-CH ₃	H	1	C	7,39 ± 0,38	5,13

Quadro B. Conjunto de 6 estruturas da classe química arilamidas com substituintes volumosos e respectivos valores de atividade biológica expressos em CI_{50} ($\mu\text{mol/L}$) e pCI_{50} .

Composto	Estrutura	CI_{50}	pCI_{50}
TBA19		$0,40 \pm 0,02$	6,40
TBA20		$0,09 \pm 0,00$	7,05
TB21		$0,20 \pm 0,01$	6,70
TBA22		$1,04 \pm 0,04$	5,98
TBA23		$1,89 \pm 0,11$	5,72
TBA24		$2,04 \pm 0,08$	5,69