

ÍNDICE DE FIGURAS

Capítulo I.

Figura I.1.1.1. Dados da OMS sobre a distribuição da doença de Chagas no mundo.....1

Figura I.1.2.1. Estruturas químicas do Nifurtimox (3-metil-4-5 (nitrofurfurilidenoamina)-tetrahydro-4H-1, tiazina 1-1 dióxido)) e do Benzonidazol (N-benzil-2-nitroimidazol acetamida).....2.

Figura I.1.3.1. Conformações da Glu205 ligada com fenilalanina (estruturas em amarelo), com arginina (estruturas em azul) ou com tirosina (estruturas em vermelho) na cavidade S₂.4

Figura I.1.3.2. Algumas estruturas selecionadas de compostos estudados como inibidores da cruzaina: a) b) c) d) e) f) g) h) i).6

Figura I.2.1. Balanço Energético da(s) interação(ões) composto-receptor: fatores entálpicos e entrópicos envolvidos.8

Figura I.4.7.1. Representação em 2 dimensões da estrutura molecular do 1-metil-2-propil-ciclobutano.25

Figura I.4.7.2. Matriz de adjacência da molécula do 1-metil-2-propil-ciclobutano. Os átomos foram numerados como atribuído na figura I.4.7.1.25

Figura I.4.7.3. Matriz de distâncias topológicas da molécula do 1-metil-2-propil-ciclobutano. Os átomos foram numerados como atribuído na figura I.4.7.1.25

Figura I.5.1.1. Estrutura dos derivados de tironina.39

Figura I.9.4.1. Valores do coeficiente de correlação de ajuste dos valores preditos da série de teste (R^2) em função dos valores de coeficiente predição interna (q^2), apresentados nos gráficos **a** e **c**; Valores observados em função dos correspondentes preditos da série de teste gráficos **b** e **d**, considerados no trabalho original de validação de modelos, proposto por Tropsha.62

Figura I.9.4.2. Gráfico entre os valores observados e preditos e o correspondente construído invertendo-se os eixos, para uma série de teste.64

Capítulo II.

Figura II.5.1.1.1. Histogramas da distribuição dos valores de atividade pIC₅₀ respectivamente das séries de treinamento e de teste IA.77

Figura II.5.1.1.2. Histogramas da distribuição dos valores de atividade pIC₅₀ respectivamente das séries de treinamento e de teste IB.78

Figura II.5.1.1.3. Histogramas da distribuição dos valores de atividade pIC₅₀ respectivamente das séries de treinamento e de teste IC.79

Figura II.5.1.2.1. Histogramas da distribuição dos valores de atividade pIC_{50} respectivamente das séries de treinamento e de teste IIA.	81
Figura II.5.1.2.2. Histogramas da distribuição dos valores de atividade pIC_{50} respectivamente das séries de treinamento e de teste IIB.	82
Figura II.5.1.2.3. Histogramas da distribuição dos valores de atividade pIC_{50} respectivamente das séries de treinamento e de teste IIC.	83
Figura II.5.1.3.1. Histogramas da distribuição dos valores de atividade pIC_{50} respectivamente das séries de treinamento e de teste IIIA.	85
Figura II.5.1.3.2. Histogramas da distribuição dos valores de atividade pIC_{50} respectivamente das séries de treinamento e de teste IIIB.	86
Figura II.5.1.3.3. Histogramas da distribuição dos valores de atividade pIC_{50} respectivamente das séries de treinamento e de teste IIIC.	87

Capítulo III

Figura III.1.2.1 Valores dos quadrados dos coeficientes de predição respectivamente Q_{cv}^2 e Q^2 , em função do número de descritores (N) considerados em cada um dos 16 modelos de <i>PLS</i> usando validação interna (série de treinamento) e validação externa (série de teste) e 3 variáveis latentes, para a série IA.	95
Figura III.1.2.2. Valores dos quadrados dos coeficientes de predição respectivamente Q_{cv}^2 e Q^2 , em função do número de descritores (N) considerados em cada um dos 16 modelos de <i>PLS</i> usando validação interna (série de treinamento) e validação externa (série de teste) e 3 variáveis latentes, para a série IB.	95
Figura III.1.2.3. Valores dos quadrados dos coeficientes de predição respectivamente Q_{cv}^2 e Q^2 , em função do número de descritores (N) considerados em cada um dos 16 modelos de <i>PLS</i> usando validação interna (série de treinamento) e validação externa (série de teste) e 3 variáveis latentes, para a série IC.	96
Figura III.1.4.1. Matriz de intercorrelação (r) das variáveis presentes no modelo de <i>QSAR</i> selecionado, expresso pela equação III.1.4.1.	102
Figura III.1.4.2. Gráficos, respectivamente, da correlação (A) e dos valores dos resíduos (B) entre os valores de pIC_{50} preditos e observados pelo modelo de <i>QSAR</i> , expresso pela equação III.1.4.1.	104
Figura III.1.4.3. Representação gráfica dos valores dos coeficientes dos descritores presentes no modelo expresso pela equação III.1.4.1.	104
Figura III.2.2.1. Valores dos quadrados dos coeficientes de predição respectivamente Q_{cv}^2 e Q^2 , em função do número de descritores (N) considerados em cada um dos 21 modelos de <i>PLS</i> usando validação interna (série de treinamento) e validação externa (série de teste) e 3 variáveis latentes, para a série IIA.	108

Figura III.2.2.2. Valores dos quadrados dos coeficientes de predição respectivamente Q_{cv}^2 e Q^2 , em função do número de descritores (N) considerados em cada um dos 21 modelos de <i>PLS</i> usando validação interna (série de treinamento) e validação externa (série de teste) e 3 variáveis latentes, para a série IIB.	108
Figura III.2.2.3. Valores dos quadrados dos coeficientes de predição respectivamente Q_{cv}^2 e Q^2 , em função do número de descritores (N) considerados em cada um dos 21 modelos de <i>PLS</i> usando validação interna (série de treinamento) e validação externa (série de teste) e 2 variáveis latentes, para a série IIC.	109
Figura III.2.2.4. Gráficos da correlação e dos resíduos entre os valores de pIC_{50} preditos e, observados a partir do modelo <i>PLS</i> (selecionado através do critério apresentado no item II.5.3.1.) para as séries de treinamento respectivamente: (A): série IIA, com 3 variáveis latentes e 30 variáveis originais; (B): série IIB, com 3 variáveis latentes e 15 variáveis originais e, (C): série IIC, com 2 variáveis latentes e 35 variáveis originais.	110
Figura III.2.2.5. Gráficos de regressão entre os valores de pIC_{50} observados e os correspondentes preditos através do modelo <i>PLS</i> (selecionado através do critério de apresentado no item II.5.3.1.) respectivamente para : (A): série IIA, com 3 variáveis latentes e 30 variáveis originais; (B): série IIB, com 3 variáveis latentes e 15 variáveis originais e, (C): série IIC, 2 variáveis latentes e 35 variáveis originais.	112
Figura III.2.4.1. Matriz de intercorrelação (r) das variáveis presentes no modelo de <i>QSAR</i> selecionado, expresso pela equação III.2.4.1.	118
Figura III.2.4.2. Gráficos, respectivamente, da correlação (A) e dos valores dos resíduos (B) entre os valores de pIC_{50} preditos e observados pelo modelo de <i>QSAR</i> , expresso pela equação III.2.4.1.	120
Figura III.2.4.3. Representação gráfica dos valores dos coeficientes dos descritores presentes no modelo expresso pela equação III.2.4.1.	120
Figura III.3.2.1. Valores dos quadrados dos coeficientes de predição respectivamente Q_{cv}^2 e Q^2 , em função do número de descritores (N) considerados em cada um dos 23 modelos de <i>PLS</i> usando validação interna (série de treinamento) e validação externa (série de teste) e 3 variáveis latentes, para a série IIA.	124
Figura III.3.2.2. Valores dos quadrados dos coeficientes de predição respectivamente Q_{cv}^2 e Q^2 , em função do número de descritores (N) considerados em cada um dos 23 modelos de <i>PLS</i> usando validação interna (série de treinamento) e validação externa (série de teste) e 2 variáveis latentes, para a série IIB.	125
Figura III.3.2.3. Valores dos quadrados dos coeficientes de predição respectivamente Q_{cv}^2 e Q^2 , em função do número de descritores (N) considerados em cada um dos 23 modelos de <i>PLS</i> usando validação interna (série de treinamento) e validação externa (série de teste) e 4 variáveis latentes, para a série IIC.	125

Figura III.3.2.4. Gráficos da correlação e dos resíduos entre os valores de pIC_{50} preditos e, observados a partir do modelo *PLS* (selecionado através do critério apresentado no item II.5.3.1.) para as séries de treinamento, respectivamente:

- (A): série IIA, com 3 variáveis latentes e 10 variáveis originais;
- (B): série IIB, com 2 variáveis latentes e 30 variáveis originais e,
- (C): série IIC, com 4 variáveis latentes e 30 variáveis originais.127

Figura III.3.2.5. Gráficos de regressão entre os valores de pIC_{50} observados e os correspondentes preditos através do modelo *PLS* (selecionado através do critério apresentado no item II.5.3.1.) respectivamente para :

- (A): série IIA, com 3 variáveis latentes e 10 variáveis originais;
- (B): série IIB, com 2 variáveis latentes e 30 variáveis originais e,
- (C): série IIC, com 4 variáveis latentes e 30 variáveis originais.129

Figura III.3.4.1. Matriz de intercorrelação (r) das variáveis presentes no modelo de *QSAR* selecionado, expresso pela equação III.3.4.1.135

Figura III.3.4.2. Gráficos, respectivamente, da correlação (A) e dos valores dos resíduos (B) entre os valores de pIC_{50} preditos e observados pelo modelo de *QSAR*, expresso pela equação III.3.4.1.137

Figura III.3.4.3. Representação gráfica dos valores dos coeficientes dos descritores presentes no modelo expresso pela equação III.3.4.1.137

Figura III.4.2.1. Valores dos quadrados dos coeficientes de predição respectivamente Q_{cv}^2 e Q^2 , em função do número de descritores (N) considerados em cada um dos 23 modelos de *PLS* usando validação interna (série de treinamento) e validação externa (série de teste) e 4 variáveis latentes, para a série IIIA.141

Figura III.4.2.2. Valores dos quadrados dos coeficientes de predição respectivamente Q_{cv}^2 e Q^2 , em função do número de descritores (N) considerados em cada um dos 23 modelos de *PLS* usando validação interna (série de treinamento) e validação externa (série de teste) e 4 variáveis latentes, para a série IIIB.141

Figura III.4.2.3. Valores dos quadrados dos coeficientes de predição respectivamente Q_{cv}^2 e Q^2 , em função do número de descritores (N) considerados em cada um dos 23 modelos de *PLS* usando validação interna (série de treinamento) e validação externa (série de teste) e 4 variáveis latentes, para a série IIIC.142

Figura III.4.2.4. Gráficos da correlação e dos resíduos entre os valores de pIC_{50} preditos e, observados a partir do modelo *PLS* (selecionado através do critério apresentado no item II.5.3.1.) para as séries de treinamento respectivamente:

- (A): série IIIA, com 4 variáveis latentes e 30 variáveis originais;
- (B): série IIIB, com 4 variáveis latentes e 25 variáveis originais e,
- (C): série IIIC, com 4 variáveis latentes e 45 variáveis originais.144

Figura III.4.2.5. Gráficos de regressão entre os valores de pIC_{50} observados e os correspondentes preditos através do modelo *PLS* (selecionado através do critério de apresentado no item II.5.3.1.) respectivamente para :

- (A): série IIIA, com 4 variáveis latentes e 30 variáveis originais;
- (B): série IIIB, com 4 variáveis latentes e 25 variáveis originais e,
- (C): série IIIC, com 4 variáveis latentes e 45 variáveis originais.145

Figura III.4.4.1. Matriz de intercorrelação (r) das variáveis presentes no modelo de <i>QSAR</i> selecionado, expresso pela equação III.4.4.1.	151
Figura III.4.4.2. Gráficos, respectivamente, da correlação (A) e dos valores dos resíduos (B) entre os valores de pIC_{50} preditos e observados pelo modelo de <i>QSAR</i> , expresso pela equação III.4.4.1.	153
Figura III.4.4.3. Representação gráfica dos valores dos coeficientes dos descritores presentes no modelo expresso pela equação III.4.4.1.	153

ÍNDICE DE TABELAS

Capítulo I

Tabela I.2.1 – Principais tipos de interações entre um composto e o sistema biológico, um exemplo e as respectivas faixas de valores de energias envolvidas.7

Tabela I.3.1. Alguns exemplos de parâmetros físico-químicos/estruturais utilizados como descritores; as correspondentes propriedades moleculares e naturezas das interações ligante-receptor (alvo biológico) envolvidas.10

Tabela 1.6.1. Alguns parâmetros estatísticos selecionados para avaliar a validade estatística das correlações/modelos gerados.42

Tabela I.9.3.1. Valores das funções R^P e R^N para alguns modelos teóricos com três variáveis independentes.60

Capítulo II

Tabela II.1.2.1. Valores da atividade inibitória da ribonucleotídeo redutase ($pIC50$)^a da série de tiossemicarbazonas substituídas^b, Série I, selecionadas da literatura^c.68

Tabela II.1.3.1. Valores da atividade inibitória frente à cruzaina ($pIC50$)^a da série de tiossemicarbazonas substituídas, série II, selecionadas da literatura.70

Tabela II.1.4.1. Valores da atividade inibitória frente à cruzaina ($pIC50$)^a da série de tiossemicarbazonas substituídas selecionadas da literatura.72

Tabela II.5.1.1.1. Números dos compostos das séries de treinamento e de teste (IA, IB e IC respectivamente) constituídas a partir dos 29 IRNR^a selecionados da literatura^b para análise *PLS*.76

Tabela II.5.1.2.1. Números dos compostos das séries de treinamento e de teste (IIA, IIB e IIC respectivamente) constituídas a partir das 37 tiossemicarbazonas substituídas (Série II) com atividade inibitória frente a cruzaina selecionadas da literatura^a para análise *PLS*.80

Tabela II.5.1.3.1. Números dos compostos das séries de treinamento e de teste (IIA, IIB e IIC respectivamente) constituídas a partir das 37 tiossemicarbazonas substituídas (Série II) com atividade inibitória frente a cruzaina selecionadas da literatura^a para análise *PLS*.84

Capítulo III

Tabela III.1.2.1. Valores dos termos estatísticos de correlação (ajuste) (r^2)^a e de predição ($(Q_{cv}^2)^b$; $(SEP_{cv})^c$) para os 16 modelos *PLS* gerados após o pré-tratamento e, a partir das 3 séries de treinamento^d respectivamente IA, IB e IC, com 3 variáveis latentes, para diferentes números de variáveis (descritores originais) (N)^e.93

Tabela III.1.2.2. Valores dos termos estatísticos de correlação (ajuste) (r^2) ^a e de predição ($(Q^2)^b$); (SEP) ^c e de predição para os 16 modelos <i>PLS</i> gerados a partir das 3 séries de teste (<i>validação externa</i>) respectivamente IA, IB e IC, com 3 variáveis latentes, para diferentes números de variáveis (descritores originais) (N) ^e	94
Tabela III.1.2.3. Número de variáveis e valores dos parâmetros estatísticos de ajuste e de predição, respectivamente, interno e externo para os modelos <i>PLS</i> gerados a partir das 3 séries de treinamento respectivamente IA, IB e IC.	97
Tabela III.1.3.1. Descritores presentes nos melhores modelos ^a e os respectivos valores de frequência: de presença absoluta e, de presença relativa, observados nas séries de treinamento IA, IB e IC, respectivamente.	98
Tabela III.1.3.2. Descritores da série I selecionados ^a para posterior análise de <i>QSAR</i> clássico e as respectivas classes.	99
Tabela III.1.4.1. Validação do modelo expresso pela equação III.1.4.1 utilizando a regra QUIK. Valores observados para os índices estatísticos descritivos de correlação multivariada, de entropias e de degeneração, respectivamente.	100
Tabela III.1.4.2. Validação do modelo expresso pela equação III.1.4.1 utilizando a regra do Q^2 assintótico. Valores observados para os índices estatísticos descritivos respectivamente de correlação linear (r), de número de parâmetros (p'), de coeficiente de predição interno (Q_{cv}^2), de coeficiente de predição assintótico (Q_{ASYN}^2) e da diferença dos valores dos coeficientes de predição ($Q_{CV}^2 - Q_{ASYN}^2$).	101
Tabela III.1.4.3. Validação do modelo expresso pela equação III.1.4.1 utilizando as regras baseadas na função R. Valores observados para os índices estatísticos descritivos respectivamente de correlação linear (r), de número de variáveis (p), de função R^P e de função R^N	102
Tabela III.1.4.4. Valores de pIC_{50} previstos pelo modelo <i>QSAR</i> expresso pela equação III.1.4.1, e os correspondentes observados, retirados da literatura ^a , bem como das respectivas diferenças para os compostos da série I, IRNR.	103
Tabela III.2.2.1. Valores dos termos estatísticos de correlação (ajuste) (r^2) ^a e de predição ($(Q_{cv}^2)^b$); (SEP_{cv}) ^c para os 21 modelos <i>PLS</i> gerados após o pré-tratamento e, a partir das 3 séries de treinamento ^d respectivamente IIA, IIB e IIC, com 3 variáveis latentes para as séries IIA e IIB e, com 2 variáveis latentes para a série IIC, para diferentes números de variáveis (descritores originais) (N) ^e	106
Tabela III.2.2.2. Valores dos termos estatísticos de correlação (ajuste) (r^2) ^a e de predição ($(Q^2)^b$); (SEP) ^c e de predição para os 21 modelos <i>PLS</i> gerados a partir das 3 séries de teste (<i>validação externa</i>) respectivamente IIA, IIB e IIC, com 3 variáveis latentes para as séries IIA e IIB e, com 2 variáveis latentes para a série IIC, para diferentes números de variáveis (descritores originais) (N) ^e	107
Tabela III.2.2.3. Número de variáveis e valores dos parâmetros estatísticos de ajuste e de predição, respectivamente, interno e externo para os modelos <i>PLS</i> gerados a partir das 3 séries de treinamento respectivamente IIA, IIB e IIC.	109

Tabela III.2.2.4. Valores observados para os índices estatísticos descritivos, respectivamente, da correlação linear externa (R^2), do coeficiente de predição interno (Q_{cv}^2), do coeficiente de correlação externo da reta que apresenta coeficiente linear igual a 0 entre os valores observados e preditos (R_0^2) e, do coeficiente angular da reta (k), respectivamente para : série de teste IIA, obtidos a partir da figura III.2.2.5.(A); série de teste IIB , obtidos a partir da figura III.2.2.5.(B) e, série de teste IIC , obtidos a partir da figura III.2.2.5 .(C).....	113
Tabela III.2.3.1. Descritores presentes nos melhores modelos ^a e os seus respectivos valores de frequência: de presença absoluta e, de presença relativa, observados nas séries de treinamento IIA, IIB e IIC, respectivamente.	114
Tabela III.2.3.2. Descritores da série II selecionados ^a para posterior análise de <i>QSAR</i> clássico e as respectivas classes.	115
Tabela III.2.4.1. Validação do modelo expresso pela equação III.2.4.1 utilizando a regra QUIK. Valores observados para os índices estatísticos descritivos de correlação multivariada, de entropias e de degeneração, respectivamente.	116
Tabela III.2.4.2. Validação do modelo expresso pela equação III.2.4.1 utilizando a regra do Q^2 assintótico. Valores observados para os índices estatísticos descritivos respectivamente de correlação linear (r), de número de parâmetros (p'), de coeficiente de predição interno (Q_{cv}^2), de coeficiente de predição assintótico (Q_{ASYN}^2) e da diferença dos valores dos coeficientes de predição ($Q_{CV}^2-Q_{ASYN}^2$).	117
Tabela III.2.4.3. Validação do modelo expresso pela equação III.2.4.1 utilizando as regras baseadas na função R. Valores observados para os índices estatísticos descritivos respectivamente de correlação linear (r), de número de variáveis (p), de função R^P e de função R^N	118
Tabela III.2.4.4. Valores de pIC_{50} previstos pelo modelo <i>QSAR</i> expresso pela equação III.2.4.1 e os correspondentes observados, retirados da literatura ^a , bem como das respectivas diferenças para os compostos da série II.	119
Tabela III.3.2.1. Valores dos termos estatísticos de ajuste (r^2) ^a e de predição ($(Q_{cv}^2)^b$); (SEP_{cv}) ^c para os 23 modelos <i>PLS</i> gerados após o pré-tratamento e, a partir das 3 séries de treinamento ^d respectivamente IIA, IIB e IIC, com 3 variáveis latentes para a série IIA, com 2 variáveis latentes para a série IIB e, com 4 variáveis latentes para a série IIC, para diferentes números de variáveis (descritores originais) (N) ^e	122
Tabela III.3.2.2. Valores dos termos estatísticos de correlação (ajuste) (r^2) ^a e de predição ($(Q^2)^b$); (SEP) ^c e de predição para os 23 modelos <i>PLS</i> gerados a partir das 3 séries de teste (<i>validação externa</i>) respectivamente IIA, IIB e IIC, com 3 variáveis latentes para a série IIA, com 2 variáveis latentes para a série IIB e, com 4 variáveis latentes para a série IIC, para diferentes números de variáveis (descritores originais) (N) ^e	123
Tabela III.3.2.3. Número de variáveis e valores dos parâmetros estatísticos de ajuste e de predição, respectivamente, interno e externo para os modelos <i>PLS</i> gerados a partir das 3 séries de treinamento respectivamente IIA, IIB e IIC.	126

Tabela III.3.2.4. Valores observados para os índices estatísticos descritivos, respectivamente, da correlação linear externa (R^2), do coeficiente de predição interno (Q_{cv}^2), do coeficiente de correlação externo da reta que apresenta coeficiente linear igual a 0 entre os valores observados e preditos (R_0^2) e, do coeficiente angular da reta (k), respectivamente para:

série de teste IIA, obtidos a partir da figura III.3.2.5.(A);
série de teste IIB, obtidos a partir da figura III.3.2.5.(B) e,
série de teste IIC, obtidos a partir da figura III.3.2.5.(C).....130

Tabela III.3.3.1. Descritores presentes nos melhores modelos^a e os respectivos valores de frequência: de presença absoluta e, de presença relativa observados nas séries de treinamento IIA, IIB e IIC, respectivamente.131

Tabela III.3.3.2. Descritores da série II selecionados^a para posterior análise de *QSAR* clássico e as respectivas classes.....132

Tabela III.3.4.1. Validação do modelo expresso pela equação III.3.4.1 utilizando a regra QUIK. Valores observados para os índices estatísticos descritivos de correlação multivariada, de entropias e de degeneração, respectivamente.134

Tabela III.3.4.2. Validação do modelo expresso pela equação III.3.4.1 utilizando a regra do Q^2 assintótico. Valores observados para os índices estatísticos descritivos respectivamente de correlação linear (r), de número de parâmetros (p'), de coeficiente de predição interno (Q_{cv}^2), de coeficiente de predição assintótico (Q_{ASYN}^2) e da diferença dos valores dos coeficientes de predição ($Q_{CV}^2 - Q_{ASYN}^2$).134

Tabela III.3.4.3. Validação do modelo expresso pela equação III.3.4.1 utilizando as regras baseadas na função R. Valores observados para os índices estatísticos descritivos respectivamente de correlação linear (r), de número de variáveis (p), de função R^P e de função R^N135

Tabela III.3.4.4. Valores de pIC_{50} previstos pelo modelo *QSAR* expresso pela equação III.3.4.1, e os correspondentes observados, retirados da literatura^a, bem como das respectivas diferenças para os compostos da série II.136

Tabela III.4.2.1. Valores dos termos estatísticos de correlação (ajuste) (r^2)^a e de predição ((Q_{cv}^2))^b; (SEP_{cv})^c para os 23 modelos *PLS* gerados após o pré-tratamento e, a partir das 3 séries de treinamento^d respectivamente IIIA, IIB e IIC, com 4 variáveis latentes, para diferentes números de variáveis (descritores originais) (N)^e.139

Tabela III.4.2.2. Valores dos termos estatísticos de ajuste (r^2)^a e de predição ((Q^2))^b; (SEP)^c e de predição para os 23 modelos *PLS* gerados a partir das 3 séries de teste (*validação externa*) respectivamente IIIA, IIB e IIC, com 4 variáveis latentes, para diferentes números de variáveis (descritores originais) (N)^e.140

Tabela III.4.2.3. Número de variáveis e valores dos parâmetros estatísticos de ajuste e de predição, respectivamente, interno e externo para os modelos *PLS* gerados a partir das 3 séries de treinamento respectivamente IIIA, IIB e IIC.143

Tabela III.4.2.4. Valores observados para os índices estatísticos descritivos, respectivamente, da correlação linear externa (R^2), do coeficiente de predição interno (Q_{cv}^2), do coeficiente de correlação externo da reta que apresenta coeficiente linear igual a 0 entre os valores observados e preditos (R_0^2) e, do coeficiente angular da reta (k), respectivamente para : série de teste IIIA, obtidos a partir da figura III.4.2.5.(A); série de teste IIIB , obtidos a partir da figura III.4.2.5.(B) e, série de teste IIIC , obtidos a partir da figura III.4.2.5 .(C).....	146
Tabela III.4.3.1. Descritores presentes nos melhores modelos ^a e os respectivos valores de frequência: de presença absoluta e, de presença relativa observados nas séries de treinamento IIIA, IIIB e IIIC, respectivamente.	147
Tabela III.4.3.2. Descritores da série III selecionados ^a para posterior análise de <i>QSAR</i> clássico e as respectivas classes.	148
Tabela III.4.4.1. Validação do modelo expresso pela equação III.4.4.1 utilizando a regra QUIK. Valores observados para os índices estatísticos descritivos de correlação multivariada, de entropias e de degeneração, respectivamente.	150
Tabela III.4.4.2. Validação do modelo expresso pela equação III.4.4.1 utilizando a regra do Q^2 assintótico. Valores observados para os índices estatísticos descritivos respectivamente de correlação linear (r), de número de parâmetros (p'), de coeficiente de predição interno (Q_{cv}^2), de coeficiente de predição assintótico (Q_{ASYN}^2) e da diferença dos valores dos coeficientes de predição ($Q_{cv}^2 - Q_{ASYN}^2$).....	150
Tabela III.4.4.3. Validação do modelo expresso pela equação III.4.4.1 utilizando as regras baseadas na função R. Valores observados para os índices estatísticos descritivos respectivamente de correlação linear (r), de número de variáveis (p), de função R^p e de função R^N	151
Tabela III.4.4.4. Valores de pIC_{50} previstos pelo modelo <i>QSAR</i> expresso pela equação III.4.4.1, e os correspondentes observados, retirados da literatura ^a , bem como das respectivas diferenças para os compostos da série III.	152

RESUMO

O desenvolvimento tecnológico observado nos últimos anos e o avanço na aquisição de dados para os sistemas tanto químicos como biológicos tem gerado um grande número de informações. Como consequência, nos últimos anos, procuram-se ferramentas, fundamentalmente matemáticas, que permitam decodificar este volume imenso de informações, em termos estruturais e biológicos.

Embora diversos parâmetros estatísticos e métodos de seleção de modelos e de variáveis tenham sido descritos, mais recentemente, são encontradas propostas de novas ferramentas visando assegurar tanto a qualidade de predição do modelo como a elucidação de algum mecanismo a partir do modelo gerado.

Com relação aos estudos de compostos antichagásicos, é grande o número de dados disponíveis, na literatura, nos diferentes aspectos pesquisados. A cruzipaina, também chamada de cruzaina, é a principal cisteína protease do *Trypanosoma cruzi* e sua inibição tem se mostrado capaz de inibir o desenvolvimento intracelular do protozoário.

Neste trabalho foram selecionados da literatura original 90 compostos, incluindo-se 29 α -(N)-heterocíclica carboxaldeído tiossemicarbazonas substituídas, com atividade inibitória frente a ribonucleotídeo redutase (IRNR) (série I; compostos I.1. a I.29); 37 tiossemicarbazonas substituídas na cadeia lateral e no anel aromático (série II; compostos II.1 a II.37) e, 61 compostos estruturalmente diferentes (série III; compostos III.1 a III.61, sendo 45 tiossemicarbazonas das quais 37 da série II e, 16 derivados de isatinas). As séries II e III apresentam atividade inibitória frente à cruzaina, uma cisteína protease do *T. cruzi*. E a série I, embora, apresente atividade inibitória frente a ribonucleotídeo redutase (IRNR) de células H.Ep.-2, é, no entanto, estruturalmente similar às séries II e III. E, ainda mais, os modelos gerados nesta dissertação para a série I foram incluídos na tese de doutorado de H. Ishiki.

Cada uma das séries foi dividida em 3 séries de treinamento (A, B e C) com suas correspondentes séries de teste.

A seguir, a partir das estruturas representadas em 3 dimensões das moléculas, foram gerados 1497 descritores através do programa DRAGON (v. 3.0) para cada série (I, II e III). Estes descritores foram submetidos a um pré-tratamento de dados, excluindo-se aqueles que não contribuiriam para as análises *PLS – Partial Least Squares* (mínimos quadrados parciais).

Através das análises *PLS* foram selecionados os descritores mais significativos das equações de regressão linear, que a seguir foram submetidos a uma análise de frequência, ou

seja, foram selecionados os descritores presentes em pelo menos em 2 dos 3 modelos obtidos das séries de treinamento (A, B e C). A partir destes descritores selecionados foram obtidos modelos de QSAR clássico com 5 descritores que foram validados por três filtros. Estes modelos de QSAR validados, apresentaram valores de coeficiente de correlação (r) e do quadrado do coeficiente de predição interna (Q_{cv}^2) maiores que 0,9 e 0,7 respectivamente.

ABSTRACT

It has been observed an enormous improvement in the methods concerning data generation, leading to a large amount of information, especially for chemical and biological systems. Through these developments, it becomes relevant to have reliable methods, mainly new mathematical tools, for structure-activity relationship (SAR) data examination, which means that there is a need for developing datasets screening tools.

Although a huge number of descriptors and methods for selection have been described in the literature, it becomes a crucial aspect to develop new concepts and tools that assure selection of relevant information as well as a high predictive power for the generated QSAR models.

Concerning antichagasic compounds, in the literature, there is a huge number of data in anti Chagas disease drug research fields. Cruzain, known also as cruzipain, is the major cysteine protease of *T. cruzi*. The protease is expressed in all life cycle stages of the parasite. Therefore, cruzain is essential for replication of the intracellular parasite. Thus, cruzain is an appealing target for new antitrypanosomal chemotherapy.

In this work, it has been taken from selected literature 90 compounds, including 29 substituted α -(N)- heterocyclic carboxaldehyde thiosemicarbazones, (set I, compounds I.1 to I. 29) and 37 substituted thiosemicarbazones (set II, compounds II.1 to II.37) and 61 structurally different compounds (set III, compounds II.1 to III.61, namely 45 thiosemicarbazones (37 from set II) and 16 isatin derivatives). Sets II and III showed inhibitory activity against cruzain, a cysteine protease of *T. cruzi*. Although set I compounds showed inhibitory activity against ribonucleotide reductase enzyme of H.Ep.-2 cells they have been included in this study taking into account that they are structurally similar to sets II and III, studied by H. Ishiki in his PhD thesis.

Each set was divided in 3 training sets (A, B and C) with its corresponding test sets. Initially, 1497 descriptors have been obtained by means of DRAGON program (v.3.0) using 3D structures of sets I, II and III compounds. In a second step to reduce the data size, all descriptors have been submitted to a pre-treatment, including the use of different filters, followed by PLS analyses with external validation.

Relevant descriptors have been selected using PLS followed by frequency analysis. That means, it has been selected descriptors, which were present at least in 2 or 3 models, generated from training sets (A, B and C). These descriptors have been used to generate QSAR models with up to 5 descriptors. The QSAR models had been validated by three filters. Values of correlation coefficient (r) and the squared correlation coefficient of internal predictions (Q_{cv}^2) were higher than 0.9 and 0.7, respectively, for the validated QSAR models.

GLOSSÁRIO

QSAR – Relação quantitativa entre a estrutura química e a atividade biológica

π - constante de hidrofobicidade de Hansch-Fujita;

σ - constante eletrônica de Hammett

E_s – parâmetro estérico de Taft

R , r - coeficiente de correlação

s-desvio padrão

F- fator de confiabilidade

Q^2 - coeficiente de predição

IC_{50} – concentração que inibe 50% do efeito

pIC_{50} – logaritmo negativo de IC_{50}

$\log P$ –coeficiente de partição

β - relação entre as fases orgânica e aquosa no sistema bilinear;

μ - média da atividade biológica com todos os substituintes na modelo Free-Wilson;

PCA – análise de componentes principais

PLS – método dos mínimos quadrados parciais

MLR – regressão linear múltipla

KNN - “*k-Nearest Neighbor analysis*”

Objetivos

Neste trabalho, destinado a dissertação de mestrado, estudamos as propriedades físico-químicas de séries de compostos, selecionados da literatura, com atividade inibitória frente à cruzafina, com o objetivo de propor e aplicar critérios de seleção para um grande número de descritores obtidos pelo programa DRAGON (v.3.0), utilizando a técnica de regressão PLS (Mínimos Quadrados Parciais) com validação externa, visando escolher parâmetros estruturais estatisticamente significativos para a obtenção de modelos QSAR válidos. Adicionalmente, estudos análogos foram realizados para séries de compostos, selecionados da literatura, com a atividade inibitória frente a ribonucleotídeo redutase (IRNR) de células H.Ep.-2, visando avaliar os critérios propostos e, considerando-se a similaridade estrutural e, a disponibilidade de informação no grupo.