

III.3. Série II – Retirando as variáveis 4 sigma

Para a série de 37 tiossemicarbazonas, com atividade frente à cruzaína do *T. cruzi*, selecionadas da literatura (Du *et al.* 2002) foram gerados 1497 descritores pelo programa DRAGON. Os dados referentes à série II estão apresentados na tabela II.1.3.1.

III.3.1. Pré-Tratamento de Dados

A partir dos valores dos descritores gerados, através da opção “*variable exclusion*” do programa DRAGON pode-se verificar que 737 (49,23 %) descritores apresentaram o mesmo valor, ou apenas um único valor diferente na série, 74 (4,94 %) descritores apresentaram valores superiores a 4 desvios padrão (4σ) em relação à respectiva média e, 328 (21,91 %) descritores apresentaram valores de coeficiente de intercorrelação, $|r|$, maiores que 0,95. E, a seguir, estes foram eliminados considerando os critérios de seleção de variáveis apresentados no item II.5.2. *Pré-tratamento de Dados*. Após este pré-tratamento o número de descritores foi reduzido para 358 descritores (23,91 %).

III.3.2. Modelos PLS

Após este pré-tratamento, 12 análises PLS com 1 até 4 variáveis latentes foram aplicadas à série dos 358 descritores juntamente com os correspondentes valores de pIC_{50} . Os modelos PLS foram obtidos através de protocolo do programa SYBYL 6.9, Tripos Inc (SYBYL program 2000) (item II.5.3.1. *Modelos PLS*). Foram considerados estatisticamente significativos aqueles apresentando os maiores valores de coeficiente de predição interno (Q_{cv}^2). Adicionalmente, observou-se que os modelos PLS estatisticamente mais significativos apresentavam 3, 2 e 4 variáveis latentes, respectivamente para as séries IIA, IIB e, para a série IIC.

Estão apresentados nas tabelas III.3.2.1 e III.3.2.2. os valores dos termos estatísticos observados para os modelos PLS estatisticamente mais significativos, gerados para as séries

de treinamento e de teste (IIA, IIB e IIC), considerando-se diferentes números de variáveis (descritores originais) em cada uma das análises *PLS*.

Tabela III.3.2.1. 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 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.

N	Séries de Treinamento IIA			Séries de Treinamento IIB			Séries de Treinamento IIC		
	r^2	Q_{cv}^2	SEP_{cv}	r^2	Q_{cv}^2	SEP_{cv}	r^2	Q_{cv}^2	SEP_{cv}
300	0,505	0,496	0,749	0,558	0,557	0,708	0,422	0,402	0,786
250	0,582	0,577	0,686	0,632	0,630	0,647	0,580	0,555	0,678
200	0,642	0,634	0,638	0,686	0,684	0,598	0,703	0,663	0,590
150	0,703	0,700	0,578	0,713	0,709	0,574	0,763	0,731	0,527
140	0,706	0,702	0,576	0,725	0,723	0,560	0,787	0,751	0,507
130	0,734	0,730	0,548	0,750	0,747	0,535	0,805	0,768	0,490
120	0,757	0,752	0,525	0,761	0,759	0,522	0,805	0,771	0,486
110	0,770	0,765	0,512	0,766	0,763	0,518	0,822	0,791	0,464
100	0,776	0,768	0,508	0,763	0,762	0,519	0,853	0,819	0,432
90	0,797	0,792	0,481	0,757	0,753	0,528	0,865	0,835	0,412
80	0,797	0,791	0,482	0,779	0,778	0,502	0,873	0,848	0,396
70	0,837	0,830	0,435	0,792	0,789	0,488	0,892	0,864	0,374
60	0,863	0,852	0,405	0,800	0,798	0,478	0,901	0,877	0,357
50	0,844	0,840	0,423	0,804	0,799	0,476	0,878	0,865	0,373
45	0,860	0,855	0,402	0,810	0,807	0,467	0,900	0,884	0,346
40	0,866	0,860	0,395	0,811	0,810	0,464	0,893	0,884	0,346
35	0,837	0,833	0,431	0,820	0,819	0,452	0,913	0,901	0,320
30	0,863	0,860	0,395	0,833^f	0,832^f	0,436^f	0,908^f	0,901^f	0,320^f
25	0,867	0,863	0,390	0,818	0,818	0,453	0,838	0,837	0,410
20	0,864	0,862	0,391	0,794	0,793	0,484	0,887	0,886	0,342
15	0,841	0,840	0,422	0,764	0,761	0,520	0,897	0,896	0,327
10	0,872^f	0,871^f	0,379^f	0,761	0,758	0,523	0,772	0,763	0,495
5	0,685	0,672	0,604	0,717	0,715	0,568	0,525	0,501	0,718

^a valor do coeficiente de correlação entre os valores preditos e os correspondentes da literatura;

^{b, c} valores do coeficiente de predição e do correspondente erro padrão de predição gerados pelo método *full cross-validation*;

^d compostos apresentados na tabela II.1.3.1;

^e número de descritores considerados no modelo *PLS*;

^f os números em negrito se referem aos termos estatísticos referentes aos modelos *PLS* selecionados.

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 *PLS* gerados a partir das 3 séries de teste (*validação externa*) 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.

N	Série de Teste IIA			Série de Teste IIB			Série de Teste IIC		
	r^2	Q^2	SEP	r^2	Q^2	SEP	r^2	Q^2	SEP
300	0,578	0,551	0,675	0,423	0,407	0,647	0,716	0,711	0,640
250	0,593	0,566	0,664	0,496	0,490	0,600	0,718	0,714	0,637
200	0,588	0,569	0,661	0,493	0,486	0,602	0,698	0,690	0,663
150	0,564	0,557	0,670	0,565	0,564	0,555	0,696	0,689	0,664
140	0,574	0,565	0,664	0,483	0,475	0,609	0,671	0,663	0,691
130	0,557	0,550	0,675	0,551	0,549	0,564	0,641	0,632	0,723
120	0,531	0,525	0,694	0,511	0,508	0,590	0,596	0,590	0,763
110	0,497	0,492	0,718	0,555	0,554	0,561	0,657	0,638	0,716
100	0,531	0,526	0,693	0,513	0,511	0,588	0,660	0,638	0,716
90	0,503	0,498	0,714	0,602	0,598	0,533	0,639	0,607	0,746
80	0,496	0,485	0,723	0,556	0,555	0,561	0,657	0,614	0,739
70	0,469	0,455	0,744	0,593	0,590	0,538	0,558	0,542	0,805
60	0,478	0,464	0,737	0,512	0,512	0,587	0,543	0,515	0,829
50	0,558	0,546	0,679	0,551	0,539	0,571	0,522	0,500	0,842
45	0,570	0,563	0,666	0,519	0,507	0,590	0,484	0,476	0,862
40	0,528	0,523	0,696	0,556	0,540	0,570	0,497	0,478	0,860
35	0,596	0,588	0,647	0,602	0,588	0,539	0,607	0,575	0,776
30	0,591	0,574	0,658	0,612^f	0,591^f	0,537^f	0,604^f	0,554^f	0,795^f
25	0,615	0,562	0,667	0,569	0,565	0,554	0,663	0,572	0,779
20	0,508	0,450	0,747	0,618	0,615	0,521	0,530	0,480	0,859
15	0,331	0,180	0,912	0,476	0,456	0,620	0,487	0,450	0,883
10	0,254^f	0,082^f	0,965^f	0,345	0,310	0,698	0,580	0,560	0,789
5	0,033	-0,685	1,307	0,332	0,309	0,699	0,760	0,669	0,685

^a valor do coeficiente de correlação entre os valores preditos e os correspondentes da literatura;

^{b, c} valores do coeficiente de predição externo e do correspondente erro padrão de predição externo;

^d compostos apresentados na tabela II.1.3.1.;

^e número de descritores considerados no modelo *PLS*;

^f Os números em negrito se referem aos termos estatísticos referentes aos modelos *PLS* seleccionados;

As análises comparativas dos 23 modelos *PLS* utilizando-se validação interna ou validação externa foram feitas por inspeção visual. Para tanto, utilizando-se os dados apresentados, respectivamente, nas tabelas III.3.2.1. (validação interna) e III.3.2.2 (validação externa) construíram-se gráficos dos valores dos quadrados dos coeficientes de predição, respectivamente, interna (Q_{cv}^2) e, externa (Q^2) em função do número de descritores originais (N). Os gráficos obtidos para as séries IIA, IIB e IIC estão apresentados nas figuras III.3.2.1. a III.3.2.3, respectivamente.

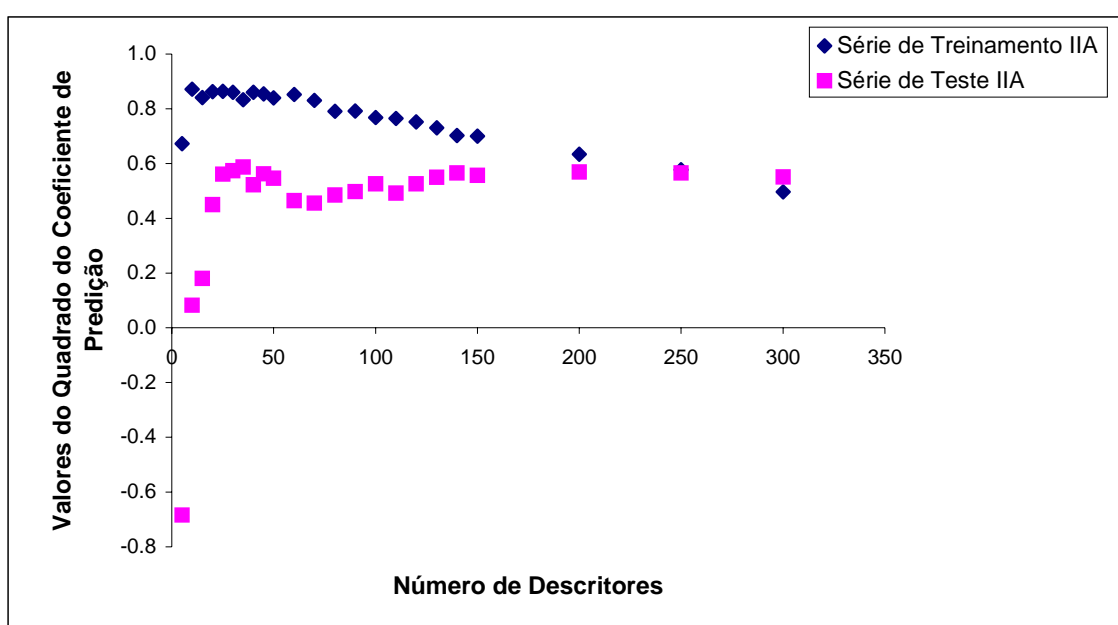


Figura III.3.2.1. Valores dos quadrados dos coeficientes de predição respectivamente Q_{cv}^2 e Q^2 , em função do número de descritores (N) considerados em cada um dos 23 modelos de *PLS* usando validação interna (série de treinamento) e validação externa (série de teste) e 3 variáveis latentes, para a série IIA.

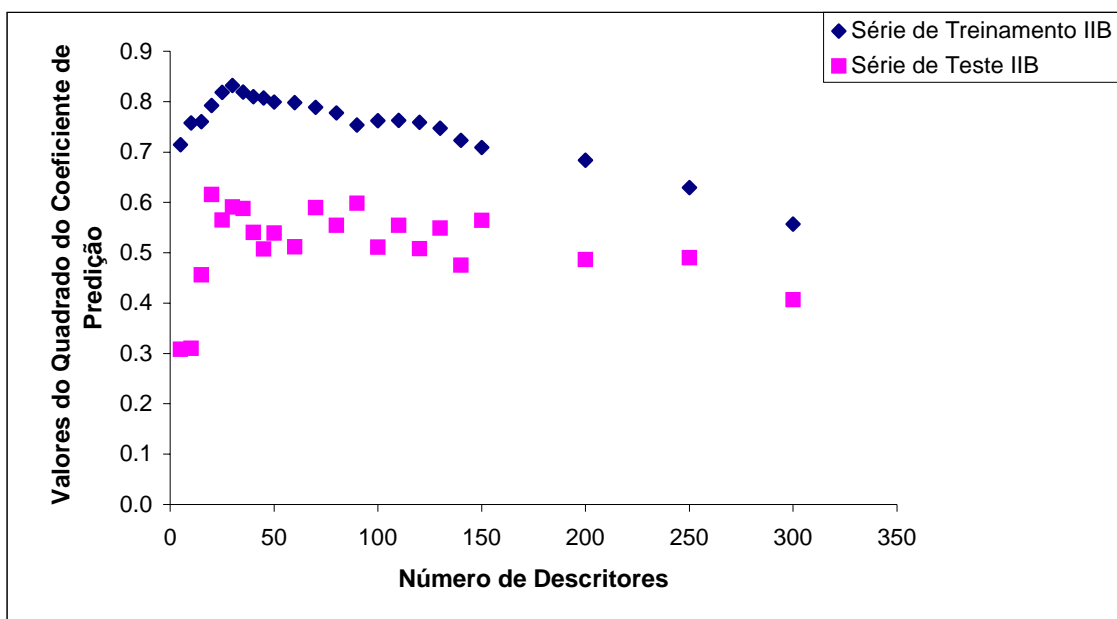


Figura III.3.2.2. Valores dos quadrados dos coeficientes de predição respectivamente Q_{cv}^2 e Q^2 , em função do número de descritores (N) considerados em cada um dos 23 modelos de *PLS* usando validação interna (série de treinamento) e validação externa (série de teste) e 2 variáveis latentes, para a série IIB.

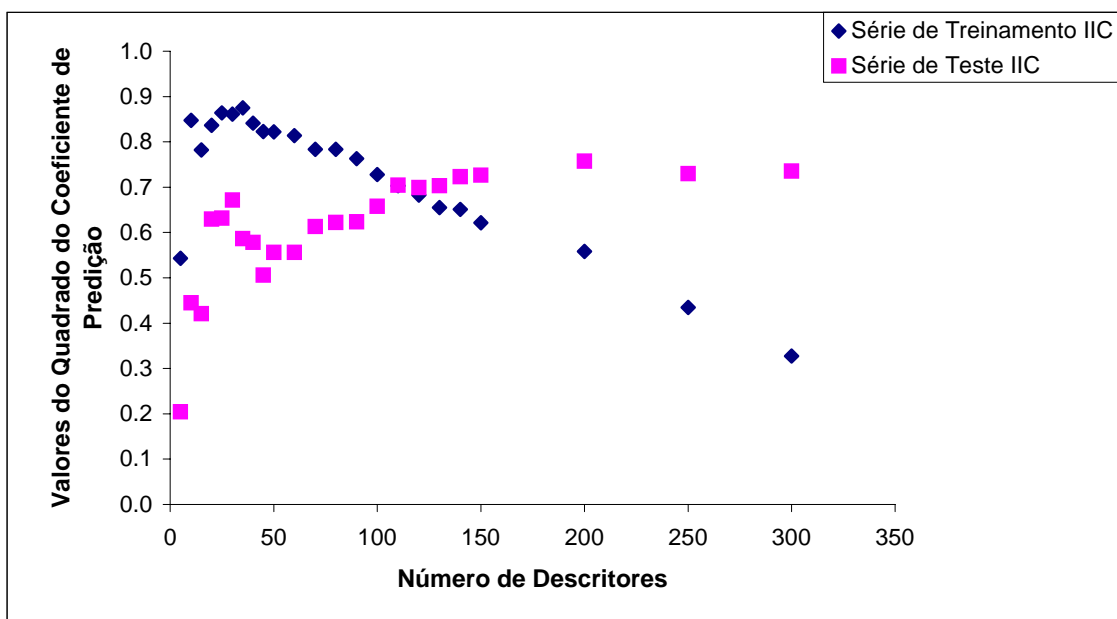


Figura III.3.2.3. Valores dos quadrados dos coeficientes de predição respectivamente Q_{cv}^2 e Q^2 , em função do número de descritores (N) considerados em cada um dos 23 modelos de *PLS* usando validação interna (série de treinamento) e validação externa (série de teste) e 4 variáveis latentes, para a série IIC.

A análise visual dos gráficos apresentados nas figuras III.3.2.1 a III.3.2.3, permitiu selecionar os modelos *PLS* mais significativos como sendo aqueles apresentando os mais altos valores do coeficiente de predição interno (série de treinamento). Os termos correspondentes aos modelos mais significativos selecionados estão assinalados em negrito nas tabelas III.3.2.1 e III.3.2.2 e sumarizados na tabela III.3.2.3.

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 *PLS* gerados a partir das 3 séries de treinamento respectivamente IIA, IIB e IIC.

Parâmetros Estatísticos	Séries de Treinamento		
	IIA	IIB	IIC
Nº de Variáveis Latentes	3	2	4
Nº de Variáveis Originais	10	30	30
R² (coeficiente de ajuste de regressão)	0,872	0,833	0,908
Q_{cv}² (coeficiente de predição interno)	0,871	0,832	0,901
SEP_{df} (desvio padrão de predição interno)	0,379	0,436	0,320
R²_{ext.} (coeficiente de ajuste de regressão externo)	0,254	0,612	0,604
Q²_{ext.} (coeficiente de predição externo)	0,082	0,591	0,554
SEP (desvio padrão de predição externo)	0,965	0,537	0,795

Com os valores de atividade preditos em função dos correspondentes observados bem como dos respectivos resíduos obtidos pelos modelos *PLS*, construíram-se os gráficos apresentados na figura III.3.2.4, respectivamente para as séries IIA, IIB e IIC.

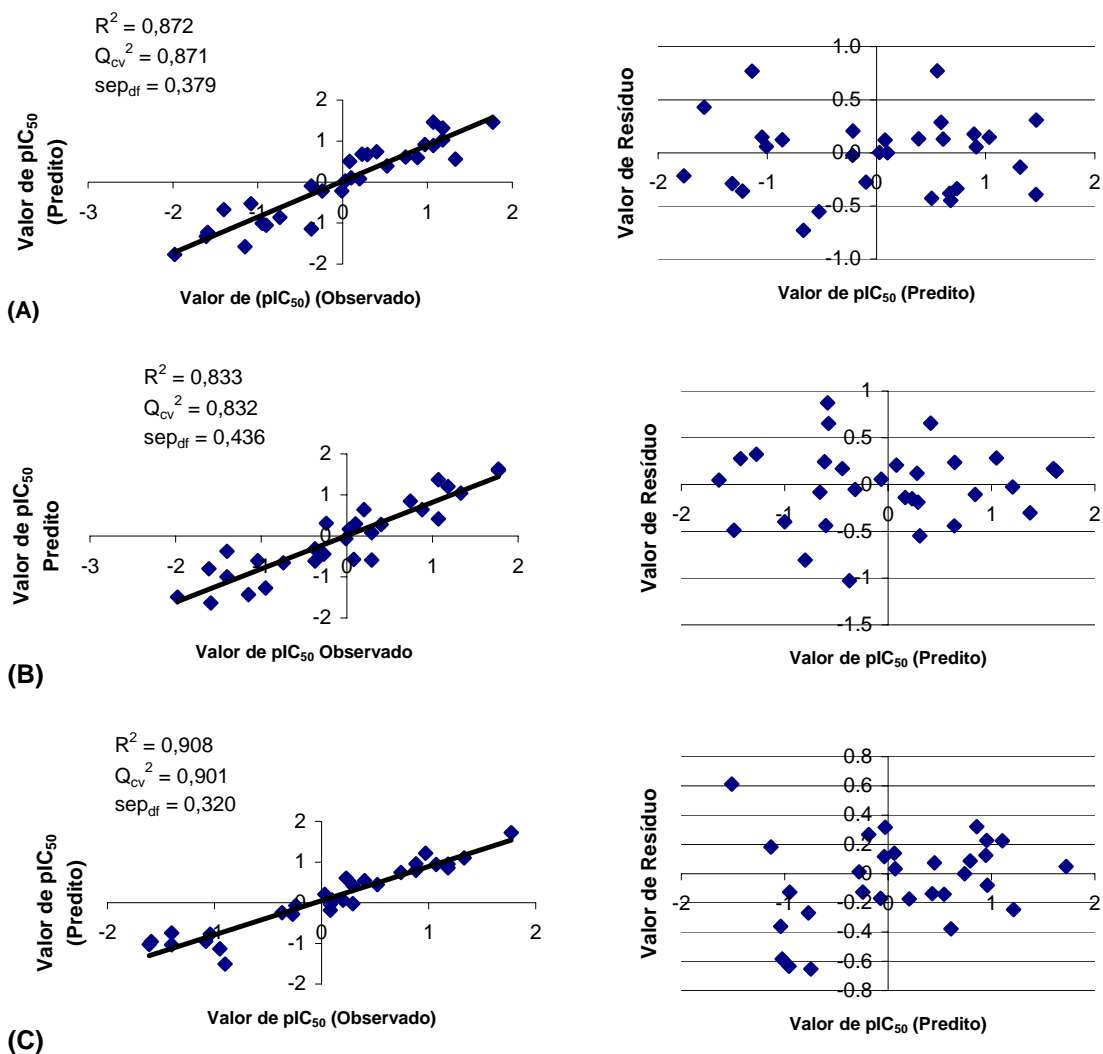


Figura III.3.2.4. Gráficos da correlação e dos resíduos entre os valores de pIC_{50} preditos e, observados a partir do modelo *PLS* (selecionado através do critério apresentado no item II.5.3.1.) para as séries de treinamento, respectivamente:

- (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.

A seguir, para cada uma das séries de teste IIA, IIB e IIC as correlações entre valores observados de atividade ($pIC_{50\text{observado}}$) em função dos correspondentes valores preditos ($pIC_{50\text{predito}}$) foram avaliadas e, representadas pelos gráficos mostrados na figura III.3.2.5, respectivamente para as séries IIA, IIB e IIC.

A seguir, utilizando os valores dos coeficientes estatísticos para cada uma das correlações, aplicou-se para as séries de teste (IIA, IIB e IIC) o critério de validação de modelos, proposto por Tropsha (Golbraikh & Tropsha 2002) e, descrito no item I.5.2.4.

Os resultados dos valores dos coeficientes estatísticos, correspondentes à 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) estão sumarizados na tabela III.3.2.4, respectivamente para as séries IIA, IIB e IIC.

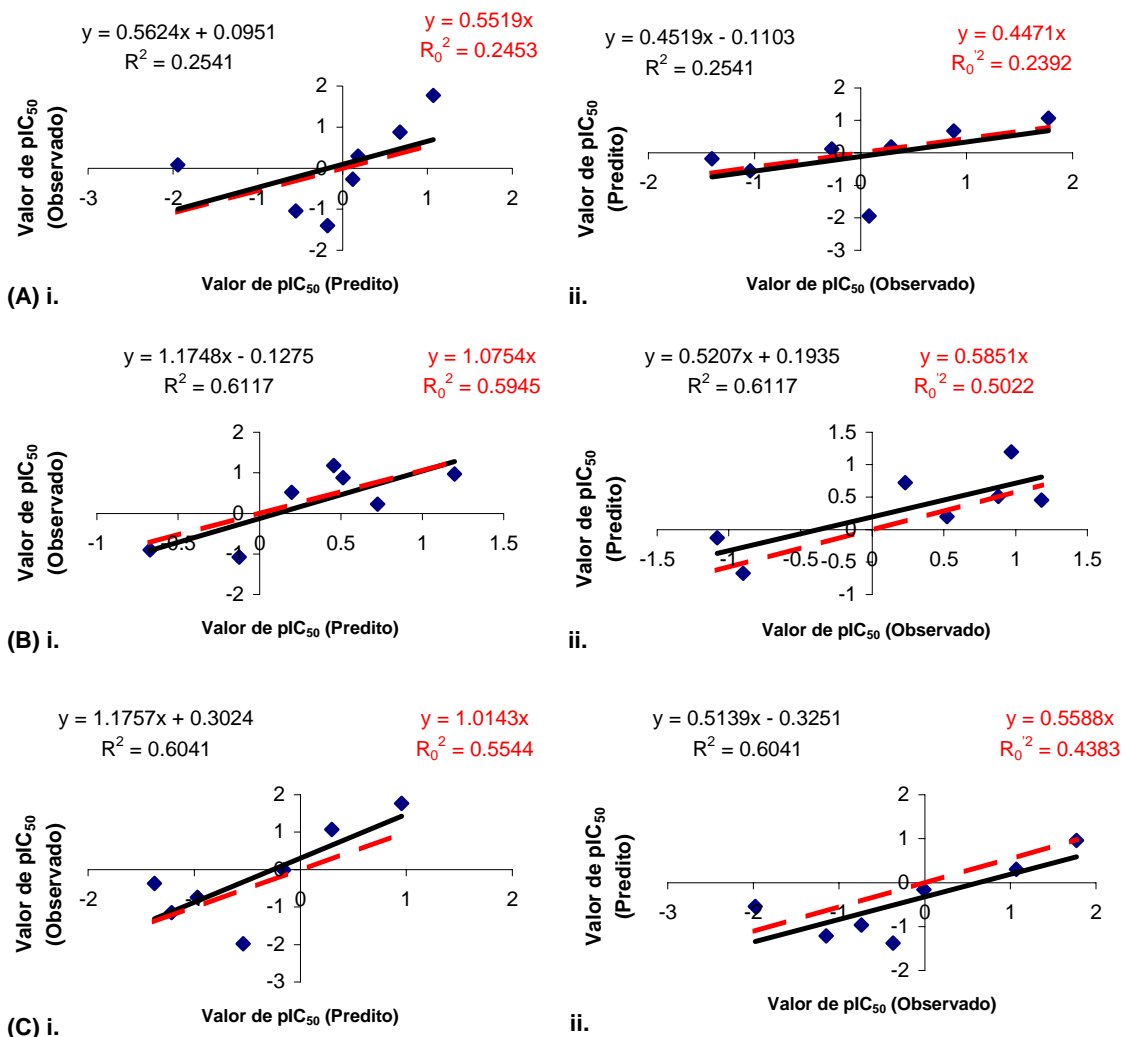


Figura III.3.2.5. Gráficos de regressão entre os valores de pIC₅₀ observados e os correspondentes preditos através do modelo *PLS* (selecionado através do critério apresentado no item II.5.3.1.) respectivamente para :

- (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.

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).

Séries de Teste	R^{2a}	$Q_{cv}^2{}^b$	$R_0^2{}^c$	$(R^2-R_0^2)/R^2{}^d$	$R_0^2{}^e$	$(R^2-R_0^2)/R^2{}^f$	k^g	k'^h
IIA	0,254	0,871	0,245	0,035	0,239	0,059	0,552	0,447
IIB	0,612	0,832	0,5945	0,028	0,5022	0,179	1,075	0,585
IIC	0,604	0,901	0,5544	0,082	0,4383	0,274	1,014	0,5588

^a valor do coeficiente de correlação de regressão (ajuste externo) observada entre os valores de pIC_{50} observados e os correspondentes preditos;

^b valor de coeficiente de predição interno;

^c valor de coeficiente de correlação (ajuste externo) da reta que apresenta coeficiente linear igual a 0 entre os valores de pIC_{50} observados e os correspondentes preditos, obtido através gráfico i. da figura III.3.2.5;

^d valor da diferença relativa entre os valores de correlação de regressão externo obtidos através gráfico i. da figura III.3.2.5;

^e valor de coeficiente de correlação (ajuste externo) da reta que apresenta coeficiente linear igual a 0 valores de pIC_{50} observados e os correspondentes preditos , obtido através gráfico ii. da figura III.3.2.5;

^f valor da diferença relativa entre os valores de correlação de regressão externo obtidos através gráfico ii. da figura III.3.2.5;

^g valor de coeficiente angular da reta que apresenta coeficiente linear igual a 0 mostrada no gráfico i. da figura III.3.2.5;

^h valor de coeficiente angular da reta que apresenta coeficiente linear igual a 0 mostrada no gráfico ii. da figura III.3.2.5.

III.3.3. Análise de Frequência de Presença

Estão apresentados na tabela III.3.3.1, os descritores e os respectivos valores de frequência, de presença observados nos melhores modelos *PLS*, selecionados de acordo com critérios descritos no item II.5.3.3. *Análise dos Valores de Frequência de Presença*, para as séries de treinamento IIA, IIB e IIC, respectivamente.

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.

Descritores	Freq.	Freq. Rel.	Série IIA	Série IIB	Série IIC
DP01	3	100,00	x	x	x
GATS7e	3	100,00	x	x	x
AROM	2	66,67	x	x	
E1m	2	66,67	x		x
GATS5e	2	66,67	x	x	
GATS6e	2	66,67		x	x
H-049	2	66,67		x	x
H2e	2	66,67		x	x
J3D	2	66,67		x	x
MATS5m	2	66,67		x	x
Mor10m	2	66,67		x	x
Mor18u	2	66,67		x	x
nROR	2	66,67		x	x
nS	2	66,67		x	x
PJ12	2	66,67		x	x
PSA	2	66,67		x	x
R1m	2	66,67		x	x
R6e	2	66,67		x	x
R6u	2	66,67		x	x
R7e	2	66,67	x	x	
RDF020p	2	66,67		x	x
RDF100v	2	66,67	x		x
SPAN	2	66,67	x	x	
C-005	1	33,33	x		
C-028	1	33,33			x
GATS7v	1	33,33		x	
GGI3	1	33,33		x	
H0e	1	33,33	x		
H1e	1	33,33		x	
H2m	1	33,33			x
H2u	1	33,33		x	
HIC	1	33,33			x
IC2	1	33,33			x
Mor07v	1	33,33			x
Mor18e	1	33,33		x	
Ms	1	33,33			x
nDB	1	33,33		x	
R3m	1	33,33		x	
R4e	1	33,33			x
R5e	1	33,33			x
R5u	1	33,33			x
RDF075m	1	33,33			x
RDF120e	1	33,33		x	
VEA1	1	33,33			x
X4v	1	33,33		x	

^a selecionados de acordo com os critérios discutidos no item II.5.3.3.

Em seguida, os descritores foram selecionados para uma análise posterior, utilizando-se os critérios de análise de frequência proposto no item II.5.3.3. e, considerando-se os valores de frequência de presença de cada descritor nas três séries apresentados na tabela III.3.3.1. Estão assinalados em negrito na tabela III.3.3.1. os 23 descritores selecionados, apresentando valores de frequência maiores que 65% nos três modelos. Na tabela III.3.3.2 são apresentados os descritores selecionados e a respectiva classe, de acordo com a classificação proposta pelo programa DRAGON.

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

Descritor	Classe
GATS5e	Autocorrelações 2D
GATS6e	Autocorrelações 2D
GATS7e	Autocorrelações 2D
MATS5m	Autocorrelações 2D
Mor10m	Descritores 3D-MoRSE
Mor18u	Descritores 3D-MoRSE
H-049	Descritores átomo centrado
nS	Descritores Constitucionais
J3D	Descritores Geométricos
SPAN	Descritores Geométricos
R1m	Descritores GETAWAY
R6e	Descritores GETAWAY
R6u	Descritores GETAWAY
R7e	Descritores GETAWAY
RDF020p	Descritores RDF
RDF100v	Descritores RDF
PJI2	Descritores Topológicos
E1m	Descritores WHIM
nROR	Grupos Funcionais
AROM	Índices de aromaticidade
GGI3	Índices de Carga Topológica Galvez
PSA	Propriedades
DP01	Randic molecular profiles

^a com valores de frequência de presença maiores que 65% nos modelos das séries de treinamento IIA, IIB e IIC.

III.3.4. QSAR Clássico

As equações de QSAR foram geradas através dos descritores previamente selecionados, de acordo com critérios descritos no II.5.4.1 Análise do Modelo de QSAR Clássico Gerado e, utilizando-se o programa Bilin (Kubinyi 1995). Para a série II, foram gerados centenas de modelos QSAR com número de descritores (p) variando de 1 a 5.

Para a série II, o modelo que apresentou o maior valor do coeficiente estatístico de predição Q_{cv}^2 e, validado pelos 3 filtros propostos nos itens I.5.2.1, I.5.2.2 e I.5.2.3 (Regra QUIK, Assintótico, R^P e R^N) foi selecionado como o melhor modelo e, está expresso pela equação III.3.4.1.

$$pIC_{50} = -0,994 (\pm 0,34) nS + 0,674 (\pm 0,31) E1m - 0,736 (\pm 0,26) RDF020p + 1,469 (\pm 0,43) R7e + 0,433 (\pm 0,35) nROR + 7,366 (\pm 0,76) \quad \text{Equação III.3.4.1}$$

$$(n = 37; r = 0,909; s = 0,304; F = 29,558; Q_{cv}^2 = 0,762; s\text{-PRESS} = 0,357)$$

Observa-se na equação III.3.4.1 que o valor do coeficiente de ajuste de correlação explica 82,6% da variância dos dados.

A seguir, foram aplicadas sucessivamente as regras QUIK, do Q assintótico, e aquelas baseadas na função R para a validação do modelo expresso pela equação III.3.4.1.

Inicialmente, a validação do modelo foi feita utilizando critérios propostos pela regra QUIK. Desta forma, como já discutido no item II.5.4.2 Aplicação da Regra QUIK, modelos com valores de $K_{xy} < K_x$ devem ser rejeitados. Assim, para a série II, os valores dos índices de intercorrelação entre as variáveis independentes (K_x) bem como entre todas as variáveis (independentes e dependente) (K_{xy}) juntamente com os valores dos índices respectivamente de degeneração multivariada (D), de entropia total (S^+), de entropia efetiva (S) e, de entropia relativa (S_R) são apresentados na tabela III.3.4.1.

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.

Índices de correlação multivariada		Índices de entropia			Índice de degeneração	Contribuição percentual de cada descritor no valor do índice de degeneração multivariada				
K_x^a	K_{xy}^b	S^{+c}	S^d	S_R^e	D	nS	E1m	RDF020p	R7e	nROR
0,1042	0,2492	26,05	15,18	0,583	0,417	49,71%	0%	0%	0,57%	49,71%

^a valor de índice de correlação multivariada entre as variáveis independentes do modelo;

^b valor de índice de correlação multivariada entre todas as variáveis do modelo;

^c valor de entropia total;

^d valor de entropia efetiva e,

^e valor de entropia relativa;

Numa segunda etapa, a validação do modelo foi feita utilizando critérios propostos pela regra do Q^2 assintótico. Desta forma, como já discutido no II.5.4.3 *Aplicação da Regra do Q^2 Assintótico*, modelos com valores de $Q_{cv}^2 - Q_{ASYN}^2 < 0$ devem ser rejeitados

Assim, para a série II os valores dos índices 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 ($Q_{cv}^2 - Q_{ASYN}^2$) são apresentados na tabela III.3.4.2

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 ($Q_{cv}^2 - Q_{ASYN}^2$).

n^a	p'^b	r^c	r^{2d}	Q_{CV}^{2e}	Q_{ASYN}^{2f}	$Q_{CV}^2 - Q_{ASYN}^{2g}$
37	6	0,909	0,826	0,762	0,753	0,009

^a número de amostras utilizadas para a obtenção do modelo;

^b número de parâmetros presentes no modelo;

^c valor do coeficiente de correlação entre os valores preditos e os observados da literatura;

^d valor do coeficiente de correlação ao quadrado entre os valores preditos e os observados da literatura;

^e valor do coeficiente de predição pelo método *full cross-validation*;

^f valor do coeficiente de predição assintótico e,

^g diferença entre os valores dos coeficientes de predição pelos métodos *full cross-validation* e o assintótico, respectivamente.

E, finalmente a validação do modelo foi feita utilizando critérios propostos pelas regras baseadas na função R. Desta forma, como já discutido no item II.5.4.4, modelos com valores de $R^P < t^P$ ou $R^N < t^N$ devem ser rejeitados.

Assim, para a série I, os valores dos índices estatísticos descritivos respectivamente de correlação linear (r), de número de variáveis (p), de função R^P e de função R^N são apresentados na tabela III.3.4.3.

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^N .

r^a	p^b	$1/p$	R^{Pc}	t^{Pd}	R^{Ne}	t^{Nf}
0,909	5	0,2	0,15	0,05	-0,131	-0,189

^a valor do coeficiente de correlação entre os valores preditos e os observados da literatura;

^b número de variáveis presentes no modelo;

^c valor da função resultante do produto das penalidades das correlações relativas positivas das variáveis presentes no modelo;

^d valor mínimo aceitável para a função R^P (Todeschini *et al.* 2004);

^e valor da função resultante da soma das penalidades das correlações relativas negativas das variáveis presentes no modelo e,

^f valor mínimo aceitável para a função R^N para um ϵ de 0,01 (Todeschini *et al.* 2004).

Estão apresentados na figura III.3.4.1 a matriz de intercorrelação (r) das variáveis presentes no modelo expresso pela equação III.3.4.1.

	pIC ₅₀	nS	E1m	RDF020p	R7e	nROR
pIC ₅₀	1,000	-0,540	0,262	-0,417	0,545	0,063
nS	-0,540	1,000	0,056	0,037	-0,106	-0,121
E1m	0,262	0,056	1,000	0,061	-0,034	-0,015
RDF020p	-0,417	0,037	0,061	1,000	-0,026	0,180
R7e	0,545	-0,106	-0,034	-0,026	1,000	-0,203
nROR	0,063	-0,121	-0,015	0,180	-0,203	1,000

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

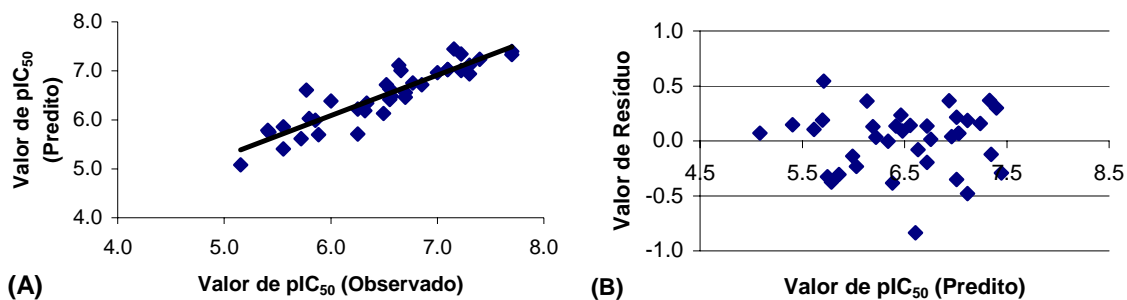
Para a série II, a partir do modelo expresso pela equação III.3.4.1 foram obtidos os respectivos valores previstos de pIC₅₀. Estes foram, a seguir comparados com os valores de pIC₅₀ observados, analisando-se os valores dos respectivos resíduos ($\Delta = \text{pIC}_{50\text{observado}} - \text{pIC}_{50\text{predito}}$) apresentados na tabela III.3.4.4 e nas figuras III.3.4.2 e III.3.4.3.

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.

Nº	pIC_{50} observado	pIC_{50} predito	$\Delta = pIC_{50\text{observado}} - pIC_{50\text{predito}}$
1	6,52	6,72	-0,20
2	6,55	6,42	0,14
3	5,16	5,08	0,07
4	6,85	6,72	0,14
5	6,55	6,63	-0,08
6	5,55	5,41	0,15
7	7,00	6,96	0,04
8	5,55	5,86	-0,31
9	6,25	5,71	0,54
10	6,66	7,01	-0,35
11	7,30	7,12	0,19
12	7,70	7,40	0,30
13	7,70	7,33	0,37
14	5,41	5,78	-0,38
15	6,25	6,22	0,03
16	5,89	5,70	0,19
17	6,70	6,56	0,14
18	6,32	6,19	0,13
19	6,77	6,76	0,02
20	6,50	6,13	0,36
21	7,22	7,35	-0,12
22	6,34	6,34	0,00
23	7,30	6,94	0,37
24	5,42	5,75	-0,33
25	5,72	5,62	0,11
26	6,70	6,46	0,24
27	7,10	7,03	0,07
28	6,57	6,48	0,09
29	6,64	7,12	-0,48
30	7,22	7,01	0,21
31	7,16	7,45	-0,29
32	7,10	7,03	0,07
33	7,40	7,24	0,16
34	5,80	6,03	-0,23
35	6,00	6,38	-0,38
36	5,77	6,61	-0,84
37	5,85	5,99	-0,14

^a (Du *et al.* 2002);

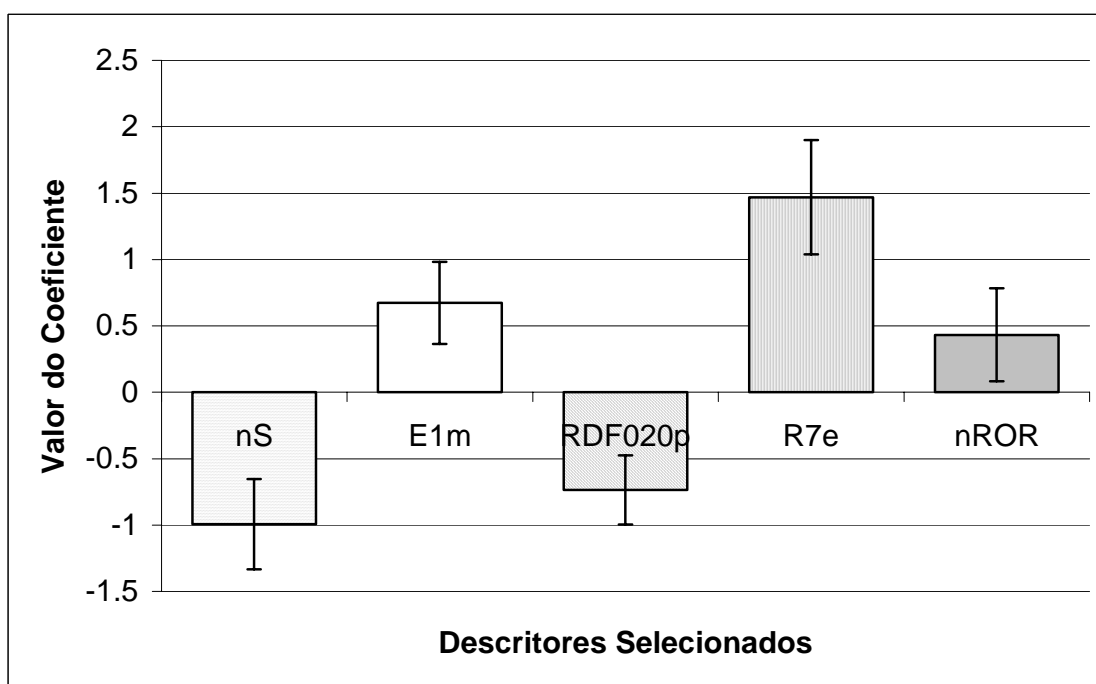
A partir dos valores apresentados na tabela III.3.4.4. foram construídos gráficos, respectivamente, da regressão linear entre os valores preditos e observados e dos respectivos resíduos em função dos valores preditos (figura III.3.4.2).



$$pIC_{50} = -0,994 (\pm 0,34) nS + 0,674 (\pm 0,31) E1m - 0,736 (\pm 0,26) RDF020p + 1,469 (\pm 0,43) R7e + 0,433 (\pm 0,35) nROR + 7,366 (\pm 0,76) \quad \text{Equação III.3.4.1}$$

($n = 37$; $r = 0,909$; $s = 0,304$; $F = 29,558$; $Q_{cv}^2 = 0,762$; $s\text{-PRESS} = 0,357$)

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.



$$pIC_{50} = -0,994 (\pm 0,34) nS + 0,674 (\pm 0,31) E1m - 0,736 (\pm 0,26) RDF020p + 1,469 (\pm 0,43) R7e + 0,433 (\pm 0,35) nROR + 7,366 (\pm 0,76) \quad \text{Equação III.3.4.1}$$

($n = 37$; $r = 0,909$; $s = 0,304$; $F = 29,558$; $Q_{cv}^2 = 0,762$; $s\text{-PRESS} = 0,357$)

Figura III.3.4.3. Representação gráfica dos valores dos coeficientes dos descritores presentes no modelo expresso pela equação III.3.4.1.