

III. Resultados

III.1. Série I

Para a série de 29 inibidores da RNR de células H.Ep.-2 (IRNR), selecionada da literatura original (French *et al.*, 1974), foram gerados 1497 descritores pelo programa DRAGON. Os dados referentes à série I estão apresentados na tabela II.1.2.1.

III.1.1. Pré-Tratamento de Dados

A partir dos valores dos descritores gerados, através da opção “*variable exclusion*” do programa DRAGON pode-se verificar que 793 (52,97 %) descritores apresentaram o mesmo valor, ou apenas um único valor diferente na série, 90 (6,01 %) 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 286 (19,10%).

III.1.2. Modelos PLS

Após este pré-tratamento, 12 análises PLS com 1 até 4 variáveis latentes foram aplicadas à série dos 286 descritores juntamente com os correspondentes valores de pIC_{50} . Os modelos PLS foram obtidos através de protocolo do programa ANALYZE (Embrechts, M. J. 2001) (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 variáveis latentes.

Estão apresentados nas tabelas III.1.2.1 e III.1.2.2. os valores dos termos estatísticos observados para os modelos PLS estatisticamente mais significativos, gerados para as séries

de treinamento e de teste (IA, IB e IC), considerando-se diferentes números de variáveis (descritores originais) em cada uma das análises *PLS*.

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.

N	Séries de Treinamento IA			Séries de Treinamento IB			Séries de Treinamento IC		
	r^2	Q_{cv}^2	SEP_{cv}	R^2	Q_{cv}^2	SEP_{cv}	r^2	Q_{cv}^2	SEP_{cv}
110	0,289	0,250	0,475	0,331	0,284	0,447	0,371	0,343	0,461
100	0,308	0,274	0,468	0,349	0,306	0,440	0,395	0,371	0,451
90	0,344	0,317	0,453	0,369	0,332	0,432	0,429	0,413	0,436
80	0,401	0,385	0,431	0,421	0,397	0,411	0,454	0,440	0,426
70	0,441	0,432	0,414	0,467	0,454	0,391	0,527	0,522	0,393
60	0,513	0,510	0,385	0,535	0,530	0,362	0,547	0,543	0,385
50	0,599	0,599	0,348	0,583	0,581	0,342	0,629	0,627	0,347
45	0,605	0,604	0,345	0,632	0,631	0,321	0,619	0,617	0,352
40	0,628	0,628	0,335	0,678	0,677	0,300	0,632	0,630	0,346
35	0,658	0,657	0,322	0,699	0,699	0,290	0,680	0,678	0,323
30	0,697	0,696	0,303	0,717	0,716	0,282	0,705	0,705	0,309
25	0,699^f	0,698^f	0,302^f	0,700	0,696	0,291	0,679	0,674	0,325
20	0,675	0,672	0,314	0,716	0,714	0,283	0,741	0,740	0,290
15	0,553	0,532	0,376	0,802^f	0,802^f	0,235^f	0,773^f	0,771^f	0,272^f
10	0,557	0,524	0,379	0,769	0,766	0,256	0,750	0,747	0,286
5	0,646	0,633	0,333	0,692	0,687	0,296	0,577	0,553	0,388

^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.1.2.1.;

^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.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 *PLS* gerados a partir das 3 séries de teste (*validação externa*) respectivamente IA, IB e IC, com 3 variáveis latentes, para diferentes números de variáveis (descritores originais) (N)^e.

N	Série de Teste IA			Série de Teste IB			Série de Teste IC		
	r^2	Q^2	SEP	r^2	Q^2	SEP	r^2	Q^2	SEP
110	0,943	0,935	0,124	0,709	0,638	0,328	0,906	0,391	0,274
100	0,944	0,933	0,125	0,724	0,659	0,318	0,904	0,374	0,277
90	0,941	0,935	0,124	0,727	0,659	0,318	0,896	0,348	0,283
80	0,943	0,934	0,124	0,709	0,650	0,322	0,897	0,328	0,287
70	0,946	0,938	0,120	0,696	0,648	0,323	0,898	0,365	0,280
60	0,949	0,943	0,116	0,724	0,662	0,316	0,884	0,383	0,275
50	0,949	0,945	0,114	0,680	0,613	0,339	0,918	0,465	0,257
45	0,950	0,944	0,114	0,684	0,611	0,339	0,921	0,420	0,267
40	0,944	0,938	0,121	0,696	0,630	0,331	0,930	0,369	0,279
35	0,969	0,956	0,102	0,692	0,623	0,334	0,934	0,453	0,259
30	0,964	0,959	0,099	0,695	0,609	0,340	0,938	0,514	0,244
25	0,948^f	0,946^f	0,113^f	0,661	0,551	0,365	0,877	0,475	0,254
20	0,940	0,928	0,130	0,646	0,459	0,400	0,865	0,327	0,288
15	0,918	0,904	0,151	0,528^f	0,338^f	0,443^f	0,870^f	0,433^f	0,264^f
10	0,971	0,933	0,126	0,656	0,322	0,448	0,823	0,602	0,221
5	0,911	0,836	0,196	0,520	0,294	0,457	0,795	0,294	0,295

^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.1.2.1.;

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

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

As análises comparativas dos 16 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.1.2.1. (validação interna) e III.1.2.2 (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 IA, IB e IC estão apresentados nas figuras III.1.2.1. a III.1.2.3, respectivamente.

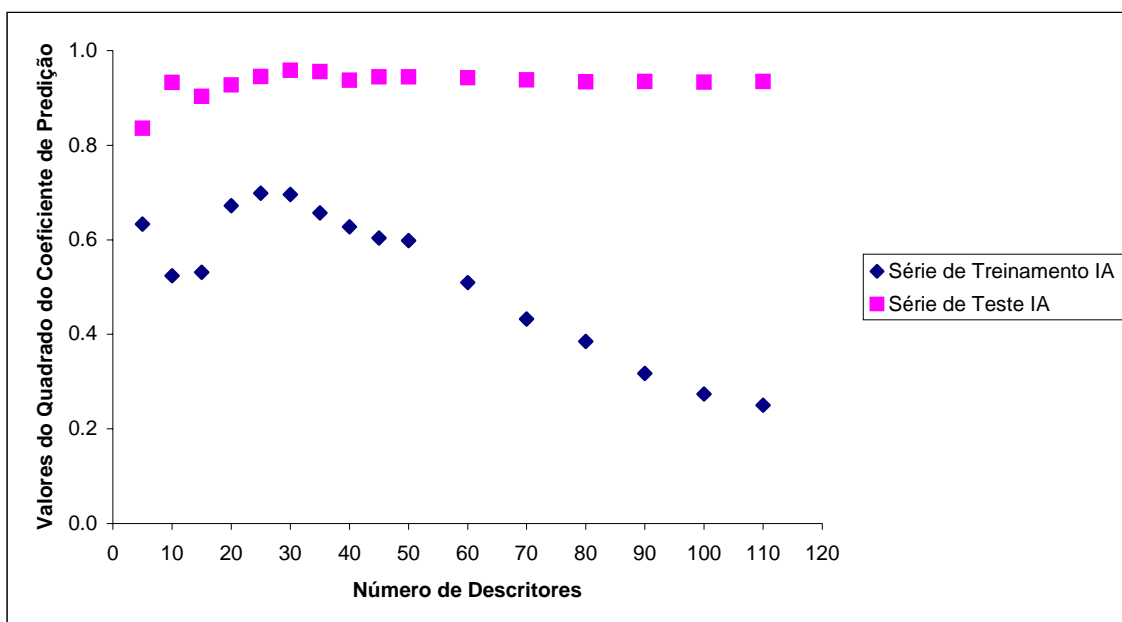


Figura III.1.2.1 Valores dos quadrados dos coeficientes de predição respectivamente Q_{cv}^2 e Q^2 , em função do número de descritores (N) considerados em cada um dos 16 modelos de *PLS* usando validação interna (série de treinamento) e validação externa (série de teste) e 3 variáveis latentes, para a série IA.

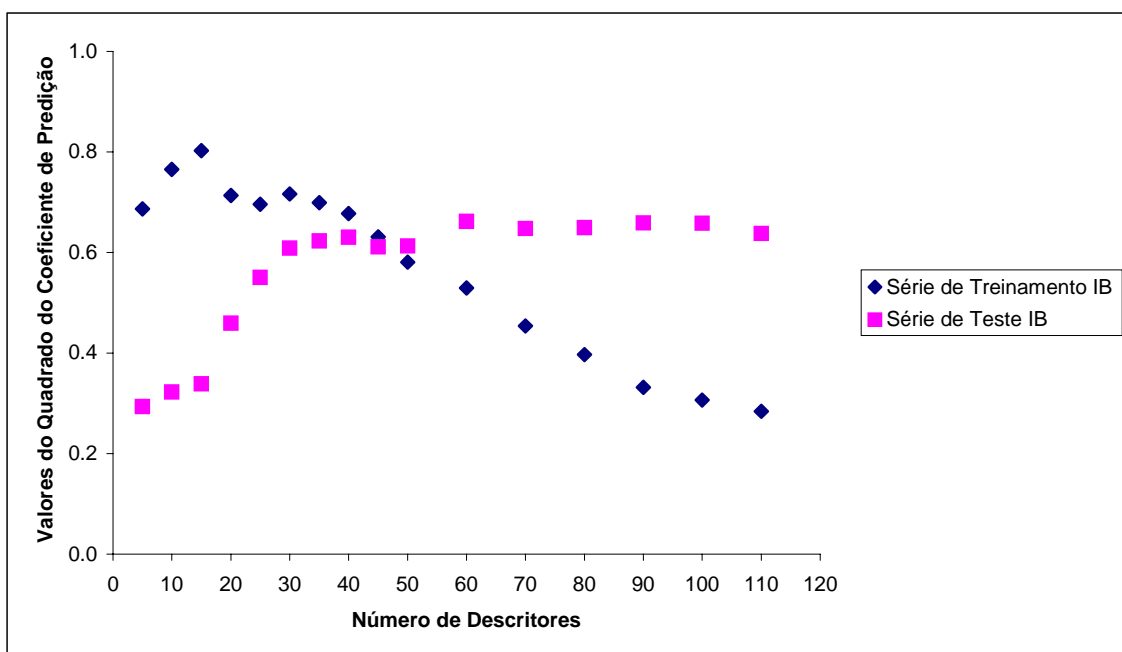


Figura III.1.2.2. Valores dos quadrados dos coeficientes de predição respectivamente Q_{cv}^2 e Q^2 , em função do número de descritores (N) considerados em cada um dos 16 modelos de *PLS* usando validação interna (série de treinamento) e validação externa (série de teste) e 3 variáveis latentes, para a série IB.

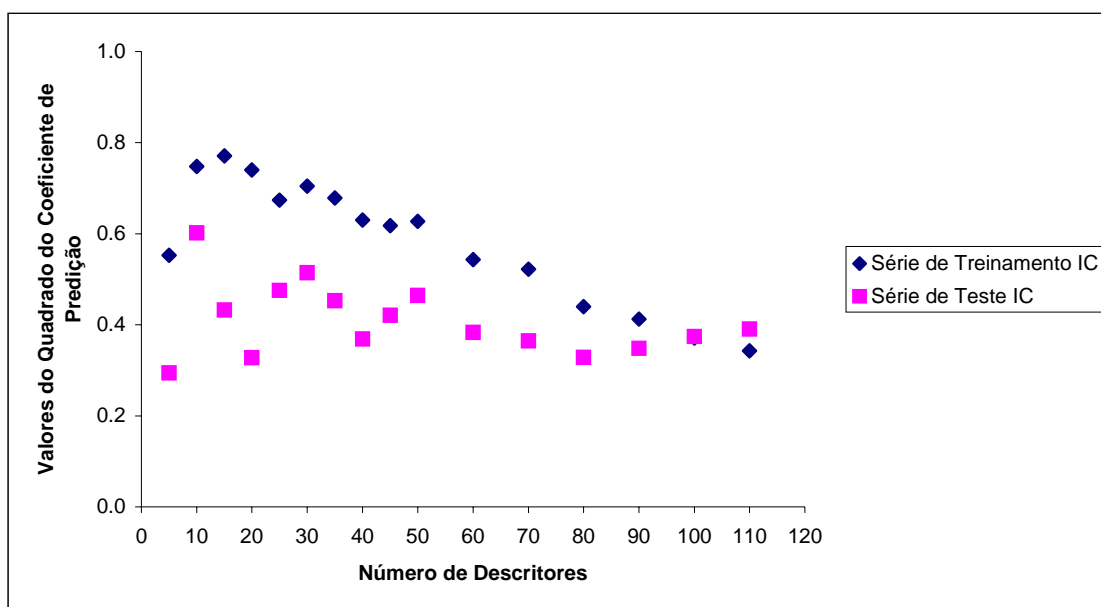


Figura III.1.2.3. Valores dos quadrados dos coeficientes de predição respectivamente Q_{cv}^2 e Q^2 , em função do número de descritores (N) considerados em cada um dos 16 modelos de *PLS* usando validação interna (série de treinamento) e validação externa (série de teste) e 3 variáveis latentes, para a série IC.

A análise visual dos gráficos apresentados nas figuras III.1.2.1 a III.1.2.3, 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.1.2.1 e III.1.2.2 e, sumarizados na tabela III.1.2.3.

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 *PLS* gerados a partir das 3 séries de treinamento respectivamente IA, IB e IC.

Parâmetros Estatísticos	Séries de Treinamento		
	IA	IB	IC
Nº de Variáveis Latentes	3	3	3
Nº de Variáveis Originais	25	15	15
R² (coeficiente de ajuste de regressão)	0,699	0,802	0,773
Q_{cv}² (coeficiente de predição interno)	0,698	0,802	0,771
SEP_{df} (desvio padrão de predição interno)	0,302	0,235	0,272
R²_{ext.} (coeficiente de ajuste de regressão externo)	0,948	0,528	0,870
Q²_{ext.} (coeficiente de predição externo)	0,946	0,338	0,433
SEP (desvio padrão de predição externo)	0,113	0,443	0,269

III.1.3. Análise da Frequência de Presença

Estão apresentados na tabela III.1.3.1, 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 IA, IB e IC, respectivamente.

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.

Descritor	Freq	Freq. Rel. %	Série IA	Série IB	Série IC
C-025	3	100,00	x	x	x
CIC2	3	100,00	x	x	x
H1p-REV	3	100,00	x	x	x
Mor07m-REV	3	100,00	x	x	x
Mor07p-REV	3	100,00	x	x	x
RDF020v	3	100,00	x	x	x
RDF065v-REV	3	100,00	x	x	x
BEHp6-REV	2	66,67	x		x
BEHv6-REV	2	66,67	x		x
GATS3e-REV	2	66,67	x	x	
GATS4e-REV	2	66,67	x		x
GATS4v	2	66,67	x	x	
H-046	2	66,67	x	x	
Jhetv	2	66,67	x	x	
MATS5m	2	66,67	x		x
MAXDP-REV	2	66,67	x		x
Mor07v-REV	2	66,67	x	x	
R4u	2	66,67	x		x
RDF065m-REV	2	66,67	x	x	
RDF115v-REV	2	66,67	x		x
Seige-REV	2	66,67	x		x
DISPm	1	33,33	x		
GATS8p	1	33,33		x	
IDDE	1	33,33		x	
IDDM-REV	1	33,33	x		
RDF115e-REV	1	33,33	x		
SEigp	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.1.3.1. Estão assinalados em negrito na tabela III.1.3.1. os 21 descritores selecionados, apresentando valores de frequência maiores que 65% nos três modelos. Na tabela III.1.3.2 são apresentados os descritores selecionados e a respectiva classe, de acordo com a classificação proposta pelo programa DRAGON.

Tabela III.1.3.2. Descritores da série I selecionados^a para posterior análise de *QSAR* clássico e as respectivas classes.

Descritor	Classe
GATS3e	Autocorrelação 2D
GATS4e	Autocorrelação 2D
GATS4v	Autocorrelação 2D
MATS5m	Autocorrelação 2D
C-025	Descritores – átomo centrado
H-046	Descritores – átomo centrado
Mor07m	Descritores 3D Morse
Mor07p	Descritores 3D Morse
Mor07v	Descritores 3D Morse
BEHp6	Descritores BCUT
BEHv6	Descritores BCUT
H1p	Descritores GATEWAY
R4u	Descritores GATEWAY
RDF020v	Descritores RDF
RDF065v	Descritores RDF
RDF065v	Descritores RDF
RDF115v	Descritores RDF
CIC2	Descritores Topológicos
Jhetv	Descritores Topológicos
MAXDP	Descritores Topológicos
Seige	Descritores Topológicos

^a com valores de frequência de presença maiores que 65% nos modelos das séries de treinamento IA, IB e IC

III.1.4. *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 I, foram gerados centenas de modelos *QSAR* com número de descritores (**p**) variando de 1 a 5. Para a série I, 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.1.4.1.

$$pIC_{50} = 0,175 (\pm 0,10) CIC2 + 0,138 (\pm 0,11) C-025 + 0,300 (\pm 0,11) Mor07m_{-REV} \\ + 0,144 (\pm 0,10) H1p_{-REV} + 0,207 (\pm 0,12) RDF065v_{-REV} + 5,791 (\pm 0,13)$$

(n = 29; r = 0,913; s = 0,248; F = 23,105; $Q_{cv}^2 = 0,758$; s-PRESS = 0,299) Equação III.1.4.1.

Observa-se na equação III.1.4.1 que o valor do coeficiente de correlação explica 83,4% da variância dos dados. O sufixo REV colocado junto a determinados descritores, por exemplo RDF065v_{-REV}, significa reversão de sinal do mesmo. Incluir esta notação nas equações de regressão se faz necessário, pois o programa ANALYZE utilizado para análise PLS, reverte automaticamente os valores dos coeficientes de regressão que não sejam positivos. Assim, por exemplo, na equação III.1.4.1 todos os valores das contribuições dos descritores são positivos. No entanto os sinais dos coeficientes dos três últimos termos estão invertidos, ou seja, os descritores Mor07m, H1p e RDF065v contribuem negativamente para a atividade inibitória RNR de células H.Ep.-2.

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.1.4.1.

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 I, 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.1.4.1.

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.

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	CIC2	C-025	Mor07m	H1p	RDF065v
0,2673	0,3633	24,29	4,984	0,2052	0,7948	25,05%	23,37%	15,62%	22,09%	13,88%

^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 I 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.1.4.2

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

n^a	p'^b	r^c	r^{2d}	Q_{CV}^{2e}	Q_{ASYN}^{2f}	$Q_{CV}^2 - Q_{ASYN}^{2g}$
29	6	0,913	0,834	0,758	0,735	0,023

^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.1.4.3

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 .

r^a	p^b	$1/p$	R^{Pc}	t^{Pd}	R^{Ne}	t^{Nf}
0,913	5	0,2	0,07	0,05	-0,046	-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.1.4.1 a matriz de intercorrelação (r) das variáveis presentes no modelo expresso pela equação III.1.4.1.

	pIC ₅₀	CIC2	C-025	Mor07m-REV	H1p-Rev	RDF065v-Rev
pIC ₅₀	1,000	0,141	0,527	0,628	0,274	0,649
CIC2	0,141	1,000	-0,124	-0,208	0,112	-0,163
C-025	0,527	-0,124	1,000	0,157	0,286	0,416
Mor07m-REV	0,628	-0,208	0,157	1,000	-0,168	0,408
H1p-REV	0,274	0,112	0,286	-0,168	1,000	-0,008
RDF065v-REV	0,649	-0,163	0,416	0,408	-0,008	1,000

Figura III.1.4.1. Matriz de correlação (r) das variáveis presentes no modelo de QSAR selecionado, expresso pela equação III.1.4.1.

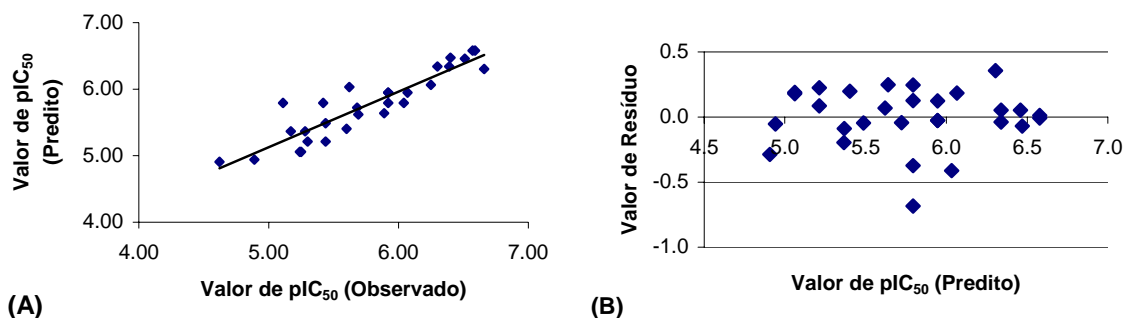
Para a série I. a partir do modelo expresso pela equação III.1.4.1 foram obtidos os respectivos valores previstos de pIC₅₀. Estes foram, a seguir comparados com os valores de pIC₅₀ observados, analisando-se os valores dos respectivos resíduos ($\Delta = \text{pIC}_{50\text{observado}} - \text{pIC}_{50\text{predito}}$) apresentados na Tabela III.1.4.4 e nas figuras III.1.4.2 e III.1.4.3.

Tabela III.1.4.4. Valores de pIC_{50} previstos pelo modelo *QSAR* 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.

Nº	pIC_{50} observado	pIC_{50} predito	$\Delta = pIC_{50\text{observado}} - pIC_{50\text{predito}}$
1	6,59	6,58	0,01
2	6,57	6,58	-0,01
3	6,51	6,46	0,05
4	6,66	6,30	0,36
5	5,11	5,79	-0,68
6	5,89	5,64	0,25
7	6,04	5,79	0,25
8	5,44	5,49	-0,05
9	5,42	5,79	-0,37
10	5,92	5,79	0,13
11	6,25	6,07	0,18
12	6,30	6,34	-0,04
13	6,39	6,34	0,05
14	5,62	6,03	-0,41
15	5,92	5,95	-0,03
16	5,60	5,40	0,20
17	6,07	5,95	0,12
18	4,62	4,91	-0,29
19	5,69	5,62	0,07
20	5,44	5,21	0,23
21	5,28	5,37	-0,09
22	5,17	5,37	-0,20
23	5,30	5,21	0,09
24	5,25	5,06	0,19
25	5,24	5,06	0,18
26	4,89	4,94	-0,05
27	5,92	5,95	-0,03
28	6,40	6,47	-0,07
29	5,68	5,72	-0,04

^a (FRENCH *et al.* 1974);

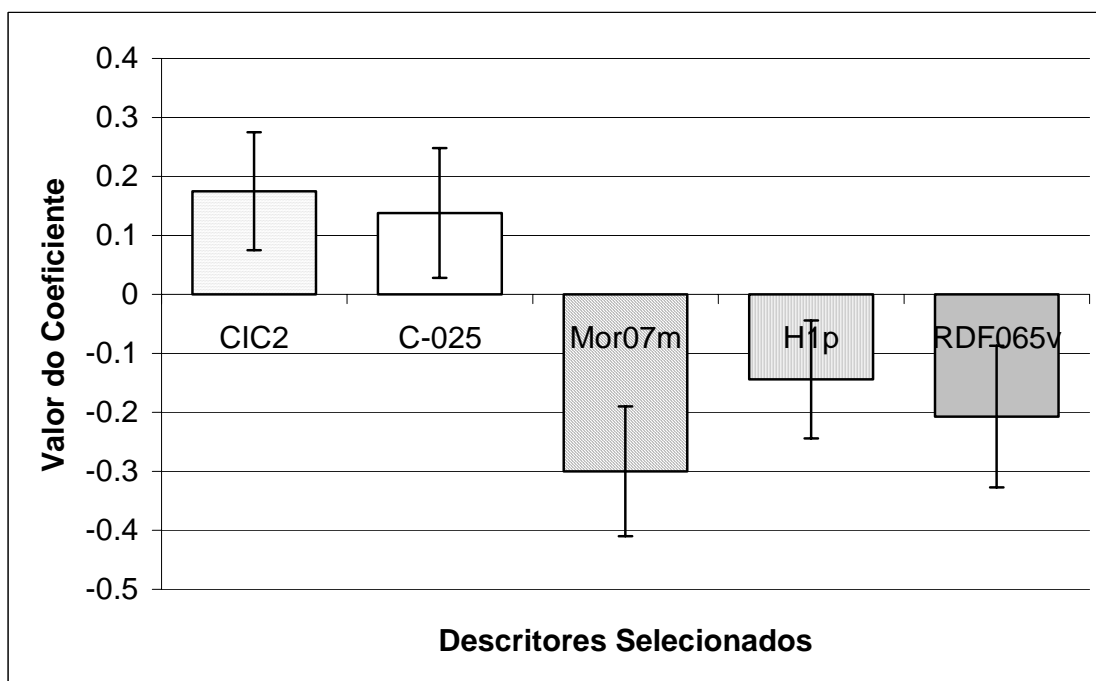
A partir dos valores apresentados na tabela III.1.4.4. foram construídos gráficos, respectivamente, da regressão linear entre os valores preditos e observados e dos respectivos resíduos em função dos valores preditos (figura III.1.4.2).



$$\text{pIC}_{50} = 0,175 (\pm 0,10) \text{ CIC2} + 0,138 (\pm 0,11) \text{ C-025} + 0,300 (\pm 0,11) \text{ Mor07m-REV} \\ + 0,144 (\pm 0,10) \text{ H1p-REV} + 0,207 (\pm 0,12) \text{ RDF065v-REV} + 5,791 (\pm 0,13)$$

(n = 29; r = 0,913; s = 0,248; F = 23,105; Q_{cv}² = 0,758; s-PRESS = 0,299) Equação III.1.4.1.

Figura III.1.4.2. Gráficos, respectivamente, da correlação (A) e dos valores dos resíduos (B) entre os valores de pIC₅₀ preditos e observados pelo modelo de QSAR, expresso pela equação III.1.4.1.



$$\text{pIC}_{50} = 0,175 (\pm 0,10) \text{ CIC2} + 0,138 (\pm 0,11) \text{ C-025} + 0,300 (\pm 0,11) \text{ Mor07m-REV} \\ + 0,144 (\pm 0,10) \text{ H1p-REV} + 0,207 (\pm 0,12) \text{ RDF065v-REV} + 5,791 (\pm 0,13)$$

(n = 29; r = 0,913; s = 0,248; F = 23,105; Q_{cv}² = 0,758; s-PRESS = 0,299) Equação III.1.4.1.

Figura III.1.4.3. Representação gráfica dos valores dos coeficientes dos descritores presentes no modelo expresso pela equação III.1.4.1.