

III.4. Série III

Para a série III (n=61) constituída: pelos compostos II.1 a II.37 (n=37) da série II (Du *et al.* 2002); pelos compostos III.38 a III.53 (n=16) (Chiyanzu *et al.* 2003) e, pelos compostos III.54 a III.61 (n= 8) (Greenbaum *et al.* 2004)}, com atividade frente à cruzaína do *T. cruzi*, selecionados da literatura, foram gerados 1497 descritores pelo programa DRAGON. Os dados referentes à série III estão apresentados na tabela II.1.4.1.

III.4.1. Pré-Tratamento de Dados

A partir dos valores dos descritores gerados , através da opção “*variable exclusion*” do programa DRAGON pode-se verificar que 250 (16,70 %) descritores apresentaram o mesmo valor, ou apenas um único valor diferente na série, 186 (12,42 %) descritores apresentaram valores superiores a 4 desvios padrão (4σ) em relação à respectiva média e, 517 (34,54 %) 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 544 descritores (36,34 %).

III.4.2. Modelos PLS

Após este pré-tratamento, 12 análises *PLS* com 1 até 4 variáveis latentes foram aplicadas à série dos 544 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 4 variáveis latentes.

Estão apresentados nas tabelas III.4.2.1 e III.4.2.2. os valores dos termos estatísticos observados para os modelos *PLS* estatisticamente mais significativos, gerados para as séries

de treinamento e de teste (IIIA, IIIB e IIIC), considerando-se diferentes números de variáveis (descritores originais) em cada uma das análises *PLS*.

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, IIIB e IIIC, com 4 variáveis latentes, para diferentes números de variáveis (descritores originais) (N)^e.

N ^a	Séries de Treinamento IIIA			Séries de Treinamento IIIB			Séries de Treinamento IIIC		
	r^2	Q_{cv}^2	SEP_{cv}	r^2	Q_{cv}^2	SEP_{cv}	r^2	Q_{cv}^2	SEP_{cv}
500	0,696	0,694	0,581	0,664	0,661	0,597	0,682	0,669	0,614
450	0,720	0,718	0,557	0,687	0,684	0,576	0,712	0,703	0,581
400	0,739	0,738	0,538	0,707	0,704	0,557	0,739	0,730	0,554
350	0,757	0,756	0,519	0,734	0,732	0,530	0,759	0,752	0,531
300	0,782	0,781	0,491	0,759	0,757	0,505	0,778	0,771	0,510
250	0,794	0,792	0,479	0,790	0,789	0,470	0,803	0,799	0,478
200	0,812	0,811	0,457	0,805	0,802	0,456	0,834	0,833	0,436
150	0,839	0,838	0,423	0,825	0,825	0,429	0,856	0,855	0,406
100	0,862	0,861	0,392	0,838	0,837	0,413	0,875	0,873	0,379
90	0,861	0,861	0,392	0,837	0,836	0,415	0,879	0,877	0,373
80	0,869	0,868	0,381	0,862	0,862	0,380	0,897	0,897	0,343
70	0,877	0,875	0,371	0,867	0,866	0,374	0,894	0,894	0,348
60	0,885	0,885	0,356	0,863	0,862	0,380	0,897	0,896	0,343
50	0,888	0,888	0,352	0,882	0,882	0,352	0,899	0,899	0,339
45	0,901	0,901	0,330	0,893	0,893	0,335	0,904^f	0,904^f	0,330^f
40	0,911	0,911	0,313	0,900	0,900	0,324	0,903	0,903	0,333
35	0,915	0,915	0,306	0,897	0,897	0,329	0,894	0,894	0,348
30	0,919^f	0,918^f	0,300^f	0,893	0,893	0,335	0,892	0,892	0,351
25	0,893	0,892	0,346	0,904^f	0,903^f	0,318^f	0,888	0,888	0,357
20	0,892	0,890	0,349	0,886	0,885	0,348	0,876	0,874	0,378
15	0,897	0,896	0,339	0,882	0,881	0,354	0,889	0,889	0,355
10	0,848	0,847	0,411	0,831	0,829	0,424	0,868	0,867	0,389
5	0,740	0,737	0,539	0,672	0,670	0,589	0,605	0,600	0,674

^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 nas tabelas II.1.3.1 e II.1.4.1;

^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.4.2.2. 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 IIIA, IIIB e IIIC, com 4 variáveis latentes, para diferentes números de variáveis (descritores originais) (N)^e.

N ^a	Série de Teste IIIA			Série de Teste IIIB			Série de Teste IIIC		
	r^2	Q^2	SEP	r^2	Q^2	SEP	r^2	Q^2	SEP
500	0,678	0,605	0,589	0,802	0,776	0,501	0,609	0,457	0,635
450	0,682	0,610	0,585	0,795	0,770	0,508	0,622	0,470	0,628
400	0,686	0,612	0,584	0,796	0,771	0,506	0,632	0,485	0,619
350	0,686	0,610	0,585	0,800	0,774	0,503	0,633	0,480	0,621
300	0,698	0,624	0,575	0,798	0,771	0,506	0,655	0,510	0,604
250	0,709	0,649	0,555	0,811	0,785	0,490	0,657	0,506	0,606
200	0,712	0,657	0,548	0,820	0,786	0,490	0,652	0,510	0,604
150	0,699	0,619	0,578	0,812	0,778	0,499	0,647	0,505	0,607
100	0,717	0,632	0,568	0,851	0,801	0,472	0,632	0,467	0,629
90	0,725	0,646	0,558	0,865	0,808	0,464	0,631	0,474	0,625
80	0,705	0,615	0,581	0,864	0,805	0,467	0,634	0,464	0,631
70	0,704	0,624	0,574	0,862	0,815	0,456	0,629	0,449	0,640
60	0,716	0,645	0,558	0,857	0,812	0,459	0,613	0,427	0,653
50	0,718	0,649	0,555	0,880	0,815	0,455	0,595	0,439	0,646
45	0,747	0,684	0,527	0,869	0,808	0,463	0,626^f	0,496^f	0,612^f
40	0,741	0,676	0,534	0,850	0,785	0,490	0,638	0,510	0,603
35	0,728	0,658	0,548	0,897	0,826	0,441	0,568	0,393	0,672
30	0,745^f	0,690^f	0,522^f	0,939	0,885	0,359	0,529	0,318	0,712
25	0,833	0,786	0,434	0,942^f	0,902^f	0,331^f	0,526	0,313	0,714
20	0,894	0,845	0,369	0,920	0,878	0,369	0,572	0,414	0,660
15	0,889	0,839	0,376	0,886	0,838	0,426	0,578	0,349	0,695
10	0,864	0,834	0,382	0,789	0,760	0,518	0,526	0,259	0,742
5	0,602	0,542	0,634	0,715	0,679	0,599	0,627	0,341	0,700

^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 nas tabelas II.1.3.1. e II.1.4.1;

^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 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.4.2.1. (validação interna) e III.4.2.2 (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 IIIA, IIIB e IIIC estão apresentados nas figuras III.4.2.1. a III.4.2.3, respectivamente.

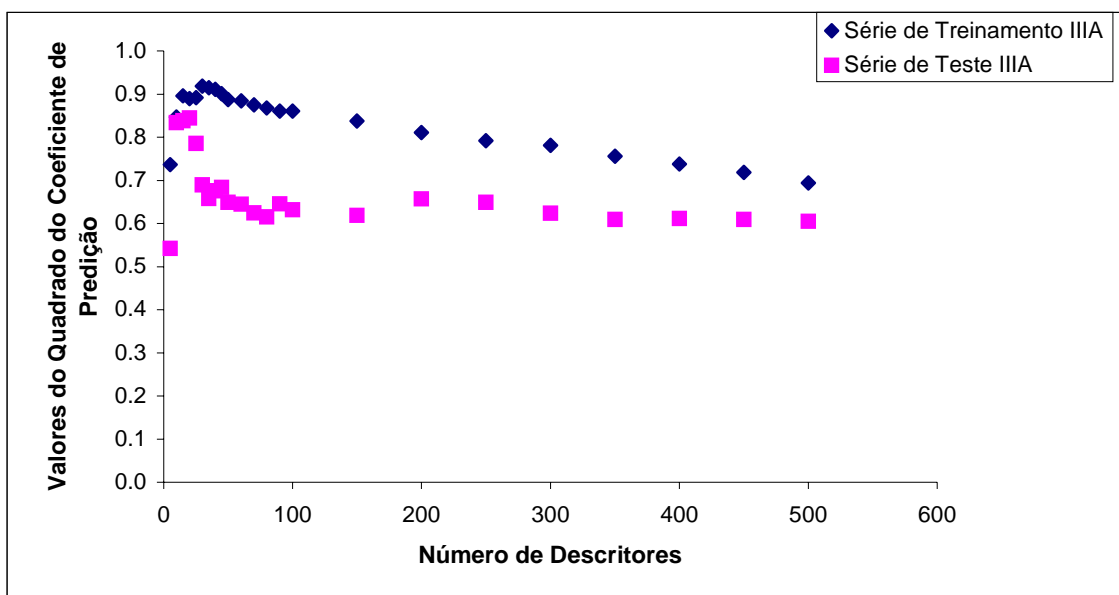


Figura III.4.2.1. Valores dos quadrados dos coeficientes de predição respectivamente Q_{cv}^2 e Q^2 , em função do número de descritores (N) considerados em cada um dos 23 modelos de *PLS* usando validação interna (série de treinamento) e validação externa (série de teste) e 4 variáveis latentes, para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

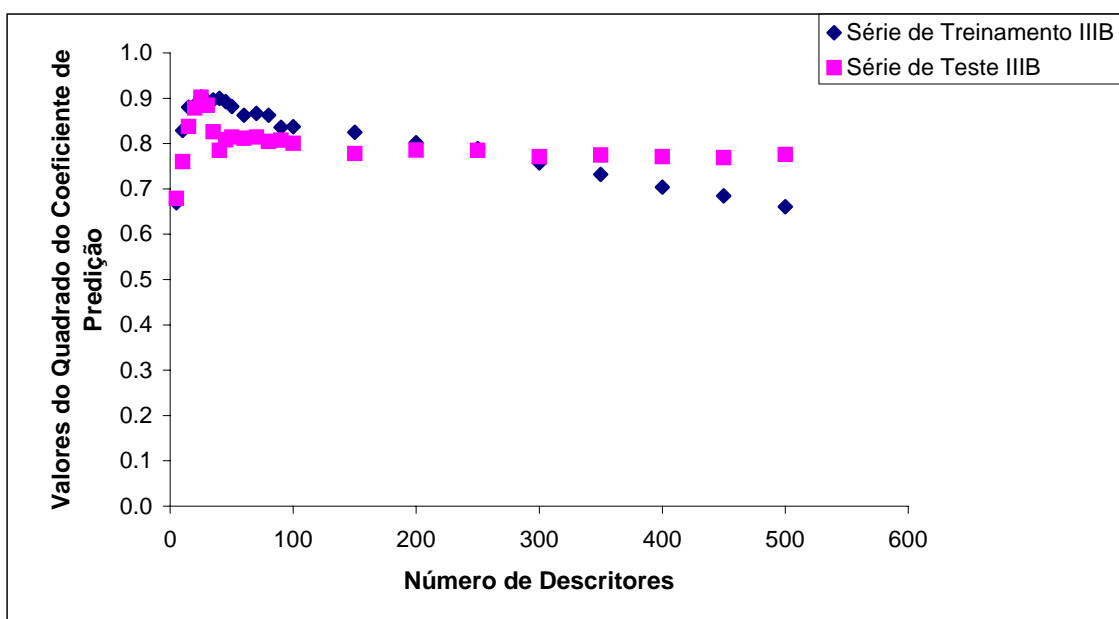


Figura III.4.2.2. Valores dos quadrados dos coeficientes de predição respectivamente Q_{cv}^2 e Q^2 , em função do número de descritores (N) considerados em cada um dos 23 modelos de *PLS* usando validação interna (série de treinamento) e validação externa (série de teste) e 4 variáveis latentes, para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

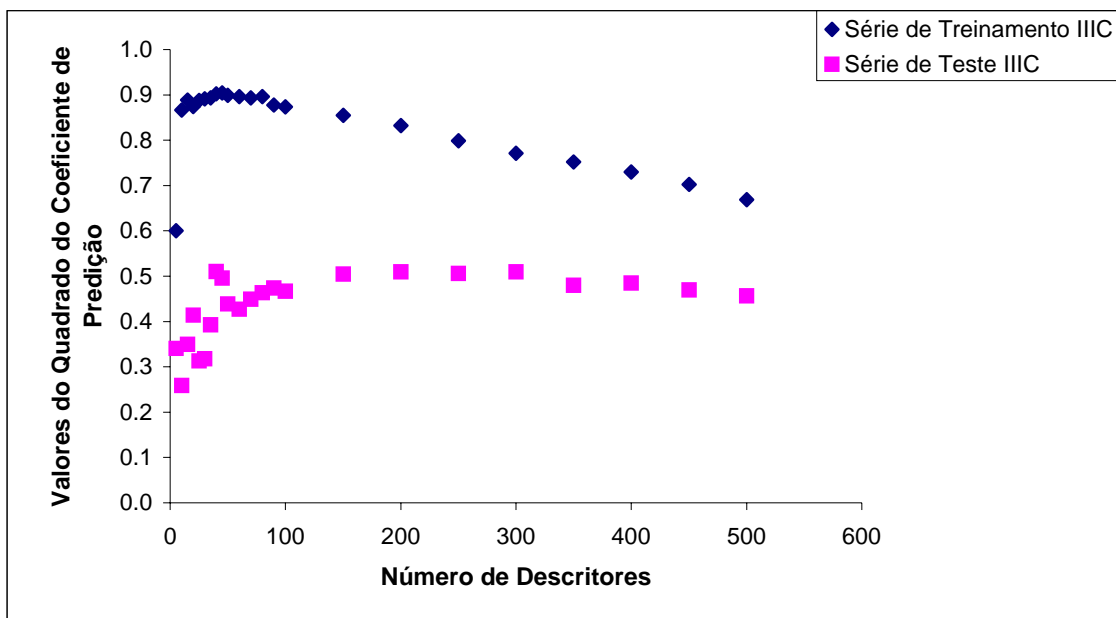


Figura III.4.2.3. Valores dos quadrados dos coeficientes de predição respectivamente Q_{cv}^2 e Q^2 , em função do número de descritores (N) considerados em cada um dos 23 modelos de *PLS* usando validação interna (série de treinamento) e validação externa (série de teste) e 4 variáveis latentes, para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

A análise visual dos gráficos, apresentados nas figuras III.4.2.1 a III.4.2.3, 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.4.2.1 e III.4.2.2 e, resumidos na tabela III.4.2.3.

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, IIIB e IIIC.

Parâmetros Estatísticos	Séries de Treinamento		
	IIIA	IIIB	IIIC
Nº de Variáveis Latentes	4	4	4
Nº de Variáveis Originais	30	25	45
R² (coeficiente de ajuste de regressão)	0,919	0,904	0,904
Q_{cv}² (coeficiente de predição interno)	0,918	0,903	0,904
SEP_{df} (desvio padrão de predição interno)	0,300	0,318	0,330
R²_{ext.} (coeficiente de ajuste de regressão externo)	0,745	0,942	0,626
Q²_{ext.} (coeficiente de predição externo)	0,690	0,902	0,496
SEP (desvio padrão de predição externo)	0,522	0,331	0,612

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.4.2.4, respectivamente para as séries IIIA, IIIB e IIIC.

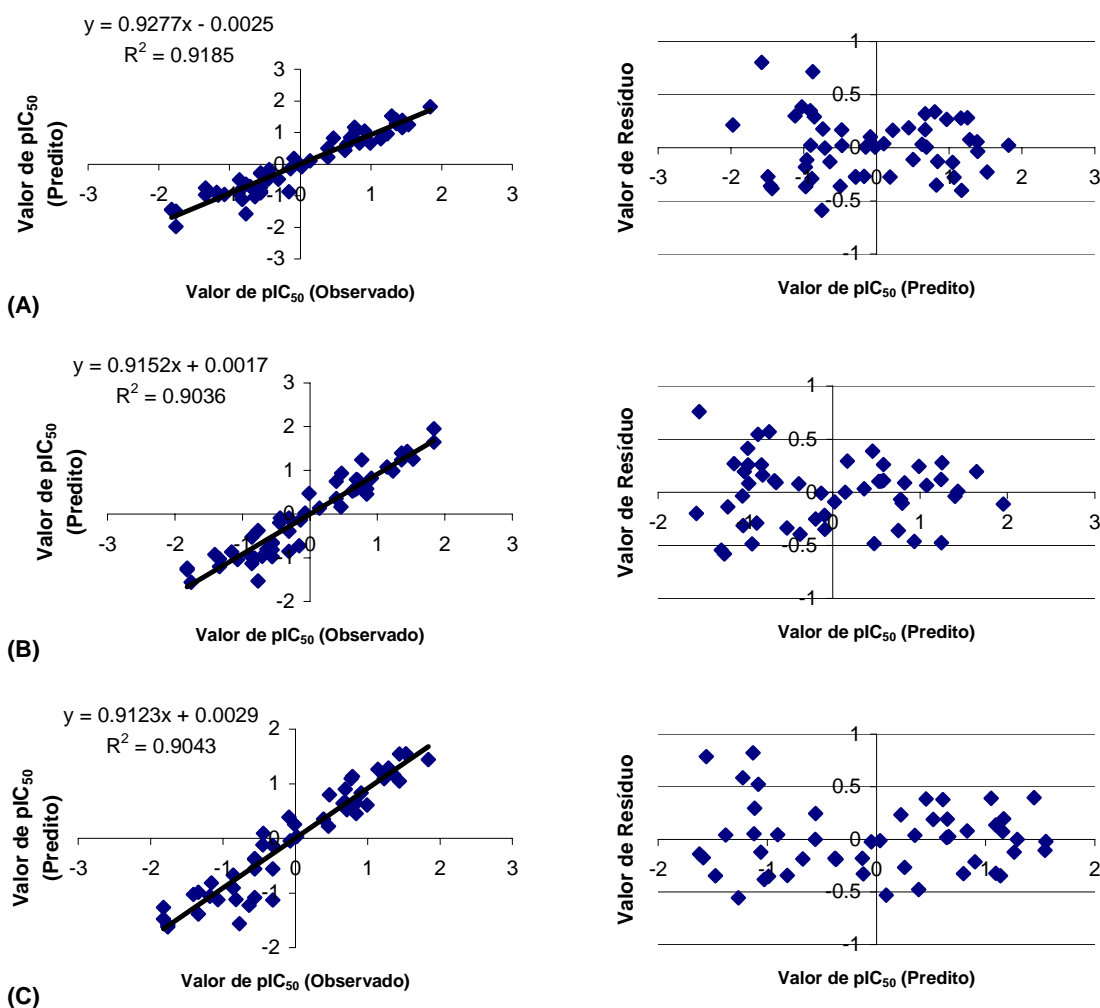


Figura III.4.2.4. Gráficos da correlação e dos resíduos entre os valores de pIC_{50} preditos e, observados a partir do modelo *PLS* (selecionado através do critério apresentado no item II.5.3.1.) para as séries de treinamento respectivamente:

- (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.

A seguir, para cada uma das séries de teste IIIA, IIIB e IIIC 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.4.2.5, respectivamente para as séries IIIA, IIIB e IIIC.

A seguir, utilizando os valores dos coeficientes estatísticos para cada uma das correlações, aplicou-se para as séries de teste IIIA, IIIB e IIIC 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.4.2.4, respectivamente para as séries IIIA, IIIB e IIIC.

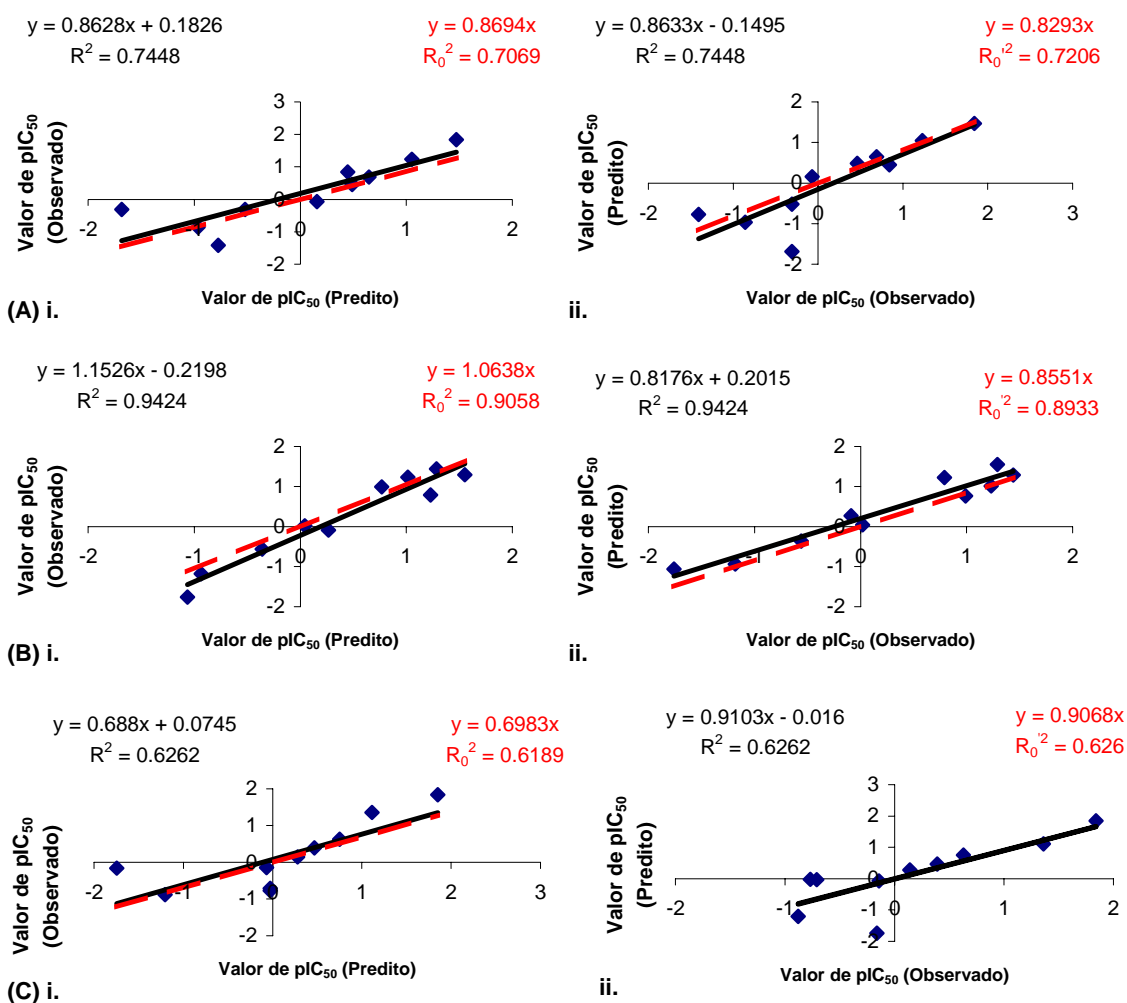


Figura III.4.2.5. Gráficos de regressão entre os valores de pIC_{50} observados e os correspondentes preditos através do modelo *PLS* (selecionado através do critério de apresentado no item II.5.3.1.) respectivamente para :

- (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.

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).

Séries de Teste	R^{2a}	$Q_{cv}^2{}^b$	$R_0^2{}^c$	$(R^2-R_0^2)/R^{2d}$	$R_0'^2{}^e$	$(R^2-R_0'^2)/R^{2f}$	k^g	$k'{}^h$
IIIA	0,745	0,918	0,7069	0,051	0,7206	0,032	0,8694	0,8293
IIIB	0,942	0,903	0,9058	0,039	0,8933	0,052	1,0638	0,8551
IIIC	0,626	0,904	0,6189	0,012	0,6260	0,000	0,6983	0,9068

^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.4.2.5;

^d valor da diferença relativa entre os valores de correlação de regressão externo obtidos através gráfico i. da figura III.4.2.5;

^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.4.2.5;

^f valor da diferença relativa entre os valores de correlação de regressão externo obtidos através gráfico ii. da figura III.4.2.5;

^g valor de coeficiente angular da reta que apresenta coeficiente linear igual a 0 mostrada no gráfico i. da figura III.4.2.5;

^h valor de coeficiente angular da reta que apresenta coeficiente linear igual a 0 mostrada no gráfico ii. da figura III.4.2.5.

III.4.3. Análise de Frequência de Presença

Estão apresentados na tabela III.4.3.1, 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.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.

Descritores	Freq.	Freq. Rel.	Série IIIA	Série IIIB	Série IIIC
BELe2	3	100.00	x	x	x
C-028	3	100.00	x	x	x
C-034	3	100.00	x	x	x
C-035	3	100.00	x	x	x
G1e	3	100.00	x	x	x
GATS8v	3	100.00	x	x	x
HATS7e	3	100.00	x	x	x
MATS8v	3	100.00	x	x	x
Mor26v	3	100.00	x	x	x
Mor27p	3	100.00	x	x	x
nNH2	3	100.00	x	x	x
nROR	3	100.00	x	x	x
PJ12	3	100.00	x	x	x
R3v+	3	100.00	x	x	x
R6e	3	100.00	x	x	x
RDF035m	3	100.00	x	x	x
BEHm3	2	66.67	x	x	
BEHp7	2	66.67		x	x
BELe1	2	66.67	x	x	
E2u	2	66.67	x		x
H7v	2	66.67	x	x	
HATS4u	2	66.67		x	x
IC2	2	66.67	x	x	
MATS8m	2	66.67	x	x	
R6u	2	66.67		x	x
BEHm4	1	33.33			x
BELe3	1	33.33			x
BIC2	1	33.33	x		
E1s	1	33.33			x
GATS5v	1	33.33			x
GATS8e	1	33.33			x
H-051	1	33.33		x	
H-052	1	33.33			x
HATS4v	1	33.33			x
HATS7p	1	33.33			x
JG11	1	33.33	x		
Mor04m	1	33.33	x		
Mor10e	1	33.33			x
Mor11e	1	33.33	x		
Mor23e	1	33.33			x
Mor23m	1	33.33			x
Mor26e	1	33.33			x
Mor27u	1	33.33			x
MR	1	33.33			x
nCIR	1	33.33			x
nCONR2	1	33.33			x
nRORPh	1	33.33	x		
O-058	1	33.33			x
PSA	1	33.33			x
R3e+	1	33.33	x		
R3p+	1	33.33			x
R4u	1	33.33			x
R7e	1	33.33			x
R7e+	1	33.33			x
RBN	1	33.33			x
RDF065m	1	33.33			x
RDF095u	1	33.33			x
RDF100v	1	33.33	x		
SPH	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.4.3.1. Estão assinalados em negrito na tabela III.4.3.1. os 25 descritores selecionados, apresentando valores de frequência maiores que 65% nos três modelos. Na tabela III.4.3.2 são apresentados os descritores selecionados e a respectiva classe de acordo com a classificação proposta pelo programa DRAGON.

Tabela III.4.3.2. Descritores da série III selecionados^a para posterior análise de QSAR clássico e as respectivas classes.

Descritor	Classe
GATS8v	Autocorrelação 2D
MATS8m	Autocorrelação 2D
MATS8v	Autocorrelação 2D
Mor26v	Descritor 3D
Mor27p	Descritor 3D
C-028	Descritores átomo centrado
C-034	Descritores átomo centrado
C-035	Descritores átomo centrado
BEHm3	Descritores BCUT
BEHp7	Descritores BCUT
BELe1	Descritores BCUT
BELe2	Descritores BCUT
H7v	Descritores GETAWAY
HATS4u	Descritores GETAWAY
HATS7e	Descritores GETAWAY
R3v+	Descritores GETAWAY
R6e	Descritores GETAWAY
R6u	Descritores GETAWAY
RDF035m	Descritores RDF
IC2	Descritores Topológicos
PJ12	Descritores Topológicos
E2u	Descritores WHIM
G1e	Descritores WHIM
nNH2	Grupos funcionais
NROR	Grupos funcionais

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

III.4.4. 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 III, foram gerados centenas de modelos *QSAR* com número de descritores (**p**) variando de 1 a 5. Para a série III, 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.4.4.1.

$$pIC_{50} = -1,287 (\pm 0,28) \text{ C-035} - 2,335 (\pm 0,75) \text{ BEHm3} - 18,66 (\pm 6,67) \text{ MATS8m} \\ + 2,693 (\pm 0,63) \text{ MATS8v} + 1,193 (\pm 0,48) \text{ nROR} + 32,98 (\pm 7,85)$$

$$(n = 61; r = 0,910; s = 0,433; F = 53,030; Q_{cv}^2 = 0,792; s\text{-PRESS} = 0,477) \quad \text{Equação III.4.4.1}$$

Observa-se na equação III.3.4.1 que o valor do coeficiente de correlação explica 82,8% 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.4.4.1.

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 III, 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.4.4.1.

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.

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	C-035	BEHm3	MATS8m	MATS8v	nROR
0,2128	0,3264	29,65	14,88	0,502	0,498	43,51%	1,40%	6,36%	1,12%	47,61%

^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 III 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.4.4.2

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

n^a	p'^b	r^c	r^{2d}	Q_{CV}^{2e}	Q_{ASYN}^{2f}	$Q_{CV}^2 - Q_{ASYN}^{2g}$
61	6	0,910	0,828	0,792	0,789	0,003

^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;

^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 III, 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.4.4.3.

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 .

r^a	p^b	$1/p$	R^{Pc}	t^{Pd}	R^{Ne}	t^{Nf}
0,910	5	0,2	0,18	0,05	-0,040	-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.4.4.1 a matriz de intercorrelação (r) das variáveis presentes no modelo expresso pela equação III.4.4.1.

	pIC50	C-035	BEHm3	MATS8m	MATS8v	nROR
pIC50	1,000	-0,685	-0,146	-0,285	0,411	0,202
C-035	-0,685	1,000	-0,205	0,344	-0,032	-0,158
BEHm3	-0,146	-0,205	1,000	-0,332	-0,134	0,204
MATS8m	-0,285	0,344	-0,332	1,000	0,313	-0,018
MATS8v	0,411	-0,032	-0,134	0,313	1,000	-0,219
nROR	0,202	-0,158	0,204	-0,018	-0,219	1,000

Figura III.4.4.1. Matriz de intercorrelação (r) das variáveis presentes no modelo de QSAR selecionado, expresso pela equação III.4.4.1.

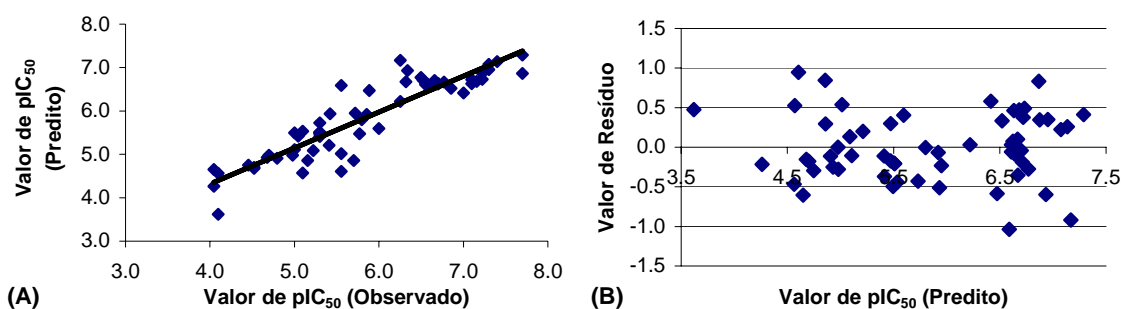
Para a série III, a partir do modelo expresso pela equação III.4.4.1 foram obtidos os respectivos valores previstos de pIC₅₀. Estes foram, a seguir comparados com os valores de pIC₅₀ observados, analisando-se os valores dos respectivos resíduos ($\Delta = \text{pIC}_{50\text{observado}} - \text{pIC}_{50\text{predito}}$) apresentados na tabela III.4.4.4 e nas figuras III.4.4.2 e III.4.4.3.

Tabela III.4.4.4. Valores de pIC_{50} previstos pelo modelo *QSAR* 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.

Nº	pIC_{50} observado	pIC_{50} predito	$\Delta = pIC_{50\text{observado}} - pIC_{50\text{predito}}$
1	6,52	6,71	-0,19
2	6,55	6,64	-0,09
3	5,16	4,86	0,29
4	6,85	6,52	0,33
5	6,55	6,61	-0,06
6	5,55	5,01	0,54
7	7,00	6,42	0,58
8	5,55	6,59	-1,04
9	6,25	6,22	0,03
10	6,66	6,70	-0,04
11	7,30	7,08	0,23
12	7,70	7,29	0,41
13	7,70	6,87	0,83
14	5,41	5,21	0,20
15	6,25	7,17	-0,92
16	5,89	6,47	-0,59
17	6,70	6,64	0,06
18	6,32	6,67	-0,35
19	6,77	6,67	0,10
20	6,50	6,77	-0,28
21	7,22	6,88	0,35
22	6,34	6,93	-0,60
23	7,30	6,95	0,35
24	5,42	5,93	-0,51
25	5,72	5,95	-0,23
26	6,70	6,62	0,08
27	7,10	6,63	0,47
28	6,57	6,63	-0,06
29	6,64	6,61	0,03
30	7,22	6,73	0,49
31	7,16	6,68	0,47
32	7,10	6,72	0,38
33	7,40	7,14	0,26
34	5,80	5,80	-0,01
35	6,00	5,59	0,41
36	5,77	5,47	0,30
37	5,85	5,92	-0,07
38	5,10	4,57	0,53
39	4,46	4,75	-0,29
40	4,52	4,70	-0,18
41	4,68	4,93	-0,25
42	4,70	4,98	-0,28
43	5,05	5,42	-0,37
44	4,52	4,68	-0,16
45	4,80	4,91	-0,11
46	4,98	4,98	0,00
47	4,10	4,57	-0,47
48	5,70	4,86	0,84
49	5,55	4,60	0,95
50	5,22	5,09	0,13
51	4,05	4,26	-0,22
52	4,10	3,62	0,48
53	4,05	4,65	-0,61
54	5,10	5,54	-0,44
55	5,30	5,47	-0,17
56	5,00	5,11	-0,11
57	5,00	5,50	-0,50
58	5,30	5,73	-0,43
59	5,30	5,41	-0,11
60	5,30	5,50	-0,20
61	5,30	5,51	-0,21

^a (Du *et al.* 2002), (Chiyanzu *et al.* 2003), (Greenbaum *et al.* 2004);

A partir dos valores apresentados na tabela III.4.4.4. foram construídos gráficos, respectivamente, da regressão linear entre os valores preditos e observados e dos respectivos resíduos em função dos valores preditos (figura III.4.4.2).



Z

$$\text{pIC}_{50} = -1,287 (\pm 0,28) \text{ C-035} - 2,335 (\pm 0,75) \text{ BEHm3} - 18,66 (\pm 6,67) \text{ MATS8m} \\ + 2,693 (\pm 0,63) \text{ MATS8v} + 1,193 (\pm 0,48) \text{ nROR} + 32,98 (\pm 7,85)$$

(n = 61; r = 0,910; s = 0,433; F = 53,030; Q_{cv}² = 0,792; s-PRESS = 0,477) Equação III.4.4.1

Figura III.4.4.2. 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.4.4.1.

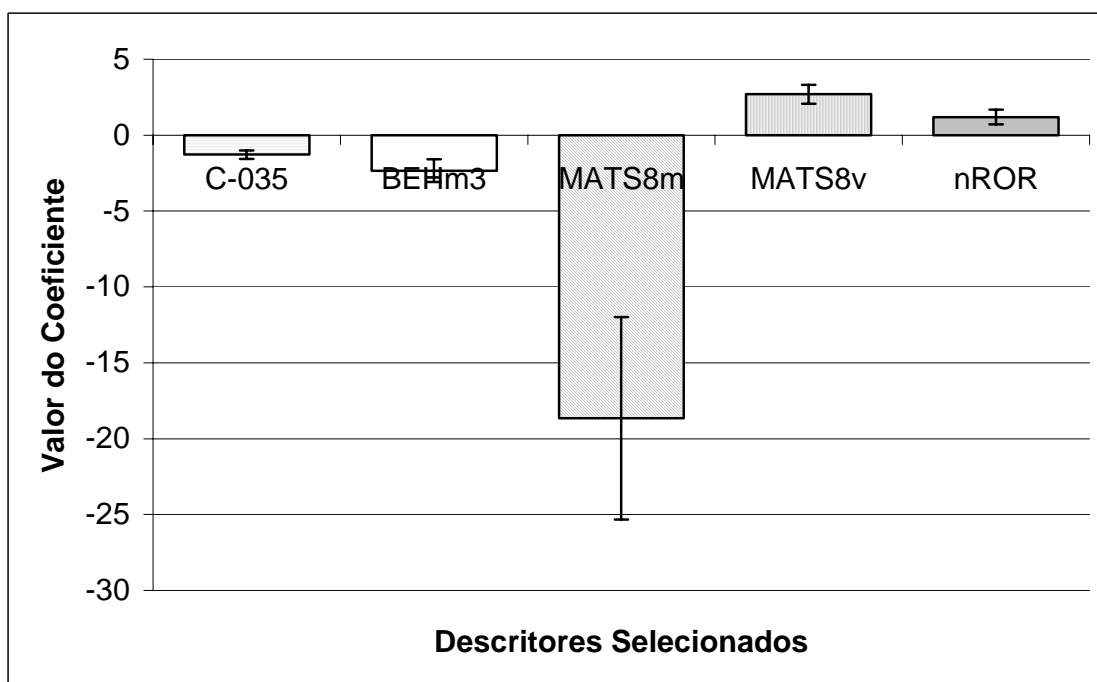


Figura III.4.4.3. Representação gráfica dos valores dos coeficientes dos descritores presentes no modelo expresso pela equação III.4.4.1.