

Influência da composição dos conjuntos de treinamento e teste na qualidade de modelos HQSAR para inibidores de DPP-4

Simone Q. Pantaleão (PG)^{1*}, Vinícius G. Maltarollo (PG)¹, Jadson C. Gertrudes (PG)², Káthia M. Honório (PQ)^{1,2}

¹Centro de Ciências Naturais e Humanas – UFABC, ²Escola de Artes, Ciências e Humanidades – USP

*simone.queiroz@ufabc.edu.br

Palavras Chave: Diabetes, HQSAR, inibidores de DPP-4.

Introdução

A enzima DPP-4 é considerada um dos alvos biológicos estratégicos para o tratamento do diabetes mellitus tipo 2. As principais vantagens do uso dos inibidores desta enzima são: ausência de hipoglicemias, preservação da massa e função das células β , efeito favorável sobre o peso corporal, efeitos colaterais mínimos e farmacodinâmica adequada.¹ Uma promissora alternativa para o desenvolvimento de novos medicamentos para esse alvo está relacionada com o emprego de técnicas bidimensionais para o estudo das relações quantitativas entre estrutura química e atividade biológica (QSAR 2D, em particular a técnica do holograma QSAR - HQSAR). Assim, o objetivo deste trabalho consiste em avaliar a influência da seleção de conjuntos de treinamento e teste empregados para a construção de modelos HQSAR preditivos.

Resultados e Discussão

Para os 60 compostos com atividade biológica conhecida²⁻⁵, foram construídos 20 modelos de QSAR 2D empregando o pacote computacional Sybyl 8.1. Tais modelos utilizaram 80% dos compostos (48 estruturas) no conjunto de treinamento e 20% dos compostos (12 moléculas) no conjunto-teste. A divisão dos compostos respeitou as seguintes faixas de atividade biológica: faixa 1 ($pIC_{50} \geq 5$ e ≤ 6), faixa 2 ($pIC_{50} > 6$ e ≤ 7), faixa 3 ($pIC_{50} > 7$ e ≤ 8) e faixa 4 ($pIC_{50} > 8$ e ≤ 9); para isso foi utilizado o programa MATLAB, sendo que dentro de cada faixa os compostos foram selecionados de forma aleatória. Após as 20 divisões em treinamento e teste, foi aplicada a técnica de HQSAR variando a distinção dos fragmentos (átomos-A, ligações-B, conexões-C, átomos de hidrogênio-H, quiralidade-Ch e grupos doadores/aceitadores de ligação de hidrogênio-DA), empregando o tamanho padrão de fragmento (4-7 átomos). A partir do modelo com maior valor de q^2 (etapa realizada para os 20 modelos), variou-se também o tamanho do fragmento. O resultado da comparação média dos 20 modelos gerados é apresentado na Tabela 1. É possível verificar a partir de alguns parâmetros estatísticos (q^2 , r^2 , r^2_{teste} e resíduos) que o modelo gerado a partir da separação 2 mostrou-se mais adequado estatisticamente. A seguir, foi realizada a variação do tamanho dos fragmentos para o modelo mais robusto (separação 2, A/B/H/Ch/DA) e o melhor resultado foi obtido com fragmento contendo

3-6 átomos ($q^2 = 0,841$, $r^2 = 0,942$). A correlação dos valores de pIC_{50} experimental e predito (Figura 1) indicou uma boa consistência interna e significativo poder preditivo. A Figura 2 ilustra os mapas de contribuição dos compostos mais e menos ativos.

Tabela 1. Valores médios obtidos para as 20 separações dos conjuntos de treinamento e teste

Separação	q^2	SEP	N	r^2	SEE	r^2_{teste}	Resíduo >1
Total	0,821	0,451	6	0,918	0,305	---	---
1	0,803	0,473	6	0,902	0,335	0,963	não
2	0,841	0,449	6	0,942	0,271	0,834	não
3	0,855	0,401	5	0,939	0,260	0,694	sim
4	0,880	0,375	6	0,956	0,228	0,627	sim
5	0,877	0,380	6	0,950	0,242	0,603	sim
6	0,820	0,462	6	0,922	0,304	0,882	sim
7	0,820	0,467	6	0,944	0,261	0,731	sim
8	0,822	0,462	5	0,929	0,292	0,821	sim
9	0,838	0,456	6	0,952	0,248	0,741	não
10	0,821	0,462	6	0,938	0,271	0,856	não
11	0,834	0,451	6	0,935	0,283	0,848	sim
12	0,882	0,389	6	0,956	0,238	0,423	sim
13	0,841	0,431	4	0,91	0,325	0,578	não
14	0,818	0,474	6	0,927	0,300	0,636	sim
15	0,834	0,430	3	0,883	0,361	0,629	sim
16	0,788	0,486	6	0,920	0,298	0,910	não
17	0,796	0,484	6	0,931	0,282	0,799	não
18	0,940	0,254	6	0,940	0,254	0,809	sim
19	0,911	0,329	5	0,911	0,329	0,937	sim
20	0,843	0,450	6	0,940	0,278	0,678	sim
MÉDIA*	0,795	0,485	6	0,910	0,317	0,750	---

SEP: erro padrão de predição; N: número de componentes principais do modelo PLS; SEE: erro padrão de validação não-cruzada; Resíduo >1: valor de resíduo de predição; *média de todos os modelos gerados 20 separações.

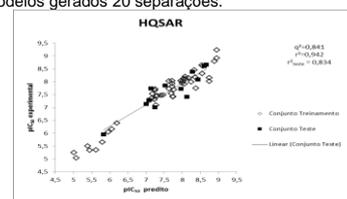


Figura 1. Valores de pIC_{50} predito e experimental (separação 2).

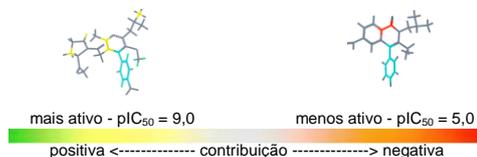


Figura 2. Mapas de contribuição HQSAR (compostos mais e menos ativos da série).

Conclusões

Neste trabalho, foi obtido um modelo de HQSAR confiável empregando diferentes conjuntos de treinamento e teste e que poderá ser utilizado para prever a atividade biológica de novos compostos.

Agradecimentos

UFABC, CAPES, CNPq e FAPESP.

- Robbins, S. L.; Cotran, R. S.; Abbas, A. K.; Kumar, V.; Fausto, N. *Patologia: Bases patológicas das doenças*. p. 610-117, 1999.
² Banno Y. *et al. Bioorg. Med. Chem.* 2011, 4953-4970.
³ Maezaki, H. *et al. Bioorg. Med. Chem.* 2011, 4482-4498.
⁴ Miyamoto, Y. *et al. Bioorg. Med. Chem.* 2011, 172-185.
⁵ Miyamoto *et al. Bioorg. Med. Chem.* 2011, 831-850.