Redes neurais não supervisionadas para classificação de séries de compostos extraídos de plantas com atividade anti-câncer.

Marcus Tullius Scotti¹ (PG)*, Hamilton Mitsugu Ishiki² (PQ), Karina de Oliveira Barbosa¹ (PG), Vicente de Paulo Emerenciano¹ (PQ)

¹ Instituto de Química – Universidade de São Paulo, ² Universidade do Oeste Paulista - *mscott@iq.usp.br

Palavras Chave: *redes neurais, descritores moleculares, programa DRAGON*

Introdução

As redes neurais artificiais alcançaram seu apogeu ao redor do mundo principalmente na última década, sendo aplicadas em diversas áreas do conhecimento humano tais como a medicina, a química e principalmente em problemas associados ao reconhecimento de padrões e não estão restritas a correções lineares ou sub-espacos lineares.

Uma variedade de 102 compostos extraídos de plantas medicinais e apresentando atividade contra linha de células humanas KB, foram estudados num recente trabalho¹. No presente trabalho, foram verificadas se diversas séries de descritores moleculares holísticos do programa DRAGON 5.0, podem ser utilizadas para classificar estes compostos em ativos e inativos utilizando mapas auto-organizáveis². Ainda mais, verificar a significância destes ao comparar com os resultados obtidos com variáveis geradas de forma aleatória.

Resultados e Discussão

As séries de treino e de teste foram divididas de forma que abrangessem toda a faixa de valores de atividade biológica (ED_{50}), e que representassem aproximadamente 80% e 20%, respectivamente, do número total de compostos estudados. Os compostos foram divididos em dois grupos: Ativos (compostos com valores de $ED_{50} \geq 0$), e Inativos (compostos com valores de $ED_{50} < 0$).

O treinamento não supervisionado foi realizado utilizando o programa SOM Toolbox para Matlab 6.5³.

Tabela 1. Resultados em porcentagem de acerto na classificação dos compostos ativos, inativos e no total dos mapas Kohonen, para as séries de treino e teste para os diferentes grupos de descritores (DRAGON 5.0, e gerados de forma aleatória).

Série de Descritor	Treino			Teste		
	Ativo	Inativo	Total	Ativo	Inativo	Total
2D ^a	77	86	81	50	73	62
Gateway	82	81	81	70	73	71
RDF	74	88	81	40	73	57
WHIM	74	88	81	60	73	67
3DMoRSE	74	90	83	50	45	48
Aleatório	74	76	75	30	36	33

^a Autocorrelação 2D.

Todos os grupos de descritores gerados pelo programa DRAGON apresentaram altos valores de índice de acerto de treinamento, mas variaram na série de teste.

A série de descritores que apresentou valores mais significativos foi os descritores GATEWAY, pois além de altos valores de índices de acertos globais de treinamento e de teste, apresentou valores próximos de índice de acerto entre os grupos ativos e inativos.

Todos os descritores apresentaram valores significativos se comparados aos resultados da média das 5 séries das variáveis geradas de forma aleatória. As variáveis aleatórias apresentaram um alto valor de índice de acerto na série de treinamento, porém inferior ao pior resultado obtido por descritores gerados através das estruturas moleculares. O valor do índice de acerto obtido utilizando as variáveis aleatórias para a série de teste também é inferior ao pior resultado obtido utilizando os descritores gerados através das estruturas moleculares.

Conclusões

Os mapa auto-organizáveis conseguiram extrair informação das séries de descritores gerados pelo programa DRAGON e utilizá-los de maneira prática, no caso, para a classificação de compostos com atividade anti-câncer frente a linhas de células humanas KB. Esta metodologia poderá ser utilizada como filtro em banco de dados de moléculas, ou para um posterior estudo dos descritores utilizando redes neurais supervisionadas para se compreender as contribuições físico-químicas importantes para a atividade anti-câncer frente as células humanas KB.

Agradecimentos

CAPES, FAPESP, CNPq, UNOESTE

¹ Xue, C. X.; Zhang, X. Y.; Liu, M. C.; Hu, Z. D.; Fan, B. T. *Pharm. and Biom Anal.* **2005**, 38(3), 497.

² Kohonen, T. *Neurocomputing*. **1998**, 21, 1.

³ <http://www.cis.hut.fi/projects/somtoolbox>. Acessado em 10/08/2005.