

Predição de tempos de retenção de produtos naturais em cromatografia líquida de alta eficiência envolvendo estudos de QSRR

Fernando B. Da Costa¹ (PQ)*, Dimitar Hristozov² (PG), Ricardo Stefani³ (PG), Johann Gasteiger² (PQ)

¹Universidade de São Paulo (USP), Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Laboratório de Farmacognosia, Ribeirão Preto, SP; ²Universidade Erlangen-Nuremberg, Computer-Chemie-Centrum, Erlangen, Alemanha; ³USP, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Departamento de Química, Ribeirão Preto, SP - * febcosta@fcrp.usp.br

Palavras Chave: descritores estruturais, quantitative structure-retention relationships, redes neurais artificiais, cromatografia líquida, produtos naturais

Introdução

Em programas de bioprospecção, é de grande interesse a rápida caracterização de produtos naturais em materiais biológicos. A cromatografia líquida acoplada a métodos espectrométricos é uma metodologia bem estabelecida para a identificação de analitos em misturas complexas. O estudo desta metodologia, aliada a bancos de dados de estruturas químicas conhecidas com diferentes propriedades, têm recebido destaque. Entretanto, caso uma substância desconhecida esteja presente na mistura complexa, porém ausente no banco de dados e na literatura, propostas plausíveis para sua estrutura química é um dilema. Uma saída para se tentar resolver este tipo de problema é a predição de tempos de retenção (t_R) aliada a bancos de dados, objetivo principal do presente trabalho.

O processo de relacionar t_R a estruturas moleculares de analitos é denominado *Quantitative Structure-Retention Relationship* (QSRR)¹.

Resultados e Discussão

A motivação para este trabalho foi testar hipóteses para auxiliar a elucidação estrutural de analitos desconhecidos em misturas. Um banco de dados composto de 51 estruturas de lactonas sesquiterpênicas, uma classe especial de substâncias naturais presentes em plantas medicinais, foi empregado para construir diferentes modelos para a predição de seus t_R em cromatografia líquida. As condições experimentais para se obter os t_R envolveu o uso de coluna C-18, eluição isocrática e dois eluentes, MeOH-H₂O 1:1 e MeCN-H₂O 3:7. Os t_R são relativos a uma substância de referência, o 2,3-dimetilfenol.

Para os estudos de QSRR, as coordenadas das estruturas em 3D foram obtidas e codificadas com vários descritores estruturais. Tais descritores são relacionados ao mecanismo de retenção em fase reversa, como momento dipolo, logaritmo do coeficiente de partição octanol-H₂O (logP),

polarizabilidade, ligações de hidrogênio e descritores 3D, dentre outros.

Para se estimar os t_R , nos dois eluentes, foram elaborados modelos empregando-se regressão linear múltipla (MLR) e redes neurais artificiais (ANN). As estruturas químicas codificadas compuseram a entrada de dados nos modelos e o t_R foi a resposta. Em um primeiro estudo, os modelos foram feitos com todas as 51 estruturas e a validação foi através de *10-fold cross-validation*. No segundo estudo, partindo-se das 51 estruturas, 22 foram selecionadas para treino dos modelos utilizando-se SOMs (*Self Organizing Maps*), restando 29 para teste (validação externa). Um terceiro estudo utilizou 18 estruturas para treino e 33 para teste, sendo a seleção realizada a partir dos seis subtipos dos esqueletos carbônicos das 51 estruturas. Os modelos finais, nos três estudos acima citados, um para cada eluente e cada um empregando MLR e ANN, continham os seguintes descritores: polarizabilidade, xlogP, número de doadores de H e número de aceptores de H. Os 12 melhores modelos foram selecionados com base na melhor capacidade de predição de t_R .

A performance dos 12 modelos foi avaliada através do erro da média da raiz quadrada (RMSE), erro absoluto médio (MAE) e coeficiente de regressão (R), que ficaram entre 0,75 e 0,88.

Conclusões

Conclui-se neste estudo de QSRR, com base nos R, que a performance dos modelos empregando ANN foi ligeiramente melhor do que aqueles empregando MLR para prever os t_R . Modelos para o eluente MeCN-H₂O apresentaram os maiores R, porém aqueles com MeOH-H₂O apresentaram os menores MAE. A divisão das 51 estruturas entre treino/teste apresentou os melhores resultados. Nos modelos com MLR, o logP contribuiu mais para a fase móvel MeOH-H₂O, enquanto que a polarizabilidade foi mais importante para a fase com MeCN-H₂O.

Agradecimentos

Alexander von Humboldt Stiftung, CAPES, FAPESP.

