

QSAR: Ferramenta *Web* para análise de descritores utilizados no cálculo de modelos QSAR

Patrick B; de S; SILVA¹; Ricardo M; da COSTA²

RESUMO

O estudo e análise das relações quantitativas estrutura-atividade (QSAR) de conjuntos de compostos diversos é de suma importância para a descoberta de novos fármacos e para o desenvolvimento de pesquisas que buscam o tratamento e até a cura de doenças. Neste trabalho, foi desenvolvido um sistema Web capaz de selecionar as melhores variáveis (descritores) para construção de modelos QSAR robustos e estatisticamente representativos. Foi implementado técnicas de agrupamento de variáveis e cálculo de correlação entre variáveis. A linguagem para o desenvolvimento da ferramenta foi o *Hypertext Preprocessor* (PHP) integrado com o software estatístico R. Os resultados mostram que foi possível implementar uma ferramenta Web capaz de realizar uma análise completa em grandes planilhas de descritores e selecionar aqueles com maior correlação com a variável dependente. Além disso, por meio da ferramenta foi possível identificar e visualizar graficamente os descritores degenerados e por meio de análise de clusters os agrupados diretamente com a variável dependente.

Palavras-chave: Descritores; Software R; modelos QSAR, ferramentas de seleção de descritores.

1. INTRODUÇÃO

A descoberta de novas substâncias e até a criação de novos fármacos são de extrema importância para a ciência. Mas para tais realizações antes dos experimentos e testes em laboratórios, são necessários estudos para a descoberta de possíveis candidatos e, a partir daí, começarem as elucubrações.

O estabelecimento de relações quantitativas entre a estrutura química e a atividade biológica é uma área de destaque hoje na comunidade científica. A predição da atividade biológica de novos compostos usando relações matemáticas baseadas em propriedades estruturais é um campo de pesquisa promissor.

O *Quantitative Structure Activity Relationship* (QSAR) ou Estudo das Relações Quantitativas Estrutura-Atividade, visa encontrar propriedades físico-químicas que tem correlação (positiva ou negativa) na atividade de um determinado composto (molécula) e dessa forma selecionar aquelas que com maior correlação para se gerar um modelo matemático que explique o

¹ Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais – Campus Muzambinho. Muzambinho/MG Email: patrickbastos00@gmail.com.

² Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais – Campus Muzambinho. Muzambinho/MG Email: ricardo.costa@muz.ifsuldeminas.edu.br.

comportamento biológico de um conjunto de compostos a partir de poucas variáveis selecionadas.

Os métodos QSAR transformam a estrutura química de um composto em uma série de descritores numéricos que representam as características mais importantes de uma atividade biológica e uma estrutura química. A partir daí, podemos estabelecer relações quantitativas entre os descritores de estrutura-atividade. Segundo Martins et al. (2013), relações QSAR são úteis para o desenvolvimento de novos compostos com propriedades biológicas desejáveis.

Segundo Golbraikh *et al.* (2003), o processo de desenvolvimento de modelos QSAR pode ser dividido em três etapas, sendo elas: Preparação dos dados, análise dos dados e validação do modelo. Sendo assim, para a análise de um modelo QSAR são necessárias duas entradas de dados, uma matriz com os valores numéricos dos descritores e o vetor contendo as atividades biológicas dos compostos sobre investigação.

Levando em conta as características que um sistema de modelagem QSAR deve ter, este trabalho teve por objetivo a construção de uma ferramenta *Web* integrada com o software estatístico R³ capaz de selecionar as melhores variáveis (descritores) para construção de modelos QSAR robustos e estatisticamente representativos.

2. MATERIAL E MÉTODOS

A aplicação foi desenvolvida com interface *Web* para melhor facilidade de acesso do usuário com a ferramenta, podendo ser acessada de qualquer dispositivo com acesso à internet e por qualquer navegador. O modelo de processo de software escolhido foi a iterativo e incremental, pois há uma maior interação entre os envolvidos no trabalho.

Inicialmente foi feita a implementação dos *scripts* em R com funções necessárias para análise de clusters dos descritores. Foram desenvolvidos dois *scripts*, um para cálculo da correlação entre as variáveis e outro para o agrupamento das variáveis (análise de *cluster*).

Seguindo a fase de implementação, foi desenvolvido o sistema web utilizando *HyperText Markup Language* (HTML) para o desenvolvimento de componentes de interface e *front-end*. Para o desenvolvimento *back-end* foi utilizado a linguagem *Hypertext Preprocessor* (PHP) que possui ferramentas para receber e apresentar os dados tanto o usuário, como também para prepará-los para os *scripts* em R.

³ Software Estatístico R: linguagem para computação estatística e visualização de dados gratuita.

Na próxima etapa foi realizada a integração do sistema Web com os *scripts* desenvolvidos em R. O sistema recebe os arquivos do usuário, prepara-os para o R e assim que solicitado pelo próprio utilizador do sistema, chama e executa os *scripts*.

Para concluir, apesar de constantes testes durante o desenvolvimento da ferramenta, será feito também ao final uma rotina de testes mais completa, para verificar a robustez da ferramenta e consistência dos resultados apresentados.

3. RESULTADOS E DISCUSSÕES

A ferramenta desenvolvida é capaz de receber dois arquivos x e y , respectivamente os descritores moleculares e a atividade biológica para a realização dos cálculos. A partir de submetido os arquivos pelo usuário, o sistema permite que o mesmo possa gerar a correlação de cada descritor molecular com a atividade biológica estudada, e também gerar o gráfico de agrupamento das mesmas.

A Figura 1 mostra a página inicial da ferramenta Web, onde o usuário poderá enviar os arquivos X e Y , que são a atividade biológica e os descritores respectivamente. Ainda nesta página o usuário poderá selecionar o separador utilizado para separar seus dados no arquivo e poderá também definir se os dados serão ou não normalizados.

Figura 1. Página inicial da ferramenta Web
Fonte: Do Autor (2018)

Para apresentar os resultados da ferramentas foram utilizadas dois tipos de variáveis como exemplos, uma que tem alta correlação com a atividade biológica e outra variável degenerada que não tem uma alta correlação com a atividade biológica.

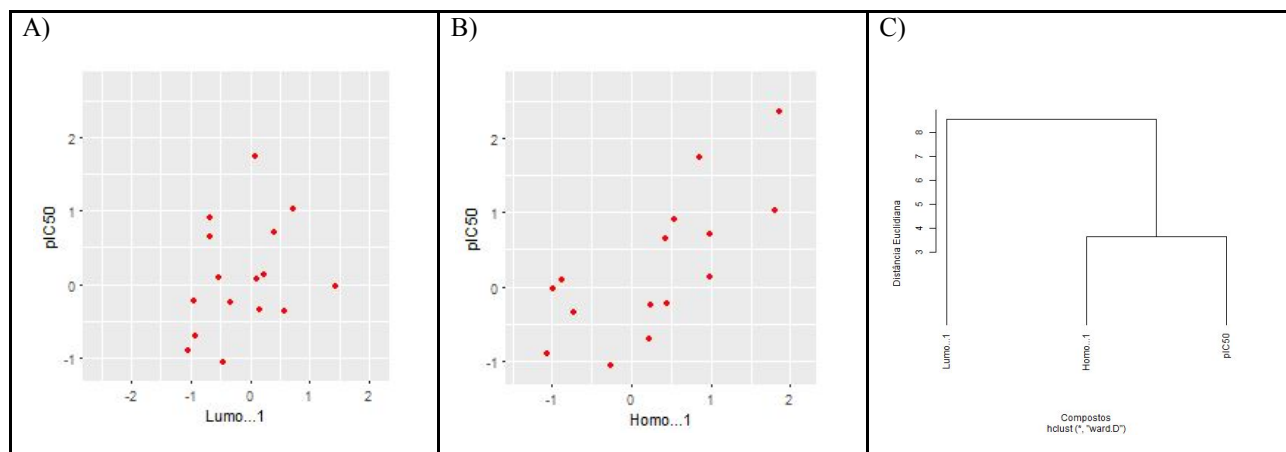


Figura 2: A) Gráfico de exibição de descritor degenerado calculado pela ferramenta desenvolvida B) Gráfico de exibição de descritor com alta correlação C) Dendograma gerado entre a atividade biológica e as demais variáveis estudadas.

A Figura 2 mostra respectivamente as correlações das variáveis Lumo 1, Homo 1 , além do dendograma que calcula a distância entre as variáveis e a atividade biológica, aqui representada pela variável pIC50 .

4. CONCLUSÕES

Com o desenvolvimento deste protótipo, a próxima etapa é a de testes e de comparação dos resultados obtidos pela ferramenta com outros softwares, além da implementação de novas funções, tornando-a mais completa e robusta. Do mais, espera-se que a ferramenta consiga atender todas as necessidades que foram propostas.

REFERÊNCIAS

GOLBRAIKH, Alexander et al. **Rational selection of training and test sets for the development of validated QSAR models.** Journal of computer-aided molecular design, v. 17, n. 2-4, p. 241-253, 2003.

MARTINS, João Paulo A. et al. **QSAR modeling: um novo pacote computacional open source para gerar e validar modelos QSAR.** Química Nova, 2013.