

## V. Conclusão

Os critérios de pré-tratamento de dados, selecionados na literatura ou propostos neste trabalho, utilizam abordagens metodológicas que consideram relevante tanto evitar ao máximo a perda de informações sobre o sistema como garantir a diversidade estrutural do dado original.

O processo de pré-tratamento de dados diminuiu significativamente o número de descritores irrelevantes para a obtenção de modelos *QSAR* estatisticamente significativos.

Em todas as séries observou-se uma diminuição do número de variáveis acima de 60%. Este fato tornou exequível, dentro das nossas condições, o tempo computacional necessário para a obtenção dos modelos *PLS* e *QSAR*.

Para a série II, avaliou-se o efeito da retirada de variáveis que apresentavam valores acima de 4 desvios padrão em relação à média, ou seja, um descritor que apresenta *outliers*, sobre a quantidade e a natureza dos descritores incluídos no modelo de *QSAR*.

Verificou-se que um dos descritores  $4\sigma$ , C-033, mostrou uma contribuição estatisticamente significativa para o modelo *QSAR* para a série II, apesar de ter contribuído significativamente para o índice de degeneração multivariada, D.

Observaram-se, para todos os modelos *PLS* selecionados, altos valores de coeficiente de predição interna ( $Q_{cv}^2 > 0,7$ ). Adicionalmente, verificaram-se, para a grande maioria dos modelos (75 %), altos valores de coeficiente de correlação externa ( $r^2 > 0,6$ ), indicando uma baixa probabilidade dos modelos *PLS* terem sido obtidos “ao acaso”.

Embora a interpretação dos descritores constitucionais e de fragmentos tenha sido extremamente simples, estes são degenerados, contribuindo significativamente para os valores dos índices de degeneração multivariada. Adicionalmente, no entanto, constatou-se que estes descritores diferenciam os compostos mais ativos dos menos ativos..

Observou-se que as classes de descritores GETAWAY e RDF, que utilizam as estruturas moleculares representadas em 3D, geraram 6 descritores (H1p, R6u, R7e, RDF065v, RDF125u, RDF020p) que mostraram contribuições estatisticamente significativas para as os modelos *QSAR* gerados.

Observou-se que a classe de descritores de autocorrelação 2D, que utiliza as estruturas moleculares representadas em 2D, gerou 3 descritores (Gats6e, Mats8m, Mats8v)

que mostraram contribuições estatisticamente significativas para os modelos *QSAR* gerados.

De modo geral, a interpretação dos descritores selecionados, dos 1497 gerados pelo programa DRAGON (v.3.0), mostrou-se complexa. Estes foram obtidos, porém, de maneira extremamente rápida e, geraram modelos *QSAR* válidos. Adicionalmente, pode-se afirmar que a abordagem metodológica utilizada neste trabalho, incluindo-se o programa DRAGON, apresenta características, que podem ser ampliadas para outras aplicações, como por exemplo, para gerar filtros para procura em bancos de dados ou estruturas.

Os critérios de seleção de descritores proposto permitiram selecionar modelos de *QSAR* clássico estatisticamente significativos, a partir de muitas variáveis. Estes apresentaram altos valores de coeficiente correlação ( $r > 0,9$ ) e de predição interna ( $Q_{cv}^2 > 0,7$ ), para as três séries de compostos estudados neste trabalho com atividade inibitória frente à ribonucleotídeo redutase (IRNR) e à cruzafina.

Os modelos *QSAR* gerados para as séries I, II e III mostraram contribuições negativas para alguns dos descritores. Estas informações são, no entanto, relevantes para o planejamento de compostos com as atividades inibitórias estudadas, considerando-se que o conhecimento das características estruturais dos compostos que não devem estar presentes também é extremamente relevante.