Applications of Artificial Neural Networks and Fuzzy Models in High Throughput Screening: Classifying the activities of various compounds towards Cobalt

Mithun Hebbar, Stan R. Svojanovsky, Qi-Zhuang Ye, Gunda Georg, Peter G. Smith, Swapan Chakrabarti, Electrical Engineering and Computer Science Department/Information and Telecommunication Technology Center/ Department of Molecular and Integrative Physiology KU Medical Center/ Department of Medicinal Chemistry/ Higuchi Biosciences Center, University of Kansas, Lawrence, KS 66045. 2006 Annual Kansas City Area Life Sciences Research Day.

The experimental High Throughput Screening (HTS) is a part of very expensive process to identify potential target molecules from a large library of chemical compounds. It requires using automated robotic equipment together with purchasing a library of chemicals. Our goal is to find a less expensive solution, which might be parallel to the existing HTS method, via Quantitative Structure-Activity Relationship (QSAR) using Artificial Neural Networks (ANNs). Since ANNs often provide significant advantages over the conventional statistical methods in solving non-linear pattern classification problems, we propose several different models of neural networks that are relevant to apply HTS data of a Methionine Aminopeptidases (MetAPs) Inhibition study.
Initially, a chemical library of 44449 organic compounds is examined by HTS and 1347 (i.e. approximately 3% of total) molecules with >40% inhibition activities towards Cobalt are selected as active. Each compound is characterized by the feature vector of 16 associated chemical and physical properties and the corresponding measured inhibition activity. Random samples of the original library are then used to train and evaluate ANNs computational models by the MATLAB toolboxes.
In spite of the facts that neural networks are applied on highly disproportional pattern classification model with excessive overlapped features, we achieved a 65% success rate when the computational model is trained with 100 trials from each class and then tested with 1247 samples from both classes. A detailed investigation reveals the exceeding degree of data polymorphism for each class of compounds in the multidimensional feature space.
Our next step is to utilize the Principal Component Analysis (PCA) in order to reduce the number of elements in the feature vector and further enhance the separation between both, active and non-active classes. In addition, the inherent ability of fuzzy logic could be applied to account for the remaining overlap between the feature vectors. The combined effect of fuzzy logic and PCA might enable us to achieve even a higher success rate for the HTS problem.

* This research was supported by the NIH Grant Number P20RR016475 from the INBRE Program and the COBRE Grant Number P20RR015563