A GMDH type net for the early screening of candidate drugs: the case of alpha-aminoamides

ID - 159

Gini Giuseppina[1], Folgheraiter Michele[2], Caccia Carla[3], Salvati Patricia[3], Dondio Giulio[4], Forlani Roberto[4]
[1]DEI, Politecnico di Milano
[2]DEI, Politecnico di Milano
[3]Newron Pharmaceuticals, Bresso
[4]NiKem Research, Baranzate

Motivation

Quantitative structure-activity relationships (QSARs) are mathematical models approximating the often complex function f between molecular properties (the independent variables x) and biological activities (the y) of compounds. Such models should predict the biological activity of untested and sometimes yet unavailable compounds.

In this case study, we examine the benefits of an approach of building accurate QSARs using advanced machine learning algorithms, integrated with the experimental platform that can produce the biological data

Methods

In the aim at identifying new sodium (Na+) channels blockers for neurological disorders, alpha-aminoamide chemical surroundings were explored building a virtual collection made of 7980 compounds. Molecular chemical space properties were defined taking into account 87 molecular descriptors belonging to topological, spatial, structural, thermodynamic, information-content typologies. Multivariate analysis of defined virtual space lead to the optimal molecular subset to be synthesised (473) and subsequently tested at the molecular target. Experimental IC50s were used along with the molecular descriptors to develop a predictive model.

Chemical space maximum diversity was considered as selection criteria to identify the final training set of 100 compounds. GA/PLS coupled method was considered to reduce the initial set of 87 molecular descriptors into a subset of most representative ones (20 descriptors). The devised method for model building was GMDH method. GMDH incrementally built a multilayer network. It started combining the neurons of the input layer, so generating an internal layer. Then it pruned the neurons keeping only the best, and continued by adding new layers and pruning them until a final criterion was met. Optimal models were evaluated by an external criteria calculated on separate part of data sample which is not used for model creation. The GMDH does not require presetting the neural network structure and allows to comprehensively present a classification rule as a concise set of short-term polynomials. Our method is implemented in poliGMDH.

Results

We developed models in poliGMDH, with various settings as limiting the number of neurons for each layer. Interesting architectures of different complexity emerged. The statistics parameter considered to evaluate the regression function y = f (x1, x2, ..xn) is R2; it represents the fraction of variability in y that can be explained by the variability in x. On the training set we reached an R2 of 0.7, paired on the external test set by 0.65. The models are saved both in an internal format to be used again by poliGMDH or in xml to be exported to other applications.

Acknowledgments Authors gratefully acknowledge European Union 6th Framework Programme Commission for financial support.

Email: gini@elet.polimi.it