ABSTRACT

A new mathematical approach that works at the level of molecular topology is proposed for characterization of structure-activity relationship of biological active compounds. A family of molecular descriptors is generated for a set of biologic active compounds and a genetic algorithm is used for identification of the best performing multivariate regression model. A series of statistical approaches are considered for model assessment (Bolboacă and Jäntsch, 2008) [8]. In order to validate the new method the performances of the obtained model will be compared through a correlated correlation analysis with other QSAR models.

INTRODUCTION

Development of information and computer technologies induces changes into research concept, leading to the development of many in silico analytical and experimental methods [1,2] used in determination and prediction of drug metabolism [3]. These methods have some advantages, from which the most important are: allows determination of metabolic profile in early stages of drug design; experiments are done into a shorter time and with fewer expenses [2].

Mathematical approach on structure-activity relationship (SAR) for biological active compounds (begun in nineteen century) lead to the concept of quantitative structure-activity relationship (QSAR, a mathematical approach that allows the identification of the quantitative link between structure and biologic activity of investigated compounds – [4]). SAR studies have been published since 1908, when Crum-Brown & Fraser stipulated the idea that the compounds activity is a function of structure and chemical composition [5].

METHODOLOGY

A mathematical approach developed starting with the information obtained from the 2D and 3D structure of a chemical compounds leads to introduction of Molecular Descriptors Family on the Structure-Activity Relationship method [7]. A family of molecular descriptors is generated for a set of biologic active compounds and a genetic algorithm is used for identification of the best performing multivariate regression model (see Figure 1 for the formal description of the approach). A series of statistical approaches [8] are considered for model assessment:

- Simple correlation analysis
- Multiple correlation analysis
- Qualitative vs. quantitative analysis (correlation coefficients: Pearson; Spearman; Semi-quantitative; Kendall tau-a; Kendall tau-b; Kendall tau-c; Goodman-Kruskal; test of significance and associated p-value)
- Leave-one-out cross-validation analysis
- Correlated correlations analysis (Steiger’s test)
- Training vs. test experiment

A series of statistical approaches [8] are considered for model assessment. Examples of the approach are given below.

EXAMPLES: BIOMATHEMATICS IN MODELLING BIOLOGICAL ACTIVE COMPOUNDS

Amino Acids Modelling [9]

+ Analysis bullelins
+ Models assessment

Histidine activity on Carbonic Anhydrase (Substituted 1,3,4-Thiadiazole- and 1,3,4-Thiadiazoline-Disulfonamides)

\[ \text{CA IV} \quad n = 40 \quad \nu = 4 \quad r = 0.3993 \quad s_w = 0.1599 \quad F (p) = 101 (< 0.001) \quad r' = 0.3034 \quad r'_\text{loo} = 0.0168 \]

\[ \text{CA II} \quad r_v = -9.9859 + 4.5643 \cdot 10^{-3} \cdot \text{IC50} + 2.945 \cdot 10^{-5} \cdot \text{IC50}^2 + 1.4832 \cdot \text{IC50}^3 \quad n = 40 \quad \nu = 4 \quad r = 0.9037 \quad s_w = 0.1706 \quad F (p) = 82 (2.7 \cdot 10^{-4}) \quad r'_\text{loo} = 0.8804 \]

CONCLUSION

The proposed mathematical model proved to have abilities in prediction and estimation of property and activity of chemical compounds in terms of estimation as well prediction.

REFERENCES


The research was partly supported by UEFISCUS Romania through project ID_458/2007.