Scientific Research and Essay Vol. 4 (13), pp. 1559-1564, December, 2009
Available online at http://www.academicjournals.org/SRE
ISSN 1992-2248 © 2009 Academic Journals

Standard Review

Computer - aided linear modeling employing QSAR for drug discovery

E. C. Ibezim[1,2]*, P. R. Duchowicz[2], N. E. Ibezim[3], L. M. A. Mullen[2], I. V. Onyishi[4], S. A. Brown[5] and E. A. Castro[2]

[1]Department of Pharmaceutics, University of Nigeria, Nsukka, Nigeria.
[2]Instituto de Investigaciones Fisicoquimicas Teoreticas y Aplicadas (inifta), Diagonal 113 Entre 63 y 64, la plata, sucursal 4 - casilla de correo 16 - (1900) la plata, Argentina.
[3]Computer Education Subdepartment, Department of VTE, University of Nigeria, Nsukka, Nigeria.
[4]Department of pharmaceutical technology and Industrial Pharmacy, University of Nigeria, Nsukka, Nigeria.
[5]Department of Pharmaceutics and Pharmaceutical Technology, University of Port Harcourt, Rivers State, Nigeria.

Accepted 21 December, 2009

Quantitative structure-activity relationship (QSAR) is a computational process that relates the chemical structure of compounds with their activities, especially biologic activities or effects. It employs series of computer-based processes to analyze quantitative experimental data of the activities of given compounds with known chemical structures in order to predict a relationship, model or equation that will help to propose the activity of known compounds with unknown activities or unknown compounds and their activities. Commonly used computer softwares in QSAR analysis include HYPERCHEM, MATLAB, DRAGON and RECKON.

Key words: QSAR, biological activity, prediction, computer software.

INTRODUCTION

Quantitative structure-activity relationship (QSAR) is the process by which chemical structure is quantitatively correlated with a well defined process, such as biological activity (Wikipedia, 27/11/09). It is the calculation of quantitative structure-activity relationship values, used to predict the activity of compounds from their structures. In this context, there is also a strong relationship to Chemometrics. Chemical expert systems are also relevant, since they represent parts of chemical knowledge as an *in silico* representation.

For example, biological activity can be expressed quantitatively as in the concentration of a substance required to give a certain biological response. Addi-tionally, when physicochemical properties or structures are expressed by numbers, one can form a mathematical relationship, or quantitative structure-activity relationship, between the two. The mathematical expression can then be used to predict the biological response of other chemical structures.

QSAR's most general mathematical form is:

Activity = f (physiochemical properties and/ or structural properties)

Methods which can be used in QSAR include various regression and pattern recognition techniques.

*Corresponding author. E-mail: ecibezim@yahoo.com.

Abbreviations: SAR, Structure-activity relationship; **QSAR,** quantitative structure-activity relationship; **3D-QSAR,** three dimensional-QSAR; **POR,** partial order ranking; **PM3,** parametric method 3; **NNRTI,** non-nucleoside reverse transcriptase inhibitors; **DFT,** hybrid density functional theory; **SMILES,** simplified molecular input line entry system; **GABA,** gamma-aminobutyric acid.

SAR AND THE SAR PARADOX

The basic assumption for all molecule based hypotheses is that similar molecules have similar activities. This principle is also called Structure-Activity Relationship (SAR). The underlying problem is therefore how to define a *small* difference on a molecular level, since each kind

of activity, e.g. reaction ability, biotransformation ability, solubility, target activity and so on, might depend on another difference. A good example was given in the bioisosterism review of Patani and LaVoie (1996).

In general, one is more interested in finding strong trends. Created hypotheses usually rely on a finite number of chemical data. Thus, the induction principle should be respected to avoid overfitted hypotheses and deriving overfitted and useless interpretations on structural/molecular data. The SAR paradox refers to the fact that it is not the case that all similar molecules have similar activities.

It has been shown that the partition coefficient, log P, of a compound can be determined by the sum of its fragments. Fragmentary log P values have been determined statistically. This method gives mixed results and is generally not trusted to have accuracy of more than ± 0.1 units (Wildman and Crippen, 1999).

3D-QSAR

3D-QSAR refers to the application of force field calculations requiring three-dimensional structures, e.g. based on protein crystallography or molecule superposition. It uses computed potentials, e.g. the Lennard-Jones potential, rather than experimental constants and is concerned with the overall molecule rather than a single substituent. It examines the steric fields (shape of the molecule) and the electrostatic fields based on the applied energy function (Leach, 2001). The created data space is then usually reduced by a following feature extraction. The following learning method can be any of the machine learning methods, e.g. support vector machines (Vert et al., 2004).

In the literature, it can be seen that chemists have a preference for partial least squares (PLS) methods, since they apply the feature extraction and induction in one step.

Data mining

For coding, usually a relatively large number of features or molecular descriptors are calculated, which can lack structural interpretation ability. In combination with the later applied learning method or as pre-processing step, occurs a feature selection problem. A typical data mining based prediction uses e.g. support vector machines, decision trees or neural networks for inducing a predictive learning model. Molecule mining approaches, a special case of structured data mining approaches, apply a similarity matrix based prediction or an automatic fragmentation scheme into molecular substructures. Furthermore, there also exist, approaches using maximum common subgraph searches or graph kernels (Gusfield, 1997; Helma, 2005).

Judging the quality of QSAR models

QSARs represent predictive models derived from application of statistical tools correlating biological activity (including desirable therapeutic effect and undesirable side effects) of chemicals (drugs/ toxicants/ environmental pollutants) with descriptors representative of molecular structure and/ or properties. QSARs are being applied in many disciplines, for example risk assessment, toxicity prediction and regulatory decisions (Tong et al., 2005) in addition to drug discovery and lead optimization (Dearden, 2003). Obtaining a good quality QSAR model depends on many factors, such as the quality of biological data, the choice of descriptors and statistical methods. Any QSAR modeling should ultimately lead to statistically robust models capable of making accurate and reliable predictions of biological activities of new compounds.

For validation of QSAR models, four strategies are usually adopted (Wold and Eriksson, 1995) namely: internal validation or cross-validation; validation by dividing the data set into training and test compounds; true external validation by application of model on external data and data randomization or Y-scrambling.

The success of any QSAR model depends on accuracy of the input data, selection of appropriate descriptors and statistical tools and most importantly, validation of the developed model. Validation is the process by which the reliability and relevance of a procedure are established for a specific purpose (Roy, 2007). Leave one-out cross-validation generally leads to an overestimation of predictive capacity and even with external validation; no one can be sure whether the selection of training and test sets was manipulated to maximize the predictive capacity of the model being published. Different aspects of validation of QSAR models that need attention include methods of selection of training set compounds (Leonard and Roy, 2006), setting training set size (Roy et al., 2008) and impact of variable selection (Roy and Roy, 2008) for training set models for determining the quality of prediction. Development of novel validation parameters for judging quality of QSAR models is also important (Roy et al., 2009).

SOME APPLICATIONS OF QSAR

Chemical applications

One of the first historical QSAR applications was to predict boiling points (Rouvray and Bonchev, 1991). It is well known for instance that within a particular family of chemical compounds, especially in organic chemistry, that there are strong correlations between structure and observed properties. A simple example is the relationship between the number of carbons in alkanes and their boiling points. There is a clear trend in the increase of

boiling point with an increase in the number carbons and this serves as a means for predicting the boiling points of higher alkanes.

A still very interesting application is the Hammett equation, Taft equation and pKa prediction methods (Fraczkiewicz, 2007).

Biological applications

The biological activity of molecules is usually measured in assays to establish the level of inhibition of particular signal transduction or metabolic pathways. Chemicals can also be biologically active by being toxic. Drug discovery often involves the use of QSAR to identify chemical structures that could have good inhibitory effects on specific targets and have low toxicity (non-specific activity). Of special interest is the prediction of partition coefficient log P, which is an important measure used in identifying "drug-likeness" according to Lipinski's Rule of Five.

While many quantitative structure activity relationship analyses involve the interactions of a family of molecules with an enzyme or receptor binding site, QSAR can also be used to study the interactions between the structural domains of proteins. Protein-protein interactions can be quantitatively analyzed for structural variations resulting from site-directed mutagenesis (Freyhult et al., 2003). It is part of the machine learning method to reduce the risk for a SAR paradox, especially taking into account that only a finite amount of data is available. In general, all QSAR problems can be divided into a coding (Timmerman et al., 2002) and learning (Strok et al., 2001).

Applicability domain

As the use of (Q) SAR models for chemical risk management increases steadily and is also used for regulatory purposes (in the EU: Registration, Evaluation, Authorisation and Restriction of Chemicals), it is of crucial importance to be able to assess the reliability of predictions. The chemical descriptor space spanned by a particular training set of chemicals is called Applicability Domain. It offers the opportunity to assess whether a compound can be reliably predicted.

PARTIAL ORDER RANKING IN QSAR

The fundamentals of QSAR are well known and they have been rationalized and systematized around ten years ago (Castro, 2006). The importance of this subject is very well documented by the apparition of several specialized journals, monographs, reviews and books and the development of numerous research works Registered in journals of general interest. The development of QSAR models are however, often based upon rather demanding statistical methods so that it has become necessary to look for more operationally simple alternatives. The Partial Order Ranking (POR) theory is one of such options and from a strict mathematical viewpoint, it seems to be extremely simple compared with the usual statistical methods. It ranks results in a very transparent and suitable way to perform comparisons among a set of objects (such as molecules) according to their attributes (molecular descriptors) values.The POR method is based on elementary procedures of Discrete mathematics which only resort to the “≥” as the basic mathematical operation. The relatively simple mathematical resources needed to apply the method have attracted the attention of a qualified group of researchers who have already gotten many relevant results within the realm of QSAR Theory.

PROCEDURES IN QSAR ANALYSIS

In a typical QSAR modeling process, a set of experimental data containing the numerical values of the activity to be modeled, from an experimental data set is divided into two sets - training and test sets, in such a way that the training set would be more than the test. Next, the structures of all the compounds in the data set are then drawn and pre-optimized with the molecular mechanics force field (MM+) procedure included in the HYPERCHEM 6.03 software package (Hypercube). After that, the resulting geometries are refined by means of a suitable semi empirical method e.g. PM3 (Parametric method 3) using the Polak-Ribiere algorithm and a suitable gradient norm limit. Suitable descriptors (that describe different relevant features of the compounds, through mathematical formula obtained from the chemical graph theory, information theory, quantum mechanics, etc) (Katritzky et al., 1995), are then proffered and calculated for the experimental data using the DRAGON (Milano Chemometrics and QSAR Researh Group) or any other suitable software, resulting in a pool usually containing D = 1497 numerical variables.

A suitable model or models is/are proposed for the experimental data, using MATLAB, REKON or any other ideal software, showing the most appropriate descriptor(s) for the experimental data or property under consideration. A suitable method, for instance, the Replacement Method (Carta et al., 2002, 2004, 2005) is then employed as variable subset selection approach. Replacement Method (RM) is an algorithm that generates multi-variable linear regression models by minimizing its standard deviation, S and whose results are quite close to the ones obtained with exact (combinatorial) search of molecular descriptors although requiring much less computational work.

The structures of the known and unknown compounds

whose activity is to be determined are now drawn using CHEMDRAW, optimized and fitted into the already established model(s) using the appropriate softwares. From this, the unknown activities can be predicted. Shown below is a typical QSAR model for the inhibitory concentration (IC_{50}) of a given antibiotic:

Log_{10} (IC_{50}) = 50.120 (±0.1) + 15.32 (± 0.2). Mor02e - 5.812 (± 1.0). MATS2m
N = 100, R = 0.9324, S = 0.1987, FIT = 20.003, $p < 10^{-5}$, $R_{loo} = 0.9130$, $S_{loo} = 0.1999$, range in Mpt: 4.00 - 6.50°C, Outliers (> 25):1.

Where Mor02e and MATS2m are molecular descriptors, N is the number of molecules in the training set, R is the correlation coefficient, S is the model's standard deviation, FIT is the Kubinyi function, p is the significance of the model, outliers, (> 25) denotes the number of molecules having a residual (res) that exceeds two standard deviations, and loo stands for the leave-one-outcross validation techniques.

RECENT APPLICATIONS OF QSAR IN DRUG DISCOVERY

Duchowicz et al. (2007) have provided QSAR models for the growth inhibition of ciliated protozoa, *Tetrahymena pyriformis* by 250 mechanistically diverse phenolic compounds. The simultaneous linear regression analysis on 1338 topological, geometrical and electronic molecular descriptors over 200 molecules led to a seven-parameter relationship with R = 0.851 and leave more out $R_{1\text{-}60\%\text{-}0}t$ = 0.730, while a model based on flexible type of descriptors improves as R = 0.900 and $R_{1\text{-}60\%\text{-}0}$ = 0.854. An external test set of 50 related derivatives demonstrates that both models behave predicatively with rms = 0.418 and rms = 0.346, respectively, comparing fairly well with previously reported Artificial Neural Networks with rms 0.352. Finally, they employed the best QSAR equation to estimate the aqueous toxicity for 74 non-yet measured structures.

An exploratory study to investigate the possible simple descriptors in order to predict relative activity of antiepileptic enaminones has been carried out by Garro-Marinez et al. (2007). In the study, a general structure, substituent and activity relationship of the following type was fitted to the available ED_{50} values of cyclic enaminone antiepileptic compounds: ED_{50} = f (structure, substituent) = f (d, σ). In this relationship, ´structure´ was quantified by (d), the distance measured between the carbonyl oxygen atom and the first atom of the aromatic ring. The ´substituent´ was quantified by the Hammett substituent constant (σ). With the aid of the above function of two independent variables, a new molecular structure was predicted by extrapolation that has shown about two orders of magnitude greater activity than the most active molecule in the original set with measured ED_{50} values.

Thomas and Castro have employed theoretical and computational techniques for the investigation into conformation of the antimalarial agent 1, 2, 3, 5 -tetroxane and some derivatives (Thomas and Castro, 2006). In this study, they attempted the conformational analysis of some compounds of 1, 2, 3, 5 -tetroxane and to evaluate three theoretically possible conformers for the compound. Among a lot of theoretical methods available, they adopted a simple theoretical method called self consistent field method because it was rather easy to perform and require moderate computational facilities and time. The results of the analysis were reliable as evident from early literature. Moreover, the softwares were available free of cost. In this study, they used a freely available software package called GAMESS created by Alex A. Granovsky. To view the geometry of the molecule, they made use of another free software ViewMol3D.exe. The input of the molecule is given in the Z-matrix format.

The anticonvulsant activity of abietic acid has been discovered through application of linear discriminant analysis (Talevi et al., 2006). Linear Discriminant Analysis was performed to derive Discriminant Functions based on 2D descriptors and capable of classifying anticonvulsant from non-anticonvulsant compounds. Through application in virtual screening of the Discriminant Function, which performed best in the validation steps, abietic acid was identified as a potential new anticonvulsant agent. The anticonvulsant activity of abietic acid at 30 and 100 mg/kg was confirmed in the Maximal Electroshock Test, both orally and intraperitoneal. Similarly, Talevi et al. (2006) have applied a new similarity-based algorithm based on a previously developed model in the classification of two sets of anticonvulsant and non-anticonvulsant drugs. Each set is composed of (a) anticonvulsant compounds that have shown moderate to high activity in the Maximal Electroshock Seizure (MES) test and (b) drugs with other biological activities or poor activity in the MES test. The results from the analysis of variance (ANOVA) indicate that the proposed algorithm is able to differentiate anticonvulsant drugs. The proposed model may therefore, be useful in the identification of new anticonvulsant agents through virtual screening of large virtual libraries of chemical structures.

Albesa et al. (2006) have undertaken a theoretical study of a family of new quinoxaline derivatives. In this study, Hybrid Density Functional Theory (DFT) calcu-lations were performed on a series of 21 new quinoxaline derivatives, which would likely exhibit important biological activities. Optimized geometries, harmonic vibrational frequencies and 1H chemicall shifts were reported and compared with available experimental data.

Duchowicz et al. (2006) have applied a QSAR treatment to model the potency plC_{90}[mM] of 154 non-nucleoside reverse transcriptase inhibitors (NNRTI) of the

wild type HIV-1 virus, considered as the second generation analogues of Efavirenz. In addition, 56 inhibitors of the K-103N viral mutant form were also investigated. A pool consisting of 1494 theoretical molecular descriptors provided mainly by the Dragon 5 software were explored resorting to different methods of variable selection: Forward Stepwise Regression, the Replacement Method and the Genetic Algorithm approach. The optimal models found included up to seven parameters with R = 0.7991, $R_{l\text{-}20\%\text{-}0}$ = 0.7233 for the case of wild type and R = 0.9261, $R_{l\text{-}5\%\text{-}0}$ = 0.8802 for the K-103N mutation.

Helguera et al. (2006) have predicted the carcinogenic potency (TD_{50}) of a set of 62 nitroso-compounds, applying the QSAR theory. A thousand of molecular descriptors obtained from the Dragon 2.1 software were used in order to model the toxicological property bioassay in female rat and considering water as route of administration. For building the regression model, three different methods of variable selection were used namely Forward Stepwise Regression Method, the Genetics Algorithms and an alternative of Elimination Method, the Replacement method. For the first tine, the Replacement Method was used for predicting the carcinogenic potency, with the achievement of the best results. The finest obtained model had seven variables and was able to explain the 84.3% of the experimental variance after removing 6 chemicals, which are considered as outliers.

The QSAR analysis of mutagenicity of 16 dental monomers has been carried out by means of optimal descriptors calculated with SMILES (Simplified Molecular Input Line Entry System) notation (Castro et al., 2007). Statistical characteristics of predictive equations are n = 11, r^2= 0.67, s = 0.59, F = 18 (training set); n = 5, r^2 = 0.87, s = 0.46, F = 20 (test set).

A virtual prediction of anticonvulsant activity in MES test of widely used pharmaceutical and food preservatives, methylparaben and propylparaben has been carried out (Talevi et al., 2009). A Discriminant Function based on topological descriptors was derived from a training set composed of anticonvulsants of clinical use or in clinical phase of development and compounds with other therapeutic uses. This model was internally and externally validated and applied in the virtual screening of chemical compounds from the Merck Index. Methyl paraben, a preservative widely used in the food, cosmetics and pharmaceutical industries was signalled as active by the Discriminant Function and tested in mice in the Maximal Electroshock (MES) test after ip administration. A discriminant function for the prediction of anticonvulsant activity in the MES test was generated and its ability to select new anticonvulsant agents was confirmed through biological tests. The discovery of the anticonvulsant activities in the MES test of methyl paraben and propylparaben might be useful in the development of new anticonvulsant medications especially considering the well known toxicological profile of current anticonvulsant drugs.

Duchowicz et al. (2008) have undertaken a QSAR modeling of the interaction of flavonoids with GABA (A) receptors. Experimentally assigned values to binding affinity constants of flavonoids ligands towards the benzodiazepine site of the GABA (A) receptor complex were compiled from several publications and enabled to perform a predictive analysis based on QSAR. The best linear model established on 78 molecular structures incorporated four molecular descriptors, selected from more than a thousand geometrical, topological, quantum-mechanical and electronic types of descriptors and calculated by Dragon software. A practical application of this QSAR equation was performed by estimating the binding affinities for some newly synthesized flavonoids displaying 2-, 7- substitutions in the benzopyrane backbone and that still have not experimentally measured potencies.

ACKNOWLEDGEMENTS

The first author, Professor E. C. Ibezim, is grateful to the Third World Academy of Sciences, Italy and CONICET, Argentina, for sponsorship to study this area of research.

REFERENCES

Albesa A, Duchowicz PR, Castaneo H, Castro EA, Diez RP (2006) A theoretical study of a family of new quinoxaline derivatives, J. Molec. Graph. Model. 25: 487 - 494.

Carta A, Corona P, Loriga M (2005). Curr. Med. Chem*,* 12: 2259.

Carta A, Paglietti G, Mattana A, Fiori PI, Mollicottin I, Sechi L, Zanetti S (2004). Eur. J. Med. Chem, 39:195.

Carta A, Paglietti G, Nikookar MER, Sanna P, Sechi L, Zanetti S (2002). J. Med. Chem. 37:355.

Castro EA, Toropova AP, Toropova AA, Maksudov SK (2007). SMILES based optimal descriptors: QSAR modeling mutagenicity. Phys. Chem: An Indian J. 1: 171 -175.

Castro, E. A. (2006) Guest Editorial on POR Approximation in QSAR/QSPR Theory, Intern. J. Mol. Sci*.,* 7: 1.

Dearden JC (2003). In silico prediction of drug toxicity. J. Computer-aided Molec. Des. 17 (2-4): 119-127.

Duchowicz PR, Llamosa MF, Caballero J, Castro EA, Fernandez FMM (2006). QSAR of non-nucleoside inhibitors of HIV-1 reverse transciptase. Bioorgan. Med. Chem. 14: 5876-5889.

Duchowicz PR, Mercader AG, Fernandez FM, Castro EA (2007). Prediction of aqueous toxicity for heterogenous phenol derivatives by QSAR. *Chemomet. Intellig. Lab. Sys*., 90 (2): 97 - 107.

Duchowicz PR, Vitale MG, Castro EA, Autino JC, Romanelli GP Bennardi, D. O. (2008) QSAR modeling of the interaction of flavonoids with GABA(A) receptor, Eu. J. Med. Chem. 43(8): 1593-1602.

Fraczkiewicz R (2007). In*: Silico* Prediction of Ionization. Comprehensive Medicinal Chemistry II. Amsterdam: Elsevier.

Freyhult EK, anderssonm K, Gustafsson MG (2003). Structural modeling extends QSAR analysis of antibody-lysozyme interactions to 3D-QSAR. Biophys. J. 84 (4): 2264-2272.

Garro-Martinez J, Andrade M, Estrada M, Castro EA, Muczi Z, Zamarbide G, Csizmadia I (2007). An exploratory study to investigate possible simple descriptors in order to predict relative activity of antiepileptic enaminones, J. Phys. Organ. Chem*.,* 21 (5): 409 - 418.

Gusfield D (1997). Algorithms on strings, trees and sequences: computer science and computational biology. Cambridge, UK: Cambridge University Press.

Helguera AM, Duchowicz PR, Cabrera Perez MAI, Castro EA, Dias Soeiro Cordeiro MN, Perez Gonzalez M (2006). Application of the

Replacement Methods as a novel selection strategy in QSAR 1 - Carcinogenic potential.Chemomim. Intellig. Lab. Sys., 81: 180-187.
Helma C (2005). Predictive toxicology. Washington, DC: Taylor & Francis.
Hyperchem 603 (Hypercube), http://www.hyper.com
Katritzky AR, Lobanov VS, Karelson M (1995). Chem. Soc. Rev. 24: 279.
Katritzky AR, Lobanov VS, Karelson M (1995). Chem. Soc. Rev. 24: 279.
Leach AR (2001). Molecular modelling: principles and applications. Englewood Cliffs, N.J: Prentice Hall.
Leonard JT, Roy K (2006). On selection of training and test sets for the development of predictive QSAR models, QSAR & Combinatorial Sci. 25 (3): 235-251.
Milano Chemometrics and QSAR Researh Group, http://michem.disat-unimib.it/chm.
Patani GA, LaVoie EJ (1996). Bioisosterism: A Rational Approach in Drug Design. Chem. Revs 96 (8): 3147-3176.
Rouvray DH, Bonchev D (1991). Chemical graph theory: introduction and fundamentals. Tunbridge Wells, Kent, England: Abacus Press.
Roy K (2007). On some aspects of validation of predictive quantitative structure- activity relationship models", Expert Opin. Drug Discov. 2(12): 1567-1577.
Roy PP, Leonard JT, Roy K (2008). Exploring the impact of size of training sets for the development of predictive QSAR models", Chemom. Intell. Lab. Syst., 90 (1): 31-42.
Roy PP, Paul S, Mitra I, Roy K (2009). On Two Novel Parameters for Validation of Predictive QSAR Models". Molec. 14 (5): 1660-1701.
Roy PP, Roy K (2008). On some aspects of variable selection for partial least squares regression models", QSAR & Combinatorial Sci. 27(3): 302-313.
Strok DG, Duda RO, Hart PW (2001). Pattern classification. Chichester: John Wiley & Sons.
Talevi A, Bellera CL, Castro EA, Bruno-Blanch LE (2009). A successful virtual screening application: Prediction of anticonvulsant activity in MES test of widely used pharmaceutical and food preservatives, methyl paraben and propyl paraben, J. Comp. Aided Mol. Des. (In Press).
Talevi A, Prieto JJ, Bruno-Blanch LE, Castro EA (2006). New similarity-based algorithm and its applications to classification of anticonvulsant compounds, J. Enzym. Inhib. Med. Lett., (In Press).
Talevi A, Sella Cravero M, Castro EA, Bruno-Blanch LE (2006). Discovery of anticonvulsant activity of abietic acid through application of linear discriminant analysis, Bioorgan. Med. Chem. Letts. (In Press).
Thomas RJ, Castro EA (2006). Theoretical and computational investigation into conformational analysis of the antimalarial agent 1,2,3,5-tetroxane and some derivatives, Acad. Revs. (In Press).
Timmerman H, Roberto T, Consonni V, Mannhold R, Kubinyi H (2002). Handbook of Molecular Descriptors. Weinheim: Wiley-VCH.
Tong W, Hong H, Xie Q, Shi L, Fang H, Perkins R (2005). Assessing QSAR Limitations - A Regulatory Perspective. Current Computer-Aided Drug Design 1 (2): 195-205.
Vert J, Schlkopf B, Schölkopf B, Tsuda K (2004). Kernel methods in computational biology. Cambridge, Mass: MIT Press.
Wikipedia, http:/en.wikipedia.org/wiki. Quantitative_structure-activity_relationship Accessed on 27/11/09
Wildman SA, Crippen GM (1999). Prediction of physicochemical parameters by atomic contributions, J. Chem. Inf. Comp. Sci. 39 (5): 868-873.
Wold S, Eriksson L (1995). Statistical validation of QSAR results. in Waterbeemd, Han van de. Chemometric methods in molecular design. Weinheim: VCH. pp. 309-318.