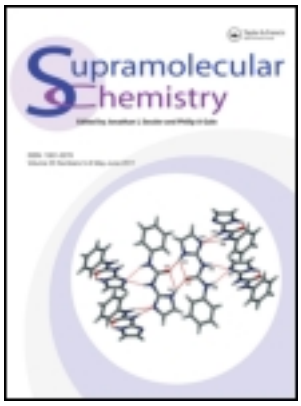


Supramolecular Chemistry
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gsch20

Review of the quantitative structure-activity relationship modelling methods on estimation of formation constants of macrocyclic compounds with different guest molecules

J. B. Ghasemi [a], M. Salahinejad [b] & M. K. Rofouei [b]

[a] Chemistry Department, Faculty of Sciences, K. N. Toosi University of Technology, Tehran, Iran

[b] Faculty of Chemistry, Tarbiat Moalem University, Tehran, Iran

Available online: 09 Jun 2011

To cite this article: J. B. Ghasemi, M. Salahinejad & M. K. Rofouei (2011): Review of the quantitative structure-activity relationship modelling methods on estimation of formation constants of macrocyclic compounds with different guest molecules, Supramolecular Chemistry, 23:9, 615-631

To link to this article: http://dx.doi.org/10.1080/10610278.2011.581281

Supramolecular Chemistry
Vol. 23, No. 9, September 2011, 615–631

Taylor & Francis
Taylor & Francis Group

Review of the quantitative structure–activity relationship modelling methods on estimation of formation constants of macrocyclic compounds with different guest molecules

J.B. Ghasemi[a]*, M. Salahinejad[b] and M.K. Rofouei[b]

[a]*Chemistry Department, Faculty of Sciences, K. N. Toosi University of Technology, Tehran, Iran;* [b]*Faculty of Chemistry, Tarbiat Moalem University, Tehran, Iran*

(*Received 11 December 2010; final version received 21 March 2011*)

There is an increasing interest in the use of quantitative structure–activity relationship (QSAR) approaches as a progressive tool in modelling and prediction of many physicochemical properties in host–guest interactions of macrocyclic complexes. A review is presented on the QSAR modelling of macrocyclic compounds formation constants, which focus on two most interesting macrocycles, e.g. crown ethers and cyclodextrins (CDs), with different guest molecules. The review starts with a short overview on experimental methods of stability constant measurement, followed by a short explanation of the QSAR methodologies. In the next section, we focus on and discuss QSAR techniques that used to predict the stability (binding) constants or free energy complexation of some most interesting macrocyclic compounds, e.g. CDs and crown ethers, with different guest molecules including anionic, cationic and neutral molecules.

Keywords: macrocyclic compounds; QSAR; formation constants; host–guest interactions; cyclodextrin; crown ethers

1. Introduction

The extensive development of host–guest chemistry started in 1967 with the discovery of crown ethers by Pedersen (*1*, *2*). Supramolecular chemistry refers to the chemistry that focuses on non-covalent interactions of molecules and is a highly interdisciplinary field covering the aspects of chemistry, physics and biochemistry. Host–guest chemistry is an important subdivision of supramolecular chemistry, in which usually two or more molecules or ions are held together to form a complex in an unique structural relationships by intermolecular forces (*3*, *4*). Host–guest inclusion complexes, as a specific type of supramolecular structure, describe complexes in such a way that a smaller guest molecule is held within the internal cavity of a larger host molecule. These complexes are composed of two or more molecules or ions that are held together by non-covalent bonds, e.g. van der Waals forces, hydrogen bonding, ionic bonding and hydrophobic interactions. Non-covalent binding provides an invisible wiring diagram for biomolecular pathways and is the essence of host–guest and supramolecular chemistry (*5*). The formation of a complex between host and guest is a basic and important process in supramolecular chemistry; therefore, stability constants are used as an important criterion for the evaluation of the host–guest complexation process (*6*).

A wide variety of experimental methods have been employed in the determination of thermodynamic parameters for the complexation of host–guest interactions. These methods are based on the measurement of changes of additive properties (concentration dependencies) including chemical reactivity, molar absorptivity and other optical properties (*7*, *8*), aqueous solubility, calorimetric titration (*9*, *10*), NMR chemical shift (*11*, *12*), pH metric method (*13*, *14*) and chromatographic retention times (*15*, *16*) between host, guest and inclusion complexes. Depending on the association strengths, a suitable method to determine the stability constants must be selected. The sensitivity of the technique must allow for the detection of free and bound species (guest, host and host–guest complexes) in solution (*6*). Principles and experimental procedures of the various methods mentioned above can be found in the references cited.

All the experimental methods for the stability constants determination should be used with the proper precautions and with appropriate skills to obtain reliable data. It should be noted that thermodynamic data such as the stability constants determined by different experimental methods for the same reaction can significantly deviate from each other (*17*, *18*). In experiments, even small inaccuracies in measuring species concentration or temperature may lead to errors in complexation constants up to several log units (*19*). The reliability, quality and accuracy of numerical data on the stability constants values are needed in many areas of chemistry such as chromatography, metals extraction, ion exchange processes, complexometric titrations and in many aspects of environmental, academic, medical and industrial researches (*20*), but these data are

*Corresponding author. Email: jahan.ghasemi@gmail.com

ISSN 1061-0278 print/ISSN 1029-0478 online

http://dx.doi.org/10.1080/10610278.2011.581281
http://www.tandfonline.com

not always experimentally available. Moreover, the experimental methods of stability constants determination often are time consuming, expensive and require the use of pure compounds. Alternatively, computational methods as useful and interesting modelling tools of estimation and prediction of stability constants and physicochemical properties provide a promising method for the calculation of log K, especially in host–guest complexation processes. They also offer a fast measure of predictability in the absence of extensive experimental or computed data on compounds properties. Quantitative structure–activity relationship (QSAR) approaches, as one of the major computational molecular modelling methodologies (*21*), act as an effective means for the prediction and estimation of activities or properties of compounds based on their structures.

2. Overview of quantitative structure–property relationship studies

All the properties of organic molecules including physical, chemical, biological and technological properties depend on their chemical structure and vary with it in a systematic way. The establishment of quantitative correlations between diverse molecular properties and chemical structure is now of great importance to the society in assessing and improving environmental, medicinal and technological aspects of life. These are expressed as quantitative structure–activity/property relationship (QSAR/QSPR) that relates physical, chemical or physicochemical properties of compounds to their structures. The advantages of QSPR approaches lie in the fact that they require only the knowledge of chemical structure and are not dependent on any experimental properties. The major goal of the QSPR studies is to find a mathematical relationship between the property of interest and one or more descriptive parameters (descriptors) derived from the structure of the molecule (*22*, *23*).

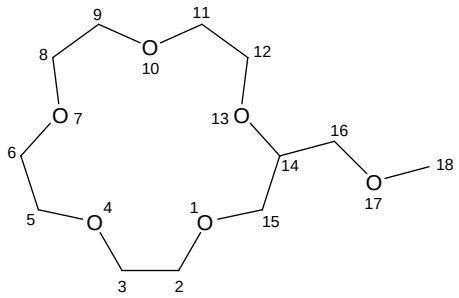
The three main steps in building QSAR models are extracting descriptors from molecular structures, choosing informative descriptors (feature selection) that are deemed to be important for explaining desired property or activity and constructions and development, validation and interpretation of QSAR models (*24–26*).

Prior to feature extraction or descriptor calculation as the first step in QSAR modelling, geometry optimisation of molecules which finds the coordinates of a molecular structure that represents a potential energy minimum is required. The correct optimisation of molecular structures to get meaningful values for the descriptors is very important (*27*). Several geometry optimisation methods are commonly used in QSAR modelling involving empirical force field methods, semi-empirical methods such as AM1, MNDO, PM3 and *ab initio* methods such as Hartree–Fock and density functional theory (DFT) methods (*28*). Drawing the molecules and optimisation of their structures are usually done with many commercial and non-commercial programs such as ChemDraw (*29*), Hyperchem (*30*), Sybyl (*31*) and Discovery Studio (*32*).

The molecular descriptor is the final result of a logic and mathematical procedure, which transforms chemical information encoded within a symbolic representation of a molecule into a useful number or the result of some standardised experiment (*33*, *34*). Based on the dependence on the information about 3D orientation and conformation of the molecule, descriptors have different kinds and thus many different QSAR approaches have been developed. Different kinds of descriptors involve 0D descriptors that include empirical properties and numbers of atoms, 1D descriptors that include substructures, 2D descriptors or topological descriptors in which the structures of compounds can be represented as graphs and 3D descriptors or geometric descriptors that encode the 3D aspects of the molecular structures that can be used in QSAR models (*33*, *35*). Based on the desired context and enough knowledge of molecular characterisations, we can also define some new descriptors, e.g. Ghasemi et al. (*36*) defined some novel descriptors to estimate lariat effect of the crown ether ring on complexation process of sodium ion with 15-crown-5 derivatives that lead to a more predictive QSAR model. For more clarification, some 0–3D descriptors of a 15-crown-5 ether are displayed in Table 1.

2D-QSAR is based on 2D descriptors that are being independent of the 3D orientation of the compound and range from measures of entities constituting the molecule, through its topological and geometrical properties to calculation of electrostatic and quantum chemical descriptors or advanced fragment counting methods (*36*). 3D-QSAR refers to the application of force field calculations requiring 3D structures. It involves obtaining numerical descriptors based on the steric property (shape of the molecule) and electrostatic fields based on the applied energy function in conformation of molecular structure from experimental data (X-ray crystallography) or theoretical methods such as molecular mechanics (*37*, *38*). Alignment-dependent 3D-QSAR descriptors require uniform alignment of molecules in space prior to the calculation of descriptors. Comparative molecular field analysis (CoMFA) (*39*, *40*) and comparative molecular similarity indices (*39*, *41*) are the most common alignment-based 3D-QSAR methods. These methods have two constraints. The correct conformation of a molecule must be used, which may not even be the lowest energy conformation, to compare structurally different compounds; second, the compounds must be properly aligned, a step that is time consuming and may introduce user bias (*42*). Alignment-independent 3D-QSAR descriptors were developed with an aim to overcome the alignment problems. The common alignment-free 3D-QSAR methods are comparative molecular moment

Table 1. Example of different kinds of descriptors for a typical compound.



0D descriptors*	Molecular weight = 250.33 Number of atoms = 39 Number of oxygen atom = 6
1D descriptors*	Number of ring secondary $CSP^3 = 9$ Hydrophilic factor = −0.134 Octanol/water partition coefficient (log *P*) = −1.29
2D descriptors*	Randic connectivity index = 8.432 Global topological charge index = 0.135 Symmetry index = 45.343
3D descriptor*	Spherocity = 0.687 1st component size directional WHIM index = 6.702 Total positive charge = 0.2
Self-defined descriptors	Distance between O_{13} and O_1 = 3.074 (A°) Number of lariat units = 1 Number of arms length of longest arms = 104

*Calculated by Dragon software.

analysis (*37*, *43*), weighted holistic invariant molecular (WHIM) descriptors (*44*, *45*), VolSurf descriptors (*46*), and grid-independent descriptors (GRIND) (*42*).

Some of the most popular software packages for generating an extensive list of 0–2D descriptors are ChemBioOffice (*47*), Dragon (*48*), Gaussian (*49*), COD-ESSA PRO (*50*), POLY (*51*), Chem-X (*52*) and TSAR (*53*), and program package generating 3D descriptor involves SYBYL (*31*), Pentacle (*54*) and Volsurf (*55*). These programs generally generate hundreds or thousands of different molecular descriptors but only some of them are significantly correlated with the desired activity. Inter-correlation of many of the descriptors, limitation of some statistical methods in handling data sets with a large number of descriptors to compound ratios and ambiguity of interpretability of the final model are the negative effects of a large number of descriptors on QSAR analysis. There are many methods based on filtering and wrapper techniques for selecting the best descriptors, or features, to be used in construction of the QSAR model. In filtering method, no model is built, and features are evaluated using some other criteria. Correlation-based methods (*56*), which use Pearson's correlation coefficients as a preliminary filter for discarding inter-correlated descriptors, statistical criteria, e.g. Fisher's ratio, ratio of the between-class variance to the within-class variance, and the Kolmogorov–Smirnov test (*57*) that measures the maximal absolute distance between cumulative distribution functions of the descriptor for individual activity classes are the common filtering methods in feature selection. Wrapper approach, which is based on constructing and evaluating a series of QSAR models to select a subset of descriptors, includes sequential backward feature elimination, sequential feature forward selection (*58*), genetic algorithm (GA) (*59*, *60*) and simulated annealing (*61*). For speed and simplicity, the combination of these techniques can also be used.

The main step of QSAR model is to derive a correlation between the activity and the values of the features. Different statistical or chemometric techniques form the mathematical foundation for building a QSAR model. Many different linear and nonlinear methods can be used to build and develop a relationship between the structure descriptors and many kinds of activities or properties of the molecules. Linear models based on multivariate analysis (*62*) include multiple linear regression (MLR) (*62*), partial least square (PLS) (*63*, *64*), principal component regression (*65*) and nonlinear methods involving artificial neural networks (ANNs) (*66*, *67*), the *k*-nearest neighbour (kNN) (*68*, *69*), GAs (*70*, *71*) and support vector machines (SVMs) (*72*, *73*).

Validation and development of the models is an important and crucial step of any QSAR study. The reliability of a QSAR model depends on how well the model can predict the interested property or activity of compounds outside the training set rather than how well the model reproduces the property/activity of compounds included in the model (*74*). A reliable and predictive QSPR model should be statistically significant and robust, be validated by making accurate predictions for external data sets not used in the model development and have a defined domain of application (*75*). Various procedures and quantitative parameters are used to express the performance of QSAR models. The most common statistical parameters are Pearson's correlation coefficient R^2 as 'model fitness' that should preferably be as close to unity as possible, the residual standard deviation (RSD), as small as possible and *F*-values (*75*, *76*). Leave-one-out cross-validated correlation coefficient Q^2 (LOO R^2) as one of the most popular validation criteria (*75*), leave-many-out (*77*), bootstrapping (*78*) and permutations of the data (randomisation test or scrambling) that represent the model robustness (*77*) are internal validation methods. Use of an external validation set is one of the most widely used methods of correlation testing. The purpose of an external validation is to evaluate how well the obtained model generalises objects in which the data have not taken part in the process of model

development (*73*, *78*). A valid model with high generalisation ability has R^2 and standard deviation (SD) for the validation set similar to those of the model. The predictive power of the QSPR models is often quantified in terms of the root mean square error (RMSE), RSD or predictive squared correlation coefficient R^2_{pred} (*79*). There are many other statistical parameters to assess the goodness of fit, robustness and predictivity of the models such as the Kubinyi function (FIT) (*80*, *81*), Akaike's information criteria (AIC) (*82*, *83*), Cook's distance, etc. (*84*). Once a reliable model is established, it could be possible to predict the property of compounds and know which structural factors play important roles in the interested property (*24*–*26*). The extraction of the structure–property relationship information encoded in the model or model descriptor interpretability is one of the important aspects of QSAR/QSPR models (*85*). A general flowchart of QSAR methodologies is summarised in Figure 1.

There are interesting developments in the area of QSAR/QSPR and related methods in the estimation of stability constants (log *K*) of the complexes of macrocycles with different cationic, anionic and neutral compounds. Here, we have discussed and reviewed QSAR/QSPR methods of stability constants prediction in host–guest complexation of some interesting macrocyclic ligands such as crown ethers and CDs with anionic, cationic or neutral guest molecules.

3. Macrocyclic compounds

As defined by IUPAC, a macrocycle is a cyclic macromolecule or a macromolecular cyclic portion of a molecule (*86*). This definition is not so clear; thus, there are other definitions of macrocyclic compounds. In supramolecular chemistry, a macrocycle is assigned as a molecule containing a number of binding sites that are arranged around the closed system (*87*). Organic chemists specify a cyclic organic molecule usually with 12 or more atoms in the ring (*88*), and coordination chemists define a macrocycle as a cyclic molecule with three or more potential donor atoms that can coordinate to a metal centre (*89*). Generally speaking, macrocycles comprise a large group of heterocyclic organic compounds that contain sizeable central holes or cavities in which a cation, anion or a neutral molecule can be encapsulated.

The ability of macrocyclic compounds to form stable complexes with various ionic and neutral spices and the recognition capabilities of these structures are their main advantages over simpler acyclic compounds. Because of their unique complexation characteristics, macrocyclic ligands have been studied for over four decades in host–guest complexation process. Applications of macrocycles that ranged from thorough investigations of fundamental principles of photophysics (*90*, *91*), electrochemistry (*92*), spectroscopy (*93*, *94*), nanotechnology (*95*, *96*), molecular

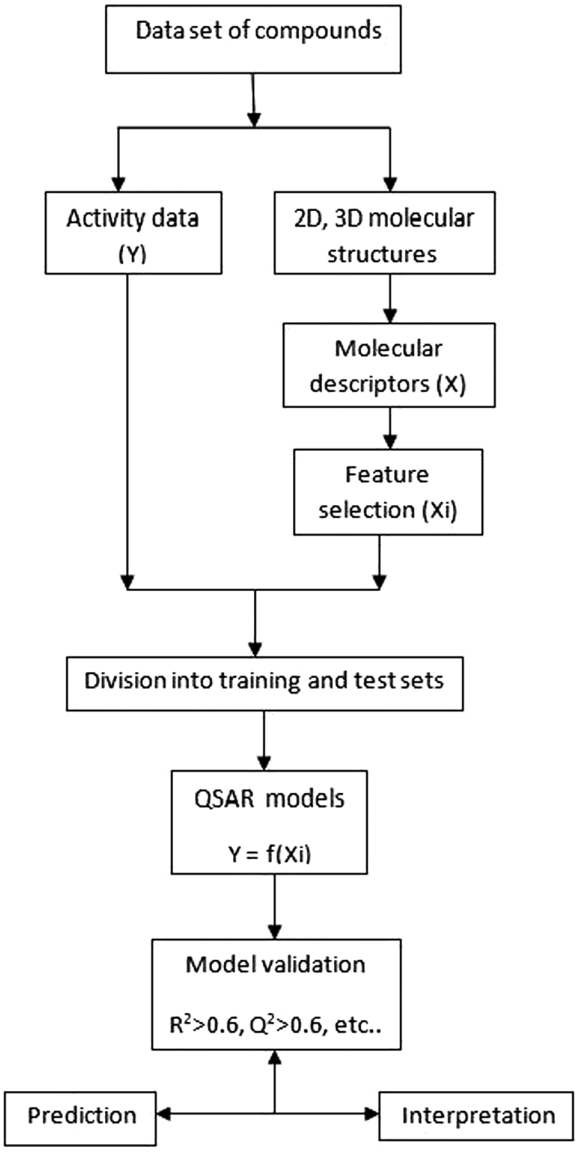


Figure 1. General flowchart of QSAR models.

recognition (*97*, *98*), separation science (*99*, *100*) and medicine (*101*) usually have host–guest interactions as a focal point.

The host–guest size-fit cavity of macrocycles and molecular symmetry play a crucial role in the complex stability. The size-fit effect appears to play a subsidiary role in the inclusion complexation of the host–guest molecules. The fit of the entire or at least a part of the guest molecule in the cyclodextrin host cavity determines the stability of the inclusion complex and the selectivity of the complexation process. The conformational flexibility of the host and the elasticity of the ligand are the important factors for successful thermodymanic and kinetic selectivity of inclusion complexation (*102*, *103*).

Izatt and his co-workers have made great contributions to the accumulation of thermodynamic data of crown

ethers complexation and have published comprehensive reviews in which extensive discussions are given based on the number of thermodynamic and kinetic data, log *K*, ΔH and ΔS, compiled for various crown ethers and different cations, anions and neutral molecules (*103–106*). Complexation thermodynamics of natural and modified α-, β- and γ-cyclodextrins (CDs) for 1:1 inclusion complexation of various organic guests was well reviewed and collected by Rekharsky et al. (*107*). The IUPAC stability constants database (*108*), National Institute of Standards and Technology (*109*) and THECOMAC (*110*) database are the most common and comprehensive databases of complexes stability constants.

3.1 Crown ethers

Crown ethers, synthesised first by Pedersen in 1967 (*1*, *2*), are the cyclic polyether molecules, such as compounds 1–3 in Figure 1, with multiple heteroatoms (three or more) incorporated in a monocyclic carbon backbone. The central feature of crown ethers is their ability to form selective well-defined complexes with a wide variety of ionic and neutral species in different solvents. Their potential to form stable complexes, exhibiting strong affinity and a high selectivity, especially for alkali and alkaline earth metal ions, is due to the nature of their multiple recognition sites and to the presence of a hydrophilic cavity delineated by a lipophilic envelope. Several factors influence on the stability of the crown ethers complexes such as the relative size of the guest (neutral or ionic) and the macrocycle cavity, the number and the nature of the binding sites and on the nature of the solvent. Many different modifications of the crown ethers, such as changing the ring size, the kinds of substituents and the types of donor atoms, have been made to enhance their complexation properties. Aza-crown ethers, thio-crown ethers and lariat crown ethers are the common types of crown ethers. Crown ethers that were extensively studied about their complexing ability, phase transfer catalysis, metal cation transport and metal cation analysis have been reported (*111*, *112*). They are also widely applied in chemical technology and analytical chemistry as ion pair extractants, membrane transfer, ion carriers, masking agents and sensors (*113–116*). Crown ethers continue to be one of the most useful parts of supramolecular (host–guest) chemistry (*117*). An overview of some of the crown ether structures is provided in Figure 2.

3.2 Cyclodextrins

CDs are cyclic oligosaccharides derived from starch containing 6 to 10 or more (α-1,4)-linked glucopyranose units. All the CDs form doughnut-shaped molecules with their hydroxyl groups on the outside of the molecule and a relatively non-polar and hydrophobic cavity in the middle, which can encapsulate a guest molecule to form an inclusion complex with a variety of guest molecules (*118*). Figure 3 represents the general structure of CDs, chemical structure and 3D structure of β-CD. In Table 2, some of the structural and physicochemical properties of natural CDs are listed (*118*). Among the natural CDs, β-CD has been used more widely in different areas because it is readily available and its cavity size is suitable for the widest range of guest molecules.

CDs have attracted tremendous interest in many different fields recently such as catalysis, separation science and technology, drug delivery, and pharmaceutical application (*119–122*). Great efforts have been devoted to the quantitative understanding of host–guest interactions due to the importance of the inclusion phenomena in biochemical systems. The enhancement of characteristics, such as stability, aqueous solubility and reduced volatility, can be modified through chemical reactions of CDs. The most common chemically modified CDs are 2-hydroxypropyl-β-CD, randomly methylated β-CD and sulphobutyl ether-β-CD respectively (*123*, *124*).

4. QSAR/QSPR prediction of the stability constants of macrocycles

4.1 0–2D QSAR/QSPR

Traditional QSAR/QSPR methods have been used for decades to correlate and predict the activity of molecules

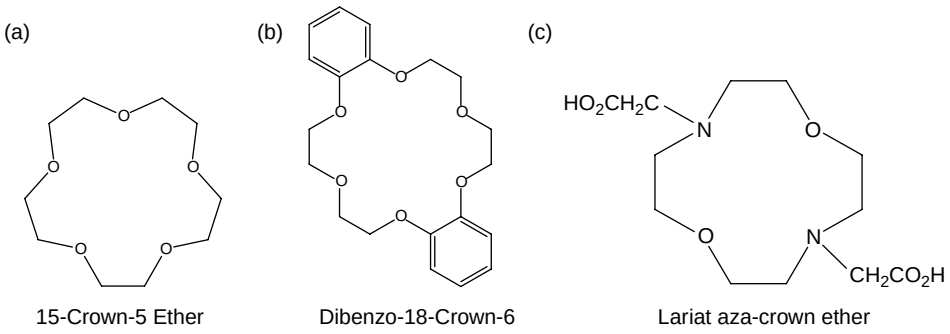


Figure 2. Structures of typical crown ethers.

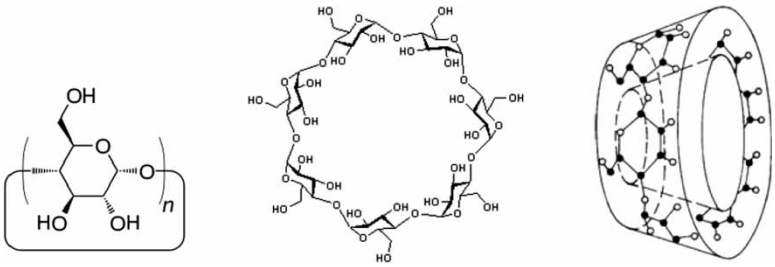


Figure 3. (a) General CDs structure, (b) chemical structure and (c) 3D structure of β-CD.

(*125*). They are fast, simple and include clearly defined physicochemical descriptors and are best suited for the analysis of a large number of compounds. Linear solvation energy relationship (LSER) (*126*) as a powerful QSAR approach is the first method for the prediction of stability constant of inclusion complexes of macrocycles.

In a significant work, Matsui and Mochida (*127*) determined the thermodynamic stabilities for α- and β-CD complexes with a variety of alcohols including 27 saturated aliphatics, 5 alicyclics and 2 aromatic alcohols. Partition coefficient of alcohol in a diethyl ether–water system was used as the only independent parameter to predict the stability constants between α- and β-CD and 1-alkanols with R (correlation coefficient) >0.99. They also applied two indices of molecular bulkiness, Taft's steric substituent constant (E_s) and ν parameter (*128*), to build QSAR models. They concluded that hydrophobic and van der Waals interactions are of primary importance in inclusion process of CD–alcohols complexation. van der Waals interactions are preferential for α-CD and hydrophobic forces for β-CD-1-alkanols inclusion complexes. As the bulkiness of alcohols increases, the stability of α-CD inclusion complex with 1-alkanols decreases owing to van der Waals repulsion, while that of β-CD-1-alkanols inclusion complexes increases due to the attainment of the close van der Waals contact of alcohols with β-CD cavity.

Lopata et al. reported a quantitative structure–stability relationship for the inclusion complexation of 17 barbituric acid derivatives with α- and β-CD. Contribution of group R_1 (an alkyl group on carbon number 7 in barbiturate ring) to hydrophobicity, Taft substituent constant, the molar refractivity and chloroform–water partition coefficient were used as descriptors. The results suggested that in α-CD–barbiturate complexes the CD cavity includes only R_1, while in β-CD complexes both R_1 and (a part of) the barbiturate ring are included (*129*).

Park et al. measured the stability constants of inclusion complexes of β-CD and some organic molecules and then used the LSER to study intermolecular forces affecting the stability of inclusion complexes between CD and the guests. It was found that increasing the guest molecular size stabilises the complex by virtue of increasing dispersive interactions between the hydrophobic interior of the CD cavity and the guest, whereas increasing guest dipolarity and hydrogen bond (HB) acceptor basicity leads to a decrease in the stability of the complex due to the stronger dipolar and hydrogen bonding interactions with water, which is more dipolar and has an acidic HB than CD (*130*).

An MLR analysis was carried out for the inclusion complexation of β-CD binding with 40 mono- and 1,4-disubstituted benzenes (*131*) from substituent molar refraction (R_m), hydrophobic constant (π) and Hammett constant (π). It was found that van der Waals forces, hydrophobic interactions and electronic effects comprise the driving forces for the binding of β-CD with mono- and 1, 4-disubstituent benzenes. Multivariate linear regression models from the R_m, π and δ of α- and β-CD with 24 mono-substituted benzenes inclusion complexes, with correlation coefficient of 0.94 for α-CD and 0.91 for β-CD were also reported (*132*, *133*).

In a significant work, Liu and Guo (*134*) demonstrated a nonlinear free energy relationship model by taking the possibility of different inclusion orientations into consideration for the molecular recognition of α- and β-CD with mono- and 1,4-disubstituted benzenes. GA was employed to optimise the model with independent variables R_m, π and δ, which reflect the volume,

Table 2. Some characteristics of natural CDs.

Cyclodextrin	α-cyclodextrin (α-CD)	β-cyclodextrin (β-CD)	γ-cyclodextrin (γ-CD)
Number of glucose unit	6	7	8
Water solubility (g/l)	145	18.5	232
Internal diameter (A°)	4.7–5.2	6–6.4	7.5–8.3
Height (A°)	6.7	7.0	7.0
Approx. cavity volume in 1 mol CD (ml)	104	157	256

hydrophobicity and electronic property of the substituents in the guest compounds, respectively. Provided that one of the substituents (*X* or *Y*) of mono- or 1,4-disubstituted benzene (X-C_6H_4-Y) was located in the cavity of CDs, they defined two microscopic binding constants (K_X and K_Y) as the following:

$$K_X = \frac{[\mathrm{CD}.X - \mathrm{C_6H_4} - Y]}{[\mathrm{CD}][X - \mathrm{C_6H_4} - Y]}$$

and

$$K_{XY} = \frac{[\mathrm{CD}.Y - \mathrm{C_6H_4} - X]}{[\mathrm{CD}][X - \mathrm{C_6H_4} - Y]}$$

Thus,

$$K_\mathrm{a} = K_X + K_Y.$$

The developed model offered quantitative information on the CD host–guest orientation and showed good agreement between the calculated results with the experimental data.

Guo et al. (*135*) described a wavelet neural network (WNN) to model construction and prediction of the binding constants for the inclusion of α-CD with mono- and 1,4-disubstituted benzene from the R_m and π. The ln K_a values calculated by the WNN model had a good predictive ability with a SD of 0.14 and a correlation coefficient of 0.995. In similar works, they also applied ANN to predict binding constants for the inclusion complexation of α-CD (*136*) and β-CD (*137*) with benzene derivatives from R_m, π and δ.

The association constants (K_a) for the inclusion complexation of α-CD with 72 mono- and 1,4-disubstituted benzenes were predicted by ANN with R_m and π as input parameters in the work of Liu and Guo. The results again indicated that the van der Waals forces and hydrophobic interactions mainly contributed to the driving forces for the inclusion complexation of α-CD with benzene derivatives (*138*). They also applied a WNN to the inclusion complexation of β-CD with 40 mono- and 1, 4-disubstituted benzenes from R_m, π and δ of the guest compounds as input parameters. WNN models compared to back propagation neural network and MLR models yielded a better model with $R > 0.99$ and SD < 0.1. The obtained results suggested that β-CD inclusion complexation is mainly driven by van der Waals force, hydrophobic interaction and electronic effects (*139*).

The combined use of potentiometry, circular dichroism, H NMR, UV spectroscopy and quantitative structure–affinity relationships of inclusion compounds of 16 *para*-substituted phenols and β-CD was reported (*140*). A linear regression based on R_m, π and δ with a good correlation between experimental and theoretical formation constants demonstrated the influence of dipolar interaction, since a withdrawing substituent favours the complex formation.

Davies and Deary used linear free energy relationships besides experimental methods for the calculation of the binding constants of complexation of different guest molecules with α-CD. They reported the determination and calculation of the stability constants of the complexes of α-CD and 4-methyl-, 4-nitro-, 4-sulphonato- and 3-chloro-substituted perbenzoic acids, perbenzoates and benzoates by iodometric titration (*141*), 22 *para*-substituted aryl alkyl sulphides, sulphoxides and sulphones by spectrophotometry (*142*), 21 *para*-substituted acetophenones and related aryl ketones spectrophotometrically or potentiometrically (*143*) and 48 substituted or 1,4-disubstituted benzene derivatives (*144*). The obtained linear free energy relationship models were described by δ_x and δ_y, which are the Hammett δ_p values for the *x*- and *y*-substituents, R_mx and R_my represented the *x*- and *y*-substituents molar refractivity, respectively, and $\delta_x\delta_y$ is an interaction term. Several mechanisms for cooperative and orientation of the different ionic or neutral guest species in the cyclodextrin cavity were suggested.

A quantitative structure–binding relationship and ANN with different back propagation algorithms of a data set of 17 barbiturates as guests to α- and β-CDs were reported (*145*). Four descriptors, derived from Lopata work (*129*), were used as the inputs of network involving contribution of group R_1 (an alkyl group on carbon number 7 in barbiturate ring) to hydrophobicity, Taft substituent constant, the molar refractivity and chloroform–water partition coefficient. The radial basis function (RBF) networks were also applied with the same descriptors on this barbiturate data set (*146*). It concluded that the effects of specific substituents were more efficient rather than the effects of the entire molecule in host–guest complexation process of barbiturates and CDs.

Table 3 represents the summary of LSER approaches with an overview of statistical methods used and model prediction performance.

The method of substructural molecular fragments (SMF) is based on the splitting of a molecular graph into a limited number of topological fragments and calculation of their contributions to a given property *X* (*147*). Solov'ev and Varnek used SMF method to build structure–property models of the stability constants for 56 complexes of crown ethers including unsubstituted macrocycles, their benzo, cyclohexyl and lariat derivatives with sodium ion in methanol. They also assessed the stabilities of the inclusion complexes of the β-CD with mono- and 1,4-disubstituted benzenes. In further work, Varnek and Wipff (*148*, *149*) applied SMF method to assess the macrocyclic effect for the complexation of crown ethers, polyethers and glymes with sodium, potassium and caesium ions in methanol. The several obtained 'best-fit' models, with

Table 3. Summary of LSER approaches and source information.

Host	Data set	Statistical method	Model prediction performance	Reference
α-CD	17 1-alkanols	MLR	$R > 0.95$, SD < 0.27	(*127*)
β-CD	18 1-alkanols		$R > 0.93$, SD < 0.38	
α-CD	17 barbituric acid derivatives	MLR	$R = 0.925$, SD $= 0.183$	
β-CD			$R = 0.95$, SD $= 0.144$	(*129*)
β-CD	20 organic molecules	MLR	$R = 0.927$, SD $= 0.27$	(*130*)
β-CD	40 mono- and 1,4-disubstituted benzenes	MLR	$R = 0.95$, SD $= 0.24$	(*131*)
α-CD	24 mono-substituted benzenes	MLR	$R = 0.94$, SD $= 0.33$	
β-CD			$R = 0.91$, SD $= 0.24$	(*133*)
α-CD	24 mono-substituted benzenes	MLR	$R = 0.96$, SD $= 0.29$	
β-CD			$R = 0.94$, SD $= 0.21$	(*132*)
α-CD	56 mono- or 1,4-disubstituted benzene	GA	$R = 0.82$	(*134*)
β-CD			$R = 0.82$	
α-CD	45 benzene derivatives	WNN	$R = 0.995$, SD $= 0.14$	(*135*)
α-CD	Benzene derivatives	ANN	–	(*136*)
β-CD	24 benzene derivatives	ANN	$R = 0.94$, SD $= 0.19$	(*137*)
α-CD	72 mono- and 1,4-disubstituted benzenes	ANN	$R = 0.961$, SD $= 0.415$	(*138*)
β-CD	40 mono- and 1,4-disubstituted benzenes	WNN	$R = 0.992$, SD $= 0.089$	(*139*)
β-CD	16 phenolic compounds	MLR	$R = 0.965$	(*140*)
α-CD	Substituted perbenzoic and benzoic acids	–	–	(*141*)
α-CD	para-substituted aryl alkyl sulphides, sulphoxides and sulphones	MLR	$R = 0.923$, SD $= 0.345$	(*142*)
α-CD	21 para-substituted aromatic ketones	MLR	–	(*143*)
α-CD	48 substituted or 1,4-disubstituted benzene derivatives	MLR	$R = 0.923$, SD $= 0.345$	(*144*)
α-CD	17 barbituric acid derivatives	ANN	$R = 0.9583$, SD $= 0.028$	(*145*)
β-CD				
α-CD	17 barbituric acid derivatives	ANN	$R^2 = 0.99$	(*146*)
β-CD				

linear and nonlinear equations, showed good agreement between calculated and experimental data of the stability constants, with $R > 0.95$ and SD < 0.5. Solov'ev et al. demonstrated a SMF method to represent structure–property models of the stability constant (log *K)* for the 1:1 (M:L) complexes of Sr^{2+} with various organic ligands including crown ethers and azacrown ether derivatives of 12-crown-4 and 18-crown-6 in water at 298 K and ionic strength 0.1 M. The obtained models that were characterised by $R^2 > 0.91$ and Q^2 (leave-one-out cross-validation correlation coefficients) > 0.85 were utilised for the generation and screening of a combinatorial library of virtual ligands (*150*).

Katritzky et al. (*151*) developed the QSPR modelling of binding energies for 1:1 complexation systems among 218 different organic compounds including aromatic hydrocarbons, alcohols, phenols, ethers, esters, aldehydes, ketones, acids, sulphur-containing compounds, nitriles, anilines, heterocyclic compounds, steroids and barbiturates with β-CD using multiple regression analysis. They used two different QSPR approaches. In descriptor approach, a large variety of molecular descriptors on the basis of the 3D geometrical or quantum chemical structure generated by CODESSA PRO were applied to yield a model with $R^2 = 0.796$ and $Q^2 = 0.779$. Fragment-based approach that used SMF descriptors has given a better fit with $R^2 = 0.943$ and $Q^2 = 0.848$. Zhokhova et al. (*152*) used the same data set of 218 organic compounds to built QSPR model using SMF descriptors and multiple linear regressions (MLRs).

Table 4 represents the summary of SMF approaches with an overview of statistical methods used and model prediction performance.

As discussed in Section 2, there are many commercial and non-commercial program packages to calculate hundreds to thousands of theoretical descriptors. Table 5 represents the summary of QSAR/QSPR approaches, based on the theoretical descriptors driven from different software packages, with an overview of statistical methods used and model prediction performance.

ANN as a powerful tool of approximation, computation and pattern recognition (*153*, *154*) have been receiving growing attention from chemists in different areas (*155–157*).

Tetko et al. used several linear and nonlinear methods including associative neural networks (ASNNs), SVMs, kNNs, maximal margin linear programming, RBF neural network (RBFNN) and MLR for QSPR of the stability constants log *K* for 1:1 (M:L) and log β for 1:2 complexes of metal cations Ag^+ and Eu^{3+} with diverse sets of macrocyclic, heterocyclic and acyclic agents bearing acidic, basic or neutral functions. Four structurally diverse data sets include 161 (Ag^+) and 241 (Eu^{3+}) log *K* and 112 (Ag^+) and 81 (Eu^{3+}) log β values for the complexation of

Table 4. Summary of SMF approaches and source information.

Host	Data set	Statistical method	Model prediction performance	Reference
Crown ethers β-CD	56 (Na^+) 29 1,4-disubstituted benzenes	Linear and nonlinear regression	$R = 0.958$, SD = 0.21 $R = 0.909$, SD = 0.28	(*147*)
Crown ethers	58 (Na^+) 106 (K^+) 28 (Cs^+)	Linear and nonlinear regression	$R = 0.978$, SD = 0.16 $R = 0.973$, SD = 0.22 $R = 0.987$, SD = 0.17	(*148*)
Crown ethers	69 (Na^+) 123 (K^+) 31 (Cs^+)	Linear and nonlinear regression	$R^2 = 0.939–0.947$ $Q^2 = 0.804–0.864$ $s = 0.48–0.61$	(*149*)
Crown ethers	130 (Sr^{+2})	Linear and nonlinear regression	$R^2 = 0.940$, $Q^2 = 0.927$	(*150*)
β-CD	218 organic compounds	MLR	$R^2 = 0.967$, $Q^2 = 0.877$	(*151*)
β-CD	218 organic compounds	MLR	$R^2 = 0.876$, $Q^2 = 0.704$	(*152*)
Crown ethers	161 (Ag^+) 241 (Eu^{3+})	ANN, SVM, KNN, MLR	$R^2 < 0.62$ (nonlinear) $R^2 < 0.77$ (nonlinear)	(*158*)
α-CD β-CD	102 organic compounds 218 organic compounds	Nonlinear regression	$R^2 = 0.868$ $R^2 = 0.917$	(*154*)

Ag^+ and Eu^{3+} with organic molecules in water at 298 K and ionic strength 0.1 M. Three types of descriptors, E-state indices and counts and SMF descriptors were used. Among the many number of significant models, nonlinear methods, with the highest value of $R^2 = 0.79$ and RMSE = 1.17, showed a significantly better performance than the models built using MLR analysis. However, the averaging of several MLR analysis models based on SMF

Table 5. Summary of QSAR/QSPR approaches and source information.

Host	Data set	Statistical method	Model prediction performance	Reference
Crown ethers	14 (Na^+) 15 (K^+) 16 (Cs^+)	ANN	– – –	(*159*)
Crown ethers	92 (Na^+) 92 (Ca^{+2}) 92 (Zn^{+2})	MLR	SD = 0.36–1.42	(*160*)
15-crown-5 ethers	58 (K^+)	MLR	$R^2 = 0.945$, SD = 0.011	(*161*)
15-crown-5 ethers	88 (Na^+)	MLR, PLS	$R^2 = 0.86$, $q^2 = 0.72$, $R^2_{pred} = 0.885$	(*36*)
15-crown-5 ethers	54 (Na^+)	GA-MLR	$R^2 = 0.93$, $q^2 = 0.88$, $R^2_{pred} = 0.83$	(*162*)
β-CD DMAB-α-CD DMAB-β-CD DMAB-γ-CD	33 10 28 13	MLR	$R^2 = 0.92$, $q^2 = 0.901$ $R^2 = 0.861$, $q^2 = 0.784$ $R^2 = 0.893$, $q^2 = 0.795$ $R^2 = 0.89$, $q^2 = 0.821$	(*163*)
α-CD β-CD	102 organic compounds 218 organic compounds	MLR	$R^2 = 0.868$, SD = 1.97 $R^2 = 0.917$, SD = 1.66	(*165*)
β-CD	18 heterocyclic compounds	PLS	$R^2 = 0.972$, $q^2 = 0.782$	(*167*)
α-CD β-CD γ-CD	179 310 51	MLR	– $R^2 > 0.515$ –	(*170*, *171*)
α-CD β-CD	56 mono- and 1,4-disubstituted benzene	MLR	$R = 0.933$, $q = 0.920$ $R = 0.94$, $q = 0.927$	(*172*)
α-CD	22 mono-substituted benzene and phenol	MLR	–	(*173*)
α-CD β-CD γ-CD	48 70 45	MLR	$R = 0.78$, $Q^2 = 0.521$ $R = 0.944$, $Q^2 = 0.861$ $R = 0.947$, $Q^2 = 0.863$	(*174*)
β-CD	258 organic compounds	PLS MLR	–	(*176*)
β-CD	233 organic compounds	MLR	$R^2 = 0.868$, $Q^2 = 0.851$	(*177*)
β-CD	86 drugs	MLR	$R^2 = 0.78$, $Q^2 = 0.67$	(*178*)
β-CD	233 organic compounds	MLR	$R^2 = 0.841$, $Q^2 = 0.821$	(*180*)
β-CD	74 chiral compounds	SVM	$R^2 = 0.99$, $Q^2 = 0.97$	(*181*)
α-CD β-CD	73 aliphatic compounds 37 aliphatic compounds	MLR	$R^2 = 0.909$, $Q^2 = 0.898$ $R^2 = 0.912$, $Q^2 = 0.91$	(*183*)
β-CD	233 organic compounds	PLS	$R^2 = 0.87$, $Q^2 = 0.75$	(*189*)

descriptors provided as good of a prediction as the most efficient nonlinear techniques (*158*).

Gakh et al. (*159*) applied neural network method for the prediction of complexation constants of simple crown ethers with alkali metal cations (Na^+, K^+ and Cs^+) in a single solvent (methanol) at 25°C. The number of aromatic rings, the size of a macrocycle and the number of oxygen atoms were used as an input vector in the training sets containing only 14–16 ligands. The obtained ANN model could predict the stability constants of crown ether complexes based on the chemical structures encoded in their chemical names with an average accuracy of ± 0.3 log *K* units.

Shi and McCullough demonstrated a simulation statistical procedure for quantitative predicting complexation equilibrium constants for different crown ethers, cryptands and spherands with sodium, calcium, zinc and ammonium ions in pure or mixed solvents system. The method involves the combination of solvent-free molecular mechanics and molecular dynamics (MD) simulations with MLRs to experimental log *K* data to incorporate solvent and other effects (*160*). Various experimental and calculated parameters such as radius and electronegativity for metal cations, dielectric constant for solvents, total energy and its components for ligands and some others as descriptors were used in MLR analysis to give standard errors in log *K* ranging from 1.42 in the largest system to 0.36 in the smallest.

Ghasemi and Saaidpour (*161*) applied a QSPR modelling of stability constants of diverse complexes of 15-crown-5 ethers with potassium cation at 25°C in methanol solution. Structure-based descriptors from CODESSA PRO software and best multilinear regression (BMLR) were used to build QSPR models. They concluded that complexation phenomenon was mainly related to cation–ligand electrostatic interactions, steric deformation of the ligand and conformational changes of the ligand accompanying the complexation and repulsion between neighbours C—H and C—C bonds. In another article, they defined some novel descriptors to consider lariat effect, the effects of different pendant arm substituents and also conformation changing (shape and symmetry) in the crown ether rings on complexation reactions. MLR and PLS techniques were used for modelling the stability constants of 15-crown-5 derivatives complexes with sodium cation at 25°C in methanol solution (*36*). Combination of new defined descriptor and Dragon descriptors could lead to relevant QSPR models in comparison with Dragon-based descriptors that allow improving the robustness of stability constants predictions. A GA-based MLR (GA-MLR) method was applied for QSPR modelling of the stability constants for 65 complexes of 15-crown-5 with sodium cation (Na^+) in methanol (*162*).

In a valuable and significant work, Suzuki et al. (*163*) demonstrated the application of classical QSAR and CoMFA to the complexation of some natural β-CD and modified α-, β- and γ-CD, which bear one *p*-(dimethylamino)benzoyl (DMAB) moiety, with guest molecules of widely varying chemical structures and properties. Classical QSAR of the binding constants of natural β-CD with 33 diverse guest molecules which were described by just two parameters, the molecular connectivity index and the octanol–water partition coefficient (log *P*), suggested a good predictive ability of the model with $R^2 = 0.92$ and $Q^2 = 0.901$. A nonlinear dependency on binding constants on the zeroth and/or first-order molecular connectivity index as a measure of size was found for the natural β-CD: guest system. The correlation analysis of mentioned descriptors with modified α-, β- and γ-CD showed good predictivity with statistical parameters of R^2 and Q^2 in a range of 0.86–0.89 and 0.78–0.82, respectively. 3D-QSAR/CoMFA models constructed for the DMAB-α- and β-CD systems yielded statistically significant values of R^2 and Q^2 which were comparable to those obtained by MLR models. In further work, Suzuki developed a group contribution method (*164*) to build QSAR models of the binding constants or the free energies of complexation between α- and β-CD based on the database consisting of 102 and 218 diverse guest molecules, respectively (*165*). He found a nonlinear relationship based on the first-order connectivity index with $R^2 = 0.868$ for α-CD and $R^2 = 0.917$ for β-CD inclusion complexes. The predictive ability of the obtained models, which was tested by a *de novo* (*166*) prediction, using the data set of host–guest systems not included in the deduction of the models, showed an accurate prediction of the free energies of the complexation.

A QSAR study of the complexation of a series of unsubstituted benzo- or dibenzo-fused heterocycles, with a single or two heteroatoms (O, S, N) in the ring, with β-CD was reported (*167*). The physicochemical properties and the spectroscopic properties such as ^{13}C NMR chemical shifts of carbons of the benzene ring and the number and type of heteroatoms were selected as descriptors. Central composite design (*168*) was chosen as a strategy for selection of the training set. The results of the PLS modelling of the stability constants as a function of the heterocyclic structure showed that separate models for heterocyclic compounds containing N, alone or with a second heteroatom, and for compounds containing O or S are needed in order to have a satisfactory predictive ability.

Structure-based parameters (*169*) such as molecular size, hydrophobicity, rotatable bonds, electronic properties and the presence or absence of functional or structural moieties were used to predict the stability constants of 1:1 complexes for 179, 310 and 51 guest molecules with unsubstituted α-, β- and γ-CDs. At first, models for all CD developed only with size (molecular volume) of molecules, but size alone cannot account for a considerable amount of variance in the complex stability data, and

this is especially true for structures larger than the limiting size requirement of the corresponding CD cavity. Thus, other structure-based parameters were used to build structure–complex stability relationships relying on molecular size-based model (*170*, *171*).

The QSPR model of association constants (K_a) for the inclusion complexation of β- and α-CD and mono- and 1,4-disubstituted benzene derivatives by multivariate linear regression analysis using a combination of 2D and 3D-connectivity indices and quantum chemical descriptors was described by Estrada et al. (*172*). The best obtained QSPR models showed a good predictability of α-CD and β-CD with $R > 0.93$, $Q > 0.92$ and $SD < 0.5$ and $R > 0.94$, $Q > 0.92$ and $SD < 0.5$, respectively. They also concluded that the main driving force for the complexation of α-CD with benzene derivatives is the electronic repulsion, while van der Waals and hydrophobic interactions were prominent in β-CD complexation process with these guest molecules.

A QSPR model for the estimation of the free energy of formation of host–guest complexes of α-CD with benzene derivatives was reported (*173*). The used independent variables were polarisability (contribution of van der Waals interactions), distribution factor in the octanol–water system (log *P*, contribution of hydrophobic interactions), the capability of H-bonding of substituent X_i with the α-CD void and the number of heavy atoms in the substituent. They derived different QSPR models of free energies of formation for mono-substituted benzene derivatives, phenol derivatives and symmetrical 1,4-disubstituted benzene derivatives. It was concluded that differences in the interactions of groups X_i, corresponding to the van der Waals and donor–acceptor interactions of group X_i, had an important role in inclusion complexation of benzene derivatives and α-CD.

Klein et al. (*174*) described a method based on MLR and PLS models for predicting the free energies of complexation between β-CD and 70 pharmaceutical compounds. They used some descriptors indicating volume, shape and lipophilicity such as molecular surface, ovality, shape index, flexibility, partition coefficient, the sum of the electrotopological indices that were derived from TSAR program package. The obtained model, with $R = 0.927$ and $Q^2 = 0.812$, showed a good predictive ability exemplified by a *de novo* prediction, using compounds not included in the deduction of the model. In the further work, linear and nonlinear regression-based TSAR generated descriptors used to build QSAR models of the free energies of complexation between α- β- and γ-CDs and pharmaceutical compounds. To support the conclusions resulting from the analysis of the regression equations, energy minimisations and MD simulations were performed using the MM2 force field (*175*). An improvement, reflected in an increased *F*-value and an increased cross-validation Q^2, was obtained by introducing explicitly nonlinearities into the models for all CDs. While van der Waals interactions were important for complexation in the case of all three CDs, dipole–dipole interactions for the α-CD, hydrophobic effect for β-CD and the HB acceptor and donor capacity of the guest for γ-CD are more dominant and might also stabilise CD–guest complexes.

Chari et al. developed QSAR models for predicting the binding constants of 1:1 inclusion complexes between 258 ligands, ranging from drug-like molecules to small polar organic compounds, and β-CD. Both PLS regression and MLR and Dragon software generated descriptors were used to derive the models. The most important descriptor in both models was the calculated log *P*, indicating that drugs with greater lipophilicity form stronger complexes with β-CD (*176*). In a similar work, Li et al. constructed a QSAR model of binding constants of 86 poorly soluble drugs with β-CD complexes based on TSAR software descriptors. MLR analysis was applied to develop the QSAR model. They applied the obtained QSAR model to a data set of 229 organic compounds, which was previously studied by Pérez-Garrido et al. (*177*), compared their results and concluded that the behaviour of the drug molecules with CD should differ from that of the organic compounds (*178*).

A significant effort was extended to build QSPR models of 233 diverse set of organic molecules, the same data set used by Suzuki (*165*), with β-CD by Pérez-Garrido et al. (*177*). In the first work, they used a number of topological, physicochemical and 3D descriptors, GA variable selection method and *k*-means cluster analysis for the selection of training and test set compounds to model construction. Many statistical parameters including Q^2, correlation coefficient of test set (R^2_{pred}), bootstrapping testing techniques, Kubinyi function (FIT) and AIC, randomisation test and applicability domain of the model were calculated to assess the goodness of fit, robustness and predictivity of the obtained model. Among several models developed, topological QSAR model was able to explain more than 84% of experimental variance and reasonable interpretation of CD complexation process. They concluded that hydrophobic and steric (van der Waals) interactions were the main driving forces for CD complexation process. In the similar work, they used the same data set and applied topological substructural molecular design (TOPS-MODE) approach (*179*) for the calculation of descriptors and correlated them with β-CD complex stability constants by linear multivariate data analysis (*180*).

Prakasvudhisarn et al. developed SVM-based QSPR models (*181*) to predict complex stability constant (ln *K*), the standard free energy (ΔG°), the enthalpy (ΔH°) and the entropy change ($T\Delta S^\circ$) of the 1:1 inclusion complexes of enantiomeric pairs of 74 selected chiral compounds and β-CD. Structural properties were calculated by the MOE program package, and particle swarm optimisation (*182*) was adopted for feature selection and linear, polynomial and Gaussian RBFs used to build QSPR models of the

chiral guest data set. The obtained models showed good performance in predicting ln K, ΔG°, ΔH° and $T\Delta S^{\circ}$ of chiral guest inclusion complexes with β-CD, by considering the major selected features with four to eight descriptors and $R^2 > 0.96$ and $R^2_{pred} > 0.89$.

Complexation free energies of neutral, anionic and cationic aliphatic compounds with α- and β-CD with a set of empirical and theoretical descriptors that were chosen to build multiple linear correlation models were described (*183*). Molecular polarisability, the maximum absolute charge located on a guest atom and the molecular dipole moment to consider electrostatic effects, a modified zero-order connectivity index to account for the steric effects and a charge transfer parameter and the octanol–water partition free energy for the hydrophobic character were chosen as the descriptors for building MLR models. From the obtained MLR models, with $R^2 > 0.9$ and $Q^2 > 0.89$ for α-CD and $R^2 > 0.91$ and $Q^2 = 0.91$ for β-CD, it was concluded that for both α- and β-CD the most relevant driving force for the inclusion complexation with aliphatic species is van der Waals interactions.

4.2 3D-QSAR/QSPR

Classical QSAR/QSPR approaches have some drawbacks, e.g. only 2D structures considered, provide no unique solutions, insufficient parameters for describing drug–receptor interactions, no representation of stereochemistry or 3D structure of molecules, higher risk of chance correlations and so on (*184*). The primary aim of 3D-QSAR techniques is to establish a correlation of biological activities of a series of structurally and biologically characterised compounds with the spatial fingerprints of numerous field properties of each molecule, such as steric demand, lipophilicity and electrostatic interactions (*185*). An advantage of 3D-QSAR method is that it takes into account the 3D structures of ligands and additionally is applicable to sets of structurally diverse compounds. Unfortunately, there are few works of 3D-QSAR approaches of stability constants prediction of macrocyclic inclusion complexes. To our knowledge, the first paper on CoMFA method, as an approach of 3D QSAR, for CDs stability constants prediction was reported by Suzuki et al. (*163*).

Topological autocorrelation descriptors, an alignment-independent approach based on the autocorrelation of certain molecular properties (*186*), were used to build 3D QSAR modelling of free energy of complexation between β-CD and 70 pharmaceutical compounds (*187*). For comparison, a variety of descriptors that were implemented in the TSAR program were used. Both TSAR and 3D topological distance-based descriptor models showed approximately the same quality of predictivity with $R^2 > 0.99$ and $Q^2 > 0.66$.

Molecular interaction fields (MIFs) were used in combination with a small number of geometrical descriptors to separate nine α-CD complexes into two classes, respectively, containing complexes having high log K (log $K > 2$) and low log K (log $K < 2$) values. Structural optimisation, conformational analysis and descriptor calculation were carried out by MOE program package. MIFs with two probes, the hydrophilic field generated by the OH_2 probe and the hydrophobic field generated by the DRY probe, were calculated using the program GRID. The distances (SL) between the centroid of the secondary oxygen's (S) and the centroid of the aromatic ring (L) of the ligands were used as geometrical descriptors. A good predictive model with $R^2 = 0.93$ and one physicochemical, the ratio between the number of dry points and the number of water points (r(DRY/OH_2)), descriptor derived from MIF calculation between log K, 1:1 of α-CD and guest molecules was achieved (*188*).

Recently, Ghasemi et al. (*189*) applied GRIND-based 3D-QSAR models to predict the stability constants of a diverse class of 233 organic guest molecules with β-CD. The variables with the highest impact on the log K values were related to TIP and DRY probes, which represent shape field and hydrophobic interactions, respectively. It was concluded that the size and shape of the molecules as well as the presence and orientation of hydrophobic groups were crucial for the stability constant of investigated compounds with β-CD. Based on the distance node of variables, it is indicated that the spatial arrangement of hydrogen bonding regions of molecules is important in complexation of guest molecule to β-CD.

4.3 Miscellaneous

Besides, QSAR/QSPR techniques of computational chemistry such as molecular mechanics, MD, quantum computation, molecular docking and simulation procedures were widely used to identify the significant factors contributing the host–guest interaction and to predict the thermodynamic stability constants of inclusion complexes of macrocyclics. These methods were commonly used to calculate and predict the host–guest complexes shapes, energies, preferred binding orientations, selectivities and so on. Although these methods are usually time consuming and need a powerful CPU, these calculations can give an adequate first guess as to the nature and strength of inclusion complexation interactions. Computational calculation in host–guest interactions was commonly used in combination with an experimental procedure such as NMR and fluorescence methods to obtain structural and dynamics information and to better understand the forces between host and guest molecules (*190*, *191*).

In a comprehensive work, applications of computational chemistry to the study of CDs have been well reviewed by Lipkowitz (*192*). The use of quantum

chemical methods to study cyclodextrin chemistry was also surveyed by Liu and Guo (*193*). Docking studies (*194–196*), Monte Carlo (MC) simulations (*197*) and DFT-based studies (*198*) were also used for the estimation of stability constants of CDs and crown ethers.

5. Concluding remarks

QSAR methods, as a progressive tool in modelling and prediction of many physicochemical properties, are in constant advancement. Well-established QSAR techniques continue to be used, providing successful results in rapid and accurate assessment of large set of compounds. As the reliability of QSAR models strongly depends on the ability of the models to accurately predict the activity/property of compounds not included in the training set, the models must be thoroughly validated both internally and externally using rigorous cross-validation techniques. Throughout the review, we have focused on QSAR techniques to predict the stability (binding) constants or free energy complexation of some macrocyclic compounds with different guest molecules including anionic, cationic and neutral molecules.

Although there is an increasing interest in computational modelling of inclusion complexes of crown ethers as an important macrocyclic compound, it should be noted that there are not many publications on structure–property modelling of stability constants of crown ethers inclusion complexes especially with neutral molecules. Because of widespread use of CDs in many different fields, they have also attracted more interests than other macrocyclic compounds in QSAR/QSPR and computational chemistry approaches. Unfortunately, there are few papers on 3D-QSAR approaches of CDs or crown ethers, and no article could be found in higher dimensional QSAR, e.g. 4D and 5D-QSAR. These may be due to lack of X-ray crystallographic 3D structures of macrocyclic complexes, especially in 3D-QSAR studies, or absence of homogenous data set of guest or host molecules for QSAR studies. However, there is an increasing interest in the use of QSAR approaches in host–guest interactions of macrocyclic complexes. Considering different approaches and the rapid growth of novel techniques, QSAR/QSPR methods provide a promising tool for predicting the physicochemical properties in a more accurate and rapid manner.

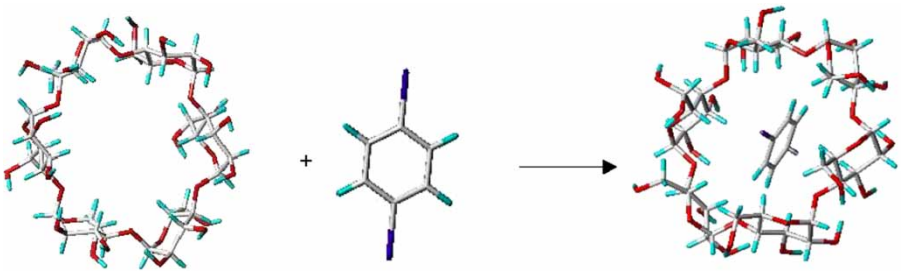
References

(1) Pedersen, C.J. *J. Am. Chem. Soc.* **1967**, *89*, 7017–7021.
(2) Pedersen, C.J. *Science* **1988**, *241*, 536–540.
(3) Cram, D.J. *Science* **1988**, *240*, 760–767.
(4) Cram, D.J.; Cram, J.M. *Science* **1974**, *183*, 803–809.
(5) Zhou, H.X.; Gilson, M.K. *Chem. Rev.* **2009**, *109*, 4092–4107.
(6) Keiji, H. *J. Inclusion Phenom. Macrocyclic Chem.* **2001**, *39*, 193–209.
(7) Douhal, A. *Cyclodextrin Materials Photochemistry, Photophysics and Photobiology*; Elsevier: USA, 2006.
(8) Mohammed-Ziegler, I.; Billes, F. *J. Inclusion Phenom. Macrocyclic Chem.* **2007**, *58*, 19–42.
(9) Arena, G.; Cali, R.; Maccarrone, G.; Purrello, R. *Thermochim. Acta* **1989**, *155*, 353–376.
(10) Willson, R.J.; Beezer, A.E.; Mitchell, J.C.; Loh, W. *J. Phys. Chem.* **1995**, *99*, 7108–7113.
(11) Lamsa, M.; Huuskonen, J.; Rissanen, K.; Pursiainen, J. *Chem. Eur. J.* **1998**, *4*, 84–92.
(12) Salvatierra, D.; Diez, C.; Jaime, C. *J. Inclusion Phenom. Mol. Recognit. Chem.* **1997**, *27*, 215–231.
(13) Carofiglio, T.; Magno, F.; Lavagnini, I. *J. Electroanal. Chem.* **1994**, *373*, 11–17.
(14) Chen, L.; Bos, M.; Grootenhuis, P.D.J.; Christenhusz, A.; Hoogendam, E.; Reinhoudt, D.N.; Van Der Linden, W.E. *Anal. Chim. Acta* **1987**, *201*, 117–125.
(15) Janos, P. *J. Chromatogr. A* **2004**, *1037*, 15–28.
(16) Skórka, M.; Asztemborska, M.; Zukowski, J. *J. Chromatogr. A* **2005**, *1078*, 136–143.
(17) Hallen, D. *Pure Appl. Chem.* **1993**, *65*, 1527–1532.
(18) Stademan, M.; Wadso, I. *Pure Appl. Chem.* **1995**, *67*, 1059–1068.
(19) Beck, M.; Nagypal, I. *Chemistry of Complex Equilibria*; Akademiai Kiado: Budapest, 1989.
(20) Ure, A.; Davidson, C. *Chemical Speciation in the Environment*, 2nd ed.; Wiley-Blackwell: Weinheim, 2002.
(21) Tropsha, A. *Mol. Inf.* **2010**, *29*, 476–488.
(22) Dudek, A.Z.; Arodzb, T.; Galvezc, J. *Comb. Chem. High T. Scr.* **2006**, *9*, 213–228.
(23) Hotje, H.D.; Sippl, W.; Rognan, D.; Folkers, G. *Molecular Modeling: Basic Principles and Applications*, 2nd ed.; Wiley-VCH: Weinheim, 2003.
(24) Fitch, W.L.; McGregor, M.; Katritzky, A.R.; Lomaka, A.; Petrukhin, R.; Karelson, M. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 830–840.
(25) Jurs, P.C. *Quantitative Structure–Property Relationships*; Wiley-VCH Verlag GmbH: Weinheim, 2008; pp 1314–1335.
(26) Karelson, M.; Lobanov, V.S. *Chem. Rev.* **1996**, *96*, 1027–1043.
(27) Rinnan, A.; Christensen, N.J.; Engelsen, S.B. *J. Comput. Aid. Mol. Des.* **2010**, *24*, 117–122.
(28) Fletcher, R. *Practical Methods of Optimization*, 2nd ed.; John Wiley: USA, 2000.
(29) ChemDraw Ultra 12.0, Cambridgesoft [Online]. http://www.cambridgesoft.com/software/ChemDraw/.
(30) HyperChem 8.0.8, Hypercube Inc. [Online]. http://www.hyper.com/.
(31) SYBYL-X 1.2, Tripos [Online]. http://www.tripos.com.
(32) Discovery Studio 3.0, Accelrys Software Inc. [Online]. http://accelrys.com/products.
(33) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, 2000.
(34) Sumathy, K. *"Chemical Descriptors" From QSAR World – A Strand Life Sciences Web Resource* [Online]. http://www.qsarworld.com/insilico-chemistry-chemical-descriptors.php.
(35) Karelson, M.; Lobanov, V.S. *Chem. Rev.* **1996**, *96*, 1027–1043.
(36) Ghasemi, J.B.; Rofouei, M.K.; Salahinejad, M. *J. Inclusion Phenom. Macrocyclic Chem.* [Online] **2010**. http://www.springerlink.com/content/pj0728511421735t (accessed Sep 13, 2010). DOI: 10.1007/s10847-010-9854-9.

(37) Greco, G.; Novellino, E.; Martin, Y.C. *Approaches to Three-Dimensional Quantitative Structure–Activity Relationships*; John Wiley & Sons, Inc.: USA, 2007; pp 183–240.
(38) Oprea, T.I.; Waller, C.L. *Theoretical and Practical Aspects of Three-Dimensional Quantitative Structure–Activity Relationships*; John Wiley & Sons, Inc.: USA, 2007; pp 127–182.
(39) Kubinyi, H. *Comparative Molecular Field Analysis (CoMFA)*; Wiley-VCH Verlag GmbH: Weinheim, 2008; pp 1555–1574.
(40) Kim, K.H.; Greco, G.; Novellinoc, E. *A Critical Review of Recent CoMFA Applications*; Kluwer Academic Publishers: Britain, 1998; pp 257–315.
(41) Klebe, G.; Abraham, U.; Mietzner, T. *J. Med. Chem.* **1994**, *37*, 4130–4146.
(42) Pastor, M.; Cruciani, G.; McLay, I.; Pickett, S.; Clementi, S. *J. Med. Chem.* **2000**, *43*, 3233–3243.
(43) Silverman, B.D. *Quant. Struct.-Act. Relat.* **2000**, *19*, 237–246.
(44) Todeschini, R.; Gramatica, P.; Provenzani, R.; Marengo, E. *Chemom. Intell. Lab. Syst.* **1995**, *27*, 221–229.
(45) Todeschini, R.; Moro, G.; Boggia, R.; Bonati, L.; Cosentino, U.; Lasagni, M.; Pitea, D. *Chemom. Intell. Lab. Syst.* **1997**, *36*, 65–73.
(46) Cruciani, G.; Clementi, S.; Crivori, P.; Carrupt, P.-A.; Testa, B. *VolSurf and its Application in Structure–Disposition Relationships*; Verlag Helvetica Chimica Acta: Zurich, 2007; pp 539–550.
(47) ChemBioOffice Ultra 12.0, Cambridgesoft [Online]. http://www.cambridgesoft.com/software/ChemBioOffice/.
(48) Dragon 6, TALETE, Italy [Online]. www.talete.mi.it/products/dragon.description.htm.
(49) Guassian 09, Guassian Inc. [Online]. http://www.gaussian.com/.
(50) CODSSAPRO; Katritzky, A.R.; Karelson, M.; Petrukhin, R. University of Florida [Online]. www.codessa-pro.com., 2001–2005.
(51) POLLY University of Minnesota, Duluth [Online]. http://www.polysoftware.com/.
(52) Chem-X, Accelrys Software Inc. [Online]. www.accelrys.com.
(53) Tsar, Accelrys Software Inc. [Online]. www.accelrys.com/products/tsar.
(54) Pentacle 1.0.6, Molecular Discovery Ltd [Online]. http://www.moldiscovery.com.
(55) Volsurf 4.1.4, Molecular Discovery Ltd [Online]. http://www.moldiscovery.com.
(56) Michalak, K.; Kwasnicka, H. *Int. J. Bio-Inspired Comp.* **2010**, *2*, 319–332.
(57) Biesiada, J.; Duch, W. Feature Selection for High-Dimensional Data: A Kolmogorov–Smirnov Correlation-Based Filter. In *Computer Recognition Systems*; Kurzynski, M., Puchala, E., Wozniak, M., Zolnierek, A., Eds.; Springer: Berlin/Heidelberg, 2005; Vol. 30, pp 95–103.
(58) Gupta, N. *"Feature Selection" From QSARWorld – A Strand Life Sciences Web Resource* [Online]. http://www.qsarworld.com/qsar-ml-feature-selection.php.
(59) Gao, H.; Lajiness, M.S.; Drie, J.V. *J. Mol. Graph. Model.* **2002**, *20*, 259–268.
(60) Sun, M.; Chen, J.; Cai, J.; Cao, M.; Yin, S.; Ji, M. *Chem. Biol. Drug Des.* **2010**, *75*, 494–505.
(61) Kirkpatrick, S.; Gelatt, C.D.; Vecchi, M.P. *Science* **1983**, *220*, 671–680.
(62) Hardle, W.; Simar, L. *Applied Multivariate Statistical Analysis*, 26th ed.; Springer Verlag: Berlin, 2004.
(63) Geladi, P.; Kowalski, B. *Anal. Chim. Acta* **1986**, *185*, 1–17.
(64) Wold, S.; Sjostrom, M.; Eriksson, L. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130.
(65) Naes, T.; Martens, H. *J. Chemom.* **1988**, *2*, 155–167.
(66) Despagne, F.; Massart, D.L. *Analyst* **1998**, *123*, 157R–178R.
(67) Cheng, B.; Titterington, D.M. *Stat. Sci.* **1994**, *9*, 2–30.
(68) Hart, P.; Cover, T. *IEEE Trans. Inf. Theory* **1988**, *13*, 21–27.
(69) Ajmani, S.; Jadhav, K.; Kulkarni, S.A. *J. Chem. Inf. Model.* **2006**, *46*, 24–31.
(70) Rogers, D.; Hopfinger, A.J. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854–866.
(71) Dunn, W.J., III.; Rogers, D. Genetic Partial Least Squares in QSAR. In *Genetic Algorithms in Molecular Modeling*; Devillers, J., Ed.; Academic Press: London, 1996.
(72) Cortes, C.V.; Varpnik, V. *Mach. Learn.* **1995**, *20*, 273–297.
(73) Ivanciuc, O. Applications of Support Vector Machines in Chemistry. In *Reviews in Computational Chemistry*; Wiley-VCH: Weinheim, 2007; Vol. 23.
(74) He, L.; Jurs, P.C. *J. Mol. Graph. Model.* **2005**, *23*, 503–523.
(75) Katritzky, A.R.; Kuanar, M.; Slavov, S.; Hall, C.D.; Karelson, M.; Kahn, I.; Dobchev, D.A. *Chem. Rev.* **2010**, *110*, 5714–5789.
(76) Lindgren, F.; Hansen, B.; Karcher, W.; Sjöström, M.; Eriksson, L. *J. Chemom.* **1996**, *10*, 521–532.
(77) Wehrens, R.; Putter, H.; Buydens, L. *Chemom. Intell. Lab. Syst.* **2000**, *45*, 35–52.
(78) Schuurmann, G.; Ebert, R.-U.; Chen, J.; Wang, B.; Kühne, R. *J. Chem. Inf. Model.* **2008**, *48*, 2140–2145.
(79) Consonni, V.; Ballabio, D.; Todeschini, R. *J. Chem. Inf. Model.* **2009**, *49*, 1669–1678.
(80) Balaz, S. *Quant. Struct.-Act. Relat.* **1994**, *13*, 381–392.
(81) Kubiny, H. *Quant. Struct.-Act. Relat.* **1994**, *13*, 285–294.
(82) Akaike, H. *IEEE T. Autom. Control* **1974**, *19*, 716–723.
(83) Wager, C.; Vaida, F.; Kauermann, G. *Aust. N Z J. Stat.* **2007**, *49*, 173–190.
(84) Pratim Roy, P.; Paul, S.; Mitra, I.; Roy, K. *Molecules* **2009**, *14*, 1660–1701.
(85) Katritzky, A.R.; Petrukhin, R.; Tatham, D. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 679–685.
(86) McNaught, A.D.; Wilkinson, A. IUPAC: Compendium of Chemical Terminology [Online]. http://goldbook.iupac.org.
(87) Steed, J.W.; Turner, D.R.; Wallace, K.J. *Core Concepts in Supramolecular Chemistry and Nanochemistry*; John Wiley & Sons, Ltd: USA, 2007.
(88) Driggers, E.M.; Hale, S.P.; Lee, J.; Terrett, N.K. *Nat. Rev. Drug Discov.* **2008**, *7*, 608–624.
(89) Fenton, D.E. *Chem. Soc. Rev.* **1998**, *28*, 159–168.
(90) Irwin, G.; Kirk, A.D.; Mackay, I.; Nera, J. *Inorg. Chem.* **2002**, *41*, 874–883.
(91) Johnson, J.M.; Chen, R.; Chen, X.; Moskun, A.C.; Zhang, X.; Hogen-Esch, T.E.; Bradforth, S.E. *J. Phys. Chem. B* **2008**, *112*, 16367–16381.
(92) Kotkar, R.; Srivastava, A. *J. Inclusion Phenom. Macrocyclic Chem.* **2008**, *60*, 271–279.
(93) Bencini, A.; Bianchi, A.; Giorgi, C.; Valtancoli, B. *J. Inclusion Phenom. Macrocyclic Chem.* **2001**, *41*, 87–93.

(94) Dean Sherry, A. *J. Alloys Compd.* **1997**, *249*, 153–157.
(95) Ariga, K.; Kunitake, T. Applications of Supermolecules – Molecular Devices and Nanotechnology. In *Supramolecular Chemistry-Fundamentals and Applications*; Springer: Berlin, 2006; pp 137–174.
(96) Barabanov, V.A.; Davydova, S.L. *Polym. Sci. USSR* **1982**, *24*, 1007–1042.
(97) Hancock, R.D.; Melton, D.L.; Harrington, J.M.; McDonald, F.C.; Gephart, R.T.; Boone, L.L.; Jones, S.B.; Dean, N.E.; Whitehead, J.R.; Cockrell, G.M. *Coord. Chem. Rev.* **2007**, *251*, 1678–1689.
(98) Korybut-Daszkiewicz, B.; Bilewicz, R.; Wozniak, K. *Coord. Chem. Rev.* **2010**, *254*, 1637–1660.
(99) Ward, T.J.; Farris Iii, A.B. *J. Chromatogr. A* **2001**, *906*, 73–89.
(100) Mewis, R.E.; Archibald, S.J. *Coord. Chem. Rev.* **2010**, *254*, 1686–1712.
(101) Comba, P.; Martin, B. Modeling of Macrocyclic Ligand Complexes. In *Macrocyclic Chemistry*; Gloe, K., Ed.; Springer: The Netherlands, 2005; pp 303–325.
(102) Comba, P.; Schiek, W. *Coord. Chem. Rev.* **2003**, *238–239*, 21–29.
(103) Izatt, R.M.; Bradshaw, J.S.; Nielsen, S.A.; Lamb, J.D.; Christensen, J.J.; Sen, D. *Chem. Rev.* **1985**, *85*, 271–339.
(104) Izatt, R.M.; Bradshaw, J.S.; Pawlak, K.; Bruening, R.L.; Tarbet, B.J. *Chem. Rev.* **1992**, *92*, 1261–1354.
(105) Izatt, R.M.; Pawlak, K.; Bradshaw, J.S.; Bruening, R.L. *Chem. Rev.* **1991**, *91*, 1721–2085.
(106) Izatt, R.M.; Pawlak, K.; Bradshaw, J.S.; Bruening, R.L. *Chem. Rev.* **1995**, *95*, 2529–2586.
(107) Rekharsky, M.V.; Mayhew, M.P.; Goldberg, R.N.; Ross, P.D.; Yamashoji, Y.; Inoue, Y. *J. Phys. Chem. B* **1997**, *101*, 87–100.
(108) SC-Database, The IUPAC Stability Constants Database [Online]. http://www.acadsoft.co.uk/.
(109) NIST Critically Selected Stability Constants of Metal Complexes [Online]. http://www.nist.gov/srd/nist46.cfm.
(110) Solov'ev, V.P.; Strakhov, N.N.; Kazachenko, V.P. *The THECOMAC (thermodynamics on complexation of macrocycles) Database*; Institute of Physiologically Active Compounds: Russia, 1991.
(111) Chandler, K.; Culp, C.W.; Lamb, D.R.; Liotta, C.L.; Eckert, C.A. *Ind. Eng. Chem. Res.* **1998**, *37*, 3252–3259.
(112) Leevy, W.M.; Weber, M.E.; Gokel, M.R.; Hughes-Strange, G.B.; Daranciang, D.D.; Ferdani, R.; Gokel, G.W. *Org. Biomol. Chem.* **2005**, *3*, 1647–1652.
(113) Bartsch, R.A.; Ivy, S.N.; Lu, J.; Huber, V.J.; Talanov, V.S. *Pure Appl. Chem.* **1998**, *70*, 2393–23400.
(114) Walkowiaka, W.; Kozlowski, C.A. *Desalination* **2009**, *240*, 186–197.
(115) Richens, D.A.; Simpson, D.; Peterson, S.; McGinn, A.; Lamb, J.D. *J. Chromatogr. A* **2003**, *1016*, 155–184.
(116) Gokel, G.W.; Leevy, W.M.; Weber, M.E. *Chem. Rev.* **2004**, *104*, 2723–2750.
(117) Lehn, J.M. *Supramolecular Chemistry: Concepts and Perspectives*; VCH: Weinheim, 1995.
(118) Szejtli, J. *Chem. Rev.* **1998**, *98*, 1743–1754.
(119) Del Valle, E.M.M. *Process Biochem.* **2004**, *39*, 1033–1046.
(120) Duchene, D.; Wouessidjewe, D. *Drug Dev. Ind. Pharm.* **1990**, *16*, 2487–2499.
(121) Ogoshi, T.; Harada, A. *Sensors* **2008**, *8*, 4961–4982.
(122) Vyas, A.; Saraf, S.; Saraf, S. *J. Inclusion Phenom. Macrocyclic Chem.* **2008**, *62*, 23–42.
(123) Bricout, H.; Hapiot, F.; Ponchel, A.; Tilloy, S.; Monflier, E. *Sustainability* **2009**, *1*, 924–945.
(124) Eastburn, S.D.; Tao, B.Y. *Biotechnol. Adv.* **1994**, *12*, 325–339.
(125) Gramatica, P. A Short History of QSAR Evolution [Online]. www.qsarworld.com/Temp_Fileupload/Shorthistoryofqsar.pdf.
(126) Taft, R.W.; Abboud, J.-L.M.; Kamlet, M.J.; Abraham, M.H. *J. Solution Chem.* **1985**, *14*, 153–186.
(127) Matsui, Y.; Mochida, K. *Bull. Chem. Soc. Jpn.* **1979**, *52*, 2808–2814.
(128) Fellous, R.; Luft, R. *J. Am. Chem. Soc.* **1973**, *95*, 5593–5595.
(129) Lopata, A.; Darvas, F.; Stadler-Szoke, A.; Szejtli, J. *J. Pharm. Sci.* **1985**, *74*, 211–213.
(130) Park, J.H.; Nah, T.H. *J. Chem. Soc. Perkin Trans. 2* **1994**, 1359–1362.
(131) Liu, L.; Li, W.-G.; Guo, Q.X. *J. Inclusion Phenom. Macrocyclic Chem.* **1999**, *34*, 413–419.
(132) Guo, Q.X.; Luo, S.H.; Zheng, X.Q.; Liu, Y.C. *Chin. Chem. Lett.* **1996**, *7*, 767–770.
(133) Guo, Q.X.; Luo, S.H.; Liu, Y.C. *J. Inclusion Phenom. Mol. Recognit. Chem.* **1998**, *30*, 173–182.
(134) Liu, L.; Guo, Q.X. *J. Phys. Chem. B* **1999**, *103*, 3461–3467.
(135) Guo, Q.X.; Liu, L.; Cai, W.S.; Jiang, Y.; Liu, Y.C. *Chem. Phys. Lett.* **1998**, *290*, 514–518.
(136) Guo, Q.X.; Luo, S.H.; Wang, H.; Zhang, M.S.; Liu, Y.C. *J. Chem. Res. (S)* **1996**, *1*, 38–39.
(137) Guo, Q.X.; Luo, S.H.; Wang, H.; Zhang, M.S.; Liu, Y.C. *Chin. Chem. Lett.* **1996**, *7*, 285–288.
(138) Liu, L.; Li, W.-G.; Guo, Q.-X. *J. Inclusion Phenom. Macrocyclic Chem.* **1999**, *34*, 291–298.
(139) Liu, L.; Guo, Q.X.J. *Chem. Inf. Comput. Sci.* **1998**, *39*, 133–138.
(140) Landy, D.; Fourmentin, S.; Salome, M.; Surpateanu, G. *J. Inclusion Phenom. Macrocyclic Chem.* **2000**, *38*, 187–198.
(141) Davies, D.M.; Savage, J.R. *J. Chem. Soc. Perkin Trans. 2* **1994**, 1525–1530.
(142) Davies, D.M.; Deary, M.E. *J. Chem. Soc. Perkin Trans. 2* **1995**, 1287–1294.
(143) Davies, D.M.; Deary, M.E.; Wealleans, D.I. *J. Chem. Soc., Perkin Trans. 2* **1998**, 193–195.
(144) Davies, D.M.; Savage, J.R. *J. Chem. Res. Synop.* **1993**, *3*, 94–95.
(145) Loukas, Y.L. *Int. J. Pharm.* **2001**, *226*, 207–211.
(146) Loukas, Y.L. *Anal. Chim. Acta* **2000**, *417*, 221–229.
(147) Solov'ev, V.P.; Varnek, A.A.; Wipff, G. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 847–858.
(148) Varnek, A.; Wipff, G. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 812–829.
(149) Solov'ev, V.P.; Varnek, A.A. *Russ. Chem. Bull.* **2004**, *53*, 1434–1445.
(150) Solov'ev, V.P.; Kireeva, N.V.; Tsivadze, A.Y.; Varnek, A.A. *J. Struct. Chem.* **2006**, *47*, 298–311.
(151) Katritzky, A.R.; Fara, D.C.; Yang, H.; Karelson, M.; Suzuki, T.; Solov'ev, V.P.; Varnek, A. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 529–541.
(152) Zhokhova, N.; Bobkov, E.; Baskin, I.; Palyulin, V.; Zefirov, A.; Zefirov, N. *Moscow Univ. Chem. Bull.* **2007**, *62*, 269–272.
(153) Presnell, S.R.; Cohen, F.E. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 283–298.

(154) Schwenker, F. *Artificial Neural Networks in Pattern Recognition*, 1st ed.; Springer: Germany, 2010; p 280.
(155) Maddalena, D.J. *Expert Opin. Ther. Pat.* **1996**, *6*, 239–251.
(156) Sumpter, B.G.; Getino, C.; Noid, D.W. *Annu. Rev. Phys. Chem.* **1994**, *45*, 439–481.
(157) Stefanek, G.; Fildes John, M. Expert Systems and Neural Networks in Food Processing. In *Food Safety Assessment*; American Chemical Society: USA, 1992; Vol. 484, pp 166–180.
(158) Tetko, I.V.; Solov'ev, V.P.; Antonov, A.V. *J. Chem. Inf. Model.* **2006**, *46*, 808–819.
(159) Gakh, A.A.; Sumpter, B.G.; Noid, D.W.; Sachleben, R.A.; Moyer, B.A. *J. Inclusion Phenom. Macrocyclic Chem.* **1997**, *27*, 201–213.
(160) Shi, Z.G.; Mccullough, E.A. *J. Inclusion Phenom. Macrocyclic Chem.* **1994**, *18*, 9–25.
(161) Ghasemi, J.; Saaidpour, S. *J. Inclusion Phenom. Macrocyclic Chem.* **2008**, *60*, 339–351.
(162) Ahmadi, S. *J. Inclusion Phenom. Macrocyclic Chem.* [Online] **2010**. http://www.springerlink.com/content/t11804241344p312/ (accessed Oct 19, 2010). DOI: 10.1007/s10847-010-9881-6.
(163) Suzuki, T.; Ishida, M.; Fabian, W. *J. Comput. Aided Mol. Des.* **2000**, *14*, 669–678.
(164) Constantinou, L.; Gani, R. *AlChE J.* **1994**, *40*, 1697–1710.
(165) Suzuki, T. *J. Chem. Inf. Comput. Sci*. **2001**, *41*, 1266–1273.
(166) Samudrala, R.; Moult, J. *J. Mol. Biol.* **1998**, *275*, 895–916.
(167) Carpignano, R.; Marzona, M.; Cattaneo, E.; Quaranta, S. *Anal. Chim. Acta* **1997**, *348*, 489–493.
(168) Khuri, A.I.; Mukhopadhyay, S. *Wiley Interdiscip. Rev.: Comput. Stat.* **2010**, *2*, 128–149.
(169) Bodor, N.; Buchwald, P. *J. Phys. Chem. B* **1997**, *101*, 3404–3412.
(170) Bodor, N.; Buchwald, P. *J. Inclusion Phenom. Macrocyclic Chem.* **2002**, *44*, 9–14.
(171) Buchwald, P. *J. Phys. Chem. B* **2002**, *106*, 6864–6870.
(172) Estrada, E.; Perdomo-López, I.; Torres-Labandeira, J.J. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1561–1568.
(173) Golovanov, I.B.; Zhenodarova, S.M.; Tsygankova, I.G. *Russ. J. Gen. Chem.* **2006**, *76*, 267–271.
(174) Klein, C.T.; Polheim, D.; Viernstein, H.; Wolschann, P. *J. Inclusion Phenom. Macrocyclic Chem.* **2000**, *36*, 409–423.
(175) Klein, C.T.; Polheim, D.; Viernstein, H.; Wolschann, P. *Pharm. Res.* **2000**, *17*, 358–365.
(176) Chari, R.; Qureshi, F.; Moschera, J.; Tarantino, R.; Kalonia, D. *Pharm. Res.* **2009**, *26*, 161–171.
(177) Pérez-Garrido, A.; Helguera, A.M.; Guillén, A.A.; Cordeiro, M.N.D.S.; Escudero, A.G. *Biorg. Med. Chem.* **2009**, *17*, 896–904.
(178) Li, H.; Sun, J.; Wang, Y.; Sui, X.; Sun, L.; Zhang, J.; He, Z. *Eur. J. Pharm. Sci.* **2010**, *42*, 55–64.
(179) Estrada, E. *SAR QSAR Environ. Res.* **2000**, *11*, 55–73.
(180) Pérez-Garrido, A.; Helguera, A.M.; Cordeiro, M.N.D.S.; Escudero, A.G. *J. Pharm. Sci.* **2009**, *98*, 4557–4576.
(181) Prakasvudhisarn, C.; Wolschann, P.; Lawtrakul, L. *Int. J. Mol. Sci.* **2009**, *10*, 2107–2121.
(182) Kennedy, J.; Eberhart, R. Proceedings of IEEE International Conference on Neural Networks. IV, 1995; pp 1942–1948.
(183) Jiménez, V.; Alderete, J.B. *Supramol. Chem.* **2008**, *20*, 317–325.
(184) Verma, J.; Khedkar, V.M.; Coutinho, E.C. *Curr. Top. Med. Chem.* **2010**, *10*, 95–115.
(185) Martin, Y.C. *Perspect. Drug Discov. Des.* **1998**, *12–14*, 3–23.
(186) Broto, P.; Moreau, G.; Vandycke, C. *Eur. J. Med. Chem.* **1984**, *19*, 66–70.
(187) Klein, C.T.; Kaiser, D.; Ecker, G. *J. Chem. Inf. Comput. Sci.* **2003**, *44*, 200–209.
(188) Caron, G.; Ermondi, G. *J. Mol. Graph. Model.* **2007**, *25*, 731–739.
(189) Ghasemi, J.B.; Salahinejad, M.; Rofouei, M.K. *J. Inclusion Phenom. Macrocyclic Chem.* [Online] **2011**. http://www.springerlink.com/content/e213gn56g826p635/ (accessed Sep 5, 2011). DOI: 10.1007/s10847-011-9927-4.
(190) Fronza, G.; Mele, A.; Redenti, E.; Ventura, P. *J. Org. Chem.* **1996**, *61*, 909–914.
(191) Oana, M.; Tintaru, A.; Gavriliu, D.; Maior, O.; Hillebrand, M. *J. Phys. Chem. B* **2001**, *106*, 257–263.
(192) Lipkowitz, K.B. *Chem. Rev.* **1998**, *98*, 1829–1873.
(193) Liu, L.; Guo, Q.X. *J. Inclusion Phenom. Macrocyclic Chem.* **2004**, *50*, 95–103.
(194) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. *Proteins: Struct. Funct. Bioinform.* **2002**, *47*, 409–443.
(195) Issaraseriruk, N.; Shitangkoon, A.; Aree, T. *J. Mol. Graph. Model.* **2010**, *28*, 506–512.
(196) Cai, W.; Xia, B.; Shao, X.; Guo, Q.; Maigret, B.; Pan, Z. *Chem. Phys. Lett.* **2001**, *342*, 387–396.
(197) Choi, Y.; Kim, S.-R.; Hwang, S.; Jeong, K. A Grid Computing-Based Monte Carlo Docking Simulations Approach for Computational Chiral Discrimination. In *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*; Slezak, D., Yao, J., Peters, J.F., Ziarko, W., Hu, X., Eds.; Springer: Berlin, 2005; Vol. 3642, pp 446–455.
(198) Alia, S.M.; Maityb, D.K.; Dea, S.; Shnoia, M.R.K. *Desalination* **2008**, *232*, 181–190.



Quantitative structure–activity relationship (QSAR) methods, are progressive and promising tools in modeling and prediction of many physiochemical properties, in host–guest interactions of macrocyclic complexes.

J.B. Ghasemi, M. Salahinejad and M.K. Rofouei

Review of the quantitative structure–activity relationship modelling methods on estimation of formation constants of macrocyclic compounds with different guest molecules

615–631