J Incl Phenom Macrocycl Chem (2012) 73:405–413
DOI 10.1007/s10847-011-0078-4

ORIGINAL ARTICLE

Docking and 3D-QSAR study of stability constants of benzene derivatives as environmental pollutants with α-cyclodextrin

Jahan B. Ghasemi · M. Salahinejad · M. K. Rofouei · M. H. Mousazadeh

Received: 7 March 2011 / Accepted: 1 November 2011 / Published online: 25 November 2011

Abstract Comparative molecular field analysis region focusing (CoMFA-RF) and VolSurf methods were employed to develop 3D-QSAR models for prediction of stability constants of mono- and 1,4-disubstituted benzenes with α-Cyclodextrin. The combination of CoMFA fields with some physicochemical descriptors ultimate to a more predictive model. We applied two effective feature selection techniques, genetic algorithm (GA) and successive projection algorithm (SPA), to extract more informative VolSurf descriptors. Partial least square and support vector machine (SVM) were used to model construction and SPA–SVM based VolSurf descriptors showed excellent performance in predicting stability constants. The predictive ability of modified CoMFA-RF and VolSurf models were determined using a test set of 18 compounds result in correlation coefficients of 0.604 and 0.889 respectively. For further model validation, the cross validation (leave one out), progressive scrambling and bootstrapping were also applied. Results of both methods showed that electrostatic and hydrophobic effects and shape parameters are main influencing factors in inclusion complexation of benzenes derivatives with α-Cyclodextrin. The results of docking study, which can predict the binding mode and orientation of guest molecules in Cyclodextrin cavity, are in agreement with of combined 3D-QSAR models results.

J. B. Ghasemi (✉)
Chemistry Department, Faculty of Sciences, K.N. Toosi University of Technology, Tehran, Iran
e-mail: Jahan.ghasemi@gmail.com

M. Salahinejad · M. K. Rofouei
Faculty of Chemistry, Tarbiat Moallem University, Tehran, Iran

M. H. Mousazadeh
Environmental Laboratory, Nuclear Science Research School, Nuclear Science & Technology Research Institute, Tehran, Iran

Keywords 3D-QSAR · α-Cyclodextrin · Benzene derivatives · Docking · Environmental pollutants

Introduction

Cyclodextrins (CD), cyclic oligomers derived from starch, have attracted tremendous interest in many different fields of host:guest applications [1]. The most common natural CDs are α-CD, β-CD and γ-CD, consisting of six, seven and eight [1–4]-linked glucose units, respectively. The CDs all form doughnut-shaped molecules with their hydroxyl groups on the outside of the molecule and a relatively nonpolar and hydrophobic cavity in the middle, which can encapsulate variety of guest molecules to form noncovalent inclusion complexes [2]. Because of their structural features, CDs have been used to improve properties such as solubility, dissolution rate, chemical and physical stability and bioavailability of poorly water-soluble compounds [3]. Therefore, the CDs are the most suitable host molecules for the recognition of hydrophobic guest molecules, such as drugs, dyes, detergents, pesticides and etc. in aqueous media in a wide range of applications in industrial, pharmaceutical, agricultural and other fields [1, 4–7].

The value of stability (association) constants of host:guest complexes, as an index of the binding strength of the inclusion complexes, are of great importance for the understanding and evaluation of the inclusion complexes formation [8]. The major interactions that have been put forward to account for the stabilities of CD inclusion complexes in aqueous solution were: release of 'high-energy' water from the CD cavity, relief of conformational strain energy possessed by the uncomplexed CD, the hydrophobic interaction, electrostatic interactions, hydrogen bonding, induction forces and the London dispersion

force [9, 10]. CDs complex formation usually results from different combinations of these forces. One approach used for addressing the contribution each of these forces makes toward complexation is to rely on quantitative structure-activity relationships (QSAR) [11]. QSAR, as a methodology that allows cost savings by reducing the laboratory resources needed and the time required to create and investigate new compounds with better complexing profile, permits one to separate one type of interaction from another and to do so quantitatively.

Widespread use of benzene derivatives in chemical industry as solvents, cooling agents, polymers, insecticides and herbicides, cause releasing of these chemical into the environment. Benzene derivatives have preferential sorption on soils and removal of them is a very difficult task [12]. As a result, adding facilitating agents such as surfactants are necessary to enhance desorption of these compounds from soils. CDs, as enhancing-solubility compounds, that negligibly adsorbed by soils [13, 14] have several advantages such as their biodegradability and their non-toxicity over solvents and surfactants [15, 16]. It's showed that the presence of CDs can significantly reduce biotoxicities of the low-polarity compounds, such as substituted benzene compounds and pesticides [17].

The inclusion complexes of CDs with the less flexible molecules, which can only take up a limited number of conformations within the CD cavity, can give a deeper understanding of host:guest interactions of CD inclusion complexation [18]. Thus, the binding of benzenes derivatives within α-CD cavity, due to the effect of the tightly fitting benzene ring offers a good system in which to investigate factors influencing inclusion complexation of CDs.

Different QSAR/QSPR approaches including linear solvation energy relationship (LSER) [19–21], neural networks [22, 24] and 2D-QSAR [25, 26] have been applied to elucidate the most important factors influencing the benzenes derivatives interactions within α-CD cavity and to predict the thermodynamic stability of their inclusion complexes. But to our knowledge, just one publication in 3D-QSAR (CoMFA) of complexation of some natural and modified CDs with guest molecules was previously reported [29].

Herein, we report a 2D-3D-QSAR analysis, performed by using the CoMFA and VolSurf methods with some modifications, for prediction of association constants of α-CD with mono- and 1,4-disubstituted benzenes. We add some physicochemical descriptors(as 2D descriptors) to COMFA fields (as 3D descriptors) and apply two efficient variable selection techniques, successive projections algorithm and genetic algorithm to select the best VolSurf based descriptors to improve models predictability. Molecular docking study also employed to find prevailing binding mode and preferred orientation of benzene derivatives in CD cavity and mechanism of encapsulation as a whole.

Materials and methods

Data set

The data sets of 72 mono- and 1,4-disubstituted benzenes were extracted from the work of Guo et al. [27] and the experimental data to be predicted are the α-CD complex stability constants (ln K_a) in water at 298 K taken from references therein. Table 1 displays a complete list of the chemicals accompany with the reported experimental data.

Subset selection and model validation

A statistical subset selection was made using largest minimum distance (LMD) method to select a balanced and chemically diverse test set. The LMD approach is based on extract a number of compounds by maximizing their mutual distances [28]. The training set of 54 molecules was used to adjust the parameters of the models, and the rest of molecules were used to evaluate models prediction ability. The obtained models were externally validated with a test set of compounds, which were not considered for QSAR model generation (test set). The progressive scrambling method (maximum: 8 bins, minimum: 2 bins and critical point: 0.85) [29], randomization y-test [30] and bootstrapping [31] were carried out for the evaluation of the sensitivity to chance correlations and prediction ability of the 3D-QSAR models.

Molecular optimization

The three dimensional structure of the guest molecules was constructed using the standard tools available in SYBYL 7.3 molecular modeling package (Tripos Inc., St. Louis, USA) running on a Red Hat Linux workstation 4.7. Energy minimization performed using the Tripos force field with a distance dependent dielectric and the Powell conjugate gradient algorithm with a convergence criterion of 0.01 kcal/mol Å. Partial atomic charges were calculated using the Gasteiger-Hückel method.

VolSurf approach

VolSurf is a computational procedure that generates useful quantitative 2D descriptors from the 3D maps of molecular interaction field (MIF) between different probes and all the atoms in a target molecule. The basic concepts and a detailed explanation of VolSurf approach have been demonstrated elsewhere [32, 33]. In the present study, VolSurf$^+$ 1.0.4 (Molecular Discovery Ltd., Oxford, UK) with four probes including the water (OH2), hydrophobic (DRY), and an H-bond donor (NH) and an H-bond

Table 1 Names, experimental and calculated stability constants (ln K_a) of mono and 1,4-dissubstituted benzene derivatives

No.	X	Y	ln $K_{a\ (exp)}$	ln $K_{a\ (CoMFA-2)}$		ln $K_{a\ (Volsurf-SPA-SVM)}$	
				Pred.	Res.	Pred.	Res.
M01	F	H	3.68	3.38	−0.30	3.67	−0.01
M02	Cl	H	4.72	4.65	−0.07	4.44	−0.28
M03[a]	Br	H	6.29	5.76	−0.53	5.36	−0.93
M04	I	H	7.09	6.25	−0.84	6.27	−0.82
M05	F	F	2.96	3.29	0.33	2.97	0.01
M06[a]	Cl	Cl	5.42	6.02	0.60	5.06	−0.36
M07	Br	Br	6.93	6.69	−0.24	6.94	0.01
M08[a]	I	I	8.34	7.58	−0.76	8.14	−0.20
M09	Cl	F	4.17	5.17	1.00	4.63	0.46
M10	Br	F	5.52	5.65	0.13	5.79	0.27
M11	I	F	6.89	6.02	−0.88	6.75	−0.14
M12	OCH_3	OCH_3	4.02	4.12	0.10	4.01	−0.01
M13	Oet	Oet	4.85	5.26	0.41	5.25	0.40
M14	CO_2Me	CO_2Me	6.14	6.10	−0.04	4.23	−1.91
M15	COMe	COMe	2.32	2.00	−0.32	3.29	0.97
M16	CN	CN	3.5	3.51	0.00	3.51	0.01
M17	NO_2	NO_2	3.58	4.34	0.76	4.32	0.74
M18[a]	COOH	COOH	7.2	6.83	−0.37	8.26	1.06
M19	NH_2	NH_2	0.83	0.90	0.07	0.84	0.01
M20	OCH_3	NH_2	1.9	2.11	0.21	3.41	1.51
M21	CH_3	NH_2	4.05	3.91	−0.14	4.04	−0.01
M22	SCH_3	OCH_3	4.7	4.80	0.10	4.17	−0.53
M23	SCH_3	CH_2OH	4.44	4.45	0.01	4.47	0.03
M24	SCH_3	Br	5.74	5.77	0.03	5.73	−0.01
M25	SCH_3	NO_2	4.81	3.55	−1.26	4.48	−0.33
M26[a]	SCH_3	Cl	5.04	5.44	0.40	5.10	0.06
M27	SCH_3	$COCH_3$	2.2	2.28	0.08	4.17	1.97
M28	SCH_3	CH_3	3.71	4.73	1.02	4.30	0.59
M29	SCH_3	NH_2	4.62	3.79	−0.84	4.61	−0.01
M30[a]	CH_3	H	3.6	3.86	0.26	3.80	0.20
M31[a]	H	H	3.35	3.49	0.14	3.18	−0.17
M32	Et	H	4.7	4.82	0.12	4.85	0.15
M33	CH_3	CH_3	4.28	4.90	0.62	4.28	0.00
M34	n-Pr	H	6.38	6.21	−0.17	5.78	−0.60
M35	i-Pr	H	4.65	5.08	0.43	5.23	0.58
M36[a]	OCH_3	H	4.95	4.16	−0.79	4.24	−0.71
M37	OEt	H	5.14	4.75	−0.39	5.02	−0.12
M38	CH_2OH	H	4.57	4.93	0.36	3.84	−0.73
M39	CH_2Cl	H	5.32	5.09	−0.23	5.33	0.01
M40	CHO	H	4.62	4.91	0.29	4.63	0.01
M41	COMe	H	4.94	5.17	0.23	4.58	−0.36
M42	CO_2Me	H	5.36	5.26	−0.10	5.32	−0.04
M43	CO2Et	H	5.89	6.32	0.43	5.90	0.01
M44	CN	H	4.36	4.97	0.61	4.35	−0.01
M45	NH_2	H	2.68	2.77	0.09	3.41	0.73
M46	Cl	NH_2	5.52	4.89	−0.63	5.51	−0.01
M47[a]	COOH	NH_2	7.2	5.26	−1.94	7.95	0.75

Table 1 continued

No.	X	Y	ln $K_{a\ (exp)}$	ln $K_{a\ (CoMFA-2)}$		ln $K_{a\ (Volsurf-SPA-SVM)}$	
				Pred.	Res.	Pred.	Res.
M48[a]	CN	NH_2	6.11	2.90	−3.21	4.79	−1.32
M49	NO_2	NH_2	6.45	5.51	−0.94	5.85	−0.60
M50	CH_2NH_2	H	2.86	3.18	0.32	2.85	−0.01
M51[a]	$CH_2CH_2NH_2$	H	3.27	2.37	−0.91	2.85	−0.42
M52[a]	I	OH	7.75	7.38	−0.37	7.53	−0.22
M53	Cl	OH	5.6	6.05	0.45	5.59	−0.01
M54	Br	OH	6.65	6.70	0.05	6.64	−0.01
M55[a]	COOH	OH	7.03	6.10	−0.93	7.06	0.03
M56	CN	OH	5.06	4.81	−0.25	5.07	0.01
M57[a]	NO_2	OH	5.5	6.05	0.55	5.81	0.31
M58[a]	COOH	$NHCH_3$	7.17	5.70	−1.47	7.29	0.12
M59	COOH	OCH_3	6.78	6.32	−0.46	6.78	0.00
M60	COOH	CH_3	6.99	6.49	−0.50	6.98	−0.01
M61[a]	COOH	H	6.53	6.32	−0.22	6.53	0.00
M62	COOH	F	6.22	6.18	−0.04	6.21	−0.01
M63[a]	COOH	CH_3CO	6.8	4.60	−2.20	6.14	−0.66
M64	COOH	CN	6.15	5.97	−0.18	6.00	−0.15
M65	COOH	NO_2	5.86	6.73	0.87	5.87	0.01
M66	NHEt	H	4.85	5.06	0.21	4.84	−0.01
M67	NHMe	H	4.42	4.46	0.04	4.52	0.10
M68	NMe_2	H	5.15	4.61	−0.54	4.20	−0.95
M69[a]	NHCOMe	H	4.63	3.91	−0.73	4.74	0.11
M70	CCH	H	4.46	4.30	−0.16	4.45	−0.01
M71	OH	H	3.7	3.57	−0.13	4.20	0.50
M72	NO_2	H	4.49	4.76	0.27	4.81	0.32

[a] Prediction set

acceptor (=O) was used to produce ninety-six Volsurf descriptors (grid spacing 0.5 Å).

Two variable selection techniques namely successive projections algorithm (SPA) [34] and genetic algorithm (GA) [35], run in the MATLAB (version 7.6.0., Math Works, Inc.), were used to extract the more informative VolSurf descriptors and generate more predictive model. The GA was carried out during 200 generations with 30 chromosomes in 1000 runs with probability of mutation 1% and probability of cross-over 50%.

Model construction

Support vector regression (SVR), which is based on support vector machines (SVMs) [36] and partial least square (PLS) [37] are used to predict the stability constant of inclusion complexation of benzene derivatives with α-CD. After variable selection process, analysis was performed using MATLAB and the PLS_ toolbox 5.8.2 (Eigenvector Research, Inc., Manson WA).

CoMFA fields calculations

The initial CoMFA models were calculated using the SYBYL7.3 molecular modeling package. The common fragment, benzene ring, of data set produced by Distill program (without including bond types in rings) in SYBYL was selected for rigid automatic alignment and the most active compound (M08) used as the template for superimposition. The aligned molecules are displayed in Fig. 1 CoMFA models are greatly sensitive to the different space orientations of the molecular collective with respect to the grid box, so all-orientation search (AOS) was also carried out on initial orientations of aligned structures by the rotation procedure written in SYBYL programming language (SPL) [38]. The steric (Lennard-Jones) and electrostatic (Coulomb) interaction energies for each molecule were calculated using a sp^3 carbon atom and a +1 charge as steric and electrostatic probes, respectively. The default value of 30 kcal/mol was set as the maximum steric and electrostatic energy cutoff. Region focusing is an iterative

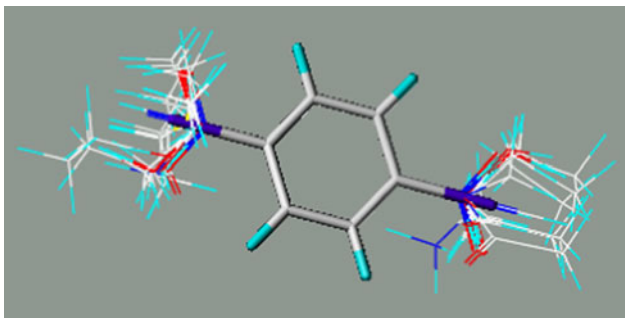

Fig. 1 Aligned compounds with highlighted of template molecule for superimposition

procedure which refines a model by improving the weight for those lattice points which are most related to the model. This enhances the resolution and predictive capability (q^2; cross validated r^2) of followed PLS analysis. Technically, this corresponds to rotate the model components during a high-order space [39, 40]. The best model of AOS procedure was picked up, then the effects of changing column filtering, cut off value and grid spacing were investigated.

Molecular docking

The structure of α-CD was extracted from crystallographic structure of α-CD-p-bromophenol complex obtained from Ref. [41]. The optimized structures of benzene derivatives, as ligands, were transferred into Discovery Studio 2.5 (Accelrys Inc, San Diego, CA, USA) and typed with CHARMm force field [42]. For preparation of α-CD, defined as a receptor, the complex typed with CHARMm force field, hydrogen atoms were added and all water molecules removed. CDOCKER, a molecular dynamics (MD) simulated-annealing-based algorithm [43], based on default protocol settings used to dock benzene derivatives into the α-CD cavity.

Results and discussions

Cyclodextrins inclusion complexation is a composite process of different kind of steric, electrostatic, van der Waals and hydrophobic interactions. Shape and preorganization of guest within the host molecule, the size-match of the host cavity to the guest, and solvation dependencies appears to play important roles in stability of complexation of the host:guest inclusion [44]. The fit of the entire or at least a part of the guest molecule in the cyclodextrin cavity determines the stability of the inclusion complex and the selectivity of the complexation process [45]. It is widely believed that 3D descriptors should provide better descriptions of the binding interactions in host:guest chemistry. VolSurf descriptors which derived from the 3D MIF of a molecule are insensitive to alignment with clear chemical meaning which quantitatively characterize size, shape, polarity, and hydrophobicity of molecules wheres CoMFA calculate steric and electrostatic interactions of aligned molecular structure.

VolSurf analysis

The multivariate data analysis of the ln K_a and a complete set of VolSurf descriptors was carried out by PLS routine of VolSurf. The PLS analysis resulted in five-latent-variables (LV) model with an $R^2_{(cal)} = 0.775$; $q^2 = 0.443$ and $R^2_{(pred)} = 0.688$. Table 2 represents statistical parameters of VolSurf models. No variable selection was implemented in the VolSurf program therefore two variable selection techniques; successive projections algorithm (SPA) and genetic algorithm (GA) were used to extract more informative descriptors and build more predictive models.

As it is obvious from Table 2, employing of both SPA and GA increased the prediction ability but SPA method give more predictive models. This may be explained by the fact that GA procedure does not take into account multicollinearity effects in the variable selection process, whereas SPA was designed to minimize such effects. As SVM analysis taking into accounts both linearity and nonlinearity in model construction, therefore it produces more predictive models than PLS analysis. The SPA-SVM analysis resulted in a model with five variables and an $R^2_{(cal)} = 0.839$; $q^2 = 0.681$and $R^2_{(pred)} = 0.889$. The predicted and residual values of ln K_a calculated by SPA-SVM presented in Table 1.

The five most significant VolSurf descriptors extracted by SPA feature selection procedure and used as SVM input are: the percentage of unionized species calculated at pH

Table 2 Summary of statistical results of VolSurf descriptor based models

Methods	R^2_{cal}	q^2	R^2_{pred}	SEE	SEP	No. of PC
VolSurf	0.775	0.443	0.688	0.649	1.021	5
SPA-SVM	0.839	0.681	0.889	0.558	0.568	–
SPA-PLS	0.772	0.603	0.859	0.652	0.715	5
GA-SVM	0.753	0.643	0.807	0.698	0.765	–
GA-PLS	0.752	0.58	0.772	0.681	0.778	5

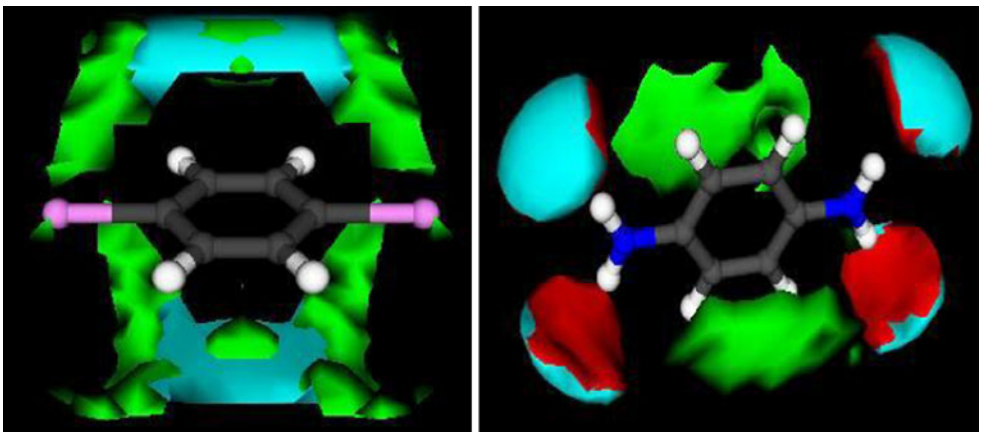

Fig. 2 GRID molecular interaction fields (MIFs) around **a** most active and **b** most inactive compounds

Table 3 Summary of the statistical results for CoMFA models

Statistical parameters	CoMFA-1	CoMFA-2
q^2	0.248	0.647
R^2_{cal}	0.701	0.881
R^2_{pred}	0.393	0.604
SEE	0.802	0.517
SEP	1.222	0.988
F-ratio	18.43	41.65
Component	6	8

4(%FU4), molecular weight (MW), solubility profiling coefficients (L4lgS) that represent the shape of the solubility profile curve, DRDRAC and DRACDO pharmacophoric descriptors that refer to the Dry–Dry-acceptor triplet and Dry-Acceptor–Donor triplet respectively. %FU4 is a charge state descriptor and MW is a structural descriptor represents the size and shape of the guest molecules. DRDRAC and DRACDO are maximum area of the triangles derived from Dry, H-bond acceptor and H-bond donor points in a molecule. These descriptors mainly represent the influence of the shape, size, electrostatic effects and hydrophobicity of guest molecules on stability constant of inclusion complexation of benzene derivatives with α-CD. Figure 2 shows the GRID molecular interaction fields (MIFs) around most active (diiodobenzene) and most inactive (diaminobenzene) compounds. Cyan color represent OH_2 probe with energy level of −1 kcal/mol and green color showed the DRY probe with energy level of −0.86 kcal/mol. As can be seen from Fig. 2, for the most active compound (M08) cyan color indicates regions favorable with hydrophobic effects in para position of benzene derivatives which caused increasing the stability constants of α-CD complexation. For the most inactive compounds, diaminobenzene, red color represent O probe that refers to H-bond acceptor, can be seen and indicate unfavorable region with hydrogen bond acceptor substituents. It could be resulted that the presences of hydrophobic group in benzene ring (mono- or 1,4-disubstituted) enhanced the stability constant whereas the presence of H-bonding groups decrease the stability constants of inclusion complexes of α-CD with benzene derivatives.

CoMFA analysis

The results of CoMFA analysis are summarized in Table 3. With a column filtering of 2 kcal/mol and cutoff value 30 kcal/mol for both steric and electrostatic fields, COMFA region focusing (COMFA-1) yielded a leave-one-out (LOO) cross validated $q^2 = 0.248$ with 6 components, non-cross-validated or calibration, $R^2_{(\text{cal})}$ of 0.701 with standard error of estimation, SEE = 0.802. To gain a more predictive model, some physicochemical descriptors were employed besides COMFA fields, as the CoMFA did not consider the hydrophobicity interactions. For this purpose we used octanol/water partition coefficient (clogP), the component of the dipole moment across y axis (Dip Y) and some charged partial surface area (CPSA) descriptors, including relative negative charge (RNCG) and total charge weighted of partial negative surface area/total area (FNSA2), all calculated by SYBYL. The statistical results of this modified COMFA model is presented in Table 3. This yielded a model (COMFA-2) with a $q^2 = 0.647$ with 8 components, $R^2_{(\text{cal})}$ of 0.881 with standard error of estimation SEE = 0.517 and $R^2_{(\text{pred})} = 0.604$ with standard error of prediction SEP = 0.988.

Figure 3 displayed contour maps of CoMFA-region focusing based on compound M08. The steric interactions are represented by green- and yellow-colored contours, while electrostatic interactions are represented by red- and blue-colored contours. The large size of sterically unfavored yellow region around para position of benzene rings indicated that the steric effects have important effects on

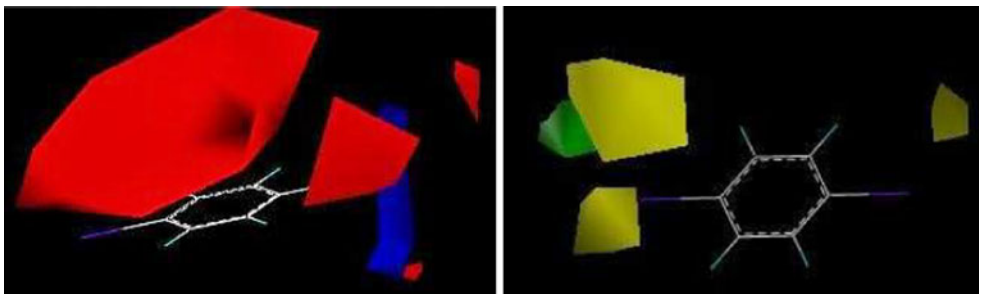

Fig. 3 Contour maps of CoMFA-region focusing based on compound M08

benzene derivatives complexation with α-CD. It could be clearly seen in M35 and M34 compounds (Table 1) in which the steric hindrance of isopropyl substitute caused a reduction of about 2 units in ln K_a with respect to propyl benzene.

Blue-colored contours represent regions where positive charge increases stability, whereas red-colored regions represent areas where negative charge enhances stability. The large red region across benzene ring indicates that stability of inclusion complexes will be enhanced by electropositive group at the mentioned positions. It's obvious from Table 1, electropositive group in benzene ring enhanced the stability complexation of α-CD with benzene derivatives, e.g., stability constant increase in halogenated benzene by increasing electropositivity of halogen group from fluoride to iodo substituent in benzene ring.

In order to check the reliability of the obtained CoMFA model to chance correlation we have used the progressive scrambling method. The obtained values of the sensitivity to the perturbation $d_q^{2'}/dr_{yy'}^2$, the prediction and cSDEP produced by a progressive scrambling analyses were 1.103, 0.897 and 0.401 for the CoMFA model respectively. These values confirm the robustness and independent of chance correlation of the model. Bootstrapping analysis was carried out with 100 runs to evaluate the statistical confidence limits of the derived model. An R_{bs}^2 (average correlation coefficient for bootstrapping) of 0.929 ± 0.034 and a SEE_{bs} (average standard error of estimate for bootstrapping) of 0.381 ± 0.224 suggested a good internal consistency of the model.

Docking study

Molecular docking is a rational method which predicts the predominant binding modes and preferred orientation of one molecule to a second when bound to each other to form a stable complex [46]. Figure 4 shows the docked structures of the most, diiodobenzene (M08), and the least, diaminobenzene (M19), stable and a typical compound with two different groups, p-iodophenol (M52), and inclusion complexes of α-CD. As previous studies are shown [2, 8], electron-withdrawing, hydrophobic and larger size groups prefer to locate near the narrower side (primary rim) of CD, while the wider side (secondary hydroxyl rim) is predominant position for electron-releasing, hydrophilic and smaller size groups. It seems diiodobenzene with two large and hydrophobic groups penetrates deeply in CD cavity thus it forms a more stable inclusion complex but p-diaminobenzene with two relatively small and hydrophilic groups tilts as amino groups stay nearly outside the cavity therefore it forms less stable inclusion complexe with α-CD. For a compound like p-iodo phenol (Fig. 4c) with two different groups, more hydrophobic iodo group gets into the cavity and the hydrophilic hydroxyl group stays outside the cavity or stay near the secondary hydroxyl rim.

Conclusions

3D-QSAR models can give different kinds of information in prediction of thermodynamic parameters. Two 3D-QSAR methods were employed to predict stability constants of inclusion complexes of mono- and 1,4-disubstituted benzenes with α-CD. VolSurf, a free alignment procedure, show a good predictive model, however, by considering major selected VolSurf descriptors of SPA approach and SVM analysis a model was obtained that showed excellent performance in predicting ln K_a. Combination of some physicochemical descriptors and CoMFA fields yielded a predictive model of α-CD complexation with benzene derivatives. Results of both methods showed that hydrophobic and electrostatic effects and shape parameters are main influencing factors in inclusion complexation of CDs. Docking study, which can predict the orientation of guest molecules in CD cavity, in agreement of CoMFA and VolSurf methods, showed that more hydrophobic groups gets into the CD cavity while hydrophilic groups preferred to stay outside the cavity. It's concluded that, electrostatic and hydrophobic interactions, by considering steric hindrance, are the mainly driving forces for α-CD complexation. The obtained 3D-QSAR models are superior to previously reported models [19–27]

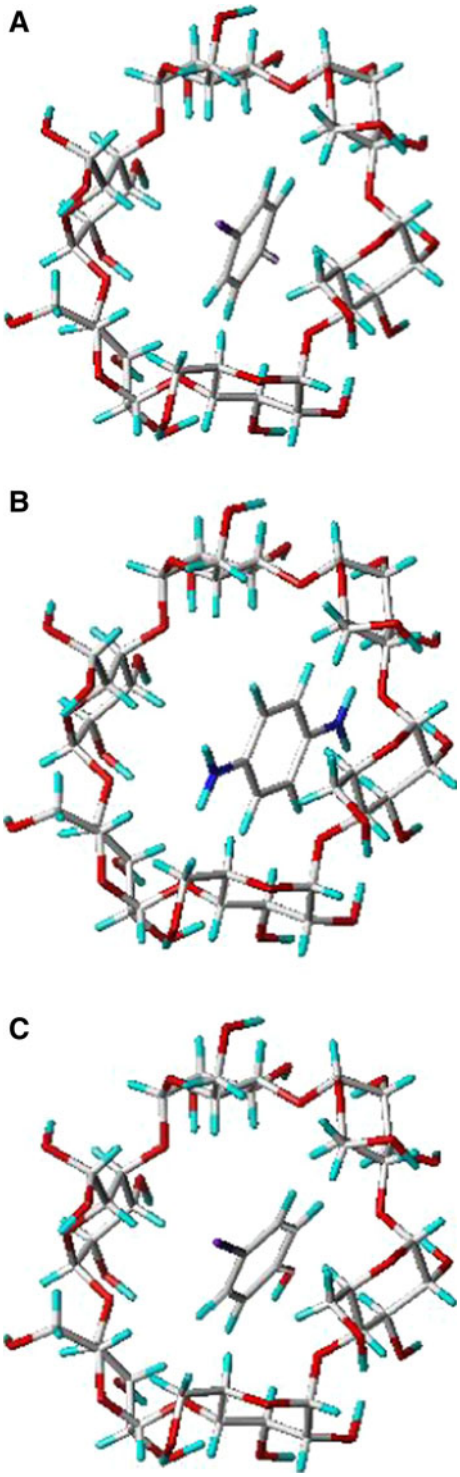

Fig. 4 Docked structures of **a** the most active compound (M08), **b** the most inactive compound (M19) and **c** a typical compound with two different groups

due to its combination of quality and information of inclusion complexation processes. Such studies help to better understanding of the mechanisms of CDs inclusion complexes with benzene derivatives and thus lead to improve removal efficiency of as such environmental pollutants.

Acknowledgments We gratefully appreciate to professor S. Kamitori (Rare Sugar Research Center and Faculty of Agriculture, Kagawa University, Japan) for granting permission to access the Crystallographic Information File (CIF) format of crystallographic structure of α-CD-p-bromophenol complex.

References

1. Valle, D.E.: Cyclodextrins and their uses: a review. Process Biochem **39**, 1033–1046 (2004)
2. Szejtli, J.: Introduction and general overview of cyclodextrin chemistry. Chem. Rev. **98**, 1743–1754 (1998)
3. Loethen, S., Kim, J., Thompson, D.H.: Biomedical applications of cyclodextrin based polyrotaxanes. Polym. Rev. **47**, 383–418 (2007)
4. Loftsson, T., Duchene, D.: Cyclodextrins and their pharmaceutical applications. Int. J. Pharm. **329**, 1–11 (2007)
5. Ogoshi, T., Harada, A.: Chemical sensors based on cyclodextrin derivatives. Sensors **8**, 4961–4982 (2008)
6. Shpigun, O.A., Ananieva, I.A., Budanova, N.Y., Shapovalova, E.N.: Use of cyclodextrins for separation of enantiomers. Russ. Chem. Rev. **72**, 1035–1048 (2003)
7. Vyas, A., Saraf, S.: Cyclodextrin based novel drug delivery systems. J. Inc. Phenom. Macro. Chem. **62**, 23–42 (2008)
8. Connors, K.A.: The stability of cyclodextrin complexes in solution. Chem. Rev. **97**, 1325–1358 (1997)
9. Astray, G., Gonzalez-Barreiro, C., Mejuto, J.C., Rial-Otero, R., Simal-Gándara, J.: A review on the use of cyclodextrins in foods. Food Hydrocolloid **23**, 1631–1640 (2009)
10. Tirapegui, C., Jara, F., Guerrero, J., Rezende, M.C.: Host–guest interactions in cyclodextrin inclusion complexes with solvatochromic dyes. J. Phys. Org. Chem. **19**, 786–792 (2005)
11. Dudek, A.Z., Arodz, T., Gálvez, T.: Computational methods in developing quantitative structure-activity relationships (QSAR): a review. Comb. Chem. High Throughput Screen **9**, 213–228 (2006)
12. Viglianti, C., Hanna, K., de Brauer, C., Germain, P.: Removal of polycyclic aromatic hydrocarbons from aged-contaminated soil using cyclodextrins: experimental study. Environ. Pollut. **140**, 427–435 (2006)
13. Ko, S.O., Schlautman, M.A., Carraway, E.R.: Partitioning of hydrophobic organic compounds to hydroxypropyl-β- cyclodextrin: experimental studies and model predictions for surfactant-enhanced remediation applications. Environ. Sci. Technol. **33**, 2765–2770 (1999)
14. Sheremata, T.W., Hawari, J.: Cyclodextrins for desorption and solubilization of 2,4,6-trinitrotoluene and its metabolites from soil. Environ. Sci. Technol. **34**, 3462–3468 (2000)
15. Crini, G.: Recent developments in polysaccharide-based materials used as adsorbents in wastewater treatment. Prog. Polym. Sci. **30**, 38–70 (2005)
16. Viglianti, C., Hanna, K., De Brauer, C., Germain, P.: Use of cyclodextrins as an environmentally friendly extracting agent in organic aged-contaminated soil remediation. J. Incl. Phenom. Macro. Chem. **56**, 275–280 (2006)
17. Wen-lu, S., Qing-guo, H., Lian-sheng, W.: β-cyclodextrin (β-CD) influence on the biotoxicities of substituted benzene compounds and pesticide intermediates. Chemosphere **38**, 693–698 (1999)
18. Sakurai, M., Kitagawa, M., Hoshi, H., Inoue, Y., Chujo, R.: A molecular orbital study of cyclodextrin (cyclomalto-oligosaccharide)

inclusion complexes. III, dipole moments of cyclodextrins in various types of inclusion complexes. Carbohydr. Res. **198**, 181–191 (1990)
19. Liu, L., Li, W.G., Guo, Q.X.: Regression analysis for the host–guest interaction of α-cyclodextrin with mono- and 1,4-disubstituted benzenes. J. Incl. Phenom. Macro. Chem. **34**, 413–419 (1999)
20. Liu, L., Guo, Q.X.: Novel prediction for the driving force and guest orientation in the complexation of β-and α-cyclodextrin with benzene derivatives. J. Phys. Chem. B **103**, 3461–3467 (1999)
21. Guo, Q.X., Luo, S.H., Liu, Y.C.: Substituent effects on the driving force for inclusion complexation of β-and α-cyclodextrin with monosubstituted benzene derivatives. J. Incl. Phenom. Macro. Recog. Chem **30**, 173–182 (1998)
22. Guo, Q.X., Liu, L., Cai, W.S., Jiang, Y., Liu, Y.: Driving force prediction for inclusion complexation of α-cyclodextrin with benzene derivatives by a wavelet neural network. Chem. Phys. Lett. **290**, 514–518 (1998)
23. Liu, L., Li, W.G., Guo, Q.X.: Association constant prediction for the inclusion of α-cyclodextrin with benzene derivatives by an artificial neural network. J. Incl. Phenom. Macro. Chem. **34**, 291–298 (1999)
24. Liu, L., Guo, Q.X.: Wavelet neural network and its application to the inclusion of α-cyclodextrin with benzene derivatives. J. Chem. Inf. Comput. Sci. **39**, 133–138 (1999)
25. Golovanov, I.B., Zhenodarova, S.M., Tsygankova, I.G.: Quantitative structure-property relationship: xxvii. estimation of the free energy of formation of host: guest complexes of α-cyclodextrin with benzene derivatives. Russ. J. Gen. Chem. **76**, 267–271 (2006)
26. Estrada, E., Perdomo-Lopez, I., Torres-Labandeira, J.J.: Combination of 2D-, 3D-connectivity and quantum chemical descriptors in QSPR. Complexation of β- and α-cyclodextrin with benzene derivatives. J. Chem. Inf. Comput. Sci. **41**, 1561–1568 (2001)
27. Suzuki, T., Ishida, M., Fabian, W.: Classical QSAR and comparative molecular field analyses of the host-guest interaction of organic molecules with cyclodextrins. J. Comput. Aided Mol. Design **14**, 669–678 (2000)
28. Marengo, E., Todeschini, R.: A new algorithm for optimal, distance based experimental design. Chemometr. Intell. Lab. Syst. **1**, 37–44 (1992)
29. Clark, R.D., Fox, P.C.: Statistical variation in progressive scrambling. J. Comput. Aided Mol. Design **18**, 563–576 (2004)
30. van der Voet, H.: Comparing the predictive accuracy of models using a simple randomization test. Chemometr. Intell. Lab. Syst. **25**, 313–323 (1994)
31. Cramer, R.D., Patterson, D.E., Bunce, J.D.: Comparative molecular-field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J. Am. Chem. Soc. **110**, 5959–5967 (1988)
32. Cruciani, C., Crivori, P., Carrupt, P.A., Testa, B.: Molecular fields in quantitative structure-permeation relationships: the VolSurf approach. Theochem **503**, 17–30 (2000)
33. Crivori, P., Cruciani, G., Carrupt, P.A., Testa, B.: Predicting blood-brain barrier permeation from three-dimensional molecular structure. J. Med. Chem. **43**, 2204–2216 (2000)
34. Araújo, M.C.U., Saldanha, T.C.B., Galvão, R.K.H., Yoneyama, T., Chame, H.C., Visani, V.: The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. Chemometr. Intell. Lab. Syst. **57**, 65–73 (2001)
35. Yasri, A., Hartsough, D.: Toward an optimal procedure for variable selection and QSAR model building. J. Chem. Inf. Comput. Sci. **41**, 1218–1227 (2001)
36. Norinder, U.: Support vector machine models in drug design: applications to drug transport processes and QSAR using simplex optimisations and variable selection. Neurocomputing **55**, 337–346 (2003)
37. Wold, S., Sjöström, M., Eriksson, L.: PLS-regression: a basic tool of chemometrics. Chemometr. Intell. Lab. Syst. **58**, 109–130 (2001)
38. Wang, R.X., Gao, Y., Liu, L., Lai, L.H.: All-orientation search and all-placement search in comparative molecular field analysis. J. Mol. Model. **4**, 276–282 (1998)
39. Baroni, M., Costantino, G., Cruciani, G., Riganelli, D., Valigi, R., Clementi, S.: Generating optimal linear PLS estimations (GOLPE): An advanced chemometric tool for handling 3D-QSAR problems. Quant. Struct. Act. Relat. **12**, 9–20 (1993)
40. Tropsha, A., Cho, S.J.: Cross-validated R^2 guided region selection for comparative molecular field analysis: a simple method to achieve consistent results. J. Med. Chem. **38**, 1060–1066 (1995)
41. Kamitori, S., Toyama, Y., Matsuzaka, O.: Crystal structures of cyclomaltohexaose (α-cyclodextrin) complexes with p-bromophenol and m-bromophenol. Carbohydr. Res. **332**, 235–240 (2001)
42. Momany, F.A., Rone, R.J.: Validation of the general purpose QUANTA 3.2/CHARMm force field. J. Comput. Chem. **13**, 888–900 (1992)
43. Guosheng, W.U., Robertson, D.H., Brooks, C.L., Vieth, M.: Detailed Analysis of grid-based molecular docking: a case study of CDOCKER-A CHARMm-based MD docking algorithm. J. Comput. Chem. **24**, 1549–1562 (2003)
44. Canceill, J., Lacombe, L., Collet, A.: Water-soluble cryptophane binding lipophilic guests in aqueous solution. J. Chem. Soc. Chem. Commun. **3**, 219–221 (1987)
45. Hirayama, F., Uekama, K.: In: Duchene, D. (ed.) Cyclodextrin and their industrial uses, pp. 131–172. Editions De Sant′e, Paris (1987)
46. Halperin, I., Ma, B., Wolfson, H., Nussinov, R.: Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins **47**, 409–443 (2002)