Journal of Computer-Aided Molecular Design, **14:** 669–678, 2000.
KLUWER/ESCOM

Classical QSAR and comparative molecular field analyses of the host-guest interaction of organic molecules with cyclodextrins

Takahiro Suzuki[a,*], Masaru Ishida[a] & Walter M.F. Fabian[b]
[a]*Research Laboratory of Resources Utilization, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama 226-8503, Japan;* [b]*Institut für Chemie, Karl-Franzens-Universität Graz, Heinrichstr. 28, A-8010 Graz, Austria*

Received 27 October 1999; Accepted 3 March 2000

Key words: CoMFA, cyclodextrin complexes, lipophilicity, molecular connectivity, molecular recognition, multivariate analysis

Summary

The application of classical QSAR and molecular modeling analysis using Comparative Molecular Field Analysis (CoMFA) to the complexation of some natural and modified cyclodextrins (CDs) with guest molecules was examined. For 1:1 complexation systems between natural β-CD, modified α-, β-, and γ-CD that bear one *p*-(dimethylamino)benzoyl (DMAB) moiety (DMAB-α-, β-, and γ-CDs) and guest molecules of widely varying chemical structures and properties, the binding constants of the complexes were successfully fitted using multiple linear regression (MLR) with hydrophobic descriptor log P (the partition coefficient between 1-octanol and water phases) and molecular connectivity indices. A non-linear dependency of binding constants on the zero-th and/or first order molecular connectivity index as a measure of size becomes apparent. The modeling performance of the CoMFA models with steric/electrostatic fields to DMAB-α- and β-CD systems was comparable to those of MLR models. However, statistically significant CoMFA models for γ-CD systems which have higher conformational flexibility of the ring could not be obtained. The CoMFA models obtained for DMAB-α- and β-CD systems showed that the predominant effects were steric for the DMAB-α-CD system and electrostatic for the DMAB-β-CD system, respectively.

Introduction

The cyclic carbohydrates, α-, β-, and γ-cyclodextrins (CDs), form cylindrical molecules with their hydrophilic groups on the outside of the molecule and a relatively non-polar hole down the middle. This hole can form noncovalent inclusion complexes with a variety of guest molecules [1]. Therefore, CDs have a wide range of applications, including improving the solubility and stability of drugs and biopharmaceuticals, and selectively binding materials that fit into the central hole in affinity purification and chromatography methods. Natural or native CDs are not used extensively in the pharmaceutical field because they are not very soluble and are rather toxic to injection. However, they may be modified by adding substituents such as alkyl or hydroxyalkyl groups on to the hydroxyls of the natural cyclodextrin, which reduces toxicity and can enhance solubility.

Several hypotheses on the driving forces of complexation have been put forward to account for the stabilities of CD inclusion complexes in aqueous solution: (1) release of 'high-energy' water from the CD cavity; (2) relief of conformational strain energy possessed by the uncomplexed CD; (3) the hydrophobic interaction; (4) electrostatic interactions; (5) hydrogen bonding; (6) induction forces; and (7) the London dispersion force [1]. However, there have been many uncertainties on the mechanisms for the CD inclusion complexation [1].

*To whom correspondence should be addressed. E-mail: tsuzuki@res.titech.ac.jp

In recent years, to identify the significant factors contributing to complex stability and predict the thermodynamic stability of CD inclusion complexes, computational chemistry techniques, molecular mechanics computation [2], linear regression [3–5], partial least squares [6], and artificial neural networks [7] have been applied. Many QSAR models have been obtained on the basis of the correlation between the measured binding constants and descriptors reflecting the molecular structure and/or properties of the guest molecules such as bulk, hydrophobicity, and electronic properties [1–8]. Although these studies have had limited success for some natural α- and β-CD complexes, little work has been performed on the quantitative analyses of natural γ-CD and chemically modified CD inclusion complexation. Besides, it is expected that the CD's molecular recognition abilities reflect their different cavity sizes: however, there is still little information on the optimum molecular size for guest molecules.

The aim of this study is to improve our understanding of CD-ligand interactions by analyzing the stabilities of two systems, natural β-CD complexes with a variety of organic molecules and modified α- and β-, and γ-CDs with oxygen-containing compounds, using QSAR techniques of classical multivariate correlations and the 3D-QSAR method of Comparative Molecular Field Analysis (CoMFA) [9]. Although great interest has been devoted to the application of CoMFA to a variety of ligand-receptor interaction problems [10, 11], CoMFA has not been employed yet to analyze the host-guest interaction of the CD system, to the best of our knowledge.

Materials and methods

Data sets for analyses

The thermodynamic stability of a CD inclusion complex is generally expressed in terms of the binding constant (or stability constant) of the complex, K. The K data of inclusion complexes at 25 °C of 20 simple organic compounds with β-CD [5], 13 heterocyclic compounds with β-CD [6], and 28 oxygen-containing compounds with modified α-, β-, and γ-CDs bearing one *p*-(dimethylamino)benzoyl (DMAB) moiety [12] were analyzed. The modified CDs act as indicators for detecting organic compounds in aqueous solution as fluorometric molecular recognition indicators, and the measurements of the K data were performed in an aqueous buffer solution at pH 7.20.

Structural descriptors for guest molecules

A total of 20 descriptors were used, of which 5 were geometric or topological, 11 were electronic or quantum chemical, 2 were hydrophobic, and the remaining 2 were miscellaneous. Geometric or topological descriptors include molecular weight, molecular volume, solvent accessible surface area, and two types of molecular connectivity indices, zero-th and first order [13]. For electronic descriptors, dipole moment, molecular polarizability [14], energies of highest occupied (HOMO) and lowest virtual orbital (LUMO), heat of formation, and 6 kinds of charges; absolute charges of the most negative and most positive atoms, sum of absolute values of negative and positive atomic charges, and relative positive and negative charges [15], were employed. The 1-octanol/water partition coefficient log P and the solvation energy in aqueous phase were used to describe hydrophobic aspects of the molecules. The remaining two descriptors were the molar refraction MR and an indicator variable for ring structure (taking the value 1 or 0 depending upon whether the ring is present or absent).

The molecular volume, the solvent accessible surface area, quantum chemical descriptors, and the solvation energy in aqueous phase were calculated with the INSIGHT II/MOPAC (using the PM3 method [16]) and SOLVATION programs [17]. Experimental log P values were taken from the LOGKOW databank [18] and the missing values were calculated by means of the CHEMICALC-2 program (QCPE program no. 608 or QCMP129) [19].

Classical QSAR analysis

The multiple linear regression analyses of the data were performed using the multivariate analysis program package [20]. For calculation of the PLS method [21], the SIMCA 3B package for IBM-PC and compatible microcomputers [22] was used.

CoMFA analysis

For the CoMFA studies, the most important facts and prerequisites of the analyses have been recommended [23]. In this study, starting structures for CD complexes were retrieved from the Cambridge Crystallographic Database [24] [BUPDEV (i.e., α-CD-3-iodopropionic acid (1:1) complex pentahydrate [25]), YIYSOO (i.e. β-CD-pentane-1,5-dio1· 6.2H_2O [26]), and DOCYID (i.e., γ-CD-12-crown-4 (1:1) complex [27])] for complexes of α-, β-, and γ-CD, respectively.

The guests contained in these X-ray structures were then modified accordingly with the aid of the SYBYL molecular modeling software package [28] to obtain the molecules investigated as guests in the present paper. While keeping the CD host fixed, all the respective guests were minimized with the TRIPOS force field [29] using Gasteiger-Hückel charges [30, 31] within the host molecule. Generally, several starting orientations and positions of the guest molecule within the host cavity were studied and the lowest energy one was used for further analysis. For the CoMFA [9, 32], the complex with the largest binding constant was used for alignment by least squares fitting of the glycosidic oxygen atoms of the CD part of all other complexes onto those of these templates. Since, as already pointed out above, both the position and orientation of the respective guest within the cyclodextrin cavity are different for the various complexes, an alternative alignment using atoms of the guest rather than of the host appeared inappropriate. After having removed the atoms of the host molecule, steric (sp^3-carbon as probe) and electrostatic fields (charge $= +1$ as probe) for the guests aligned in the described manner were computed. A box with dimensions $13\times12\times16$ Å (α-CD:guests) and $17\times20\times22$ Å (β- and γ-CD:guests), respectively, and a grid spacing of 2 Å were used. Statistical analysis was done by the PLS method with cross-validation (leave-one-out [33]) to determine the optimal numbers of components to be used in the final PLS without cross-validation. Because of the lack of data, the substituent (DMAB moiety) in the modified CDs was omitted in all CoMFA calculations.

Results and discussion

Natural β-CD:guest system

The PLS and MLR analysis showed that the binding constants (log K) can be described by just two parameters, the molecular connectivity index ($^0\chi$ or $^1\chi$) and the octanol-water partition coefficient (log P) among a variety of 20 descriptors prepared for guests. The log K values for the 1:1 complexes between natural β-CD and 33 organic compounds containing C, H, N, O, S, and Cl together with the corresponding $^0\chi$ and log P values are listed in Table 1. The optimized model is obtained (the figures at the 95% confidence interval are given in parentheses) as follows:

$$\log K = -0.095(\pm0.014)\,{}^0\chi^2 + 1.30(\pm0.19)\,{}^0\chi + 0.43(\pm0.07)\log P - 2.85(\pm0.51), \quad (1)$$
$$n = 33,\ r^2 = 0.920,\ Q^2 = 0.901,\ s = 0.35,\ F = 111.1$$

where n = number of compounds, r = correlation coefficient, Q = cross-validated correlation coefficient (by leave-one-out), s = standard error, and F = Fisher statistic. The relative significance of the terms in the

Table 1. Observed binding constants (K) for the 1:1 complexation system between β-CD and organic compounds at 25 °C, together with descriptor values included in Equation 1

No.	Guest	log K	$^0\chi$	log P
Subset 1				
1	Acetaldehyde	−0.64	2.7071	0.45
2	Acetone	0.42	3.5773	−0.24
3	Acetonitrile	−0.27	2.7071	−0.34
4	Tetrahydrofuran	1.47	3.5355	0.46
5	Benzene	1.83	4.2426	2.13
6	Toluene	2.09	5.1128	2.73
7	Nitrobenzene	2.04	6.6902	1.85
8	Benzaldehyde	1.78	5.8199	1.48
9	Aniline	1.60	5.1128	0.90
10	Benzyl alcohol	1.71	5.8199	1.05
11	Methanol	−0.96	2.0000	−0.74
12	Ethanol	−0.027	2.7071	−0.30
13	2-Propanol	0.63	3.5773	0.05
14	1-Butanol	1.17	4.1213	0.84
15	Cyclohexanol	2.76	5.1128	1.23
16	Trichloromethane	1.43	3.5773	1.97
17	Tetrachloromethane	2.20	4.5000	2.83
18	Diethylamine	1.36	4.1213	0.58
19	Dimethylsulfoxide	0.16	3.5773	−1.35
20	Dimethylformamide	0.43	4.2844	−1.01
Subset 2				
21	Quinoline	2.12	6.8115	2.03
22	Benzothiazole	2.38	6.1044	2.02
23	Acridine	2.33	9.1209	3.40
24	Carbazole	2.44	8.6733	3.72
25	Phenazine	2.41	9.1209	2.84
26	Phenothiazine	2.73	9.1209	4.15
27	Phenanthridine	2.57	9.3804	3.29[a]
28	Phenoxazine	2.69	9.1209	3.85
29	Thianaphthene	3.23	6.1044	3.12
30	Dibenzofuran	2.97	8.6733	4.12
31	Dibenzothiophene	3.48	8.6733	4.38
32	Thianthrene	3.57	9.1209	4.47
33	Xanthene	2.71	9.1209	4.23

[a]Estimated value.

above model was in the following order from F test: $^0\chi > {}^0\chi^2 >$ log P. The employment of $^1\chi$ (the pairwise correlation coefficient (r) between $^1\chi$ and $^0\chi$ was 0.9867) instead of $^0\chi$ gave a slightly inferior r^2 value of 0.896, as shown in the following equation.

$$\log K = -0.12(\pm 0.02)\,{}^1\chi^2 + 1.07(\pm 0.20)\,{}^1\chi + 0.52(\pm 0.08)\log P - 1.11(\pm 0.33), \quad (1')$$
$$n = 33,\ r^2 = 0.896,\ Q^2 = 0.847,\ s = 0.40,\ F = 83.5.$$

The squared term in $^0\chi$ is included in an attempt to account for non-linear relationships in molecular size of the guests. The $^0\chi$ and $^1\chi$, topological indices derived from graph theory, incorporate structural and topological information in terms of the atomic and bond levels, respectively. $^0\chi$ is sensitive to the degree of saturation of an organic compound. The value of $^1\chi$ gives information about branching in the molecular structure.

For the pairwise correlations among the 20 descriptors, $^0\chi$ was highly correlated ($r > 0.95$) with three descriptors, molecular volume, MR, and molecular polarizability in addition to $^1\chi$. The combinations of such descriptors which have been frequently used for a comparative descriptor of molecular bulkiness in the QSAR and log P gave just comparable or slightly inferior fits as compared to the combination of $^0\chi$ and log P. For the case of the molecular volume, the use gave the r^2 value of 0.919.

From a practical point of view, the generality of the derived Equation 1 was tested by the leave-50%-out cross-validation, giving the following correlation between observed log K and predicted log K ($\log K_{cal,cv}$):

$$\log K = 1.05 \log K_{cal,cv} - 0.13, \quad (2)$$
$$n = 33,\ r_{cv} = 0.936,\ s_{cv} = 0.42.$$

As can be seen from the statistical indices of Equation 2, the prediction offered by the QSAR model from Equation 1 is excellent, with a low decrease of the correlation coefficient r_{cv} and a small increase of the standard deviation s_{cv}.

The observed values of log K vs. the calculated values according to Equation 1 are shown in Figure 1. This model shows that there exists an optimum value of $^0\chi$ for which binding of the guest to the cyclodextrin is strongest. Furthermore, at a given value of $^0\chi$, the larger the log P of the guests, the greater the stability of the inclusion complexes. An example of the plotting of the parabolic relationship between log K and $^0\chi$ as a function of log P (Equation 1) is given in Figure 2. From this figure, the optimum size of

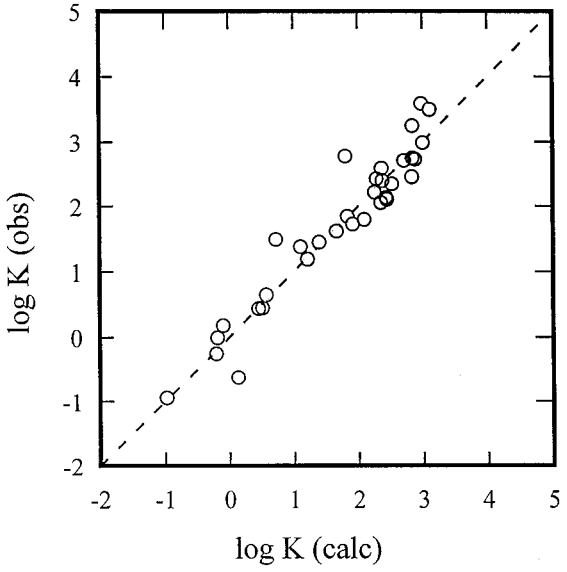

Figure 1. Comparison between observed and fitted log K values by Equation 1.

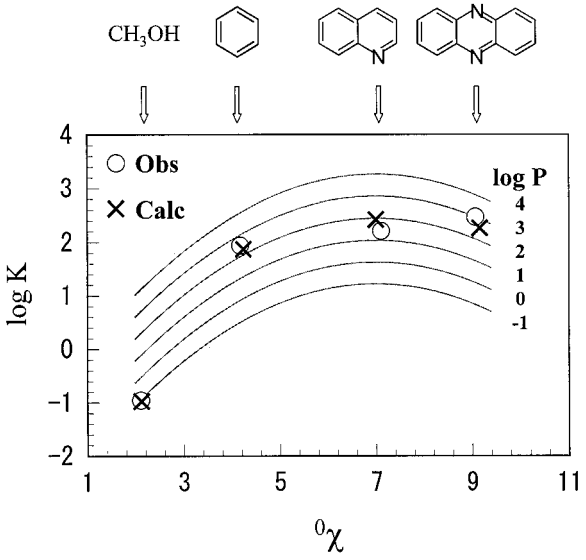

Figure 2. Relationship between log K and $^0\chi$ for natural β-CD:33 guests system at various log P values. The curves are calculated from Equation 1.

the guests for the cavity of β-CD can be assessed by means of a size parameter based on $^0\chi$. The optimum values for $^0\chi$ and $^1\chi$ are found to be 6.878 and 8.413, respectively, from Equations 1 and 1′.

In contrast, Park and Nah [5] have developed a QSAR model for 20 simple compounds (subset 1 in Table 1) complexing with β-CD using solvatochromic parameters [34]. The authors found that for ligands

Table 2. Observed binding constants (K) for the 1:1 complexation system between modified α, β, and γ-CDs bearing one *p*-(diethylamino)benzoyl (DMAB) moiety and organic compounds at 25 °C, together with descriptor values

No.	Guest	log K			$^0\chi$	$^1\chi$	log P[a]
		α	β	γ			
1	1-Butanol	2.48	1.75		4.1213	2.4142	0.84
2	2-Methyl-1-propanol		2.04		4.2844	2.2700	0.76
3	2-Methyl-2-propanol		2.18		4.5000	2.0000	0.35
4	1-Pentanol	2.78	2.00		4.8284	2.9142	1.51
5	3-Methyl-1-butanol	2.40	2.40		4.9916	2.7701	1.28
6	2-Methyl-2-butanol		2.45		5.2071	2.5607	0.89
7	1-Hexanol	2.91	2.51		5.5355	3.4142	2.03
8	Butanoic acid	2.43	2.11		4.9915	2.7700	0.79 (pKa=4.82)
9	2-Methylpropanoic acid	2.26	2.18		5.1547	2.6427	0.94[a]
10	Pentanoic acid	2.53	2.41		5.6986	3.2700	1.39 (pKa=4.83)
11	3-Methylbutanoic acid	2.34	2.75		5.8618	3.1258	1.16[b]
12	2,2-Dimethylpropanoic acid	2.30	3.60		6.0773	2.9433	1.14[b]
13	Hexanoic acid	2.72	2.70		6.4057	3.7700	1.92 (pKa=4.89)
14	Cyclohexanol		3.30		5.1128	3.3938	1.23
15	Cyclooctanol		4.70	2.30	6.5270	4.3938	2.38[b]
16	Nerol		3.60	2.78	8.6902	5.1639	3.59[b]
17	Geraniol		3.48	2.70	8.6902	5.1639	3.59[b]
18	*d*-Menthol		4.00	2.92	7.4307	5.1090	3.23[b]
19	*l*-Menthol		4.26	2.92	7.4307	5.1090	3.23[b]
20	1-Adamantanol		5.11	3.78	7.4746	5.2348	2.16[b]
21	1-Adamantancarboxylic acid		5.34	4.04	9.0520	6.1782	2.76[b]
22	Cyclododecanol		4.45	4.56	9.3555	6.3938	4.52[b]
23	*l*-Borneol		4.77	4.30	8.2760	4.9830	2.63[b]
24	Cholic acid		4.43	3.61	21.1375	13.5851	2.02 (pH=1)
25	Ursodeoxycholic acid		5.25	4.54	20.2672	13.1644	3.00 (pH=1)
26	Deoxycholic acid		4.67	4.34	20.2672	13.1744	3.50 (pH=1)
27	Chenodeoxycholic acid		4.71	4.76	20.2672	13.1644	3.28 (pH=1)
28	Lithocholic acid		5.20	4.92	19.3970	12.7537	3.08[b]

[a]Recommended value (not experimental) by LOGKOW databank [18].
[b]Estimated values.

whose sizes are small enough to be included in the cavity of β-CD, increasing guest molecular size stabilizes the complex by virtue of increasing dispersive interactions between the hydrophobic interior of the CD cavity and the guest, whereas increasing guest hydrogen bond acceptor basicity and dipolarity lead to a decrease in the stability of the complex due to stronger dipolar and hydrogen bonding interactions with water, which is more dipolar and hydrogen bond acidic than CD. However, it would not be so easy to prepare the empirical solvatochromic parameters for most organic compounds. There are many typical compounds whose solvatochromic parameters have not been determined or assigned.

For the 13 heterocyclic compounds contained in subset 2, the results of the PLS modeling of log K as a function of the heterocycle structure (global molecular properties and indicator variables) showed that separate models for compounds containing N, alone or with a second hetero atom, and for compounds containing O or S, were needed in order to have a satisfactory predictive ability [6].

Therefore, the present results clearly demonstrate that Equation 1 is useful for a generalized predictive equation for the thermodynamic stability of the natural β-CD complexes. Since $^0\chi$ reflects a measure of molecular volume or size and molecular volume is a major component of the van der Waals forces, it can be concluded that β-CD inclusion complexation is

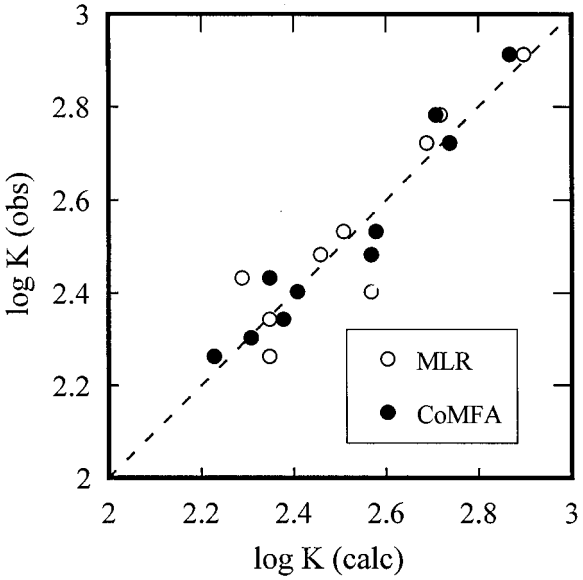

Figure 3. Comparison between observed and fitted log K values by MLR and CoMFA models for DMAB-α-CD complexes with 10 guest molecules.

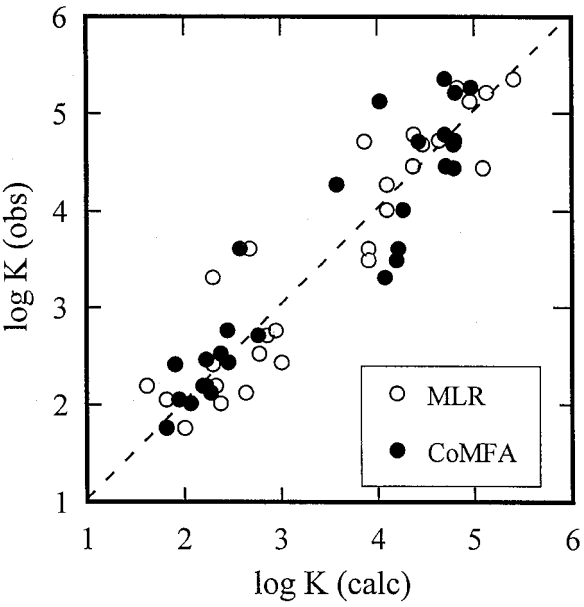

Figure 4. Comparison between observed and fitted log K values by MLR and CoMFA models for DMAB-β complexes with 28 guest molecules.

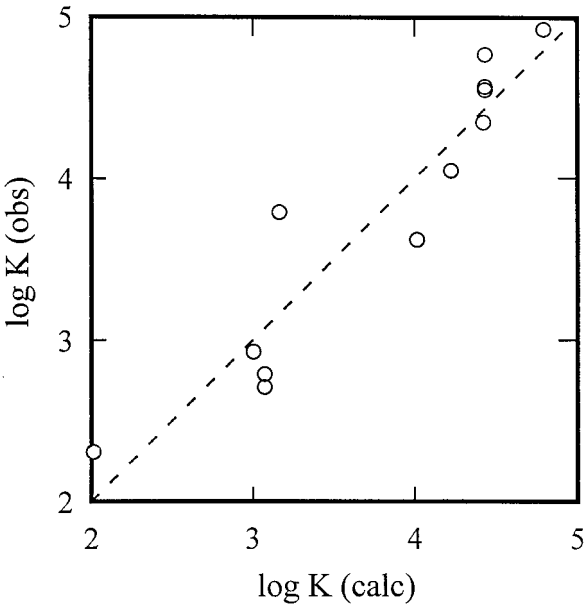

Figure 5. Plot of observed log K versus MLR calculated log K for DMAB-γ-CD complexes with 13 guest molecules.

mainly driven by van der Waals force and hydrophobic interaction (log P).

With respect to the Gibbs free energy change (ΔG) of binding, it can be imagined that entropic contributions are covered by molecular connectivity indices from Equations 1 and 1′ since a major factor arises from the water-to-β-CD transfer and this portion approximately correlated with the size of the hydrophobic surface area of the guest molecules [10]. In addition, differences in the guest's conformational flexibility have to be considered since the immobilization at the binding site involves significant entropy changes.

According to a collection of the thermodynamic parameters, ΔG, the enthalpy change (ΔH) and the entropy change (ΔS), for β-CD complexes in water at 25 °C [35], available combined ΔH and ΔS values (ΔH in kJ mol^{-1}, ΔS in J mol^{-1}) are (2.9, 33.0), (−10.0, 17.7) and (−1.88, 36.4) for 1-butanol, cyclohexanol and benzene, respectively (since the data sources are different from that of Table 1, their ΔG values are somewhat different from the calculated ones from the log K values in Table 1). From the data, the so-called enthalpy-entropy compensation effect [1, 35, 36] can be observed. In the cases of 1-butanol and benzene the ΔH values are unfavorable and larger ΔS values compensate the energy to increase the binding constants or −ΔG. On the other hand, the favorable

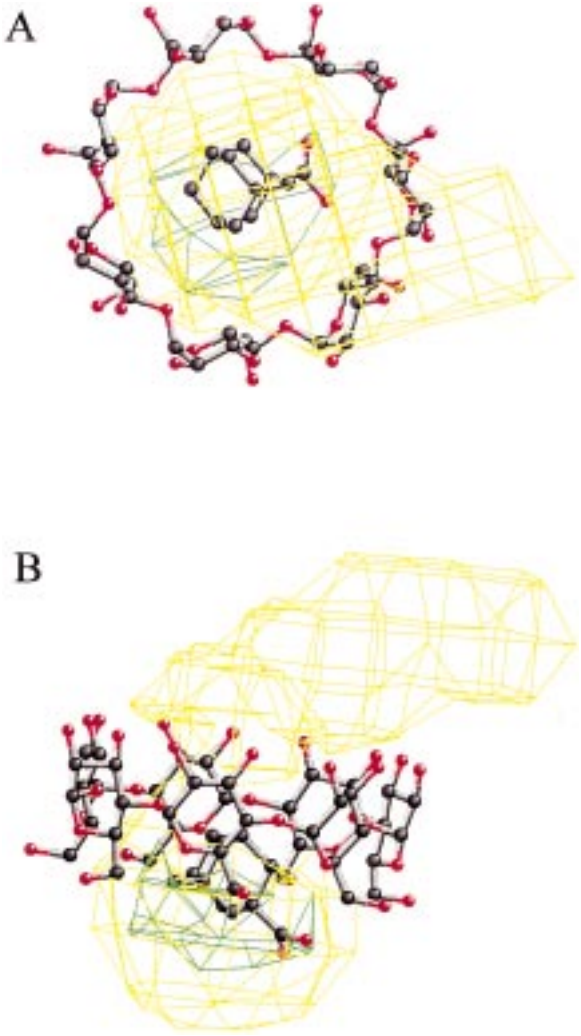

Figure 6. Top (A) and side (B) views of the contribution (standard deviation* coefficients) of the steric CoMFA field for the 1-adamantanecarboxylic acid:DMAB-β-CD system; the yellow- and green-colored areas indicate regions where a decrease and increase, respectively, in steric bulk will enhance the binding.

ΔH would result in a smaller ΔS value in the case of cyclohexanol, having more rotatable bonds than benzene. Janssen [37] discussed the relationship between the experimentally observed ligand–receptor binding constant, conformational flexibility of the ligand in solution, and the receptor-bound conformation in thermodynamic terms. The author concluded that higher energy conformations have a higher affinity for a receptor than lower energy conformations, independent of the conformation in which the ligands bound to the receptor.

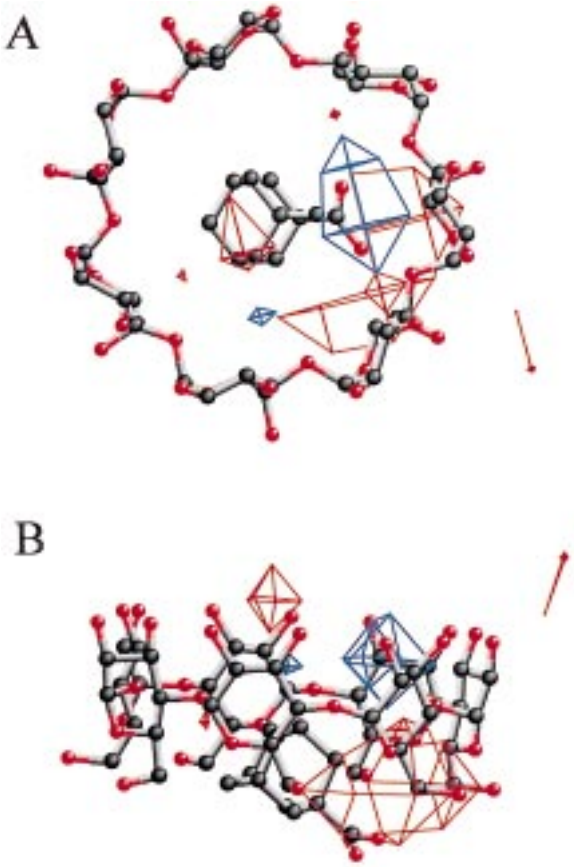

Figure 7. Top (A) and side (B) views of the contribution (standard deviation* coefficients) of the electrostatic CoMFA field for the 1-adamantanecarboxylic acid:DMAB-β-CD system; the red- and blue-colored areas indicate regions where an increase of negative and positive charges, respectively, will increase the binding affinity.

DMAB-CDs:guest systems

It is observed that the three DMAB-CDs show characteristic molecular recognition abilities reflecting their different cavity sizes. DMAB-α-CD is effective to bind the smaller chain molecules such as aliphatic alcohols but showed no response to cyclic compounds. DMAB-γ-CD has high affinities to the large cyclic compounds and steroidal compounds but showed no response to the smaller aliphatic compounds. DMAB-β-CD is effective in detecting all types of guest molecules examined in this study. Therefore, the available numbers of complexes to be correlated are 10, 28 and 14, respectively, for DMAB-α-, β-, and γ-CD systems.

The correlation analysis with the descriptors $^{0}\chi$, $^{1}\chi$ and log P (Table 2) gave the following optimized

Table 3. Comparison of the performance of MLR and CoMFA approaches to modeling the binding constants (K) for the 1:1 complexation system between DMAB-α-, β- and γ-CD and organic compounds at 25 °C

Host	MLR			CoMFA			
	n[a]	r^{2} [b]	Q^{2} [c]	n	comp[d]	r^{2}	Q^{2}
DMAB-α-CD	10	0.861	0.784	10	3	0.938	0.631
DMAB-β-CD	28	0.893	0.795	28	2	0.854	0.720
DMAB-γ-CD	13	0.890	0.821	14	3	0.981	0.231

[a]Number of molecules.
[b]Correlation coefficient in fitting.
[c]Cross-validated correlation coefficient.
[d]Number of components.

equations for DMAB-α- (Equation 3), β- (Equation 4) and γ-CD (Equation 5) systems:

$$\log K = -0.17(\pm 0.05)\,{}^{0}\chi + 0.56(\pm 0.09)\log P + 2.67(\pm 0.25), \quad (3)$$
$$n = 10,\ r^{2} = 0.861,\ Q^{2} = 0.784,\ s = 0.09,\ F = 21.7,$$

$$\log K = -0.12(\pm 0.02)\,{}^{1}\chi^{2} + 2.30(\pm 0.33)\,{}^{1}\chi - 0.69(\pm 0.20)\log P - 2.25(\pm 0.62), \quad (4)$$
$$n = 28,\ r^{2} = 0.893,\ Q^{2} = 0.795,\ s = 0.42,\ F = 67.3,$$

$$\log K = -0.14(\pm 0.02)\,{}^{1}\chi^{2} + 2.70(\pm 0.44)\,{}^{1}\chi - 7.18(\pm 1.67), \quad (5)$$
$$n = 13,\ r^{2} = 0.890,\ Q^{2} = 0.821,\ s = 0.33,\ F = 40.4.$$

For the DMAB-γ-CD system, one guest molecule, *l*-borneol, was left out from the analysis as an outlier based on the 'jackknife test', although the reason is not known yet. The ${}^{0}\chi^{2}$ and log P terms were insignificant in Equations 3 and 5, respectively, according to the t-test. This can be attributed to the limited number of available data of the host-guest systems. Relatively high Q^{2} values for Equations 3–5 show the statistical stability of the models. Introduction of a H-bonding indicator variable [3] did not contribute to further improvement of the models for the above three systems.

For the case of Equation 4, the order of the relative significance of the three terms is the same as the model of Equation 1, however, the contribution of the log P term to the binding affinity is less remarkable than that for log P in Equation 1. Since the data set for DMAB-β-CD includes the data of the 14 bulky ligands which could fit to DMAB-γ-CD, they must fit closely into the cavity of the DMAB-β-CD and then the complexation may be mainly governed by van der Waals interaction rather than by hydrophobic interaction.

Besides, it can be seen that the larger the log P, the weaker the stability of the inclusion complexes from the negative coefficient for the log P in Equation 4. This trend is opposite to the cases of Equations 1, 1′ and 3. The positive coefficients for log P or other hydrophobic parameters have been reported for some natural α- and β-CD:guest systems by other QSAR models [3, 7]. It appears that the negative coefficient for log P in Equation 4 results from a specific interaction between the DMAB moiety and the ligands. It has been reported that the action (interaction with guests) of the DMAB moiety of the β-CD system is different from that of the α-CD system [12]. On this problem of the negative sign for the log P term in the above QSAR model, further study to find more evidence would be required.

From Equations 4 and 5, the optimum ${}^{1}\chi$ values of guest molecules were determined to be 9.426 and 9.779 for DMAB-β- and γ-CDs, respectively. It is interesting to note that the optimum ${}^{1}\chi$ value for DMAB-β-CD is rather bigger than that (8.413) for natural β-CD. The difference may come from the behavior of the DMAB moiety. According to the observed fluorescence and circular dichroism spectra of DMAB-β-CD [12], the DMAB moiety is included in the cavity in the absence of a guest. However, the guest addition causes the expulsion of the DMAB moiety from the interior of the cavity by forming an intermolecular inclusion complex with the guest molecule. This fact could contribute to an increase of the hydrophobic surface of the host and the resulting increase of the optimum molecular size of guests as compared with the case of natural β-CD.

3D-QSAR/CoMFA models constructed for the DMAB-α- and β-CD systems yielded correlations comparable to those of MLR models as shown in Table 3. In Figures 3–5, scatter plots of experimental and calculated log K values for the three CD complexes are shown. For both α- and β-CD systems

statistically significant values of r^2 and Q^2 (see Table 3) were obtained with CoMFA models including steric and electrostatic fields, indicating good self-consistency and predictive ability. The contributions from the steric and electrostatic fields were 55% and 45% for DMAB-α-CD, and 42% and 58% for the DMAB-β-CD system, respectively.

To take an example, the corresponding CoMFA field contributions in the 1-adamantanecarboxylic acid (guest 21 in Table 2):DMAB-β-CD system are shown in Figures 6 and 7. The yellow and green contours characterize regions where a decrease and increase, respectively, in steric bulk would lead to enhanced binding (see Figure 6). For the electrostatic field, the red and blue contours indicate regions where addition of negative and positive charges, respectively, would increase the binding affinity (Figure 7). The areas where an increase of negative charge would be beneficial for binding closely match those regions where the carboxy group of 1-adamantanecarboxylic acid is situated. For the steric field a region where an increase of steric bulk should enhance binding is discernible near the lower rim of the β-CD. Thus, some increase in the molecular size should lead to more stable complexes, indicating the importance of van der Waals forces for complex formation. Just three and two principal components were required in the CoMFA models for DMAB-α- and β-CD, respectively, to optimally explain the variation of log K values. However, in the case of the DMAB-γ-CD system, CoMFA could not give any statistically significant (rather low Q^2) models for the complexation behavior of this system. This may be attributable to the higher conformational flexibility of the γ-CD ring, which can adapt its shape much easier than the more rigid α- and β-CDs to fit to the surface of the guest. In this context it should be mentioned that the minimization of the guests within the host cavity was done by keeping the structure of the host fixed. Thus, any possible adaptation of the host conformation to the guest was not taken into account.

Conclusions

Traditional QSAR and CoMFA approaches have been applied to derive quantitative relationships between the binding constants and guest molecular structures for natural β-CD, DMAB-α-, β-, and γ-CD systems. The observed variations in the CD binding affinity among a variety of guest molecules were successfully explained by differences in their size and hydrophobic/hydrophilic character by MLR models even in the case of β- and γ-CD systems, for which little QSAR work has been performed. Although CoMFA modeling did not work well for the highly flexible γ-CD system – at least with the approximation of a fixed host structure used here – this study demonstrates the usefulness of this approach to the analysis of host-guest interactions of CD systems.

References

1. Connors, K.A., Chem. Rev., 97 (1997) 1325.
2. Pérez, F., Jaime, C. and Sánchez-Ruiz, X., J. Org. Chem., 60 (1995) 3840.
3. Matsui, Y., Nishioka, T. and Fujita, T., Topics Curr. Chem., 128 (1985) 61.
4. Davies, D.M. and Savage, J.R., J. Chem. Res. (S), (1993) 94.
5. Park, J.H. and Nah, T.H., J. Chem. Soc. Perkin Trans. 2, (1994) 1359.
6. Carpignano, R., Marzona, M., Cattaneo, E. and Quaranta, S., Analytica Chim. Acta, 348 (1997) 489.
7. Liu, L. and Guo, Q.-X., J. Chem. Inf. Comput. Sci., 39 (1999) 133.
8. Lipkowitz, K.B., Chem. Rev., 98 (1998) 1829.
9. Cramer, III, R.D., Patterson, D.E. and Bunce, J.D., J. Am. Chem. Soc., 110 (1988) 5959.
10. Klebe, G., Abraham, U. and Mietzner, T., J. Med. Chem., 37 (1994) 4130.
11. Tong, W., Lowis, D.R., Perkins, R., Chen, Y., Welsh, W.J., Goddette, D.W., Heritage, T.W. and Sheehan, D.M., J. Chem. Inf. Comput. Sci., 38 (1998) 669.
12. Hamasaki, K., Ueno, A., Toda, F., Suzuki, I. and Osa, T., Bull. Chem. Soc. Jpn., 67 (1994) 516.
13. Kier, L.B. and Hall, L.B. Molecular Connectivity in Structure-Activity Analysis, Research Studies Press, Letchworth, 1986.
14. Miller, K.L. and Savchik, J.A., J. Am. Chem. Soc., 101 (1979) 7206.
15. Stanton, D.T. and Jurs, P.C., Anal. Chem., 62 (1990) 2323.
16. Stewart, J.J.P., J. Comput.-Aided Mol. Design, 4 (1990) 1.
17. INSIGHT II/MOPAC and SOLVATION, Biosym Technologies, San Diego, CA, 1994.
18. Sangster, J., LOGKOW – a Databank of Evaluated Octanol-Water Partition Coefficients, Sangster Research Laboratories, Montreal, 1993.
19. Suzuki, T., J. Comput.-Aided Mol. Design, 5 (1991) 149.
20. EXCEL MULTIVARIATE ANALYSIS, Ver. 3.0, Esumi Co. Ltd., Tokyo, Japan, 1998.
21. Wold, S., Ruhe, A. and Dunn, III, W.J., Siam J. Sci. Stat. Comput., 5 (1984) 735.
22. SIMCA 3B, Ver. SIMCA-3XQ., Principal Data Components, Columbia, MO, 1987.
23. Thibaut, U., Folkers, G., Klebe, G., Kubinyi, H., Merz, A. and Rognan, D., Quant. Struct.-Act. Relat., 13 (1994) 1.
24. Allen, F.H., Davies, J.E., Galloy, J.J., Johnson, O., Kennard, O., Macrae, C.F., Mitchell, E.M., Mitchell, G.F., Smith, J.M. and Watson, D.G., J. Chem. Inf. Comput. Sci., 31 (1991) 187.
25. Harata, K., Uekama, K., Otagiri, M. and Hirayama, F., Nippon Kagaku Kaishi, (1983) 173.

26. Steiner, T., Koellner, G., Gessler, K. and Saenger, W., J. Chem. Soc., Chem. Commun., (1995) 511.
27. Kamitori, S., Hirotsu, K. and Iliguchi, T., J. Chem. Soc., Chem. Commun., (1986) 690.
28. SYBYL Molecular Modeling Software, Ver. 6.0, Tripos Associates, Inc., St. Louis, MO, 1992.
29. Clark, M., Cramer III, R.D. and van Opdenbosch, N., J. Comput. Chem., 10 (1989) 982.
30. Gasteiger, J. and Marsili, M., Tetrahedron, 36 (1980) 3219.
31. Streitwieser, A., Molecular Orbital Theory for Organic Chemists, Wiley, New York, NY, 1961.
32. Marshall, G.R and Cramer III, R.D. Trend Pharmacol. Sci., 9 (1988) 285.
33. Cramer III, R.D., Bunce, J.D., Patterson, D.E. and Frank, E.I., Quant. Struct.-Act. Relat., 7 (1988) 18.
34. Kamlet, M.J., Abboud, J.-L.M., Abraham, M.H. and Taft, R.W., J. Org. Chem., 48 (1983) 2877.
35. Inoue, Y., Hakushi, T., Liu, Y., Tong, L.-H., Shen, B.-J. and Jin, D.-S., J. Am. Chem. Soc., 115 (1993) 475.
36. Searle, M.S., Westwell, M.S. and Williams, D.H., J. Chem. Soc. Perkin Trans., 2 (1995) 141.
37. Janssen, L.H.M., Bioorg. Med. Chem., 6 (1998) 785.