Predicting the Free Energies of Complexation Between Cyclodextrins and Guest Molecules: Linear versus Nonlinear Models

Christian Th. Klein,[1,4] Diether Polheim,[2] Helmut Viernstein,[3] and Peter Wolschann[1]

Received November 11, 1999; accepted December 9, 1999

Purpose. In the present paper, linear and nonlinear models for complexation of α- β- and γ-cyclodextrin with guest molecules are developed, with the aim of free energy prediction and interpretation of the association process.

Methods. Linear and nonlinear regression is used to correlate experimental free energies of complexation with calculated molecular descriptors. Molecular modeling supports the interpretation of the results.

Results. Highly predictive models are obtained, although the structural variability of the compounds used for their deduction is large, reaching from synthetic heterocycles to steroids and prostaglandins.

Conclusions. The scaled regression coefficients give insight to the complexation mechanisms, which appear to be different for the three types of cyclodextrins.

KEY WORDS: QSPR, Quantitative Structure-Property Relationship; correlation analysis; regression models; molecular modeling.

INTRODUCTION

Degradation of starch [α(1 → 4) linked poly-glucose] by α-1,4-glycosyltransferases yields cyclic oligosaccharides, cyclodextrins (CDs) with 6 (α-CD), 7 (β-CD) or 8 (γ-CD) glucose units. Their shape resembles that of cones, with a more or less hydrophobic cavity (1,2). The different types of CDs are obtained by glycosyltransferases with different specificities under special conditions (3).

Due to their ability to include organic molecules (guests) into the cavity, CDs are used more and more in pharmacy, technical and environmental chemistry and biotechnology, since complexation may protect the included compound against light or oxidation (4), or may improve its solubility (5) or bioavailability (6).

Being composed of chiral subunits (α-D-glucose), CDs interact differently with the enantiomers of the same compound, and can thus be used in enantioselective chromatography (7).

To decide whether a host-guest complexation is useful in a particular application, the complexation ability of the guest must be known. However, if the solubility of the guest molecules is low, the experimental determination of the stability constants is difficult, regardless of the method used. This is a strong motivation for developing theoretical models to estimate the host-guest complexation strength.

In a previous paper (8) we have described a method to predict the free energies of complexation between β-cyclodextrin and guest molecules, based on linear correlation analysis [multivariate linear regression (MLR) and partial least squares (PLS)]. The obtained models have high predictive power and allow conclusions on the host-guest interaction.

In the present work, we develop linear and nonlinear models for α- and γ-cyclodextrin. To compare the improvement introduced by nonlinearities for all three types of cyclodextrins, nonlinear models are deduced also for β-cyclodextrin. From the statistically significant regression coefficients, obtained by scaled variables (zero mean, unity variance), conclusions on the complexation mechanisms are inferred.

METHODS

All molecular properties used in the correlation analysis are calculated by the TSAR 3.1 (tools for structure-activity relationships) program package (9), starting from structures built and minimized with the HyperChem 5.0 program (10). The regression models are deduced also by TSAR using the two way stepping algorithm (11) implemented in the program: variables are included into and eliminated from the model via their partial *F*-test. The *F*-test checks the consistence of two variances, s_1^2 and s_2^2: if they do not differ significantly (i.e., they are consistent), their ratio s_1^2/s_2^2 should be *F* (Fisher)-distributed. In this way the significance of an individual regression coefficient a_i is tested by the ratio $F_i = a_i/se(a_i)$ [where $se(a_i)$ denotes the standard error of the regression coefficient a_i]: because $se(a_i)$ should be as small as possible a_i and $se(a_i)$ should differ significantly, i.e. F_i should not be *F*-distributed, and should exceed the percentage point of the *F*-distribution, $F_i > F_{1\text{-}P,1,n\text{-}k\text{-}1}$. n denotes the number of measurements, k the number of variables in the model, and P the confidence probability. For example, in a model with 4 variables, deduced from 30 compounds (measurements), F_i should exceed the point $F_{0.05,1,30-4-1} = 2.76$, to be significant with 0.95 probability. In a similar way, an overall F_0 value is defined, according to which the regression equation is significant if $F_0 > F_{I\text{-}P,k,n\text{-}k\text{-}1}$.

Two way stepping represents an attempt of finding best models without deducing all possible regression equations, which in the case of many independent variables is almost impossible. Thus, all variables included into the regression equations via the stepping algorithm are statistically significant. In the present study a correlation limit of 0.85 is applied, i.e., only variables with $r < 0.85$ are included into the models. Cross-validation is performed by leaving out one compound in turn.

The quality of the *significant* models (proper F_i and F_0 values) is estimated additionally by the correlation coefficient *r*, which should be close to 1 for good models, and the cross validated r^2, r_{cv}^2, which is a measure of the predictive power of the model. It is obtained by leaving out one compound in turn during cross-validation, and should be close to r^2.

The molecular descriptors considered are basically the same as in the previous study (8):

(i) surface (*S*), volume (*V*) and ovality (*O*) (12) of the molecules;

[1] Institut für Theoretische Chemie und Molekulare Strukturbiologie, Währinger Straße 17, A-1090 Wien.
[2] F. Joh. Kwizda Ges.m.b.H, Vienna.
[3] Institut für Pharmazeutische Technologie, Althanstraße 14, A-1090 Wien.
[4] To whom correspondence should be addressed.

0724-8741/00/0300-0358$18.00/0

(ii) the logarithm of the octanol:water partition coefficient (log$P$);

(iii) the molecular refractivity (MR);

(iv) number of hydrogen bond donors (n_{HD}) and number of hydrogen bond acceptors (n_{HA});

(v) number of chlorine (n_{Cl}) and of nitrogen (n_N) atoms as indicator variables;

(vi) different topological and connectivity indices: the Balaban index (J) (13), the χ, κ indices and the flexibility (ϕ) as defined by Hall and Kier (14) (the original notation is used);

(vii) the total dipole moment (D);

(viii) the sum of electrotopological indices (E) (15).

The molecular descriptors are correlated with free energies of complexation, obtained from the respective stability constants ($-RT \ln K_{\text{complex}}$) (16–22). In the situation where a compound at neutral pH occurs in a charged form, the stability constant of the uncharged form is considered, which has, generally, a lower ΔG-value (a higher complexation constant): being less soluble, the driving force of the hydrophobic effect will be more pronounced than for the charged compound.

Due to the high tendency of γ-CD to form higher order complexes, only about the half of available data stem from 1:1 complexes. Therefore, in contrast to the β-CD models previously presented (8), for γ-CD also higher order complexes are considered. The obtained models are capable of good predictions, regardless of the host-guest stoichiometry.

To support the conclusions resulting from the analysis of the regression equations, energy minimizations and molecular dynamics (MD) simulations are performed using the MM2 force field as implemented into the HyperChem (10) software. Within the MD simulations Newton's equations of motion are integrated, the solutions representing trajectories (motions) of the molecular system. Simulations of 50 and 200 ps are performed at a constant temperature of 298 K (this is the temperature at which the experimentally used stability constants were determined).

RESULTS AND DISCUSSION

α-Cyclodextrin

For the set of used compounds (Table I), the obtained models have rather low predictive power.

The best linear model found with the stepping algorithm is:

$$\Delta G = -14.829(1.099) \cdot {}^{5}\chi^{v}_{ch} - 1.318(0.031) \cdot J - 0.066(0.001) \cdot D - 0.458(0.121) \cdot n_{Cl} + 0.434(0.110) \quad (1)$$

$$r = 0.780, \quad s = 0.491, \quad F_0 = 16.737, \quad r^2_{cv} = 0.521$$

Although the regression equation is statistically significant at 95% level [$F_{0.05,4,43} \in (2.53, 2.61) < F_0$], the F_0-value is rather at the limit of significance, since according to some authors F_0 should exceed the percentage point at least four times (23). The partial F- or t-test additionally indicates the significance of individual regression coefficients. In what follows, only significant models will be presented.

An improvement, reflected in an increased F_0-value and an increased cross-validation r^2 (r^2_{cv}) is obtained by introducing explicitly nonlinearities into the model. Following equation is obtained:

Table I. Experimental and Predicted [with Eqn. (2)] Values of Free Energies of Complexation Between Guest Molecules and α-cyclodextrin

Compound	$\Delta G_{\text{experimental}}$	$\Delta G_{\text{predicted}}$
1 Prostaglandin E_1^{16}*	−4.286	−3.969
2 Prostaglandin E_2^{16}	−3.733	−3.914
3 Prostaglandin $F_{2\alpha}^{16}$	−3.257	−4.034
4 Prostacyclin[16]	−3.617	−3.845
5 Progesterone[16]	−2.956	−3.013
6 Testosterone[16]	−2.872	−2.915
7 Hydrocortisone[16]	−2.415	−3.026
8 Prednisolone[16]	−3.365	−2.975
9 Beclomethasone dipropionate[16]	−3.456	−3.794
10 Triamicinolone[16]	−2.852	−2.932
11 Betamethasone[16]	−3.182	−2.985
12 Spironolactone[16]	−4.051	−3.799
13 Diazepam[16]	−1.899	−2.195
14 Fludiazepam[16]	−2.006	−2.243
15 Indomethacin[16]	−3.095	−2.740
16 Flurbiprofen[16]	−1.767	−1.960
17 Fenbufen[16]	−2.006	−1.940
18 Ketoprofen[16]	−1.513	−2.101
19 Ibuprofen[16]	−1.922	−2.697
20 Piroxicam[16]	−2.628	−1.856
21 Phenobarbital[16]	−3.030	−2.220
22 Thiopental[16]	−3.280	−3.234
23 Phenythoin[16]	−2.655	−2.243
24 Sulphaphenazole[16]	−1.358	−2.237
25 Acetohexamide[16]	−2.308	−2.291
26 Tolbutamide[16]	−2.506	−2.737
27 Clofibrate[16]	−3.545	−3.087
28 Menadion[16]	−2.176	−2.137
29 p-Aminobenzoate[16]	−3.345	−3.302
30 p-Butylbenzoate[16]	−3.764	−3.407
31 p-Ethylhydroxybenzoate[16]	−3.064	−3.425
32 p-Butylhydroxybenzoate[16]	−3.678	−3.533
33 Medazepam[16]	−2.258	−2.198
34 Prednisolone acetate[16]	−3.303	−3.010
35 Cortisone[16]	−2.444	−3.002
36 Cortisone acetate[16]	−2.628	−3.036
37 Triamicinolone acetonide[4]	−3.272	−3.269
38 Deaxamethasone[4]	−3.026	−2.985
39 Fluocinolol acetonide[4]	−3.359	−3.306
40 Hydrocortisone acetate[4]	−2.642	−3.059
41 Picotamide[17]	−1.823	−1.772
42 Proscillaridin[18]	−2.773	−2.899
43 Prostaglandin A_1[19]	−4.230	−3.768
44 Prostaglandin B_1[19]	−4.138	−3.957
45 Digitoxigenin[20]	−4.388	−3.436
46 Dehydrocholic acid[20]	−3.030	−3.269
47 Betamethasone-17-valerate[4]	−3.369	−3.228
48 Paramethasone[4]	−3.653	−2.955

* References.

$$\Delta G = -13.032(2.915) \cdot {}^{5}\chi^{v}_{ch} - 1.322(0.281) \cdot J - 0.941(0.223) \cdot \sqrt{{}^{2}\chi^{v}_{p}} - 0.674(0.059)\frac{1}{{}^{4}\chi^{v}_{pc}} - 0.464(0.175) \cdot n_{Cl} + 2.958(0.529) \quad (2)$$

$$r = 0.834, \quad s = 0.438, \quad F_0 = 19.249, \quad r^2_{cv} = 0.614$$

The regression coefficients for all models, obtained by using

scaled variables (zero mean and unity variance), are given in Table II, where the importance of individual variables within a model is thus directly comparable.

One can see that for the linear model the highest contribution to the complexation energy stems from the $^5\chi^v_{ch}$ index and the Balaban index (J). The $^m\chi^v_t$ are weighted counts of subgraphs of type t, consisting of m joined bonds. t can be p (path), pc (path/cluster), c (cluster) or ch (ring).

$^5\chi^v_{ch}$ thus indicates the presence of five-membered rings in the molecule. The negative coefficient of $^5\chi^v_{ch}$ suggests that five-membered rings are included into CD cavity (if present and accessible) and thus stabilize the complex.

Correlation of the total dipole moment (D) of the guest molecules with the free energy of complexation indicates that dipole-dipole interactions might be important in the association process.

From Eqn. (2) (Table II) results that besides $^5\chi^v_{ch}$, the nonlinear terms have the highest contributions to the stabilization of the complex. Although the physical meaning of the nonlinear terms [$(^2\chi^v_p)^{1/2}$ and $1/(^4\chi^v_{pc})$] is rather difficult to be assessed, it is obvious, that in a way or the other they reflect contributions due to the topology (shape) of the molecule, thus suggesting the importance of van der Waals forces in the association process: a shape which favors van der Waals contacts with the host will stabilize the complex.

J contains structural information: the more branched a molecule is, the higher is its J-value. Branching thus favors the guest's interaction with α-cyclodextrin because it offers the possibility that individual branches are included into the CD cavity.

Table II. Regression Coefficients Obtained from Variables Scaled to Zero Mean and Unity Variance of the Obtained Models.

	α-Cyclodextrin		β-Cyclodextrin		γ-Cyclodextrin	
Descriptors	Eqn. (1)	Eqn. (2)	Eqn. (3)[a]	Eqn. (4)	Eqn. (5)	Eqn. (6)
S			−1.561			
V				−0.816	−1.569	
logP			−0.252	−0.401		
O			0.852	0.469		
E			0.779			
n_{HD}			−0.145		−0.495	−0.374
n_{HA}					0.805	0.941
n_N			0.546	0.417	0.542	0.261
n_{Cl}	−0.149	−0.155	0.326	0.348		
J	−0.474	−0.483				
D	−0.187					
$^4\chi^v_c$						−0.724
$^5\chi^v_{ch}$	−0.583	−0.512				
$^6\chi_{ch}$				−0.194		0.261
$^2\kappa$				1.121		
$^3\kappa$			−0.539			
$^3\kappa_\alpha$				−0.948		
ϕ			0.691		0.684	
$(^2\chi^v_p)^{1/2}$		−0.532				
$1/(^4\chi^v_{pc})$		−0.515				
$(^4\chi_p)^2$						−1.029
$(^4\chi_c)^3$				0.315		
r^2_{cv}	0.521	0.614	0.812	0.861	0.772	0.863

[a] The values of Eqn.(3) are not deduced in the present work, but taken from ref.(8).

These considerations, together with the observation that in non of the models the whole volume of the guest correlates significantly with the free energy of complexation, and the fact that the predictive ability of the models is rather low, suggest that (for this set of considered compounds) the complexes are not well defined inclusion complexes for all guest molecules. In other words, molecules are included only partially or not at all, if their volume is too large. This hypothesis is supported by molecular modeling: for triamicinolone, which has a larger volume (295.8 Å^3) than the average volume of the guests considered (252.6 Å^3), complexes are built by positioning the guests at random with their central part into the CD cavity. After conjugate-gradient energy minimization, three types of complexes are obtained, shown in Fig. 1: (i) complexes which have a favorable interaction energy ($\Delta E < 140.43$ kcal/mol, which is the sum of energies of the two isolated molecules) with the cyclohexadienone ring partially included into the CD cavity (Fig. 1, left); (ii) complexes with a favorable interaction energy where the hydroxyacetyl chain of the molecule is partially included (Fig. 1, middle); (iii) complexes where the central part of the molecule is included into the CD cavity, but which have a unfavorable interaction energy ($\Delta E > 140.43$ kcal/mol, Fig. 1, right); The complexes of type (iii) represent metastable states: performing 10 runs of MD simulations at 298 K, starting from fully minimized type (iii) structures, shows that after short times of 1.1 to 28.5 ps the guest leaves completely the host cavity, as shown in Table III.

A typical situation is presented in Figure 2 (run 9, Table III).

While the guest leaves the cavity at 12.7 ps rather quickly, i.e., within 0.8 ps, the total (potential and kinetic) energy decreases in average about 25 kcal/mol. The time interval from 12.7 to 24 ps corresponds to complexes of the type 1b (Fig. 1) where the hydroxyacetyl group is partially included in the CD cavity. After 24 ps the guest leaves the cavity completely, accompanied by a slight decrease of the potential energy, forming low energy complexes where the guest is oriented with its longitudinal axis parallel to the longitudinal axis of the CD. Such a complex is presented in Fig. 3.

The rather large variation of the distance at a relatively constant potential energy in the interval from 24 to 50 ps is a consequence of the fact, that the guest rotates in a plane parallel to the longitudinal axis of the host. Complexes of the type presented in Fig. 3 are found in 8 out of 10 runs.

These results in vacuo do not consider solvation effects, which are important for complexation. On the other hand, in cyclodextrin-guest complexations the enthalpy change is always negative (24). The total energy from MD simulations, being the sum of potential and kinetic (vibrational, translational and rotational) energy can be regarded as the enthalpic contribution to the complexation energy. Thus, the simulations only indicate that the triamicinolone: α-CD is unstable because exclusion of the guest leads to a decrease of the total energy.

β-Cyclodextrin

The best linear model found in our previous work (8) reads:

$$\Delta G = -0.018(0.003) \cdot S - 0.176(0.067) \cdot \log P + 7.109(0.955)$$

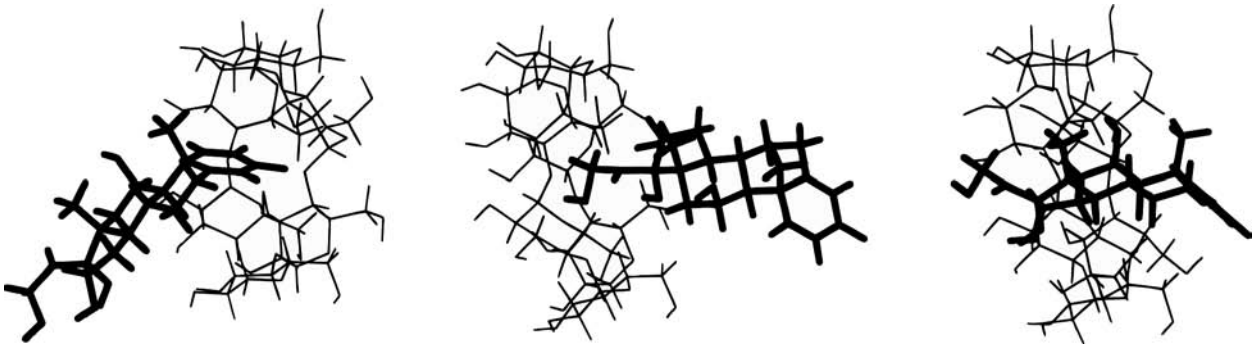

Fig. 1. Three types of complexes between triamicinolone and α-CD obtained by full minimization: (left) stable complex of 117.68 kcal/mol with the cyclohexadienone ring included into the cavity; (middle) stable complex of 126.29 kcal/mol where the hydroxyacetyl group is included into the cavity; (right) metastable complex with higher energy (165.78 kcal/mol) than the sum of energies of the two isolated molecules (140.43 kcal/mol), where the central part of the molecules is included into the cavity.

$$\cdot O + 0.330(0.118) \cdot \phi + 0.044(0.010) \cdot E - 0.144(0.081)$$
$$\cdot n_{HD} + 0.392(0.043) \cdot n_N \qquad (3)$$
$$+ 0.925(0.154) \cdot n_{Cl} - 0.292(0.115) \cdot {}^3\kappa - 12.749(1.256)$$
$$r = 0.927,\ s = 0.377,\ F_0 = 40.989,\ r^2_{cv} = 0.812$$

The 70 compounds employed for the deduction of the models are given in (8). A single cubic nonlinearity improves the model substantially:

$$\Delta G =$$
$$- 0.009(0.001) \cdot V - 0.281(0.002) \cdot \log P + 3.920(0.842)$$
$$\cdot O - 2.691(0.515) \cdot {}^6\chi_{ch} + 0.448(0.108) \cdot {}^2\kappa - 0.552(0.142)$$
$$\cdot {}^3\kappa_\alpha + 0.299(0.015) \cdot n_N + 0.988(0.202) \cdot n_{Cl} + 3.620(0.875)$$
$$\cdot ({}^4\chi_c)^3 - 9.006(1.304)$$
$$r = 0.944, \quad s = 0.331, \quad F_0 = 55.167, \quad r^2_{cv} = 0.861 \qquad (4)$$

From both equations [Eqn. (3) and Eqn. (4), Table II] one can see that steric and hydrophobic properties of the guest are most significant for complexation. Volume (*V*) and ovality (*O*) of the guest molecules are the most weighty descriptors in the linear model, followed by lipophilicity descripors (*logP, E*).

Because the ovality is calculated from the surface and the volume it is a relative quantity and must therefore be interpreted together with either *S* or *V*. In Eqn. (3) or Eqn. (4) (Table II), *S* or *V* have a high favorable contribution to complexation, but which is limited by the opposite sign of the coefficient of *O*: if the volume (or surface) of a guest is large, it can only enter the CD cavity if its ovality is also high (e.g. a linear polymer which would have a very high volume could enter the cavity due to a very high ovality).

The sum of electrotopological indices, *E*, is decreased by less electronegative atoms buried in the skeleton, and increased

Table III. Simulation Time Needed (ps) for the Guest Molecules to Leave Completely the Cavity of α- and β-CD, Respectively

α-Cyclodextrin		β-Cyclodextrin	
Energy (kcal/mol)[a]	Time	Energy (kcal/mol)[a]	Time (ps)
176.92	4.3	103.95	100 ps; reenters the cavity after 3 ps
165.78	28.5	106.34	80 ps; reenters after 1–2 ps
159.78	14.7	113.59	never leaves completely
173.64	6.1	107.58	never leaves completely
161.82	2.2	106.47	never leaves completely
168.78	5.7	113.32	never leaves completely
171.76	3.9	105.68	70 ps; reenters after 2 ps
163.51	1.1	108.55	never leaves completely
171.82	12.7	106.56	never leaves completely
174.52	1.9	109.76	never leaves completely

[a] Energy of the fully minimized starting structure.

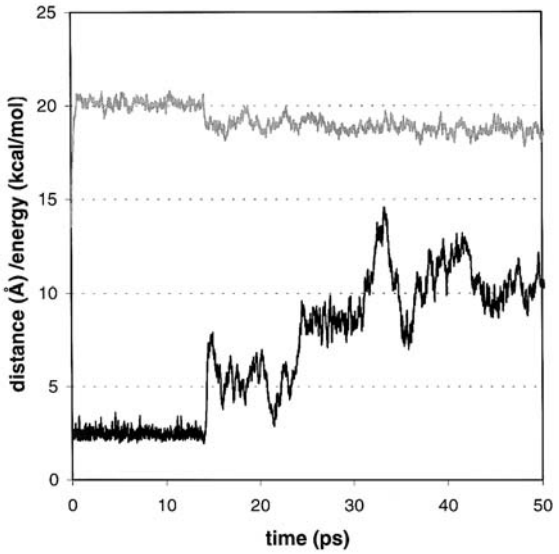

Fig. 2. Evolution of the distance (Å, black) and total energy [$E_{tot} \times 0.04$ (kcal/mol), grey] in the course of a 50 ps MD simulation of a triamicinolone: α-CD complex (run 9, Table III). The distance is measured between the carbonyl group of the hydroxyacetyl moiety of the guest and the nearest glucose hydrogen of the host.

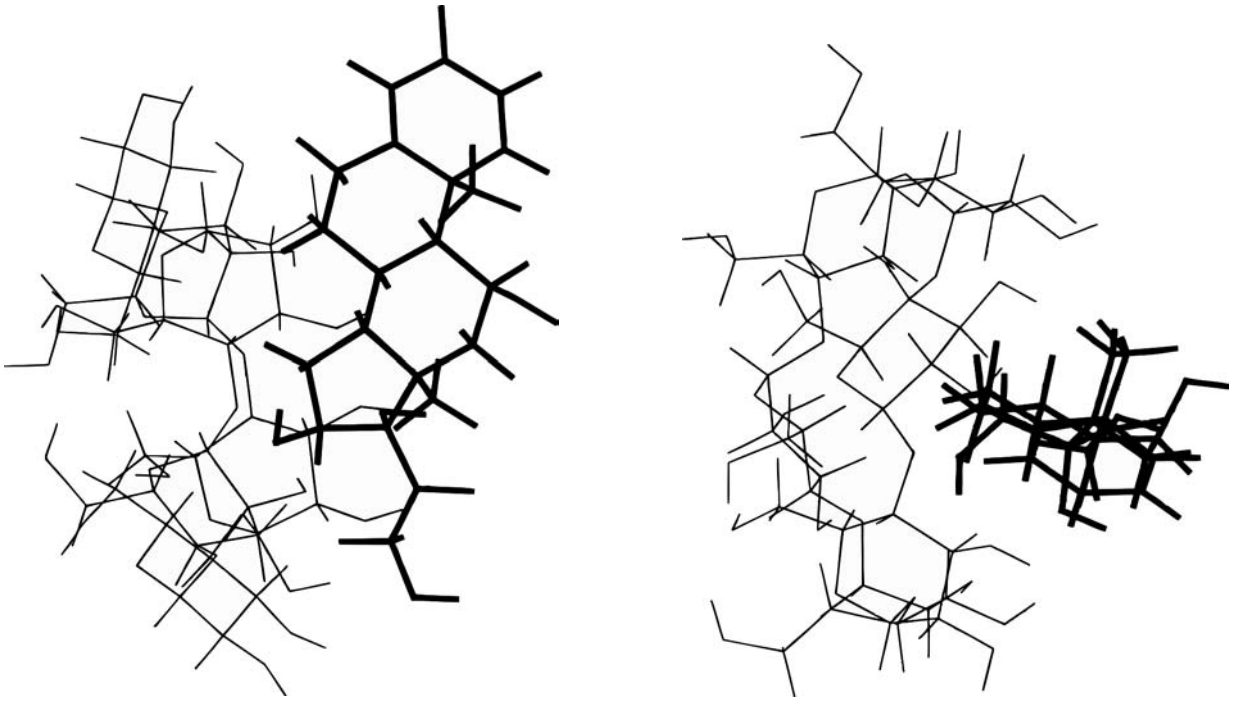

Fig. 3. Structure of the fully minimized complex (113.45 kcal/mol) from the end of the simulation from Fig. 2. Left: longitudinal axis of the guest is the paper plane; right: longitudinal axis of the guest is perpendicular to the paper plane.

by terminal (i.e., generally more exposed) atoms of high electronegativity. An increased *E*-value will thus occur in molecules with a rather hydrophilic surface. Hence, the destabilizing effect of increased *E*-values can be explained by the above considerations. The $^3\kappa$ shape index is related to the degree and centrality of branching in the guest molecule: it is the larger when branching is nonexistent or when it is located at the extremities of the molecular graph. The negative sign of the regression coefficient indicates that non or terminally branched guests should have increased complexation ability. The same considerations hold in principle also for $^3\kappa_\alpha$ from the nonlinear model. In the latter the largest non-favorable contribution to the complexation energy stems from $^2\kappa$. $^2\kappa$ encodes information related to the degree of star graph-likeness and linear graph-likeness, and is the higher the more linear the molecules is. The large positive sign of $^2\kappa$ indicates that nonbranched molecules will have better complexation ability. This is in agreement with the discussion about $^3\kappa$.

The lowest contribution to the complexation energy in Eqn. (3) stems from the hydrogen bond donor capacity of the guest molecule.

The high predictive ability of the models for β-CD, and the fact that either volume (*V*) or surface (*S*) correlates significantly with the complexation free energy, suggests that the inclusion complexes are well defined. Molecular modeling studies as those described for α-CD with triamicinolone complex support this conclusion: a β-CD:triamicinolone complex is built by positioning the guest at random with the central part into the CD cavity. After full minimization always complexes with a favorable interaction energy are obtained, as presented in Fig. 4.

Molecular dynamics simulations (Table III) show that in most of the cases the guest does not leave the CD cavity completely within 200 ps. In the cases where it leaves the total energy does not decrease, and it reenters after short time (from 1 to 3 ps). In Fig. 5 a typical run is presented for the β-CD:triamicinolone complex.

The rather large variations of the distance at a constant average total energy, come from large fluctuations of the guest in the β-CD cavity. As in the α-CD case, the simulations, being performed in vacuo, only show that the triamicinolone: β-CD complex is in principle stable. The significant minima in vacuo must not be necessary the relevant minima in solution: the hydrophobic effect will favor, for example, those minimum structures where the hydrophobic groups will be burried in the cavity, while the hydrophilic moieties will be exposed to the solvent. In other words, from the in vacuo geometries of the complexes, no conclusions can be inferred on the structure of the solvated complex.

γ-Cyclodextrin

The compounds used for the deduction of the regression models are presented in Table IV.

The best linear model for the γ-cyclodextrin:guest interaction has a substantially better quality (higher r^2_{cv}, higher statistical significance) than the models for α-cyclodextrin:

$$\Delta G = -0.019(0.002)\cdot \mathrm{V} + 0.334(0.071)\cdot \phi - 0.381(0.079) \cdot n_{HD} + 0.372(0.087)\cdot n_{HA} + 0.440(0.071)\cdot n_N - 1.913(0.319)$$

$$r = 0.906, s = 0.592, F_0 = 36.075, r^2_{cv} = 0.772 \quad (5)$$

Nonlinearities do improve the models significantly. The best nonlinear regression equation, involving a quadratic term reads:

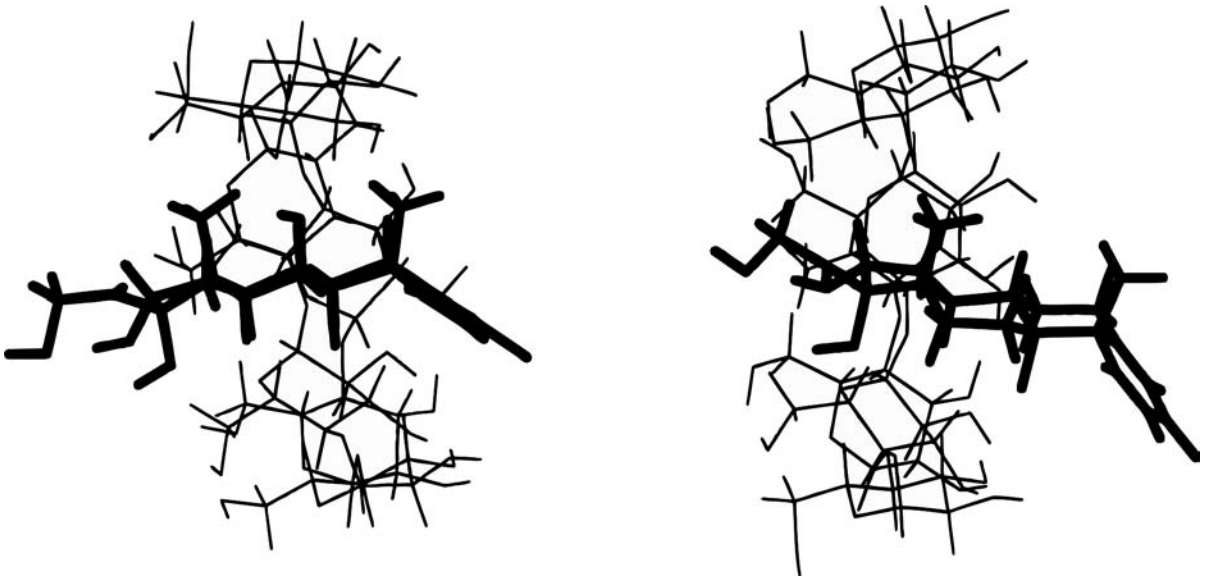

Fig. 4. Two complexes between triamicinolone and β-CD obtained by full minimization of a built complex by positioning the guest at random with the central part into the cavity. Both complexes (left, 106.34 kcal/mol) and (right 103.95 kcal/mol) have a favorable interaction energy (the sum of individual potential energies of host and guest, respectively, is 128.49 kcal/mol).

$$\Delta G = -6.157(1.867)\cdot {}^4\chi_c^v + 3.268(2.319)\cdot {}^6\chi_{ch} - 0.015(0.002)\cdot ({}^4\chi_p)^2 - 0.288(0.047)\cdot n_{HD} + 0.435(0.042)\cdot n_{HA} + 0.212(0.082)\cdot n_N - 4.102(0.346)$$

$$r = 0.947, \quad s = 0.454, \quad f_0 = 55.828, \quad r_{cv}^2 = 0.863 \qquad (6)$$

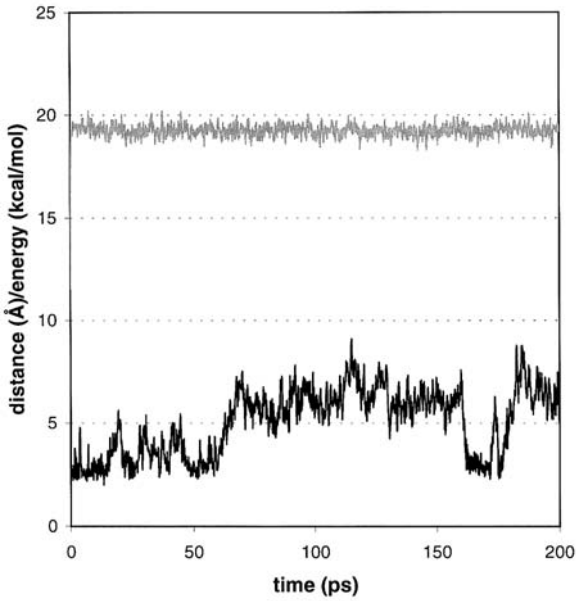

Fig. 5. Evolution of the distance (Å, black) and total energy [$E_{tot} \times 0.04$ (kcal/mol), grey] in the course of a 200 ps MD simulation of a triamicinolone: β-CD complex (run 4, Table III). The distance is measured between the carbonyl group of the hydroxyacetyl moiety of the guest and the nearest glucose hydrogen of the host.

The predictive power of the nonlinear model is fairly high, judged by r_{cv}^2.

In the case of the linear model [Eqn. (5), Table II], the largest influences on the complexation energy stem from the volume (V) and from the hydrogen bond acceptor capacity (n_{HA}) of the guest molecules, the hydrogen bond donor capacity (n_{HD}) being significant as well. It is interesting that an increased hydrogen bond acceptor capacity of the guest has a substantial destabilizing effect on the complex, while an increased donor capacity has the contrary effect.

More rigid molecules (decreased value of ϕ) should have better complexation abilities than flexible ones, as the positive sign of the respective regression coefficient shows, because in this case the host-guest interaction is better defined, i.e. inclusion into the CD cavity is facilitated. In contrast, more flexible molecules can favorably interact with the host also without being included.

As in the case of α- and β-CD models, the importance of indicator variables (n_N in this case) can be explained by the large variability of the considered compounds.

The nonlinear model [Eqn. (6), Table II] suggests that higher branching is favorable for host:guest complexation (negative sign of ${}^4\chi_c^v$.) ${}^4\chi_c^v$ represents a weighted count of all subgraphs of four bonds joined to a cluster (e.g. a tertiary carbon atom). This situation is contrary to β-CD where branching seems to be unfavorable; in the case of the larger cavity of γ-CD (diameter of 7.8 Å instead of 6.2 Å) branching appears to increase the van der Waals contacts with the host.

The presence of six-membered rings seems to have a destabilizing effect on the complexes, reflected in a positive sign of ${}^6\chi_{ch}$.

The highest stabilizing contribution to the complexation energy in the nonlinear model comes from the nonlinear term $({}^4\chi_p)^2$. As in the previous cases, the physical meaning of $({}^4\chi_p)^2$ is rather difficult to be established, but, since ${}^4\chi_p$ sensitively

Table IV. Experimental and Predicted [with Eqn. (6)] Values of Free Energies of Complexation Between Guest Molecules and γ-Cyclodextrin

Compound	$\Delta G_{experimental}$	$\Delta G_{predicted}$
1 Prostaglandin E_1^{16}*	−3.700	−3.249
2 Prostaglandin F_{2a}^{16}	−3.642	−3.537
3 Prostacyclin[16]	−2.944	−3.395
4 Progesterone[16]	−5.951	−5.232
5 Testosterone[16]	−5.729	−5.298
6 Hydrocortisone[16]	−4.551	−5.282
7 Prednisolone[16]	−4.769	−5.166
8 Beclomethasone dipropionate[16]	−5.161	−5.152
9 Triamicinolone[16]	−5.429	−5.405
10 Betamethasone[16]	−5.888	−5.552
11 Spironolactone[16]	−5.272	−5.289
12 Diazepam[16]	−2.824	−2.829
13 Fludiazepam[16]	−3.095	−2.961
14 Indomethacin[16]	−2.717	−2.854
15 Flurbiprofen[16]	−3.617	−3.315
16 Fenbufen[16]	−3.064	−2.812
17 Ketoprofen[16]	−2.773	−2.839
18 Ibuprofen[16]	−2.585	−3.403
19 Piroxicam[16]	−2.773	−3.174
20 Thiopental[16]	−3.403	−3.051
21 Phenythoin[16]	−2.824	−3.558
22 Sulphaphenazole[16]	−3.928	−3.753
23 Clofibrate[16]	−2.770	−3.322
24 Menadion[16]	−2.585	−3.054
25 Medazepam[16]	−2.594	−3.227
26 Prednisolone acetate[16]	−4.875	−4.583
27 Cortisone[16]	−4.533	−4.994
28 Cortisone acetate[16]	−4.609	−4.410
29 Triamicinolone acetonide[4]	−6.000	−5.998
30 Deaxamethasone[4]	−6.011	−5.551
31 Fluocinolol acetonide[4]	−6.118	−6.183
32 Hydrocortisone acetate[4]	−4.559	−4.698
33 Nitrazepam[4]	−2.006	−2.195
34 Nimetazepam[4]	−2.146	−1.970
35 Picotamide[17]	−2.054	−1.621
36 Proscillaridin[18]	−5.013	−5.253
37 Dehydrocholic acid[20]	−3.944	−4.581
38 Betamethasone-17-valerate[4]	−5.427	−5.419
39 Paramethasone[4]	−5.325	−5.466
40 p-Butylaminobenzoat[3]	−3.965	−3.158
41 Cephaelin[21]	−4.137	−3.815
42 Disopyramid[22]	−2.705	−2.838
43 Estriol[20]	−5.747	−4.748
44 Griseofulvine[20]	−2.956	−2.519
45 Digitoxin[3]	−6.525	−6.535

* References.

depends on the topology (and implicitly on the shape) of a molecule, $({}^4\chi_p)^2$ should reflect an interaction contribution which comes from van der Waals contacts.

CONCLUSIONS

The present study gives rise to models capable for good predictions of the free energy of complexation in the case of β- and γ-CD:guest complexes. The α-CD:guest complexes, for the series of guest molecules used, seem to be less well defined than those for β- and γ-CD, thus leading to models with significant lower predictive power. This hypothesis is supported by the analysis of the regression equations, as well as by molecular modeling studies.

Moreover, the analysis of the regression models gives an insight to the complexation mechanisms, which seem to differ for the three types of CDs. The significant correlation of steric descriptors (molecular surface, molecular volume, topological indices) suggests that van der Waals interactions are important for complexation in the case of all three CDs. In the case of α-CD dipole-dipole interactions might also stabilize the complexes.

For the β-CD:guest complexes the hydrophobic effect seems to have an important contribution to their stabilization, reflected in high weights of regression coefficients for *logP* and *E*, the hydrogen bond donor capacity of the guest playing only a minor role in the complexation process. This is in agreement with the fact that in β-CD most of the hydrogen bonds are intramolecular, forming a bend around the molecule (therefore also the lowest solubility of β-CD, compared to α- and γ-CD).

In contrast, the γ-CD:guest association appears to be crucially influenced by the hydrogen bond acceptor and donor capacity of the guest, while the hydrophobic effect seems to be of lower importance.

Simpler regression models of cyclodextrin:guest association have been reported (25,26). They are derived from small data sets (20 and 24 compounds, respectively), with high similarity. In (26) for example, monosubstituted benzene derivatives are employed in the deduction of the models. However, for modeling larger data sets of high complexity (variability), also more complex models, as those presented above, are necessary.

ACKNOWLEDGMENTS

The present work was performed within a project of the European Union (FAIR-CT96-1436). The authors thank Luckhana Lawtrakul and Pornpan Pungpo for reading the manuscript carefully.

REFERENCES

1. J. Szejtli. *Topics in Inclusion Sciences—Cyclodextrin Technology*, Kluwer Academic Publisher, 1988.
2. G. Le Bas, and N. Rysanek. Structural aspects of cyclodextrins. In Duchêne, D. (ed.), *Cyclodextrins and their Industrial Uses*, Editions de Santé 1987, pp. 107–130.
3. P. J. Sicard and M.-H. Saniez. Biosynthesis of cycloglycosyltransferases and obtention of its enzymatic reaction products. In Duchêne, D. (ed.), *Cyclodextrins and their Industrial Uses*, Editions de Santé, 1987, pp. 77–103.
4. D. Duchêne. Pharmaceutical applications of cyclodextrins. In D. Duchêne (ed.), *Cyclodextrins and their Industrial Uses*, Editions de Santé, 1987, pp. 211–257.
5. C. T. Klein, G. Köhler, B. Mayer, K. Mraz, S. Reiter, H. Viernstein, and P. Wolschann. Solubility and molecular modeling of triflumizole–β-cyclodextrin inclusion complexes. *J. Incl. Phenom.* **22**:15–32 (1995).
6. J. Szejtli. The metabolism, toxicity and biological effects of cyclodextrins. In Duchêne, D. (ed.), *Cyclodextrins and their Industrial Uses*, Editions de Santé, 1987, pp. 175–212.
7. E. Smolková-Keulemansová. Cyclodextrins in chromatography. In Duchêne, D. (ed.), *Cyclodextrins and their Industrial Uses*, Editions de Santé, 1987, pp. 259–295.
8. C. T. Klein, D. Polheim, H. Viernstein, and P. Wolschann. A method for predicting the free energy of complexation between β-cyclodextrin and guest molecules. *J. Incl. Phenom*, in press.
9. TSAR 3.1, Oxford Molecular Limited, 1997.

10. HyperChem 5.0, Hypercube Inc. 1996.
11. D. C. Montogomery and E. A. Peck. *Introduction into Linear Regression Analysis*, J. Wiley & Sons (1992), Chapter 7, pp. 290–298.
12. N. Bodor, Z. Gabanyi, and C.-K. Wong. A new method for the estimation of partition coefficients. *J. Am. Chem. Soc.* **111**:3783–3786 (1989).
13. A. T. Balaban. Chemical graphs XXXIV. Highly discriminating distance-based topological index. *Chem. Phys. Lett.* **89**:399–404 (1982).
14. L. H. Hall and L. B. Kier. The molecular chi connectivity indexes and kappa shape indexes in structure-property modeling. In K. B. Lipkowitz and D. B. Boyd (eds.), *Reviews in Computational Chemistry, Vol. 2*, VCH Publishers Inc, 1991, pp. 367–422.
15. L. B. Kier and L. H. Hall. The electrotopological state: structural information at atomic level for molecular graphs. *J. Chem. Inf. Comput. Sci.* **31**:76–82 (1991).
16. F. Hirayama and K. Uekama. Methods of investigating and preparing inclusion compounds. In Duchêne, D. (ed.), *Cyclodextrins and their Industrial Uses*, Editions de Santé, 1987, pp. 142–144.
17. P. Mura, A. Liguori, G. Bramanti, and L. Poggi. Improvement of dissolution characteristics of picotamide by cyclodextrin complexation. *Acta Pharm. Technol.* **34**:77–79 (1988).
18. K. Uekama, T. Fujinaga, M. Otagiri, N. Matsuo, and Y. Matsuoka. Improvement of dissolution and chemical stability of proscillaridin by cyclodextrin complexation. *Acta Pharm. Sued.* **20**:287–294 (1983).
19. K. Uekama, F. Hirayama, and T. Irie. A new method for the determination of the stability constants of cyclodextrin-prostaglandin inclusion complexes by liquid chromatography. *Chem. Lett.* 661–664 (1978).
20. Y. Okada, Y. Kubota, K. Koizumi, S. Hizukuri, T. Ohfuji, and K. Ogata. Some properties and the inclusion behavior of branched cyclodextrins. *Chem. Pharm. Bull.* **36**:2176–2185 (1988).
21. D. Teshima, K. Otsubo, S. Higuchi, F. Hirayama, K. Uekama, and T. Aoyama. Effects of cyclodextrins on degradations of emetine and cephaeline in aqueous solution. *Chem. Pharm. Bull.* **37**:1591–1594 (1989).
22. Y. Takahashi, T. Tsukuda, C. Izumi, K. Ikemoto, K. Kokubun, N. Yagi, and M. Takada. Preparation of solid dispersion systems of disopyramid with polyvinylpyrrolidone and γ-cyclodextrin. *Chem. Pharm. Bull.* **36**:2708–2710 (1988).
23. N. R. Drapper and H. Smith. Applied Regression Analysis, J. Wiley & Sons, 1981, p. 93.
24. W. Saenger. Cyclodextrin-Einschlußverbindungen in Forschung und Industrie. *Angew. Chem.* **92**:343–361 (1980).
25. J. H. Park and T. H. Nah. Binding forces contributing to the complexation of organic molecules with β-cyclodextrin in aqueous Solution. *J. Chem. Soc. Perkin Trans.* **2**:1359–1362 (1994).
26. Q.-X. Guo, S.-H. Luo, and Y.-C. Liu. Substituent effects on the driving force for inclusion complexation of α- and β-cyclodextrin with monosubstituted benzene derivatives. *J. Incl. Phenom.* **30**:173–182 (1998).