Novel Prediction for the Driving Force and Guest Orientation in the Complexation of α- and *β*-Cyclodextrin with Benzene Derivatives

Lei Liu[†] **and Qing-Xiang Guo***[,†,‡]

Department of Chemistry, University of Science and Technology of China, Hefei 230026, P. R. China and National Laboratory of Applied Organic Chemistry, Lanzhou University, Lanzhou 730000, P. R. China

Received: November 30, 1998; In Final Form: February 24, 1999

Taking the possibility of different inclusion orientations into consideration, a nonlinear free energy relationship model has been established by means of a genetic algorithm for the molecular recognition of α- and *β*-CD with mono- and 1,4-disubstituted benzenes. The association constants (K_a) for the inclusion complexation of α- and *β*-cyclodextrin with a number of benzene derivatives were evaluated by the model from the substituent molar refraction R_m, hydrophobic constant π, and Hammett constant σ, which respectively reflect the volume, hydrophobicity, and electronic property of the substituents in the guest compounds. The K_a values calculated by the model are quite close to those determined experimentally. It suggested that the van der Waals force, hydrophobic interactions, and electronic effects comprise the main driving forces for CD molecular recognition. Furthermore, van der Waals force plays a dominant role in α-CD complexation; on the other hand, van der Waals force and hydrophobic interactions play the major roles in *β*-CD complexation. The model is capable of quantitatively estimating the orientation of guest compounds in CD cavities. The predictions in both driving force and the orientation are in good agreement with the experimental studies.

1. Introduction

α- and *β*-cyclodextins (α- and *β*-CD) are macrocyclic oligosaccharides containing six and seven α-D-glucose units. They are shaped like truncated cones with primary and secondary hydroxyl groups crowning the narrower rim and wider rim, respectively. CD has a hydrophobic internal cavity of appropriate dimension and can bind with various organic and inorganic molecules to form inclusion complexes.[1] This molecular recognition property of CD has attracted much attention, not only due to its wide applications to pharmaceutical science, catalysis, and separation technology[2−4] but also because the inclusion process is considered an ideal model mimicking the enzyme−substrate interaction.[5] Furthermore, CD inclusion complexes are the most valuable models for understanding noncovalent intermolecular interactions in aqueous solution.

The geometry of the inclusion system and the driving forces leading to the complexation are the two aspects of main concern in supramolecular chemistry. Information from model studies on the complexation of CD has afforded a reasonable picture of the nature of molecular recognition. Up to now, the driving forces for the inclusion complexation of CD with substrates are attributed to several factors, such as van der Waals force,[6] hydrophobic interactions,[7] hydrogen bonding,[8] dipole−dipole, and dipole−induced dipole interactions,[9] steric effects[10] and so on.[11] But the most important driving forces contributed to CD complexation in aqueous solution were generally considered to consist of van der Waals forces, hydrophobic interactions, and electronic effects.[11] However, there still remains no clear agreement on the CD inclusion mechanism.[11−15] The relative contributions of the different driving forces are not well-known, and a general model capable of separating these effects is yet to be developed.[16]

Recently, quantitative studies of the inclusion complexation of CD based on quantum or molecular mechanics calculations,[12] multiple linear regression (MLR),[13−15] and artificial neural networks (ANN)[17] were carried out extensively. Considerable interests were devoted to the molecular recognition of mono- and 1,4-disubstituted benzenes with α- and *β*-CD,[10−18] because the complexation model possesses a relatively clear conformation[18] and is therefore desirable for direct investigation of the driving forces in the molecular recognition.

The quantum mechanics or molecular mechanics calculations can evaluate both the stability and the geometry of the complexes, which generally concluded that van der Waals force is the most important driving force.[12] However, the restriction of many calculations to the gas phase greatly discounted their reliability.[11,12] Linear regression analysis correlated the stability of CD complexes with molecular or substituent size, surface area, polarizability, hydrophobicity, etc.[11,13−15] But since there were not enough available data, the results were limited. Davies et al.[14] has collected a number of association constants of α- and *β*-CD complexation with benzene derivatives. The logarithm of the association constants for 48 α-CD complexes showed a good correlation coefficient (0.935) with the Hammett σ-constants and molar refraction of substituents. But the same method failed when applied to *β*-CD complexation.[14] In our previous work, multiple linear regression and stepwise regression analyses were performed on the binding of α- and *β*-CD with 24 monosubstituted benzenes. The correlation coefficients were 0.94 and 0.96 for α-CD complexation, and 0.91 and 0.94 for *β*-CD complexation, respectively.[15] Although these regressions have yielded some useful information, their empirical assignment of the orientation of the complexation was neither convenient nor soundly founded. Artificial neural networks (ANN) enjoyed the advantage of automatic and quick predictions. Our previous studies using ANN in the study of CD complexation indicated that there exists a certain nonlinear

[†] University of Science and Technology of China.
[‡] Lanzhou University.

10.1021/jp984545f CCC: $18.00
Published on Web 04/13/1999

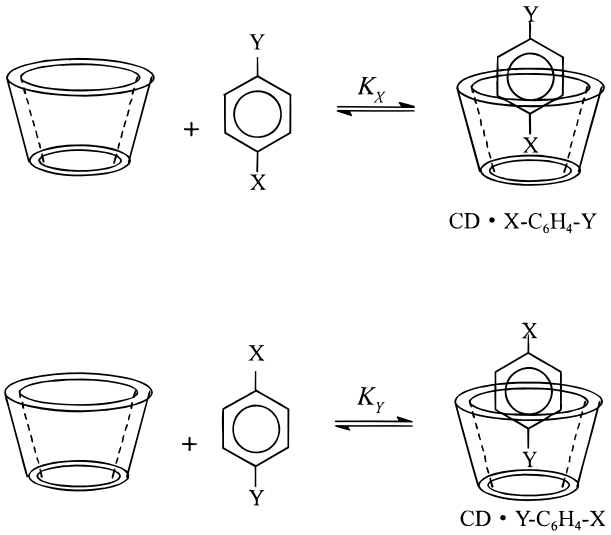

Figure 1. Two isomeric complexes of CD with mono- and 1,4-disubstituted benzenes.

relationship between the stability of CD complexes and the properties of the guests.[17] This nonlinear relationship, as revealed by the present study, is caused by the existence of different complexation orientations.

In the present study, we deduced a novel and reliable model for the inclusion complexation of α- and *β*-CD with mono- and 1,4-disubstituted benzenes by taking the possibility of different inclusion orientation into consideration. A genetic algorithm (GA) was employed to optimize the model due to its nonlinear nature. The association constants (K_a) for the inclusion of both α- and *β*-CD with a number of mono- and 1,4-disubstituted benzenes were successfully calculated by the optimized model. The results are in good agreement with the experimental data. Interestingly, this model can well predict the favorable orientations in CD binding with benzene derivatives, and the prediction is in accordance with the experimental studies.

2. Theory and Method

2.1. Inclusion Models. CD usually functions as a one-site ligand, and this site is the wider rim of the cavity.[11] CD inclusion complexes with guest molecules mainly have a stoichiometric ratio of 1:1.[11] The equilibrium equation is

$$\mathrm{CD} + \mathrm{guest} \overset{K_a}{\rightleftharpoons} \mathrm{CD{\cdot}guest}$$

Therefore,

$$K_a = \frac{[\mathrm{CD{\cdot}guest}]}{[\mathrm{CD}][\mathrm{guest}]}$$

However, binding of a mono- or 1,4-disubstituted benzene (X−C_6H_4−Y) with α- and *β*-CD can form two isomeric complexes, CD·X−C_6H_4−Y and CD·Y−C_6H_4−X, according to which substituent is located in the cavity[11,18] as shown in Figure 1. Based on this view, the following equation was obtained.

$$[\mathrm{CD{\cdot}guest}] = [\mathrm{CD{\cdot}X{-}C_6H_4{-}Y}] + [\mathrm{CD{\cdot}Y{-}C_6H_4{-}X}]$$

Then,

$$K_a = \frac{[\mathrm{CD{\cdot}X{-}C_6H_4{-}Y}] + [\mathrm{CD{\cdot}Y{-}C_6H_4{-}X}]}{[\mathrm{CD}]\,[\mathrm{X{-}C_6H_4{-}Y}]}$$

TABLE 1: Substituent R_m, π, and σ Constants Used in This Study[a]

no.	substituent	R_m	π	σ
1	H	0.34	0.00	0.00
2	F	5.17	0.14	0.06
3	Cl	9.83	0.71	0.30
4	Br	17.25	0.86	0.27
5	I	24.38	1.12	0.30
6	OCH_3	9.44	−0.02	−0.32
7	OEt	13.86	0.44	−0.24
8	CO_2Me	16.42	−0.01	0.39
9	COMe	13.00	−0.55	0.50
10	CN	7.79	−0.57	0.66
11	NO_2	12.29	−0.28	0.78
12	COOH	13.07	−0.28	0.45
13	NH_2	5.25	−1.23	−0.66
14	CH_3	5.07	0.56	−0.17
15	OH	5.05	−0.67	−0.37
16	SCH_3	14.94	0.62	0.00
17	CH_2OH	9.31	−1.03	0.08
18	Et	9.77	1.02	−0.15
19	*i*-Pr	14.48	1.40	−0.15
20	CH_2Cl	14.11	0.17	0.18
21	NHEt	14.69	0.13	−0.61
22	NHMe	9.96	−0.36	−0.84
23	NMe_2	14.48	0.18	−0.83
24	NHCOMe	15.04	−0.97	0.00
25	C≡CH	8.52	0.40	0.23
26	CHO	8.78	−0.65	0.22
27	CO_2Et	20.59	0.45	0.45

[a] The parameters are cited from ref 19.

If we define microscopic binding constants as the following:[11]

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

and

$$K_Y = \frac{[\mathrm{CD{\cdot}Y{-}C_6H_4{-}X}]}{[\mathrm{CD}]\,[\mathrm{X{-}C_6H_4{-}Y}]}$$

Thus,

$$K_a = K_X + K_Y$$

From the above equations, it is clear that except for *syn*-1,4-disubstituted benzenes, K_X and K_Y are not generally equal. However, the case that one of them is overwhelmingly greater than the other is not common either. Consequently, a quantitative model for binding constants of CD binding with mono- and 1,4-disubstituted benzene is not safely right, if only the binding position with larger microscopic binding constant is taken into account. Unfortunately, most studies,[10,11,13,14] as well as our previous work,[15] overlooked this point. This argument prompted us to seek a more general model, which takes both possible binding sites into consideration.

The driving force ($\Delta G_{binding}$) has been postulated for the binding of CD with guest molecules.[6,15]

$$\Delta G_{binding} = \Delta G_V + \Delta G_H + \Delta G_E$$

where ΔG_V, ΔG_H, and ΔG_E represent the contribution of van der Waals force, hydrophobic interactions, and electronic effects to the stability of the complexes, which are respectively correlated linearly with the substituent molar refraction R_m, hydrophobic constant π, and Hammett constant σ (Table 1).[19]

CHART 1

```
Genetic Algorithm ( )
{
    Initialize population;
    Evaluate population;
    While termination criterion not reached
    {
      Select solutions for next population;
      Perform crossover and mutation;
      Evaluate population;
    }
}
```

For CD binding with mono- and 1,4-disubstituted benzenes, we can obtain

$$\ln K_X = -\frac{\Delta G_{binding}}{RT} = aR_{mX} + b\pi_X + c\sigma_X + d\sigma_Y + f$$

where X represents the substituent inside the CD cavity and f reflects the contribution of the phenyl ring. Similarly, when Y is the substituent inside the CD cavity,

$$\ln K_Y = aR_{mY} + b\pi_Y + c\sigma_Y + d\sigma_X + f$$

As a result,

$$K_a = K_X + K_Y = e^{aR_{mX}+b\pi_X+c\sigma_X+d\sigma_Y+f} + e^{aR_{mY}+b\pi_Y+c\sigma_Y+d\sigma_X+f}$$

Obviously, this equation is nonlinear and its optimization is different from linear cases. Although several traditional methods are applicable, they are mostly based on gradient descent techniques and hence easily trapped in the local minimum of the hypersurfaces. Thus, as an efficient approach, the genetic algorithm is employed in this study.

2.2. Genetic Algorithm. Recently, the genetic algorithm (GA), founded upon the principles of natural evolution,[20] has attracted tremendous interest in many different fields. As a high-performance stochastic search algorithm, GA is capable of handling combinatorial and numerical globe minimum optimization problems. It is also convenient to the users that GA does not require the calculation of the parameter derivatives. Actually, GA can be seen as an evolutionary process in which a set of candidate solutions to a problem evolves over a sequence of generations. During each generation, the "fitness" values of the set of solutions are evaluated and the better solutions with higher "fitness" values will enjoy a higher possibility of existence in the evolution.

The main idea and procedure of the genetic algorithm can be summarized as shown in Chart 1. In this procedure, the initialization step randomly creates an initial population of solutions, which are usually defined as certain bytes of binary or decimal codes, respectively. The "fitness" values of the candidate solutions are evaluated by a certain problem-specific fitness function. The solutions with higher "fitness" values are selected to form the next generation according to a certain proportion. Some pairs of candidate solutions are randomly chosen from the newly selected generation. Parts of their codes are exchanged and thus generate new solutions. This step is generally called crossover. Meanwhile, a new solution can also be generated by randomly changing some codes of an old candidate solution to random values. This procedure is often named mutation. All the newly engendered solutions by crossover and mutation are added to the new generation of solutions. The new generation is then subjected to the next sequence of evolution until certain termination criterion is satisfied.

In the domain of chemistry, GA has been successfully applied to various areas, such as conformational analysis, atomic and molecular clusters, chemical dynamics, protein folding, QSAR, etc.[21] Herein, GA strategy is used to optimize the coefficients in the above equations. The program, written in Borland C^{++} 3.1, was run on an 80586 personal computer. The fitness function is taken as

$$\text{Fitness}(a,b,c,d,f) = \sum_{i=1}^{N}[\ln K_a(\text{obs})_i - \ln K_a(\text{calc})_i]^2$$

The equations optimized by GA for calculation of K_a values as follows:

$$K_a(\alpha - \text{CD}) = e^{0.166R_{mX}+0.139\pi_X+1.44\sigma_X-1.27\sigma_Y+2.51} + e^{0.166R_{mY}+0.139\pi_Y+1.44\sigma_Y-1.27\sigma_X+2.51} \quad (1)$$

and

$$K_a(\beta - \text{CD}) = e^{0.073R_{mX}+0.640\pi_X+0.507\sigma_X-0.506\sigma_Y+4.02} + e^{0.073R_{mY}+0.640\pi_Y+0.507\sigma_Y-0.506\sigma_X+4.02} \quad (2)$$

3. Results and Discussion

3.1. Driving Force for CD Inclusion Complexation. The K_a values for the inclusion complexation of α- and β-CD with benzene derivatives were calculated by eqs 1 and 2 from the substituent parameters R_m, π, and σ (see Table 1). The results are summarized in Tables 2 and 3.

Plotting the ln K_a values calculated by eqs 1 and 2 against experimental data gives straight lines (Figures 2 and 3), which fit the following equations:

For α-CD,

$$\ln K_a(\text{obs}) = 1.00 \ln K_a(\text{calc}) + 0.01 \quad (r = 0.92, sd = 0.46, n = 56)$$

For β-CD,

$$\ln K_a(\text{obs}) = 0.94 \ln K_a(\text{calc}) + 0.37 \quad (r = 0.92, sd = 0.25, n = 46)$$

From Figures 2 and 3, it can be seen that the slope of the line for α-CD is unity and for β-CD is approximately unity both with good correlation coefficients. It is obvious that the calculation results are remarkable. Moreover, due to their simplicity and clear physical meaning, the models offer valuable insights into the roles played by the driving forces in the process of host−guest complexation.

TABLE 2. ln K_a Values Calculated by eq 1 and the Experimental Data for the Inclusion Complexation of α-CD with Mono- and 1,4-Disubstituted Benzenes

no.	X	Y	ln K_a(obs)	ref	ln K_a(calc)	ln K_X	ln K_Y
1	F	H	3.68	15	3.79	3.48	2.49
2	Cl	H	4.72	15	4.76	4.68	2.19
3	Br	H	6.29	15	5.92	5.89	2.22
4	I	H	7.09	15	7.16	7.16	2.19
5	Cl	Cl	5.42	22	4.99	4.30	4.30
6	Br	Br	6.93	22	6.24	5.55	5.55
7	I	I	8.34	22	7.47	6.78	6.78
8	Cl	F	4.17	22	4.80	4.60	3.10
9	Br	F	5.52	22	5.88	5.81	3.13
10	I	F	6.89	22	7.10	7.08	3.10
11	OCH_3	OCH_3	4.02	23	4.72	4.02	4.02
12	OEt	OEt	4.85	23	5.53	4.84	4.84
13	CO_2Me	CO_2Me	6.14	23	6.00	5.31	5.31
14	CH_3	NH_2	4.05	23	4.22	4.02	2.47
15	Cl	NH_2	5.53	23	5.54	5.52	1.88
16	CN	NH_2	6.11	23	5.54	5.52	1.42
17	NO_2	NH_2	6.45	23	6.49	6.48	1.27
18	I	OH	7.75	23	7.63	7.63	2.34
19	Cl	OH	5.61	23	5.21	5.15	2.34
20	Br	OH	6.56	23	6.38	6.36	2.38
21	CN	OH	5.06	23	5.19	5.15	1.89
22	NO_2	OH	5.99	18	6.13	6.11	1.73
23	COOH	$NHCH_3$	7.17	23	6.38	6.36	2.33
24	COOH	F	6.22	23	5.31	5.22	2.91
25	NO_2	COOH	5.86	23	5.45	5.07	4.31
26	SCH_3	CH_2OH	4.44	10	5.31	4.98	4.03
27	Br	SCH_3	5.74	10	6.17	5.89	4.74
28	SCH_3	Cl	5.04	10	5.38	4.70	4.68
29	SCH_3	NH_2	4.62	10	4.90	4.25	4.17
30	CH_3	H	3.60	15	3.70	3.19	2.78
31	H	H	3.35	15	3.26	2.57	2.57
32	Et	H	4.60	15	4.30	4.06	2.76
33	CH_3	CH_3	4.28	15	4.10	3.40	3.40
34	*i*-Pr	H	4.56	15	5.01	4.90	2.76
35	OCH_3	H	4.95	15	4.04	3.62	2.97
36	OEt	H	5.14	15	4.71	4.53	2.87
37	CH_2OH	H	4.57	15	4.22	4.03	2.47
38	CH_2Cl	H	5.32	15	5.20	5.14	2.34
39	CHO	H	4.62	15	4.34	4.20	2.29
40	COMe	H	4.94	15	5.35	5.32	1.93
41	CO_2Me	H	5.36	15	5.83	5.81	2.07
42	CO_2Et	H	5.89	15	6.66	6.65	2.00
43	CN	H	4.36	15	4.73	4.68	1.73
44	NHEt	H	4.85	15	4.49	4.09	3.34
45	H	NHMe	4.42	15	4.03	3.63	2.91
46	NMe_2	H	5.15	15	4.38	3.75	3.62
47	NHCOMe	H	4.63	15	4.97	4.88	2.57
48	C≡CH	H	4.46	15	4.44	4.32	2.28
49	CH_3	OH	3.92	24	4.05	3.65	2.93
50	H	OH	3.61	25	3.58	3.03	2.72
51	OH	OH	3.18	25	3.89	3.19	3.19
52	COOH	H	5.82	26	5.33	5.29	1.99
53	COOH	OH	6.07	26	5.79	5.76	2.15
54	COOH	NH_2	6.22	27	6.14	6.13	1.69
55	CHO	OH	4.20	18	4.76	4.66	2.44
56	H	NH_2	4.03	28	3.68	3.40	2.26

The inclusion complexation of α- and β-CD with benzene derivatives is affected by the substituent molar refraction R_m of the guest molecules. Since R_m well reflects the volume and polarizability of the substrate, it has been expected that the CD inclusion complexation is affected by van der Waals force, which primarily consists of induction and London dispersion which depend on the guest volume and polarizability.[6,15] The larger the R_m value, the larger the volume and the polarizability of the compound, and hence the larger the van der Waals force. According to eqs 1 and 2, the coefficients for the R_m (0.166 for α-CD and 0.073 for β-CD) are positive; therefore, increasing R_m leads to increasing stability of the CD complexes in this work.

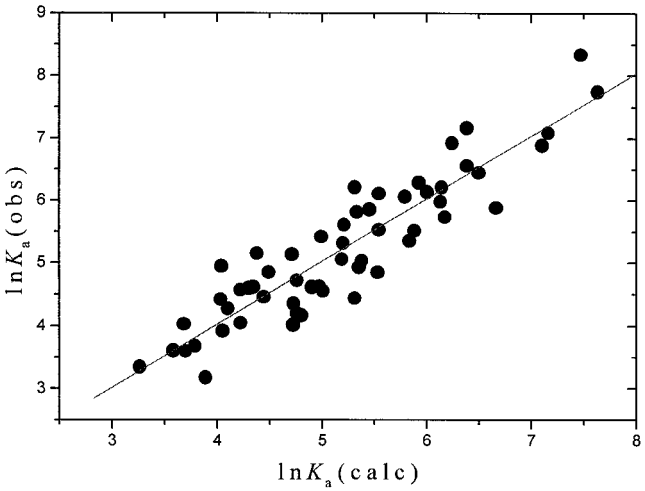

Figure 2. ln K_a values calculated by eq 1 vs those determined experimentally.

TABLE 3. ln K_a Values Calculated by eq 2 and the Experimental Data for the Inclusion Complexation of β-CD with Mono- and 1,4-Disubstituted Benzenes

no.	X	Y	ln K_a(obs)	ref	ln K_a(calc)	ln K_X	ln K_y
1	H	H	4.79	22	4.74	4.05	4.05
2	CH_3	H	4.94	22	5.13	4.67	4.13
3	C_2H_5	H	5.80	22	5.58	5.31	4.12
4	C≡CH	H	5.44	15	5.31	5.02	3.93
5	H	OH	4.77	29	4.73	4.24	3.78
6	OCH_3	H	5.34	15	5.08	4.54	4.21
7	OC_2H_5	H	5.66	29	5.50	5.19	4.17
8	CH_2OH	H	4.96	15	4.74	4.08	4.01
9	CH_2Cl	H	5.63	15	5.50	5.25	3.96
10	CHO	H	5.01	15	4.86	4.36	3.94
11	COMe	H	5.24	15	5.17	4.87	3.80
12	CO_2Me	H	5.76	15	5.60	5.41	3.85
13	CO_2Et	H	6.29	15	6.15	6.04	3.82
14	CN	H	5.14	15	4.92	4.56	3.71
15	H	NH_2	4.83	30	4.67	4.38	3.29
16	H	$NHCH_3$	4.87	15	5.00	4.47	4.09
17	NHC_2H_5	H	5.38	15	5.35	4.89	4.36
18	$N(CH_3)_2$	H	5.38	29	5.33	4.77	4.47
19	$NHCOCH_3$	H	5.06	15	4.99	4.50	4.05
20	NO_2	H	5.63	15	5.34	5.14	3.65
21	F	H	4.51	15	4.99	4.52	4.02
22	Cl	H	5.17	15	5.56	5.35	3.90
23	Br	H	5.77	15	6.09	5.97	3.91
24	I	H	6.74	15	6.73	6.67	3.90
25	COOH	H	5.88	31	5.29	5.03	3.82
26	CH_3	CH_3	5.48	22	5.45	4.75	4.75
27	Cl	Cl	5.77	22	5.89	5.20	5.20
28	Br	Br	6.85	22	6.53	5.83	5.83
29	I	I	7.31	22	7.21	6.52	6.52
30	CH_2OH	OH	4.98	32	4.91	4.68	3.33
31	C_2H_5	OH	6.20	32	5.68	5.50	3.85
32	NO_2	OH	5.50	32	5.46	5.32	3.38
33	I	OH	6.86	33	6.90	6.86	3.62
34	I	F	6.23	22	6.74	6.64	4.37
35	NO_2	NH_2	5.72	32	5.54	5.47	2.89
36	CH_3CO	OH	5.02	34	5.26	5.06	3.52
37	Br	OH	6.10	35	6.23	6.16	3.64
38	CH_3O	OH	5.09	35	5.10	4.73	3.94
39	OH	OH	4.73	25	4.66	3.96	3.96
40	Cl	OH	5.55	18	5.67	5.53	3.62
41	CH_3	OH	5.34	18	5.17	4.85	3.86
42	COOH	OH	5.06	32	5.39	5.21	3.55
43	Cl	NO_2	5.04	36	5.66	4.95	4.98
44	NO_2	COOH	5.39	37	5.47	4.91	4.63
45	CHO	OH	4.96	18	4.89	4.54	3.66
46	COOH	NH_2	5.21	27	5.45	5.36	3.05

It is well-known that CD has remarkably large dipole moments which point from the wider rim toward the narrower rim.[9,11] Therefore, the influence of the electronic effect on the

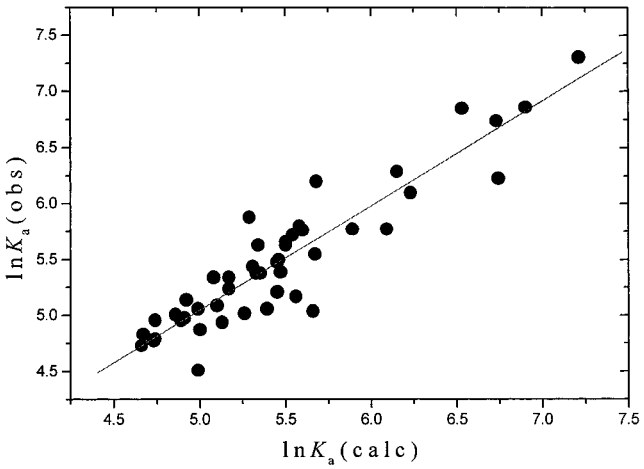

Figure 3. ln K_a values calculated by eq 2 vs those determined experimentally.

complexation of CD with guest compounds is expected, and the antiparallel arrangement of the dipoles between the host and the guest favors binding.[18] Hammett σ constants of the substituents were usually chosen as a measure of the electronic properties of the guest molecules.[11] Generally speaking, the substituent located near the narrower rim of the CD cavity with larger σ value or located near wider rim of the CD cavity with smaller σ value favors the inclusion binding.[23] Interestingly, herein the coefficients for σ_X and σ_Y obtained in eqs 1 and 2 are positive and negative, respectively (1.44 and −1.27 for α-CD, and 0.507 and −0.506 for β-CD), indicating that the substituent X with positive σ value and thus electron withdrawing will positively contribute to the complexation, while the substituent Y with positive σ value will negatively contribute to the complexation. This perfectly offers additional confidence in the above argument.

It has been noted that the hydrophobicity of the guest plays an important role in CD complexation. The contribution of hydrophobicity to the complexation of CD with guest compounds mainly comes from the entropy gain in the inclusion process when a hydrophobic guest breaks down the water cluster and penetrates into the apolar cavity of CD.[11] Since the substituent hydrophobic constant π is a good measure of hydrophobicity, it is feasible to be employed in the study of the role of the hydrophobic effect in the molecular recognition. The substituents with larger π values are more hydrophobic and therefore are strongly driven into the hydrophobic cavity of CD from the water environment. The positive sign of the π parameters in eqs 1 and 2 (0.139 for α-CD and 0.640 for β-CD) indicates that the more hydrophobic the substituents, the better the binding. This result is in perfect agreement with previous studies.[8]

The constant terms in eqs 1 and 2 are 2.51 and 4.02 for α- and β-CD, respectively. It indicates that the phenyl ring exerts more effects in β-CD than in α-CD complexation, since the bigger cavity of β-CD allows the aromatic ring to be embedded deeply, and thus a stronger interaction between the phenyl ring and the internal wall of the cavity occurred.

Admittedly, there would be still other driving forces contributing to the binding of CD with guest compounds. The most significant one of them is hydrogen bonding. However, it has been indicated that the contribution of the hydrogen bonding to CD inclusion in aqueous solution is not important.[13] The semiempirical formulas (eqs 1 and 2) were successful in calculation of the K_a values, although only van der Waals forces, hydrophobic interactions, and electronic effects were concerned. Therefore, the contribution of the hydrogen bonding is ommitted in this work.

3.2. Relative Importance of Each Factor. To examine the importance of the role played by each factor, the following equations that use R_m, σ, and π as input parameters, respectively, were also optimized by the genetic algorithm for the same samples. The optimum coefficients and the correlation coefficients are summarized as follows:

$$\ln K_a(\alpha\text{-CD}) = \ln[e^{0.180R_{mX}+2.61} + e^{0.180R_{mY}+2.61}] \quad (r = 0.82, n = 56)$$

$$\ln K_a(\alpha\text{-CD}) = \ln[e^{0.720\pi_X+4.38} + e^{0.720\pi_Y+4.38}] \quad (r = 0.28, n = 56)$$

$$\ln K_a(\alpha\text{-CD}) = \ln[e^{2.17\sigma_X-0.10\sigma_Y+4.21} + e^{2.17\sigma_Y-0.10\sigma_X+4.21}] \quad (r = 0.48, n = 56)$$

$$\ln K_a(\beta\text{-CD}) = \ln[e^{0.114R_{mX}+3.75} + e^{0.114R_{mY}+3.75}] \quad (r = 0.82, n = 46)$$

$$\ln K_a(\beta\text{-CD}) = \ln[e^{1.15\pi_X+4.71} + e^{1.15\pi_Y+4.71}] \quad (r = 0.72, n = 46)$$

$$\ln K_a(\beta\text{-CD}) = \ln[e^{0.496\sigma_X-0.361\sigma_Y+4.79} + e^{0.496\sigma_Y-0.361\sigma_X+4.79}] \quad (r = 0.28, n = 46)$$

Based on the analyses above, it can be seen that the R_m plays the most significant role in both α- and β-CD complexation, which confirmed the importance of the van der Waals force in the CD inclusion complexation.[6]

However, the relative importance of the π and σ constants are different between α- and β-CD. In α-CD complexes, the electronic effect is more important than the hydrophobic effect. However, the hydrophobic effect contributes more than the electronic effect in β-CD complexation because the cavity of β-CD is larger than that of α-CD.[11] A tighter contact results in stronger electronic interactions, while a deeper inclusion of a guest in an apolar cavity is more affected by the hydrophobic properties.

3.3. Guest Orientation in the Complexation. The study on the orientation of the guest in α- and β-CD cavities is very important in the molecular recognition and the chemical reactions catalyzed by CD and its derivatives.[38] Experimental methods,[11,39] such as NMR and X-ray analyses, can only be applied to a rather limited range of cases. However, the study of the dependence of the complex stability on the host−guest structure can alternatively provide useful information about the orientation.[11] It has been suggested that the nonpolar group of the guest compound would locate near the narrower rim of the CD cavity, and the polar group of the guest compounds would locate near the wider rim of the CD cavity.[11,23] Davies et al.[10,14] considered the electronic effect to be more decisive and therefore suggested that uncharged guests are embedded in the CD cavity so that the guest dipole is antiparallel to the host, with electron-withdrawing substituents located in the narrower end of the cavity. Lichtenthaler and Immel[40] argued that the complementarity of hydrophobic and hydrophilic sites is important in determining host−guest orientation. Although all of these empirical rules can be helpful to decide the orientations in the complexation, they are of qualitative nature and only applicable to limited circumstances.

The model developed in this study offers quantitative information on the CD host−guest orientation. The microscopic binding constants K_X and K_Y reflect the stability of the two possible isomeric complexes. Comparison between K_X and K_Y can determine which substituent (X or Y) of the guest (X−C_6H_4−Y) is more favorable to be included within the CD cavity. Therefore, the orientation of the guest compound in the α- and β-CD cavities can be quickly predicted by eqs 1 and 2 from the substituent R_m, π, and σ constants. The K_X and K_Y values calculated for each complex are listed in Tables 2 and 3 for α- and β-CD systems, respectively.

The *p*-substituted benzoic acids generally have much larger microscopic binding constants for COOH group (K_{COOH}) than the *p*-substituent, which indicates that the carboxyl group locates deeply in the CD cavity more favorably. The same observation was also reported by Connors.[23] The NMR studies showed that α-CD and β-CD complexation with benzoic acid took the orientation in which the carboxyl group located near the narrower rim of the cavity.[41] It is also reported that the carboxyl group in the *p*-hydroxybenzoic acid directed the narrower rim of the CD cavity.[41c]

The nitro-substituted benzenes generally take the orientation that the nitro group is inserted into the cavity, mainly because the nitro group is strongly electron withdrawing.[18,42]

For binding of *p*-substituted phenols and *p*-substituted anilines with α- and β-CD, the dominant binding site is the *p*-substituent. The results are also in agreement with the studies by Connors et al.[11,23] These phenomena are mainly due to the small size, the high hydrophilicity, and the high electron-releasing properties of OH and NH_2 groups. ^1H and ^{13}C NMR experiments and CPK model studies supported this interpretation.[18,33]

Interestingly, for inclusion complexation of α-CD with *p*-fluorophenol, $K_F = 51.4$ L/mol and $K_{OH} = 14.1$ L/mol from calculation by the models. It indicates that F mainly locates inside the α-CD cavity and OH outside the α-CD cavity. This finding is not in agreement with the report by Shibakami and Sekiya based on their X-ray study.[43] However, it has been noted that the guest orientation in the solid state is different within the solution. A very recent NMR study indicated that in aqueous solution multiple configurations are possible in a supramolecular complex and hence should be given much caution.[44]

4. Conclusion

Cyclodextrin complexation is one of the most valuable models for understanding the molecular recognition. We presented novel and reasonable equations to predict the driving forces and the guest orientation for the complexation of α- and β-CD with mono- and 1,4-disubstituted benzenes. The calculation results, in great agreement with the experimental studies, suggested that the van der Waals forces, hydrophobic interactions, and electronic effects comprise the primary driving forces for the inclusion process; the relative contributions of these driving forces are somewhat different for α-CD and β-CD systems.

Acknowledgment. We are grateful to Professor Jean-Marie Lehn for discussions and insights. This research was supported by the National Science Foundation of China.

References and Notes

(1) (a) Bender, M. L.; Komiyama, M. *Cyclodextrin Chemistry*; Springer-Verlag: Berlin, 1978. (b) Szejtli, J. *Cyclodextrin Technology*; Kluwer: Netherlands, 1988. (c) Wenz, G. *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 803. (d) Szejtli, J. *Chem. Rev.* **1998**, *98*, 1743.

(2) (a) Cranados, A.; de Rossi, R. H. *J. Am. Chem. Soc.* **1995**, *117*, 3690. (b) Fernandez, M. A.; de Rossi, R. H. *J. Org. Chem.* **1997**, *62*, 7554.

(3) (a) Schurig, V.; Nowotny, H.-P. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 939. (b) Li, S.; Purdy, W. C. *Chem. Rev.* **1992**, *92*, 1457. (c) Easton, C. J.; Lincoln, S. F. *Chem. Soc. Rev.* **1996**, *25*, 163.

(4) Hedges, A. R. *Chem. Rev.* **1998**, *98*, 2035.

(5) (a) Breslow, R. *Pure Appl. Chem.* **1994**, *66*, 1573. (b) Breslow, R. *Acc. Chem. Res.* **1995**, *28*, 146. (c) Breslow, R.; Dong, S. D. *Chem. Rev.* **1998**, *98*, 1997.

(6) (a) Tabushi, I.; Kiyosuke, Y.; Sugimoto, T.; Yamamura, R. *J. Am. Chem. Soc.* **1978**, *100*, 916. (b) Tabushi, I. *Acc. Chem. Res.* **1982**, *15*, 66.

(7) Komiyama, M.; Bender, M. L. *J. Am. Chem. Soc.* **1978**, *100*, 2259.

(8) (a) Blokzijl, W.; Engberts, J. F. B. N. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1545. (b) Ross, P. D.; Rekharsky, M. V. *Biophys. J.* **1996**, *71*, 2144.

(9) (a) Kitagawa, M.; Hoshi, H.; Sakurai, M.; Inoue, Y.; Chujo, R. *Bull. Chem. Soc. Jpn.* **1988**, *61*, 4225. (b) Kano, K.; Mori, K.; Uno, B.; Goto, M.; Kubota, T. *J. Am. Chem. Soc.* **1990**, *112*, 8645.

(10) (a) Davies, D. M.; Deary, M. E. *J. Chem. Soc., Perkin Trans. 2* **1995**, 1287. (b) Davies, D. M.; Deary, M. E.; Wealleans, D. I. *J. Chem. Soc., Perkin Trans. 2* **1998**, 193.

(11) (a) Connors, K. A. *Chem. Rev.* **1997**, *97*, 1325. (b) Rekharsky, M. V.; Inoue Y. *Chem. Rev.* **1998**, *98*, 1875.

(12) (a) Lipkowitz, K. B. *Chem. Rev.* **1998**, *98*, 1829. (b) Pérez, F.; Jaime, C.; Sánchez-Ruiz, X. *J. Org. Chem.* **1995**, *60*, 3840. (c) Alvira, E.; Mayoral, J. A.; García, J. I. *Chem. Phys. Lett.* **1997**, *271*, 178. (d) Huang, M. J.; Watts, J. D.; Bodor, N. *Int. J. Quantum. Chem.* **1997**, *65*, 1135. (e) Madrid, J. M.; Mendicuti, F.; Mattice, W. L. *J. Phys. Chem. B* **1998**, *102*, 2037.

(13) Park, J. H.; Nah, T. H. *J. Chem. Soc., Perkin Trans. 2* **1994**, 1359.

(14) Davies, D. M.; Savage, J. R. *J. Chem. Res. (S)* **1993**, 94; *J. Chem. Res. (M)* **1993**, 660.

(15) (a) Guo, Q.-X.; Luo, S.-H.; Zheng, X.-Q.; Liu, Y.-C. *Chin. Chem. Lett.* **1996**, *7*, 767. (b) Guo, Q.-X.; Luo, S.-H.; Liu Y.-C. *J. Inclusion Phenom.* **1998**, *30*, 173.

(16) (a) Rekharsky, M. V.; Schwarz, F. P.; Tewari, Y. B.; Goldberg, R. N. *J. Phys. Chem.* **1994**, *98*, 10282. (b) Rekharsky, M. V.; Mayhew, M. P.; Goldberg, R. N.; Ross, P. D.; Yamashoji, Y.; Inoue, Y. *J. Phys. Chem. B* **1997**, *101*, 87

(17) (a) Guo, Q.-X.; Luo, S.-H.; Wang, H.; Zhang, M.-S.; Liu, Y.-C. *J. Chem. Res. (S)* **1996**, 38; *Chin. Chem. Lett.* **1996**, *7*, 285. (b) Guo, Q.-X.; Liu, L.; Cai, W.-S.; Jiang, Y.; Liu, Y.-C. *Chem. Phys. Lett.* **1998**, *290*, 514. (c) Liu, L., Guo, Q.-X. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 133.

(18) (a) Inoue, Y.; Okuda, H.; Miyata, Y.; Chujo, R. *Carbohydr. Res.* **1984**, *125*, 65. (b) Yamamoto, Y.; Onda, M.; Takahashi, Y.; Inoue, Y.; Chujo, R. *Carbohydr. Res.* **1988**, *182*, 41. (c) Sakurai, M.; Kitagama, M.; Hoshi, H.; Inoue, Y.; Chujo, R. *Crabohydr. Res.* **1990**, *198*, 181

(19) (a) Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; Wiley: New York, 1979. (b) Hansch, C.; Leo, A.; Taft, R. W. *Chem. Rev.* **1991**, *91*, 165.

(20) (a)Vose, M. D. *Artif. Intell.* **1991**, *50*, 385. (b) Forrest, S. *Science* **1993**, 872.

(21) (a) Apostolakis, J.; Pluckthun, A.; Caflisch, A. *J. Comput. Chem.* **1998**, *19*, 21. (b) Topaler, M. S.; Truhlar, D. G.; Chang, X. Y.; Piecuch, P.; Polanyi, J. C. *J. Chem. Phys.* **1998**, *108*, 5349. (c) Mitchell, B. E.; Jurs, P. C. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 489. (d) Hasegawa, K.; Funatsu, K. *J. Mol. Struct. (THEOCHEM)* **1998**, *425*, 255. (e) Eliseev, A. V.; Nelen, M. I. *Chem. Eur. J.* **1998**, *4*, 825. (f) Vedani, A.; Dobler, M.; Zbinden, P. *J. Am. Chem. Soc.* **1998**, *120*, 4471.

(22) (a) Sanemasa, I.; Mizoguchi, T.; Deguchi, T. *Bull. Chem. Soc. Jpn.* **1984**, *57*, 1358. (b) Sanemasa, I.; Akamine, Y. *Bull. Chem. Soc. Jpn.* **1987**, *60*, 2059. (c) Takuma, T.; Deyuchi, T.; Sanemasa, I. *Bull. Chem. Soc. Jpn.* **1990**, *63*, 1246; **1991**, *64*, 480; **1991**, *64*, 1979.

(23) (a) Connors, K. A.; Lin, S.-F.; Wong, A. B. *J. Pharm. Sci.* **1982**, *71*, 217. (b) Wong, A. B.; Lin, S.-F.; Connors, K. A. *J. Pharm. Sci.* **1983**, *72*, 388. (c) Lin, S.-F.; Connors, K. A. *J. Pharm. Sci.* **1983**, *72*, 1333. (d) Connors, K. A.; Pendergast, D. J. *J. Am. Chem. Soc.* **1984**, *106*, 7607.

(24) Krysl, S.; Smolková-Keulemansová, E. *J. Chromatogr.* **1985**, *349*, 167.

(25) Bertrand, G. L.; Faulkner, J. R.; Han, S. M.; Armstrong, D. W. *J. Phys. Chem.* **1989**, *93*, 6863.

(26) (a) Harata, K. *Bioorg. Chem.* **1981**, *10*, 255. (b) Harata, K. *J. Inclusion Phenom.* **1992**, *13*, 77.

(27) Kawaguchi, Y.; Tanaka, M.; Nakae, M.; Funazo, K.; Shono, T. *Anal. Chem.* **1983**, *55*, 1852.

(28) Buvári, A.; Barcza, L. *J. Chem. Soc., Perkin 2* **1988**, 543.

(29) Hoshino, M.; Imamura, M.; Ikehara, K.; Hama, Y. *J. Phys. Chem.* **1981**, *85*, 1820.

(30) Kamiya, M.; Mitsuhashi, S.; Makino, M.; Yoshioka, H. *J. Phys. Chem.* **1992**, *96*, 95.

(31) Sumer, E.; Kobu, M.; Kurvits, M. *Thermochim. Acta* **1990**, *170*, 89.

(32) Shimizu, H.; Kaito, A.; Hatano, M. *Bull. Chem. Soc. Jpn.* **1979**, *52*, 2678.

(33) Rudiger, V.; Eliseev, A.; Simova, S.; Schneider, H.-J.; Blandamer, M. J.; Cullis, P. M.; Meyer, A. J. *J. Chem. Soc., Perkin 2* **1996**, 2119.
(34) Liu, Y.; You, C.-C.; Han, B.-H.; Li, Y.-M.; Chen, R.-T. *Chem. J. Chin. University* **1997**, *18*, 1316.
(35) Matsui, Y.; Nishioka, T.; Fujita, T. *Top. Current Chem.* **1985**, *128*, 61.
(36) Taraszewska, J.; Piasecki, A. K. *J. Electronal. Chem.* **1987**, *226*, 137.
(37) Pitchumani, K.; Vellayappan, M. *J. Inclusion Phenom.* **1992**, *14*, 157.
(38) (a) Easton, C. J.; Kassara, S.; Lincoln, S. F.; May, B. L. *Aust. J. Chem.* **1995**, *48*, 269. (b) Gadosy, T. A.; Tee, O. S. *Can J. Chem.* **1996**, *74*, 745. (c) Takahashi, T. *Chem. Rev.* **1998**, *98*, 2013.
(39) (a) Inoue, Y. *Ann. Rep. NMR Spectrosc.* **1993**, *27*, 59. (b) Schneider, H. J.; Hacket, F.; Rudiger, V.; Ikeda, H. *Chem. Rev.* **1998**, *98*, 1755.
(40) Lichtenthaler, F. W.; Immel, S. *Starch/Staerke* **1996**, *48*, 145.
(41) (a) Gelb, R. I.; Schwartz, L. M.; Johnson, R. F.; Laufer, D. A. *J. Am. Chem. Soc.* **1979**, *101*, 1869. (b) Salvatierra, D.; Jaime, C.; Virgili, A.; Sánchez-Ferrando, F. *J. Org. Chem.* **1996**, *61*, 9578. (c) Inoue, Y.; Hoshi, H.; Sakurai, M.; Chujo, R. *J. Am. Chem. Soc.* **1985**, *107*, 2319.
(42) (a) Wood, D. J.; Hruska, F. E.; Saenger, W. *J. Am. Chem. Soc.* **1977**, *99*, 1735. (b) Bergeron, R. J.; Channing, M. A.; Gibeily, G. J.; Pillor, D. M. *J. Am. Chem. Soc.* **1977**, *99*, 5146.
(43) Shibakami, M.; Sekiya, A. *J. Chem. Soc., Chem. Commun.* **1992**, 1742.
(44) Alderfer, J. L.; Eliseev, A. V. *J. Org. Chem.* **1997**, *62*, 8225.