A Nonlinear Group Contribution Method for Predicting the Free Energies of Inclusion Complexation of Organic Molecules with α- and β-Cyclodextrins

Takahiro Suzuki†

Chemical Resources Laboratory, Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama 226-8503, Japan

Received April 17, 2001

A group-contribution method was developed for calculating the binding constants or the free energies of complexation between native α- or β-cyclodextrin (CD) and organic guest molecules. The nonlinear models for binary (1:1 stoichiometry) complexes of α- and β-CDs were derived with squared correlation coefficients (r^2) of 0.868 and 0.917 based on a database consisting of 102 and 218 diverse guest molecules, respectively. The parameters used in the models are first-order molecular connectivity index as a measure of molecular bulk and atom/group counts in the guest molecules. The models allow accurate estimations for the wide range of guests containing C, H, N, O, S, and/or all halogens by summing the contribution values of each defined group present in the chemical structure of the guest along with guest's molecular size factors (linear and square terms) and then the summation to a constant coefficient value. The predictive performance of the models was tested by extra set of 27 compounds which were not included in the original data set. The predicted values by the models are in good agreement with the experimentally determined data.

INTRODUCTION

Cyclodextrins (CDs) are cyclic carbohydrates made of six, seven, or eight glucose molecules joined in a ring, to form α-, β-, and γ-CD, respectively. The CDs all form cylindrical or doughnut-shaped molecules with their hydroxyl groups on the outside of the molecule and a relatively nonpolar hole down the middle. This hole can accommodate another molecule, called "guest", and form noncovalent host−guest inclusion complexes with a variety of organic and inorganic molecules *via* molecular recognition in aqueous solution.[1] This allows them to be used in a wide range of applications in industrial, pharmaceutical, agricultural, and other related fields, including improving the solubility and stability (against the effects of light, heat, and oxidation) of drugs and selectively binding materials that fit into the central hole in affinity purification and chromatography methods.[2−4] The CDs also provide an excellent model system mimicking the substrate-specific interaction of enzymes.[2]

A variety of intermolecular interactions and solvation effects were proposed to explain the stability of the inclusion complexes formed with a great diversity of substrates. Hydrophobic desolvation and van der Waals interactions are among the most significant driving forces for CD complexation besides hydrogen bonding between polar groups of the substrate and the OH groups of the macrocycles. In addition, release of "high energy water" from the CD cavity, relief of conformational strain energy possessed by the uncomplexed CD, and molecular shape and size of guest have been postulated as key driving forces.[1] However, there have been many uncertainties on the mechanisms for the CD inclusion complexation.[1,5]

Until now, to identify the significant factors contributing the host−guest interaction and predict the thermodynamic stability of CD complexes, computational chemistry and QSAR/QSPR techniques, molecular mechanics computation,[6] linear regression,[7−10] partial least squares,[11] and artificial neural networks[12] have been applied. In a previous paper,[13] traditional QSAR and the 3D-QSAR method of CoMFA[14] were applied to derive quantitative relationships between the binding constants and guest molecular structures for natural β-CD, modified α-, β-, and γ-CD that bear one *p*-(dimethylamino)benzoyl moiety. However, the applicability range of these models to predict the complexation properties of guest molecules is rather limited. Experimental determination of the complex binding constant is often difficult and time-consuming because of the low solubility of the guest molecules in aqueous solution. For instance, 10 days were required for the equilibrium system of digitoxin: β-CD.[15] The divergence of the reported binding constants is often observed. Therefore, a reliable and convenient method for predicting the thermodynamic stability of CD complexes is desirable.

The purpose of the present study is to propose a group-contribution model (GCM) for predicting the free energies of complexation between guest molecules with α- and β-CDs based on a collection of reported binding data. GCMs or additive schemes for the estimation of thermochemical quantities have had wide usage over a long period.[16] To our knowledge, development of GCMs to the study of inclusion complexes of CDs has not been reported in the literature.

METHODS

Data Sets. The thermodynamic stability of a CD inclusion complex is generally expressed in terms of binding constant (or stability constant) of the complex, *K*. The 218 organic

† Corresponding author phone and fax: (+81 45) 924-5255; e-mail: tsuzuki@res.titech.ac.jp.

10.1021/ci010295f CCC: $20.00
Published on Web 08/11/2001

Table 1. Experimental and Calculated Values of Free Energies of Complexation (kJ/mol) and Estimation Error

			α-CD:guest system				β-CD:guest system			
no.	guest	formula	ΔGobsd	ΔGcalcd	Δ	ref	ΔGobsd	ΔGcalcd	Δ	ref
1	carbon tetrachloride	CCl4	−9.25	−10.88	1.63	20	−12.56	−11.78	−0.78	9
2	chloroform	CHCl3	−9.25	−11.25	2.00	20	−8.16	−6.80	−1.36	9
3	methanol	CH4O	0.17	−0.58	0.75	7	2.80	2.96	−0.16	7
4	acetonitrile	C2H3N	−4.28	−5.95	1.67	20	1.54	1.07	0.47	9
5	acetaldehyde	C2H4O					3.65	2.04	1.61	9
6	ethanol	C2H6O	−4.28	−3.52	−0.76	7	0.17	−0.16	0.33	7
7	1,2-ethanediol	C2H6O2	0.57	−0.47	1.04	7	1.08	0.67	0.41	7
8	acetone	C3H6O					−2.40	−1.77	−0.63	9
9	1-propanol	C3H8O	−7.82	−7.07	−0.75	7	−3.26	−3.37	0.11	7
10	2-propanol	C3H8O	−3.94	−3.70	−0.24	7	−3.60	-3.84	0.24	9
11	1,3-propanediol	C3H8O2	−3.03	−3.80	0.77	7	−3.82	−2.37	−1.45	7
12	tetrahydrofuran	C4H8O					-8.39	−9.16	0.77	9
13	cyclobutanol	C4H8O	−9.08	−6.81	−2.27	7	−6.74	−5.69	−1.05	7
14	1-butanol	C4H10O	−11.13	−10.40	−0.73	7	−6.99	−6.41	−0.58	17
15	2-butanol	C4H10O	−8.10	−7.44	−0.66	7	−6.79	−7.02	0.23	7
16	2-methyl-1-propanol	C4H10O	−8.22	−7.44	−0.78	7	−9.25	−7.02	-2.23	7
17	2-methyl-2-propanol	C4H10O	−3.65	−3.01	−0.64	7	−9.59	−9.18	−0.41	7
18	1,4-butanediol	C4H10O2	−5.20	−6.88	1.68	7	−3.65	−5.25	1.60	7
19	diethylamine	C4H11N					−7.77	−8.15	0.38	9
20	cyclopentanol	C5H10O	−9.48	−9.32	−0.16	7	−11.87	−9.27	−2.60	7
21	1-pentanol	C5H12O	−14.33	−13.49	−0.84	7	−10.29	−9.29	-1.00	17
22	2-pentanol	C5H12O	−12.16	−10.59	−1.57	7	−8.51	−9.95	1.44	7
23	3-pentanol	C5H12O	−11.08	−10.90	−0.18	7	−7.71	−10.01	2.3.0	7
24	2-methyl-1-butanol	C5H12O	−11.64	−10.90	−0.74	7	−11.87	−10.01	−1.86	7
25	2-methyl-2-butanol	C5H12O	−8.73	−6.80	−1.93	7	−10.90	−12.31	1.41	7
26	3-methyl-1-butanol	C5H12O	−10.67	−10.59	−0.08	7	−12.84	−9.95	−2.89	7
27	3-methyl-2-butanol	C5H12O	−7.25	−7.82	0.57	7	−10.96	−10.62	−0.34	7
28	2,2-dimethyl-1-propanol	C5H12O	−8.39	−6.80	−1.59	17	−15.48	−12.31	−3.17	17
29	1,5-pentanediol	C5H12O2	−8.56	−9.74	1.18	7	−6.97	−7.97	1.00	7
30	1,4-dibromobenzene	C6H4Br2	−17.07	−19.67	2.60	20	−16.98	−15.60	−1.38	12
31	1,4-diiodobenzene	C6H4I2					−18.12	−19.34	1.22	12
32	3,5-dibromophenol	C6H4Br2O	−17.81	−17.54	−0.27	7	−14.61	−16.48	1.87	7
33	3,5-dichlorophenol	C6H4Cl2O	−14.44	−10.56	−3.88	7	−11.82	−13.61	1.79	7
34	1-chloro-4-nitrobenzene	C6H4ClNO2	−10.45	−13.40	2.95	20	−12.27	−13.31	1.04	12
35	fluorobenzene	C6H5F	−8.73	− 7.98	−0.75	20	−11.18	−11.27	0.09	12
36	bromobenzene	C6H5Br	−15.47	−16.68	0.21	20	−14.28	−13.57	−0.71	12
37	iodobenzene	C6H5I	−17.35	−19.64	2.29	20	−16.71	−15.44	−1.27	12
38	3-fluorophenol	C6H5FO					−9.70	−12.25	2.55	7
39	4-fluorophenol	C6H5FO	−5.20	− 5.98	0.78	7	−9.87	−12.25	2.38	7
40	3-chlorophenol	C6H5ClO	−12.39	−10.19	−2.20	7	−13.01	−13.11	0.10	7
41	4-chlorophenol	C6H5ClO	−13.99	−10.19	−3.80	7	−14.92	−13.11	−1.81	12
42	3-bromophenol	C6H5BrO	−14.04	−13.68	−0.36	7	−14.33	−14.55	0.22	7
43	4-bromophenol	C6H5BrO	−16.44	−13.68	−2.76	7	−15.12	−14.55	−0.57	12
44	3-iodophenol	C6H5IO	−17.81	−17.64	−0.17	7	−16.73	−16.42	−0.31	7
45	4-iodophenol	C6H5IO	−19.01	−17.64	−1.37	7	−17.00	−16.42	−0.58	12
46	nitrobenzene	C6H5NO2	−17.81	−13.22	−4.59	7	−11.64	−12.95	1.31	9
47	4-nitrophenol	C6H5NO3	−12.55	−10.90	−1.65	17	−14.76	−13.69	−1.07	12
48	benzene	C6H6	−14.35	−11.53	−2.82	17	−12.72	−11.44	−1.28	17
49	phenol	C6H6O	−6.17	−9.68	3.51	7	−11.28	−12.52	1.24	12
50	hydroquinone	C6H6O2	−6.85	−7.69	0.84	20	−11.72	−13.49	1.78	12
51	4-nitroaniline	C6H6N2O2	−15.47	−15.20	−0.27	20	−14.18	−13.02	−1.16	12
52	aniline	C6H7N					−9.13	−11.84	2.71	9
53	sulfanilamide	C6H8N2O2S					−15.79	−14.81	−0.98	10
54	cyclohexanol	C6H12O	−10.33	−11.60	1.27	17	−15.27	−12.68	−2.59	17
55	1-hexanol	C6H14O	−16.84	−16.34	−0.50	7	−13.31	−12.01	−1.30	17
56	2-hexanol	C6H14O	−14.56	−13.52	−1.04	7	−11.30	−12.71	1.41	7
57	2-methyl-2-pentanol	C6H14O					−11.36	−15.32	3.96	7
58	3-methyl-3-pentanol	C6H14O					−12.27	−15.23	2.96	7
59	4-methyl-2-pentanol	C6H14O	−9.82	−10.67	0.85	7	−11.64	−13.40	1.76	7
60	3,3-dimethyl-2-butanol	C6H14O	−7.42	−7.17	−0.25	7	−15.70	−15.85	0.15	7
61	1,6-hexanediol	C6H14O2	−11.25	−12.37	1.12	7	−9.65	−10.52	0.87	7
62	benzonitrile	C7H5N					−12.74	−9.83	−2.91	12
63	benzothiazole	C7H5NS					−13.59	−13.02	−0.57	11
64	4-nitrobenzoic acid	C7H5NO4	−12.97	−16.72	3.75	17	−13.36	−11.17	−2.19	12
65	benzaldehyde	C7H6O	−19.35	−15.76	−3.59	20	−10.16	−8.85	−1.31	9
66	benzoic acid	C7H6O2	−17.15	−16.61	−0.54	17	−12.13	−10.77	−1.36	17
67	4-hydroxybenzaldehyde	C7H6O2	−9.99	−13.58	3.59	7	−9.99	−9.69	−0.30	7
68	4-hydroxybenzoic acid	C7H6O3	−17.52	−14.28	−3.24	20	−12.54	−11.52	−1.02	12
69	benzyl chloride	C7H7Cl					−13.95	−14.73	0.78	12
70	toluene	C7H8	−12.27	−11.87	−0.40	20	−11.93	−12.49	0.56	9
71	benzyl alcohol	C7H8O	−7.59	−8.18	0.59	7	−9.76	−11.05	1.29	9
72	anisole	C7H8O	−12.22	−10.58	−1.64	20	−13.24	−12.43	−0.81	12
73	*m*-cresol	C7H8O	−6.68	−9.87	3.19	18	−11.30	−13.47	2.17	7
74	*p*-cresol	C7H8O	−8.62	−9.87	1.25	7	−13.68	−13.47	−0.21	12
75	4-methoxyphenol	C7H8O2	−9.08	−8.39	−0.69	7	−12.61	−13.27	0.66	12

Table 1 (Continued)

			α-CD:guest system				*β*-CD:guest system			
no.	guest	formula	ΔGobsd	ΔGcalcd	Δ	ref	ΔGobsd	ΔGcalcd	Δ	ref
76	3-methoxyphenol	C7H8O2	−5.88	−8.39	2.51	7	−12.05	−13.27	1.22	7
77	4-hydroxybenzyl alcohol	C7H8O2					−12.34	−11.89	−0.45	12
78	hydrochlorothiazide	C7H8ClN3O4S2					−10.04	−10.56	0.52	10
79	*N*-methylaniline	C7H9N					−12.07	−12.79	0.72	12
80	1-butylimidazole	C7H12N2	−12.77	−14.72	0.95	7	−12.50	−14.70	2.20	18
81	1-heptanol	C7H16O	−19.18	−18.97	−0.21	7	−16.27	−14.56	−1.71	7
82	phenylacetylene	C8H6					−13.48	−13.48	0.00	12
83	thianaphthene	C8H6S					−18.44	−15.72	−2.72	11
84	4-fluorophenyl acetate	C8H7FO2					−12.05	−13.02	0.97	7
85	3-fluorophenyl acetate	C8H7FO2	−6.56	−6.53	−0.03	7	−10.90	−13.02	2.12	7
86	4-chlorophenyl acetate	C8H7ClO2	−10.28	−10.74	0.46	20	−14.27	−13.89	−0.38	7
87	3-chlorophenyl acetate	C8H7ClO2	−12.90	−10.74	−2.16	7	−13.93	−13.89	−0.04	7
88	4-bromophenyl acetate	C8H7BrO2					−15.30	−15.32	−0.02	7
89	3-bromophenyl acetate	C8H7BrO2	−15.98	−14.23	−1.75	7	−15.24	−15.32	0.08	7
90	4-iodophenyl acetate	C8H7IO2					−17.13	−17.19	0.06	7
91	3-iodophenyl acetate	C8H7IO2	−18.95	−18.19	−0.76	7	−17.53	−17.19	−0.34	7
92	4-nitrophenyl acetate	C8H7NO4	−10.46	−10.27	−0.19	17	−12.16	−13.64	1.48	7
93	acetophenone	C8H8O					−12.96	−12.18	−0.78	12
94	phenyl acetate	C8H8O2	−7.02	−10.74	3.72	7	−11.99	−13.65	1.66	7
95	methyl benzoate	C8H8O2					−14.28	−13.67	−0.61	12
96	3-hydroxyacetophenone	C8H8O2					−11.76	−12.92	1.16	7
97	4-hydroxyacetophenone	C8H8O2					−12.44	−12.92	0.48	12
98	acetoanilide	C8H9NO					−12.54	−12.13	−0.41	12
99	*p*-xylene	C8H10	−10.33	−12.06	1.73	20	−13.58	−13.45	−0.13	12
100	ethylbenzene	C8H10	−11.65	−14.78	3.13	20	−14.77	−15.09	0.32	12
101	phenetole	C8H10O					−14.20	−14.82	0.62	12
102	2-phenylethanol	C8H10O					−12.27	−13.44	1.17	7
103	3-ethylphenol	C8H10O	−10.84	−12.59	1.75	7	−14.84	−15.93	1.09	7
104	4-ethylphenol	C8H10O	−12.16	−12.59	0.43	7	−15.37	−15.93	0.56	12
105	4-ethoxyphenol	C8H10O2	−10.67	−10.60	−0.07	7	−13.30	−15.52	2.22	7
106	3-ethoxyphenol	C8H10O2	−9.65	−10.60	0.95	7	−13.41	−15.52	2.11	7
107	3,5-dimethoxyphenol	C8H10O3	−6.79	−6.30	−0.49	7	−13.36	−13.47	0.11	7
108	*N*-ethylaniline	C8H11N					−13.33	−15.18	1.85	12
109	*N*,*N*-dimethylaniline	C8H11N					−13.48	−13.48	0.00	12
110	barbital	C8H12N2O3					−10.19	−10.63	0.44	10
111	cyclooctanol	C8H16O	−12.84	−12.76	−0.08	7	−18.84	−18.69	−0.15	7
112	1-octanol	C8H18O	−21.69	−21.37	−0.32	7	−18.10	−16.95	−1.15	7
113	2-octanol	C8H18O	−17.98	−18.67	0.69	7	−17.87	−17.75	−0.12	7
114	quinoline	C9H7N	−8.33	−9.42	1.09	20	−12.10	−13.36	1.26	11
115	3-cyanophenyl acetate	C9H7NO2	−9.93	−8.26	−1.67	7	−8.51	−10.95	2.44	7
116	4-hydroxycinnamic acid	C9H8O3	−18.84	−20.13	1.29	20	−16.15	−16.47	0.32	10
117	ethyl benzoate	C9H10O2					−15.59	−15.76	0.17	12
118	4′-hydroxypropiophenone	C9H10O2					−15.01	−15.06	0.05	7
119	3′-hydroxypropiophenone	C9H10O2					−14.90	−15.06	0.16	7
120	*p*-tolyl acetate	C9H10O2	−11.19	−10.42	−0.77	20	−14.21	−14.25	0.04	7
121	3-methylphenyl acetate	C9H10O2	−8.39	−10.42	2.03	7	−12.61	−14.25	1.64	7
122	4-methoxyphenyl acetate	C9H10O3					−13.99	−13.56	−0.43	7
123	4-propylphenol	C9H12O	−15.70	−14.79	−0.91	7	−20.26	−18.19	−2.07	7
124	3-propylphenol	C9H12O	−14.33	−14.79	−0.46	7	−18.72	−18.19	−0.53	7
125	4-isopropylphenol	C9H12O					−20.43	−19.03	−1.40	7
126	3-isopropylphenol	C9H12O					−19.64	−19.03	−0.61	7
127	4-isopropoxyphenol	C9H12O2					−16.33	−18.50	2.17	7
128	2-norbornaneacetate	C9H14O2	−12.55	−12.96	0.41	17	−20.50	−18.18	−2.32	17
129	1-benzylimidazole	C10H10N2	−10.15	− 8.66	−1.49	18	−14.92	−17.34	2.42	18
130	*m*-methylcinnamic acid	C10H10O2	−23.58	−22.32	−1.26	20	−16.75	−16.45	−0.30	10
131	4-ethylphenyl acetate	C10H12O2					−16.15	−16.22	0.07	7
132	3-ethylphenyl acetate	C10H12O2	−10.90	−12.44	1.54	7	−15.30	−16.22	0.92	7
133	4-ethoxyphenyl acetate	C10H12O3					−14.50	−15.36	0.86	7
134	3-ethoxyphenyl acetate	C10H12O3	−9.87	−9.80	−0.07	7	−14.21	−15.36	1.15	7
135	allobarbital	C10H12N2O3					−11.32	−11.32	0.00	10
136	4-*n*-butylphenol	C10H14O	−18.61	−16.76	−1.85	7	−22.66	−20.28	−2.38	7
137	3-*n*-butylphenol	C10H14O	−16.56	−16.76	0.20	7	−21.46	−20.28	−1.18	7
138	3-isobutylphenol	C10H14O					−24.03	−21.16	−2.87	7
139	4-*sec*-butylphenol	C10H14O					−23.86	−21.17	−2.69	7
140	3-*sec*-butylphenol	C10H14O					−23.18	−21.17	−2.01	7
141	4-*tert*-butylphenol	C10H14O	−10.96	−11.03	0.07	20	−26.03	−23.86	−2.17	7
142	3-*tert*-butylphenol	C10H14O	−8.33	−11.03	2.70	20	−25.18	−23.86	−1.32	7
143	menadion	C11H8O2					−12.95	−12.50	−0.45	10
144	sulfapyridine	C11H11N3O2S					−15.43	−16.58	1.15	10
145	sulfamonomethoxine	C11H12N4O3S					−14.16	−11.79	−2.37	10
146	sulfisoxazole	C11H13N3O3S					−13.25	−13.04	−0.21	10
147	4-*n*-propylphenyl acetate	C11H14O2					−17.98	−18.02	0.04	7
148	3-*n*-propylphenyl acetate	C11H14O2	−13.82	−14.00	0.18	7	−18.72	−18.02	−0.70	7
149	4-isopropylphenyl acetate	C11H14O2					−16.44	−18.98	2.54	7
150	3-isopropylphenyl acetate	C11H14O2					−19.18	−18.98	−0.20	7

Table 1 (Continued)

			α-CD:guest system				*β*-CD:guest system			
no.	guest	formula	Δ*G*obsd	Δ*G*calcd	Δ	ref	Δ*G*obsd	Δ*G*calcd	Δ	ref
151	4-*n*-amylphenol	C11H16O	−19.18	−18.50	−0.68	7	−23.92	−22.21	−1.71	7
152	4-*tert*-amylphenol	C11H16O					−26.83	−25.90	−0.93	7
153	carbutamide	C11H17N3O3S					−13.08	−16.57	3.49	10
154	pentobarbital	C11H18N2O3					−17.22	−18.15	0.93	10
155	amobarbital	C11H18N2O3					−17.53	−18.24	0.71	17
156	thiopental	C11H18N2O2S					−18.71	−18.77	0.06	10
157	dibenzofuran	C12H8O					−16.95	−14.65	−2.30	11
158	dibenzothiophene	C12H8S					−19.87	−18.45	−1.42	11
159	phenazine	C12H8N2					−13.76	−12.77	−0.99	11
160	thianthrene	C12H8S2					−20.38	−22.01	1.63	11
161	carbazole	C12H9N					−13.93	−13.47	−0.46	11
162	phenoxazine	C12H9NO					−15.36	−13.23	−2.13	11
163	phenothiazine	C12H9NS					−15.59	−17.04	1.45	11
164	furosemide	C12H11ClN2O5S					−10.19	−11.38	1.19	10
165	phenobarbital	C12H12N2O3					−18.37	−14.74	−3.63	17
166	sulfisomidine	C12H14N4O2S					−12.02	−13.28	1.26	10
167	sulfamethomidine	C12H14N4O3S					−13.31	−11.30	−2.01	10
168	sulfadimethoxine	C12H14N4O4S					−12.89	−9.13	−3.76	10
169	4-*n*-butylphenyl acetate	C12H16O2					−20.66	−19.65	−1.01	7
170	3-*n*-butylphenyl acetate	C12H16O2	−16.27	−15.32	−0.95	7	−20.89	−19.65	−1.24	7
171	3-isobutylphenyl acetate	C12H16O2					−21.86	−20.67	−1.19	7
172	4-*tert*-butylphenyl acetate	C12H16O2					−21.98	−21.99	0.01	7
173	cyclobarbital	C12H16N2O3					−15.47	−17.25	1.78	10
174	hexobarbital	C12H16N2O3					−17.60	−16.06	−1.54	10
175	1-adamantaneacetate	C12H18O2	−12.97	−12.97	0.00	17	−24.69	−24.03	−0.66	17
176	acridine	C13H9N					−13.30	−15.47	2.17	11
177	phenanthridine	C13H9N					−14.67	−15.45	0.78	11
178	xanthene	C13H10O					−15.47	−16.92	1.45	11
179	*N*-phenylanthranilic acid	C13H11NO2					−16.53	−15.92	−0.61	17
180	mephobarbital	C13H14N2O3					−18.05	−15.73	−2.32	10
181	4-*n*-amylphenyl acetate	C13H18O2					−21.69	−21.13	−0.56	7
182	flufenamic acid	C14H10F3NO2					−17.70	−15.78	−1.92	17
183	meclofenamic acid	C14H11Cl2NO2					−15.27	−14.28	−0.99	17
184	nitrazepam	C15H11N3O3					−11.26	−10.66	−0.60	10
185	flurbiprofen	C15H13FO2					−21.07	−18.88	−2.19	10
186	sulfaphenazole	C15H14N4O2S					−13.42	−13.34	−0.08	10
187	bendroflumethiazide	C15H14F3N3O4S2					−10.83	−12.05	1.22	10
188	mefenamic acid	C15H15NO2					−14.24	−15.41	1.17	10
189	acetohexamide	C15H20N2O4S					−16.77	−15.60	−1.17	10
190	fludiazepam	C16H12ClFN2O					−13.31	−12.34	−0.97	10
191	nimetazepam	C16H13N3O3					−9.89	−11.08	1.19	10
192	fenbufen	C16H14O3					−15.02	−16.17	1.15	10
193	ketoprofen	C16H14O3					−16.31	−17.25	0.94	10
194	medazepam	C16H15ClN2					−13.72	−17.66	3.94	10
195	progabide	C17H16ClFN2O2					−14.48	−14.48	0.00	10
196	griseofulvin	C17H17ClO6					−8.39	−9.62	1.23	10
197	tolnaftate	C19H17NOS					−21.90	−21.84	−0.06	10
198	prostacyclin	C20H32O5					−16.79	−16.85	0.06	10
199	triamcinolone	C21H27FO6					−19.27	−18.40	−0.87	19
200	cortisone	C21H28O5					−19.10	−22.35	3.25	10
201	prednisolone	C21H28O5					−20.30	−20.59	0.29	19
202	hydrocortisone	C21H30O5					−20.58	−21.85	1.27	10
203	corticosterone	C21H30O4					−22.00	−21.62	−0.38	19
204	dexamethasone	C22H29FO5					−20.84	−22.43	1.59	10
205	betamethasone	C22H29FO5					−21.31	−22.43	1.12	19
206	paramethasone	C22H29FO5					−19.44	−17.37	−2.07	19
207	cortisone-21-acetate	C23H30O6					−20.65	−19.76	−0.89	19
208	prednisolone-21-acetate	C23H30O6					−21.47	−18.00	−3.47	19
209	hydrocortisone-21-acetate	C23H32O6					−20.05	−19.26	−0.79	19
210	fluocinolone acetonide	C24H30F2O6					−19.85	−18.13	−1.72	19
211	triamcinolone acetonide	C24H31FO6					−20.03	−20.86	0.83	19
212	spironolactone	C24H32O4S					−25.34	−25.34	0.00	19
213	dehydrocholic acid	C24H34O5					−19.29	−22.28	2.99	19
214	chenodeoxycholic acid	C24H40O4					−24.90	−26.04	1.14	19
215	ursodeoxycholic acid	C24H40O4					−25.71	−26.04	0.33	19
216	cholic acid	C24H40O5					−19.97	−21.26	1.29	19
217	hydrocortisone-17-butyrate	C25H36O6					−18.43	−17.01	−1.42	19
218	cinnarizine	C26H28N2					−20.77	−20.66	−0.11	10

guests used in developing the GCMs, their experimental free energies of complexation for α- and *β*-CDs at 25 °C, and literature sources are presented in Table 1.[7,9,10−12,17−20] For the α-CD system, available data is limited since its cavity size is small, and resulting molecular recognition is effective to bind the smaller molecules such as aliphatic alcohols and monosubstituted benzene derivatives but showed no response to large cyclic compounds. Otherwise a missing value in the database means that a particular compound was not tested or reliable experimental value could not be found.

The set of guests is structurally diverse and includes a large number of classes of organic compounds: aromatic hydrocarbons, alcohols, phenols, ethers, aldehydes, ketones, acids, esters, nitriles, anilines, halogenated compounds, heterocycles, nitro, and sulfur compounds. Besides structurally complex compounds, such as steroids and barbitals are included. The measured K data were converted into free energies ΔG ($= -RT \ln K$) in kJ/mol.

Model Development. The complex stability, ΔG, can be expressed as a function of a set of group contributions G_1, G_2, ..., G_m

$$\Delta G = -\mathrm{RT} \ln K = f(G_1, G_2, ..., G_m) \quad (1)$$

where m is the number of group-contributions considered in a particular group-contribution table. The group or fragment contributions are numerical quantities associated with single atoms, atom pairs, or multiatom groups or functional groups.[21] The GCM function may be linear or nonlinear. In our previous study,[13] a nonlinear dependency of binding constants on the zeroth and/or first-order molecular connectivity index ($^0\chi$, $^1\chi$)[22] as a measure of size was found for the natural β-CD:guest system. The free energy of complexation according to our approach can, therefore, be expressed by follows

$$\Delta G = c_0 + c_1 D + c_2 D^2 + \sum n_i G_i \quad (2)$$

where c_0 is a model-specific constant, D is a molecular descriptor as a measure of bulkiness, c_1 and c_2 are the regression coefficients for D and its quadratic term, respectively, G_i is the contribution of the group of type i, n_i is the number of times this group occurs in the molecule, and the summation is carried over all type i. Regression coefficients were estimated by the method of least squares, using Excel Multivariate Analysis v3.0 (Esumi Co., Ltd., Tokyo, Japan) on a microcomputer running Windows 95 as its operating system.

The quality of this model (i.e., the precision of its predictions and their reliability) is critically dependent on the definition of a set of groups and on the values used. Any structure can be divided into groups, but their number can be considerable. Generally, as the size of the groups increases, more precise estimates would be available. However, the sheer volume of experimental data required to determine their contribution values is a prohibitive factor from a viewpoint of statistical reliability. The definition of what constitutes the best value of a group contribution is not very rigid. After considering the available number of data points, easy identification of constitutional groups in a molecule, and implicit inclusion of the constitutive factors, a similar set of groups used earlier by Lydersen for estimating critical properties[23] was employed in this study. This type of group was successfully applied to estimate 11 kinds of fundamental physical properties of pure materials as a common set of structural groups by Joback and Reid.[24]

RESULTS AND DISCUSSION

α-CD:Guest System. The experimentally determined ΔG values at 25 °C for 1:1 complexes between natural α-CD and 102 organic compounds containing C, H, N, O, and all halogens were used to obtain the model. As readily available and computable descriptors for D, $^0\chi$, $^1\chi$ and molecular weight (Mw) were attempted to introduce. Although other geometrical or topological descriptors such as molecular volume, molecular surface area, and molecular refraction could be also used as candidates for D, the former three descriptors are much easier to prepare for any compounds without calculation errors. The best GCM using $^1\chi$ was given by the following equation

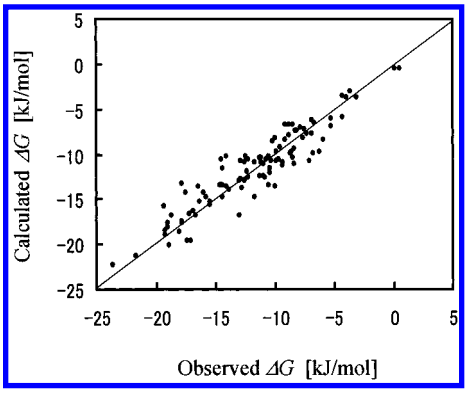

Figure 1. Correlation between the calculated (eq 3) versus experimental free energies of binding for α-cyclodextrin with 102 guest molecules.

$$\Delta G[\mathrm{kJ/mol}] = -1.13 - 10.76^1\chi + 0.463^1\chi^2 + \sum n_i G_i \quad (3)$$

$$(n = 102, r^2 = 0.868, SD = 1.97, ME = 1.32)$$

where n = number of compounds, r^2 = explained variance, SD = standard deviation, and ME = absolute mean error. Table 2 lists the coefficient or contribution (G_i) for 22 groups together with the coefficients for $^1\chi$ (−10.76) and $^1\chi^2$ (0.463) parameters. The statistical quality of the above model was slightly inferior ($r^2 = 0.828$) without the two terms for $^1\chi$ and $^1\chi^2$. Good models were also obtained with $^0\chi$ ($r^2 = 0.848$) or Mw ($r^2 = 0.854$), respectively, instead of $^1\chi$. A plot of the experimental ΔG values against the calculated ΔG values based on eq 3 for all the compounds in the data database is shown in Figure 1.

***β*-CD:Guest System.** On the basis of measured ΔG values at 25 °C for 1:1 complexes of β-CD with 218 organic compounds containing C, H, N, O, S, and halogens, the GCM was obtained as follows:

$$\Delta G[\mathrm{kJ/mol}] = 8.48 - 3.53^1\chi + 0.326^1\chi^2 + \sum n_i G_i \quad (4)$$

$$(n = 218, r^2 = 0.917, SD = 1.66, ME = 1.20)$$

The contributions (G_i) for 39 groups are also listed in Table 2. The experimental values are plotted versus the predicted ones in Figure 2. Omission of two parameters, $^1\chi$ and $^1\chi^2$, gave a substantially inferior fit with $r^2 = 0.771$. The employment of $^0\chi$ or Mw in place of $^1\chi$ gave slightly inferior r^2 values of 0.908 and 0.907, respectively, as well as the case of α-CD.

Group Contributions to CD Inclusion Complexation. The importance of individual contributions to the complexation process can be related to the value of the respective regression coefficients in Table 2. The negative sign of the regression coefficient indicates that the groups should enhance complexation ability. For the case of β-CD:guest

Table 2. Structural Fragments and Their Contributions to the Free Energies of Compexiation

fragment or parameter	α-CD:guest system			β-CD:guest system		
	no. of compds	freq of occurrence	contribution	no. of compds	freq of occurrence	contribution
constant			−1.13			8.48
$^1\chi$	102	102	−10.76	218	218	−3.53
$^1\chi^2$	102	102	0.463	218	218	0.326
			Nonring Increments			
-CH_3	62	96	2.12	144	261	−3.18
-CH_2-	41	101	1.06	95	190	−1.98
>CH-	13	15	1.71	34	36	−1.69
>C<	7	7	2.70	13	13	−3.23
=CH_2				1	2	−1.49
=CH-				5	11	−2.41
=C<				3	3	−2.09
-C≡CH				1	1	−3.55
			Ring Increments			
-CH_2-	6	29	1.64	38	172	−2.67
>CH-	6	10	3.77	29	106	−0.44
>C<	1	1	5.59	30	68	−3.08
=CH-	65	277	2.95	165	783	−2.04
=C<	63	116	3.57	167	379	0.65
			Halogen Increments			
-F	3	3	6.01	16	21	−1.95
-Cl	8	14	1.81	16	23	−2.82
-Br	6	8	−1.69	7	9	−4.25
-I	4	4	−5.65	6	7	−6.12
			Oxygen Increments			
-OH (alcohol)	33	38	8.72	56	86	0.86
-OH (phenol)	31	32	4.31	44	45	−3.20
-O− (nonring)	7	8	5.25	19	25	0.67
-O− (ring)				8	8	−0.17
>C=O (nonring)				27	28	1.25
>C=O (ring)				30	53	-0.08
-CHO (aldehyde)	2	2	2.19	3	3	1.08
-COOH (acid)	5	5	3.93	18	18	−0.53
-COO− (ester)	16	16	10.84	35	35	0.05
			Nitrogen Increments			
-NH_2				15	16	−2.52
>NH (nonring)				18	20	0.32
>NH (ring)				14	23	1.01
>N− (nonring)				1	1	3.12
>N− (ring)				7	8	2.39
-N= (nonring)				1	1	1.68
-N= (ring)	3	5	5.93	19	26	0.66
-CN	2	2	7.35	3	3	0.11
-NO_2	6	6	7.32	8	8	−2.70
			Sulfur Increments			
-S−(nonring)				1	1	−1.86
-S−(ring)				5	6	−2.64
-SO_2-				13	15	−1.55
>C=S				2	2	−3.97

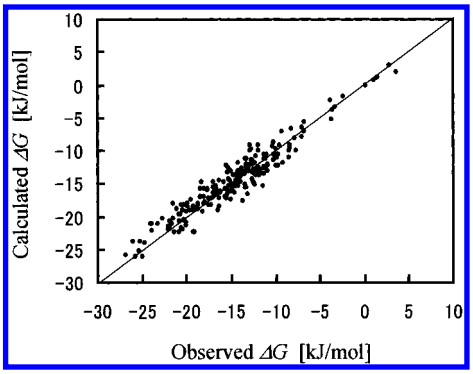

Figure 2. Correlation between the calculated (eq 4) versus experimental free energies of binding for β-cyclodextrin with 218 guest molecules.

system, the presence of carbon (except only =C < (ring)), halogen, and sulfur increments with negative coefficients results in an increase of the stability of the complex, while the presence of most oxygen and nitrogen containing groups (except −OH (phenol), −O−(ring), >C=O(ring), −COOH, −NH_2, and −NO_2) decreases the complex stability according to their positive coefficients. On the contrary, a markedly different trend can be found for the α-CD:guest system. The contribution values have positive signs with the exception of −Br and −Cl. The reason can be attributable to the difference in the number of possible configurations and resulting interaction effects between the guests and the host cavity in the two complex systems. The major difference in the composition of the driving forces for α- and β-CD can be caused by the cavity dimensions.

The significant contribution of van der Waals interactions can be assessed from the negative signs of the coefficients for the fragments containing carbon and halogens in the β-CD system. It seems that the contribution values for F, Cl, Br, and I in both the α-CD and the β-CD system are correlated with the effective van der Waals radii of their

Table 3. De Novo Pediction of Free Energies of Complexation (kJ/mol) Using Eqs 3 and 4

host	guest	formula	ΔGobsd	ΔGcalcd	Δ	ref
α-	pyridine	C5H5N	−12.55	−8.56	−4.00	17
α-	3-cyanophenol	C7H5NO	−9.99	−8.41	−1.58	7
α-	4-cyanophenol	C7H5NO	−10.62	−8.41	−2.21	7
α-	cycloheptanol	C7H14O	−10.84	−14.50	3.66	7
α-	3,5-dichlorophenyl acetate	C8H6Cl2O2	−14.16	−10.59	−3.57	7
α-	1-phenylimidazole	C8H8N2	−8.00	−7.08	−0.92	18
α-	3-methybenzoic acid	C8H8O2	−13.81	−16.47	2.66	17
α-	3,5-dimethylphenol	C8H10O	−6.85	−9.92	3.07	7
α-	3-methoxyphenyl acetate	C9H10O3	−7.88	−8.26	0.38	7
α-	3,5-dimethylphenyl acetate	C10H12O2	−8.79	− 9.97	1.18	7
α-	3,5-dimethoxylphenyl acetate	C10H12O4	−7.31	−4.96	−2.35	7
α-	L-α-*O*-benzylglycerol	C10H14O3	−6.66	−4.77	− 1.89	18
β-	2-methoxyethanol	C3H8O2	−1.26	−3.75	2.50	7
β-	cycloheptanol	C7H14O	−18.44	−15.92	− 2.52	7
β-	3-hydroxycinnamic acid	C9H8O3	−14.63	−16.46	1.83	10
β-	ethyl 4-hydroxybenzoate	C9H10O3	−17.19	−16.21	−0.98	10
β-	ethyl 4-aminobenzoate	C9H11NO2	−15.34	−15.53	0.19	10
β-	4-methylcinnamic acid	C10H10O2	−15.13	−16.44	1.31	10
β-	sulfadiazine	C10H10N4O2S	−14.40	−13.52	0.88	10
β-	L-α-*O*-benzylglycerol	C10H14O3	−12.03	−15.60	3.59	18
β-	sulfamerazine	C11H12N4O2S	−11.28	−13.28	2.00	10
β-	butyl 4-hydroxybenzoate	C11H14O3	−19.37	−19.64	0.27	10
β-	butyl 4-aminobenzoate	C11H15NO2	−18.24	−18.96	0.72	10
β-	benzidine	C12H12N2	−19.11	−19.25	0.14	10
β-	triflumizole	C15H15ClF3N3O	−15.19	−19.21	4.02	10
β-	diazepam	C16H13ClN2O	−13.31	−13.96	0.65	10
β-	prostaglandine E_2	C20H21O5	−17.64	−20.37	2.73	10
	absolute mean error				1.92	

atoms. A recent QSAR study on inclusion complexes between 16 *para*-substituted phenols and natural *β*-CD showed that the molar refractivity as a measure of van der Waals interactions represents an approximation for host/guest complementarity.[25] The study identified that hydrophobic and dipolar interactions were minor factors to binding for the phenolic series. For the binding of α- and *β*-CD with 24 monosubstituted benzenes, it was concluded that the inclusion by α-CD was predominantly driven by van der Waals force and with *β*-CD by both van der Waals force and hydrophobic interactions.[26]

The positive coefficients for oxygen and nitrogen containing groups in the *β*-CD system can be assigned to the possibility of the competitive interactions with the solvent as discussed by Park and Nah.[9] It is worth to note the difference of the signs of −OH in alcoholic and phenolic types in the *β*-CD system. In contrast, both have positive coefficients in the α-CD system. Only phenolic OH in the *β*-CD system has a negative coefficient and should have an increased complexation ability, suggesting the possibility to stabilize the complex by forming hydrogen bonding with the host's hydroxyl groups. However, the contribution by the phenolic OH moiety of the guests may not play a major role according to the magnitude of their coefficients.

The model can be further developed incrementally by determining the missing contributions for the α-CD:guest system and adding new substructures as experimental data become available.

Comparison with Molecular Modeling Approaches. The orientation and alignment of the guest molecule in the CD cavity is crucial in molecular modeling analysis or many 3D-QSAR methods. In fact, various probable starting orientations and positions of the guest molecule within the CD cavity were studied, and the lowest energy one was used in our previous CoMFA analyses.[13] For the case of disubstituted benzenes, two and three orientational isomers for the complexes with *p*- and *m*-substituted phenols, respectively, were postulated, and the observed binding constants were divided into the corresponding dissociation constants of the isomers involved.[7] In contrast, the present GCM approach does not depend on structural alignment of guest molecules to CDs, thus eliminating a substantial measure of conjecture.

Predictive Performance of the GCMs. The predictive capabilities of eqs 3 and 4 were tested by a *de novo* prediction, using the data set of host−guest systems not included in the deduction of the models, as shown in Table 3. The absolute mean error for the 27 compounds is 1.92. This value is somewhat larger than the values of 1.32 for eq 3 and 1.22 for eq 4. The results of the table confirm that the models have a good and reasonable predictive ability.

To illustrate a typical estimation of Δ*G*, we will explain the calculations necessary for 4-*tert*-butylphenyl acetate ($^1\chi$ = 6.3929): *β*-CD complex using the GCM. Using Table 2 and eq 4, we have

$$\Delta G = \text{equation constant} + (-3.53)(^1\chi) + (0.326)(^1\chi^2) + (-3.18)(\text{four } -CH_3) + (-3.23)\ (\text{one nonring } >C<) + (-2.04)\ (\text{four ring } =CH\text{-}) + (0.65)\ (\text{two ring } =C<) + (0.05)\ (\text{one } -COO\text{-}) = -21.99 \text{ kJ/mol}$$

The measured value is −21.98 kJ/mol as listed in Table 1.

CONCLUSIONS

Group-contribution models for the estimation of the thermodynamic stability of the complexation of a large variety of organic compounds with natural α- and *β*-cyclodextrins, respectively, were established. Our results demonstrate that a single set of chemical substructures and the descriptor incorporating molecular size, first-order molecular

connectivity index, allow an accurate prediction of the free energies of the complexation. The models proposed in this work provide a basis on the evaluation of the reliability of the experimentally determined data as well as the prediction for the complexation properties of guest molecules for which the experimental determination is difficult or exceedingly time-consuming.

REFERENCES AND NOTES

(1) Connors, K. A. The stability of cyclodextrin complexes in solution. *Chem. Rev.* **1997**, *97*, 1325−1357.

(2) Szejtli, J. *Cyclodextrin Technology*; Kluwer Academic Publishers: Dordrecht, 1988.

(3) Szejtli, J. Introduction and general overview of cyclodextrin chemistry. *Chem. Rev.* **1998**, *98*, 1743−1753.

(4) Hedges, A. R. Industrial applications of cyclodextrins. *Chem. Rev.* **1998**, *98*, 2035−2044.

(5) Lipkowitz, K. B. Applications of computational chemistry to the study of cyclodextrins. *Chem. Rev.* **1998**, *98*, 1829−1873.

(6) Pérez, F.; Jaime, C.; Sánchez-Ruiz, X. MM2 calculations on cyclodextrins: Multimodel inclusion complexes. *J. Org. Chem.* **1995**, *60*, 3840−3845.

(7) Matsui, Y.; Nishioka, T.; Fujita, T. Quantitative structure−reactivity analysis of the inclusion mechanism by cyclodextrins. *Topics Curr. Chem.* **1985**, *128*, 61−89.

(8) Davis, D. M.; Savage, J. R. Correlation analysis of the host−guest interaction of α-cyclodextrin and substituted benzenes. *J. Chem. Res. (S)* **1993**, 94−95.

(9) Park, J. H.; Nah, T. H. Binding forces contributing to the complexation of organic molecules with β-cyclodextrin in aqueous solution. *J. Chem. Soc., Perkin Trans. 2.* **1994**, 1359−1362.

(10) Klein, C. Th.; Polheim, D.; Viernstein, H.; Wolschann, P. A method for predicting the free energies of complexation between β-cyclodextrin and guest molecules. *J. Incl. Phenom. Macro. Chem.* **2000**, *36*, 409−423.

(11) Carpignano, R.; Marzona, M.; Cattaneo, E.; Quaranta, S. QSAR study of inclusion complexes of heterocyclic compounds with β-cyclodextrin. *Anal. Chim. Acta* **1997**, *348*, 489−493.

(12) Liu, L.; Guo, Q.-X. Wavelet neural network and its application to the inclusion of β-cyclodextrin with benzene derivatives. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 133−138.

(13) Suzuki, T.; Ishida, M.; Fabian, W. M. F. Classical QSAR and comparative molecular field analyses of the host−guest interaction of organic molecules with cyclodextrins. *J. Comput.-Aided Mol. Design* **2000**, *14*, 669−678.

(14) Cramer, R. D., III.; Patterson, D. E; Bunce, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959−5967.

(15) Yoshida, A.; Yamamoto, M.; Hirayama, F.; Uekawa, K. Improvement of chemical instability of digitoxin in aqueous solution by complexation with β-cyclodextrin derivatives. *Chem. Pharm. Bull.* **1988**, *36*, 4075−4080.

(16) Benson, S. W. *Thermochemical Kinetics*, 2nd ed.; Wiley: New York, 1976.

(17) Inoue, Y.; Hakushi, T.; Liu, Y.; Tong, L.-H.; Shen, B.-J.; Jin, D.-S. Thermodynamics of molecular recognition by cyclodextrins. 1. Calorimetric titration of inclusion complexation of naphthalenesulfonates with α-, β-, and γ-cyclodextrins: Enthalpy−entropy compensation. *J. Am. Chem. Soc.* **1993**, *115*, 475−481.

(18) Rekharsky, M. V.; Goldberg, R. N.; Schwarz, F. P.; Tewari, Y. B.; Ross, P. D.; Yamashoji, Y.; Inoue, Y. Thermodynamic and nuclear magnetic resonance study of the interactions of α- and β-cyclodextrin with model substances: phenethylamine, ephedrines, and related substances. *J. Am. Chem. Soc.* **1995**, *117*, 8830−8840.

(19) Wallimann, P.; Marti, T.; Fürer, A.; Diederich, F. Steroids in molecular recognition. *Chem. Rev.* **1997**, *97*, 1567−1608.

(20) Connors, K. A. Population characteristics of cyclodextrin complex stabilities in aqueous solution. *J. Pharm. Sci.* **1995**, *84*, 843−848.

(21) Reinhard, M.; Drefahl, A. *Handbook for Estimating Physicochemical Properties of Organic Compounds*; Wiley: New York, 1999.

(22) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure−Activity Analysis*; Research Studies Press: Letchworth, Herts, U.K., 1986.

(23) Lydersen, A. L. Estimation of critical properties of organic compounds. *University Wisconsin Coll. Exp. Stn., Rep. 3*; Madison, WI, April, 1955.

(24) Joback, K. G.; Reid, R. C. Estimation of pure-component properties from group- contributions. *Chem. Eng. Comm.* **1987**, *57*, 233−243.

(25) Landy, D.; Fourmentin, S.; Salome, M.; Surpateanu, G. Analytical improvement in measuring formation constants of inclusion complexes between β-cyclodextrin and phenolic compounds. *J. Incl. Phenom. Macro. Chem.* **2000**, *38*, 187−198.

(26) Guo, Q.-X.; Luo, S.-H.; Liu, Y.-C. Substituent effects on the driving force for inclusion complexation of α- and β-cyclodextrin with monosubstituted benzene derivatives. *J. Incl. Phenom. Mol. Recog. Chem.* **1998**, *30*, 173−182.

CI010295F