Journal of Inclusion Phenomena and Macrocyclic Chemistry **34:** 291–298, 1999.

Association Constant Prediction for the Inclusion of α-Cyclodextrin with Benzene Derivatives by an Artificial Neural Network

LEI LIU[1], WEN-GUANG LI[2] and QING-XIANG GUO[1,⋆]
[1]*Department of Chemistry, University of Science and Technology of China, Hefei 230026, P. R. China*
[2]*Institute of Chemical Metallurgy, Chinese Academy of Scinces, Beijing 100080, P. R. China*

(Received: 8 June 1998; in final form: 8 September 1998)

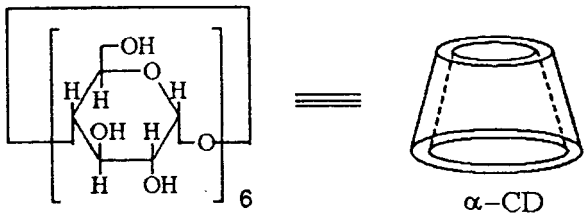
Abstract. The association constants (K_a) for the inclusion complexation of α-cyclodextrin (α-CD) with 72 mono- and 1,4-disubstituted benzenes were predicted successfully by an artificial neural network (ANN) with molar refraction (R_m) and hydrophobic constant (π) as input parameters, which reflect the volume and hydrophobicity of the substituents respectively. The predictions strongly suggested that the inclusion complexation of α-CD with guest molecules was mainly driven by van der Waals forces and hydrophobic interactions.

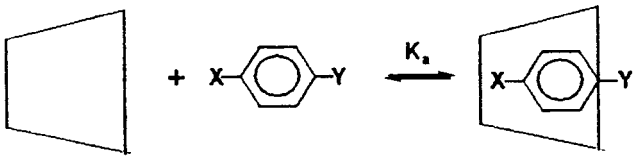
Key words: artificial neural network, driving force, α-cyclodextrin, inclusion.

1. Introduction

Cyclodextrin (CD), an ideal model for mimicking enzyme-substrate interactions, has attracted tremendous interest in many different fields recently, such as catalysis, separation science and technology, pharmaceutical application etc. [1]. Great efforts have been devoted to the quantitative understanding of host-guest interactions due to the importance of the inclusion phenomena in biochemical systems. To date, several driving forces have been postulated for the inclusion complexation of CD with guest molecules [1–3]: van der Waals forces, hydrophobic interactions, steric effects and conformational energy, hydrogen bonding etc. However, there still remains no clear agreement on the mechanism for the inclusion of CD species.

⋆ Author for correspondence.





An artificial neural network (ANN) proves an excellent tool for nonlinear analysis. During the past decade, it has been extensively applied to the domain of chemistry [4]. In our previous reports [5], the association constants for the inclusion of α- and β-CD with monosubstituted benzenes and some 1,4-disubstituted benzenes were evaluated by ANN calculations with substituent molar refraction R_m, hydrophobic constant π, and Hammett constant σ as input parameters. These parameters reflect van der Waals forces, hydrophobic interactions and electronic effects, respectively. However, our very recent study [6] indicated that the Hammett constant σ plays a trivial role in CD inclusion complexation. Herein, we report the association constants of α-CD with 72 mono- and 1,4-disubstituted benzenes only with R_m and π constants as input parameters, in order to widen the understanding of the driving forces and develop a convenient method to evaluate the association constants for the molecular recognition of cyclodextrin.

2. Method

The theory and application of ANN have been well documented [7]. In this work, a neural network with three layers (one input layer, one hidden layer and one output layer) was employed. The Sigmoid function is chosen as the transfer function and the back-propagation (BP) algorithm was employed in training the neural network. The objective function E is taken as

$$E = \frac{1}{2}\sum_{n=1}^{N}(v_n^T - v_n)^2,$$

and the weights connecting the nodes are adjusted by using

$$u_{ti}^{\text{new}} = u_{ti}^{\text{old}} + \Delta u_{ti}^{\text{old}},$$

in which

$$\Delta u_{ti}^{\text{new}} = -\eta \frac{\partial E}{\partial u_{ti}^{\text{old}}} + \alpha \Delta u_{ti}^{\text{old}},$$

η is the learning rate and α the momentum. The program was written in *Borland C^{++} 3.1*, and run on an 80586 personal computer.

It is well known that the determination of the orientation of CD complexation is still a controversial problem both experimentally [8] and theoretically [3]. In this study, the orientation was determined as follows:

(1) The substituents in mono-substituted benzenes stay in the α-CD cavity, since they are larger than hydrogen and therefore favored by the α-CD cavity *via* van der Waals forces.
(2) The OH and NH_2 groups stay outside the α-CD cavity since they are highly hydrophilic.
(3) The SCH_3 and COOH groups stay in the cavity of α-CD since they are generally larger than other groups and therefore favored by the α-CD cavity *via* van der Waals forces.

The substituent located inside the α-CD cavity was denoted as group X in this paper. Despite the empirical nature of the above decision, the good prediction of the ANN offered some further confidence.

The molar refraction R_{mX}, R_{mY} and hydrophobic constant π_X and π_Y of the substituents X and Y in disubstituted benzenes were taken from the compilation of Hansch and Leo [9] and used as input parameters. The experimental ln K_a values for the aqueous host-guest complexes were chosen as the target output. The parameters of 57 inclusion complexes were used in the network training, and the association constants for 72 inclusion complexes were predicted.

Different 4-n-1 ANNs were respectively trained. It was found that the ANN with a topological structure of 4-12-1 was the best in quickly and well generalizing the model. Therefore, in this study, the number of the hidden neurons was selected as 12. In order to avoid the chance effect of the neural network calculation, the predictions over five independent networks running were statistically averaged [10].

3. Results and Discussion

The values calculated by the ANN are listed in Table 1. Plotting the ln K_a(obs) values determined experimentally versus the ln K_a(ANN) values calculated by ANN gives a straight line (Figure 1), which fits the following equation:

$$\ln K_a(\text{obs}) = 1.006 \ln K_a(\text{calc}) - 0.017$$

$$(n = 72, r = 0.961, sd = 0.415)$$

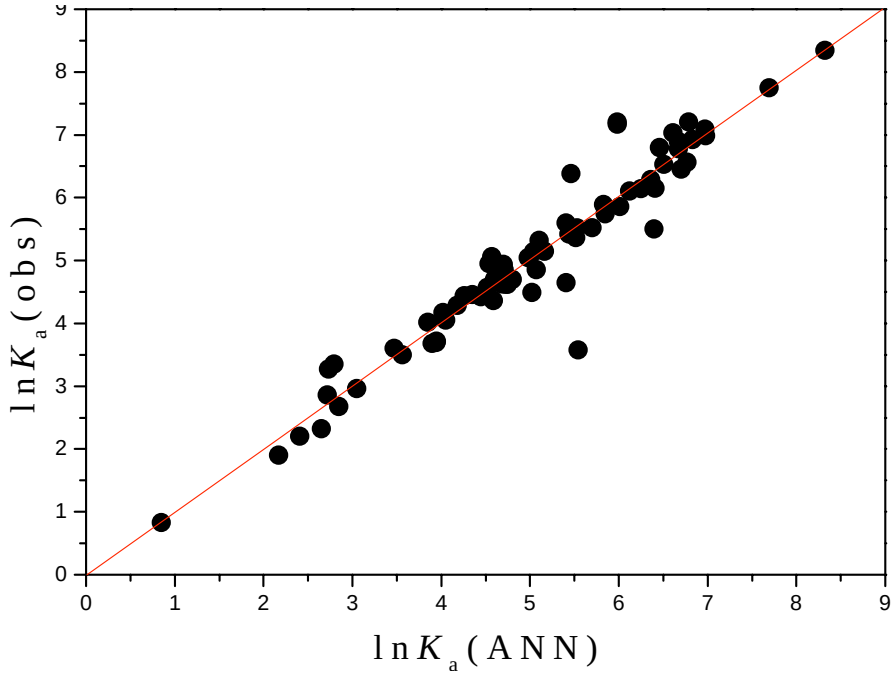


Figure 1. The ln K_a values predicted by ANN versus the observed ln K_a values.

From the above equation it can be seen that the ln K_a values evaluated by ANN are in good agreement with those determined experimentally. This result suggested that the inclusion complexation of α-CD with guest molecules is greatly correlated with the R_m and π constants, and hence, mainly driven by van der Waals forces and hydrophobic interaction [6].

The dependence of ln K_a for α-CD-guest complexes on the R_m of the guest compound has been observed previously [14]. It was generally accepted that the substituent molar refractivity (R_m) reflects the volume and polarizibility of the substrate. The greater the substituent R_m value, the larger the volume and the higher the polarizibility of the compound. In α-CD inclusion complexation, the substituent with appropriate volume fits more snugly into the α-CD cavity *via* van der Waals contact, and the substituent with high polarizibility favors binding through London dispersion and dipole/induced dipole interactions. Therefore, van der Waals forces play important roles in α-CD inclusion complexation [6].

In addition, another important factor affecting the inclusion of α-CD with guest compound is the hydrophobicity of the substituents [15]. The hydrophobic interaction is mainly due to the effects of entropy produced on the water molecules. In aqueous solution, a hydrophobic guest compound is restricted by the water shell made of the hydrogen bonding network. It tends strongly to break down the water cluster and penetrate the apolar cavity of CD [6].

Table I. ln K_a values predicted by ANN from R_m and π and ln K_a values determined experimentally for inclusion of α-CD with mono-and 1,4-disubstituted benzenes

No.	X	Y	ln K_a(obs.)	Ref.	ln K_a(ANN)	Error(%)
1	F	H	3.68	6	3.90	6.0
2	Cl	H	4.72	6	4.74	0.4
3	Br	H	6.29	6	6.36	1.1
4	I	H	7.09	6	6.97	−1.7
5	F	F	2.96	11	3.05	3.0
6	Cl	Cl	5.42	11	5.44	0.4
7	Br	Br	6.93	11	6.83	−1.4
8	I	I	8.34	11	8.32	−0.2
9	Cl	F	4.17	11	4.02	−3.6
10	Br	F	5.52	11	5.70	3.3
11	I	F	6.89	11	6.67	−3.2
12	OCH_3	OCH_3	4.02	12	3.85	−4.2
13	OEt	OEt	4.85	12	4.71	−2.9
14	CO_2Me	CO_2Me	6.14	12	6.25	1.8
15	COMe	COMe	2.32	12	2.65	14.2
16	CN	CN	3.50	12	3.56	1.7
17	NO_2	NO_2	3.58	12	5.54	−0.7
18	COOH	COOH	7.20	12	5.98	−16.9
19	NH_2	NH_2	0.83	12	0.85	2.4
20	OCH_3	NH_2	1.90	12	2.17	14.2
21	CH_3	NH_2	4.05	12	4.05	0.0
22	SCH_3	OCH_3	4.70	13	4.80	2.1
23	SCH_3	CH_2OH	4.44	13	4.26	−4.0
24	SCH_3	Br	5.74	13	5.85	1.9
25	SCH_3	NO_2	4.81	13	4.64	−3.5
26	SCH_3	Cl	5.04	13	4.98	−1.2
27	SCH_3	$COCH_3$	2.20	13	2.41	9.5
28	SCH_3	CH_3	3.71	13	3.95	6.5
29	SCH_3	NH_2	4.62	13	4.75	2.8
30	CH_3	H	3.60	6	3.47	−3.6
31	H	H	3.35	6	2.79	−16.7
32	Et	H	4.70	6	4.60	−2.1
33	CH_3	H	4.28	6	4.18	−2.3
34	*n*-Pr	H	6.38	6	5.46	−14.4
35	*i*-Pr	H	4.65	6	5.41	16.3
36	OCH_3	H	4.95	6	4.54	−8.3

Table I. Continued

No.	X	Y	ln K_a(obs.)	Ref.	ln K_a(ANN)	Error(%)
37	OEt	H	5.14	6	5.04	−1.9
38	CH_2OH	H	4.57	6	4.52	−1.1
39	CH_2Cl	H	5.32	6	5.10	−4.1
40	CHO	H	4.62	6	4.72	2.2
41	COMe	H	4.94	6	4.70	−4.8
42	CO_2Me	H	5.36	6	5.52	3.0
43	CO_2Et	H	5.89	6	5.83	−1.0
44	CN	H	4.36	6	4.59	5.3
45	NH_2	H	2.68	6	2.85	6.3
46	Cl	NH_2	5.52	12	5.53	0.2
47	COOH	NH_2	7.20	12	6.79	−5.7
48	CN	NH_2	6.11	12	6.12	−0.2
49	NO_2	NH_2	6.45	12	6.70	3.9
50	CH_2NH_2	H	2.86	12	2.96	3.5
51	$CH_2CH_2NH_2$	H	3.27	12	3.77	15.3
52	I	OH	7.75	12	7.69	−0.8
53	Cl	OH	5.60	12	5.41	−3.4
54	Br	OH	6.56	12	6.77	3.1
55	COOH	OH	7.03	12	6.61	−6.0
56	CN	OH	5.06	12	4.57	−9.7
57	NO_2	OH	5.50	12	6.40	16.4
58	COOH	$NHCH_3$	7.17	12	5.98	−16.6
59	COOH	OCH_3	6.78	12	6.67	−1.6
60	COOH	CH_3	6.99	12	6.98	−0.1
61	COOH	H	6.53	12	6.51	−0.3
62	COOH	F	6.22	12	6.36	2.2
63	COOH	CH_3CO	6.80	12	6.46	−5.0
64	COOH	CN	6.15	12	6.41	4.2
65	COOH	NO_2	5.86	12	6.01	2.6
66	NHEt	H	4.85	6	5.07	4.5
67	NHMe	H	4.42	6	4.45	0.7
68	NMe_2	H	5.15	6	5.16	0.2
69	NHCOMe	H	4.63	6	4.68	1.1
70	C≡CH	H	4.46	6	4.35	−2.5
71	OH	H	3.70	6	3.94	6.5
72	NO_2	H	4.49	6	5.02	11.8

Since van der Waals forces and hydrophobic interactions operate simultaneously in CD inclusion complexation, the interaction of these driving forces is readily anticipated [3, 6]. Therefore, there exist nonlinear relations between the driving forces and association constants. From the above calculation, it is obvious that the ANN has well generalized these relations. Since there is some correlation between the substituent R_m, π and σ constants [9], it is reasonable to infer that the substituent σ constant can also affect the stability of CD complexes. However, through a statistical method, our previous paper has demonstrated clearly that this influence was trivial. It is also noted that the predictions of the ANN from only R_m and π constants in this study are as good as the predictions of the ANN from R_m, π and σ constants [5c]. Therefore, in the inclusion complexation of α-CD with benzene derivatives, the R_m and π constants have sufficiently reflected the major characters of the substituents.

Admittedly, the ANN also showed a relatively bad performance on some compounds (see compound No.15, 18, 20, 31, 34, 35, 51, 57, 58, 72 in Table1). Generally, α-CD complexation with these compounds are probably affected by hydrogen bonding. Since the role of hydrogen bonding is difficult to describe quantitatively, it was not chosen as the input parameter in this study and could bring certain deviation. Besides, it is possible that the association constants of these compounds are not very precise due to the error in the experimental measurements.

4. Conclusion

An ANN method has been successfully applied to the prediction of α-CD inclusion complexation. The results indicated that the van der Waals forces and hydrophobic interactions mainly contributed to the driving forces for the inclusion complexation of α-CD with benzene derivatives. The method is satisfactory, convenient and instructive, and can be used in other kinds of inclusion complexes.

Acknowledgement

This work was supported by the National Natural Science Foundation of China.

References

1. (a) G. Wenz: *Angew. Chem. Int. Ed. Engl.* **33**, 803 (1994); (b) R. Breslow: *Acc. Chem. Res.* **28**, 146 (1995); (c) K. A. Connors: *Chem Rev.* **97**, 1325 (1997); (d) M. V. Rekharsky and Y. Inoue: *Chem. Rev.* **98**, 1875 (1998).
2. (a) D. M. Davies and J. R. Savage: *J. Chem. Res. (S)* 94 (1993); *J. Chem. Res. (M)*, 660 (1993); (b) J. H. Park and T. H. Nah: *J. Chem. Soc. Perkin Trans. 2*, 1359 (1994).
3. (a) E. Alvira, J. Mayoral and J. Garcia: *Chem. Phys. Lett.* **245**, 335 (1995); (b) K. A. Connors: *J. Pharm. Sci.* **85**, 796 (1996); (c) P. Ross and M. Rekbarsky: *Biophys. J.* **71**, 2144 (1996); (d) M. Huang, J. Watts and N. Bodor: *Inter. J. Quan. Chem.* **65**, 1135 (1997).
4. (a) T. Aoyama, Y. Suzuki and H. Ichikuwa: *J. Med. Chem.* **33**, 2583 (1990); (b) N. Bodor, A. Harget and M. Huang: *J. Am. Chem. Soc.* **112**, 216 (1990); (c) B. Mayer, T. Hansen, D. Nute,

P. Albersheim, A. Darvill, W. York and J. Sellers: *Science* **251**, 542 (1991); (d) G. M. Silva: *J. Comput. Chem.* **18**, 1407 (1997); (e) T. R. Cundari and E. W. Moody: *J. Chem. Inf. Comput. Sci*. **37**, 871 (1997).

5. (a) Q.-X. Guo , S.-H. Luo, H. Wang, M.-S. Zhang and Y.-C. Liu: *J. Chem. Res. (S)*, 38 (1996); (b) Q.-X. Guo, S.-H. Luo, H. Wang, M.-S. Zhang and Y.-C. Liu: *Chin. Chem. Lett.* **7**, 285 (1996); (c) Q.-X. Guo, S.-D. Chu, Y.-M. Wang, Z.-X. Pan, M.-S. Zhang, Y.-C. Liu: *Chin. Chem. Lett*. **8**, 145 (1997); (d) Q.-X. Guo, L. Liu, W.-S. Cai, Y. Jiang and Y.-C. Liu: *Chem. Phys. Lett*. **290**, 514 (1998).
6. Q.-X. Guo, S.-H. Luo and Y.-C. Liu: *J. Incl. Phenom.* **30**, 173 (1998).
7. (a) J. Gasteiger and J. Zupan: *Angew. Chem. Int. Ed. Engl*. **32**, 503 (1993); (b) S. Burns and G. M. Whitesides: *Chem. Rev.* **93**, 2583 (1993); (c) B. Sumpter, C. Getino and D. Noid: *Ann. Rev. Phys. Chem.* **45**, 439 (1994).
8. J. L. Alderfer, A. V. Eliseev: *J. Org. Chem.* **62**, 8225 (1997).
9. (a) C. Hansch and A. Leo: *Substituent Constants for Correlation Analysis in Chemistry and Biology* (Wiley, New York, 1979); (b) C. Hansch, A. Leo and R. W. Taft: *Chem. Rev.* **91**, 165 (1991).
10. D. J. Livingstone and D. T. Manallack: *J. Med. Chem.* **36**, 1295 (1993).
11. T. Takuma, T. Deyuchi and I. Sanemasa: *Bull. Chem. Soc. Jpn.* **63**, 1246 (1990); **64**, 480 (1991); **64**, 1979 (1991).
12. (a) K. A. Connors, S.-F. Lin and A. Wong: *J. Pharm. Sci.* **71**, 217 (1992); (b) S.-F. Lin and K. A. Connors: *J. Pharm. Sci.* **72**, 1333 (1983); (c)A. Wong, S.-F. Lin and K. A. Conors: *J. Pharm. Sci*. **72**, 388 (1983); (d) K. A. Connors and D. Pendergast: *J. Am. Chem. Soc.* **106**, 7607 (1984).
13. D. M. Davies and M. R. Deary: *J. Chem. Soc. Perkin Trans. 2*, 1287 (1995).
14. R. Etten, J. Sebastian, G. Clowes and M. Bender: *J. Am. Chem. Soc.* **89**, 3242 (1967).
15. Y. Mattsui and K. Mochida: *Bull. Chem. Soc. Jpn.* **52**, 2808 (1979).