

# A Computer Simulation – Statistical Procedure for Predicting Complexation Equilibrium Constants

ZHONG-GUO SHI and E. A. McCULLOUGH, Jr.\*

*Department of Chemistry and Biochemistry, Utah State University, Logan, Utah 84322–0300, U.S.A.*

(Received: 29 October 1993; in final form: 17 May 1994)

**Abstract.** The complexing properties of macrocyclic ligands have been quantitatively studied by the combined use of molecular mechanics, molecular dynamics, and multiple linear regression. The dependent variables in the regression equations are experimental equilibrium constants for known macrocycle complexes in various solvents. The independent variables are theoretical simulation results on the solvent-free ligand and its complex and additional physically motivated empirical variables to describe solvent and other important effects.

The systems studied were: (a) 314 metal ion–macrocyclic–solvent (including 3 mixed solvents) combinations; (b) 88 ammonium ion–crown ether–solvent (including 1 mixed solvent) combinations; (c) 24 hydrogen ion–crown ether–H<sub>2</sub>O combinations; (d) 26 Na<sup>+</sup> ion–spherand–CDCl<sub>3</sub> combinations; (e) 78 ammonium ion–spherand–CDCl<sub>3</sub> combinations; and (f) 73 complicated host–guest–solvent (including 1 mixed solvent) combinations.

For each system, we report the best regression equation obtained using the AMBER force field. The standard errors in log *K* range from 1.42 in the largest system to 0.36 in the smallest. Regression equations were determined for several of the systems using the MMP2 force field as well, and the equations are shown to be relatively insensitive to the force field.

The predictive ability of the method was tested by predicting log *K* for 20% of the cases chosen at random using equations derived from the remaining 80%. The errors in the predicted values are shown to be consistent with the statistical assumptions of the model.

Regression equations obtained with this method can be used to predict the equilibrium constants for new complexes involving some combination of new, possibly unknown macrocycle, new host and, in certain cases new solvent. No X-ray or other structural data for the macrocycle is needed.

**Key words:** Complexation equilibrium constant, prediction, computer simulation.

## 1. Introduction

As complexing agents for anions, neutral molecules, and especially for cations, macrocyclic ligands are of great scientific and practical interest. In addition to physical and chemical methods [1–4], computational chemistry is increasingly employed in the study of macrocyclic ligands and their complexes [5–19]. Most of this computational work has involved reproducing and interpreting the properties of known complexes. The more difficult task of predicting the properties of unknown complexes has received less attention.

The major difficulty lies in the quantitative treatment of the solvent effects. While many researchers [15–19] have shown that detailed computer simulations which explicitly include many solvent molecules are potentially capable of accurately reproducing experimental data, such simulations are exceedingly time-

\* Author for correspondence.

consuming and expensive. At present, such an approach is impossible to apply rapidly and routinely to a large number of different host–guest–solvent combinations, as would be required for the design of new complexing systems.

In the present study we attempt to develop a method which can be used to predict  $K$  (the equilibrium constant for complex formation) for new, unknown complexes in various solvents. The method is not based on detailed, first principles simulations, but instead is more empirical in nature, and assumes an existing database of experimental information. The procedure is as follows.

First, on chemical and physical grounds we try to identify variables which might be important in determining  $K$ . For the free or complexed macrocycle only theoretical results from molecular mechanics and molecular dynamics simulations without solvent are considered. In particular, no X-ray crystallographic or experimental structural data for the macrocycle is used. For simple cations, empirical data such as crystal radius and metal electronegativity are used; for more complicated guests, results from solvent-free simulations are employed. For the solvent, the variables consist of experimental properties such as dielectric constant and various empirical solvent parameters.

We then use multiple linear regression to determine the statistical relationship between these variables and the experimental  $\log K$  for a large number of known complexes. From the regression results we are able to determine which of the variables are most important and eliminate the least important ones. The final regression equations may be used to predict  $\log K$  for unknown complexes. Only experimental data for the solvent and computer simulation results for the solvent-free macrocycle and complex are needed. Since the solvent is not included, the simulation can be done rapidly; furthermore only one simulation is needed for any host–guest pair, regardless of solvent.

In this paper we present the results of this approach for a number of different systems ranging from simple hydrogen ion–crown ether–water complexes to complicated host–guest–mixed solvent complexes. In each case a regression equation based on a minimum number of variables is determined. The predictive ability of the resulting equations is demonstrated.

## 2. Method and Variable Selection

The general form of a 1 : 1 complexation reaction is



where  $M$  is a cation or neutral molecule,  $[ ]$  represents the macrocycle, and  $[M]$  denotes the complex. The equilibrium constant satisfies the equation

$$\log K = -\Delta G^0/2.303 \cdot RT \quad (2)$$

The temperature could be included as an additional variable, but in the present work we treat only the constant temperature case, in which  $\log K$  is directly proportional

to the free energy change. We consider the free energy change to be the sum of two terms

$$\Delta G = \Delta G_g + \Delta G_s \quad (3)$$

where  $\Delta G_g$  is the free energy change in the gas phase and  $\Delta G_s$  is the free energy increment due to the transfer of the reaction from the gas phase to the solvent. As mentioned, we do not attempt to calculate  $\Delta G_g$  and  $\Delta G_s$  from a detailed physical theory. Instead, we attempt to identify physically based variables upon which  $\Delta G_g$  and  $\Delta G_s$  might depend and determine the actual dependence by multivariate regression.

The free energy change  $\Delta G_g$  can be written as

$$\Delta G_g = \Delta H_g - T\Delta S_g \quad (4)$$

For the 1 : 1 complexes considered in this study,  $\Delta H$  differs from  $\Delta E$ , the energy change in the reaction, by only  $RT$ . Thus, to a first approximation we may replace  $\Delta H_g$  by  $\Delta E_g$ . Furthermore, work by Inoue *et al.* [20] has shown that there is an approximate proportionality between  $T\Delta S$  and  $\Delta H$  for a wide variety of complexes. Therefore, we take  $\Delta E_g$  as our starting point in developing the variables which determine  $\Delta G_g$ .

In the force field approximation, the total energy,  $E_{\text{total}}$ , of the molecule can be expressed as a sum of bond stretch,  $E_b$ , angle bending energy,  $E_\Theta$ , torsional energy,  $E_\Phi$ , van der Waals energy,  $E_{\text{vdw}}$ , electrostatic energy,  $E_{\text{elec}}$ , and hydrogen bond energy,  $E_h$ . In order to decrease the number of variables and improve the physical interpretability of certain variables, we can take each type of energy differences between the reactants and products (such as  $\Delta E_b$ ,  $\Delta E_\Theta$ ,  $\Delta E_\Phi$ ,  $\Delta E_{\text{vdw}}$ ,  $\Delta E_{\text{elec}}$ ) as a set of variables. Also we may define new variables which are linear combinations of original variables. For example, we may define  $\Delta E_{\text{inter}} = \Delta E_b + \Delta E_\Theta + \Delta E_\Phi$ , where  $\Delta E_{\text{inter}}$  is the internal strain energy of the molecule. Thus, we have several possible sets of energy variables to use in the  $\Delta G_g$ . The choice of which set among possible sets of energy variables is determined by the compromise of maximizing the statistical accuracy and minimizing the number of variables.

We compute these energy values as follows. Computer graphics (BioGraf [21] version 2.1) are used to create models of the molecular structures on the Stardent Titan workstation. The atomic charges in each molecule are calculated by the Gasteiger method [22]. A molecular mechanics program [21] is utilized to refine these coordinates and to evaluate their energies employing the AMBER [23], [24] or MMP2 [25] force field. We also use molecular quenched dynamics [26] to search the lowest energy conformation at a fixed temperature. A 10–15 ps dynamics simulation is used for crown ethers, cryptands, and their complexes; 30 ps for spherands and their complexes; and 50 ps for cyclophanes and their complexes. Extracting the best conformer, we again utilize molecular mechanics to refine these coordinates and to evaluate their energies.

The electrostatic interaction between simple cations and the macrocycle is quite large and also well described in the force field approximation. However, for some metal cations the interaction can have a significant covalent component which may not be treated adequately. Finding variables which accurately reflect the tendency for covalent bonding is not easy, and the physical basis is less clear. One of the simplest possibilities is electronegativity. Of the many different definitions of electronegativity, we have chosen the Mulliken–Jaffe system [27], since it is the most fundamental and is based directly on energetics.

For  $\Delta G_s$  we choose a model to treat the solute–solvent interactions with reference to the methods of Still *et al.* [28] and Cramer and Truhlar [29]. The solvation free energy,  $\Delta G_s$ , is assumed to be of the form

$$\Delta G_s = \Delta G_{\text{cav}} + \Delta G_{\text{st}} + \Delta G_{\text{pol}} + \Delta G_{\text{h}} \quad (5)$$

where  $\Delta G_{\text{cav}}$  is a solvent–solvent cavity term,  $\Delta G_{\text{st}}$  is a surface tension term involving work against the surface tension of the solvent and includes the free energy of creating cavity plus the solute–solvent dispersion interaction,  $\Delta G_{\text{pol}}$  is a solute–solvent electrostatic polarization term, and  $\Delta G_{\text{h}}$  is a solvation energy term involving hydrogen bonding.

We consider the cavity term  $\Delta G_{\text{cav}}$  as a single term that is differently dealt with in references [28] and [29]. They combined the cavity term  $\Delta G_{\text{cav}}$  into the surface tension term  $\Delta G_{\text{st}}$ . Our  $\Delta G_{\text{cav}}$  term has the following form

$$\Delta G_{\text{cav}} \propto \delta_t^2 \quad (6)$$

where  $\delta_t$ , the cohesive energy, is the solubility parameter [30] of the solvent, which is considered as a measure of the energy to create a suitably sized cavity in the solvent for the solute [31].

We define the surface tension term  $\Delta G_{\text{st}}$  as

$$\Delta G_{\text{st}} \propto \sigma A \quad (7)$$

where  $A$  is the surface area of the solute, and  $\sigma$  is the bulk surface tension of the solvent.

Based on the continuum model of the solvent, a flexible interpretation of the surface tension term,  $\Delta G_{\text{st}}$ , is that  $\Delta G_{\text{st}}$  is involved in not only cavity and dispersion interaction terms but also the structure change of the solvent (i.e., change of entropy of the solvent) [32]. In this research, each of the solutes in solution is regarded as a sphere of radius  $r$ , thus  $A = 4\pi r^2$ . So Equation (7) can be taken in the form

$$\Delta G_{\text{st}} \propto 4\sigma\pi r^2 \quad (8)$$

where  $r$  is the approximate radius of the solute.

For the  $\Delta G_{\text{pol}}$  term, the following interactions are considered:

- (a) the free energy of transfer of a spherical ion from vacuum to the solvent of dielectric constant  $\varepsilon$  [31]

$$\Delta G_{\text{tr}}^0 \propto -z_i^2 (e^2/2r_i) (1 - 1/\varepsilon) \quad (9)$$

where  $z_i$  is the charge of ion  $i$ , and  $r_i$  is the appropriate radius of ion  $i$ ;

- (b) the free energy of transfer of a spherical dipole from vacuum to the solvent of dielectric constant  $\varepsilon$  [31]

$$\Delta G_{\text{tr}}^0 \propto -(\mu_j^2/r_j^3) (\varepsilon - 1)/(2\varepsilon + 1) \quad (10)$$

where  $\mu_j$  and  $r_j$  are the appropriate dipole moment and radius, respectively;

- (c) the ion–dipole interaction [33]

$$\Delta G_{\text{int}}^0 \propto -[Z_A|\mu_B/r_{AB}^2] \quad (11)$$

where  $Z_A$  is the charge of ion  $A$ ,  $\mu_B$  is the dipole moment of  $B$ , and  $r_{AB}$  is the distance between an ion  $A$  and a dipole  $B$ ; and

- (d) the dipole–dipole interaction [33]

$$\Delta G_{\text{int}}^0 \propto -2\mu_A\mu_B/r^3 \quad (12)$$

where  $\mu_A$  and  $\mu_B$  are the dipole moments of molecules  $A$  and  $B$ , respectively, and  $r = r_A + r_B$ , the distance between the two dipolar molecules.

For the free and complexed macrocycle, the radii (and also the dipole moments) which appear in Equations (8) and (9)–(12) must be evaluated from the molecular mechanics and dynamics simulations since no experimental structural data are allowed. Three different definitions of radius were investigated. The first ( $r_1$ ) is one half the maximum distance between the O and N atoms of the macrocycle. This can be viewed as an approximate cavity radius since the complexing interaction mainly involves the O and N atoms. The second ( $r_2$ ) is one half the maximum distance between atoms in the macrocycles; whereas  $r_3$  is the same, except it excludes hydrogen atoms. These are effective radii of the macrocycle itself. More accurate definitions are possible, of course, but elaborate definitions which require extensive computation are probably not justified given the many other approximations.

Hydrogen bonding is similar to covalency in the difficulty of representing this effect with variables with a clear physical basis. The regression results showed that  $\delta_{\text{h}}$ , which is a component of the solubility parameter  $\delta_{\text{t}}$ , was an important variable, even if the relation to  $\Delta G_{\text{h}}$  is less than straightforward.

A list of the major variables we considered is shown in Table I.

In multiple linear regression [34], a dependent variable  $Y$  is modeled as a linear function of  $m$  independent variables  $\{X_1, \dots, X_m\}$ ,

$$Y = B_0 + B_1X_1 + \dots + B_mX_m. \quad (13)$$

Given  $n (> m)$  independent values of  $Y$  and the associated  $\{X_1, \dots, X_m\}$ , the coefficients in the function are determined by least squares.

For large  $n$  the standard error of estimate of  $Y$ ,  $SE$ , is given by Equation (14)

$$SE^2 = \Sigma(Y_j - \hat{Y}_i)^2 / (n - m - 1) \quad (14)$$

TABLE I. List of major variables considered in each term

	Term	Variables
$\Delta G_g$ :	First set:	$E_{\text{elec},[ ]}, E_{\text{elec},[M]}, E_{\Phi,[ ]}, E_{\Phi,[M]}, E_{\text{inter},[ ]},$ $E_{\text{inter},[M]}, E_{\text{vdw},[ ]}, E_{\text{vdw},[M]}, \chi$
	Second set:	$\Delta E_{\text{elec}}, \Delta E_{\Phi}, \Delta E_{\text{inter}}, \Delta E_{\text{vdw}}, \chi$
	Third set:	$\Delta E_{\text{total}}, \chi$
$\Delta G_s$ :	$\Delta G_{\text{cav}}$	$\delta_t^2$
	$\Delta G_{\text{st}}$	$\sigma r_M^2, \sigma r_{[ ]}^2, \sigma r_{[M]}^2$
	$\Delta G_{\text{pol}}$	$1/\epsilon, z_M^2/r_M, z_M^2/r_{[M]}, \mu_M^2/r_M^3, \mu_{[ ]}^2/r_{[ ]}^3,$ $\mu_{[M]}^2/r_{[M]}^3, z_M \mu_{[ ]}/r_{[ ]}^2, z_M \mu_{[ ]}/r_{[M]}^2,$ $\mu_M \mu_{[ ]}/r_{[M]}^3$
	$\Delta G_h$	$\delta_h$

where  $Y_i$  is the  $i$ th value of  $Y$  and  $\hat{Y}_i$  is the value predicted from Equation (13). For normally distributed data, there is approximately a 68% probability that a predicted value lies within  $\pm SE$  of the true value and a 95% probability that it lies within  $\pm 2SE$ .

The multiple correlation coefficient,  $R$ , is a measure of the adequacy of fit. When the model fits the data well, the  $R$  value is close to unity; when the regression model gives a poor fit, the  $R$  value will be close to zero. In our case, deviations of  $R$  from unity almost certainly are due to limitations of our model rather than errors in the experimental log  $K$  values.

If prediction and not interpretation is the major goal, then a small  $SE$  is more important than an  $R$  close to unity since the probable error in the predicted value is of more interest than how well the model fits the data.

The  $t$  ratio is employed to do significance testing. Its value is a test of the hypothesis that there is no linear relation between the dependent variable and a given independent variable. In general, the bigger the  $t$  ratio is, the more significant the coefficient  $B$  is.

In order to select the 'best' set of variables for predictive purposes we employed the stepwise (SW) method. The variables are entered into the equation one at a time. At each step, the variable chosen is the one with the largest  $t$  ratio. A variable entered at one step may be deleted at a later step if its  $t$  ratio becomes insignificant.

The variables in Table I are certainly not all independent in the mathematical sense. We examine whether any two variables are independent or not from the correlation matrix of the variables. If there are no large correlations between any two variables, we roughly think these variables are all independent. If there are large correlations between two variables, we drop one of the strongly intercorrelated variables from the regression model to eliminate collinearity.

Note that for our purposes it is not necessary for the variables in Table I to be expressed in the same units since any conversion factors can be absorbed into the multiple regression coefficients. To avoid the possibility of unit conversion errors

we used the units in which the variables were generated by the computer programs: energy in Kcal/mol, distance in Ångstroms, dipole moment in Debyes, charge in multiples of the proton charge.

For each case, the residual is the difference between the actual value of  $Y$  and the value calculated from the regression equation. Cases with extremely large residuals are called outliers. There are many techniques, such as the residual plots, available for detecting outliers. In order to detect more damaging outliers which may not always show up in residual plots, we rerun the regression analysis without the potential outliers, and compare this regression result with that including the potential outliers. If deleting certain points gives a better regression coefficient, the deleted, high-influence points can be considered to be outliers. Otherwise, we will retain these.

The predictive ability of the final regression model is tested by selecting randomly 20% of the cases in a system as a test set. A regression equation is developed on the remaining 80% using the same variables as in the equation for 100% of the cases. The new regression equation is then used to predict the  $\log K$  values for the test set as if they were unknown, and the residuals are compared to a normal distribution. For a normal distribution, approximately 68% of the predicted values will lie within  $\pm SE$  and 95% within  $\pm 2SE$  of the observed values.

The multiple linear regression analyses were done using SPSS-X [35] (Statistical package for the Social Sciences, release 3.0). SPSS-X was run on a Digital Equipment Corporation VAX (8600 series).

### 3. Results and Discussion

#### 3.1. METAL ION-MACROCYCLE (CROWN ETHERS, CRYPTANDS, ETC.) SYSTEM

This system consists of 3 metal ions ( $\text{Na}^+$ ,  $\text{Ca}^{2+}$ ,  $\text{Zn}^{2+}$ ), 92 crown ethers, 10 cryptands, 4 antibiotics, 17 pure solvents, and 3 mixed solvents. Equilibrium constant data are from References [36, 37 and 38]. The  $\log K$  values range all the way from 0.3 to 21. The regression results are shown in Table II. We find that the electronegativity  $\chi$  is indeed an extremely important variable. A comparison of the unstandardized coefficients of the regression equations for 100% of the cases versus 80% of the cases also is presented in Table II. We see that the regression equation is not very sensitive to the number of cases; that is, this regression equation is stable.

In contrast to the other systems examined, the outliers in the metal ion-macrocycle system show some definite patterns. Table III lists the fourteen outliers in the order in which they were deleted. Four outliers consist of  $\text{Zn}^{2+}$  ion and mixed donor crown ethers. For each of these our model *overestimates* the equilibrium constant by several orders of magnitude. According to Inoue's work [3], the regression coefficient of  $\Delta H$  versus  $T\Delta S$  for 1 : 1 complexation of heavy/transition metal ions with crown ethers possessing nitrogen/sulfur donors is low with  $R = 0.36$ .

TABLE II. The unstandardized coefficients ( $B$ ) and  $t$  ratios in the metal ion–macrocycle system

Variables	$B$		$t$ ratio
	80% of Cases	100% of Cases	100% of Cases
$\Delta E_{\text{inter}}$	0.637	0.670	14.520
$\chi$	2.151	2.103	23.300
$z_M^2/r_{1,[M]}$	-9.873	-9.801	-14.415
$z_M\mu_{[ ]}/(r_{1,[M]})^2$	3.222	2.849	13.621
$\delta_h$	-0.065	-0.063	-8.053
$\Delta E_{\Phi}$	-0.576	-0.622	-7.642
$\sigma(r_{2,[ ]})^2$	-0.010	-0.009	-6.423
$z_M\mu_{[ ]}/(r_{1,[ ]})^2$	-6.515	-6.323	-5.729
$\mu_{[ ]}^2/(r_{1,[ ]})^3$	1.753	2.003	2.371
$\Delta E_{\text{vdw}}$	0.088	0.110	2.619
$\Delta E_{\text{elec}}$	0.010	0.014	1.703
Constant	-6.462	-6.126	-10.312

<sup>a</sup>  $R = 0.90$ ;  $SE = 1.42$ ;  $n = 300$  (14 outliers excluded).

TABLE III. Outliers in the metal ion–macrocycle system

Case	Macrocycle/metal ion/solvent
1	1,10A <sub>2</sub> 18C6/Zn <sup>2+</sup> /H <sub>2</sub> O
2	1,7A <sub>2</sub> 18C6/Zn <sup>2+</sup> /H <sub>2</sub> O
3	B2.2.2/Ca <sup>2+</sup> /DMF
4	2.2.2/Ca <sup>2+</sup> /Me <sub>2</sub> SO
5	1,4A <sub>2</sub> 15C5/Zn <sup>2+</sup> /H <sub>2</sub> O
6	2.2.1/Na <sup>+</sup> /PC
7	1,7A <sub>2</sub> 15C5/Zn <sup>2+</sup> /H <sub>2</sub> O
8	2.2.1/Na <sup>+</sup> /MeCN
9	(CbMA) <sub>4</sub> 14C4/Ca <sup>2+</sup> /H <sub>2</sub> O
10	2.2.1/Ca <sup>2+</sup> /PC
11	2.2.1/Ca <sup>2+</sup> /MeOH
12	2.2.1/Ca <sup>2+</sup> /95%MeOH
13	DodecOM18C6/Na <sup>+</sup> /MeOH
14	B <sub>2</sub> 21C7/Na <sup>+</sup> /Me <sub>2</sub> CO

This means that it may not be adequate to estimate the  $\Delta G$  change by the  $\Delta H$  change. On the other hand, the strict geometric requirement of coordination sites of the cation used is very important for covalently interacting complexes. How to estimate the effects of type, number, and arrangement of the coordinating donor



atoms that are forced to conform to the rigid direction requirement is a very difficult problem, especially for the mixed donor macrocycles.

An additional five outliers involve either  $\text{Ca}^{2+}$  or  $\text{Na}^+$  complexes with the cryptand 2.2.1 in various solvents. The equilibrium constants for all of these are *underestimated* by several orders of magnitude. Some limitation in our method causes difficulty with the size-match selectivity of this particular cryptand in these solvents.

The experimental and predicted  $\log K$  values and the residuals of the randomly selected cases for the metal ion-macrocycle system are respectively listed in Table IV. For this system the test set is 60 cases (20% of the 300 total cases). Examining this table, we find that 41 residuals (68% of 60) lie between  $-1.45$  and  $1.44$ , and 57 (95% of 60) between  $-2.72$  and  $2.7$ , so the statistical prediction based on SE (1.42) for this regression model is satisfied.

### 3.2. AMMONIUM ION-CROWN ETHER SYSTEM

This system consists of 33 crown ethers and 29 ammonium ions, 3 pure solvents, and 1 mixed solvent. Equilibrium constant data are from References [36, 37]. The  $\log K$  values range from 0.7 to 6.36. We assume that the van der Waals volume of an ammonium ion roughly equals the sum of the van der Waals volumes of each group composing the ion. From the formula  $V = (4/3)\pi r^3$ , we estimate the radius  $r$  of the ion if the  $V$  value is known. The necessary van der Waals volumes of groups are from References [39, 40]. From the regression result (Table V), we see that the variable  $\delta_1^2$  is the most important variable, that is, there is a large solvent dependency of binding strength of complexes between ammonium ions and crown ethers. The outliers are  $\text{B18C6/C}_4\text{H}_9\text{NH}_3^+/\text{H}_2\text{O}$ ,  $24\text{C8/NH}_4^+/\text{MeOH}$  and  $\text{Cb}_418\text{C6/NH}_4^+/\text{H}_2\text{O}$ .

The test of the prediction ability of the regression model is shown in Table VI. We find that the agreement with theory is good.

### 3.3. HYDROGEN ION-CROWN ETHER SYSTEM

This system consists of 21 mixed donor crown ethers possessing N atoms, and 1 solvent (water). Stability constant data are from References [36, 37]. The  $\log K$  values range from 6.68 to 12.6. Owing to the simple hydrogen ion guest and the single solvent, we find that  $\Delta E_{\text{total}}$  has a good linear relationship with  $\log K$ . The best equation obtained is:

$$\log K = 0.067\Delta E_{\text{total}} + 12.514 \quad (15)$$

$R = 0.94$ ,  $SE = 0.36$ ,  $n = 21$  (3 outliers excluded).

The outliers are  $(\text{AcetM})_2\text{A}_218\text{C6}$ ,  $\text{Me}_6\text{A}_414\text{C4}$  and  $\text{Me}_4\text{A}_414\text{C4}$ .

TABLE IV. Experimental and predicted log  $K$  values and residuals for the randomly selected cases in the metal ion–macrocycle system<sup>a</sup>

Case	Macrocycle/metal ion/solvent	Experimental log $K$	Predicted log $K$	Residual
1	(Meb) <sub>2</sub> 30C10/Na <sup>+</sup> /MeCN	3.6	3.49	0.11
2	B <sub>2</sub> 18C6/Na <sup>+</sup> /MeCN	4.95	5.09	-0.14
3	2.2.2/Ca <sup>2+</sup> /MeCN-H <sub>2</sub> O ( $X_{\text{MeCN}} = 0.05$ )	5.43	5.26	0.17
4	MeOEA18C6/Ca <sup>2+</sup> /MeOH	4.83	4.50	0.33
5	Cy <sub>2</sub> 18C6/Na <sup>+</sup> /MeOH	4.21	3.88	0.33
6	B18C6/Na <sup>+</sup> /Me <sub>2</sub> CO	4.72	4.37	0.35
7	OctOM12C4/Na <sup>+</sup> /MeOH	1.32	1.69	-0.37
8	(Meb) <sub>2</sub> 18C6/Na <sup>+</sup> /DiOX	3.69	3.27	0.42
9	18C6/Na <sup>+</sup> /MeCN	4.3	3.88	0.42
10	B15C5/Na <sup>+</sup> /20%MeOH (MeOH-H <sub>2</sub> O)	0.72	1.16	-0.44
11	1,13-B <sub>2</sub> 24C8/Na <sup>+</sup> /PC	4.16	3.70	0.46
12	1,13-B <sub>2</sub> 24C8/Na <sup>+</sup> /MeCN	4.0	4.47	-0.47
13	B15C5/Na <sup>+</sup> /40%MeOH (MeOH-H <sub>2</sub> O)	1.17	1.64	-0.47
14	K <sub>2</sub> Py2.2.1/Na <sup>+</sup> /H <sub>2</sub> O	4.58	5.13	-0.55
15	Cy15C5/Na <sup>+</sup> /H <sub>2</sub> O	0.3	-0.32	0.62
16	B18C6/Na <sup>+</sup> /MeCN	4.9	4.26	0.64
17	18C6/Na <sup>+</sup> /PY	3.0	3.68	-0.68
18	2MeOPhOM15C5/Na <sup>+</sup> /MeOH	3.24	3.95	-0.71
19	B15C5/Na <sup>+</sup> /PC	4.35	3.64	0.71
20	PhOM15C5/Na <sup>+</sup> /MeOH	3.07	3.81	-0.74
21	2.2.2/Ca <sup>2+</sup> /95%MeOH (MeOH-H <sub>2</sub> O)	7.55	6.76	0.79
22	(HOE) <sub>2</sub> A <sub>2</sub> 18C6/Ca <sup>2+</sup> /H <sub>2</sub> O	4.08	4.88	-0.80
23	B18C6/Ca <sup>2+</sup> /MeCN	5.2	4.40	0.80
24	B18C6/Na <sup>+</sup> /MeOH	4.27	3.43	0.84
25	2.1.1/Na <sup>+</sup> /DMF	5.23	6.08	-0.85
26	18C6/Ca <sup>2+</sup> /MeOH	3.88	3.03	0.85
27	18C6/Na <sup>+</sup> /DiOX	4.54	3.66	0.88
28	Pent2.2/2.2/Na <sup>+</sup> /95%MeOH (MeOH-H <sub>2</sub> O)	3.20	2.27	0.93
29	18C6/Na <sup>+</sup> /EtOH	4.13	3.17	0.96
30	2.2.2/Ca <sup>2+</sup> /MeCN-H <sub>2</sub> O ( $X_{\text{MeCN}} = 0.6$ )	8.12	7.14	0.98
31	18C6/Na <sup>+</sup> /DMF	2.31	3.36	-1.05
32	Meb15C5/Na <sup>+</sup> /DiOX	4.28	3.16	1.12
33	15C5/Ca <sup>2+</sup> /MeOH	2.18	1.05	1.13
34	1,10-A <sub>2</sub> 18C6/Ca <sup>2+</sup> /DMF	2.70	3.84	-1.14
35	B18C6/Na <sup>+</sup> /DMF	2.5	3.67	-1.17

TABLE IV. (continued)

Case	Macrocycle/metal ion/solvent	Experimental log $K$	Predicted log $K$	Residual
36	Monactin/Na <sup>+</sup> /MeCN	4.28	2.91	1.37
37	1,10-A <sub>2</sub> 18C6/Ca <sup>2+</sup> /Me <sub>2</sub> SO	2.35	3.72	-1.37
38	(Meb) <sub>2</sub> 18C6/Na <sup>+</sup> /MeCN	5.1	3.70	1.40
39	B2.2.2/Ca <sup>2+</sup> /PC	10.10	8.66	1.44
40	2.2.2/Na <sup>+</sup> /PrOH	8.39	6.95	1.44
41	B <sub>2</sub> 18C6/Na <sup>+</sup> /Me <sub>2</sub> SO	2.62	4.07	-1.45
42	2.1.1/Ca <sup>2+</sup> /MeCN-H <sub>2</sub> O ( $X_{\text{MeCN}} = 0.4$ )	5.0	6.51	-1.51
43	Oct18C6/Na <sup>+</sup> /MeOH	3.91	2.38	1.53
44	15C5/Na <sup>+</sup> /MeCN	5.09	3.54	1.55
45	7,9-Me <sub>2</sub> B215C5/Na <sup>+</sup> /MeOH	2.0	3.59	-1.59
46	Dinactin/Na <sup>+</sup> /EtOH	3.6	2.00	1.60
47	2.2.2/Na <sup>+</sup> /Me <sub>2</sub> SO	5.3	7.00	-1.70
48	B15C5/Na <sup>+</sup> /DMF	1.6	3.33	-1.73
49	Dodec18C6/Na <sup>+</sup> /MeOH	3.93	2.02	1.91
50	Dinactin/Na <sup>+</sup> /MeCN	4.44	2.35	2.09
51	2.2.1/Ca <sup>2+</sup> /Me <sub>2</sub> SO	3.9	6.43	-2.53
52	Ph18C6/Na <sup>+</sup> /MeOH	4.17	1.62	2.55
53	6,10-Me <sub>2</sub> B <sub>2</sub> 15C5/Na <sup>+</sup> /MeOH	1.3	3.85	-2.55
54	2,3,3,3-A <sub>4</sub> 15C4/Zn <sup>2+</sup> /H <sub>2</sub> O	15.0	12.36	2.64
55	Nonactin/Na <sup>+</sup> /EtOH	3.27	5.96	-2.69
56	2.2.2/Ca <sup>2+</sup> /MeCN	10.5	7.80	2.70
57	B <sub>2</sub> 21C7/Na <sup>+</sup> /PY	2.56	5.28	-2.72
58	(CbMA) <sub>4</sub> 12C4/Ca <sup>2+</sup> /H <sub>2</sub> O	17.23	14.41	2.82
59	(CbMA) <sub>4</sub> 13C4/Ca <sup>2+</sup> /H <sub>2</sub> O	12.09	15.48	-3.39
60	(1.1/1.1)Ca <sup>2+</sup> /H <sub>2</sub> O	6.53	1.83	4.70

<sup>a</sup> The  $SE$  is 1.42; 68% of 60 is 41; 95% of 60 is 57.

TABLE V. The unstandardized coefficients ( $B$ ) and  $t$  ratios in the ammonium ion-crown ether system<sup>a</sup>

Variables	$B$	$t$ ratio
$\delta_t^2$	-0.001	-7.205
$1/\varepsilon$	2.501	3.738
$\mu_{[1]}^2 / (r_{1,[1]}^3)^3$	-1.901	-3.785
$\Delta E_{\Phi}$	-0.142	-2.309
$\Delta E_{\text{vdw}}$	0.102	2.916
$\sigma r_M^2$	0.015	1.818
constant	4.508	16.147

<sup>a</sup>  $R = 0.91$ ;  $SE = 0.64$ ;  $n = 85$  (3 outliers excluded).

TABLE VI. Experimental and predicted  $\log K$  values and residuals for the randomly selected cases in the ammonium ion-crown ether system<sup>a</sup>

Case	Crown ether/ammonium ion/solvent	Experimental $\log K$	Predicted $\log K$	Residual
1	Cb <sub>4</sub> 18C6/CH <sub>3</sub> (CH <sub>2</sub> ) <sub>3</sub> NH <sub>3</sub> <sup>+</sup> /H <sub>2</sub> O	2.3	2.42	-0.12
2	B18C6/ <i>t</i> -C <sub>4</sub> H <sub>9</sub> NH <sub>3</sub> <sup>+</sup> /C <sub>6</sub> H <sub>6</sub>	5.0	5.13	-0.13
3	18C6/3,5-(CH <sub>3</sub> ) <sub>2</sub> phNH <sub>3</sub> <sup>+</sup> /MeOH	3.74	3.50	0.24
4	Cb <sub>4</sub> 18C6/(CH <sub>3</sub> ) <sub>3</sub> CNH <sub>3</sub> <sup>+</sup> /H <sub>2</sub> O	1.6	1.90	-0.30
5	Me(OE) <sub>5</sub> A18C6/NH <sub>4</sub> <sup>+</sup> /90%MeOH (MeOH-H <sub>2</sub> O)	4.05	3.74	0.31
6	B18C6/ <i>n</i> -C <sub>4</sub> H <sub>9</sub> NH <sub>3</sub> <sup>+</sup> /C <sub>6</sub> H <sub>6</sub>	5.18	5.55	-0.37
7	18C6/C <sub>2</sub> H <sub>5</sub> NH <sub>3</sub> <sup>+</sup> /MeOH	3.99	3.57	0.42
8	18C6/NH <sub>4</sub> <sup>+</sup> /H <sub>2</sub> O	1.23	1.69	-0.46
9	18C6/ <i>n</i> -C <sub>3</sub> H <sub>7</sub> NH <sub>3</sub> <sup>+</sup> /MeOH	3.97	3.50	0.47
10	(GluCba) <sub>4</sub> 18C6/NH <sub>4</sub> <sup>+</sup> /H <sub>2</sub> O	2.4	1.91	0.49
11	B18C6/ <i>sec</i> -C <sub>4</sub> H <sub>9</sub> NH <sub>3</sub> <sup>+</sup> /C <sub>6</sub> H <sub>6</sub>	4.88	5.40	-0.52
12	18C6/C <sub>2</sub> H <sub>5</sub> OC(O)CH <sub>2</sub> NH <sub>3</sub> <sup>+</sup> /MeOH	3.84	3.27	0.57
13	B18C6/ <i>i</i> -C <sub>3</sub> H <sub>7</sub> NH <sub>3</sub> <sup>+</sup> /C <sub>6</sub> H <sub>6</sub>	5.16	5.74	-0.58
14	Cb <sub>4</sub> 18C6/HO(CH <sub>2</sub> ) <sub>2</sub> NH <sub>3</sub> <sup>+</sup> /H <sub>2</sub> O	2.7	1.92	0.78
15	18C6/CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup> /MeOH	4.25	3.40	0.85
16	18C6/ <i>n</i> -C <sub>3</sub> H <sub>7</sub> NH <sub>3</sub> <sup>+</sup> /C <sub>6</sub> H <sub>6</sub>	6.36	5.39	0.97
17	18C6/ <i>n</i> -C <sub>5</sub> H <sub>11</sub> NH <sub>3</sub> <sup>+</sup> /C <sub>6</sub> H <sub>6</sub>	6.36	5.21	1.15

<sup>a</sup> The *SE* is 0.64; 68% of 17 is 12; 95% of 17 is 16.

TABLE VII. The unstandardized coefficients (*B*) and *t* ratios in the Na<sup>+</sup> ion-spherand system<sup>a</sup>

Variables	<i>B</i>	<i>t</i> ratio
$z_M \mu_{[ ]} / (r_{2,[M]})^2$	92.313	5.275
$\mu_{[M]}^2 / (r_{2,[M]})^3$	47.823	6.065
$E_{\text{vdw},[ ]}$	-1.231	-6.028
$E_{\text{elec},[M]}$	-0.222	3.919
$\sigma(r_{2,[ ]})^2$	0.111	4.675
$E_{\Phi,[M]}$	0.110	2.688
$E_{\text{vdw},[M]}$	0.365	1.609
$E_{\text{elec},[ ]}$	0.147	1.251
constant	-21.568	-4.791

<sup>a</sup> *R* = 0.95; *SE* = 1.2; *n* = 22 (4 outliers excluded).

### 3.4. NA<sup>+</sup> ION-SPHERAND SYSTEM

The system consists of 22 spherands and 1 solvent (CDCl<sub>3</sub> saturated with D<sub>2</sub>O). Equilibrium constant data are from References [41–45]. The  $\log K$  values range

TABLE VIII. The unstandardized coefficients ( $B$ ) and  $t$  ratios in the ammonium ion–spherand system<sup>a</sup>

Variables	$B$	$t$ ratio
$E_{\text{elec},[M]}$	-0.181	-6.540
$E_{\text{inter},[M]}$	0.082	3.786
$\mu_{[ ]}^2/(r_{2,[ ]})^3$	-24.211	-3.419
$z_M\mu_{[ ]}/(r_{2,[M]})^2$	66.370	3.857
$E_{\text{vdw},[M]}$	-0.232	-2.788
$E_{\text{elec},[ ]}$	-0.062	-1.129
$\sigma(r_{2,[M]})^2$	-0.016	-1.754
$\mu_{[M]}^2/(r_{2,[M]})^3$	1.445	1.869
$z_M\mu_{[ ]}/(r_{2,[ ]})^2$	-33.996	-1.911
$E_{\Phi,[M]}$	-0.089	-2.133
$E_{\text{vdw},[ ]}$	-0.128	-1.516
$\sigma r_M^2$	0.035	1.364
constant	3.179	2.013

<sup>a</sup>  $R = 0.86$ ;  $SE = 0.96$ ;  $n = 70$  (8 outliers excluded).

from 4.69 to 15.4. The regression result is shown in Table VII. The outliers are spherand-1, spherand-2, spherand-3 and spherand-4.

### 3.5. AMMONIUM ION–SPHERAND SYSTEM

This system consists of 30 spherands and 1 solvent ( $\text{CDCl}_3$  saturated with  $\text{D}_2\text{O}$ ), and 3 ammonium ions ( $\text{NH}_4^+$ ,  $\text{CH}_3\text{NH}_3^+$ ,  $(\text{CH}_3)_3\text{CNH}_3^+$ ). Equilibrium constant data are from References [41, 44, 45]. The  $\log K$  values range from 4.69 to 14.8. The outliers are spherand-5/ $\text{tBuNH}_3^+$ , spherand-6/ $\text{NH}_4^+$  and  $\text{CH}_3\text{NH}_3^+$ , spherand-7/ $\text{CH}_3\text{NH}_3^+$ , spherand-8/ $\text{NH}_4^+$  and spherand-9/ $\text{NH}_4^+$ ,  $\text{CH}_3\text{NH}_3^+$  and  $\text{tBuNH}_3^+$ .

This regression result (Table VIII) is not as good as for the  $\text{Na}^+$  ion–spherand system when only considering  $R$ , but, it is better in terms of  $SE$ . Furthermore, there are quite a lot more cases in the ammonium ion–spherand system. Thus, the regression results either in the  $\text{Na}^+$  ion–spherand or in the ammonium ion–spherand system all show a good correlation of equilibrium constants with the variables, even though the spherands are very highly preorganized, and the spherands employ different preorganized structures for different guests in which this special effect is difficult to estimate effectively. Note that seven of the eight variables selected as most important in the  $\text{Na}^+$  ion–spherand system are also selected in the ammonium ion–spherand system. Also note that the outliers in the two systems have no spherand in common.

From Table IX, we find that the predictive power of the resulting model is good.

TABLE IX. Experimental and predicted  $\log K$  values and residuals for the randomly selected cases in the ammonium ion–spherand system<sup>a</sup>

Case	spherand/ammonium ion	Experimental $\log K$	Predicted $\log K$	Residual
1	(Spherand)-6 <sup>b</sup> /NH <sub>4</sub> <sup>+</sup>	6.09	6.13	-0.04
2	(Spherand)-5 <sup>b</sup> /NH <sub>4</sub> <sup>+</sup>	6.82	7.03	-0.21
3	(Spherand)-37 <sup>c</sup> /NH <sub>4</sub> <sup>+</sup>	5.87	6.23	-0.37
4	(Spherand)-19 <sup>c</sup> /CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	5.21	5.63	-0.43
5	(Spherand)-26 <sup>c</sup> /CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	7.19	7.67	-0.49
6	(Spherand)-38 <sup>c</sup> /NH <sub>4</sub> <sup>+</sup>	6.16	6.72	-0.56
7	(Spherand)-32 <sup>c</sup> /CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	6.01	5.30	0.72
8	(Spherand)-12 <sup>c</sup> /NH <sub>4</sub> <sup>+</sup>	9.31	10.09	-0.78
9	(Spherand)-3 <sup>d</sup> /NH <sub>4</sub> <sup>+</sup>	8.07	9.31	-1.24
10	(Spherand)-13 <sup>c</sup> / <i>t</i> -BuNH <sub>3</sub> <sup>+</sup>	6.09	4.68	1.41
11	(Spherand)-26 <sup>c</sup> /NH <sub>4</sub> <sup>+</sup>	6.97	8.66	-1.69
12	(Spherand)-2 <sup>c</sup> / <i>t</i> -BuNH <sub>3</sub> <sup>+</sup>	9.68	7.86	1.82
13	(Spherand)-8 <sup>b</sup> /NH <sub>4</sub> <sup>+</sup>	5.43	7.31	-1.88
14	(Spherand)-13 <sup>c</sup> /CH <sub>3</sub> NH <sub>3</sub> <sup>+</sup>	6.60	4.68	1.92

<sup>a</sup> The solvent is CDCl<sub>3</sub> saturated with D<sub>2</sub>O; the *SE* is 0.96; 68% of 14 is 10; 95% of 14 is 13.

<sup>b</sup> These compounds were only numbered in reference 45.

<sup>c</sup> These compounds were only numbered in reference 41.

<sup>d</sup> These compounds were only numbered in reference 44.

### 3.6. HOST-GUEST SYSTEM

In this system, hosts are of the cyclophane type, whereas guests are naphthalene derivatives, benzene derivatives, steroids, paracyclophanes, and alicyclic and aromatic hydrocarbons. The solvents are H<sub>2</sub>O, D<sub>2</sub>O, D<sub>2</sub>O/CD<sub>3</sub>OD, and methanol-*d*<sub>4</sub>. The solvent parameters of methanol and H<sub>2</sub>O/CH<sub>3</sub>OH are approximately applied to methanol-*d*<sub>4</sub> and D<sub>2</sub>O/CD<sub>3</sub>OD, respectively. The necessary solvent parameters of D<sub>2</sub>O are from Reference [46]. Equilibrium constant data are from References [47–50]. The  $\log K$  values range from 1.3 to 7.16.

The regression results and the predictive ability of the final regression model are shown in Tables X and XI, respectively. Again, the predictive ability is seen to be good. The outliers are cyclophane-5/perylene/H<sub>2</sub>O, cyclophane-5/pyrene/H<sub>2</sub>O and cyclophane-2/*p*-diaminobenzene/D<sub>2</sub>O.

### 3.7. COMPARISON OF THE REGRESSION RESULTS USING THE AMBER AND MMP2 FORCE FIELDS

In this study, we also did the multivariate regression for the metal ion–macrocycle, ammonium ion–crown ether, and hydrogen ion–crown ether systems using energies

TABLE X. The unstandardized coefficients ( $B$ ) and  $t$  ratios in the host-guest system<sup>a</sup>

Variables	$B$	$t$ ratio
$E_{\text{inter},[\text{M}]}$	0.028	5.965
$E_{\text{inter},[ ]}$	0.061	3.288
$E_{\text{vdw},[\text{M}]}$	-0.028	-3.147
$\sigma(r_{3,\text{M}})^2$	0.008	5.827
$1/\epsilon$	174.931	6.566
$\sigma(r_{3,[ ]})^2$	0.010	6.521
$\mu_{[\text{M}]}^2 / (r_{3,[\text{M}]})^3$	-0.117	-4.209
$E_{\text{elec},[\text{M}]}$	0.036	4.135
$\sigma(r_{3,[\text{M}]})^2$	-0.006	-4.118
constant	-9.488	-8.114

<sup>a</sup>  $R = 0.95$ ;  $SE = 0.44$ ;  $n = 70$  (3 outliers excluded).

computed with the MMP2 force field, but with energy expressions and ‘best’ sets of variables determined for AMBER. For the metal ion-macrocyclic system, the MMP2 results were slightly inferior to AMBER ( $SE$  of 1.70 versus 1.42) whereas for other systems the  $SE$  were identical for the two force fields. Thus, we may say that the regression results are not very sensitive to the force field, especially for restricted classes of complexes which do not involve metals.

#### 4. Summary

We have presented a method for the quantitative prediction of the complexing properties of macrocyclic ligands. The method involves the combination of solvent-free molecular mechanics and molecular dynamics simulations with multiple linear regression to experimental  $\log K$  data to incorporate solvent and other effects. The predictive power of the method is good; for the largest and most varied system considered, the standard error in  $\log K$  is 1.42 which corresponds to a factor of 26 in  $K$  itself.

The major advantages of this approach over detailed simulations which include the solvent are speed and economy. Only a few hours of computer time are required on our Stardent Titan; on modern workstations this might be reduced to a few minutes. Furthermore, no experimental structural information for the macrocycle is needed, so the method is applicable to macrocycles which have not been yet synthesized. New solvents also can be treated if sufficient input data for the regression are available.

The need for an experimental database is only a minor disadvantage given the very large number of complexes for which  $\log K$  has been determined. The major disadvantage of the method is that the predictions are only statistically accurate. It is not possible to guarantee the accuracy of the predicted  $\log K$  for any single

TABLE XI. Experimental and predicted log  $K$  values and residuals for the randomly selected cases in the host-guest system<sup>a</sup>

Case	Host/guest/solvent	Experimental log $K$	Predicted log $K$	Residual
1	Cyclophane-2 <sup>b</sup> /16 <sup>b</sup> /D <sub>2</sub> O-CD <sub>2</sub> OD	2.56	2.51	0.05
2	Cyclophane-2 <sup>b</sup> /17 <sup>b</sup> /D <sub>2</sub> O-CD <sub>3</sub> OD	2.64	2.73	-0.09
3	Cyclophane-20 <sup>c</sup> /1,5- Bis(dimethylamino)naphthalene/H <sub>2</sub> O	4.03	4.16	-0.13
4	Cyclophane-20 <sup>c</sup> /1,5- Naphthalenesulfonate/H <sub>2</sub> O	5.67	5.83	-0.16
5	Cyclophane-2 <sup>b</sup> /Cholic acid/D <sub>2</sub> O- CD <sub>3</sub> OD	2.16	2.32	-0.16
6	Cyclophane-1 <sup>d</sup> /21 <sup>d</sup> /methanol- <i>d</i> <sub>4</sub>	2.03	2.20	-0.17
7	Cyclophane-2 <sup>c</sup> / <i>p</i> -xylene/D <sub>2</sub> O	3.98	3.78	0.20
8	Cyclophane-20 <sup>c</sup> /2,7- Naphthalenediol/H <sub>2</sub> O	4.25	4.03	0.22
9	Cyclophane-2 <sup>b</sup> /13 <sup>b</sup> /D <sub>2</sub> O-CD <sub>3</sub> OD	1.99	1.76	0.23
10	Cyclophane-2 <sup>b</sup> /12 <sup>b</sup> /D <sub>2</sub> O-CD <sub>3</sub> OD	2.92	2.62	0.30
11	Cyclophane-2 <sup>b</sup> /18 <sup>b</sup> /D <sub>2</sub> O-CD <sub>3</sub> OD	2.64	3.04	-0.40
12	Cyclophane-20 <sup>c</sup> / <i>p</i> - Toluenesulfonate/H <sub>2</sub> O	3.88	4.35	-0.47
13	Cyclophane-2 <sup>b</sup> /ursodeoxycholic acid/D <sub>2</sub> O-CD <sub>3</sub> OD	3.24	2.12	1.12
14	Cyclophane-20 <sup>c</sup> /1- (Trimethylammonium)naphthalenefluor <i>o</i> -sulfonate/H <sub>2</sub> O	3.21	4.38	-1.17

<sup>a</sup> The *SE* is 0.44; 68% of 14 is 10; 95% of 14 is 13.

<sup>b</sup> These compounds were only numbered in reference 50.

<sup>c</sup> These compounds were only numbered in reference 49.

<sup>d</sup> These compounds were only numbered in reference 47.

<sup>e</sup> These compounds were only numbered in reference 48.

complex, only the average accuracy of a large number of them. Thus, this approach is complementary to, but certainly no replacement for, more detailed simulations. It would be most useful for applications such as the rapid screening of large numbers of possible complexes to identify promising candidates, which might then be subjected to experimental characterization or more accurate simulation.

We view the present work only as a preliminary study which demonstrates the feasibility of this type of approach. We certainly make no claim that the predictive accuracy of the equations we have presented cannot be improved upon. Of course, as the results for the hydrogen ion-crown ether-water system suggest, this can always be done by restricting the class of complexes sufficiently. For example, splitting the metal ion-macrocycle-mixed solvent system into separate systems for



each cation, or each macrocycle, or each solvent would probably lead to smaller  $SE$ , but at the expense of generality.

More interesting is the possibility of finding other variables which will improve the accuracy while retaining the generality. From this study, we have found that representing the covalent energy of metal complexes presents one potential problem area. Another is the macrocycle–solvent surface interaction, which is highly oversimplified in our treatment. It is quite likely that the results of accurate simulations including solvent can provide guidance in describing the latter effect.

## Acknowledgement

We wish to thank Professor M.P. Windham of the Department of Mathematics and Statistics for consultations regarding the statistical analysis.

## References

1. *Synthetic Multidentate Macrocyclic Compounds*, R.M. Izatt and J.J. Christensen Eds., Academic, New York (1978).
2. L.F. Lindoy: *The Chemistry of Macrocyclic Ligand Complexes*, Cambridge University, New York (1989).
3. *Cation Binding by Macrocycles*, Y. Inoue and G.W. Gokel Eds., Marcel Dekker, New York (1990).
4. G.W. Gokel: *Crown Ethers and Cryptands*, The Royal Society of Chemistry, Cambridge (1991).
5. A. Pullman, C. Giessner-Prettre, and Y.V. Kruglyak: *Chem. Phys. Lett.* **35**, 156 (1975).
6. T. Yamabe, K. Hori, K. Akagi, and K. Fukkui: *Tetrahedron* **35**, 1065 (1979).
7. K. Hori, H. Yamada, and T. Yamabe: *Tetrahedron* **39**, 67 (1983).
8. M.J. Bovill, D.J. Chadwick, I.O. Sutherland, and D. Watkin: *J. Chem. Soc. Perkin Trans. 2*, 1529 (1980).
9. G. Wipff, P. Weiner, and P. Kollman: *J. Am. Chem. Soc.* **104**, 3249 (1982).
10. D. Gehin, P.A. Kollman, and G. Wipff: *J. Am. Chem. Soc.* **111**, 3011 (1989).
11. P. Auffinger and G. Wipff: *J. Am. Chem. Soc.* **113**, 5976 (1991).
12. R.D. Hancock: *Prog. Inorg. Chem.* **37**, 187 (1989).
13. R.D. Hancock and A.E. Martell: *Chem. Rev.* **89**, 1875 (1989).
14. R. Dharanipragada, S.B. Ferguson, and F. Diederich: *J. Am. Chem. Soc.* **110**, 1679 (1988).
15. K.S. Kim and E.J. Clementi: *J. Phys. Chem.* **89**, 3655 (1985).
16. P.K. Mehrota and D.L. Beveridge: *J. Am. Chem. Soc.* **102**, 428 (1980).
17. J. Chandrasekhar, D.C. Spellmeyer, and W.L. Jorgensen: *J. Am. Chem. Soc.* **106**, 903 (1984).
18. G. Alagona, C. Ghio, and P.A. Kollman: *J. Am. Chem. Soc.* **108**, 185 (1986).
19. K. Heinzinger: *Pure Appl. Chem.* **57**, 1031 (1985).
20. Y. Inoue, T. Hakushi, Y. Liu: *Thermodynamics of Cation–Ligand Complexation*, (Cation Binding by Macrocycles, Eds. Y. Inoue and G.W. Gokel), pp. 1–110, Marcel Dekker, New York (1990).
21. BioGraf Reference Manual Version 2.1, Biodesign, Inc., Pasadena, CA 91101.
22. J. Gasteiger: *Tetrahedron* **36**, 3219 (1980).
23. S.J. Weiner, U.C. Singh, and T.J. O'Donnell: *J. Am. Chem. Soc.* **106**, 765 (1984).
24. S.J. Weiner, P.A. Kollman, D.T. Nguyen, and D.A. Case: *J. Comput. Chem.* **7**, 230 (1986). This force field implementation was made by BioDesign, Inc.
25. J.T. Sprague, J.C. Tai, Y. Yuh, and N.L. Allinger: *J. Comput. Chem.* **8**, 581 (1987). This force field implementation was made by BioDesign, Inc.
26. S.W. Fesik, T.J. O'Donnell, R.T. Gampe, Jr., and E.T. Olejniczak: *J. Am. Chem. Soc.* **108**, 3165 (1986).
27. J.E. Huheey: *Inorganic Chemistry*, 3rd ed., Harper & Row, New York (1983).
28. W.C. Still, A. Tempczak, R.C. Hawley, and T. Hendrickson: *J. Am. Chem. Soc.* **112**, 6127 (1990).

29. C.J. Cramer and D.G. Truhlar: *J. Am. Chem. Soc.* **113**, 8305 (1991).
30. A.F.M. Barton: *CRC Handbook of Solubility Parameters and Other Cohesion Parameters*, CRC, Boca Raton, Florida (1983).
31. K.A. Connors: *Chemical Kinetics*, VCH, New York (1990).
32. C.J. Cramer and D.G. Truhlar: *Science* **256**, 213 (1992).
33. W.B. Jensen: *The Lewis Acid-Base Concepts*, John Wiley & Sons, New York (1980).
34. S. Chatterjee and B. Price: *Regression Analysis by Example*, 2nd ed., John Wiley & Sons, New York (1991).
35. Statistical Package for the Social Science, SPSS-X Users Guide, 3rd ed., SPSS Inc., Chicago (1988).
36. R.M. Izatt, J.S. Bradshaw, S.A. Nielsen, J.D. Lamb, and J.J. Christensen: *Chem. Rev.* **85**, 271 (1985).
37. R.M. Izatt, K. Pawlak, J.S. Bradshaw, and R.L. Bruening: *Chem. Rev.* **91**, 1721 (1991).
38. A.E. Martell and R.M. Smith: *Critical Stability Constants*, Vols. 5-6, Plenum, New York (1982, 1989).
39. E.J. King: *J. Phys. Chem.* **74**, 4590 (1970).
40. A. Bondi: *J. Phys. Chem.* **68**, 441 (1964).
41. D.J. Cram: *Angew. Chem. Int. Ed. Engl.* **25**, 1039 (1986).
42. D.J. Cram and S.P. Ho: *J. Am. Chem. Soc.* **108**, 2998 (1986).
43. D.J. Cram, R.A. Carmack, and R.C. Helgeson: *J. Am. Chem. Soc.* **110**, 571 (1988).
44. J.A. Bryant, S.P. Ho, C.B. Knobler, and D.J. Cram: *J. Am. Chem. Soc.* **112**, 5837 (1990).
45. J.A. Bryant, R.C. Helgeson, C.B. Knobler, M.P. deGrandpre, and D.J. Cram: *J. Org. Chem.* **55**, 4622 (1990).
46. Y. Marcus: *Ion Solvation*, John Wiley & Sons, New York (1985).
47. S.B. Ferguson and F. Diederich: *Angew. Chem. Int. Ed. Engl.* **25**, 1127 (1986).
48. S.B. Ferguson, E.M. Seward, F. Diederich, E.M. Sanford, A. Chou, P. Inocencio-Szweda, and C.B. Knobler: *J. Org. Chem.* **53**, 5593 (1988).
49. F. Diederich: *Angew. Chem. Int. Ed. Engl.* **27**, 362 (1988).
50. D.R. Carcanague and F. Diederich: *Angew. Chem. Int. Ed. Engl.* **29**, 769 (1990).