Molecular Recognition in Terms of a Dimensionless Index. 1. A Theoretical Model of the Orientation-Based Fit and Its Applicability to Ligand−Macromolecule Systems

Jian Chen*

College of Chemical Engineering, Zhejiang University of Technology, Hangzhou, 310014, P. R. China

Received: July 19, 2002; In Final Form: February 9, 2003

Large quantities of thermodynamic parameters that characterize the formation of weakly bound complexes have dramatically accumulated in the literature. Parameters such as enthalpy and entropy changes that accompany ligand−receptor interactions contain the underlying information that richly describes the mechanism of molecular interaction and have been instrumental in revealing the mechanism of molecular recognition. Molecular recognition is ultimately achieved via intermolecular orientation; the latter has beenspecifically defined as "orientation-based fit" in the present article. In the mean field approximation, a model showed that the compensatory enthalpy−entropy relationship operates as a consequence of a complexation series being studied that occurs far enough away from the ideal complexation. Then, a dimensionless factor was proposed to measure the orientation-based fit between constituent components, in terms of the enthalpy and entropy associated with intermolecular interaction. In addition, the applicability of this dimensionless factor has been satisfactorily demonstrated in the ligand−macromolecule system that consists of small molecules and two poly(ethylene glycol) (PEG) compounds (PEG-6000 and PEG-1000). All the small-molecule−PEG systems are assigned to one of three classes, on the basis of the mechanisms of their individual weak interaction with PEG. The flexibility of PEG has conflicting effects on molecular recognition.

1. Introduction

Weak noncovalent interactions—a "soft" constraint, with respect to chemical bond binding—generally operate between the ligand and the receptor. Individual molecules aggregate via intermolecular interaction, forming the so-called "supramolecular structure", which then can exert specific physicochemical functions that the constituent molecules cannot afford. In chemistry, noncovalent interactions are now exploited for the synthesis of large supramolecular aggregates in solution.[1] However, "mere binding is not recognition, although it is often taken as such." Molecular recognition "implies a structurally well-defined pattern of intermolecular interactions".[2] The "soft" constraint and the large number of conformational states available to a ligand−macromolecule system during molecular recognition make it very difficult to assess a structurally well-defined pattern of intermolecular interactions. Molecular recognition that occurs during transient interaction processes is especially difficult to identify. Understanding the structural basis of molecular recognition presents a fundamental theoretical challenge and has significant practical importance in drug and material design.[3−5]

From a theoretical point of view, intermolecular interactions involve complex many-body systems, and molecular recognition is ultimately achieved via intermolecular orientation between constituent components (also called "site-specific interaction"[6]). Supramolecular assemblies generally include a combination of several local orientations (multisite interaction) between constituent components. Every minimum on a potential interface is responsible for a possible fit mode. For many systems of interest, high energy barriers separate the basins (local minima). Although a large variety of docking methods, including the widely used GOLD, FLEXx, and DOCK programs, and empirical free-energy functions have been developed in recent years and have been favorable tools for making progress toward more-reliable computational supramolecular structure prediction,[5] these barriers can make it difficult to achieve global or local minima of interest. This problem is called "broken ergodicity".[7]

Recently, large quantities of thermodynamic parameters that characterize the formation of weakly bound complexes have dramatically accumulated in the literature. Parameters such as enthalpy and entropy changes that accompany ligand−receptor interactions have been instrumental in revealing the mechanism of molecular recognition.[8,9] Moreover, it was found that the conformational change and the desolvation upon formation of the weakly bound complex can be revealed by the slope and intercept of the compensatory enthalpy−entropy relationship,[10,11] although the compensatory relationship has been observed empirically in thermodynamic quantities determined so often for a wide variety of reactions and equilibria that its absence is more striking—and, therefore, more interesting—than its presence.[12] Currently, the structural basis of the enthalpy and entropy changes in selective recognition processes still is not well-known. Determination of this basis is the object of this study.

The path to this goal is complicated, because of at least two factors: the aforementioned problem of broken ergodicity involved in computer simulation calculations for molecular conformations and the absence of an explicit, compensatory enthalpy−entropy relationship. An approach to simple ligand−macromolecule systems that contain no solvent (gas−liquid systems) is developed, and a dimensionless index is presented for the measurement of the orientation-based fit responsible for molecular recognition. In addition, its applicability to the understanding of the nature of intermolecular interactions operating within supramolecular assemblies, and also the factors governing the complexation by macromolecules, both are illustrated.

E-mail address: chenzhz@sohu.com.

10.1021/jp026570y CCC: $25.00
Published on Web 03/21/2003

2. Theoretical Model

2.1. Background. There are several methods available to correlate thermodynamic parameters with weak noncovalent interaction (for instance, statistic thermodynamic approaches,[13,14] linear solvation energy relationship (LSER) theory,[15] and mean field approximation (MFA) and perturbation theory[16]). Among them, less attention was paid to the use of the MFA in conjunction with simulation calculations of liquids, although its theoretical bases are well-known.[17−19] The methodology of MFA has been successfully applied to the calculation of thermodynamic functions for some organic liquids[20] and to the computation of quantum mechanics/molecular mechanics with the solvent effect.[19] MFA is used as the building block for modeling orientation-based fits that exist in the gas−liquid system.

For a receptor−ligand (1:1) complexation,

$$\mathrm{H(liq)} + \mathrm{G(gas)} = \mathrm{C(soln)}$$

where H, G, and C represent the host, guest, and 1:1 molecular complex molecules, respectively. Its solution thermodynamic functions, free energy, and entropy, are expressed as

$$\begin{aligned}\Delta F &= F(\mathrm{C}) - F(\mathrm{H}) - F(\mathrm{G})\\ &= -kT(\ln Q(\mathrm{C}) - \ln Q(\mathrm{H}) - \ln Q(\mathrm{G}))\\ &= -kT[(\ln Q_{\mathrm{trans}}(\mathrm{C}) - \ln Q_{\mathrm{trans}}(\mathrm{H}) - \ln Q_{\mathrm{trans}}(\mathrm{G}))\\ &\quad + (\ln Q_{\mathrm{rot}}(\mathrm{C}) - \ln Q_{\mathrm{rot}}(\mathrm{H}) - \ln Q_{\mathrm{rot}}(\mathrm{G}))\\ &\quad + (\ln Q_{\mathrm{vib}}(\mathrm{C}) - \ln Q_{\mathrm{vib}}(\mathrm{H}) - \ln Q_{\mathrm{vib}}(\mathrm{G}))\\ &\quad + (\ln Q_{\mathrm{bkg}}(\mathrm{C}) - \ln Q_{\mathrm{bkg}}(\mathrm{H}) - \ln Q_{\mathrm{bkg}}(\mathrm{G}))]\\ &= \Delta F_{\mathrm{trans}} + \Delta F_{\mathrm{rot}} + \Delta F_{\mathrm{vib}} + \Delta F_{\mathrm{bkg}} \qquad (1)\end{aligned}$$

$$\begin{aligned}\Delta S &= \frac{\partial \Delta F}{\partial T}\\ &= \frac{\partial \Delta F_{\mathrm{trans}}}{\partial T} + \frac{\partial \Delta F_{\mathrm{rot}}}{\partial T} + \frac{\partial \Delta F_{\mathrm{vib}}}{\partial T} \qquad (2)\end{aligned}$$

Here,

$$Q_{\mathrm{bkg}} = \exp\left(-\frac{\langle V_{\mathrm{N}}\rangle}{kT}\right) \qquad (3)$$

In the previous equations, the variables k and T have their usual thermodynamic meanings. In this model, the perturbation due to the semi-independent kinetic motion of chain segments of the macromolecule is introduced in an average way. More specifically, the quantity Q_{bkg} that is introduced into the model is the averaged value of the condensed phase potential.

2.2. Ideal Complexation. The ideal complexation is now defined, assuming that a weakly bound complex is featured in terms of the vibrational, rotational, and translational motions of the entire body of the complex, without the respective independent relative motions between the constituent components. In the ideal complexation, constituent components are rigidly constrained together to form the entire body by both binding interactions and geometrical steric constraints, with the complex translating and rotating as a rigid body; vibrational modes for the constituent components are less affected, because weakly binding interactions operate in supramolecular assemblies. The vibrational entropy is generally less than half of the translational entropy, because of the wider interval between vibrational energy levels. It is assumed that the vibrational motions for the constituent components in the structure of a weakly bound complex are independent of each other and that $Q_{\mathrm{vib}}(\mathrm{C})$ can be expressed as the product of $Q_{\mathrm{vib}}(\mathrm{H})$ and $Q_{\mathrm{vib}}(\mathrm{G})$. Thus, ΔF_{vib} term is equal to zero.

Liquid macromolecules behave similar to a troupe of disordered threads. For instance, poly(ethylene gylcol) (PEG) with a mean molecular weight of 6000 (PEG-6000), which is used in the present work, comprises approximately 139 motifs. They can wrap around each other. As an approximation, we assume that the receptor macromolecules remain still and fixed in the supported liquid film. Thus, in the ideal complexation, the previously described expression for ΔS is simplified as shown below:

$$\begin{aligned}\Delta S &= \frac{\partial \Delta F_{\mathrm{trans}}}{\partial T} + \frac{\partial \Delta F_{\mathrm{rot}}}{\partial T}\\ &= -\frac{\partial}{\partial T}(kT \ln Q_{\mathrm{trans}}(\mathrm{G})) - \frac{\partial}{\partial T}(kT \ln Q_{\mathrm{rot}}(\mathrm{G}))\\ &= -3R - k(\ln Q_{\mathrm{trans}}(\mathrm{G}) + \ln Q_{\mathrm{rot}}(\mathrm{G}))\\ &= \Delta S_{\max} \qquad (4)\end{aligned}$$

R has its usual thermodynamic meaning. When the organic vapor phase G can be treated as an ideal gas, the aforementioned expression can be specifically written as

$$\begin{aligned}\Delta S_{\max} = -\Bigg(3R - R\ln\left(\frac{2\pi kT}{h^2}\right)^{3/2} eV_{\mathrm{eff}} -\\ R\ln\left[8\pi^2\left(\frac{2\pi kT}{h^2}\right)^{3/2}\right]\Bigg) - R\ln\left[m^{3/2}\frac{(I_A I_B I_C)^{1/2}}{\sigma}\right] \qquad (5)\end{aligned}$$

where V_{eff} is the effective free volume; I_A, I_B, and I_C are the moments of inertia; and σ is the symmetry factor.

It is interesting to note that, in this expression, the first term involves T and V_{eff} and the second term has contributions specifically from inertia properties of the vapor ligand molecules, such as m, I_A, I_B, and I_C. The $\Delta S_{\max}$ term for the formation of ligand−receptor macromolecule complexes directly refers to the kinetic motion of constituent components under the experimental conditions of T and V_{eff}. This result is in accordance with the available understanding that the entropic cost associated with receptor−ligand interaction is a consequence of the degrees of freedom of motion that are lost when two molecules are rigidly constrained with a weakly bound complex.[22] It seems feasible to find an alternative solution to estimating the value of $\Delta S_{\max}$, in light of Newtonian mechanics.

In fact, the ideal complexation seldom occurs. Components in the structure of the complex always tend to move in relatively independent ways, because of the weak binding constraint between the ligand and the receptor. Although macromolecules seem immobile in their adhesive liquid phase, complexed organic ligand molecules may translate on the film in one- or two-dimensional space and may simultaneously rotate around their individual one or two axes. Evidence has suggested that rotational motions of the ligand in a weakly bound complex have significant importance in recognition processes.[23] Therefore, the hypothesis of $Q_{\mathrm{vib}}(\mathrm{C})$ being a product of $Q_{\mathrm{vib}}(\mathrm{H})$ and $Q_{\mathrm{vib}}(\mathrm{G})$ is ultimately a good approximation to this reality. If we can neglect the internal rotation of the chain segments of the macromolecules, a general formula of ΔS for the calculation of interaction entropy is expressed as

$$\begin{aligned}\Delta S &= \frac{\partial \Delta F_{\text{trans}}}{\partial T} + \frac{\partial \Delta F_{\text{rot}}}{\partial T} \\ &= \frac{\partial}{\partial T}[kT(\ln Q_{\text{trans}}(\text{C}) - \ln Q_{\text{trans}}(\text{G}))] \\ &\qquad + \frac{\partial}{\partial T}[kT(\ln Q_{\text{rot}}(\text{C}) - \ln Q_{\text{rot}}(\text{G}))] \\ &= R\left(\frac{t+r}{2} - 3\right) + k\left[\ln\left(\frac{Q_{\text{trans}}(\text{C})}{Q_{\text{trans}}(\text{G})}\right) + \ln\left(\frac{Q_{\text{rot}}(\text{C})}{Q_{\text{rot}}(\text{G})}\right)\right] \\ &= R\left(\frac{t+r}{2} - 3\right) + R[\ln(q_{\text{trans}}^{t-3}(\text{G})) + \ln(q_{\text{rot}}^{r-3}(\text{G}))] \end{aligned} \tag{6}$$

where, for the purpose of compact representation, the term $(q_{\text{trans}}(\text{G}))^{3N}$ is used to denote $Q_{\text{trans}}(\text{G})$ (but does not mean that they are mathematically identical to each other); $(q_{\text{rot}}(\text{G}))^{3N}$ is used to denote $Q_{\text{rot}}(\text{G})$. The variable t represents the degrees of freedom of motion of the ligand molecules that are translating in the structure of the complex, and r represents the degrees of freedom of motion of the ligand molecules that are rotating in the structure of the complex. The formula presents the structural basis of molecular recognition. In the present article, this structural basis is discussed specifically in terms of "orientation-based fit". The fit refers to electronic and geometrical structures at binding sites. Total orientation-based fit between constituent components is the resultant vector of local orientation-based fits. Molecular recognition starts with local orientation-based fit of the ligand to the receptor and results from the cooperative interaction of site−site pairing. Although the total orientation-based fit remains well-defined, this situation is referenced as the molecular recognition.

On the other hand, in regard to the enthalpy change, we have

$$\Delta H = \Delta U - RT \tag{7}$$

where ΔU represents the inner energy change and is expressed as

$$\begin{aligned}\Delta U &= \Delta U(\text{C}) - \Delta U(\text{H}) - \Delta U(\text{G}) \\ &= \Delta U_{\text{trans}} + \Delta U_{\text{rot}} + \Delta U_{\text{bkg}} \\ &= \left(\frac{t-3}{2}\right)RT + \left(\frac{r-3}{2}\right)RT + (\Delta U_{\text{bkg}}(\text{C}) - \Delta U_{\text{bkg}}(\text{H})) \\ &\approx \left(\frac{t+r}{2} - 3\right)RT + \Delta U(\text{G/H}) \end{aligned} \tag{8}$$

where ΔU(G/H) represents the interaction energy between the ligand and the receptor, which is considered to be approximately equal to the quantity $\Delta U_{\text{bkg}}(\text{C}) - \Delta U_{\text{bkg}}(\text{H})$. This expression indicates that ΔU includes receptor−ligand interactions that are operating during complexation processes. The first term represents the dependent property of non-independent constituent components within a system.

2.3. Orientation-Based Fit. To summarize, entropic costs associated with complexation occur as a consequence of intermolecular interactions. For a series of ligands, if the weak interactions between the ligand and the receptor are attractive enough to maintain the ideal complexation—that is, weakly bound complexes with the constraints of $t = 0$ and $r = 0$—then, we have

$$\Delta S = -3R + R(\ln q_{\text{trans}}^{-3}(\text{G}) + \ln q_{\text{rot}}^{-3}(\text{G})) \tag{9a}$$

$$\Delta U = -3RT + \Delta U(\text{G/H}) \tag{9b}$$

Here, the partition functions, q_{trans}(G) and q_{rot}(G), are invariant at the given temperature T for an individual small molecule; therefore, ΔS achieves its "full load" and cannot positively vary as the weak interaction (represented by ΔU(G/H)) increases. Enthalpies in a series will then not exhibit a simple linear relationship (compensatory relationship) with their respective corresponding entropies, because the stronger intermolecular interactions between the constituent components in a series cannot change the modes of motion of the components in the complex anymore. Thus, it is concluded that the compensatory enthalpy−entropy relationship occurs if and only if a series of complexations of interest remain far enough away from the ideal complexation. The enthalpy−entropy relationship actually reveals a type of mechanical relationship between the strength of the intermolecular interactions and the extent that the constituent components are rigidly constrained together by their interactions. This seems to imply that the linear relationship is probably a special case of the mechanical relationship. Incidentally, for chemical reactions, the formation of new covalence bonds has a great impact on vibration modes for reactants in the formed structure. Bond springs always are coordinated with the chemically binding interaction.

Of course, the acquisition of t and r by performing the calculations of ΔH and ΔS is a straightforward way to determine the nature of the intermolecular interaction, as well as binding sites, in the complex. However, calculations associated with ΔS are difficult, because ΔS refers to the receptor macromolecule and the flexible ligands. Here, we introduce a dimensionless index that is associated with entropy changes (or enthalpy changes), as a probe into molecular recognition, which is defined as

$$\text{index} = -\left(\frac{\Delta S_{\text{obs}} - \Delta S_{\text{calc}}}{\Delta S_{\text{calc}}}\right) \times 100 \tag{10}$$

where ΔS_{obs} is the experimental entropy change and ΔS_{calc} is the calculated entropy change, which is determined using the following procedure.

Suppose that the enthalpy−entropy relationship is mathematically represented by the following formula:

$$\Delta S^\circ = \text{Slope}(\Delta H^\circ) + \text{Intercept} \tag{11}$$

The thermodynamic parameters for the formation of a weakly bound complex are given (ΔH_{obs}, ΔS_{obs}). The calculated entropy change ΔS_{calc}, which is responsible for ΔH_{obs}, is then calculated by

$$\Delta S_{\text{calc}} = \text{Slope}(\Delta H_{\text{obs}}) + \text{Intercept} \tag{12}$$

In the case of the ideal complexation, constituent components remain still, with respect to one another; ΔS_{obs} has the same value as ΔS_{max}, which is more negative than ΔS_{calc}. Thus, the index possesses a negative value, indicating that the highly orientation-based fit occurs upon complexation and that the fit of the entire body of the ligand molecule to the receptor molecule is bad. From the objectivistic point of view, however, it is also possible that ΔS_{max} is probably less negative than ΔS_{calc}. In this situation, the index value is positive. Fortunately, the situation seems to seldom happen and can be theoretically distinguished through comparison of the experimental value with the calculated value of ΔS_{max}.

When intermolecular interactions are so weak that the ligand molecules possess almost the same degrees of freedom of motion, regardless of complexation, the value of ΔS_{obs} is very

close to zero. Then, its index is equal to 100, which means that the complete fit of ligand molecules to receptor molecules operates without resorting to the intermolecular orientation. In other words, the larger the index, the better the fit between the ligand and the receptor, but the worse the orientation-based fit.

Several works reported the phenomenon that the enthalpy and entropy changes that accompany receptor−ligand complexation reveal information about microscopic complexation mechanisms. For 24 enantiomeric pairs in retention on derivatized cyclodextrin gas chromatographic (GC) chiral stationary phases, it was found that the compounds could be arranged in the two groups by the scope of values for enthalpy, entropy, free energy, and the corresponding difference parameters between enantiomers. Compounds belonging to the second group follow an enthalpy−entropy compensation regression, whereas the first group of compounds do not. Researchers believed that there might be at least two different chiral recognition mechanisms with the chiral stationary phases.[24] In a different manner, on the basis of many observations on the receptor−ligand complexation phenomena, Inoue et al. proposed that the slope and intercept of compensation regression could be used as quantitative measures of the conformational change and the extent of desolvation upon complex formation, respectively.[10]

Finally, it should be stressed that the introduction of the dimensionless index is a compromising methodology, based on the knowledge of the enthalpy−entropy relationship, for addressing the extraction of r and t from the ΔH and ΔS expressions. In principle, its application could be extended to a variety of systems involved with thermodynamics. Although values of the index may vary with modifications of the species and the number of species, (ΔH_{obs}, ΔS_{obs}) pairs are irrelevant to the modification. The modifications seem to act as a scale ruler or a distorting mirror, in regard to the index: the orientation-based fit operating with a complex is similarly measured by the index. The study of transient interactions between individual components can benefit from this dependence of the index on the (ΔH_{obs}, ΔS_{obs}) pairs set.[21]

3. Applicability of the Dimensionless Factor

PEG is a type of noncyclic structure in which oxygen atoms and ethylene units are alternately linked together. The flexibility of the PEG chain may result in many complicated configurations in the structure. In the 1960s, it was reported that PEG can selectively receive organic vapor molecules through intermolecular forces.[25] Curled chains of PEG can construct many tiny "water pools", to accommodate hydrophilic substances, as well as form many tiny "oil pools", to solubilize lipophilic substances.[26,27] In addition, noncyclic PEG is supposed to be able to form cationic complexes with alkali and alkaline-earth metal ions in the same manner as crown ethers.[28] Furthermore, both the numerical simulation and experimental observation on model compounds (methoxy-terminated oligo(ethylene oxide), $(EG)_n$ where $n = 1-4$) suggest that the conformational distribution in PEG probably changes significantly when the intermolecular interactions change.[29,30] Surface modification of liposomes with PEG has provided a major advance in drug delivery applications, because of the ability of this polymer to reduce protein binding and plasma elimination of liposomes.[31] The mechanism by which the PEG polymer protects the liposome surface has yet been not clarified.[32−34] We are curious about the influence of chain length on the flexibility.

Small-molecule−PEG systems have been selected as model systems to demonstrate the applicability of the dimensionless index in probing the mechanism of molecular recognition. Gas−liquid chromatography (GLC) is an ideal choice for determining thermodynamic quantities for gas−liquid film systems. The PEG compounds chosen as the stationary liquid films have mean molecular weights of 6000 (PEG-6000) and 1000 (PEG-1000). In addition, small organic molecules are chosen as flowing gas phases in GLC, to examine their individual interactions with PEG. Experimental details are provided in the following Experimental Section.

3.1. Experimental Section. All the reagents used in the present experiments were purchased from Shanghai Chemical Co. (Shanghai, P.R. China). The two types of PEGs have GC purity. The two reagents were used as obtained. Twenty-six types of compounds are under consideration: carbon tetrachloride; chloroform; methylene dichloride; 1,2-dichloroethane; methanol; ethanol; 1-propanol; 2-propanol; 1-butanol; 1-pentanol; acetone; butanone; 2-pentanone; nitromethane; nitroethane; 2-nitropropane; cyclopentanone; acetonitrile; tetrahydrofuran (THF); benzene; toluene; *p*-, *o*-, and *m*-xylene; 1,4-dioxan; and cyclohexane. These reagents were chemically or analytically pure; among them, the chemically pure reagents were further purified by fractional distillation or reduced-pressure distillation. The alcohols, ketones, nitroalkanes, and THF were stored above a 4 Å molecular sieve, to remove the residual water.

PEG was coated on the white diatomaceous supporter (60−80 mesh, specific surface area of 0.3 $m^2\ g^{-1}$), which was silanized by hexamethyldisilazane (the trade name is 102 white silanized supporter; purchased from Shanghai Chemical Co.). The weight ratio of PEG to the supporter was 1: 10. A stainless steel column (the length and inner diameter of which were 1.5 and 3.0 mm, respectively) was then packed with the coated supporter particles. Details about the apparatus and procedure were found in the paper by Sun et al.[39] Before commencing the measurements, the PEG was dried in the column, using a slow flow of dried pure nitrogen (99.999%) at 396.15 K for 60 h. A similar drying procedure was applied to the column that was packed with the 102 white silanized supporter, which will be used to examine the supporter effects on the retention of organic vapors.

For the column that was packed with the noncoated supporter, the average of the retention times for all the ligands under consideration is almost identical to the GC dead time (0.643−0.738 min), over the present experimental temperature interval. The standard deviations are 0.0271 min at 402.2 K, 0.0175 min at 424.6 K, and 0.0191 min at 451.0 K, of which all are less than that of the molecular weights of these ligands. The supporter can be regarded as an inert carrier. Therefore, thermodynamic behavior in dissolution can be completely regarded as a consequence of intermolecular interactions between the individual ligand and the stationary PEG liquid phase.

The enthalpy and entropy changes for dissolution of compounds on PEG liquid films are directly derived from van't Hoff plots. Attention should be paid, however, to the fact that the errors in $\Delta H°$ and $\Delta S°$ values obtained in this manner are correlated with each other. Otherwise, even if no real correlation exists, a correlation between enthalpy and entropy changes will be observed, because of the mutual dependence of the errors in the two quantities. A data analysis method was proposed for overcoming this possible source of an artifactual correlation.[40]

3.2. Thermodynamic Properties. The enthalpy and entropy changes that we obtained by means of the GLC technique are listed in Tables 1 and 2. All the enthalpy and entropy changes possess negative values. The entropy change partially cancels the enthalpic gain obtained from weak interactions between the

TABLE 1: Free Energies ($\Delta G°$), Enthalpies ($\Delta H°$), Entropies ($\Delta S°$), and Indexes for Weak Interactions of Individual Ligands with PEG-6000, Together with the Classification (T_{hm} = 362.85 K, Unless Noted Otherwise)

ligand	$\Delta G°$ (kJ mol^{-1})[f]	$\Delta H°$ (kJ mol^{-1})[f]	$\Delta S°$ (J mol^{-1} K^{-1})	index	class
cyclohexane	−3.95 ± 0.10	−19.2 ± 2.1	−42.1	−2.0	B
benzene	−8.14 ± 0.13	−27.3 ± 2.9	−52.7	3.7	A
toluene	−10.02 ± 0.13	−29.2 ± 2.9	−52.9	8.8	A
o-xylene	−12.49 ± 0.05	−38.9 ± 1.2	−72.8	1.8	A
m-xylene	−11.72 ± 0.07	−38.0 ± 1.8	−72.5	0.3	A
p-xylene	−11.58 ± 0.04	−38.2 ± 1.0	−73.5	−0.5	C
methanol	−6.80 ± 0.14	−30.7 ± 3.0	−66.0	−9.0	B
ethanol	−7.50 ± 0.05	−33.8 ± 1.2	−72.4[b]	−10.3	B
n-propanol	−9.58 ± 0.11	−35.8 ± 2.4	−72.4	−4.8	B
2-propanol	−7.39 ± 0.13	−32.0 ± 2.8	−67.9	−8.3	B
n-butanol	−11.61 ± 0.14	−36.3 ± 3.3	−68.0[a]	2.6	C
n-pentanol	−13.50 ± 0.03	−44.8 ± 1.2	−86.1[c]	−2.6	B
nitromethane	−11.91 ± 0.13	−35.6 ± 3.1	−65.2[a]	5.0	A
nitroethane	−12.03 ± 0.12	−34.6 ± 2.8	−62.3[a]	7.1	A
2-nitropropane	−11.39 ± 0.11	−31.9 ± 2.4	−56.5	9.5	A
1,2-dichloroethane	−10.32 ± 0.11	−31.5 ± 2.4	−58.3	5.6	A
methylene dichloride	−7.54 ± 0.1	−29.1 ± 2.2	−59.4	−2.8	C
chloroform	−9.34 ± 0.11	−31.2 ± 2.4	−60.3	1.7	A
carbon tetrachloride	−6.66 ± 0.04	−30.9 ± 1.0	−66.8	−9.9	C
acetonitrile	−9.29 ± 0.12	−30.1 ± 2.6	−57.3	3.6	A
acetone	−5.49 ± 0.10	−26.4 ± 2.3	−57.7	−8.2	C
2-butanone	−7.25 ± 0.12	−27.8 ± 2.6	−56.6	−1.8	C
2-pentanone	−8.76 ± 0.12	−30.0 ± 2.5	−58.4	1.4	A
cyclopentanone	−12.54 ± 0.05	−38.6 ± 1.2	−71.8[e]	2.6	A
THF	−6.51 ± 0.02	−29.3 ± 0.6	−62.9[b]	−8.1	C
1,4-dioxan	−10.11 ± 0.11	−27.0 ± 3.2	−46.6[b]	14.2	A

[a] The entropy value was derived at T_{hm} = 364.07 K. [b] The entropy value was derived at T_{hm} = 360.01 K. [c] The entropy value was derived at T_{hm} = 361.84 K. [d] The entropy value was derived at T_{hm} = 370.53 K. [e] The entropy value was derived at T_{hm} = 361.74 K. [f] Values given with standard deviations.

TABLE 2: Free Energies ($\Delta G°$), Enthalpies ($\Delta H°$), Entropies ($\Delta S°$), and Indexes for Weak Interactions of Individual Ligands with PEG-1000, Together with the Classification (T_{hm} = 359.52 K)

ligand	$\Delta G°$ (kJ mol^{-1})[a]	$\Delta H°$ (kJ mol^{-1})[a]	$\Delta S°$ (J mol^{-1} K^{-1})	index	class
cyclohexane	1.93 ± 0.08	−22.5 ± 0.7	−68.07	−1.8	B
benzene	−2.37 ± 0.08	−29.8 ± 0.8	−76.35	1.4	A
toluene	−4.18 ± 0.08	−32.7 ± 0.7	−79.43	2.7	A
o-xylene	−6.72 ± 0.04	−36.1 ± 0.5	−81.79	5.5	A
m-xylene	−5.95 ± 0.07	−35.6 ± 0.9	−82.52	3.9	A
p-xylene	−5.88 ± 0.11	−35.3 ± 1.4	−81.86	4.1	A
methanol	−1.87 ± 0.02	−36.6 ± 0.2	−96.71	−10.7	B
ethanol	−2.65 ± 0.08	−36.4 ± 0.7	−93.79	−7.9	B
n-propanol	−4.58 ± 0.07	−38.0 ± 0.6	−92.98	−4.1	B
2-propanol	−2.44 ± 0.12	−35.8 ± 1.1	−92.64	−7.7	B
n-butanol	−6.54 ± 0.09	−40.9 ± 1.1	−95.51	−2.2	B
n-pentanol	−8.34 ± 0.15	−42.0 ± 1.9	−93.78	1.5	C
nitromethane	−6.38 ± 0.08	−37.5 ± 1.1	−86.58	2.3	A
nitroethane	−6.54 ± 0.09	−37.2 ± 1.1	−85.31	3.2	A
2-nitropropane	−5.85 ± 0.07	−35.1 ± 0.9	−81.37	4.4	A
1,2-dichloroethane	−4.51 ± 0.13	−35.3 ± 1.2	−85.74	−0.4	C
methylene dichloride	−1.80 ± 0.11	−31.7 ± 1.0	−83.14	−3.7	C
chloroform	−3.53 ± 0.09	−34.7 ± 0.8	−86.83	−2.6	C
carbon tetrachloride	−1.02 ± 0.08	−28.7 ± 0.8	−76.92	−1.5	C
acetonitrile	−3.87 ± 0.09	−32.9 ± 0.8	−80.66	1.5	A
acetone	−0.26 ± 0.08	−27.9 ± 0.8	−76.87	−3.0	C
butanone	−1.86 ± 0.07	−30.4 ± 0.6	−79.40	−1.4	C
2-pentanone	−3.33 ± 0.09	−32.2 ± 0.8	−80.21	0.8	A
cyclopentanone	−7.30 ± 0.05	−35.0 ± 1.2	−77.02	9.3	A
THF	−1.23 ± 0.07	−27.3 ± 0.7	−72.57	1.7	A
1,4-dioxan	−4.99 ± 0.07	−33.4 ± 0.9	−78.98	4.4	A

[a] Values listed with standard deviations.

ligand and the PEG. Therefore, the interaction thermodynamic behavior of the ligand with the PEG is enthalpically controlled over the present experimental temperature interval.

The regression equations for the enthalpy−entropy relationship are listed in Table 3. The slopes of the regression lines, which represent the so-called "compensation temperature", are 496.7 K for the complex with PEG-6000 and 496.5 K for the complex with PEG-1000. However, the intercept of the ΔH vs ΔS plot is −0.93 kJ mol^{-1} for the complex with PEG-6000 but 7.49 kJ mol^{-1} for the complex with PEG-1000.

TABLE 3: Regression Equations for the Enthalpy−Entropy Relationship

receptor	slope (K)	intercept (kJ mol^{-1})	correlative coefficient
PEG-6000	496.7	−0.93	0.881
PEG-1000	496.5	7.49	0.720

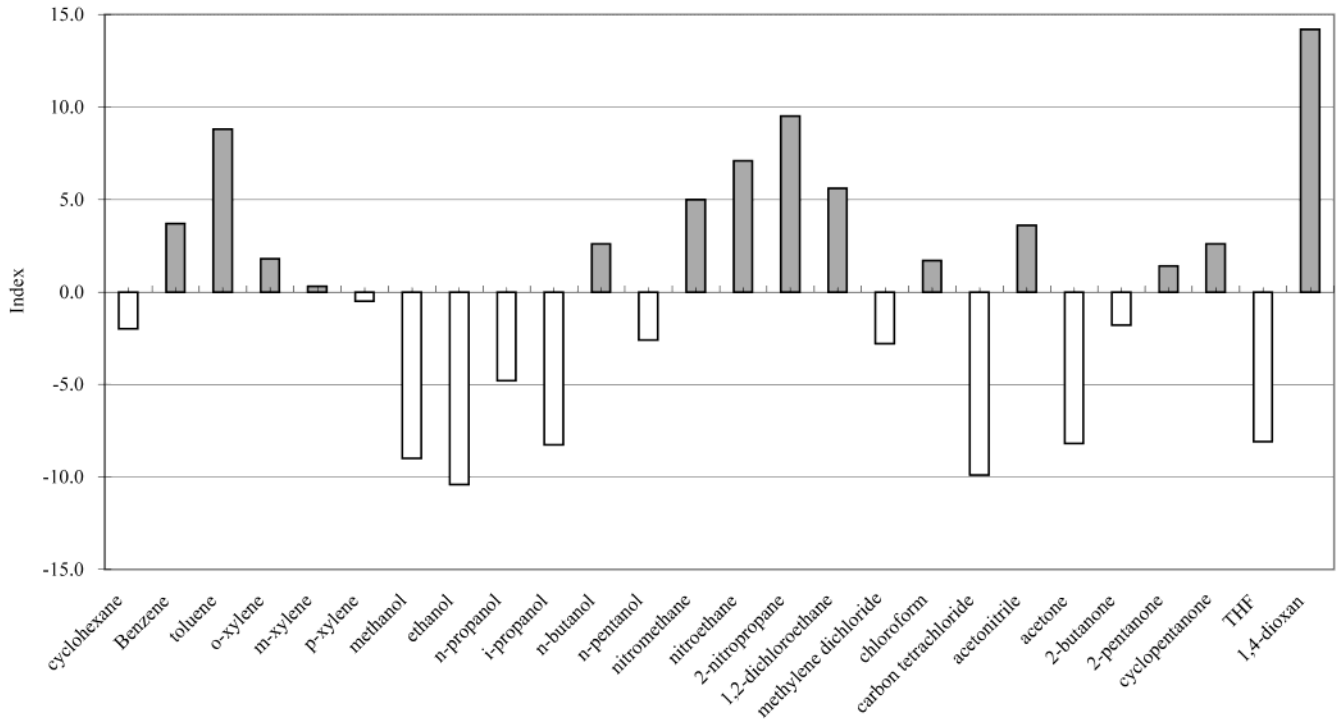

Figure 1. Column plot of index vs name of ligand. Indexes are calculated for orientation-based fits of individual ligands to PEG-6000.

According to the Flory−Huggins model, the total partition function for chain macromolecules is contributed from the distribution of conformations of chain molecules in solution.[35] Therefore, the ΔS value for complexation by flexible PEG will become more negative than that obtained for complexation by the less-flexible PEG. Compared with that of PEG-6000 liquid film, the regression line on PEG-1000 is clearly translated upward or positively along the ΔH axis or downward along the ΔS axis. Clearly, the difference is due to the fact that chain segments of the lighter PEG (PEG-1000) have more-flexible movement kinetically than those of PEG-6000. PEG-1000 is active, with respect to PEG-6000.

3.3. Classification of Small-Molecule−PEG Systems. The calculated indexes for all the various types of ligands are also listed in Tables 1 and 2. All the index values are much less than 100, suggesting that ligand−PEG interactions exhibit, more or less, orientation-based fits. The column plots of index versus ligand name are shown in Figures 1 and 2. IThe column plots show that the indexes clearly discriminate between these small-molecule−PEG systems. The various types of systems can be separated by their indexes into two categories: systems that possess positive indexes and systems that possess negative indexes. Almost the same assignments are available for the two types of receptors.

PEG consists of (−CH_2CH_2O−) motifs. Within the PEG structure, possible binding sites lie at methylene groups and ether oxygen atoms. They are capable of exerting short-range London forces or polar interactions, both of which, in the enthalpy term, are responsible for receptor−ligand complexation.[36,37] The two species must complement each other in size, shape, and binding or functionality.[38] It is interesting to note that almost every type of ligand that possesses a positive index contains an electrophobic group, such as a nitro, cyano, carbonyl, or chloro group and benzene, which are good candidates for the interaction with methylene groups in PEG, whereas almost every type of ligand that has a negative index contains an electrophilic group, such as a hydroxyl group or a hydrogen atom being bonded to a carbon atom, each of which are capable of exerting the polar interaction with ether oxygen atoms in the PEG.

However, several exceptional cases are available. For the orientation-based fits of the alcohol series to PEG-6000, for instance, the homologues, methanol through *n*-propanol and *n*-pentanol, contain a hydroxyl group and have negative indexes, but *n*-butanol possesses a positive index. However, in contrast, for interactions with PEG-1000, the homologues, methanol through *n*-butanol, all have negative indexes, whereas *n*-pentanol possesses a positive index. These exceptional cases probably result from complicated fits of individual ligands to the potential binding sites in PEG; ligands exhibit comparably significant interactions with both ether oxygen and ethylene units in the PEG structure. We use the character "C", which is listed in Tables 1 and 2 as the label to these exceptional cases. The character "A" in the tables denotes the former, in which ligands that contain electrophobic groups possess positive indexes. The character "B" in the tables denotes the latter, in which ligands of electrophilic groups possess negative indexes. The orientation-based fits of the systems being studied are actually expected to be centrally distributed around zero, because the two types of possible binding sites that lie in the PEG structure—the methylene group and the ether oxygen atom—are responsible for local orientation-based fits between the small molecule and the PEG. The behavior of orientation-based fits of the alcohol series to PEG is examined in detail in an upcoming paper.[21]

3.4. Flexibility of PEG Chain. When Table 2 is compared to Table 1, special attention is paid to the following fact: when the receptor is changed from PEG-6000 to PEG-1000, increases (more negative) in the enthalpy change correspond to large increments of entropic costs (more negative), whereas decreases (less negative) in the enthalpy change result in merely a few increments of entropic costs. The following ligands correspond to the latter case: THF; cyclopentanone; *p*-, *m*-, and *o*-xylene; *n*-pentanol; and carbon tetrachloride. A discussion regarding this observation, in terms of the dimensionless index, follows.

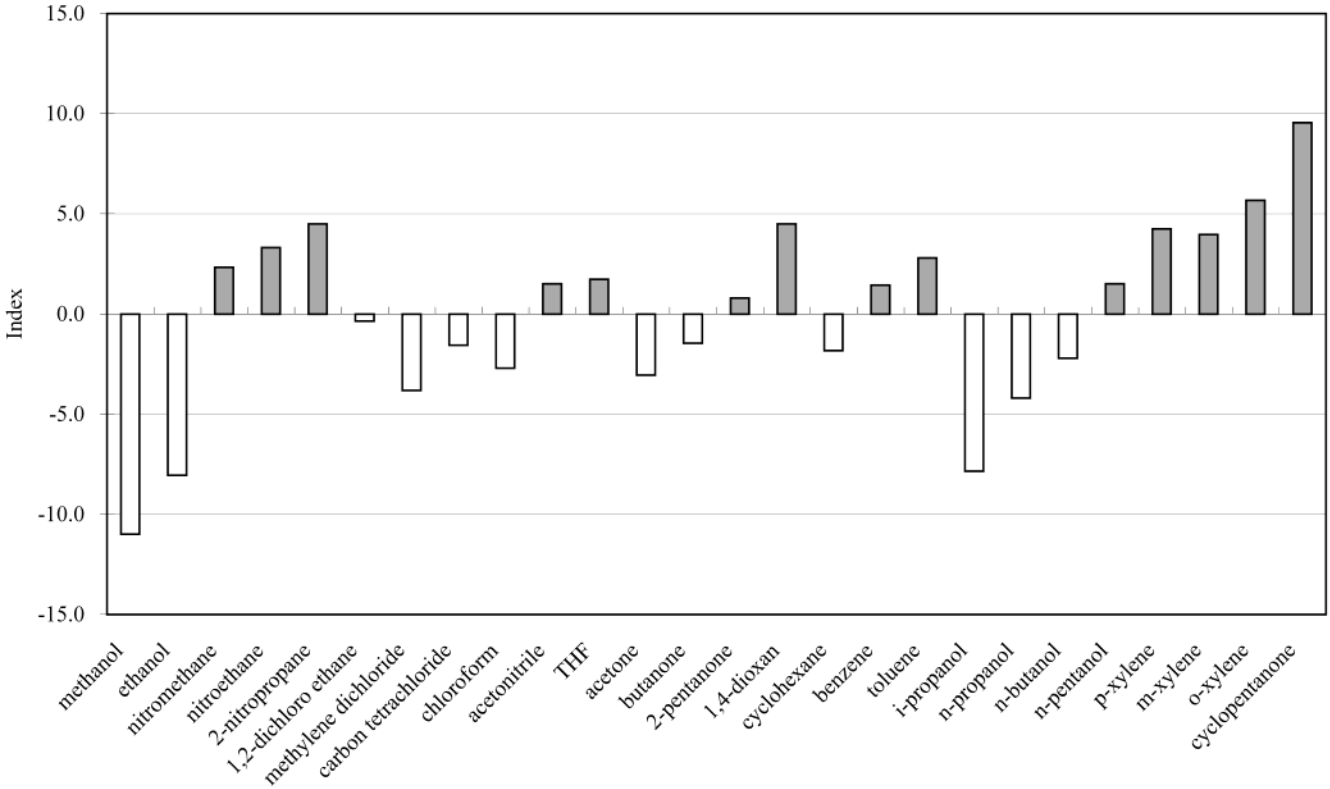

Figure 2. Column plot of index vs name of ligand. Indexes are calculated for orientation-based fits of individual ligands to PEG-1000.

TABLE 4: Orientation-Based Fit Analyses for Weak Interactions of Individual Ligands with PEG as the PEG Weight Shifts to 1000 from 6000[a]

category	class A	class B	class C
I	benzene, toluene, nitromethane, nitroethane, 2-nitropropane, 1,2-dichloroethane (C), chloroform (C), acetonitrile, 2-pentanone, 1,4-dioxan	methanol	methylene dichloride, *n*-butanol (B)
II	cyclopentanone, *o*-xylene, *m*-xylene	cyclohexane, ethanol, *n*-propanol, 2-propanol, *n*-pentanol (C)	acetone, 2-butanone, carbon tetrachloride, THF (A), *p*-xylene (A)

[a] The classes for the ligand system with PEG-6000 are shown. For systems that changed classes when the PEG weight shifted from 1000 to 6000, the class for the ligand system with PEG-1000 is given in parentheses immediately after the ligand name.

First, when the PEG molecular weight decreased from 6000 to 1000, the index values decreased for some systems (category I) but increased for the other systems (category II). In particular, the ligands assigned to the former are distributed almost entirely in class A and a few are distributed in class C, whereas the ligands assigned to the latter are distributed almost entirely in classes B and C and a few are distributed in class A (Table 4). This phenomenon is very interesting to chemists. It means that orientation-based fits between electrophobic sites of the ligand and the methylene group of PEG generally benefit from the flexibility of PEG-1000 (and, thus, the resultant enthalpies and entropies are more negative), whereas orientation-based fits between electrophilic sites of the ligand and the ether oxygen atoms of PEG suffer from this flexibility. Thus, are the resultant enthalpies and entropies less negative? No.

In the resultant distribution of classes of all the systems (ligand−PEG-1000), three category II systems changed their individual classes: the THF−PEG-6000 and *p*-xylene−PEG-6000 systems changed from class C to class A (when associated with PEG-1000), and the *n*-pentanol−PEG-6000 system changed from class B to class C (when associated with PEG-1000) (Table 4). These clearly suggest that, although the chain segments of PEG move more freely, their perturbation to local orientation-based fit (or site-specific interaction) has a more-significant impact on weak interactions. In addition, orientation-based fits occurring between PEG and the ligand of a rigid ring, such as THF and xylene, suggest that more-suitable fits are available in the PEG-6000 structure.

Second, the indexes for the interactions of PEG-6000 with ligands of rigid five-member rings, such as THF and cyclopentanone, have small or negative values (−8.1 and 2.6, respectively) (Table 1), indicating that quite orientation-based molecular fits do exist in the supramolecular assemblies being formed on the PEG liquid film. Therefore, it is reasonable to guess that PEG-6000 in the liquid state contains binding pockets that are responsible for the fit to such ring structures.

However, the formation of these binding pockets in the PEG-6000 structure (i) may be lured by weakly binding interactions between the ligand and PEG (induced fit), (ii) may result from the intertwisting of PEG chains (preorganized conformer), or (iii) may result from both factors. If the pockets form during the complexation procedure, the flexiblity of PEG-1000 should make their formation occur more readily and subsequently will improve orientation-based fits between the ligand and the PEG molecules. In contrast, if the pocket formation results from an intertwisting of chains, regardless of complexation, the flexibility of PEG-1000 will cause the pockets to disband. Consequently, the geometrical and electronic complementarity between the

pocket and the rigid five-member ring will be destroyed, or at least will be lessened.

As we observe in Tables 1 and 2, the index for the orientation-based fit of cyclopentanone varies from 2.6 for PEG-6000 to 9.3 for PEG-1000, and the index for the fit of THF varies from −8.1 for PEG-6000 to 1.7 for PEG-1000. These observations suggest that the orientation-based fits between the pockets and the five-member rings will be markedly lessened in PEG-1000. Therefore, it is concluded that the pockets that match the rigid five-member ring exist prior to complexation in the PEG-6000 structure and complexation by PEG-1000 is not strong enough to restrain the thermal motions (also called thermal perturbation) of the chain segments of PEG-1000 to form the pockets. As seen in Table 2, the thermodynamic parameters for cyclopentanone are similar to those of the noncyclic ketone series; this fact seems to support the conclusion. Therefore, interaction enthalpies of individual ligands associated with PEG-1000 are less negative than those associated with PEG-6000.

The indexes for orientation-based fits of xylene isomers to PEG behave analogously to THF and cyclopentanone; therefore, similar considerations apply to xylene isomers. The flexibility of PEG-1000 cannot allow the formation of binding pockets that fit to the structures of xylene isomers. Consequently, the enthalpy changes that accompany the formation of complexes are less negative. These conclusions are in accordance with the aforementioned observation that the weakly bound complexes involved with ligands such as THF and *p*-xylene are dominantly restrained by the interaction between the electrophobic sites of the ligand and the methylene groups of PEG-1000. However, the chainlike molecule *n*-pentanol behaves in a different way, because of the complicated combination of several local orientation-based fits (multisite interaction).[21]

In conclusion, PEG-1000 is more flexible than PEG-6000 in the liquid state. The flexible chain of PEG-1000 makes its complexation with some ligands stronger, but that with the other ligands is made weaker. The flexibility appears to have conflicting effects on molecular recognition.

4. Conclusions

For the formation of a weakly bound complex, thermodynamic parameters can be regarded as a combined consequence of various molecular local orientation-based fits between constituent components in the supramolecular assembly. By means of a mean field approximation method, a statistical model is proposed for exploring the mechanism of the enthalpy−entropy relationship associated with noncovalent interaction. The model shows that the enthalpy−entropy relationship occurs as a consequence of a series of weak complexations being studied that occur far enough away from the ideal complexation. The enthalpic and entropic changes associated with individual ligand−receptor interactions are correlated with the orientation-based fit via the enthalpy−entropy relationship. The introduction of a dimensionless factor provides a measure of the orientation-based fit between the ligand and the macromolecule. This factor has been successfully applied to understand the structural basis of weak interactions between small organic molecules and poly(ethylene glycol) (PEG). All the small-molecule−PEG systems are assigned to three classes, and these assignments are based on the mechanisms of their respective weak interactions with PEG.

In class A, ligands are detained mainly through nucleophilic interactions between their electrophobic sites and PEG ethylene groups. Class B ligands mainly contain the electrophilic interaction with PEG ether oxygen atoms. Class C ligands are detained through a combination of comparably significant intermolecular interactions associated with both ethylene groups and ether oxygen atoms in the PEG structure.

Electrophilic sites involved in the last two classes, which interact with the ether oxygen atoms of PEG, easily suffer from thermal perturbation, from motions of the chain segments of PEG. It is found that PEG-1000 improves its binding interaction with some organic ligands, whereas its flexibility causes binding pockets to disband, lessening its complexation with the other ligands. The PEG flexibility has conflicting effects on molecular recognition.

Acknowledgment. The present work was partly performed at the Department of Chemistry, Xi Xi Campus, Zhejiang University.

References and Notes

(1) Reinhoudt, D. N.; Crego-Calama, M. *Science* **2002**, *295*, 2403−2407.
(2) Lehn, J.-M. *Angew. Chem., Int. Ed. Engl.* **1988**, *27*, 89.
(3) McCammon, J. A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 245−249.
(4) Muller-Dethlefs, K.; Hobza, P. *Chem. Rev.* **2000**, *100*, 143−167.
(5) Verkhivker, G. M.; Bouzida, D.; Gehlhaar, D. K.; Rejto, P. A.; Freer; S. T.; Rose, P. W. *Curr. Opin. Struct. Biol.* **2002**, *12*, 197−203
(6) Cera, E. D. *Chem. Rev.* **1998**, *98*, 1563.
(7) Berne, B. J.; Straub, J. E. *Curr. Opin. Struct. Biol.* **1997**, *7*, 181.
(8) Jelesarov, I.; Bosshard, H. R. *J. Mol. Recognit.* **1999**, *12*, 3−18.
(9) Lin, F.; Chen, W.; Hearn, M. T. W. *J. Mol. Recognit.* **2002**, *15*, 55−93.
(10) Inoue, Y.; Hakushi, T.; Liu, Y.; Tong, L. H.; Shen, B. J.; Jin, D. S. *J. Am. Chem. Soc.* **1993**, *115*, 475−481 and references therein.
(11) Rekharsky, M. V.; Inoue, Y. *Chem. Rev.* **1998**, *98*, 1875−1917 and references therein.
(12) Connors, K. A. *Chem. Rev.* **1997**, *97*, 1325.
(13) Flory, P. J. *Discuss. Faraday Soc.* **1970**, *49*, 7−29.
(14) Panayiotou, C.; Sanchez, I. C. *Macromolecules* **1991**, *24*, 6231.
(15) Abraham, M. H.; Andonian-Haftran, J.; Du, C. M.; Diart, V.; Whiting, G. S.; Grate, J. W.; McGill, R. A. *J. Chem. Soc., Perkin Trans. 2* **1995**, 369−378.
(16) Donohue, M.; Iconomou, I. G. E. *Fluid Phase Equilib.* **1996**, *116*, 518.
(17) Tapia, O. In *Theoretical Treatment of Large Molecules and Their Interactions*; Maksic, Z. B., Ed.; Springer-Verlag: Berlin, 1991; Vol. 4, p 435.
(18) Angyán, J. G. *J. Math. Chem.* **1992**, *10*, 93.
(19) Sánchez, M. L.; Martín, M. E.; Aguilar, M. A.; Olivares, F. J. *J. Comput. Chem.* **2000**, *21*, 705−715.
(20) Kawai, S.; Arakawa, K. *Bull. Chem. Soc. Jpn.* **1990**, *63*, 307−314.
(21) Chen, J., submitted to *J. Phys. Chem.*
(22) Searle, M. S.; Williams, D. H. *J. Am. Chem. Soc.* **1992**, *114*, 10690.
(23) Kirchhoff, P. D.; Dutasta, J.-P.; Collet, A.; McCammon, J. A. *J. Am. Chem. Soc.* **1999**, *121*, 381−390.
(24) Berthod, A.; Li, W.; Armstrong, D. W. *Anal. Chem.* **1992**, *64*, 873−879.
(25) Rohrschneider, L. *J. Chromatogr.* **1966**, *22*, 6.
(26) Regen, S. L.; Besse, J. J.; McLick, J. *J. Am. Chem. Soc.* **1979**, *101*, 116.
(27) Regen, S. L.; Besse, J. J. *J. Am. Chem. Soc.* **1978**, *100*, 7117.
(28) (a) Kuhl, T. L.; Berman, A. D.; Hui, S. W.; Israelachvili, J. N. *Macromolecules* **1998**, *31*, 8258. (b) Kikuchi, Y.; Sakamoto, Y.; Sawada, K. *J. Chem. Soc., Faraday Trans.* **1998**, *94*, 105.
(29) Yang, X.; Su, Z.; Wu, D.; Hsu, S. L.; Stidham, H. D. *Macromolecules* **1997**, *30*, 3796.
(30) Wang, R. L. C.; Kreuzer, H. J.; Grunze, M. *Phys. Chem. Chem. Phys.* **2000**, *2* (16), 3613−3622.
(31) Chirila, T. V.; Rakoczy, P. E.; Garrett, K. L.; Lou, X.; Constable, I. J. *Biomaterials* **2002**, *23*, 321−342.
(32) Tirosh, O.; Barenholz, Y.; Katzhendler, J.; Priev, A. *Biophys. J.* **1998**, *74*, 1371−1379.
(33) Torchilin, V. P.; Papisov, M. I. *J. Lipid Res.* **1994**, *4*, 725−739.
(34) Chiu, G. N. C.; Bally, M. B.; Mayer, L. D. *Biochim. Biophys. Acta* **2001**, *1510*, 56−69.
(35) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Wiley−Interscience: New York, 1969.
(36) Lipkowitz, K.; Green, K. M.; Yang, J.; Pearl, G.; Peterson, M. A. *Chirality* **1993**, *5*, 57.

(37) (a) Armstrong, D. W.; Ward, T. J.; Armstrong, R. D.; Beesly, T. E. *Science* **1986**, *232*, 1132. (b) Colonna, S.; Manifredi, A.; Annunziata, R.; Gaggero, N.; Casella, L. *J. Org. Chem.* **1990**, *55*, 5862.

(38) (a) Lehn, J. M. *Science* **1985**, *227*, 849. (b) Rebek, J. *Science* **1987**, *235*, 1478−1484.

(39) Sun, D.-Z.; Chen, J.; Lu, W.-M.; Zheng, X.-M. *Thermochim. Acta* **2000**, *351* (1−2), 1−7.

(40) (a) Krug, R. R.; Hunter, W. G.; Grieger, R. A. *J. Phys. Chem.* **1976**, *80*, 2335; **1976**, *80*, 2341. (b) Grunwald, E.; Steel, C. *J. Am. Chem. Soc.* **1995**, *117*, 5687.