J|A|C|S
ARTICLES
Published on Web 09/18/2004

Free Energy, Entropy, and Induced Fit in Host−Guest Recognition: Calculations with the Second-Generation Mining Minima Algorithm

Chia-En Chang and Michael K. Gilson*

Contribution from the Center for Advanced Research in Biotechnology, 9600 Gudelsky Drive, Rockville, Maryland 20850

Received May 17, 2004; E-mail: gilson@umbi.umd.edu

Abstract: This study applies a novel computational method to study molecular recognition for three sets of synthetic hosts: molecular clips, molecular tweezers, and a synthetic barbiturate receptor. The computed standard free energies of binding for the 12 binding reactions agree closely with experiment and provide insight into the roles of configurational entropy, preorganization, and induced fit in the systems studied. The computed changes in configurational entropy are comparable in magnitude to the changes in mean potential plus solvation energy, and they result primarily from changes in the average width of the energy wells upon binding. A strong correlation is observed between the changes in configurational energy and configurational entropy upon binding, resulting in near-linear compensation analogous to classical entropy−enthalpy compensation.

1. Introduction

Molecular recognition of guest molecules by synthetic hosts is of great practical interest in applications such as chemical detection, separation and encapsulation, and enantioselective synthesis. These systems also represent elegant, simplified models of biological molecular recognition, because the same physical principles are operative and they display the subtle and often unpredictable structure-affinity relationships that make a detailed understanding of biomolecular interactions so difficult to achieve. However, despite significant advances in molecular modeling techniques and the comparative simplicity of host−guest systems, there is still a need for a tractable and theoretically sound computational method to interpret experimental data and ultimately to help with the design of new hosts for targeted molecular guests.

Like many macromolecular systems (see ref 1 and references therein), host−guest systems exhibit marked entropy−enthalpy compensation (ref 2 and references therein). That is, a chemical change that leads to a more negative enthalpy of association typically, though not always,[3] is accompanied by a loss in entropy that moderates the enhancement of affinity due to the enthalpy change. Such compensation has been observed for noncovalent associations in both water and organic solvents.[2] It seems intuitively reasonable that strengthening attractive forces and hence making enthalpy more negative should reduce conformational freedom and thus incur a greater entropy penalty. On the other hand, an illusion of enthalpy−entropy compensation can result from thermodynamic measurements in which $-T\Delta S^\circ$ is computed as $\Delta G^\circ - \Delta H^\circ$, if the measurements of ΔH° are imprecise.[4] Furthermore, it has been suggested that enthalpy−entropy compensation within a given series of host−guest complexes might be an uninteresting mathematical consequence of the relative constancy of ΔG°.[2] Thus, there are significant questions regarding the generality and physical significance of enthalpy−entropy compensation in host−guest systems, and more generally.

The present study applies a second-generation form of the Mining Minima[5,6] algorithm, termed M2, to analyze the binding reactions of 12 host−guest complexes in an organic solvent. The molecular clips[7] (**1**, **2**, **3**) and tweezers[8,9] (**5**, **6**) are rather rigid, aromatic hosts, some of which are additionally outfitted with small, rotatable polar moieties at the edge of the binding cavity. These compounds exploit primarily aryl−aryl interactions to bind aromatic guests with standard free energies of −1.6 to −4.5 kcal/mol. In contrast, the barbiturate receptor[10] (**10**) is a flexible, partly heteroaromatic macrocycle that was designed to bind its guests primarily through formation of up to six hydrogen bonds and that achieves binding free energies as strong as −8.3 kcal/mol. Preorganization and configurational entropy are likely to be particularly important for this flexible receptor, so it is essential to use a computational method that can address these issues. The present method uses well-developed theory

(1) Sharp, K. *Protein Sci.* **2001**, *10*, 661−667.
(2) Houk, K. N.; Andrew, G. L.; Kim, S. P.; Zhang, X. *Angew. Chem., Int. Ed.* **2003**, *42*, 4872−4897.
(3) Gallicchio, E.; Kubo, N. M.; Levy, R. M. *J. Am. Chem. Soc.* **1998**, *120*, 4526−4527.
(4) Lumry, R.; Rajender, S. *Biopolymers* **1970**, *9*, 1125−1227.
(5) Chang, C.-E.; Potter, M. J.; Gilson, M. K. *J. Phys. Chem. B* **2003**, *107*, 1048−1055.
(6) Head, M. S.; Given, J. A.; Gilson, M. K. *J. Phys. Chem.* **1997**, *101*, 1609−1618.
(7) Klarner, F.-G.; Panizky, J.; Blaser, D.; Boese, R. *Tetrahedron* **2001**, *57*, 3673−3687.
(8) Ruloff, R.; Seelbach, U. P.; Merbach, A. E.; Klarner, F.-G. *J. Phys. Org. Chem.* **2002**, *15*, 189−196.
(9) Klarner, F.-G.; Lobert, M.; Naatz, U.; Bandmann, H.; Boese, R. *Chem.−Eur. J.* **2003**, *9*, 2−13.
(10) Chang, S.; Hamilton, A. *J. Am. Chem. Soc.* **1988**, *110*, 1318−1319.

10.1021/ja047115d CCC: $27.50

and validated numerics to compute the standard free energy of binding along with information on conformational preferences in the free and bound states and to compute changes in configurational entropy upon binding. We are thus able to examine not only the detailed pattern of interactions in each host−guest system, but also the broader picture of tradeoffs between attractive forces and configurational entropy losses across all 12 systems.

2. Theory

The binding constant K_b of a receptor R and ligand L which form the noncovalent complex RL is given by:[11,12]

$$-RT \ln K_b = \Delta G^\circ \equiv \mu^\circ_{RL} - \mu^\circ_R - \mu^\circ_L \tag{1}$$

where μ°_X, the standard chemical potential of species X = R, L, RL, may be written in terms of the configuration integral Z_X:[11]

$$\mu^\circ_X = -RT \ln\left(\frac{8\pi^2}{\sigma_X C^\circ} Z_X\right) \tag{2}$$

$$Z_X = \int e^{-(U_X(\mathbf{r}_X)+W_X(\mathbf{r}_X))/RT}\, d\boldsymbol{r}_X \tag{3}$$

where R is the gas constant, T is the absolute temperature, C° is the standard concentration (typically 1 M), σ_X is the symmetry number, and $U_X(\boldsymbol{r}_X)$ and $W_X(\boldsymbol{r}_X)$ are, respectively, the vacuum potential energy and the solvation energy of X as a function of its coordinates $\boldsymbol{r}_X$. The factors of $8\pi^2$ and $(C^\circ)^{-1}$ relate to the rotational and translational freedom of the molecule.[11] The bound complex includes all configurations of the ligand in the binding site such that it blocks any other ligand from occupying a position of lower energy in the binding site.[12]

3. Methods

3.1. Method of Predominant States. Numerical evaluation of the configuration integral Z can be computationally daunting. However, substantial acceleration can be achieved by using a fast implicit solvent model (see, for example, refs 13−18) to evaluate W without explicitly integrating over the coordinates of solvent molecules. We have shown that, when the solvent is treated in this way, Z can be accurately approximated for many molecular systems as a sum of contributions from the system's predominant low-energy conformations.[5,6] Thus, given M local energy minima, the configuration integral Z is approximated as a sum of M local integrals z_j:

$$Z \approx \sum_{j=1}^{M} z_j$$

$$z_j \equiv \int_j e^{-(U(\boldsymbol{r})+W(\boldsymbol{r}))/RT}\, d\boldsymbol{r} \tag{4}$$

where $\int_j$ implies an integral whose domain is restricted to energy well j. For the complex, we include all energy wells for which the ligand is within the binding cleft of the host. The next two subsections describe the methods used to identify local energy minima $j = 1...\ M$ and to evaluate the local configuration integral z_j in each well.

3.2. Conformational Search. The Tork algorithm[19] is used to identify the most stable conformations $j = 1...\ M$ of the free molecules and their complex. A starting conformation is identified and energy-minimized. For the host−guest complexes, the starting conformation for Tork is generated by using the docking program Vdock[20,21] to fit the ligand into an open conformation of the receptor. Normal modes are computed to identify natural molecular motions that tend to correspond to "mountain passes" between energy minima, as in the low-mode search algorithm.[22−24] The molecule is then distorted along the torsional components of these modes and pairwise combinations thereof, and the distorted conformations are energy-minimized to yield new low-energy conformations. Tork gains efficiency by operating in a molecular coordinate system consisting of bond lengths, bond angles, and bond torsions, termed BAT coordinates,[5,25−28] rather than in Cartesian coordinates.[19] The most stable conformations found by Tork are used to start successive generations of Tork searching, where stability is assessed on the basis of z_j, whose calculation is detailed below. This procedure is iterated until no new conformations are discovered in a generation for the simpler systems or, for the more complex barbiturate receptor, until including the new conformations in two successive generations causes the free energy of the system to change by $<10^{-4}$ kcal/mol. (The Supporting Information includes graphs of the cumulative free energy as a function of the number of energy wells for two of the most complex systems studied here, the free barbiturate receptor and the complex of this receptor with mephobarbital.)

Each conformation generated by Tork is compared with its predecessors and eliminated if it is a repeat, where dihedral angles associated with freely rotating bonds are considered distinct if different by greater than 60° and dihedrals within rings are considered distinct if different by greater than 15°. To speed the calculations and avoid double counting of low-energy conformations, a recently developed algorithm[29] is used to detect topological symmetries[29,30] of the molecule and eliminate conformations that are the same after rotational interchange of chemically equivalent atoms. For example, two conformations that differ by a 180° flip of a phenyl group are considered identical, as are all the chair conformations of cyclohexane that can be generated by circular permutation of the dihedral angles.

3.3. Local Configuration Integrals. In BAT coordinates, the configuration integral in an energy well j can be written as:[5,25,31]

$$z_j = \int_j b_2^{\,2} \prod_{i=3}^{N_{atom}} (b_i^{\,2} \sin\theta_i)\, e^{E(\vec{b},\vec{\theta},\vec{\phi})/RT}\, d\vec{b}\, d\vec{\theta}\, d\vec{\phi} \tag{5}$$

where $\int_j$ indicates an integral in energy well j, $E \equiv U + W$, and b_i, θ_i, and ϕ_i are, respectively, the bond length, bond angle, and bond torsion associated with atom i.[5,31] The prefactor of the exponential, which is the Jacobian determinant for BAT coordinates, can be moved outside the integrand because bond lengths and angles do not vary much.

(11) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. *Biophys. J.* **1997**, *72*, 1047−1069.
(12) Mihailescu, M.; Gilson, M. K. *Biophys. J.* **2004**, *87*, 23−36.
(13) Gilson, M. K.; Honig, B. *J. Comput.-Aided. Mol. Des.* **1991**, *5*, 5−20.
(14) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(15) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem.* **1997**, *101*, 3005−3014.
(16) Onufriev, A.; Case, D. A.; Bashford, D. *J. Comput. Chem.* **2002**, *23*, 1297−1304.
(17) Lee, M. S.; Salsbury, F. R., Jr.; Brooks, C. L., III. *J. Chem. Phys.* **2002**, *116*, 10606−10614.
(18) Zhang, L. Y.; Gallicchio, E.; Friesner, R. A.; Levy, R. M. *J. Comput. Chem.* **2001**, *22*, 591−607.
(19) Chang, C.-E.; Gilson, M. K. *J. Comput. Chem.* **2003**, *24*, 1987−1998.
(20) David, L.; Luo, R.; Gilson, M. K. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 157−171.
(21) Kairys, V.; Gilson, M. K. *J. Comput. Chem.* **2002**, *23*, 1656−1670.
(22) Kolossvary, I.; Guida, W. *J. Am. Chem. Soc.* **1996**, *118*, 5011−5019.
(23) Kolossvary, I.; Guida, W. *J. Comput. Chem.* **1999**, *20*, 1671−1684.
(24) Keserû, G. M.; Kolossvary, I. *J. Am. Chem. Soc.* **2001**, *123*, 12708−12709.
(25) Pitzer, K. S. *J. Chem. Phys.* **1946**, *14*, 239−243.
(26) Herschbach, D. R.; Johnston, H. S.; Rapp, D. *J. Chem. Phys.* **1959**, *31*, 1652−1661.
(27) Go, N.; Scheraga, H. A. *Macromolecules* **1976**, *9*, 535−542.
(28) Abagyan, R.; Totrov, M.; Kuznetsov, D. *J. Comput. Chem.* **1994**, *15*, 488−506.
(29) Chen, W.; Huang, J.; Gilson, M. K. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1301−1313.
(30) Ivanov, J.; Schûûrmann, G. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 728−737.
(31) Potter, M. J.; Gilson, M. K. *J. Phys. Chem. A* **2002**, *126*, 563−566.

In the standard harmonic approximation, E is approximated by a multidimensional parabola whose shape is determined by the second derivative matrix of E with respect to the conformational coordinates, and the integrand $e^{-E/RT}$ thus is approximated as a multidimensional Gaussian that is integrated as all the conformational variables range over $[-\infty,\infty]$. This integral is always finite because the harmonic approximation to the energy rises without bound away from the energy minimum. However, the use of infinite integration ranges can lead to partial double-counting of neighboring conformations.[5,32] Also, the harmonic approximation is subject to error because the actual energy E can deviate substantially from a parabola.

These problems are addressed by a new method[5] that involves diagonalizing the second derivative matrix of E and carrying out a separate integral along each eigenvector. Double-counting of neighboring energy wells is avoided by limiting the integration range along each eigenvector to the lesser of three standard deviations (3σ) of the associated 1-dimensional Gaussian and $\pm 60°$ in any torsional angle. The imposition of such integration limits follows the earlier work of Kolossvary.[32] The present method differs from prior work in correcting for anharmonicity by mode scanning,[5] in which the molecule is distorted stepwise along each eigenvector i that has a low force constant, and computing the energy of each distorted conformation. The resulting energy scan is then used to compute a one-dimensional numerical integral S_i by the trapezoidal rule. If this integral deviates significantly from the harmonic approximation along the same mode, then the numerical integral is substituted for the harmonic approximation. The resulting configuration integral is computed partly harmonically and partly via one-dimensional numerical integrals, and the computational method is therefore called harmonic approximation/mode-scanning (HA/MS). The configuration integral for energy well j becomes:

$$z_j \approx {b_2}^2 b_3 \sin\theta_3 \prod_i^{N_{\text{scan}}} S_i \prod_k^{N_{\text{harm}}} \left[\sqrt{\frac{2\pi RT}{K_k}} \operatorname{erf}\left(\frac{w_k}{\sqrt{\frac{2RT}{K_k}}} \right) \right] \tag{6}$$

where the index i ranges over the N_{scan} numerically integrated modes, k ranges over the N_{harm} modes treated as harmonic, K_k is the eigenvalue of mode k, w_k is the integration range of mode k, and erf is the error function. Modes with harmonic force constants <10 kcal/mol/Å^2 are scanned for anharmonicity, and the numerical integral is substituted for the harmonic approximation when the numerical integral deviates from the harmonic >1 kcal/mol. The accuracy of this method has been established by comparisons with analytic and semianalytic results for tractable model systems.[5] Interestingly, the method is accurate when BAT coordinates are used, as in the present study, but not when Cartesian coordinates are used, for reasons discussed previously.[5]

3.4. Calculation of Configurational Entropy and Mean Energy. Binding free energies are rigorously separable[11] into the change in average energy $\Delta\langle E\rangle \equiv \Delta\langle U + W\rangle$ and the change in solute configurational entropy $\Delta S^\circ_{\text{config}}$ via the following expressions:

$$\langle E\rangle = \frac{\sum_i^M \int_i E(\boldsymbol{r})\, e^{-E(\boldsymbol{r})/RT}\, d\mathbf{r}}{\sum_i^M z_i} \tag{7}$$

$$-TS^\circ_{\text{config}} = RT \int p(\boldsymbol{r}) \ln p(\boldsymbol{r})\, d\boldsymbol{r} \tag{8}$$

where z_i is defined in eq 5, $p(\boldsymbol{r})$ is the probability density $e^{-E(\boldsymbol{r})/RT}/\int_i e^{-E(\boldsymbol{r})/RT}\, dr$, M is the number of energy minima, and E is obtained by the harmonic approximation with the integration ranges described above. There is a direct analogy between $\Delta G° = \Delta H° - T\Delta S°$ and $\Delta G° = \Delta\langle U + W\rangle - T\Delta S^\circ_{\text{config}}$ where G will be calculated by eq 12; in effect, the present separation puts the problem into the form of a gas-phase theory in which the potential of mean force $\langle U + W\rangle$ plays the role of potential energy and the configurational entropy of the solutes S°_{config} plays the role of the total entropy.[11,33,34] It should be emphasized that S°_{config} is not the *total* entropy because it omits solvent contributions implicit in W, so the values of S°_{config} reported here should not be compared directly with experimental entropies of binding.[35] However, S°_{config} does correctly account for the change in entropy upon binding due to changes in the configurational freedom of the host and guest themselves.

As previously noted,[11] any separation of the change in configurational entropy upon binding into rotational, translational, and internal (or vibrational) components must use an arbitrary separation of internal solute coordinates from external coordinates, and thus such entropy separations are themselves arbitrary, rather than fundamental, in nature. Therefore, we do not provide separate results for rotational and translational entropy.

3.5. Energy Model. The internal potential energy $U(\boldsymbol{r})$, which includes bonded, Coulombic, and van der Waals components, is computed with the CHARMM force field,[36] with parameters assigned by the program Quanta[37] based upon the CHARMM 22 parameter set.[36] The solvation energy $W(\boldsymbol{r})$, which is the work of transferring a molecule in conformation $\boldsymbol{r}$ from vacuum to solvent, is considered to occur in two steps:[38] formation of a nonpolar cavity in the solvent that has the shape of the solute and has van der Waals interactions with the solvent, followed by transfer of the partial charges of the molecule into the prepared cavity. The overall solvation energy is the sum of the work of these two steps:

$$W = W_{\text{elec}} + W_{\text{np}} \tag{9}$$

The electrostatic component, W_{elec}, is always negative because electrostatic energy is always reduced by moving charges into a region of higher dielectric constant, and every solvent has a higher dielectric constant than vacuum. The nonpolar component, W_{np}, is negative for chloroform, reflecting the fact that chloroform is a good solvent for nonpolar compounds.

During conformational searching and the evaluation of configuration integrals, W_{elec} is computed with a simplified but fast generalized Born model.[15] The electrostatic solvation energy of each energy well is then corrected toward a more accurate but time-consuming finite-difference solution of the Poisson equation (see Computational Details, section 3.6). The dielectric cavity radius of each atom is set to the mean of the solvent probe radius (2.4 Å for chloroform)[39] and the atom's van der Waals radius, and the dielectric boundary between the molecule and the solvent is the solvent-accessible molecular surface.[40]

The nonpolar term ΔW_{np} is computed as a linear function of the molecular surface area A:[38]

$$W_{\text{np}} = aA + b \tag{10}$$

where a and b are adjustable parameters. Their values were established via a least squares linear fit of $W_{\text{expt},i} - W_{\text{elec},i}$ versus A_i, where $W_{\text{expt},i}$ is the experimental solvation energy of molecule i in chloroform, and $W_{\text{elec},i}$ is computed by finite difference solution of the Poisson equation. Three series of molecules were used for the fitting:[39] aliphatic alcohols,

(32) Kolossvary, I. *J. Phys. Chem. A* **1997**, *101*, 9900−9905.
(33) McMillan, W. G.; Mayer, J. E. *J. Chem. Phys.* **1945**, *13*, 276−305.
(34) Hill, T. L. *An Introduction to Statistical Thermodynamics*; Dover: New York, 1986.
(35) It would be possible to fully separate entropy from enthalpy in these calculations if the implicit solvation model separated solvation enthalpy and entropy.
(36) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(37) *Quanta*; Accelrys, Inc.: San Diego, CA.
(38) Honig, B.; Sharp, K.; Yang, A.-S. *J. Phys. Chem.* **1993**, *97*, 1101−1109.
(39) Luo, R.; Head, M. S.; Given, J. A.; Gilson, M. K. *Biophys. Chem.* **1999**, *78*, 183−193.
(40) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151−176.

amines, and esters. The resulting values of a and b are -0.04 kcal/(mol Å^2) and 1.41 kcal/mol, respectively.

3.6. Computational Details. For each molecule, an initial conformation generated with the program Quanta[37] is energy-minimized by the conjugate gradient method and then the Newton−Raphson method until the energy gradient is $<10^{-3}$ kcal/mol/Å. The parameters for Tork conformational search are as described previously.[19] During the Tork and HA/MS calculations, the solvation energy is estimated by the generalized Born (GB) model because more accurate finite difference solutions of the Poisson equation are relatively slow, as is the calculation of the molecular surface area required for W_{np}. Once an energy well has been found and its free energy evaluated ($G_i \equiv -RT \ln z_i$), its free energy is adjusted by replacing the GB energy by the more accurate finite-difference Poisson result[41] computed with UHBD[42] and adding in the nonpolar component of the solvation energy, to yield a corrected free-energy $G_{corr,i}$:

$$G_{corr,i} = G_i - W_{GB,i} + W_{Poisson,i} + W_{np,i} \quad (11)$$

The overall free-energy G of the molecular system then is

$$G \approx -RT \ln \sum_{i=1}^{M} (e^{-G_{corr,i}/RT}) \quad (12)$$

The solvation calculations use a chloroform dielectric constant of 4.8 and a chloroform radius of 2.4 Å. All calculations were carried out on Linux computers with Athlon 1.67 GHz processors. A few CPU hours are required to complete processing of the simpler systems studied here, and about a day is required for the more complex barbiturate receptor systems.

4. Results

4.1. Molecular Clips and Tweezers. Klärner and co-workers have developed and characterized a series of relatively rigid benzene- and naphthalene-spaced receptors to study the interplay of steric constraints and aryl−aryl ($\pi-\pi$) interactions. This section analyzes the association reactions of these molecular "clips"[7] and "tweezers"[8,9] with aromatic guests in chloroform.

4.1.1. Molecular Clips. Table 1 compares the computed ($\Delta G°$) and experimental ($\Delta G°_{expt}$) affinities of three clips **1**, **2**, **3** for the ligand 1,2,4,5-tetracyanobenzene (TCNB), **4** (Figure 1). The calculated results are within 1.5 kcal/mol of experiment, and they provide a correct ranking of the affinities.

As shown in Figure 2, all of the complexes involve stabilizing aryl−aryl interactions. Clip 1 in addition forms O−H···N hydrogen bonds (Figure 2A) with the ligand, helping to account for this host's relatively high affinity. For the free receptor, these hydroxyls are not perfectly positioned for binding, but adopting the binding conformation only requires one hydroxyl to rotate 60°. A brief additional calculation indicates that this rotation costs less than 1 kcal/mol, so clip 1 is relatively well preorganized to form hydrogen bonds with the TCNB guest, **4**.

Clip 3 also has flexible side chains, but these OAc groups do not form stabilizing interactions with the ligand. Instead, they tend to oppose binding. Thus, in the most stable conformation of free clip 3, its OAc groups partly block the binding site (Figure 2C), and an additional calculation indicates that rotating these groups to the most stable form of the bound state (Figure 2C) costs 2 kcal/mol. This lack of preorganization helps to account for the surprising experimental observation that clip 2 binds the ligand more strongly than does clip 3, even though the binding cleft of clip 2 has two sides rather than three, and therefore forms less extensive contacts with the bound ligand.

Table 1. Experimental ($\Delta G°_{expt}$) and Computed ($\Delta G°$) Standard Free Energies (kcal/mol) of Binding[a]

		computed			
	$\Delta G°_{expt}$	$\Delta G°$	$\Delta\langle U+W\rangle$	$-T\Delta S°_{config}$	$\Delta G°'$
Clips					
TCNB−clip 1[b]	−4.5	−4.0	−12.8	8.8	−6.9
TCNB−clip 2[b]	−3.4	−3.2	−9.6	6.4	−4.9
TCNB−clip 3[b]	−2.9	−1.4	−8.7	7.3	−2.9
Tweezers					
guest 7/tweezers 5[b]	−1.8	−1.9	−10.2	8.3	−5.2
guest 8/tweezers 5[b]	−2.1	−2.6	−12.9	10.3	−7.3
guest 9/tweezers 5[b]	−2.8	−3.4	−12.2	8.8	−7.1
guest 7/tweezers 6[c]	−1.6	−1.1	−9.8	8.7	−6.0
guest 8/tweezers 6[d]	−1.7	−2.2	−13.5	11.4	−10.6
guest 9/tweezers 6[e]	−2.8	−3.7	−13.5	9.8	−9.0
Macrocyclic Barbiturate Receptor					
barbital 10/11[c]	−8.3	−9.1	−26.6	17.5	−20.5
phenobarbital 10/12[c]	−7.3	−7.5	−28.1	20.6	−14.2
mephobarbital 10/13[c]	−3.8	−3.3	−18.1	14.8	−13.1

[a] $\Delta\langle U + W\rangle$: change in mean potential plus solvation energy upon binding. $-T\Delta S°_{config}$: contribution of change in configurational entropy to computed free energy of binding. $\Delta G°'$: Binding free energies computed based on the corrected energy at the base of each energy well but without the HA/MS integration; see Section 4.3. Calculations and experiments were carried out at the following temperature (K): [b]294, [c]298, [d]293, and [e]283.

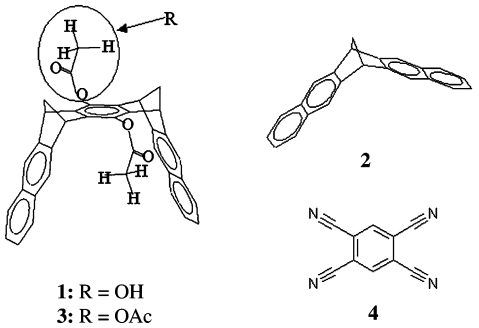


Figure 1. **1**: clip 1. **2**: clip 2. **3**: clip 3. **4**: ligand TCNB.

4.1.2. Molecular Tweezers. Table 1 compares the measured and computed affinities of two molecular tweezers, compounds **5** and **6** in Figure 3, with three guest ligands, compounds **7**, **8**, **9** in Figure 3. The calculated results all are within 1.1 kcal/mol of experiment, and they provide a correct ranking of the affinities. Overall, the tweezers tend to interact more strongly with their ligands than do the clips, in the sense that the values of $\Delta\langle U + W\rangle$ are lower. On the other hand, the tweezers tend to pay a higher cost in configurational entropy ($-T\Delta S°_{config}$). The probable explanation for both of these differences is that the tweezers encase their ligands more completely than the clips, and therefore they form more attractive interactions with the ligand but also immobilize the ligand more completely.

Like the clips, the tweezers form stabilizing aryl−aryl interactions with their guest ligands, as illustrated in Figure 4. Interestingly, however, the rotatable OAc groups of tweezers 6 are not found to oppose binding, based upon a comparison with tweezers 5, which lacks the OAc moieties. This contrasts with the OAc groups of clip 3, which tend to block the binding site, as discussed in the previous section. The explanation is again related to preorganization: inspection of the global energy minimum of the free receptor shows that the OAc groups tend

(41) Gilson, M. K.; Honig, B. *Proteins: Struct., Funct., Genet.* **1988**, *4*, 7−18.
(42) Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. *Comput. Phys. Commun.* **1991**, *62*, 187−197.

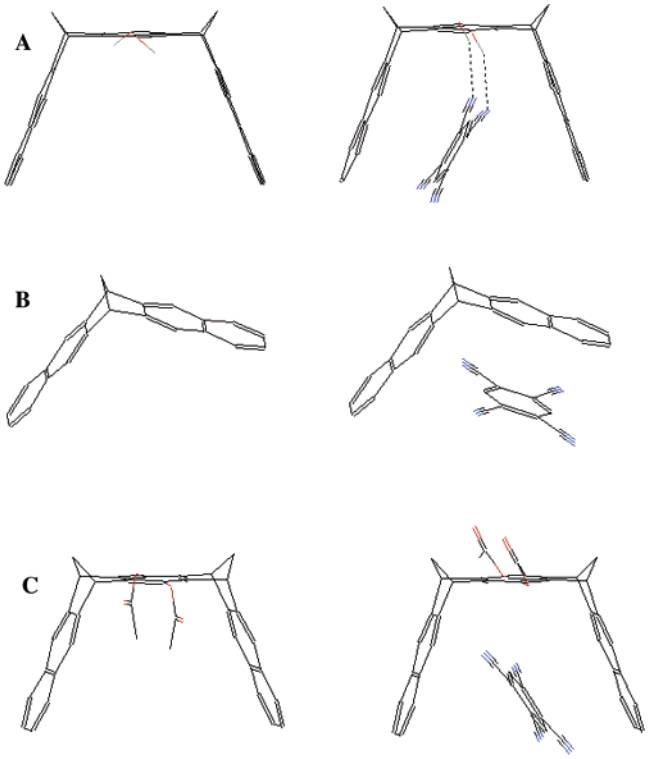
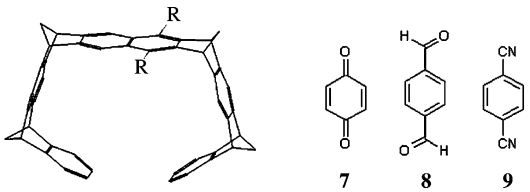


Figure 2. Computed global energy minima of free receptors (left) and complexes with TCNB (**4**) guest. (A) Clip 1. (B) Clip 2. (C) Clip 3.



5: R = H
6: R = OAc

7 **8** **9**

Figure 3. **5**: Tweezers 5. **6**: Tweezers 6. **7**: Guest 7. **8**: Guest 8. **9**: Guest 9.

to be already rotated out of the binding site, leaving the site ready to accept the ligand (Figure 4). This conformational preference is accounted for by a significant rise in molecular surface area (∼100 Å^2) when the OAc groups are rotated out of the binding site, whereas the rise in surface area for the corresponding motion in clip 3 is smaller (∼30 Å^2). In addition, the OAc groups tend to provide stabilizing interactions with the guest, as reflected by the more negative interaction energies, $\Delta\langle U + W\rangle$, for tweezers 6 relative to tweezers 5 (Table 1), for guests 8 and 9. As shown in Figure 4B, the ligand tilts in the binding site to form these interactions, while it is more symmetrically positioned in the absence of the OAc groups. These stabilizing interactions are compensated by the greater loss of configurational entropy for tweezers 6 versus tweezers 5 (Table 1), presumably due to the reduced rotational freedom of the OAc groups upon binding and perhaps also to a further reduction in the rotational and translational freedom of the bound ligand. The net result is that tweezers 6 binds the guests about as well as tweezers 5.

4.2. Macrocyclic Barbiturate Receptor. The macrocyclic barbiturate receptor **10** (Figure 5) is an early example of a synthetic host that binds guests via hydrogen bonding in chloroform.[10,43] The design is based upon hydrogen-bonding complementarity between the diamidopyridine moiety of the receptor and the imide components of the barbiturate guests, **11**, **12**, and **13** (Figure 5). Aromatic spacers X and Y (Figure 5) link two diamidopyridines to correctly position hydrogen-bonding groups while preventing the receptor from forming intramolecular H-bonds and thereby closing the binding site.

As shown in Table 1, the computed binding free energies agree with the experimental data to within 1 kcal/mol and provide the correct ranking of affinities. The affinities are much stronger than those for the clips and tweezers, with binding free energies more negative than −8 kcal/mol. This difference is traceable to more favorable changes in the mean energy, $\Delta\langle U + W\rangle$, up to ∼ −28 kcal/mol, due to the formation of multiple intermolecular hydrogen bonds. These strongly favorable energy changes are partly compensated by strongly unfavorable losses in configurational entropy, $-T\Delta S^\circ_{\text{config}}$, up to ∼21 kcal/mol.

Although the design seeks to preorganize the receptor for binding of its barbiturate guests, the calculations indicate that the free receptor is quite flexible, much more so than the clips and tweezers. This is reflected in Table 2, which lists the numbers of energy minima with free energies within *RT* (0.6 kcal/mol) and 5 kcal/mol of the global minimum of each free molecule and complex. The most stable conformations of the free barbiturate receptor appear collapsed rather than open, and the H-bonding components of the two diamidopyridines point in opposite directions, rather than converging to form the intended binding site; see, for example, Figure 6A. The free receptor does access open conformations with convergent H-bonding components (e.g., Figure 6B), but these lie greater than 3.5 kcal/mol above the global minimum and they differ significantly from the most stable conformation of the bound complexes (Figure 8A). No crystallographic structure is available for this receptor, but another receptor in this series, with a naphthalene spacer instead of a diphenylmethane group in the Y position,[43] has been shown to adopt an open, flat conformation in the solid state, consistent with the design. Calculations for this variant receptor yield a global energy minimum (Figure 7) that matches the design (Figure 5) and agrees with the crystallographic analysis. Thus, the precise nature of the Y linker appears to significantly influence conformational preferences in this series of receptors.

With barbital **11** and phenobarbital **12** bound, the global energy minimum of the complex still does not adopt the expected flat, open conformation. Instead, the Y component of the receptor curves back from the X component and the diamidopyridine moieties to create a 3D binding cleft that allows formation of the 6 expected H-bonds, along with additional host−guest contacts out of the plane of the H-bonding groups (Figure 8A and B). This receptor conformation differs by at least 3.0 Å RMSD from the four most stable conformations of the free receptor, implying an induced fit upon binding of these ligands.

Interestingly, the R_2 phenyl substituent of phenobarbital **12** forms weak interactions with the receptor **10** that help to stabilize the complex (see $\langle U + W\rangle$ in Table 1). However, these interactions also reduce the flexibility of the complex versus the unbound state and thus leads to a penalty in configurational

(43) Chang, S. K.; Vanengen, D.; Fan, E.; Hamilton, A. *J. Am. Chem. Soc.* **1991**, *113*, 7640−7645.

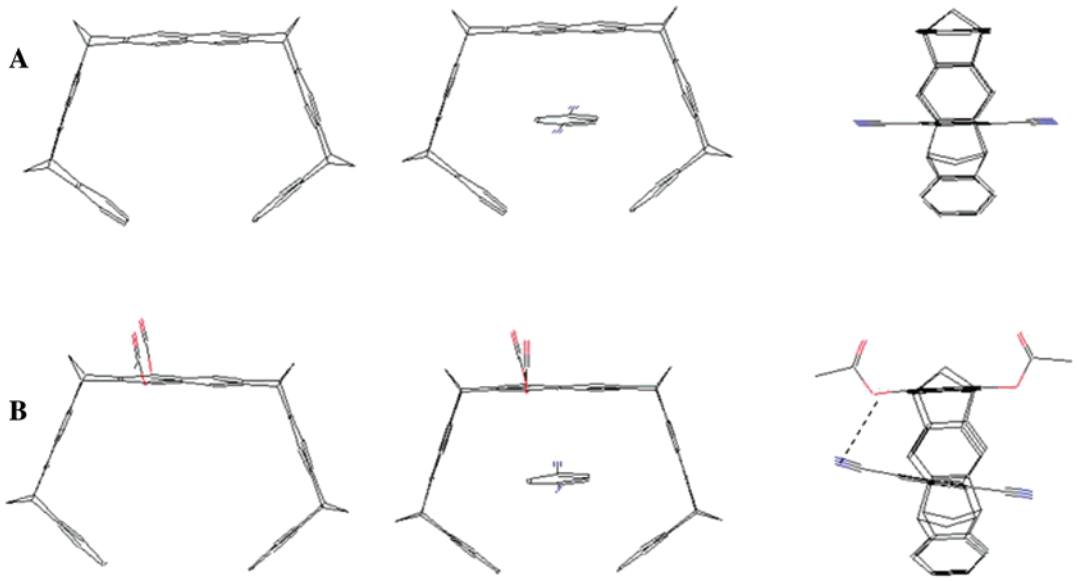
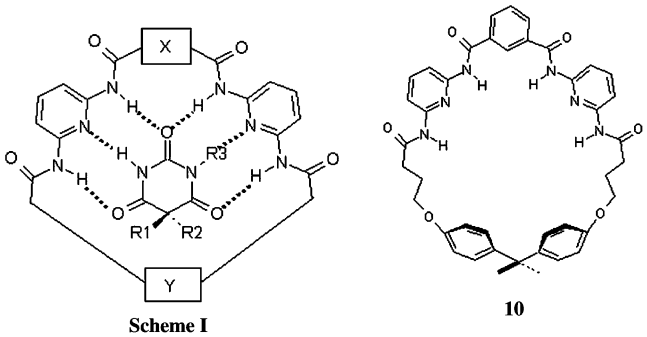


Figure 4. Computed global energy minima of free molecular tweezers and their complexes with guest **9**. (A) Tweezers 5. (B) Tweezers 6. Left: free receptors. Middle: complexes, front view. Right: complexes,side view.



Scheme I

11: Barbital; $R_1 = R_2 = Et$, $R_3 = H$
12: Phenobarbital; $R_1 = Et$, $R_2 = Ph$, $R_3 = H$
13: Mephobarbital; $R_1 = Et$, $R_2 = Ph$, $R_3 = CH_3$

Figure 5. Macrocyclic barbiturate receptor **10** (left) and schematic of the design of the receptor (right) with spacers X and Y. Barbital **11**: $R_1 = R_2 = CH_2CH_3$, $R_3 = H$. Phenobarbital **12**: $R_1 = CH_2CH_3$, $R_2 = Ph$, $R_3 = H$. Mephobarbital **13**: $R_1 = CH_2CH_3$, $R_2 =$ phenyl, $R_3 = CH_3$.

entropy, which more than compensates for the gain in interaction energy (see $-T\Delta S^\circ_{config}$ in Table 1).

The calculations confirm that the diamidopyridine components of the receptor select for the hydrogen-bonding pattern of the barbituric acid core. Thus, while barbital **11** and phenobarbital **12** bind tightly via six hydrogen bonds to the receptor **10**, the methyl in the R_3 position of mephobarbital **13** interrupts the hydrogen-bonding pattern, reducing the binding affinity by a calculated value of >4 kcal/mol, in accord with experiment. Moreover, the lack of strong interactions between mephobarbital **13** and the receptor **10** prevents the induced fit seen for the other two ligands. Thus, the most stable conformation of the receptor in this complex deviates by only 0.65 Å RMSD from the global minimum of the free receptor. Mephobarbital binds approximately on top of one benzene and forms only two H-bonds with a diamidopyridine moiety, as shown in Figure 8C.

4.3. Configurational Entropy. Binding is usually accompanied by a loss of configurational entropy on the part of the reactants, and the present calculations afford a perhaps unique view of the balance between this entropic cost and the driving force for binding, $\Delta\langle U + W\rangle$, which is here termed the configurational energy. As shown in Table 1, configurational entropy strongly opposes binding in every instance considered here, contributing 6−16 kcal/mol to the computed values of ΔG°, while the configurational energy drives binding by contributing −9 to −24 kcal/mol to ΔG°.

The losses in configurational entropy on binding could result from a reduction in the number of low-energy wells on binding and/or from a reduction in the average width of the wells. This issue can be addressed semiquantitatively by examining the number of energy wells N whose standard chemical potential

Table 2. Number of Energy Minima with Chemical Potential within 5 kcal/mol and RT (0.6 kcal/mol) of the Most Stable Conformation Found

	number of energy minima	
	within 5 kcal/mol	within RT
clip **1**	6	2
clip **2**	1	1
clip **3**	20	4
TCNB (**4**)	1	1
TCNB−clip 1	3	2
TCNB−clip 2	6	4
TCNB−clip 3	37	7
tweezers **5**	1	1
tweezers **6**	6	1
guest **7**	1	1
guest **8**	3	3
guest **9**	1	1
guest 7/tweezers 5	2	2
guest 8/tweezers 5	3	2
guest 9/tweezers 5	2	2
guest 7/tweezers 6	9	2
guest 8/tweezers 6	26	2
guest 9/tweezers 6	11	2
receptor (**10**)	63	4
barbital (**11**)	8	3
phenobarbital (**12**)	5	2
mephobarbital (**13**)	7	1
barbital/receptor	27	5
phenobarbital/receptor	18	1
mephobarbital/receptor	26	1

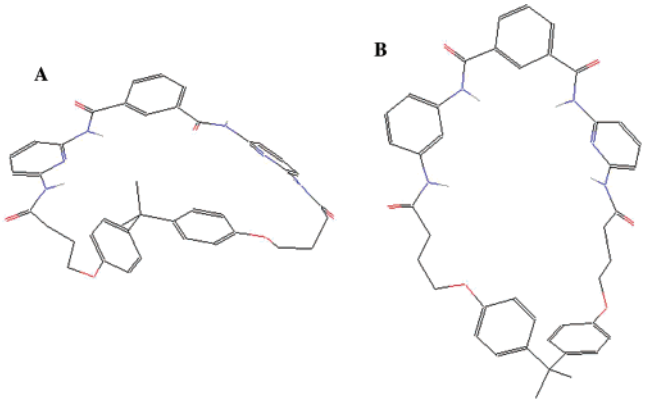


Figure 6. Global free-energy (or chemical potential) minimum (A) and one open conformation (B) of the free barbiturate receptor **10**.

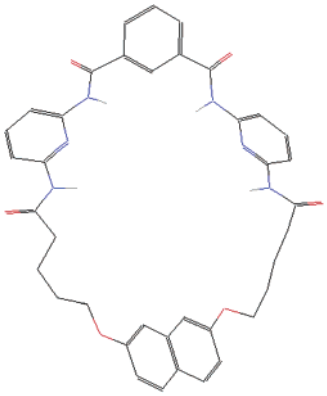


Figure 7. Global free-energy energy minimum of a variant barbiturate receptor.[43]

μ° is within thermal energy ($RT = 0.6$ kcal/mol) of the most stable conformation. If we make the approximation that these energy wells are equally stable and that less stable wells contribute trivially to the overall stability, then the entropy change on binding can be estimated as $-RT \ln(N_{RL}/N_R N_L)$. The numbers of energy wells are shown in Table 2, and the resulting values of the configurational entropy range between -1.25 and 0.83 kcal/mol for the 12 systems examined here. Since these entropy losses are far smaller than the full changes in configurational entropy (Table 1), it may be concluded that the entropy losses upon binding are due primarily to narrowing of energy wells rather than to a drop in the number of wells, in the cases examined here.[44]

There is a striking correlation between the configurational entropy (S°_{config}) and the configurational energy ($\Delta\langle U + W\rangle$), as shown in Figure 9. This result echoes the common experimental observation of entropy−enthalpy compensation, though it is important to note again that ΔS°_{config} omits entropic contributions from the solvent that are implicit in $\Delta\langle W\rangle$ and that $\Delta\langle U + W\rangle$, by the same token, is not purely enthalpic. Despite this caveat, the present results support the view that stronger binding typically results in a greater loss of configurational entropy. It has previously been emphasized that experimental observations of entropy−enthalpy compensation can be illusory, so it is important to critically assess the significance of such results.

(44) The narrowing of energy wells includes the reduction in translational and rotational freedom on going from two free molecules to a single complex. In effect, binding reduces the accessible translational and rotational volume accessible to one of the molecules from its standard solution value of $8\pi^2/C^\circ$.

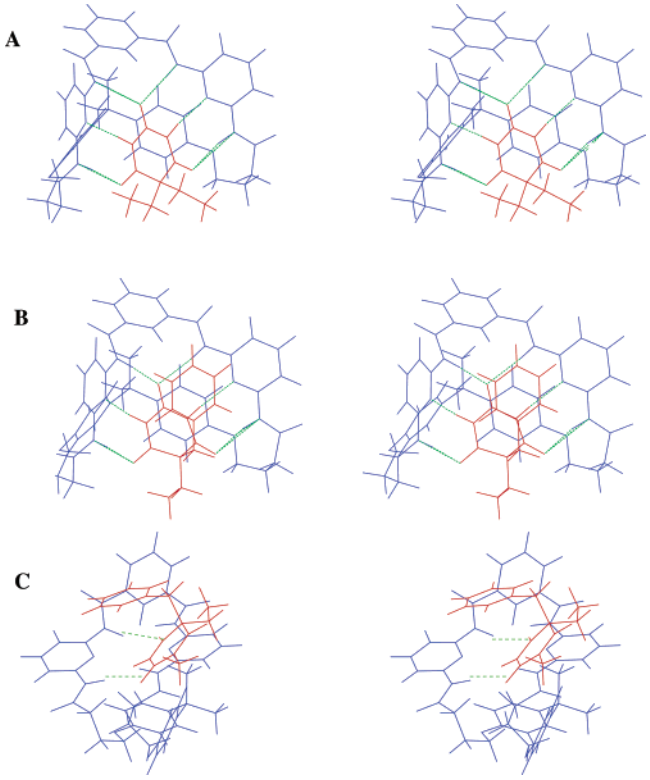


Figure 8. Global free-energy minima of complexes of the barbiturate receptor **10** with (A) barbital **11**, (B) phenobarbital **12**, and (C) mephobarbital **13**, shown as stereopairs for wall-eyed viewing.

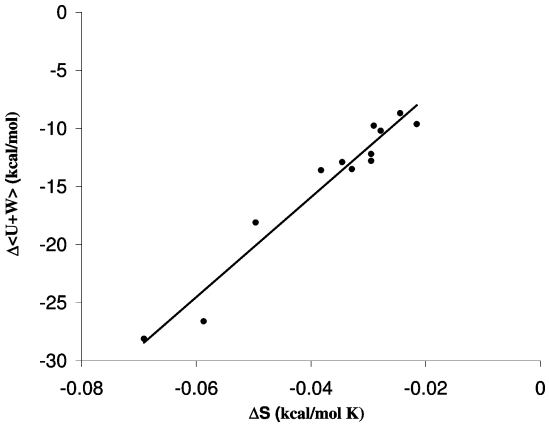


Figure 9. Scatter plot of computed changes in configurational entropy versus changes in configurational energy, $\langle U + W\rangle$, upon binding, for all 12 binding reactions studied here.

The criterion for significance is that the slope of the graph (Figure 9), which is known as the compensation temperature T_c, must differ from the actual temperature by more than twice the standard error of T_c.[1,4,45] The compensation seen here is significant by this criterion, because $T_c = 430$ K with a standard error of 31 K, while the ambient temperature is ~300 K.

It is not clear why the relationship between configurational energy ($\Delta\langle U + W\rangle$) and configurational entropy (ΔS°_{config}) so often has a near-linear form. We conjectured that linearity might result from a correlation between the depth of an energy well and its narrowness. However, no such correlation is found in the present data, as illustrated in Figure 10, for the conformations of one of the tweezers complexes, compound **6** with **7**. Here,

(45) Krug, R.; Hunter, W.; Grieger, R. *Nature* **1976**, *261*, 566−567.

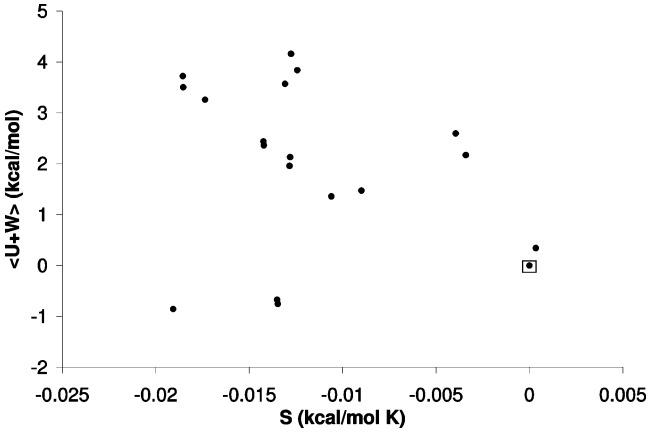


Figure 10. Scatter plot of computed configurational entropy versus configurational energy, $\langle U + W \rangle$, for the conformations of the complex of compounds **6** and **7**. Both quantities are referenced to their values at the global energy minimum, marked as an open square. The apparent pairing of some conformations results from the presence of mirror images or near mirror images and, in at least one case, from a small adjustment in the distance between the tips of the tweezers.

the conformations with the lowest values of $\langle U + W \rangle$ are those with both OAc arms directed upward and out of the binding site, as noted in a previous section and illustrated in Figure 4B. Interestingly, these conformations also are high in entropy, presumably because the OAc arms are relatively mobile when directed into the solvent. Thus, the most stable conformations here are associated with energy wells that are both deep and wide.

Given the correlation between $-T\Delta S^\circ_{\text{config}}$ and $\Delta\langle U + W \rangle$, it is of interest to determine whether $\Delta\langle U + W \rangle$ by itself predicts or correctly ranks binding free energies. As shown in Table 1, $\Delta\langle U + W \rangle$ by itself markedly overestimates the actual binding free energies. Moreover, it does not rank the affinities as accurately as the calculated standard free energies of binding ΔG°. For example, $\Delta\langle U + W \rangle$ is more favorable for guest 8 with tweezers 5 than for guest 9, largely because of more favorable Coulombic interactions, though guest 9 actually binds more tightly. The calculated values of ΔG° capture this relationship correctly because $-T\Delta S^\circ_{\text{config}}$ is less unfavorable for guest 9 than for guest 8, presumably because of the entropic cost of partly immobilizing the aldehyde moieties of guest 8, along with the additional configurational restriction imposed by their polar interactions with the tweezers.

4.4. Detailed Analysis of Energy Wells. Finally, we inquire whether the detailed treatment of each energy well in the present method really is necessary to achieve good agreement with experiment. Two issues are considered. The first is whether accounting for the width of individual energy wells by the HA/MS method is important in obtaining accurate results, since this step adds somewhat to the complexity of the method. This issue can be addressed by replacing $G_{\text{corr},i}$ in eq 12 with $U(\boldsymbol{r}_i) + W(\boldsymbol{r}_i)$; that is, with the energy at the base of the energy well, including the finite difference Poisson−Boltzmann and surface area correction terms. This is essentially the same approach taken by Lipkowitz in pioneering calculations of a similar type.[46] The resulting binding free energies, listed as $\Delta G^{\circ\prime}$ in Table 1, are too negative and incorrectly rank the binding energies. Fortunately, the HA/MS step requires at most a few seconds of computer time per energy minimum for the systems studied here; most of the computer time for the M2 method is spent on energy-minimization during the Tork conformational search. The second issue is whether it is important to adjust the electrostatic solvation energy of each energy well by subtracting the generalized Born contribution and replacing it with the result of a finite difference solution of the Poisson equation. This correction does appear to be necessary; for example, the computed binding free energies of the three clips (**1**, **2**, **3**) with the TCNB ligand (**4**) are −1.4, −1.8, and 0.6 kcal/mol, respectively, when the electrostatic correction is omitted.

(46) Lipkowitz, K. B.; Demeter, D. A.; Zegarra, R.; Larter, R.; Darden, T. *J. Am. Chem. Soc.* **1988**, *110*, 3446−3452.

5. Discussion

5.1. General Comments. This study examines a range of host−guest systems, including some driven primarily by aryl−aryl interactions and others by hydrogen-bonding, with affinities modulated by strain and preorganization. Overall, the accuracy of the present calculations is promising and suggests that the methodology will be more broadly useful for interpreting experimental studies of host−guest binding and for the design of hosts targeted to bind chosen guests. Although this study focuses on binding in an organic solvent, a separate study shows similarly accurate results for binding in an aqueous environment (Chen, W.; Chang, C.-E.; Gilson, M. K. *Biophys. J.*, in press). The present method should also be helpful for testing and ultimately improving force field parameters, since it brings a large collection of experimental binding data into the realm of the calculable, and thus greatly expands the set of experimental data that can be used for validating energy models.

The calculations also provide reasonable interpretations of the experimental data, including some unexpected observations. For example, although clip 3 encloses the guest more completely than clip 2, its affinity is weaker. As detailed in the Results section, this reduction in affinity appears to result from the energy cost of rotating the OAc groups out of their preferred position obstructing the binding site. In contrast, the OAc groups in tweezers 6 prefer to reside outside the binding site and therefore have little effect upon affinity. Interestingly, then, even a simple chemical component at the periphery of a binding site can modulate the thermodynamics of binding via a nontrivial interplay among stabilizing intermolecular interactions, steric blockade, and mobility. Chemical intuition may allow one to recognize that these forces are at work, but their net effect is much more difficult to assess. It is encouraging that the calculations strike the right balance in these systems.

It is also of interest that the two tighter-binding ligands of the barbiturate receptor **10** induce restructuring of the host to form a complementary binding site, while mephobarbital **13** binds more weakly to what is essentially the free conformation of the receptor. These observations suggest that the affinity of this particular receptor for barbital **11** and phenobarbital **12** might be enhanced by a chemical modification that would stabilize its bound conformation. These results also indicate that, in designing a synthetic receptor, one ought to model not only the free receptor, but also the receptor−ligand complex. The Tork search algorithm is well-suited to this purpose because it is effective for bimolecular complexes involving guests and receptors with complex, flexible ring systems.

5.2. Entropy of Binding. The computed changes in configurational entropy are large, comparable in magnitude to the

changes in configurational energy, $\langle U + W \rangle$, and omitting them therefore leads to large errors in $\Delta G°$. The losses in entropy upon binding are found, in the cases studied here, to result primarily from a reduction in the average width of energy wells, rather than a drop in the number of energy wells in the bound versus the free state. This pattern is consistent across the systems studied here and may well be fairly general for host−guest systems and for other systems without many flexible degrees of freedom. However, changes in the number of low-energy conformations upon binding may be more important for macromolecules, since they typically access many more energy wells at room temperature.

It has been proposed that the change in configurational entropy for protein−ligand binding should correlate with the number of rotatable bonds in the ligand (e.g., ref 47), but no such correlation is observed here. This lack of correlation is consistent with our observation that the entropy changes are attributable more to a narrowing of energy wells than to a drop in the number of energy wells, since the logic connecting the number of rotatable bonds to the configurational entropy is based upon a counting of energy wells. On the other hand, the guests studied here have few rotatable bonds, and most of the rotatable bonds in the relatively flexible barbiturate receptor are restricted by involvement in the macrocyclic ring. For more flexible, chainlike molecules, the binding entropies may correlate with the number of rotatable bonds, as indeed seen for a series of cyclic urea inhibitors of HIV-1 protease with simple alkyl chains, which were studied with the first generation of Mining Minima.[48] In general, then, the presence or absence of such a correlation is likely to depend on the nature of the series of compounds. Thus, it is unlikely that models of binding can achieve so-called "chemical accuracy" of ±1 kcal/mol without a reasonably sophisticated treatment of configurational entropy.

The present calculations show a striking correlation between the change in configurational entropy upon binding and what we have termed the configurational energy, $\Delta\langle U + W \rangle$, in a manner analogous to entropy−enthalpy compensation. This observation is consistent with the intuitive concept that stronger intermolecular attractive forces pull the receptor and ligand more tightly against each other's repulsive van der Waals cores, thus reducing mobility and entropy. Experimentally, a spurious entropy−enthalpy compensation can appear if the experimental noise in $\Delta H°$ is considerably greater than that in $\Delta G°$, when $\Delta S°$ is computed from $\Delta G°$ and $\Delta H°$.[1,4,45] However, there is no basis for attributing greater numerical noise to $\Delta\langle U + W \rangle$ versus $\Delta S°_{\text{config}}$ in the present calculations; moreover, the results meet the recognized criterion for significance[1,45] since the compensation temperature is 430 K with a standard error of only 30 K. It has also been argued that entropy−enthalpy compensation may, in some cases, amount to no more than a trivial statement that $\Delta G°$ varies little across a series of measurements.[2] However, the present calculations place no constraint on $\Delta G°$ and, as discussed in the Results section, analysis of the thermodynamics in the context of structure yields physically meaningful interpretations for drops in $\Delta\langle U + W \rangle$ opposed by compensatory changes in $-T\Delta S°_{\text{config}}$. We thus conclude that the compensation between $\Delta\langle U + W \rangle$ and $\Delta S°_{\text{config}}$ seen here is significant, and that the tendency of $\Delta G°$ not to change much as $\Delta\langle U + W \rangle$ varies is best understood not as the mathematical cause of the compensation seen here, but rather as its physical consequence.

Acknowledgment. We thank Dr. Ken Houk for providing data reported in ref 2, Dr. Frank-Gerrit Klarner for helpful discussions and for providing preprints and reprints regarding the tweezers and clips, Dr. Andrew D. Hamilton for helpful discussions regarding the barbiturate receptor, Dr. Wei Chen for providing the symmetry detection methodology, and Dr. Hillary S. R. Gilson for valuable comments on the manuscript. This publication was made possible by Grant No. GM61300 from the National Institute of General Medical Sciences (NIGMS) of the NIH. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the NIGMS.

Supporting Information Available: Graphs of the cumulative free energy as a function of the number of energy wells for two of the most complex systems studied here, the free barbiturate receptor and the complex of this receptor with mephobarbital (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA047115D

(47) Böhm, H.-J. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 243−256.
(48) Mardis, K. L.; Luo, R.; Gilson, M. K. *J. Mol. Biol.* **2001**, *309*, 507−517.