Conformational and Enantioselectivity in Host−Guest Chemistry: The Selective Binding of Cis Amides Examined by Free Energy Calculations

Richard H. Henchman,† Jeremy D. Kilburn, David L. Turner, and Jonathan W. Essex*

School of Chemistry, University of Southampton, Southampton, SO17 1BJ, U.K.

Received: August 3, 2004

Relative binding free energies of amino acid derivatives to the host macrobicycle **12** in chloroform are calculated to determine the effect of stereochemistry and amide bond conformation. Simulations are performed for three different amino acid derivatives using Monte Carlo simulations, free energy perturbations, and the generalized Born/surface area solvation model. The free energy results support previous experimental findings that the L enantiomers of *N*-acetyl phenylalanine carboxylate and *N*-acetyl alanine carboxylate bind preferentially over the D enantiomers and that the amide bond is in the cis conformation for the bound L enantiomers. The enantioselectivity arises because of better shape complementarity, and the cis stabilization is achieved by a selective hydrogen bond to the cavity rim of the host and a remote steric stabilization of the side-chain χ_1 dihedral angle. This computational approach was also applied to *N*-acetyl glycine carboxylate and indicated that, for this amino acid derivative, both the trans and the cis forms of the acetyl amide can be accommodated within the macrobicycle cavity. Experimental evidence to support this finding was obtained by detailed NMR experiments on a 1:1 complex of M**12** and *N*-acetyl glycine. The study demonstrates two important points regarding binding and conformation: First, the destabilization of the χ_1 dihedral angle of *N*-acetyl phenylalanine away from the most stable conformation appears to stabilize the cis conformation of the preceding amide bond, suggesting that amino acids other than proline might be able to increase the likelihood of cis amide bonds. More generally, internal conformational coupling is an important factor to consider in molecular recognition. Second, the comprehensive sampling protocol undertaken for this relatively small system that allows the host to adjust properly to each guest is shown to be crucial in obtaining converged free energies and different binding modes. Subtle changes in host geometry have profound effects on binding. Conversely, different binding modes arising from small changes in guest conformation, chirality, or side chain serve as a reminder that binding modes are not always transferable between apparently similar molecules.

Introduction

The host molecule macrobicycle **12** has been observed using NMR methods to exhibit the unusual property of selectively binding *N*-acetyl-L-phenylalanine carboxylate and *N*-acetyl-L-alanine carboxylate (as their tetrabutylammonium salts) in the cis conformation in chloroform, whereas D derivatives were bound with the amide bond in the more stable trans conformation.[1] The primary cause of this difference was that the L derivative was bound inside the cavity, whereas the D derivative was bound outside. Molecular modeling indicated that the cis conformation for the guest was being stabilized by a hydrogen bond with the cavity rim of the host.[1] Building from these initial results, the study here is a comprehensive examination of the selectivity and stabilization of such systems that reveals a number of aspects important to binding. First, the stabilization of unstable conformations can be a factor in molecular recognition or activation. Second, elucidating general principles of selectivity is useful for the design of molecules, particularly those relating to amino acids[2−9] and requiring enantioselectivity.[10−13] Third, the effect of host and guest flexibility on binding, the good sampling necessary to observe this, and the determination of binding modes are important factors in rational drug design.[14−17] Adequate sampling is currently one of the major limitations facing computational modeling, and the small size of host−guest systems is particularly advantageous in assessing sampling. Finally, the cis/trans equilibrium of the amide bond is important in the folding, structure, and function of many peptides and proteins.[18,19] Enzymes that catalyze cis/trans isomerization such as cyclophilin and FKBP are targeted therapeutically to induce immunosuppression, and their roles in other pathologies such as AIDS, neurodegenerative diseases, and cancer make them ongoing subjects of interest.[20,21] More generally, host−guest systems are important to study because their small size and ready chemical modification make them more amenable to the systematic determination of the role of the factors affecting binding.[22] Their small size also renders them more suitable to simulation studies because they can be treated more thoroughly and are better suited as test cases for methodological testing and development, of which there are many recent examples.[23−32]

The objectives of this study are to calculate the relative free energies of binding of amino acid derivatives to the host, macrobicycle **12** (M**12**), and determine and understand the effects of amide bond conformation, stereochemistry, type of side chain, and host flexibility on binding. M**12**, shown in Figure 1, is a cup-shaped molecule made up of two rings that binds amino acid derivatives by virtue of strong hydrogen bonds

* To whom correspondence should be addressed. E-mail: jwe1@soton.ac.uk.

† Current address: Howard Hughes Medical Institute and Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0365.

10.1021/jp046520o CCC: $27.50
Published on Web 10/19/2004

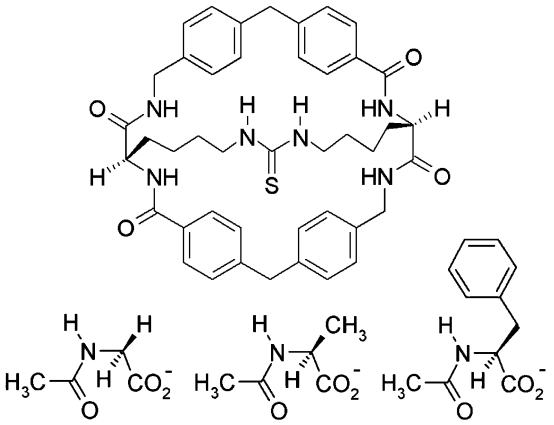

Figure 1. The host molecule macrobicycle **12** (M**12**) and the three guest amino acid derivatives, *N*-acetyl−glycine, *N*-acetyl-L-alanine, and *N*-acetyl-L-phenylalanine, abbreviated in the text as Gly, Ala, and Phe.

between the thiourea and amide units of the host and the carboxylate group of the guest. The three amino acid derivatives studied in this work are glycine, alanine, and phenylalanine, all capped by acetyl groups on their N termini, as shown in Figure 1. Each derivative will hereafter be denoted by its conventional amino acid abbreviation, Gly, Ala, and Phe, respectively. A simulation protocol is developed to calculate the free energies as follows: The host−guest system is parametrized. Two chloroform solvent models are implemented and compared in the free energy calculations. One is an explicit chloroform model, and the other is a continuum generalized Born/surface area (GBSA) model, also parametrized here. Equilibrium configurations are generated using the Monte Carlo (MC) Metropolis method with a range of move types to ensure extensive sampling. Free energy differences are calculated using the free energy perturbation (FEP) method. The free energy results for Ala and Phe are compared with the experimental findings and rationalized by comparing a range of properties including binding geometry, number of close contacts, hydrogen bonding, host conformation, and guest conformation. The simulations indicated that a different mode of binding was taking place in the complex formed between M**12** and Gly. Subsequent NMR experiments were found to confirm this prediction.

Method

Force Field. The system was modeled using the all-atom OPLS force field[33,34] with a 10-Å cutoff for nonbonded interactions. Owing to the presence of nonstandard functionality in this system, additional force field parameters had to be derived. The thiourea moiety needed full parametrization, and dihedral angle parameters were required for the guest and parts of the host. The full listing of the parameters is given in the Supporting Information (Table S1 and Figures S1 and S2). Briefly, equilibrium bond lengths and angles for the thiourea moiety were taken from MP2/6-31$^+$G*-optimized structures of *N*,*N*′-dimethylthiourea produced using Gaussian 94.[35] Force constants for bonds and angles were not required for thiourea as this moiety is held rigid in the MC simulations. Dihedral angle parameters were obtained by fitting a three-term Fourier series plus the nonbonded contribution to the energies of HF/6-31G* ab initio minima and transition states along the dihedral angle energy profiles of small representative molecules containing the dihedral angle and capped by methyl groups, again using Gaussian 94. Charges for *N*,*N*′-dimethylthiourea were obtained using REPD charges[36] for the conformation that this molecule predominantly adopts in the host. As with the RESP method[37] used for the AMBER force field,[38] these charges were fitted to the molecule's HF/6-31$^+$G* electrostatic potential with a quadratic charge restraint that was parametrized to reproduce OPLS charges. REPD charges have been shown to reproduce the free energy of solvation for a range of small molecules.[39] Finally, the charges of two carbon types at the junctions of two different functionalities were adjusted to accommodate both functionalities. One carbon is the tertiary carbon connecting three chains, and the other connects the two aryl rings.

Solvent Model. Two types of solvent model were implemented for chloroform. Explicit-solvent simulations were performed using the OPLS model[40] for chloroform. Guest perturbations were performed in a cubic box of side 33 Å containing 265 chloroform molecules, while host−guest perturbations were done in a larger 41 × 44 × 45 Å^3 box containing 592 molecules. Implicit solvent simulations used the generalized Born/surface area (GBSA) model.[41,42] The required GBSA code was taken from the Tinker software package[43] and implemented into MCPRO 1.5,[44] the simulation program used in this work. Because a GBSA parametrization for all-atom OPLS in chloroform could not be found in the literature, it was necessary to perform the parametrization in this work. The GB free energy of solvation is given by

$$\Delta G_{\mathrm{GB}} = -166\left(1 - \frac{1}{\epsilon}\right)\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{q_i q_j}{\sqrt{r_{ij}^{\,2} + \alpha_{ij}^{\,2}\mathrm{e}^{-D_{ij}}}} \qquad (1)$$

where ϵ is the dielectric constant of the solvent, r_{ij} is the distance between atom i and j, $\alpha_{ij} = \sqrt{(\alpha_i \alpha_j)}$, and $D_{ij} = r_{ij}^2/(2\alpha_{ij})^2$. The pairwise descreening approximation[45] was used for the calculation of the Born radii, α_i. ΔG_{GB} requires the screening parameter, S, that scales the Born radii because of the pairwise descreening approximation. To parametrize S, the electrostatic part of the free energies of solvation of 20 small organic molecules (listed in the Supporting Information) was fitted to the electrostatic component of the free energies of solvation, calculated using the Poisson method.[46] Finite-difference Poisson calculations were carried out using UHBD[47] with a 65 × 65 × 65 grid of spacing 0.3 Å. A dielectic constant of 4.81 was used for chloroform and 1 for the cavity, and the separating boundary was taken as the solvent-accessible surface area of the molecule, defined as the atom's van der Waals radius plus the 2.5-Å probe radius of chloroform.[48] Polar hydrogens, having zero radius in the OPLS force field, were assigned a van der Waals radius of 1.2 Å. The electrostatic potential of each molecule must be calculated in the dielectric, $\phi^{\mathrm{sol}}(\mathbf{r})$, and in a vacuum, $\phi^{\mathrm{gas}}(\mathbf{r})$ by solving the Poisson equation

$$\nabla\cdot\epsilon(\mathbf{r})\nabla\phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \qquad (2)$$

over all space $\mathbf{r}$, where $\epsilon(\mathbf{r})$ is the dielectric constant and $\rho(\mathbf{r})$ is the charge density. Poisson free energies are then given by

$$\Delta G_{\mathrm{Poisson}} = \frac{1}{2}\sum_i q_i(\phi_i^{\mathrm{sol}} - \phi_i^{\mathrm{gas}}) \qquad (3)$$

where the sum is over all grid points. $\Delta G_{\mathrm{Poisson}}$ was calculated and averaged for 10 random orientations of each molecule to remove rotational dependence due to the cubic grid. A least-squares fit of the two free energies obtained using the Poisson and generalized Born models provided the value of $S = 0.56$. The average error for this fit was only 0.3 kcal mol^{-1}.

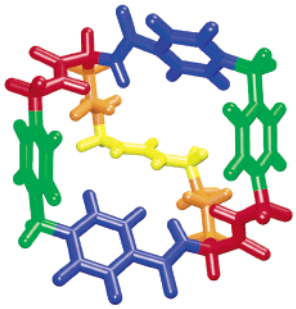

Figure 2. The nine "residues" of M**12**, colored according to type, that are moved independently in the MC sampling. The figure was made in VMD.[64]

The surface area free energy component is calculated using

$$\Delta G_{SA} = \sum_i \gamma_i (SASA)_i \qquad (4)$$

where the sum is over all atoms and γ_i and $SASA_i$ are the atomic solvation parameter and the solvent-accessible surface area, respectively, for each atom. To parametrize eq 4, two γ parameters were found to be sufficient; any more would be overfitting to the 20 molecules. These parameters are γ_{NO} for polar nitrogen and oxygen atoms and γ_{rest} for the remainder. $SASA_i$ terms were calculated using the method of Richmond for multiple overlapping spheres.[48] The parametrization involved a least-squares fit between the experimental free energy of solvation[49−51] and the total calculated free energy for the same 20 molecules. The parameters obtained are $\gamma_{NO} = 16.0$ cal Å^{-2} and $\gamma_{rest} = -9.0$ cal Å^{-2}. The average error fitting to experiment was 0.7 kcal mol^{-1}. All experimental and calculated free energies in the parametrization are listed in Supporting Information, Table S2. Using this three-parameter fit, the total GBSA free energy of solvation, ΔG_{sol}, is given by

$$\Delta G_{sol} = \Delta G_{GB}(S) + \Delta G_{SA}(\gamma_{NO}, \gamma_{rest}) \qquad (5)$$

Because the SASA calculation for the whole complex is time-intensive and SASA changes relatively slowly, SASA was updated only every 100 configurations.[52]

Host and Guest Z Matrix and Starting Configuration. Because much of the MC sampling was done by varying internal coordinates (angles and dihedral angles), the success of this sampling is strongly dependent on how these internal coordinates are defined in the initial **Z** matrix.[24] The means to achieve good sampling is to make as many atoms move simultaneously a long way, while at the same time minimizing changes in energy to achieve good acceptance. By analogy to the way proteins are partitioned into residues, the host was partitioned into nine "residues". These residues are illustrated in Figure 2. Atoms in each residue are defined with respect to other atoms in the same residue or to three anchor dummy atoms at the center of the host. The exceptions to this rule are the hydrogens attached to carbons at the junction between two residues. These hydrogens are defined in the **Z** matrix with respect to both residues to minimize strain when either residue moves. Each residue was allowed to rotate and translate, albeit marginally because of the constraining bonds that connect them, and was constructed so that the most flexible dihedral angles were defined explicitly in the **Z** matrix. For example, the flipping of aryl groups can be directly sampled by defining a dihedral angle about the axis connecting the two carbons at either end. Each guest was manually docked to bring the carboxylate group adjacent to the thiourea polar hydrogens in accordance with NMR data.[1] Using these initial structures, a large number of configurations was generated with simulated annealing for each type of guest. Considerable variation in the resulting structures was found principally in the conformation of the hydrocarbon chain. As described below, using different structures in explicit chloroform resulted in an undesirable dependence of the free energy on the starting structure for the host−guest mutations. However, once the sampling in continuum solvent was shown to be adequate to remove this dependence, the same starting structure was used thereafter for all mutations.

MC Simulations. Equilibrium configurations were generated at a temperature of 293 K and a pressure of 1 atm (explicit solvation model) using Metropolis MC sampling implemented in the simulation program MCPRO1.5.[44] The types of moves employed were rotations and translation of the solvent, guest, and host residues; changes to host and guest internal coordinate angles and dihedral angles defined in the **Z** matrix; extra-large dihedral angle moves for the side-chain dihedral angles; and volume, conrot,[53] and flip[54] moves. No bonds were directly sampled, nor were any degrees of freedom in the thiourea and aryl moieties, so as to keep them planar. An internal coordinate move consisted of randomly selecting a residue and changing all of its variable internal coordinates. There were three types of guest move: internal coordinates, translation, and rotation. Conrot moves involve the concerted rotation of seven adjacent dihedral angles in a 10-atom chain while keeping both ends fixed, and hence, this is a very efficient way of achieving a large conformational change for constrained systems such as M**12**. A flip move brings about more local changes by moving a single atom about the axis connecting its two bonded neighbors.[54] The move preserves the bond lengths with the neighboring atoms but distorts the angles centered at these two atoms. The conrot code used by Boone et al.[53] and the flip move were implemented into MPCRO. The conrot and flip moves were applied to all dihedral angles of the hydrocarbon chain containing the thiourea moiety, but not to the thiourea moiety itself, to avoid sampling its two stiff C−N−C−N dihedral angles. This was achieved by placing a single dummy atom at the intersection of the two C−N bond vectors (see Figure 3) to replace the two nitrogens and carbons of thiourea in the sequence of conrot atoms. Conventional internal coordinate sampling was sufficient to sample the rest of the molecule as judged from a comparison with structures generated by simulated annealing.

Move sizes were chosen to give ∼40% acceptance rates for solvent, volume, and internal coordinate moves. However, for the other move types, larger move sizes with lower acceptances were chosen because these moves have to traverse larger energy barriers. Conrot moves using a driver dihedral angle maximum amplitude of 5° yielded 10% acceptance; flip moves with a maximum move size of 50° yielded 30% acceptance; additional large dihedral angle moves of maximum size 180° for the two dihedral angles of the Phe side chain yielded 10% acceptance; and additional large guest translation or rotation moves with maximum sizes of 0.4 Å and 30°, respectively, yielded 10% acceptance. These acceptance rates are for continuum solvent. Acceptance rates were found to be considerably lower in explicit solvent. The total number of configurations sampled and the percentages of each type of move are listed in Table 1 for the seven different simulation protocols. The first, simulated annealing, was used to determine the range of structural variation in the host with guest present. Each host−guest complex, held together by a small harmonic restraint, was simulated for 0.5 $\times$ 10^6 configurations at 2000 K in the gas phase and then cooled in 20 steps of 20 K, each lasting for 0.1 $\times$ 10^6 moves. The

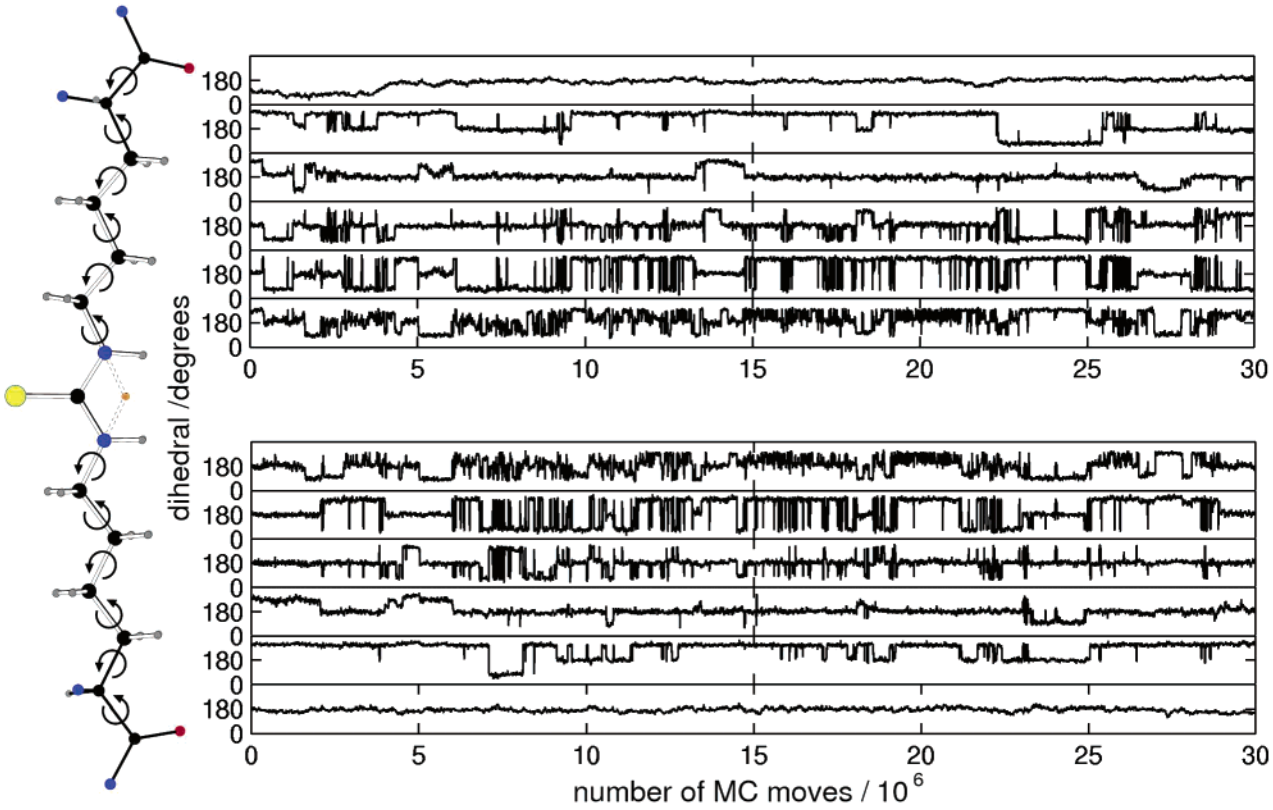

Figure 3. Trajectories of the 12 dihedral angles of the hydrocarbon chain for the host alone over 30 × 10^6 configurations. The chain illustrated on the left aligns with the particular dihedral angle. The molecule was drawn in ChemEdit.[65]

TABLE 1: Total Number and Frequency of Move Type for Each Simulation Protocol

	annealing complex	gas guest	explicit chloroform		continuum chloroform		
			guest	complex	guest	complex	host
			total number of moves (×10^6)				
equilibration	2.5	3	3	5	1	1	30
data collection		5	10	20	5	8	–
			move frequency (%)				
solvent			92	92			
volume			0.7	0.2			
guest	10	95	5.3	2.5	95	26	
host residue	90			3.3		36	67
large dihedral angle[a]		5	2		5	5	
conrot				1		20	20
flip				1		13	13

[a] Phe only.

structure was then minimized to 0 K using the Fletcher−Powell algorithm. Ten structures were generated for each host−guest complex. Simulations were also performed for the guest in the gas phase, explicit-solvent chloroform, and continuum-solvent chloroform, as well as for the complex in explicit-solvent chloroform and continuum-solvent chloroform, to allow the calculation of relative solvation and binding free energies for the two solvent models. Simulations of the host−guest complexes at the end points of the FEP simulations were extended from 8 × 10^6 to 10 × 10^6 configurations, and configurations sampled every 10 × 10^5 configurations yielded 100 representative structures for detailed analysis. Finally, a simulation was run of the host alone in continuum solvent to assess its flexibility without the guest.

Free Energy Calculations. Relative free energies, $\Delta\Delta G$, were calculated for the guest molecules using free energy perturbation (FEP)[55] and thermodynamic cycles.[56] The relative binding free energy, $\Delta\Delta G_{bind}$, of molecule B with respect to A was obtained by subtracting the free energy for perturbing from A to B in chloroform from the free energy for the same perturbation when the perturbed molecule is bound to the host. The relative solvation free energies, $\Delta\Delta G_{solv}$, were obtained similarly except that the free energy in the gas phase was subtracted from the free energy in chloroform. There are three types of mutation between molecules. The first is the cis−trans mutation, which changes the carbon of the methyl group capping the amide bond to an oxygen, with the three hydrogens becoming dummy atoms, each with bond lengths of 0.2 Å to the oxygen, and vice versa for the amide oxygen. Second, Gly-Ala mutations involve growing a hydrogen on the C_α carbon with three dummy atoms into a methyl group. Third, Ala-Phe mutations transform a methyl hydrogen with 10 dummy atoms into a phenyl group. In the gas-phase and unbound perturbations, calculations of only one stereochemistry were necessary because the environment is achiral. The complex in continuum chloroform requires the full set of perturbations involving these three mutations for both L and D enantiomers. However, for the complex in explicit solvent, the only perturbation attempted was

the Gly cis−trans mutation. Widely varing free energies dependent on the starting structure in explicit solvent made clear the insufficient sampling under these conditions. The number of windows for the gas-phase and continuum-chloroform perturbations ranged from 2 for Gly-Ala mutations to 10 for Phe cis−trans mutations. A larger number, 10−15 windows, was used for mutations in explicit solvent or for the complex. The spacing of the windows was chosen such that the surface area of the molecule increased linearly with λ, the scaling parameter, to make the mutation as smooth as possible. For each window, free energies were taken as the average of the forward and reverse free energies, with the difference between these data defining the error.

Glycine Measurements. The results of the simulation studies encouraged a reexamination of the Gly NMR data and the performing of additional experiments. The solution conformations of the 1:1 complex of M**12** and Gly (tetrabutylammonium salt) were investigated by detailed NMR experiments (at 294 and 318 K) using a 1:1 mixture of Gly and M**12** (4 mM) in $CDCl_3$. Spectral assignments were obtained using DQF−COSY[57] and TOCSY[58] experiments to define spin systems that were connected using data from ROESY[59] spectra.

Results

Explicit versus Implicit Solvent Free Energy Perturbations. The perturbations performed for the guest alone in explicit chloroform were relatively straightforward, with small errors and good thermodynamic closure (data not shown). However, this was not the case for perturbations of the complex in explicit chloroform. The free energies obtained for the Gly cis−trans mutation bound to the host were calculated in explicit chloroform for a range of different starting structures from the conformations generated in the simulated annealing. It was found that the free energies obtained differed considerably. In total, 17 different starting structures were tested and values of $\Delta\Delta G_{\text{bind}}^{\text{explicit}}$ for *trans*-Gly with respect to *cis*-Gly ranged from −1 to 4 kcal mol^{-1}, clearly an unacceptable variation. Structures that were annealed with a cis guest were responsible for the larger $\Delta\Delta G_{\text{bind}}^{\text{explicit}}$ values, implying that they were biased to stabilize the cis conformation. Further, an inspection of the sampling of the host revealed that virtually none of the dihedral angles in the hydrocarbon chain of the host underwent any significant conformational change so as to access other structures generated in the simulated annealing. The difficulty was that the large moves necessary to explore configuration space successfully had an almost negligible acceptance probability in explicit solvent.[60] Because good conformational sampling was achieved by the host in the gas phase, primarily through the conrot move, it was the explicit solvent that seemed directly responsible for the poor sampling observed. In light of these data, the free energy calculations were repeated using the GBSA continuum solvation model.

In Figure 3, the sampling of the dihedral angles in the hydrocarbon chain for just the host in GBSA continuum solvent over 30×10^6 configurations is reported. The improvement in sampling is significant, considering that none of these large conformational changes were observed for any dihedral angles in explicit solvent. The sampling is particularly good for the inner six dihedral angles, while the outer dihedral angles change less often. The outermost dihedral angle is not expected to vary much because the rotatable bond lies beyond the junction carbon from the hydrocarbon chain. One interesting observation for the host-only simulation was that the thiourea inverted so that the polar hydrogens of the thiourea moiety pointed away from the binding cavity, which itself underwent a moderate collapse. Upon inclusion of the guest in the cavity, acceptance probabilities were observed to decrease slightly, yet conformational transitions in the hydrocarbon chain still occurred at a significant rate.

TABLE 2: Free Energies of Solvation (kcal mol^{-1}) Relative to *trans*-Gly Comparing Explicit and Implicit Solvent

guest	$\Delta\Delta G_{\text{solv}}^{\text{explicit}}$	$\Delta\Delta G_{\text{solv}}^{\text{continuum}}$
cis-Gly	0.40	−0.35
trans-Gly	0	0
cis-Ala	−0.26	−0.37
trans-Ala	−0.34	0.09
cis-Phe	−3.23	−0.71
trans-Phe	−2.52	−0.52

To determine how the better sampling affects the free energy calculations, four different starting structures that gave $\Delta\Delta G_{\text{bind}}^{\text{explicit}}$ values of *cis*-Gly with respect to *trans*-Gly of −0.49, 0.77, 3.09, and 3.94 kcal mol^{-1} with a maximum error 0.8 kcal mol^{-1} were considered. In continuum solvent, the same four structures gave $\Delta\Delta G_{\text{bind}}^{\text{continuum}}$ values of 0.06, −0.23, −0.39, and 0.11 kcal mol^{-1}, respectively, with a maximum error of 0.45 kcal mol^{-1}. Clearly, within the bounds of error, these results are independent of starting structure. Thereafter, one of the structures generated from the *trans*-Gly simulated annealing was selected and was used as the starting structure for all windows and for all perturbations for the host in implicit solvent.

Before explicit solvent is abandoned for this system altogether, it is worthwhile comparing the relative free energies of solvation, $\Delta\Delta G_{\text{solv}}^{\text{explicit}}$ and $\Delta\Delta G_{\text{solv}}^{\text{continuum}}$ using *trans*-Gly as the reference. These values are listed in Table 2. The agreement between the two solvent models is fair for Gly and Ala. In these cases, the cis conformation has marginally greater stability in the continuum model than in the explicit model. However, it is clear that $\Delta\Delta G_{\text{solv}}$ does show some sensitivity to the solvent model for Phe. Phe is stabilized by ∼2 kcal mol^{-1} in explicit solvent relative to the continuum. This might be due to inaccurate GBSA parametrization, particularly for the aromatic ring that differentiates Phe from Ala. It has been noted[61] that the atomic solvation parameters for aryl rings should have a more negative dependence on SASA to compensate for the fact that cyclic structures, being more compact, have a smaller SASA. However, such a trend was not evident for aromatic molecules in the parametrization in this work. The discrepancy for Phe might also arise from the use of the continuum solvation model itself. Explicit chloroform molecules might be able to make multiple favorable and specific interactions with Phe that cannot be captured in a continuum solvation model. Curiously, *cis*-Phe is stabilized in explicit solvent relative to continuum solvent in opposition to the trend for Gly and Ala. These differences between solvation models might affect $\Delta\Delta G_{\text{solv}}$ calculations, particularly for the comparison of different amino acid derivatives, but they should cancel out to a large extent when stereoisomers or conformers are compared.

Relative Binding Free Energies. $\Delta\Delta G_{\text{bind}}^{\text{continuum}}$ energies were calculated for each amide bond conformation and sterochemistry of all three guests in continuum chloroform. Figure 4 summarizes the results. Also shown are the closures of thermodynamic cycles in the boxes. The closures range from 0.32 to 1.17 kcal mol^{-1}. The simulation errors range from 0.2 to 0.7 kcal mol^{-1} and tend to be worse for the cis-to-trans perturbations. The rather large errors provide an indication of the size of the configuration space sampled by the host−guest system. Nevertheless, the reassuring feature is that the errors

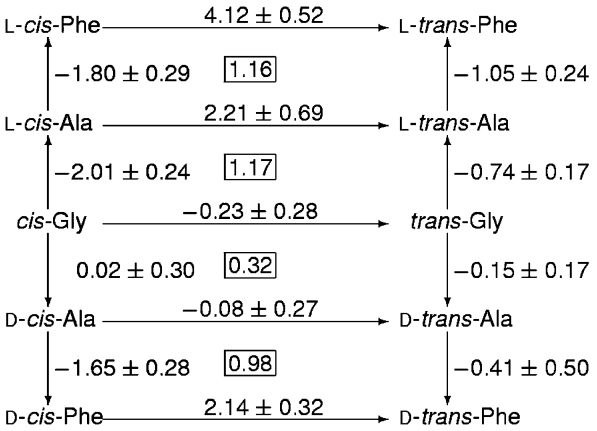

Figure 4. Relative free energies of binding (kcal mol^{-1}) calculated using FEP in continuum chloroform for the different stereochemistries and peptide conformations of each guest. The error is taken as the RMS total of the hysteresis of each window. The closure of each thermodynamic cycle is given in the box in the center.

TABLE 3: Relative Binding Free Energies (kcal mol^{-1}) from Simulation and Experiment

	$\Delta\Delta G_{bind}^{sim}$		
guest	cis	trans	$\Delta\Delta G_{bind}^{expt}$
Gly	0.23	0	0
L-Ala	−1.78	−0.74	0.81
D-Ala	0.25	−0.15	0.91
L-Phe	−3.58	−1.79	0.67
D-Phe	−1.40	−0.66	0.96

are sufficiently small compared to the size of the calculated free energies to allow meaningful qualitative deductions.

As anticipated by the experimental findings, a wealth of information is available addressing the two main points of study in this work regarding enantioselectivity and conformational stabilization. The first major point of interest is the stabilization of the cis conformation relative to the trans by ∼2−4 kcal mol^{-1} for L-Ala, L-Phe, and apparently D-Phe, but not Gly or D-Ala. The second major point of interest is the stronger binding by ∼2 kcal mol^{-1} of the L enantiomer compared to the D enantiomer for both Ala and Phe, particularly for the cis compounds. These results are in exact agreement with the major findings of the previously reported experiments.[1] A similar trend is seen in the potential energies of the complexes that are listed in Table S3 of the Supporting Information.

Concerning the selectivity of the host for a given side chain, there is more uncertainty. To observe more clearly $\Delta\Delta G_{bind}^{continuum}$ of each amino acid derivative, the data from Figure 4 can be used to construct a table of $\Delta\Delta G_{bind}^{continuum}$ values, with *trans*-Gly as the reference, by following the shortest path connecting the two molecules. If more than one shortest path was found, then the average of the free energies for each path was taken. Table 3 contains these data, together with the experimental values.[1] On first examination, the comparison does not appear very good. One source of error might lie in the continuum solvent model used ,which appears to give moderately different solvation free energies, particularly for Phe as discussed earlier. The stronger binding predicted for Phe might result from an underestimation of its solvation free energy found in continuum solvent. Another difficulty in making comparisons with experiment is that experiment is measuring binding of the guest to the host anywhere, either inside or outside the cavity, whereas simulation measures only binding inside the cavity. There is evidently some penalty in opening up the cavity, given that the collapsed host structure was more stable with no guest bound. Presumably, binding outside the cavity is likely to be less selective, so the overall experimental $\Delta\Delta G_{bind}^{expt}$ values are likely to be much less discriminating. Qualitatively, experiment and simulation agree in predicting stronger binding for L enantiomers than D and stronger binding for Phe over Ala, but they differ for Gly. Experiment indicates that Gly is the strongest-bound ligand, but this is not reproduced by the simulations. Finally, experiment did not indicate whether any cis stabilization was taking place in the amide bond. In light of this discrepancy and the fact that no stabilization of the cis conformation was predicted, the experimental NMR results for Gly were reexamined, and additional experiments were performed.

Glycine NMR Results. The 1:1 mixture of Gly with M**12** in chloroform presents a complicated proton NMR spectrum with broad peaks of nonintegral intensity. The temperature dependence of line widths (294−318 K) together with exchange peaks observed in the rotating-frame NOE (ROE), COSY, TOCSY, and NOESY spectra show that the macrocycle occurs in two forms with a concentration ratio of ∼5:1, neither coinciding with the free form. The ratio of the two forms of the complex indicates that they differ in free energy by ∼1.0 kcal mol^{-1}. The chemical shifts of the amide protons in the two forms differ by up to 1 ppm, showing slow exchange on the NMR time scale at 298 K, whereas other peaks are coalesced and show lifetime broadening. The signals of the biarylmethane unit of the macrocycle remain sharp. The amide protons for both forms of the macrocycle are shifted downfield with respect to the free macrocycle by up to 3 ppm, indicating strong hydrogen bonding, particularly in the major form. Signals for the Gly guest molecule are overlapped by signals from the macrocycle, Bu_4N, and residual water, with the exception of one broad signal at 5.8 ppm, which has a relative intensity slightly less than 1, and a sharper signal at 0.6 ppm, which has a relative intensity of 0.5. The temperature-dependent line widths (294−318 K, 87−44 Hz at 5.8 ppm, 22−15 Hz at 0.6 ppm) show that these signals are exchange broadened but coalesced. The signal at 5.8 ppm shows one very weak (because both signals are broad) TOCSY cross-peak connecting to 2.6 ppm, close to the residual water. No other signals were detected for this proton in any of the spectra. The signal at 0.6 ppm shows weak ROE connections to the aromatic protons and to three other weak signals at 1.6, 1.7, and 3.9 ppm. These three signals are connected to each other by TOCSY cross-peaks. The signals at 5.8 and 2.6 ppm can thus be assigned to the NH and CH_2 protons of *trans*-Gly in the major form of the complex, but it was not possible to find the acetyl methyl signals because of the line width and relative weakness of the NOE spectra. The signals at 1.6 and 1.7 ppm can be assigned to the CH_2 of *cis*-Gly in the minor form of the complex, the signal at 3.9 ppm is assigned as the NH, and the signal at 0.6 ppm is from the methyl group, which should then have a relative intensity of 0.6 (0.5 observed). These patterns can be observed despite the low intensity of the signals because the signals are sharper than those of *trans*-Gly in the major form.

Unlike the two forms of the macrocycle, there are no exchange peaks connecting the two forms of Gly. The exchange broadening in this case is likely to result from exchange with free Gly, with the sharpness of the minor component resulting from a very much lower concentration of the minor cis form of free Gly than exists for the major trans form. This agrees with a very weak signal (at 6.3 ppm with a relative intensity of ∼0.03) in pure Gly salt, which is presumably from the NH in a minor cis form. The NH signal for the major trans form is at

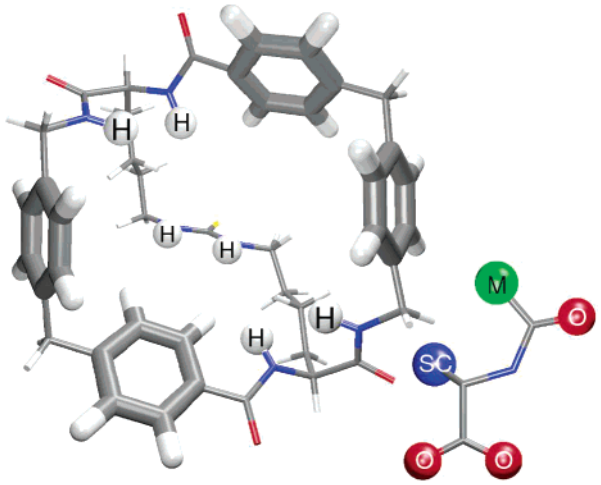

Figure 5. Schematic emphasizing the key components of the host and guest. The host has six polar hydrogens and four aryl groups lining the binding cavity. At each end, the benzamide polar hydrogen lies further from the cavity center and higher than that on *N*-benzylamide, and the benzamide aryl ring tilts horizontally while the *N*-benzylamide aryl ring tilts vertically. The guest has three oxygens, a methyl group (M), and a side chain (SC). The figure was made in VMD.[64]

6.7 ppm. Thus, the exchange peaks observed for the macrocycle involve dissociation of the complex followed by complexation of free Gly in its alternative form. Any exchange between bound and free macrocycle would result only in additional broadening of the coalesced peaks. The broadening of the Gly signals arises from exchange between the free and bound forms, with a negligible contribution from cis−trans isomerization. The Gly amide signals are shifted upfield by 1−3 ppm in both forms of the complex, as are the CH_2 signals of the major form and, presumably, the CH_2 and CH_3 signals of the minor form. This result can be attributed to a ring-current shift from binding in the interior of the macrocycle. In the latter case, ROE connections to all four aromatic signals support this view. Significantly, the ratio of major and minor forms of the complex (∼ 5:1) is different from that of free Gly salt (∼30:1), which implies a binding energy approximately 1.1 kcal mol^{-1} greater for the putative cis form than for the trans. Thus, the major form is assigned as the M**12** *trans*-Gly complex, and the minor form is assigned as the M**12** *cis*-Gly complex, with the guest bound to the interior of the macrocycle in both cases. That the dominant bound form of Gly is trans is in agreement with simulation. However, a moderate level of cis stabilization is still observed for Gly in the host by experiment, stronger than the negligible stabilization that simulation predicts. This small discrepancy might be due to insufficient sampling or taken into account by the estimated error, although another likely contribution is the solvation model. As discussed earlier, *cis*-Gly is predicted to be 0.8 kcal mol^{-1} less stable in explicit solvent than in continuum solvent relative to *trans*-Gly. Therefore, $\Delta\Delta G_{bind}$ of *cis*-Gly might be more favorable for the presumably more accurate hypothetical explicit-solvent calculation, assuming that the guests are less affected by the solvation model when complexed to the host.

Host and Guest Features. A detailed description of the host and guest is of assistance to help interpret the free energies. Figure 5 details the main structural elements of the host and guest. The host has C_2 symmetry, giving two equivalent binding modes for each guest. The binding cavity is surrounded by four moderately flexible aryl rings and lined by six polar hydrogens, four from amide groups and two from the thiourea moiety. A closer examination of these units in Figure 5 reveals that the face of the *N*-benzylamide aryl group is tilted almost perpendicularly to the mouth of the host cup, while the face of the benzamide aryl group lies flatter on the mouth of the cup. These different aryl orientations are designated as "vertical" and "horizontal", respectively. Thus, the rim of the cup is lowest over the amide hydrogens, intermediate for the benzamide aryl groups, and highest for the *N*-benzylamide aryl groups. This is also significant for the amide polar hydrogens of the host. *N*-Benzylamide amide hydrogens lie both further from the cavity center and higher than the benzamide hydrogens. For the guest, the negatively charged carboxylate group, a constant for all guests, can strongly hydrogen bond to any of the polar hydrogens in the host. The conformation of the amide bond determines whether the methyl group or the oxygen atom is presented to the rim of the guest. The guest's stereochemistry determines the direction of the side chain.

TABLE 4: Frequency of Binding Motif for Each Guest[a]

binding motif	L-Phe		L-Ala		Gly		D-Ala		D-Phe	
	cis	trans	cis	trans	cis	trans	cis	trans	cis	trans
1	**94**		**60**		7					
2	6	**86**	35	**71**	25			27		
3		14	5	29	**68**	28	**100**	**73**	**100**	**100**
4						**72**				

[a] Predominant motif for each guest denoted in bold face.

Observed Binding Motifs. Four main binding motifs were observed in the simulations of the 10 guests. These modes are illustrated in Figure 6 for representative guests, and the distribution of each guest in each motif is given in Table 4. The motifs are classified according to the number and type of hydrogen bonds (oxygen−polar hydrogen distance of ≤2.5 Å), which are listed in Figure 7. Hydrogen bonds in Figure 6 are marked as dashed purple lines. To distinguish the two ends of the C_2-symmetric host, "forward" denotes the side of the guest polar hydrogen and "backward" the other side. Here is a summary of the motif characteristics:

Motif 1 has the guest tipping forward moderately with the carbonyl oxygen hydrogen bonding to one or possibly two amide hydrogen atoms of the host. One oxygen of the guest carboxylate forms a double hydrogen bond to the thiourea, while the other forms a double hydrogen bond to the amide hydrogens of the host on the opposite side to the guest's nitrogen. It is this hydrogen bond that might be stabilizing the cis conformation as observed previously by Pernia et al.[1] The side chain typically lies over the horizontal benzamide and low polar hydrogen. Such a bonding pattern is observed by L-cis guests.

Motif 2 has both carboxylate oxygens hydrogen bonding to at least one amide hydrogen on both sides of the cavity and usually one or two thiourea hydrogens. Guests in this motif are fairly upright and are unable to form carbonyl−amide hydrogen bonds but are able to place their side chains over the benzamide unit so as to prevent tipping. Such a bonding pattern is found for L-trans guests.

Motif 3 has the guest tipping backward such that the base of the side chain descends partially into the cavity. One carboxylate oxygen forms a double hydrogen bond to the thiourea, while the other forms a double hydrogen bond to the amide group this time on the same side as the guest's nitrogen. This motif is accessible to most guests, and it is characteriztized by the absence of any additional hydrogen bonds that define the other three motifs. It is particularly common for the D isomers whose side chains are obstructed by the high polar hydrogen and vertical aryl ring of *N*-benzylamide. Because motif 3 appears

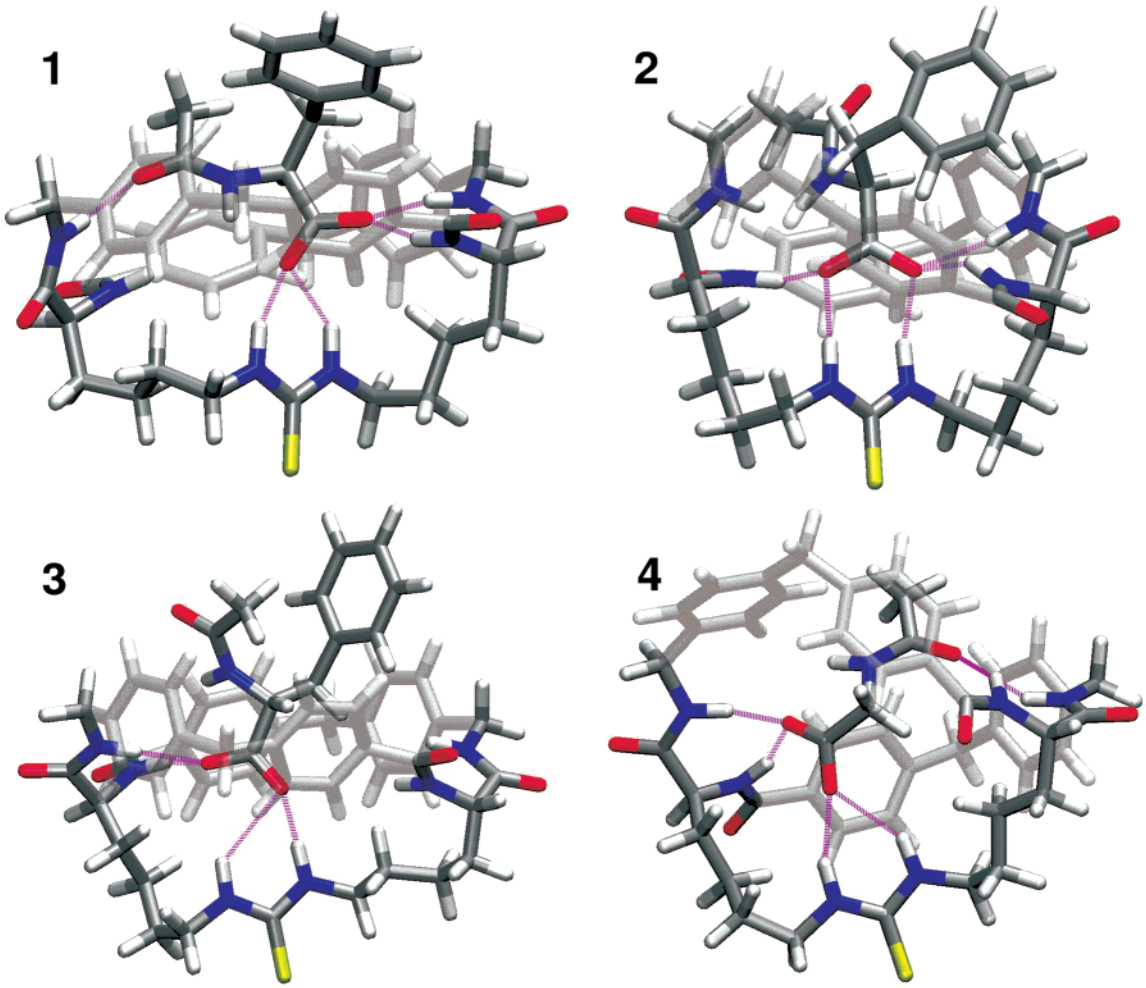

Figure 6. The four main binding motifs of the guest in the host shown for representative molecules. Residues making up the main ring of the host are rendered transparent for clarity. The preference of each guest for a given motif is listed in Table 4. The four guests shown here in order are L-*cis*-Phe, D-*trans*-Phe, D-*cis*-Phe, and *trans*-Gly. Typical hydrogen bonds are shown as dashed purple lines. The figure was made in VMD.[64]

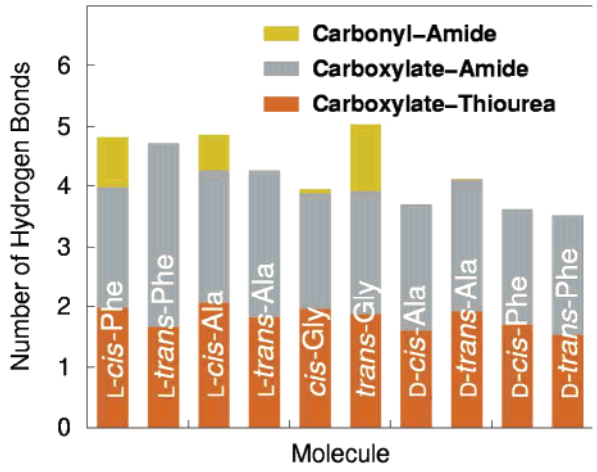

Figure 7. The three types of hydrogen bond between each guest and the host. All guests achieve two hydrogen bonds between their carboxylate group and the thiourea (bronze). The number of hydrogen bonds between the carboxylate and host amide groups ranges from two to three (silver). Only three of the guests manage a hydrogen bond between their acetyl carbonyl oxygen and the host amide (gold).

to be accessible to almost all guests including L-cis, this suggests that this motif is less stable, involves less complementarity, and is fairly nonselective. However, the fact that *cis*-Gly also prefers this motif suggests that a side chain is necessary for cis guests to form motif 1.

Motif 4 has the guest tipped backward completely onto its side with the carbonyl oxygen forming one or two hydrogen bonds to the amide hydrogens on the side of the cavity opposite to the guest polar hydrogen. This motif appears to be accessible only to Gly because it lacks a side chain.

The origin of these motifs can be understood with the simple schematic shown in Figure 8. This schematic illustrates how the four permutations of amide bond conformation and stereochemistry would fit ideally to the host. The ticks or crosses indicate these favorable or unfavorable interactions, respectively, with the host. The guest's amide oxygen forms a more favorable hydrogen bond with the better-placed high hydrogen. The methyl group or side chain, on the other hand, prefers to lie over the low hydrogen and the horizontal aryl ring as it experiences greater steric clash with the high hydrogen and vertical aryl ring. The model is consistent with experiment and the free energy calculations that L binds stronger than D and that cis binds more strongly than trans, although it does not predict alternative motifs in the cases of unfavorable interactions.

A more detailed analysis of how each individual guest binds to the host cavity is shown in Figure 9. The coloring is used to indicate the part of the guest making the contact: blue for the side chain, green for the methyl group, and red for the carbonyl oxygen. In this analysis, the host is divided into 11 residues, described in more detail in the figure caption. The figure was constructed by measuring the shortest distance between atoms in the guest unit and host residue; subtracting the Lennard-Jones

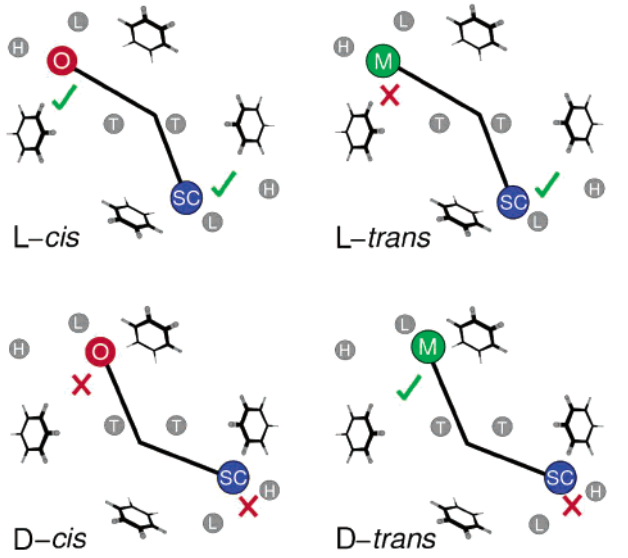

Figure 8. Binding motifs predicted for the four types of guest differing in cis or trans amide bond conformation and L or D stereochemistry. Illustrated are the four aryl rings, thiourea (T), and low amide (L) and high amide (H) hydrogens of the host and the guest side chain (SC) and either methyl group (M) or oxygen (O) depending on which interacts with the host according to the amide bond conformation. The ticks and crosses indicate favorable and unfavorable interactions, respectively. An oxygen makes a favorable hydrogen bond with the high hydrogen and a less favorable one with the low hydrogen. The methyl group and side chain show the opposite effect, requiring the extra space above the low hydrogen for the interaction to be favorable. The figure was made in ChemEdit.[65]

radii of the two atoms to produce a contact distance, r_C; and binning for all 100 stored structures into three bins: $r_C \leq 0.3$ Å, $0.3 < r_C \leq 0.6$ Å, and $r_C > 0.6$ Å. The closest bin is in bold color, the next in light color, and the third is gray. The coloring is according to atom type, and if a residue has two colors, the stronger is shown. In these structures, the guest is always oriented with its polar amide hydrogen facing the corner of the host containing the residue abbreviations suffixed by "1". The main points displayed in this figure are that the interacting parts are consistent with the motifs, larger guests have more interactions than smaller guests, and the better-fitting L guests have more interactions than the D guests.

Host Conformation. A particularly striking result from this study is that the host conformation was found to vary significantly for each guest, primarily in the hydrocarbon chain that determines both the shape of the cavity and the position of the thiourea moiety. For the 10 × 100 configurations saved, 92 unique conformations were counted for all 12 dihedral angles for all guests, binned according to whether they trans (*t*), gauche+ (*g*+), or gauche− (*g*−). Many more are likely as conformations were saved at only 1×10^5 intervals, but they represent the major ones. Of the 12 dihedral angles, four were found to differ significantly between the guests. In Figure 2, these dihedral angles are numbered 2, 4, 9, and 11. Of the 26 subconformations observed for these dihedral angles, six accounted for 77% of all 1000 structures. The populations of each guest in these six subconformations are reported in Table 5. Evidently, there are two dominant subconformations. The *g*−*ttg*− subconformation dominates for the guests at the L-*cis*-Phe end of the table, while the *tg*−*g*−*t* subconformation on the other hand is the most common for guests at the L-*trans*-Phe end. What distinguishes these two conformations is the width of the cavity. The average distance between the tertiary junction carbons is 10.6 Å for *g*−*ttg*− but only 9.5 Å for *tg*−*g*−*t*. This is in agreement with the previous observation that the L-trans guests prefer narrower cavities that compensate for their inability to form the cis-stabilizing hydrogen bond and allow their carboxylate group to span the binding cavity and hydrogen bond to amides at both ends as well as the thiourea group. Smaller guests also prefer the narrow host, presumably to increase their interactions with the host. On the other hand, L-cis guests require wider cavities to allow the guests to tip forward to form the cis-stabilizing hydrogen bond, while D guests need wide cavities to fit at all. An analysis of the rim of the cavity showed that this change in width comes about not by any dramatic single changes but by many small variations in angles and dihedral angles. Also interesting to note in Table 5 is the crossover region for the Gly, D-*trans*-Ala, and L-*cis*-Ala guests. Wide numbers

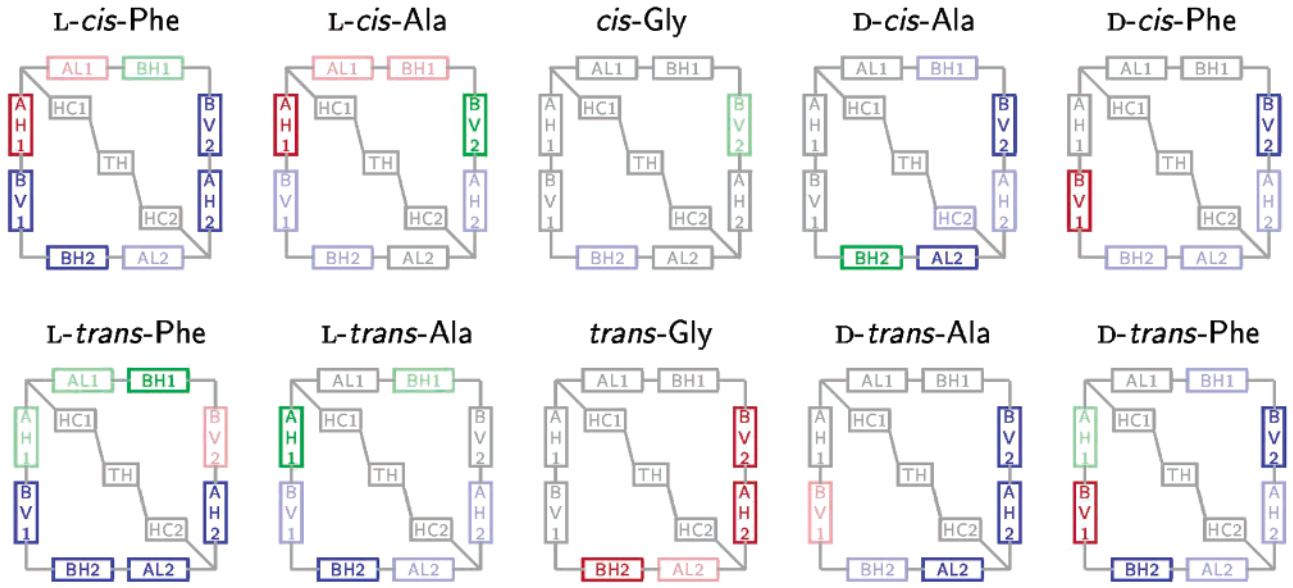

Figure 9. Close contacts with the host for all 10 guests. The color of the host indicates which part of the guest makes close contact. As in Figure 5, the side chain is represented by blue, the methyl group by green, and the oxygen by red. The darker color indicates a closer contact. The coding for the host parts is as follows. For the first part of the name, A, B, HC, and TH denote amide, benzene ring, hydrocarbon, and thiourea, respectively. For A, H and L denote high and low, and for B, H and V denote horizontal and vertical. The final 1 means that the guest's polar hydrogen lies on this side, and 2 indicates the other side.

TABLE 5: Populations of Each Host Subconformation for Each Guest[a]

conformation	total	D-Phe trans	L-Phe cis	D-Ala cis	D-Phe cis	Gly cis	Gly trans	D-Ala trans	L-Ala cis	L-Ala trans	L-Phe trans
g−*ttg*−	363	**92**	**65**	**64**	**51**	**27**	5	**17**	34	8	
g−*tg*−*t*	66		28		2	12		10	1	8	5
tg−*g*−*t*	168					11	1	7	**42**	**43**	**64**
tg−*tg*−	64			12		14	**13**	6	4	4	11
g−*g*+*tg*−	75	2	3	3	4	19	**13**	**17**	10	4	
g−*tg*+*g*−	53	1	4	2	11	13	1	6	3	5	7
% of total	77	95	100	81	68	96	33	72	94	72	87

[a] Predominant subconformation for each guest denoted in bold face.

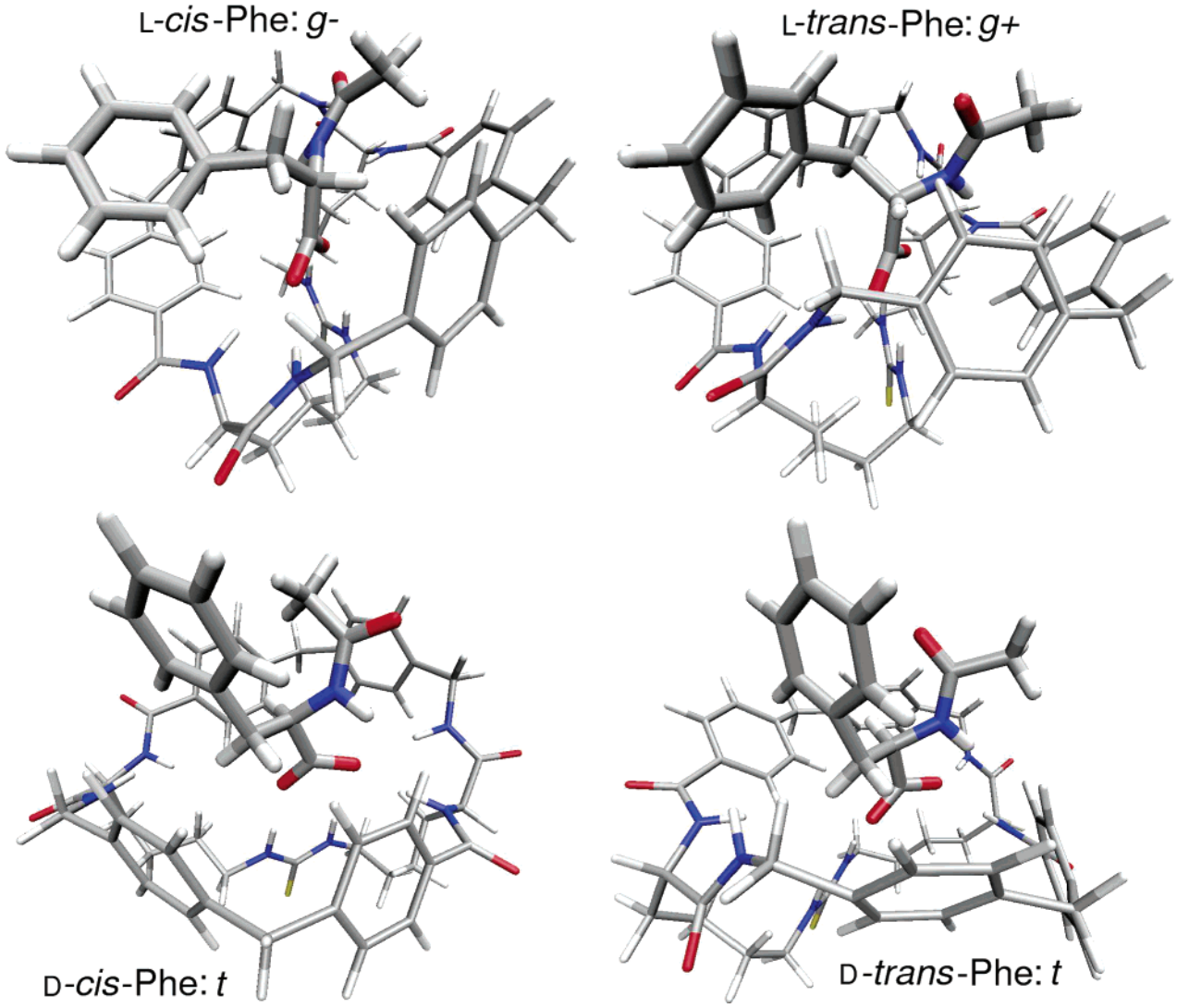

Figure 10. Complexes for the four different Phe guests. Depending on the stereochemistry and peptide conformation, the preferred value of the side-chain dihedral angle χ_1 varies from *g*− for L-*cis*-Phe, *g*+ for L-*trans*-Phe, to *t* for D-*cis*-Phe and D-*trans*-Phe. The figure was made in VMD.[64]

of conformations are associated with these guests, particularly *trans*-Gly, for which the dominant six subconformations make up only 33% of its structures. Presumably, this reflects a weaker interaction of their carboxylate groups with the thiourea moiety.

Guest Conformation. Apart from the amide bond, the only guest dihedral angle whose flexibility has substantial implications for ligand conformation is χ_1 on the side chain of Phe. The ϕ dihedral angle is held in place by an internal hydrogen bond between the polar hydrogen and a carboxylate oxygen, while the ψ dihedral angle flips the carboxyl group and methyl group. In this system, the preferred conformation of the side-chain dihedral angle χ_1 was found to depend on both stereochemistry and amide bond conformation. Figure 10 illustrates the typical binding geometries, and Table 6 lists the χ_1 conformation population of each Phe guest in the host and in solution. The preferred conformation in solution is *g*+ with the

TABLE 6: Conformational Distribution of χ_1 for the Phe Guests in Complex and Solution

	complex				solution	
	L-Phe		D-Phe		Phe	
conformation	cis	trans	cis	trans	cis	trans
g+	18	**66**			15	**97**
t	5	34	**99**	**99**	**78**	1
g−	**77**		1	1	7	2

[a] Predominant conformation for Phe guests denoted in bold face.

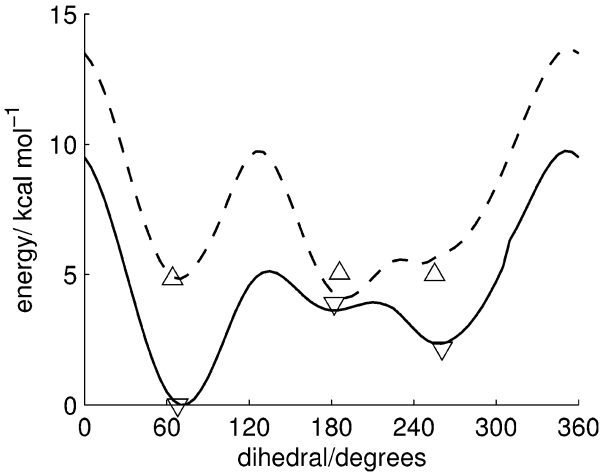

Figure 11. Dihedral angle energy profiles for the χ_1 dihedral angle in Phe for cis (dashed line) and trans (solid line) conformations. The triangles are quantum mechanical energies (HF/6-31G*) with the same minimum energy zero and with △ being cis and ▽ trans.

amide bond in the trans conformation. This is also the most stable pair of conformations in explicit solvent and in the gas phase. However, differences arise in the complex. The D guests adopts the *t* conformation almost exclusively, regardless of the amide bond conformation. This must be a result of the backward tipping that draws the side chain deeper into the cavity. This leaves the phenyl nowhere to go but straight up. L-*trans*-Phe prefers to adopt the *g*+ conformation, the most stable one in chloroform. However, there is still some inclination for it be *t*. L-*cis*-Phe, on the other hand, adopts the *g*− conformation, although it is also able to access the other two. The forward tipping must raise the phenyl group out of the host to a small degree, making all three conformations accessible. Therefore, steric effects and possible $\pi-\pi$ interactions appear able to stabilize less stable conformations.

To investigate this issue further, it is instructive to examine the dihedral angle energy profile of χ_1. This profile is plotted in Figure 11 together with ab initio minimum energies from HF/6-31G* calculations of the χ_1 dihedral angle for a gas-phase Phe molecule in the cis and trans conformations. The *g*+ is indeed the lowest-energy conformation when the amide bond dihedral angle is trans. However, what is more interesting is that the cis−trans energy difference is lower for the less stable *t* and *g*− χ_1 conformations. Although this ab initio calculation was not performed on the full complex, evidently, the steric constraint on the χ_1 dihedral angle is indirectly stabilizing the amide bond in the cis over the trans conformation for both L and D guests. This also explains the more favorable relative binding free energy in Figure 4 for D-*cis*-Phe over D-*trans*-Phe, as well as the stronger stabilization of the cis conformation for L-Phe than L-Ala. These results therefore imply that the conformational restrictions imposed on χ_1 by binding are directly affecting the cis−trans conformational energy difference of the nearby amide bond. It is well-known that the amide bond preceding a proline has a greater tendency to occur in the cis conformation.[62] This fact has traditionally been attributed to steric effects[18] and also, recently, electronic effects.[63] The results here for Phe are consistent with this interpretation. For both Phe and proline, the cis conformation is favored more when the χ_1 dihedral angle is destabilized away from *g*+ and indeed the χ_1 dihedral angle of proline is constrained to be *t*. The more general postulate arising from this result is that the cis conformation can be stabilized for the amide bond preceding not only proline but other amino acids that have larger, flexible side chains. This stabilization arises when the χ_1 dihedral angle is destabilized away from the *g*+ conformation. Further, this internal coupling of conformational states might prove to be a useful general principal worthy of exploitation in future design strategies.

Conclusion

The relative binding free energies to M**12** of the cis and trans conformers and L and D enantiomers of three amino acid derivatives, namely, Gly, Ala, and Phe, have been calculated using FEP and a rigorous MC sampling scheme in continuum chloroform modeled with the GBSA method. Necessary parameters that were missing in the OPLS force field were derived. Charges for the thiourea moiety were obtained using the REPD method, and parameters for dihedral angles for certain parts of the host and guest were calculated by fitting a three-term Fourier series to the ab initio dihedral energy profile. A sophisticated MC procedure was adopted to comprehensively sample the guest. Three aspects of this procedure were essential to achieve this sampling: First, and principally, the conrot move produced large conformational changes in the hydrocarbon segments that proved crucial in controling host shape. Second, the **Z** matrix of the host was constructed so that the most flexible parts of the rest of the host were directly sampled by internal dihedral angle moves. Third, the continuum nature of the GBSA solvent model, parametrized in this work for chloroform, allowed free passage for these large moves that would be otherwise obstructed in explicit solvent. This level of sampling was essential to reproduce the range of minimized structures produced in simulated annealing runs. More importantly, allowing the host to conform to the shape of each guest yielded free energies independent of starting structure. Explicit solvent, although more physically realistic, was unable to produce the desired level of sampling, consequently giving free energies that were sensitive to starting structure.

The free energies obtained reproduce the earlier experimental findings of the host selectivity for the L enantiomer and the stabilization of the L enantiomer's amide bond in the cis conformation relative to the trans. The enantioselectivity of ∼2 kcal mol^{-1} is rationalized by a more favorable placement of the side chain in the host. The cis conformation stabilization of ∼2−4 kcal mol^{-1} arises because of a hydrogen bond between the carbonyl oxygen of the guest and the amide rim of the host and, in the case of Phe, to a remote steric interaction between the side chain and the host. The dependence observed here of the amide bond's conformation on the side-chain conformation has profound implications for amino acid cis−trans equilibria and suggests a means by which this equilibrium might be influenced for residues other than proline. Good sampling in the simulation was found to be essential for capturing the structural rearrangements and yielding converged free energies for different guests. Sufficient sampling was not achieved using an explicit solvent model, and adopting a continuum solvent model was therefore required to accomplish this. Indeed, the different guests adopted quite different binding motifs, a reminder that binding modes between similar molecules are not necessarily transferable. Accompanying the different binding motifs are different host conformations, loosely characterized as either wide or narrow and controlled by the conformations of the hydrocarbon chain. A prediction was made that Gly bound in the trans conformation inside the host, quite different from the conformation of Ala and Phe. This prediction was tested experimentally and subsequently verified, lending further support to the simulation approach.

Acknowledgment. We are especially grateful to William L. Jorgensen for generously providing the MCPRO code, Doros N. Theodorou for the conrot code, and Jay W. Ponder for the GBSA code. Thanks are also due to Richard D. Taylor and Chistopher J. Woods for assistance with the GBSA and Poisson calculations and Ian D. Wall for helpful discussions. R.H.H. was supported by a Commonwealth Scholarship.

Supporting Information Available: Additional force field parameters for the system not in the OPLS force field. Free energy components for the 20 small molecules used in the GBSA parametrization. Potential energies of the complexes. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Pernia, G. J.; Kilburn, J. D.; Essex, J. W.; Mortishire-Smith, R. J.; Rowley, M. *J. Am. Chem. Soc.* **1996**, *118*, 10220−10227.
(2) Barboiu, M. D.; Hovnanians, N. D.; Luca, C.; Cot, L. *Tetrahedron* **1999**, *55*, 9221−9232.
(3) Jullian, V.; Shepherd, E.; Gelbrich, T.; Hursthouse, M. B.; Kilburn, J. D. *Tetrahedron Lett.* **2000**, *41*, 3963−3966.
(4) Wehner, M.; Schrader, T.; Finocchiaro, P.; Failla, S.; Consiglio, G. *Org. Lett.* **2000**, *2*, 605−608.
(5) Aït-Haddou, H.; Wiskur, S. L.; Lynch, V. M.; Anslyn, E. V. *J. Am. Chem. Soc.* **2001**, *123*, 11296−11297.
(6) Grawe, T.; Schrader, T.; Finocchiaro, P.; Consiglio, G.; Failla, S. *Org. Lett.* **2001**, *3*, 1597−1600.
(7) Sasaki, S.; Hashizume, A.; Citterio, D.; Fujii, E.; Suzuki, K. *Angew. Chem., Int. Ed.* **2002**, *41*, 3005−3007.
(8) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210−1250.
(9) Tsubaki, K.; Tanaka, H.; Morikawa, H.; Fuji, K. *Tetrahedron* **2003**, *59*, 3195−3199.
(10) Zhang, X. X.; Bradshaw, J. S.; Izatt, R. M. *Chem. Rev.* **1997**, *97*, 3313−3361.
(11) Lipkowitz, K. B. *Acc. Chem. Res.* **2000**, *33*, 555−562.
(12) Kim, H. J.; Asif, R.; Chung, D. S.; Hong, J. I. *Tetrahedron Lett.* **2003**, *44*, 4335−4338.
(13) Breccia, P.; Van Gool, M.; Pérez-Fernández, R.; Martín-Santamaría, S.; Gago, F.; Prados, P.; de Mendoza, J. *J. Am. Chem. Soc.* **2003**, *125*, 8270−8284.
(14) Carlson, H. A. *Curr. Opin. Chem. Biol.* **2002**, *6*, 447−452.
(15) Bursavich, M. G.; Rich, D. H. *J. Med. Chem.* **2002**, *45*, 541−558.
(16) Teodoro, M. L.; Kavraki, L. E. *Curr. Pharm. Des.* **2003**, *9*, 1635−1648.
(17) Teague, S. J. *Nat. Rev. Drug Discovery* **2003**, *2*, 527−541.
(18) Fischer, G. *Chem. Soc. Rev.* **2000**, *29*, 119−127.
(19) Dugave, C.; Demange, L. *Chem. Rev.* **2003**, *103*, 2475−2532.
(20) Hamilton, G. S.; Steiner, J. P. *J. Med. Chem.* **1998**, *41*, 5119−5143.
(21) Dugave, C. *Curr. Org. Chem.* **2002**, *6*, 1397−1431.
(22) Houk, K. N.; Leach, A. G.; Kim, S. P.; Zhang, X. *Angew. Chem., Int. Ed.* **2003**, *42*, 4872−4897.
(23) Kirchhoff, P. D.; Dutasta, J. P.; Collet, A.; McCammon, J. A. *J. Am. Chem. Soc.* **1999**, *121*, 381−390.
(24) van Hoorn, W. P.; Jorgensen, W. L. *J. Org. Chem.* **1999**, *64*, 7439−7444.
(25) Damodaran, K. V.; Banba, S.; Brooks, C. L. *J. Phys. Chem. B* **2001**, *105*, 9316−9322.
(26) Varady, J.; Wu, X. W.; Wang, S. M. *J. Phys. Chem. B* **2002**, *106*, 4863−4872.
(27) Chaumont, A.; Wipff, G. *J. Comput. Chem.* **2002**, *23*, 1532−1543.
(28) Bartoli, S.; Roelens, S. *J. Am. Chem. Soc.* **2002**, *124*, 8307−8315.
(29) Woods, C. J.; Camiolo, S.; Light, M. E.; Coles, S. J.; Hursthouse, M. B.; King, M. A.; Gale, P. A.; Essex, J. W. *J. Am. Chem. Soc.* **2002**, *124*, 8644−8652.
(30) Bea, I.; Jaime, C.; Kollman, P. *Theor. Chem. Acc.* **2002**, *108*, 286−292.
(31) Blas, J. R.; Marquez, M.; Sessler, J. L.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2002**, *124*, 12796−12805.
(32) Grabuleda, X.; Ivanov, P.; Jaime, C. *J. Phys. Chem. B* **2003**, *107*, 7582−7588.
(33) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1666−1671.
(34) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(35) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; T. Keith, G. A. P.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94*, revision d4; Gaussian, Inc.: Pittsburgh, PA, 1995.
(36) Henchman, R. H.; Essex, J. W. *J. Comput. Chem.* **1999**, *20*, 483−498.
(37) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(38) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049−1074.
(39) Henchman, R. H.; Essex, J. W. *J. Comput. Chem.* **1999**, *20*, 499−510.
(40) Jorgensen, W. L.; Briggs, J. M.; Contreras, M. L. *J. Phys. Chem.* **1990**, *94*, 1683−1686.
(41) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(42) Bashford, D.; Case, D. A. *Annu. Rev. Phys Chem.* **2000**, *51*, 129−152.
(43) Ponder, J. W. *TINKER*, version 3.6; Washington University: St. Louis, MO, 1998.
(44) Jorgensen, W. L. *MCPRO*, version 1.5; Yale University: New Haven, CT, 1997.
(45) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122−129.
(46) Warwicker, J.; Watson, H. C. *J. Mol. Biol.* **1982**, *157*, 671−679.
(47) Madura, J. D.; Davis, M. E.; Gilson, M. K.; Wade, R. C.; Luty, B. A.; McCammon, J. A. *Reviews in Computational Chemistry V*; VCH Publishers: New York, 1994; pp 229−267.
(48) Richmond, T. J. *J. Mol. Biol.* **1984**, *178*, 63−89.
(49) McDonald, N. A.; Carlson, H. A.; Jorgensen, W. L. *J. Phys. Org. Chem.* **1997**, *10*, 563−576.
(50) Giesen, D. J.; Chambers, C. C.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1997**, *101*, 2061−2069.
(51) Li, J. B.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1998**, *288*, 293−298.
(52) Potter, M. J.; Kirchhoff, P. D.; Carlson, H. A.; McCammon, J. A. *J. Comput. Chem.* **1999**, *20*, 956−970.
(53) Dodd, L. R.; Boone, T. D.; Theodorou, D. N. *Mol. Phys.* **1993**, *78*, 961−996.
(54) Mavrantzas, V. G.; Theodorou, D. N. *Macromolecules* **1998**, *31*, 6310−6332.
(55) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420−1426.
(56) Tembe, B. L.; McCammon, J. A. *Comput. Chem.* **1984**, *8*, 281−283.
(57) Rance, M.; Sorensen, O. W.; Bodenhausen, G.; Wagner, G.; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1983**, *117*, 479−485.
(58) Braunschweiler, L.; Ernst, R. R. *J. Magn. Reson.* **1983**, *53*, 521−528.
(59) Bothner-By, A. A.; Stephens, R. L.; Lee, J. M.; Warren, C. D.; Jeanloz, R. W. *J. Am. Chem. Soc.* **1984**, *106*, 811−813.
(60) Senderowitz, H.; Still, W. C. *J. Comput. Chem.* **1998**, *19*, 1736−1745.
(61) Sitkoff, D.; Bental, N.; Honig, B. *J. Phys. Chem.* **1996**, *100*, 2744−2752.
(62) Stewart, D. E.; Sarkar, A.; Wampler, J. E. *J. Mol. Biol.* **1990**, *214*, 253−260.
(63) Hinderaker, M. P.; Raines, R. T. *Protein Sci.* **2003**, *12*, 1188−1194.
(64) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics.* **1996**, *14*, 33−38.
(65) Lim, D.; Jorgensen, W. L. In *The Encyclopedia of Computational Chemistry*; von Ragué Schleyer, P., Ed.; Wiley: Athens, 1998; pp 3295−3302.