A Computational Study of the Role of Solvation Effects in Reverse Turn Formation in the Tetrapeptides APGD and APGN

Donald Bashford,* David A. Case,* Chris Choi, and Garry P. Gippert

Contribution from the Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037

Received October 7, 1996[⊗]

Abstract: The tetrapeptides APGD and APGN are known by NMR analysis to adopt reverse turn conformations to a significant degree in aqueous solution. We have carried out a 7.7 ns molecular dynamics simulation of Ace-APGD-NHMe in explicit water, and have analyzed the energetics of snapshots from this simulation in terms of a molecular mechanics energy function, estimates of solvation free energy based on numerical solutions of the Poisson−Boltzmann equation (in which the solvent is treated as a high-dielectric continuum), and an estimate of chain entropy effects derived from a systematic search procedure. In the unconstrained trajectory, 17 transitions occur between turn and extended conformers, suggesting that the free energy profile is nearly flat and that the simulation is moderately-well-equilibrated with respect to this transition; the turn population found is within the experimental range. The potential of mean force, constructed as the sum of solute force-field energies, continuum solvation, and hard-sphere chain entropy, agrees with that computed directly from the simulation to within 2 kcal/mol across the entire range of configurations sampled. The study has been extended to the tetrapeptide APGN by repeating the energetic analysis with an Asn side chain replacing the Asp in the APGD snapshots. A comparison of energetics with Asp and Asn side chains shows a complex interplay among vacuum electrostatic terms, dielectric screening terms, and solvation free energy terms such that the net effect of side chain substitution on turn formation is very small. Prospects for application of this sort of analysis to other peptide and protein conformational problems are discussed.

1. Introduction

Potential molecular mechanisms of protein folding have received widespread attention. The formation of local secondary structure by short-range interactions appears at an early stage,[1−3] but direct experimental identification of such transient early folding intermediates of native proteins is difficult since many polypeptide chains fold rapidly and completely into a final structure. The study of native-like secondary structure in peptide analogues of protein fragments can provide information about the ability of local forces to produce secondary structural intermediates in protein folding. Reverse turns have been postulated to play an important role in initiating the protein folding process and in directing the subsequent pathway of folding.[4,5] Some early attempts to find structures in peptide fragments in proteins were not successful[6−8] and led to a general conclusion that short peptides do not contain secondary structure in water solution, but advances in circular dichroism and NMR spectroscopy now provide a means to identify peptides that have distinct structural propensities in aqueous solution and to obtain direct experimental information on their preferred conformations.[9] This has led to the discovery of a number of peptides with a tendency to form nascent secondary structures such as β-turns and helices. Perhaps the simplest of these are a series of tetra- and pentapeptides found by Dyson *et al.* to form reverse β-turns.[10−12]

Here we focus on two tetrapeptides (APGD and APGN) that contain the core of one of the first peptide sequences for which strong secondary structural preferences were identified by solution NMR spectroscopy.[10] A characteristic pattern of cross-peaks in the ROESY spectrum (especially connecting protons in residues 2 and 4) along with the temperature dependence of the amide proton chemical shift in residue 4 led Dyson *et al.* to conclude that the conformational ensemble of these and similar peptides includes a significant fraction of β-turn-like conformers, with the turn population APGD being slightly higher than that in APGN. Since these peptides are so short and include a proline residue, they have a small number of dihedral degrees of freedom, and it can be hoped that their energetically accessible conformational space can be explored by molecular dynamics techniques even when the extra computational time is spent to provide an explicit account of aqueous solvation. For example, in APGD, the presence or absence of a turn hydrogen bond (between the carbonyl of residue 1 and the amide proton of residue 4) is determined by just three "soft" degrees of freedom: ψ of residue 2 and ϕ and ψ of residue 3, assuming that the proline ϕ value remains fixed near −60°. We present the results of a 7.7 ns simulation of Ace-APGD-NHMe in a box of explicit water molecules, concentrating our analysis on the question of conformational equilibrium between turn-like and extended conformers. Since there will still be many larger peptides and proteins for which it will remain impossible for some time to explore folding using molecular dynamics simulations with explicit water molecules, it is also of interest to have relatively simple methods of comparing the energetics of individual conformers in which solvent effects are handled by

[⊗] Abstract published in *Advance ACS Abstracts,* May 1, 1997.

(1) Karplus, M.; Weaver, D. *Protein Sci.* **1994,** *3,* 650−668.
(2) Baldwin, R. L. *Bioessays* **1994,** *16,* 207−210.
(3) Ptitsyn, O. B. *Protein Eng.* **1994,** *7,* 593−596.
(4) Lewis, P. N.; Momany, F. A.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1971,** *68,* 2293−2297.
(5) Zimmerman, S. S.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1977,** *74,* 4126−4129.
(6) Epand, R. M.; Scheraga, H. A. *Biochemistry* **1968,** *7,* 2864−2872.
(7) Taniguchi, H.; Anfinsen, C. B. *J. Biol. Chem* **1968,** *244,* 3864−3875.
(8) She, L. L.; Herman, J. *Biochemistry* **1972,** *11,* 1836−1841.
(9) Dyson, H. J.; Wright, P. E. *Annu. Rev. Phys. Chem.* **1996,** *47,* 369−395.
(10) Dyson, H. J.; Cross, K. J.; Houghten, H. J.; Wilson, I. A.; Wright, P. E.; Lerner, R. A. *Nature (London)* **1985,** *318,* 480−483.
(11) Dyson, H. J.; Merutka, G.; Waltho, J. P.; Lerner, R. A.; Wright, P. E. *J. Mol. Biol.* **1992,** *226,* 795−817.
(12) Dyson, H. J.; Sayre, J. R.; Merutka, G.; Shin, H.-C.; Lerner, R. A.; Wright, P. E. *J. Mol. Biol.* **1992,** *226,* 819−835.

S0002-7863(96)03516-0 CCC: $14.00

a less computationally demanding approximation. To this end, we have analyzed conformational "snapshots" from the simulation trajectory in the MEAD (macroscopic electrostatics with atomic detail) approximation in which the solvent region is modeled as a region of high dielectric constant. This analysis of sample conformers also allows us to bring a comparison of APGN with APGD into the study by a computational substitution, D → N, for the fourth side chain in the conformers from the APGD simulation.

Some aspects of the energetics of turn formation can be anticipated in advance. The main factor stabilizing turn conformations is expected to be a hydrogen bond involving backbone amide groups, similar to those seen in type I or type II β-turns in proteins,[13] along with the increased dispersion interactions that typically characterize compact structures. Solvation effects will generally oppose turn formation, since less of the polar surface of the peptide will be exposed to solvent. One of the goals of the current study is to explore the extent to which this sort of energy balance can be treated in a quantitative fashion that could allow similar analyses of much larger systems.

2. Methods

Three basic types of computation were carried out. The first was a 7.7 ns molecular dynamics simulation in explicit water, which is described in section 2.1. Second, 3000 snapshots from this simulation were then analyzed by an implicit solvation scheme that is outlined in sections 2.2 and 2.3. In this model, the free energy for a particular range of reaction coordinate (which we take to be the 1−4 hydrogen bond distance) can be written as

$$G_{\mathrm{conf}} = E_{\mathrm{vac}} + \Delta G_{\mathrm{elec}} + \Delta G_{\mathrm{nonp}} - TS_{\mathrm{conf}} \qquad (1)$$

The first three terms on the right-hand side are averages over all snapshots that have the hydrogen bond distance within a certain range or "bin"; ΔG_{elec} and ΔG_{nonp} estimate electrostatic and nonpolar solvation effects, and are free energies in the sense that they are implicitly averages over the solvent degrees of freedom. Finally, the last term of eq 1 is an estimate of the configurational entropy of the solute within a certain bin, described in section 2.4 below.

2.1. Molecular Dynamics Simulation. The simulation was carried out on the blocked peptide Ace-Ala-Pro-Gly-Asp-NHMe using the charged state of the aspartate side chain. The peptide was described using the AMBER/OPLS potential functions[14] and the water with the TIP3 function.[15] All calculations used the AMBER 4.0 suite of programs,[16] with temperature coupling as described by Berendsen *et al.*,[17] using a relaxation time τ_{T} of 0.25 ps. Bonds to hydrogen were constrained by the SHAKE procedure, and a time step of 0.5 fs was used for the dynamics integration. Nonbonded interactions were limited to an 8 Å interaction zone with a residue-based cutoff procedure, and the nonbonded list was updated every 20 steps.

The initial conformation was that of an idealized type II turn, with $(\phi, \psi) = (-60, 120)$ for the proline and (90, 0) for glycine, with the remainder of the backbone in an extended conformation. The peptide was immersed in a box of TIP3 water, and molecules overlapping the peptide were removed, leaving 419 water molecules in a periodic box. The system was slowly heated to 300 K over 20 ps, with harmonic constraints of 1 kcal/(mol Å^2) anchoring the peptide group to its original position. This was followed by 15 ps of constant pressure simulation, with a time constant of 0.4 ps for pressure coupling. This yielded box dimensions of 25.2 × 24.2 × 21.9 Å and a density of 0.99 g/cm^3. At this point the harmonic restraints were removed, and the simulation continued at constant volume for 7.7 ns. Coordinates were saved every 100 steps (0.05 ps) for subsequent analysis; most results presented below are based on an analysis of 3078 snapshots at 2.5 ps intervals.

2.2. Electrostatic Contribution to Solvation Free Energy. Even with the fairly extensive sampling of conformational space provided by a multi-nanosecond simulation of a short peptide, it can be difficult to understand the origins of solvation contributions to conformational preferences, both because a large number of small contributions are involved and because of the "noise" arising from fluctuations in the solvent−solvent potential energy contributions. As mentioned above, we have chosen here to carry out the energetic analysis using a continuum solvation model in which an estimate of the solvation contribution is added to a "vacuum" molecular mechanics energy.

We assume that the electrostatic potential is described by the macroscopic Poisson equation

$$\nabla\epsilon(r)\nabla\phi(r) = -4\pi\rho(r) \qquad (2)$$

where ϕ is the electrostatic potential due to the charge density, ρ, and ϵ is the dielectric constant which takes on the value ϵ_{m} in the peptide interior and $\epsilon_{\mathrm{s}} = 80$ in the solvent. For comparison to (and analysis of) conventional microscopic simulations, where the peptide solute is described by a nonpolarizable fixed charge distribution, setting ϵ_{m} to 1 is the most appropriate procedure, which we follow here. The boundary between the interior and exterior regions is defined by the contact and reentrant surfaces[18] of a 1.4 Å sphere rolling on the atomic radii of the atoms in the molecule. The atomic radii we have used are discussed elsewhere,[19] and are essentially the "Bondi" radii[20] with a slightly larger value of 1.2 Å used for hydrogen. This choice is similar to that of other investigations where (limited) investigations of adjusting radii have been attempted.[19,21,22] The Poisson equation is solved using a finite difference method using the MEAD programming suite.[23] For these calculations, a coarse cubic grid of dimensions 61^3 and 1.0 Å spacing was used to obtain the long-range behavior of the potential, followed by focusing to a finer 81^3 grid of spacing 0.25 Å for potentials closer to the atoms. The AMBER/OPLS charges were used to generate the potential.

For the sake of discussion, it is convenient to describe the electrostatic potentials in terms of the Green function of eq 2, $\Phi(\mathbf{r}_i,\mathbf{r}_j)$, which is defined as the potential produced at point $\mathbf{r}_i$ by a unit charge at point $\mathbf{r}_j$. In the interior region, the effect of the exterior medium can be separated out by writing

$$\Phi(\mathbf{r}_i,\mathbf{r}_j) = \frac{1}{\epsilon_{\mathrm{m}}|\mathbf{r}_i - \mathbf{r}_j|} + \Phi^*(\mathbf{r}_i,\mathbf{r}_j) \qquad (3)$$

where the first term is the direct Coulomb-like potential and the second term, Φ^*, arises due to the dielectric interface between the interior and exterior regions.

The process of taking a polar molecule from vacuum to solution can be broken into three hypothetical steps: First, all the charges in the molecule are reduced to zero; next, the resulting apolar molecule is transferred to solution—a step which has no electrostatic component (its cavity-forming or hydrophobic components are discussed later); finally, the molecular charges are restored to their original value. The electrostatic solvation energy is then the sum of the electrostatic work of charging the solute molecule in solution and discharging it in vacuum. In terms of the MEAD approximation, the incremental work, δW, to add increments of charge, δq_i, to a set of atoms indexed by i in an environment whose electrostatic properties are characterized by the Green function, Φ, is

(13) Richardson, J. S. *Adv. Protein Chem.* **1981,** *34,* 167−339.
(14) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988,** *110,* 1657−1671.
(15) Jorgensen, W. L. *J. Chem. Phys.* **1982,** *77,* 4156.
(16) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995,** *91,* 1−41.
(17) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984,** *81,* 3684−3690.
(18) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977,** *6,* 151−176.
(19) Bashford, D.; Case, D. A.; Dalvit, C.; Tennant, L.; Wright, P. E. *Biochemistry* **1993,** *32,* 8045−8056.
(20) Bondi, A. *J. Chem. Phys.* **1964,** *64,* 441.
(21) Lim, C.; Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991,** *95,* 5610−5620.
(22) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994,** *98,* 1978−1988.
(23) Bashford, D.; Gerwert, K. *J Mol. Biol.* **1992,** *224,* 473−486.

$$\delta W = \sum_i \delta q_i \, \phi(\mathbf{r}_i) = \sum_{i,j} \delta q_i q_j \, \Phi(\mathbf{r}_i, \mathbf{r}_j) \tag{4}$$

Integrating eq 4 over the q_i from 0 to the full charges, Q_i, gives the work of charging as

$$W = (1/2) \sum_{i,j} Q_i Q_j \Phi(r_i, r_j) \tag{5}$$

The subtraction to obtain ΔG_{Born} is then

$$\Delta G_{\text{Born}} = W_{\text{solv}} - W_{\text{vac}} = (1/2) \sum_{i,j} Q_i Q_j [\Phi^*_{\text{solv}}(\mathbf{r}_i, \mathbf{r}_j) - \Phi^*_{\text{vac}}(\mathbf{r}_i, \mathbf{r}_j)] \tag{6}$$

where the subscripts solv and vac refer to solvent and vacuum calculations, respectively, which differ only in that ϵ_w has been set to 80 in the former and 1.0 in the latter. Note that Φ^* rather than Φ appears in eq 7 since we have made use of eq 2 and the constant value of the interior dielectric constant, ϵ_m, to cancel the Coulomb terms. Computer programs for finite difference electrostatics, such as those in the MEAD suite, do not directly calculate Green functions, but rather electrostatic potentials, so for practical calculations, eq 6 is written as

$$\Delta G_{\text{Born}} = (1/2) \sum_i Q_i [\phi_{\text{solv}}(\mathbf{r}_i) - \phi_{\text{vac}}(\mathbf{r}_i)] \tag{7}$$

where ϕ is the electrostatic potential calculated for the molecule with its full set of charges Q_i.

2.3. Nonpolar Contributions to Solvation Free Energies. There are a variety of approaches to estimating nonpolar contributions to solvation free energies, most based on proportionality to the solvent-accessible (SA) surface area.[22,24] For the conformational changes of a tetrapeptide considered here, from turns to more extended forms, the changes in surface area are relatively modest (less than 100 Å^2), so that the nonpolar contribution to unfolding potentials of mean force is minor, and the details of how the nonpolar contribution is estimated are relatively unimportant. We have chosen to use the procedure arising from the PARSE algorithm,[22] which has been shown to give a good account of hydration free energies for small molecules, and which was parametrized in combination with an electrostatic model that is similar to the one used here. In this model, the nonpolar contribution to the solvation free energy is given by

$$\Delta G_{\text{nonp}} = \gamma(\text{SA}) \tag{8}$$

where $\gamma = 0.0054$ kcal/(mol Å^2) and SA is the solvent-accessible surface computed with a probe radius of 1.4 Å and the atomic radii listed above. This value of γ was designed to be appropriate for analysis of gas-to-water transfer free energies, which is the transfer we are considering here.

A variety of other methods have attempted to include both polar and nonpolar contributions to solvation into a surface-area term, generally by ascribing different coefficients to different types of atoms.[24−28] These theories do not attempt to compute directly any electrostatic interactions, but favor the exposure of polar groups by having negative values of γ for certain classes of atoms. We used the ASC program of Eisenhaber and Argos[29] to carry out the calculations. Below, we report tests of several of these models for the β-turn formation considered here.

2.4. Peptide Configurational Entropy. Our estimates of configurational entropy are based on a systematic search for "allowed" conformations in a model where only the torsion angles are allowed to vary (bond lengths and angles are kept at fixed positions) and where a set of atomic hard-sphere radii discriminate allowed from disallowed conformations. Conformers are placed in bins defined by ranges of values of our reaction coordinate, which is the distance between the carbonyl oxygen of residue 1 and the amide hydrogen of residue 4. Calculations were carried out using the DTAGS (distributed torsion angle grid search) programs described elsewhere,[30,31] with grid spacings of 30, 24, 20, and 18° for backbone degrees of freedom within the four-residue peptide, and rigid (*trans*) peptide bonds. The proline ring conformation was modeled using a pseudorotation scheme, with uniform sampling of the pseudorotation phase angle at the grid spacings given above. Other side chain degrees of freedom in the peptide were neglected in the search, as are the corresponding bump-checks involving side chain atoms beyond Cβ, with the result that the number of allowed backbone conformations produced by the systematic search may be slightly overestimated.

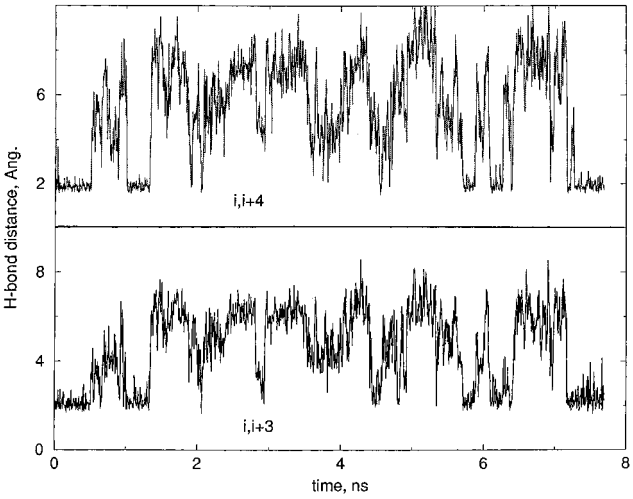

Figure 1. Hydrogen bond distance between the carbonyl oxygen of residue 1 and the amide proton of residues 4 (bottom) and 5 (top).

Hard-sphere radii that are appropriate for this sort of calculation are smaller than typical van der Waals radii, since no overlaps are allowed at all. The values we used[30,32] have been calibrated to approximate the shapes of dipeptide maps calculated with more elaborate molecular mechanics potentials: 0.95 Å for hydrogen, 1.25 Å for nitrogen, 1.35 Å for aromatic carbons, 1.40 Å for all other carbons, and 1.20 Å for oxygen. More details of the computational procedure are given elsewhere.[30]

Since, in a hard-sphere model, all allowed conformations are of equal energy, only entropy distinguishes one range of conformers from another, so that the configurational entropy can be extracted from the logarithm of the number of allowed states in each bin,

$$S_{\text{conf},i} = -k \ln(\Omega_i) \tag{9}$$

where $S_{\text{conf},i}$ is the configurational entropy of the ith bin, and Ω_i is the number of allowed states found with the appropriate range of reaction coordinate (hydrogen bond distance).

3. Results and Discussion

3.1. Molecular Dynamics Trajectory. As we mentioned above, one of the attractive features of the APGD peptide is the relatively small number of floppy degrees of freedom, so that to a first approximation the overall conformation can be characterized by one or two distances that describe how closely the two ends of the peptide come together. Figure 1 plots the length of potential hydrogen bonds between residues 1 and 5

(24) Eisenberg, D.; McLachlan, A. D. *Nature* **1986,** *319,* 199−203.
(25) Wesson, L.; Eisenberg, D. *Protein Sci.* **1992,** *1,* 227−235.
(26) Ooi, T.; Oobatake, M.; Ne'methy, G.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987,** *84,* 3086−3090.
(27) Vila, J.; Williams, R. L.; Vasquez, M.; Scheraga, H. A. *Proteins* **1991,** *10,* 199−218.
(28) Juffer, A. H.; Eisenhaber, F.; Hubbard, S. J.; Walther, D.; Argos, P. *Protein Sci.* **1995,** *4,* 2499−2509.
(29) Eisenhaber, F.; Argos, P. *J. Comput. Chem.* **1993,** *14,* 1272−1280.
(30) Gippert, G. P. New computational methods for 3D NMR data analysis and protein structure determination in high-dimensional internal coordinate space. Ph.D. Thesis, The Scripps Research Institute, 1995.
(31) Gippert, G. P.; Wright, P. E.; Case, D. A. Recursive torsion angle grid search in high dimensions: A systematic approach to NMR structure determination. Submitted for publication.
(32) Iijima, H.; Dunbar, J. B., Jr.; Marshall, Garland R. *Proteins: Struct., Funct., Genet.* **1987,** *2,* 330−339.

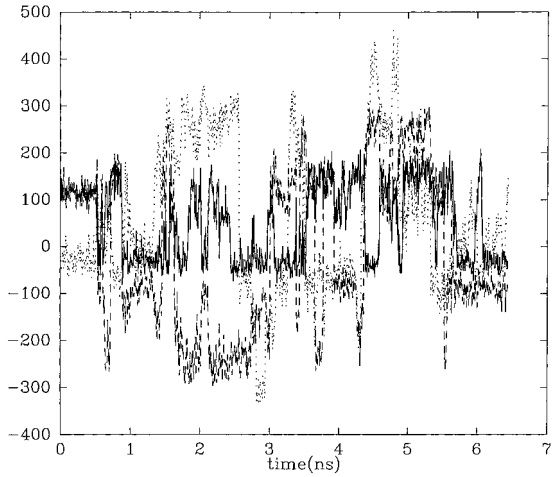

Figure 2. Backbone dihedral angles as a function of time: Solid line, ψ of residue 2 (Pro); dashed and dotted lines, ϕ and ψ of residue 3 (Gly), respectively.

(top) and 1 and 4 (bottom) as a function of time. There are clear time segments where one or both hydrogen bonds are formed (with O−H distances near 2 Å), separated by segments with longer distances. In this simple classification, there are nine "unfolding" transitions (near 0.7, 1.4, 2.0, 2.5, 3.9, 4.3, 5.8, 6.1, and 6.3 ns) and eight "folding" transitions (near 1.0, 1.9, 2.4, 3.3, 4.2, 5.7, 5.9, and 7.0 ns). It is clear that in this simulation the free energy barrier for forming and breaking internal hydrogen bonds is quite low, so that transitions occur about every 1/2 ns; a more quantitative analysis of the potential of mean force will be given below. This flexibility is in rough accord with simulations by Tobias *et al*. on shorter models for reverse-turn formation,[33] and with the rate of similar conformational transitions seen in solvated simulations of YPGDV,[34] which has a similar sequence to the one considered here. The time course of the backbone dihedral angles is shown in Figure 2, which shows one to two dozen major transitions from one favored area to another for each of the dihedral angles shown; the value of ϕ for residue 2 (Pro) stays near −60°, and is not shown.

It is clear from Figure 1 that simultaneous formation of $i \rightarrow i + 3$ and $i \rightarrow i + 4$ hydrogen bonds is a prominent feature of the dynamics. This "bifurcated" hydrogen bonding pattern allows an especially compact structure to be formed, as illustrated in the snapshot shown in Figure 3. The importance of structures like this as transient intermediates in the formation and dissolution of turns and helices has been noted in several earlier studies,[35−39] and appears to have an important effect on the way in which short peptides fold and unfold.[40,41]

3.2. Energetics of Electrostatic Solvation. In order to analyze the energetic components that might distinguish folded

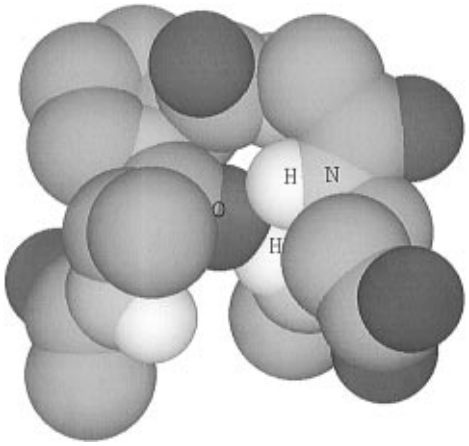

Figure 3. Space-filling model of the conformation at $t = 0.3$ ns, showing amide protons from both residues 4 and 5 forming a hydrogen bond to the carbonyl oxygen of residue 1.

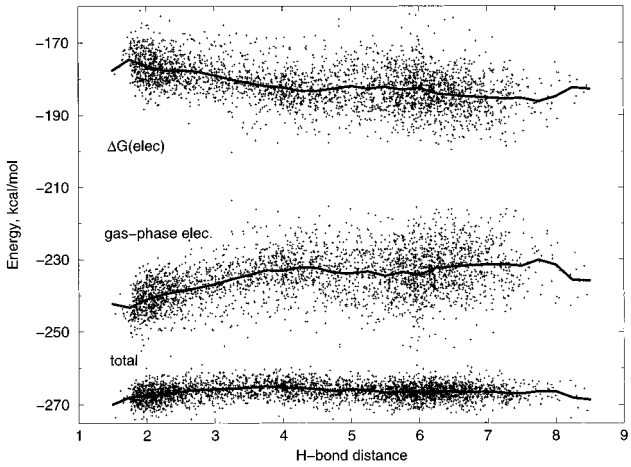

Figure 4. OPLS electrostatic energy (top), the electrostatic contribution to the solvation free energy (middle), and their sum (bottom) (+). Values are for 3078 snapshots selected at 2.5 ps intervals from the molecular dynamics trajectory. Solid lines show binned averages, using 0.25 Å bins. For plotting purposes, the top curve has been shifted down by 50 kcal/mol, and the middle curve down by 100 kcal/mol.

from unfolded structures, we selected 3078 snapshots from the trajectory, taken every 2.5 ps. For each snapshot we computed three quantities: the AMBER/OPLS force-field energy for the peptide (which we call the *gas-phase* energy), the electrostatic component of the solvation free energy, estimated by the finite-difference Poisson−Boltzmann method outlined above, and the nonpolar contribution to the solvation energy, estimated from the solvent-accessible surface areas.

Figure 4 shows the gas-phase electrostatic energy and the electrostatic part of the solvation free energy (ΔG_{elec}) plotted against the value of the "turn" hydrogen bond distance, *i.e.*, the distance from the oxygen atom of residue 1 to the amide proton of residue 4. The solid curve shows the averages in 0.25 Å wide bins. The significant scatter about the averages reflects the fact that there are many detailed ways of making a conformation with a particular H bond distance, and that these have a spread in energy that is as large or larger than the conformational energy differences we are trying to track. Nevertheless, there are clear trends in the average energies shown in Figure 4. As in our earlier studies of solvent effects on hydrogen bond formation,[42] the gas-phase energetics favor hydrogen bond formation (by 6−7 kcal/mol), an effect that

(33) Tobias, D. J.; Sneddon, S. F.; Brooks, C. L., III. *J. Mol. Biol.* **1990,** *216,* 783−796.
(34) Tobias, D. J.; Mertz, J. E.; Brooks, C. L., III. *Biochemistry* **1991,** *30,* 6054−6058.
(35) Wright, P. E.; Dyson, H. J.; Feher, V. A.; Tennant, L. L.; Waltho, J. P.; Lerner, R. A.; Case, D. A. In *Frontiers of NMR in Molecular Biology*; Live, D., Armitage, I. M., Patel, D., Eds.; Alan R. Liss: New York, 1990; pp 1−13.
(36) Soman, K. V.; Karimi, A.; Case, D. A. *Biopolymers* **1991,** *31,* 1351−1361.
(37) Soman, K. V.; Karimi, A.; Case, D. A. *Biopolymers* **1993,** *33,* 1567−1580.
(38) Sundaralingham, M.; Sekharudu, Y. C. *Science* **1989,** *244,* 1333−1337.
(39) Tirado-Rives, J.; Jorgensen, W. L. *Biochemistry* **1991,** *30,* 3864−3871.
(40) Choi, C.; Elber, R. *J. Chem. Phys.* **1991,** *94,* 751−760.
(41) Brooks, C. L., III; Case, D. A. *Chem. Rev.* **1993,** *93,* 2487−2502.
(42) Osapay, K.; Young, W.; Bashford, D.; Brooks, C. L., III; Case, D. A. *J. Phys. Chem.* **1996,** *100,* 2698−2705.

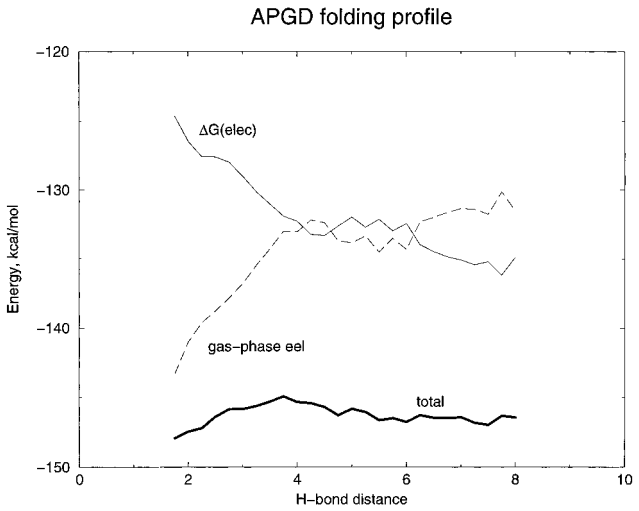

Figure 5. Average values (for 0.25 Å bins) for the electrostatic portion of the gas-phase energy, the solvation free energy, and their sum. For plotting purposes, the bottom curve has been shifted upward by 120 kcal/mol.

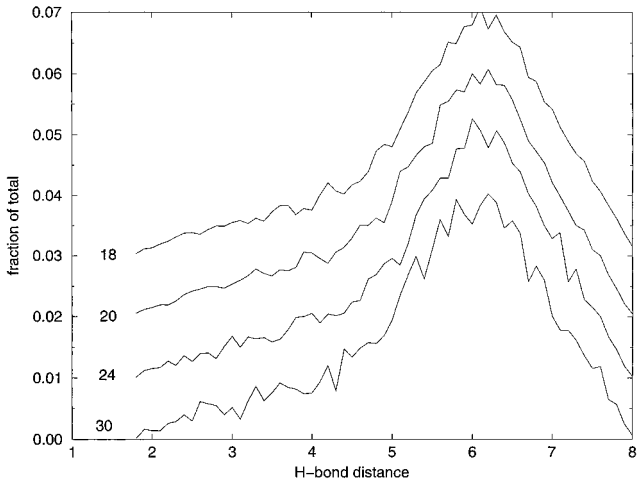

Figure 6. Fractional populations (in 0.1 Å bins) for systematic searches at 30, 24, 20, and 18° grid spacings. Each successive plot is shifted upward by 0.01.

mostly arises from the favorable electrostatic interactions that arise when the amide hydrogen (with a positive partial charge) comes in close contact with the carbonyl oxygen bearing a negative partial charge. A principal effect of solvation is to screen these favorable electrostatic interactions, so that the solvation free energy profile disfavors hydrogen bond formation. This cancellation is illustrated in the bottom part of Figure 4, which shows that the scatter about the sum of these electrostatic energies is significantly less than the scatter of either part individually: standard deviations about the mean for the gas phase and solvation energies shown are 6.6 and 5.8 kcal/mol, respectively, whereas their sum has a standard deviation of 2.3 kcal/mol.

A more detailed look at the electrostatic interactions is given in Figure 5, which plots the binned averages for the electrostatic part of the gas-phase energy, the electrostatic component of the solvation free energy, and their sum. It is clear that there is a close mirror symmetry in the first two curves, so that for most distances changes in the gas-phase electrostatic interactions are almost completely canceled out by opposite changes in the electrostatic contribution to the solvation free energy. This behavior is expected in a high-dielectric solvent for conformations where most of the charged and polar groups are well-exposed to solvent. The compensation of these two terms is expected to break down somewhat at short distances, where some of the polar groups are slightly sequestered from solvent, and the favorable electrostatic interactions are not completely shielded. This can happen to only a slight extent in the small peptide considered here, but similar effects are expected to be considerably more important in large peptides and proteins, where burial of polar groups can be much more complete.[42,43]

3.3. Solute Configurational Entropy. The profiles presented in Figures 4 and 5 cannot be directly compared to the potential of mean force determined from the microscopic simulation because the configurational or chain entropy of the solute peptide is not included in this analysis. The solvation free energy estimates correspond to carrying out a Boltzmann average over the solvent degrees of freedom, but the solute degrees of freedom have been frozen at the snapshot configurations.

Estimates of the loss of configurational entropy upon formation of secondary structure form an important part of theories of protein folding, and a variety of attempts have been made to provide a theoretical or empirical framework for these effects.[44−46] At best, however, these approaches provide estimates for the entropy differences at the end points of a folding/unfolding transition, and would not provide information about entropies at intermediate stages along the folding profile. The approach we have taken here, as outlined above, uses a hard-sphere model for peptide conformations, in which conformations are either allowed or disallowed. The model uses hard-sphere radii calibrated to appropriate dihedral-angle distributions for dipeptide fragments. A systematic search is used to enumerate all conformations on a discrete torsion-angle grid. Since all of the energies of the allowed states are the same, all of the thermodynamics in this model are entropic, and an entropy profile can be obtained by taking the logarithm of the observed probability distributions. We assume that entropic effects in this model system are close enough to those of the AMBER/OPLS solvated peptide that the former can be used to estimate the latter.

Torsion-angle grid searches were carried out, as described in the Methods, with grid spacings of 30, 24, 20, and 18°. Because of the combinatorial nature of the problem, the number of allowed states grows quickly as the torsion increment decreases: the number of allowed states at these grid spacings was 3, 19, 55, and 235 million, respectively. Nevertheless, the probability profiles (shown in Figure 6) are nearly the same for each calculation, suggesting that the gross distribution of allowed states (as a function of the hydrogen bond distance) is roughly converged, even though the details of the distribution will continue to change slightly as the grid spacing becomes smaller. Since the entropy is related to the logarithm of the probability distributions shown in Figure 6, the estimated uncertainty in the entropy curve due to grid spacing is less than 0.3 eu, or 0.1 kcal/mol, at $T = 300$ K. This is less than the uncertainty arising from the replacement of the AMBER/OPLS potential with a hard-sphere model (see below).

3.4. Overall Folding Profile. Figure 7 summarizes the conclusions from these calculations, showing the individual averages described above, along with their sum; these are compared to the potential of mean force derived directly from the solvated dynamics trajectory as $-kT \ln(\rho)$, where ρ

(43) Honig, B.; Nicholls, A. *Science* **1995,** *268,* 1144−1149.
(44) Sternberg, M. J. E.; Chickos, J. S. *Protein Eng.* **1994,** *7,* 149−155.
(45) Lee, K. H.; Xie, D.; Freire, E.; Amzel, L. M. *Proteins: Struct., Funct. Genet.* **1994,** *20,* 68−84.
(46) Saven, J. G.; Wolynes, P. G. *J. Mol. Biol.* **1996,** *257,* 199−216.

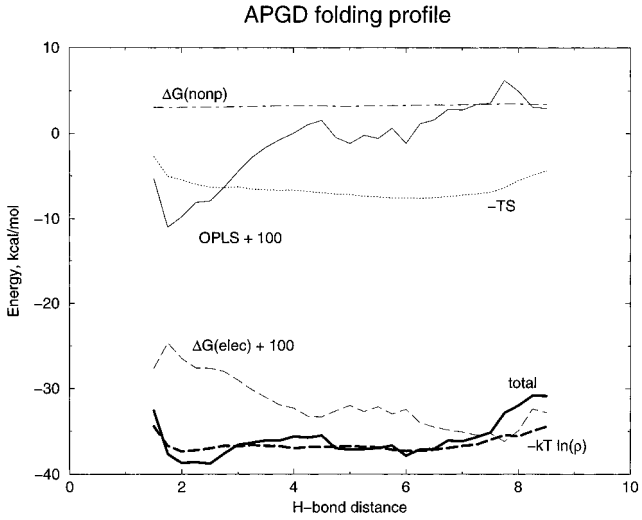

Figure 7. Binned averages for gas-phase energies (labeled "OPLS", and shifted upward by 100 kcal/mol), the electrostatic component of the solvation free energy (ΔG_{elec}, also shifted by 100 kcal/mol), the nonpolar contribution to solvation free energy (ΔG_{nonp}), the solute configurational entropy ($-TS$), and their sum ("total", shifted by 200 kcal/mol). The heavy dashed line shows the potential of mean force derived from the data in Figure 1.

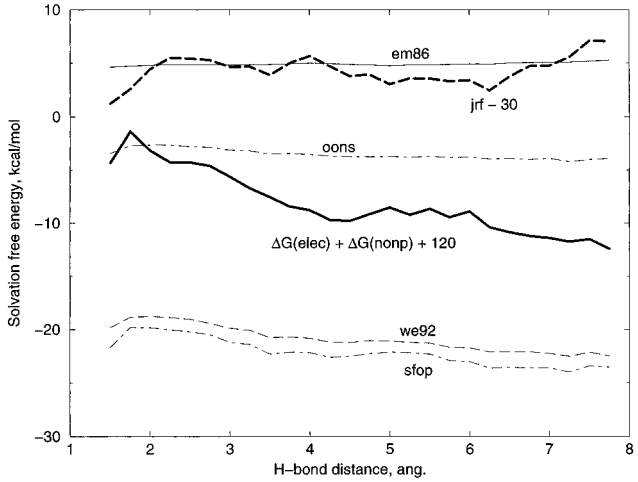

Figure 8. Same as Figure 5, comparing different solvation models. Key "em86" from Eisenberg and McLachlan;[24] "we92" from Wesson and Eisenberg;[25] "oons" from Ooi *et al.*;[26] "jrf" and "sfop" from Vila *et al.*[27] For convenience in plotting, the jrf values have been shifted downward by 30 kcal/mol, and the continuum electrostatic results have been shifted upward by 120 kcal/mol.

represents that fraction of time spent at each value of the H bond distance, evaluated in 0.5 Å bins. Many of the basic principles described above are illustrated in this figure. As expected, the gas-phase energetics strongly favors turn hydrogen bond formation, with an average energy nearly 10 kcal/mol higher for the extended conformations than for the most optimal hydrogen bonded configurations. This difference includes not only the intrinsic hydrogen bond strength, but also the more favorable dispersion terms that stabilize compact structures like that shown in Figure 3 compared to more extended structures. The electrostatic contribution to solvation free energy destabilizes the turn, whereas the nonpolar contribution has a nearly flat profile as a function of hydrogen bond distance. The chain configurational entropy terms favor unfolded states, by about 1−2 kcal/mol for equal-sized bins.

The heavy lines at the bottom of Figure 7 compare the sum of these energy terms to the potential of mean force determined directly from the molecular dynamics trajectory. Overall, the agreement between the two curves is quite good, with no deviation greater than 2 kcal/mol except for the extreme conformations beyond 8 Å. The "flatness" of this potential of mean force is something of a coincidence, arising from the nature of the particular sequence chosen and molecular mechanics potentials we have chosen to use. Nevertheless, the close agreement between microscopic and continuum solvent models supports earlier conclusions[42,47,48] that these two models provide approximately equivalent theoretical descriptions of the available conformational space.

3.5. Results Using Surface-Area Models for Solvation. The close agreement shown in Figure 7 between results for the continuum solvent model and the OPLS/Amber/TIP3 molecular dynamics trajectory suggests that this continuum model provides about the same description of the effects of solvation on hydrogen bond formation as is provided by the explicit TIP3 water model. This result is in close agreement with earlier studies on hydrogen bond formation in simple systems like dimers of amides or dipeptides.[42,47,49] A popular alternative model for solvation postulates that solvent effects depend upon exposed surface area, generally with different coefficients for different types of atoms.[24−27] Figure 8 compares the continuum results to those that would have been calculated from a variety of such surface-area-based theories. Basically, none of the nonelectrostatic solvation theories destabilize the hydrogen bond (turn) conformers by more than 1 kcal/mol, and hence end up having solvated hydrogen bonds that are nearly as strong as those in vacuum (or other low-dielectric environments). Since the amount of exposed area buried on formation of a single hydrogen bond is not large, these surface-area models tend to have very flat profiles as a function of hydrogen bond strength; by contrast, a variety of experimental and theoretical studies indicate that hydrogen bonds are much weaker in aqueous solution than in less polar environments.[41,50−54] This apparent deficiency of solvation models based only on molecular areas may significantly hinder their use in detailed studies of peptide conformational energetics.

3.6. Results for APGN. It is clear from peptide studies and surveys of protein structures that the Pro-Gly sequence is the most favorable one for the central two residues of a β-turn,[13,55,56] but the effect of outer residues is less clear. Of interest here is the role of the charge on the side chain of the aspartate at position 4. NMR results show a reduction in turn formation if this charge is neutralized, either by going to low pH or by an asparagine at position 4, but these effects are small, and probably reflect changes in folding/unfolding free energy of less than 1 kcal/mol.[10,11]

Figure 9 shows results for a simple model in which peptide configurations are taken from the APGD simulation, but electrostatic energies are computed for a neutralized side chain. Only the electrostatic components are shown, since in this model all other contributions are the same as for APGD. Although

(47) Marrone, T. J.; Gilson, M. K.; McCammon, J. A. *J. Phys. Chem.* **1996,** *100,* 1439−1441.
(48) Wang, L.; O'Connell, T.; Tropsha, A.; Hermans, J. *Biopolymers* **1996,** *39,* 479−489.
(49) Sharp, K. *J. Comput. Chem.* **1991,** *12,* 454−468.
(50) Klotz, I. M.; Farnham, S. B. *Biochemistry* **1968,** *7,* 3879−3882.
(51) Wolfenden, R. *Biochemistry* **1978,** *17,* 201−204.
(52) Roseman, M. A. *J. Mol. Biol.* **1988,** *201,* 621−623.
(53) Sneddon, S. F.; Tobias, D. J.; Brooks, C. L., III. *J. Mol. Biol.* **1989,** *209,* 817−820.
(54) Jorgensen, W. L. *J. Am. Chem. Soc.* **1989,** *111,* 3770−2771.
(55) Wilmot, C. M.; Thornton, J. M. *J. Mol. Biol.* **1988,** *203,* 221−232.
(56) Sibanda, B. L.; Blundell, T. L.; Thornton, J. M. *J. Mol. Biol.* **1989,** *206,* 759−777.

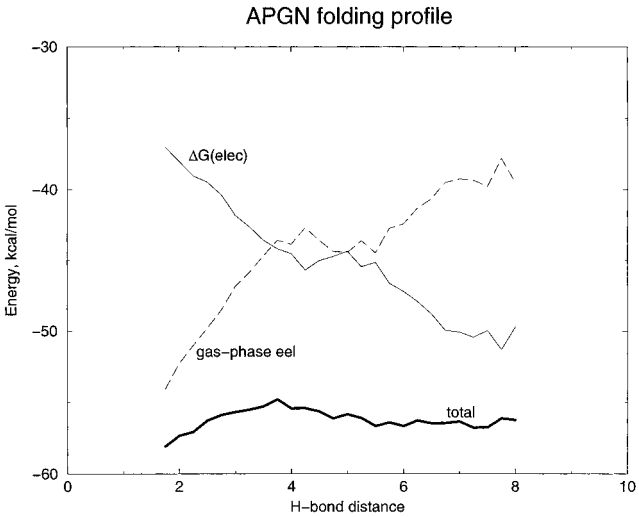

Figure 9. Same as Figure 5, but for AGPN rather than AGPD. For plotting purposes, the gas-phase curve has been shifted upward by 94 kcal/mol and the total curve shifted upward by 127 kcal/mol.

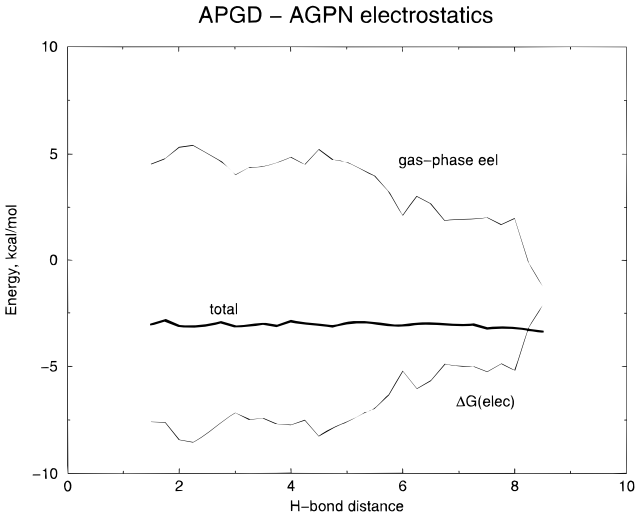

Figure 10. Difference of Figures 5 and 9. For plotting purposes, the light lines have been shifted upward by 80 kcal/mol.

the overall electrostatic energies are much smaller (*e.g.*, solvation energies near −45 kcal/mol rather than values near −130 kcal/mol seen in Figure 5) the near cancellation of gas-phase and solvation electrostatic profiles is nearly the same in the APGN model as in APGD (compare Figures 5 and 9). In particular, the "Born" solvation energy arising directly from polarization of the solvent by the aspartate side chain is nearly independent of H bond distance (data not shown), since the extent of solvent exposure of this side chain is nearly unaffected by the conformation of the backbone. In turn conformations, there is a repulsive interaction between the carbonyl of residue 1 and the charged side chain of residue 4, so that, in the gas phase, turn formation is favored by neutralizing the aspartate, as shown in the difference curves in Figure 10. In solution, however, this electrostatic repulsion is mostly screened, so that the net effect of neutralization on the folding profile nearly vanishes. This is in rough accord with experimental findings, although it is clear that a detailed analysis of this mutation would require separate molecular dynamics simulations on each species. As above, however, use of solvation theories that depend only on the exposed area would underestimate the screening effects and hence more resemble the gas-phase values, where neutralization favors turn formation by a large amount.

4. Discussion

When these peptides were first analyzed nearly 10 years ago, they were notable for having measurable amounts of β-turn content, given their short length and a general belief that short peptides (and denatured proteins) failed to possess such local structure. However, it should be recognized that even such "structured" peptides have a close free energy match between folded and unfolded conformers, so that significant populations of both forms are present at equilibrium. This circumstance offers an excellent opportunity to evaluate the ability of simulation methods to model fairly large conformational changes without the need to impose artificial, biasing potentials to ensure good sampling of conformational space. Indeed, our molecular dynamics simulation shows 17 spontaneous folding/unfolding transitions, occuring roughly once each 1/2 ns.

One principal goal of this project was to investigate the extent to which continuum models for solvent energetics matched that seen experimentally and in microscopic molecular dynamics simulations. It is known that models based on the Poisson−Boltzmann approximation can give a good account of solvation free energies of small organic molecules,[22,57,58] but less is known about the predictions of such models for conformational changes. Earlier studies[42,47] on model hydrogen bonding systems in peptides indicated that general trends were quite well preserved between microscopic and continuum models, and were generally consistent with other applications of continuum solvent models to problems of secondary structure formation in peptides.[48,59−61] The present study, which covers a wider range of peptide conformations, plus results on the sequence AYPYD[62,63] further supports the notion that peptide solutes respond to TIP3 water (in molecular dynamics simulations) nearly the same way they respond to a continuum solvent (in Poisson−Boltzmann simulations). Models for solvation based solely on exposed surface area fare much more poorly in such comparisons.

It is, of course, of considerable interest to make comparisons to experiment as well. NMR and CD measurements can be used to rank various peptide sequences in terms of their propensities to form particular types of secondary structure, but are less quantitative in providing estimates for particular populations. For AGPD, NMR evidence suggests that the fraction of turn conformer is about 50%,[64] but values from 20% to 70% could probably be accommodated. For a two-state model, this range of populations corresponds to a folding free energy near zero, with an uncertainty of ±0.3 kcal/mol. In our MD simulation, 27% of the structures had 4 → 1 H-bond distances shorter than 3.0 Å, with 20% being shorter than 2.5 Å, in rough agreement with experimental findings. It is not clear that any better agreement should be expected without major improvements in simulation methodology and in the quality of force fields.

Recently, Yang *et al*. have studied β-turn propensities in a blocked dipeptide system using an energy methodology similar to the one used here (*i.e.*, gas-phase force field + Poisson−Boltzmann solvation), sampling configurational space with a Monte Carlo method.[61] Like us, they find turns to be slightly destabilized with respect to more extended conformations; for

(57) Cramer, C. J.; Truhlar, D. G. *Science* **1992,** *256*, 213−217.
(58) Tomasi, J.; Persico, M. *Chem. Rev.* **1994,** *94*, 2027−2094.
(59) Yang, A.-S.; Honig, B. *J. Mol. Biol.* **1995,** *252*, 351−365.
(60) Yang, A.-S.; Honig, B. *J. Mol. Biol.* **1995,** *252*, 366−376.
(61) Yang, A.-S.; Hitz, B.; Honig, B. *J. Mol. Biol.* **1996,** *259*, 873−882.
(62) Demchuk, E.; Bashford, D.; Case, D. A. *Folding Des.* **1997,** *2*, 35−46.
(63) Demchuk, E.; Bashford, D.; Gippert, G. P.; Case, D. A. *J. Mol. Biol.*, in press.
(64) Dyson, H. J.; Rance, M.; Houghten, R. A.; Lerner, R. A.; Wright, P. E. *J. Mol. Biol.* **1988,** *201*, 161−200.

a Pro-Gly sequence, they predict about 7% population of turn, slightly lower than what we find here for a longer sequence. (In addition to the longer length, this difference could also result from details of the force field [CHARMM *vs* OPLS/Amber] or from the definition of what counts as a turn conformer.) Tobias *et al*. have used specialized sampling techniques to map out free energy profiles for folding/unfolding transitions in blocked dipeptides.[33,65] Their results show a somewhat larger penalty for turn formation, with 2−3% of Pro-Gly sequences predicted to populate turn regions. Both groups studying dipeptides postulate that additional effects in larger peptides could stabilize turn conformers, which is consistent with what we observe, although no consistent computation study of this size dependence is yet available.

While it is clear that computational results that require accuracy to greater than 1 or 2 kcal/mol often vary from one study to another, we find it encouraging that results from different approaches often lead to the same trends,[41] and that a continuum solvent model appears to give good results across a spectrum of folded and unfolded conformers. We show elsewhere that the same approach can produce good results for a somewhat more complicated conformational problem where oppositely charged groups can approach each other in certain conformations.[63] The method used here to estimate configurational entropy effects along reaction pathways should also be applicable to other peptide systems, and the combination of techniques outlined here appears to provide a powerful new approach to the analysis of conformational problems in peptides and proteins.

Acknowledgment. This work was supported by NIH Grant P01-GM38794. We thank Peter Wright, Jane Dyson, Doree Sitkoff, and Eugene Demchuk for helpful discussion, and Frank Eisenhaber for the ASC program.

JA963516K

(65) Tobias, D. J.; Sneddon, S. F.; Brooks, C. L., III. *Advances in Biomedical Simulations*; American Institute of Physics: Obernai, France, 1991; pp 174−199.