Thermodynamics of a Reverse Turn Motif. Solvent Effects and Side-chain Packing

Eugene Demchuk, Donald Bashford*, Garry P. Gippert and David A. Case

Department of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA

The linear pentapeptide, Ala-Tyr-cis-Pro-Tyr-Asp-NMA (AYPYD) is known to have a significant population of type VI turn conformers in aqueous solvent. We have carried out theoretical studies of the conformational energetics of this peptide using a potential of mean force (PMF) consisting of the AMBER/OPLS empirical potential energy function, a macroscopic electrostatic model of polar solvation, and a surface area-based model of non-polar solvation. Conformers were taken from molecular dynamics simulations reported elsewhere, or generated by a random search method reported here. The chain entropy of folding was calculated by a systematic search of accessible dihedral angle space. The intra-peptide component was found to strongly favor folding and was nearly cancelled by the polar solvation term which disfavored folding. The non-polar solvation term had little effect. Fluctuations about the average value of the PMF were small and in accord with estimates from a simple harmonic model. When applied to conformers generated by a random search, the PMF selected a conformer close to the NMR-determined structure as the lowest energy conformer. The conformer with the second-lowest energy was extended, but was found to fold rapidly to the turn state in a subsequent molecular dynamics study, and may be an important state on the folding–unfolding pathway. Averages of the PMF were combined with the entropy estimates to provide an estimate of the free energy of folding that is in reasonable agreement with experimental results. In terms of the interplay between backbone electrostatic interactions and the packing of apolar side-chains, this peptide provides a model for the energetics of protein folding, and therefore makes a useful test case for calculations.

*Corresponding author

Keywords: folding; peptide; solvation free energy; Poisson–Boltzmann equation; electrostatics

Introduction

Local regions of a protein’s primary sequence often take on nascent secondary structural forms similar to their conformation in the native tertiary structure. This propensity is of interest for the prediction of protein folds, the elucidation of the folding process and for their role in immune system recognition (Kim & Baldwin, 1982; Dyson et al., 1988b; Fersht, 1993; Dobson et al., 1994). A number of examples have been found in which small peptide analogs of these local regions take on, at least nascently, a secondary structure characteristic of the native protein (Dyson & Wright, 1991). In many such cases, the independently forming fragments are large to medium-sized units of structure such as one or more helices or a helix–strand motif (Dyson et al., 1992a; Waltho et al., 1993; Yang et al., 1995), but in a few cases, they are much shorter sequences that form reverse turns (Dyson et al., 1992b, 1988a). Reversal of the direction of the polypeptide chain commonly occurs over a short span of residues in protein structures. These tight reverse turns involve nearly one-third of the residues in...
globular proteins (Creighton, 1993) and have been classified into six to eight types (Richardson, 1981; Wilmot & Thornton, 1990). It is generally believed that short peptides, including many homologs to turns in proteins, have little or no specific structure in isolation, and some theoretical studies (Tobias et al., 1991a; Lazaridis et al., 1991; Yang et al., 1996; Avbelj & Moult, 1995) have been consistent with this expectation. On the other hand, the existence of at least some protein reverse turns whose peptide homologs have detectable turn population suggests that such turns may play a role in nucleating or guiding the protein folding process (Wright et al., 1988). Furthermore, these turn forming peptides are short enough that quite detailed computational studies can carried out at reasonable cost, and can thus serve as test cases for techniques that could also be applied to larger units of structure.

Here, we present a computational study of the conformational energetics of one such peptide, AYPYD, which belongs to a class of peptides fitting the motif, X-Ar-cis-Pro-Ar-Hp, where X is any amino acid, Ar an aromatic residue, and Hp a small hydrophilic residue. Peptides of this class have been found by NMR measurements to contain significant populations of type VI turn conformers in aqueous solution (Yao et al., 1994b). Unlike the type I and type II β-turns, which have been the subject of several recent theoretical studies (Tobias et al., 1991a,b; Lazaridis et al., 1991; Scully & Hermans, 1994; Yang et al., 1996) the type VI turn does not generally have a cross-turn hydrogen bond as a defining or stabilizing feature. Rather, its features include a cis-proline and a set of characteristic backbone dihedral angles (Wilmot & Thornton, 1990); and in the present case, its stability appears to be related to the packing of the aromatic side-chains against the proline ring (Yao et al., 1994a).

Previously (Demchuk et al., 1997), we reported three long molecular dynamics simulations of this peptide. One simulation started near a type VI turn conformation remained so, while two additional simulations started in different “unfolded” conformations spontaneously folded to a turn in one case, while going to a “non-native” but stable or meta-stable structure in the other. This is in contrast to our previous results with the smaller peptide, APGD, for which a single simulation provided a number of folding and unfolding transitions and a wide diversity of conformers which could serve as an ensemble for calculating free energy as a function of turn hydrogen-bond distance (Bashford et al., 1997). The present work is a study of the energetics of turn formation in AYPYD, with a particular emphasis on the interplay of intra-peptide, solvation and chain entropic contributions. The availability of experimental estimates of the population of type VI turn for peptides within this motif provides an opportunity to compare calculated free energies with experiment and to gain some insight into the applicability of the methods used here to studies of the stability of other structural units.

Calculations of folding thermodynamics can be divided into two parts: the generation of an ensemble of conformers and the calculation of a suitable potential of mean force or weighting function for each conformer. Here, we use systematic searches of dihedral angle space for an estimate of chain entropy; and we carry out randomized generation of conformers (Smith & Honig, 1994; Chan & Lim, 1994) for energetic analysis. We also make use of the conformational ensembles generated previously by molecular dynamics to explore smaller regions of conformational space and to test the consistency of our methods.

To make the per-conformer free energy calculation tractable, it is desirable to replace the explicit microscopic representation of the solvent with a computationally less expensive semi-macroscopic or parameterized model of solvation. Simplified solvation models have a long history, and include models based on the treatment of the solvent as a dielectric continuum (Still et al., 1990; Tomasi & Persico, 1994; Honig & Nicholls, 1995) or parameterized models based on cavity surface area or volume (Hermann, 1972; Eisenberg & McLachlan, 1986; Ooi et al., 1987). Here, we focus on the technique of Macroscopic Electrostatics with Atomic Detail (MEAD) for the response of solvent to solute polarity and a surface area-based model for the apolar contribution to solvation. Models of this kind have been used successfully to calculate the hydration free energies of small-molecule analogs of peptides and amino acid side-chains (Sitkoff et al., 1994). In comparisons of microscopic-solvent and MEAD/surface-area calculations of solvent effects on the interactions of peptide units, reasonable agreement was found between the two methods (Osapay et al., 1996; Wang et al., 1996; Marrone et al., 1996); and a MEAD/surface-area analysis of the above-mentioned APGD trajectory gave results consistent with the folding free energy profile derived from molecular dynamics (Bashford et al., 1997). Several other studies of secondary structure formation have been done using solvation models of this kind (Yang & Honig, 1995a,b; Yang et al., 1996). For comparison, we also test the use of several popular solvation energy models based more heavily on surface areas (Eisenberg & McLachlan, 1986; Ooi et al., 1987; Wesson & Eisenberg, 1992; Vila et al., 1991).

Methods

The peptide modeled is Ala-Tyr-cis-Pro-Tyr-Asp-NMA, where NMA denotes an N-methylamide-blocked C terminus; the N-terminal ammonium is protonated; and the aspartate side-chain is deprotonated so that the overall system is neutral. Conformers were taken from molecular dynamics studies described elsewhere (Demchuk et al., 1997), and from a random build-up procedure described here. For comparison purposes, a
conformer modeled on the NMR-determined structure of the closely analogous peptide, SYFPDV, was used (Yao et al., 1994b).

**Free energy and the potential of mean force**

Calculation of free energy implies statistical averaging over the entire phase space of the system, but in liquid simulations, calculations over solvent degrees of freedom often consume most of the computational resources. In order to focus computational efforts on the polypeptide molecule rather than solvent, a potential of mean force (PMF) can be defined in terms of the full set of peptide coordinates, \( q_{\text{mol}} \):

\[
U_{\text{pmf}}(q_{\text{mol}}) = U_{\text{gas}}(q_{\text{mol}}) + \Delta G_{\text{sol}}^*(q_{\text{mol}})
\]

where \( U_{\text{gas}} \) is the gas-phase potential energy of the peptide and \( \Delta G_{\text{sol}}^* \) is the free energy of solvation that the molecule would have if its coordinates are fixed at \( q_{\text{mol}} \). The superscript is a reminder that it is not a complete solvation free energy of the molecule. Formally this corresponds to a simple rearrangement of the configurational integral and an integration over solvent degrees of freedom. Here we have used the AMBER/OPLS energy function (Jorgensen & Tirado-Rives, 1988; Jorgensen & Severance, 1990) without solvent molecules or cut-offs for the calculation of \( U_{\text{gas}} \). The way in which \( \Delta G_{\text{sol}}^* \) was calculated is discussed below.

If the conformational space, \( q_{\text{mol}} \), is divided into folded and unfolded regions, the free energy of folding will be:

\[
\Delta G_{\text{folding}} = (U_{\text{pmf}}(\text{folded}) - (U_{\text{pmf}}(\text{unfolded}) - T\Delta S_{\text{mol,conf}})
\]

where the angle brackets indicate averaging with weights, \( e^{-\beta U_{\text{pmf}}} \), over the folded or unfolded regions of \( q_{\text{mol}} \), and \( T \) and \( S \) are temperature and entropy, respectively. Note that \( \Delta S_{\text{mol,conf}} \) refers only to configurational entropy of the solute molecule, and not to solvent configurations. The \( U_{\text{pmf}} \) terms can be obtained from suitable molecular dynamics simulations which are presumed to produce correctly weighted ensembles, but entropic terms generally require more conformational sampling for convergence (Beveridge & DiCapua, 1989). A method for estimating \( \Delta S_{\text{mol,conf}} \) will be described.

**Conformation-dependent solvation free energy**

For the purpose of calculating \( \Delta G_{\text{sol}}(q_{\text{mol}}) \), the process of transferring a solute molecule fixed in a conformation, \( q_{\text{mol}} \), from vacuum to solvent can be divided into three steps: the removal of all charges, \( Q_s \), from the molecule in vacuum; the transfer of the resulting, non-polar pseudo-molecule from vacuum to solvent; and the restoration of the charges, \( Q_s \), to the molecule in solvent. We can then write:

\[
\Delta G_{\text{sol}} = \Delta G_{\text{es}} + \Delta G_{\text{np}}
\]

where \( \Delta G_{\text{es}} \) is the free energy of the charging and discharging steps and \( \Delta G_{\text{np}} \) is the free energy of the non-polar transfer. Although both terms formally involve integration over solvent degrees of freedom, we will make the approximations that \( \Delta G_{\text{es}} \) can be obtained from quasi-macroscopic electrostatics, and \( \Delta G_{\text{np}} \) is proportional to the solute’s surface area (Osapay et al., 1996).

**Calculation of the electrostatic term**

Electrostatic effects were treated in the dielectric continuum approximation (Bushford & Karplus, 1991; Honig et al., 1993). It implies a model in which: (1) the polypeptide molecule is considered a low dielectric cavity surrounded either by vacuum (\( \epsilon = 1 \)) or by the high dielectric constant of bulk solvent; (2) the dielectric boundary is determined from the atomic coordinates and radii of the molecules, and a solvent probe radius; (3) partial point charges are assigned to polypeptide atoms; and (4) solvent counterion charges are modeled according to a linearized Boltzmann distribution. We refer to this model as Macroscopic Electrostatics with Atomic Detail (MEAD).

The electrostatic solvation component of the PMF, \( \Delta G_{\text{es}}^* \), was calculated from the Born formulation (Born, 1920) as the electrostatic work required to bring peptide atomic charges from zero to their full values in solvent versus the gas phase:

\[
\Delta G_{\text{es}}^* = \frac{1}{2} \sum Q_i (\phi_i^s - \phi_i^g)
\]

where \( Q_i \) is the partial point charge at the nucleus of the \( i \)th atom of the molecule and \( \phi_i^s \) and \( \phi_i^g \) are the electrostatic potentials at the position of the atom in the gas phase and in solvent, respectively.

The OPLS non-bonded parameter set (Jorgensen & Tirado-Rives, 1988) was used to assign atomic partial charges and radii. The radii were set to \( 2^{-\frac{1}{6}} \sigma \), where \( \sigma \) is the Lennard–Jones parameter, i.e. the radii correspond to the minimum in the Lennard–Jones potential. The dielectric boundary was determined from the protein three-dimensional structure by the algorithm of You & Bushford (1995) which implements the “molecular surface” definition of the boundary (the contact and reentrant surfaces of the molecule (Richards, 1977; Connolly, 1983)). A spherical probe radius of 1.4 Å was used in these calculations. The dielectric constant of the molecular interior was set to 1.0, the gas phase dielectric constant; and the dielectric constant of solvent was 78.5. The molecular interior was surrounded by a 2 Å ion exclusion layer. An ionic strength of 0.1 M was assigned to the solvent outside the exclusion layer and a temperature of 298 K was used in the Boltzmann term. The above set of radii, charges and dielectric constants is one of several parameterizations of MEAD models tested by Sitkoff et al. (1994) for the ability to reproduce small molecule solvation energies. Among the molecular mechanics parameters tested, it was found to produce results in closest agreement with experiment.

The electrostatic potentials for this model were determined from the linearized Poisson–Boltzmann equation. They were solved numerically using the finite difference methods provided in the MEAD program suite (Bushford et al., 1993). The potentials at the boundary of the grid were assigned from the Debye (or Coulomb) law using the exterior solvent (or vacuum) dielectric constant. The electrostatic potentials were solved on a nested set of finite-difference grids. Four cubic grids with spacings of 2.6, 1.2, 0.6 and 0.25 Å and corresponding dimensions of 25³, 31³, 53³ and 111³ were used. They were centered at the geometrical center of the molecules.

**Calculation of the non-polar hydration term**

The non-polar term of the PMF, \( \Delta G_{\text{np}} \), is analogous to a gas-to-water transfer free energy of acyclic alkanes,
which has been approximated as a linear function of the surface area of the cavity occupied by the molecule (Hermann, 1972; Simonson & Brünger, 1994; Sitkoff et al., 1994). We therefore use:

$$\Delta G_{\text{tip}} = \gamma \Omega + b$$

(5)

where $\Omega$ is the surface area and the empirical coefficients, $\gamma$ and $b$, are derived by fitting experimental transfer free energies. These parameters have been estimated as $\gamma = 5(\pm 0.5)$ cal/mol per $\AA^2$ and $b = 860(\pm 100)$ cal/mol (Sitkoff et al., 1994) when the scaled OPLS atomic radii, the Lee & Richards (1971) definition of the accessible surface and a 1.4 Å water probe radius are used. Solvent accessible surface areas were calculated using the SURFV program (Sridharan et al., 1992).

Peptide conformational entropy of folding

Torsional degrees of freedom were assumed to be the source of solute conformational entropy changes upon folding and unfolding. We therefore use a systematic search of the backbone and side-chain $\phi$, $\psi$, $\chi^1$, and $\chi^2$ angles. An evaluation of $U_{\text{pred}}$ at each search point is not feasible, but techniques are available to find all sterically allowable points on a specified multidimensional dihedral angle lattice for peptides of the present size (Gippert, 1995). If the dihedral space points are further classified as representing folded or unfolded conformations, the folding entropy can be estimated as:

$$\Delta S_{\text{mol,conf}} = k_B \ln \frac{W_{\text{folded}}}{W_{\text{unfolded}}}$$

(6)

where $W_{\text{folded}}/W_{\text{unfolded}}$ is the ratio of the number of allowed folded dihedral points to allowed unfolded dihedral points. If the dihedral sampling grid is sufficiently fine, the calculated $\Delta S_{\text{mol,conf}}$ should be insensitive to further refinement of the sampling. The approximation used here amounts to the replacement of $U_{\text{pred}}$ by a square well; one may hope that the consequences of this replacement are similar in the folded and unfolded states, resulting in some cancelation of errors.

The search for sterically allowed dihedral points was carried out using the DTAGS (Distributed Torsion Angle Grid Search) program suite (developed by G.P.G.). The programs perform a systematic search for all conformations that satisfy excluded volume constraints. The search algorithm is similar to the “build-up” method (Howard et al., 1975) except that all satisfactory conformations are retained at each step of the calculation, and conformations are accepted on an “all-or-nothing” basis, i.e. explicit energy terms are not included in the evaluation of conformations.

Systematic grid searches of the $\phi$, $\psi$, $\chi^1$, and $\chi^2$ dihedral angles were carried out using fixed dihedral increments through $360^\circ$. The $\chi^2$ angle for tyrosyl aromatic rings was varied in a $180^\circ$ range due to twofold degeneracy of the ring. Proline ring conformers were generated using a pseudo-rotation scheme, with sampling of the pseudo-rotation phase angle at uniform increments. In this scheme, ring conformers are generated on the pseudo-rotation trajectory defining correlated $\phi$, $\psi$, $\chi^1$, and $\chi^2$. Peptide $\omega$ bond angles were fixed at planar values: $180^\circ$ for trans and $0^\circ$ for cis.

As to means of classifying conformers as folded or unfolded, we considered the cross-turn interatomic distances, Tyr2H$^\alpha$-TyrH and Tyr2H$^\alpha$-AspH. For these pairs, NMR constraints corresponding to a type VI turn have been established by Yao et al. (1994b). We also considered backbone root-mean-square deviation from the NMR structure, and ranges of backbone dihedral angles characteristic of type VI turns.

Alternative estimates of solvation energy

For comparison purposes we also estimated solvation energy directly from surface area. The approach is based on “atomic solvation parameters” (ASPs) rather than equations (4) and (5). It is of the form:

$$\Delta G_{\text{sol}}^{\text{ASP}} = \sum \Delta \sigma_i \Omega_i$$

(7)

where $\Delta \sigma_i$ and $\Omega_i$ are the ASPs and atomic solvent-accessible areas of the molecule, respectively. $\Delta G_{\text{sol}}^{\text{ASP}}$ has been thought of as a computationally inexpensive alternative to the solvation free energy, $\Delta G_{\text{sol}}^{\text{IP}}$. We have used the ASP models of Eisenberg & McLachlan (1986), Ooi et al. (1987) and Wesson & Eisenberg (1992). We also tested the model of Vila et al. (1991), which is an ASP model with an exponentially decaying, distance-dependent correction for the interactions of solvent-exposed groups. For these calculations, the implementations in the computer program, ASC (Eisenhaber & Argos, 1993) were used.

Random conformational sampling

Random sampling of locally minimized molecular configurations was carried out using a build-up procedure similar to that of Howard et al. (1975), but with random initial building block conformations. For each consecutive residue with the exception of proline, an alanine backbone with random $\phi$ and $\psi$ dihedral angles was generated and its energy was minimized. Then the $\text{C}^\beta$ atom of the residue was mutated to the proper side-chain. All dihedral angles of the side-chain rotamer were assigned at random. The new residue was energy minimized and its conformation was fixed, so that the next residues to be added could not affect a conformation of the previously built polypeptide fragment. For cis-proline, only the dihedral angle, $\psi$, was assigned at random.

The minimizations in the above procedure were carried out by the following protocol. First, up to 250 steps of minimization were performed without electrostatic forces in order to relieve steric clashes. Second, up to 6000 steps of minimization were performed using an electrostatic term with a distance-dependent dielectric constant of $\varepsilon = 15\varepsilon_0$, as a rough approximation of solvent screening effects. In both stages, Powell conjugate gradient minimization was used and the minimization stages terminated when the RMS energy gradient became less than 0.05 kcal/mol per Å or when the maximum number of iterations was reached, whichever came first.

The final step of the build-up procedure, after all of the residues were added, was to subject the hydrogen atoms of the molecule to energy minimization. This minimization was done in the same way as the second stage of the one-residue minimizations.

The above procedure was carried out using the SYBYL version 6.03 molecular modeling software package (Tripos Inc., St. Louis, MO) and its implementation of the AMBER all-atom force field (Weiner et al., 1986).
Results and Discussion

Systematic search and chain entropy

Three systematic searches of the dihedral angle space were carried out using the DTAGS procedure described. In order from coarsest to finest, the increments of \((\phi, \psi, \chi', \chi'')\), in degrees, were \((36, 36, 120, 90)\), \((30, 30, 120, 60)\) and \((30, 30, 60, 60)\). The coarse, medium and fine searches generated \(10,205,153, 40,577,236\) and \(277,521,293\) sterically allowed conformers, respectively. Further reduction of the search increments was not feasible because of increasing cost of the calculations. Entropies calculated from these searches using equation (6) and various definitions of the folded state are presented in Table 1 in terms of \(-T\Delta S_{\text{mol,conf}}\). Variation of the search angle intervals gave results varying by up to 0.4 kcal/mol. This gives some indication of the degree of convergence of the calculation with respect to search intervals.

Wilmot & Thornton (1990) define the type VIa and type VIb turns found in protein structures in terms of the backbone dihedral angles of the residues corresponding to Tyr2 and Pro3 in AYPYD. Yao et al. (1994b) found average dihedral angles in their distance geometry structures close to the Wilmot and Thornton definition of a type VIa turn, so their distance geometry structures close to the Wilmot & Thornton definition of a type VIa turn, so their distance geometry structures close to the Wilmot & Thornton (1990) central angles for type VIa except: \(\phi_2 = 60^\circ, \psi_2 = 120^\circ, \phi_3 = 90^\circ, \psi_3 = 0^\circ\); for type VIb, same as type VIa except: \(\phi_2 = 120^\circ, \phi_3 = 60^\circ\); n.a., not available.

<table>
<thead>
<tr>
<th>Folded state criteria*</th>
<th>(-T\Delta S_{\text{mol,conf}}) (kcal/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coarse</td>
</tr>
<tr>
<td>Type VIa (NMR)</td>
<td>2.5</td>
</tr>
<tr>
<td>Types VIa + VIb (std.)</td>
<td>1.4</td>
</tr>
<tr>
<td>2H(^{-})-4H</td>
<td>0.4</td>
</tr>
<tr>
<td>2H(^{-})-5H</td>
<td>2.6</td>
</tr>
<tr>
<td>Backbone RMSD</td>
<td>0.8</td>
</tr>
</tbody>
</table>

* Criteria as given described in the text and legend to Figure 1. The type VI-like NMR average angles are \(\phi_2 = -55^\circ, \psi_2 = 139^\circ, \phi_3 = -81^\circ, \psi_3 = -7^\circ\) (Yao et al., 1994b). The standard (Wilmot & Thornton, 1990) central angles for type VIa are \(\phi_2 = -60^\circ, \psi_2 = 120^\circ, \phi_3 = 90^\circ, \psi_3 = 0^\circ\); for type VIb, same as type VIa except: \(\phi_2 = 120^\circ, \phi_3 = 60^\circ\); n.a., not available.

b Dihedral angle intervals for DTAGS searches are given in text.

lower value of \(-T\Delta S_{\text{mol,conf}}\) was found, as expected from a broadening of the definition. A nearly identical value was obtained in a calculation (not shown) using a 45\(^\circ\) range about average angles from the molecular dynamics simulation of the folded peptide (Demchuk et al., 1997). The definition based on the Cartesian RMSD of backbone atoms from the NMR model gave somewhat lower \(-T\Delta S_{\text{mol,conf}}\) and further examination of the conformers showed that this definition produced a set containing about two-thirds of those produced by the VIa + VIb definition. The definition based on 2H\(^{-}\)-4H distance produced the lowest value of \(-T\Delta S_{\text{mol,conf}}\) (largest set of folded conformers).

Three of the definitions used involved a description of the structure in terms of a single parameter (a distance or an RMSD) which might be thought of as “reaction coordinates” for folding. Histograms of these parameters versus conformer population are presented in Figure 1. The sparseness of the histogram for the 2H\(^{-}\)-4H distance is a reflection of the discreteness of the grid search and the fact that this distance depends on only three dihedral angles, one of which is a proline \(\psi\) angle. The histogram for the RMSD of residues 2 to 4 has a broad minimum near the threshold value which is suggestive of a conformational “bottleneck” between two regions of conformational space, while the histogram for the 2H\(^{-}\)-5H distance is more ambiguous in this regard. These results are consistent with the results of our 20 ns molecular dynamics simulation of the turn (Demchuk et al., 1997) in which the NMR distance constraints were often exceeded and transitions were readily made between VIa and VIb, but the backbone atoms of residues 2 to 4 remained within 1.15 Å RMSD of the NMR-derived model.

Random sampling

A set of 500 conformers were generated by the random build-up procedure. The breadth of the conformational space sampled is illustrated by plots of the RMS deviation of the backbone atoms of residues 2 to 4, which we take as a folding coordinate, versus frequency of occurrence and versus several other structural parameters.

The frequency histogram, Figure 2, shows that a number of folded (RMSD <1.15 Å) and unfolded conformers are sampled. The overall shape and range of the histogram is similar to Figure 1c suggesting that the random and systematic searches are not too different in breadth, even if the random search is much sparser. In terms of a different structural parameter, distances between the C\(^\alpha\) atoms of Ala1 and Tyr4 varied from 3.67 Å, a distance even shorter than that of a typical \(\alpha\)-helix, to 8.97 Å, a value specific for extended conformations (Lewis et al., 1973). Figure 3, in which the y-axes are angular RMS deviations from standard turn VI definitions, shows that conformers near both the a and b subtypes are found as well as conformers far away from either, and that con-
formers quite far from either standard type VI turn can have central backbone Cartesian coordinates that are not very far from the NMR structure. This is consistent with the observation from the systematic search results that RMSD-based and dihedral angle-based definitions of the turn produce only partly overlapping sets of folded conformers. In Figure 4, the \( y \)-axes are distances between the central proline ring and the aromatic side-chains of the flanking tyrosine residues. Proline–aromatic interactions are thought to be important for the stability of the turn (Yao et al., 1994b). The Figures show that conformations are generated both with and without these interactions.

The energetic components, \( U_{\text{gas}} \) and \( \Delta G_{\text{sol}}^{*} \), are plotted in Figures 5a and b, respectively; and \( U_{\text{pmf}} \), which is their sum, is shown in Figure 5c (see also equation (1)). The gas phase free energy term and

Figure 1. Distribution of configurations found in a systematic search of the dihedral angle space of the cis-AYPYD peptide. The \( x \)-axes are: distance between 2H\( ^{\alpha} \) and 4H (a), and 5H (b); backbone RMSD to the three central residues, Tyr2, Pro3 and Tyr4, in the NMR structure (c). Vertical broken lines indicate NMR constraints of 3.5 Å (a) and 4.0 Å (b), and the limit of fluctuations observed in MD simulations (c), 1.15 Å (Demchuk et al., 1997). Data are from the finest of the three searches described in the text.

Figure 2. Histogram of RMSDs between the NMR structure and random conformations of the AYPYD polypeptide. The RMSD was calculated for the backbone atoms of Tyr2, Pro3 and Tyr4.

Figure 3. Angular RMS deviation of random AYPYD conformers from the four backbone angles defining standard type VIa (a) and of type VIb (b) turns (see Table 1). Points that cluster at the distance less than 45° around the standard turn VIb (see b) are shown with concentric filled circles in a. Diamonds represent the unminimized (empty) and the minimized (filled) NMR structure. The Cartesian RMSD (\( x \)-axes) was calculated as for Figure 2.
the solvation energy show opposing trends: conformers with a low $U_{\text{gas}}$ component are clustered near the NMR conformation, while the lowest $\Delta G_{\text{sol}}$ values are found for unfolded conformations (Figure 5b). The net effect, shown in Figure 5c, is that 197, the conformer with the lowest $U_{\text{gas}}$, also has the lowest $U_{\text{pmf}}$ (−340.5 kcal/mol) and is very close to the NMR structure (see Figure 6a). Moving up along the energy axis, the next several lowest energy structures are far from the NMR structure. These conformers are not among those with the lowest $U_{\text{gas}}$, nor are they the ones with the lowest $\Delta G_{\text{sol}}$. Rather, they are extended conformers which have a favorable balance of internal and solvation energetics. RMSD calculations (not shown), examination of side-chain rotamers and visual inspection demonstrate that these low-PMF unfolded structures, although similar in their distance from the NMR structure, and similar in their energetics, are not conformationally similar to one another. The tendency of vacuum electrostatic and solvation terms to stabilize folded and unfolded conformers, respectively, has also been observed in calculations of randomly generated tetrapeptide conformers (Chan & Lim, 1994).

Conformer 401, the second-lowest-energy conformer, was singled out for further study and is depicted in Figure 6b. Its backbone does not make a reverse turn because of the extended orientation of the C-terminal residues. This extension of the C terminus prevents the stacking interaction between Pro3 and the side-chain of Tyr4, although the Pro3 to Tyr2 stacking interaction is present. The gas-phase energy of 401 is considerably higher than that of folded conformers: −168 kcal/mole compared to −207 and −251 kcal/mol for the minimized NMR and 197 conformers, respectively. On the other hand, the electrostatic solvation energy of 401 is lower: −176 kcal/mol compared to −128 and −95 kcal/mol for the NMR and 197 conformers; and the resulting $U_{\text{pmf}}$ for conformer 401 is

![Figure 4. Distances between aromatics for random conformations of AYPYD polypeptide versus RMSD of backbone atoms of the three residues to the NMR structure. The distance between Cα atom of Tyr2 and Cα atom of Pro3 (a); and between Cα atom of Pro3 and Cα atom of Tyr4 (b). Circles, diamonds and triangles correspond to gauche⁺, gauche⁻ and trans orientations of $\chi^1$ dihedrals. The enlarged characters are the values for the unminimized (open characters) and the minimized (filled characters) NMR structure.](image)

![Figure 5. PMF and its components calculated for random conformers of AYPYD polypeptide. a, $U_{\text{gas}}$; b, $\Delta G_{\text{sol}}$; c, $U_{\text{pmf}}$, the sum of a and b. Only conformers in the lower energy ranges are shown. Lower energy conformers are labeled with their index numbers from the random generation process. The large filled diamonds are the values for the minimized NMR structure.](image)
only 2 kcal/mol higher than that of 197. The comparison suggests that intra-peptide interactions stabilize the turn while solvent effects, particularly the electrostatic component of solvation, destabilize the turn. These competing effects combine to produce net energetics slightly favorable to folding. In a molecular dynamics simulation reported elsewhere (Demchuk et al., 1997) conformer 401 was found to fold rapidly and spontaneously. The energetics of this process are analyzed in terms of $U_{pmf}$ in the next section.

Energetics of conformers from trajectories

Three molecular dynamics trajectories that are detailed elsewhere (Demchuk et al., 1997) are analyzed in terms of $U_{pmf}$ and its components here. They are: (1) the NMR trajectory, a 20 ps trajectory which starts from the NMR-determined turn-like conformer and remains folded throughout; (2) the A trajectory, a 5.2 ns dynamics trajectory starting from the low-energy unfolded conformer, 401, described earlier and undergoing a transition to a folded form at approximately 2.7 ns; and (3) the B trajectory, a 4.7 ps trajectory that starts in a cis-proline, but non-type-VI, conformation and quickly becomes trapped in a "misfolded" state. We have calculated values of $U_{pmf}$ for snapshots taken at 10 ps intervals from these trajectories. A total of 2990 conformers have been analyzed.

A trajectory in the folded state

Figure 7a shows a striking compensatory effect between $U_{gas}$ and $\Delta C_{cal}$ during the NMR trajectory. Each significant deviation of the gas-phase energy term from its mean is almost exactly counterpoised by the solvation energy. As a result, the RMS deviation of $U_{pmf}$ from its mean is 6.0 kcal/mol even though the corresponding gas-phase RMS is 16.2 kcal/mol. Such an effect is expected since solvation is known to flatten the backbone dihedral energy map, weaken hydrogen bonds, and generally screen electrostatic interactions; and these effects are seen in both explicit microscopic solvent models and in MEAD models (Smith & Honig, 1994; Chan & Lim, 1994; Bashford et al., 1977; Osapay et al., 1996; Marrone et al., 1996). There is no general expression from statistical physics for the expected RMS of $U_{pmf}$ for a peptide, but the harmonic approximation provides a useful estimate of the order of magnitude.
expected. A system of $N$ independent oscillators at temperature $T$ has an RMS deviation given by:

$$\langle (U_{\text{pmf}}) - \langle U_{\text{pmf}} \rangle^2 \rangle^{1/2} = k_BT(N/2)^{1/2}$$

(8)

For the 63-atom peptide, an estimate of 5.9 kcal/mol is obtained which is quite similar to the fluctuations of $U_{\text{pmf}}$ of the simulation snapshot. This suggests that our estimate of $\Delta G^*_{\text{gas}}$ is damping the fluctuations of $U_{\text{gas}}$ to about the right degree. This property of $\Delta G^*_{\text{gas}}$ is almost entirely due to $\Delta G^*_{\text{sol}}$ the MEAD component. The $\Delta G^*_{\text{traj}}$ component fluctuates very little, remaining between 4.8 and 5.8 kcal/mol throughout the simulation. In Figure 7b four different surface area-based empirical hydration energy functions are plotted against time. Three of them clearly do not possess the compensatory effects needed to dampen the fluctuations of $U_{\text{gas}}$, to the extent observed for the $\Delta G^*_{\text{sol}}$ of equation (3). The function of Vila et al. (1991) comes close to $U_{\text{gas}}$ in the magnitude of its fluctuations, presumably because it includes corrections based on interatomic distances in addition to surface area terms, but its fluctuations are not correlated so as to cancel the fluctuations of $U_{\text{gas}}$.

Table 2 summarizes the energetics of the NMR trajectory, including averages restricted to periods when the peptide is within the specifications of a type VIa or type VIb turn. The VIb form is favored by approximately 1 kcal/mol in $U_{\text{pmf}}$. (For a more detailed discussion of the VIa–VIb equilibrium, see Demchuk et al. (1997).)

**A spontaneously folding trajectory**

The folding event at 2.7 ns in the A trajectory is accompanied by a sharp negative fluctuation of $U_{\text{pmf}}$ (thin line of Figure 8) and a more gradual decline of the average such that $\langle U_{\text{pmf}} \rangle$ over the last 2 ns of the simulation is approximately 4 kcal/mol lower than the average over the first 2 ns (Table 2). There is no noticeable barrier in the plot of $U_{\text{pmf}}$ versus time corresponding to the folding transition. The change in $\langle U_{\text{pmf}} \rangle$ is modest in comparison to both the fluctuations of $U_{\text{pmf}}$ and the changes in individual components of $U_{\text{pmf}}$. Throughout the simulation the standard deviation of $U_{\text{pmf}}$ is 5 to 6 kcal/mol. The average $U_{\text{gas}}$ component of the PMF decreases by approximately 15 kcal/mol between the first and last 2 ns while the $\Delta G^*_{\text{sol}}$ component increases by 11 kcal/mol. This is consistent with the earlier observation that the unfolded conformer, 401, is stabilized more by solvation whereas the folded form is stabilized more by intra-peptide interactions. The non-polar term, $\Delta G^*_{\text{traj}}$, makes no significant contribution to folding.

**A trajectory to a misfolded state**

A second trajectory, which we have named the B trajectory, begins from a more compact unfolded state and quickly falls into a misfolded state that does not resemble a type VI turn (Demchuk et al., 1997). Figure 9 shows $U_{\text{pmf}}$ versus time for this simulation. The energy drop in the first 0.5 ns corresponds to the structural relaxation into the misfolded state, and the relatively flat energy profile over the remaining 4.5 ns of simulation corresponds to the lack of significant structural change during this period. The average value of $U_{\text{pmf}}$ during the trajectory is approximately 10 kcal/mol higher than the average over the NMR trajectory or the latter part of A trajectory indicating that the misfolded conformer is only meta-stable (Table 2). Although the PMF of the misfolded state is higher, the relative distribution of its components, in which $U_{\text{gas}}$ contributes more stability than does $\Delta G^*_{\text{sol}}$, resembles the component distribution of the

---

**Table 2. Energetic characteristics of type VI turns**

<table>
<thead>
<tr>
<th>Source $^*$</th>
<th>$U_{\text{gas}}$</th>
<th>$\Delta G^*_{\text{traj}}$</th>
<th>$\Delta G^*_{\text{traj}}$</th>
<th>$U_{\text{pmf}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMR</td>
<td>−154.0</td>
<td>−139.2</td>
<td>5.4</td>
<td>−287.9</td>
</tr>
<tr>
<td>NMR, min</td>
<td>−207.1</td>
<td>−127.6</td>
<td>5.3</td>
<td>−293.5</td>
</tr>
<tr>
<td>NMR, max</td>
<td>−160.2</td>
<td>−146.4</td>
<td>5.3</td>
<td>−301.3</td>
</tr>
<tr>
<td>NMR, traj/VIa</td>
<td>−161.4</td>
<td>−144.7</td>
<td>5.6</td>
<td>−301.0</td>
</tr>
<tr>
<td>NMR, traj/VIb</td>
<td>−162.0</td>
<td>−145.3</td>
<td>5.4</td>
<td>−302.0</td>
</tr>
<tr>
<td>A start</td>
<td>−250.8</td>
<td>−95.1</td>
<td>5.4</td>
<td>−340.5</td>
</tr>
<tr>
<td>A start (401)</td>
<td>−168.5</td>
<td>−175.7</td>
<td>5.6</td>
<td>−338.6</td>
</tr>
<tr>
<td>A, traj &lt; 0.5 ns</td>
<td>−131.7</td>
<td>−173.2</td>
<td>5.6</td>
<td>−298.9</td>
</tr>
<tr>
<td>A, traj &lt; 2 ns</td>
<td>−138.5</td>
<td>−164.8</td>
<td>5.5</td>
<td>−297.9</td>
</tr>
<tr>
<td>A, traj &lt; 4 ns</td>
<td>−153.5</td>
<td>−153.8</td>
<td>5.3</td>
<td>−302.0</td>
</tr>
<tr>
<td>B start</td>
<td>−255.1</td>
<td>−53.3</td>
<td>4.9</td>
<td>−303.5</td>
</tr>
<tr>
<td>B start</td>
<td>−164.2</td>
<td>−131.9</td>
<td>5.1</td>
<td>−291.0</td>
</tr>
</tbody>
</table>

All energies are in kcal/mol.

$^*$ NMR refers to the coordinate set derived from NMR data (Yao et al., 1994a); B refers to the starting conformer of the B trajectory; numbers refer to the conformers. Unsubscripted names refer to starting conformations and min subscripts refer to minimized conformers. Subscripts: traj refers to averages over trajectories; traj followed by a positive or negative time refers to a time segment at the beginning or end of the trajectory, respectively.
folded state (NMR trajectory from Table 2) more closely than that of the extended unfolded state ($A_{\text{traj}} + 2$ ns from Table 2). However, the $U_{\text{gas}}$ component is slightly more favorable for the B trajectory than for the NMR trajectory, while the $\Delta G_{\text{es}}$ component is less favorable for the B trajectory.

**Overall energetics of folding**

Combining the earlier results, it is possible to make some estimates of overall folding free energy using equation (2). If the folded state is defined by dihedral angle criteria as including both type VIa and VIb conformers, then the entropy calculations of Table 1 give a $-T\Delta S_{\text{mol,conf}}$ contribution of 1.3 kcal/mol. The average of $U_{\text{pmf}}$ over the NMR trajectory is an obvious choice for $(U_{\text{pmf}})_{\text{folded}}$ since the trajectory makes several transitions between VIa-like and VIb-like regions of the folded conformational space (Demchuk et al., 1997). On the other hand, the only candidates available for $(U_{\text{pmf}})_{\text{unfolded}}$ are the averages from either the unfolded part of the A trajectory or the B trajectory. The latter is approximately 7 kcal/mol higher in energy than the former and would not contribute significantly to a Boltzmann average. Therefore, we choose the average of $U_{\text{pmf}}$ over the first 2 ns of the A trajectory as $(U_{\text{pmf}})_{\text{unfolded}}$. The differences of $(U_{\text{pmf}})$ values is then $-3.4$ kcal/mol and subtracting $T\Delta S_{\text{mol,conf}}$ gives a $\Delta G_{\text{folding}}$ value of $-2.1$ kcal/mol. If the dihedral angle criteria defining the folded state are instead based on the type VIa-like NMR model, the $-T\Delta S_{\text{mol,conf}}$ contribution is 2.1 kcal/mol, and since the value of $(U_{\text{pmf}})$ was 0.3 kcal/mol higher over the VIa-like parts of the NMR trajectory than that for the overall trajectory (Table 2), the difference of $(U_{\text{pmf}})$ values is $-3.1$ so that $\Delta G_{\text{folding}} = -1.0$ kcal/mol. Apart from the limitations of the method of estimating chain entropy and ambiguity in defining the folded state, the major limitation of the above estimates is the use of the unfolded part of the A trajectory as an ensemble over the unfolded conformational space. Other unfolded trajectories with higher $(U_{\text{pmf}})$ values (such as B) would have less impact on the result than other trajectories with lower $(U_{\text{pmf}})$. Therefore a more thorough exploration of unfolded space would be likely to lead to less negative values.

The population of turn conformers in the cis form of the peptide SYPFDV has been estimated as approximately 70% on the basis of NMR studies; and the AYPYD peptide studied here is expected to have similar properties (Dyson et al., 1988a; Yao et al., 1994a) although in MD simulations it had less VIa and more VIb population than SYPFDV is thought to have (Demchuk et al., 1997). The $\Delta G_{\text{folding}}$ value corresponding to 70% is $-0.5$ kcal/mol. Our calculated values are of the same sign and magnitude, but more negative, as one would expect from the argument above.

**Conclusions**

The combination of a gas-phase empirical potential function, a MEAD model of electrostatic solvation, and an area-based model of non-polar solvation to create a potential of mean force gave very encouraging results in this study. The fluctuations of $U_{\text{pmf}}$ over a dynamics trajectory are small and consistent with simple models from statistical physics; this suggests that the large $\Delta G_{\text{es}}$ component cancels the large gas-phase electrostatic component to about the right extent. This cancellation is not achieved by purely area-based models of polar solvation which do not include the screening of long-range electrostatic interactions (Figure 7b). The average value of $U_{\text{pmf}}$ for an unfolded form that appears to be on a folding–unfolding pathway is slightly higher than that of the folded form, and lower than that of a trapped unfolded form that does not resemble the NMR-determined structure. The observation that the intrapeptide interactions ($U_{\text{gas}}$) tend to favor folding, while aqueous solvation opposes it is consistent with similar studies of other peptides (Smith & Honig, 1994; Chan & Lim, 1994; Bashford et al., 1997; Yang & Honig, 1995a,b; Wang et al., 1996; Marrone et al., 1996). When combined with entropy estimates, the $(U_{\text{pmf}})$ values give an estimate of $\Delta G_{\text{folding}}$ that is in reasonable agreement with experiment.

The AYPYD peptide studied here presents a more difficult conformational sampling problem than the tetrapeptides we have studied previously (Bashford et al., 1997), in that a single molecular dynamics trajectory cannot be relied upon to adequately explore both the folded and unfolded states. Furthermore, individual simulations of the unfolded state explore only limited regions of the unfolded conformational space, although it appears that the folded state can be explored adequately by a simulation on the order of 10 ns. This
makes it difficult to calculate $\Delta G_{\text{folding}}$ particularly its chain entropy component. The variation of the systematic dihedral angle search results for different dihedral lattice spacings and the very large number of conformers generated in the finest search suggest that with a pentapeptide, we are close to the limits of this particular method for estimating chain entropy. Other sources of error in the entropy estimate include uncertainty about the exact conformational descriptors distinguishing folded from unfolded states and the underlying assumption that the energy surface of the accessible conformational space is approximately flat. Even so, the entropies estimated by this method have the expected sign and order of magnitude and are similar to the results for the tetrapeptide which could be searched more thoroughly.

Another approach we have taken to the conformational problem is random conformer generation followed by $U_{\text{em}}$ calculations. Although the method is not directly applicable to the calculation of $\Delta G_{\text{folding}}$, it is useful for finding important conformers that might not be found by other means. Without any a priori information as to the expected turn structure, this method found as its lowest energy structure, one closely resembling that of the turn as determined by NMR. Its second-lowest-energy structure, 401, is a good candidate for a significant conformer in a folding pathway since it was found to fold rapidly and spontaneously in a subsequent molecular dynamics study, while a conformer generated by a different method did not fold. The 401 conformer could not have been found by ordinary molecular dynamics simulations started from the folded form (Demchuk et al., 1997).

Much of the methodology tested here should be useful for studies of larger units of protein structure. As noted above, the combination of an empirical gas-phase potential energy with a continuum model for electrostatic solvation gives a potential of mean force whose behavior is consistent with molecular dynamics results and leads to folding energetics in reasonable agreement with available experimental data. The computational cost of this PMF is significantly less than that of equivalent calculations using explicit water, but greater than the cost of evaluating the empirical gas-phase potential energy function. Although the particular method used for estimating entropy has a cost that increases rapidly with the size of the peptide, larger peptides could be accommodated either by coarser sampling or improved efficiency of the systematic search, or by substitution of alternative methods for entropy estimation (D’Aquino et al., 1996; Makhadzhe & Privalov, 1996). In the present study we have relied partly upon microscopically solvated molecular dynamics simulations (Demchuk et al., 1997) as a source of conformational samples, and these simulations were quite expensive (CPU times on the order of seven hundred hours on eight processors of a Convex/HP SPP machine). Clearly it would be advantageous to use $U_{\text{em}}$ directly in a dynamics or Monte Carlo calculation. To this end, there has been significant progress in incorporating Poisson solvers into molecular dynamics (Gilson et al., 1995; Sharp, 1991) or structural optimization (Cortis et al., 1996) programs. Another alternative is to find a simple modification of the gas phase potential that mimics $U_{\text{em}}$ reasonably well, to carry out dynamics or Monte Carlo on this potential, and to use MEAD calculations on a set of sample points in a more detailed energy analysis. In this connection, we note that Lazaridis et al. (1991) found that setting the dielectric constant, $\varepsilon$, to 50, gave reasonable reaction paths and preliminary energy profiles for turn formation in an alanine tripeptide, in comparison to subsequent calculations with microscopic solvent; and Yang et al. (1996) found that $\varepsilon=10$ gave similar results to a PMF using a MEAD-like continuum solvation model for the Ramachandran maps of the alanine and glycine dipeptides.

The present work, as well as experimental measurements (Dyson et al., 1988a; Yao et al., 1994a,b), show significantly greater turn propensity for AYPYD than one would generally expect for small peptides. However, calculations of type I or II turn formation in minimal peptide models of the form Ace-X-X-NMe have suggested that turn formation is unfavorable by 2 to 5 kcal/mol (Lazaridis et al., 1991; Tobias et al., 1991a; Yang et al., 1996); so a question arises as to the cause of this difference. In the latter peptides, the primary interaction stabilizing the turn would appear to be the hydrogen bond between the amides of the terminal blocking groups. These groups are more exposed to solvent than amide groups in a longer peptide would be, and both microscopic and macroscopic solvent models have found that solvent effects greatly weaken hydrogen bonds between amides, especially in more solvent-exposed cases (Sneddon et al., 1989; Osapay et al., 1996). One therefore expects the contribution of the $i$ to $i+3$ hydrogen bond to be substantially weakened by solvation. Furthermore, Scully & Hermans (1994) find that in microscopically solvated simulations of type I and II turns of Ace-Cys-Pro-X-Cys-NMe, most of the turn-like conformers seen do not have this hydrogen bond. In the case of AYPYD, hydrogen bonding across the turn is not observed during most of a molecular dynamics trajectory although a minor species includes a bifurcated hydrogen bond between the carbonyl group of the first residue and the NH groups of the fourth and fifth residues (Demchuk et al., 1997). The crucial turn-stabilizing interactions in AYPYD appear to be the stacking of the aromatic side-chains against the proline ring, a favorable interaction that is not diminished by solvation. One is reminded of the general picture of protein folding in which backbone hydrogen bonding confers only marginal stability on units of secondary structure and apolar interactions that occur upon the packing together of secondary units is
necessary for overall stability. Here, both kinds of interactions are present in a single small peptide.

Acknowledgments

We thank Dr Frank Eisenhaber for the program ASC, and P. E. Wright, J. Yao and H. J. Dyson for many helpful discussions. This work was supported by the National Institutes of Health (GM38794).

References


*Edited by B. Honig* (Received 7 October 1996; received in revised form 7 February 1997; accepted 13 March 1997)