A novel method reveals that solvent water favors polyproline II over β-strand conformation in peptides and unfolded proteins: conditional hydrophobic accessible surface area (CHASA)

PATRICK J. FLEMING,[1] NICHOLAS C. FITZKEE,[1] MIHALY MEZEI,[2] RAJGOPAL SRINIVASAN,[3] AND GEORGE D. ROSE[1]

[1]Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, Maryland 21218, USA
[2]Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York University, New York, New York 10029, USA
[3]Advanced Technology Centre, Tata Consultancy Services, Hyderabad, India

(RECEIVED August 10, 2004; FINAL REVISION September 3, 2004; ACCEPTED September 3, 2004)

Abstract

In aqueous solution, the ensemble of conformations sampled by peptides and unfolded proteins is largely determined by their interaction with water. It has been a long-standing goal to capture these solute-water energetics accurately and efficiently in calculations. Historically, accessible surface area (ASA) has been used to estimate these energies, but this method breaks down when applied to amphipathic peptides and proteins. Here we introduce a novel method in which hydrophobic ASA is determined after first positioning water oxygens in hydrogen-bonded orientations proximate to all accessible peptide/protein backbone N and O atoms. This *conditional* hydrophobic accessible surface area is termed CHASA. The CHASA method was validated by predicting the polyproline-II (P_{II}) and β-strand conformational preferences of non-proline residues in the coil library (i.e., non-α-helix, non-β-strand, non-β-turn library derived from X-ray elucidated structures). Further, the method successfully rationalizes the previously unexplained solvation energies in polyalanyl peptides and compares favorably with published experimentally determined P_{II} residue propensities.

We dedicate this paper to Frederic M. Richards.

Keywords: Solvation energy; conditional hydrophobic accessible surface area; CHASA; polyproline-II; coil library; probability density map

Of the many factors that determine the conformation of polypeptides, the interaction with water is one of the most important. Intrapeptide backbone interactions limit available conformational space (Pappu and Rose 2002), and sidechain interactions specify conformational preferences (Chou and Fasman 1978; Creamer and Rose 1992; Penel and Doig 2001), but neither works in isolation. Rather, such effects act in concert with solvation preferences, which are measured by the free energy of interaction between peptides and water (Luo and Baldwin 1999; Thomas et al. 2001).

It is often assumed, usually implicitly, that protein backbone solvation is uniform in the unfolded state. However, computational studies indicate that the solvation energetics differ among conformations in both peptides (Anderson and Hermans 1988; Pettitt and Karplus 1988; Tobias and Brooks 1992; Brooks and Case 1993; Bartels and Karplus 1997; Resat et al. 1997; Smart et al. 1997; Han et al. 1998; Scarsi et al. 1998; Tazaki and Shimizu 1998; Apostolakis et al.

Reprint requests to: George D. Rose, Jenkins Department of Biophysics, Johns Hopkins University, 3400 N. Charles St., Baltimore, MD 21218, USA; e-mail: grose@jhu.edu; fax: (410) 516-4118.

Article published online ahead of print. Article and publication date are at http://www.proteinscience.org/cgi/doi/10.1110/ps.041047005.

1999; Smith 1999; Hu et al. 2003; Drozdov et al. 2004) and longer chains (Avbelj et al. 2000; Garcia 2004; Kentsis et al. 2004; Mezei et al. 2004).

Accessible surface area (ASA) (Lee and Richards 1971) has a long and successful history in estimating the energetics of solvation in small molecules and peptides (Ooi et al. 1987; Wimley et al. 1996; Chan and Dill 1997) and desolvation during protein folding, unfolding, and association (Horton and Lewis 1992; Makhatadze and Privalov 1993; Murphy et al. 1993; Privalov and Makhatadze 1993; Hilser et al. 1996; Baker and Murphy 1998; Vallone et al. 1998). Indeed, this method has even been used as a potential function in protein simulations (Ferrara et al. 2002; Rathore et al. 2003).

Unfortunately, the ASA approximation breaks down eventually (Avbelj et al. 2000; Gallicchio et al. 2000; Mezei et al. 2004) because typical ASA calculations assume the entire surface in question is available for solvation simultaneously, and in an undifferentiated way, with no distinction made between one accessible site and another. This is a particularly poor assumption for a chemically heterogeneous moiety like a peptide, which has both polar and apolar sites in close proximity. Given that interactions between water and polar sites are much stronger than the corresponding interactions between water and apolar sites, the Boltzmann-weighted distribution of water-peptide interactions favors solvation of N-H and C=O groups over apolar groups. The residence times of water molecules at polar solvation sites are known to be two to three times longer than those at apolar sites (Russo et al. 2004), and therefore effective apolar solvation will frequently depend on prior polar solvation. In other words, a water molecule hydrogen-bonded to a backbone polar site can inhibit the close approach of other water molecules at nearby apolar sites.

In practice, this distribution can be approximated simply by pre-solvating sterically accessible polar sites. Subsequent calculation of hydrophobic ASA under the prior condition that these polar sites are already solvated provides a method to differentiate among these distinct solvation environments. Here we compare the conventional hydrophobic ASA to the hydrophobic ASA *conditional* upon prior water occupancy at polar solvation sites.

This conditional hydrophobic accessible surface area, termed CHASA, is a modification of conventional hydrophobic ASA, and it captures experimental propensities effectively. The CHASA modification was validated by comparison of predicted conformational propensities for all residues (except proline) to the experimentally observed natural propensities in the coil library, a data base of non-α-helix, non-β-strand, non-β-turn fragments derived from X-ray crystallographic structures (www.roselab.jhu.edu/coil). This library is thought to represent the unfolded population of proteins (Swindells et al. 1995; Fiebig et al. 1996; Smith et al. 1996).

Results

Position-specific solvation sites differ between polar and apolar groups. In Figure 1, solvation sites around polyalanine, calculated from an all-atom simulation in explicit solvent (Mezei et al. 2004), are displayed as probability density map contours. Both P_{II} (top) and antiparallel β-strand conformations (bottom) are shown. Probability densities are contoured at a level of 2.5%, i.e., there is a 2.5% probability that the center of a water oxygen is present in a 0.5 Å cube within these solid green contours. At this probability level, backbone N-H groups already exhibit significant position-specific solvation, and distinct solvation sites around backbone C=O groups are beginning to emerge. Decreasing the probability to 1.5% does not change the N-H contours significantly, but the carbonyl oxygen contours become more pronounced. Conversely, upon increasing the probability, only the N-H contours persist, and these survive to probability levels of 10%. In mean bulk solvent, probability density in the 0.5Å volume element was 0.45%.

It can be seen from these contours that N-H solvation localizes to a single site, while C=O solvation localizes to multiple positions around the oxygen. Solvation sites around Cβ atoms only emerge at very low contour levels, less than 1.5%; these are seen to be partially localized in antiparallel β-strand but not in P_{II} conformation.

The position of each N-H solvation site is consistent with an N-H···OH_2 hydrogen bond; i.e., the site is situated at a distance of 2.95 Å from the nitrogen and approximately in-line with the N-H bond vector (Taylor and Kennard 1984). At probability density levels between 2.5 and 1.5, C=O groups usually exhibit two predominant solvation sites, situated ~2.95 Å from the carbonyl oxygen and with some dependence on polypeptide conformation. The locations and densities of solvation sites in P_{II} conformation (Fig. 1) resemble those found by Garcia (2004) using unconstrained molecular dynamics of a blocked polyalanyl 21-residue peptide and a different force field.

These results, extracted from detailed simulations in explicit water, indicate that solvation is positionally specific

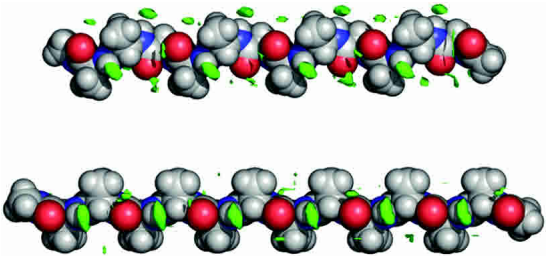

Figure 1. Molecular graphics images of solvation sites around polyalanine as probability density map contours. *Top*, P_{II} conformation; *bottom*, antiparallel β-strand conformation. The probability densities are depicted as green solid contours at the 2.5% level. The images were created using the program PyMOL (DeLano 2003).

for backbone polar atoms in polyalanine but much less so for the nonpolar Cβ group. The observed positional specificities are consistent with water hydrogen bonds to both N-H and C = O groups. Our measure of solvation positional specificity is related to the residence time of solvating water molecules at these sites. A recent study combining quasi-elastic neutron scattering and molecular simulation of a blocked leucinyl peptide concluded that water residence time near polar groups exceeds that of apolar groups by a factor of 2–3 (Russo et al. 2004). The positional specificity of the long-residence, low-energy sites will influence the location of more dynamic solvation positions around the rest of the molecule, especially the high-energy positions around hydrophobic groups.

Estimating solvation free energy using CHASA

Methods that estimate the hydrophobic contribution to solvation free energy by simply summing surface area fail to take site-dependent positional specificity into account. In these methods, the area is calculated as if the entire surface were simultaneously available for interaction with the solvent (Lee and Richards 1971), with all accessible positions treated equivalently. However, both experimental findings (Russo et al. 2004), other computational results (Garcia 2004; Kentsis et al. 2004), and our results (Fig. 1) indicate that polar and apolar solvent sites are not equivalent. The availability of high-energy hydrophobic solvation positions depends in part on prior solvation of low-energy hydrogen-bonded sites, which have longer water residence times. The CHASA method was developed to take these critical differences into account.

The predictive value of CHASA was tested by using the method to calculate the solvation free energies and resulting preferences of residues in both P_{II} and β-strand conformations in a polyalanyl host-guest peptide. The calculated P_{II}/β-strand preferences were then compared to natural preferences in a data set derived from experimentally determined structures.

The difference in CHASA between the two respective conformations in a pure polyalanyl peptide is 59.5 – 38.0 = 21.5 $Å^2$ per residue (Table 1 with X = Ala). This difference stands in marked contrast to the conventional, nonconditional difference in hydrophobic ASA, which is almost identical for the two conformations: 68.8 and 68.3 $Å^2$ per residue, respectively (data not shown). CHASA distinguishes between these two conformations while conventional hydrophobic ASA fails to measure a difference.

In detailed free energy simulations of polyalanine, the difference in solvation free energy, ΔΔA, between β-strand and P_{II} is 0.7 kcal/mol/residue (β-strand = −4.0 kcal/mol/residue; P_{II} = −4.7 kcal/mol/residue) (Mezei et al. 2004). Using these values and assuming the difference in solvation free energies is due entirely to differences in hydrophobic surface, the free energy of nonpolar surface solvation

$$\Delta\Delta A_{nonpolar} = \gamma_{np}\Delta CHASA \quad (1)$$

gives γ_{np} = 0.03 kcal/$Å^2$, remarkably close to both early (Hermann 1972) and recent (Chan and Dill 1997) values in the literature. Furthermore, this value of γ_{np} is probably an underestimate, owing to the fact that the polar solvation condition would be more realistically derived as a Boltzmann-weighted probability, in which case the effective ΔCHASA would be less than indicated in Table 1.

We can then derive

$$-\Delta A_{polar} = \Delta A_{solv} - \Delta A_{nonpolar} \quad (2)$$

using the above values for ΔA_{solv} (−4.0, −4.7) together with a calculated $\Delta A_{nonpolar}$ from surface areas in Table 1 and γ_{np} in Equation 1. Doing so, we obtain a total solvation free energy of −6 kcal/mol/residue for peptide polar groups (ΔA_{polar}). This estimate compares favorably with the value of −7.9 kcal/mol/residue calculated by Avbelj and Baldwin (Avbelj et al. 2000) using a finite difference Poisson-Boltzmann method. From our analysis, there is no basis on which to separate the respective contributions of N-H and C = O groups to the total polar solvation free energy. Therefore, we have approximated the individual contributions by splitting the difference equally, assigning a value of −3 kcal/mol/residue for the solvation free energy of each polar group that has free access to solvent water within its respective cone of approach, as described in Materials and Methods. Steric limitations in the close approach of water molecules are assessed by attempting to situate a water oxygen at multiple positions within the solvent approach cone of each polar group. That group is assigned to have a favorable solvation free energy if and only if it can form a clash-free hydrogen bond with at least one water molecule, with the solvation free energy becoming increasingly favorable as the cone volume becomes increasingly accessible.

Detailed results of the host-guest peptide solvation free energy and P_{II}/β-strand preference calculations are listed in Table 1. In every case, hydrophobic solvation free energy favors P_{II} conformation, i.e., conditional hydrophobic accessible surface area is significantly less in P_{II} than in β-strand. In contrast, polar group solvation free energy is slightly favored in β-strand conformation for 17 residue types, and equal to P_{II} conformation in the remaining two. Total solvation free energies, a balance between hydrophobic and polar contributions, favor P_{II} conformation for all residue types.

The correlation between our calculated P_{II}/β-strand preferences and the corresponding P_{II}/β-strand preferences observed in the coil library is plotted in Figure 2. Although imperfect, our calculations capture important features that

Table 1. *Calculated conformational energetics of host-guest peptides*

	β-strand[a]				P_{II}[a]					
X Res.	CHASA[b] Å^2	Apolar solv. free energy kcal/mol[c]	Polar solv. free energy kcal/mol[c]	Total solv. free energy kcal/mol[c]	CHASA[b] Å^2	Apolar solv. free energy kcal/mol[c]	Polar solv. free energy kcal/mol[c]	Total solv. free energy kcal/mol[c]	P_{II}[d] pref. %	Coil library pref.[e] %
ALA	297.4	8.9	−30.0	−21.1	189.9	5.7	−30.0	−24.3	75.0	76.0
ARG	296.5	8.8	−30.0	−21.2	197.2	5.9	−29.8	−23.9	72.3	57.2
ASN	269.5	8.0	−30.0	−22.1	174.5	5.2	−29.2	−24.0	62.5	49.0
ASP	271.1	8.0	−29.9	−22.1	171.9	5.1	−29.1	−24.2	67.0	55.3
CYS	325.5	9.8	−30.0	−20.2	216.8	6.5	−29.3	−22.9	70.9	57.3
GLN	281.3	8.3	−29.8	−21.7	185.4	5.3	−29.1	−23.9	68.2	56.1
GLU	278.4	8.3	−29.9	−21.8	186.5	5.3	−29.1	−24.0	68.1	63.7
GLY	264.5	7.9	−30.0	−22.1	172.6	5.2	−30.0	−24.8	71.9	79.1
HIS	303.2	9.0	−29.8	−20.9	217.4	6.5	−28.7	−22.2	61.0	52.7
ILE	345.0	10.3	−30.0	−19.7	239.4	7.0	−28.5	−21.5	64.6	46.8
LEU	347.3	10.4	−30.0	−19.6	241.1	7.2	−29.8	−22.4	71.9	67.6
LYS	323.9	9.6	−30.0	−20.3	222.1	6.6	−29.8	−23.2	71.1	59.1
MET	324.0	9.5	−30.0	−20.3	220.7	6.5	−29.8	−23.3	73.3	62.3
PHE	357.0	10.6	−29.8	−19.2	261.5	7.8	−29.4	−21.4	67.9	53.0
SER	266.2	7.9	−30.0	−22.1	174.5	5.2	−29.9	−24.8	71.0	63.2
THR	283.7	8.3	−30.0	−21.7	188.8	5.6	−29.2	−23.8	67.3	50.5
TRP	369.4	10.9	−29.8	−18.5	291.7	8.7	−29.2	−20.7	68.2	58.8
TYR	320.7	9.5	−29.8	−20.5	235.5	7.0	−29.4	22.2	64.1	52.1
VAL	325.7	9.8	−30.0	−20.2	215.1	6.4	−28.5	−22.2	66.2	46.6

The host peptide model (Acetyl-ALA_5-X-ALA_6-N-methylamide) was constructed with backbone torsions set to $\phi = -120°$, $\Psi = 130°$ (β-strand) or $\phi = -78°$, $\Psi = 149°$ (P_{II}). Different guest residues were constructed in the X position with the same backbone conformations. All sterically allowed rotamers were analyzed. The constructed peptide models were tested for accessibility to hydrogen-bonding water oxygens, and solvation free energy was calculated as the Boltzmann-weighted sum of hydrophobic and polar solvation terms as described in the text.

[a]Values shown are the sums for 5 residues (4–8) of the peptide.

[b]Conditional hydrophobic accessible surface area (CHASA) is the hydrophobic accessible surface area calculated conditional upon prior solvation of the backbone N and O atoms.

[c]Solvation free energies are Boltzmann-weighted averages according to the number of sterically allowed rotamers.

[d]Preference was calculated from the relative solvation free energies taking into account the number of ways (rotamers) that contributed to the energies as described in the text.

[e]Preference was calculated from a non-α-helix, non-β-strand, non-β-turn fragment database extracted from the PDB (Berman et al. 2000), as described in the text.

agree with these experimentally observed distributions. First, alanine and glycine have high P_{II} preferences in both the coil library and model host-guest system. Second, the β-branched residues, valine, isoleucine, and threonine have low P_{II} preferences in both the coil library and calculations. For these residues, the CHASA values, and thus hydrophobic solvation free energies, are favorable in P_{II} conformation, but this contribution is partially counterbalanced by less favorable polar solvation. Histidine, tyrosine, and tryptophan have the least favorable CHASA differences between P_{II} and β-strand (85.7, 85.2, 77.7 Å^2 per five residues, respectively) with correspondingly low calculated P_{II} preferences.

A more comprehensive exploration of conformational space may change the exact values of these calculated P_{II} preferences, but the trends seen in the coil library are captured effectively with our straightforward analysis. The P_{II}/β-strand preferences for the coil library reported here largely agree with previously published coil library preferences (Avbelj and Baldwin 2003), although the two coil libraries are not strictly comparable. Specifically, we have removed β-turn residues, unlike the coil library used by Avbelj and Baldwin (2003); this difference may explain the discrepancy between the low P_{II} preferences of aspartate and asparagine observed here and those published previously.

Comparison with experimentally determined scales

Additional experimental validation can be seen by comparing the trends in Figure 2 with recent estimates of P_{II} preferences in host-guest experiments using circular dichroism (Chellgren and Creamer 2004). In these experiments, alanine, glutamine, asparagine, and valine were measured in a proline-based host. In this system, the P_{II} propensity of alanine is high, glutamine is less, and asparagine and valine are the lowest. The relative order of valine and asparagine is reversed from that in Figure 1, but the overall trends are consistent. Our results also are in general agreement with

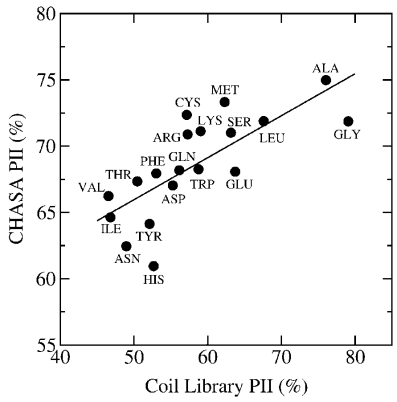

Figure 2. Calculated and natural P_{II} preferences compared to β-strand for different amino acid residues. Calculated preferences (CHASA P_{II}) are from Table 1, and natural preferences (Coil Library P_{II}) were calculated from a coil library derived from the PDB as described in the text. The line represents a linear regression with a correlation coefficient of 0.75.

those of Eker et al. (2004), who used spectroscopic methods to obtain the dihedral angles of ALA-X-ALA tripeptides, although the relative propensities of TYR and TRP reported here are different. A strict correlation is not expected in either case, because the experimental systems report on the P_{II} propensities with respect to all available conformational space, while we have singled out the P_{II} to β-strand preferences.

Discussion

The CHASA method is easily extended to any other range of peptide conformations, such as that shown in Figure 3. The conformational space examined in the figure represents an area in the upper left quadrant of the Ramachandran plot (Ramachandran et al. 1963), covering the extended strand and P_{II} regions. The top plot shows that traditional hydrophobic ASA is relatively insensitive to conformation. In contrast, the middle and bottom plots show that the CHASA, conditional upon N-H and C=O solvation, respectively, varies considerably with conformation. In both CHASA plots, the hydrophobic accessible surface is minimal within the region $-85° \leq \phi \leq -55°$ and $130° \leq \psi \leq 180°$, which includes P_{II} but excludes β-strand.

In developing the CHASA method, we started with the simplifying assumption that backbone polar groups are solvated equivalently in both P_{II} and β-strand conformers of polyalanine, and therefore the main difference in solvation free energy between these two conformations would arise from differences in hydrophobic solvation. Similarly, Avbelj and Baldwin found that the free energy of backbone group solvation is dependent on whether the polar group is solvent-accessible, but not on the extent of polar ASA (Avbelj et al. 2000), earlier findings notwithstanding (Spolar et al. 1992; Makhatadze and Privalov 1993; Privalov and Makhatadze 1993). In both β-strand and P_{II}, the N-H and C=O are openly available for solvation in a polyalanine peptide (Fig. 1). Very recently, Petukov et al. (2004) reported that polar group solvation free energies are proportional to polar ASA but with a hyperbolic relationship, based on analysis of molecular dynamics simulations. Their results suggest that after a minimal exposure to solvent is attained (5–10 $Å^2$ ASA), solvation free energy becomes only slightly dependent on additional polar ASA. We assess this nonlinear proportionality by probing backbone polar groups with water oxygens at multiple positions and quantifying the number of successful attempts.

In other related work, Hamburger et al. (2004) and Ferreon and Hilser (2004) use an experimental peptide/protein binding system to assess the thermodynamics of P_{II} formation. For alanine in a proline-based host peptide, they find that P_{II} conformation is enthalpy-driven, implicating backbone-solvent interactions as the molecular origin of the P_{II} preference. How can this evidence be reconciled with the fact that such preferences are also predicted by differences in conditional hydrophobic accessible surface areas, an entropy-driven effect at 25°C?

The substantial CHASA-based reduction in apolar surface favors P_{II} over β-strand in all cases (Table 1). However, once in P_{II}, individual residue preferences may be dominated instead by differences in polar accessibility to solvent. For example, within P_{II} an ALA to VAL mutation would increase ΔA_{apolar} by 0.7 kcal/mol but ΔA_{polar} by 1.5 kcal/mol. These within-structure differences may predominate in experimental peptide mutation studies.

Clearly, our understanding of these solvation effects is incomplete. The interplay between polar and apolar solvation in an amphipathic molecule like a peptide is complicated by their spatial juxtaposition and the solvation dynamics of water. In a comprehensive thermodynamic analysis of hydration during protein unfolding, Privalov and Makhatadze used ASA parametized with small-molecule experimental data to successfully calculate hydration enthalpies, entropies, and Gibbs energies on a macromolecular scale for four proteins (Makhatadze and Privalov 1993; Privalov and Makhatadze 1993). Despite this success, the atomic scale contributions—including hydrogen bonding, van der Waals forces, polar and apolar hydration enthalpies and entropies—remain difficult to quantify.

Three recent studies (Drozdov et al. 2004; Garcia 2004; Mezei et al. 2004) found that a significant component of P_{II} preference is contributed by the fact that this conformation is less disruptive of bulk solvent organization than β-strand. This contribution does not come from backbone:water hydrogen bonding per se; instead, it is manifest indirectly in the energetics of water:water interactions. The success of CHASA may reflect the compensating effects of decreased apolar group–water interaction with increased polar group–

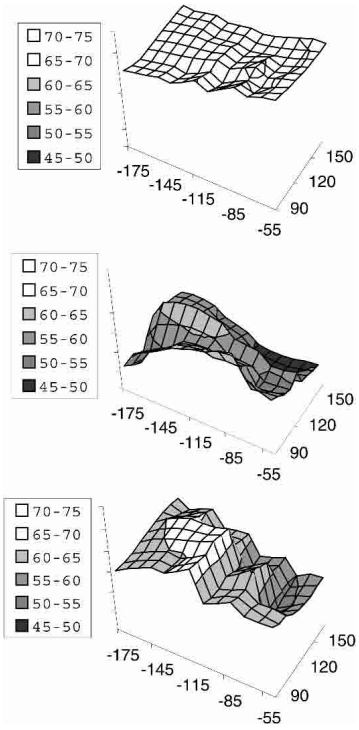

Figure 3. Per residue hydrophobic accessible surface areas as a function of polyalanine ϕ, ψ values. The surface areas in Å^2 are indicated in the boxes to the *left* of each plot; ϕ is along the *bottom* axis, and ψ is along the axis on the *right* in each plot. Conformational distributions of 12-residue polyalanyl peptides were generated as described in Materials and Methods. The nonconditional hydrophobic ASA (*top*), the CHASA conditional upon N-H solvation (*middle*), and the CHASA conditional upon C = O solvation (*bottom*) were calculated for each residue in each polypeptide conformation as described in the text. The surface area values for individual residues in each polypeptide conformation were accumulated into 10° × 10° ϕ, ψ bins, and the mean hydrophobic ASA or CHASA value within the bin is plotted. The number of residues represented in each 10° × 10° ϕ, ψ bin is 890 ± 134 (S.D.).

water interaction, coincident with β-strand to P_{II} conformational change. In any event, the fact that differences in hydrophobic solvation can rationalize the preferred conformations of peptides suggests that we must look beyond the backbone to understanding this phenomenon.

Summary

Previous applications of the ASA method assume that the entire solvent-accessible surface is available simultaneously. We challenge this assumption by showing that the accessibility of hydrophobic solvation sites may depend on prior solvation of proximate polar sites. For solutes such as polypeptides, computing the hydrophobic ASA after first pre-solvating backbone polar groups can correct for this dependence. Solvation free energy calculations using CHASA successfully predict individual residue preferences for P_{II} conformations over β-strand conformations in a coil library derived from known protein structures. To the extent that the coil library represents the unfolded population of proteins, these results add to the growing body of evidence that P_{II} is a prevalent conformation in unfolded proteins.

Materials and methods

The conditional hydrophobic accessible surface area (CHASA) of model peptides was calculated using modules from the LINUS suite of programs (Srinivasan and Rose 1995; Srinivasan et al. 2004) with an effective probe radius of 1.4 Å after first placing virtual oxygen atoms at suitable positions around backbone N and O atoms, viz., a distance of 2.95 Å from the peptide N or O atom and positioned within a 30° cone around the N-H bond vector or an 80° cone around the C = O bond vector. The algorithm described by Shrake and Rupley (1973) was used with 960 sampling points per atom surface to calculate accessible surface area. Hydrophobic area is defined as the accessible surface area associated with carbon atoms; hydrogen atoms were not included in the calculation.

Conformational preferences of residues were predicted in a model host-guest system. Specifically, each of 19 different residue side chains was modeled at position 6 of a blocked, 12-residue polyalanyl host constructed in either β-strand or P_{II} backbone conformation. All rotamers described by Lovell et al. (2000) were constructed, and van der Waals overlap was removed, if possible, by varying the χ angles up to 10°. Only models without steric clash were used in the analysis. CHASA was calculated for residues 4–8 in each host-guest model as described above.

CHASA values, together with estimates of polar solvation, were used to calculate a solvation free energy (ΔA_{solv}) as

$$\Delta A_{solv} = \Delta A_{npolar} - \Delta A_{polar} \tag{3}$$

$$\Delta A_{solv} = [\gamma_{np} * \text{CHASA}] - [\varepsilon_{pol} * (\text{Number solvated backbone polar groups})] \tag{4}$$

where γ_{np} is the solvation free energy per Å^2 of conditional hydrophobic surface area, and ε_{pol} is the free energy contributed by each backbone N-H or C = O group for which a hydrogen bond to water is sterically allowed. As discussed in the Results, values for these parameters (in kcal/mol) were determined to be: $\gamma_{np} = 0.03$ and $\varepsilon_{pol} = 3.0$. In addition, placement of a hydrogen-bonded water oxygen was probed at five uniformly distributed positions within each cone of approach. The number of successful, clash-free attempts was used to calculate the final polar solvation free energy, with $\varepsilon_{pol} = (3.0/5)$ each.

The Boltzmann-weighted P_{II} preference (P_{PII}) of each residue type in the guest position of host-guest peptides was calculated from ΔA_{solv} as:

$$P_{PII} = \frac{g_{PII} e^{-\Delta A_{PII}/kT}}{g_{\beta-strand} e^{-\Delta A_{\beta-strand}/kT} + g_{PII} e^{-\Delta A_{PII}/kT}} \tag{5}$$

where g_{PII} is the number of sterically possible rotamers in P_{II} conformation, ΔA_{PII} is the Boltzmann-weighted average solvation free energy of the P_{II} conformation, $g_{\beta\text{-}strand}$ is the number of sterically possible rotamers in β-strand conformation, and $\Delta A_{\beta\text{-}strand}$ is the Boltzmann-weighted average solvation free energy of the β-strand conformation.

For the flexible backbone calculations shown in Figure 3, sterically allowed distributions of polyalanine conformations were made by constrained torsion angle Monte Carlo simulations of a blocked 12-residue polyalanyl model using modified modules from the LINUS suite of programs (Srinivasan and Rose 1995; Srinivasan et al. 2004) and www.roselab.jhu.edu/dist/. The region of the Ramachandran plot bounded by $-175° \leq \phi \leq -45°$ and $90° \leq \psi \leq 180°$ was subdivided into 15 equal, $30° \times 30°$ bins. Fifteen separate hard-sphere Monte Carlo simulations were performed, with ϕ, ψ-values constrained to each respective bin. Conformations were saved after every 300 successful attempts; 1000 conformations were accumulated for each bin. Virtual oxygen atoms were placed at backbone solvation sites only where sterically feasible, as described above. The accessible surface area for residues 3–10 was then calculated with and without virtual oxygens and indexed by conformation, resulting in 8000×2 values per bin.

Probability density maps of solvent water oxygen atoms were calculated from previously described molecular simulations (Mezei et al. 2004) using the MMC program (inka.mssm.edu/~mezei/mmc). Briefly, blocked polyalanyl peptides modeled as either β-strand ($\phi = -139°$, $\psi = 135°$) or P_{II} ($\phi = -78°$, $\psi = 149°$) were solvated with ~2600 TIP3P waters (Jorgensen et al. 1983) under periodic boundary conditions and simulated with the CHARMM22 force field (MacKerell et al. 1998) at 300°K. The simulations used the Metropolis algorithm (Metropolis et al. 1953) with force-bias sampling (Rao et al. 1979) scaled down near the solute (Mezei 1991); solute positions were held fixed. Our previous simulations (Mezei et al. 2004) of 10^8 Monte Carlo steps were extended 5×10^8 steps. Configurations were saved every 25,000 steps; in all, 20,000 configurations were included. Every configuration was embedded in a cubic 0.5 Å grid, and each grid volume element was scored for occupancy by a water oxygen center. Occupancy probabilities were calculated for each grid element and formatted as a CNS density map (Brunger et al. 1998), commonly used in X-ray crystallography. Probability density maps and molecular models were displayed using PyMOL (DeLano 2003).

Natural P_{II}/β-strand preferences were calculated from the Protein Coil Library (www.roselab.jhu.edu/coil) a non-α-helix, non-β-strand, non-β-turn fragment data base extracted from the PDB (Berman et al. 2000). In this library, secondary structures for each protein model in the PDB are classified solely by dihedral angles (Srinivasan and Rose 1999): an α-helix is defined as having at least five consecutive residues in helical conformation, a β-strand as having at least three consecutive residues in strand conformation, and a β-turn as having two consecutive residues in one of the eight major turn conformations (Rose et al. 1985). Residues not included in these three categories comprise the coil library. In the current study, all coil library residues from a data set of nonhomologous protein X-ray crystallographic structures were classified individually into right-handed helix, left-handed helix, strand, P_{II}, or coil regions of the Ramachandran map (Ramachandran et al. 1963). A nonhomologous protein data set with sequence identity $\leq 90\%$, resolution ≤ 2 Å, and R-factor ≤ 0.25 was obtained from the PISCES Web site (Wang and Dunbrack 2003). Residues from this data set found in the coil library ranged from 250,830 for proline to 4889 for tryptophan, with a mean of 30,200.

Acknowledgments

We thank Nicholas Panasik, Timothy Street, Haipeng Gong, and Buzz Baldwin for critical discussions. Support from the Mathers Foundation (G.D.R.) and NIH grant CA-63317 (M.M.) is gratefully acknowledged.

Note added in proof

CHASA software and a Web service to calculate CHASA-related parameters can be found at www.roselab.jhu.edu/chasa.

References

Anderson, A.G. and Hermans, J. 1988. Microfolding: Conformational probability map for the alanine dipeptide in water from molecular dynamics simulations. *Proteins* **3:** 262–265.

Apostolakis, J., Ferrara, P., and Caflisch, A. 1999. Calculation of conformational transitions and barriers in solvated systems: Application to the alanine dipeptide in water. *J. Chem. Phys.* **110:** 2099–2108.

Avbelj, F. and Baldwin, R.L. 2003. Role of backbone solvation and electrostatics in generating preferred peptide backbone conformations: Distributions of ϕ. *Proc. Natl. Acad. Sci.* **100:** 5742–5747.

Avbelj, F., Luo, P., and Baldwin, R.L. 2000. Energetics of the interaction between water and the helical peptide group and its role in determining helix propensities. *Proc. Natl. Acad. Sci.* **97:** 10786–10791.

Baker, B.M. and Murphy, K.P. 1998. Prediction of binding energetics from structure using empirical parameterization. *Methods Enzymol.* **295:** 294–315.

Bartels, C. and Karplus, M. 1997. Multidimensional adaptive umbrella sampling: Applications to main chain and side chain peptide conformations. *J. Comp. Chem.* **18:** 1450–1462.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Res.* **28:** 235–242.

Brooks, C.L. and Case, D.A. 1993. Simulations of peptide conformational dynamics and thermodynamics. *Chem. Rev.* **93:** 2487–2502.

Brunger, A., Adams, P., Clore, G., DeLano, W., Gros, P., Grosse-Kunstleve, R., Jiang, J., Kuszewski, J., Nilges, M., Pannu, N., et al. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54:** 905–921.

Chan, H. and Dill, K. 1997. Solvation: How to obtain microscopic energies from partitioning and solvation experiments. *Annu. Rev. Biophys. Biomol. Struct.* **26:** 425–459.

Chellgren, B.W. and Creamer, T.P. 2004. Short sequences of non-proline residues can adopt the polyproline II helical conformation. *Biochemistry* **43:** 5864–5869.

Chou, P.Y. and Fasman, G.D. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol. Relat. Areas Mol. Biol.* **47:** 45–148.

Creamer, T.P. and Rose, G.D. 1992. Side-chain entropy opposes α-helix formation but rationalizes experimentally determined helix-forming propensities. *Proc. Natl. Acad. Sci.* **89:** 5937–5941.

DeLano, W. 2003. *The PyMOL molecular graphics system.* DeLano Scientific LLC, San Carlos, CA.

Drozdov, A.N., Grossfield, A., and Pappu, R.V. 2004. Role of solvent in determining conformational preferences of alanine dipeptide in water. *J. Am. Chem. Soc.* **126:** 2574–2581.

Eker, F., Griebenow, K., Cao, X., Nafie, L.A., and Schweitzer-Stenner, R. 2004. Preferred peptide backbone conformations in the unfolded state revealed by the structure analysis of alanine-based (AXA) tripeptides in aqueous solution. *Proc. Natl. Acad. Sci.* **101:** 10054–10059.

Ferrara, P., Apostolakis, J., and Caflisch, A. 2002. Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins* **46:** 24–33.

Ferreon, J.C. and Hilser, V.J. 2004. Thermodynamics of binding to SH3 domains: The energetic impact of polyproline II (PII) helix formation. *Biochemistry* **43:** 7787–7797.

Fiebig, K.M., Schwalbe, H., Buck, M., Smith, L.J., and Dobson, C.M. 1996. Toward a description of the conformations of denatured states of proteins. Comparison of a random coil model with NMR measurements. *J. Phys. Chem. B* **100:** 2661–2666.

Gallicchio, E., Kubo, M.M., and Levy, R.M. 2000. Enthalpy-entropy and cavity decomposition of alkane hydration free energies: Numerical results and implications for theories of hydrophobic solvation. *J. Phys. Chem. B* **104:** 6271–6285.

Garcia, A.E. 2004. Characterization of non-α helical conformations in Ala peptides. *Polymer* **45:** 669–676.

Hamburger, J.B., Ferreon, J.C., Whitten, S.T., and Hilser, V.J. 2004. Thermo-

dynamic mechanism and consequences of the polyproline II structural bias in the denatured states of proteins. *Biochemistry* **43:** 9790–9799.

Han, W.-G., Jalkanen, K.J., Elstner, M., and Suhai, S. 1998. Theoretical study of aqueous N-Acetyl-L-alanine N′-Methylamide: Structures and Raman, VCD, and ROA spectra. *J. Phys. Chem. B* **102:** 2587–2602.

Hermann, R.B. 1972. Theory of hydrophobic bonding. 2. Correlation of hydrocarbon solubility in water with solvent cavity surface-area. *J. Phys. Chem.* **76:** 2754–2759.

Hilser, V.J., Gomez, J., and Freire, E. 1996. The enthalpy change in protein folding and binding: Refinement of parameters for structure-based calculations. *Proteins* **26:** 123–133.

Horton, N. and Lewis, M. 1992. Calculation of the free energy of association for protein complexes. *Protein Sci.* **1:** 169–181.

Hu, H., Elstner, M., and Hermans, J. 2003. Comparison of a QM/MM force field and molecular mechanics force fields in simulations of alanine and glycine "dipeptides" (Ace-Ala-Nme and Ace-Gly-Nme) in water in relation to the problem of modeling the unfolded peptide backbone in solution. *Proteins* **50:** 451–463.

Jorgensen, W.L., Chandrasekhar, J., and Madura, J.D. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79:** 926–935.

Kentsis, A., Mezei, M., Gindin, T., and Osman, R. 2004. Unfolded state of polyalanine is a segmented polyproline II helix. *Proteins* **55:** 493–501.

Lee, B. and Richards, F.M. 1971. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* **55:** 379–400.

Lovell, S.C., Word, J.M., Richardson, J.S., and Richardson, D.C. 2000. The penultimate rotamer library. *Proteins* **40:** 389–408.

Luo, P. and Baldwin, R.L. 1999. Interaction between water and polar groups of the helix backbone: An important determinant of helix propensities. *Proc. Natl. Acad. Sci.* **96:** 4930–4935.

MacKerell, A., Bashford, D., Bellott, M., Dunbrack, R., Evanseck, J., Field, M., Fischer, S., Gao, J., Guo, H., Ha, S., et al. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem.* **102:** 3586–3616.

Makhatadze, G.I. and Privalov, P.L. 1993. Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration. *J. Mol. Biol.* **232:** 639–659.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21:** 1087–1092.

Mezei, M. 1991. Distance-scaled force biased Monte Carlo simulation for solutions containing a strongly interacting solute. *Molecular Simulations* **5:** 405–408.

Mezei, M., Fleming, P.J., Srinivasan, R., and Rose, G.D. 2004. Polyproline II helix is the preferred conformation for unfolded polyalanine in water. *Proteins* **55:** 502–507.

Murphy, K.P., Xie, D., Garcia, K.C., Amzel, L.M., and Freire, E. 1993. Structural energetics of peptide recognition: Angiotensin II/antibody binding. *Proteins* **15:** 113–120.

Ooi, T., Oobatake, M., Nemethy, G., and Scheraga, H.A. 1987. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. *Proc. Natl. Acad. Sci.* **84:** 3086–3090.

Pappu, R.V. and Rose, G.D. 2002. A simple model for polyproline II structure in unfolded states of alanine-based peptides. *Protein Sci.* **11:** 2437–2455.

Penel, S. and Doig, A.J. 2001. Rotamer strain energy in protein helices—quantification of a major force opposing protein folding. *J. Mol. Biol.* **305:** 961–968.

Pettitt, B.M. and Karplus, M. 1988. Conformational free-energy of hydration for the alanine dipeptide&mash;thermodynamic analysis. *J. Phys. Chem.* **92:** 3994–3997.

Petukhov, M., Rychkov, G., Firsov, L., and Serrano, L. 2004. H-bonding in protein hydration revisited. *Protein Sci.* **13:** 2120–2129.

Privalov, P.L. and Makhatadze, G.I. 1993. Contribution of hydration to protein folding thermodynamics. II. The entropy and Gibbs energy of hydration. *J. Mol. Biol.* **232:** 660–679.

Ramachandran, G.N., Ramakrishnan, C., and Sasisekharan, V. 1963. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7:** 95–99.

Rao, M., Pangali, C., and Berne, B.J. 1979. On the force bias Monte Carlo simulation of water: Methodology, optimization and comparison with molecular dynamics. *Mol. Phys.* **37:** 1773.

Rathore, N., Knotts, T.A.t., and de Pablo, J.J. 2003. Configurational temperature density of states simulations of proteins. *Biophys. J.* **85:** 3963–3968.

Resat, H., Maye, P.V., and Mezei, M. 1997. The sensitivity of conformational free energies of the alanine dipeptide to atomic site charges. *Biopolymers* **41:** 73–81.

Rose, G.D., Gierasch, L., and Smith, J.A. 1985. *Turns in peptides and proteins*, pp. 1–109. Academic Press, New York.

Russo, D., Hura, G., and Head-Gordon, T. 2004. Hydration dynamics near a model protein surface. *Biophys. J.* **86:** 1852–1862.

Scarsi, M., Apostolakis, J., and Caflisch, A. 1998. Comparison of a GB solvation model with explicit solvent simulations: Potentials of mean force and conformation preferences of alanine dipeptide and 1,2-dicloroethane. *J. Phys. Chem.* **102:** 3637–3641.

Shrake, A. and Rupley, J.A. 1973. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* **79:** 351–371.

Smart, J.L., Marrone, T.J., and McCammon, J.A. 1997. Conformational sampling with Poisson-Boltzmann forces and a stochastic dynamic Monte Carlo method: Application to alanine dipeptide. *J. Comp. Chem.* **18:** 1750–1759.

Smith, P.E. 1999. The alanine dipeptide free energy surface in solution. *J. Chem. Phys.* **111:** 5568–5579.

Smith, L.J., Bolin, K.A., Schwalbe, H., MacArthur, M.W., Thornton, J.M., and Dobson, C.M. 1996. Analysis of main chain torsion angles in proteins: Prediction of NMR coupling constants for native and random coil conformations. *J. Mol. Biol.* **255:** 494–506.

Spolar, R.S., Livingstone, J.R., and Record, M.T. 1992. Use of liquid hydrocarbon and amide transfer data to estimate contributions to thermodynamic functions of protein folding from the removal of nonpolar and polar surface from water. *Biochemistry* **31:** 3947–3955.

Srinivasan, R. and Rose, G.D. 1995. LINUS: A hierarchic procedure to predict the fold of a protein. *Proteins* **22:** 81–99.

———.1999. A physical basis for protein secondary structure. *Proc. Natl. Acad. Sci.* **96:** 14258–14263.

Srinivasan, R., Fleming, P.J., and Rose, G.D. 2004. Ab initio protein folding using LINUS. *Methods Enzymol.* **383:** 48–66.

Swindells, M.B., MacArthur, M.W., and Thornton, J.M. 1995. Intrinsic ϕ, ψ propensities of amino acids, derived from the coil regions of known structures. *Nat. Struct. Biol.* **2:** 596–603.

Taylor, R. and Kennard, O. 1984. Hydrogen-bond geometry in organic crystals. *Acc. Chem. Res.* **17:** 320–326.

Tazaki, K. and Shimizu, K. 1998. Molecular dynamics simulations in aqueous solution: Application to free energy calculation of oligopeptides. *J. Phys. Chem. B* **102:** 6419–6424.

Thomas, S.T., Loladze, V.V., and Makhatadze, G.I. 2001. Hydration of the peptide backbone largely defines the thermodynamic propensity scale of residues at the C′ position of the C-capping box of α-helices. *Proc. Natl. Acad. Sci.* **98:** 10670–10675.

Tobias, D.J. and Brooks, C.L. 1992. Conformational equilibrium in the alanine dipeptide in the gas-phase and aqueous-solution: A comparison of theoretical results. *J. Phys. Chem.* **96:** 3864–3870.

Vallone, B., Miele, A.E., Vecchini, P., Chiancone, E., and Brunori, M. 1998. Free energy of burying hydrophobic residues in the interface between protein subunits. *Proc. Natl. Acad. Sci.* **95:** 6103–6107.

Wang, G. and Dunbrack Jr., R.L. 2003. PISCES: A protein sequence culling server. *Bioinformatics* **19:** 1589–1591.

Wimley, W.C., Creamer, T.P., and White, S.H. 1996. Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides. *Biochemistry* **35:** 5109–5124.