Modeling Unfolded States of Peptides and Proteins†

Trevor P. Creamer, Rajgopal Srinivasan, and George D. Rose*

Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205

Received October 5, 1995; Revised Manuscript Received November 13, 1995®

ABSTRACT: The hydrophobic effect is the major factor that drives a protein molecule toward collapse and folding. In this process, residues with apolar side chains associate to form a solvent-shielded hydrophobic core. Often, this hydrophobic contribution to folding is quantified by measuring buried apolar surface area, reckoned as the difference in area between hydrophobic groups in the folded protein and in a standard state. Typically, the standard state area of a residue is obtained from tripeptide models, with the results taken to implicitly represent values appropriate for the unfolded state. Here, we show that a tripeptide is a poor descriptor of the unfolded state, and its widespread use has prompted erroneous conclusions about folding. As an alternative, we explore two limiting models, chosen to bracket the expected behavior of the unfolded chain between reliable extremes. One extreme is represented by simulated hard-sphere peptides and shown to behave like a homopolymer with excluded volume in a good solvent. The other extreme is represented by fragments excised from folded proteins and conjectured to approximate the time-average behavior of a thermally denatured protein at T_m, the midpoint of the unfolding transition. Using these models, it is shown that the area buried by apolar side chains upon folding is considerably less than earlier estimates. For example, upon transfer from the unfolded state to the middle of an α-helix, an alanine side chain buries negligible area and, surprisingly, a valine side chain actually gains area. Among other applications, an improved model of the unfolded state can be used to better evaluate the driving force for helix formation in peptides and proteins.

Protein folding is defined as the transition of a protein from an unfolded state to the native fold under suitable conditions (Anfinsen, 1973). Most structural studies have focused on the native fold, both because biological properties of proteins are a consequence of their structure and because unfolded states are harder to characterize. Nonetheless, the folding process cannot be understood without a quantitative description of the unfolded states of proteins (Dill & Shortle, 1991).

Since the influential review by Kauzmann (1959), the hydrophobic interaction has been accepted as the driving force in protein folding (Dill, 1990). During folding, hydrophobic groups are expelled spontaneously from water, thereby engendering a sequestered, solvent-shielded core. To quantify this effect, Lee and Richards (1971) and Shrake and Rupley (1973) devised algorithms to calculate the solvent-accessible surface area (ASA) of protein molecules. Early numerical approximations of the ASA were followed by analytic treatments (Connolly, 1983; Richmond, 1984). Detailed definitions of the ASA and a closely related quantity called the molecular surface area are found in the authoritative review by Richards (1974).

The reduction of apolar surface area is central to the folding process, and thousands of studies have utilized the Lee and Richards algorithm since its inception. Yet, the ASA lacks the thermodynamic rigor of the other geometric variable of interest—volume. This comparative deficit is due to the fact that the area of the unfolded state has never been uniquely defined. Further, no instrument has been developed to measure the area change upon folding, although the calorimeter shows promise of becoming one (Murphy & Freire, 1992; Privalov & Makhatadze, 1992; Spolar *et al.,* 1992). Indeed, one is hard-pressed to think of another quantity of such central importance in protein thermodynamics that cannot be determined by direct measurement. Given these uncertainties, it is hardly surprising that debate is ongoing over the interpretation of area changes during protein folding (Sharp *et al.,* 1991; Rose & Wolfenden, 1993; Lee, 1994).

To calculate the area lost by residue Xaa upon folding, the area in both native and unfolded states is required. When the three-dimensional structure is known, the ASA of the native state can be obtained by straightforward application of the Lee and Richards (1971) or an equivalent algorithm. In contrast, the ASA of the unfolded state is ill-defined, and typically, a standard state is used instead. Such standard state models include a single blocked residue (Sneddon & Tobias, 1992), tripeptides of sequence Gly-Xaa-Gly (Shrake & Rupley, 1973; Rose *et al.,* 1985a; Miller *et al.,* 1987; Lesser & Rose, 1990), Ala-Xaa-Ala tripeptides (Zielenkiewicz & Saenger, 1992), and Xaa in the context of the protein with flanking backbone and side chains fully extended (Livingstone *et al.,* 1991). All are valid standard state models that provide a defined reference against which the buried area of a residue in a folded protein can be compared. But, none model the unfolded states of proteins.

In this work we examine two extreme reference states: (i) simulated flexible peptides modeled using a hard-sphere approximation and (ii) peptide fragments extracted from high-resolution protein structures. These models bracket the expected behavior of short unfolded peptides between reliable

† Supported by a grant from the National Institutes of Health (GM 29458).

* Corresponding author. Phone: (410) 614-3970. Fax: (410) 614-3971.

0006-2960/95/0434-16245$09.00/0

extremes. Data from hard-sphere simulations serve as upper bounds for both accessible surface area and chain dimensions. Interactions in such peptides are confined to excluded volume effects, resulting in chains that are more expanded than would be expected for unfolded peptides of heterogeneous composition in any actual solvent. At the other extreme, protein fragments excised from fully folded structures define corresponding lower bounds because they will be more compact than would be expected for unfolded peptides. Under attainable unfolding conditions, the surface area of a peptide will lie somewhere between these two limits. These results are used to show that tripeptide and extended state models do not adequately represent the unfolded state of a protein and, if used as such, greatly overestimate the area loss upon folding.

As a specific example, we show that, on average, an alanine side chain in the center of an 11-residue, unfolded polyalanyl peptide loses little or no area upon helix formation, and a valine side chain can actually *gain* area in the helix. In sharp contrast, use of a tripeptide to model the unfolded state leads to the erroneous conclusion that both alanine and valine side chains lose substantial surface area upon helix formation.

Finally, we propose an alternative to the tripeptide model. No generic substitute can be expected because residue accessibility in the unfolded state will depend on denaturation conditions. However, segments abstracted from proteins of known structure may provide plausible—albeit crude—models for estimating the side-chain accessibility of a thermally denatured protein at the midpoint of the unfolding transition.

METHODS

In this work, peptide unfolded states were modeled using both computer simulations and data extracted from high-resolution X-ray structures. Monte Carlo sampling and a hard-sphere approximation were used in the computer simulations. The hard-sphere approximation treats atoms as having only excluded volume effects with no attractive components. In detail, acetylated and methylamidated polyalanine and polyvaline peptides of varying lengths N (N = 3, 5, 7, ..., 15) were modeled using methods developed in earlier work (Creamer & Rose, 1994). A heterogeneous peptide of sequence Ace-$(Ala)_x$-Val-$(Ala)_x$-Nme ($2x + 1 = N$, N = 3, 5, 7, ..., 15) was also modeled. In the simulations, peptides were allowed to sample available conformational space freely: random rotations were made about backbone dihedral angles, ϕ and ψ, and side-chain torsion angles, χ. Bond lengths and angles were fixed at the values listed in Table 1, and peptide units were held rigid and planar (ω = 180°). The united atoms approximation was used in which CH, CH_2, and CH_3 moieties are treated as single atoms with inflated radii. Atomic radii are given in Table 1.

In analysis of heterogeneous peptides excised from proteins of known structure, all peptide fragments of length N = 5, 6, 7, ..., 40 were extracted from 43 well-refined, high-resolution protein chains (R-factor of 20% or better, resolution of 2.0 Å or better) from the Brookhaven Protein Data Bank (PDB) (Bernstein *et al.*, 1977). Extracted fragments start with the amide nitrogen of the first residue and end with the carbonyl group of the last residue. Side chains for all residues within the fragment were included. The four-letter PDB codes for the 43 protein structures used are 1bp2, 1crn, 1ecd, 1gcr, 1gd1(O), 1gp1(A), 1hmq(A), 1hoe, 1lz1, 1mbo, 1ppt, 1rdg, 1sn3, 1snc, 1tpp, 1ubq, 2act, 2aza(B), 2ca2, 2cdv, 2cts, 2lhb, 2ovo, 2pcy, 2rhe, 2wrp(R), 351c, 3app, 3grs, 3ins(C), 3ins(D), 3lzm, 3rnt, 3tln, 4dfr(B), 4fxn, 4pep, 5cha(A), 5cpa, 5cyt(R), 5pti, 7rsa, and 9pap. In multimeric structures, the subunits used in the analyses are shown in parentheses.

Table 1: Atomic Radii and Covalent Geometry Used in the Hard-Sphere Monte Carlo Simulations[a]

	Atomic Radii (Å)[b]		
N	1.35	CA	1.65
HN	1.00	CB	1.65
C	1.50	CG	1.65
O	1.35		
	Bond Lengths (Å)[c]		
N−CA	1.46	N−HN	1.10
CA−C	1.51	CA−CB	1.54
C−N	1.32	C−O	1.24
CB−CG	1.54		
	Bond Angles (deg)[c]		
C−N−CA	122	N−CA−CB	108
N−CA−C	111	C−CA−CB	113
CA−C−N	116	CA−CB−CG	116
C−N−HN	116	CA−C−O	122
CA−N−HN	122	N−C−O	122

[a] N = amide nitrogen, HN = amide hydrogen, CA = α-carbon, C = carbonyl carbon, O = carbonyl oxygen, CB = β-carbon, and CG = γ-carbon. [b] From Hopfinger (1973). [c] From the Biopolymer module of InsightII, Biosym Technologies.

Accessible surface areas were calculated according to the method of Richmond (1984) using the atomic radii of Richards (1977) and with a probe radius set to that of a water molecule (1.4 Å). The mean square end-to-end distance, $\langle r^2 \rangle$, of all simulated peptides and protein fragments was calculated. In simulated peptides, the end-to-end distance was measured between the methyl moieties of the terminating acetyl and methylamide groups. In protein fragments, the end-to-end distance was measured between the terminal α-carbons.

RESULTS

Conformational Behavior. The conformational behavior of both simulated peptides and protein fragments can be assessed from their average chain dimensions (Flory, 1969). Figure 1 illustrates the behavior of the mean square end-to-end distance, $\langle r^2 \rangle$, of simulated polyalanine, polyvaline, and Ace-$(Ala)_x$-Val-$(Ala)_x$-Nme. Figure 2 is a similar plot of protein fragments. Flory (1969) showed that simple homopolymers obey a scaling law of the form

$$\langle r^2 \rangle \sim N^{\upsilon} \qquad (1)$$

where N is the number of repeat units in the polymer chain. The Flory exponent υ is 6/5 for a polymer chain with excluded volume in a good solvent and 2 for rigid rods (Flory, 1969; Chan & Dill, 1991). The data from Figure 1 have been fit to eq 1, with the resulting exponents given in Table 2. Polyalanine peptides have an exponent of υ = 1.18, indicating a homopolymer with excluded volume. This result is expected for a hard-sphere model peptide with small side chains. Simulated polyvaline peptides have a slightly larger Flory exponent of 1.31 (Table 2), indicative of the somewhat stiffer polymer chain that results from steric effects imposed by the β-branched side chains upon the peptide backbone.

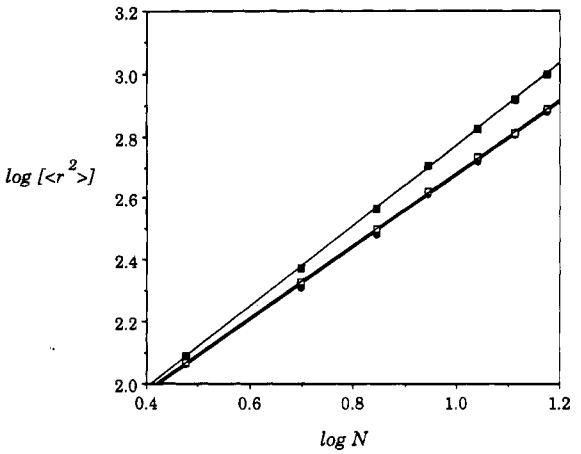

FIGURE 1: log of mean square end-to-end distances, $\langle r^2 \rangle$, of simulated peptides against log of the number of residues in the peptide chain, N. Polyalanine peptides are represented by the filled circles, polyvaline by the filled squares, and Ace-$(Ala)_x$-Val-$(Ala)_x$-Nme by the open squares.

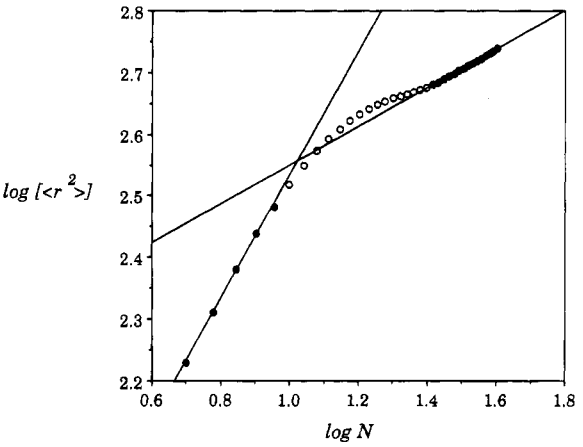

FIGURE 2: log of mean square end-to-end distances, $\langle r^2 \rangle$, of protein fragments against log of the number of residues in the fragments, N. Filled circles denote those fragments used to calculate Flory exponents (see text).

Table 2: Flory Exponents of Simulated Peptides and Protein Fragments

model	υ
simulated peptides	
Ace-$(Ala)_N$-Nme	1.18
Ace-$(Val)_N$-Nme	1.31
Ace-$(Ala)_x$-Val-$(Ala)_x$-Nme	1.18
protein fragments	
$5 \leq N \leq 9$	1.00
$26 \leq N \leq 40$	0.31

Although the mean square end-to-end distance computed in the simulations will be sensitive to the covalent geometry and atomic radii used, the Flory exponents are expected to be relatively insensitive.

The Ace-$(Ala)_x$-Val-$(Ala)_x$-Nme peptides are not strict homopolymers, but they too have a Flory exponent of 6/5, that of a homopolymer with excluded volume in a good solvent. The polyalanine and Ace-$(Ala)_x$-Val-$(Ala)_x$-Nme plots are virtually coincident in Figure 1, implying that insertion of a single valine in the center of a hard-sphere polyalanine peptide has a negligible effect upon the backbone conformational behavior of the peptide as a whole. Of course, different results would be expected in force fields with an attractive component.

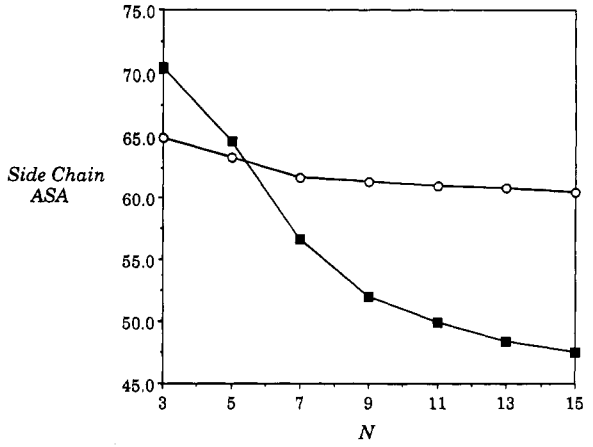

FIGURE 3: Side-chain ASA of the central alanine residue as a function of chain length for the modeled polyalanine peptides (open circles) and for alanine residues at the center of protein fragments (filled squares).

Protein fragments display both short-range and long-range behavior (Figure 2). Emphatically, these are not homopolymers. Nonetheless, eq 1 was fit separately to shorter (N = 5−9) and longer (N = 26−40) fragments to assess average backbone conformational properties (Table 2). The short-range exponent, υ = 1.00, is indicative of Gaussian chain behavior, a freely jointed chain without excluded volume constraints. Fragments drawn from protein structures are clearly not Gaussian—they possess significant excluded volume and are not freely jointed. In this case, they are highly collapsed because they are excised from folded proteins, which are known to be exquisitely packed (Richards, 1974).

Longer protein fragments are even more compact. The Flory exponent obtained for $26 \leq N \leq 40$ is very small, υ = 0.31. These fragments exceed the average length of secondary structure elements and will encompass supersecondary structure or even small domains. (Supersecondary structure consists of two or more interacting elements of secondary structure.) Such fragments are expected to have compactness that rivals whole proteins (Zehfus & Rose, 1986), more compact by far than unfolded hard-sphere peptides.

Surface Area. In Figure 3, the average accessible surface area (ASA) of the side chain in the central residue of simulated hard-sphere polyalanyl peptides is plotted as a function of peptide length. The ASA of a central alanine side chain in protein fragments is also plotted. In simulated peptides, the central alanine side chain loses about 4.5 Å^2 between N = 3 and N = 15, and its value becomes asymptotic near N = 11. In protein fragments, the central alanine side chain loses more than 20 Å^2 as the length increases from N = 3 to N = 15, with values that do not level off.

Figure 4 is a similar plot of simulated polyvaline and Ace-$(Ala)_x$-Val-$(Ala)_x$-Nme peptides, together with the average ASA of central valine side chains in protein fragments. In simulated hard-sphere polyvalyl peptides, the average ASA of the central valine side chain becomes asymptotic near N = 9, losing about 15 Å^2 between N = 3 and N = 15. The ASA of the Val side chain in the Ace-$(Ala)_x$-Val-$(Ala)_x$-Nme peptides also becomes asymptotic near N = 9 but loses only about 10 Å^2 between N = 3 and N = 15. The central valine side chain in protein fragments loses ~40 Å^2 from N = 3 to

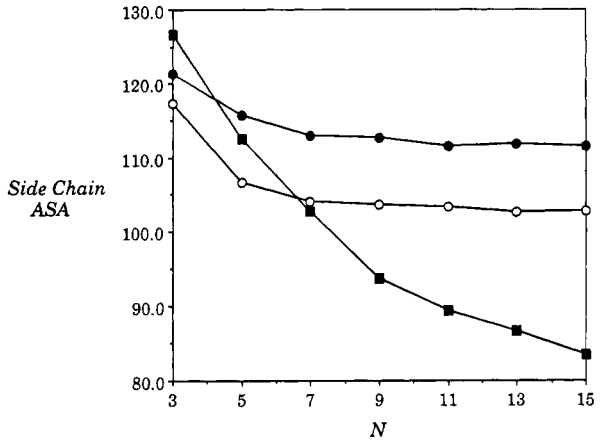

FIGURE 4: Side-chain ASA of the central valine residue as a function of chain length for modeled polyvaline peptides (open circles), modeled Ace-$(Ala)_x$-Val-$(Ala)_x$-Nme peptides (filled circles), and valine residues at the center of protein fragments (filled squares).

$N = 15$, with values that do not level off.

The central side chains of shorter simulated peptides, $N = 3$ or 5, are less exposed than their counterparts in protein fragments (Figures 3 and 4), presumably because turnlike conformations are more heavily populated in protein fragments than in simulated peptides. Of necessity, side chains in turns point outward, with minimal shielding from neighboring residues (Rose *et al.*, 1985b). In contrast, simulated peptides spend a substantial fraction of time with one or more residues in extended conformations. In the hard-sphere approximation, all allowed conformations are equiprobable, and as shown by Ramachandran and Sasisekharan (1968), extended conformations dominate the allowed conformational space. In comparison with a turn, a side chain at position i in an extended conformation is more likely to be partially shielded from solvent, either by the peptide backbone or by side chains at positions $i \pm 2$. The fact that three-residue protein fragments bury less surface than corresponding side chains in simulated peptides was noted by Zielenkiewicz and Saenger (1992).

The ASA behavior of alanine and valine side chains in protein fragments typifies that of all residues (Figures 3 and 4). Figure 5 shows ASA data for 19 side-chain types (excluding Gly which lacks a side chain) derived from protein fragments, for lengths of $N \leq 25$. In all cases, the ASA becomes asymptotic with fragment length. However, at longer lengths, $\sim N = 40$, a pronounced decrease in side-chain ASA appears as fragments increasingly encompass supersecondary structure and small subdomains (data not shown).

DISCUSSION

Understanding the behavior of unfolded peptides and proteins is prerequisite to understanding the folding process (Dill & Shortle, 1991). Short extended peptides—often just tripeptides—are used commonly to model peptides and proteins (Shrake & Rupley, 1973; Rose *et al.*, 1985a; Miller *et al.*, 1987; Lesser & Rose, 1990; Zielenkiewicz & Saenger, 1992). Alternatively, the protein itself is used, with backbone and side chains extended fully (Livingstone *et al.*, 1991). Both models overestimate the surface area buried upon folding. In this work, two models were used to provide

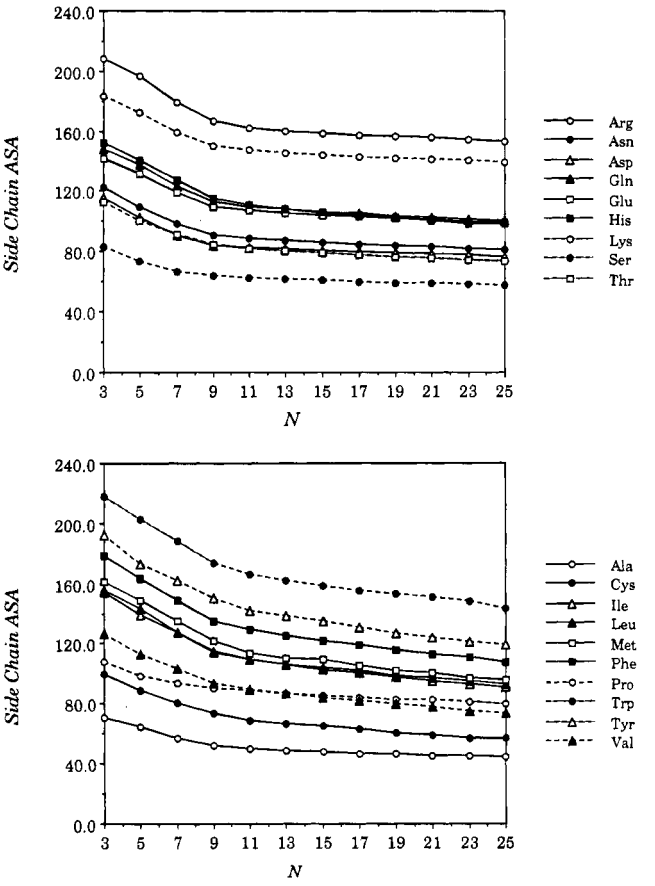

FIGURE 5: Side-chain ASA for the central side chains in the protein fragments as a function of chain length: (a, top) polar residues and (b, bottom) apolar residues.

upper (hard-sphere peptides) and lower (protein fragments) bounds for the behavior of unfolded peptides in solution.

These two models appear to bracket the expected conformational behavior of unfolded peptides in solution, consistent with their Flory exponents (Table 2, Figure 1). In simulations, use of the hard sphere results in expanded peptides that explore available conformational space freely. Actual peptides would experience intramolecular attractive forces, leading to further chain collapse. Although simulated peptides are more expanded than actual unfolded peptides, they are nonetheless more compact than fully extended peptides.

At the other extreme, protein fragments are very compact (Table 2 and Figure 2) because they were excised from folded structures, where the mean packing density resembles that of small organic crystals (Richards, 1974). By definition, unfolded peptides and proteins do not contain long-lived structural elements such as turns, helices, strands, and loops; on average, they will be more expanded in solution than corresponding protein fragments. Therefore, the conformational behavior of unfolded peptides will lie somewhere between that of simulated hard-sphere peptides and excised protein fragments.

Differences in conformational properties of the models give rise to differences in ASA. For lengths of $N > 5$, the central side chains in all three hard-sphere peptides—polyalanine, polyvaline, and Ace-$(Ala)_x$-Val-$(Ala)_x$-Nme—are more exposed to solvent than in protein fragments of equal length, for two reasons. (i) On average, hard-sphere peptides are more expanded than their counterparts in proteins, and

Table 3: Mean Accessible Surface Areas of Side Chains from Protein Fragments

	N								
residue	3	5	7	9	11	13	17	21	25
Ala	70.4	64.6	56.7	52.0	49.9	48.4	46.6	45.3	44.1
Arg	208.5	196.4	179.1	166.7	161.9	160.0	156.9	155.4	152.9
Asn	122.2	109.2	98.2	90.9	88.8	87.0	84.5	82.8	80.8
Asp	115.2	102.4	90.4	84.0	82.1	81.3	79.2	78.5	76.3
Cys	99.7	88.6	80.0	73.1	68.5	66.6	62.9	59.0	56.4
Gln	148.2	137.6	123.0	112.4	109.0	107.6	105.0	102.2	100.6
Glu	141.8	131.7	118.8	109.5	106.9	105.1	102.8	100.8	99.2
His	152.4	140.1	127.5	115.1	110.5	107.9	103.9	100.4	98.2
Ile	154.0	139.3	127.3	114.6	109.3	105.6	100.1	95.1	90.9
Leu	155.9	143.0	127.1	114.3	109.0	105.5	101.4	96.6	92.8
Lys	183.0	172.5	159.0	149.8	147.1	145.3	142.5	141.4	139.1
Met	161.1	148.9	135.1	121.4	113.5	110.1	105.3	100.3	95.3
Phe	178.5	163.2	149.0	134.9	129.0	125.1	118.7	112.7	107.4
Pro	108.2	98.4	93.4	90.1	88.4	86.7	83.5	82.1	79.5
Ser	82.7	73.0	66.7	63.7	62.4	61.5	59.7	58.5	57.5
Thr	112.6	100.2	91.5	84.6	81.7	80.0	77.3	75.3	73.4
Trp	218.1	202.9	188.2	173.3	165.8	161.7	154.7	150.7	143.4
Tyr	192.0	173.1	162.0	150.4	142.0	138.4	131.0	124.1	119.1
Val	126.7	112.6	102.8	93.7	89.4	86.7	81.8	77.2	73.0

(ii) protein fragments are not homopolymers; neighboring residues with heterogeneous side chains provide solvent shielding for the central residue. Given these considerations, the two models will also establish upper (hard-sphere peptides) and lower (protein fragments) boundaries for surface area exposure.

From Figures 3–5, it is clear that a tripeptide standard state overestimates the ASA lost by a side chain on folding. For example, Lesser and Rose (1990) calculate the average ASA of the alanine side chain to be 72 Å^2 in their Gly-Xaa-Gly stochastic standard state and 21 Å^2 in proteins. The difference implies an area loss of 51 Å^2 on folding, if the standard state is taken as the unfolded state. A protein fragment or model peptide of length $N = 15$ would afford a better estimate of the ASA in the unfolded state (Figures 3 and 5). In protein fragments, an alanine side chain has an ASA of ~48 Å^2 and, in simulations, an ASA of 61 Å^2. These extremes bracket the average loss of ASA on folding between 27 and 40 Å^2. Consequently, use of a tripeptide overestimates the area loss of the methyl group by 25%, at least.

It has been suggested that the burial of side-chain hydrophobic surface (Blaber *et al.*, 1994), the β-carbon in particular (Yang & Honig, 1995), is a major determinant of the helix-forming tendencies of naturally occurring amino acids. The average ASA of an alanine side chain in the center of an 11-residue polyalanine model helix is 51.5 Å^2 (Creamer & Rose, 1994). A corresponding side chain has an average ASA of 61.0 Å^2 in hard-sphere peptides and 49.9 Å^2 in protein fragments (for $N = 11$). From the difference, an alanine side chain loses between 10 and 0 Å^2 upon transfer from an unfolded state to a midhelical position. It seems unlikely that this difference makes a substantial contribution to the helix-forming tendency of alanine.

The situation is even more pronounced for valine. The average ASA of a valine side chain in the center of an 11-residue polyalanine model helix is 106.6 Å^2 (Creamer & Rose, 1994). The central valine side chain in an Ace-(Ala)$_x$-Val-(Ala)$_x$-Nme peptide and in protein fragments has average ASA values of 103.3 and 89.4 Å^2, respectively (for $N = 11$). Consequently, a valine side chain *exposes* between 0 and 17 Å^2 of hydrophobic surface upon transfer from an unfolded state to a midhelical position. This difference represents a small to negligible *unfavorable* contribution to the helix-forming tendency of valine.

From the foregoing, we argue that the extended tripeptide and related cognates do not adequately model solvent accessibility in the unfolded state. Simulation is unlikely to provide an improved model in the immediate future because differing force fields give widely varying results, and inclusion of explicit solvent leads to further divergence. Eventually, with suitable parameterization, it may be possible to deconvolute calorimetric data into polar and apolar areas (Murphy & Freire, 1992; Spolar *et al.*, 1992; Makhatadze & Privalov, 1995). Meanwhile, in response to colleagues requesting a practical alternative to the extended tripeptide, we hesitantly propose the following.

A Hesitant Conjecture. The term *unfolded state* is imprecise (Dill & Shortle, 1991). Protein residues are expected to have differing patterns of solvent accessibility under different unfolding conditions. What we seek is a well-defined experimental state in which equivalent chemical groups have comparable patterns of solvation from one molecule to another. The midpoint of the thermal unfolding transition, T_m, represents such a state. At this unique point, the equilibrium constant between folded and denatured forms is unity and the free energy difference is naught.

We propose the use of segments abstracted from proteins of known structure (Figures 4 and 5; Table 3) to estimate side-chain accessibility at T_m during thermal denaturation. The proposal is based on the conjecture that the unfolded state at the transition midpoint will still be highly self-associated, with many transient, native-like regions but diminished longer range interactions. To the degree that this conjecture is valid, protein fragments are expected to be plausible models, for several reasons. Unlike tripeptides, fragments are extensively self-associated. Further, fragments are devoid of longer range solvent-shielding interactions, although vestigial long-range interactions may affect these structures (*i.e.*, global "ghost" forces may still inhabit local segments). The Flory exponents (Table 2) show greater compactness in larger segments, as would be expected. Finally, Monte Carlo simulations have been effective at predicting secondary structure with interaction intervals limited to ~25 residues (Srinivasan & Rose, 1995), indicating

that secondary structure conformations and their resultant solvent accessibilities are determined by intrafragment interactions.

ACKNOWLEDGMENT

We thank Rajeev Aurora and George Makhatadze for useful discussions and helpful suggestions.

REFERENCES

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., & Tasumi, M. (1977) *J. Mol. Biol. 112*, 535−542.
Blaber, M., Zhang, X.-J., Lindstrom, J. D., Pepoit, S. D., Baase, W. A., & Matthews, B. M. (1994) *J. Mol. Biol. 235*, 600−624.
Chan, H. S., & Dill, K. A. (1991) *Annu. Rev. Biophys. Biophys. Chem. 20*, 447−490.
Connolly, M. L. (1983) *J. Appl. Crystallogr. 16*, 548−558.
Creamer, T. P., & Rose, G. D. (1994) *Proteins: Struct., Funct., Genet. 19*, 85−97.
Dill, K. A., & Shortle, D. (1991) *Annu. Rev. Biochem. 60*, 795−825.
Flory, P. J. (1969) *Statistical Mechanics of Chain Molecules*, John Wiley and Sons, New York.
Ginsburg, A., & Carroll, W. R. (1965) *Biochemistry 4*, 2159−2174.
Hopfinger, A. J. (1973) *Conformational Properties of Macromolecules*, Academic Press, New York.
Kauzmann, W. (1959) *Adv. Protein Chem. 14*, 1−64.
Lee, B. K. (1994) *Biophys. Chem. 51*, 263−269.
Lee, B. K., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379−400.
Lesser, G. J., & Rose, G. D. (1990) *Proteins: Struct., Funct., Genet. 8*, 6−13.
Livingstone, J. R., Spolar, R. S., & Record, M. T., Jr. (1991) *Biochemistry 30*, 4237−4244.
Makhatadze, G. I., & Privalov, P. L. (1995) *Adv. Protein Chem.* (in press).
Miller, S., Janin, J., Lesk, A. M., & Chothia, C. (1987) *J. Mol. Biol. 196*, 641−656.
Murphy, K. P., & Freire, E. (1992) *Adv. Protein Chem. 43*, 313−361.
Privalov, P. L., & Makhatadze, G. I. (1992) *J. Mol. Biol. 224*, 715−723.
Ramachandran, G. N., & Sasisekharan, V. (1968) *Adv. Protein Chem. 23*, 283−438.
Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng. 6*, 151−176.
Richmond, T. J. (1984) *J. Mol. Biol. 178*, 63−89.
Rose, G. D., & Wolfenden, R. (1993) *Annu. Rev. Biophys. Biomol. Struct. 22*, 381−415.
Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H., & Zehfus, M. H. (1985a) *Science 229*, 834−838.
Rose, G. D., Gierasch, L. M., & Smith, J. A. (1985b) *Adv. Protein Chem. 37*, 1−109.
Sharp, K. A., Nicholls, A., Friedman, R., & Honig, B. (1991) *Biochemistry 30*, 9686−9697.
Shrake, A., & Rupley, J. A. (1973) *J. Mol. Biol. 79*, 351−372.
Sneddon, S. F., & Tobias, D. J. (1992) *Biochemistry 31*, 2842−2846.
Spolar, R. S., Livingston, J. R., & Record, M. T., Jr. (1992) *Biochemistry 31*, 3947−3955.
Srinivasan, R., & Rose, G. D. (1995) *Proteins: Struct., Funct., Genet. 22*, 81−99.
Yang, A.-S., & Honig, B. (1995) *J. Mol. Biol. 252*, 351−365.
Zehfus, M. H., & Rose, G. D. (1986) *Biochemistry 25*, 5759−5765.
Zielenkiewicz, P., & Saenger, W. (1992) *Biophys. J. 63*, 1483−1486.

BI952385H