Residue solvent accessibilities in the unfolded polypeptide chain

Piotr Zielenkiewicz and Wolfram Saenger*
Institute of Biochemistry and Biophysics, Polish Academy of Sciences, 02532 Warszawa, Poland, *Institut für Kristallographie, Freie Universität Berlin, 1000 Berlin 33, Germany

ABSTRACT The difference of solvent accessibilities in the native and unfolded states of the protein is used as a measure of the hydrophobic contribution to the free energy of folding. We present a new approximation of amino acids solvent accessibilities in the unfolded state based on the 1-ns molecular dynamics simulation of Ala-X-Ala tripeptides at a temperature of 368 K. The standard accessibility values averaged from the molecular dynamics study are significantly lower from those previously obtained by considering only selected conformations of Ala-X-Ala tripeptides.

INTRODUCTION

The hydrophobic effect, described for proteins first by Kauzmann (1959), plays a key role in the formation and stability of the native protein structures. It is often semi-quantitatively measured by the loss of hydrophobic surface area accessible to the solvent upon folding (Chothia, 1975; Richards, 1977; Eisenberg and Mclachlan, 1986). This type of approach can also be successfully used for thermodynamic interpretation of stability and recognition processes of protein–protein complexes (Chothia and Janin, 1975).

For protein folding, the loss of accessible surface area is defined as the difference of accessibilities in the native state and the unfolded state. It is often assumed (Chothia, 1976; Janin, 1979; Eisenberg and McLachlan, 1986; Ooi et al., 1987), although the validity of this assumption has been questioned (Wood and Thompson, 1990), that the hydrophobic free energy is proportional to the surface area buried upon folding.

Several algorithms have been described which allow to calculate solvent accessibilities from atomic coordinates (e.g., Lee and Richards, 1971; Shrake and Rupley, 1973; Lavery and Pullman, 1981; Connolly, 1981; Richmond, 1984). It is therefore straightforward to calculate accessibilities in the native state if the tertiary structure of the protein under consideration has been determined.

The unfolded state, however, is poorly defined and the solvent accessibilities of the amino acids in the denatured polypeptide chain have to be approximated. This was done in earlier studies on a residue basis by calculating the solvent accessibility value for the central residue in Gly-X-Gly tripeptides.

Miller et al. (1987) define the standard state accessibility of residue X to be the surface area of X in an extended Gly-X-Gly tripeptide. Shrake and Rupley (1973) and Rose et al. (1985) used a stochastic standard state defined as mean solvent accessibility of X in Gly-X-Gly tripeptides with dihedral angles reflecting the observed distribution in the protein structures data base. Both these approaches tend to overestimate the standard state solvent accessibilities as in a real polypeptide chain the nearest neighbors of a residue are on average more bulky than glycine. They also do not account (although the stochastic method better than extended) for the real behavior of even the tripeptide under denaturing conditions. None of the methods, including the one presented below, can account for the behavior of the denatured polypeptide chain as a whole which, to an unknown extent, reduces solvent accessibilities of residues in the denatured protein.

In this paper we use a new method for calculating the standard state residue solvent accessibilities. It is based on averaging the surface area of a central residue in Ala-X-Ala tripeptides from a molecular dynamics trajectory at a temperature of 368 K.

METHODS

The computational procedure applied was the same for all the 20 Ala-X-Ala tripeptides studied. The starting structures for the simulations were prepared with the molecular editor program MOLEDT (Biosym Technologies, San Diego, CA). The tripeptides were constructed from the residue library of the molecular editor. Fully extended chains were taken as starting conformations. To avoid strong electrostatic attraction of charged termini and to simulate the polypeptide chain, they were methylated at the ends, CH_3-NH- and -$COOCH_3$. The X residues were taken in their electrostatically neutral forms. Hydrogen atoms were added using the molecular editor. The correctness of the starting structures was controlled on an Evans and Sutherland PS390 graphics device.

The Consistent Valence Force Field (Maple, Dinur, and Hagler, 1988) as implemented in DISCOVER (Ver. 2.6.0, Biosym) was used throughout the calculations. The dielectric constant was set to 80. To adjust the positions of hydrogen atoms the systems were initially treated with 100 steps of tethered steepest descents minimization with harmonic constraint of 20 kcal/Å on heavy atoms. This was followed by 30 steps of minimization with all the atoms free to move in order to reduce gradients before dynamics initialization. The molecular dynamics was run at a temperature of 368 K. After 20-ps equilibration the molecular dynamics trajectory was recorded for a period of 1 ns. Every 1,000 dynamics steps (1 ps) a snapshot of coordinates was taken giving 1,000 conformations for each tripeptide. For each of 1,000 conformations the solvent accessibility of the central residue was calculated and then averaged. For these calculations the program of Lavery and Pullman (1981) was used with the radius of the probe sphere at 1.4 Å and 987 points on each sphere. The solvent accessibility calculations were performed with three different sets of group radii (those of Shrake and Rupley, 1973; Chothia, 1975; and Lee and Richards, 1971) to allow

0006-3495/92/12/1483/04 $2.00

TABLE 1 **Average areas (Å²) exposed to solvent during molecular dynamics simulation. If multiple values are given, e.g., for δ, the first refers to δ1 atom, the second to δ2, etc**

X	N	Cα	C	O	β	γ	δ	ε	ρ	η	Σ
Ala	4.2	16.6	0.5	23.9	66.5						111.6
Arg	3.6	5.9	0.3	25.6	28.3	28.4	33.9	9.5	3.3	51.4 41.3	231.4
Asn	2.6	10.7	0.4	21.7	38.3	1.2	32.8 43.6				151.2
Asp	3.6	8.5	0.3	26.7	36.3	1.9	38.1 39.3				154.7
Cys	2.0	11.4	0.3	19.1	43.4	60.7					136.9
Glu	4.1	7.1	0.3	25.3	28.5	33.8	2.1	39.7 39.1			179.9
Gln	3.9	7.4	0.3	25.2	28.5	35.1	1.3	35.1 45.6			183.2
Gly	4.9	28.7	1.2	40.9							75.6
His	3.2	9.2	0.4	23.8	33.8	3.6	13.2 30.3	48.9 20.7			187.2
Ile	2.2	5.6	0.3	21.6	7.8	25.1 56.2	69.7				188.4
Leu	3.8	5.1	0.3	24.4	22.1	5.8	63.9 67.0				192.2
Lys	3.9	6.6	0.3	25.9	26.4	25.4	31.6	47.4	42.3		209.9
Met	2.4	10.1	0.4	21.2	30.1	27.3	27.3	77.7			196.6
Phe	4.3	6.8	0.7	24.0	29.2	2.8	23.7 19.8	35.0 29.8	34.4		210.6
Pro	0.2	14.0	0.7	22.2	39.4	43.2	26.5				146.2
Ser	3.7	11.2	0.4	24.8	52.4	30.8					123.2
Thr	2.3	9.7	0.3	21.6	18.8	27.6 65.4					145.8
Trp	4.1	7.8	0.7	20.7	28.1	2.9	30.2 3.4	17.3 7.8 18.6	33.9 31.7	34.9	242.1
Tyr	3.5	8.0	0.4	23.3	30.5	2.2	22.4 18.4	27.4 27.4	9.5	38.9	218.0
Val	2.4	6.6	0.3	22.9	10.5	61.7 60.5					164.8

direct comparison of these results with the previously calculated standard solvent accessibility values.

Several controls on the sensitivity of the obtained results to simulation parameters (described in the following section) were performed. The procedures used were the same as described above except simulations which explicitly included water molecules. For the solvent simulations we used a cut-off value of 10 Å for the non-bonding interactions, employing a switching function above 8.5 Å. The list of neighboring residues was regenerated every 20 iterations. The dielectric constant was set to 1. The tripeptides studied were soaked with a water box so that the distance between any of the peptide atoms and the box boundary was not less than 7 Å. A typical box had the dimensions of 25 × 25 × 25 Å and ~350 water molecules around the peptide. The calculations were performed on a Convex 220 and a local VAX cluster at the Institute of Crystallography, Freie Universitaet Berlin.

RESULTS AND DISCUSSION

The values of solvent accessibility for individual atoms (calculated with radii of Shrake and Rupley, 1973) of the central residue in Ala-X-Ala peptides are presented in Table 1. They are the averages over 1,000 conformations from the 1-ns molecular dynamics simulation at 368 K.

Two of the peptides studied: Ala-Arg-Ala and Ala-Ser-Ala were additionally subjected to a 500-ps molecular dynamics simulation in which solvent molecules were explicitly included as described above. The total residue accessibilities calculated for the central residues from the trajectories are in very good agreement with those presented in Table 1: 231.8 Å² for Arg and 123.4 Å² for Ser. For the latter residue, two additional simulations in vacuo at higher temperatures were performed giving the total solvent accessibilities of 123.5 Å² at 418 K and 123.0 Å² at 468 K. A good agreement of the total residue solvent accessibilities for the vacuum and solvent simulations and the lack of temperature dependence of the results indicates that the behavior of the peptide under the simulation conditions is similar to the one expected for the unfolded polypeptide chain. The visual inspection of the trajectories on the graphics display also confirmed no preference for single hydrogen-bonded conformations. This is, however, not true for room temperature simulations (the 268 K dynamics of Ala-Asp-Ala peptide gives a smaller accessibility of the central residue of 150.1 and this effect results from the preference of the peptide for hydrogen-bonded conformations as judged from the inspection of the trajectories). The distribution of bond

TABLE 2 **Standard state solvent accessibilities calculated in this work with group radii published by three different authors (columns 2, 4, 6). The obtained values are compared with corresponding data of the respective authors as given in columns 3, 5, 7**

X	This work	Shrake and Rupley, 1975	This work	Rose et al., 1985	This work	Miller et al., 1987
Ala	111.6	124	109.8	118.1	106.1	113
Arg	231.4	244	239.7	256.0	235.7	241
Asn	151.2	161	153.2	165.5	151.0	158
Asp	154.7	154	153.1	158.7	153.3	151
Cys	136.9	94	139.2	146.1	133.2	140
Glu	179.9	187	178.3	186.2	177.7	183
Gln	183.2	190	186.4	193.2	184.2	189
Gly	75.6	89	76.9	88.1	75.2	85
His	187.2	201	185.2	202.5	185.4	194
Ile	188.4	194	186.4	181.0	180.0	182
Leu	192.2	198	190.7	193.1	184.2	180
Lys	209.9	214	213.4	225.8	208.5	211
Met	196.6	215	196.0	203.4	188.6	204
Phe	210.6	221	200.1	222.8	204.4	218
Pro	146.2	150	143.2	146.8	137.6	143
Ser	123.2	126	123.0	129.8	119.3	122
Thr	145.8	152	145.2	152.5	140.3	146
Trp	242.1	265	234.0	266.3	237.8	259
Tyr	218.0	236	213.1	236.8	213.2	229
Val	164.8	169	162.5	164.5	156.7	160

angles is very similar for both room and high temperatures (the high temperature simulation does not produce excessive structural changes in the peptide). The application of 368 K temperature and a dielectric constant of 80 seems to be a good choice for approximating the behavior of the unfolded state. It should also be mentioned that for Ala-Ser-Ala the average total accessibility changes by only 0.1 $Å^2$ if the simulation time is doubled to 2 ns.

The average accessibilities obtained from the trajectories are not sensitive to the charge state of the central residue if calculated with the same group radii. For Ala-Lys^+-Ala the total average value differs from the one presented in Table 1 only by 0.2 $Å^2$. However, if one uses the system of group radii of Lee and Richards (1971), in which the radius for tetrahedral nitrogen atom is larger by 0.3 Å, the respective solvent accessibility value rises from 213.4 to 223.5 $Å^2$.

In Table 2 the average solvent accessibilities from the molecular dynamics simulation calculated with the three most commonly used sets of group radii are presented. The values obtained by us are usually smaller than those calculated previously by other authors. It is to be expected that the average solvent accessibility in the ensemble of allowed conformations is smaller than that obtained for selected conformations and, in principle, should be closer to reality. However, one should consider the values presented here to lie in the upper bound region of solvent accessibilities in the denatured protein as, on average, the neighbors of a certain residue along the polypeptide chain are more bulky than alanine. For the tripeptide Phe-Ser-Phe the solvent accessibility of the central residue drops to 112.0 $Å^2$ compared with 123.2 $Å^2$ for Ala-Ser-Ala. The respective value for the central serine in Gly-Ser-Gly tripeptide is 128.8 $Å^2$ (radii of Lee and Richards, 1971), and is still smaller than the stochastic standard state solvent accessibility of Rose et al. (1985). Alanine instead of glycine was taken as the neighboring residue in this series of tripeptides because 19 of 20 side chains do possess a $C\beta$ atom and the values obtained are overestimated due to the reasons given above and probable behavior of the chain as a whole.

The procedure of averaging the solvent accessibilities of the central residues for unconstrained allowed tripeptide conformations should provide values that reflect the characteristics of an extended polypeptide chain better than the models used previously. As the concept of solvent accessibility is often used to assess the hydrophobic contribution to the free energy of folding (e.g., Lesser and Rose, 1990; Eisenberg and McLachlan, 1990; Privalov and Maktahadze, 1990), much attention should be paid to have the standard state values as correct as possible. It should be borne in mind that an error in these values is usually multiplied by the number of residues in the protein under consideration. Recent successes of the thermodynamic perturbation approach in quantitative explanation of mutant protein stability (e.g., Dang, Merz, and Kollman, 1989), indirectly shows that the dynamic behavior of the central residue in the tripeptide of the native sequence correctly approximates its behavior in the denatured polypeptide chain. As it is computationally impossible to consider all the tripeptide sequences, choosing the Ala-X-Ala sequence seems to be a reasonable approximation to date.

The authors acknowledge financial support by the Deutsche Forschungsgemeinschaft (Schwerpunktprogramm "Protein Design" and Leibnitz-Programm), and KBN grant No. 663809203.

Received for publication 1 April 1992 and in final form 25 August 1992.

REFERENCES

Chothia, C. 1975. Structural invariants in protein folding. *Nature (Lond.)*. 254:304–308.

Chothia, C. 1976. The nature of accessible and buried surfaces in proteins. *J. Mol. Biol.* 105:1–14.

Chothia, C., and J. Janin. 1975. Principles of protein-protein recognition. *Nature (Lond.)*. 256:705–708.

Connolly, M. L. 1981. Quantum Chemistry Program Exchange Bulletin, Indiana University Chemistry Department, Bloomington, Indiana.

Dang, L. X., M. M. Merz, and P. A. Kollman. 1989. Free energy calculations on protein stability: Thr157-Val157 mutation of T4 lysozyme. *J. Am. Chem. Soc.* 111:8505–8508.

Eisenberg, D., and A. D. McLachlan. 1986. Solvation energy in protein folding and binding. *Nature* 319:199–203.

Hagler, A. T. 1979. Consistent force field studies of intermolecular forces in hydrogen bonded crystals. *J. Am. Chem. Soc.* 101:5122–5130.

Janin, J. 1979. Surface and inside volumes in globular proteins. *Nature (Lond.)*. 277:491–492.

Kauzmann, W. 1959. Some factors in the interpretation of protein denaturation. *Adv. Prot. Chem.* 14:1–64.

Lavery, R., and B. Pullman. 1981. Molecular electrostatic potentials on the surface envelopes of macromolecules. *Int. J. Quant. Chem.* 20:251–272.

Lesser, G. J., and G. D. Rose. 1990. Hydrophobicity of amino acid subgroups in proteins. *Proteins.* 8:6–13.

Maple, J. R., U. Dinur, and A. T. Hagler. 1988. Derivation of force fields for molecular mechanics and dynamics from ab initio energy surfaces. *Proc. Natl. Acad. Sci. USA.* 85:5350–5354.

Miller, S., J. Janin, A. M. Lesk, and C. Chothia. 1987. Interior and surface of monomeric proteins. *J. Mol. Biol.* 196:641–656.

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

Privalov, P. L., and G. I. Makhatadze. 1990. Heat capacity of proteins. Partial molar heat capacity of the unfolded polypeptide chain of proteins: protein unfolding effect. *J. Mol. Biol.* 213:385–391.

Richards, F. M. 1977. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* 6:151–176.

Richmond, T. J. 1984. Solvent accessible surface area and excluded volume in proteins. Analytical equations for overlapping spheres and implications for the hydrophobic effect. *J. Mol. Biol.* 178:63–72.

Rose, G. D., A. R. Geselowitz, G. J. Lesser, R. M. Lee, and M. M. Zehfus. 1985. Hydrophobicity of amino acid residues in globular proteins. *Science (Wash. DC).* 229:834–838.

Shrake, A., and J. A. Rupley. 1973. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* 79:351–372.

Wood, R. H., and P. T. Thompson. 1990. Differences between pair and bulk hydrophobic interactions. *Proc. Natl. Acad. Sci. USA.* 87:946–949.