Solvent Accessibilities in Glycyl, Alanyl and Seryl Dipeptides

P. MANAVALAN and P. K. PONNUSWAMY*

Department of Physics, Postgraduate Centre, Tiruchirappalli-620 020, Tamilnadu, India
and A. R. SRINIVASAN

*Department of Crystallography and Biophysics, University of Madras,
Guindy Campus, Madras-600 025, India*

(Received 9 March 1977)

Theoretical studies on glycyl, alanyl and seryl dipeptides were performed to determine the probable backbone and side-group conformations that are preferred for solvent interaction. By following the method of Lee & Richards [(1971) *J. Mol. Biol.* 55, 379–400], a solute molecule is represented by a set of interlocking spheres of appropriate van der Waals radii assigned to each atom, and a solvent (water) molecule is rolled along the envelope of the van der Waals surface, and the surface accessible to the solvent molecule, and hence the solvent accessibility for a particular conformation of the solute molecule, is computed. From the calculated solvent accessibilities for various conformations, solvation maps for dipeptides were constructed. These solvation maps suggest that the backbone polar atoms could interact with solvent molecules selectively, depending on the backbone conformation. A conformation in the right-handed bridge (ζ_R) region is favoured for both solvent interaction and intrachain hydrogen-bonding. Also the backbone side-chain hydrogen-bonding within the same dipeptide fragment in proteins is less favoured than hydrogen-bonding between side chain and water and between side chain and atoms of other residues. Solvent accessibilities suggest that very short distorted α_R -helical and extended-structural parts may be stabilized via solvent interaction, and this could easily be possible at the surface of the protein molecules, in agreement with protein-crystal data.

The spatial structure of an amino acid residue in a polypeptide or protein chain is governed by a variety of factors, such as steric and polar interactions (including hydrogen-bonding) and solvent–solute interactions (including hydrophobic bonding). No unified theoretical model incorporating all these factors to predict the possible stable conformational states of amino acid residues has been hitherto proposed. Most of the theoretical studies (Brant & Flory, 1965; Ramachandran & Sasisekharan, 1968; Scheraga, 1968; Ponnuswamy & Sasisekharan, 1971; Popov & Lipkind, 1971; Lewis *et al.*, 1973; Chandrasekaran *et al.*, 1973; Pullman & Pullman, 1974) were based on semi-empirical or quantum-mechanical-energy calculations on dipeptide models of amino acid residues. Only a very few workers have attempted to include solvent contribution to the stability of conformational states (Gibson & Scheraga, 1967; Venkatachalam & Krimm, 1973; Forsythe & Hopfinger, 1973; Perricaudet & Pullman, 1973). These attempts were also very preliminary and approximate and could not be used directly to explain the observed features in the protein molecules. The polymer–solvent interaction is as important as the other types of interactions. Solvent interactions through hydrogen-bonding with the amino acid

residues in protein molecules could give added stability to the native conformation. Studies by Kuntz (1972) have revealed that most of the surface residues are polar and they strongly interact with solvent (water) molecules. A possible source of information on solvent involvement in the stability comes from a comparison of low-energy conformations of amino acid residues computed by excluding solvent interactions with those actually observed for these residues in proteins.

Although the observed conformations of most of the residues in a protein lie well within the low-energy regions in the (ϕ – ψ) potential-energy maps (the definitions of the conformational parameters ϕ and ψ are given in the Method of Calculation section), it is noticed that a significant number of residues adopt conformations disallowed on the map (Burgess *et al.*, 1974). The disagreement may be due to a number of reasons: (i) the dipeptide approximation ignores stability contributions from residue interactions originating beyond the dipeptide range; (ii) the potential functions used may themselves be inaccurate (predicting wrong minima and omitting right minima); (iii) the omitted solvent contribution does have appreciable influence on the stability of some conformational states. Attempts to incorporate longer-range interactions (Ponnuswamy *et al.*, 1973) and to modify potential-energy expressions (Kolaskar

* To whom reprint requests should be addressed.

et al., 1975) have been reported, and these methods are yet to be refined for interpreting data from protein molecules in general. We have done extensive studies on some dipeptides and oligopeptides to determine the probable backbone and side-chain conformations preferred for solvent interactions. We followed the method proposed by Lee & Richards (1971) to calculate the accessibility of a solvent molecule to an atom or group of atoms in the model molecular structures and compiled solvation maps which were then used to interpret the observed conformations of amino acid residues in proteins, taking into consideration the (ϕ - ψ) potential-energy maps already available. We have already reported some preliminary results of that study (Ponnuswamy & Manavalan, 1976). In the present paper we describe the results in detail for the glycyl, alanyl and seryl dipeptides.

Method of Calculation

An atom (group) in the solute molecule is said to be accessible to a solvent molecule if the solvent molecule can be brought into van der Waals contact with it, without penetrating any other atom of the molecule (Lee & Richards, 1971). The molecule is represented by a set of interlocking spheres of appropriate van der Waals radii assigned to each atom and is sectioned into a set of parallel planes with a known spacing. The solvent molecule is rolled along the envelope of the van der Waals surface for the molecule in each plane (one such plane for the glycyl dipeptide is shown in Fig. 1) and the maximum permitted contact arc is calculated. The total accessible surface area and the accessibility for an atom are then computed by the use of formulae:

$$a = \sum (R/\sqrt{R^2 - Z_i^2}) DL_i \quad (1)$$

and

$$S = 100a/4\pi R^2 \quad (2)$$

Fig. 1. *van der Waals and accessibility contours of a section of glycyl dipeptide and a water molecule*
The details are given in the text.

where L_i is the length of the arc drawn on a given section i , Z_i is the perpendicular distance from the centre of the sphere to the section i , D is the spacing between sections [the value of D has been taken to be 0.25 nm for our computations; for reasons see Lee & Richards (1971)] and R is the sum of the van der Waals radius of the atom and the radius of the solvent molecule. $\sum$ indicates that the summation is over all of the arcs drawn for the given atom. The number S is the static accessibility or static accessible area for an atom in the solute molecule, which in turn reflects the probable extent of surface available for free interaction with a solvent molecule.

We consider here the dipeptides of the form

where R is the side chain of glycine, alanine or serine. The water molecule is considered as solvent. The bond lengths and bond angles were taken from Momany *et al.* (1975) and the van der Waals radii were the same as those used by Lee & Richards (1971), except for O'H of the serine side chain, for which the value was taken as 0.152 nm. For glycyl (*N*-acetylglycine-*N*-methylamide) and alanyl (*N*-acetylalanine-*N*-methylamide) dipeptides, accessibility values were computed by varying the conformational angles ϕ (N-C α) and ψ (C α -C') at 20° intervals and represented in the ϕ - ψ plane. For the seryl dipeptide a number of ϕ - ψ grid points were considered by varying χ_1 (C α -C β) from -180° to 160° at 20° intervals. The conformational angles are positive for a right-handed rotation; when looking along the relevant bond, the far end rotates clockwise, relative to the near end. The zero values are ascribed as: $\phi = 0^\circ$ when the C α -C' bond is *cis* to N-C' bond, $\psi = 0^\circ$ when the N-C α bond is *cis* to the C'-N bond and $\chi_1 = 0^\circ$ when the N-C α bond is *cis* to the C β -O' bond (see IUPAC-IUB Commission on Biochemical Nomenclature, 1970).

Results

The computed accessibility maps for atoms O $_{(1)}$, N $_{(1)}$, O $_{(2)}$ and N $_{(2)}$ and a superposition map obtained by superimposing these four individual maps for glycyl and alanyl dipeptides respectively are presented in Figs. 2 and 3. The accessibilities of these four atoms and their combinations have been considered in order to discuss the potentialities for solvation of these atoms and groups individually and together in different situations, and also because of their polar and directional properties promoting hydrogen-bonding with solvent molecules. The accessibilities of the central α -carbon atom and carbonyl carbon atoms were found to be small and invariant with conformation, and those of the terminal α -carbon atoms were found to decrease when the dipeptide was elongated on either side to become an oligopeptide; hence

these carbon atoms, as well as the β -carbon, are not considered here.

The superposition maps (Figs. 2e and 3e) were obtained as follows. First, the boundary regions for the commonly observed conformational states of amino acid residues in protein crystals, i.e. right- and left-handed α -helical (α_R and α_L), extended structure (β) and right- and left-handed bridge conformations (ζ_R and ζ_L) on the (ϕ - ψ) plane, were marked. [The boundary regions for α_R , α_L and β are those defined by Burgess *et al.* (1974), and those for ζ_R and ζ_L have been chosen so as to include the maximum number of observed conformations with ϕ around $\pm 90^\circ$ and ψ around 0° in ten proteins: ζ_R : $\phi = -120^\circ$ to -50° , $\psi = -40^\circ$ to 5° and $\phi = -145^\circ$ to -85° ; $\psi = -5^\circ$ to 40° ; ζ_L : $\phi = 50^\circ$ to 120° , $\psi = -5^\circ$ to $+40^\circ$ and $\phi = +85^\circ$ to 145° ; $\psi = -40^\circ$ to $+5^\circ$.] Each of the four atomic accessibility maps (Figs. 2a-2d and 3a-3d) was then individually superimposed on the boundary map and the areas of α_R -, α_L -, β -, ζ_R - and ζ_L -regions which fall within the iso-accessibility contour of X-5, where X is the global maximum, were shaded so as to identify the accessible areas of the four atoms individually. This superposition map is then used to express the total accessibility potential of the respective dipeptide with reference to one or more polar atoms involved in solvent association.

Also marked on these superposition maps is the energy contour of 12.552 kJ/mol above the global minimum as computed by Ponnuswamy (1970), to permit discussion of the accessibility maps more realistically.

Glycyl dipeptide

For this dipeptide, results are presented only for the left half of the (ϕ - ψ) map, as the right half is related by the symmetry (ϕ, ψ) = ($-\phi, -\psi$).

$O_{(1)}$ map (Fig. 2a). The entire region of $\phi \simeq -100^\circ$ to $+100^\circ$ for the whole of the ψ range is included in the area enclosed by the X-2 contour. The maximum accessibility occurs around $\psi \simeq 0^\circ$ and around $\psi \simeq 180^\circ$, and in the vicinity of these conformations the accessibility is almost invariant.

$N_{(1)}$ map (Fig. 2b). This atom has maximum accessibility at $\phi = \pm 20^\circ$ and $\psi = \pm 60^\circ$ or $\pm 120^\circ$, and the X-2 contour encompasses the region $\phi \simeq -60^\circ$ to $+60^\circ$ through 0° , and the entire ψ range which is practically disallowed in the ϕ - ψ potential-energy map for glycyl dipeptide (see Fig. 2e). It is noteworthy that the region around the polyglycine-II structure ($\phi \simeq -90^\circ$, $\psi \simeq 140^\circ$) is more accessible to solvent than is the extended-structure region ($\phi \simeq -140^\circ$, $\psi \simeq 140^\circ$), in contrast with the $O_{(1)}$ map.

$O_{(2)}$ map (Fig. 2c). In contrast with the $N_{(1)}$ map, the maxima occur at $\phi = \pm 80^\circ$ and $\psi = \pm 40^\circ$, and the entire ϕ range with $\psi \simeq 0$ to $\pm 60^\circ$ becomes preferred for solvent interaction with $O_{(2)}$. The interesting part of this map is that it predicts that conformations with

$\psi \simeq 0^\circ$ will permit this carbonyl O atom to have the maximum accessibility to solvent molecules.

$N_{(2)}$ map (Fig. 2d). Maxima occur at different locations on the ϕ range, and the higher accessibility contour encompasses the region $\phi \simeq -180^\circ$ to $+180^\circ$ and $\psi \simeq -100^\circ$ to $+100^\circ$.

The superposition map of accessibility for glycyl dipeptide (Fig. 2e) indicates that a triangular strip in the upper left quarter and another such triangular strip in the lower left quarter of the ϕ - ψ plane permit the atoms $O_{(1)}$, $O_{(2)}$ and $N_{(2)}$ simultaneously to have higher solvent accessibilities. The upper triangular strip includes the part (ϕ, ψ) $\simeq (-140^\circ \pm 40^\circ, 120^\circ \pm 30^\circ)$ of the β -region and the lower triangular strip includes a small part of the α_R -region with $\phi < -80^\circ$. Considering the other sections of the β -region limit, the conformational ranges (ϕ, ψ) $\simeq (-140^\circ \pm 40^\circ, 180^\circ \pm 40^\circ)$, $(-50^\circ \pm 20^\circ, 180^\circ \pm 40^\circ)$ and $(-60^\circ \pm 20^\circ, 110^\circ \pm 10^\circ)$ permit respectively the atom pairs ($O_{(1)}, N_{(2)}$), ($N_{(1)}, N_{(2)}$) and ($N_{(1)}, O_{(2)}$) to have larger solvent accessibilities. Within the ζ_R -region, conformations with $\phi < -80^\circ$ allow the atoms $O_{(1)}$ and $O_{(2)}$ to have larger accessibility values, as do those with $\phi > -80^\circ$ for the atoms $N_{(1)}$ and $O_{(2)}$. Also, a small strip of ϕ - ψ range in the middle of the ζ_R -region enjoys the higher accessibilities of three atoms $O_{(1)}$, $N_{(1)}$ and $O_{(2)}$. This three-atom strip extends through almost the whole of the ψ range in a zig-zag way. Considering the α_R -region, conformations in the far left of this region, i.e. around (ϕ, ψ) $\simeq (-100^\circ, -60^\circ)$, permit the three atoms $O_{(1)}$, $O_{(2)}$ and $N_{(2)}$ to possess higher accessibilities, whereas a conformation near (ϕ, ψ) $\simeq (-57^\circ, -48^\circ)$ permits only two atoms, namely $N_{(1)}$ and $O_{(2)}$, to have higher accessibilities. The region (ϕ, ψ) $\simeq (-50^\circ \pm 30^\circ, -80^\circ \pm 20^\circ)$ allows the atoms $N_{(1)}$ and $N_{(2)}$ to have the best accessibilities.

Alanine dipeptide

The β -carbon atom of alanine affects the accessibilities of the N and O atoms significantly at selective ϕ - ψ regions.

$O_{(1)}$ map (Fig. 3a). No significant change in accessibility is noticed in the left half of the map compared with the corresponding glycyl dipeptide $O_{(1)}$ map. The β -carbon affects the solvent accessibility of $O_{(1)}$ in the right-hand side, making it completely inaccessible. The maxima occur at (ϕ, ψ) = $(-120^\circ, 0^\circ)$ and at $(-120^\circ, 160^\circ)$.

$N_{(1)}$ map (Fig. 3b). The presence of C^β interferes with the solvent accessibility of $N_{(1)}$ in such a way that region $\phi \simeq -60^\circ$ to 120° over the entire ψ range becomes less accessible to solvent. This map predicts that conformations observed in the right-hand side of the map, including the α_L -region, could be associated with solvent molecules via this N atom.

$O_{(2)}$ map (Fig. 3c). The β -carbon affects the accessibility of $O_{(2)}$ in the upper half of the map, making it less accessible to solvent. The main feature of this

map is that the maximum accessibility occurs near the axial C_7 -conformation ($\phi \simeq 80^\circ$, $\psi \simeq -60^\circ$) and conformations with $\psi \simeq -30^\circ \pm 30^\circ$ allow $O_{(2)}$ to have better solvent accessibilities.

$N_{(2)}$ map (Fig. 3d). This atom has maximum accessibility at $(\phi, \psi) \simeq (20^\circ, 120^\circ)$, and the higher accessibility areas (covered by contours X-2 and X-5) occur in the upper quarter of the map, whose overall shape is very similar to that of the corresponding glycyl dipeptide $N_{(2)}$ map.

The superposition accessibility map (Fig. 3e) indicates that a triangular strip in the upper left quarter permits the atoms $O_{(1)}$ and $O_{(2)}$ and another such triangular strip in the lower left quarter of the map permits the atoms $O_{(1)}$ and $O_{(2)}$ to enjoy better solvent accessibility. The upper-left-quarter triangle occupies approximately more than half of the low-energy region, whereas the lower-left-quarter triangle occupies a considerable energetically forbidden area. The upper right quarter and the lower right quarter permit respectively the atom pairs $(N_{(1)}, N_{(2)})$ and $(N_{(1)}, O_{(2)})$ to have higher accessibilities, but these regions, except the α_L -region, are energetically completely forbidden. Considering the β -region, the β -carbon has decreased the accessibility potential of $O_{(2)}$ in the region $(\phi, \psi) \simeq (-120^\circ \pm 50^\circ, 120^\circ \pm 40^\circ)$, compared with the corresponding glycyl dipeptide map. The accessibilities of the atoms $O_{(1)}$, $N_{(1)}$ and $O_{(2)}$ were also decreased in the range $(\phi, \psi) \simeq (-60^\circ \pm 10^\circ, 160^\circ \pm 40^\circ)$ and thus the only atom which has higher accessibility in this region is $N_{(2)}$. Considering the ζ_R -region, the range $(\phi, \psi) \simeq (-90^\circ \pm 20^\circ, -20^\circ \pm 20^\circ)$ orients the atoms $O_{(1)}$ and $O_{(2)}$ so as to give them the best solvent accessibilities. Conformations with $\psi > 10^\circ$ permit only $O_{(1)}$, and those with $\phi > -80^\circ$ permit only $O_{(2)}$, to have higher accessibilities. As far as the α_R -region is concerned, conformations with $\phi < -80^\circ$ (distorted α_R -conformation) orient the atoms $O_{(1)}$ and $O_{(2)}$ and those with $\phi > -40^\circ$ (again distorted α_R -conformation) orient the atoms $N_{(1)}$ and $O_{(2)}$ to have higher accessibilities. Only the atom $O_{(2)}$ has higher accessibility near the classical α_R -helical conformation with $(\phi, \psi) = (-57^\circ, -48^\circ)$. Approximately half of the α_L -region allows the atoms $N_{(1)}$

and $N_{(2)}$ to enjoy higher accessibilities, the other half being accessible only to atom $N_{(1)}$. The lower half of the ζ_L -region permits the atoms $N_{(1)}$ and $O_{(2)}$ and the upper half permits the atom $N_{(1)}$ to have higher accessibilities.

Seryl dipeptide

The O'H group of the seryl dipeptide affects the solvent accessibilities of the backbone atoms characteristically. The O' atom approaches different backbone atoms selectively to overlapping distances for specific χ_1 values, and hence the accessibilities of the corresponding atoms are drastically decreased. The variation of accessibility with the conformational angle $\chi_1(C^\alpha-C^\beta)$ for the atoms $O_{(1)}$, $N_{(1)}$, $O_{(2)}$, $N_{(2)}$ and O' is given in Fig. 4 for four selected backbone conformations. When the backbone is in the β -conformation and χ_1 is around 60° , the γ -atom approaches the $N_{(1)}$ and $O_{(2)}$ atoms to closer distances, and hence their accessibilities are decreased for this conformational situation. Similarly the hindrance caused by the O' atom largely decreases the accessibility values of $N_{(2)}$ and $O_{(1)}$ near $\chi_1 \simeq 180^\circ$ and $\simeq -60^\circ$ respectively (Fig. 4a). When the backbone is in the α_R -conformation (Fig. 4c), O' affects the accessibility values of $O_{(2)}$ near $\chi_1 \simeq 180^\circ$ and of $N_{(2)}$ near $\chi_1 \simeq 60^\circ$. For the α_L -conformation (Fig. 4d), O' approaches $O_{(1)}$ when $\chi_1 \simeq 60^\circ$, $O_{(2)}$ when $\chi_1 \simeq 180^\circ$ and $N_{(1)}$ when $\chi_1 \simeq -60^\circ$ respectively to closer distances. For a conformation in the ζ_R -region (Fig. 4b), the accessibility of $N_{(2)}$ is decreased near $\chi_1 \simeq 60^\circ$, that of $O_{(2)}$ around $\chi_1 \simeq 180^\circ$ and that of $N_{(1)}$ near $\chi_1 \simeq -20^\circ$. The accessibilities of O' consequently vary along with the backbone atoms. These χ_1 -ranges, where the accessibilities of one of the backbone atoms and the O' atom become drastically decreased, suggest the possibility of a backbone side-chain hydrogen-bonding, and those ranges where the accessibility is maximum and unaffected suggest the possibilities of interaction between (i) backbone and water (or -X), (ii) backbone, water and backbone, (iii) side chain and water and (iv) side chain, water and backbone. Here X represents any potential polar atom (or group) of the main chain (for a protein) which is in the vicinity of the dipeptide in consideration. From

Fig. 2. Solvent-accessibility contour maps for the atoms of a glycyl dipeptide

The location of maximum accessibility is marked by X. The area enclosed by the values X and X-2 is denoted by number 2 and numbers 5, 10 and 20 have similar meanings. (a) Map for the atom $O_{(1)}$ (X = 39.38); (b) map for the atom $N_{(1)}$ (X = 18.07); (c) map for the atom $O_{(2)}$ (X = 40.36); (d) map for the atom $N_{(2)}$ (X = 18.23); (e) superposition map for glycyl dipeptide. $\square$, $\blacksquare$, $\blacksquare$, $\blacksquare$ represent respectively the areas of high accessibilities for the atoms $O_{(1)}$, $N_{(1)}$, $O_{(2)}$ and $N_{(2)}$. Boundaries for β -, α_R -, α_L -, ζ_R - and ζ_L -regions are marked. The total potential-energy contour of 12.552 kJ/mol above the global minimum is also marked (see the text). The observed (ϕ, ψ) values of glycyl residues at the second and third positions of bends ($\blacktriangle$) and at irregular regions ($\bullet$) are plotted for ten proteins, namely α -chymotrypsin (Birktoft & Blow, 1972), lysozyme (Browne *et al.*, 1969), subtilisin BPN' (Alden *et al.*, 1971), staphylococcal nuclease (Arnone *et al.*, 1971), bovine ribonuclease S (Wyckoff *et al.*, 1970), ferricytochrome C (Dickerson *et al.*, 1971), myoglobin (Watson, 1969), carboxypeptidase A (Quiocho & Lipscomb, 1971), carp myogen (Nockolds *et al.*, 1972) and cytochrome b_5 (Mathews *et al.*, 1972).

Fig. 4. Variation of accessibility of $O_{(1)}$, $N_{(1)}$, $O_{(2)}$, $N_{(2)}$ and O' with the angle $\chi_1(C^\alpha-C^\beta)$ in seryl dipeptide for various backbone conformations

(a) β : $(\phi, \psi) = (-140^\circ, 140^\circ)$; (b) ζ_R : $(\phi, \psi) = (-80^\circ, -20^\circ)$; (c) α_R : $(\phi, \psi) = (-57^\circ, -48^\circ)$; (d) α_L : $(\phi, \psi) = (57^\circ, 48^\circ)$.

a model and the accessibility graphs shown in Figs. 4(a)–4(d), the probable hydrogen-bonding situations in the seryl dipeptide have been scrutinized, and the most likely ones are given in Table 1. Table 1 suggests that, irrespective of the backbone conformation, all of the three low-energy side-chain conformations orient the O' atom so as to interact with a solvent or external solute polar atom, and the conformation with $\chi_1 \approx 60^\circ$ is favourable for internal backbone–side-chain hydrogen-bonding interaction (within the dipeptide), and that with $\chi_1 \approx -60^\circ$ is the preferred conformation for hydrogen-bonding interaction via a solvent molecule or an external solute group. An analysis of the involvement of the $O'H$ group of the serine residue in two proteins, α -chymotrypsin and lysozyme, was made, and the observed pattern is given in Table 2. Comparison of Tables 1 and 2 indicates that the backbone–side-chain hydrogen-bonding within the same dipeptide fragment is not favoured in proteins as expected, compared with hydrogen-bonding between side-chain and water and between side chain and atoms of other residues.

A comparison of the accessibility curves in Figs. 4(a)–4(d) brings out an interesting fact that the back-

bone conformations ζ_R and α_L are the more favoured than β - and α_R -conformations as far as solvent accessibility is concerned. This selectivity is actually reflected in the protein crystals, as we notice that more serine residues occupy the ζ_R -region (Burgess *et al.*, 1974).

Discussion

The accessibility maps for glycyl and alanyl dipeptides suggest that the N and O atoms of the backbone could interact with solvent molecules selectively in the most commonly observed structural regions α , β and ζ . Considering the maps for individual atoms, $O_{(1)}$ is more accessible to solvent in the ζ_R - and β -regions in both glycyl and alanyl dipeptides. N_1 in glycyl dipeptides is more accessible in the β - and ζ -regions, whereas the α_L -region is preferred in the alanyl dipeptide by this atom for solvent association. $O_{(2)}$ has higher accessibility in the ζ_R - and ζ_L -regions and N_2 only in the β -region in both glycyl and alanyl dipeptides. The superposition maps reflect the potentiality of the different conformational regions for solvent association via the four polar backbone atoms. Depending on the actual ϕ – ψ location, differ-

Fig. 3. Solvent-accessibility contour maps for the atoms of an alanyl dipeptide

(a) Map for the atom $O_{(1)}$ ($X = 39.32$); (b) map for the atom $N_{(1)}$ ($X = 15.58$); (c) map for the atom $O_{(2)}$ ($X = 40.06$); (d) map for the atom $N_{(2)}$ ($X = 17.97$); (e) superposition map for alanyl dipeptide. The observed (ϕ, ψ) values of non-glycyl residues at the second and third positions of bends and at irregular regions in the same ten proteins as in Fig. 2 are plotted. Definitions of the symbols are the same as in Fig. 2.

Table 1. *Probable hydrogen-bonding schemes in seryl dipeptides including water molecules*

The possible hydrogen-bonding of the hydroxyl group of a seryl dipeptide with the backbone polar atoms of the same dipeptide and with a water molecule was scrutinized, by using a visual model of seryl dipeptide, for the preferred backbone and side-chain conformations. W represents the water molecules and X is any polar atom of the neighbouring residues.

Backbone conformation	Side-chain conformation ($\sim\chi_1$)	Probable hydrogen-bonding interaction
β	60°	O ^y ...N ₍₁₎
		O ^y ...O ₍₂₎
		O ^y ...W...N ₍₁₎
		O ^y ...W(X)
	180°	O ^y ...W...N ₍₂₎
		O ^y ...W(X)
α_R	-60°	O ^y ...O ₍₁₎
		O ^y ...W...O ₍₁₎
		O ^y ...W(X)
	60°	O ^y ...N ₍₂₎
		O ^y ...W(X)
		O ^y ...W(X)
α_L	-60°	O ^y ...W...N ₍₁₎
		O ^y ...W(X)
		O ^y ...O ₍₁₎
	60°	O ^y ...O ₍₂₎
		O ^y ...W(X)
		O ^y ...O ₍₂₎
ζ_R	-60°	O ^y ...W(X)
		O ^y ...W...HN ₍₁₎
		O ^y ...W(X)
	60°	O ^y ...W...N ₍₁₎
		O ^y ...W...N ₍₂₎
		O ^y ...W(X)
	180°	O ^y ...O ₍₂₎
		O ^y ...W...O ₍₂₎
	-60°	O ^y ...W(X)

ent atoms contribute individually and collectively to the total solvent accessibility of the dipeptides. Pysh (1974) has constructed, by visual inspection of space-filling models, a qualitative dipeptide solvation map by considering the ϕ - ψ regions, where one amide group overlaps or blocks a solvation region of the neighbouring amide group, and as such the map is not able to distinguish specific contributions from individual atoms. Our superposition maps, apart from distinguishing the atomic contributions, differentiate the solvatable regions more quantitatively.

In order to compare the solvent-accessibility results with the experimental observations in protein crystals, it is essential to consider additional effects introduced by solute atoms that may approach closer to the polar atoms of the dipeptide in question. The accessibilities of atoms N and O given in the superposition maps will be drastically decreased in the regular α -helical

and parallel/anti-parallel pleated structures as these atoms take part in intra- and inter-chain hydrogen-bonding when the dipeptide unit forms part of such chain segments. For example, the accessibility of atom O₍₂₎ of the central fifth residue in a nine-residue α_R -helical alanyl oligomer (Lee & Richards, 1971) is only 5.5, whereas in the dipeptide it is approx. 35. On the other hand, if one or two residues alone adopt α -helical or extended conformation at some location on the chain, the likelihood of backbone hydrogen-bonding may be remote, thus leaving the N and O atoms to interact with either the solvent molecules or side-chain polar atoms of the same or other residues. In the 10 proteins considered in this study, 45 out of 69 distorted α_R -helical conformational residues and 83 out of 129 distorted extended conformational residues occur as surface residues. [The residues whose side chains are completely or partially accessible to the surrounding solvent medium of the protein molecule have been considered as surface residues; see Browne *et al.* (1969) and Birktoft & Blow (1972).] This fact suggests that very short distorted α_R -helical or extended-structural parts may be stabilized via solvent interaction, and this could be easily possible at the surface of the protein molecule. With regard to ζ_R -conformation, i.e. around $\phi \simeq -90^\circ$ and $\psi \simeq 0^\circ$, the two polar atoms O₍₁₎ and O₍₂₎ are completely, and the atom N₍₁₎ is partially, exposed outwards without any interference from neighbouring atoms, thus having maximum potential for solvent interaction in both glycyl and alanyl dipeptides. As this conformation results in a hinge, the likelihood of hydrogen bonds of the type observed in regular α_R -helical or parallel/anti-parallel structure is again very remote. Moreover, as the chain extends on both sides, the N₍₂₎ atom, which is completely inaccessible to solvent in this region, could take part in hydrogen-bonding with the carbonyl O atom of the third preceding residue, resulting in the type I β -bend (Venkatachalam, 1968); this energetically stabilizes this conformation further. This situation, however, does not affect the solvent-accessibility potentialities of O₍₁₎, N₍₁₎ and O₍₂₎ atoms. Therefore a conformation in the ζ_R -region is favoured both by the external solvent and by internal hydrogen-bonding. Refined potential-energy calculations (Ponnuswamy, 1970) indicate that this conformation ($\phi \simeq -90^\circ$, $\psi \simeq 0^\circ$) is less stable by about 12.552 kJ/mol with respect to the global minimum in the β -region. A slight departure from $\psi \simeq 0^\circ$ to $\psi \simeq -20^\circ$ appreciably decreases this energy barrier, and the acquired additional interactions via solvent association and a relatively weak intrachain hydrogen bond stabilize this conformation, making it a very probable conformation for an amino acid residue. In addition, if the residue contains a potential polar group in the side chain, the probability of the side group pointing away from the backbone atoms of the residue, so as to have further

Table 2. *Observed hydrogen-bonding schemes in serine residues from α -chymotrypsin and lysozyme molecules*

The types of interactions observed for the hydroxyl group of the serine side chain in the two proteins α -chymotrypsin and lysozyme are given. The sequence number denotes the residue number along the sequence of the protein chain. The situation of the side chain is as follows: I, internal, inaccessible to external water; E, external, accessible to external water; S, surface, less than half of the side-chain surface is accessible to water. In the last column the number in parenthesis denotes the sequence number of the residue whose atom is involved in the hydrogen bond.

Protein	Sequence number	Backbone conformation $\psi(\phi^\circ, \psi^\circ)$ β	Side-chain conformation (χ_1°)	Situation of side chain	Observed hydrogen-bonding interactions
α -Chymotrypsin	32	-129, 134	-178	S	O γ ...N ⁶¹ (40)
	45	-153, 139	-55	I	O γ ...W
	75	-63, 159	-97	E	O γ ...CO (72)
	77	-179, 132	-161	E	O γ ...W
	113	-94, 136	53	E	O γ ...O ⁶¹ (81)
	115	-160, 176	53	E	O γ ...NH (117) O γ ...O ⁷¹ (117) O γ ...NH (118) O γ ...CO (118) O γ ...W
	119	177, 168	-42	E	O γ ...W
	125	-136, 138	-52	E	O γ ...W
	159	-117, 150	-62	E	O γ ...CO (136)
	164	-84, 136	-66	E	
	189	-159, 153	49	E	O γ ...CO (220) O γ ...CO (190)
	190	-74, 145	-61	E	O γ ...O ⁴ H (228) O γ ...W
	217	-59, 140	180	E	O γ ...NH (219) O γ ...NH (221)
	221	-73, 154	172	E	O γ ...O γ (223) O γ ...NH (223)
		α_R			
	63	-71, -32	-169	E	O γ ...W
	76	-68, -33	-46	E	O γ ...W
	92	-33, -34	-56	E	O γ ...W
	96	-10, -22	-70	E	O γ ...W
	109	-54, -70	173	E	O γ ...O ⁷¹ (110)
	214	-115, -74	-66	I	O γ ...O ⁶¹ (102) O γ ...W
	218	-67, -51	-45	E	O γ ...W
		ζ_R			
	26	-86, 6	80	I	O γ ...CO (23) O γ ...W
	127	-78, 7	56	E	
	223	-107, -2	58	E	O γ ...O γ (221)
	186*	-143, 26	68	E	
	195*	-37, 135	-74	E	O γ ...W O γ ...NH (195) O γ ...NH (193) O γ ...W
Lysozyme		β			
	24	-71, 135	171	S	O γ ...NH (26)
	72	-57, 130	-77	S	O γ ...CO (69)
		α_R			
	36	-119, -22	92	I	O γ ...CO (55)
	81	-39, -48	111	E	O γ ...W
	91	-60, -60	-58	I	O γ ...W
		ζ_R			
	60	-110, 14	51	I	O γ ...NH (51)
	86	-90, 11	140	E	
	100	-125, 28	-62	S	O γ ...CO (96)
	50*	-0.4, 114	140	S	O γ ...O ⁶ (46) O γ ...N ⁶ (59)
	85*	-27, 153	103	E	O γ ...CO (87)

* The conformations of these residues do not come under α , β and ζ designations.

Fig. 5. Histograms showing the distributions of ϕ and ψ values for the *N*- and *C*-termini of helices in ten proteins. See the caption to Fig. 2(e) for the names of the proteins. The ordinate N denotes the number of occurrences in each range. (a) Distribution of ϕ values for *N*-terminal residues; (b) distribution of ϕ values for *C*-terminal residues; (c) distribution of ψ values for *N*-terminal residues; (d) distribution of ψ values for *C*-terminal residues.

solvent interaction, is greater, as for the seryl dipeptide. These facts indicate that the conformations in the ζ_R -region are those expected on the surface of a protein molecule. This is supported by the fact that 32 out of 42 ζ_R -residues in the ten protein crystals occur on the surface.

The strong support for the above selectivity of conformational regions for solvent interaction comes from the actual observation of conformational states of amino acid residues in the irregular or coil regions and in the bend regions of protein crystals. The observed (ϕ , ψ) values of glycyl and non-glycyl surface residues (one to three in the α_R - and β -regions without the α -helix/ β -structure hydrogen bonds and residues in β -bends) in ten proteins are plotted in the glycyl and alanyl dipeptide superposition maps respectively (Figs. 2e and 3e). In both these maps, residues cluster at regions where solvent accessibilities of two or three polar atoms of the backbone are simultaneously higher. This demonstrates that the N and O atoms of the backbone themselves prefer certain conformational states in order to have solvent association.

Another interesting feature emerges from the analysis of the conformational states of the end residues of the helical and extended-structure segments in proteins. The distribution of the ϕ and ψ values of the *N*- and *C*-terminal residues of helices with six or more residues, which occur in the ten proteins, are shown in Fig. 5 in the form of histo-

grams. The ϕ and ψ ranges have been divided into 10° intervals and the number of examples (N) in each interval has been counted and plotted. The histograms (Figs. 5a and 5b) that correspond to the ϕ values of the *N*- and *C*-terminal residues respectively show that the maximum frequency is in the range -60° to -50° for the *N*-terminal residues and -70° to -50° for the *C*-terminal residues. For the *N*-terminal residues the frequency of occurrence of the ψ value is high between -60° and -50° (Fig. 5c) and for the *C*-terminal residues it is high between -60° and -40° (Fig. 5d). A comparison of Figs. 5(c) and 5(d) reveals that the ψ values of the *C*-terminal residues are more frequently found in the range -50° to -10° than those of the *N*-terminal residues.

Robson & Pain (1971), who have analysed the dihedral angles of the ends of the helices for two proteins, showed that the conformational state of the *C*-terminal residue differs significantly from the rest of the residues of a helical segment. (For the sake of our analysis, the first three residues at the *N*-terminus are denoted by $N_{(i)}$, $N_{(i+1)}$ and $N_{(i+2)}$ and the last three residues at the *C*-terminus are denoted by $C_{(i)}$, $C_{(i+1)}$ and $C_{(i+2)}$ respectively. The residues are counted from either the *N*-terminus or the *C*-terminus.) The results of our analysis of ten proteins are shown in Fig. 6, where the distributions of the (ϕ , ψ) values of the $C_{(i)}$ (filled circle), $C_{(i+2)}$ (open circle) and the $C_{(i+3)}$ (open square) residues of helices are plotted. The ϕ values of the $C_{(i+3)}$ residue significantly differ from

Fig. 6. (ϕ, ψ) plot showing the distribution of (ϕ, ψ) values for the residues occurring at $C_{(i)}$ position ($\bullet$), $C_{(i+2)}$ position ($\circ$) and $C_{(i+3)}$ position ($\square$)

See the text for the definition of the symbols. Residues whose dihedral angles have been plotted are from the helices containing seven residues or more in the same ten proteins as in Fig. 5.

those of the $C_{(i)}$ and $C_{(i+2)}$ residues. Many of these residues have $\phi > -60^\circ$. Similarly there is a noticeable difference in the ψ values of the residues located at the $C_{(i)}$, $C_{(i+2)}$ and $C_{(i+3)}$ positions. The residues in the $C_{(i)}$ position have comparatively higher ψ values than those of the residues occupying the $C_{(i+2)}$ and $C_{(i+3)}$ positions. It is noteworthy that in most of the helices the carbonyl O atom of the C-terminal residue is exposed to solvent medium (Shrake & Rupley, 1973). Our superposition map (Fig. 3e) also shows that the region in which the C-terminal residues frequently occur is more accessible to solvent than the regular helical conformational region.

With respect to the β -structure, the first and the last residues of a strand do not form the β -sheet hydrogen-bonding. However, it has been observed in the crystal structures of several proteins that there are a number of structural regions in which four or more successive residues are in the extended conformation, without participating in the β -structure hydrogen-bonding (Burgess *et al.*, 1974). The analysis of the dihedral angles of the first and the last residues (also the other two relevant residues at the N- and C-termini as for helices) at either end of the extended-structure segments shows that the (ϕ, ψ) values are scattered over a wide range (ϕ from -180° to -45° ; ψ from 100° to 180°). The distributions of the (ϕ, ψ) values of the terminal and middle residues do not show any significant difference, except for the fact

that the ϕ and ψ ranges for the middle residues are comparatively small (ϕ from -180° to -75° and ψ from 115° to 175°).

An interesting aspect of solvent accessibilities of amino acid residues has been dealt with by Chothia (1974, 1975), who has related the solvent accessibilities of amino acid residues to their hydrophobicities, and thus attempted to calculate the hydrophobic contribution to the free energy of chain folding. We have tried here to relate the solvent accessibilities to the probable conformations of amino acid residues that acquire stabilization via solvent interaction.

The results presented here are to be interpreted cautiously for the following reasons.

(1) The numerical value of the calculated accessibility does not impart any direct meaning, such as a strong hydrogen bond with a solvent molecule at the locations of highest accessibility; it simply suggests the possibility of accommodation of a solvent molecule. The strength of the solute-solvent hydrogen bond depends mostly on the proper directional approach of donor and acceptor groups. Also, a smaller accessible area may suffice to have a hydrogen bond with a solvent molecule and it is at present unknown how small an accessibility value is sufficient for such a bond. The relationship between the accessible surface area and the number of water molecules will indicate the number of water molecules necessary for the solvation of different conformations. As hydrogen atoms have not been considered individually in the present study, we have not attempted to study this problem here.

(2) An arbitrary boundary of number 5 with respect to the maximum accessibility value has been chosen, as this range was inferred to be sufficient for interpreting the experimental results satisfactorily. We assume in our discussions that the actual numerical value and the areas covered by the accessibility contour of X-5 reflect, in one sense, the probable extent of solvent mobility and conformational range for the particular atom(s). Therefore we speak of 'probable conformational regions' which are preferred for solvent interaction without invoking the solvent-solute interaction-energy criteria.

(3) The model itself could be compared with the earlier treatment by the hard-sphere model (Ramachandran & Sasisekharan, 1968), in which the allowed and disallowed conformational regions in the (ϕ, ψ) space were deduced without considering the relative stabilities of various conformations within the allowed region.

P. M. thanks the University Grants Commission, New Delhi, India, for financial assistance. We are grateful to Professor F. M. Richards for providing the subroutine for calculating the solvent accessibilities, and Professor P. S. Manisundaram for his encouragement.

References

- Alden, R. A., Birktoft, J. J., Kraut, J., Roberts, J. D. & Wright, C. S. (1971) *Biochem. Biophys. Res. Commun.* **45**, 337-344
- Arnone, A., Bier, C. J., Cotton, F. A., Day, V. W., Hazen, E. E., Richardson, D. C., Richardson, J. S. & Yonath, A. (1971) *J. Biol. Chem.* **246**, 2302-2316
- Birktoft, J. J. & Blow, D. M. (1972) *J. Mol. Biol.* **68**, 187-240
- Brant, D. A. & Flory, P. J. (1965) *J. Am. Chem. Soc.* **87**, 2788-2791
- Browne, W. J., North, A. C. T., Philips, D. C., Brew, K., Vanaman, T. C. & Hill, R. L. (1969) *J. Mol. Biol.* **42**, 65-86
- Burgess, A. W., Ponnuswamy, P. K. & Scheraga, H. A. (1974) *Isr. J. Chem.* **12**, 239-286
- Chandrasekaran, R., Lakshminarayanan, A. V., Pandya, U. V. & Ramachandran, G. N. (1973) *Biochim. Biophys. Acta* **303**, 14-27
- Chothia, C. H. (1974) *Nature (London)* **248**, 338-339
- Chothia, C. H. (1975) *Nature (London)* **254**, 304-308
- Dickerson, R. E., Takano, T., Eisenberg, D., Kallai, O. B., Samson, L., Cooper, A. & Margoliash, E. (1971) *J. Biol. Chem.* **246**, 1511-1535
- Forsythe, K. H. & Hopfinger, A. J. (1973) *Macromolecules* **6**, 423-437
- Gibson, K. D. & Scheraga, H. A. (1967) *Proc. Natl. Acad. Sci. U.S.A.* **58**, 420-427
- IUPAC-IUB Commission on Biochemical Nomenclature (1970) *Biochemistry* **9**, 3471-3479; also in *Biochem. J.* (1971) **121**, 577-585
- Kolaskar, A. S., Sarathy, K. P. & Sasisekharan, V. (1975) *Curr. Sci.* **44**, 35-38
- Kuntz, I. D. (1972) *J. Am. Chem. Soc.* **94**, 4009-4012
- Lee, B. & Richards, F. M. (1971) *J. Mol. Biol.* **55**, 379-400
- Lewis, P. N., Momany, F. A. & Scheraga, H. A. (1973) *Isr. J. Chem.* **11**, 121-152
- Mathews, F. S., Levine, M. & Argos, P. (1972) *J. Mol. Biol.* **64**, 449-464
- Momany, F. A., McGuire, R. F. & Scheraga, H. A. (1975) *J. Phys. Chem.* **79**, 2361-2381
- Nockolds, C. E., Kretsinger, R. H., Coffee, C. J. & Bradshaw, R. A. (1972) *Proc. Natl. Acad. Sci. U.S.A.* **69**, 581-584
- Perricaudet, M. & Pullman, A. (1973) *FEBS Lett.* **34**, 222-226
- Ponnuswamy, P. K. (1970) Ph.D. Thesis, University of Madras
- Ponnuswamy, P. K. & Manavalan, P. (1976) *J. Theor. Biol.* **60**, 481-486
- Ponnuswamy, P. K. & Sasisekharan, V. (1971) *Biopolymers* **10**, 565-582
- Ponnuswamy, P. K., Warme, P. K. & Scheraga, H. A. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **60**, 830-833
- Popov, E. M. & Lipkind, G. M. (1971) *Mol. Biol. (Moscow)* **5**, 624-636
- Pullman, B. & Pullman, A. (1974) *Adv. Protein Chem.* **28**, 347-526
- Pysh, E. S. (1974) *Biopolymers* **13**, 1557-1562
- Quiocho, F. A. & Lipscomb, W. N. (1971) *Adv. Protein Chem.* **25**, 1-78
- Ramachandran, G. N. & Sasisekharan, V. (1968) *Adv. Protein Chem.* **23**, 283-438
- Robson, B. & Pain, R. H. (1971) *J. Mol. Biol.* **58**, 237-259
- Scheraga, H. A. (1968) *Adv. Phys. Org. Chem.* **6**, 103-184
- Shrake, A. & Rupley, J. A. (1973) *J. Mol. Biol.* **79**, 351-372
- Venkatachalam, C. M. (1968) *Biopolymers* **6**, 1425-1436
- Venkatachalam, C. M. & Krimm, S. (1973) *Jerusalem Symp. Quantum Chem. Biochem.* **5**, 141-154
- Watson, H. C. (1969) *Prog. Stereochem.* **4**, 299-333
- Wyckoff, H. W., Tsernoglou, D., Hanson, A. W., Knox, J. R., Lee, B. & Richards, F. M. (1970) *J. Biol. Chem.* **245**, 305-328