Influence of Solvent Molecules on the Stereochemical Code of Glycyl Residues in Proteins

Narayanan Eswar, H.A. Nagarajaram, C. Ramakrishnan, and N. Srinivasan*
Molecular Biophysics Unit. Indian Institute of Science. Bangalore. India

ABSTRACT The Ramachandran steric map and energy diagrams of the glycyl residue are symmetric. A plot of (ϕ,ψ) angles of glycyl residues in 250 nonhomologous and high-resolution protein structures is also largely symmetric. However, there is a clear aberration in the symmetry. Although there is a cluster of points corresponding to the righthanded α -helical region, the "equivalent" cluster is clearly shifted to in and around the (ϕ,ψ) values of (90°, 0°) instead of being centered at the left-handed α-helical region of (60°, 40°). This lack of symmetry exists even in the (ϕ,ψ) distribution of residues from non-α-helical regions in proteins. Here we provide an explanation for this observation. An analysis of glycyl conformations in small peptide structures and in "coil" proteins, which are largely devoid of helical and sheet regions, shows that glycyl residues prefer to adopt conformations around (±90°, 0°) instead of right- and left-handed α -helical regions. By using theoretical calculations, such conformations are shown to have highest solvent accessibility in a system of two-linked peptide units with glycyl residue at the central C^{α} atom. This finding is consistent with the observations from 250 nonhomologous protein structures where glycyl residues with conformations close to (±90°, 0°) are seen to have high solvent accessibility. Analysis of a subset of nonhomologous structures with very high resolution (1.5 Å or better) shows that water molecules are indeed present at distances suitable for hydrogen bond interaction with glycyl residues possessing conformations close to $(\pm 90^{\circ}, 0^{\circ})$. It is suggested that water molecules play a key role in determining and stabilizing these conformations of glycyl residues and explain the aberration in the symmetry of glycyl conformations in proteins. Proteins 2002;49:326-334. \odot 2002 Wiley-Liss, Inc.

Key words: glycyl residues; protein structures; Ramachandran map; solvation

INTRODUCTION

The conformational space used by the backbone of an amino acid residue in proteins and peptides is best described by the Ramachandran map. The map defines the ranges of the backbone dihedral angles, (φ,ψ) , which are "allowed" under a stereochemical criteria. The earliest form of the steric map, based on the hard-sphere model of a system of two-linked peptide units, with an alanyl or glycyl

residue at the central C^{α} , had spatially distinct "allowed" regions in the (ϕ,ψ) space. A "bridge region" between the β-sheet and right handed α-helical regions of the map [around the (ϕ,ψ) point $(-90^{\circ}, 0^{\circ})$] and between the analogous regions of the glycyl map [around (±90°, 0°)] was subsequently included for two reasons. First, it was observed that there were some (ϕ,ψ) points of residues from peptide structures and the only protein structure available at that time, that of myoglobin, which occurred in those regions. Second, it was realized that the conformations in the "bridge region" presented stereochemically disallowed contacts of very small magnitude that could be relieved by giving a 5° allowance in the bond angle $\tau(N-C^{\alpha}-C)$, from 110° to 115°, in the theoretical calculations. The recalculated stereochemical map would then demarcate the bridge regions as allowed.4 But from further analysis of protein structures, where a good number of residues have backbone dihedral angles which fall in the bridge region, it was realized that owing to the complex tertiary interactions in a protein, it does not necessarily mean that all these residues do indeed have wider τ-angles.⁵

On the other hand, energy contours drawn for the same system of two-linked peptide units and superposed on the steric map showed that there is an almost exact correspondence between the energy map and the map based on the hard-sphere approximation. The energy map also revealed that there are very distinct minima that correspond to the repeating units of secondary structures, namely, right- and left-handed $\alpha\text{-helices}$ and $\beta\text{-strands}$. These minima agreed well with the clusters of the (φ,ψ) points observed in proteins and peptides.

Although the points mentioned above are generally true for the non-glycyl amino acid residues, the glycyl residues showed a contrasting difference. The inherent symmetry in the steric or the energy map of glycyl residue is not clearly supported by the observations on the known three-dimensional structures. Although a cluster of glycyl conformations is known to exist centred around $(-60^{\circ}, -40^{\circ})$, the

N. Eswar's present address is Laboratory of Molecular Biophysics, The Rockefeller University, 1230 York Avenue, New York, NY 10021-6399.

 $[\]rm H.S.$ Nagarajaram's present address is Centre for DNA Fingerprinting and Diagnostics, Nacharam, Hyderabad 500 076, India.

^{*}Correspondence to: N. Srinivasan, Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India. E-mail: ns@mbu.iisc.ernet.in

Received 14 April 2002; Accepted 25 June 2002

corresponding cluster centered around (60°, 40°) is nearly absent. This aberration in the symmetry persists even after removing the glycyl residues in the right-handed α -helical regions in proteins. Karplus suggested that a form of evolutionary pressure might be a reason for this asymmetrical distribution of glycyl (ϕ, ψ) points.

It is of interest that the preferred conformational region of the glycyl residue is also shifted closer to the ψ -axis. Glycyl residues in the crystal structures of proteins are observed to preferentially adopt conformations around (90°, 0°), which does not correspond to an energy minimum of the Ramachandran map. ^{7,9-10} A direct evidence for the preference of glycyl residues to adopt conformations around (90°, 0°) is provided by Nicholson et al., 11 who mutated a non-glycyl (Asn55) residue in the left-handed α -helical conformation [with backbone dihedral angles (58°, 49°)] in T4 lysozyme to a glycyl residue. Crystal structure of the mutant clearly showed a shift in the conformation at the mutated glycyl residue that adopts (ϕ,ψ) values of $(82^{\circ},\psi)$ 29°). Most of the reported energy calculations of conformations in the (ϕ,ψ) space do not specify an energy minimum for glycine residues in the bridge region, although most glycine residues have backbone dihedral angles that fall into this region. 12-20 Kolaskar et al. 21 suggested the existence of a twofold torsional potential for ψ , which shows a minimum at 0° and $\pm 180^{\circ}$. But although this form of the potential supported the existence of the bridge region, it did not stay consistent with the more populated α -helical region.

The bridge region of the map mainly consists of examples of residues participating in β-turns, some types of which require one of the residues to be confined to this region to maintain the associated $4 \rightarrow 1$ hydrogen bond.²² Nevertheless, the question of how such conformations, in the bridge regions that represent a strained backbone,8 are stabilized in the absence of an energy minimum remains unanswered.²³ Ponnuswamy et al.^{24,25} suggested that solvent interactions could play a role in determining unusual backbone conformations. Solvent interactions have been postulated in helping to maintain unusual mainchain conformations, such as kinks in the middle of α -helices^{26–28} or a sheet,⁵ solvent-mediated helix termination,29 and so forth. Many disallowed conformations of residues are also known to be stabilized by interactions with the surrounding solvent.³⁰

In this analysis, we describe a plausible hypothesis in which the solvent interactions could be the key factor of stabilization of glycyl conformations in the bridge regions of the Ramachandran map.

MATERIALS AND METHODS Analysis of Protein Structures

A data set of 250 nonhomologous and highly resolved (≥ 2 Å resolution) protein structures derived from the Protein Data Bank (PDB)³¹ was used for the analysis. When a protein had multiple homologous or identical chains, only one of them was used for the analysis. The data set comprises proteins involved in a range of functions and adopting a variety of folds from various structures.

tural classes. The list of proteins used along with the chain identifiers is given in the Supplementary Material.

The Subset of Proteins Considered as "Coil Proteins"

A subset of 21 proteins were identified from these 250 proteins, based on the secondary structural content of the protein, to represent the coil proteins. The secondary structural content of the proteins was calculated from the percentage of residues forming α -helices and β -strands. Any other type of conformation was considered a coil. A protein was classified as a coil protein if >50% of the residues in its polypeptide chain were identified as coil and neither of the categories of α -helices and β -strands contained >40% of the residues. The list of the codes of such proteins used is: 1fus, 1tgsI, 1gpr, 4gcr, 1frrA, 2sarA, 1cbn, 2cdv, 1fxd, 1fdn, 9wgaA, 2sn3, 2bbkL, 2ovo, 2cpl, 1isuA, 1pk4, 1hip, 1hpi, 1rdg, and 4mt2.

The Data Set of Linear Peptide Structures

Glycyl conformations in the crystal structures of small peptides were analyzed. We did not include the cyclic peptides in our analysis because the constraint of ring closure could influence the conformation of the main chain. A total of 111 linear peptide structures were identified from the Cambridge Structure Database (CSD) that contained at least one glycyl residue. Out of these, 15 structures were omitted from the analysis because they contained glycyl residues either as the first or the last residue of the sequence, for which the backbone (φ,ψ) angles could not be calculated.

Recognition of Secondary Structures in Proteins

A stretch of at least four consecutive residues was identified as an α -helix if all the (φ,ψ) values in this region lie within the region of α_R (defined by $-140^\circ \le \varphi \le -30^\circ, -90^\circ \le \psi \le 45^\circ)$. β -strands were identified in a manner similar to the α -helices with a criterion that at least four contiguous residues were in the β -region (defined by $-180^\circ \le \psi \le -30^\circ, 60^\circ \le \psi \le 180^\circ$ or $-180^\circ \le \psi \le -150^\circ$). Furthermore, a stretch of at least four consecutive residues that does not fall under any of the categories described above was classified as a loop, and the remaining nonsecondary structural non-loop residues were termed as "random coil residues."

Definition of Bridge Regions

For the sake of this analysis, the bridge regions of the Ramachandran map were defined to lie within an interval described by the region: $-120^{\circ} \leq \varphi \leq -60^{\circ}$ and $-30^{\circ} \leq \psi \leq +30^{\circ}$ for the left half of the map (b_R-region) and a corresponding region on the right half for the b_L region (defined by $120^{\circ} \leq \varphi \leq 60^{\circ}$ and $-30^{\circ} \leq \psi \leq +30^{\circ}$).

Solvent Accessibility Calculations

All the solvent accessibility calculations were made by using the algorithm of Lee and Richards.³² The accessible surface area (ASA) from the program output was converted to accessibilities by normalizing with the sum of

van der Waals radius of the atom and that of the water, in which single atoms were considered. In residue accessibilities, the ASA values were normalized with standard values reported by Miller et al. 33

RESULTS AND DISCUSSION Conformation of Glycine Residues in Proteins

A data set of 250 nonhomologous and well-resolved protein structures derived from the PDB31 was used to analyze glycyl conformations. A total of 4330 glycyl residues were identified from this data set whose backbone torsion angles (ϕ,ψ) are shown plotted in Figure 1. It is obvious that the glycyl residues adopt (ϕ, ψ) values that are normally disallowed for the other 19 amino acid residues. By using an interval defined in the section describing methods, the examples falling in the bridge regions were classified as the left bridge (called b_R owing to its proximity to the α_R region) and right bridge (b_L), corresponding to the left and right halves, respectively, of the Ramachandran map. A little more than one third (\sim 36%, N = 1557) of all the glycyl residues adopt conformations in the bridge regions of the (ϕ,ψ) map. Out of these, 294 examples fall into the b_B-region, and 1263 examples fall into the corresponding region on the right half of the map, the b_L region. It is clear that glycyl residues adopt conformations more often in the $b_{\rm L}$ region than in $b_{\rm R}$ region. This distinct preference for the b_L region may be attributed to the formation of the type II and type I' $\beta\text{-turns},$ which require the second residue of the turn to possess a b_L-conformation. 34,35 It has also been shown that when a non-glycyl residue in α_L -conformation is changed to glycine, the (ϕ, ψ) values of the glycyl residue move down into the b_Lregion. 11 Nevertheless, the question of how the glycyl conformation in these regions, which does not represent a local energy minimum, is stabilized remains unanswered. In the analysis presented here, we discuss the role of solvent molecules that could be responsible for the stabilization of the conformation of glycyl residues in the bridge region.

Conformation of Glycine Residues in Peptides and Coil Proteins

The exact sequence of events that transform the amino acid sequence of a protein to its finally folded form is not completely known. In terms of the conformations of individual residues and their side-chains it has been shown that, for instance, the formation of secondary structures affects a few residues in its vicinity. 36,37 In addition, in the folded form of the protein, the packing of the secondary structures influences the side-chain conformations of the residues that get buried in the process.38 Examples of preferred local conformations have been reported in many cases like helix-helix packing³⁹ and interstrand correlation between residues in a β -hairpin,⁴⁰ intersecondary structural packing,^{41,42} and so forth. To rule out the effect of formation of helical and β-sheet regions on backbone conformation of glycyl residues, we made an analysis by using two data sets. The first consisted of peptide structures derived from the Cambridge Structure Database

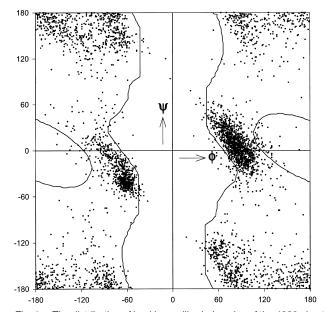


Fig. 1. The distribution of backbone dihedral angles of the 4330 glycyl residues derived from the data set. The Ramachandran Gly map is shown superimposed on the distribution to highlight the clustering of points in the bridge regions of the map.

(CSD), and the other is a set of protein structures that are largely devoid of helical and β -sheet regions.

The advantages of analyzing glycyl conformations from peptide structures are twofold. The peptide structures provide a means of analyzing very high-resolution data, which is not possible in the case of proteins because of the inherent features of structure determination. Second, peptide structures are usually devoid of the effects of secondary structure packing and other tertiary interactions observed in proteins. Thus, the properties of glycyl residues derived from peptide structures can be thought of as an inherent feature of the residue that is not influenced by the complexities of structures observed in proteins. The second data set consisted of a subset of 21 proteins derived from the original data set of 250 proteins, which were largely devoid of secondary structures such as α -helices and β-strands (criteria for the selection of these proteins is detailed in Materials and Methods). As in peptide structures, the rationale behind choosing such a data set was to minimize the influence of secondary structures on the conformation of glycyl residues. By considering such proteins, which are made up of predominantly irregular regions, we expect to identify intrinsic preferences of glycyl conformations. It has indeed been previously reported that intrinsic (ϕ,ψ) preferences of amino acid residues can be derived from analyzing the nonsecondary structural regions of proteins.37

The distribution of the backbone torsion angles of 99 glycyl residues from the 96 linear peptide structures is shown in Figure 2(a). Approximately 53% (N = 52) of the examples have (φ,ψ) values that fall into the bridge region of the map. Among these, 20 examples possess the b_R conformation, whereas the remaining 32 examples have the b_L conformation. It is evident from this that the

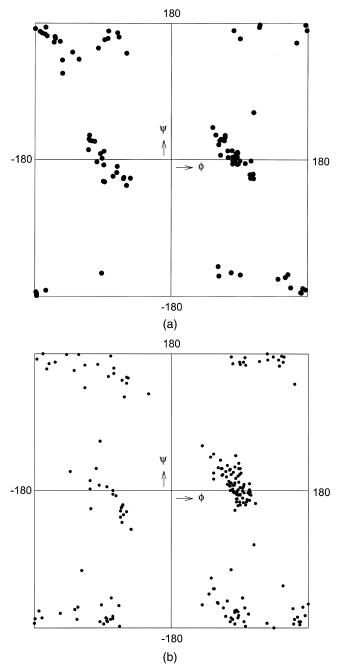


Fig. 2. A plot of the (ϕ,ψ) values from (a) 99 glycyl residues derived from 96 peptide structures derived from the CSD and (b) 202 glycyl residues from the 21 proteins, which were identified as coil proteins.

clustering of glycyl points near the bridge region is an intrinsic feature of glycyl residue. These results are in correspondence with those of Ashida et al., who also pointed out that glycine residues derived from peptides that possess either the $b_{\rm R}$ or $b_{\rm L}$ conformation do indeed have higher values of the $\tau(N\text{-}C^{\alpha}\text{-}C)$ angle. Work of Ramachandran and coworkers showed that slight strain involved in the bridge conformations is relieved when τ is increased to about 115°.

A total of 202 glycyl residues were identified from the 21 proteins whose structures are almost completely devoid of helices and extended strands. The distribution of these points is shown in Figure 2(b). Close to 42% (N = 84) of the examples fall into the bridge region. Among these, 13 examples fall into the $b_{\rm R}$ region, and as many as 71 examples fall into the $b_{\rm L}$ region. It can again be seen that the clustering of points is greater near the $b_{\rm L}$ region than that of the $b_{\rm R}$. It is also found that 54 examples of the total of 202 glycines are part of β -turns. The type II and I' β -turns, which are described by the motif $\beta/\alpha_{\rm L}$ -b_L account for 40 examples. The remaining 14 examples stem from the type II' or I β -turns described by the motif $\beta'/\alpha_{\rm R}$ -b_R [β' corresponds to (φ,ψ) angles in the region (60°, -120°)].

This finding shows that the clustering of the (φ,ψ) values of glycyl residues in the bridge region is an intrinsic property of the glycyl residues. The participation in the formation of β -turns could be one of the factors responsible in holding the conformation of these residues in either the b_R or b_L region, which is necessary to complete the $4\to 1$ hydrogen bond, a characteristic of these turns.

Accessibility of Main-Chain Polar Groups of Glycyl Residue

The intrinsic preference of glycyl residues to adopt conformations in the bridge regions of the steric map that do not enclose an energy minimum, was investigated for the possible sources of stabilization of such conformations. Interactions of the main-chain polar groups with the surrounding solvent molecules were hypothesized as a possible stabilization scheme. To this end, by using a system of two linked peptide units with glycyl residue at the central C^{α} atom, the accessibility of the main-chain amides and carbonyls were calculated by scanning the entire conformational space in intervals of 10° for both ϕ and ψ rotations. 24,32 The results are shown in Figure 3(a and b). Figure 3(a) shows the equi-accessibility contours of the amide hydrogen superposed on the conformational (ϕ,ψ) space. Because of the inherent symmetry of the residue, the contours are centro-symmetric about the (0°, 0°) point. The region enclosed by the first contour of 25% roughly extends from -60° to $+60^{\circ}$ of ϕ and encompasses the entire range of ψ in this interval. On the other hand, the equi-accessibility contours of the carbonyl oxygen [shown in Fig. 3(b)] shows a different trend from that of the amide hydrogen. These contours are also centro-symmetric about the central point of the (ϕ,ψ) space but are inverted along the ϕ -axis, in contrast to the contours of the amide hydrogen which are inverted along the ψ -axis. The region of maximum accessibility is enclosed by the 40% contour that approximately extends from -60° to $+60^{\circ}$ of ψ and encompasses the entire range of φ in this interval. The carbonyl oxygen shows maximum accessibility when the backbone torsion angle ψ is close to 0°.

To arrive at possible conformations that render the backbone polar groups with maximum accessibility, the sum of the equi-accessibility contours of the amide hydrogen and the carbonyl oxygen were superimposed on the steric map for glycine, as shown in Figure 4. The set of

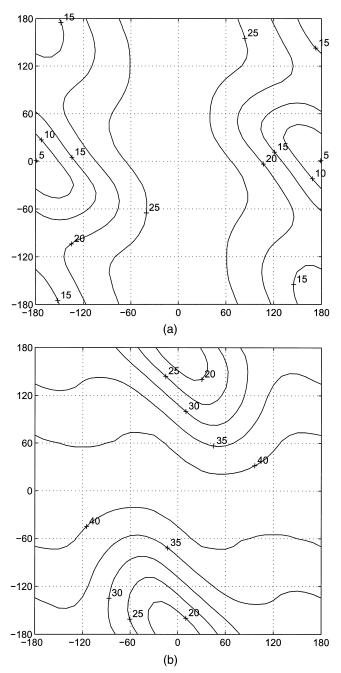


Fig. 3. The equi-accessibility contours of (a) the amide hydrogen of the glycine residue in a system of two-linked peptide units with the accessibility (%) represented by the contours marked with (+) and (b) the carbonyl oxygen of the glycine residue. The two sets of contours are symmetric with respect to the ψ - and φ -axis, respectively.

 (ϕ,ψ) points that belong to the intersection of the areas enclosed by the maximum accessibility contours of the polar groups and the steric map would be the points that maximally expose these polar atoms. It can be seen that the region enclosed by the maximum accessibility contour (65%) of the polar groups taken together encloses largely only the disallowed regions of the steric map. However, the next contour of 60% accessibility encloses a substantial portion of the allowed region of the steric map. It is of much

interest that it can be seen that the region of intersection between the accessibility contours and the allowed regions of the steric map that represents the combination of the torsion angles, which render the polar groups maximally exposed, is indeed the bridge regions of the map. For the sake of comparison, the ideal (φ,ψ) values of the right- and left-handed $\alpha\text{-helices}$ are also indicated in the figure (as crosses). It can be seen that these points fall outside the region enclosed by the maximum accessibility contour.

As a control to the glycine residue, similar calculations were performed on the system of two-linked peptide units with L-alanyl residue at the central C^{α} atom. The equiaccessibility contours, representing the sum of the accessibilities of the backbone polar groups of the residue, were superimposed on the steric map of alanine, and the result is shown in Figure 5. Remarkably, the bridge region of the steric map (only on the left half of the map in this case) does not fall inside the area enclosed by the maximum accessibility contour. It is surprising that the maximum accessibility contour encloses the region defined in the steric map for a y-turn. 44 It is also interesting to note that although the 60% contour does not cut across any allowed region of the map, the 55% cuts across the bottom portion of the region defined as the left-handed α -helical region. It may be recalled that the (ϕ, ψ) points of non-glycyl residues adopting a left-handed α -helical conformation, fall only in this region of the map. 45

From the above calculations it may be suggested that solvent interactions play an important role in maintaining these partially allowed main-chain conformations. The above discussion shows that the conformations lying at the bridge region for glycine, the γ -turn conformation for non-glycyl non-protein residues, and the left-handed α -helical conformations of non-glycyl residues tend to expose their backbone polar groups maximally to the solvent. It is already known, from analysis of protein structures that the residues with these backbone conformations are indeed on the surface of the protein and are exposed to the solvent. ⁴⁵ Thus, it may be suggested that such partially allowed conformations are stabilized by hydrogen-bonding interactions between the backbone polar groups and the surrounding solvent.

Solvent Accessibility of Glycines in the Known Structures of Proteins

The solvent accessibilities of the 4330 glycine residues identified from the data set of 250 proteins were computed by using the Lee and Richards algorithm³² and were normalized with tripeptide values listed by Miller et al.³³ These 4330 residues were then divided into three classes based on the values of accessibility, into <10% (low), between 10 and 50% (intermediate), and >50% (high). The distribution of the (φ,ψ) points of the residue in each of these classes is shown in Figure 6(a–c). In Figure 6(a), which represents the examples with low accessibility, the predominant cluster of points is at the α_R region, and the density of points near the b_R and b_L regions is very sparse compared to the cluster at α_R . There are only 103 residues in the b_R region and 215 examples in the b_L region. In

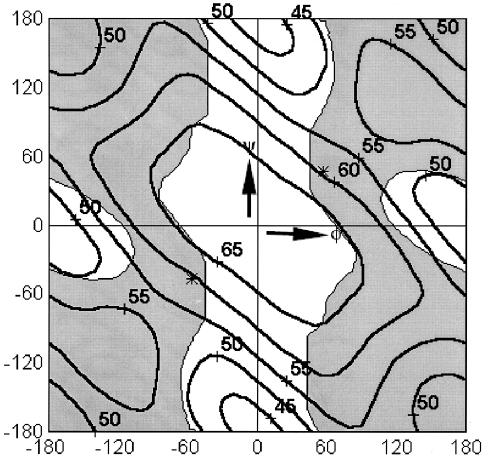


Fig. 4. The equi-accessibility contours of the total accessibility of the amide hydrogen and carbonyl oxygen of the glycyl residue in a system of two-linked peptide units. These contours are superimposed on the Ramachandran Gly map. The characteristic (ϕ,ψ) values of the right-handed and left-handed α -helices are also shown as (+).

contrast, in Figure 6(c), the strong cluster of points at the $\alpha_{\rm R}$ region has been dispersed, and it extends into the $b_{\rm R}$ region with about 100 examples in this region. In addition, the density of points in the b_L region is very high with as many as 632 examples in this region. Figure 6(b), which represents the points with intermediate accessibilities, shows a convenient transition between the characteristics of low- and high-accessibility points. This category of examples has 91 examples in the $b_{\rm R}$ region and 418 examples in the $b_{\rm L}$ region. There appears to be greater preference for the b_L region compared to the b_R for the glycyl residues, a fact that has already been reported by Ramakrishnan and Srinivasan.⁹ This goes to show that glycine residues with backbone dihedral angles in the bridge regions are indeed exposed to the solvent even in the context of proteins, with all their tertiary interactions included.

Observed Interactions of Bridge Glycines

Because interactions with the surrounding solvent molecules could be the major stabilizing feature of glycine residues occurring in the $b_{\rm L}$ and $b_{\rm R}$ regions, we confirm the existence of such interactions by a more detailed analysis

of protein structures. Because of the inherent problems in fixing the exact position of water during refinement of protein structures due to its high mobility, a subset of the original 250 proteins was devised; the subset consisted of 26 crystal structures of proteins with a resolution of ≤ 1.5 A and had positions of water molecules available. From this data set, 352 glycyl residues were identified, out of which, 124 examples occurred in the bridge regions (96 examples in the b_L region and 28 examples at the b_R). A glycyl residue was considered to be potentially interacting with water if a molecule of water was identified anywhere within a circle of radius 3.6 Å centered around any one of its backbone polar groups. It was observed that almost close to 70% (N = 67) of the examples occurring in the b_L region and close to 82% (N = 23) of the examples from the b_R region were interacting with water molecules. Of the 90 examples in the bridge regions that were interacting with water, most (51%, N = 46) of the examples were found to be interacting through the carbonyl oxygen of the glycyl residue. This correlates very well with the theoretical results described earlier where it was shown that the carbonyl oxygen atoms of glycyl residues show high accessibility in the bridge regions with a maximum around $\psi =$

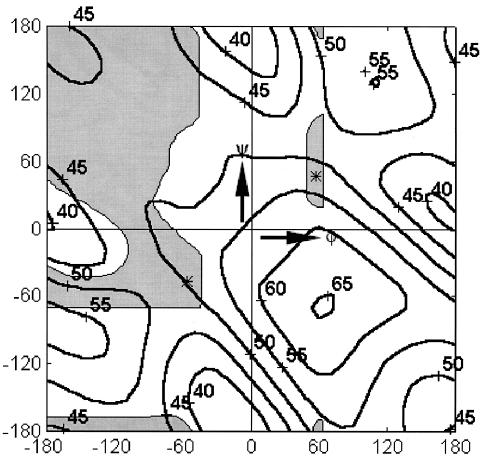


Fig. 5. Equi-accessibility contours of the total accessibility of backbone polar groups of an alanine residue superimposed on the Ramachandran Ala map. The maximum accessibility contour (65%) encloses the region that defines the classic γ-turn.

0°. About 21 examples had water molecules attached to the amide nitrogen, and 23 examples had interacting water molecules near either of the polar groups.

Although we observe that most of the glycyl residues adopting conformations in the bridge regions interact with water, it is also possible that glycyl residues adopting other conformations and not present in helices and β -sheets interact with water. This is also suggested by the solvent accessibility calculations on two-linked peptide system with the glycyl residue in the middle. As seen earlier, high solvent accessibility is not only seen for bridge glycines but also for some other regions such as extended conformation.

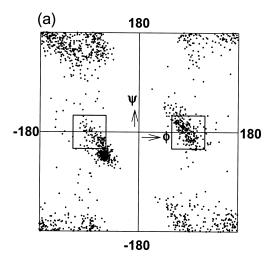
An examination of the glycyl residues with bridge conformation in the peptides suggests that, in general, the main-chain amide and carbonyls have high solvent accessible surface area enabling their interaction with water or with a neighboring molecule and within the molecule.

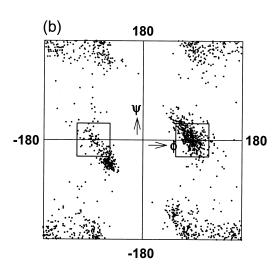
It is apparent from the peptide structures that the main-chain N at the glycyl residue at position i with its ψ close to 0° could form a favorable electrostatic interaction with the amide N-H group at the position i+1. However, this requires the amide N of Gly to be slightly pyrimidalized. Although such interactions are sometimes observed in the peptide structures, the main-chain polar atoms at

glycyl residues in the bridge region could interact with water without a requirement of marked deviation of the bond lengths and angles from the ideal values.

CONCLUSIONS

From the analysis of the glycyl residues in the known structures of proteins, it has been confirmed that there is an aberration to the symmetry of glycyl conformations. This is caused by the tendency of the glycyl residues to adopt conformations in the b_L region more often than in the b_R region. The conformations represented in the bridge regions of the steric map usually could be those that arise from residues participating in β-turns. But from the theoretical analysis presented in this article, it is clear that whatever the causes of the glycyl residues adopting the 'bridge' conformations are, in the absence of an energy minimum, the solvent interactions seems to provide a stabilizing effect. This is exhibited by the fact that the accessibility of the backbone polar groups of the glycyl residues is a maximum only for conformations in the bridge regions. The tendency of the backbone polar atoms at Gly with b_L conformation to interact with water is consistent with observations in crystal structures of pep-





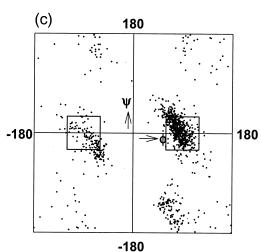


Fig. 6. The distribution of (ϕ,ψ) points of the 4330 glycines observed in the data set in the three classes (a) low accessibility (\leq 10%), (b) medium accessibility (between 10 and 50%), and (c) high accessibility (\geq 50%). The b_B and b_I regions are represented by rectangles.

tides and coil proteins where the influence of helices and β -sheets on the conformation of a residue does not exist.

Hence, although the accessible conformational space of glycyl residue is large and the stereochemical code of the glycyl residue is determined by the preferred structural motifs (helices and sheets) in proteins, it can sometimes be highly influenced by interactions with solvent.

ACKNOWLEDGMENT

N.S. is supported by a senior research fellowship from the Wellcome Trust, London.

REFERENCES

- Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. J Mol Biol 1963;7: 95–99.
- Ramakrishnan C, Ramachandran GN. Stereochemical criteria for polypeptide and protein chain conformations. Part II. Allowed conformations for a pair of peptide units. Biophys J 1965;5:909– 933.
- Ramachandran GN, Sasisekharan V. Conformation of polypeptides and proteins. Adv Protein Chem 1968;23:283–437.
- Ramakrishnan C. X-ray studies on biological fibres. Ph.D. thesis. Department of Physics, University of Madras, India, 1965.
- Richardson JS, Richardson DC. Principles and patterns of protein conformation. In: Fasman GD, editor. Prediction of protein structure and the principles of protein conformation. New York: Plenum Press; 1989. p 1–98.
- Richardson JS. Anatomy and taxonomy of protein structures. Adv Protein Chem 1981;34:167–330.
- Ramakrishnan C, Srinivasan N, Prashanth D. Conformation of glycyl residues in globular proteins. Int J Pept Protein Res 1987;29:629-637.
- Karplus PA. Experimentally observed conformation-dependent geometry and hidden strain in proteins. Protein Sci 1996;5:1406– 1420.
- Ramakrishnan C, Srinivasan N. Glycyl residues in proteins and peptides: an analysis. Curr Sci 1990;59:851–861.
- Ramakrishnan C, Srinivasan N. Conformational preference of glycyl residues in proteins—a further confirmation. Proc Ind Natl Sci Acad 1993;B59:45–54.
- Nicholson H, Soderlind E, Tronrud DE, Matthews BW. Contributions of left-handed helical residues to the structure and stability of bacteriophage T4 lysozyme. J Mol Biol 1989;210:181–193.
- Ramachandran GN, Venkatachalam CM, Krimm S. Sterochemical criteria for polypeptide and protein chain conformations. Biophys J 1966;6:849–872.
- Brant DA, Miller WG, Flory PJ. Conformational energy estimates for statistically coiling polypeptide chains. J Mol Biol 1967;23:47– 65
- Zimmerman SS, Pottle MS, Nemethy G, Scheraga HA. Conformational analysis of the 20 naturally occurring amino acid residues. Macromolecules 1977;10:1–9.
- Nemethy G, Scheraga HA. Protein folding. Q Rev Biophys 1977;10: 239–352.
- Dunfield LG, Burgess AW, Scheraga HA. Energy parameters in polypeptides. Empirical potential energy algorithm for the conformational analysis of large molecules. J Phys Chem 1978;82:2609– 2616
- 17. Madison V, Kopple KD. Solvent dependent conformational distribution of some dipeptides. J Am Chem Soc 1980;102:4855–4863.
- Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization and dynamics calculations. J Comp Chem 1983;4: 187–217.
- Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S Jr, Weiner P. A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 1984;106: 765–784.
- Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins,

nucleic acids and organic molecules. J Am Chem Soc 1995;117: 5179-5197.

- Kolaskar AS, Sarathy KP, Sasisekharan V. The need for a modified psi-potential in the dipeptide model. Curr Sci 1975;44:35– 38.
- 22. Venkatachalam CM. Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. Biopolymers 1968;6:1425–1436.
- O'Connell TM, Wang L, Tropsha A, Hermans J. The "random-coil" state of proteins: comparison of database statistics and molecular simulations. Proteins 1999;36:407–418.
- Ponnuswamy PK, Manavalan P. Preferred backbone conformations of amino acid residues for solvent interactions. J Theor Biol 1976;60:481–486.
- Manavalan P, Ponnuswamy PK, Srinivasan AR. Solvent accessibilities of glycyl, alanyl and seryl dipeptides. Biochem J 1977;167: 171–182.
- Blundell TM, Barlow DJ, Borkakoti N, Thornton JM. Solvent induced distortions and the curvature of α-helices. Nature 1983; 306:281–283.
- Baker EN, Hubbard RE. Hydrogen bonding in globular proteins. Prog Biophys Mol Biol 1984;44:97–179.
- Thanki N, Umrania Y, Thornton JM, Goodfellow JM. Analysis of protein main-chain solvation as a function of secondary structure. J Mol Biol 1991;221:669–691.
- Gunasekaran K, Nagarajaram HA, Ramakrishnan C, Balaram P. Stereochemical punctuation marks in protein structures: glycine and proline containing helix stop signals. J Mol Biol 1998;275:917– 932.
- Gunasekaran K, Ramakrishnan C, Balaram P. Disallowed Ramachandran conformations of amino acid residues in protein structures. J Mol Biol 1996;264:191–198.
- 31. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: a computer based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.

- 32. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol 1971;55:379-400.
- Miller S, Janin J, Lesk AM, Chothia C. Interior and surface of monomeric proteins. J Mol Biol 1987;196:341–356.
- 34. Wilmot CM, Thornton JM. Analysis and prediction of the different types of β-turns in proteins. J Mol Biol 1988;203:221–232.
- 35. Wilmot CM, Thornton JM. β-turns and their distortions: a proposed new nomenclature. Protein Eng 1990;3:479–493.
- Munoz V, Serrano L. Intrinsic secondary structure propensities of the amino acids, using statistical phi-psi matrices: comparison with experimental scales. Proteins 1994;20:301–311.
- 37. Swindells MB, MacArthur MW, Thornton JM. Intrinsic φ, ψ propensities of amino acids, derived from the coil regions of known structures. Nat Struct Biol 1995;2:596–603.
- Dunbrack RL Jr, Karplus M. Backbone dependent rotamer library for proteins. Application to side-chain prediction. J Mol Biol 1993;230:543–574.
- 39. Chothia C, Levitt M, Richardson, D. Helix to helix packing in proteins. J Mol Biol 1981;145:215–250.
- Lifson S, Sander C. Specific recognition in the tertiary structure of beta sheets of proteins. J Mol Biol 1980;139:627–639.
- Ponder JW, Richards FM. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. J Mol Biol 1987;193:775–791.
- 42. Chothia C. Principles that determine the structure of proteins. Annu Rev Biochem 1984;53:537–572.
- Ashida T, Tsunogae Y, Tanaka I, Yamane T. Peptide chain structure parameters, bond angles and conformational angles from the Cambridge Structural Database. Acta Crystallogr 1987; B43:212.
- Milner-White EJ. Situations of gamma turns in proteins. Their relation to alpha helices, beta-sheets and ligand binding sites. J Mol Biol 1990;216:385–397.
- Eswar N, Ramakrishnan C. Occurrences of left-handed α-helical conformations in protein structures. In: Vijayan M, Yathindra N, Kolaskar AS, editors. Perspectives in structural biology. Bangalore: Indian Academy of Sciences; 1999. p 181–195.