Debnath Pal
Pinak Chakrabarti
Department of Biochemistry,
Bose Institute,
P-1/12 CIT Scheme VIIM,
Calcutta 700 054, India

Received 6 June 1999;
accepted 23 November 1999

Terminal Residues in Protein Chains: Residue Preference, Conformation, and Interaction

Abstract: *The known protein structures have been analyzed to find out if there is any pattern in the type of residues used and their conformation at the two terminal positions of the polypeptide chains. While the N-terminal position is overwhelmingly occupied by Met (followed by Ala and Ser), the preference for the C-terminal is not as distinct, the residues with highest propensities being Lys, Arg, Gln, and Asn. Only one main-chain torsion angle, ψ, can be defined for the N-terminal residue, which is found to be in the extended conformation due to a favorable electrostatic interaction between the charged amino group and the carbonyl oxygen atom. The distribution of the angle ϕ for the C-terminal residue, on the other hand, is not much different from that of the nonterminal residues. There are some differences in the distribution of the side-chain torsion angle χ_1 of both the terminal residues from the general distribution. The terminal segments are generally flexible and there is a tendency for the more ordered residues to have lesser solvent exposure. About 40% of the terminal groups form a hydrogen bond with protein atoms—a slight preference is observed for the side-chain atoms (more than half of which belong to charged residues) over the main-chain ones. Although the terminal residues are not included in any regular secondary structure, the adjacent ones have a high preference to occur in the β conformation. There is a higher chance of a β-strand rather than an α-helix to start within the first 6 positions from the N-terminal end. It is suggested that the extended conformation observed for the N-terminal residue propagates along the chain leading to the formation of β-strand. In the C-terminal end, on the other hand, as one moves upstream the α and β structures are encountered in proportion similar to the average value for these structures in the database. The cleavage site of the zymogen structures has a conformation that can be retained by the N-terminal residue of the active enzyme.* © 2000 John Wiley & Sons, Inc. Biopoly 53: 467–475, 2000

Keywords: *terminal residues; hydrogen bonding; residue preference; residue conformation; zymogen activation*

INTRODUCTION

The backbone of a protein molecule is governed by the two main-chain torsion angles, ϕ and ψ, of individual residues.[1] For the two terminal residues, however, one of the angles cannot be defined (Figure 1). Moreover, at the pH values normally used for crystallographic experiments, the carboxy- (C-)terminal carries a negative charge and the amino- (N-)terminal, with the usual pK_a in the range 6.8–8.0,[2] may also be positively charged. Consequently, both the steric and electrostatic factors prevailing on the terminal residues are different from the rest of the polypeptide chain. Though potential energy calculations have been carried out to determine the possible conformations of N- and C-terminal Gly and Ala residues,[3] we

Correspondence to: Pinak Chakrabarti; email: pinak@boseinst.ernet.in.

Biopolymers, Vol. 53, 467–475 (2000)

© 2000 John Wiley & Sons, Inc.

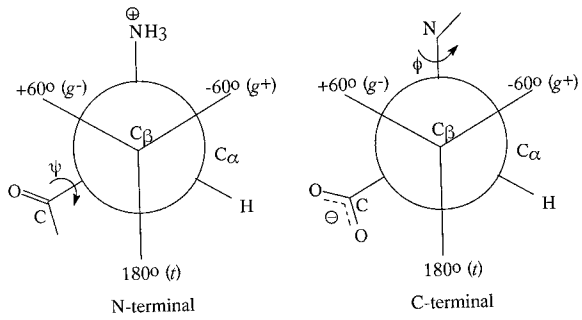

FIGURE 1 Newman projections down the C_β—C_α bond for the two terminal residues showing the ϕ and ψ torsion angles and the three positions of the γ atom corresponding to the χ_1 angle of $+60°$, $-60°$, and 180° (conformational states g^-, g^+, and t, respectively).

have recently shown that for non-Gly/Ala residues the distribution of ϕ, ψ angles and the side-chain conformation (χ_1) are correlated.[4] As such, we have carried out an analysis of the preference of residues to occupy the two terminal positions, their solvent accessibility and hydrogen-bonding features, the distribution of their main- and side-chain conformations, and compared these to the general pattern. The preference for the secondary structural elements in the terminal regions (consisting of 10 residues) and the location of the chain termini in the three-dimensional structure have been studied,[5,6] but no attempt has been made to understand the origin of the secondary structural preference. Along this line we also analyze the secondary structural elements that are observed closest to the chain termini and if the conformation of the terminal residues has any bearing on their formation.

The proteolytic enzymes of the eukaryotic organisms are synthesized as inactive precursors called zymogens, which have a polypeptide chain as an N-terminal extension of the active proteinases.[7] Activation of a zymogen is commonly accomplished by limited proteolysis that removes the proenzyme segment extension. Many biologically important processes, such as blood clotting cascade, are critically dependent upon this precursor to active enzyme conversion.[8] We have studied the conformational features of the N-terminal residue of these enzymes vis-à-vis those that exist in the corresponding zymogen structures.

METHODS

A dataset of 393 polypeptide sequences, from 385 files of Brookhaven Protein Data Bank (PDB[9]; January, 1999 version) was selected[10] such that they had a pairwise sequence identity of $\leq$25%, resolution better than 2.0 Å and R factor ≤ 0.20. In 10 cases, the N-terminal residue was derivatized with acetyl (label ACE) and pyroglutamyl (PCA) groups (in equal numbers). Records for atomic coordinates were searched for nonderivatized termini, and were found only for 255 and 279 free N- and C-termini residues. Some of these had large thermal parameters that make their coordinates less reliable. As we are concerned with ϕ, ψ, χ_1 angles, we found out the average B factor for the terminal atoms (TA) that define these torsion angles (Figure 1), and found a value of 34($\pm$19) Å^2 for N-TA, 37($\pm$22) Å^2 for C-TA, and 35($\pm$21) Å^2 overall. The data were then subdivided into two sets—those having an average $B > 40$ Å^2 (83 at N-terminus, and 96 at C-terminus), and with $\leq$40 Å^2 (172 and 183 cases, respectively, each group having an average B factor of 23($\pm$9) Å^2.

The torsion angles ϕ, ψ, and χ_1, following the recommendations of IUPAC-IUB Commission,[11] were calculated using the program DIHDRL provided by the PDB; for Val, however, 120° was added to χ_1 so as to make its atomic positions equivalent to those of Thr and Ile.[4] Depending on the value of χ_1 near $+60°$, $-60°$, or 180°, the side-chain conformation is designated as g^- (*gauche*$^-$), g^+ (*gauche*$^+$), or t (*trans*), respectively (Figure 1). In order to have a continuous distribution of points in the plots, the ranges used were -240 to 120° for χ_1, and -120 to 240° for ψ. The secondary structures were assigned in accordance with the algorithm DSSP.[12] The notations H and G represent helical structures (α and 3_{10}, respectively); E, β-strand; S and T, turns; and C, regions of no regular structure. The hydrogen bonding at the N- and C-terminal groups was found out by noting the presence of acceptor or donor moieties (at least one residue away) within 3.5 Å; as proton positions in the —NH_3^+ group cannot be fixed unambiguously, no distance or angle criteria involving the proton could be applied.

The propensity of each residue to occupy the terminal position was calculated first by finding out the percentage residue composition at the two terminal positions (considering all the 393 polypeptide chains) and in the whole database, and then dividing the former sets of values by the latter.

The solvent accessible surface area of a terminal residue was calculated using the program ACCESS[13] based on the algorithm of Lee and Richards.[14] A probe size of 1.4 Å for water was used. As a measure of maximum possible accessibility, values were calculated for X-Gly and Gly-X fragments (models for a residue, X, at the N- and C-termini, respectively), with the main chain in the extended conformation and the side chain at the most prevalent rotameric state.[15] The Gly residues in the two peptides were capped with N and C atoms, respectively.

The zymogen structures in PDB were found out by searching for protein names prefixed by "PRO". If more than one structure was available, the one with better resolution was retained. These were then paired up with the corresponding enzyme, if the structural data on the cleavage site was available in both the proteins.

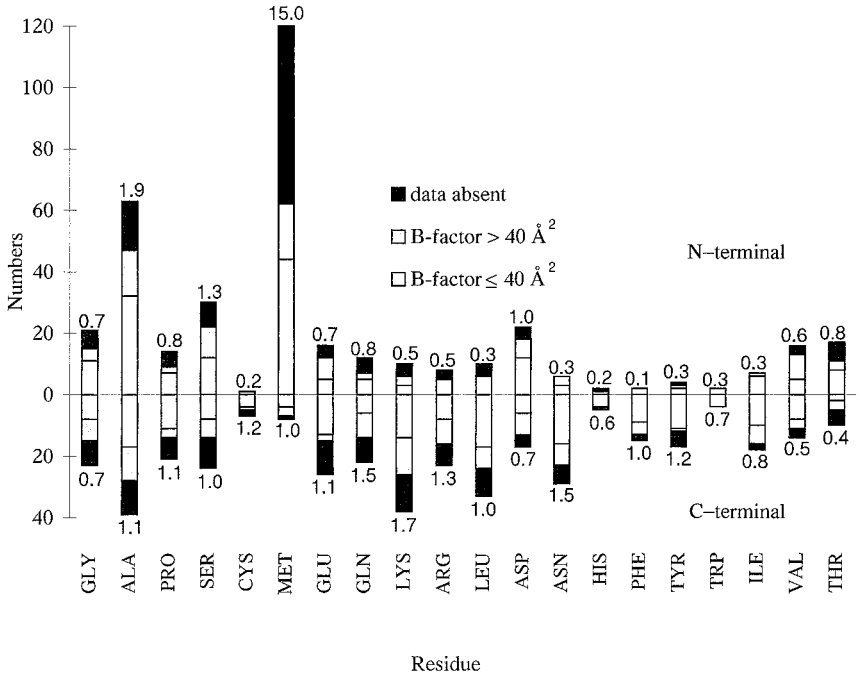

FIGURE 2 Histogram showing the residue preference for the terminal position. Each bar has three components: (i) cases, which according to the sequence records of the PDB file, occupy a terminal position, but for which no atomic coordinates are available; when the positional information is present, those with average *B* factor (ii) >40 Å^2 or (iii) ≤40 Å^2 are separated. The number against each bar corresponds to the propensity of the residue to occupy the terminal position.

RESULTS AND DISCUSSION

Our database contained 393 distinct polypeptide chains, not all of which had the positions of the terminal residues determined in the crystallographic analysis. Consequently, only 255 and 279 cases of free N- and C-terminal residues could be identified. Additionally, as the terminal regions are generally more flexible, a cut-off (40 Å^2) on the average thermal parameter has been used to distinguish the more resolved termini (172 and 183, respectively) from the more mobile ones.

Residue Preference

The occurrence of various residues at the two termini is presented in Figure 2. In eukaryotes all proteins are initiated[16] with a methionine (*N*-formyl methionine in prokaryotes), which is usually removed by aminopeptidases.[17] In PDB, Met is overwhelmingly the first residue of the chain. Ala has the next highest occurrence, and is also a prominent residue at the other end. Residues like Gly, Ser, and Asp are found in nearly equal numbers in both the ends. Unlike the N-terminus, the C-terminal position does not have any marked preference for any particular residue. Residues like Gly, Ser, and Asp are found in nearly equal numbers in both the ends. However, more illuminating than the absolute number of occurrence would be the propensity of a residue to occupy the terminal position, which is provided against each bar in Figure 2. A value of 1 represents the average propensity. Residues occurring more frequently than the average have propensity values greater than unity. Consideration of propensities indicate that Ala and Ser (besides the clear favorite, Met) have a distinct preference to occupy the N-terminal position, while long-chain basic residues (Lys and Arg) and those with amide side chains (Gln and Asn) have marked inclinations to occur at the other end.

Considering the thermal parameter based demarcation, it is found that Ala, Pro, Leu, Ile, and aromatic groups have a greater percentage of residues that are ordered, suggesting that the hydrophobic residues can be expected to make the chain termini less mobile.

Conformation

For non-Gly/Ala residues the interdependence between the main-chain torsion angle (ψ or ϕ) and the side-chain torsion χ_1 is shown in Figure 3(a and b). Considering only the ordered residues, the most strik-

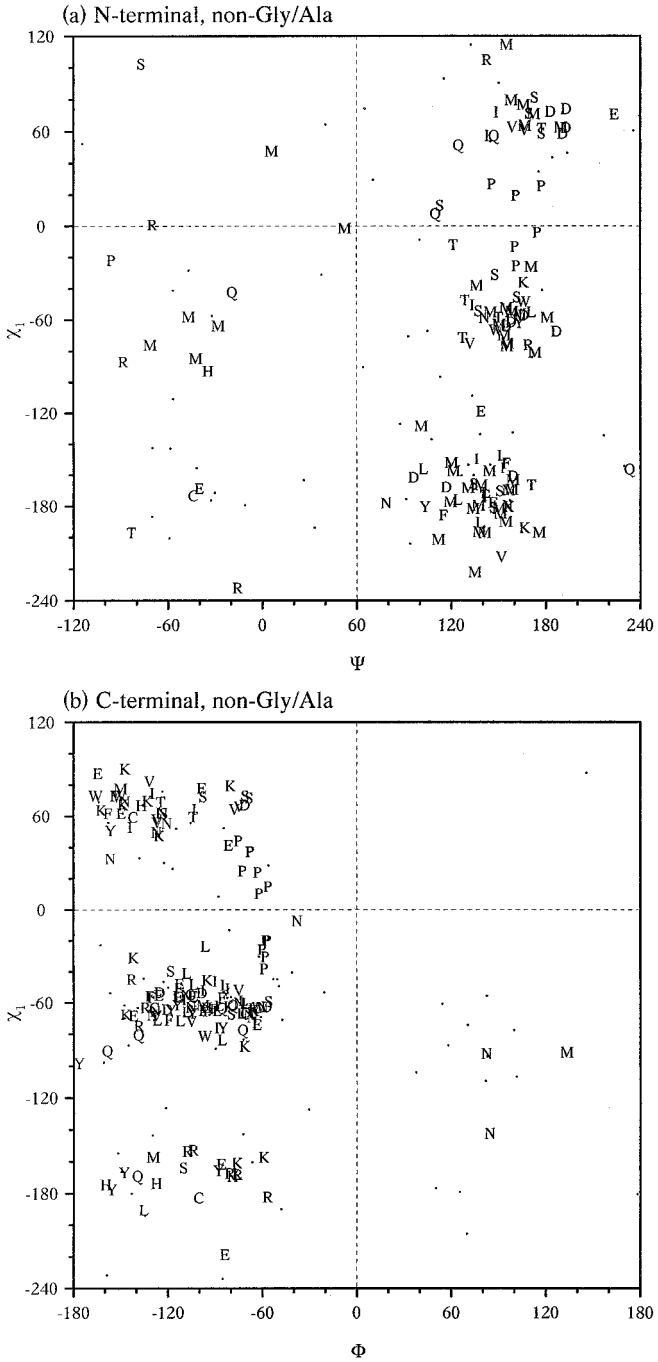

FIGURE 3 Joint distribution of (a) χ_1 and ψ, and (b) χ_1 and ϕ (in °), for residues occupying the first (127 cases) and the last (158) positions of the polypeptide chains. Positions are indicated by the one-letter amino acid code of the corresponding residue if the average B factor is ≤ 40 Å^2; otherwise, a dot is used. For Gly and Ala (with no χ_1) the distribution of ψ and ϕ angles are given in (c) and (d); the shading used in (d) is the same as in (c).

ing feature of Figure 3(a) is that there are only a few points below ψ of 60°; these are mostly in the range 110°–200° corresponding to an extended conformation. The preferred range for Ala is 140°–170°, and for Gly, 170°–210° [Figure 3(c)]. The origin for the inclination of the N-terminal residue to assume a value of ψ close to 180° must be electrostatics, as this conformation is stabilized by the *syn* orientation of the $—NH_3^+$ group and the carbonyl oxygen (carrying a partial negative charge; Figure 1). The pK_a of the α-amino group is 6.8–8.0,[2] depending on its environment and the identity of the terminal residue, and

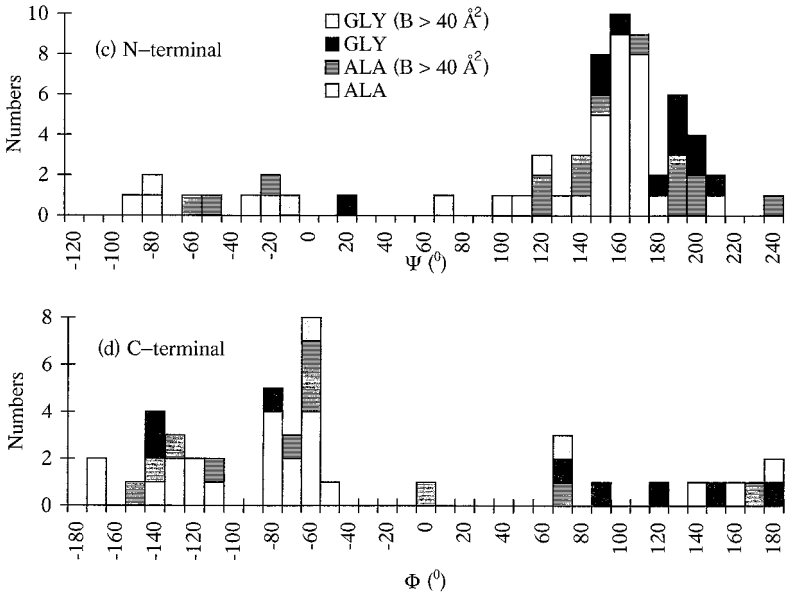

FIGURE 3 (*Continued from the previous page.*)

depending on the pH of the crystallization medium, even if it exists as —NH_2, rather than —NH_3^+, an amino proton may be suitably placed to interact favorably with the carbonyl oxygen atom when $\psi \approx 180°$.

As found for most of the nonterminal residues,[4] the mean of the ψ distribution shifts toward a more extended value [136(25)°, 151(15)°, and 160(25)°, using points with $\psi > 60°$] as the side-chain conformation is changed from t to g^+ to g^- states. The total numbers of points occurring in the t, g^+, and g^- states are 51, 47, and 29, respectively. Unlike the general distribution, where the population decreases in the order $g^+ > t > g^-$,[3] for the N-terminal residue the maximally occupied state is t, which places the side-chain atoms opposite to the —NH_3^+ group (Figure 1). As most of the residues are Met, it is unlikely that electrostatics is the primary reason for this observation. On the contrary, it is plausible that the relative preference for the t state increases because it offers the long side chains to have van der Waals contacts with the rest of the molecule, whereas in the other two states (especially, in g^+) these would point away from the main body of the molecule.

For the C-terminal residues, although ψ cannot be defined, the steric interaction may not be much different from a nonterminal residue as there is a carboxylate oxygen in place of the N atom (Figure 1). The χ_1, ϕ plot [Figure 3(b)] is essentially identical to similar plots for nonterminal residues.[4] However, the t state is the least occupied (the t, g^+, and g^- states are occupied with numbers 22, 90, and 46, respectively). The reason offered for the higher occurrence of the t state at the N-terminal may also be applied to explain its lower occurrence here. This conformation of the side chain, as compared to the other two conformations, would provide it with the least opportunity to come in contact with the rest of the molecule. This need for the optimum surface to pack against explains why polypeptide chains can crystallize only when they have some threshold length. For Ala, the ϕ values are distributed in two ranges, $-170°$ to $-110°$ and $-90°$ to $-50°$, whereas for Gly the points are widely spread, including the positive region of the ϕ.

As compared to the nonterminal residues, the proportion occurring in the g^- state increases relative to the other two states for both the terminal residues. This behavior is the same as has been found in small oligopeptide structures,[18,19] which have, because of the small length of the peptide chain, a relatively larger percentage of residues occurring at the two termini.

Secondary Structural Features

An analysis of the terminal regions had shown that the N-terminal preferentially adopts a β-sheet conformation and the C-terminal is usually helical, and this led to the suggestion that $\beta\alpha$ is the basic unit using which all α/β proteins are constructed.[5] In our database, the ratio of residues with β-sheet and α-helical conformation, $\beta : \alpha$, is 0.70. If the ten terminal residues are classified as α or β, depending on whether the first secondary structure encountered from the terminus is an α-helix or β-sheet, then the $\beta : \alpha$ ratio at the C-terminus (0.90) is close to the overall value, but at

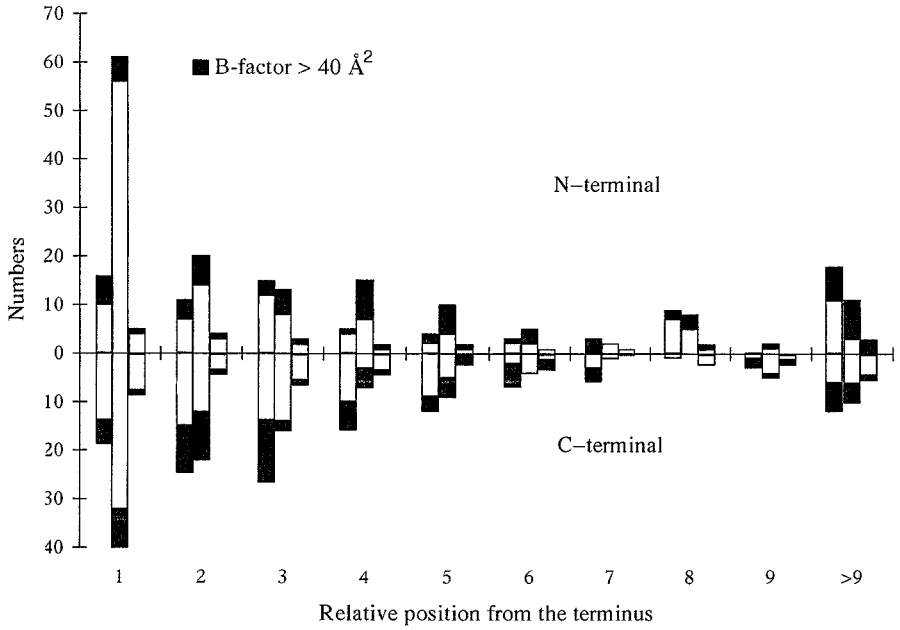

FIGURE 4 Histogram showing the variation of the sequence gap between the terminal position and the starting position of the nearest regular secondary structural element (α-helix, β-strand, and 3_{10}-helix, shown in this order against each position) along the polypeptide chain.

the N-terminus it is 2.1. This indicates that only the N-terminal region has a secondary structural preference different from the rest of the protein.

None of the residues at the two ends is a part of any regular secondary structure. We then found out at what position along the sequence is the first secondary structure encountered (Figure 4). Interestingly, the residue occurring next to the terminal has a very high propensity to be in the β conformation ($\beta : \alpha = 3.8$ and 2.1 for the two termini). The greater proclivity toward taking up the β conformation continues in the N-terminal region until the relative position of 6, beyond which for about 3 positions there is no particular preference, and then $\beta : \alpha$ ratio nears the average value. In contrast, in the C-terminus the preference for the β over the α structure shown at the relative position 1 is reversed in the next residue, and beyond 2 the $\beta : \alpha$ ratio approximates the average value.

It is possible to seek for an explanation for the higher occurrence of β structure in the N-terminal region in terms of the conformation of the terminal residue. It has been pointed out that because of electrostatics the N-terminal residue assumes an extended conformation [Figure 3(a and c)], which is propagated along the chain so that the next residue in particular has a very high probability to occur in the β conformation. On the other hand, the ϕ value of the C-terminal residue is not restricted to the extended conformation only [Figure 3(b and d)], and therefore can lead to both helix and sheet. As to why the preference for helical structure becomes conspicuous from position 3 onward, it is plausible that 2–3 residues are needed at the free end to satisfy the capping requirement of the helical C-terminus. It may be mentioned in this connection that a 12-residue synthetic peptide was found to be helical in the stretch 1–9, with an extended region near the C-terminus of the chain.[20]

Because of its role, the free amino end of a polypeptide chain can be described as a β-strand initiator, like the influence Pro exerts in initiating an α-helix.[21] Moreover, it is not yet clear whether proteins fold during biosynthesis or whether translation and folding are separated in time.[22] Whatever might be the mechanism it is obvious that the positive charge on the amino end controls the formation of the first secondary structure in the polypeptide chain.

Our observation has implications for the helix dipole model[23,24] that has been invoked to explain many observations related to protein structure and function, although there has also been argument[25] suggesting that instead of any long distance electric field exerted by the macroscopic dipole of the helix it is the specific hydrogen bond that may be important. A helix at either end of the polypeptide chain should be unstable because the positive charge developed at the helical N-terminus would have a repulsive interaction with the positive charge on the amino end of the chain, and similarly, destabilization will ensue for a helix formed

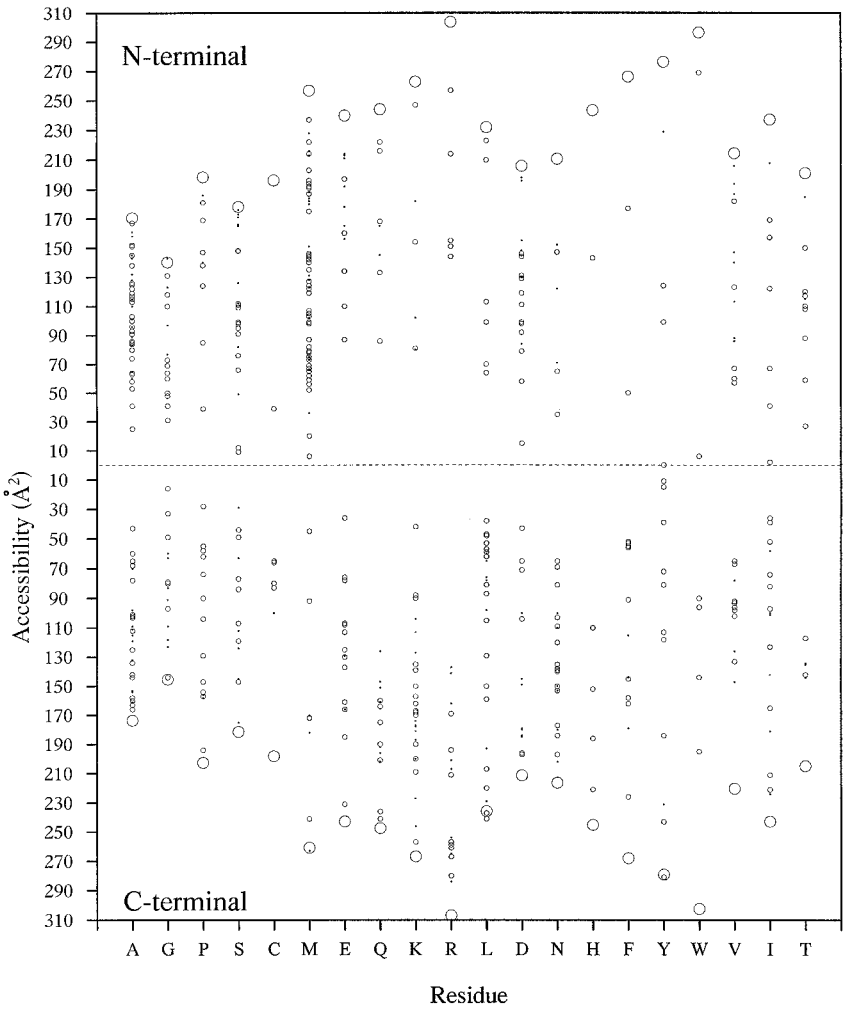

FIGURE 5 Absolute accessibility of the terminal residues. The large open circles indicate the values for individual residues in model dipeptides, X-Gly and Gly-X (representing N- and C-termini, respectively. The residues with average *B* factor > 40 Å^2 have been marked with dots, while more ordered ones have been represented by medium-sized open circles.

at the other end of the chain. Nonetheless, a considerable number of α-helices are observed near the chain ends, especially the C-terminus (though less frequent, 3_{10}-helices are almost evenly distributed between the two ends). Thus the effect of helix dipole is not enough to prevent the occurrence of helices near the chain termini.

Solvent Accessibility

The maximum accessible surface areas (ASA) of individual residues terminally located, as obtained from model peptide fragments, are indicated by large open circles in Figure 5. In real structures, the ASA values are smaller than these, but there is a wide variation. However, residues with greater flexibility (represented by dots), in general, have values closer to the maximum than the ordered ones (small, open circle). This is along expected lines, as the residues having more contact with the rest of the molecule (and thus with lesser ASAs) are more ordered.

Hydrogen Bonding

The difference in the chemical nature of the two ends gives rise to some differences in their hydrogen-bonding features (Table I). While two-thirds of them have no hydrogen-bond interaction at all at the N-terminus, the corresponding figure for the C-terminus is close to

Table I Statistics on Hydrogen-Bonding Involving the Amino and Carboxyl Groups at Protein Termini[a]

Hydrogen Bond	Amino	Carboxyl
Total number	255 (172,62,14,6,1)[b]	279 (158,74,35,5,4,2,1)[b]
Partner		
Main chain	50	93
Side chain[c]	62 (R-3,H-4,D-15,N-9,E-17,Q-5,T-2,S-6,Y-1)	98 (R-30,K-28,H-6,D-2,N-8,E-1,Q-4,T-4,S-10,Y-5)

[a] Only the protein partners are considered. If contact with water molecules is also included, 115 and 153 amino and carboxyl groups, respectively, have water molecules as neighbors.

[b] The number of groups, followed by those having 0, 1, 2 . . . number of hydrogen bonds (in parentheses) are given.

[c] One-letter amino acid codes of the residues involved and their numbers (after a hyphen) are given.

a half. As a hydrogen-bond partner, the side chain is preferred (1.2 times) to the main chain at the N-terminus, while both are equally observed at the C-terminus. Involvement with a charged side chain (salt bridge) is slightly more at the C-terminus (65% of all side-chain interactions) than the N-terminus (52%). There have been discussions on the implications of the proximity of N- and C-terminals for protein folding.[6,26,27] Interestingly enough, examples of salt bridges are observed in PDB files 1MRO (subunit B), 1XNB, 2PII, and 3CHY, where the two ends of the polypeptide chain are linked. The C-terminus exhibits a slightly higher chance of short-range contacts—in 24 cases the difference in sequence numbers between the terminal and the partner residues is less than 5, whereas only 9 examples are observed at the opposite end.

In the PDB file 1AH7, a zinc ion is found at a distance 2.07 Å from the amino group, while in structures (1XNB and 3NUL) there is a sulfate anion. Interestingly, there is a negatively charged side chain near the C-terminal in 3 cases, and a positively charged group near the N-terminal in 7. In the former cases, one of the acid groups may have an abnormal pK_a value (sharing a proton between them), while in the latter, the amino group must be in the neutral state (with Arg as partner) or its positive charge may be shared with the partner His ring.

Conformational Change During Zymogen Activation

In order to study if during the activation of zymogen the newly formed N-terminal residue of the enzyme has a conformation that already existed for the residue in the zymogen, we extracted a few pairs of such structures from PDB. Table II shows that the residue in zymogen that would form the N-terminal residue in the active

Table II Inactive (Zymogen) and Catalytic Forms of Some Enzymes and the Structural Features at the Cleavage Site[a]

Sl. No.	PDB Code	Protein	Cleavage Site (Residues *i*–*j*)	Residue *j* Conformation			Structure[b]	Relative Accessibility[c]	
				ϕ	ψ	χ_1		*i*	*j*
1	1SLM	Fibroblast stromelysin-1	P87-G88	−76	−150		CC	47	39
	1HFS				−114		C		86
2	1PFZ	Plasmepsin II	G51-S52	−157	−176	66	CB	24	18
	1SME				113	64	C		72
3	1PCI	Caricain	N106-L107	−126	−177	74	CS	85	23
	1MEG				137	−80	C		60
4	3PBH	Cathepsin B	K62-L63	−92	143	−101	CC	88	21
	1CSB				147	−71	C		47
5	2PSG	Pepsinogen	L44-I45	−118	107	177	SS	85	47
	4PEP				128	88	C		55

[a] Under each serial number, the two entries are for the zymogen and the active enzyme, respectively.

[b] The secondary structure from the DSSP[12] output: S = bend, B = β-ladder, and C = nonregular structure.

[c] The relative accessibility (%) of a residue was evaluated by the ratio of the summed atomic accessible surface areas of that residue in protein to that of the same residue (X) in an extended Gly-X-Gly tripeptide (for the zymogen) or the X-Gly fragment (active molecule).

enzyme already has an extended ψ (more extended than ϕ, which gets undefined on cleavage and is thus of no consequence to the conformation of the enzyme). This suggests that the spliced region has a conformation predisposed toward a value it would assume [Figure 3(a)] on cleavage of the peptide bond. The cleavage is also facilitated by the higher solvent exposure at the sites and the absence of significant secondary structure.

SUMMARY

An analysis of the conformational features of the terminal residues in polypeptide chains provides clues to some aspects of protein folding pattern. The electrostatic field due to the amino end may cause the main-chain torsion angle ψ of the first residue to be in an extended conformation, which in turn makes the next residue a more likely location for the start of a β-strand rather than an α-helix (which needs the chain to be folded). The greater preference for the β-conformation is observed for six locations from the end. Although the residue preceding the C-terminal end is more likely to have a β, rather than α, conformation, as one moves further upstream, the preference changes to the helical structure and the $\alpha : \beta$ ratio approaches the average value.

The side-chain rotamer population is different from the normal distribution in proteins and resembles preferences found in small polypeptides; the terminal charges may be a causative factor for this. Because of its known role as the chain initiator, Met, not unsurprisingly, is overwhelmingly found at the N-terminus, followed by Ala and Ser. The long positively charged residues (Lys and Arg), Gln and Asn are more favored at the other end. Although flexible in general, the terminal residues can also make considerable contact with the rest of the molecule, and the accessible surface areas show a large variation. Only about 40% of the terminal groups show any hydrogen-bond contact with the rest of the molecule. Because of the preference of the N-terminal residue for an extended ψ value, the residue that would form the N-terminus in the active enzyme is also made to have an extended ψ in the precursor zymogen structure, so that on cleavage there is very little change in the main-chain conformation.

DP is recipient of the NET fellowship and PC a grant from the Council of Scientific and Industrial Research. We also thank the Department of Biotechnology for supporting the computational facilities.

REFERENCES

1. Ramachandran, G. N.; Sasisekharan, V. Adv Protein Chem 1968, 23, 283–437.
2. Creighton, T. E. Proteins: Structure and Molecular Properties; W. H. Freeman and Company: New York, 1993; chap 1.
3. Ponnuswamy, P. K.; Sasisekharan, V. Int J Protein Res II 1970, 47–57.
4. Chakrabarti, P.; Pal, D. Protein Eng 1998, 11, 631–647.
5. Thornton, J. M.; Chakauya, B. L. Nature 1982, 298, 296–297.
6. Thornton, J. M.; Sibanda, B. L. J Mol Biol 1983, 167, 443–460.
7. Kassell, B.; Kay, J. Science 1973, 180, 1022–1027.
8. Neurath, H. Science 1984, 224, 350–357.
9. Abola, E. E.; Sussman, J. L.; Prilusky, J.; Manning, N. O. Methods Enzymol 1997, 277, 556–571.
10. Hobohm, U.; Sander, C. Protein Sci 1994, 3, 522–524.
11. IUPAC-IUB Commission on Biochemical Nomenclature. Biochemistry 1970, 9, 3471–3479.
12. Kabsch, W.; Sander, C. Biopolymers 1983, 22, 2577–2637.
13. Hubbard, S. ACCESS: A Program for Calculating Accessibilities; Department of Biochemistry and Molecular Biology, University College of London, 1992.
14. Lee, B.; Richards, F. M. J Mol Biol 1971, 55, 379–400.
15. Ponder, J. W.; Richards, F. M. J Mol Biol 1987, 193, 775–791.
16. Varshney, U.; Rajbhandary, U. L. Proc Natl Acad Sci USA 1990, 87, 1586–1590.
17. Taylor, A. Trends Biochem Sci 1993, 18, 167–171.
18. Chakrabarti, P. J Mol Biol 1994, 239, 306–314.
19. Chakrabarti, P. Int J Peptide Protein Res 1994, 43, 284–291.
20. Hill, C. P.; Anderson, D. H.; Wesson, L.; DeGrado, W. F.; Eisenberg, D. Science 1990, 249, 543–546.
21. Richardson, J. S.; Richardson, D. C. Science 1988, 240, 1648–1652.
22. Alexandrov, N. Protein Sci 1993, 2, 1989–1991.
23. Wada, A. Adv Biophys 1976, 9, 1–63.
24. Hol, W. G. J. Prog Biophys Mol Biol 1985, 149–195.
25. Chakrabarti, P. Protein Eng 1994, 7, 471–474.
26. Ptitsyn, O. B. FEBS Lett 1981, 131, 197–202.
27. Christopher, J. A.; Baldwin, T. O. J Mol Biol 1996, 257, 175–187.