Article No. jmbi.1999.3217 available online at http://www.idealibrary.com on IDEAL® J. Mol. Biol. (1999) 294, 271–288

JMB

# *Cis* Peptide Bonds in Proteins: Residues Involved, their Conformations, Interactions and Locations

**Debnath Pal and Pinak Chakrabarti***

*Department of Biochemistry, P-1/12 CIT Scheme VIIM Bose Institute, Calcutta 700 054, India*

**Corresponding author*

An analysis of a non-redundant set of protein structures from the Brookhaven Protein Data Bank has been carried out to find out the residue preference, local conformation, hydrogen bonding and other stabilizing interactions involving *cis* peptide bonds. This has led to a reclassification of turns mediated by *cis* peptides, and their average geometrical parameters have been evaluated. The interdependence of the side and main-chain torsion angles of proline rings provided an explanation why such rings in *cis* peptides are found to have the DOWN puckering. A comparison of *cis* peptides containing proline and non-proline residues show differences in conformation, location in the secondary structure and in relation to the centre of the molecule, and relative accessibilities of residues. Relevance of the results in mutation studies and the *cis-trans* isomerization during protein folding is discussed.



*Keywords:* *cis* peptide; pyrrolidine ring puckering; turn conformation; residue preference in *cis* peptides; weak interactions

## Introduction

In proteins, the partial double bond character of the peptide bond results in two conformations depending on the value of the dihedral angle, ω [$C_\alpha$(1)-C(1)-N(1′)-$C_\alpha$(1′)]: *cis* and *trans* (with $\omega = 0$ and 180°, respectively) (Pauling, 1960; Ramachandran & Sasisekharan, 1968) (Figure 1(a)). The isomer with the two $C^\alpha$ atoms *trans* to each other is favoured overwhelmingly due to the lesser steric conflict involving the substituents at these positions, and only when a Pro residue is in position (1′) is there a substantial steric clash involving the $C^\alpha$ atom at position (1) and $C^\delta$ atom of Pro at position (1′), even in the *trans* conformation, to give the *cis* imide bond, X-Pro, a higher frequency of occurrence than what is observed for the amide bond, X-Xnp.

A difference in energy of approximately 2.5 kcal/mol between the *trans* and the *cis* isomers (corresponding to only 1.5% occurrence of the *cis* form), regardless of the solvent, and a rotational barrier of about 20 kcal/mol have been found for the peptide bond analog *N*-methylacetamide (LaPlanche & Rogers, 1964; Christensen *et al*., 1970; Drakenberg & Forsén, 1971; Perricaudet & Pullman, 1973; Radzicka *et al*., 1988; Jorgensen & Gao, 1988; Schnur *et al*., 1989; Scherer *et al*., 1998). For an imide bond in Pro-containing peptides, however, the *trans* isomer is favoured over the *cis* by only 0.5 kcal/mol (Maigret *et al*., 1970), so that a higher abundance (10-30%) of the *cis* form is observed (Brandts *et al*., 1975; Grathwohl & Wüthrich, 1976; Juy *et al*., 1983); the activation energy barrier for *cis-trans* isomerization is also less, 13 kcal/mol (Schulz & Schirmer, 1984). Using conformational energy calculations, Ramachandran & Mitra (1976) found expected frequencies for the *cis* isomer to be 0.1% and 30% (corresponding to an enthalpy difference of 4.0 and 0.5 kcal/mol, respectively) for an Ala-Ala and Ala-Pro peptide bond, respectively. A survey of protein structures by Stewart *et al*. (1990) found only 0.05% of all X-Xnp, but 6.5% of all X-Pro peptide bonds to occur in the *cis* conformation. The analysis of MacArthur & Thornton (1991) provided a value of 5.7% for the latter group, whereas a recent work (Weiss *et al*., 1998; Jabs *et al*., 1999) gave values of 0.03% and 5.2%, respectively, for the two types of peptide bonds.

Due to the energy barrier, *cis-trans* isomerization of peptide bond is a rather slow process at room temperature and has been shown to play an important role in protein folding (Brandts *et al*., 1975; Creighton, 1978; Schmid & Baldwin, 1978; Cook *et al*., 1979; Lin & Brandts, 1984; Brandts &

Abbreviations used: X, any amino acid residue; Xnp, any non-Pro amino acid residue; Ar, any aromatic residue.

E-mail address of the corresponding author: pinak@boseinst.ernet.in

0022-2836/99/460271–18 $30.00/0

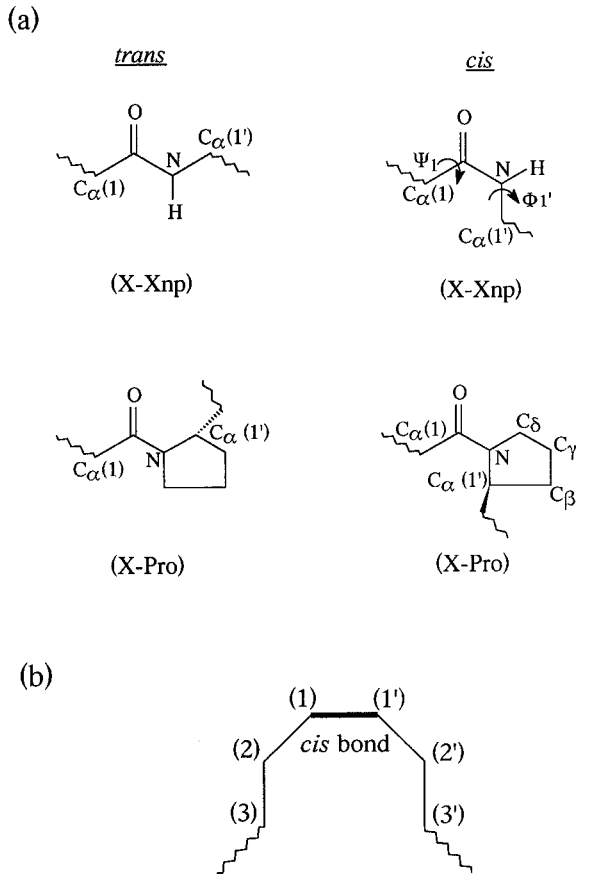



**Figure 1.** Schematic representation of *cis* and *trans* conformations around X-Xnp and X-Pro peptide bonds (where X = any residue, Xnp = any non-Pro residue). (b) Convention for numbering residues flanking a *cis* peptide bond.

Lin, 1986; Kim & Baldwin, 1990). An enzyme, prolyl isomerase is known to catalyze the *cis-trans* isomerization of X-Pro bonds (Schmid *et al*., 1993). Experimental data have been derived on the thermodynamics and kinetics of *cis-trans* isomerization by substituting a Pro at (1′) by a non-Pro residue (Schultz & Baldwin, 1992; Mayr *et al*., 1993; Tweedy *et al*., 1993; Odefey *et al*., 1995; Vanhove *et al*., 1996). However, to understand the structural effect of such mutations it is important to know the conformational features of residues in and around (Figure 1(b)) the X-Xnp *cis* peptide linkages *vis-à-vis* the X-Pro bond. Moreover, although a *cis* peptide bond can cause reversal of chain direction (Lewis *et al*., 1973) leading to two types of turns with the two central residues having canonical $\phi,\psi$ (degree) values of (−60,120; −90,0) and (−120,120; −60,0) (Richardson, 1981; Rose *et al*., 1985). These values, though widely quoted in literature (Wilmot & Thornton, 1988), need a reassessment as $\psi$ of the second residue in the second set is usually significantly off the idealized value. Consequently, we thought it important to make a reclassification of *cis* peptide mediated turns and an evaluation of the torsion angles of the involved residues.

We have recently investigated the interrelationship between the side-chain and the main-chain conformational angles in residues involved in the *trans* peptide units (Chakrabarti & Pal, 1998), and from this perspective it is worthwhile to study the relationship in *cis* peptide bonds. The pyrrolidine ring of Pro can be associated with two types of puckering, designated UP and DOWN, depending on the ring torsion angles (Ramachandran *et al*., 1970; Ashida & Kakudo, 1974). It has been noted by Milner-White *et al.* (1992) that the puckering of the ring when it is involved in the *cis* linkage is DOWN, and an explanation may be sought in terms of the interaction between the main-chain and side-chain atoms.

The *cis* peptide bonds, especially the ones with non-Pro residues, are located near the active sites or are implicated to have roles in the function of the protein molecule (Herzberg & Moult, 1991; Stoddard & Pietrokovski, 1998; Jabs *et al*., 1999). Though important, some of the *cis* peptide bonds might have gone unreported in the structures determined at lower resolution (Weiss *et al*., 1998). To facilitate the identification of such overlooked *cis* peptide bonds it is important to characterize the location of known *cis* peptide units (both X-Pro and X-Xnp) in the three-dimensional structures and their solvent accessibility. A comprehensive analysis of these issues is made here, so as to understand the interactions that stabilize a *cis* peptide bond and possibly identify regions/sequences in protein structures that are likely to adopt a *cis* peptide linkage.

## Results and Discussion

### Residues forming the *cis* peptide linkage

A total of 50 % (147 out of 294) of well-defined protein structures contain one or more *cis* peptide bonds; 0.3 % of all the bonds in the database exist in the *cis* form (231 in total). Most of them (87 %) are preceding Pro residues (5.7 % of X-Pro bonds have the *cis* conformation). The intrinsic probability of a residue (X) to cause a *cis* conformation of the X-Pro linkage, given by the fraction of occurrence of the bond in the *cis* form, is provided in Figure 2. Stewart *et al*. (1990) found Tyr-Pro sequence to be *cis* 25 % of time, while *cis* Trp-Pro was absent. A high occurrence (19 %) of Tyr in *cis* bonds was also reported by MacArthur & Thornton (1991). From a larger database we find that the percentage of Tyr occurring in *cis* bonds has been reduced considerably (9.7 %), and Trp has become equally conspicuous (10.4 %). A Pro-Pro bond has the highest frequency (11.2 %) to be in the *cis* form. The residue X in X-Pro that causes the bond to be *cis* at least 6 % of the time belongs to one of the following four groups: (i) aromatic residues, (ii) small residues, Gly and Ala; (iii) polar residues Ser, Gln and Arg; and (iv) Pro provide

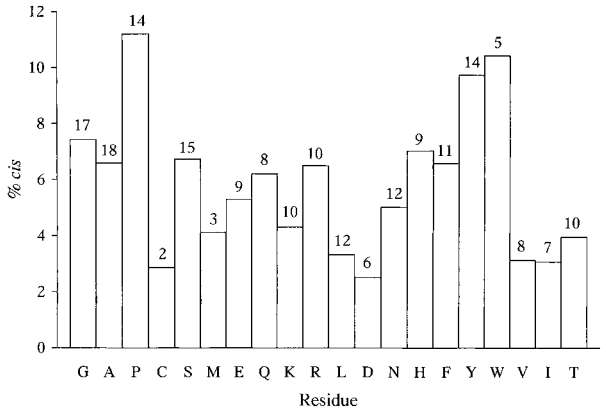



**Figure 2.** Histogram showing the percentage of occurrence of various residues in the *cis* conformation of the X-Pro peptide bond; the numbers of *cis* cases are given on top of each bar.

61 % of the data points. Branched aliphatic residues Val, Ile, Thr and Leu are less frequent. Recently, Reimer *et al*. (1998) have also calculated the amino acid frequency of *cis* prolyl bonds for every single amino acid preceding Pro. Some of their values are smaller than ours, possibly because of their inclusion of lower resolution (3.5 Å) data, where *cis* bonds are underestimated (Weiss *et al*., 1998).

The number of observations of the X-Xnp bond in the *cis* form is rather small to make any definite statement. Out of 29 cases (Table 1) Gly, Ser, Trp, Ala and Asp have higher occurrences at position (1) and Asn, Ala, Thr, Asp and Phe at (1′) (Table 2B).

## Residue preferences in the neighbourhood of *cis* peptide bonds

When considering the neighbours (±six residues) of prolyl residues and their physicochemical properties, Frömmel & Preissner (1990) found six different patterns which contained 75 % of known X-Pro cases. To see if the local sequence has any influence on the occurrence of a X-Pro, Pro-Pro or Xnp-Xnp bond in the *cis* conformation the percentage composition of residues at each position, from (3) to (3′) (Figure 1(b)), was calculated, and the preferred residues are given in Table 2. Being most abundant, the X-Pro cases were analyzed after grouping them into two main turn types (VIa and VIb), as well as the four subgroups (VIa-1, VIa-2, VIb-1 and VIb-2) of the above types. In addition to considering individual residues we also analyzed the occurrence of groups of residues, like small (Gly and Ala), aromatic (Phe, Tyr, Trp and His), β-branched (Val, Ile and Thr) and short polar (Ser, Asp and Asn).

There are interesting trends considering groups of residues (Table 2C). Taking X-Pro (VIa) as an example, aromatic residues have high occurrences at positions (1) and (2′), which decrease sharply on moving outward. On the contrary, the β-branched residues are less at positions (1) and (2′) (especially in the former position, which is also indicated in Figure 2), and increase along the outward locations (especially upstream). X-Pro (VIb) and Xnp-Xnp cases have very similar position-specific variations of these two groups of residues. In Pro-Pro cases, the aromatic residues are abundant at position (2′) and the branched residues at position (2). Short, polar residues (Ser, Asp and Asn) are likely to be a constituent of the *cis* Xnp-Xnp bond, and also be a part of X-Pro (type VIb) bond.

Small residues have relatively higher occurrences in all the positions of Xnp-Xnp, and also in position (2′) of X-Pro (Table 2C). Although taken together as small residues, Gly and Ala are not always found in similar numbers. For example, Gly is more abundant in the location (1) of Xnp-Xnp, whereas Ala predominates in locations (1′) and (2′) (Table 2B). Likewise in X-Pro (turn type VIa), Gly is prominent at position (2′) and is exclusively found in position (2), but does not occur at all in position (1). As to be discussed later, because of the conformations being distinct from other X-Pro cases, Gly-Pro *cis* peptides belong to different turn categories, VIb-3, VIc and VId (Table 3). However, if taken together, they have equal preferences of aromatic and β-branched residues at positions (3) and (2′).

Among the different turn types involving the X-Pro cases (Table 2A), VIa-2 type, in comparison to VIa-1, has a high proportion of small residues, notably Gly in position (2′). Relative to the above two types, VIb-1 has a greater presence of Pro

**Table 1.** Percentage occurrence in the *cis* conformation of X-Xnp sequences

| Range (%) | Sequence[a] |
|---|---|
| 0.0-0.2 | AA,GA |
| 0.2-0.4 | DA,GT,AD,LT,AT,SV,VN,EI,GF,GG(2),DD |
| 0.4-0.6 | QL,FS,RD,SY,DN,SF,SR,PN,NY |
| 0.6-0.8 | PY |
| 0.8-1.0 | WA,HT |
| 1.0-1.2 | HF,CA |
| 4.3 | WN(3) |

[a] Within a range, the sequences are in an ascending order of occurrence. The number of cases, if more than one, is given in parenthesis.

**Table 2.** Preference of amino acid residues around various categories of *cis* peptide units

| Categories | (3) | (2) | (1) | (1′) | (2′) | (3′) |
|---|---|---|---|---|---|---|
| A. *Xnp-Pro: different turn types*[a] | | | | | | |
| VIa-1 (39) | G(15), A,V(10) *[F,E,P]* | D(15),G(13) *[A,N,R]* | **A**(23),Y(13), E(10) *[G,T,N]* | P | G,F(15), T,D,N(10) *[I,K,P]* | S,I(13), D,N,A(10) *[L]* |
| VIa-2 (13) | T(23),V,I(15) | V(31),G(15) | L,S,W(15) | P | **G**(46),A(15) | S(23),Q(15) |
| VIb-1 (100) | I(10),V(9) | T,P(11),G(9) *[E]* | S(10) | P | **A**(18) | **P**(14) |
| VIb-2 (12) | I(25),A,D,S(17) | **G**(42) | N(25),T,H(17) | P | **G**(25),A,I(17) | Q,R(17) |
| B. *Different possible sequences forming* cis *peptide*[b] | | | | | | |
| X-Pro (VIa) (52) | G(13),V,T(12) *[P]* | G,D(13),V(12) *[A,N,R]* | A(19),L,Y(10) *[G,N]* | P | **G**(23),F(12),T, D(10) *[K,P]* | S(15),I(12), P(10) *[L]* |
| X-Pro (VIb) (116) | I(11),A,V(9) | G(12),T,P(9) *[E]* | N,S,T,Y(9) | P | A(17),V,Y(9) | P(13),T(9) |
| Xnp-Xnp (29) | **I**(21),A(14), V,R(10) *[D,K,P]* | V,I,L,N(14) *[S,E,K,P]* | G(17),W,S(14), A,D(10) *[I,T,K]* | A,N(17),T(14), D,F(10) *[E,K]* | **A**(24),G,L(14) *[V,F,P]* | **G**(21),Q(14), L,S(10) *[T,E]* |
| Pro-Pro (14) | K(21) | I,T,K,D(14) | P | P | **F**(29),T,Q(14) | L,T(21),P(14) |
| Gly-Pro (16) | L(25),V,Y(19) | Q(19),G,S(12) | G | P | V,F(19) | G(19) |
| C. *Preference of groups of residues*[c] | | | | | | |
| X-Pro (VIa) (52) | Sm(21),Ar(8), **Bb**(32),Sp(18) | Sm(13),Ar(16), **Bb**(24),Sp(17) | Sm(19),**Ar**(26), Bb(10),Sp(10) | P | **Sm**(29),Ar(24), Bb(16),Sp(20) | Sm(12),Ar(10), Bb(20),**Sp**(31) |
| X-Pro (VIb) (116) | Sm(15),Ar(11), **Bb**(27),Sp(17) | Sm(18),Ar(15), Bb(18),Sp(13) | Sm(9),**Ar**(23), Bb(16),Sp(21) | P | **Sm**(23),Ar(13), Bb(20),Sp(11) | Sm(9),Ar(12), Bb(20),Sp(20) |
| Xnp-Xnp (29) | Sm(21),Ar(7), **Bb**(38),Sp(10) | Sm(20),Ar(13), **Bb**(31),Sp(17) | **Sm**(27),Ar(24), Bb(3),**Sp**(27) | Sm(24),Ar(17), Bb(20),**Sp**(30) | **Sm**(38),Ar(7), Bb(10),Sp(17) | **Sm**(28),Ar(6), Bb(10),Sp(20) |
| Pro-Pro (14) | Sm(14),Ar(7), Bb(14),Sp(14) | Sm(7),Ar(7), **Bb**(35),Sp(14) | P | P | Sm(0),**Ar**(43), Bb(21),Sp(0) | Sm(7),Ar(7), Bb(21),Sp(14) |
| Gly-Pro (16) | Sm(0),**Ar**(25), **Bb**(25),Sp(6) | Sm(18),Ar(12), Bb(18),Sp(18) | G | P | Sm(12),**Ar**(31), **Bb**(31),Sp(12) | Sm(19),Ar(12), Bb(18),Sp(12) |
| D. *Most (and least) likely residues*[d] | | | | | | |
| X-Pro (VIa) | Bb,Sm *[P]* | Bb,G *[A,N,R]* | Ar,A *[G,N]* | P | Sm,Ar *[K,P]* | Sp,Bb *[L]* |
| X-Pro (VIb) | Bb | Bb,Sm *[E]* | Ar,Sp | P | Sm,Bb | Sp,Bb |
| Xnp-Xnp[e] | Bb,Sm | Bb,Sm | Sm,Sp,W | Sp,Sm | Sm | Sm |
| Pro-Pro | - | Bb | P | P | Ar | - |
| Gly-Pro | Ar,Bb | - | G | P | Ar,Bb | - |

At each position (Figure 1(b)), the percentage residue composition is calculated and the residues having high values are entered with the percentage composition given in parentheses (when multiple residues have the same value the number is given after the last entry). If the first entry has a distinctly higher value than the next, it is given in bold and underlined. Residues whose average occurrence in protein structures is greater than 4% (Pal & Chakrabarti, 1999a), but are not found at all in a given position, are given in italics within square parentheses. The number of cases in each category is given in column 1.

[a] Given in Table 3 (sparsely populated types are excluded).

[b] X-Pro sequences are broken into two classical VIa and VIb turns.

[c] Residues are grouped as: Sm, small (G, A); Ar, aromatic (F, Y, W, H); Bb, β-branched (V, I, T); and Sp, short polar (S, D, N).

[d] Indicated by either one-letter amino acid code, or a two-letter group designation.[c]

[e] Less likely to have Pro and Lys all throughout.

around the *cis* peptide. Because of steric factors, VIb-2 type needs to have a small residue (Gly in particular) at either position (2) or (2′). Based on the above, the notable presence (or absence) of various residues around the *cis* peptide moieties are summarized in Table 2D. Interestingly, there are only two examples of Pro-X *cis* peptides: 2CTC (PDB file) with sequence, Leu-Tyr-Pro-Tyr-Gly-Tyr and 1MKA, Pro-Ala-Pro-Asn-Met-Leu.

## Possible role of neighbouring residues in *cis-trans* isomerization

Data in Table 2C show a contrast in the relative presence of aromatic and β-branched residues around the *cis* peptide units. For X-Pro cases, while aromatic residues have a higher presence at position (1), their numbers decline as one moves out along the sequence from the *cis* bond. On the other hand, the branched residues, show the opposite trend and have the maximum presence at position (3) (even for Xnp-Xnp cases). This observation is suggestive of the steric requirement for the isomerization of a *trans* peptide bond into *cis*. The residues with two bulky alkyl groups at $C^{\beta}$ (close to main-chain) if located at position (1) hinders the isomerization process. Support for the steric clash having an inhibitory role on the isomerization process also comes from nature of the residue preceding Pro-Pro *cis* peptides. In the sequence X-Pro-Pro, one may ask what determines the second bond to be in the *cis* peptide conformation rather than the first. It appears that a large percentage of these cases have a β-branched residues for X (and in addition, aromatics at position (2′)). Even Xnp-Xnp *cis* peptides have a few such residues at either position (the relatively higher number at position (1′) is due to a large contribution from Thr which, due to its polar features, acts in a different way, as discussed below). The β-branched residues, however, may have a beneficial role when located at position (2) or (3). Because of the larger steric clash

**Table 3.** Types of turns mediated by *cis* peptide bonds and their geometries

| Turn type[a] | Conf.[b] | No. | $\phi_1$ | $\psi_1$ | $\phi_{1'}$ | $\psi_{1'}$ | Dist (Å) (2)-(2′) | Secondary structure[c] |
|---|---|---|---|---|---|---|---|---|
| A. *Xnp-P* | | | | | | | | |
| VIa-1 | BA | 39 | −74(24) | 141(9) | −93(9) | 12(16) | 5.9(6) | TT(39) |
| VIa-2 | BA | 13 | −131(24) | 145(16) | −79(9) | −16(24) | 6(1) | SS(5),CS(3),ET(2),BS,ES,II |
| VIb-1 | BB | 100 | −117(26) | 138(16) | −77(10) | 158(17) | 6.3(8) | SS(50),CS(16),SC(10),ES(8), CC(8),BS(4),EC(2),BC,EE |
| VIb-2 | BB | 12 | −134(12) | 98(23) | −78(12) | 165(9) | 4.5(7) | TT(12) |
| VIb-3 | BB | 4 | −100(20) | 183(8) | −72(10) | 154(2) | 7.7(2) | SC(2),SS,EC |
| VIc | RA | 5 | 104(38) | 188(8) | −83(9) | −16(7) | 8.4(4) | CS(4),SS |
| VId | RB | 7 | 102(20) | 186(25) | −69(8) | 171(23) | 8.3(3) | CC(3),EE(3),BC |
| B. *P-P* | | | | | | | | |
| VIa-1 | BA | 7 | −54(5) | 147(5) | −81(5) | 9(10) | 5.6(3) | TT(7) |
| VIb-1 | BB | 6 | −69(6) | 160(8) | −77(11) | 149(14) | 7.4(7) | SS(3),CC,SS,EE |
| VIb-2 | BB | 1 | −84 | 149 | −96 | 115 | 6.3 | TT |
| C. *Xnp-Xnp* | | | | | | | | |
| VIa-1 | BA | 5 | −89(21) | 134(30) | −111(17) | 14(36) | 6.4(9) | TT(5) |
| VIa-2 | BA | 3 | −113(41) | 149(9) | −106(7) | −15(17) | 7(1) | CC,ET,EE |
| VIb-1 | BB | 15 | −108(29) | 121(23) | −134(21) | 168(15) | 8(1) | EC(5),EE(4),SC(3),ES(2),CC |
| VIb-2 | BB | 2 | −123(6) | 121(57) | −102(23) | 152(26) | 6(1) | TT(2) |
| VIb-3 | BB | 1 | −155 | 176 | −102 | 129 | 8.6 | EE |
| VId | RB | 3 | 131(30) | 174(11) | −91(2) | 202(13) | 9.1(6) | CC(2),EC |

Data for eight cases are not included in the Table: two Pro-Xnp cases (with conformations BA and BB); one C-terminal *cis* peptide; one Gly-Pro sequence (LB); and four sterically strained non-Gly-Pro sequence (LB(2), AB(1), AA(1)). Representative diagrams are given in Figure 7.

[a] VIb-3, VIc and VId turns have Gly at position (1). The hydrogen bond (Figure 6) is usually between residues (2) and (2′) (providing CO and NH groups, respectively) in VIa-1, (3) and (3′) in VIb-2, and (1) and (1′) (providing CH and CO, respectively) in VIb-3 and VId.

[b] Conformation based on the location of the two residues in the Ramachandran plot (see Materials and Methods and Figure 5).

[c] Of positions (1) and (1′) as specified by the program DSSP (Kabsch & Sander, 1983): H, α-helix; I, π-helix; E, strand; T, hydrogen bonded turn; S, non-hydrogen bonded turn; C, non-regular structure. The number of observations, if more than one, is given in parentheses.

between the main and side-chain atoms, the $\phi,\psi$ angles of these residues lie in a limited range (Chakrabarti & Pal, 1998), and thus they can act as a tether or a wrench to hold the chain in position while an adjacent bond is being isomerized.

A corollary of the above hypothesis is that the small residues offering the minimum steric resistance should facilitate the *cis* form. Indeed, a large number of Gly and Ala residues are found in positions (1) of X-Pro, (1) and (1′) of Xnp-Xnp, and (2′) of both. In the case of Xnp-Xnp there may be another factor operating during the *trans* to *cis* isomerization. Most of these have polar residues, Ser, Thr, Asp and Asn at position (1′) and their side-chains are usually within the hydrogen bonding distance of the main-chain NH group at the same position (although the angles, in the range 60-120°, do not fulfill the usual hydrogen bond criterion). Even though the geometry may not be optimum, it is quite plausible that during isomerization such interaction may satisfy the hydrogen bonding potential of the NH group, and thus lower the activation energy of the process. Participation of a nearby residue facilitating the *cis-trans* isomerization is known (Reimer *et al*., 1997). Once formed, the *cis* peptides may be stabilized by interactions (discussed later) involving aromatic residues which are found in large numbers at positions (1) of X-Pro and (2′) of Pro-Pro and X-Pro (turn type VIa).

## Correlation between main-chain and side-chain conformations

Recently, we have shown how the side-chain torsion angle $\chi_1$ is correlated with the backbone angles $\phi$ and $\psi$ of residues held by *trans* peptide linkage, and how the result can be used to classify the amino acid residues (Chakrabarti & Pal, 1998). The paucity of data for *cis* peptides does not allow one to study the interrelationships of angles for individual residues. However, some general trends can be deciphered (Figure 3). For example, in Xnp-Pro cases, the means of the distributions of the $\psi$ values of Xnp get changed ($130° \rightarrow 135° \rightarrow 148°$) as $\chi_1$ goes from $-180°$ to $-60°$ to $60°$ (conformational states *t*, $g^+$ and $g^-$, respectively, which occur in the ratio ≈3:5:1; Figure 3(a)). For a Pro residue in this position, though any value of $\chi_1$ from $-30$ to $+30°$ is possible, negative values predominate (in the ratio ≈2:1). As noted earlier (MacArthur & Thornton, 1991), $\psi$ is above $60°$ for a residue in this position. Considering $\phi$ (Figure 3(b)), the points are rare below $-140°$ in the $g^+$ state, whereas in the other two states, although the spread is from ca $-60$ to $-170°$, most of the points are closer to the latter value.

Pro in *cis* X-Pro has a noteworthy dependence of $\chi_1$ on $\phi$ and $\psi$ (Figure 3(c) and (d)). Residues predominantly have a positive $\chi_1$ (positive:negative ≈6:1). Notably, however, when $\psi$ is less than $60°$,

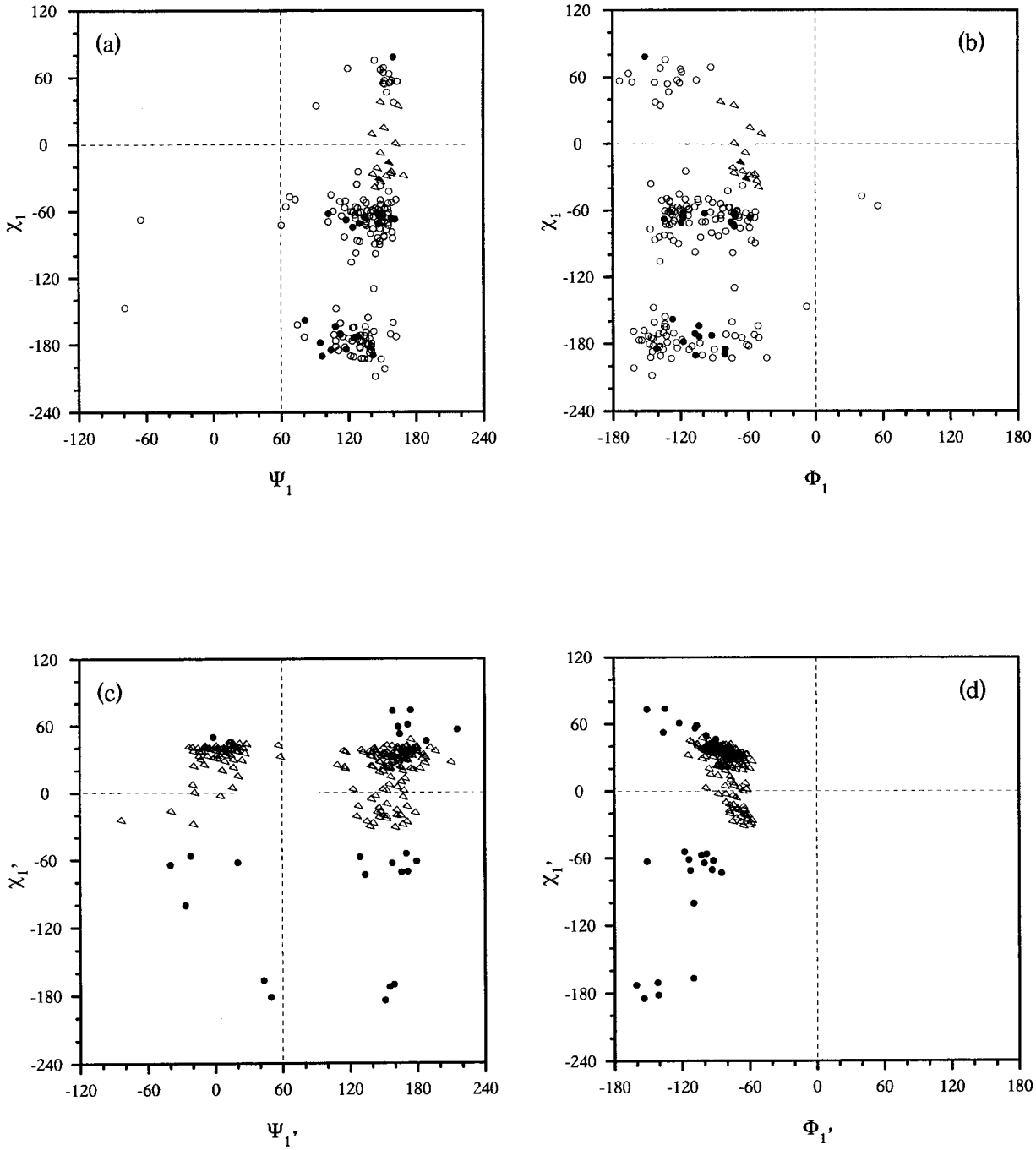



**Figure 3.** Joint distributions of $\chi_1$ with $\phi$ and $\psi$ for residues at positions (1) and (1′). Symbols used: △, Pro, ○, non-Pro, and these are open for X-Pro and filled for X-Xnp cases.

a positive value of $\chi_1$ is the norm, and only when $\psi$ is ~120° or more a few points are also observed in the negative range of $\chi_1$. Starting at −60° the $\phi$ values go up to −80° when $\chi_1$ is negative, whereas for positive $\chi_1$ it can extend upto −110°.

Although the residue X in both X-Pro and X-Xnp peptide units has similar conformational features, those for Pro and Xnp are considerably different. The most conspicuous but obvious difference is the $\chi_1$ angles, which are restricted in the range −30 to +40° for Pro, whereas for the non-Pro residues there are three conformational states. Additionally, however, compared to the former, the $\phi$ values of the latter are shifted towards more negative region (Figure 3(d)). Without the constraint on $\phi$ imposed by the pyrrolidine ring, non-Pro residues, by taking a more extended value of $\phi$, reduces the steric clash between $C^\alpha$ of position (1) and the carbonyl group of position (1′).

A striking feature of the $\chi_{1'}$ $\phi_{1'}$ plot (Figure 3(d)) is the near linear relationship between the two parameters (irrespective of whether it is a Pro or a

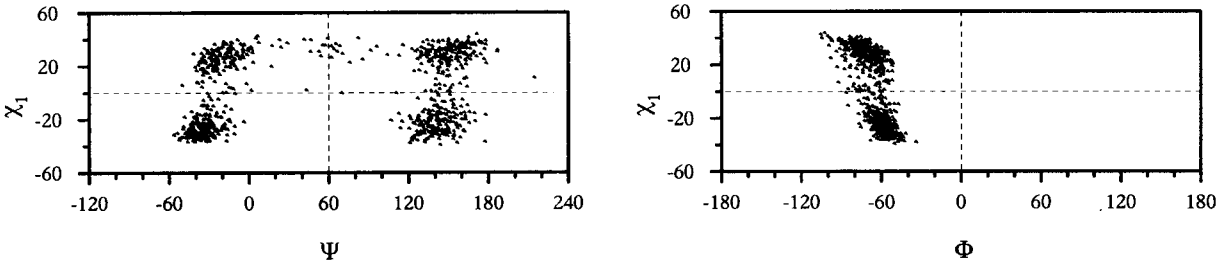



**Figure 4.** $\chi_1$,$\psi$ and $\chi_1$,$\phi$ plots for *trans* proline residues.

non-Pro residue) when $\chi_{1'}$ has a positive value. In fact, if one excludes the Pro rings with small puckering ($\chi_{1'} < 10^\circ$) then the correlation coefficient between the two parameters is $-0.65$ (equation: $\chi_{1'} = -0.42\ \phi_{1'} + 1.59$).

### Pyrrolidine ring puckering in *cis* peptides

The pyrrolidine ring of the Pro residue invariably occurs in puckered conformations, which are essentially of two types, UP (or A or $C^\gamma$-*exo*) and DOWN (or B or $C^\gamma$-*endo*) depending on the placement of the $C^\gamma$ atom and the CO group of Pro on the opposite or the same the side of the plane defined by the remaining ring atoms (N, $C^\alpha$, $C^\beta$ and $C^\delta$) (Ramachandran *et al.*, 1970; Ashida & Kakudo, 1974; Milner-White *et al.*, 1992; Chakrabarti & Chakrabarti, 1998). The UP conformation is characterized by negative $\chi_1$ and $\chi_3$ and positive $\chi_2$ and $\chi_4$ values, and the opposite holds good for the DOWN conformation. Both are isoenergetic when Pro is involved in a *trans* X-Pro bond. For the *cis* isomer however, 89% of the Pro residues in proteins exhibit DOWN pucker with average values for the four side-chain torsion angles being 30, $-36$, 24 and $-8^\circ$ (Milner-White *et al.*, 1992). To find out the reason for such an occurrence we have carried out a conformational analysis (in terms of $\phi$, $\psi$ and $\chi_1$) of Pro residues involved in *cis* (Figure 3(c) and (d)) and *trans* (Figure 4) peptide bonds. MacArthur & Thornton (1991) had observed that compared to *trans*, *cis* proline residues show a displacement to a more negative $\phi$ values in both the A and B regions, and a more positive $\psi$ value in the A region so as to reduce the steric clash between the $C^\alpha$ group of the preceding residue and Pro carbonyl group. While validating the earlier observations our results indicate the striking dependence of the main-chain torsion angles on $\chi_1$. *Cis* proline residues in the A region ($\psi = -60$ to $0^\circ$) with negative value of $\chi_1$ are almost non-existent; though rather uncommon, residues with $\psi$ in the range of 10 to $120^\circ$ are only found for *trans* residues if $\chi_1$ is positive, whereas for *cis* such points are absent. More remarkable, however, is the interdependence of $\phi$ and $\chi_1$ as both the torsion angles are around bonds in the pyrrolidine ring, and can thus contribute to the observed puckering of the ring. If residues with positive and negative values of $\chi_1$ are considered separately, within each group, as $\phi$ is reduced $\chi_1$ tends to increase. As already mentioned, the unfavourable main-chain contacts around the *cis* bond are reduced by making $\phi$ more negative as compared to the *trans* bond. As a result, when $\chi_1$ is negative, while $\phi$ varies from $-75$ to $-40^\circ$ for *trans*, the range is $-80$ to $-60^\circ$ for the *cis* proline residues. A shortened range of $\phi$ means only a few *cis* Pro residues can have negative $\chi_1$ angles. On the other hand, a more negative and wider range of $\phi$ ($-110$ to $-60^\circ$) is available when $\chi_1$ is positive, which is thus the predominant state of the side-chain conformation (DOWN puckering) observed for Pro residues involved in *cis* peptide bonds. Thus local steric interaction (resulting in more negative $\phi$, which in turn causes $\chi_1$ to be positive) explain the DOWN puckering of the *cis* Pro residues, whereas in the case of the UP puckering observed in Pro residues in the middle of $\alpha$-helices it was a specific C-H$\cdots$O interaction involving the $C^\delta$H groups that was responsible (Chakrabarti & Chakrabarti, 1998).

### Conformations delineating *cis* peptide mediated turns

The classic VIa and VIb turns formed by *cis* proline residues (Lewis *et al.*, 1973) can be described by the two residues (1) and (1′), residing in regions B and A, respectively, of the Ramachandran plot (see Materials and Methods), and both occupying the region B, respectively (MacArthur & Thornton, 1991). Consequently, we have constructed $\phi$,$\psi$ plots for pairs of residues forming *cis* peptides, X-Pro and X-Xnp, in three groups corresponding to type VIa and VIb turns, and those falling outside (Figure 5).

In Figure 5(a), BA conformation (type VIa) of the two residues are shown. When $\phi_1$ is greater than $\sim -90^\circ$ there is a hydrogen bond between residues (2) and (2′) (sometimes between (2) and (3′)) (Figure 6). As $\phi$ is decreased (below $-90^\circ$), the CO group of position (2) moves away from the NH group of (2′) and the hydrogen bond between them is lost (the same thing can also be achieved, though the number of cases is not many, by decreasing $\psi_{1'}$ below $-30^\circ$ so as to turn away the NH group of (2′)). Consequently, we have subdivided type VIa turn type into two groups, VIa-1 and VIa-2, the former with hydrogen bonding and

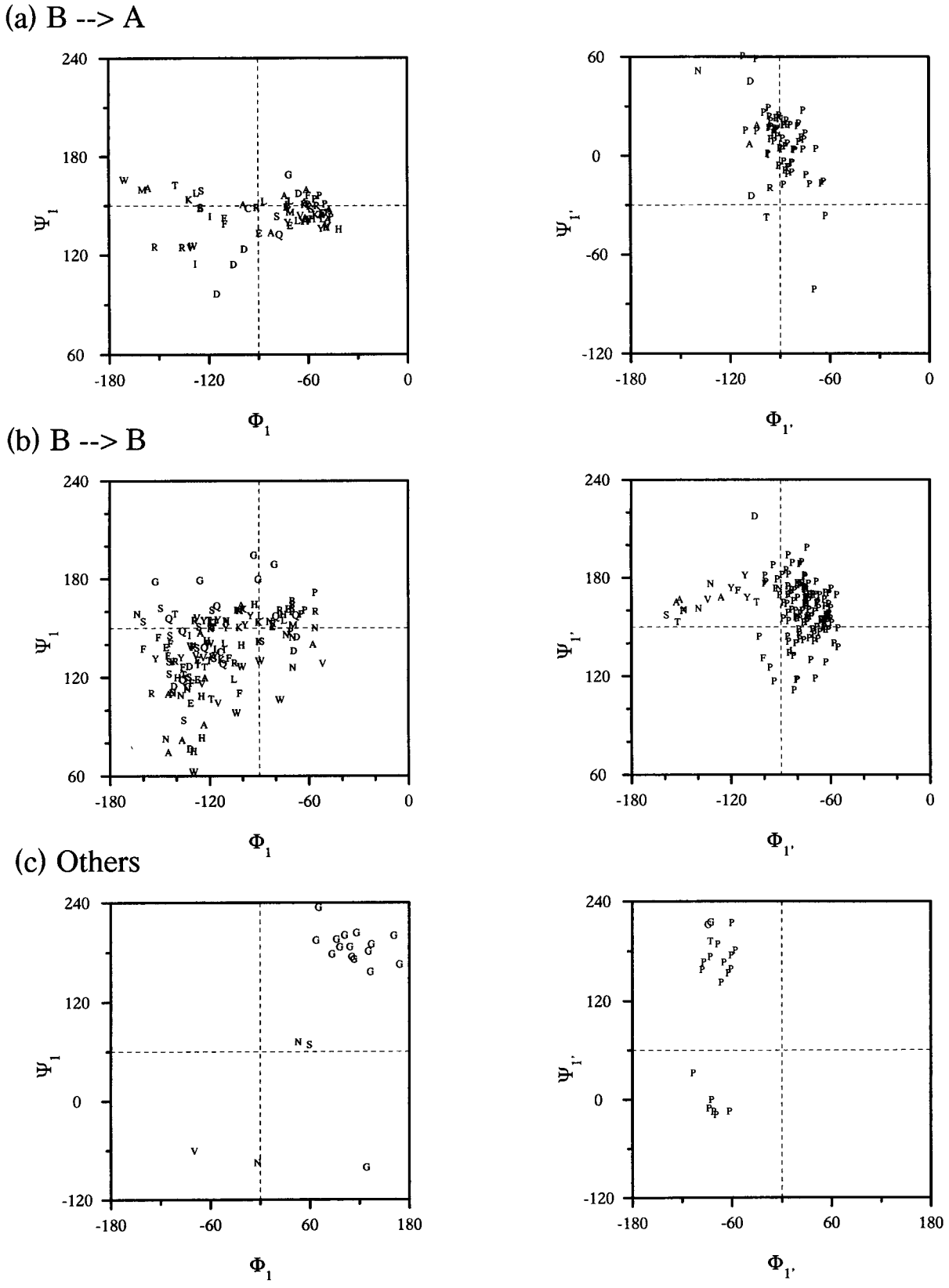



**Figure 5.** $\phi,\psi$ plots for residue pairs at positions (1) and (1′) (each point is indicated by the one-letter amino acid code of the corresponding residue). (a) The first residue is in the region B and the second in A; (b) both are in the region B; and (c) the rest (except five cases, the first residue is in R, whereas the second is either in A or B region). Only the specified regions of the Ramachandran plot are shown in (a) and (b).

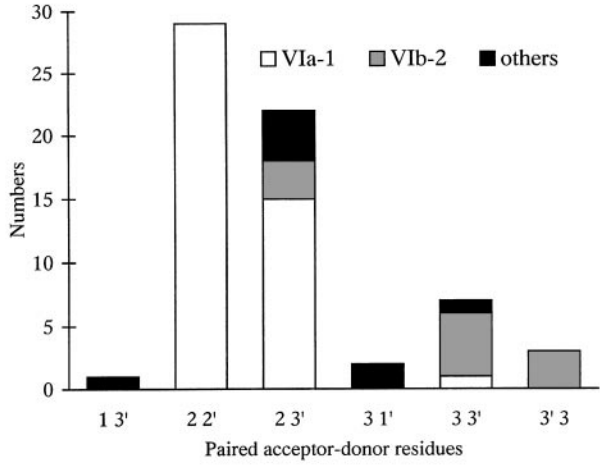



**Figure 6.** Histogram showing how the residues around the *cis* peptide are hydrogen bonded through the main-chain atoms; the bonded pairs (the first providing the CO group, and the second the NH group) are specified along the horizontal axis. (In cases where more than one pair of hydrogen bonding site is available, the one with the shortest distance is retained.)

the latter without (Figure 7), and their average $\phi,\psi$ values are listed in Table 3.

Type VIb turn with central residues in the extended conformation (B) (Figure 5(b)) are without any hydrogen bonds. However, when $\psi_1$ is below 100°, depending on the conformation of the flanking residues, there is the possibility of hydrogen bonding between residues (3) and (3′) or (2) and (3′) (Figure 6). Such cases, also identified by DSSP as hydrogen-bonded turns, are classified as type VIb-2 turn to distinguish them from the non-hydrogen-bonded, but predominant, type VIb-1 turn (Figure 7 and Table 3). Gly at position (1) stands out from the other residues in having $\psi$ close to 180° (Figure 5(b)), which could be aided by the formation of a C-H···O hydrogen bond between the $C^{\alpha}H$ of Gly and residue (1′) carbonyl group (Figure 7). As a result, these cases constitute a separate class of turn, type VIb-3.

Considering the cases which are not included in the broad categories of type VIa and VIb turns, the residue (1) is overwhelmingly Gly and belongs to the region R of the Ramachandran plot, while position (1′) occurs in the region A or B (Figure 5(c)). To encompass these cases two new categories of turn type can be introduced, VIc and VId. Like the type VIb-3 turn, VId also has a favourable geometry for C-H···O interaction involving the $C^{\alpha}H$ of Gly.

As has been discussed earlier and also seen in Figure 5(b), the non-Pro residues at position (1′) have a more negative $\phi$ value than proline residues. Hence, the average values (Table 3) are calculated separately for Xnp-Xnp cases, as also for Pro-Pro cases (which have a more restricted conformational parameters). Because of the more extended nature of $\phi$, the turn opens up in Xnp-Xnp cases (in Figure 7, compare (c) and (h), both having the same turn type, but different sequences), which thus have a longer (2)-(2′) distance (between $C^{\alpha}$ atoms) than what is observed in the corresponding X-Pro motif. The (2)-(2′) distance is usually restricted below 7 Å for a β-turn (Wilmot & Thornton, 1988). However, type VIc, VId and VIb-3 turns have longer distances (and may be termed as pseudo turns). This is because although the *cis* peptide causes a sharp turn between residues (1) and (1′), preceding (1) there is also another turn caused by a positive $\phi_1$ (in VIc and VId) and a very extended $\psi_1$, one nullifying the effect of the other and nearly aligning the chain directions beyond positions (2) and (2′). The reversal of the chain direction in a turn can be shown by using a virtual torsion angle defined in Figure 8. As expected, the peak for the distribution of X-Pro *cis* peptides occurs at a small angle (30°), but values extending up to 180° are found. The relatively less restricted nature of the turn in X-Xnp case is indicated by a shift of the peak to a higher value (60°).

## Position relative to the protein centre

The global position of the *cis* bond is of vital interest due to various reasons. The speed of protein folding is believed to be controlled kinetically by the rate of *cis-trans* isomerization. Proteases have been isolated which are selective for the Tyr-Pro bond only after it has been isomerized to the *trans* conformation (Vance *et al.*, 1997). Similarly, the membrane-binding conformation of bovine prothrombin is generated following the *trans* → *cis* isomertization of an X-Pro bond (Evans & Nelsestuen, 1996). For these proteins the proper exposure of the bond concerned should have a direct bearing to the function.

To address the question of the location of the *cis* peptides in the three-dimensional structure we have carried out two types of calculations. First is the radial distribution of the *cis* units relative to the centre of mass of the polypeptide chain. Such a depiction is provided in Figure 9(a) where the position of the *cis* peptide in concentric shells (obtained by dividing the distance from the centre of mass to the outermost atom in the structure into ten equal parts) is shown. The X-Pro peak occurs at shell number 7, which may indicate a position in shallow crevices close to the surface of the protein. The distribution has a broad shoulder at 4, which is suggestive of a group of *cis* peptides, possibly with functional role, that are found deeper inside the structure. X-Xnp *cis* peptides, on the other hand, are more buried (peak at shell number 3) with no case observed at the two outermost shells. The above behaviour is reproduced in our second type of calculation involving the solvent accessibility of the *cis* peptide units. The average accessibility for the two residues making up the X-Pro bond

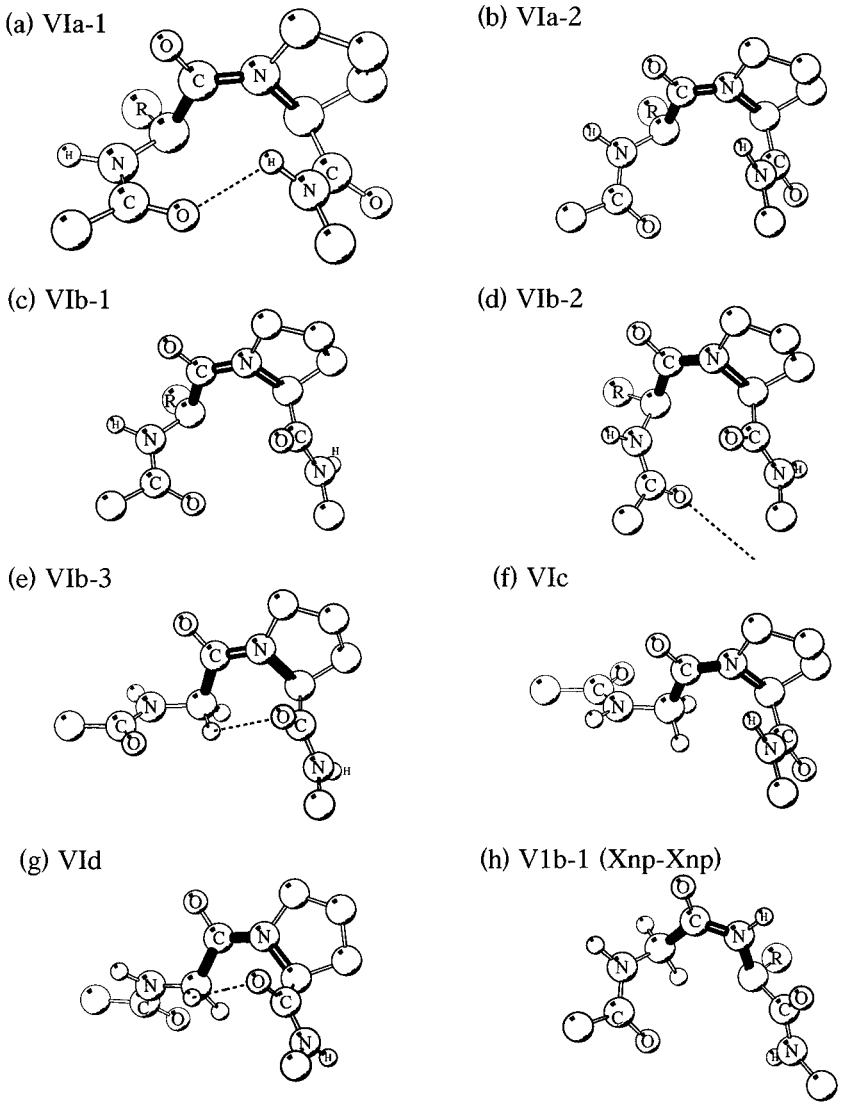



**Figure 7.** Molecular representations of the different classes of β-turns around (a)-(g) *cis* Xnp-Pro bond, and (h) one case of Xnp-Xnp bond. The *cis* peptide is shown in thick lines; hydrogen bond, if present, is shown in broken lines.

has a bimodal distribution with the main and the minor peaks appearing at 50 % and 10 % accessibility (Figure 9(b)). For X-Xnp the only peak at 10 % average accessibility indicates a more buried location of such groups in protein structure. Being more buried, an X-Xnp *cis* peptide bond is likely to be formed early in the folding process, as otherwise the isomerization of a bond not on the surface would involve a greater rearrangement of the structure.

To assess the local geometry from the accessibility we have found out how the average accessibility varies as one moves from the central pair to the pairs of residues on either side. A ++ sign (the average accessibility of the central pair of residues is greater than those of the pairs on either side), shown by the maximum number of X-Pro *cis* peptides (Figure 9(c)), indicates a convex nature of the surface in general, with the two residues forming the *cis* peptide being near the bulge. The diagram also indicates that X-Xnp *cis* peptides have different characteristics.

## Secondary structural features

As can be expected from the conformation of the two predominant types of turn, VIa-1 and VIb-1 (Table 3), the two residues (X and Pro) making up the *cis* bond mostly have turn (hydrogen bonded, T or non-hydrogen bonded, S) as the secondary structure (Table 4). Residues on either side are not found in any regular secondary structure in 40 % of the cases (60 % for Pro-Pro *cis* units). There are, however, clear distinctions between the X-Pro and Xnp-Xnp cases. While the neighbours of the former are mostly without any secondary structure (only 20-30 % are located in strands) the latter can be accommodated completely in (or preceded by) strands, or be at the N-terminal ends of helices. This suggests that the occurrence of Xnp-Xnp *cis* peptides may be dictated to a greater extent by the

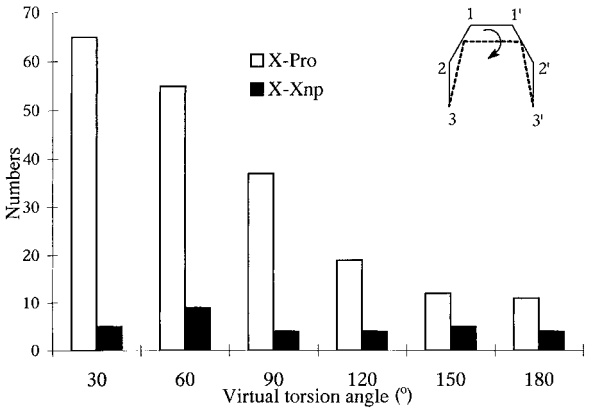



**Figure 8.** Frequency of occurrence of the virtual torsion angle (defined using the $C^{\alpha}$ positions of (3), (2)-(1) mid-point, (1′)-(2′) mid-point and (3′)).

secondary structure around them, whereas X-Pro *cis* peptides are controlled more by surrounding residues.

There are four structures (PDB files: 2EBN, 1NAR, 1LUC and 1XYZ) in which all the six residues surrounding the Xnp-Xnp *cis* peptide are in a strand (in another, 1CNV, only the last residue is not) (as defined by DSSP, which does not always match with the information provided in the PDB files). In these the strand containing the *cis* peptide is part of an eight-stranded β-barrel (Figure 10(a)). Unlike a typical β-strand, where the $C^{\beta}$ atom projects out (up and down, alternatively) perpendicular to the β-sheet, the side-chains of the two residues are facing to the same side. This also creates a wider groove, which interestingly enough, fits in nicely against the convex surface formed by a turn of an adjacent helix, thus showing how the formation of a *cis* peptide can lead to surface complementarity between secondary structural elements.

An example of a *cis* peptide leading to the start of a helix is present in 1NAR (Figure 10(b)). The six adjacent residues have the secondary structure, EEECHH (in the DSSP notation), which is also found in 1CNV; in both, the *cis* peptide is formed between residues Trp and Asn. The latter residue has a dual role. At the helix N-cap position its side-chain carbonyl group can form hydrogen bond within the helix (Richardson & Richardson, 1988). Additionally, the amino group interacts with the π electrons at the N atom of the preceding Trp side-chain. Another example of an Asn residue at the N-cap position of a helix (residues 99-115) constituting a *cis* peptide (Asp97-Asn98) is found in 1XJO.

An interesting example of X-Pro *cis* peptide is found in 1VID, where the sequence Val(171)-Ile-Val-Pro-Gly forms a π-helix that leads to an

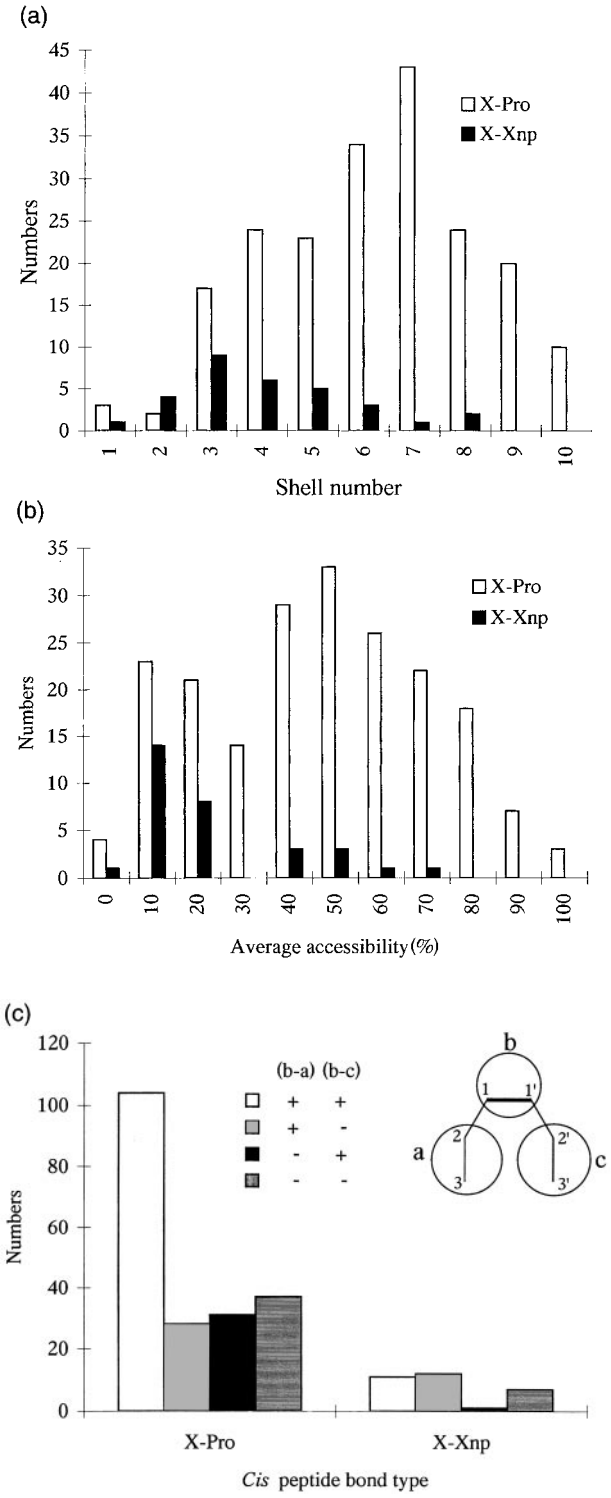



**Figure 9.** (a) Radial distribution of *cis* peptides in protein structures. (Starting at the center of mass, the span to the outermost atom in the structure was divided into ten equal parts, and the shell containing the *cis* peptide was calculated.) (b) Distribution of the average accessibility of residues (1) and (1′). (c) Histogram showing the variation of accessibility along the polypeptide chain around the *cis* peptide. (Assuming a, b and c to be the average accessibilities of residues (3) and (2), (1) and (1′), and (2′) and (3′), respectively, values of (b − a) and (b − c) were calculated, and depending on their sign one of the four combinations was assigned).

**Table 4.** Secondary structural features at positions around *cis* peptide bonds

| Secondary structure | (3) | (2) | (1) | (1′) | (2′) | (3′) |
|---|---|---|---|---|---|---|
| A. *X-Pro (200 cases)* | | | | | | |
| H | 10 | 5 | 0 | 0 | 10 | 13 |
| E | 22 | 23 | 13 | 3 | 19 | 28 |
| S | 16 | 22 | 37 | 50 | 16 | 13 |
| T | 15 | 14 | 31 | 32 | 16 | 10 |
| C | 37 | 37 | 19 | 16 | 41 | 37 |
| B. *Xnp-Xnp (29 cases)* | | | | | | |
| H | 7 | 0 | 0 | 0 | 24 | 31 |
| E | 72 | 65 | 52 | 21 | 28 | 24 |
| S | 3 | 3 | 10 | 7 | 14 | 3 |
| T | 14 | 17 | 24 | 28 | 24 | 21 |
| C | 3 | 14 | 14 | 45 | 10 | 21 |

The number at each position corresponds to the percentage of occurrence of different secondary structural elements, as defined by the program DSSP (Kabsch & Sander, 1983) except that H (helix) includes all residues marked H, G, I and P, E (strand) stands for both E and B, and C represents residues with no regular structure. The *cis* peptides are grouped into two classes.

α-helix. There are four structures (2ER7 and 1MPP with sequence, Ile(20)-Gly-Thr-Pro-Ala/Gly-Gln; 7RSA, Glu(111)-Gly-Asn-Pro-Tyr-Val; and 3TGL, Asp(226)-Asn-Ser-Pro-Glu-Thr) with $\pi_\beta$-turn (hydrogen bond between the CO group of position (3) and the NH group of residue (3′), both of which being part of an antiparallel β-sheet). The DSSP notation of secondary structure for the first three is ETTTTE and for the last, EETTEE.

## Hydrogen bonding across the *cis* peptide linkage

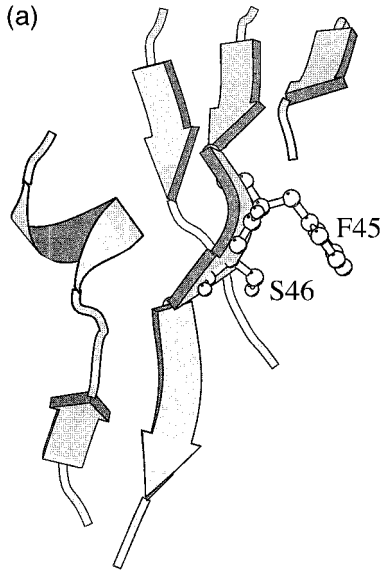
We considered if hydrogen bonding (Figure 6) has a significant role in stabilizing the six-residue *cis* peptide loop, and found that in only 28% of cases is there a hydrogen bond connecting the main-chain atoms across the *cis* bond (the value does not increase much (35%) even if the window size is doubled). Also, 43% of the positions in the loop are comprised of residues capable of forming hydrogen bond through the side-chain, of which only 10% are actually engaged across the *cis* bond within the loop. Of all the *cis* peptide-mediated turn types, only VIa-1 and VIb-2 have hydrogen bonds connecting the two halves. However, in Figure 6, a few other *cis* peptides are also shown to have hydrogen bonds. This is because in the classification of turns we used hydrogen bonds as defined by the program DSSP, whereas in Figure 6 it was based on geometric criteria and in a few borderline cases the two definitions do not match.

Unlike X-Pro, the Xnp-Xnp *cis* peptide has a free NH group available for hydrogen bonding which in about 50% cases is with a protein atom (Table 5). Interestingly, two donor and acceptor sites sequentially close can simultaneously satisfy the hydrogen bonding potential of the *cis* peptide unit; such cyclic motifs are found in 25% cases. The protein environment around *cis* peptide can also create a pocket for binding an anion (Figure 11).

## C-H···π and C-H···O interactions

Aromatic residues preceding Pro have a higher chance of making an Ar-Pro *cis* bond; however, there is no specific explanation for this in literature (Grathwohl & Wüthrich, 1981). We propose that the C-H···π interaction (Nishio *et al*., 1998; Chakrabarti & Samanta, 1995; Samanta *et al*., 1998) may have a role. This interaction, like the C-H···O interaction (Desiraju & Steiner, 1999; Derewenda *et al*., 1995; Chakrabarti & Chakrabarti, 1998) is facile if the CH group is made more acidic by an adjacent electron-withdrawing nitrogen atom, as in the $C^\alpha$ and $C^\delta$ positions of Pro ring, both of which, as discussed below, can be involved in conferring stability to the *cis* peptide units.

There are 39 cases of *cis* peptide involving Ar-Pro bond, 26 of which are of classical VIb turn type and 13 of type VIa. In 16 of the former and ten of the latter, the $C^\alpha$ atom of Pro has a close contact (3.6 (±1) Å) with the $C^\gamma$ atom of the aromatic residue and interacting with its face (Figure 12(a)). The Figure shows that there is not much overlap between the Pro and Trp rings; it appears that it is the specific orientation of the $C^\alpha H$ proton, rather than the stacking, which stabilizes the *cis* isomer. An aromatic residue following Pro can also have face-specific interactions. A total of 21 cases of turn type VIb and 12 of VIa have an aromatic residue at position (2′), whose side-chain usually comes in close contact with the main-chain atoms of residues (2) and (1) in the former (Figure 12(b)). In all but one case of the latter, the aromatic residue nearly stacks against the Pro ring with the $C^\delta$ atom of Pro showing the shortest contact distance (average, 4.1 (±3) Å) (Figure 12(c)). There are six examples of two aromatic residues flanking the *cis* Pro, of which three are of type VIa in which the CH groups of Pro interact with the π-face of the aromatic residues on either side (Figure 12 (c)). The stability conferred by two C-H···π interactions to the type VIa turn is exemplified by the occurrence of a high population of the *cis* isomeric form in sol-



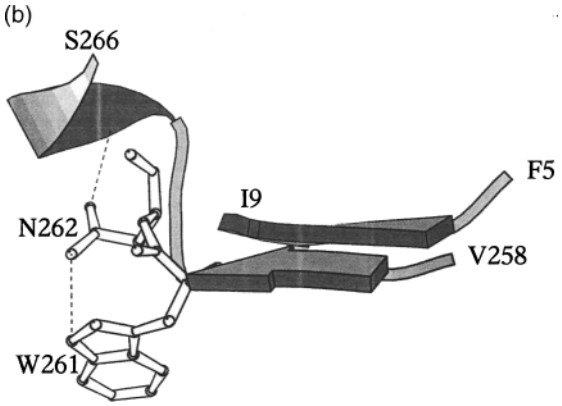




**Figure 10.** (a) Phe45-Ser46 *cis* peptide in the longest strand (residues 41-52) of the eight parallel-stranded β-barrel structure of 2EBN; two adjacent strands are shown, as well as another strand that forms an antiparallel β sheet with the C-terminal end of the strand. The turn of a nearby helix (residues 64-71) has a curvature that matches the bulge formed by the *cis* peptide. (b) Trp261-Asn262 *cis* peptide at the junction of a strand (residues 257-261) (constituent of a nine-stranded β-sheet) and a helix (263-266) in 1NAR. The side-chain of Asn262 is involved in a hydrogen bonding with N-Asn264 (2.97 Å) in the helix, and an N-H$\cdots\pi$ interaction (3.32 Å) with NE1-Trp261 across the *cis* peptide.

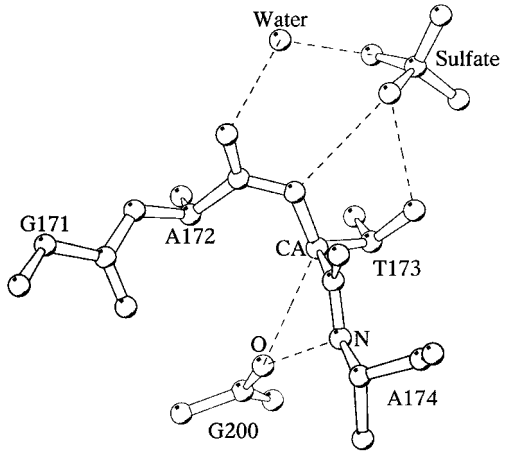



**Figure 11.** C-H$\cdots$O interaction at a distance of 3.25 Å, between CA-Thr173 and O-Gly200, with a central *cis* bond in the fragment Gly(171)-Ala-Thr-Ala, from 1NBA (subunit A). Also shown are the hydrogen bonds involving the *cis* peptide group, a sulfate anion and a water molecule.

ution of the polypeptide, Ser-Tyr-Pro-Tyr-Asp-Val (Yao *et al.*, 1994).

It has already been mentioned that when the *cis* bond is between two Pro residues, an aromatic residue is favoured at the (2′) position (Table 2). There are six such examples; in four of them $C^{\alpha}$ of Pro at (1) interacts with the π-face of the aromatic residue with the closest contact distance being 3.8 (±3) Å (Figure 12(d)).

Jabs and co-workers (1999) have analyzed the existence of C-H$\cdots\pi$ interaction in Xnp-Ar and Ar-Xnp cases, where a $C^{\beta}$-H group of a non-aromatic residue points directly to the centre of the aromatic residue at distances ranging from 3.4 to 4.4 Å. In the light of our recent work on the face-specific interactions of Trp residues, where we observed that of all the ring atoms NE1 has the maximum number of CH groups interacting with it (Samanta *et al.*, 1999), it is worth looking into the interaction present in Trp-Xnp *cis* peptides which have the maximum number of occurrence in Table 1. Interestingly, in three cases, $C^{\beta}$ of Xnp has the shortest contact with NE1 of Trp, and in the fourth an NH from an Asn side-chain points toward the π electrons of Trp at NE1 giving rise to an N-H$\cdots\pi$ interaction (Figure 10(b)). The possibility of a strong interaction when a CH or NH group is directed towards the π electrons of a nitrogen atom may be the reason why a relatively larger number of Trp-Xnp bonds occur in the *cis* form.

While dealing with the turn conformations it was mentioned that type VIb-3 and VId turns have a C-H$\cdots$O interaction between the $C^{\alpha}$H at position (1) and CO at position (1′). Indeed, 65 % of Gly-X *cis* peptides in these conformations are characterized by a C$\cdots$O distance of 3.4 (±2) Å, H$\cdots$O of 2.6 (±3) Å and a C-H$\cdots$O angle of 126 (±4)°.

The above observation prompted us to look for the existence of other C-H$\cdots$O interactions, especially involving the $C^{\alpha}$H group at position (1′), which can impart stability to the *cis* bond. Using a cut-off distance of 3.6 Å, we found that in 16 of 29 Xnp-Xnp cases the $C^{\alpha}$H (1′) interacts with an oxygen atom with the following average parameters: C$\cdots$O, 3.3 (±1) Å; H$\cdots$O, 2.4 (±1) Å and C-H$\cdots$O, 145 (±16)°. The partner oxygen atom usually also participates in conventional hydrogen bonding with the NH group of the residue at position (2′) (Figure 11). This scenario of the $C^{\alpha}$H and

**Table 5.** Hydrogen bond interaction involving NH at (1′) in Xnp-Xnp *cis* peptides

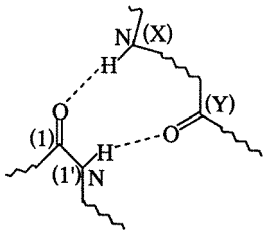
| Type of interaction/hydrogen bond partner[a] | Number of cases (total 29) |
|---|---|
| No interaction | 4 |
| Water | 10[b] |
| Anion (sulfate) | 2 |
| Main-chain O atom | 5 |
| Side-chain O atom | 1 |
| $X = Y$ | 2 |
| $X = i + 2$, $Y = i$ | 2 |
| $X = i + 3$, $Y = i$ | 1 |
| $X = i$, $Y = i + 3$ | 2[c] |

Considering the carbonyl group at (1) it shows no interaction in two cases, binds water in 16 and protein atoms in 11 (seven of which are cyclic structures as given in the diagram below).

[a] As shown in the adjacent diagram, both the CO and NH groups may interact with the main-chain atoms of residues (with numbers X and Y) close in sequence; the relative sequence numbers ($i$, $i + 2$, etc.) of the two groups are provided in the last four entries which exhibit such binding motifs.

[b] In two examples the carboxylate side-chain at position (1) binds Zn.

[c] In both the cases, instead of a CO group the hydroxyl group of Thr is located at (Y).




NH groups of two neighbouring residues interacting with the same oxygen atom has been observed in β-sheets (Derewenda *et al*., 1995; Fabiola *et al*., 1997). A similar C-H···O interaction is also observed in 55 of 200 X-Pro cases, with values 3.3 (±2) Å; 2.5 (±2) Å and 137 (±11)°; in these *cis* peptides, however, only those $C^{\alpha}H$ groups at position (1′), which are not already engaged by C-H···π interaction, can participate in the C-H···O interaction.

## Mutation of residues involved in the *cis* peptide bond

There have been mutational studies in which the proline residue in a *cis* X-Pro has been changed to a non-prolyl residue to assess if the three-dimensional structure around the new X-Xnp bond can preferentially stabilize its *cis* conformation. Tweedy *et al*. (1993) have shown that the *cis* Pro201-Pro202 bond is retained in the structure of a single amino acid variant, Pro202 → Ala (P202A) carbonic anhydrase II, but the substitution causes a reduction in the stability by 5 kcal/mol. Though the destabilization has been attributed mainly to the less favourable *cis-trans* equilibrium of X-Ala bonds compared to X-Pro bonds, our study suggests an additional factor. For wild-type protein (PDB file, 2CAB), the φ,ψ values of Pro202 in type VIa-1 turn are (−77°, 8°) (as can be expected (Table 3) for a Pro-Pro *cis* bond), which remain almost unaltered (−74°, 11°) in Ala202 of the mutant, whereas the average values in an Xnp-Xnp *cis* bond with the same turn conformation are (−111°, 14°). This shows that the three-dimensional structure has not been able to accomodate a shift towards a more negative $\phi_{1'}$ required to change an X-Pro to an X-Xnp *cis* bond. In another study involving ribonuclease A (Schultz & Baldwin, 1992) a destabilization of 2.7 kcal/mol has been reported for the Pro93 → Ala mutant. Apparently the *cis* Tyr92-Ala93 bond is retained, but because of greater mobility in the loop region the value of $\phi_{1'}$ could not be ascertained (Pearson *et al*., 1998). Another point can be made with reference to Figure 3(c). Although a Xnp residue with a $\chi_1 \approx +60°$ is unusual at ψ below 30°, Pro residues with similar $\chi_1$ and ψ are quite common. Hence a substitution by one another may cause a change in the ψ value also.

Like the replacement of Pro to a non-Pro residue at position (1′) leading to local structural changes, alterations can also be expected with some mutations at position (1). For example, as the conformation involving Gly-Pro sequence is quite distinct from the non-Gly-Pro sequence (Figure 5 and Table 3), a substitution by one another may not be isostructural. Staphylococcal nuclease contains a single *cis* peptide bond between residues Lys116 and Pro117. The structure of K116A mutant is indistinguishable from that of the wild-type, but in the structure of K116G mutant, the Gly116-Pro117 bond is found in the *trans* conformation (Hodel *et al*., 1993). It is noteworthy that the *cis* peptide in the wild-type structure (1SNC) exists in VIa turn type for which no X-Pro bond is known to have Gly at position (1) (Table 2D). Even for a VIb turn the Gly residues at this position have conformation significantly different from all non-Gly residues (Figure 5(b)). It is possible that a non-Gly-Pro to Gly-Pro substitution would involve a significant local adjustment in the structure and the energy cost for this (retaining the *cis* bond) is more than what is required to make the bond *trans*.

### Miscellaneous observations

The *cis* peptide can occur anywhere along the polypeptide chain including the C terminus (1EUR, Ala-Pro(407)); an X-Pro *cis* peptide is found within the first and the last ten residues of the polypeptide chain in ten and eight cases, respectively (only one Xnp-Xnp case is found near the C-terminal end). Three protein subunits have four *cis* peptides each (1APY, 1CNV, 3TGL), and 22 have three. The maximum numbers of X-Xnp *cis* peptides are found in 2CTC (all three present are of this type), 1CNV (two out of four) and 1NAR (two out of three). In two structures, there are two *cis* peptides one residue apart: 8FAB: Asp(152)-Tyr-Phe-Pro-Glu-Pro-Val-Thr; 2NAC: Val(309)-Trp-Phe-Pro-Gln-Pro-Ala-Pro (the *cis* bond preceds the underlined residue). There are three *cis* peptides located at constant sequence intervals (eight in 1APY and 49 in 1CNV). In nine cases the difference in sequence number between the consecutive *cis* peptides lies in the range eight to 12, and a short helix or a strand can be accomodated in the intervening region, as also turns and regions of no secondary structure.

In 1PGS, the Cys residue involved in Cys(204)-Ala *cis* peptide is also a constituent of a disulfide linkage (to Cys208). In 1APY, Cys residues preceding two *cis* peptides (Cys(140A)-Gln-Pro and Cys(156A)-Gly-Pro) are connected by a disulfide bond. Moreover, Cys residues following the *cis* peptides can also form disulfide bonds (1LKI, His-Pro-Cys(18); 1SVB, Lys-Pro-Cys(338); and 9PAP, Gly-Pro-Cys(153)). If we consider positions (2) to (2′) there are ten Cys residues overall, out of which six are involved in disulfide linkages. Thus, though Cys residues are not very common in and around *cis* peptides, when present, in 60% cases they are also involved in a disulfide linkage. *Cis-trans* isomerization and disulfide bond formation are known to have profound influence on the rate of protein folding, and in the above instances we have the two important groups adjacent to each other.

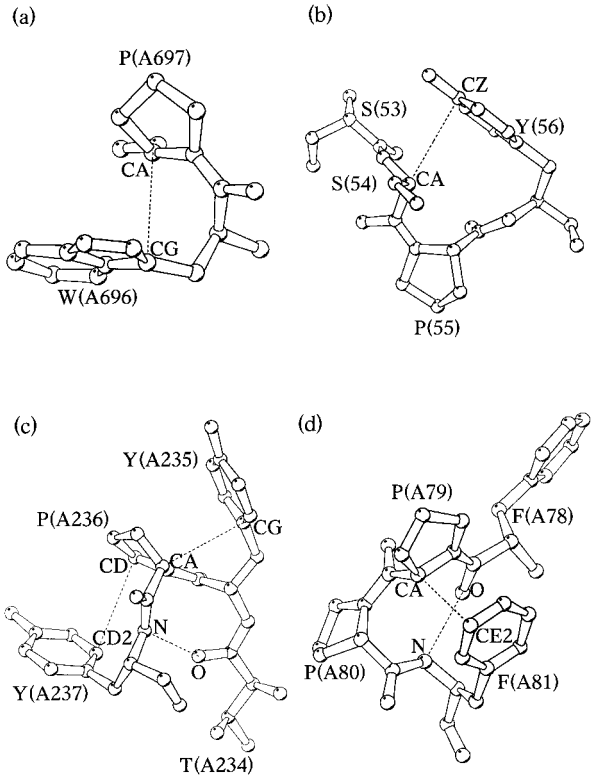



**Figure 12.** C-H$\cdots\pi$ interaction in X-Pro *cis* peptide units; the atoms involved in the shortest contact (broken line) are labelled, as are the residues (the subunit name, if present, is given in front of the residue number). (a) Trp696-Pro697 *cis* peptide in the PDB file 1OAC; the contact distance is 3.54 Å. (b) In 1RGA, the ring of Tyr56 following the Ser54-Pro55 *cis* peptide stacks against the Ser53-Ser54 peptide bond and is in contact (3.73 Å) with the $C^{\alpha}$ atom of Ser54. (c) Thr(234)-Tyr-Pro-Tyr peptide segment from 1ADE having the type VIa turn with a hydrogen bond of length 2.74 Å. The two Tyr rings are aligned with the Pro ring in the middle with the shortest distances of contact being 3.50 and 3.88 Å. (d) Phe(78)-Pro-Pro-Phe peptide fragment from 1AOC with the central bond in the *cis* conformation. The contact distance between Pro79 and Phe81 is 3.55 Å. The hydrogen bond (2.80 Å) in the turn is also shown.

## Conclusions

A comparative analysis of *cis* peptides, both X-Pro and X-Xnp (Figure 1(a)) and their neighbouring residues (Figure 1(b)) has been made in terms of conformation, sequence preference, secondary structural features, local interactions, location in the three-dimensional structure and solvent accessibility. *Cis* peptide-mediated turns are usually designated as types VIa and VIb based on the conformations of residues (1) and (1′) (Figure 5). Depending on the presence (or absence) of hydrogen bonding (Figure 6), C-H$\cdots$O interaction, the existence of a Gly residue at position (1), these turns have been further subdivided into VIa-1 and VIa-2, VIb-1, VIb-2, and VIb-3, VIc and VId with distinct $\phi,\psi$ angles and (2)-(2′) distances (Table 3 and Figure 7), although relatively longer values of the last parameter in a few turn categories, as also the existence of the virtual torsion angle (as defined in Figure 8) beyond 100°, suggest that some of these are really pseudo turns. Compared to X-Pro, Xnp-Xnp *cis* peptides have a more negative $\phi_{1'}$ value which helps to relax the steric clash across the bond. A similar shift in the $\phi$ value in going from a *trans* X-Pro (Figure 4) to a *cis* bond (Figure 3) causes, because of the interrelationship of $\phi$ and $\chi_1$ angles, the pyrrolidine ring of the latter mostly have the DOWN puckering ($\chi_1$ positive).

Aromatic residues have more chance of occurrence preceding (Figure 2) or following (Table 2) Pro in X-Pro *cis* peptides, so that there could be a C-H$\cdots\pi$ interaction involving the CH group of

Pro at $C^{\alpha}$ or $C^{\delta}$ position (Figure 12). With an aromatic residue at position (1) or (1′), a similar interaction can also occur in Xnp-Xnp *cis* peptides (Figure 10(b)), which, however, have more of Gly, Ala or small polar residues at these positions (Tables 1 and 2). In general, β-branched residues are preferred at positions (2) and (3). A proper hydrogen bonding environment (Table 5), C-H···O interaction (Figure 11), shape complementarity and tertiary interaction (Figure 10) stabilize Xnp-Xnp *cis* peptides.

X-Pro and X-Xnp *cis* peptides have differences in their preferred location in the structure, the former lying mostly close to the surface, and the latter more buried (Figure 9). A greater percentage of Xnp-Xnp *cis* peptides are located in regular secondary structures (Table 4), especially strands, and may lead to helices (Figure 10).

## Materials and Methods

The analysis was carried out using 147 protein structures containing *cis* peptide bonds, out of a dataset of 294 X-ray structures from the Brookhaven Protein Data Bank (PDB) (Sussman *et al*., 1998) with resolution $\leqslant$2.0 Å and *R*-factor of $\leqslant$0.2, and which had a homology of $\leqslant$25% on pairwise alignment (Hobohm & Sander, 1994).

The torsion angles ($\phi$, $\psi$ and $\chi_1$) were calculated using the program DIHDRL from PDB. The definitions follow the standard IUPAC-IUB convention except for Val, where 120° was added to $\chi_1$ so as to make its atomic positions equivalent to those of Thr and Ile at any given $\chi_1$ (Chakrabarti & Pal, 1998). The $\chi_1$ and $\psi$ angles were shifted to the range −240 to 120° and −120 to 240°, respectively, to keep the distributions continuous for plotting purposes. The Ramachandran map was divided into four regions A, B, L and R ($A = -180 \leqslant \phi \leqslant 0°$, $-120 \leqslant \psi \leqslant 60°$; $B = -180 \leqslant \phi \leqslant 0°$, $60° < \psi \leqslant 240°$; $L = 0 < \phi \leqslant 180°$, $-90 \leqslant \psi \leqslant 90°$; $R = 0 < \phi \leqslant 180°$, $90 < \psi \leqslant 270°$) corresponding to the $\alpha_R$, β, $\alpha_L$ and the remaining region of the map, respectively (Pal & Chakrabarti, 1999b). The peptide bond was defined as *cis* when the torsional angle ω was found to be $-40° < \omega < 40°$. For comparison, conformational parameters for *trans* Pro residues were also calculated; the larger number of such cases allowed us to choose the more ordered ones (all atoms with temperature factors $\leqslant$15 Å$^2$). The secondary structures were marked using the DSSP program by Kabsch & Sander (1983). For a hydrogen bond the N···O distance was $\leqslant$3.5 Å and the N-H···O angle $\geqslant$120°.

The solvent-accessible surface area (ASA) was calculated using the program ACCESS (Hubbard, 1992), which is an implementation of the Lee & Richards (1971) algorithm. We used the default van der Waals radii in the program and the solvent probe size was 1.4 Å. The solvent accessibility of a residue was evaluated by the ratio of the summed atomic accessible surface areas of that residue in the protein to that of the same residue (X) in an extended Ala-X-Ala tripeptide. Only one subunit was considered while performing these calculations. The molecular plots were made using MOLSCRIPT (Kraulis, 1991).

The following PDB files were used (subunit name (if present) and the number of *cis* peptides (if more than one) are given after the hyphen): 1ADE-A; 1AKZ-2; 1ALO; 1AMP; 1AOC-A; 1AOZ-A3; 1APY-A4; 1AYL; 1BDM-B; 1BEC-2; 1CEL-A; 1CEM; 1CEO-3; 1CHD; 1CLC-2; 1CNS-A; 1CNV-4; 1CPO-3; 1CVL; 1DIN; 1DJA; 1DKZ-A; 1DOR-A2; 1DPE; 1DYR-2; 1ECA; 1EDE-2; 1EDG; 1EUR-3; 1FBA-A; 1FNC; 1FUA; 1GAI-3; 1GDO-A; 1GEO-2; 1GOF-3; 1GP1-A2; 1GSA-2; 1HA1; 1HCZ; 1HGX-A; 1HSL-A; 1HXN; 1I1B; 1ILK; 1IOW-3; 1ISC-A; 1ISO; 1JAP-A; 1JER; 1JPC; 1KNB; 1LCL; 1LCP-A; 1LEN-A; 1LFA-A; 1LIT; 1LKI-2; 1LTS-A,D; 1LUC-A2; 1MHY-G; 1MKA-A2; 1MLA; 1MPP; 1MSK-2; 1NAR-3; 1NBA-A; 1NIF; 1NSJ; 1NUL-A2; 1OAC-A; 1OBW-A; 1OYC; 1PBE; 1PBN; 1PGS-3; 1PHG-3; 1PNK-B3; 1PUD-3; 1QBA-3; 1RBU; 1RCY-3; 1RGA-2; 1RSY; 1RYC; 1SAC-A; 1SFT-A; 1SMD-2; 1SNC; 1SRA; 1SBV; 1TCA-2; 1TF4-A3; 1TGX-A; 1THV; 1THX; 1U9A-A2; 1UAE-2; 1VHH; 1VID; 1VOM; 1VPS-A; 1VSC-A3; 1VSD; 1WBA-3; 1WHO; 1WHT-A3; 1XEL; 1XER; 1XGS-A; 1XJO; 1XNB; 1XYZ-A; 1ZIA; 2AK3-A; 2AYH; 2CBA-2; 2CMD; 2CTC-3; 2EBN-2; 2ER7-E2; 2FHA; 2GST-A3; 2HMZ-A; 2KAU-C3; 2MYR-3; 2NAC-A2; 2OLB-A; 2PGD; 2PRK; 2SIL; 2TGI; 2TYS-A, B2; 3BCL; 3CHY; 3DFR-2; 3GRS-2; 3PTE; 3TGL-4; 4ENL-2; 4RHN; 5RUB-A; 7RSA-2; 8ACN; 8FAB-B2; 8TLN-E; and 9PAP. Two structures (1NIF and 1XER) had a break in the chain in the six-residue window (Figure 1(b)), and were excluded. Further information on residues and their conformational parameters is available as a text file (pub/pinak/cis/cisdata) in boseinst.ernet.in.

## Acknowledgments

The authors thank the Department of Science and Technology, Council of Scientific and Industrial Research for financial assistance, and the Department of Biotechnology for supporting the computational facilities.

## References

Ashida, T. & Kakudo, M. (1974). Conformations of prolyl residues in oligopeptides. *Bull. Chem. Soc. Jpn.* **47**, 1129-1133.

Brandts, J. F. & Lin., L.-N. (1986). Proline isomerization studied with proteolytic enzymes. *Methods Enzymol.* **131**, 107-126.

Brandts, J. F., Halvorson, H. R. & Brennan, M. (1975). Consideration of the possibility that the slow step in protein denaturation reactions is due to *cis-trans* isomerism of proline residues. *Biochemistry,* **14**, 4953-4963.

Chakrabarti, P. & Chakrabarti, S. (1998). C-H···O hydrogen bond involving proline residues in α-helices. *J. Mol. Biol.* **284**, 867-873.

Chakrabarti, P. & Pal, D. (1998). Main-chain conformational features at different conformations of side-chains in proteins. *Protein Eng.* **11**, 631-647.

Chakrabarti, P. & Samanta, U. (1995). CH/π interaction in the packing of adenine ring in protein structures. *J. Mol. Biol.* **251**, 9-14.

Christensen, D. H., Kortzeborn, R. N., Bak, B. & Led, J. J. (1970). Results of *ab initio* calculations on formamide. *J. Chem. Phys.* **53**, 3912-3922.

Cook, K. H., Schmid, F. X. & Baldwin, R. L. (1979). Role of proline isomerization in folding of ribonuclease A at low temperatures. *Proc. Natl Acad. Sci. USA,* **76**, 6157-6161.

Creighton, T. E. (1978). Possible implications of many proline residues for the kinetics of protein unfolding and refolding. *J. Mol. Biol.* **125**, 401-406.

Derewenda, Z. S., Lee, L. & Derewenda, U. (1995). The occurrence of C-H···O hydrogen bonds in proteins. *J. Mol. Biol.* **252**, 248-262.

Desiraju, G. R. & Steiner, T. (1999). *The Weak Hydrogen Bond in Structural Chemistry and Biology*, Oxford University Press. Oxford.

Drakenberg, T. & Forsén, S. (1971). The barrier to internal rotation in monosubstituted amides. *J. Chem. Soc. Chem. Commun.* **000**, 1404-1405.

Evans, T. C., Jr & Nelsestuen, G. L. (1996). Importance of *cis*-proline 22 in the membrane-binding conformation of bovine prothrombin. *Biochemistry,* **35**, 8210-8215.

Fabiola, G. F., Krishnaswamy, S., Nagarajan, V. & Pattabhi, V. (1997). C-H···O hydrogen bonds in β-sheets. *Acta Crystallog, sect. D,* **53**, 316-320.

Frömmel, C. & Preissner, R. (1990). Prediction of prolyl residues in *cis*-conformation in protein structures on the basis of the amino acid sequence. *FEBS Letters,* **277**, 159-163.

Grathwohl, C. & Wüthrich, K. (1976). NMR studies of the molecular conformations in the linear oligopeptides H-(L-Ala)$_n$-L-Pro-OH. *Biopolymers,* **15**, 2043-2057.

Grathwohl, C. & Wüthrich, K. (1981). NMR studies of the rates of proline *cis-trans* isomerization in oligopeptides. *Biopolymers,* **20**, 2623-2633.

Herzberg, O. & Moult, J. (1991). Analysis of the steric strain in the polypeptide backbone of protein molecules. *Proteins: Struct. Func. Genet.* **11**, 223-229.

Hobohm, U. & Sander, C. (1994). Enlarged representative set of protein structures. *Protein Sci.* **3**, 522-524.

Hodel, A., Kautz, R. A., Jacobs, M. D. & Fox, R. O. (1993). Stress and strain in staphylococcal nuclease. *Protein Sci.* **2**, 838-850.

Hubbard, S. (1992). *ACCESS: A Program for Calculating Accessibilities*, Department of Biochemistry and Molecular Biology. University College of London.

Jabs, A., Weiss, M. S. & Hilgenfeld, R. (1999). Non-proline *cis* peptide bonds in proteins. *J. Mol. Biol.* **286**, 291-304.

Jorgensen, W. L. & Gao, J. (1988). *Cis-trans* energy difference for the peptide bond in the gas phase and in aqueous solution. *J. Am. Chem. Soc.* **110**, 4212-4216.

Juy, M., Lam-Thanh, H., Lintner, K. & Fermandjian, S. (1983). Conformation and mobility of tyrosine side chain in tetrapeptides. Specific effects of *cis*- and *trans*-proline in Tyr-Pro- and Pro-Tyr-segments. *Int. J. Pept. Protein Res.* **22**, 437-449.

Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure. Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers,* **22**, 2577-2637.

Kim, P. S. & Baldwin, R. L. (1990). Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* **59**, 631-660.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.

LaPlanche, L. A. & Rogers, M. T. (1964). *Cis* and *trans* configurations of the peptide bond in *N*-monosubstituted amides by nuclear magnetic resonance. *J. Am. Chem. Soc.* **86**, 337-341.

Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379-400.

Lewis., P. N., Momany, F. A. & Scheraga, H. A. (1973). Chain reversals in proteins. *Biochim. Biophys. Acta,* **303**, 211-229.

Lin, L.-N. & Brandts, J. F. (1984). Involvement of prolines-114 and -117 in the slow refolding phase of ribonuclease A as determined by isomer-specific proteolysis. *Biochemistry,* **23**, 5713-5723.

MacArthur, M. W. & Thornton, J. M. (1991). Influence of proline residues on protein conformation. *J. Mol. Biol.* **218**, 397-412.

Maigret, B., Perahia, D. & Pullman, B. (1970). Molecular orbital calculations on the conformation of polypeptides and proteins. IV. The conformation of the prolyl and hydroxyprolyl residues. *J. Theor. Biol.* **29**, 275-291.

Mayr, L. M., Landt, O., Hahn, U. & Schmid, F. X. (1993). Stability and folding kinetics of ribonuclease $T_1$ are strongly altered by the replacement of *cis*-proline 39 with alanine. *J. Mol. Biol.* **231**, 897-912.

Milner-White, J. E., Bell, L. H. & Maccallum, P. H. (1992). Pyrrolidine ring puckering in *cis* and *trans*-proline residues in proteins and polypeptides. Different puckers are favoured in certain situations. *J. Mol. Biol.* **228**, 725-734.

Nishio, M., Hirota, M. & Umezawa, Y. (1998). *The CH-π Interaction. Evidence, Nature, and Consequences*, Wiley-VCH, New York.

Odefey, C., Mayr, L. M. & Schmid, F. X. (1995). Non-prolyl *cis-trans* peptide bond isomerization as a rate-determining step in protein unfolding and refolding. *J. Mol. Biol.* **245**, 69-78.

Pal, D. & Chakrabarti, P. (1999a). Estimates of the loss of main-chain conformational entropy of different residues on protein folding. *Proteins: Struct. Func. Genet.* **36**, 332-329.

Pal, D. & Chakrabarti, P. (1999b). Graphical representation of salient conformational features of residues in protein. *Protein Eng.* **12**, 523-526.

Pauling, L. (1960). *The Nature of the Chemical Bond*, 3rd edit., pp. 281, Cornell University Press, Ithaca, NY.

Pearson, M. A., Karplus, P. A., Dodge, R. W., Laity, J. H. & Scheraga, H. A. (1998). Crystal structures of two mutants that have implications for the folding of bovine pancreatic ribonuclease A. *Protein Sci.* **7**, 1255-1258.

Perricaudet, M. & Pullman, A. (1973). An ab initio quantum-mechanical investigation on the rotational isomerism in amides and esters. *Int. J. Pept. Protein Res.* **5**, 99-107.

Radzicka, A., Pedersen, L. & Wolfenden, R. (1988). Influences of solvent water on protein folding: free energies of solvation of *cis* and *trans* peptides are nearly identical. *Biochemistry*, **27**, 4538-4541.

Ramachandran, G. N. & Mitra, A. K. (1976). An explanation for the rare occurrence of *cis* peptide units in proteins and polypeptides. *J. Mol. Biol.* **107**, 85-92.

Ramachandran, G. N. & Sasisekharan, V. (1968). Conformation of polypeptides and proteins. *Advan. Protein Chem.* **23**, 283-437.

Ramachandran, G. N., Lakshminarayanan, A. V., Balasubramanium, R. & Tegoni, G. (1970). Energy calculations on proline residues. *Biochim. Biophys. Acta,* **221**, 165-181.

Reimer, U., El Mokdad, N., Schutkowski, M. & Fischer, G. (1997). Intramolecular assistance of *cis/trans* isomerisation of the His-Pro moiety. *Biochemistry,* **36**, 13802-13808.

Reimer, U., Scherer, G., Drewello, M., Kruber, S., Schutkowski, M. & Fischer, G. (1998). Side-chain

effects on peptidyl-prolyl *cis/trans* isomerization. *J. Mol. Biol.* **279**, 449-460.
Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Advan. Protein Chem.* **34**, 167-339.
Richardson, J. S. & Richardson, D. C. (1988). Amino acid preferences for specific locations at the ends of α helices. *Science,* **240**, 1648-1652.
Rose, G. D., Gierasch, L. M. & Smith, J. A. (1985). Turns in peptides and proteins. *Advan. Protein Chem.* **37**, 1-109.
Samanta, U., Chakrabarti, P. & Chandrashekhar, J. (1998). *Ab initio* study of energetics of X-H$\cdots\pi$ (X = N, O, and C) interactions involving a heteroaromatic ring. *J. Phys. Chem. A.* **102**, 8964-8969.
Samanta, U., Pal, D. & Chakrabarti, P. (1999). Environment of tryptophan side-chains in proteins. *Proteins: Struct. Funct. Genet.* In the press.
Scherer, G., Kramer, M. L., Schutkowski, M., Reimer, U. & Fischer, G. (1998). Barriers to rotation in secondary amide peptide bonds. *J. Am. Chem. Soc.* **120**, 5568-5574.
Schmid, F. X. & Baldwin, R. L. (1978). Acid catalysis of the formation of the slow-folding species of RNase A: evidence that the reaction is proline isomerization. *Proc. Natl Acad. Sci. USA,* **75**, 4764-4768.
Schmid, F. X., Mayr, L. M., Mücke, M. & Schönbrunner, E. R. (1993). Prolyl isomerases: role in protein folding. *Advan. Protein Chem.* **44**, 25-66.
Schnur, D. M., Yuh, Y. H. & Dalton, D. R. (1989). Molecular mechanics study of amide conformations. *J. Org. Chem.* **54**, 3779-3785.
Schultz, D. A. & Baldwin, R. L. (1992). *Cis* proline mutants of ribonulease A. I. Thermal stability. *Protein Sci.* **1**, 910-916.
Schulz, G. D. & Schirmer, R. H. (1984). *Principles of Protein Structure*, pp. 25, Springer-Verlag, New York.
Stewart, D. E., Sarkar, A. & Wampler, J. E. (1990). Occurrence and role of *cis* peptide bonds in protein structures. *J. Mol. Biol.* **214**, 253-260.
Stoddard, B. L. & Pietrokovski, S. (1998). Breaking up is hard to do. *Nature Struct. Biol.* **5**, 3-5.
Sussman, J. L., Lin, D., Jiang, J., Manning, N. O., Prilusky, J., Ritter, O. & Abola, E. E. (1998). Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. *Acta Crystallog. sect. D,* **54**, 1078-1084.
Tweedy, N. B., Nair, S. K., Paterno, S. A., Fierke, C. A. & Christianson, D. W. (1993). Structure and energetics of a non-proline *cis*-peptidyl linkage in a proline-202 → alanine carbonic anhydrase-II variant. *Biochemistry*, **32**, 10944-10949.
Vance, J. E., LeBlanc, D. A., Wingfield, P. & London, R. E. (1997). Conformational selectivity of HIV-1 protease cleavage of *X*-Pro peptide bonds and its implications. *J. Biol. Chem.* **272**, 15603-15606.
Vanhove, M., Raquet, X., Palzkill, T., Pain, R. H. & Frere, J. M. (1996). The rate-limiting step in the folding of the *cis*-Pro167Thr mutant of TEM-1 beta-lactamase is the *trans* to *cis* isomerization of a non-proline peptide bond. *Proteins: Struct. Funct. Genet.* **25**, 104-111.
Wilmot, C. M. & Thornton, J. M. (1988). Analysis and prediction of the different types of β-turn in proteins. *J. Mol. Biol.* **203**, 221-232.
Weiss, M. S., Jabs, A. & Hilgenfeld, R. (1998). Peptide bonds revisited. *Nature Struct. Biol.* **5**, 676.
Yao, J., Feher, V. A., Espejo, B. F., Reymond, M. T., Wright, P. E. & Dyson, H. J. (1994). Stabilization of a type VI turn in a family of linear peptides in water solution. *J. Mol. Biol.* **243**, 736-753.

***Edited by J. M. Thornton***

*(Received 15 June 1999; received in revised form 26 August 1999; accepted 20 September 1999)*