PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

ω-Turn: A novel β-turn mimic in globular proteins stabilized by main-chain to side-chain C—H···O interaction

Jesmita Dhar,[1] Pinak Chakrabarti,[1,2]* Harpreet Saini,[3] Gajendra Pal Singh Raghava,[3] and Raghuvansh Kishore[4]*

[1] Bioinformatics Centre, Bose Institute, Kolkata, West Bengal 700 054, India

[2] Department of Biochemistry, Bose Institute, Kolkata, West Bengal 700 054, India

[3] Bioinformatics Centre, CSIR-Institute of Microbial Technology, Chandigarh, 160 036, India

[4] Protein Science and Engineering Division, CSIR-Institute of Microbial Technology, Chandigarh, 160 036, India

ABSTRACT

Mimicry of structural motifs is a common feature in proteins. The 10-membered hydrogen-bonded ring involving the main-chain C═O in a β-turn can be formed using a side-chain carbonyl group leading to Asx-turn. We show that the N—H component of hydrogen bond can be replaced by a C^{γ}-H group in the side chain, culminating in a nonconventional C—H···O interaction. Because of its shape this β-turn mimic is designated as ω-turn, which is found to occur ~three times per 100 residues. Three residues (i to $i+2$) constitute the turn with the C—H···O interaction occurring between the terminal residues, constraining the torsion angles ϕ_{i+1}, ψ_{i+1}, ϕ_{i+2} and $\chi'_{1(i+2)}$ (using the interacting C^{γ} atom). Based on these angles there are two types of ω-turns, each of which can be further divided into two groups. C^{β}-branched side-chains, and Met and Gln have high propensities to occur at $i+2$; for the last two residues the carbonyl oxygen may participate in an additional interaction involving the S and amino group, respectively. With Cys occupying the $i+1$ position, such turns are found in the metal-binding sites. N-linked glycosylation occurs at the consensus pattern Asn-Xaa-Ser/Thr; with Thr at $i+2$, the sequence can adopt the secondary structure of a ω-turn, which may be the recognition site for protein modification. Location between two β-strands is the most common occurrence in protein tertiary structure, and being generally exposed ω-turn may constitute the antigenic determinant site. It is a stable scaffold and may be used in protein engineering and peptide design.

Proteins 2015; 83:203–214.

Key words: ω-turn; C—H···O interaction; β-turn mimicry; weak interactions in protein structures; N-linked glycosylation.

INTRODUCTION

Hydrogen bonds stabilize not only the secondary structural elements,[1] but also various nonrepetitive local structures and loops leading to unique globular topology of proteins.[2,3] β-turns that lead to the reversal of chain direction are ubiquitous in protein structures, and were characterized based on hydrogen bond between the main-chain carbonyl group of residue i and the main-chain NH of residue $i+3$ in a four-residue segment spanning positions i to $i+3$ (Fig. 1).[4–9] As the carbonyl group is also found in the side chain of Asx (Asp or Asn) residue, a variation of the above interaction can also occur in nature where the main-chain carbonyl is substituted by the γ-carbonyl of Asx.[6] This so-called Asx turn has three residues (i to $i+2$) with Asx occupying the first position, and like β-turn this also forms a 10-atom hydrogen-bonded ring.[10–15] The backbone mimicry by polar side chains is not restricted to only Asx; Ser and Thr residues (i) can also form hydrogen bond with the main-chain NH group of residues two

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: Council of Scientific and Industrial Research (CSIR), Department of Biotechnology (DBT), and the Department of Science and Technology (DST; JC Bose National Fellowships to P.C. and G.P.S.R.). IMTech Communication No. 050/2014.

The paper is dedicated to Prof. P. Balaram on the occasion of his 65th birthday.

Financially support was provided by the in the form of JC Bose National Fellowships to PC and GPSR. This is.

*Correspondence to: Pinak Chakrabarti; Department of Biochemistry, Bose Institute, Kolkata, West Bengal 700 054, India. E-mail: pinak@jcbose.ac.in or Raghuvansh Kishore; Protein Science and Engineering Division, CSIR-Institute of Microbial Technology, Chandigarh, 160 036, India. E-mail: kishore@imtech.res.in

Received 11 July 2014; Revised 28 October 2014; Accepted 3 November 2014
Published online 12 November 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/prot.24720

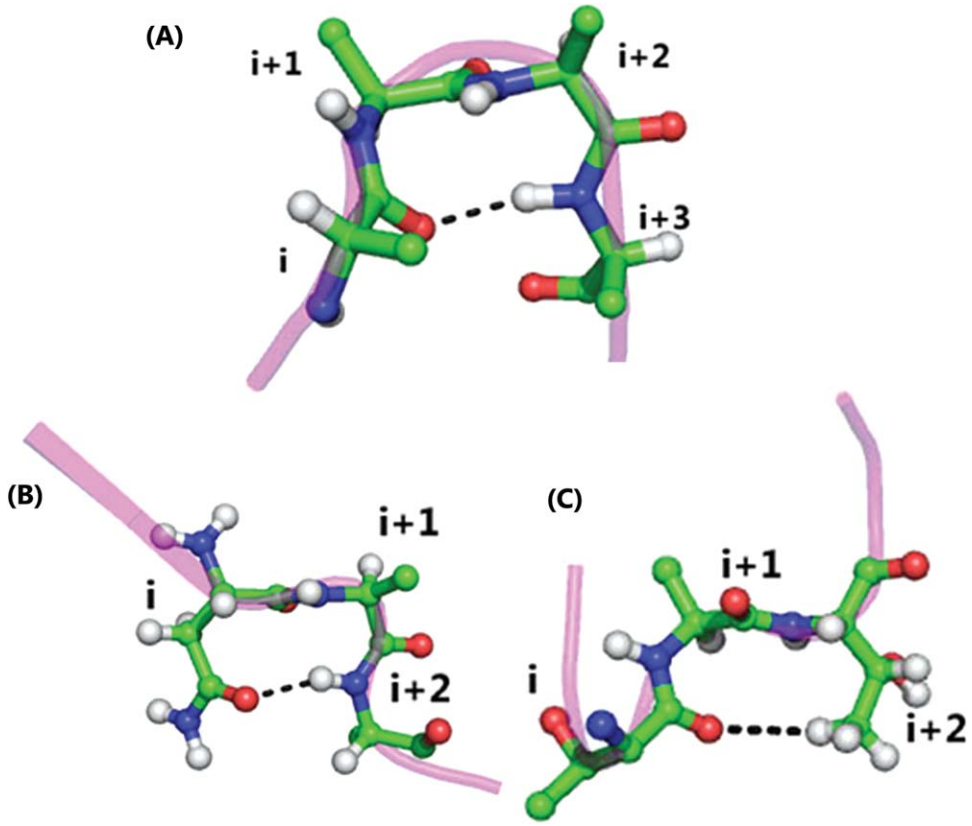

Figure 1

The overall topological resemblance of (A) a classical β-turn [from the PDB file 1CSE, sequence (39)His-Pro-Asp-Leu], (B) Asx-turns [1NEY, (10)Asn-Phe-Lys], and (C) the newly proposed ω-turn [1DP4, (12)Thr-Asn-Thr]. Dotted lines indicate intraturn hydrogen bonds (N/C—H···O). The propagation of the polypeptide chain around the turns is shown in pink. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

ahead ($i+2$), though now the hydrogen-bonded ring contains nine atoms (rather than 10). These have been designated as ST-turns.[12,13,16] The four residues have often been found to replace each other indicating the interchangeability between Asx- and ST-turns.[16,17] There is interesting similarity in conformation and the frequency of occurrence in the four classes of turns that are observed within β-, Asx-, and ST-turns.[18]

Just as the interaction involving the main-chain carbonyl in β-turns can be substituted by side chains in Asx- or ST-turns, here we show that the NH group can also be replaced by a $C^{\gamma}H$ group from a side chain, leading to what we term ω-turn, held together by a C—H···O interaction. Again the mimicry involves residues at i and $i+2$ [Fig. 1(C)] harboring a 10-membered ring motif.

Ever since Taylor and Kennard[19] presented an unambiguous crystallographic evidence for the existence of weak nonconventional C—H···O type attractive interactions based on statistical analysis of high-resolution organic crystal structures, there has been remarkable improvement in our understanding of the structural importance of these weak interactions and their use in the constructive creation of a variety of supramolecular architectures (i.e., crystal engineering).[20,21] Subsequently, these interactions have also been found to be important in the structure and function of proteins.[22–35] The structural significance of C—H···O type attractive forces in stabilization of folding motifs and scaffolds has also been recognized in membrane proteins.[36,37] Weak nonconventional C—H···π interactions[38] have been found to play a very important role in the formation of β-turns around the *cis* peptide bond.[39] Likewise, C—H···O hydrogen bonds are shown here to define β-turn mimics with distinct structural features, based on which they can be termed as ω-turn.

The impetus for this analysis in proteins is rooted in the systematic X-ray diffraction analysis of model peptides of the type: Boc-Xaa-Thr-NH_2 (where Boc = *t*-butoxycarbonyl group and Xaa = C^{β}- or C^{γ}-branched proteinogenic residue) that revealed the existence of distinct secondary structural motifs, stabilized by conventional intramolecular hydrogen bonding interactions.[40,41] Of particular interest,

the crystal structure analysis of Thr containing model peptides.[42,43] Boc-Thr-Thr-OCH_3 and Boc-Thr-Thr-NH_2 demonstrated the subsistence of a significantly "flat" architecture across the Thr^1-Thr^2 segment with the backbone torsion angles of $\phi_{i+1} \sim -60°$, $\psi_{i+1} \sim 140°$, $\phi_{i+2} \sim -130°$, $\psi_{i+2} \sim 0°$, in conjunction with favored *gauche*$^-$ orientation of the Thr^2 side-chain, that is, N—C^α—C^β—C^γ $\sim -60°$ that facilitated the formation of a nonconventional main-chain to side-chain C—H···O type of hydrogen bond, that is, C_i═O···H—C^γ_{i+2} interaction. As to be discussed later (Table II), the conformation of this hydrogen bonded structure mimics quite well the well-known Type II β-turn and the Asx-turn or pseudoturn. As such the starting point of our analysis was centered around the identification of three-residue stretch in protein structures centered around the two pairs of ϕ, ψ angles observed above, with Thr at $i+2$ positions; if the C^γH of this residue formed a C—H···O hydrogen bond with the carbonyl group at *i*, it was identified as the ω-turn. Subsequently, we expanded the scope of ω-turn by including any residue with a C^γH group and following the earlier work on the classifications of Asx-turns,[18] the different combinations of the main- and side-chain torsion angles that might lead to a ω-turn.

METHODS

Dataset

Atomic coordinates were obtained from the Protein Data Bank (PDB; version March, 2013).[44] A representative data set of 4114 protein chains (Supporting Information Table S1) present in 3976 X-ray structures (resolution $\leq$ 2.0 Å, R-factor $\leq$ 20%) were selected using PISCES server[45] such that the sequence identity between any pair was less than 25%. The analysis was also extended to structures determined using neutron diffraction, and the list of 21 chains selected using the same set of conditions is provided in the table.

Stereochemical criteria and the identification of C_i═O···H—C^γ_{i+2} interaction

We searched for two-residue fragments (Thr being the second residue) satisfying the main-chain conformation of the ω-turn as observed in the crystal structures of terminally protected model peptide, Boc-Thr-Thr-OCH_3,[42] having $\phi_{i+1} = -62.1°$, $\psi_{i+1} = 137.1°$ for Thr^1, $\phi_{i+2} = -130.3°$, $\psi_{i+2} = 5.7°$ for Thr^2 and the *gauche*$^-$ orientation of the Thr^2 $C^\gamma H_3$ group (i.e., N-C^α-C^β-C^γ) being $-52.9°$. Of these four backbone torsion angles the ones for the position $i+1$ correspond to Type II β-turn and those for $i+2$ to Type I (or II′) β-turn (Table II). After some initial testing we restricted our search with the stereochemical constraints: $\phi_{i+1} = -73 \pm 30°$, $\psi_{i+1} = 124 \pm 30°$ and $\phi_{i+2} = -110 \pm 30°$, $\psi_{i+2} = -9 \pm 30°$. The peptide fragments identified were then checked for the occurrence of C—H···O hydrogen bond between the Thr C^γ donor atom and the main-chain carbonyl O (at *i*) acceptor atom such that the distance. $d_{C\cdots O} \leq 4.0$ Å.[20] REDUCE[46] was used to fix the hydrogen positions.

At this stage we realized that the ω-turns in proteins need not necessarily be restricted to the backbone torsion angles as observed in a model peptide structure. As such we looked for any peptide fragment (*i* to $i+2$) satisfying the C—H···O hydrogen bond involving the carbonyl group at *i* and a C^γH at $i+2$ (which could be occupied, in addition to Thr, by Val, Ile, Leu, Met, Arg, Glu, Lys, and Gln). The side-chain torsion angle, χ_1 was determined using MMTSB tool set.[47]

ϕ, ψ angle pairs (at $i+1$ and $i+2$) were clustered into four major groups using a Python script to invoke the k-means algorithm within a C clustering library.[3,48] Secondary structures of the residues present in and around ω-turns were identified using DSSP[49]; structural designations G (for 3_{10}-helix), H (α-helix) and I (π-helix) were grouped as helix; B (β-sheet) and E (extended strand) as β-strand; T (turn) and S (bend) as turn; and the remaining cases as belonging to irregular region (C) in structure. The molecular diagrams were made using PyMol.[50] Relative accessibility was calculated using the program NACCESS.[51]

Residue propensity

The propensity of a residue (with a C^γ—H group) to occur at $i+2$ position in ω-turn was calculated as the ratio of the percentage of occurrence of the residue in ω-turns (Table I, last column) to its percentage in the whole database. The propensity of Ser was calculated as a control, by considering those cases that have a normal O—H···O (instead of C—H···O) hydrogen bond. The propensity of all the 20 residues to occur at *i* and $i+1$ positions were also calculated in a similar fashion, except that the percentage in the numerator was computed relative to the total number of ω-turns observed (the denominator remaining the same).

Prediction of antigenic determinant site

The epitope mapping around the $i+2$ residue in ω-turns was done using DiscoTope server.[52] As this is a structure-based algorithm that needs manual curation we considered only Type IIb turns containing Val/Ile/Thr at $i+2$ position—this type has a reasonable number of occurrences (although not the highest among all the subclasses, Table II) of ω-turns. DiscoTope uses a combination of amino acid statistics and surface exposure. The statistics of amino acids in epitopes and non-epitopes are based on the epitope log-odd ratios (indicating the probability of a peptide being part of an epitope). The sequentially averaging epitope log-odds

Table I
Statistics on the Occurrence of Different Residues in ω-Turn

Residue (at $i+2$)	Propensity	Number of observations[a]	% Out of total occurrence
Val	1.34	6327 (CG2, 4271; CG1, 2056)	9.2
Thr	1.10	3411 (CG2)[b]	6.1
Ile	1.46	4475 (CG2, 1179; CG1, 2659; CD1, 637)	8.1
Leu	0.54	4428 (CG, 2816; CD1, 623; CD2, 989)	4.9
Met	2.25	705(CG)	4.0
Arg	0.99	2536 (CG)	5.0
Glu	0.37	1671 (CG)	2.5
Lys	0.94	3021 (CG)	5.3
Gln	1.44	2113 (CG)	5.5
Ser	0.19	753 (OG)	1.2

Ser, with no C^{γ} atom and thus incapable of forming C—H···O interaction, is used as a control residue to see how many of them are involved in normal hydrogen bonds instead. Branched aliphatic residues with higher propensity to occur in ω-turns are given in the top of the table, followed by those with linear chains.
[a]In square brackets are the numbers corresponding to the different atoms at the γ (and also δ for Ile) positions that participate in the C—H···O interaction.
[b]For Thr, 1303 cases were found with normal hydrogen bond involving OG1.

ratios are calculated by sliding window of nine residues through the sequence and then for a given residue, these scores are summed up from all residues within 10 Å to give epitope propensity value. The residue contact number is the number of C^{α} atoms in the antigen within a distance of 10 Å. Lower contact number indicates that the residue is either laid close to the surface or at a protruding region. The propensity value is combined with contact number to derive the DiscoTope prediction score. A score above the threshold value of −7.7 (corresponding to specificity of 75% and sensitivity 47%) indicates positive prediction.

Conservation of residues

While trying to assign functional relevance to ω-turns we found out that Cys at $i+1$ could be involved in metal binding. As such we calculated the sequence variability in terms of entropy for both $i+1$ and $i+2$ positions, when a Cys occupies the former location. Shanon entropy of each residue present in ω-turn was calculated,[53] the method uses the multiple sequence alignment of all available homologs (sequence identity > 60%) for the protein structure under consideration, as available in the HSSP (homology-derived secondary structures of proteins) database.[54] The entropy was compared to the average entropy of that protein chain, and was assumed to be conserved if the entropy value was less than the average.

RESULTS AND DISCUSSION

In this analysis, we have looked for a three-residue motif (positions i to $i+2$) such that there is a C—H···O hydrogen bond between the carbonyl group at i and a $C^{\gamma}H$ of the side chain at $i+2$. We term this as ω-turn because the reversal of chain direction generates a shape that resembles the Greek symbol (Fig. 1). Although model peptides exhibiting this motif have been found containing Thr at $i+2$, and this is indeed one of the preferred residues in protein structures, we have considered all residues having a C—H proton donor at the γ position in the side chain (Table I). 28,687 ω-turns have been identified in 4114 protein chains (Supporting Information Table S1)— thus there are ∼7 turns per chain. This number translates to 2.9(±1.4) occurrences per 100 residues. Val with two C^{γ} atoms has the highest percentage of occurrence (9.2%) of any amino acid in ω-turns,

Table II
A Comparison of the Average Main-Chain Torsion Angles (°) of Ideal β-Turns with ω-Turn and Asx-Turns

	Residue position[a] and its torsion angles					
Turn and type (occurrence)	Position	ϕ	ψ	Position	ϕ	ψ
(A) ω-turn with Thr/Val/Ile at $i+2$						
Type IIa (7549)	$i+1$	−74(±12)	128(±19)	$i+2$	−119(±17)	137(±21)
Type IIb (3338)	$i+1$	−75(±10)	116(±26)	$i+2$	−101(±28)	−15(±23)
Type I'a (2598)	$i+1$	65(±11)	28(±20)	$i+2$	−110(±20)	138(±20)
Type I'b (728)	$i+1$	60(±30)	19(±26)	$i+2$	−99(±36)	−14(±30)
(B) ω-turn with Met/Arg/Glu/Lys/Gln/Leu is at $i+2$						
Type IIa (5616)	$i+1$	−72(±14)	130(±16)	$i+2$	−118(±21)	141(±22)
Type IIb (3338)	$i+1$	−72(±16)	120(±21)	$i+2$	−100(±21)	−13(±26)
Type I'a (4544)	$i+1$	66(±11)	24(±18)	$i+2$	−110(±21)	103(±70)
Type I (976)	$i+1$	−74(±19)	−14(±21)	$i+2$	58(±20)	−28(±49)
β-turn Type I/I'	$i+1$	−60/60	−30/30	$i+2$	−90/90	0/0
Type II/II'	$i+1$	−60/60	120/−120	$i+2$	80/−80	0/0
Asx-turn Type I/I'	i	−58/53[b]	−55/70	$i+1$	−72/98	−13/−18
Type II/II'	i	−62/65[b]	128/−123	$i+1$	80/−73	5/−14

Values for β- and Asx-turns are from (18); types I and I', II and II' are given in the same rows.
[a]The residue positions are indicated in Fig. 1.
[b]It has been named ϕ because of its conformational equivalence to the main-chain torsion angle ϕ in β-turn; however, it is related to side-chain torsion angle as, $\phi = \chi_1 - 120$ (see Fig. 1 of Ref. 18).

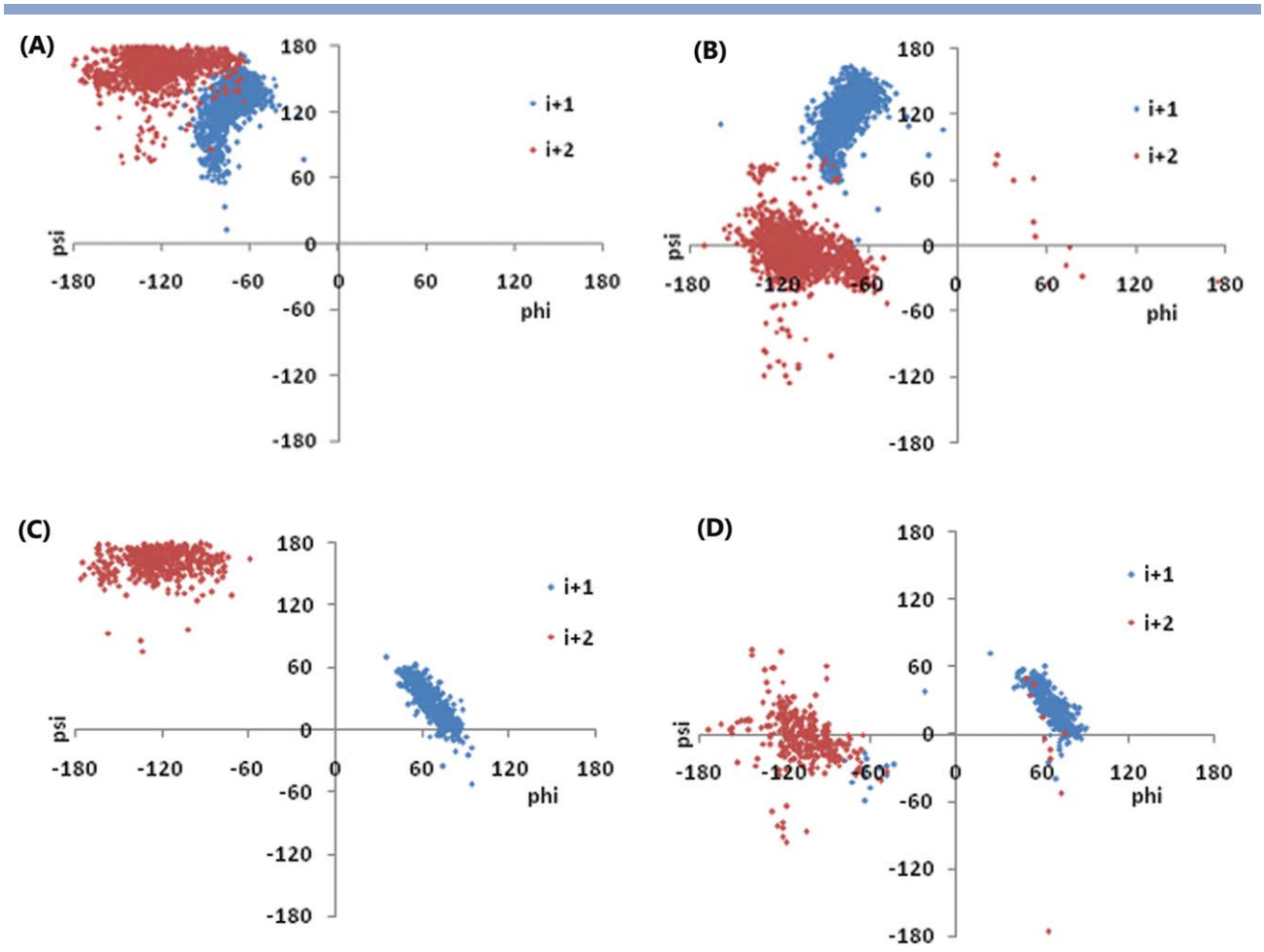

Figure 2

Distribution of ϕ, ψ torsion angles (°) at $i+1$ (blue) and $i+2$ (red) positions for Thr residues involved in ω-turns, types (A) IIa, (B) IIb, (C) I'a and (D) I'b (see Table II for classifications).

comparable to 10.1% observed for Asx/Ser/Thr to occur in β-turn mimics.[18] The C—H···O hydrogen bond is quite linear and an average value of 158(±14)° is observed for the cases involving Thr.

As we have used stereochemically fixed hydrogen positions for the X-ray structures, we also used an additional 21 protein chains whose structures were determined using neutron scattering and thus had accurate information on hydrogen positions. This dataset contained 104 ω-turns – five per chain and 3.7(±1.5) per 100 residues. The geometrical parameters for these turns are: C···O and H···O distances, 3.8(±0.2) Å and 2.9(±0.3) Å, respectively, and the C—H···O angle, 151(±19)°.

Conformational features and classification of ω-turns

When Thr is at $i+2$, the distribution of ϕ, ψ angles at positions $i+1$ and $i+2$ is shown in Figure 2. We could identify four major clusters from the distribution of the torsion angles; these are enumerated in Table II along with the prominent types of β- and Asx-turns. As can be seen from Figure 3, the ω-turn imposes constraint on the ϕ, ψ angles of $i+1$ residue, and only the ϕ angle of $i+2$ (ψ of which is beyond the 10-membered hydrogen bonded ring defining the turn). The fourth angle that is under constraint is the χ_1' angle – a value of ~180° would place the $C^{\gamma}H$ group away from the carbonyl group, limiting the value to be ~ ±60°. If ψ_{i+2} is not considered, the four clusters collapse into two clusters; however, to retain the details of ψ_{i+2} we distinguish them by adding 'a' or 'b' after the name of the ω-turn class in Table II. While trying to match with the β-turn types we find that only the ϕ, ψ angles at $i+1$ can be matched, and there is no correspondence with the values at $i+2$ residue. One distinctive feature is that ϕ_{i+2} is always negative in ω-turn, whereas depending on the turn type the value can be either positive or negative in β-turn. Considering the values at $i+1$ we find that the major ω-turn is of Type II, followed by Type I′. However,

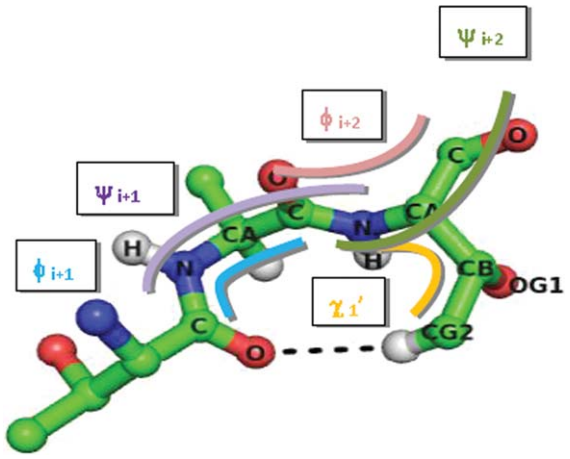

Figure 3

Torsion angles that define ω-turn [shown against residues, (12)Thr-Asn-Thr, in the PDB file 1DP4]. χ_1' is defined with respect to the side-chain C atom that is involved in C—H···O interaction (CG2 here), whereas the standard χ_1 is defined relative to OG1 for Thr. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the long chain residues show some difference with the branched aliphatic residues in that these have a minor cluster that resembles Type I β-turn (Table IIB).

As already mentioned χ_1' angle is constrained in ω-turns; its relationship to χ_1 is depicted in Figure 3. Supporting Information Figure S1 provides the distribution of χ_1 angles. The three idealized values for χ_1 and their equivalent χ_1' angles for different residues are given in Supporting Information Table S2. It can be seen that the value is predominantly $-60°$.

The use of C—H vs O—H group at the γ position for the intraturn interaction

Different moieties in protein structures have been shown to be participating in C—H···O interactions, the strength of which can be modulated by the changes in the C—H bond polarity[55]: the C^{δ}—H of Pro,[25] the $C^{\epsilon 1}$—H and $C^{\delta 2}$—H groups of His[22,26] and the C^{ϵ}—H group of Lys and C^{γ}—H group of Val.[56] It has been shown that C^{β}—H and C^{γ}—H in side chains (at *i*) are favorably oriented to participate in intrahelical C—H···O interactions with the carbonyl group at $i-4$ (or $i-3$) position.[33] The interaction linking the aromatic and the peptide C—H groups is quite common in protein structures and protein–protein interfaces.[34] Here, we observe that C^{γ}—H group (mostly from the β-branched side chains) can function as a potential nonconventional donor. Recent theoretical calculations, performed using hybrid DFT methods at the B3LYP/D95(d,p) level, suggest the involvement of alkyl side-chains of Val, Leu and Ile residues in the formation of C—H···O hydrogen bonds, which in a cooperative fashion lead to the stabilization of α-helices relative to β-strands.[57] High level *ab initio* calculations have indicated that the magnitude of C—H···O interactions involving C^{α}—H donors lies in the range between 1.9 and 2.5 kcal/mol, roughly one-half the strength of the N—H···O hydrogen bond.[58,59]

Val has two C^{γ} positions; CG2 is found to be involved more than CG1 (Table I). For Ile, the interaction is mostly with the C^{γ} positions, with C^{δ} showing only a meager participation. Thr has both C^{γ}H as well as O^{γ}H groups at the γ position, and we wondered why C—H···O is preferred over O—H···O in the formation of the turn. We found that the average C···O distance is 3.8(±0.2) Å, whereas for the cases having normal hydrogen bond the O···O distance is 3.4(±0.4) Å. The latter distance is rather long for a typical hydrogen bond. Thus it appears that the conformation prevailing in ω-turn is such that it does not allow the formation of a optimum, conventional O—H···O hydrogen bond, and the Thr hydroxyl group is better off forming hydrogen bond with other protein or solvent molecule, leaving the C^{γ}H to stabilize the turn geometry with a C—H···O interaction. This is also the reason why Ser has a low propensity (Table I) – with a very similar O···O distance 3.5(±0.4) Å to what is seen in Thr and is not a common residue in such a turn.

Secondary structural features around ω-turns

Consideration of the secondary structure of the residue at $i+2$ position [Fig. 4(A)] shows that for Val (and Ile) β-sheet is the preferred structure, whereas for the rest the preference is not to occur in any regular structural element. If we consider the structural milieu in which such a turn occurs it is observed that the most common occurrence is between two β-strands [Fig. 4(B)]. Some examples of ω-turns with the flanking secondary structural elements are shown in Figure 5.

Propensities of residues to occur in ω-turns

Based on the propensity values (Table I) the side chains with β-branching (Thr, Val and Ile) are preferred, as compared to long chain residues, at $i+2$ position. This could be because of the lesser degrees of freedom associated with the branched side chains of these residues, and the consequent smaller loss of side-chain conformational entropy due to the formation of C—H···O interaction. However, the highest propensity value is exhibited by the long chain hydrophobic residue, Met. While trying to ascribe a reason for this observation we came across the existence of an additional interaction involving the S^{δ} atom in the side chain. It has been noted that S in Met or disulfide bridges in protein structures can have stabilizing stereospecific interaction with the carbonyl oxygen such that the O atom lies along the

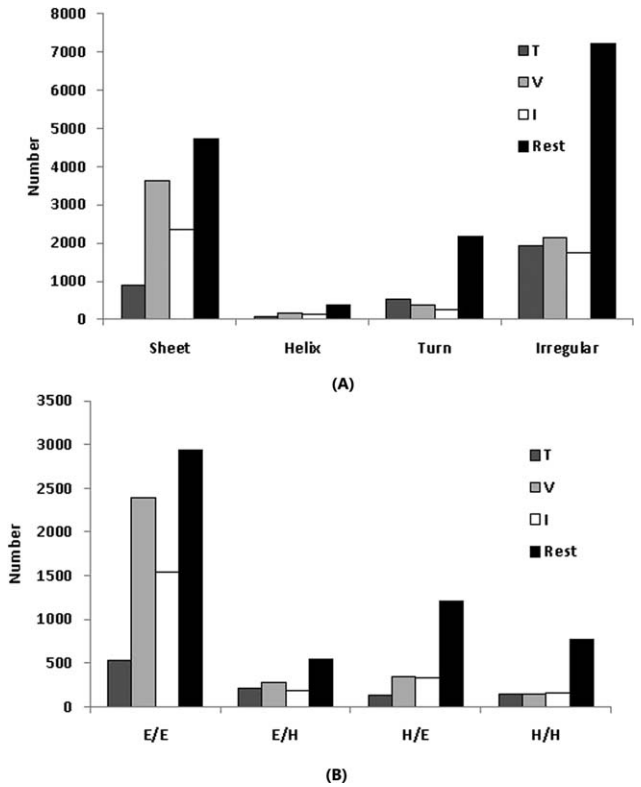

Figure 4

Secondary structural features, (A) for residues at $i+2$ position, and (B) on either side of this position [only helix (H) and strand (E) are considered, occurring in the range $i-1$ to $i+1$ upstream and $i+3$ to $i+5$ downstream].

extension of a C—S bond with the S···O distance being < 4 Å.[60,61] In 12% of the cases involving Met (Table I), there exists a $C_i{=}O\cdots S^{\delta}{-}C^{\gamma}{}_{i+2}$ type interaction satisfying the above criteria [as exemplified in Fig. 6(A)]. Gln is another residue having a high propensity to occur in ω-turns. As in Met this can be attributed to the simultaneous occurrence of another interaction (in 24% of Gln residues), this time conventional hydrogen bonding (distance < 3.5 Å) involving the side-chain $N^{\epsilon}H_2$ group as illustrated in Figure 6(B).

We then looked for residue propensities at other positions in ω-turns. Considering all the turns only four residues (Asp, Asn, Gly, and Pro) had propensities >1 (but below 2) at $i+1$ (data not shown). As the propensity values are known to depend on the type of β-turns,[8] we divided ω-turns into two groups, major (Types IIa + IIb) and minor (I′a + I′b), and calculated their propensities separately. Now the values were larger and distinct (Fig. 7)—the former having Pro as the preferred residue, as in Type II β-turn,[8] whereas the latter indicates strong preference for Gly, Asn, and Asp, as in Type I′ β-turn. We also calculated the propensities of different residues to occur at position *i*. Although not very strong, there is a preference for Gly, Asn, and Lys (Supporting Information Fig. S2). It is plausible to consider ω-turn to be made up of two sharp turns centered at *i* and $i+2$ positions (Fig. 5), and the residues like Gly and Asn can induce the topological features at *i*.

Functional relevance of ω-turns

We analyzed if ω-turns can occur at the enzyme active site. As the number of cases is too large we restricted ourselves only to those having Thr at $i+2$ position and belonging to Type IIb turn (Table IIA) as such a ω-turn was initially characterized in Thr containing peptides.[42,43] For this group Cys was found to have a propensity > 1 to occur at $i+1$ position (Supporting Information Fig. S3). We analyzed such Cys residues and found that these are never involved in disulfide bonds and many of them are part of iron-sulfur cluster in metalloproteins or bind metal ions (Fig. 8), indicating the importance of ω-turns in proteins. Being part of the active site would imply that these residues are conserved among homologous proteins. This indeed is the case, as

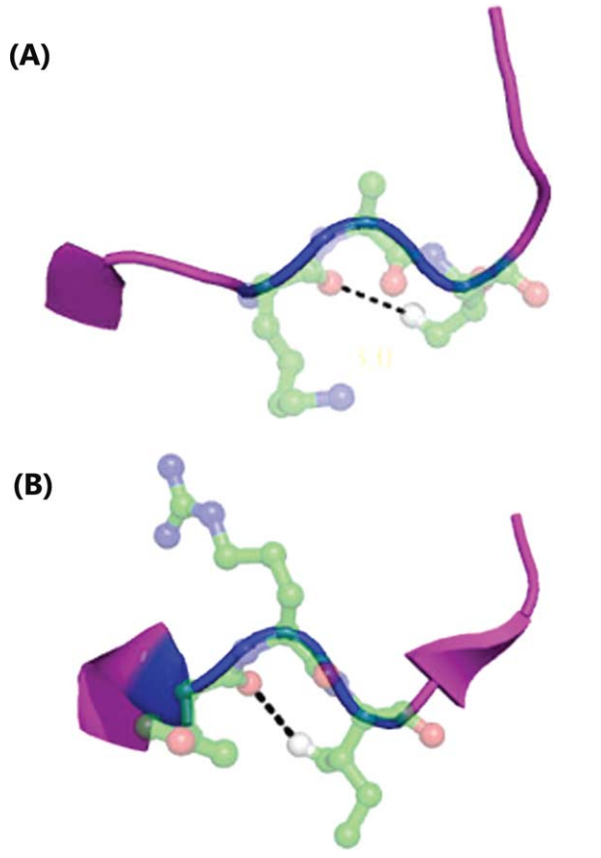

Figure 5

ω turn is shown using two representative structures, (A) sequence (488)Lys-Ala-Thr in blue with flanking turn regions (magenta) in PDB file 1KQF, and (B) (52)Tyr-Arg-Ile between a helix and a β strand in 1J7G. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

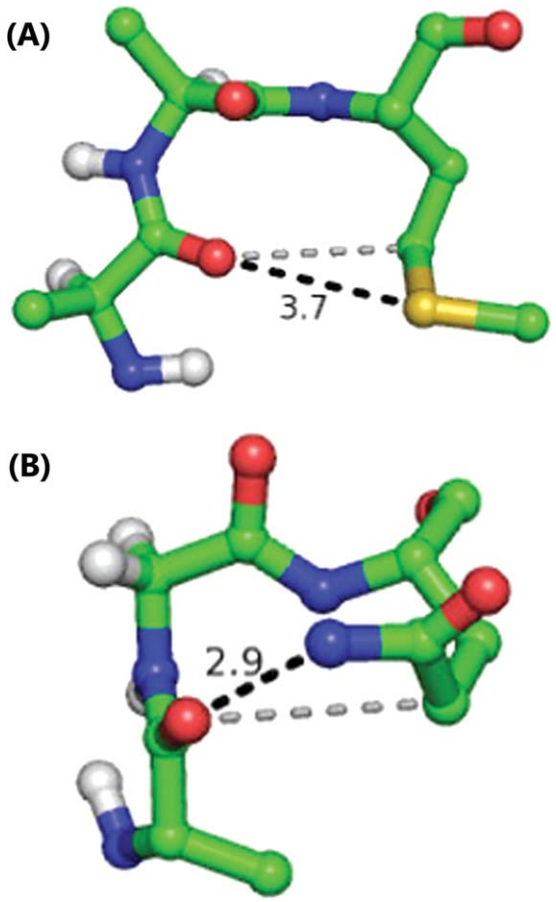

Figure 6

Examples of Met and Gln occurring at $i+2$, and having an additional interaction involving the side chain with the carbonyl group at *i*. (A) Sequence (292)Gln-Trp-Met in PDB file 1CCW, and (B) (209)Thr-Gly-Gln in 1AE9. The additional S···O interaction or hydrogen bonding is shown in bold dashed line (with distance in Å), whereas the typical C—H···O interaction is in grey shade. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

can be seen in Supporting Information Table S3, not only for Cys occurring at $i+1$ but also for any other residue at $i+2$ (in particular when it is a hydrophobic residue).

ω-turns can be important in the context of asparagine (N)-linked glycosylation that results in the covalent attachment of an oligosaccharide onto Asn residues of polypeptide chains and can occur in all domains of life.[62] The glycosylation occurs before the nascent polypeptide has completely folded, indicating that the native protein structure has no role in the recognition of the peptide/protein substrate by oligosaccharyltransferase which catalyzes the transfer of an oligosaccharide to the Asn side chain—the only requirement being the presence of the residue in the consensus sequence, Asn-Xnp-Thr/Ser (where Xnp is a nonproline residue). In conformity with this pattern we observe that Asn and Thr have high propensity to occur at i and $i+2$ positions, respectively (Table I and Supporting Information Fig. S2), in ω-turns, and indeed this tripeptide sequence is observed in 189 structures (out of 3411 cases having Thr at $i+2$ position) – 68% having Type II conformation, and the rest belonging to Type I′. Although it has been suggested that the peptide sequence that undergoes glycosylation adopts an Asx-turn,[63] it is quite likely that the possibility to adopt the ω-turn conformation is the driving force that led to the selection of the consensus sequence as the recognition element. Support for the role of ω-turn also stems from the fact that many nonconsensus sequences that are also glycosylated have Val,[64] which is again an important residue at $i+2$ position of the ω-turn.

Although Ser and Thr are O-glycosylated, there is no apparent sequence signature for such residues[65]; an enhanced occurrence of Pro is observed at -1 and $+3$ positions relative to the glycosylated residue,[66,67] which need to be surface accessible.[68] The amphiphilic Thr side-chains in ω-turns are generally solvent exposed (discussed later), and interestingly, for Type II turns Pro is the preferred residue preceding Thr [Fig. 7(A)]. The Thr $O^{\gamma}H$ group is not engaged in any interaction within ω-turn, and should be amenable for O-glycosylation.

Location of ω-turn relative to the antigenic determinant sites in proteins

It has been shown that many short peptides that are structured in aqueous environment are immunogenic

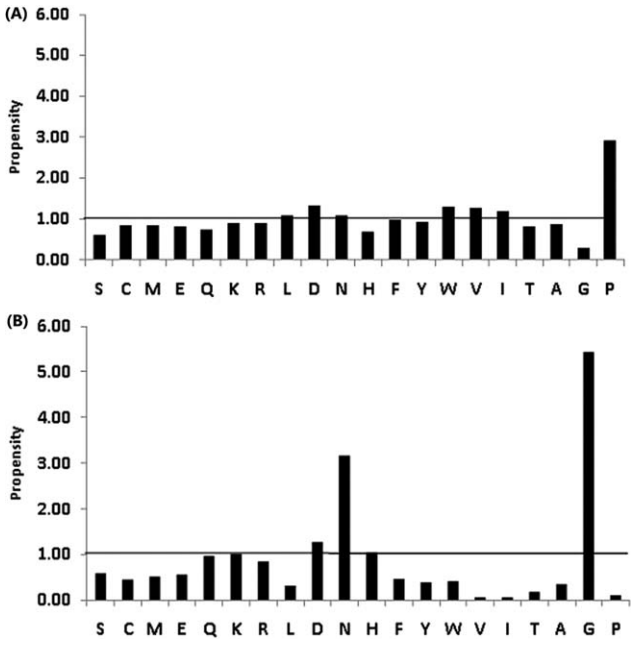

Figure 7

Propensities of residues to occur at $i+1$ position dividing all the ω-turns into two groups, types (A) IIa + IIb and (B) I'a + I'b.

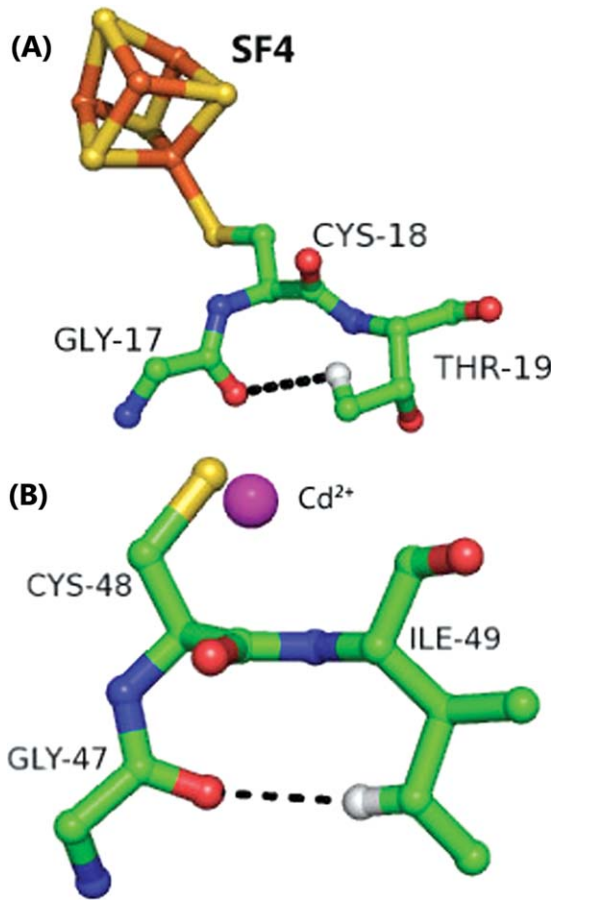

Figure 8

Examples showing the importance of ω-turns in protein function with Cys at $i+1$. (A) (17)Gly-Cys-Thr (PDB, 1GNL), hybrid cluster protein from *Desulfovibrio desulfurican*. The iron-sulfur cluster, $(Fe_4S_4)^{2+}$, SF4 is also coordinated by three other Cys residues at position 6, 9, and 24. (B) (47)Gly-Cys-Ile (PDB, 4MT2), rat metallothionein-2 . The Cd^{2+} ion also binds two other Cys residues at position 33 and 34. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

and these have β-turns as the characteristic feature of the antigenic determinant or epitope site on proteins from which the peptides were extracted.[69] Thus the structural studies on the hypervariable region (V3) of human immunodeficiency virus Type 1 (HIV-1) with its corresponding Fab fragment[70,71] and the solution NMR results of the V3 region of HIV-1[72] suggest that the epitope region lies near the conserved sequence (Gly-Pro-Gly-Arg) which samples a Type II β-turn. Experimental and computer simulations investigations on a short peptide sequence (Pro-Asp-Thr-Arg-Pro) from the antigenic determinant site of MUC1 mucin glycoprotein suggest that this peptide significantly takes up a Type I β-turn.[73,74] The structure of the 11-residue peptide epitope of Type 1 poliovirus in complex with Fab fragment of its antibody shows two Type I β-turns in tandem.[75] A recent neural network study on different conformational propensities of amino acids suggests that Chou's β-turn values are strong determinants for the prediction of linear B-cell epitopes.[76] As the β-turns are known to play important role in peptide immunogenicity we wanted to examine the manifestation of such a possibility with ω-turn.

Except for Ile the vast majority of other residues at $i+2$ appear to be exposed (Supporting Information Fig. S4), as expected when a residue belongs to an antigenic site. As the vast majority of B-cell epitopes are discontinuous,[52] we investigated if such sites are found in or near ω-turns. It is observed (Fig. 9) that Thr has the highest % of cases (44%) with the residue being located within two residues of the antigenic site; considering Thr, Val and Ile together 79% are found within six residues, indicating thereby that these residues when located in ω-turn can be identified as epitopes.

CONCLUSIONS

In conclusion, we have characterized ω-turn that is stabilized by a main-chain to side-chain $C_i{=}O\cdots H{-}C^{\gamma}_{i+2}$ type intramolecular interaction, instead of the main-chain to main-chain conventional $C_i{=}O\cdots H{-}N_{i+3}$ hydrogen bond seen in typical β-turn, both encompassing a noncovalent 10-membered ring (Fig. 1). As hydrogen positions are important in the definition of the turn, we depict a ω-turn with hydrogen coordinates determined using neutron diffraction in Supporting Information Figure S5. As a mimic of β-turn, the Asx-turn reproduces the hydrogen bond interaction involving the Asx side chain; however, the conformation defined by this turn does not really lead to the chain reversal as seen in β-turn. The ω-turn on the other hand does indeed cause a chain reversal and has been found in the metal-binding locations, N-linked glycosylation motif and can constitute the antigenic determinant site. Thr has a high propensity to occur in such a turn; though it

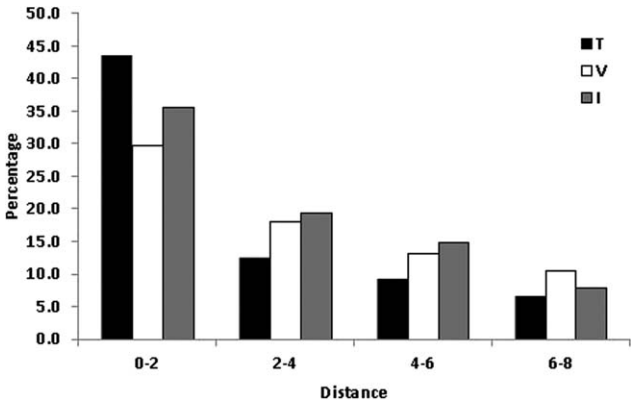

Figure 9

Distance (in terms of residues) of antigenic sites from the $i+2$ position in ω-turns [(considering only type IIb turns (Table IIA)].

has both C—H and O—H groups in its side chain it is the former that is used predominantly in the formation of intraturn hydrogen bond. Along this line the hydroxyl-containing Ser is not a favored residue to occur in such a turn. ω-turn is yet another manifestation of weak C—H···O interaction in protein structures. The interaction is placed like a crown under which the peptide chain reverses its direction. As these are also seen in short peptide structures, it is likely that these are formed as transient, local secondary structural elements during folding pathways. Like the role played by C—H···O interactions in stabilizing helix and its termini in globular[27,33] and transmembrane proteins,[37] β-sheets,[24,28] collagen triple-helical stability,[23] this work highlights the importance of weak C—H···O interactions in the context of turn conformation. Being a widely occurring motif in protein structures and because of the flat topology, ω-turns should be important for protein and peptide design.

It may be mentioned that the capital Greek symbol, Ω has been used in connection with Ω-loops,[77,78] where the symbol stands for a loop of ~6–16 residues connecting two regular secondary structural elements. There is no conformational preference in such loops, and no generality in the number and occurrence of hydrogen bonds within them. The small Greek symbol, ω has been used by us to specify the shape of the three-residue β-turn mimic, with a specific pattern of hydrogen bond, identified here.

Although our analysis involves only globular proteins it may be mentioned that ω-turns may also occur in membrane proteins. A cursory study revealed the occurrence of such a turn in the structure of halorhodopsin (Supporting Information Fig. S6), an archeal rhodopsin from *Halobacterium salinarum* that uses light energy to pump chloride through biological membranes.[79] The backbone torsion angles: $\phi_{i+1} = -72^\circ$, $\psi_{i+1} = 132^\circ$, $\phi_{i+2} = -132^\circ$, $\psi_{i+2} = -10^\circ$, across the Thr-Thr segment, indicate the turn to be of Type IIb. The presence of the turn at the edge of the lipid bilayer may be functionally relevant.

ACKNOWLEDGMENT

We are grateful to Prof. E.J. Milner-White for useful discussion.

REFERENCES

1. Baker EN, Hubbard RE. Hydrogen bonding in globular proteins. Prog Biophys Mol Biol 1984;44:97–179.
2. Golovin A, Henrick K. MSDmotif: exploring protein sites and motifs. BMC Bioinform 2008;9:312.
3. Torrance GM, Leader DP, Gilbert DR, Milner-White EJ. A novel main chain motif in proteins bridged by cationic groups: the niche. J Mol Biol 2009;385:1076–1086.
4. Venkatachalam CM. Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. Biopolymers 1968;6:1425–1436.
5. Lewis PN, Momany FA, Scheraga H. Chain reversals in proteins. Biochim Biophys Acta 1973;303:211–229.
6. Richardson JS. The anatomy and taxonomy of protein structure. Adv Protein Chem 1981;34:167–339.
7. Rose GD, Gierasch LM, Smith JA. Turns in peptides and proteins. Adv Protein Chem 1985;37:1–109.
8. Hutchinson EG, Thornton JM. A revised set of potentials for β-turn formation in proteins. Protein Sci 1994;3:2207–2216.
9. Chakrabarti P, Pal D. The interrelationships of side-chain and main-chain conformations in proteins. Prog Biophys Mol Biol 2001;76:1–102.
10. Tainer JA, Getzoff ED, Beem KM, Richardson JS, Richardson DC. Determination and analysis of the 2 Å structure of copper zinc superoxide dismutase. J Mol Biol 1982;160:181–217.
11. Rees DC, Lewis M, Lipscomb WN. Refined crystal structure of carboxypeptidase A at 1.54 Å resolution. J Mol Biol 1983;168:367–387.
12. Eswar N, Ramakrishnan C. Secondary structures without backbone: an analysis of backbone mimicry by polar side-chains protein structures. Protein Eng 1999;12:447–455.
13. Eswar N, Ramakrishnan C. Deterministic features of side-chain to main-chain hydrogen bonds in globular protein structures. Protein Eng 2000;13:227–238.
14. Wilson DR, Finlay BB. The 'Asx-Pro turn' as a local structural motif stabilized by alternative patterns of hydrogen bonds and a consensus-derived model of the sequence Asn-Pro-Asn. Protein Eng 1997;10:519–529.
15. Wan WY, Milner-White EJ. A natural grouping of motifs with an aspartate or asparagine residue forming two hydrogen bonds to residues ahead in sequence: their occurrence at α-helical N-termini and in other situations. J Mol Biol 1999;286:1633–1649.
16. Wan WY, Milner-White EJ. A recurring two-hydrogen-bond motif incorporating a serine or threonine residue is found both at alpha-helical N termini and in other situations. J Mol Biol 1999;286:1633–1649.
17. Vijayakumar M, Quian H, Zhou HZ. Hydrogen bonds between short polar side chains and peptide backbone: prevalence in proteins and effects on helix-forming propensities. Proteins 1999;34:497–507.
18. Duddy WJ, Nissink JWM, Allen F, Milner-White EJ. Mimicry by asx- and ST-turns of the four main types of β-turn in proteins. Protein Sci 2004;13:3051–3055.
19. Taylor R, Kennard O. Crystallographic evidence for the existence of C−H···O, C−H···N and C−H···Cl hydrogen bonds. J Am Chem Soc 1982;104:5063–5070.
20. Desiraju GR. The C–H···O hydrogen bond: structural implications and supramolecular design. Acc Chem Res 1996;29:441–449.
21. Desiraju GR, Steiner T. The weak hydrogen bond in structural chemistry and biology. New York: Oxford University Press; 1999. pp 10–25.
22. Derewenda ZS, Lee L, Derewenda U. The occurrence of C–H···O hydrogen bonds in proteins. J Mol Biol 1995;252:248–262.
23. Bella J, Berman HM. Crystallographic evidence for C^{α}-H···O=C hydrogen bonds in a collagen triple helix. J Mol Biol 1996;264:734–742.
24. Fabiola GF, Krishnaswamy S, Nagarajan V, Pattabhi V. C−H···O hydrogen bonds in β-sheets. Acta Crystallogr 1997;53:316–320.
25. Chakrabarti P, Chakrabarti S. C−H···O hydrogen bond involving proline residues in α-helices. J Mol Biol 1998;284:867–873.
26. Ash EL, Sudmeier JL, Day RM, Vincent M, Torchilin EV, Haddad KC, Bradshaw EM, Sanford DG, Bachovchin WW. Unusal ^{1}H NMR chemical shifts support (His) Cε1-H···O=C H-bond: proposal for reaction-driven ring flip mechanism in serine protease catalysis. Proc Natl Acad Sci USA 2000;97:10371–10376.
27. Babu MM, Singh SK, Balaram P. A C−H···O hydrogen bond stabilized polypeptide chain reversal motif at the C-terminus of helices in proteins. J Mol Biol 2002;322:871–880.

28. Ho BK, Curmi PMG. Twist and shear in β-sheets and β-ribbons. J Mol Biol 2002;317:291–308.
29. Klaholz BP, Moras D. C–H···O hydrogen bonds in the nuclear receptor RARγ – a potential tool for drug selectivity. Structure 2002;10:1197–1204.
30. Pierce AC, Sandretto KL, Bemis GW. Kinase inhibitors and the case for C–H···O hydrogen bonds in protein-ligand binding. Proteins 2002;49:567–576.
31. Jiang L, Lai L. C–H···O hydrogen bonds at proten-protein interfaces. J Biol Chem 2002;277:37732–37740.
32. Singh SK, Babu MM, Balaram P. Registering α-helices and β-strands using backbone C–H···O interactions. Proteins: Struct Func Bioinf 2003;51:167–171.
33. Manikandan K, Ramakumar S. The occurrence of C–H···O hydrogen bonds in α-helices and helix termini in globular proteins. Proteins 2004;56:768–781.
34. Saha RP, Bhattacharyya R, Chakrabarti P. Interaction geometry involving planar groups in protein-protein interfaces. Proteins 2007; 67:84–97.
35. Chakrabarti P, Bhattacharyya R. Geometry of nonbonded interactions involving planar groups in proteins. Prog Biophys Mol Biol 2007;95:83–137.
36. Senes A, Ubarretxena-Belandia I, Ebgelman DM. The C^{α}-H···O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. Proc Natl Acad Sci USA 2001;98: 9056–9061.
37. Loll B, Raszewski G, Saenger W, Biesiadka J. Functional role of C^{α}-H···O hydrogen bonds between transmembrane α-helices in photosystem I. J Mol Biol 2003;328:737–747.
38. Nishio M, Umezawa Y, Fantini J, Weiss MS, Chakrabarti P. The CH-π hydrogen bonds in biological macromolecules. Phys Chem Chem Phys 2014;16:12648–12683.
39. Pal D, Chakrabarti P. Cis peptide bonds in proteins: residues involved, their conformations, interactions and locations. J Mol Biol 1999;294:271–288.
40. Ashish A, Kishore R. 1H NMR spectroscopic signatures of an intra-residue hydrogen bonded C5-structure. Tetrahedron Lett 1997;38: 2767–2770.
41. Thakur AK, Ashish G, Kishore R. Novel hydrogen bonding ring motifs in a model peptide: crystal and molecular conformation. Chem Commun 1999;17:1643–1644.
42. Thakur AK, Kishore R. Crystallographic characterization of novel β-turn like folds in a model peptide: stabilization by main-chain to side-chain interactions. Tetrahedron Lett 2001;42:4691–4694.
43. Thakur AK, Kishore R. Characterization of a β-turn and Asx-turns *mimicry* in a model peptide: stabilization *via* C−H···O interaction. Biopolymers 2006;81:440–449.
44. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
45. Wang G, Dunbrack RL Jr. PISCES: recent improvements to a PDB sequence culling server. Nucleic Acids Res 2005;33:W94–W98.
46. Word JM, Lovell SC, Richardson JS, Richardson DC. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. J Mol Biol 1999;285: 1735–1747.
47. Feig M, Karanicolas J, Brooks III CL. MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. J Mol Graph Model 2004; 22: 377–395.
48. de Hoon MJL, Imoto1 S, Nolan2 J, Miyano S. Open source clustering software. Bioinformatics 2004;20:1453–1454.
49. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Bioplymers 1983;22:2577–2637.
50. De Lano WL. The PyMOL molecular graphics system. 2002. Available at: http://www.pymol.org.
51. Hubbard SJ. NACCESS: Program for calculating accessibilities. 1992. http://www.bioinf.manchester.ac.uk/naccess/.
52. Andersen PH, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. Protein Sci 2006;15:2558–2567.
53. Guharoy M, Chakrabarti P. Conservation and relative importance of residues across protein-protein interfaces. Proc Natl Acad Sci USA 2005;102:15447–15452.
54. Sander C, Schneider R. Database of homology-derived protein structures and the structural meaning of sequence alignment. Proteins 1991;9:56–68.
55. Musah RA, Jensen GM, Rosenfeld RJ, McRee DE, Goodin DB. Variation in strength of an unconventional CH to O hydrogen bond in an engineered protein cavity. J Am Chem Soc 1997;119:9083–9084.
56. Steiner T, Saenger W. Role of C–H···O hydrogen bonds in the coordination of water molecules. Analysis of neutron diffraction data. J Am Chem Soc 1993;115:4540–4547.
57. Wieczorek R, Dannenberg JJ. H-bonding cooperativity and energetic of α-helix formation of five 17-amino acids peptides. J Am Chem Soc 20031;25:8124–8129.
58. Vargas R, Garza J, Dixon DA, Hay BP. How strong is the C^{α}-H···O=C hydrogen bond? J Am Chem Soc 2000;122:4750–4755.
59. Scheiner S, Kar T, Gu Y. Strength of the C^{α}−H···O hydrogen bond of amino acid residues. J Biol Chem 2001;276:9832–9837.
60. Pal D, Chakrabarti P. Non-hydrogen bond interactions involving the methionine sulfur atom. J Biomol Struct Dyn 2001;19:115–128.
61. Bhattacharyya R, Pal D, Chakrabarti P. Disulfide bonds, their stereospecific environment and conservation in protein structures. Protein Eng Des Sel 2004;17:795–808.
62. Schwarz F, Markus A. Mechanisms and principles of N-linked protein glycosylation. Curr Opin Struct Biol 2011;21:576–582.
63. Imperiali B, O'Connor SE, Hendrickson T, Kellenberger C. Chemistry and biology of asparagin-linked glycosylation. Pure Applm Chem 1999;71:777–787.
64. Zielinska DF, Gnad F, Wiśniewski JR, Mann M. Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. Cell 2010;141:897–907.
65. Steen PVD, Rudd PM, Dwek RA, Opdenakker G. Concepts and principles of O-linked glycosylation. Crit. Rev Biochem Mol Biol 1998;33:151–208.
66. Wilson IB, Gavel Y, Von Heijne G. Amino acid distributions around O-linked glycosylation sites. Biochem J 1991;275:529–534.
67. Nishikawa I, Nakajima Y, Ito M, Fukuchi S, Homma K, Nishikawa K. Computational prediction of O-linked glycosylation sites that preferentially map on intrinsically disordered regions of extracellular proteins. Int J Mol Sci 2010;11:4991–5008.
68. Hansen JE, Lund O, Tolstrup N, Gooley AA, Williams KL, Brunak, S. NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. Glycoconj J 1998;15: 115–130.
69. Dyson HJ, Wright PE. Antigenic peptides. FASEB J. 1995;9:37–42.
70. Rini JM, Stanfield RL, Stura EA, Salinas PA, Profy AT, Wilson IA. Crystal structure of a human immunodeficiency virus type 1 neutralizing antibody, 50.1, in complex with its V3 loop peptide antigen. Proc Natl Acad Sci USA 1993;90:6325–6329.
71. Stanfield R, et al. Dual conformations for the HIV-1 gp120 V3 loop in complexes with different neutralizing fabs. Structure 1999;7:131–142.
72. Chandrasekhar K, Profy AT, Dyson HJ. Solution conformational preferences of immunogenic peptides derived from the principal neutralizing determinant of the HIV-1 envelope glycoprotein gp120. Biochemistry 1991;30:9187–9194.
73. Scanlon MJ, Morley SD, Jackson DE, Price MR, Tendler SJ. Structural and computational investigations of the conformation of

antigenic peptide fragments of human polymorphic epithelial mucin. Biochem J. 1992;284:137–144.
74. Schuman J, Campbell AP, Koganty RR, Longenecker BM. Probing the conformational and dynamical effects of O-glycosylation within the immunodominant region of a MUC1 peptide tumor antigen. J Pept Res 2003;61:91–108.
75. Wien MW, Filman DJ, Stura EA, Guillot S, Delpeyroux F, Crainic R, Hogle JM. Structure of the complex between the Fab fragment of a neutralizing antibody for type 1 poliovirus and its viral epitope. Nature Struct Biol 1995;2:232–243.
76. Su CH, Pal NR, Lin KL, Chung IF. Identification of amino acid propensities that are strong determinants of linear B-cell epitope using neural networks. PLoS One 2012;7:e30617.
77. Leszczynski JF, Rose GD. Loops in globular proteins: a novel category of secondary structure. Science 1986;234:849–855.
78. Pal M, Dasgupta S. The nature of the turn in omega loops of proteins. Proteins: Struct Func Bioinf 2003;51:591–606.
79. Kolbe M, Besir H, Essen LO, Oesterhelt D. Structure of the light-driven chloride pump halorhodopsin at 1.8 Å resolution. Science 2000;288:1390–1396.