Protein Engineering, Design & Selection, 2015, vol. 28 no. 6, pp. 153–161
doi: 10.1093/protein/gzv017
Advance Access Publication Date: 13 April 2015
Original Article

OXFORD

Original Article

Defining the loop structures in proteins based on composite β-turn mimics

Jesmita Dhar[1] and Pinak Chakrabarti[1,2,*]

[1]Bioinformatics Centre, Bose Institute, P1/12, CIT Scheme VIIM, Kolkata 700 054, India and [2]Department of Biochemistry, Bose Institute, P1/12, CIT Scheme VIIM, Kolkata 700 054, India

*To whom correspondence should be addressed. E-mail: pinak@boseinst.ernet.in, pinak.chak@gmail.com
Edited by Valerie Daggett

Received 11 December 2014; Revised 26 February 2015; Accepted 26 February 2015

Abstract

Asx- and ω-turns are β-turn mimics, which replace the conventional main-chain hydrogen bonds seen in the latter by those involving the side chains, and both involve three residues. In this paper we analyzed the cases where these turns occur together—side by side, with or without any gap, overlapping and in any order. These composite turns (of length 3–15 residues), occurring at ~1 per 100 residues, may constitute the full length of many loops, and when the residues in the two component turns overlap or are adjacent to each other, the composite may take well-defined shape. It is thus possible for non-regular regions in protein structure to form local structural motifs, akin to the regular geometrical features exhibited by secondary structures. Composites having the order ω-turns followed by Asx-turns can constitute N-terminal helix capping motif. Ternary composite turns (made up of ω-, Asx- and ST-turns), some with characteristic shape, have also been identified. Delineation of composite turns would help in characterizing loops in protein structures, which often have functional roles. Some sequence patterns seen in composites can be used for their incorporation in protein design.

Key words: β-turn mimicry, protein loops, shapes in loops, structural motifs

Introduction

Loops in protein structures are important not only because they interconnect secondary structural elements and change the direction of propagation of the polypeptide chain, but they usually also harbor the active site residues. They are, however, recalcitrant for structural characterization as their residues do not follow any pattern in main-chain torsion angles or hydrogen bonding. This poses problem in generating loop structures in homology modeling (Fiser *et al.*, 2000). Shorter loops, which are usually called turns, have been grouped into classes such as β-turns (Hutchinson and Thornton, 1994), α-turns (Dasgupta *et al.*, 2004), π-turns (Dasgupta and Chakrabarti, 2008), usually in terms of ϕ, ψ angles or hydrogen bonding involving the CO and NH groups of the terminal residues.

In recent years, there have been attempts to increase the scope of turns by including the interaction involving side-chain atoms also. Thus, the Asx-turns use the interaction involving the side-chain carbonyl group of Asx (Asp or Asn) at position *i* with the main-chain NH at *i* + 2 (Richardson, 1981; Duddy *et al.*, 2004). An ST-turn uses the hydroxyl group of Ser or Thr, the relative positions of the interacting groups being the same as in Asx-turn (Eswar and Ramakrishnan, 1999; Wan and Milner-White, 1999). These are β-turn mimics. Very recently, another β-turn mimic has been characterized, called ω-turn, because of its shape being like the Greek symbol (ω) and it involves a non-conventional C–H···O interaction between the main-chain carbonyl group at *i* and the C^{γ}–H group in the side chain of Thr, Val, Ile, Leu, Met, Arg, Glu, Lys or Gln, located at *i* + 2 position (Dhar *et al.*, 2015). With these background material it is now possible to revisit the protein loops to find out if some of these could be defined as composite of β-turn mimics, and like the ω-turn also have well-defined shape/structure that can visually be identified. It may be mentioned that a morphologic definition that focuses on the linearity and planarity of loops has been utilized to reduce the geometric complexity of

loops 4–20 residues long (Ring *et al.*, 1992). A sub-class of general loops has been termed as Ω-loops, where the two ends of the loop are spatially proximal and attached to regular secondary structural elements (Leszczynski and Rose, 1986). Here we analyze the occurrence of ω- and Asx-turns in protein loops in peptide stretch ≤15 residues long, if such loops are bounded by secondary structures, and if any combination of the two turn types leads to any specific geometric shape. In addition to binary composites of ω- and Asx-turns, we have analyzed the ternary composites by including ST-turns, but such combinations become too complicated for a detailed analysis.

At the outset we would like to delineate the convention followed in the work. Both Asx- and ω-turns are three-residue turns and can be represented by the three-letter codes, **A**sx and Om**g**, with '**A**' and '**g**' being in bold to indicate the location of the side chain that is involved in hydrogen bonding in the respective turn. In a short polypeptide stretch, the three residues each in **A**sx and Om**g** can occur adjacent to each other in any order, and with or without any intervening residues. They can also overlap, singly, doubly or triply, and the resulting peptide containing both the turns would be short stretches of lengths 5, 4 and 3, respectively. To convey the relative position of the two turns, the three residues in **A**sx are numbered i to $i+2$, and those in Om**g**, $m-2$ to m, i and m being the positions of the side chains involved in hydrogen bonding in the two cases (Fig. 1). The sequence difference, $\Delta = i - m$, indicates the position of one with respect to the other. Thus, Δ is −1 in the sequence, Om**gA**sx shown in Fig. 1. Using the three-letter codes to represent three consecutive residue positions in turns, some short stretches of composite turns with different Δ values are shown in Table I.

Results and discussion

A total of 28 687, 18 639 and 16 883 ω-, Asx- and ST-turns were observed in our dataset of 4114 non-redundant protein chains. A total of 9347 composite (made up of ω and Asx) turns were observed, which corresponds to 1.00 (±0.8) per 100 residues. Two ω-turns can also constitute composites, almost to the extent as above (10 945 cases in peptide stretches of length ≤15 residues). Likewise, 8277 binary composites of ω- and ST-turns were found. As Asn residues are found in locations close to the ω-turns, and with functional implications (the possibility of getting glycosylated) (Dhar *et al.*, 2015), we considered the composites of ω- and Asx-turns in greater details. The number reduces drastically if ternary composites are considered; in stretches of 4–11 residues (the minimum number was decided based on the maximum possible overlap between the three turn types, and the maximum from the consideration of three residues per turn and one intervening residue between a pair of turns). A total of 1213 such turns were found in 828 (20%) protein chains, of which 240 have multiple occurrences.

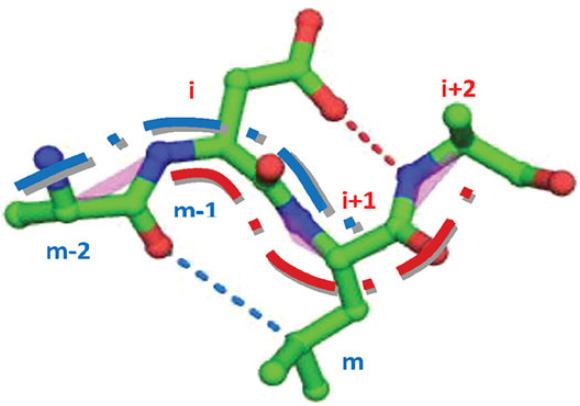

Fig. 1 Schematic representation of the simultaneous occurrence of Asx- and ω-turns. For the former, the residues are labeled i to $i+2$, with the Asx residue being at i, while for the latter the labeling is $m-2$ to m, with the residue at m providing the C^{γ}–H group for the C–H···O interaction. The residue labels and hydrogen bonding in Asx-turn are in red, those in ω-turn are in cyan; the curves indicate the chain direction in the two cases. In this example, the sequence difference, $\Delta = i - m = -1$.

Classification of composite turns depending on the Δ values

The first and the last residue are unique in the three-residue Asx- and ω-turns, respectively. If these are represented by i and m, the relative position of the two turns in a sequence can be specified by $\Delta = i - m$ (Fig. 1 and Table I). In a stretch of 15 residues, Δ can range from −10 to +10, a minus value indicating that the Asx-turn precedes the ω-turn (except when $\Delta = -1$), a plus value indicating the opposite order (Table II). The sequences corresponding to the two turns overlap when Δ is −4 to −1 (the value is −4 with one residue overlapping, −3 or −1 with 2, and −2 when there is complete overlap between the three-residue patches of the two turns). There cannot be any sequence with $\Delta = 0$ (as an Asx residue, obligatory at position i in Asx-turn, cannot be at position m in ω-turn). When there is no overlap the number of the intervening residues varies between 1 and 9.

The numbers of composite turns for each value of Δ with a representative example (with sequence and the residue secondary structures, including those of the flanking residues) are given in Table II. When |Δ| is large, the polypeptide stretch encompassing the composite is large, and some residues are seen being located in regular secondary structural elements. Thus, the % of loops containing such composites is less; the value is <20% if the length of the stretch is ≥11. On the other hand, due to the short length, >90% of the composite turns

Table I. Relative position of the three-residue **A**sx and Om**g** exhibiting Asx- and ω-turns in loops, with the corresponding Δ values and their distinct shape

Δ	Relative position[a]	Shape[b]
−4	**A**sx Om**g**	U
−3	**A**sx Om**g**	S
−2	**A**sx Om**g**	Bulge
−1	Om**g** **A**sx	Da[c]
+1	Om**gA**sx	M
+2	Om**g** – **A**sx	–
+3	Om**g** – – **A**sx	–

[a]The three letters in **A**sx and Om**g** represent three residues that define the two turn types; 'A' and 'g' are in bold to show the uniqueness of these positions in the two turns. A '–' indicates a residue not belonging to the two turns. When there is an overlap of residues in the two turns, **A**sx and Om**g** are shown one over the other. For example, when $\Delta = -3$, the two turns are such that the last two residues in Asx-turn overlaps with the first two in ω-turn.

[b]In addition the ternary composite turn, Asx–ST–Om**g** (Table III) has the shape of mushroom.

[c]'Da' corresponds to the Bengali letter ঢ়.

Table II. Number of composite turns and the percentage of these occurring in loops in peptide stretch 3–15 residues long and having different Δ values

No. of cases (% in loops)[a]	Δ	Representative example				
		PDB chain[c]	Sequence[b]	Structure[b] of turns with flanking residues[d]	Turn type[e] for	
					ω-	Asx-
418 (17.5)	−10	1DC1A	*DSR*dkkytNWM	E*ETT*sttccCEEE	IIa	II′
222 (23)	−9	3PPMA	*DPT*vpplPFR	H*CTT*scccCCCH	IIa	II′
190 (32.1)	−8	1H1NA	*DTD*ney*H*DM	E*ECC*sccCSSCH	IIa	II′
259 (47.5)	−7	1W6SA	*DVK*tgEQV	E*ETT*tcCEEE	IIa	II′
416 (68.3)	−6	3BONA	*NGY*gSTQ	ECCCCSSCCTTT*SS*CcCCEE	IIa	II′
514 (79.2)	−5	3H51A	*DKN*GKK	EC*TT*SCEE	IIa	II′
544 (94.9)	−4	1I4UA	*NP*YQL	ECC*SS*CCE	IIb	II′
285 (97.2)	−3	1P3DA	*N*NER	ECT*TCC*EE	IIa	I′
354 (96.6)	−2	1T92A	**DGV**	ECCS**SSC**CE	I′a	I′
1063 (94.5)	−1	1HDHA	RDQ*S*	HT**S**C**TT**SCE	IIb	II′
906 (66)	+1	1O22A	PKE*NAV*	EC*CCS*C*CTT*CCSSSCCCCCSCE	IIa	II′
629 (52.6)	+2	2ZM9A	EGRi*DPA*	HTTSc*CTT*SB	IIa	II′
574 (35)	+3	1D5TA	TVEtt*DPE*	E*ECC*ss*CHH*H	IIa	II′
534 (28.3)	+4	2QW5A	GATrtf*DPS*	EC*CCC*sss*CTT*CSSH	IIa	II′
446 (20.9)	+5	1XQOA	DYVpnle*DLG*	HT*TCC*cctt*CHH*H	IIa	II′
417 (15.8)	+6	3N0UA	GVEdlail*DKH*	HTT*TCC*scccc*CHH*H	IIa	II′
386 (14)	+7	1EEXA	DNMfagsne*DAE*	G*GCT*tsccss*CGG*G	IIa	II′
381 (9.4)	+8	2BFFA	YRIghhstsd*DSS*	EC*CCC*scssttc*CGG*G	IIa	II′
294 (6.5)	+9	1FT5A	AKLdpakdytq*DKD*	H*TTC*cttcccttC*TT*TG	IIa	II′
515 (6.6)	+10	1M55A	GQIwelppesdl*NLT*	H*HCC*ccccttccc*CGG*G	IIa	II′

Each category is exemplified by a representative sequence from a PDB file, showing the secondary structures of all the residues in the stretch along with those of the flanking residues. The individual turn classes of ω-, and Asx-turns are indicated.

[a]In loops there is no intervening regular secondary structural element (helix or strand) between the two turns.

[b]Residues in Asx- and ω-turns are in italics and underlined, respectively. Overlapping residues are in bold; intervening residues between the two 3-residue turns are in lowercase. A color version of this table is available as Supplementary data at *PEDS* Online.

[c]Indicated by 4-lettered PDB codes followed by chain identifier.

[d]The flanking residues are shown till a residue with the secondary structure of H or E is found.

[e]The details of torsion angles are given in Table S1.

Table III. The ternary composite turns categorized based on the order of the individual turns (Asx, ω and ST)

Order of individual turns (number)	1–2					2–3				
	G0	G1	O1	O2	O3	G0	G1	O1	O2	O3
Asx–Omg–ST (273)	46	59	101	37	30	85	57	86	45	0
Asx–ST–Omg (283)	198	26	48	11	0	14	18	64	44	143
Omg–Asx–ST (171)	50	41	0	80	0	71	29	56	15	0
Omg–ST–Asx (146)	45	29	44	28	0	36	39	64	70	0
ST–Asx–Omg (146)	28	15	92	11	0	13	11	27	15	80
ST–Omg–Asx (194)	37	33	45	51	28	60	52	0	82	0

Each category is split into cases depending on the presence of gaps (0 or 1, indicated by G0 or G1) or overlap of (1, 2 or 3 residues, indicated by O1, O2 or O3) between the first two (1–2) or the last two (2–3) individual turns.

with overlapping residues are located within a single loop. Of all the composites, the one with Δ = −1 occurs most often, and Fig. 1 shows a schematic representation of such a turn with the two central residues overlapping. As the two turns in composites can occur in any order (Table I), it is instructive to find out if any order is preferred. For example, a six-residue stretch (with the two turns being adjacent to each other) occurs for both Δ = +1 and −5, and the former occurs 1.8 times more than the latter. Thus, the sequence Om**g**A**s**x is preferred to **A**s**x**O**m**g. Again considering the cases with Δ of +2 and −6, corresponding to seven-residue stretch of composite turns with one intervening residue, the former (Om**g** – **A**s**x**, the convention as specified in Table I) occurs 1.5 times more than the latter (**A**s**x** – Om**g**). Even with the component residues overlapping, the cases with ω-turn proceeding the Asx-turn are more frequent (e.g. cases with Δ = −1 vs. those with −3, exhibit the highest discrimination, 3.7 times). The reason why ω-turn has a tendency to precede Asx-turn may be because such composites constitute a capping motif at helix N-termini (discussed later).

Shapes of composite turns

Loops in protein structures are most variable in nature. Loops may be classified as (i) short loop (≤10 amino acids long), connecting two adjacent regular secondary structures and (ii) long loop (>10 amino acids

long) which connects two distant secondary structures (Martin *et al.*, 1995). In our analysis, 51% of the total (9347) composite turns are located within a loop and out of them 94% can be categorized as short. We asked the question if these short loops have any specific shape. Figure 2A and B and Table I show the four-residue turn with $\Delta = -1$ looks like a Bengali letter 'ঘ'; it is as if the backbone has taken a detour, with the two hydrogen bonds in the composite being almost coplanar, with the central peptide group being perpendicular to it. This structure is likely to be stable as it has the highest numbers (Table II). The residues constituting the ω- and Asx-turns fully overlap when $\Delta = -2$, and result in a bulge shape, usually occurring between two β-strands (Fig. 2C and D). The two hydrogen bonds cross each other. When $\Delta = -3$, the four-residue turn takes the shape of 'S' with the two hydrogen bonds being nearly aligned (Fig. 2E and F). There may be some distortion of the S-shape and the motif may deviate from planarity. Superposition of the five-residue turns ($\Delta = -4$) indicates two limiting orientations (Fig. 3A and B), with the first four residues taking up the 'U' shape. ω- and Asx-turns are adjacent to each other when $\Delta = +1$, and the shape resembles the letter 'M' (which may have some distortion) (Fig. 3C and D). As in $\Delta = +1$, $\Delta = -5$ also has the two turns side by side, but in the reverse order. But these do not exhibit any particular shape.

Although DSSP identifies many residues as being irregular (not belonging to any regular secondary structure), many of these, as part of composite β-turns, may not really be irregular. Thus, if we consider $\Delta = -5$ to $+1$ cases (overlapping or adjacent ω- and Asx-turns), 37% of the residues in the composite turns have 'irregular' conformation (according to DSSP), but may actually be part of a motif with well-defined shape.

Secondary structural features around composite turns

When in loop how far are the flanking secondary structures from the two ends of the composite turn? The answer can be found in Fig. 4. A large number of cases with very small sequence differences (46% of the total are found for 0–0, 0–1, 1–0) suggest that a secondary structure starts almost immediately, or even from the terminal one or two residues at each end. The trend is very similar irrespective of the order

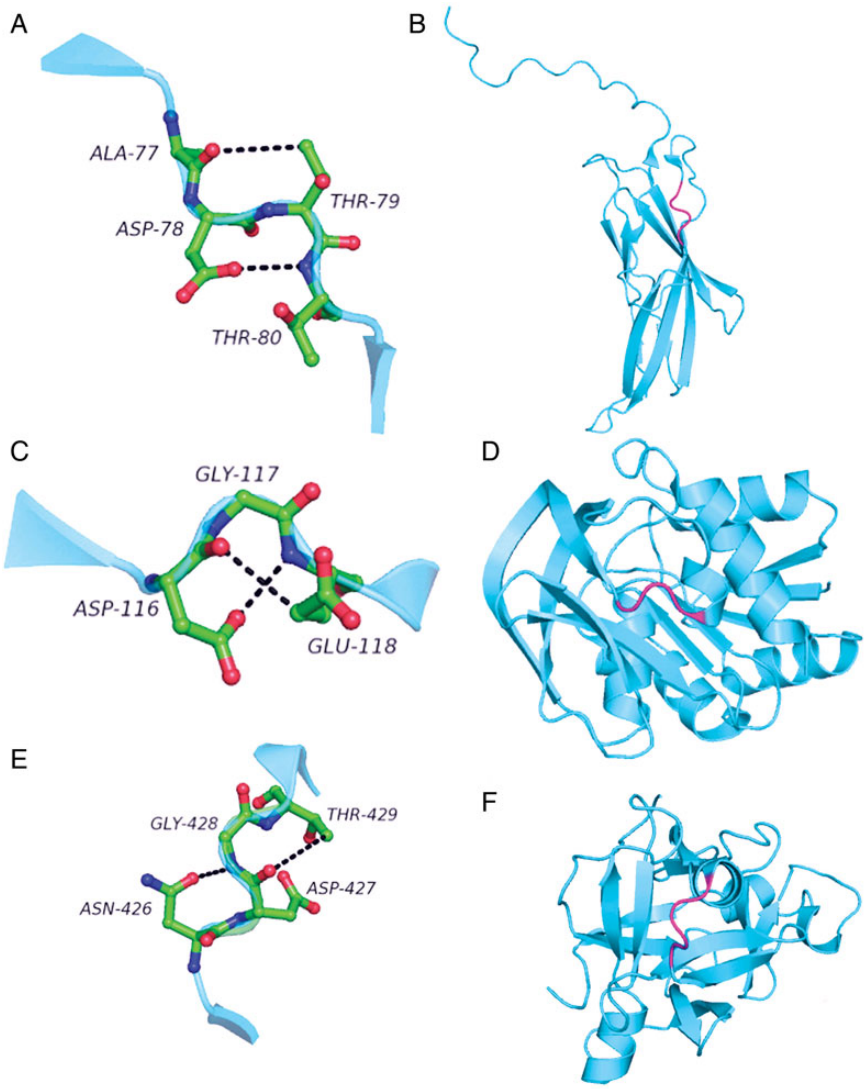

Fig. 2 Some examples of composite turns having distinct shape and with Δ values of (**A**) −1, (**C**) −2 and (**E**) −3. The tertiary structures (or domain) where these (in darker shade) are located are shown in (**B**) for PDB file 3KLQ, (**D**) for 1I7Q and (**F**) for 1WD3, respectively. The broken lines indicate hydrogen bonds. A color version of this figure is available as Supplementary data at *PEDS* Online.

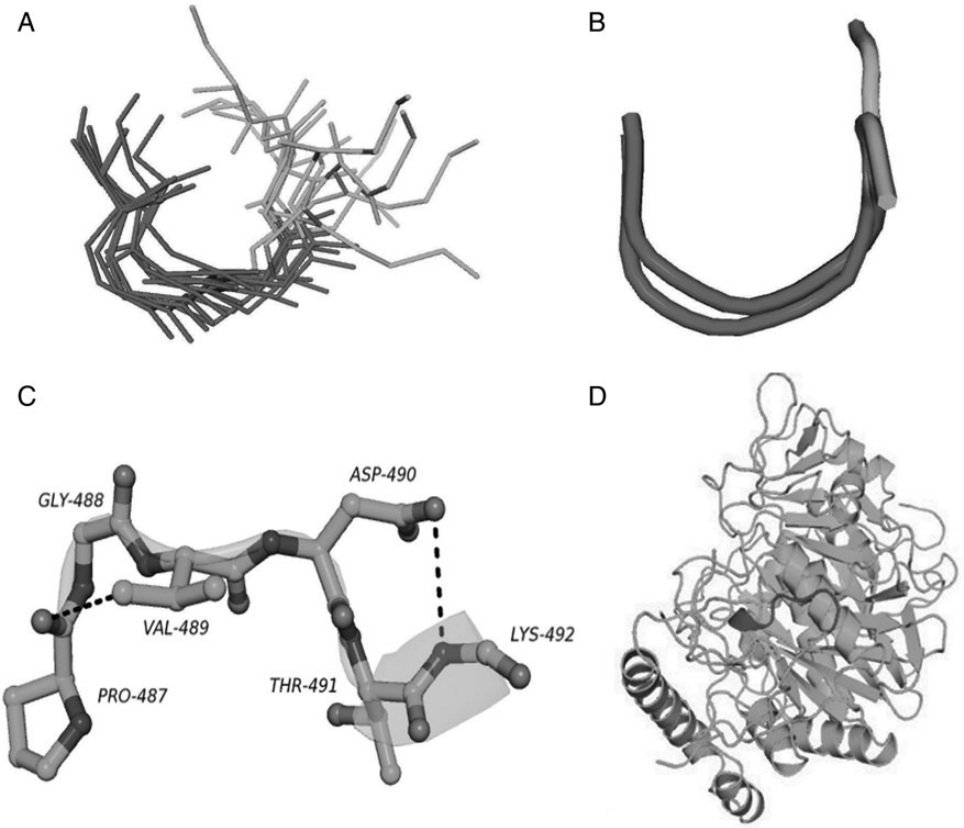

Fig. 3 Gross shape of composite turns with (**A**, **B**) Δ = −4 and (**C**, **D**) Δ = +1. (B) is the cartoon diagram (the Asx-turn is in darker shade) representing two limiting orientations corresponding to (A) the superimposed structures of (10 randomly selected) composite turns (shown in stick model). The two files used in (B) are 1P3C [(177) Asp–Gln–Asn–Gln–Gln] and 2J6G [(20) Asp–Tyr–Arg–Gln–Lys], and the one in (C) and (D) is 1AYL. A color version of this figure is available as Supplementary data at *PEDS* Online.

of the component turns within the composite. However, as the figure does not discriminate between the categories based on the Δ value, we made Supplementary Fig. S1 corresponding to cases with Δ = −1. The similarity between the two plots indicates that the conclusion derived from Fig. 4 is applicable to individual Δ values also. To understand if any of the two structural elements [helix (H) or β-strand (E)] is preferred, it would be instructive to first know the preference of secondary structures around ω- and Asx-turns individually. It has been reported that ω-turn is mainly found between two β-strands (Dhar *et al.*, 2015), whereas the most common occurrence of Asx-turn is between a strand and a helix. Looking at the nearest H or E on either side, E is found in higher number (~78% cases) at the upstream region of the Asx-turn, whereas it is H (62%) at the downstream region. Restricting to only three residues on either side of the Asp/Asn residue in the turn, the discrimination at the two ends becomes more prominent—84% E upstream and 81% H in the downstream positions. The occurrence of helix immediately following Asp or Asn is the consequence of these residues being involved in capping at the N-terminal end of helix (Richardson and Richardson, 1988; Aurora and Rose, 1998; Wan and Milner-White, 1999; Leader and Milner-White, 2010).

If we divide the composite turns into two categories, Asx-turn followed by ω-turn (these would include all turns with Δ = −10 to −2, Table II) and the other with the opposite order (Δ = −1, or +1 to +10), we find that in the former (represented by Asx_Omg in Fig. 5), the most common occurrence is in between two β-strands. When the order is Omg_Asx, the composite usually precedes a helix as expected from the preference of Asp and Asn residues to occur at the N-terminal region of α-helices. Thus, these composite turns can be considered to constitute an extended N-terminal capping motif, an example of which can be seen in Fig. 3C and D. The secondary structure on the other side can be either E or H (Fig. 5A); this can also be seen clearly in Fig. 5B (where the two ends of the composite turn are part of the flanking secondary structural elements) and Supplementary Fig. S2, where the results for only the cases with Δ = −1 have been displayed.

Conformational features of composite turns

The conformation of each of the ω- and Asx-turns can be defined based on the torsion angles at two positions. The average , angles at four positions for composite turns (with different Δ values) are provided in Supplementary Table SI. The turns can be further categorized into sub-classes (in conformity with the nomenclature followed for β-turns) based on these angles (Duddy *et al.*, 2004; Dhar *et al.*, 2015). For ω-turns, the major types are IIa and IIb, and the minor types are I′ and I. For Asx-turns, the frequency of occurrence is type II′ (major) > type I > type II > type I′. From Supplementary Table SI, one can see that both the types usually belong to the major type of ω- or Asx-turns; for ω-turn it is either type IIa (mainly) or type IIb (for the cases with Δ = −1 and −4) and for Asx-turn it is mainly type II′. In a few cases, one of the turn type is of minor category; Δ = −5 has type I′a for ω-turn, and Δ = +7 and −3 exhibit types II and I′, respectively, for Asx-turn. When Δ = −2 (complete overlap of ω- and Asx-turns, and the same residue corresponds to the positions

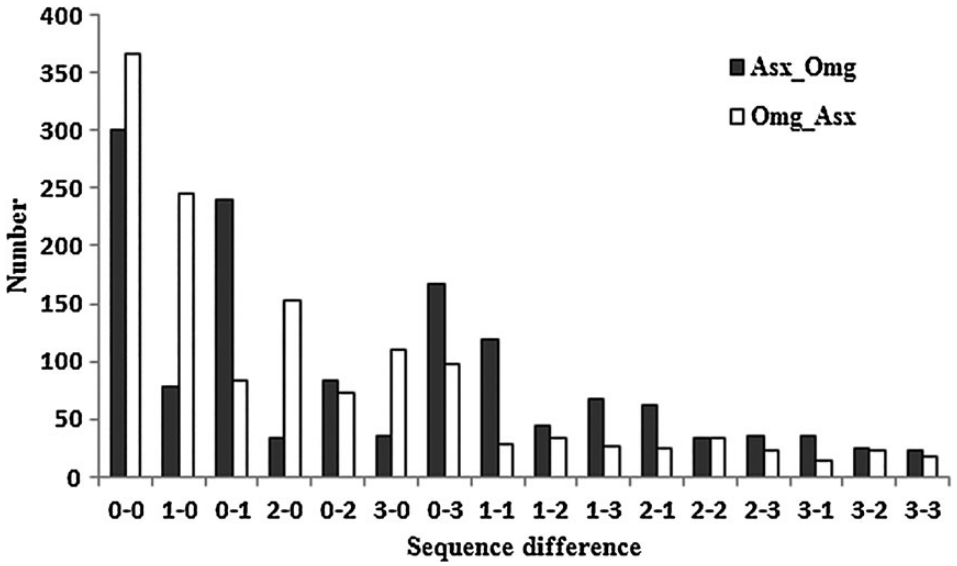

Fig. 4 Histogram showing the number of intervening residues between the two ends of the composite turn and the nearest secondary structural element (H or E); to indicate directionality cases have been split into two categories, one with the Asx-turn followed by ω-turns, and the other with the reverse order. A sequence difference of $x - y$ indicates that 'x' is the distance of the nearest H or E from the ω-turn, whereas 'y' is for the secondary structure at the other end from the Asx-turn. For a six-residue stretch (e.g. OmgAsx), a value of '0–0' would indicate that one structural element had continued up to the position 'O' or 'm' in the sequence, and the other had started from 's' or 'x'. Because of scarce observations, data beyond 3–3 are not included.

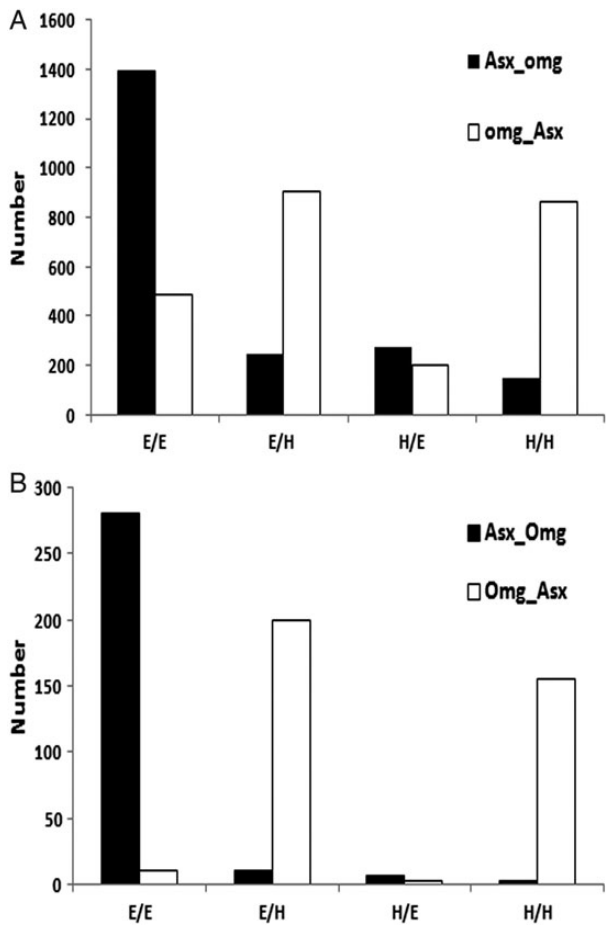

Fig. 5 Combination of secondary structures on either side of composite turns, separated into two groups based on the order of occurrence of ω- and Asx-turns. In (**A**), all the loops made up of composite turns are considered, whereas in (**B**) are the cases with the sequence difference of '0–0' (as given in Fig. 4).

$m - 1$ of the former and $i + 1$ of the latter) both the types are from the minor categories (types I′a and I′, respectively).

Composite turn consisting of three β-turn mimics: ω-, Asx- and ST-turns

If we also include ST-turns in the composites, the situation gets complicated as the three constituent turns may be in any order (six possibilities), with or without overlapping residues, or different gaps in sequence between them. We restricted ourselves to only one intervening residue. Also it may be noted that the complete overlap between Asx- and ST-turns is not possible, as these are defined by unique residues (Asn/Asp or Ser/Thr), which are not interchangeable. Considering the possibility of all the overlaps and one intervening residue, the composite turns we analyzed may vary in length from 4 to 11 residues. A total of 1213 composite turns were identified. Following the procedure given in 'Materials and methods' section, 854 (70%) were found to constitute loops. A total of 607 (50%) were found to have a regular secondary structure located within three residues on either side of the composites, the ends of which are usually encompassed within the structural element (as indicated by the peak at '0–0' in Supplementary Fig. S3). Such ternary composites, irrespective of the order of occurrence of individual turns, are mainly found between β-strands (Supplementary Fig. S4).

As shown in Table III, among all the possibilities, Asx–ST–Omg is the most abundant (283 cases). Among these there are 112 cases (40%), where the Asx-turn is immediately followed by completely overlapping ST- and ω-turns. Figure 6 shows such a ternary composite turn. Interestingly, this six-residue turn takes the characteristic shape of a mushroom.

Position-wise preferences of residues within composite turns

The position-wise probability value of a residue was calculated for all composite turns with different Δ values to identify any pattern in the sequence (Supplementary Fig. S5). If we consider the ω- and Asx-turns

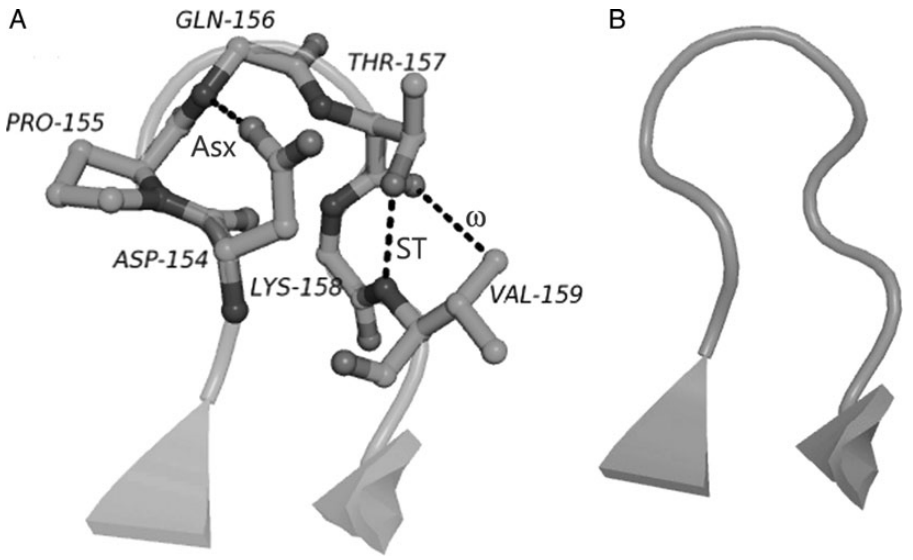

Fig. 6 Example of the most abundant ternary composite turn (Asx–ST–Omg, Table III) with no intervening residue between the first two turns and complete overlap between ST- and ω-turns. Hydrogen bonds (broken lines) are shown in (**A**) for the sequence, (154) Asp–Pro–Gln–Thr–Lys–Val (159), located between two β-strands, (**B**) taken from the file, 1UCD. A color version of this figure is available as Supplementary data at *PEDS* Online.

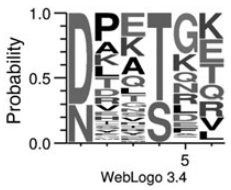

Fig. 7 Sequence pattern of ternary composite turns (Asx–ST–Omg, considering 112 cases satisfying the conditions given in Fig. 6). A color version of this figure is available as Supplementary data at *PEDS* Online.

individually and specify their position labels as given in Fig. 1, we find a higher level of occurrence of Asp (D) than Asn (N) at position i in Asx-turn. Pro seems to be the most favored residue at $i+1$ in all cases, except with $\Delta = -3$, -2 or -1, where there is overlap (Table I) between ω- and Asx-turns. Gly is seen in greater numbers in the first two cases, but is forbidden to occupy the position (which is the position m in the ω-turn, and needs to be occupied by a residue with a C^{γ}–H group) in the last case. Looking at the next position ($i+2$), although Ser is forbidden here (because of the aforementioned reason) when $\Delta = -2$, with all other Δ values the probability of Ser is very high (either of the top two positions; exceptions $\Delta = -5$ or -4). The trend is less clear with ω-turns, except some preference for Val at position m, and Gly at $m-1$. Of all the Δ values, $\Delta = -2$ (when the residues in the two turns overlap completely) has the clearest pattern of sequence, the most probable being Asp–Gly–Lys. Similarly, the ternary composites (considered in Fig. 6) also give rise to a clear pattern (Fig. 7). In the stretch of six residues, the first three correspond to Asx-turn, and the next three have completely overlapping ST- and ω-turns. Akin to the case mentioned above, the last two positions of the overlapping region are occupied by Gly–Lys, preceded by Thr (or Ser) whose side chain is involved in hydrogen bonding in the β-turn mimic. This category of ternary composites resembles the binary composite with $\Delta = -5$, where the Asx- and ω-turns are adjacent to each other, and the patterns of residues making the Asx-turn are similar in the two cases.

Retention of composite turns among SCOP families

We analyzed if the composite turn occurs among the different members of the same protein family—for this the four-residue composite turns with $\Delta = -1$, which have the highest number of occurrences (Table II), were considered. However, the different levels of filtering (Supplementary Fig. S6) reduced the number of domains that could be analyzed to only 12, the structure-based sequence alignment of which (from PALI) is given in Supplementary Table SII. We checked if the composite turn occurs in the matched sequences and that all of them constitute loops in the respective structures. In 58% domains, the composite turn is found to occur in >50% of the aligned structures (Table IV). In 25% more domains only the Asx-turn can be seen, but the ω-turn is missing, as also the typical '↯' shape of the composite turn.

Materials and methods

Atomic coordinates were obtained from the Protein Data Bank (PDB) (version March 2013) (Berman *et al.*, 2000) at the Research Collaboratory for Structural Bioinformatics (RCSB). A non-redundant dataset of 4114 protein chains (in 3976 X-ray structures) was selected (Dhar *et al.*, 2015) with *R*-factor ≤20%, resolution ≤2.0 Å and the sequence identity between any pair <25%.

Identification of composite of β-turn mimics

We searched the dataset for the three-residue ω-turns, which were characterized based on the hydrogen bond between the side-chain C^{γ}–H group of residue (Thr, Val, Ile, Leu, Met, Arg, Glu, Lys or Gln) at position m and the main-chain carbonyl group of residue at $m-2$, such that $d_{C\cdots O} \leq 4.0$ Å (Fig. 1). Likewise, the Asx-turns were identified based on the presence of hydrogen bond ($d_{O\cdots N} \leq 3.5$ Å) between the side-chain carbonyl group of Asx (Asn or Asp) at position i and the main-chain NH group of residue at position $i+2$; the average O···N distance was found to be 3.08 ± 0.27 Å. ST-turns were also identified where the side-chain hydroxyl group of Ser or Thr was involved in similar hydrogen bonding. We then identified the composite

Table IV. Statistics on the occurrence of composite turns (with $\Delta = -1$) across SCOP protein families

Number of domains	Average number of aligned sequences	Number of domains with % retention of the composite turn in the range		
		≥90%	70–90%	50–70%
12	5.6 (±3.4)	2	1	4

(binary) turns by finding out if Asx- and ω-turns occur together in a peptide stretch of length ≤15 residues. The distance (in terms of residue), Δ between the first residue (at *i*) of Asx-turn and the last residue (at *m*) of ω-turn, was calculated as $\Delta = i - m$ (Fig. 1 and Table I). Ternary composite turns containing all the three turns (overlapping, adjacent or with one intervening residue between any pair of turns) in peptide stretches of length ≤11 residues were also identified.

Identification of loops containing composite turns

Secondary structural information of the residues presents in and around ω-turns and Asx-turns were identified using DSSP (Kabsch and Sander, 1983); structural designations G (for 3_{10}-helix), H (α-helix) and I (π-helix) were grouped as helix (H); B (β-sheet) and E (extended strand) as β-strand (E); T (turn) and S (bend) as turn (T); and the remaining cases as belonging to irregular region (C) in structure. We next identified how many of the composite turns are located in loops (bounded by H or E). Depending on the value of Δ, the two types of turns in a composite may overlap (singly, doubly or triply) have different order of occurrence (Asx followed by Omg, or vice versa) with different number of sequences in between (Table I). For consistency, we had to consider the value of 'Δ' while defining the loop. For cases with $\Delta = -1$ or -3, corresponding to a four-residue stretch of composite turns, one (or both) of the two central residues would have to be a residue with non-regular structure. $\Delta = -2$ corresponds to a three-residue stretch (complete overlap of two turn types), and the central residue should be of non-regular structure. For all other Δ values, the stretch would be at least five-residue long, and the two residues at each end were allowed to have a regular secondary structure (either H or E), leaving one (or more) residues in the center with irregular structure. Thus, the loops that are essentially defined by the composite turns were identified.

We also identified the sub-classes of the individual turns in the composites on the basis of the torsion angles within the hydrogen-bonded turn. This needed the (derived) ϕ_e, ψ_e angles of residue at *i* position in Asx-turn, and (the standard) ϕ, ψ angles at $i + 1$ of Asx-turn, and both at $m - 1$ and *m* positions of ω-turn; for the Asx-turn, the derived angles are defined as $\phi_e = \chi_1 - 120°$ and $\psi_e = \psi + 120°$, respectively (Duddy *et al.*, 2004).

Position-wise preferences of residues in composite turns

The residue preferences for each position of the composite turns were determined from sequence logos made using WebLogo 3 server (Crooks *et al.*, 2004). PDB files were used to extract the sequence of the polypeptide fragments corresponding to a given value of Δ. The sequences were merged to generate a 'multiple sequence alignment' file in FASTA format, which was then used as input to the server. A logo plot represents each column of the alignment by a stack of letters where the height of each letter is directly proportional to the observed frequency of the corresponding amino acid at that particular position. The shape of specific composite turns for different values of Δ was visualized, and the molecular diagrams made using PyMol (DeLano, 2002).

Analysis of the occurrence of composite turns across members in a protein family

To study if the composite turns (only a representative class with $\Delta = -1$ was considered) are retained among members in a protein family, we used SCOPe (The Structural Classification of Proteins) 2.04 database to identify the domains containing the composite turns (Fox *et al.*, 2014). The non-redundant protein domains (with <40% identity to each other) corresponding to each of these domains were identified using Astral database 2.04 (Chandonia *et al.*, 2004). The structural alignment files were obtained from PALI server (Balaji *et al.*, 2001). Only the domains/sequences common to both Astral and PALI could be used further. In majority of the cases, the aligned region had gaps for most of the entries, which were excluded. Only those domains with three or more protein chains aligned without gap in the region corresponding to the composite turn were analyzed.

Conclusion

Protein loops are enriched in residues, such as Asp and Asn, which are known to be involved in Asx-turns, Gln, which is abundant in ω-turns, and Ser, which can participate in ST-turns and form local hydrogen bonds in loops (Dasgupta *et al.*, 2014). As such, our aim of this paper was to see if these three turn motifs can occur together to define the structure of protein loops. Although the ternary turns have been identified (Table III and Fig. 6), the situation is rather complicated because of the different possibilities of the order in which the individual turns can occur, whether they overlap (wholly or partially), if they are contiguous, or occur with intervening residues. As such detailed analysis was done only for binary composites consisting of ω- and Asx-turns. The relative position of the two three-residue turns in composites is specified by $\Delta = i - m$, where *i* and *m* are the positions of the residues whose side chains are involved in hydrogen bonding in Asx- and ω-turns, respectively (Fig. 1 and Table I). The composites were distinguished based on their Δ values (Table II), and some of the categories, especially where the component turns overlap, take up well-defined shapes (Table I, Figs 2 and 3). Protein structures contain many short loops, which are essentially composite turns (Fig. 4). There are some preferences in the secondary structural milieu within which the composite turns are found (Fig. 5). When the order of the individual turns is Asx-turn, followed by ω-turn, the composite is usually bound between two β-strands. With the reverse order, the composite is usually followed by α-helix. These can have positional preferences for amino-acid residues analogous to those seen in β-turns (Fig. 7 and Supplementary Fig. S4), and these can be used in protein/peptide design. ω-turns are found in active sites (Dhar *et al.*, 2015). There are indications that the composite turns may be retained across members in protein families (Table IV), and the delineation of composite turns would help in modeling protein loops having functional implications.

Supplementary data

Supplementary data are available at *PEDS* online.

Acknowledgements

We are thankful to Dr R. Kishore with whom the work on ω-turn was initiated, and to Prof. George D. Rose for discussion.

Funding

P.C. is supported by JC Bose National Fellowship.

References

Aurora,R. and Rose,G.D. (1998) *Protein Sci.*, **7**, 21–38.

Balaji,S., Sujatha,S., Kumar,S.S.C. and Srinivasan,N. (2001) *Nucleic Acids Res.*, **29**, 61–65.

Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) *Nucleic Acids Res.*, **28**, 235–242.

Chandonia,J.M., Hon,G., Walker,N.S., Lo Conte,L., Koehl,P., Levitt,M. and Brenner,S.E. (2004) *Nucleic Acids Res.*, **32**, D189–D192.

Crooks,G.E., Hon,G., Chandonia,J.M. and Brenner,S.E. (2004) *Genome Res.*, **14**, 1188–1190.

Dasgupta,B. and Chakrabarti,P. (2008) *BMC Struct. Biol.*, **8**, 39.

Dasgupta,B., Pal,L., Basu,G. and Chakrabarti,P. (2004) *Proteins*, **55**, 305–315.

Dasgupta,B., Dey,S. and Chakrabarti,P. (2014) *Biopolymers*, **101**, 441–453.

DeLano,W.L. (2002). The PyMOL molecular graphics system. http://www.pymol.org.

Dhar,J., Chakrabarti,P., Saini,H., Raghava,G.P.S. and Kishore,R. (2015) *Proteins*, **83**, 203–214.

Duddy,W.J., Nissink,J.W.M., Allen,F. and Milner-White,E.J. (2004) *Protein Sci.*, **13**, 3051–3055.

Eswar,N. and Ramakrishnan,C. (1999) *Protein Eng.*, **12**, 447–455.

Fiser,A., Do,R.K. and Sali,A. (2000) *Protein Sci.*, **9**, 1753–1773.

Fox,N.K., Brenner,S.E. and Chandonia,J.M. (2014) *Nucleic Acids Res.*, **42**, D304–D309.

Hutchinson,E.G. and Thornton,J.M. (1994) *Protein Sci.*, **3**, 2207–2216.

Kabsch,W. and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637.

Leader,D.P. and Milner-White,E.J. (2010) *Proteins*, **79**, 1010–1019.

Leszczynski,J.F. and Rose,G.D. (1986) *Science*, **234**, 849–855.

Martin,A.C.R., Toda,K., Stirk,H.J. and Thornton,J.M. (1995) Protein Eng., 8, **1093–1101**.

Richardson,J.S. (1981) *Adv. Protein Chem.*, **34**, 167–339.

Richardson,J.S. and Richardson,D.C. (1988) *Science*, **240**, 1648–1652.

Ring,C.S., Kneller,D.G., Langridge,R. and Cohen,F.E. (1992) *J. Mol. Biol.*, **224**, 685–699.

Wan,W.Y. and Milner-White,E.J. (1999) *J. Mol. Biol.*, **286**, 1633–1649.