ExploreTurns: A web tool for the exploration and analysis of beta turns and their contexts in proteins; application to beta-bulge and Schellman loops, Asx helix caps, and other hydrogen-bonded motifs

Nicholas E. Newell, Independent Researcher, 21 Parkview Road, Reading, MA 01867, USA
nenewell@comcast.net

Abstract

The most common type of protein secondary structure after the alpha helix and beta sheet is the four-residue beta turn, which plays many key structural and functional roles. Existing tools for the study of beta turns operate almost exclusively in backbone dihedral-angle (Ramachandran) space, which presents challenges for the visualization, comparison and analysis of the wide range of turn conformations. A recent study has introduced a turn-local Euclidean-space coordinate system and global alignment for turns, along with geometric parameters describing their backbone shapes, and these features and others are incorporated here into ExploreTurns, a web facility for the exploration, analysis, geometric tuning and retrieval of beta turns and their structural contexts which combines the advantages of Ramachandran- and Euclidean-space representations. Due to the widespread occurrence of beta turns in proteins, this facility, supported by its interpreter for a new general nomenclature for short H-bonded loops, can serve as an exploratory browser and analysis tool for most loop structure. ExploreTurns is applied here to detect new H-bonded loops, including a "short Schellman loop" and a large family of motifs satisfying a generalized definition of the beta-bulge loop, map Asx N-cap sequence preferences, profile Schellman loop/beta-turn conformations, and investigate the depth dependence of turn geometry. The tool, available at www.betaturn.com, should prove useful in research, education, and applications such as protein design, in which an enhanced Euclidean-space picture of turn/motif structure and the ability to identify and tune structures suited to particular requirements may improve performance.

Keywords

Beta turn, beta-bulge loop, Schellman loop, helix cap, protein structure, motif detection, sequence motif, H-bonded loop, compound-turn nomenclature, short Schellman loop, beta bracket, pi-alpha loop, beta-gamma loop

Abbreviations

AA: Amino acid

BB: Protein backbone

SC: Amino acid side-chain

Position index: Specifies residue positions in the turn and its four-residue N-/C-terminal BB neighborhoods according to the key: -4-3-2-1<1234>+1+2+3+4, where the brackets delineate the turn.

BB H-bonds: BB H-bonds are specified directly in the beta-turn frame by indicating the donor and acceptor positions, separated by '>'. For example, "4>-1" specifies an H-bond donated by the BB NH group at turn residue 4 and accepted by the BB carbonyl oxygen of the residue just before the turn. BB H-bonds can also be specified in the frame of a BB loop, using the compound-turn notation (Supplementary Section 2).

AA sequence motif: AA sequence motifs are labelled with the single-letter codes of their included AAs, each followed by its position in the turn frame. For example, "D-1T3" specifies Asp just before the turn and Thr at the third turn position.

PDB address: PDB addresses for structures are given in the format ABCD_X_N, where ABCD is the 4-character PDB file name, X is the chain letter, and N is the position in the chain of the first turn residue.

BBL: Beta-bulge loop

1 | Introduction

Of the five types of tight turns[1] in proteins, the four-residue beta turn, first described by Venkatachalam[2] in 1968, is by far the most common, and it represents the most common type of protein secondary structure after the alpha helix and beta sheet. Due to the wide variety of backbone (BB) conformations and side-chain (SC) interactions they exhibit, beta turns play key roles in multiple contexts in proteins; for an overview, see *Roles of beta turns in proteins* at www.betaturn.com.

Since their identification, beta turns have been the subject of a large body of work that has evolved their definition and classification (see brief reviews in[3,4]). The beta-turn definition used here[3] describes a four-residue BB segment in which the alpha carbons of the first and last residues are separated by no more than 7Å and the two central turn residues lie outside helices or strands; in about half of these structures the fourth turn residue donates a BB H-bond to the first residue. Three beta-turn classification systems are employed here, each specifying BB geometry at a different level of precision in the Ramachandran space of BB dihedral angles; in increasing order of precision these are referred to as the "classical" types[5], BB clusters[3] and Ramachandran types[3,6,7] (see the ExploreTurns online help for definitions).

As far as the author is aware, the only tools now available that specifically support the exploration, analysis, and retrieval of beta turns are the web facility Motivated Proteins[8] and its associated standalone application Structure Motivator[9], both of which employ a Ramachandran-space description of turn structure and treat turns as one of multiple small H-bonded motifs. The derivation, in a recent study by the present author[10], of a common local Euclidean-space coordinate system and global alignment for beta turns, together with a set of geometric parameters that describe their bulk BB shapes, enabled the development of ExploreTurns, which combines the advantages of Ramachandran- and Euclidean-space representations in support of the exploration of turns and their associated structures in a redundancy-screened, PDB-derived database.

ExploreTurns' turn-local coordinate system provides a common framework for the visualization and comparison of the full range of turn BB geometries, and also supports analyses of the many recurrent contexts that incorporate turns, including local H-bonded motifs, ligand-binding sites and supersecondary structures. The tool's incorporation of geometric turn parameters enhances the Euclidean-space representation of turns, characterizing meaningful modes of structural variation not explicitly captured by Ramachandran-space classification systems and complementing those systems[10]. Parameters enable structural discrimination within classical types and BB clusters, yielding major improvements in the specificity of measurements of

sequence preferences in turns and enabling the tuning of turn geometry for compatibility with key SC interactions and contexts.

Since beta turns constitute close to two-thirds of all residues in loops[3], and ExploreTurns also encompasses the four-residue N-/C-terminal BB "tails" of turns, the tool gives access to most loop structure in proteins; in addition to beta turns, turns of all lengths can be explored, along with the "compound turns" that constitute short H-bonded BB loops. ExploreTurns includes an interpreter for a new compound-turn nomenclature that allows the facility to serve as a general browser and analysis tool for H-bonded loops, and this feature is applied here to explore the structures of loop motifs and their variants and detect new motifs, including short and long variants of the Schellman loop and a large family of structures which satisfy a generalized definition of the beta-bulge loop[11] proposed here and exhibit a range of geometries and H-bonding topologies.

2 | Tool description

2.1 General description

ExploreTurns is a database selection screen, structure profiler and graphical browser, integrated onto a single web page, which renders a redundancy-screened, PDB[12]-derived dataset of 102,192 beta turns structurally addressable by a wide range of criteria describing the turns and their four-residue BB tails. Criteria include classical or Ramachandran turn type, BB cluster, inclusionary/exclusionary patterns of BB H-bonds, ranges of the geometric parameters ("span", N- and C-terminal "half-spans", "bulge", "aspect", "skew", and "warp"), BB dihedral angles, the *cis/trans* status of the turn's peptide bonds, sequence motif content, the electrostatic energy of the turn's potential 4>1 BB H-bond, DSSP[13] secondary structure codes (and extensions), the approximate orientations of the tails with respect to the turn, and the depth of the turn beneath the protein's solvent-accessible surface. H-bond patterns can be specified either directly (in the position frame of the beta turn), in the frame of a BB loop via the compound-turn notation, or via descriptive shorthand names for H-bonded motifs.

ExploreTurns supports the comprehensive exploration and analysis of structure and interaction in beta turns and their BB contexts and the identification, tuning and retrieval of structures tailored to suit particular requirements. Browsing with the tool reveals the characteristic geometries of the turn types and BB clusters and the conformational variation within types and clusters. With the guidance of the tool's built-in parameter distribution and motif overrepresentation plots, browsing also reveals the modes of variation that correspond to the geometric parameters, and the user can

select structures with BB geometries compatible with a particular SC interaction, by adjusting parameter ranges to maximize the fractional overrepresentation and/or abundance of the sequence motif associated with the interaction in the selected set. Geometric tuning also supports the selection of turns suitable for particular structural contexts, and the context can itself be tuned (using the "context vectors") to select desired approximate tail vs. turn orientations or identify the orientations that optimize interactions, such as helix-capping H-bonds, that occur between turns and adjacent structures.

ExploreTurns enables the analysis and retrieval of sets of examples of any BB motif that can be defined by the tool's selection criteria operating on a twelve-residue window centered on a beta burn, including many types of beta hairpins[14,15], supersecondary structures in which beta turns link helices/strands into particular geometries, some beta bulges[16] and nests[17], and all types of short H-bonded loops known to the author, including the alpha-beta loop[8], beta-bulge loops[11], crown bridge loop[18], BB mediated helix caps such as the Schellman loop[19,20,21,22] and its variants, and others (Supplementary Section 2).

The tool also gives access to examples of sequence motifs defined by any combination of amino acids (AAs) at any position(s) in the turn/tails, including motifs which commonly form characteristic structures involving SC/BB or SC/SC H-bonds. The most common SC interactions include Asx/ST turns/motifs[23,24] and Asx/ST helix caps[20,22,23,24], but turns and their BB contexts host many recurrent SC structures that can be identified with the aid of ExploreTurns' motif detection tools, which rank sequence motifs by fractional overrepresentation or statistical significance within the selected set and include options that focus the search on particular residue positions or position ranges. In addition to H-bonds, interactions associated with recurrent SC structures in turns/tails include salt bridges, hydrophobic interactions, a wide range of aromatic-Pro and pi-stacking interactions, and cation-pi and pi-(peptide bond) relationships.

ExploreTurns supports the further investigation of SC motifs by providing direct access to the motif maps generated by the MapTurns[25] tool (available at www.betaturn.com), which enable the comprehensive exploration of the BB and SC structure, H-bonding and contexts associated with single-AA, pair, and many triplet sequence motifs in turns and their BB neighborhoods.

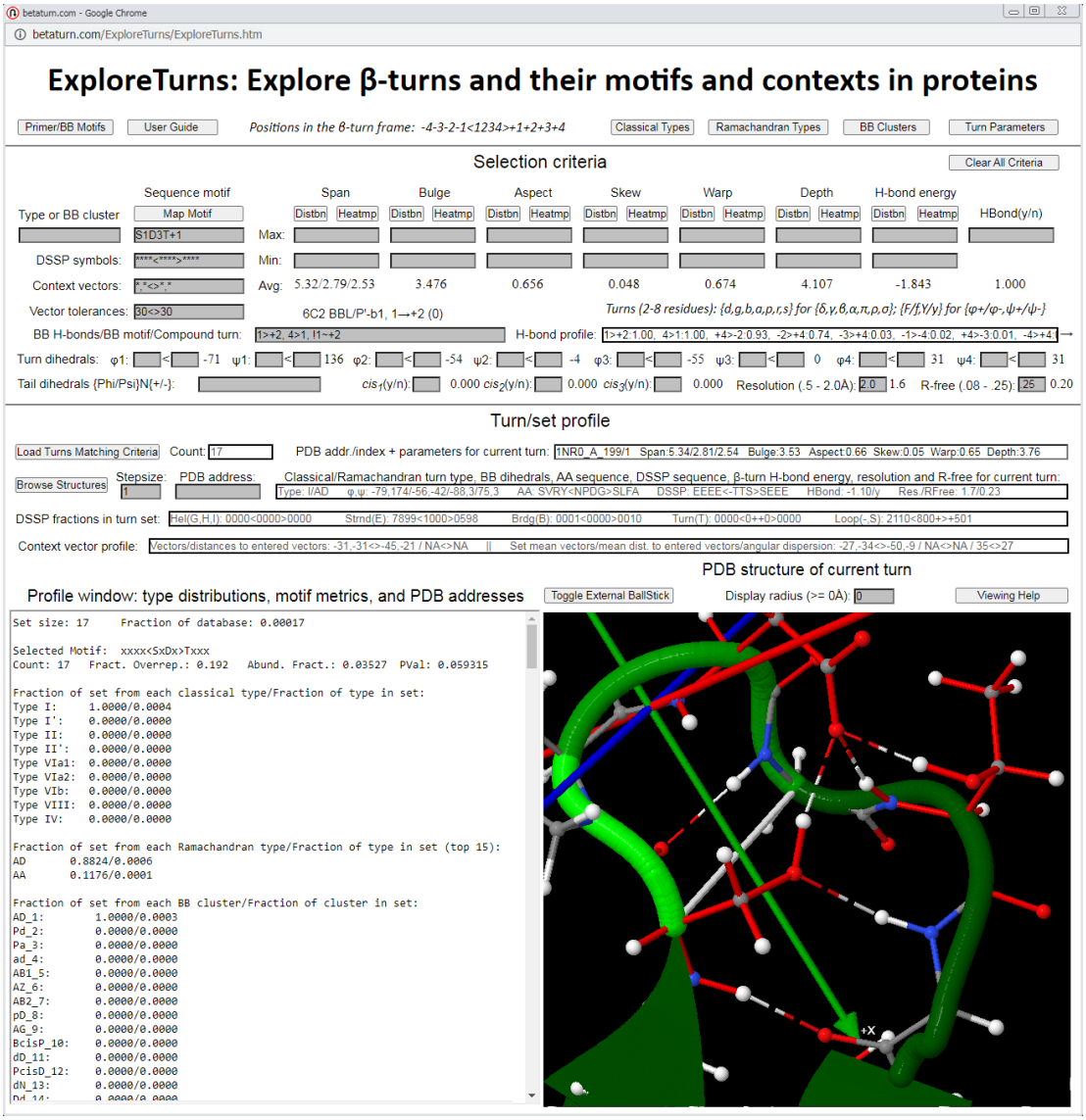

Figure 1. The ExploreTurns screen. Selection criteria have been entered for the sequence motif specifying Ser at turn position 1, Asp at position 3 and Thr just after the turn in the 6C2 beta-bulge loop (BBL), one of the generalized BBL types described here. This SC motif, which mediates a network of H-bonds linking the loop's turn to its "bulge", is specified with the ***Sequence motif*** entry *S1D3T+1*. The ***BB H-bonds/BB motif/Compound turn*** entry *1>+2, 4>1, !1~+2*, which was auto-filled by ExploreTurns to replace the shorthand "6C2" originally entered, specifies the BBL's defining H-bond pair in the β-turn's position frame, while excluding structures with additional BB H-bonds involving the loop (*!1~+2*). The structure is displayed in the turn-local coordinate system, with the turn's $C_{\alpha 1} \rightarrow C_{\alpha 4}$ span shown as a white bar and the motif's SCs highlighted in red. The radius for display of structure outside the turn and its four-residue BB "tails" is set to zero to remove clutter. The text displayed above the ***BB H-bonds...*** box, *6C2 BBL/P'-b1, 1→+2 (0)*, gives the loop's shorthand name, its (Latinized) label in the "compound-turn" nomenclature for short H-bonded loops defined here, its position range in the β-turn's frame, and its position offset from the β-turn. H-bonded loops can be selected using either direct H-bond entry in the frame of the β-turn, the compound-turn notation in the frame of the loop, or descriptive motif shorthands; when one of the latter two options is chosen, the BB motif must be loaded first, before any other criterion (such as a sequence motif) is specified, since the loop's position in the turn frame is not in general known until it is loaded.

2.2 The ExploreTurns screen

The top half of the ExploreTurns web page (Figure 1) is divided vertically into three sections. The six buttons at the top of the screen provide access to extensive online help: a primer introduces the tool and provides examples with usage notes and a catalog of BB motifs, while a comprehensive user guide documents all features of the application, and buttons also give access to documentation for turn types, BB clusters and turn parameters.

The *Selection criteria* section in the middle of the top half of the page is a database selection screen which accepts input in the shaded boxes, and average values of numerical criteria for structures in the selected set are displayed beneath or to the right of these boxes once a set of structures is loaded. Also displayed, in the ***H-bond profile*** box, are the ranked occurrence frequencies of all BB/BB H-bonds that occur in the turns/tails in the selected set, expressed in the frame of the selected beta turn.

The *Turn/set profile* section at the bottom of the top half of the page profiles the structure from the selected set that is currently displayed in the viewer, reporting values for all selection criteria and also displaying information on the selected set as a whole, including profiles of DSSP secondary structure and context vectors.

At bottom left on the page, the profile window provides additional information on the selected set. The distributions of the set's turns across the classical types, Ramachandran types and BB clusters are displayed, along with ranked statistics for all possible single-AA sequence motifs in the turns/tails (if no specific sequence motif is entered as a selection criterion), statistics for the selected single or multiple-AA motif (if a motif is entered), or ranked statistics for motifs that have been specified by "wild-card" sequence motif entries, which support motif detection at particular positions or position ranges in the selected set. A list of PDB addresses for all turns in the set is displayed at the bottom of the profile window.

At bottom right on the page, a JSmol viewer displays turn/tail structures from the selected set in the turn-local coordinate system, in the contexts of their PDB files. By default, structure external to the turn/tails is displayed in translucent cartoon only, out to a radius of 30Å from the turn, but the user can add a ball/stick representation and change the display radius.

2.3 Distribution plots

ExploreTurns gives access to three types of distribution plots which each involve the geometric turn parameters, H-bond energy or depth beneath the solvent-accessible surface and cover the classical types, BB clusters or the global turn set. Parameter

distribution plots show the variation of each parameter within and between types and clusters and guide turn selection by parameter value. Also provided, for single-AA and pair sequence motifs with sufficient abundance within the turn and the first two residues of each tail, are plots of motif fractional overrepresentations vs. parameter value, which identify the parameter regimes most compatible with each motif. Lastly, Ramachandran-space scatterplot heatmaps chart the distributions of each parameter across the dihedral-angle spaces of the two central turn residues, identifying the bond rotations that drive parameter variation.

2.4 Comparison with existing tools

Supplementary Section 1 compares ExploreTurns to existing tools.

3 | Applications

3.1 Proposed generalization of the beta-bulge loop

The current definition of the beta-bulge loop (BBL)[11], a five-residue (type 1) or six-residue (type 2) chain-reversing motif characterized by a pair of BB H-bonds, describes structures which commonly occur at the loop ends of beta hairpins, in which a residue is added at the beginning of the hairpin's C-terminal strand which splits the pair of beta-ladder H-bonds adjacent to the chain-reversing turn, forming a bulge where the turn meets the strand. However, despite the motif's name, the classification of these structures as BBLs does not depend on the presence of a beta ladder, since the loops can also occur independently of hairpins. Accordingly, the most salient feature of BBLs is that they include a loop-terminal residue with a split pair of BB H-bonds, in which one member of the pair closes the loop, defining the structure's extent, while the other links to an interior loop residue, simultaneously forming an internal H-bonded turn at one end of the loop and a bulge at the other, which is constrained at its ends by the H-bond pair. In both type 1 and 2 BBLs, the bulge contains only one residue and occurs at the loop's C-terminus, but neither of these conditions is structurally required. These observations motivated a search for structures that would satisfy a generalized BBL definition requiring only the presence of a split pair of BB H-bonds closing a loop and forming a turn and bulge, without limiting the bulge's length to one residue or requiring that it occur at the loop's C-terminus.

ExploreTurns' BB H-bond selection features (including its interpreter for the compound-turn notation, fully described in Supplementary Section 2) made the search for examples of generalized BBLs straightforward, and Tables 1 and 2 report the results for the most abundant motifs found, which include examples with bulges of lengths up

to three residues that occur at both ends of the loops. The newly identified motifs are all less common than type 1 or 2 BBLs and some are very rare, but like the original pair they occur either at the loop ends of beta hairpins or independently, can occur at ligand-binding and active sites, may contain BB H-bonds in addition to the pair that define their "baseline" structures, and show geometric variation that depends on these additional H-bonds as well as their internal SC interactions and the conformations of their turns. In the longer BBL types, increased BB freedom can give rise to a greater conformational range than that seen in the original types, but recurrent approximate geometries can be identified.

A natural nomenclature for the expanded family of BBLs labels the types with three descriptors: the overall length of the loop, the location of the bulge at the loop's N- or C-terminus, and the bulge length. The motifs are also labelled here using the compound turn notation, which describes short H-bonded loops by the types (lengths), H-bond directions and start positions (in the loop frame) of their component H-bonded turns, together with any constraints on the signs of their BB dihedral angles; the notation is entered in ExploreTurns using the Beta Code[26] romanization of the Greek letters for turn types and dihedral angles.

Table 1 lists selection criteria for four BBL types which exhibit bulges at their C-termini, and Figure 2 displays examples of these motifs, which are labelled {5C1, 6C2, 7C3, 6C1} in the BBL nomenclature. The 5C1 BBL corresponds to type 1 in the existing classification, while the 6C2 and 7C3 BBLs can be formed by successive insertions of additional residues into the 5C1 BBL's bulge. The 6C1 (type 2) BBL can be formed by an insertion of a residue into the 5C1 BBL's beta turn, converting it to a five-residue alpha turn. Each BBL in Table 1 can be further classified into variants by the sign of the φ BB dihedral angle of the last residue in its beta turn, which correlates with whether the bulge orients above (+) or below (-) the beta-turn's plane (defined as the plane which passes through $C_{\alpha 1}$, $C_{\alpha 2}$ and the midpoint of the turn's middle peptide bond[10]).

The BBLs in Table 1 are selected using BB H-bond criteria with exclusions that specify only "baseline" examples, ruling out structures with additional H-bonds involving BB groups within the span of the loop's outermost "framing" H-bond. If exclusions are dropped and a BBL is loaded, the ***H-bond profile*** output box displays all BB H-bonds that can occur in the motif (expressed in the frame of the selected beta turn), and the structures of variants that include particular additional BB H-bonds can be explored by restoring the exclusions, then adding the desired entries from the H-bond profile to the BBL's H-bond list in the ***BB H-bonds...*** box (inclusions selectively override exclusions). Alternatively, H-bonds can be specified in the loop's frame, by adding their entries to the motif's compound-turn label and entering the augmented label in the box. Note that overriding H-bond exclusions can result in the selection of BBLs that simultaneously

Table 1. Examples of beta-bulge loops with C-terminal bulges.

BBL type/compound turn type	Selection criteria and notes (top-line entries select both dihedral variants)	Additional notes
5C1 (type 1) α'-β1 turn Variants 5C1+: α'-β1φ^4 ($\varphi_4 > 0$) 5C1-: α'-β1φ_4 ($\varphi_4 < 0$) The most common BBL type.	***BB H-bonds***: **5C1** or **A'-b1** or 1>+1, 4>1, !1~+1 5C1+: **A'-b1-F4** Very common; "reclined chair", 98% type I turns, bulge towards +z. 5C1-: **a'-b1-f4** Bulge projects towards -z. 76% add a +1>1 (α1) H-bond in bulge. In the compound-turn notation, a BB motif is described by its content of "normal" or "reverse"[1] (primed) H-bonded component turns of lengths 2-8 {δ,γ,β,α,π,ρ,σ}, listed with their start positions in the loop (except in the case of the "framing" turn, which is listed first). The notation is entered in ExploreTurns using Latin-script equivalents, and +/- BB dihedral constraints can be appended using {F/f, Y/y} for {φ+/φ-, ψ+/ψ-}.	BB H-bonds are specified using BBL notation, a compound-turn label, or directly, by indicating the donor and acceptor positions in the β-turn's position frame, separated by '>'. H-bonds can be excluded with direct entry using the '!' prefix, used here with the range specifier *1~+1* to exclude structures with additional H-bonds within the loop's "framing" H-bond (uppercase for the framing reverse α-turn in ***A'-b1*** specifies this exclusion).
6C2 π'-β1 turn Variants 6C2+: π'-β1φ^4 ($\varphi_4 > 0$) 6C2-: π'-β1φ_4 ($\varphi_4 < 0$)	***BB H-bonds***: **6C2** or **P'-b1** or 1>+2, 4>1, !1~+2 6C2+: **P'-b1-F4** Bulge towards +z, 96% type I. 6C2-: **P'-b1-f4** Bulge projects towards -z, 85% type II/II' turns.	Adds a residue to the 5C1 BBL's bulge. As above, the use of uppercase for the framing turn (in **P'-b1**) selects only "baseline" examples.
7C3 ρ'–β1 turn Variants 7C3+: ρ'-β1φ^4 ($\varphi_4 > 0$) 7C3-: ρ'-β1φ_4 ($\varphi_4 < 0$)	***BB H-bonds***: **7C3** or **R'-b1** or 1>+3, 4>1, !1~+3 7C3+: **R'-b1-F4** Very rare; bulge towards +z. 7C3-: **R'-b1-f4** Bulge projects towards -z.	Adds a residue to the 6C2 BBL's bulge.
6C1 (type 2) π'-α1 turn Variants 6C1+: π'-α1φ^5 ($\varphi_5 > 0$) 6C1-: π'-α1φ_5 ($\varphi_5 < 0$)	***BB H-bonds***: **6C1** or **P'-a1** or -1>+1, 4>-1, !-1~+1 6C1+: **P'-a1-F5** "Upright chair"; to select metal-binding "rubredoxin knuckle"[27], enter *C-1C3* into the ***Sequence motif*** box after loading and reload. 6C1-: **p'-a1-f5** Commonly forms a "hook".	Adds a residue to the 5C1 BBL's β-turn, forming an H-bonded α-turn.

Notes: motifs are labelled in bold using both a BBL-specific notation and a general compound-turn nomenclature (Supplementary Section 2). To browse a loop's structures, enter its selection criteria and click ***Load Turns Matching Criteria***; successive clicks of ***Browse Structures*** then step through the set by the interval specified in the ***Stepsize*** box (negative values specify back-steps). To change the radius for the display of structure external to the turn/tails, enter a new value in the ***Display radius*** box and click ***Browse Structures***. The first turn residue is flashed in light green to aid in consistent orientation; orientations with +x toward the viewer and +z generally upwards are suggested in most cases (+z towards the viewer and -x upwards is preferred for the 6C1 BBL). Click ***Clear All Criteria*** before entering each new example.

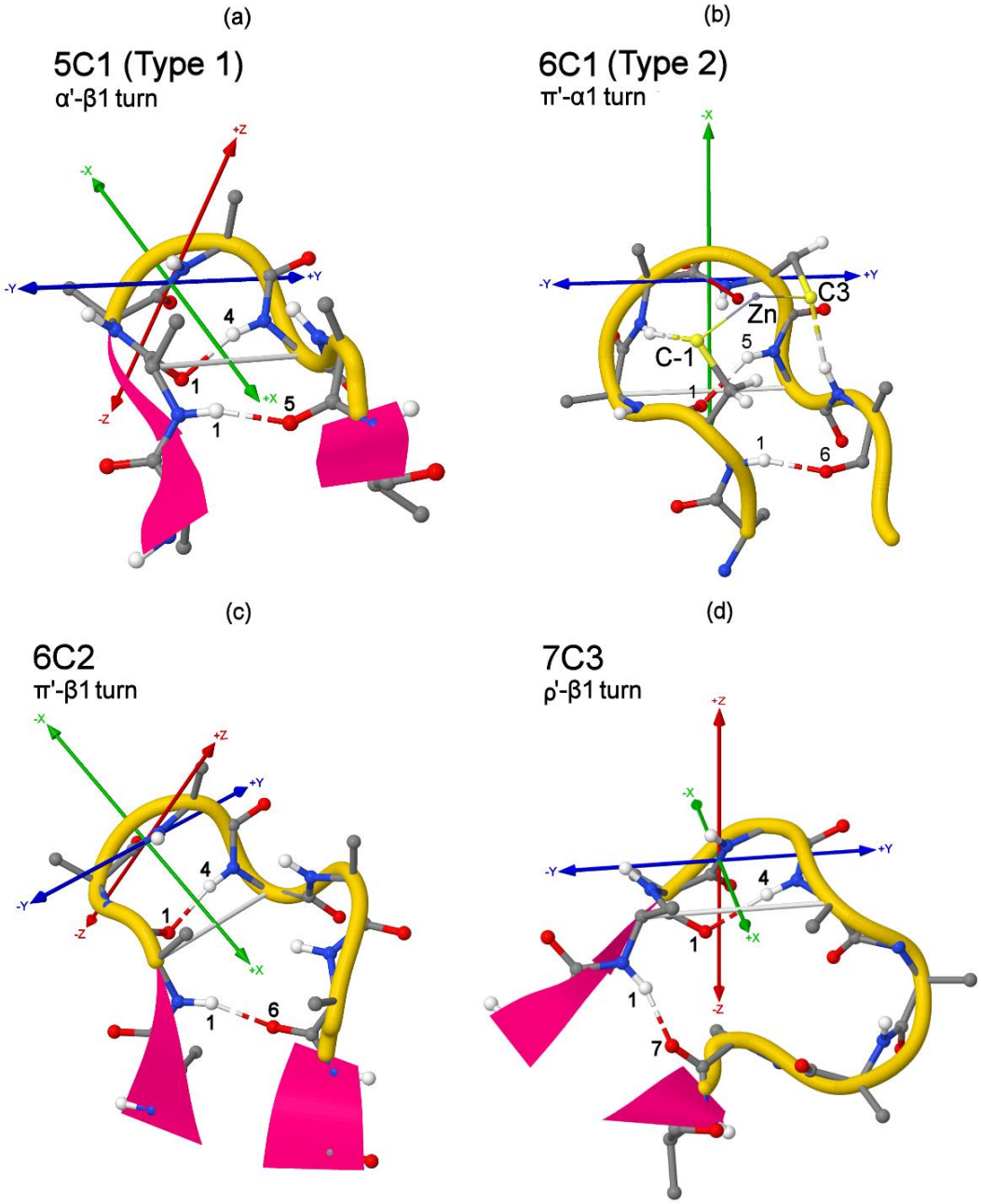

Figure 2. Examples of beta-bulge loops with C-terminal bulges. Structures are labelled using both a BBL-specific notation, which specifies overall loop length, bulge location (N/C terminal) and bulge length, and a general "compound-turn" nomenclature for all short H-bonded loops based on the types (lengths) and start positions of their component H-bonded turns (Supplementary Section 2). Loop-defining H-bonds are labelled at each end with residue positions in the loop. See Table 1 for loop selection criteria; to view the individual structures shown here, load the entire database by clicking ***Load Turns Matching Criteria*** without any criteria, then enter a structure's address into the ***PDB address*** box and click ***Browse Structures***. **(a)** 5C1 (type 1); common α'-$\beta 1\varphi^4$ variant (superscript indicates $\varphi_4 > 0$), 5MBX_A_265. **(b)** 6C1 (type 2), common π'-$\alpha 1\varphi^5$ variant, in its "rubredoxin knuckle"[27] metal-binding form found in Zn fingers/ribbons, 4XB6_D_242. **(c)** 6C2, π'-$\beta 1\varphi^4$ variant, 4HTG_A_278. **(d)** 7C3, ρ'-$\beta 1\varphi_4$ variant (subscript indicates $\varphi_4 < 0$), 2OXG_B_51.

satisfy the definition of more than one type; for example, the addition of a +1>1 BB H-bond (in the beta-turn frame) to the H-bond list for the 6C2 BBL results in the selection of a 6C1C2 combination loop (or, expressed in the compound-turn notation, adding an α1 H-bond to a π'-b1 turn yields a π'-α1-β1 turn, entered in ExploreTurns as P'-a1-b1, where the framing π-turn is capitalized to specify the loop's baseline form).

The effects on BBL conformations of SC interactions, including those that commonly shape the loops with H-bonds or constrain them with proline, can be investigated by identifying the sequence motifs that occur in a loop using ExploreTurns' motif detection tools, then browsing/profiling examples that contain those motifs (see example 14 in the online primer, or section 1 in the user guide).

Table 2 lists selection criteria for four BBL types with bulges at their N-termini {5N1, 6N2, 7N3, 7NC3}, and Figure 3 displays examples of these motifs. In its hairpin context, the 5N1 BBL can be formed by the insertion of a single residue into a 2:2 hairpin immediately before its chain-reversing turn, while the 6N2 and 7N3 BBLs can result from successive insertions into the 5N1 BBL's bulge. Like the C-type motifs, the 6N2 BBL can be classified into variants based on the sign of the φ angle of the last residue in its beta turn. The 7NC3 BBL is classified as both N- or C-type, since it contains split pairs of BB H-bonds at each terminus.

Table 2. Examples of beta-bulge loops with N-terminal bulges.

BBL type/compound turn type	Selection criteria and notes (direct H-bond entries are specified in the β-turn frame)	Additional notes
5N1 α'-β2 turn	***BB H-bonds***: **5N1** or **A'-b2** or -1>4, 4>1, !-1~4	In its hairpin context, this motif can be formed by the insertion of a residue just before the chain-reversing turn of a 2:2 hairpin.
6N2 π'-β3 turn <u>Variants</u> 6N2-: π'-β3φ_6 ($\varphi_6 < 0$) 6N2+: π'-β3φ^6 ($\varphi_6 > 0$)	***BB H-bonds:*** **6N2** or **P'-b3** or -2>4, 4>1, !-2~4 <u>6N2-</u>: **P'-b3-f6** Tails commonly project towards -z. <u>6N2+</u>: **p'-b3-F6** C-tail projects towards +z.	Adds a residue to the 5N1 BBL's bulge. The use of lowercase for the first (framing) turn of the 6N2+ variant drops the general exclusion of additional H-bonds specified by uppercase.
7N3 ρ'-β∗1-β4 turn (β∗1 excludes 7NC3 below)	***BB H-bonds***: **7N3** or **r'-b*1-b4** or -3>4, 4>1, !1>-3	Adds a residue to the 6N2 BBL's bulge. *b*1* specifically excludes an H-bonded β-turn beginning at position 1, which would yield the 7NC3 BBL below.
7NC3 ρ'-β1-β4 turn	***BB H-bonds***: **7NC3** or **R'-b1-b4** or 1>+3, 4>1, +3>4, !1~+3	Since this loop contains split pairs of BB H-bonds at both termini, it is classified as both N- and C-type.

Note: for browsing directions, see the notes for Table 1.

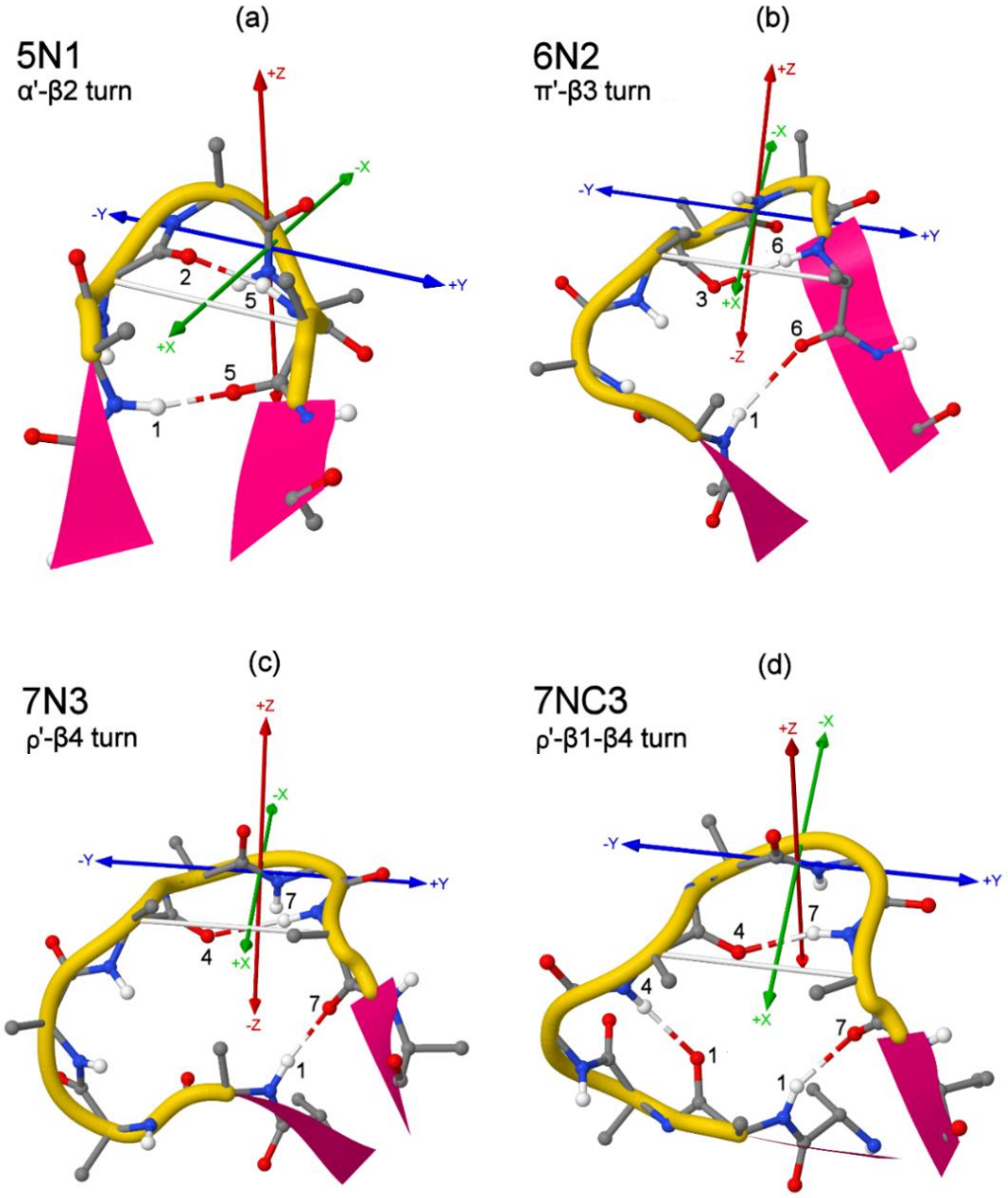

Figure 3. Examples of beta-bulge loops with N-terminal bulges. Table 2 lists the ExploreTurns selection criteria for these motifs. See the Figure 2 caption for loop nomenclature and viewing directions. **(a)** 5N1 (4R1D_B_107). **(b)** 6N2, π'-β3φ_6 variant, 4PXW_A_1104. **(c)** 7N3 (5KAZ_A_36). **(d)** 7NC3; split pairs of BB H-bonds at both termini classify this BBL as both N- and C-type (3W6W_A_182).

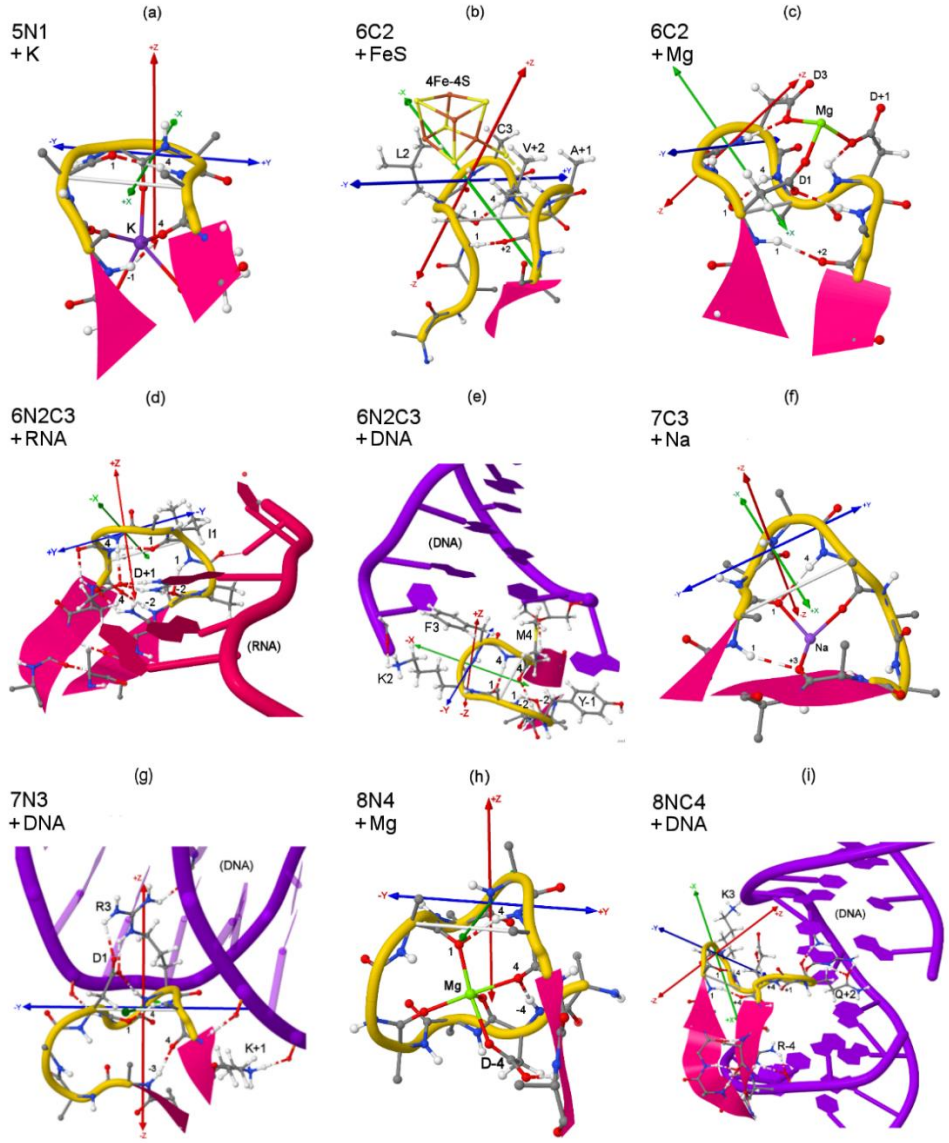

Figure 4. Examples of ligand-binding/active site beta-bulge loops of the generalized types. H-bonded BB atoms and interacting SCs are labelled with residue positions in the frame of the selected beta turn. See the Figure 2 caption for structure viewing directions. **(a)** A 5N1 BBL binds K in the fatty-acid binding protein AtFAP1 (4DOO_A_35). **(b)** A 6C2 BBL binds an active-site 4Fe-4S cluster with Cys3 (C3) in acetyl-CoA synthase/carbon monoxide dehydrogenase. All other SCs in the loop are nonpolar (1OAO_C_526). **(c)** A 6C2 BBL binds Mg with an Asp triplet (D1, D3, D+1) at the active site of phosphoglucomutase from Xanthomonas citri (5BMN_A_259). **(d)** A 6N2C3 BBL binds RNA with interactions including BB and SC (D+1) H-bonds, a hydrophobic interaction (I1), and a possible pi-peptide bond interaction, in the HutP antitermination complex (1WPU_A_133). **(e)** A 6N2C3 BBL at the tip of a "beta wing" motif acts as a "separating knife", wedging between the first and second base pairs at the blunt end of a DNA duplex, performing a function crucial to the helicase reaction in human Werner syndrome protein (WRN)[28] (3AAF_A_1035). **(f)** A 7C3 BBL binds Na in methionine R-sulfoxide reductase B1 (3MAO_A_60). **(g)** A 7N3 BBL binds DNA with BB H-bonds, a Lys SC (K+1), and an Asp/Arg salt bridge (D1R3) in RovA, a transcriptional activator of Yersinia invasin (4AIK_A_84). **(h)** An 8N4 BBL binds Mg with four BB carbonyls and an Asp SC (D-4) (4WA0_A_477). **(i)** An 8NC4 BBL and its hairpin bind DNA with BB H-bonds and three SCs (R-4, K3, Q+2) in the de novo DNA methyl transferase DNMT3B (5CIY_A_232).

Like the C-type BBLs, the 5N1, 6N2 and 7NC3 BBLs are selected with BB H-bond criteria that contain exclusions ruling out additional H-bonds, and the structural effects of particular extra H-bonds can be explored by selectively overriding these exclusions.

Figures S3 and S4 in Supplementary Section 3 guide the exploration of a larger set of BBLs, including those presented in Tables 1 and 2 and 11 additional motifs, with schematic diagrams that show the H-bond topology of each type and its relationships with other types.

BBLs combine the constraints enforced by their defining BB H-bonds with conformational range in their bulges and turns that is configurable by SC H-bonding or additional BB H-bonds, giving them versatility for structural or functional roles. Figure 4 shows nine examples of generalized BBL types at binding or active sites, with ligands that include metal ions, an FeS cluster, RNA and DNA (x3).

3.2 Mapping Asx N-cap sequence preference vs. helix/turn geometry

At the helical N-terminus, unsatisfied BB NH groups are commonly "capped" by H-bonding with the SCs of Asp or Asn (Asx) or Ser or Thr (ST) in Asx/ST N-caps[20,22,23,24]. Beta turns are also found at the helical N-terminus[22], and the Asx AAs in these turns show greatest overrepresentation when they occur at the third turn position and this position coincides with the NCap residue just before the helix. ExploreTurns "context vectors" specify the approximate directions of the N- and C-tails in the turn frame, and when a helix begins within two residues of the turn, tail direction is measured by the orientation of the helix axis, so the tool can be used to select sets of structures in which the helix has particular orientations with respect to the turn. Since the tool also computes SC motif statistics for each selected set, it can be used to map N-cap sequence preference vs. helix/turn geometry, by evaluating motif overrepresentation and abundance in sets of structures spanning the range of helix/turn orientations.

The context vector for each tail is specified in a (longitude, latitude) format, with the "equator" lying in the turn plane, the zero of longitude at the equator corresponding to the +x direction, positive longitudes occurring in the +y half-space, and positive latitudes occurring in the +z half-space (see primer example 5 or section 1 of the user guide). In this analysis, sets of structures are selected by specifying C-tail context vectors at 10° intervals of longitude and latitude, and the ***Vector tolerances*** field restricts the selected structures to those lying within 5° of a specified context vector. Figure 5 plots the abundance and overrepresentation of the Asp3 motif (D3, see example structure in Figure S8b) for each orientation.

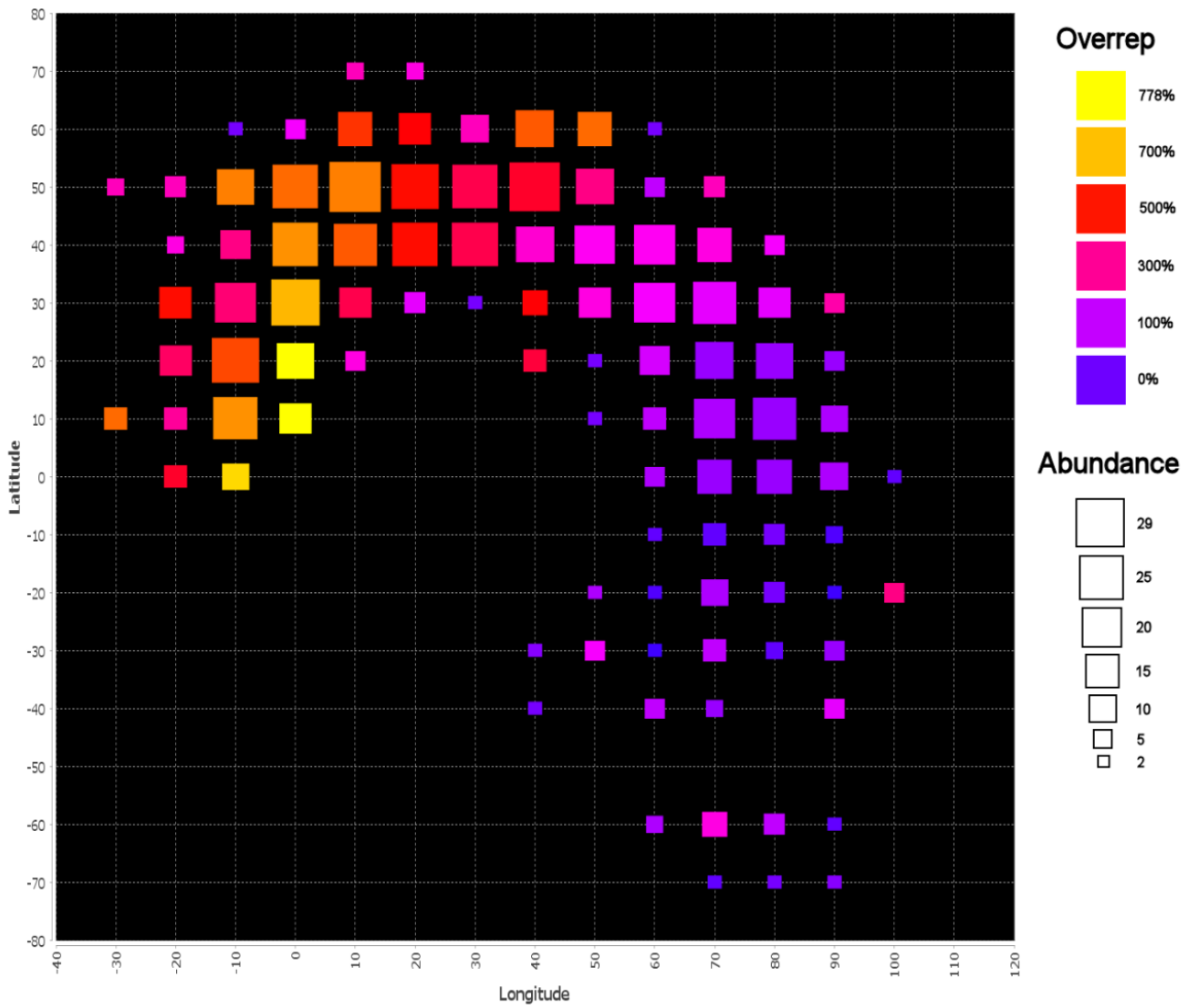

Figure 5. Fractional overrepresentation and abundance vs. helix/turn orientation for the Asx N-cap sequence motif specifying Asp at turn position 3. Beta turns in which the third turn residue lies at the NCap position in an α-helix are selected using the ***DSSP symbols*** entry *****<***H>*****, and the orientations of the helix with respect to the turn are sampled at 10° increments by varying the direction of the C-tail (here the helix axis), using ***Context vectors*** entries of the form **,*<>longitude,latitude*. The ***Vector tolerances*** entry **<>5* is applied to select structures with helix axes that lie within 5° of each context vector, and the Asp3 motif is specified with the ***Sequence motif*** entry *D3*. Squares are plotted for all orientations that yield at least 10 structures, with at least one structure containing D3. D3's fractional overrepresentation is depicted as a heatmap colored according to the upper legend, while the area of each square is set proportional to the occurrence of D3 in each orientation according to the lower legend. Not all orientations with high overrepresentation show frequent SC capping of helix BB groups, as Asp's hydrophilicity is likely also favorable at this commonly exposed position, and SC/SC H-bonding (including interactions with the helix), as well as non-capping SC/BB H-bonding also likely contribute.

Figure 5 shows that D3's peak overrepresentation occurs at zero longitude, with the helix axis parallel to the xz plane, and latitudes just above the turn plane; overrepresentation maintains very high values as the helix rotates upward to 50° above the plane. As the helix rotates into negative longitudes, towards -y and across the mouth of the turn, overrepresentation and abundance fall off past -10°, while at positive longitudes, where the helix rotates towards +y and away from the turn, the two measures remain high at mid-latitudes, extending out to 50° longitude. At longitudes past 50°, overrepresentation drops as the helix rotates downwards, crossing below the plane, and D3's SC loses access to its face.

Figure S5 in Supplementary Section 4 maps abundance and overrepresentation for the Asn3 N-cap motif, and shows that peak overrepresentation occurs not when the helix orients at zero longitude, as seen for Asp3, but when it straddles the meridian at longitudes of +/-10°, and peak overrepresentation is more focused in latitude. Figures 5 and S5 can guide ExploreTurns investigations of the structure and H-bonding associated with Asx N-caps formed by beta turns.

3.3 Classifying, exploring and comparing short H-bonded loops

Supplementary Section 2 gives a complete description of the compound-turn nomenclature for short H-bonded loops and its interpreter in ExploreTurns, and applies the notation to explore the structures of non-BBL loops and their variants and demonstrate their relationships (Figure S1). New motifs are identified, including a "short Schellman loop", two new seven-residue Schellman loop variants, short and long forms of the alpha-beta loop[8] (the "beta-gamma" and "pi-alpha" loops), and a seven-residue "open" loop dubbed the "beta bracket".

3.4 Profiling Schellman loop/beta turn conformations

Supplementary Section 5 applies ExploreTurns to profile the geometries of Schellman loop/beta turn combinations at the helical C-terminus. Four principal conformations are identified, and Figure S6 presents example structures and selection criteria for each.

3.5 Investigating the depth dependence of turn geometry

Supplementary Section 6 applies ExploreTurns to investigate the relationship between the distribution of turns across the classical types or BB clusters and a turn's depth beneath the solvent-accessible surface. The proportion of structurally unclassified

turns is found to increase markedly with depth at the expense of classical type I, which may reflect distortions due to increasing contacts with structures in the protein external to turns, increasing turn involvement in local structures such as ligand-binding/active sites, or other factors.

3.6 Exploring sequence motifs, supersecondary structures and beta hairpins

Supplementary Section 7 provides selection criteria and sample images for examples of sequence motifs in beta turns; some motifs are tuned by BB geometry to maximize overrepresentation (and likely compatibility with SC H-bonding).

Supplementary Section 8 gives criteria for the selection of supersecondary structures in which turns link helices and strands into particular orientations, and Figure S9 displays example structures.

Sets of examples of 12 types of beta hairpins across 4 classes[14,15] can be browsed and profiled by entering type names into the ***BB H-bonds...*** box; for a hairpin catalog, see *H-bonded backbone motifs/Beta hairpins* in the ExploreTurns primer.

3.7 Geometric parameter ranges and distribution plots

Supplementary Section 9 illustrates the modes of structural variation described by geometric turn parameters by providing selection criteria for examples of turns that exhibit parameter values near their limits in the database, and Figure S10 presents examples structures.

Supplementary Section 10 gives directions for the display of notable examples of the distribution plots available in ExploreTurns, and Figure S11 displays six plots.

4 | Discussion

The utility of ExploreTurns is demonstrated here in several areas, including the detection of new H-bonded loops. Although many of these "compound turns" are rare in the redundancy-screened ExploreTurns database, they can nevertheless play key roles, as demonstrated by the ligand-binding/active site BBLs in Figure 4, and the large family of generalized beta-bulge loops explored here, customizable by its BB and SC H-bonded variants, may support loop design efforts targeted at binding and active sites. For example, BBLs might serve as configurable "tools" at the ends of beta-ladder "arms" which can project them into interaction sites, as seen in the beta-wing motif that

separates DNA strands with a 6N2C3 BBL in human Werner syndrome helicase (Figure 4e, or enter *6N2C3* in the ***BB H-bonds...*** box).

A reservoir of new examples of compound turns which could provide a much more complete coverage of the range of BB conformations and SC interactions in short H-bonded loops may already be available, in the form of the large sets of protein structures predicted from sequence by tools such as AlphaFold 2[29]. These structures may enable the classification of the geometries of some compound-turn motifs into "types" analogous to beta-turn types. However, while AlphaFold 2 has shown good accuracy in the prediction of short loops[30], the rarity, irregularity, and/or complexity of some compound turns (e.g. the 7N3C1C3 BBL in Figure S4) may present challenges to prediction algorithms. If all algorithms do not accurately predict all loop motifs, a set of loops could be used in benchmark testing to rank the tools by their ability to predict small, non-repetitive H-bonded BB structures. The conformations of any compound-turn motifs that cannot be accurately predicted can be further explored with computational modelling like that applied by Venkatachalam to beta turns[2], and the structure and energetics of all the new loops can be investigated experimentally with mutagenesis and other methods.

5 | Conclusions

ExploreTurns supports the study of beta turns with multiple new features, including a turn-local coordinate system, a global turn alignment, and a set of geometric parameters that complement Ramachandran-space classification systems by discriminating structure within types or clusters. The widespread occurrence of beta turns in proteins also allows the tool to serve as an exploratory browser for most loop structure which introduces a framework for analysis and comparison into the many motifs and contexts associated with turns.

ExploreTurns should prove useful in any application in which an enhanced Euclidean-space picture of turns and the ability to identify and tune structures suited to particular requirements can improve understanding or performance. The facility also supports education, since it allows a user to ask and immediately answer a wide range of questions concerning the structures, interactions, and roles of beta turns and their motifs.

6 | Methods

The ExploreTurns dataset is described in Supplementary Section 11. Biopython[31] was used to extract the structural data from PDB[12] files. Measurements of the depth of beta turns beneath the protein's solvent-accessible surface were computed with Depth[32]. ExploreTurns was written in HTML/CSS and Javascript, with an embedded JSmol viewer[33]. The tool is tested for compatibility with the Chrome, Edge, and Firefox browsers; for best performance, browse with Chrome or Edge.

The turn-local coordinate system[10], the global turn alignment which it implicitly establishes, the methods used to compute sequence motif overrepresentation and p-value[10,22,34,35], the derivation of context vectors, and the H-bond definition are all described in the online user guide, while geometric turn parameters[10] are described in dedicated online help.

Supplementary Material

Supplementary file 1 (PDF)**.** Supplementary sections, with table of contents.

Declarations

Availability of data and materials

ExploreTurns is available at www.betaturn.com. For best performance, browse with Chrome or Edge. All data supporting the conclusions of this article are available from the Protein Data Bank at www.rcsb.org.

Competing interests

The author declares that he has no competing interests.

Acknowledgements

The author wishes to thank Athena Newell for useful discussions and work as a research assistant and Kit Newell for work as a research assistant. The constructive suggestions and support of the attendees of ISMB/3DSig 2019 (Basel) and the 2019 and 2023 Annual Symposia of The Protein Society (Seattle and Boston) also contributed to the development of this project.

References

1) Koch O, Klebe G. Turns revisited: A uniform and comprehensive classification of normal, open, and reverse turn families minimizing unassigned random chain portions. *Proteins*. 2009;74(2):353-67.

2) Venkatachalam CM. Stereochemical Criteria for Polypeptides and Proteins. V. Conformation of a System of Three Linked Peptide Units. *Biopolymers*. 1968;6(10):1425-36.

3) Shapovalov M, Vucetic S, Dunbrack RL. A new clustering and nomenclature for beta turns derived from high-resolution protein structures. *PLOS Computational Biology*. 2019;15(3).

4) D'Arminio N, Ruggiero V, Pierri G, Marabotti A, Tedesco C. Emerging role of carbonyl-carbonyl interactions in the classification of beta turns. *Protein Science*. 2023;33(2).

5) Hutchinson EG, Thornton JM. PROMOTIF--a program to identify and analyze structural motifs in proteins. *Protein Science*. 1996;5(2):212-220.

6) Wilmot CM, Thornton JM. β-turns and their distortions: A proposed new nomenclature. *Protein Eng*. 1990;3(6):479-93.

7) Hollingsworth SA, Karplus PA. A fresh look at the Ramachandran plot and the occurrence of standard structures in proteins. *Biomol Concepts*. 2010;1(3-4):271-83.

8) Leader DP, Milner-White EJ. Motivated Proteins: A web application for studying small three-dimensional protein motifs. *BMC Bioinformatics*. 2009;10:60.

9) Leader DP, Milner-White EJ. Structure Motivator: A tool for exploring small three-dimensional elements in proteins. *BMC Struct Biol*. 2012;12:26.

10) Newell NE. A geometric parameterization for beta turns. *bioRxiv*. doi: https://doi.org/10.1101/2024.01.01.573818

11) Milner-White EJ. Beta-bulges within loops as recurring features of protein structure. *Biochim Biophys Acta*. 1987;911(2):261-5.

12) Berman HM, WestBrook J, Feng Z, Gilliland G, Bhat TN et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235-242.

13) Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*. 1983;22(12):2577-637.

14) Milner-White, EJ, Poet R. Four classes of beta hairpins in proteins. *Biochem J*. 1986; 240(1): 289-292.

15) DuPai CD, Davies BW, Wilke CO. A systematic analysis of the beta hairpin motif in the Protein Data Bank. *Protein Science*. 2021;30(3):613-23.

16) Richardson JS, Getzoff ED, Richardson DC. The beta bulge: A common small unit of nonrepetitive protein structure. *Proceedings of the National Academy of Sciences*. 1978;75(6):2574-8.

17) Watson JD, Milner-White EJ. A novel main-chain anion-binding site in proteins: the nest. A particular combination of phi,psi values in successive residues gives rise to anion-binding sites that occur commonly and are found often at functionally important regions. *Journal of Molecular Biology*. 2002;315(2):171-82.

18) Leader DP, Milner-White EJ. Bridging of partially negative atoms by hydrogen bonds from main-chain NH groups in proteins: The crown motif. *Proteins*. 2015;83(11):2067-76.

19) Schellman C. The R-L conformation at the ends of helices. In: Jaenicke R, editor. Protein folding. NewYork: Elsevier/North-Holland; 1980. p. 53-61.

20) Aurora R, Rose GD. Helix capping. *Protein Science*. 1998;7(1):21-38.

21) Newell NE. Cascade detection for the extraction of localized sequence features; specificity results for HIV-1 protease and structure-function results for the Schellman loop. *Bioinformatics*. 2011;27(24):3415-22.

22) Newell NE. Mapping side chain interactions at protein helix termini. *BMC Bioinformatics*. 2015;16:231.

23) Wan W-Y, 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. *Journal of Molecular Biology*. 1999;286(5):1633-49.

24) Wan W-Y, Milner-White EJ. A recurring two-hydrogen-bond motif incorporating a serine or threonine residue is found both at α-helical N termini and in other situations. *Journal of Molecular Biology*. 1999;286(5):1651-62.

25) Newell, NE. MapTurns: mapping the structure, H-bonding and contexts of beta turns in proteins. *bioRxiv*. doi: https://doi.org/10.1101/2024.01.02.573949

26) The TLG Beta Code Manual. [Accessed February 26 2024]. https://stepahnus.tlg.uci.edu/encoding/BCM.pdf

27) Wang B, Alam SL, Meyer HH, Payne M, Stemmler TL, Davis DR, Sundquist WI. Structure and Ubiquitin Interactions of the Conserved Zinc Finger Domain of NpI4. *J. Biol Chem*. 2010;278(22):20225-20234.

28) Kitano K, Kim, S-Y, Hakoshima T. Structural Basis for DNA Strand Separation by the Unconventional Winged-Helix Domain of RecQ Helicase WRN. Structure. 2003;18(2):177-187.

29) Mi T, Nguyen D, Burgess K. Bicyclic Schellman Loop Mimics (BSMs): Rigid Synthetic C-Caps for Enforcing Peptide Helicity. *ACS Cent. Sci.* 2023;9: 300-306.

30) Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Zidek A, et al. Highly accurate protein structure prediction for the human proteome. *Nature*. 2021; 596:590-596.

31) Stevens AO, He Y. Benchmarking the Accuracy of AlphaFold2 in Loop Structure Prediction. *Biomolecules*. 2022;12:985.

32) Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*. 2009;25(11):1422-3.

33) Tan KP, Nguyen BT, Patel S, Varadarajan R, Madhusudhan MS. Depth: a web server to compute depth, cavity sizes, detect potential small molecule ligand-binding cavities and predict the pKa of ionizable residues in proteins. *Nucleic Acids Res*. 2013;41:W314-21.

34) Herráez A. Biomolecules in the computer: Jmol to the rescue. *Biochemistry and Molecular Biology Education*. 2006;34(4):255-61.

35) Bishop YM, Fienberg SE, Holland PW. Discrete Multivariate Analysis: Theory and Practice. New York: Springer; 2007.

36) Hu JC, Newell NE, Tidor B, Sauer RT. Probing the roles of residues at the e and g positions of the GCN4 leucine zipper by combinatorial mutagenesis. *Protein Science*. 1993;2(7):1072-84.