ExploreTurns: A web tool for the exploration, analysis and retrieval of beta turns and their contexts in proteins; application to beta-bulge loops and helix capping

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

Abstract

The most common type of protein secondary structure after the alpha helix and beta strand is the beta turn, in which the backbone (BB) chain abruptly changes direction over four amino acid residues. Existing tools for the study of beta turns characterize their structures exclusively in the Ramachandran space of BB dihedral angles, which presents challenges for the visualization, comparison and analysis of the wide variety of turn geometries. A recent study has introduced a turn-local Euclidean-space coordinate system and a global alignment for turns, along with a set of geometric parameters for their bulk BB shape which describe modes of structural variation not explicitly captured by Ramachandran-space systems. These features, and others, are incorporated here into ExploreTurns, a tool specialized for the exploration, analysis, geometric tuning and retrieval of beta turns and their local contexts which combines the advantages of Ramachandran- and Euclidean-space representations. ExploreTurns is a database selection screen, structure profiler and 3D viewer, integrated onto a single web page, that renders a redundancy-screened, PDB-derived beta-turn dataset structurally addressable by a wide range of criteria describing the turns and their BB neighborhoods. The tool is applied here to a proposed generalization of the beta-bulge loop, a mapping of helix-capping sequence preferences, a conformational profile of the Schellman loop, and an analysis of the depth dependence of turn geometry, and it should prove generally useful in research, education, and applications such as protein design, in which an enhanced Euclidean-space description of turns and structural tuning may improve performance.

Keywords

Beta turn, beta-bulge loop, helix cap, Schellman loop, beta clip, protein structure, motif detection, secondary structure, sequence motif, structural motif

Abbreviations

AA: Amino acid

BB: Protein backbone

SC: Amino acid side-chain

Position index: Describes residue positions in the turn or its BB neighborhood 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 by indicating the donor and acceptor positions in the turn frame, 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.

AA sequence motif: AA sequence motifs are labelled with the single-letter codes of the included AAs, each followed by its position index. 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; also BBL1 (type 1), BBL2 (type 2).

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 third most common type of protein secondary structure after helices and beta strands. Beta turns play key roles in multiple contexts in proteins due to the wide variety of backbone (BB) conformations and side-chain (SC) interactions they exhibit; 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 review in[3]). 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[4], BB clusters[3] and Ramachandran types[3,5,6] (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[7] and its associated standalone application Structure Motivator[8], 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[9], of a common local Euclidean-space coordinate system for beta turns, together with a set of geometric parameters that describe their bulk BB shape, has enabled the development of ExploreTurns, a tool specialized for the study and retrieval of beta turns which combines the advantages of Ramachandran- and Euclidean-space representations. 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 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[9]. Parameters enable structural discrimination within the classical and Ramachandran types, yielding major improvements in the specificity and accuracy of measurements of sequence preferences in turns and enabling the geometric tuning of turn structure for compatibility with key SC interactions and BB contexts.

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[10]-derived dataset of 102,192 beta turns structurally addressable by a wide range of criteria describing the turns and their neighboring four-residue BB "tails". Criteria include classical type, Ramachandran type, BB cluster, 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[11] secondary structure codes, inclusionary/exclusionary patterns of BB H-bonds, the approximate orientations of the tails with respect to the turn, and the depth of the turn beneath the protein's solvent-accessible surface.

ExploreTurns supports the comprehensive exploration and analysis of structure and interaction in beta turns and their local 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 particular SC interactions, by adjusting parameter ranges to maximize the fractional overrepresentation and/or abundance of the sequence motifs associated with these interactions 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 structural motif that incorporates a beta turn and can be defined by the selection criteria. The most common of these motifs include multiple types of beta hairpins[12,13], beta-bulge loops[14] and some beta bulges[15], alpha-beta loops[7], BB-mediated helix caps such as the Schellman loop[16,17,18,19], some nests[20], and helix/helix, helix/strand, or strand/strand supersecondary structures linked by beta turns into particular geometries.

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[21,22] and Asx/ST helix caps[17,19,21,22], 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 motif maps generated by the MapTurns[23] tool (also available at www.betaturn.com), which support 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.

2.2 The ExploreTurns page

The top half of the ExploreTurns web page (Figure 1) is divided vertically into three sections. The six buttons at the top of the page provide access to extensive online help: a primer, structured as a set of 25 examples, serves as an introduction to the roles beta turns play in proteins as well as the tool itself, while a comprehensive user guide documents all features of the application, and separate buttons 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 possible BB/BB H-bonds in the turns/tails.

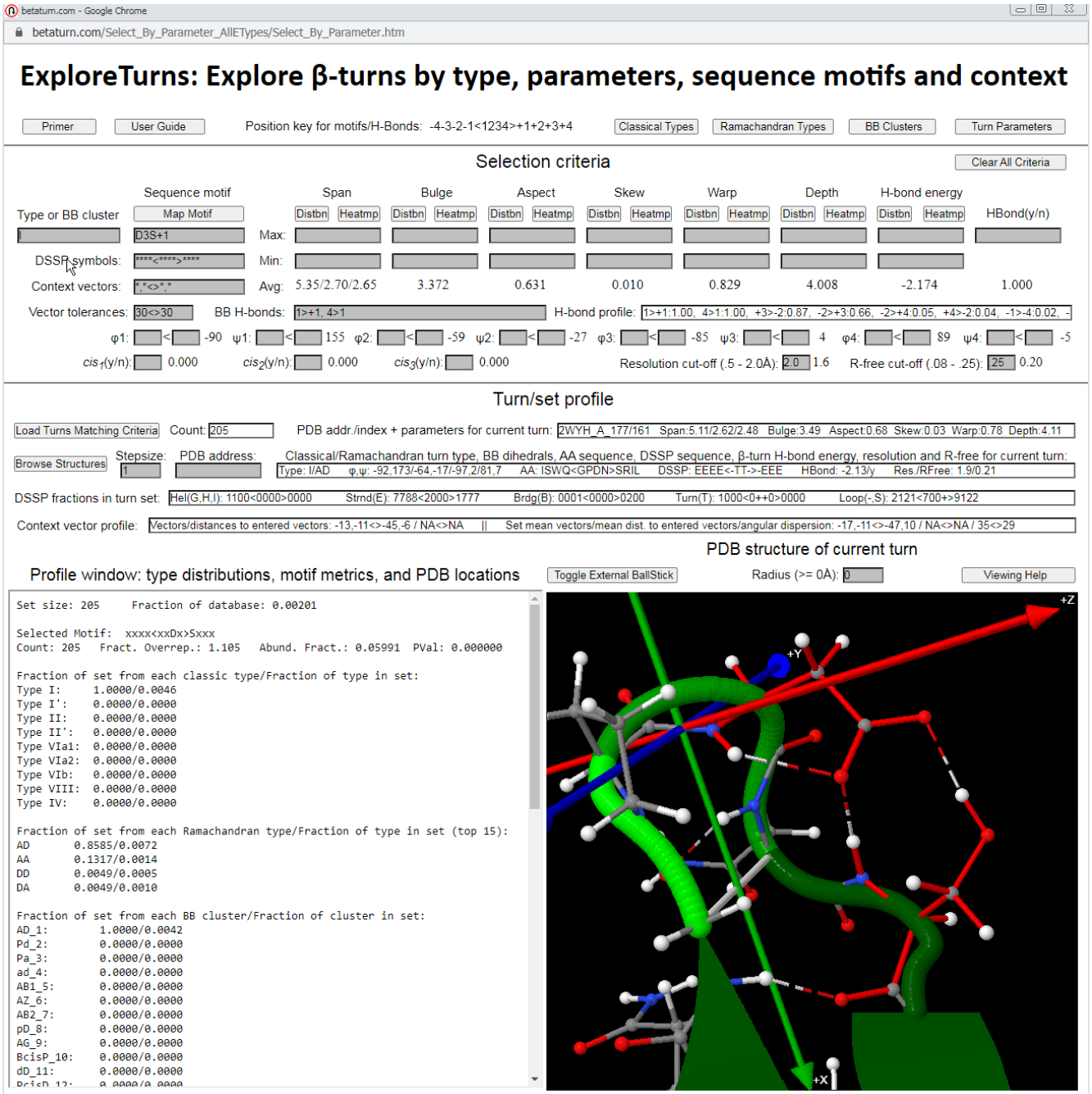

Figure 1. The ExploreTurns page. Selection criteria have been entered for the sequence motif which specifies Asp at turn position 3 and Ser just after the turn in turns of classical type I in the type 1 beta-bulge loop (BBL1). The sequence motif, which is associated with SC/SC and SC/BB H-bonds linking the loop's turn to its "bulge", is specified with the ***Sequence motif*** entry *D3S+1*, while the ***BB H-bonds*** entry *1>+1, 4>1* specifies the loop's characteristic pair of H-bonds (in the position frame of the turn) and the ***Type of BB cluster*** entry *I* selects the turn type. The structure is displayed in the turn-local coordinate system, with the turn's span shown as a white bar and the motif's SCs highlighted in red. The radius for the display of structure outside the turn and its four-residue BB "tails" has been set to zero to remove clutter.

The *Turn/set profile* section at the bottom of the top half of the page profiles the structure in 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 context of their PDB files. By default, structure external to the turn/tails is displayed in translucent cartoon 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 covering the classical types, BB clusters and the global turn set which can guide structure selection and reveal the relationships between the geometric parameters and the Ramachandran-space turn representation. Plots of the parameter distributions 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 values of each geometric 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

Section 1 in Supplementary Material 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)[14], 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 joins the ends of the loop, defining the structure's extent, while the other links its residue to an interior loop residue, forming a bulge. In both type 1 and 2 BBLs, this 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 delineating the loop and forming the bulge, without limiting the bulge's length to one residue or requiring that it occur at the loop's C-terminus.

ExploreTurns' BB H-bonds selection feature made a search for examples of the generalized BBL straightforward for structures which contain beta turns, 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 structure. The newly identified motifs are all less common than type 1 or 2 BBLs, but like the original pair they occur either at the loop ends of beta hairpins or independently, may contain 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 the conformations of the turns which they include. In the types in which the bulge contains more than one residue, increased BB freedom can give rise to a greater conformational range within the bulge 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.

Table 1 lists ExploreTurns selection criteria for four BBL types which exhibit bulges at their C-termini, and Figure 2 displays examples of these motifs. Using the generalized nomenclature, these motifs are termed 5C1, 6C2, 7C3 and 6C1. 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 6C2 and 7C3 BBLs can be further classified by whether the bulge extends above (+) or below (-) the plane of the loop's N-terminal turn, which is defined as the plane which passes through $C_{\alpha 1}$, $C_{\alpha 2}$ and the midpoint of the turn's middle peptide bond[9]. The 6C1 BBL, which corresponds to type 2 in the existing classification, can be formed by an insertion of a residue into the 5C1 BBL's chain-reversing beta turn, converting it to a five-residue alpha turn.

With the exception of the 6C1 BBL, which is selected here in its "rubredoxin knuckle"[24] metal-ion binding variant, in which additional BB H-bonds are inherently restricted, the BBLs in Table 1 are selected using BB H-bond criteria with exclusions that rule out H-bonds in addition to those that define the loops, thereby specifying only "baseline" examples. If all exclusions are dropped and a set of BBL structures is loaded, the ***H-bond profile*** output box displays all BB H-bonds found in the BBL, and the structures of variants that include particular additional BB H-bonds can be explored by restoring the exclusions, then adding ***BB H-bonds*** entries that correspond to the desired H-bonds, selectively overruling the exclusions. Note that overriding H-bond exclusions can result in the selection of BBLs that simultaneously satisfy the definition of more than one type; for example, the explicit inclusion of the +1>1 BB H-bond in the criteria for the 6C2 BBL results in the selection of a set of structures which meet both the 6C1 and 6C2 definitions.

Table 2 lists selection criteria for four BBL types with bulges at their N-termini {5N1, 6N2, 7N3, 7(N/C)3}, 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 be formed by successive insertions into the 5N1 BBL's bulge. The 7(N/C)3 BBL is both N- and C-type, since it contains split pairs of BB H-bonds at both termini.

Table 1. Selection criteria for beta-bulge loops with C-terminal bulges.

BBL type	Selection criteria	Notes
5C1: Type 1 in the existing classification; the most common BBL type.	***BB H-bonds***: 1>+1, 4>1, !1>*, !2>*, !3>*, !4>*, !+1>*,!*>1, !*>2, !*>3, !*>4, !*>+1	BB H-bonds are specified by indicating the donor and acceptor positions in the turn frame, separated by '>'. H-bonds are excluded with the '!' prefix, used here in combination with the wildcard '*' to select only "baseline" structures. In its common beta-hairpin context, this motif can be formed by the insertion of a residue just after the chain-reversing turn of a 2:2 hairpin.
6C2(+): More common than (-) variant. Bulge projects above the turn plane; the turn is most commonly of type I. 6C2(-): Bulge projects below the turn plane; the turn is most commonly of type II.	***BB H-bonds***: 1>+2, 4>1, !1>*, !2>*, !3>*, !4>*, !+1>*, !+2>*, !*>1, !*>2, !*>3, !*>4, !*>+1, !*>+2	5C1 BBL with an additional residue inserted into its bulge.
7C3(-): More common than (+) variant. Bulge projects below the plane of the loop's N-terminal turn. 7C3(+): Bulge projects above the plane of the loop's N-terminal turn; the turn is of type I.	***BB H-bonds***: 1>+3, 4>1, !1>*, !2>*, !3>*, !4>*, !+1>*, !+2>*, !+3>*, !*>1, !*>2, !*>3, !*>4, !*>+1, !*>+2, !*>+3	6C2 BBL with an additional residue inserted into its bulge.
6C1: Type 2 in the existing classification. These criteria select the "rubredoxin knuckle"[24], a metal ion-binding variant; to select the general loop, drop the sequence motif criterion.	***Sequence motif***: C-1C3 ***BB H-bonds***: -1>+1, 4>-1	5C1 BBL with a residue inserted into its beta turn, forming an alpha turn. In this variant, a Cys pair binds metal ions (commonly Zn) in contexts that include Zn fingers and Zn ribbons.

Note: to view a set of BBL structures in ExploreTurns, enter the selection criteria for the type and click ***Load Turns Matching Criteria***. Successive clicks of ***Browse Structures*** then step through the loaded set by the interval specified in the ***Stepsize*** box (negative values specify back-steps), displaying each structure along with its profile information. To change the radius for the display of structure external to the turn/tails, enter a new value in the ***Radius*** box and click ***Browse Structures*** again. Before entering the selection criteria for a new set of structures, clear the criteria from any previous set by clicking ***Clear All Criteria***.

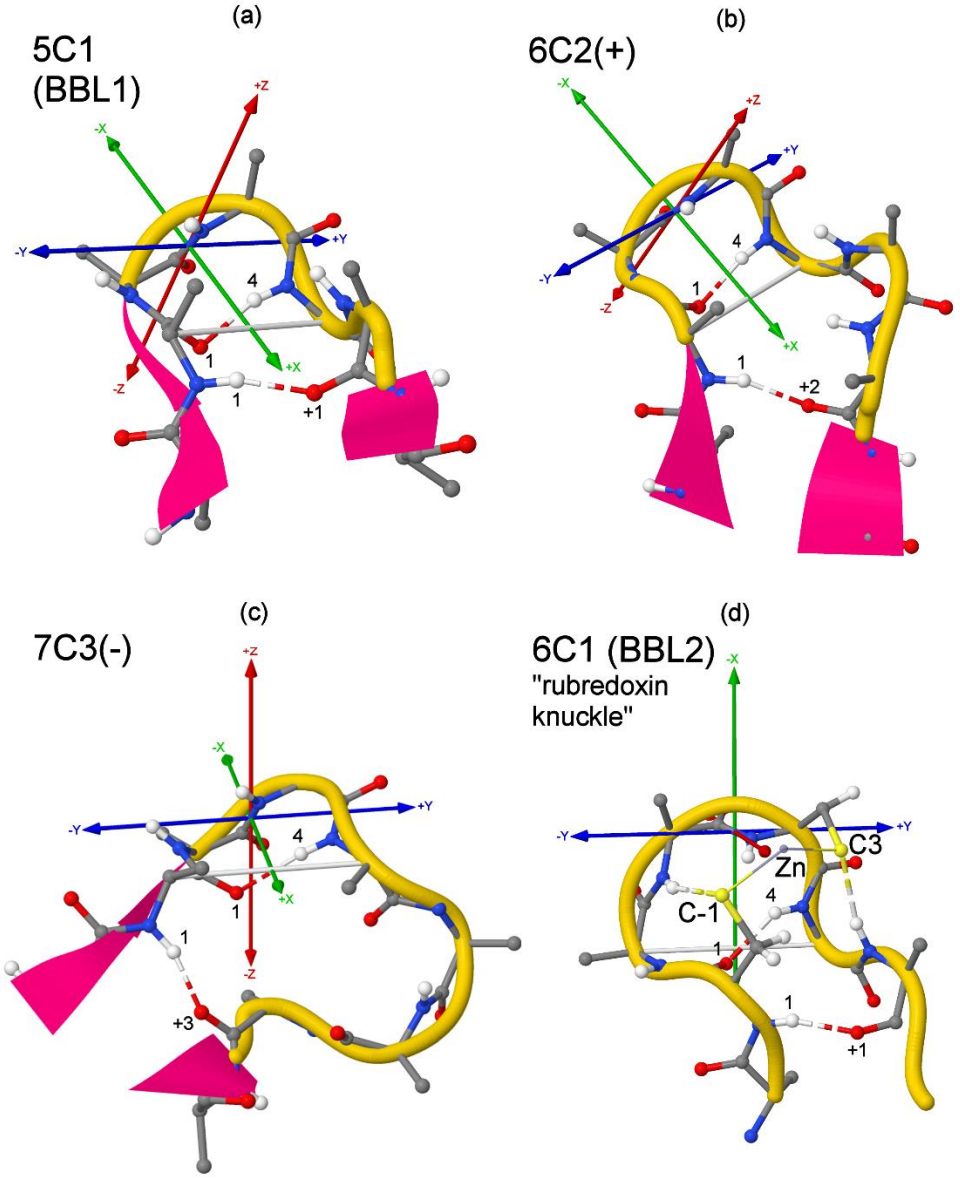

Figure 2. Generalized beta-bulge loop types with C-terminal bulges. Structures are labelled using a nomenclature that specifies overall loop length, bulge location (N/C terminal), bulge length, and (for the 6C2 and 7C3 types) whether the bulge projects above (+) or below (-) the plane of the loop's N-terminal beta turn. Loop-defining H-bonds are labelled at their ends with residue positions in the frame of the selected turn. To view a structure in ExploreTurns, load the entire turn database by clicking ***Load Turns Matching Criteria*** without any selection criteria, then copy the structure's PDB address to the ***PDB address*** box and click ***Browse Structures***. See Table 1 for the selection criteria for sets of structures of each BBL type. **(a)** 5C1 (type 1 in the existing classification, 5MBX_A_265). **(b)** 6C2(+) (4HTG_A_278). **(c)** 7C3(-) (2OXG_B_51). **(d)** 6C1 (type 2 in the existing nomenclature), shown here in its "rubredoxin knuckle" variant[24], found in contexts including Zn fingers/ribbons, in which a Cys pair binds metal ions (4XB6_D_242).

As in Table 1, the BBLs in Table 2 are selected with BB H-bond criteria that contain exclusions to rule out additional H-bonds, specifying only baseline structures, and the geometries of BBLs with particular additional BB H-bonds can be explored by overriding these exclusions.

BBL motifs, and particularly the generalized types with more than one residue in their bulges, combine the constraints enforced by their BB H-bonds with the flexibility of their bulges, giving them versatility for structural and functional roles, including ligand binding. Figure 4 shows six examples of generalized BBL types at binding/active sites, with ligands that include metal ions, a GTP complex, a FeS cluster and DNA.

ExploreTurns' wildcard motif detection tools can be used to identify and rank the most significant single-AA and pair sequence motifs in each BBL type (see example 20 in the primer).

Table 2. Selection criteria for beta-bulge loops with N-terminal bulges.

BBL type	Selection criteria	Notes
5N1	***BB H-bonds***: -1>4, 4>1, !-1>*, !1>*, !2>*, !3>*, !4>*,!*>-1, !*>1, !*>2, !*>3, !*>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	***BB H-bonds:*** -2>4, 4>1, !-2>*,!-1>*, !1>*, !2>*, !3>*, !4>*,!*>-2,!*>-1, !*>1, !*>2, !*>3, !*>4	5N1 BBL with an additional residue inserted into its bulge.
7N3	***BB H-bonds***: -3>4, 4>1, !-3>*,!-2>*,!-1>*, !1>*, !2>*, !3>*, !4>*,!*>-3,!*>-2,!*>-1, !*>1, !*>2, !*>3, !*>4	6N2 BBL with an additional residue inserted into its bulge.
7(N/C)3	***BB H-bonds***: 1>+3, 4>1, +3>4, !1>*, !2>*, !3>*, !4>*, !+1>*, !+2>*, !+3>*, !*>1, !*>2, !*>3, !*>4, !*>+1, !*>+2, !*>+3	Since this loop contains split pairs of BB H-bonds at both termini, it is both N- and C-type. Note that this BBL is selected in the frame of its N-terminal beta-turn.

Note: for instructions on browsing/profiling structures, see the note below Table 1.

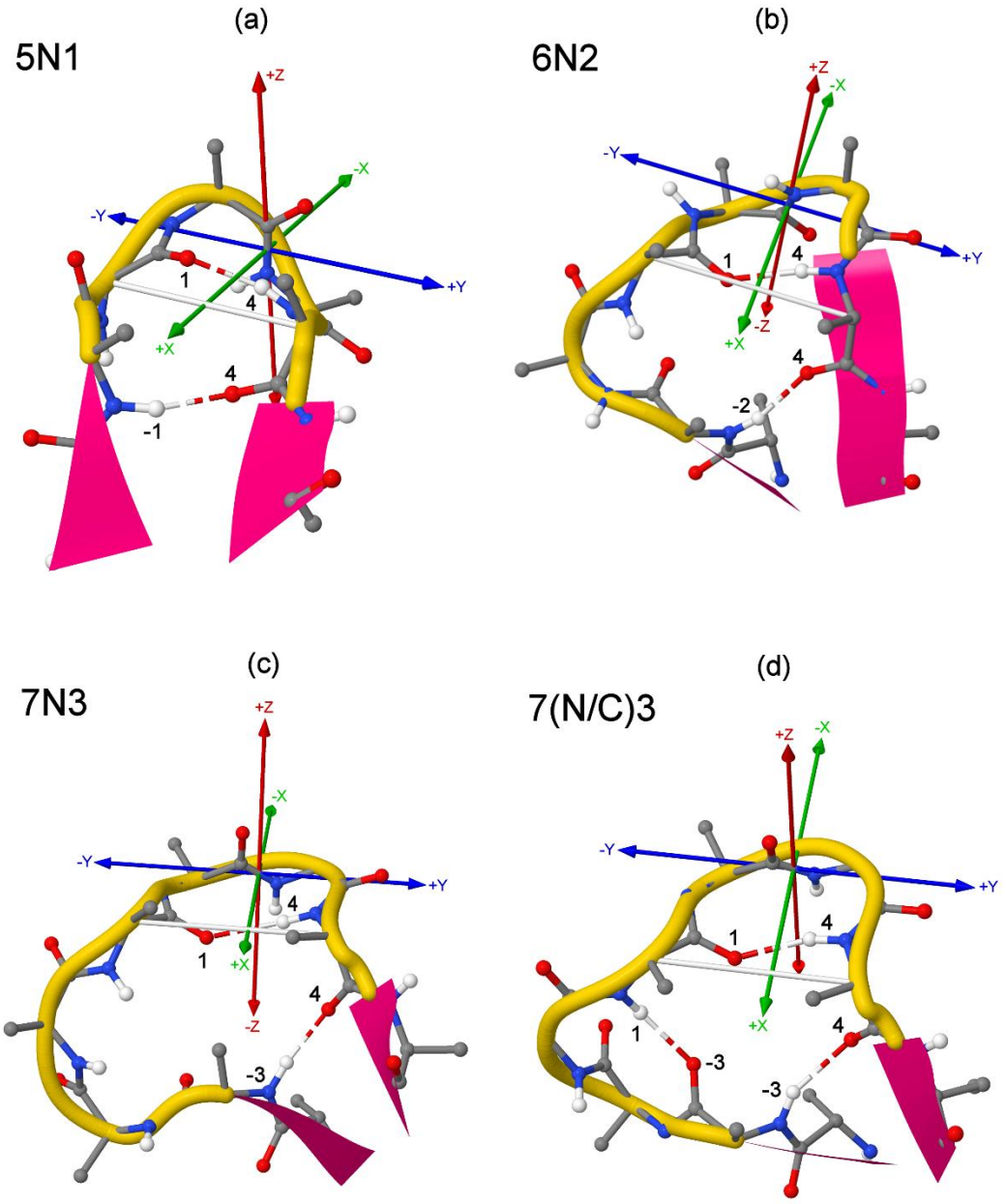

Figure 3. Examples of beta-bulge loops with N-terminal bulges. Table 2 lists the criteria needed to select sets of these structures in ExploreTurns. See the Figure 2 caption for loop nomenclature and structure viewing directions. **(a)** 5N1 (4R1D_B_107). **(b)** 6N2 (4PXW_A_1104). **(c)** 7N3 (5KAZ_A_36). **(d)** 7(N/C)3, which is both N- and C-type (3W6W_A_182).

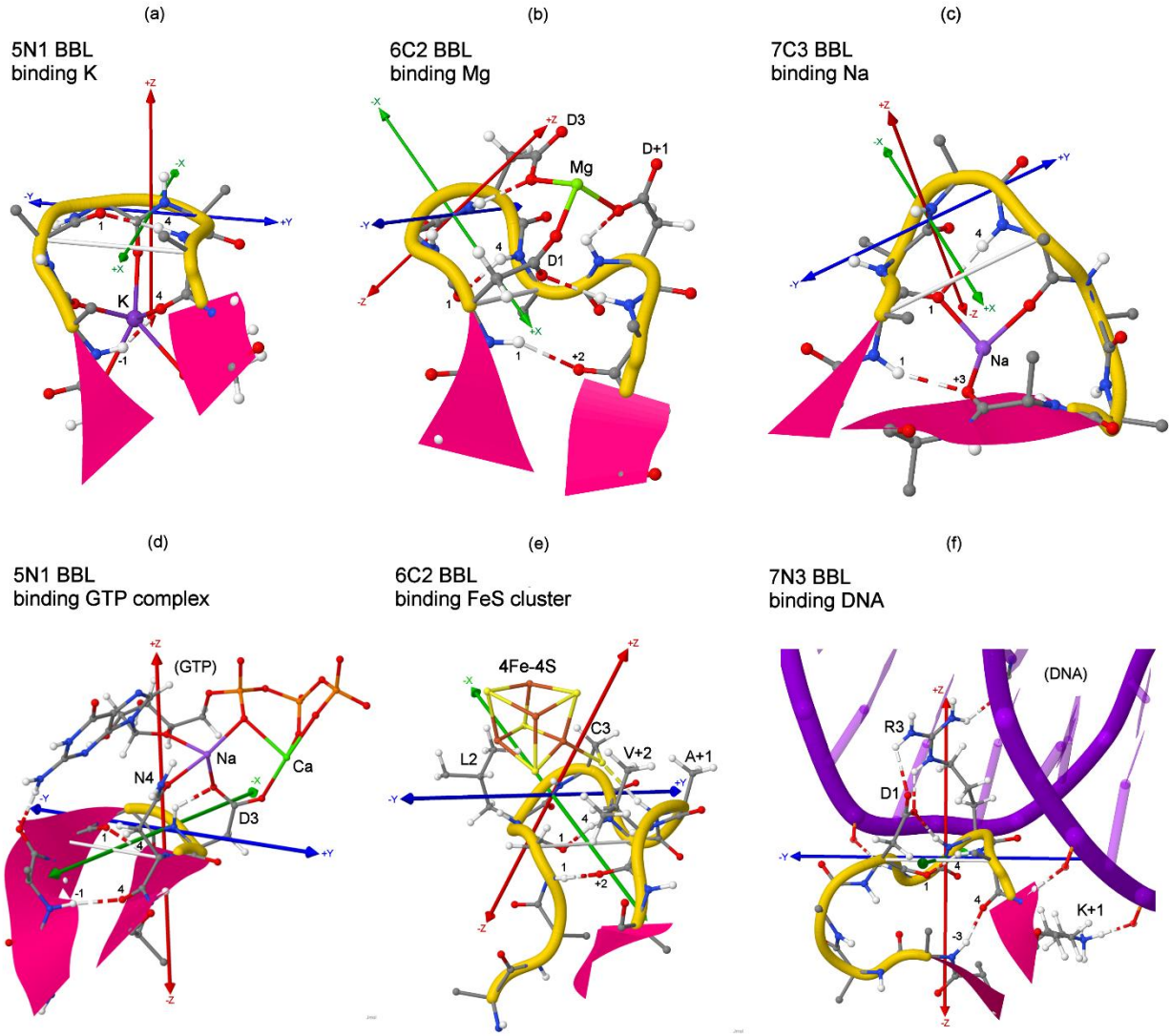

Figure 4. Examples of generalized BBL types at ligand binding/active sites. See the Figure 2 caption for structure viewing directions. **(a)** A 5N1 BBL binds potassium in the fatty-acid binding protein AtFAP1 (4DOO_A_35). **(b)** A 6C2 BBL binds Mg with an Asp triplet (D1, D3, and D+1, with positions indexed in the turn frame) at the active site of phosphoglucomutase from Xanthomonas citri (5BMN_A_259). **(c)** A 7C3 BBL binds Na in methionine R-sulfoxide reductase B1 (3MAO_A_60). **(d)** A 5N1 BBL binds a GTP/Na/Ca complex with BB H-bonds, Asp3 and Asn4 (D3, N4) at the active site of GTP cyclohydrolase III (2QV6_A_181). **(e)** 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). **(f)** 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).

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 or ST N-caps[17,19,21,22]. Beta turns are also found at the helical N-terminus[19], 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. Context vectors in ExploreTurns specify the approximate local directions of the turn's N- and C-tails as they point away from the ends of the turn, and when a helix begins within two residues of the turn, tail direction is measured by the orientation of the helical 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 *Context vectors* in the user guide). In this analysis, sets of structures are selected by specifying C-tail context vectors at 10° intervals of longitude and latitude. The ***Vector tolerances*** field is used to restrict the selected structures to those lying within 5° of a specified context vector, and Figure 5 plots the abundance and overrepresentation of the Asp3 motif (D3, see example structure in Figure S5b) for each orientation.

Figure 5 shows that D3's peak overrepresentation occurs at zero longitude, with the helical 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 axis 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 fall off abruptly in the low latitudes but remain high at higher latitudes, extending out to 50° longitude. At longitudes past 50°, overrepresentation drops as the helix rotates downwards, crossing below the plane, and its face becomes inaccessible to D3's SC.

Figure S1 maps abundance and overrepresentation for the Asn3 N-cap motif, and shows that peak overrepresentation occurs not at zero longitude, as seen for Asp3, but straddling the meridian at longitudes of +/-10°, and the maxima are more focused on

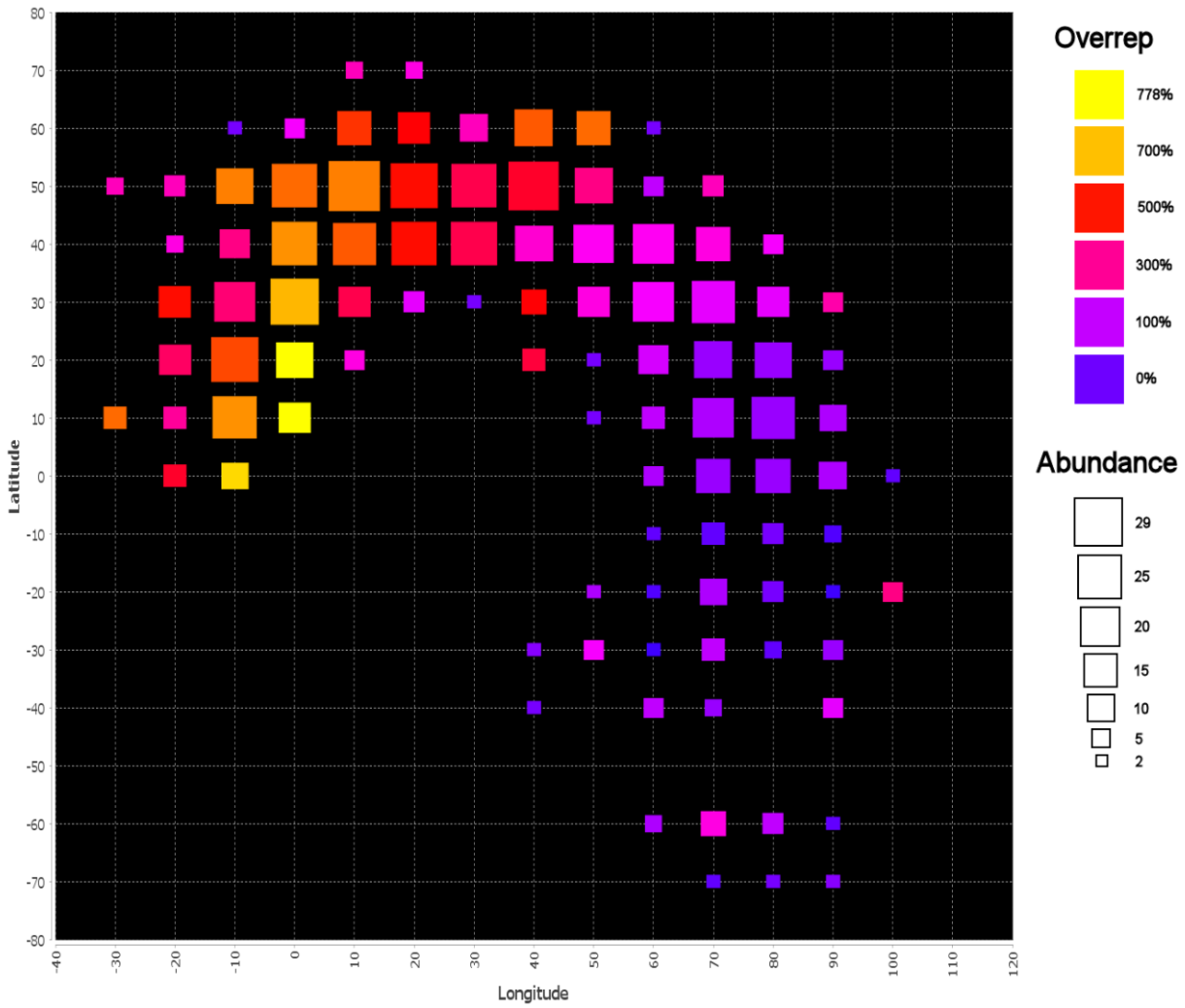

Figure 5. Fractional overrepresentation and abundance vs. helix/turn orientation for the Asx N-cap sequence motif that specifies Asp at turn position 3. Beta turns in which the third turn residue lies at the NCap position (just before the helix) are selected using the ***DSSP symbols*** entry *****<***H>*****, 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 helical axis) using ***Context vectors*** entries of the form **,*<>longitude,latitude*. The ***Vector tolerances*** entry **<>5* is applied to select structures with helical 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. 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. The map shows that D3's overrepresentation peaks at zero longitude (parallel to the xz plane), with the helix lying close to the turn plane, and remains high at that longitude as the helix rotates up to 50° above the plane. Overrepresentation and abundance fall off rapidly as the helix rotates into negative longitudes past -10° across the mouth of the turn, but persist, favoring higher latitudes, as it rotates away from the turn. At longitudes past 50°, overrepresentation drops as the helix rotates downwards and crosses below the plane and its face becomes inaccessible to D3's SC. Note that not all orientations with high overrepresentation show frequent SC capping of helical BB groups, as Asp's hydrophilicity is likely also favorable, and SC/SC H-bonding (including interactions with the helix) and non-capping SC/BB H-bonding may also contribute.

latitudes near the turn plane. Figures 5 and S1 can serve as guides for ExploreTurns investigations of the structure and H-bonding associated with Asx N-caps in beta turns.

3.3 Helix-terminal Schellman loop/beta turn conformations

Section 3 in Supplementary Material describes the use of ExploreTurns to profile the geometries of Schellman loop/beta turn combinations at the helical C-terminus. Four principal conformations are identified, and Figure S2 presents example structures and selection criteria for each.

3.4 Depth dependence of turn geometry

In section 4 in Supplementary Material, ExploreTurns is used to investigate the relationship between the distribution of turns across the classical types or BB clusters and depth beneath the solvent-accessible surface. The interpretation of the results is complicated by the effects of SC-related depth bias, but results for both classification systems show a trend that is unlikely to be produced by bias: the proportion of structurally unclassified turns increases markedly with depth, largely at the expense of the type I or type I-like geometries. This effect may result from distortions due to increasing contacts with structures external to turns as they become more deeply embedded in a protein, and may also reflect increasing turn involvement in local structures such as ligand-binding sites.

3.5 Exploring, profiling and tuning structural motifs

Section 5 in Supplementary Material provides selection criteria and example images for a selection of motifs which incorporate beta turns, including beta hairpins, local BB H-bonded motifs (including one motif, dubbed the "beta clip", that has not previously been described), supersecondary structures and sequence motifs.

3.6 Viewing parameter ranges in structures

Section 6 in Supplementary Material illustrates the modes of geometric variation described by the geometric turn parameters using examples of turns with parameter values near their limits in the database.

4 | Conclusions

ExploreTurns is a web-based graphical tool that supports the comprehensive exploration, analysis, and retrieval of beta turns and the structures that incorporate them. The facility combines Euclidean- and Ramachandran-space features and renders a redundancy-screened, PDB-derived database of beta turns and their contexts structurally addressable by a wide range of descriptors.

A turn-local coordinate system provides a common framework for the representation of structure, while geometric parameters characterize the bulk BB shapes of turns and complement the Ramachandran-space-based classification systems, enabling the selection of turn conformations within classical types or BB clusters for compatibility with particular SC interactions or contexts.

Context vectors support the selection of supersecondary structures linked by beta turns into particular geometries, or structures in which the BB adjacent to the turn is oriented to optimize interactions, such as helix capping H-bonds, which occur between turns and their BB neighborhoods.

As demonstrated here, ExploreTurns is well-suited for research, and it 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. The tool should prove valuable in applications such as protein design, in which an enhanced Euclidean-space description of turn geometry along with structural tuning may improve performance.

5 | Methods

Beta turns are defined here as four-residue BB segments with a distance of no more than 7Å between the alpha carbons of the first and fourth residues and central residues with DSSP codes outside the set {H, G, I, E} that specifies helix or strand[3]. Turn/tail structures were extracted from their PDB files[10] with the aid of a list of peptide chains with a maximum mutual identity of 25% obtained from PISCES[25]. The final, redundancy-screened dataset of 102,192 turns was compiled with resolution and R-value cutoffs of 2.0Å and .25 respectively (average values in the dataset are 1.6Å and .20), and turn structures were further screened for quality by requiring that there be no missing residues within 5 residues of any turn and excluding structures with bond lengths that constituted outliers from established values.

Biopython[26] was used to extract the structural data, which was then processed in Java[27]. Measurements of the depth of beta turns beneath the protein's solvent-accessible surface were computed with Depth[28]. ExploreTurns was written in HTML/CSS and Javascript, with an embedded JSmol viewer[29]. The tool is tested for compatibility with the Chrome, Edge, and Firefox browsers; Chrome and Edge provide the best performance.

The turn-local coordinate system[9], the Euclidean-space alignment which it implicitly establishes, the methods used to compute sequence motif overrepresentation and p-value[9,19,30,31], the derivation of context vectors, and the H-bond definition are all described in the online user guide, while turn parameters are described in dedicated online help.

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.

Supplementary Material

Supplementary file 1 (PDF)**.** Supplementary sections.

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) Hutchinson EG, Thornton JM. A revised set of potentials for β-turn formation in proteins. *Protein Science*. 1994;3(12):2207-16.

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

6) 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.

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

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

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

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

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

12) Sibanda BL, Blundell TL, Thornton JM. Conformation of β-hairpins in protein structures. *Journal of Molecular Biology*. 1989;206(4):759-77.

13) 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.

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

15) 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.

16) 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.

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

18) 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.

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

20) 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.

21) 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.

22) 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.

23) 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

24) 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*. 2003;278(22):20225-20234.

25) Wang G, Dunbrack RL. PISCES: A protein sequence culling server. *Bioinformatics*. 2003;19(12):1589-91.

26) 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.

27) Arnold K, Gosling J, Holmes D. The Java Programming Language. 4th ed. Boston. Addison-Wesley Professional; 2005.

28) 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.

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

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

31) 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.