ExploreTurns: A web tool for the exploration, analysis, and classification of beta turns and structured loops in proteins; application to beta-bulge and Schellman loops, Asx helix caps, beta hairpins 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. In this work, a new turn-local coordinate system and structural alignment, together with a set of geometric descriptors for turn backbone shape, are incorporated into ExploreTurns, a web facility for the exploration, analysis, geometric tuning and retrieval of beta turns and their contexts which combines the advantages of Ramachandran- and Euclidean-space representations. Due to the prevalence of beta turns in proteins, this facility, supported by its interpreter for a new general nomenclature which classifies H-bonded loop motifs and beta hairpins, serves as an exploratory browser and analysis tool for most loop structure. The tool is applied to the detection of new H-bonded loop motifs, including a "short Schellman loop" and a large family of beta-bulge loops with a range of geometries and H-bond topologies. Other applications presented here include the mapping of sequence preferences in Asx helix N-caps and an investigation of the depth dependence of beta-turn geometry. ExploreTurns, 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 and 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, beta hairpin, sequence motif, H-bonded loop, compound-turn nomenclature, short Schellman loop, beta bracket, pi-alpha loop, beta-gamma loop

1 | Introduction

Of the five reported 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. Either individually or in overlapping multiples[3], beta turns constitute close to two-thirds of all loop residues[4]. In short loops, turns can link helices or strands together into "supersecondary" structures which contribute to a protein's tertiary fold, and turns are also common in the extended loops of intrinsically disordered regions, where they may form phase-separation nucleation points[5].

Beta turns are also features of beta sheets, for which they can form entrance or exit structures or all or part of the chain-reversing loops between the antiparallel strands of beta hairpins[6,7,8], and turns also occur in beta bulges[9,10], which add irregularities at the edges of sheets that hinder the aggregation associated with the plaque diseases[11]. Beta turns are also found in beta-bulge loops[12] (BBLs), which occur when beta bulges are found at the loop ends of beta hairpins, but also exist as independent structures. The occurrence of beta turns in multiple

contexts associated with beta sheets suggests that mutations within turns may be a driver of plaque formation, a possibility highlighted by the finding that cross-alpha crystals can be converted to cross-beta fibrils by flipping a beta turn's type[13].

Beta turns are associated with helices as well as sheets; they are found at both ends of helices, where they can "cap" unsatisfied BB polar groups with H-bonds[14,15]. The six-residue Schellman loop[16,17] is a common C-terminal capping structure that can incorporate or overlap a beta turn, and a seven-residue "wide" form[18] has also been identified. Beta turns can also interrupt helices, forming helix kinks[19] that introduce design flexibility and allow a greater variety of SC placements[20].

Other structures in which beta turns occur include beta helices[21] and other solenoidal folds, beta[22] and TIM[23] barrels, beta propellers[24], and multiple local motifs such as nests[25], Asx and ST turns/motifs[26,27] and ST staples[28]. Beta turns are components of all reported local H-bonded BB structures, including the alpha-beta[28] and crown bridge[29] motifs in addition to the Schellman, wide Schellman and beta-bulge loops.

Beta turns may play important roles in protein folding, including as nucleation sites, established by stable intra-turn interactions, which help initiate folding and enforce chain registration[30-34], and it has been shown that interactions involving beta-turn SCs can make a substantial contribution to overall protein stability[35], and that the stability of proteins can be increased via AA substitutions in beta turns[36].

Since beta turns often change the direction of the BB as it reaches the protein surface, they are frequently exposed to the surrounding solvent, and polar, hydrophilic SCs are overrepresented in turns. The common occurrence of turns at the protein surface also positions

them well to interact with external ligands, including other proteins, and turns also participate in ligand binding within the protein body. Key beta-turn binding partners include metal ions, ADP/ATP, nucleic acids and heme groups. Beta turns play important roles in determining the conformation and specificity of enzyme active sites and antibody-combining sites[37].

Beta turns are found in peptides as well as proteins; they can provide chain-reversal in the beta hairpins of cyclic antimicrobial bioactive peptides[38], and it has been postulated that in many cases they represent the bioactive conformation of linear peptides[39]. Chemically modified beta-turn mimics can serve as asymmetric catalysts[40]. Beta turns and their mimics are employed in drug development studies[41]; it has been demonstrated that beta-turn mimics can inhibit plaque formation[42].

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[4,43]). The beta-turn definition used here[4] describes a four-residue BB segment in which the alpha carbons of the first and last residues are separated by no more than 7 Å 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 (4>1 H-bond). ExploreTurns employs three beta-turn classification systems which specify BB geometry at different levels of precision in the Ramachandran space of BB dihedral angles; in increasing order of precision, these systems are referred to here as the "classical" types[44], BB clusters[4] and Ramachandran types[4,45,46] (see the ExploreTurns online help for definitions).

As far as the author is aware, the only currently available tools that specifically support the study of beta turns are the web facility Motivated Proteins[28] and its associated stand-alone

application Structure Motivator[47], which present turns as one of multiple small H-bonded motifs, including SC structures. ExploreTurns has a different focus, taking a beta-turn centered approach to the exploration of loop structure in a redundancy-screened, PDB-derived database. The tool incorporates multiple new features, including a turn-local Euclidean-space coordinate system and structural alignment for all beta turns (see Sections 5.2), a set of geometric descriptors[48] that describe the bulk BB shapes of turns (Section 5.3), statistical tools for the detection of SC motifs, and other features. In addition, ExploreTurns is the first tool to encompass beta turns without the 4>1 H-bond, which comprise about half of the ExploreTurns database and include classical types VIa2, VIb and VIII, as well as 13 of the 18 BB clusters. Also new are ExploreTurns' integration of structure selection, profiling, and 3D viewing onto a single web page and its support for rapid comparative structure browsing.

ExploreTurns' turn-local coordinate system provides a common framework for the visualization and comparison of the full range of beta-turn BB geometries, and also supports analyses of the many recurrent contexts that incorporate turns, including local BB structures, ligand-binding and active sites and supersecondary structures. The tool's incorporation of geometric turn descriptors 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[48]. When used in the tool, descriptors enable structural discrimination within the classical types or BB clusters, yielding major improvements in the specificity of the measurements of sequence preferences in turns (which the tool computes), and enabling the Euclidean-space 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[4], and ExploreTurns also encompasses the four-residue BB segments N- and C-terminal to 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 short loops shaped by multiple BB H-bonds into local structural motifs such as the Schellman, alpha-beta and beta-bulge loops. This work introduces a general description of all of these H-bonded loop motifs as "compound turns" composed of overlapping "simple" H-bonded turns of all types (lengths), and defines a nomenclature which classifies the motifs by their content of simple turns. ExploreTurns includes an interpreter for this notation which is applied here to explore the structures of H-bonded loop motifs and their variants and detect new motifs, including a "short Schellman loop" and a large family of motifs that satisfy a generalized definition of the beta-bulge loop and exhibit a range of geometries and H-bond topologies. The compound-turn nomenclature is also applied here to the chain-reversing loops of beta hairpins, where the notation provides a hairpin classification that is more complete than the X:Y hairpin nomenclature[8], since it can describe the H-bond configuration within a hairpin's loop as well as the configuration at the loop end of the hairpin's beta ladder.

ExploreTurns is also applied to investigate the relationship between beta-turn geometry and depth beneath the solvent-accessible surface, map sequence preference vs. cap geometry in Asx helix N-caps, and profile the conformations of helix-terminal Schellman loop/beta-turn combinations.

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 supports the analysis and retrieval of structures from a redundancy-screened, PDB[49]-derived dataset of 102,192 beta turns and their four-residue BB neighborhoods (their "tails"), using a wide range of selection criteria. Criteria include turn BB geometry classified in the Ramachandran space of BB dihedral angles (the classical types, Ramachandran types and BB clusters), ranges of the geometric turn descriptors ("span", N- and C-terminal "half-spans", "bulge", "aspect", "skew", and "warp"; see Section 5.3), inclusionary/exclusionary patterns of BB H-bonds, BB dihedral angles, the *cis/trans* status of the turn's peptide bonds, sequence motif content, the electrostatic energy of the turn's potential BB H-bond, DSSP[50] secondary structure codes, the approximate orientations of the tails with respect to the turn (the "context vectors"), and the depth of the turn beneath the protein's solvent-accessible surface. Required patterns of BB H-bonds in structures can be specified either in the position frame of the beta turn at the center of the 12-residue ExploreTurns structural window, in the frame of an H-bonded loop motif via the compound-turn notation, or via descriptive shorthand names for loop motifs. A summary of all ExploreTurns selection criteria, with examples of their use and links to the relevant help pages, is accessible from the tool page via the "Feature Summary/Examples" button.

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 plots of the distributions of the geometric descriptors, browsing reveals the modes of Euclidean-space structural variation characterized by the descriptors. Plots of sequence motif overrepresentation vs. descriptor value allow a user to "tune" the selected structures for compatibility with particular SC interactions, by adjusting descriptor ranges to maximize the fractional overrepresentation of the sequence motifs associated with the 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 supersecondary structures with particular orientations between their helix/strand components (see Supplementary Section 7), or identify the BB geometries that optimize interactions, such as helix-capping H-bonds, that occur between turns and adjacent structures (Supplementary Section 3).

ExploreTurns supports the analysis and retrieval of sets of examples of any structural motif that can be defined by the tool's selection criteria operating on its twelve-residue window centered on a beta burn, including beta hairpins from the four classes[6,7,8], supersecondary structures, some beta bulges[9,10] and nests[25], and all H-bonded loop motifs that occur within four residues of a beta turn, including all previously reported motifs as well as multiple new types.

The tool also gives access to sequence motifs defined by any combination of amino acids (AAs) at any position(s) in the ExploreTurns window, including motifs which form characteristic recurrent SC structures, which are referred to here as SC motifs. The most common SC motifs

include Asx/ST turns/motifs[26,27] and Asx/ST helix caps[14,15,26,27], 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 SC motifs 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[51] tool (also available at www.betaturn.com), which enable the comprehensive, interactive 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 screen

The top half of the ExploreTurns screen (Figure 1) is divided vertically into three sections. The buttons at the top of the screen provide access to extensive online help, including an introduction, a feature summary with examples, an in-depth user guide, and definitions of the compound-turn nomenclature, BBL shorthand notation, turn types, BB clusters and geometric descriptors. The ***Explore Loop Motifs*** button opens a browser for H-bonded loop motifs and beta hairpins which includes schematics of structure and H-bonding as well as links which load sets of motif structures into the tool.

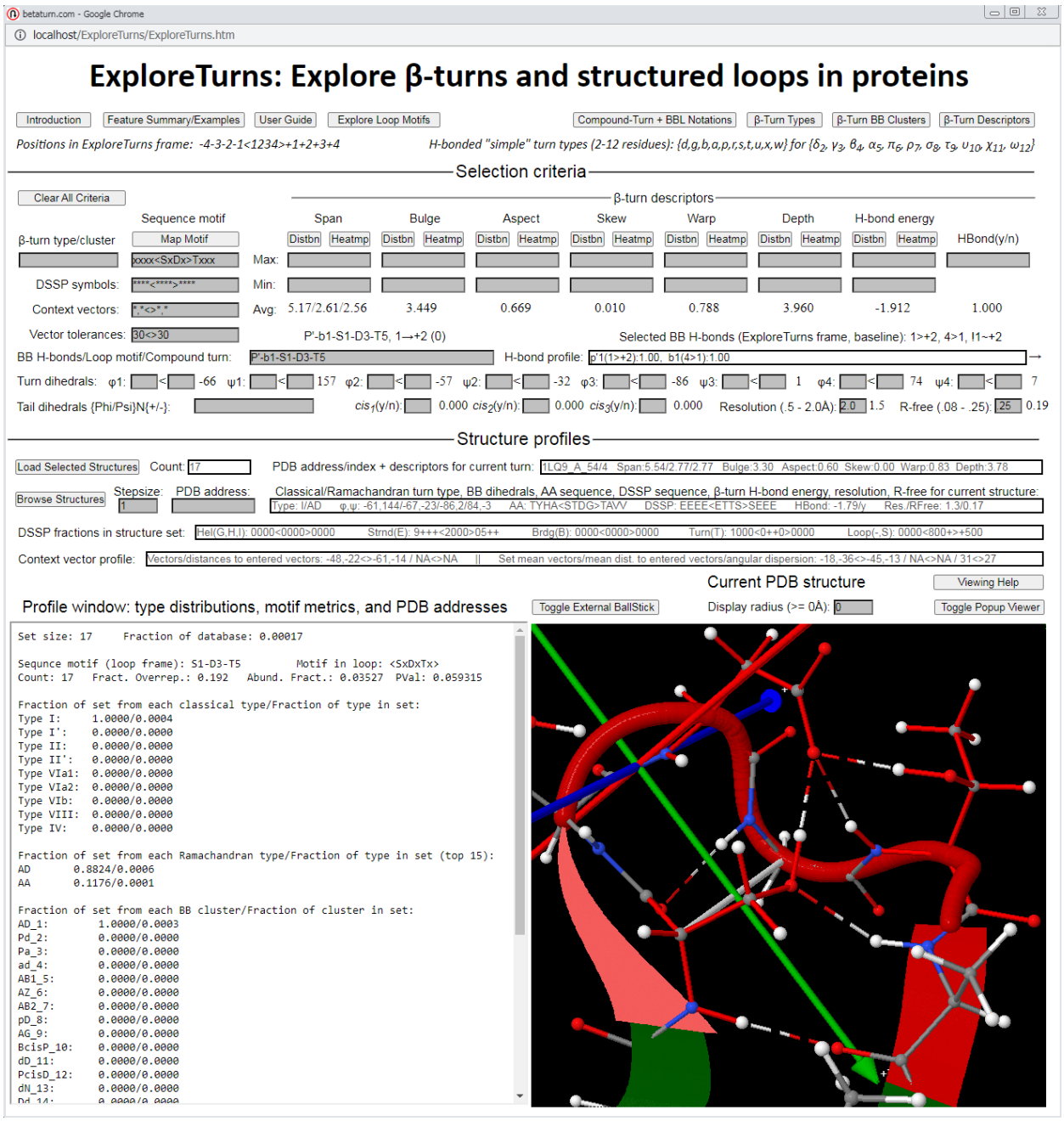

Figure 1. The ExploreTurns screen. The compound-turn label *P'-b1-S1-D3-T5* (see sections 5.4 and 5.5 or click *Compound-Turn + BBL Notations*) has been entered into the *BB H-bonds/Loop motif/Compound-Turn* box to select the sequence motif that specifies Ser at position 1, Asp at position 3 and Thr at position 5 in the 6C2 beta-bulge loop (BBL), one of the new BBL motifs described here (Section 3.1.6). The structure is displayed in the turn-local coordinate system, with a white bar drawn between $C_{\alpha 1}$ and

$C_{\alpha 4}$ in the β-turn to show the turn's span. The sequence motif's SCs and the loop's cartoon backbone are highlighted in red. The radius for the display of structures outside the 12-residue ExploreTurns window (which is always centered on a beta turn) has been set to zero in the *Display radius* box to remove clutter.

The *Selection criteria* section in the middle of the top half of the tool page is a database selection screen which accepts input in the shaded boxes. Average values of numerical criteria for the selected structures are displayed beneath or to the right of the input boxes when a set of structures is loaded. Also displayed, in the *H-bond profile* box, are the ranked occurrence frequencies of all BB H-bonds that occur in the turns/tails in the selected set, with the H-bonds labelled by the positions of their donor and acceptor residues expressed in the frame of the beta turn at the center of the ExploreTurns window in the format "M>N", using the position key: -4-3-2-1<1234>+1+2+3+4 (where the brackets delineate the turn). When an H-bonded loop motif is selected using a compound-turn label (see sections 3.1.4, 5.4 and 5.5), the H-bond profile window displays only the H-bonds that occur within the motif, and the H-bonds are also labelled with their associated simple turn types, which can be copied into the compound-turn label to investigate the structures of the motif's BB H-bond variants.

The *Structure profiles* section at the bottom of the top half of the tool page profiles the structure from the selected set that is currently displayed in the viewer, reporting values for all selection criteria, and also displays information on the selected set of structures 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 turns in the set across the classical types, Ramachandran

types and BB clusters are displayed, along with ranked statistics for all single-AA sequence motifs that occur in the 12-residue ExploreTurns window (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 SC motif detection at particular positions or position ranges in the selected set (see *Sequence motif selection…/Examples/Ex. 3* in the feature summary). If an H-bonded loop motif is selected with a compound-turn label, ranked statistics for the single-AA and pair sequence motifs that occur in the loop are listed at the top of the profile window, and sequence motif labels can be copied into the compound-turn label to investigate the effects of SC interactions on the loop's structure. A list of PDB addresses for all structures in the selected set is displayed at the bottom of the profile window.

At bottom right on the page, a JSmol viewer displays the structures from the selected set in the turn-local coordinate system (see Section 5.2 and Figure 9) in the contexts of their PDB files. By default, structure in the PDB files external to the ExploreTurns window is displayed in translucent cartoon only, out to a radius of 30 Å from the turn, but the user can add a ball/stick representation and change the display radius.

2.3 Distribution plots

ExploreTurns provides three types of distribution plots which involve the geometric turn descriptors, H-bond energy or depth beneath the solvent-accessible surface, and cover the classical types, BB clusters or the global turn set. Distribution plots are available via the *Distbn* and *Heatmp* buttons in each descriptor's column. Interpolated histograms of the distributions

of descriptor values within types and clusters are provided in order to guide structure selection by descriptor value. Also available, 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 overrepresentation vs. descriptor value, which identify the descriptor regimes most compatible with each motif (and therefore most likely to be compatible with any associated SC motif), and support the tuning of structure sets for compatibility with particular SC interactions (see *Sequence motif selection…/Examples/Ex. 4* in the feature summary). Lastly, Ramachandran-space scatterplot heatmaps chart the distributions of each descriptor across the dihedral-angle spaces of the two central turn residues, identifying the bond rotations that drive descriptor variation.

3 | Applications

3.1 Classifying and exploring H-bonded loop motifs

3.1.1 Roles of H-bonded loop motifs

Helices and sheets, stabilized by repetitive patterns of BB H-bonds, provide the extended "scaffolds" of protein structure, but there is also a role in protein architecture for local H-bonded motifs which complement the repetitive structures by shaping the BB into irregular local conformations which can serve as "accessories" to the scaffold. Motifs such as the Schellman, alpha-beta and crown bridge loops are often found at the ends of helices, where they can stabilize the helix and shape its entrance and exit loops, contributing to the geometries of supersecondary structures (and, by extension, to a protein's overall fold). The beta-bulge loops are also found in association with repetitive structure; they often occur at the

loop ends of beta hairpins, where they shape the BB for structural roles or support function in contexts such as ligand-binding or active sites (see Figure 8). In addition to their associations with helices and strands, H-bonded loop motifs play independent structural and functional roles; they are versatile structures which combine the constraints imposed by their BB H-bonds with a degree of conformational flexibility, filling a niche intermediate between the "simple" turns characterized by a single H-bond and the highly constrained repetitive structures.

3.1.2 H-bonded loop motifs as "compound turns"

The BB geometry of a simple H-bonded turn is restricted by the constraint imposed by its H-bond to a set of conformations characteristic of its type (length), and loops formed of overlapping H-bonded turns[3] are constrained to some degree along their full lengths by overlapping H-bonds (when the overlap is sufficient). All reported H-bonded loop motifs in proteins are formed of overlapping H-bonded turns, and although the component turns in these motifs may be shaped by the H-bonds associated with their overlapping partners, each turn remains an example of its type, since its own constraining H-bond is still present. Accordingly, H-bonded loop motifs can be described as "compound turns" and classified by the types of their component simple turns and the positions of those turns within the motifs. Before formally defining compound turns and introducing a nomenclature which classifies them and supports their exploration in ExploreTurns, it is helpful to review what is known about their component simple turns.

3.1.3 Review of "simple" H-bonded turns

H-bonded "tight" turns are classified into the types {δ, γ, β, α, π}, with lengths of two to six residues. The four longest types exhibit sub-types in which the H-bond that links the turn's

terminal residues is donated in either the C→N ("normal") or N→C ("reverse") direction[1] (delta turns are found only in the reverse form), and the normal and reverse turn types are further classified based on the BB dihedral angles of their central residues.

Figure 2 presents example structures and schematics for the five simple turn types. The two-residue, H-bonded delta turn (Figure 2a), found only in reverse form, is extremely rare: just six nonredundant prototype structures were identified in proteins in a 2015 study[52]. Four of these examples exhibit "p*cis*P" Ramachandran-space conformations, with 'p' (+φ polypro) and 'P' (-φ polypro) conformations at their first and second residues respectively, and their single peptide bonds in *cis* conformation (click *β-Turn Types* on the ExploreTurns page for a map of Ramachandran space). The p*cis*P conformation, shown in figure 2, is also the most common delta-turn type in the ExploreTurns database, with four examples out of the 15 turn structures. However, while previous theoretical and observational studies have identified *cis* as the exclusive conformation for the delta-turn peptide bond[52], six of the delta turns in the ExploreTurns database exhibit *trans* peptide bonds, and three of these (4PDX_A_416, 4QHW_A_253 and 5HDM_A_219) show no outliers in the PDB's validation plots[53].

The three-residue, H-bonded gamma turn (Figure 2b) is found in either a "classic" conformation (shown), with its central residue falling in the 'g' region of Ramachandran space adjacent to the left-handed helical region, or the "inverse" conformation, in which that residue lies in the 'G' region, related by inversion symmetry to 'g'[54,55].

The four-residue, H-bonded beta turn (Figure 2c) is found in abundance in five of the eight structurally classified classical beta-turn types[44], including types {I, I', II, II', VIa1}. Classical type I

(shown in the figure), with right-handed helical conformation at both central residues, contains about 60% of all examples.

Five-residue, H-bonded alpha turns (Figure 2d) have been grouped into three structurally classified types and an unclassified category. Alpha turns of the dominant structurally classified type (shown in the figure), which represent about three-quarters of all examples, exhibit right-handed helical conformation at all three central residues[56].

Six-residue, H-bonded pi turns (Figure 2e) have been grouped into eleven overall classes based on the conformations at their four central residues[57]. The dominant pi-turn conformation exhibits right-handed helical conformation at the first three of central residues and left-handed helical conformation at the fourth[58], as seen in the Schellman loop.

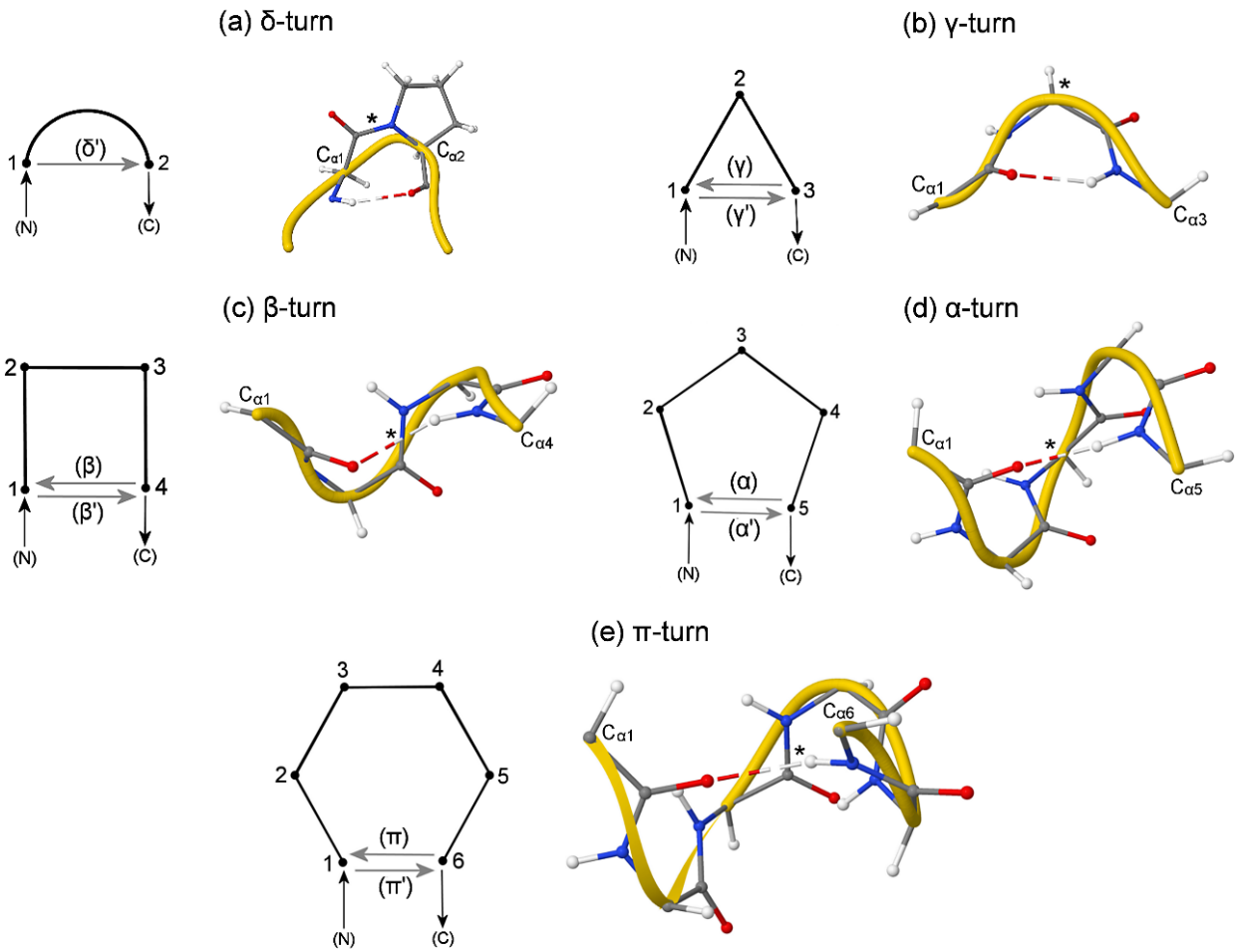

Figure 2. Schematics and example structures for the five types of "simple" H-bonded "tight" turns which overlap to form compound turns. Schematics and example structures are shown for H-bonded turns of types {δ, γ, β, α, π}, with lengths of 2-6 residues. Schematics show H-bonds as arrows oriented in the direction of H-bond donation: arrows pointing left represent the "normal"[1] turns of each type, with H-bond donation in the C→N direction, while arrows pointing right represent "reverse" (primed) turns, which are donated in the N→C direction. With the exception of the δ-turn, which exists only in the reverse form, all types are shown in their normal forms. Each turn is labelled at the α-carbons of its terminal residues, and its "turn center", which lies either in the center of its middle peptide bond (when the turn contains an even number of residues), or at the α-carbon of its central residue (when the turn

has an odd length) is marked with an asterisk. To aid in consistent orientation, a "turn plane" (not shown) is defined which passes through the three labelled points, and the relatively flat δ- and γ-turns are viewed from above this plane, while the longer turns are viewed approximately "edge-on" to the plane to best illustrate their conformations. **(a)** The two-residue reverse δ-turn, in its most common 'p*cis*P' conformation (2FYF_A_25). **(b)** The three-residue γ-turn, in its "classic" form, with 'g' conformation at its central residue (4D6Z_A_263). **(c)** The four-residue β-turn, in its most common geometry (classical type I), with right-handed helical conformation at both central residues (3LY1_A_228). **(d)** The five-residue α-turn, in its most common geometry, with right-handed helical conformation at all three central residues (3SUK_A_46). **(e)** The six-residue π-turn, in its most common geometry, with right-handed helical conformation at the first three central residues and left-handed helical conformation at the last central residue, as seen in the Schellman loop (1W07_A_361).

3.1.4 Compound-turn definition and nomenclature

The compound turn (CT) is defined here as a short loop incorporating two or more H-bonded turns, in which both of the loop's terminal residues are linked by a BB H-bond to at least one other loop residue, and all covalent bonds in the main-chain path between the loop's first and last H-bonded atoms lie within the H-bond arc of at least one H-bonded turn. This definition guarantees a degree of constraint across the entire loop by preventing free rotation around any internal covalent bond, while allowing "open" loops, in which the terminal residues are not linked by a shared BB H-bond (see the "Schellman+α" loop and "beta bracket" in Figure 3). In the more common, "closed" CTs, one or more shorter H-bonded turns lie embedded within the longest, "framing" H-bonded turn (they may share terminal residues with that turn). A CT's conformation is constrained to a degree which depends on factors including the ratio of

the number of BB H-bonds in the loop to the number of loop residues, the degree of overlap of BB H-bonds, and the presence of SC interactions, BB-constraining proline residues, or BB-freeing Gly residues in the loop. Recurrent CTs exhibit "baseline" forms, which contain only the BB H-bonds that define the motif, as well as BB H-bond variants, in which the structures are constrained by additional BB H-bonds, and structures can also be shaped by SC interactions into SC variants.

Including previously reported motifs and variants and those identified here, more than forty H-bonded loops are presented in this work, and the abundance and variety of loop types highlights the need for a nomenclature that classifies all CTs. In the concise notation introduced here (summarized in Section 5.4, or click *Compound-Turn + BBL Notations*), the Greek letters that specify the types (lengths) of a loop's component H-bonded turns are concatenated with dashes, along with the start positions within the loop of each component turn (except in the case of the loop's framing turn, which is listed first, without a position specifier). A component turn's letter is primed if the H-bond that defines it is donated in the N→C (reverse) direction along the chain, rather than the C→N (normal) direction, and H-bonds oriented in either direction can be specifically excluded by appending '*' to a component turn's entry, after any prime character and before the component turn's position specifier, which terminates the entry. The length of a loop is conveyed by the type of its framing turn: if this turn is specified with '*' exclusions that rule out both its normal and reverse forms, then it serves only to frame an open loop. To avoid confusion with the names of supersecondary structure motifs (e.g. βαβ), a CT label always concludes with the word "turn".

Since the existing classification for simple H-bonded turns only encompasses structures with lengths up to six residues, additional turn types are needed to cover motifs such as the wide Schellman loop[18] (seven residues) as well as the longer beta-bulge loops identified below, which contain up to eight residues. In addition, beta hairpins, which are classifiable by the lengths of their chain-reversing loops and the H-bond configurations at the loop ends of their beta-ladders using the X:Y nomenclature[8], can also be described in the CT notation (see Section 3.1.8 and the loop motif browser), and the availability of longer component turn types allows hairpins with longer chain-reversing loops to be described by the notation and selectable in ExploreTurns. For these reasons, the Greek letters following π are included in the notation, with the exception of φ and ψ, which are reserved for the specification of dihedral angles (see below). The complete set of turn specifiers is $\{\delta, \gamma, \beta, \alpha, \pi, \rho, \sigma, \tau, \upsilon, \chi, \omega\}$, which includes component turns with lengths of 2-12 residues, covering the full ExploreTurns window and enabling the classification of hairpins with loop lengths up to ten residues (the 10:10, or ω-ω' hairpin).

For loops which include constraints on their BB dihedral angles in their definitions (e. g. the crown bridge loop), or show conformational variants for which dihedral information is a useful discriminator (e. g. the Schellman and short Schellman loops), the CT notation adds dihedral-angle specifiers: 'φ' or 'ψ' is appended to the CT label for each constrained angle, with the constrained residue position(s) (in the loop's frame) indicated either as superscript(s), if the angle is restricted to positive values, or subscripts for negative values.

ExploreTurns includes an interpreter for the CT notation which supports the exploration of H-bonded loop motifs and beta hairpins and the detection of new recurrent motifs. The

interpreter extends the notation to include entries which specify sequence motifs or secondary structures within the loop, or support the selection of loops that occur in particular position frames with respect to the beta turn at the center of the ExploreTurns window. CT labels are entered in ExploreTurns after translation using the Beta Code[59] romanization of the Greek script. The CT interpreter is described in section 5.5 (or click *Compound-Turn + BBL Notations*), and Table 1 presents examples of CT labels which fully demonstrate its features.

3.1.5 Classifying and comparing H-bonded loop motifs

Figure 3 compares the structures of 12 H-bonded loop motifs, including eight previously described loops and four new motifs detected with ExploreTurns using the CT interpreter. Previously reported structures include the six-residue Schellman loop[16.17], which very commonly forms a C-terminal helix cap with a pair of BB H-bonds, as well as three related structures: a common Schellman variant[28], referred to here as "Schellman+β", which adds a second overlapping H-bonded beta turn to the "classic" loop, a second variant, dubbed "Schellman+α"[28], in which an H-bonded alpha turn precedes and overlaps the loop, and the wide Schellman loop[18], which adds a residue inside the classic motif. Also included are the common five-residue alpha-beta loop[28] as well as the crown bridge loop[29] and two variants referred to here as the "crown nest"[29], which adds an H-bonded alpha turn, and the "crown pyramid"[60], which adds a reverse H-bonded beta turn to the crown nest, yielding a motif which binds Mg and substrate at the active sites of Mg phosphatases in all three kingdoms.

Motifs detected with ExploreTurns and displayed in Figure 3 include the "short Schellman loop", which removes a residue from the Schellman loop's internal beta turn (forming a gamma turn), the "pi-alpha" and "beta-gamma" loops, which add and subtract residues from the alpha-

beta loop respectively, and the "beta bracket", a seven-residue open loop in which two overlapping beta turns oriented at approximately 90° to each other are linked by a reverse alpha-turn. Figure 3 does not include beta-bulge loops[12], as they are treated separately below.

Table S1 lists the romanized CT labels used to select all of the motifs shown in Figure 3 in ExploreTurns, as well as labels for two additional seven-residue BB H-bond variants of the classic Schellman loop identified with the tool. Table S1 also lists the CT labels for the left- and right-handed conformational variants of the classic Schellman loop and the analogous variants of the short Schellman loop.

The loops in Figure 3 are labelled with their motif names and also classified by the CT notation, which facilitates their comparison by explicitly indicating the differences in their component-turn profiles. Simple transformations are suggested in the figure caption which relate the structures to each other by addition or subtraction of loop residues or component turn H-bonds. Sets of loop structures can be explored and profiled in the ExploreTurns loop motif browser, where schematics are also displayed, or by entering a motif's name into the *BB H-bonds/Loop motif/Compound turn* box on the tool page and clicking *Load Selected Structures* ("loop" is optional for all motif names, and Greek letters are translated to English words ("alpha", "beta").

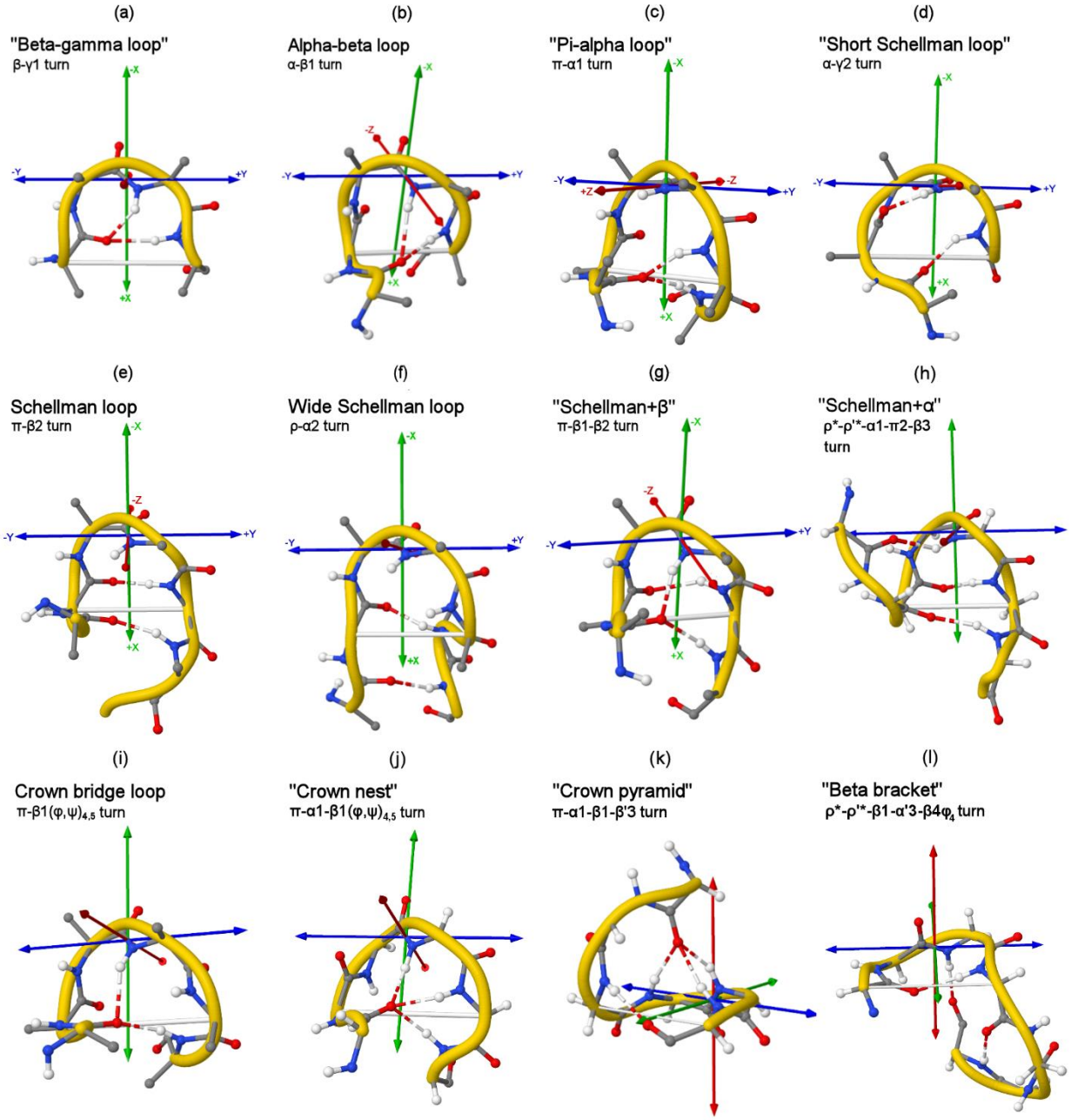

Figure 3. Classifying and comparing H-bonded loop motifs (excluding beta-bulge loops). Loops are labeled with motif names as well as compound-turn labels that classify the structures by the lengths (types) and start positions of their component H-bonded turns. Simple transformations relating the loops to each other are suggested. **(a)** The four-residue beta-gamma loop shows a high frequency of outliers in PDB validation checks[53], but some examples pass checks, including 3Q23_A_949, which binds

Mn/G2P at the active site of bacteriophage N4 RNA polymerase. **(b)** The common alpha-beta loop[28] adds a residue inside the beta-gamma loop's γ1-turn, yielding a β1-turn. **(c)** The pi-alpha loop adds a residue inside the alpha-beta loop's β1-turn, forming an α1-turn. **(d)** Bond rotations that transition the alpha-beta loop's β1-turn to a γ2-turn yield the short Schellman loop. **(e)** The classic Schellman loop[16,17] adds a residue inside the short Schellman's γ2-turn, forming a β2-turn. **(f)** The wide Schellman loop[18] adds a residue inside the Schellman's β2-turn, forming an α2-turn. **(g)** A rotation of the Schellman loop's fourth BB NH group adds a β1-turn, yielding the Schellman+β loop[28]. **(h)** An α-turn precedes and overlaps the Schellman loop, forming the seven-residue, open Schellman+α loop[28]. **(i)** Bond rotations disrupt the Schellman+β loop's β2-turn and redirect its final residue, yielding the crown bridge loop[29]. **(j)** A rotation of the crown bridge loop's fifth BB NH group adds an α1-turn, yielding the crown nest[29]. **(k)** Major conformational changes in residues (1-3, 6) of the crown nest form a β'3-turn, yielding the crown pyramid[60], which binds Mg and substrate at the active sites of Mg phosphatases in all three kingdoms. **(l)** Two overlapping β-turns oriented at approximately 90° to each other are linked by an α'-turn, forming the beta bracket.

3.1.6 Generalizing the beta-bulge loop

The "classical" definition of the type 1 beta-bulge loop (BBL1)[12] describes a five residue motif which is commonly found at the loop ends of beta hairpins, where it can be formed by the insertion of a residue that splits the pair of reciprocal BB H-bonds at the loop end of the C-terminal strand of a 2:2 hairpin, forming a new chain-reversing loop containing an H-bonded beta turn and a "bulge" at the loop's C-terminus constrained at its ends by the split H-bond pair (to view a 2:2 hairpin, enter *2:2* into the *BB H-bonds/Loop motif/Compound turn* box in ExploreTurns and click *Load Selected Structures*; to view a BBL1 in a hairpin, enter "type 1 BBL"

and click *Browse Structures* after loading). The classical BBL definition also describes a second motif, which can be formed by adding a residue inside the BBL1's beta turn, forming an alpha turn (enter and load "type 2 BBL").

Although the names of the BBLs refer to the beta bulges that occur in beta sheets/hairpins, their definitions do not depend on the presence of a beta ladder, since the loops can also be found as independent structures. 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 links the motif's terminal residues, closing the loop and defining the structure's extent, while the other member links to an interior loop residue, simultaneously forming an interior H-bonded turn and a bulge constrained at its ends by the split H-bond pair. In both the type 1 and 2 BBLs, the bulge contains only one added 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 a bulge, without limiting the bulge's length to one residue or requiring that it occur at the loop's C-terminus.

A search for generalized BBLs using the CT interpreter and other features of ExploreTurns yielded a large family of motifs with lengths up to eight residues and bulges which add up to four residues and occur at either the N- or C-terminus of the loop. These BBLs are less common than the classical pair, but like those motifs they occur either at the loop ends of beta hairpins or independently and can be found at ligand-binding and active sites (see Figure 8). The longer BBL types exhibit greater BB freedom in their bulges than that seen in the classic pair, but

recurrent approximate geometries can be identified, and the motifs show variants constrained by additional BB H-bonds or shaped by internal SC/BB and SC/SC interactions.

Since the CT nomenclature encompasses all short H-bonded loops, it applies to BBLs and classifies them by their H-bonded component turns. However, since the structures of BBLs have a characteristic general form, it is also possible to classify them in a simple, motif-specific notation: loop types are labelled in a three-component format which specifies the motif's overall length, the location of the bulge at the N- or C-terminus, and the bulge length (the number of residues added to form the bulge). BBLs can be selected in ExploreTurns by entering either their CT labels or their BBL type names into the *BB H-bonds...* box (or they may be explored in the loop motif browser).

Figure 4 displays example structures for four BBLs with C-terminal bulges, including types {5C1, 6C2, 7C3, 6C1}. The 5C1 BBL corresponds to type 1 in the classical nomenclature, while the 6C2 and 7C3 BBLs can be formed by successive insertions of additional residues into the 5C1 BBL's bulge. As stated above, the 6C1 BBL (classical type 2) can be formed by an insertion of a residue into the 5C1 BBL's beta turn. Each BBL in Figure 4 can be further classified into conformational 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 $C_{\alpha 1}$, $C_{\alpha 4}$ and the midpoint of the turn's middle peptide bond[48]; see Section 5.2).

Table S2 supports the exploration of the variants of the BBLs displayed in Figure 4 by listing the compound-turn labels for both conformations of each loop; the variants can also be explored in the tool by entering a BBL's type followed by "+/-". Table S2 also lists labels for

"extended" forms of the 5C1, 6C2 and 6C1 BBLs which lie embedded within longer H-bonded turns.

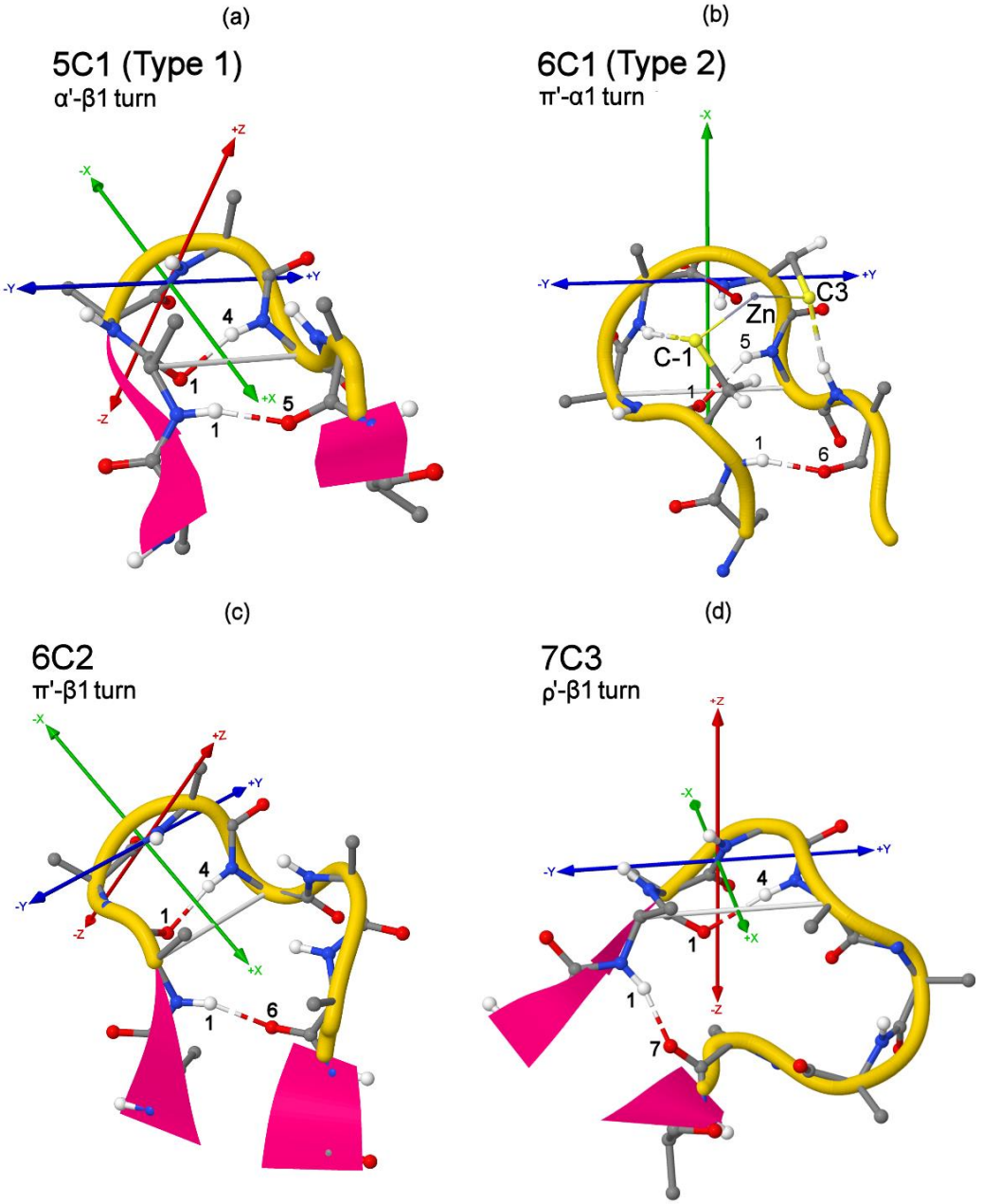

Figure 4. Examples of beta-bulge loops with C-terminal bulges. Structures are labelled using both a BBL-specific notation that specifies overall loop length, bulge location (N/C terminal) and bulge length, and

the compound-turn nomenclature for all H-bonded loop motifs, which classifies structures by the types and start positions of their component H-bonded turns (see Section 5.4). Each motif's defining H-bonds are labelled at their ends with residue positions in the loop. To view examples of a BBL, enter its type name into the *BB H-bonds/Loop motif/Compound turn* box and click *Load Selected Structures*. **(a)** 5C1 (type 1) BBL; common 5C1+ variant. **(b)** 6C1 (type 2) BBL; common 6C1+ variant, in its "rubredoxin knuckle"[61] metal-binding form found in Zn fingers/ribbons. This loop adds a residue inside 5C1's β-turn, forming an α-turn. **(c)** 6C2 BBL; 6C2+ variant, adds a residue to 5C1's bulge. **(d)** 7C3 BBL; 7C3- variant, adds a residue to 6C2's bulge.

Figure 5 presents schematic diagrams of the structures and H-bond topologies of a larger set of BBLs with C-terminal bulges, including four additional types with bulge lengths of up to four residues.

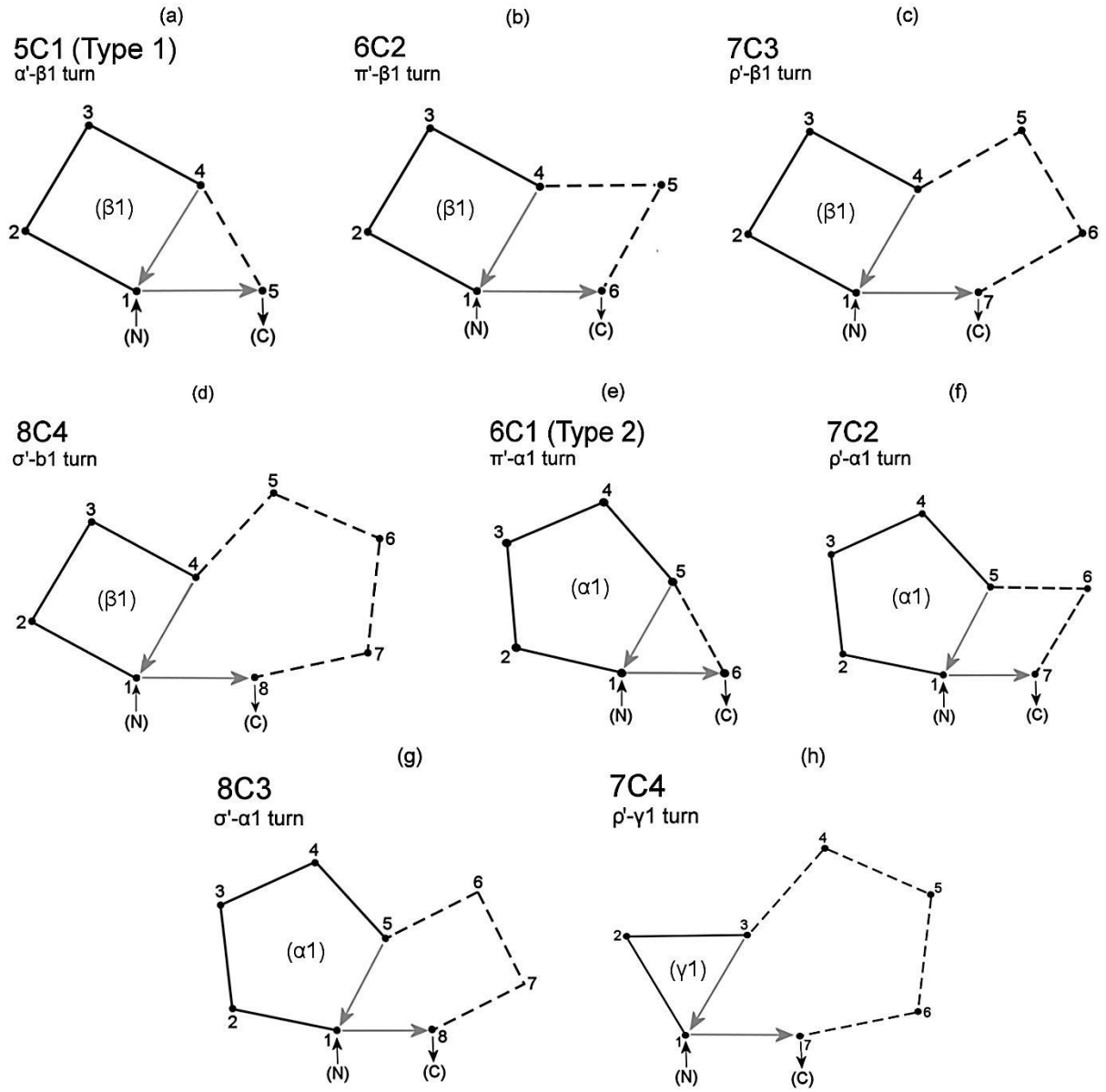

Figure 5. Schematics for eight generalized BBLs with C-terminal bulges. Internal H-bonded turns are drawn with solid lines and bulges are drawn with dashed lines. H-bonds are displayed as arrows oriented in the direction of donation. Loops are labelled using both BBL-specific and compound-turn nomenclatures. To browse a BBL's structures in ExploreTurns, enter its type names into the *BB H-bonds...* box and click *Load Selected Structures*. For the BBLs marked below with asterisks, variants can be explored using the compound-turn labels listed in Tables S2 and S4. BBLs not selected in "baseline"

form (with a capitalized framing turn in the romanized CT label) may show additional BB H-bonds. **(a)*** In its hairpin context, the 5C1 (type 1) BBL can be formed by the insertion of a residue at the loop end of the C-terminal strand of a 2:2 β-hairpin. **(b-d)*** Successive additional insertions into the 5C1's bulge yield the 6C2-8C4 BBLs. **(e)*** The type 2 (6C1) BBL can be formed by the insertion of a residue into the 5C1's β1-turn, yielding an α1-turn. **(f, g)** Successive insertions of residues into the 6C1's bulge yield the 7C2 and 8C3 BBLs. **(h)*** Conversion of the 7C3's β1-turn to a γ1-turn yields the 7C4 BBL.

Figure 6 displays example structures for four BBLs with bulges at their N-termini, including types {5N1, 6N2, 7N3, 7NC3}. In its hairpin context, the 5N1 BBL can be formed by the insertion of a residue at the loop end of the N-terminal strand of a 2:2 hairpin, while the 6N2 and 7N3 BBLs can result from successive additional insertions into the 5N1 BBL's bulge. Like the C-type BBLs, 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 both termini. Table S3 lists the CT labels used to select the BBLs in Figure 6.

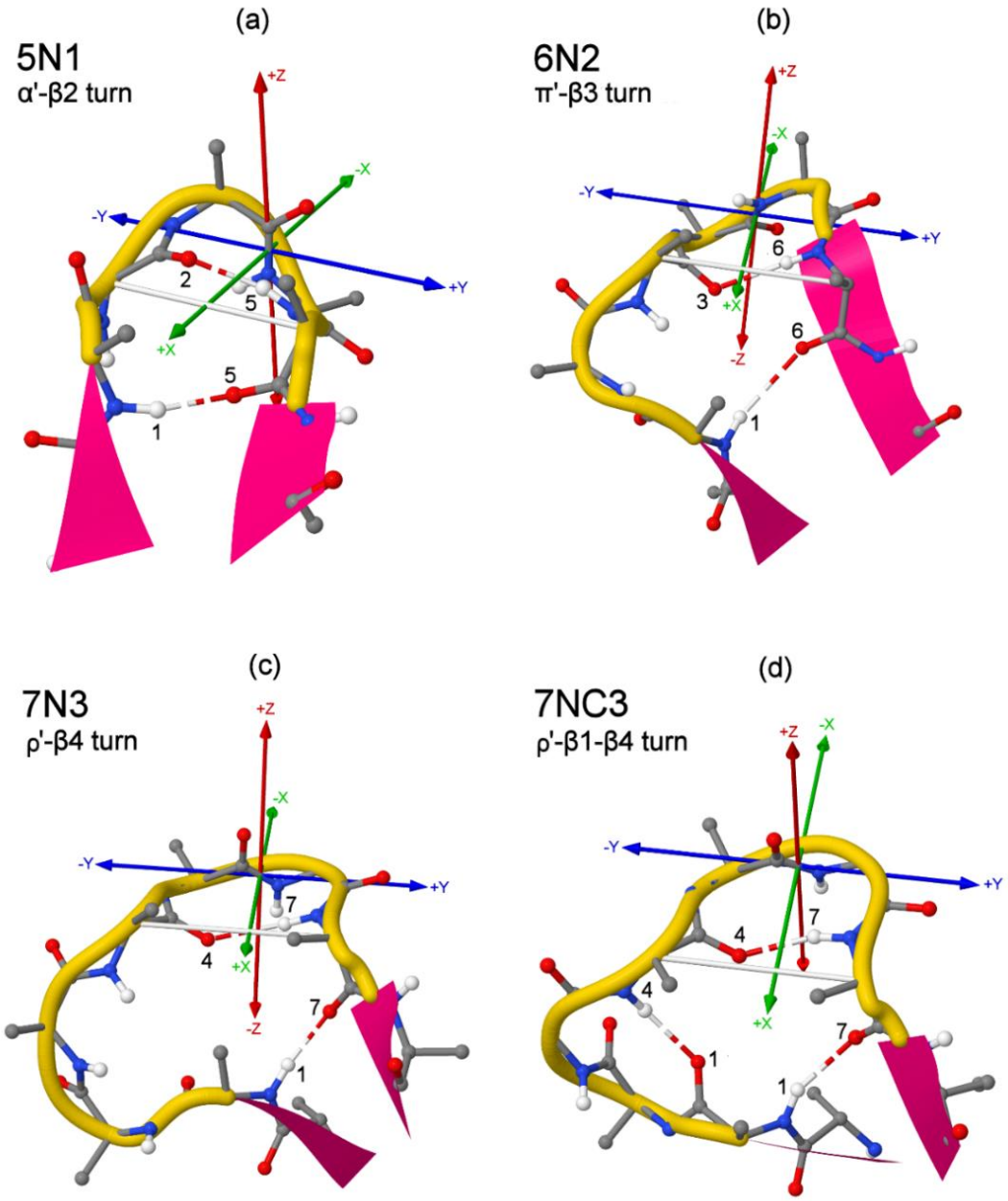

Figure 6. Examples of beta-bulge loops with N-or N/C-terminal bulges. See the Figure 4 caption for loop nomenclature and structure viewing directions. **(a)** 5N1 BBL. **(b)** 6N2 BBL; 6N2- variant, adds a residue to 5N1's bulge. **(c)** 7N3 BBL; adds a residue to 6N2's bulge. **(d)** 7NC3 BBL; adds a second β-turn H-bond to 7N3. Split pairs of BB H-bonds at both termini classify this BBL as both N- and C-type.

Figure 7 presents schematic diagrams of the structures and H-bond topologies of a larger set of BBLs with N- and N-/C-terminal bulges, including seven additional types with bulge lengths of up to four residues.

With the exception of the 7N3 BBL, the loops displayed in Figures 4 and 6 were selected with CT labels in the strict "baseline" CT format, specified by the capitalization of the loop's framing turn, which rules out structures with additional BB H-bonds within the loop or between loop residues and other residues in the ExploreTurns window (Section 5.5). If a BBL (or any CT) is instead loaded with a lowercase entry for its framing turn, then any variants with additional BB H-bonds are included in the selected set, and the corresponding additional H-bonded component turns are listed in the *H-bond profile* box. The effects of these additional H-bonds on loop structure can be explored by copying the corresponding component-turn labels from the *H-bond profile* box into the CT label in the *BB H-bonds...* box and loading the set of structures of the resulting BB H-bond variant (see Table 1 and *H-bonded loop motifs/Examples/Ex. 3* in the tool's feature summary).

ExploreTurns also supports the exploration of the SC motifs which can shape the BB conformations of CTs or play functional roles. When a CT label is loaded, the tool lists the single-AA and pair sequence motifs present in the loop at the top of the profile window, ranked by fractional overrepresentation (for single-AA motifs), or the chi-square measure of statistical significance (for pairs), and the SC structures in a loop and their effects on loop conformation can be explored by copying the corresponding sequence motif labels from the profile window into the CT label in the *BB H-bonds...* box and loading the set of structures (see Table 1 and *H-bonded loop motifs/Examples/Ex. 4* in the feature summary).

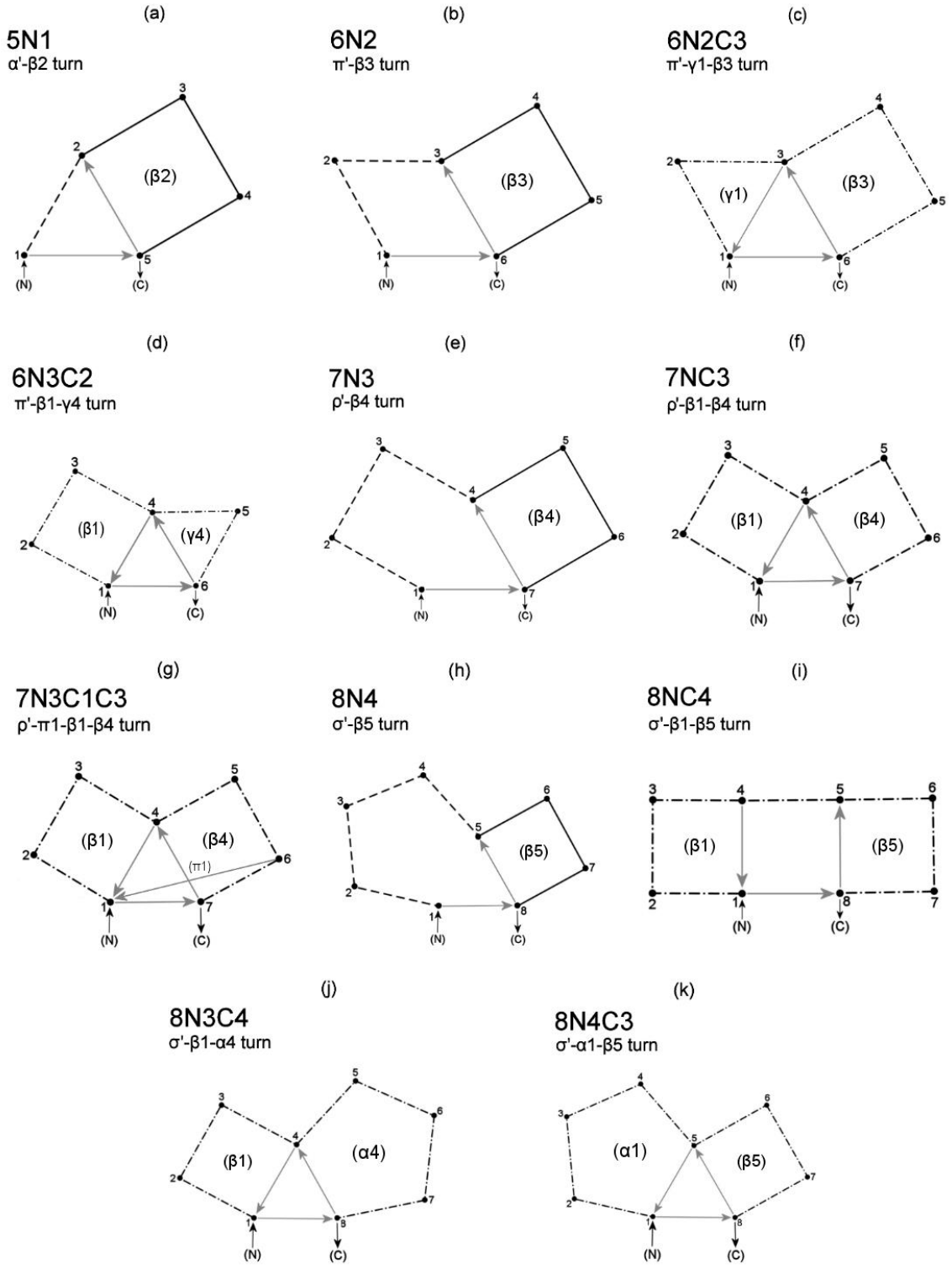

Figure 7. Schematics for eleven generalized BBLs with N- or N/C-terminal bulges. See the Figure 5 caption for display conventions. Dual N/C-type BBLs are drawn entirely with dash-dot lines. Additional

BB H-bonds can occur. BBLs marked with asterisks below have variants which can be explored using the compound-turn labels in Tables S3 and S4. **(a)** The 5N1 BBL can be formed by the insertion of a residue at the loop end of the N-terminal strand of the 2:2 β-hairpin. **(b)*** The 6N2 BBL adds a residue to 5N1's bulge. **(c)*** Addition of a γ1 H-bond to 6N2 yields 6N2C3, with split pairs of BB H-bonds at each terminus that classify it as both N-type, with a two-residue bulge, and C-type with a three-residue bulge. Examples of this motif bind RNA (Figure 8d) and separate DNA strands in the helicase reaction (Figure 8e). **(d)** Addition of a γ4 H-bond to 6C2 (Figure 5b) yields 6N3C2, classified as both N-type with a three-residue bulge and C-type with a two-residue bulge. **(e)** 7N3 adds a residue to 6N2's bulge (see DNA-binding example in Figure 8g). **(f)** Addition of a β1 H-bond to 7N3 yields 7NC3, with three-residue N- and C-terminal bulges. **(g)** Addition of a π1 H-bond to 7NC3 yields 7N3C1C3, with overlapping one- and three-residue C-terminal bulges; the π1 H-bond draws the β-turn "lobes" closer together. DSSP calls 3_{10} helices within the loop in multiple examples; helices can be excluded by loading the motif with the CT label *r'-p1-b1-b4-\h2*. **(h)*** 8N4 adds a residue to 7N3's bulge; 3_{10} helices can be excluded by loading the motif with the CT label *s'-b*1-b5-\h3*. (**i**) Addition of a β1 H-bond to 8N4 yields 8NC4 (see DNA-binding example in Figure 8i). **(j)** Conversion of 8NC4's β5-turn to α4 yields 8N3C4. **(k)** Conversion of 8NC4's β1-turn to α1 yields 8N4C3.

3.1.7 Roles of beta-bulge loops

BBLs are hybrid structures that combine the constraints imposed by their defining BB H-bonds with a degree of conformational range configurable by additional BB H-bonds or SC interactions, giving them versatility for structural or functional roles. Figure 8 displays nine examples of generalized BBLs at binding and active sites, with ligands that include metal ions, an FeS cluster, RNA and DNA (x3).

BBLs often form the loop ends of beta-hairpins, and they can be projected by hairpin "arms"

into interaction sites, as in Figure 8e, where 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 in the helicase reaction in human Werner syndrome protein (WRN)[62] (3AAF_A_1033). In another example of BBL projection (Figure 8g), a 7N3 BBL binds DNA with BB H-bonds, a Lys SC, and an Asp/Arg salt bridge in RovA, a transcriptional activator of Yersinia (plague) invasin (4AIK_A_81).

BBLs can bind metal ions for the purposes of catalysis, as in Figure 8c, where 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). But Figure 8 also shows examples in which metals might be bound in order to supplement the loop's internal H-bonds in the shaping or stabilization of its structure, as in Figure 8a, where a 5N1 BBL and its hairpin coordinate potassium with 5 covalent bonds in the fatty-acid binding protein AtFAP1 (4DOO_A_34), as well as Figure 8f, where A 7C3 BBL coordinates Na with three bonds in methionine R-sulfoxide reductase B1 (the loop is not at the catalytic site) (3MAO_A_60). In Figure 8h, an 8N4 BBL is shaped by its coordination of Mg with five covalent bonds as it joins a helix in the "capping" of a beta solenoid in a possible adhesin domain (4WA0_A_473).

Figure 8b shows another type of interaction between BBLs and metals: a 6C2 BBL tethers an active-site 4Fe-4S cluster with a Cys SC, and it may also "cradle" the cluster with hydrophobic SCs, which could help shield the cube from degradation[63] (1OAO_C_526).

The utility of a BBL's combination of structural constraint and flexibility is shown to good effect in Figure 8i, in which an 8NC4 BBL and its hairpin bind DNA with BB and SC H-bonds in the de novo DNA methyl transferase DNMT3B. This σ'-β1-β5 turn inserts its β5-turn into the major

groove for interaction with the DNA's bases, while its β1-turn orients at about 90° to the β5-turn, positioning its Lys SC for H-bonding with the DNA backbone from outside the double helix (5CIY_A_232). The versatility of BBLs is also demonstrated in Figure 8d, in which a 6N2C3 BBL (π'-γ1-β3 turn), binding RNA in the HutP antitermination complex, partially "wraps" a guanine base, stacking its β3-turn above the base in a likely pi-peptide bond interaction, while its γ1-turn interacts with the guanine by forming a BB H-bond with the base in its plane (1WPU_A_131). These two examples suggest that the H-bonded simple turns in BBLs can impart modularity to the structures, serving as stabilized components that cooperate to accomplish multipart binding or interaction tasks.

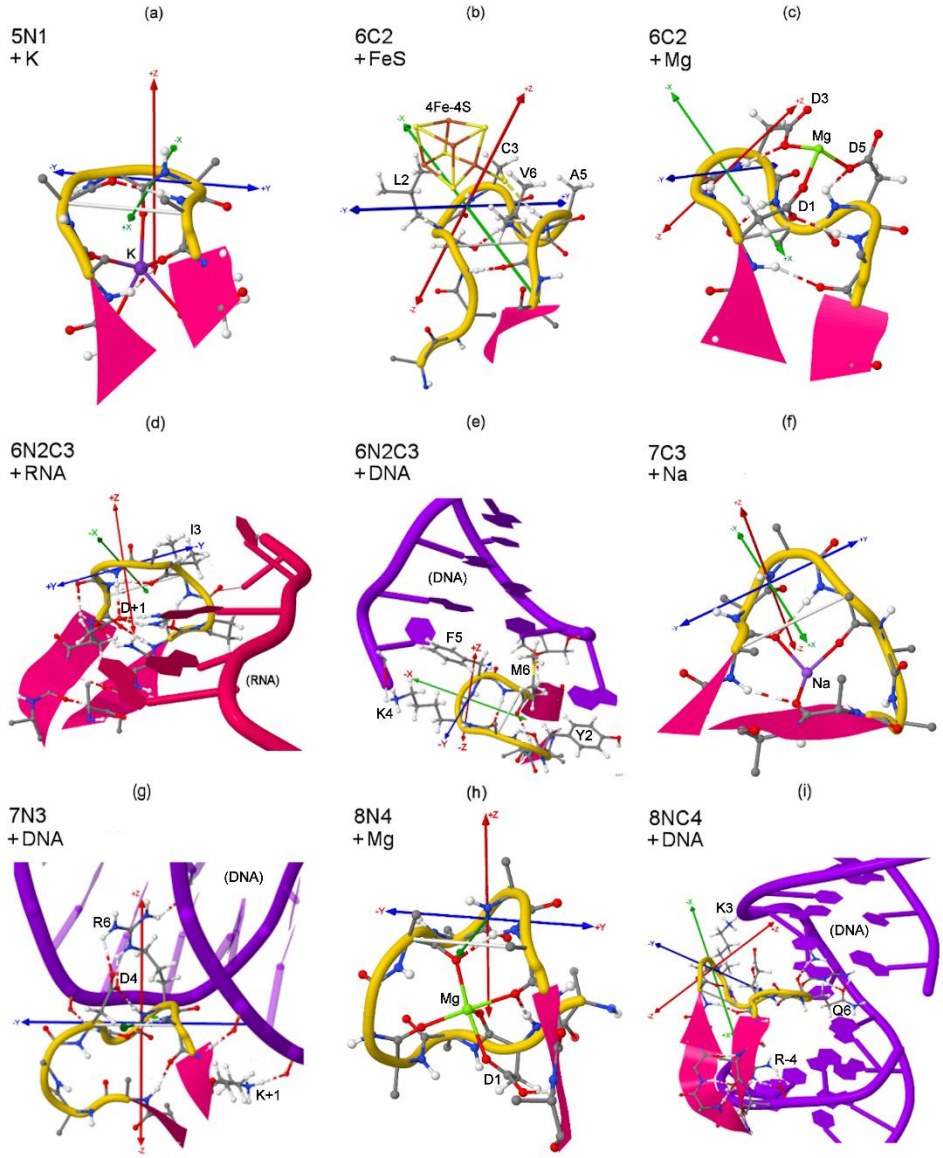

Figure 8. Examples of generalized BBLs at ligand-binding and active sites. Selected SCs are shown, labelled with residue positions in the loop frames. View structures by loading their CT labels in ExploreTurns (the SCs specified in the labels serve to uniquely identify the loops in the database, and they may also highlight interactions of interest). **(a)** A 5N1 BBL binds K in the fatty-acid binding protein AtFAP1 (4DOO_A_34, CT label *A'-b2-L1-L2-G3*). **(b)** A 6C2 BBL tethers an active-site 4Fe-4S cluster with Cys3 (C3) in acetyl-CoA synthase/carbon monoxide dehydrogenase. The loop also may help shield the cube from oxidation using its hydrophobic SCs[63] (1OAO_C_526, CT label *P'-b1-G1-L2-C3*). **(c)** A 6C2 BBL

binds Mg with an Asp triplet (D1, D3, D5) at the active site of phosphoglucomutase from Xanthomonas citri (5BMN_A_259, CT label *P'-b1-D1-D3-F4*-D5). **(d)** A 6N2C3 BBL partially "wraps" an RNA base in the HutP antitermination complex, stacking its β3-turn above the base in a likely pi-peptide bond interaction while its γ1-turn forms a BB H-bond with the guanine. Asp (D+1) also H-bonds with the base, and Ile (I3) packs with an adjacent base (1WPU_A_131, CT label *p'-g1-b3-A1-P2*). **(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 in the helicase reaction in human Werner syndrome protein (WRN)[62] (3AAF_A_1033, CT label *p'-g1-b3-R1*). **(f)** A 7C3 BBL binds Na in methionine R-sulfoxide reductase B1 (3MAO_A_60, CT label *R'-b1-E1-H2*). **(g)** A 7N3 BBL and its hairpin binds DNA with BB H-bonds, a Lys SC (K+1), and an Asp/Arg salt bridge (D4R6) in RovA, a transcriptional activator of Yersinia (plague) invasin (4AIK_A_81, CT label *r'-b*1-b4-D4-R5-R6*). **(h)** An 8N4 BBL binds Mg with four BB carbonyls and an Asp SC (D1) (4WA0_A_473, CT label *s'-b*1-b5-D1-S5*). **(i)** An 8NC4 BBL and its hairpin bind DNA with BB H-bonds and three SCs (R-4, K3, Q6) in the de novo DNA methyl transferase DNMT3B (5CIY_A_232, CT label *s'-b1-b5-K3-Q6-E8*).

3.1.8 Classifying beta hairpins

A beta hairpin can be classified by the length of its chain reversing loop and the H-bond configuration at the loop end of its beta ladder using the nomenclature X:Y, where X is the loop length according the IUPAC definition of a strand residue and Y is its length according to the DSSP definition[8]. This notation conveys loop length directly and ladder H-bond configuration indirectly, using the contrast between the two conventions. Since the CT nomenclature also describes the length of a loop and the H-bond configuration at its ends (or anywhere in the loop), the notation can also classify hairpins, if it is applied to the hairpin's chain-reversing turn

(the chain-reversing loop inclusive of the pair of H-bonded residues at the loop end of the hairpin's beta ladder). Since all hairpins classified in the X:Y nomenclature exhibit either a single reverse H-bond or a pair of reciprocal normal and reverse H-bonds at the loop ends of their ladders, all X:Y hairpins with loop lengths up to 10 residues (corresponding to a chain-reversing turn length of 12 residues, the longest covered by the CT notation) are classifiable with CT labels composed of either a single entry for a reverse H-bonded turn, or a pair of entries for normal and reverse turns of the same type, so that the X:Y hairpin types {1:1, 1:3, 2:2, 2:4, 3:3, 3:5,...} become {γ-γ', γ', β-β', β', α-α', α',...} in the CT hairpin notation.

The CT hairpin notation has the disadvantage, compared to the X:Y system, that it requires knowledge of the lengths of the simple turn types. But it has the advantage that it directly specifies H-bonds rather than relying on the contrast between two conflicting conventions (IUPAC and DSSP), and more importantly, the CT notation is more complete, since it can describe the H-bond configuration in the hairpin's chain-reversing loop as well as its ladder. For example, the common structure described in the X:Y notation as a "3:5 hairpin with a type 1 beta-bulge loop" is simply an "α'-β1 hairpin" in the CT hairpin notation, while a "4:4 hairpin with a Schellman loop" is a "π-π'-β2 hairpin".

The ExploreTurns loop motif browser uses preset CT labels to select sets of examples of 16 types of beta hairpins from the four classes[6] in "baseline" configuration, with no BB H-bonds in their chain-reversing loops, and the browser includes hairpin schematics and classifies the structures using both the X:Y and CT hairpin notations. Each hairpin schematic is also annotated with the CT label used to select it, which must encompass residues N- and C-terminal to the chain-reversing turn as well as the residues within the turn, in order to specify the required BB

H-bonds in the hairpin's beta ladder. A shorthand CT hairpin notation that automatically embeds a CT label entered by the user into a longer label with the required additional beta-ladder H-bonds (e. g. "a'-b1 hairpin" entered by the user becomes the CT label "t-t'-a'3-b3", which is then used to select structures) is not yet live in the ExploreTurns CT interpreter; this is a priority for a future ExploreTurns release with a structural window longer than 12 residues, which can better support the selection of hairpins with longer chain-reversing loops.

3.2 Mapping Asx N-cap sequence preference vs. cap geometry

Figures S1 and S2 in Supplementary Section 3 apply the motif selection, DSSP secondary structure selection, and "context vector" features of ExploreTurns to heat-map the sequence preferences of Asp/Asn (Asx) N-cap motifs in beta turns at the alpha-helical N-terminus vs. helix/turn geometry. The maps, which show contrasting patterns for Asp and Asn, reveal the cap geometries most compatible with these motifs, and they can be used to guide ExploreTurns investigations of the structure and H-bonding associated with Asx N-caps formed by beta turns.

3.3 Investigating the depth dependence of turn geometry

Supplementary Section 4 applies ExploreTurns to investigate the long-standing question of the extent to which beta-turn BB geometry is affected by context, by profiling the changes that occur in the distribution of beta-turn geometries as turns become more buried in proteins. The proportion of structurally unclassified (type IV) turns is found to increase by more than 50% with depth at the expense of classical type I, reaching almost half of all turns (Figure S3). This

finding may reflect distortions due to increasing or more impactful interactions with structures in the protein external to turns, or increasing turn involvement in local structures such as ligand-binding/active sites, and it supports the hypothesis that many type IV turns represent distortions of the structurally classified types rather than intrinsically stable turn conformations.

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 C-termini of alpha helices. Four principal conformations are identified, and Figure S4 presents example structures and ExploreTurns selection criteria for each.

3.5 Exploring sequence motifs and supersecondary structures

Supplementary Section 6 provides selection criteria (Table S5) and sample images (Figure S5) for examples of sequence motifs in beta turns; some motifs are tuned by BB geometry to maximize overrepresentation (and likely compatibility with SC interactions).

Supplementary Section 7 demonstrates the application of context vectors to select supersecondary structures in which beta turns link helices and strands into particular orientations. Table S6 gives selection criteria for these structures, and Figure S6 presents sample images.

3.6 Geometric descriptor ranges and distribution plots

Supplementary Section 8 provides selection criteria (Table S7) that demonstrate the range of geometric descriptor values which occur in beta turns, by specifying examples of turns with descriptor values near their limits in the database (Figure 9 juxtaposes examples of these structures).

Supplementary Section 9 gives directions for the display of notable examples of the distribution plots involving geometric descriptors which are available in ExploreTurns (Table S8), and Figure S7 displays six plots.

4 | Discussion

ExploreTurns supports the study of beta turns and most loop structure with multiple new features, including a turn-local coordinate system which provides a useful framework for the analysis of turns and their many contexts, a global turn alignment that supports structure comparisons, a set of geometric descriptors for beta turns that complement the Ramachandran-space classification systems, and new nomenclatures that classify H-bonded loop motifs and beta hairpins as well as BBLs.

The utility of ExploreTurns is demonstrated here in areas including the exploration of H-bonded loop motifs, the detection of new motifs, and the analysis of turn and loop structure. The new loops characterized here represent significant new contexts for beta turns, and though many of them are relatively rare in the redundancy-screened dataset, they can nevertheless play key roles, as demonstrated in Figure 8.

The large family of BBLs, which demonstrate a range of geometries and H-bond topologies, could provide models for loop design. BBLs can play structural, regulatory, binding or catalytic roles (Figures 4b, 8a-c, 8f, 8h), and since they commonly constitute the chain-reversing loops of beta hairpins, they may prove particularly useful in the design of these motifs, which are a common design target because they are key components of a wide range of bioactive molecules, from antimicrobial peptides to beta-barrel transporter proteins[7] (see e. g. the 7C3 BBL, with CT label *R'-b1-E1*-S3, which forms the L3 hairpin loop in the alginate transporter AlgE). BBLs may find application as configurable "tools" at the ends of beta-ladder "arms" that project them into interaction sites, as seen with the 6N2C3 and 7N3 BBLs that interact with DNA in Figures 8e and 8g.

CTs are already employed in design; for example, the beta-turn motif used in a recent design of "buttressed" loops for function[64] was a β-β' turn extracted from a structural database, and as design efforts increasingly encompass detailed, nonrepetitive BB structures tailored to particular contexts and functions, the availability of a library of versatile loop motifs may prove useful. Design efforts involving CTs could focus not only on employing existing motifs, but also on designing new loops; motifs of up to eight residues constrained by up to four BB H-bonds are described here, but design work could probe the limits of the size and complexity of local H-bonded BB structures.

A reservoir of new examples of compound turns which could provide much more complete coverage of the BB and SC variants of the CTs detected here (and yield examples of new recurrent motifs) could already be available, in the form of the large sets of protein structures predicted from sequence by tools such as AlphaFold 2[65]. However, while AlphaFold 2 has shown

good accuracy in the prediction of short loops[66], the rarity and/or complexity of some CTs (e.g. the 7N3C1C3 BBL in Figure 7g or the "crown pyramid"[60] in Figure 3k) may present challenges to prediction algorithms. If all algorithms cannot accurately predict all loops, a set of loops could be used in benchmark testing to rank the tools by their ability to predict local, irregular H-bonded BB structures. The conformations of any CTs that cannot be accurately predicted could be further explored with computational modelling like that applied by Venkatachalam to beta turns[2], and the structures and energetics of all the newly-identified loops could be investigated experimentally using combinatorial mutagenesis and other methods.

Other potential applications of the CT nomenclature and the motif library presented here include structural annotation, protein classification, and the study of protein evolution. The CT notation could be applied to profile the local H-bonded loop structures in all proteins, and this information could be incorporated into tools such as the PDB's secondary structure annotator. CT profiles could be computed for protein families classified by structure or function, and differences in these profiles could be informative regarding the relationships between local H-bonded structure and fold or function, and could also shed light on the evolution of loop motifs.

5 | Methods

5.1 The ExploreTurns database

Beta turns are defined here as four-residue BB segments with a distance of no more than 7 Å 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[4]. Turn/tail structures were

extracted from their PDB files[49] with the aid of a list of peptide chains with a maximum mutual identity of 25% obtained from PISCES[67]. Since sequence motif detection identified multiple motif "artifacts" generated by remaining local redundancy in the data, an additional step of turn-local redundancy screening was applied: structures were clustered and filtered to reduce redundancy within five residues of turns to 40% or less; this threshold was chosen because it effectively reduced artifacts while preserving a dataset of sufficient size.

Turn structures were 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. The final dataset of 102,192 turns was compiled with resolution and R-value cutoffs of 2.0 Å and .25 respectively; average values in the dataset are 1.6 Å and .20.

5.2 Turn-local coordinate system and alignment

A common local Euclidean-space coordinate system for beta turns was established via a set of geometric definitions[48] (see Figure 9). The **span line** is the line between a turn's first and last alpha carbons ($C_{\alpha 1}$and $C_{\alpha 4}$), and the span, already used in the beta-turn definition, is its length. The **turn center** is the midpoint of ($C_2 \rightarrow N_3$), the middle peptide bond in the turn. The **turn plane** contains the span line and passes through the turn center, and the perpendicular dropped from the turn center to the span line defines the **turn axis**. Using these definitions, a local, orthogonal right-handed coordinate system is established. The system's origin lies at the turn center, while the x axis lies in the turn plane, collinear with the turn axis, with its positive sense oriented from the center towards the span line. The y axis lies in the turn plane, parallel

to the span, with its positive sense oriented towards the C-terminal half of the turn. Finally, the z axis is perpendicular to the turn plane, with its positive sense determined by the cross-product of the x and y axes.

The atomic coordinates of each turn are transformed from the global protein system in the turn's PDB file to the turn-local system, implicitly aligning the structures. The alignment's accuracy was tested by a comparison between implicit pairwise turn alignments and the best alignments generated by an RMSD-based perturbative search procedure. In this process, the average distance between the corresponding BB atoms of randomly chosen, implicitly aligned turn pairs were compared to the distance obtained for the best alignments generated by a perturbative search applied to the pairs which seeks the minimum average interatomic distances by introducing small variations in the relative position and orientation of the two turns. This procedure was applied to a random sample of 100 million turn pairs to generate an estimate of the average positional error of the implicit alignments. Since this estimate (0.17 Å) is comparable to the errors expected for the positions of well-determined atoms in well-refined structures (0.1 Å - 0.2 Å; overall RMSDs for independently-determined 2 Å structures have been measured as 0.5 Å - 0.8 Å)[68], the accuracy of the implicitly established alignment was judged to be satisfactory.

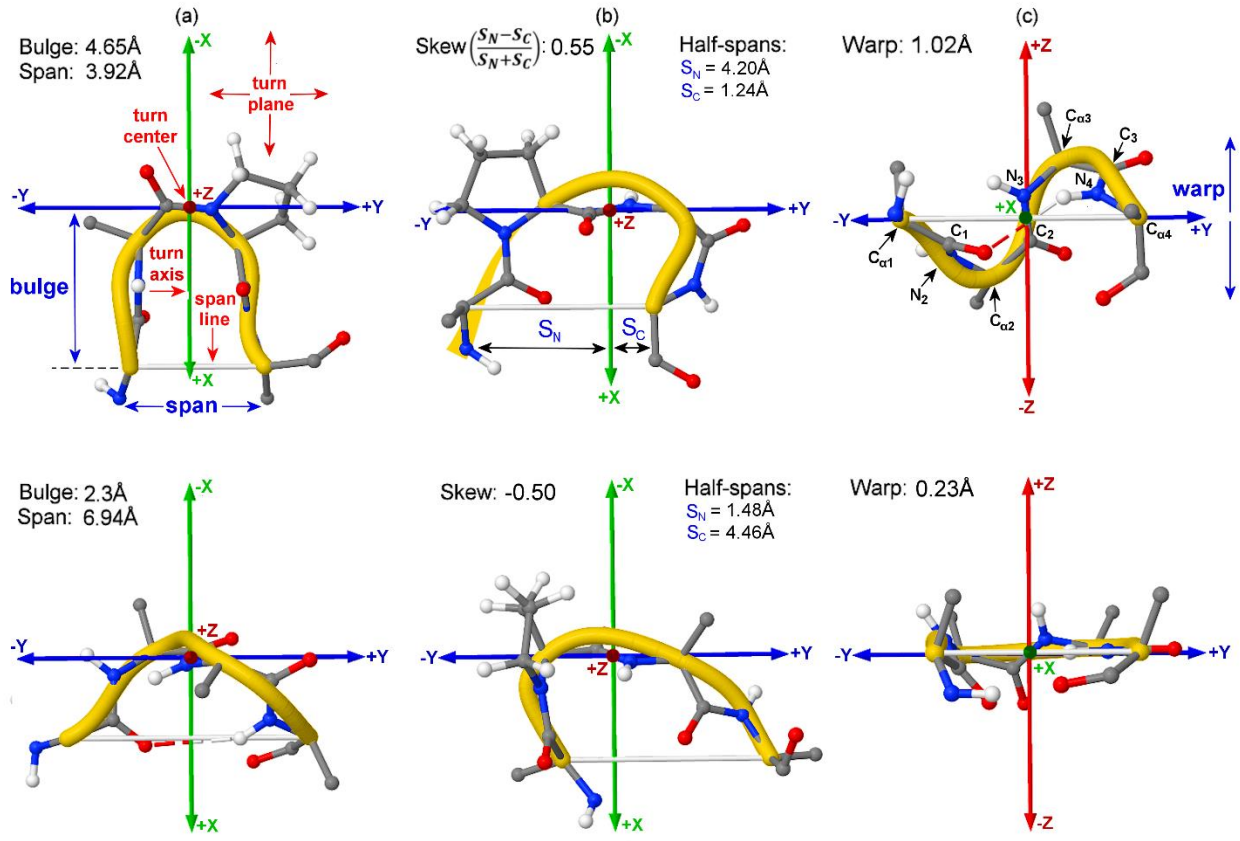

Figure 9. Ranges of the geometric descriptors in structures, displayed in the turn-local coordinate system, with geometric definitions labelled in red and descriptors labelled in blue. This figure illustrates the geometric definitions used to establish the turn-local coordinate system, the coordinate system itself, and the geometric descriptors for beta turns. The figure also demonstrates the modes of variation measured by the descriptors, by displaying turn structures with descriptor values near their limits in the dataset. A white bar representing the span line is added to all structures **(a)** 5HKQ_I_74 (top) and 4YZZ_A_239 (bottom) represent the ranges of both bulge and span, since the two descriptors exhibit strong negative correlation. **(b)** 5TEE_A_23 (top), and 3NSW_A_44 (bottom) represent strong positive (C-terminal) and negative (N-terminal) skew. N-and C-terminal half spans are also given. **(c)** 5BV8_A_1285 (top) and 3S9D_B_117 (bottom) represent the limits of warp in classical type I, and demonstrate the wide warp range that occurs within this type. The BB atoms in the main-chain path of

the turn between $C_{\alpha 1}$ and $C_{\alpha 4}$ are labelled; all labelled atoms except $C_{\alpha 1}$ and $C_{\alpha 4}$ are used to compute warp, as the average of the absolute values of the distances between these atoms and the turn plane.

5.3 Geometric descriptors

Geometric descriptors for beta turns[48] are defined here and illustrated in Figure 9.

5.3.1 Span and half-spans

The pre-existing span descriptor measures the distance [Å] between $C_{\alpha 1}$ and $C_{\alpha 4}$ in the turn, and a 7 Å threshold is part of the beta-turn definition. Two new span-related descriptors are defined: the N- and C-terminal half-spans measure the distances between the turn axis and $C_{\alpha 1}$ and $C_{\alpha 4}$ respectively, and quantify the separate "widths" of the N- and C-terminal halves of the turn, providing an additional degree of freedom in the measurement of span. Span values in the dataset range from a minimum of 3.13 Å to the 7 Å limit set by the beta-turn definition, while N-terminal half-spans range from 0 to 4.76 Å, and C-terminal half-spans from 0.76 Å to 4.82 Å.

5.3.2 Bulge

Bulge measures the magnitude of the excursion that the BB makes from the turn span as it bends to form the turn. Bulge is defined as the distance [Å] from the turn center to the span line. Bulge values in the dataset range from 1.28 Å to 4.7 Å.

5.3.3 Aspect

Aspect, which approximates the aspect ratio of the projection of the turn's BB onto the turn plane, is computed as the ratio of bulge to span.

5.3.4 Skew

Skew is a dimensionless measure of the asymmetry of the projection of a beta turn's BB onto the turn plane. Skew is defined as the directed displacement along the span line between the line's center and the point of intersection between the turn axis and the line, divided by half the span. Skew is negative when the axis intersects the span line in its N-terminal half, indicating a "lean" in the N-terminal direction, or positive when the axis intersects the line in its C-terminal half, producing a C-terminal lean. Skew is expressed in terms of the N- and C-terminal half-spans (S_N and S_C) as:

$$skew = \frac{S_N - S_C}{S_N + S_C} \tag{1}$$

Skew is positive when S_N is larger than S_C, negative when the reverse is true, or zero when the half-spans are equal. A turn that skews so strongly N-terminally that its axis intersects the span line at the position of $C_{\alpha 1}$ (S_N = 0) would exhibit a skew of -1, while a turn skewed C-terminally so that its axis intersects the line at $C_{\alpha 4}$ (S_C = 0) would show a skew of +1. Skew values in the dataset range from -1 to 0.78.

5.3.5 Warp

Warp measures the departure from flatness of a beta turn's BB in the z dimension of the turn coordinate system, perpendicular to the turn plane. Warp [Å] is defined as the average excursion of the turn's BB above or below the turn plane between $C_{\alpha 1}$ and $C_{\alpha 4}$, excluding BB carbonyl oxygen atoms as well as $C_{\alpha 1}$ and $C_{\alpha 4}$ themselves (since they lie within the plane). Warp is computed as the average absolute value of the z-components of the BB atoms (C_1, N_2, $\mathrm{C}_{\alpha 2}$, C_2, N_3, $\mathrm{C}_{\alpha 3}$, C_3, N_4). Warp values in the dataset range from 0.14 Å to 1.25 Å.

5.4 The compound-turn nomenclature for loop classification

The compound-turn (CT) nomenclature treats all short H-bonded loop motifs as overlapping H-bonded turns, and classifies them by the types (lengths) of these component turns and the start positions of the turns in the loops. H-bonded Loop motifs are classified by CT labels composed of entries separated by dashes. The first entry in a label, which indicates the length of the "framing" turn that embodies the entire loop, has the form:

$$(\delta|\gamma|\beta|\alpha|\pi|\rho|\sigma|\iota|\upsilon|\chi|\omega)['][*]$$

where the parentheses are used here for grouping and the brackets indicate optional components; neither punctuation appears in the notation. The Greek letters, separated by 'OR' bars here, indicate the component turn types, which have lengths of 2-12 residues. The single-quote character is used when a component turn's defining H-bond is donated in the N→C direction, indicating a "reverse" turn, rather than the default C→N direction seen in "normal" turns, while '*' is used when the turn's defining H-bond is to be excluded from the loop rather than included. If a CT's label contains entries which exclude both normal and reverse H-bonds in its framing turn (e.g. π*-π'*), then it describes an "open" loop, with termini not linked by a shared H-bond.

Subsequent dash-separated entries in the CT label specify the loop's internal H-bonded turns (which may share a terminus with the framing turn), using the form:

$$(\delta|\gamma|\beta|\alpha|\pi|\rho|\sigma|\iota|\upsilon|\chi|\omega)['][*]M$$

which adds M, an integer that specifies a component turn's start position in the loop.

Constraints that require positive or negative values for the (ϕ, ψ) dihedral angles of a CT's residues may be appended to its dash-separated list of component-turn specifiers, using the forms:

$(\varphi|\psi)_{N,P,\ldots}$ or $(\varphi|\psi)^{N,P,\ldots}$

where the subscripts constrain the dihedral angles of loop residues (N,P,...) to negative values, while the superscripts constrain the angles to positive values. Dihedral constraints are concatenated to the end of the list of component-turn entries in a label without dashes or spaces.

To avoid confusion with the names of supersecondary structure motifs (e.g. βαβ), a compound-turn label is always followed by the word "turn".

5.5 The ExploreTurns compound-turn interpreter

ExploreTurns includes a CT interpreter which adds features to the notation that facilitate loop exploration. A CT label is entered into the *BB H-bonds/Loop motif/Compound turn* box as a list of dash-separated entries, after translation from the Greek alphabet using the Beta Code[59] romanization (see examples in Table 1). The entries for a loop's component H-bonded turns take the form:

(d|g|b|a|p|r|s|t|u|x|w)['][*][M]

where the brackets, which enclose optional elements, and the parentheses and 'OR' bars, which present the choice of turn types, are not entered. The single-quote character specifies a 'reverse' H-bonded turn, while the asterisk excludes the specified turn. The integer M, which specifies the start position of a component turn in the loop, is required for internal component turns, which must be specified in lowercase, but is not entered for the framing turn, which may be capitalized when a "baseline" loop is specified (see below).

The CT interpreter supports the selection of loops that contain particular sequence motifs, which are defined as the requirement for one or more specific AAs at one or more loop positions. An AA is specified at loop position M with an entry of the form:

(A|C|D|E|F|G|H|I|K|L|M|N|P|Q|R|S|T|V|W|Y)M

Any (ϕ, ψ) BB dihedral-angle constraints in a CT label are translated, using the romanizations (f, y), into dash-separated entries of the form:

\(F|f|Y|y)M

where an uppercase letter requires a positive dihedral angle while a lowercase letter requires a negative angle at loop position M. The backslash is required as an "escape" character which signals the tool not to interpret a letter as an AA.

Optional, escaped dash-separated entries of the form:

\(E|e|H|h|O|o)M

apply secondary-structure conditions to loop residues, which either require (uppercase) or exclude (lowercase) residues of strand type (E/e), helix type (H/h, for helices of any type), or any repetitive (strand or helix, "ordered") structure (O/o) at loop position M.

Capitalization of the letter specifying the first (framing) turn in a CT specifies the loop in its strict baseline form, without additional BB H-bonds either within the loop or linking loop residues to any other residue in the 12-reside ExploreTurns window. An alternative, more permissive baseline loop definition, which only excludes structures with additional BB H-bonds that occur entirely within the loop, may be specified by prepending '@' to the CT label. The choice of baseline format can have a large impact on the number of structures selected if the loop is commonly in contact with other secondary structure: the Schellman loop, for example, frequently caps helical C-termini and receives additional H-bonds from the helix, and the strict baseline CT label "P-b2" reduces the size of the selected set by 73% compared to the unrestricted label "p-b2", while the permissive baseline label "@p-b2" reduces the set size by just 24%.

Loop variants with additional BB H-bonds can be ruled out more selectively using entries of the form:

*M[~N]

which exclude examples with additional BB H-bonds donated or received at loop position M, or (when ~N is included) within the position range from M to N inclusive.

CTs can occur at any position within the 12-residue ExploreTurns window, which is always centered on a beta turn. By default, a CT label selects the structures found in the loop's most common position frame, but alternative frames can be selected by appending the suffix:

_(P|*)

to the end of the label to select either the set of structures from the frame ranked P in the list of frames ordered by loop abundance, or (with "_*") select the merged set of unique loop structures from all frames (the same structures can occur in multiple frames due to overlapping turns/tails). Alternative position frames show the loop in different structural contexts, or they may be used to orient the loop in a more natural way in the turn-local coordinate system. Note that adding a sequence motif, dihedral constraint, or DSSP constraint to a CT label may alter the ordering of a loop's position frames, resulting in a shifted default frame.

A loop's position in the ExploreTurns window is indicated in the text displayed above the *BB H-bonds/Loop motif/Compound turn* box after the loop is loaded, in the format "M→N (P)", where M and N indicate the start and end positions of the loop in the ExploreTurns position frame centered on a beta turn, and P indicates the loop's offset from the central β-turn. If structures are selected from a single frame, then the loop position is the same for all structures in the set, but if the merged set of all structures is selected with "_*", then loop position is in general different for each structure, and the text above the *BB H-bonds...* box updates as each example is loaded.

The complete, romanized CT notation for entry into ExploreTurns is summarized, using single entries of each type, as:

(d|g|b...)['][*][M] - [A|C|D...]M - [\(F|f|Y|y)M] - [\(E|e|H|h|O|o)M] - [*M[~N]] ... [_(P|*)]

where the parentheses, brackets and bars are not entered, and white space is optional. The residue position specifier M is omitted from the entry for the framing turn (which must be listed first), but all other entries (except the frame-selection suffix) must include M. Capitalization of the framing turn applies the strict baseline format, while prepending '@' specifies the permissive form. If only a framing turn is specified, then a "simple" H-bonded turn is selected (note that the longer types of simple turns usually contain additional internal H-bonds). Apart from the entry that specifies the framing turn, entries may be listed in any order, but a convention followed here is that internal turns are ordered by increasing position in the loop followed by decreasing component-turn length, so that a β2-turn is listed before a β3-turn, which is itself listed before a (shorter) γ3-turn.

Although CT labels specify structure only within a loop's framing turn, adjacent structures can be specified or excluded by extending the label to cover neighboring residues. This technique can be used to search for variants with additional BB H-bonds that extend the structure of a given loop. For example, some 5C1 (type 1) BBLs (α'-β1 turns) are embedded within nine-residue H-bonded ι-turns, since these motifs commonly occur in the chain-reversing loops of 3:5 beta hairpins, where the ι-turn H-bond is the first beta-ladder H-bond outside the BBL. However, since 5C1 BBLs enclosed by ι-turns also occur independently of beta hairpins, this extended loop (which may have originated in a hairpin) qualifies as a structural motif in its own right, and the CT label "t-a'3-b3-\e1-\e2-\e8-\e9" selects extended 5C1 BBLs

which do not overlap beta ladders. Other loop motifs can also be extended, including the 6C1 (type 2) and 6C2 BBLs and the Schellman loop (see the loop browser).

Two ExploreTurns display features support the exploration of CT variants. When a loop is loaded with a CT label, the BB H-bonds displayed in the *H-bond profile* box (which include only those that occur within the loop) are labelled with their associated simple turn types, so that they can be copied into the CT label to investigate the structures of the loop's BB H-bond variants (see *H-bonded loop motifs/Examples/Ex. 3* in the feature summary). In addition, ranked statistics for the single-AA and pair sequence motifs that occur in the loop are displayed at the top of the profile window when a CT is loaded, so that sequence motif labels can be copied from the profile window into the CT label to investigate the loop's SC structures and their effect on its BB geometry (see *H-bonded loop motifs/Examples/Ex. 4* in the feature summary).

Table 1. Examples of compound turns

CT label	Motif	Translation
γ turn	γ-turn, most common position frame	g
α-β1 turn	Alpha-beta loop	
	All BB H-bond variants, most common frame	a-b1
	Heme binding with Cys and His, all frames	a-b1-C1-C4-H5_*
α'-β1 turn	5C1 (type 1) β-bulge loop (BBL):	
	All BB H-bond variants, most common frame	a'-b1
	BB H-bond variant adds a1, a "normal" α-turn	a'-a1-b1
	Adds D3T5 sequence motif	a'-b1-D3-T5
	"Extended" 5C1 BBL (\e's exclude β-ladder)	t-a'3-b3-\e1-\e2-\e8-\e9
π'-a1 turn	6C1 (type 2) BBL:	
	2nd most common position frame aligns the loop's "upright chair back" with the β-turn at the center of the ExploreTurns window	p'-a1_2
	Metal-ion binding in Zn fingers/ribbons	p'-a1-C1-C4
π'- β1 turn	6C2 BBL:	
	Baseline (all BB H-bond variants excluded)	P'-b1
	Excludes only 6C1 BBLs, using a*1	p'-a*1-b1
	"6C2-" variant (\f4), with T4 H-bond bridge	P'-b1-\f4-T4
	Metal-ion binding with an Asp triplet	P'-b1-D1-D3-D5
ρ'-β1-β4 turn	7NC3 BBL, baseline, all position frames	R'-b1-b4_*
τ-τ' turn	7:7 β-hairpin:	
	BB H-bonds excluded from loop with *1~9	t-t'-*1~9-\E1-\E9
	Ser, Asp H-bonds shape loop into a "hook"	t-t'-*1~9-\E1-\E9-S2-D4
π-β2 turn	Schellman loop:	
	All BB H-bond variants	p-b2
	Strict baseline	P-b2
	Permissive baseline	@p-b2
	Restricted to "dual" helix caps	p-b2-\H1-\H2
π-β2$\phi^{3,4}\phi_5$ turn	"Left-handed"	p-b2-\F3-\F4-\f5
π-α1-β1-β'3 turn	"Crown pyramid", 3rd position frame aligns "base" with the central ExploreTurns β-turn	p-a1-b1-b'3_3
ρ*-ρ'*-β1-α'3-β4ϕ_4 turn	"Beta bracket" (open loop), 2nd β-turn orients towards -z in the frame of the first (\f4)	r*-r'*-b1-a'3-b4-\f4

Notes: When a CT is loaded without specifying a baseline mode, all additional H-bonded component turns that can occur in the loop are listed in the *H-bond profile* box, and these turns may be copied to the CT label to investigate the effects of their BB H-bonds on loop structure. The loop's sequence motifs, ranked in the profile window, may also be copied into the label.

5.6 Beta-bulge loop nomenclature

Since beta-bulge loops (BBLs) have a characteristic general form (Section 3.1.6), they can be classified with a motif-specific notation in addition to the CT nomenclature. BBL types are labelled in the three-component format:

M(N|C|NC)P

which specifies the overall length of the loop (M), the location of the bulge(s) at the N-terminus, C-terminus, or both (for loops which have split pairs of BB H-bonds at each terminus), and bulge length (the number of residues added to form the bulge). If an NC-type BBL has bulges of different lengths, or bulges defined by overlapping pairs of split BB H-bonds, then the length of each bulge is given separately; e.g. 6N2C3, 7N3C1C3.

5.7 Other methods

The computation of sequence motif overrepresentation and p-value[15,48,69,70], the derivation of context vectors, and the H-bond definition are all described in the online user guide.

5.8 Languages and tools used

Biopython[71] was used to extract the structural data from PDB[49] files. The depths of beta turns beneath the protein's solvent-accessible surface were computed with Depth[72]. ExploreTurns was written in HTML/CSS and Javascript, with an embedded JSmol viewer[73]. The tool is tested for compatibility with the Chrome, Edge, and Firefox browsers; for best performance, browse with Chrome or Edge.

Supplementary Material

Supplementary file 1 (PDF)**.** Supplementary sections as described in the main text, 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.

Funding

This work was funded entirely by the author.

Acknowledgements

The author wishes to thank Athena Newell for her work as a research associate and Kit Newell for his work as a research assistant. The constructive suggestions 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. Isogai Y, Némethy G, Rackovsky S, Leach SJ, Scheraga HA. Characterization of multiple bends in proteins. *Biopolymers*. 1980;19(6):1183-210.
4. 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).
5. Paiz EA, Allen JH, Correia JJ et al. Beta turn propensity and a model polymer scaling exponent identify intrinsically disordered phase-separating proteins. *Journal of Biological Chemistry*. 2021;297(5).
6. Milner-White, EJ, Poet R. Four classes of beta hairpins in proteins. *Biochem J*. 1986; 240(1): 289-292.
7. 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.
8. Sibanda BL, Blundell TL, Thornton JM. Conformation of beta-hairpins in Protein Structures. A Systematic Classification with Applications to Modelling by Homology, Electron Density Fitting and Protein Engineering. *Journal of Molecular Biology*. 1989;206:759-777.

9. 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.

10. Craveur P, Joseph AP, Rebehmed J, de Brevern AG. β-bulges: extensive structural analyses of β-sheets irregularities. *Protein Science*. 2013;22(10):1366-78.

11. Richardson JS, Richardson DC. Natural β-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proceedings of the National Academy of Sciences*. 2002;99(5):2754-9.

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

13. Sudipta M, Jacoby G, Sawaya MR, Arnon ZA, Adler-Abramovich LA et al. Transition of Metastable Cross-α crystals into Cross-β Fibrils by β-turn flipping. *J Am Chem Soc*. 2019;141(1):363-369.

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

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

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. Milner-White EJ. Recurring loop motif in proteins that occurs in right-handed and left-handed forms. Its relationship with alpha-helices and beta-bulge loops. *Journal of Molecular Biology*. 1988;199(3):503-11.

18. 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.

19. Wilman HR, Shi J, Deane CM. Helix kinks are equally prevalent in soluble and membrane proteins. *Proteins*. 2014;82(9):1960-70.

20. Kessel A, Ben-Tal N. Introduction to Proteins: Structure, Function and Motion. 2nd ed. Boca Raton: Chapman & Hall/CRC; 2018.

21. Buchko GW, Ni S, Robinson H, Welsh EA, Pakrasi HB, Kennedy MA. Characterization of two potentially universal turn motifs that shape the repeated five-residues fold-Crystal structure of a lumenal pentapeptide repeat protein from cyanothece 51142. *Protein Science.* 2006;15(11):2579-95.

22. Fairman JW, Noinaj N, Buchanan SK. The structural biology of β-barrel membrane proteins: a summary of recent reports. Curr Opin Struct Biol. 2011; 21(4): 523-531.

23. Wierenga RK. The TIM-barrel fold: A versatile framework for efficient enzymes. *FEBS Letters*. 2001;492(3):193-8.

24. Pons T, Gómez R, Chinea G, Valencia A. Beta-propellers: associated functions and their role in human diseases. *Current Medicinal Chemistry*. 2003;10(6):505-24.

25. 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.

26. 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.

27. 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.

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

29. 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.

30. Marcelino AM, Gierasch LM. Roles of β-turns in protein folding: From peptide models to protein engineering. *Biopolymers*. 2008;89(5):380-91.

31. Gu H, Kim D, Baker D. Contrasting roles for symmetrically disposed β-turns in the folding of a small protein. *Journal of Molecular Biology*. 1997;274(4):588-96.

32. Alba ED, Rico M, Jiménez MA. The turn sequence directs β-strand alignment in designed β-hairpins. *Protein Science*. 2008;8(11):2234-44.

33. Fuller AA, Du D, Liu F, Davoren JE, Bhabha G, et al. Evaluating β-turn mimics as β-sheet folding nucleators. *Proceedings of the National Academy of Sciences*. 2009;106(27):11067-72.

34. Rose GD. From propensities to patterns to principles in protein folding. *Proteins*. doi: 10.1002/prot.26540

35. Lee J, Dubey VK, Longo LM, Blaber M. A logical OR redundancy within the ASX-pro-asx-gly type I β-turn motif. *Journal of Molecular Biology*. 2008;377(4):1251-64.

36. Fu H, Grimsley GR, Razvi A, Scholtz JM, Pace CN. Increasing protein stability by improving beta-turns. *Proteins*. 2009;77(3):491-8.

37. Hynes TR, Kautz RA, Goodman MA, Gill JF, Fox RO. Transfer of a beta-turn structure to a new protein context. *Nature*. 1989;339(6219):73-6.

38. Kokryakov VN, Harwig SSL, Panyutich EA, Shevchenko AA, Aleshina GM et al. Protegrins: Leukocyte antimicrobial peptides that combine features of corticostatic defensins and Tachyplesins. *FEBS Letters*. 1993;327(2):231-6.

39. Rose GD, Glerasch LM, Smith JA. Turns in peptides and proteins. *Adv Protein Chem*. 1985;37:1-109.

40. Metrano AJ, Abascal NC, Mercado BQ, Paulson EK, Hurtley AE, Miller SJ. Diversity of Secondary Structure in Catalytic Peptides with β-turn Biased Sequences. *J Am Chem Soc*. 2017;139:492-516.

41. Hruby VJ. Conformational and topographical considerations in the design of biologically active peptides. *Biopolymers*. 1993;33(7):1073-82.

42. Deike S, Rothemund S, Voigt B, Samantray S, Strodel B, Binder WH. β-turn mimetic synthetic peptides as amyloid-β aggregation inhibitors. *Bioorg Chem*. 2020;101.

43. 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).

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

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

46. 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.

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

48. Newell NE. Geometric descriptors for beta turns. *Protein Science*. 2024;33(9):e5159.

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

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

51. Newell, NE. 2024. MapTurns: mapping the structure, H-bonding and contexts of beta turns in proteins. *bioRxiv*. doi: https://doi.org/10.1101/2024.01.02.573949

52. Toniolo C, Crisma M, Moretto A, Peggion C, Formagggio F et al. Peptide delta-Turn: Literature Survey and Recent Progress. *Chemistry*. 2015;21(40):13866-77.

53. Gore S, Garcia ES, Hendrickx PMS, Gutmanas A, Westbrook JD et al. Validation of Structures in the Protein Data Bank. *Structure*. 2017;25(12):1916-1927.

54. Nemethy G, Printz MP. The gamma turn, a possible folded conformation of the polypeptide chain. Comparison with the beta turn. *Macromolecules*. 1972;5:755-758.

55. Mathews BW. The gamma turn. Evidence for a new folded conformation in proteins. *Macromolecules*. 1972;5:818-819.

56. Dasgupta B, Pal L, Gautam B, Chakrabarti P. Expanded turn Conformations: Characterization and Sequence-Structure Correspondence in alpha-Turns With Implications in Helix Folding. *Proteins*. 2004;55:305-315.

57. Rajashankar KR, Ramakumar S. Pi-turns in proteins and peptides: classification, conformation, occurrence, hydration and sequence. *Protein Sci*. 1996;5:932-946.

58. Dasgupta B, Chakrabarti P. pi-Turns: types, systematics and the context of their occurrence in protein structures. *BMC Structural Biology*. 2008;8:39.

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

60. Leader DP, Milner-White EJ. The conserved crown bridge loop at the catalytic center of enzymes of the haloacid dehalogenase superfamily. *Current Research in Structural Biology*. 2023;6.

61. 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 Npl4. *J. Biol Chem*. 2010;278(22):20225-20234.

62. 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.

63. Lu Z, Imlay JA. A conserved motif liganding the [4Fe-4S] cluster in [4Fe-4S] fumarases prevents irreversible inactivation of the enzyme during hydrogen peroxide stress. 2019. *Redox Biology.* doi: https://doi.org/10.1016/j.redox.2019.101296

64. Jiang H, Jude K, Wu K, Fallas J et al. De novo design of buttressed loops for sculpting protein functions. *Nature chemical biology*. 2024;20:974-980.

65. 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.

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

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

68. Richardson JS. The anatomy and taxonomy of Protein structure. *Adv. Protein Chem*. 1981; 34:167-339.

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

70. 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.

71. 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.

72. 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.

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