Extension of the classical classification of $\beta$-turns
Alexandre de Brevern

To cite this version:
Alexandre de Brevern. Extension of the classical classification of $\beta$-turns. Scientific Reports, 2016, pp.33191 10.1038/srep33191 . inserm-01374651

HAL Id: inserm-01374651
https://inserm.hal.science/inserm-01374651v1
Submitted on 30 Sep 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Extension of the classical classification of \(\beta\)-turns

Alexandre G. de Brevern

The functional properties of a protein primarily depend on its three-dimensional (3D) structure. These properties have classically been assigned, visualized and analyzed on the basis of protein secondary structures. The \(\beta\)-turn is the third most important secondary structure after helices and \(\beta\)-strands. \(\beta\)-turns have been classified according to the values of the dihedral angles \(\varphi\) and \(\psi\) of the central residue. Conventionally, eight different types of \(\beta\)-turns have been defined, whereas those that cannot be defined are classified as type IV \(\beta\)-turns. This classification remains the most widely used. Nonetheless, the miscellaneous type IV \(\beta\)-turns represent 1/3rd of \(\beta\)-turn residues. An unsupervised specific clustering approach was designed to search for recurrent new turns in the type IV category. The classical rules of \(\beta\)-turn type assignment were central to the approach. The four most frequently occurring clusters defined the new \(\beta\)-turn types. Unexpectedly, these types, designated IV\(_{1}\), IV\(_{2}\), IV\(_{3}\) and IV\(_{4}\), represent half of the type IV \(\beta\)-turns and occur more frequently than many of the previously established types. These types show convincing particularities, in terms of both structures and sequences that allow for the classical \(\beta\)-turn classification to be extended for the first time in 25 years.

The functional properties of a protein primarily depend on its three-dimensional (3D) structure. These properties have classically been assigned, visualized and analyzed on the basis of protein secondary structures, which are composed of repetitive parts (\(\alpha\)-helices represent 1/3rd of residues, and \(\beta\)-strands represent 1/5th of residues) connected by coils. This simplification of 3D structure into a unidimensional representation of secondary structure is often regarded as a resolved question. In fact, this simplification conceals the difficulty of precisely defining and assigning repetitive structures, thus explaining the large number of alternative assignment approaches. For instance, comparison of different approaches emphasizes their major discrepancies. Another limitation of this type of simplification is that the coil state is neglected, although it represents almost 50% of all residues and a large set of distinct local protein structures. Loop analyses cannot provide a complete representation of the coil state because their classification is usually limited to 8 residues. More precise descriptions are needed to comprehensively describe their diversity.

Helical and extended regions are the most frequently occurring repetitive structures. However, two other local protein conformations have also been characterized: the polyproline II helix and turns. The former is a left-handed helical structure with an overall shape resembling a triangular prism. It represents 5% of all protein residues, contributes to coiled coil super secondary structure formation and is present in fibrous proteins. Because polyproline II helices do not have strong hydrogen bond patterns, they have not been studied in as much detail as the other local conformations.

\(\beta\)-turns have been the most analyzed among the turn conformations. Apart from the distance between \(\cos\) of residues \(i\) and \(i+\) must be smaller than 7 \(\AA\) (or 7.5 \(\AA\), according to some authors). The turns are composed of \(\gamma\)-turns (\(n = 3\)), \(\beta\)-turns (\(n = 4\)), \(\alpha\)-turns (\(n = 5\)) and \(\pi\)-turns (\(n = 6\)). The restrictive distance between \(\cos\) applies a particular geometry to the backbone, thereby causing it to turn back on itself. \(\beta\)-turns have been the most analyzed among the turn conformations. Apart from the distance between \(\cos\), a second rule applies to the characterization of their secondary structure; because helices can easily be confused with a succession of turns, the central residues of \(\beta\)-turns, i.e., \(i+1\) and \(i+2\), should not be helical. Similarly, \(\beta\)-turn residues must not consist solely of \(\beta\)-strand residues. \(\beta\)-turns have been classified according to the values of their central residue dihedral angles, \(\varphi\) and \(\psi\). A deviation of \(\pm 30^\circ\) from these canonical values is allowed on 3 of these angles, whereas the fourth can deviate by \(\pm 45^\circ\).

1INSERM, U 1134, DSIMB, F-75739 Paris, France. 2Univ Paris Diderot, Sorbonne Paris Cité, UMR_S 1134, F-75739 Paris, France. 3Institut National de la Transfusion Sanguine (INTS), F-75739 Paris, France. 4Laboratoire d’Excellence GR-Ex, F-75739 Paris, France. Correspondence and requests for materials should be addressed to A.G.d.B. (email: alexandre.debrevern@univ-paris-diderot.fr)
The $\beta$-turns, as defined by C.M. Venkatachalam, are characterized by a hydrogen bond between the N-H and C=O of residues $i$ and $i+3$. Venkatachalam has also defined types I, II, and III, and their corresponding mirror image types, I', II' and III'. Crawford and collaborators have proposed a more strict definition in terms of distance. Lewis and co-workers have added types V and V'. $\beta$-turn type VI is characterized by the presence of a proline; type VII is associated with a kink; and type IV corresponds to all other non-classified $\beta$-turns. Different turns have been excluded for various reasons: $\beta$-turns III and III' are too close to the $\beta_3$-helix and types I and I', whereas turns V, V' and VII are rare, and their definitions are inaccurate. Type VI is divided into 2 sub-types, VIa and VIb. Hutchinson and Thornton have divided type VI into the 2 sub-types VIa and VIb. Wilmot and Thornton have precisely defined type VIII, which is based on Richardson's type IX, and was proposed after the removal of type VII. The definitions used by Thornton's group are currently considered to be the standard (see Supplementary Information 1). The $\beta$-turn assignment program PROMOTIF assigns $\beta$-turns on the basis of these standards. Studies have shown that repetitive structure assignment approaches have a direct effect on decreasing or increasing the number of residues associated with $\beta$-turns.

The difficulty with using such an approach is the 'strict' rule(s) used to define the $\beta$-turn types. Efimov has used a Ramachandran plot simplified to 6 and 8 regions: $(\beta_3, \beta_8, \alpha, \epsilon, \delta, \alpha)$ and $\gamma_5$ and $\gamma_6$). This rough clustering allows various classes to be defined, with some being associated with amino acid specific behaviours. The turns are also divided into full turns (with a polypeptide chain reversal of 180°) and half turns (with a polypeptide chain angle of 90°). The first category represents 7 major clusters, and the second one represents 8 major clusters. This system has widely been used to define super-secondary elements and structural trees of protein superfamilies. In a similar way, Wilmot and Thornton have also used a simplification of the Ramachandran plot for the following 6 major regions: $(\beta_3, \beta_8, \alpha, \epsilon, \delta, \alpha)$ and $\gamma_5$. They observed 12 combinations in their dataset. The most frequent turns were easily detected, whereas the two most interesting non-classical turns were $\beta_3 \rightarrow \gamma_5$ (8%) and $\gamma_6 \rightarrow \alpha$. The 6 other clusters represented only 1% each.

More recently, Koch and Klebe have proposed a combination of turns of different lengths ranging from 3 to 6 residues; the turns sometimes overlap, thus leading to complex categorizations. Koch and Klebe trained a very large modified Self-Organizing Map and extracted new types from the map. The assignment is provided as part of Sebbase, an extension module of Relibase. Koch and Klebe have used the identified new types in a second step to perform a prediction from the sequence. This approach is innovative, but it has not been implemented as a web tool and is therefore less used. George Rose's group has conducted research with a focus on the rationalization of turn types. Rose's group has extended this approach to redraw the Ramachandran plot.

However, none of these approaches has succeeded in superseding the classical definition of $\beta$-turns. A major shortcoming of past $\beta$-turn classification concerns the classification of type IV $\beta$-turns, i.e., the miscellaneous category, because it represents $1/3$ of all $\beta$-turn residues and is the second most common type of $\beta$-turn. To locate potentially new recurrent conformations in this miscellaneous type, an automatic clustering approach based on the rules of $\beta$-turn assignment was designed. It is related to Self-Organizing Maps and takes into account the specificity of $\beta$-turn assignment rules. All type IV $\beta$-turns were clustered. The four most occurring clusters were chosen as new types and analysed. Unexpectedly, these sub-types, denoted IVa, IVb, IVc and IVd, represent half of the type IV $\beta$-turns and occur more frequently than many of the classical types.

Methods

Data sets. To remove representative bias regarding protein resolution or sequence identity, non-redundant datasets were used. These datasets were generated using the PISCES database. As previously performed in, 10 sets of proteins were defined. Each contained no more than 3% pairwise sequence identity (with x ranging from 20 to 90%). The selected chains had X-ray crystallographic resolutions less than 1.6 Å or 2.5 Å and R-factors less than 0.25 or 1.0. They comprised 2,542 and 23,943 protein chains. Each chain was automatically examined with geometric criteria to avoid bias from zones with missing density. The main purpose of such diversification was to examine (i) the poorly populated turns and (ii) the stability of the clustering approach (see below).

Secondary structure assignment. Secondary structure assignment was performed with DSSP (CMBI version 2000) using the default parameters. DSSP yields more than three states, so we reduced them to the following: the $\alpha$-helix, containing $\alpha_1$ and $\pi$-helices; the $\beta$-strand, containing only the $\beta$-sheet; and the coil, comprising everything else ($\beta$-bridge, hydrogen bond turn, bend, and coil). Turn assignment was performed as described previously using the following classical rules: the distance between residues $i$ and $i+3$ should be less than 7 Å; the central residues of the turns must be non-helical; and in the case of strands, at least one residue must be associated with a coil. The types of turns (I, I', II, III', V, V', VIa, VIb, and VIII) were assigned according to the classical definition by using the $\varphi$ and $\psi$ dihedral angles of the central residues (see Supplementary Information 1). The turns were required to be less than 30° from the canonical values (at most one angle was allowed to deviate by $+/\sim 45^\circ$). Types V, VIa, and VIb were characterized by a cis-proline at position i+2. Turns that did not fit any of the above criteria were classified as type IV. The turns were also classified into two classes according to their function as described by Efimov: full turns resulting in a chain reversal of 180° and half turns that change everything else.

Protein Blocks. Protein Blocks (PBs) corresponded to a set of 16 local prototypes, labelled from a to p, of 5 residue length that were described on the basis of dihedral angles ($\varphi$, $\psi$). The PBs were obtained with an unsupervised classifier similar to Kohonen Maps and hidden Markov models. The PBs m and d are prototypes...
for the central regions of α-helix and β-strands respectively. PBs α through e primarily represent the N-cap of a β-strand, whereas e and f correspond to the C-caps; PBs g through j are specific to coils, PBs k and l correspond to the N cap of an α-helix, and PBs n through p correspond to C-caps. PBs were assigned by using in-house Python software, although similar assignment can be performed through the PBE web server45 or PBxplore (https://github.com/pierrepo/PBxplore66).

Specific clustering approach. A specific clustering approach was designed to cluster type IV β-turns by using the classical rule, allowing +/− 30° for all angles, with the exception of one at +/− 45° for the defined values. The clustering derived from Self-Organizing Maps (SOM, without diffusion between the clusters53,54).

The training was carried out in 2 successive parts: the first one limited the potential bias of initialization, and the second refined the clustering by using the specific rules for β-turn types. The type IV β-turns were selected from a dataset D. Thus, each dataset was associated with T type IV β-turns.

Step one:

1. k clusters were created and were vectors v of length 2M = 4, representing the dihedral angles (ϕi+1, ωi+1, ψi+2, and ωi+3), k type IV β-turns were taken randomly to initialize the clusters.
2. One of the T type IV β-turns was randomly selected from the dataset D (denoted Vc) and compared with each of the k clusters.
3. The dissimilarity measure between two vectors Vi (representing the clusters) and Vc of dihedral angles was defined as the Euclidean distance among the M links, the RMSDA (root mean square deviations on angular values67):

$$\text{RMSDA}(V_i, V_c) = \sqrt{\frac{\sum_{t=1}^{T-2}[(\Phi_i(V_t) - \Phi_i(V_c))^2 + |\Psi_i(V_t) - \Psi_i(V_c)|^2]}{2M}}$$  

where {Φi(Vt), Ψi(Vt)} (resp. Φi(Vc), Ψi(Vc)) denotes the series of the (2M) dihedral angles for Vi (resp. Vc). The angle differences were computed modulo 360°. Thus, in the training, this distance was used for assessing the dissimilarity of any fragment in the database with the different clusters.

4. The minimal RMSDA value was used to define the winning cluster W, i.e., the closest to the observation. W values were modified according to the learning coefficient α:

$$\Phi_j(V_0)^{t+1} = \Phi_j(V_0)^t + (\Phi_j(V_0) - \Phi_j(V^*)^t) \times \alpha^t$$  

$$\Psi_j(V_0)^{t+1} = \Psi_j(V_0)^t + (\Psi_j(V_0) - \Psi_j(V^*)^t) \times \alpha^t$$  

where {Φj(V0)} and Ψj(V0) are the values of the winner at time t, with j ranging from 1 to 2, similar to the values of the real data (i.e., dihedral angles i+1 and i+2, modulo 360°).

$$\alpha^t = \frac{\alpha_0}{1 + \frac{t}{T}}$$

The decrease of α was performed similarly to that for SOM53,54. T represents the total amount of data to learn (here the number of type IV β-turns). T represents the number of β-turns already used. The process goes back to step 2. One cycle of training corresponds to the learning of the whole dataset α0, which is then equal to α0/2; after 5 cycles, it is equal to α0/5, etc. Initially, α0 = 0.35, as in68,69. The process was iterated for 20 cycles, i.e., 20 times T; these steps were important to diminish the potential effect of the initialization.

Step two:

1. The final values of the k clusters were used as initial values. α0 was still equal to 0.35.
2. One of the T type IV β-turns was randomly selected from the dataset D (denoted Vc) and compared with each of the k clusters. Instead of using only RMSDA, the β-turn rule was used: 3 angles can be at +/− 30° and 1 angle at +/− 45°.
3. The winner positively applied this rule; otherwise no training was performed.
4. Modification of the winner weights was performed as in step one − 3.
5. The process was iterated for 20 cycles.

An important point is the choice of k. k was first set at 50 and then reduced. The obtained clusters were compared in the order of largest to smallest k values.

Z-score. The amino acid occurrences for each local structure conformation were normalized into a Z-score:

$$Z(n_{ij}) = \frac{n_{ij}^{obs} - n_{ij}^{cl}}{\sqrt{n_{ij}^{cl}}}$$
where $n_{ij}^{th}$ is the observed number of occurrences of amino acid $i$ in position $j$ for a given secondary structure, and $n_{ij}^{th}$ is the expected number. The product of the occurrences in position $j$ with the frequency of amino acid $i$ in the entire databank equals $n_{ij}^{th}$. Positive Z-scores (respectively negative) corresponded to overrepresented amino acids (respectively underrepresented); threshold values of 4.42 and 1.96 were chosen (probability less than $10^{-5}$ and $5.10^{-5}$, respectively). The same computation was also performed for the protein blocks.

**Analysis.** Most of the quantitative analysis was performed using in-house Python scripts, and statistics and visualization were performed with R software (version 3.2.2).

**Results and Discussion**

**Protein structure dataset.** The different amino acid datasets showed the expected amino acid and protein block occurrences, with no peculiarities in the rate of redundancy and the resolution quality (see Supplementary Information 2). As noted previously, the occurrence of $\beta$-turns is highly dependent on the way in which the assignment is performed. Following the work of Fuchs and Alix, we assigned secondary structures to the different protein datasets by using DSSP. The DSSP provided 8 classes that were reduced to 3 classes (helix, strand and coil) or 4 classes (helix, strand, turn and coil, see Supplementary Information 3) for practicality. Helical structures represented more than 37.3% of the residues and the $\beta$-sheets represented 22.5%, whereas the remaining coil class covered 42.7% of the residues and included 20.4% of the $\beta$-turns (11.9% were turns and 8.5% were bends). Our $\beta$-turn assignment in the coil regions provided a slightly different number, with 21.9% being $\beta$-turns (difference: 1.5%). In total, 71.8% were similar to the DSSP assignment (45.6% were turns, and 23.0% were bends), whereas 28.1% and 1.9% were associated with coils and bridges, respectively. These proportions were comparable to the results of previous studies. The $\beta$-turn types were then assigned by using classical definitions (described in the methods section, see Supplementary Information 1). Type I $\beta$-turns were the most frequent (38.2%), followed by the miscellaneous type IV (31.7%), and types II (11.8%), VIII (9.8%), I' (4.1%), II' (2.5%) and the different sub-types of the type VI $\beta$-turns (ranging from 0.9 to 0.2%, see Table 1). Henceforth, the type IV $\beta$-turns will be denoted type IVori to differentiate them from the new types in the current analyses. Figures 1 and 2 show the different types of $\beta$-turns in 3D and the distribution of their dihedral angles in the Ramachandran plot.

**Analyses of discarded types.** As a first step, before searching for new types, the previously discarded types were analysed.

Notably, type III and III' $\beta$-turns had been included by Venkatachalam, but have been discarded because they are considered to be too close to the $\beta_\parallel$ helices and to type I (and I') $\beta$-turns. The type V $\beta$-turn has been considered to be a rather unusual departure from the type II $\beta$-turn (see Figures 35 and 36 of ref. 35). If the type III $\beta$-turn was still recognized, it would represent 9.6% of the residues; i.e., it would be the third most frequently occurring type. The obsolete type III' $\beta$-turn represented approximately 1.5% of the turns, whereas the type V and V' $\beta$-turns represented only 0.03 and 0.02%, respectively (see Supplementary Information 4), and were associated with type IV $\beta$-turns (see Supplementary Information 5), but they were negligible.

For the type III and III' $\beta$-turns, the overlap with type I and I' $\beta$-turns remained as expected, with 88.7% of the type III $\beta$-turns assigned as type I $\beta$-turns, 87.6% of the type III' $\beta$-turns assigned as type I' $\beta$-turns (see Supplementary Information 4 and 6), and the remaining 11–12% associated with type IV $\beta$-turns. Interestingly, 60% of type I $\beta$-turns were also assignable to type III, and 83.9% of type I' were assignable to type III' (see Supplementary Information 7). Therefore, the decision to remove this particular definition was clearly reasonable.

**Searching for new types.** From the above section, it is apparent that nearly 1/3rd of residues are not associated with a defined type. Moreover, as presented in the methods section, learning was performed on the type IV $\beta$-turns, the clustering was conducted on the basis of dihedral angles with an unsupervised approach similar to the approaches used for protein blocks. The first step of learning was entirely unsupervised and was performed to properly define the initial values of the clusters, whereas the second step dictated the specific rules of the $\beta$-turns (e.g., +/− 30° and one dihedral angle at +/− 45°).

| $\beta$-turn | (%) |
|--------------|-----|
| I            | 38.21|
| II           | 11.81|
| VIII         | 9.84 |
| I'           | 4.10 |
| II'          | 2.51 |
| Vb           | 0.88 |
| Vb'          | 0.73 |
| Vb''         | 0.20 |
| IVori        | 31.72|
| Sum          | 100.00|

Table 1. $\beta$-turn frequencies. Classical types and their frequencies. Type Vb, Vb' and Vb'' $\beta$-turns are characterized by a cis-proline at position $i$+2. Type IV is denoted IVori to distinguish it from the new classification.
A major difficulty in every classification approach is the choice of the clusters. Here, it was slightly different; the idea was not to have an optimal number of clusters but to assess the most frequently occurring and recurrent clusters to define the new pertinent types. In related research, Micheletti and collaborators have decided to take the largest cluster each time and iteratively repeat the clustering, each time removing the largest cluster. This clustering is slightly unstable because each repetition removes a large amount of data. Thus, it did not seem pertinent to use it here. Moreover, with a large initial number of clusters, determining the clusterability of the data was manageable.

The training was performed with different datasets beginning with a large number of clusters (50 at first), which was progressively reduced (to 10). A notable feature of the learning was that four clusters appeared at the beginning and remained the most frequently occurring cluster for each of the different datasets. The deviation in the dihedral angle values between the different simulations (and different datasets) was never higher than 0.3°, thus indicating that the clustering was reasonably stable (a more detailed description is provided in Supplementary Information 8).

Figure 1. β-turn representation (beginning). (a–d) Type I, (e–h) type II, (i–l) type VIII, and (m–p) type I’. A turn close to the ideal values of its type (a,e,i,m) within a protein and (b,f,j,n) a close-up of the turn. Type I is represented by PDB id 2BK9, type II by PDB id 1H16, type VIII by PDB id 1SU8 and type I’ by PDB id 1KKO. (c,g,k,o) Ramachandran plot (ϕ, ψ) of residue i+1 and (d,h,l,p) of residue i+2; red dots are the ideal values. The number of observations of both residues is strictly identical.
The four new type IV β-turn sub-types were named IV₁, IV₂, IV₃, and IV₄. They represent half of the original type IV structure, composing 16.1, 12.4, 11.2 and 8.5% of the IV₁ type, respectively. In regards to all of the

| new β-turn | (%) | (%) of β-turn IV₁  |
|------------|-----|-------------------|
| IV₁        | 5.10| 16.08             |
| IV₂        | 3.95| 12.44             |
| IV₃        | 3.53| 11.15             |
| IV₄        | 2.70| 8.50              |
| IVmisc     | 16.44| 51.83             |
| Sum        | 31.72| 100.00            |

Table 2. β-turn frequencies. The four new β-turns are denoted IV₁ to IV₄, and the remaining residues are assigned to type IVmisc. Their frequencies in regards to the turns and to the original type IVmisc are provided.
defined types, they were the 4th, 6th, 7th and 8th most frequent turns (5.10%, 3.9%, 3.5% and 2.7%, respectively). These numbers are reasonable because they were highly consistent across all of the datasets. Figure 3 shows these four new categories. The remaining clusters were not selected because (i) their occurrences were very low (largely less than those of type VI $\beta$-turns) and (ii) they were often dependent on the number of clusters (see Supplementary Information 9). They were not useful for either protein structure or sequence–structure relationship analyses. The rest of the type IV $\beta$-turns were classified as IV.misc.

**Figure 3. New $\beta$-turn representation.** (a–d) Type IV$_1$, (e–h) type IV$_2$, (i–l) type IV$_3$, and (m–p) type IV$_4$ (see Fig. 1 for legend). Type IV$_1$ is represented by PDB id 1JYK, type IV$_2$ by PDB id 1URS, type IV$_3$ by PDB id 1PA7, and type IV$_4$ by PDB id 1QWG.

**Table 3. New $\beta$-turns.** Dihedral angle values of the four new turns.

| $\beta$-turn | $\phi_{1+1}$ | $\psi_{1+1}$ | $\phi_{1+2}$ | $\psi_{1+2}$ |
|--------------|---------------|---------------|---------------|---------------|
| Type IV$_1$   | $-120.0$      | 130.0         | 55.0          | 41.0          |
| Type IV$_2$   | $-85.0$       | $-15.0$       | $-125.0$      | 55.0          |
| Type IV$_3$   | $-71.0$       | $-30.0$       | $-72.0$       | $-47.0$       |
| Type IV$_4$   | $-97.0$       | $-2.0$        | $-117.0$      | $-11.0$       |
Table 3 provides the observed angles. Because the clustering approach was based on the specific clustering of type IV, no overlap could be found with the existing types. Figures 4a,b show the relative position of each turn. A relationship was observed between type IV1 and type II β-turns (see Fig. 4c) and between type IV2 and VIII β-turns (see Fig. 4d, see Supplementary Information 10). In terms of dihedral angle values, the type IV1 β-turn resembled a slightly displaced conformation of the type II β-turn, whereas the type IV2 β-turn appeared to be a less extended type VIII β-turn. Type IV3 and IV4 were much more specific, with very particular dihedral angles in the helical regions (see Supplementary Information 11).

New turns in regards to DSSP. To describe the type IV β-turns more precisely, we examined their former DSSP assignments (hydrogen bond estimation) as turns or bends. Interestingly, more than 2/3 of the residues of IVori were identified by DSSP as turns, with 35% being bends and 37% being hydrogen bond turns, and the rest were mainly associated with coils and β-sheets. The type IVmisc was more associated with non-hydrogen bond, stabilized local structures, with a 41% enrichment in bends and 31% fewer hydrogen bond turns. This evolution is mainly associated with the newer and less frequent type IV β-turns (e.g., type IV1 and IV4), which comprise 70% and 49% fewer hydrogen bond turns. The evolution was strikingly lower for the type IV1 β-turn, with less than 30% of residues associated with hydrogen bond turns. Although all the new type IV β-turns were linked to neither α-helices nor β-sheets, type IV1 β-turns were often observed at the ends of β-sheets (in nearly 2/3 of the cases).

Comparison with previous analyses. As mentioned in the introduction, two major efforts were made in the 1980s and 1990s to define β-turns. Both were based on a Ramachandran plot divided into 6 to 8 large regions. The size and shape of these regions were largely different from the strict rule of +/- 30° (and 45°). Notably, these
previous classifications were performed with all turns, whereas in the current analyses the classification was performed on only a subset of type IV $\beta$-turns.

Table 4 shows the new turns classified using a Ramachandran plot division scheme similar to that described above. Efimov has proposed a very precise definition of turns and half-turns with 7 and 8 types of turn$^{44,45}$. Interestingly, type IV$_1$ might seem as if it could be characterized as $\beta\delta\alpha\gamma$, because it looks like the proposed $\beta\alpha\delta$-half-turn; however, the type IV$_1$ $\beta$-turn is not a half-turn but a complete turn. The type IV$_2$ $\beta$-turn is the only local conformation that can be described as a half-turn, but instead of being a $\alpha\gamma$-half-turn, it is mainly $\alpha\beta\gamma\alpha$. Type IV$_3$ $\beta$-turns can be described as a $\beta\delta$-turn, but here it is mainly a turn, whereas the previously described types were half-turns. In fact, the type IV$_2$ $\beta$-turns were the only ones that seemed to be directly related to Efimov’s analyses, because they could be characterized by a $\gamma\delta$ connection between $\alpha$-helices, as described in$^{45}$. The percentage of turns and half-turns observed correctly correlated with the distance threshold proposed by Crawford and co-workers$^{37}$.

Wilmot and Thornton have also used a simplification of the Ramachandran plot in 6 major regions, with 12 combinations$^{51}$. Because the size of the different regions is higher than Efimov’s, the number of types is relatively limited. The region $\alpha\gamma\delta\beta$ represents the $\gamma\delta\beta\alpha$ regions; very diverse conformations were found in type IV$_3$ and IV$_4$ $\beta$-turns as well as type I $\beta$-turns (i.e., $\alpha\gamma\delta\beta$). Type IV$_2$ $\beta$-turns had the same description as type VIII (i.e., $\alpha\gamma\delta\beta$). Interestingly, only two non-classical turns, $\beta\delta\alpha\gamma$ (8%) and $\gamma\delta\alpha\beta$ (4%)$^{51}$, were defined by Wilmot and Thornton. One could expect that one of these two types might be associated with the most frequent new turn. However, this was not the case, because the type IV$_1$ $\beta$-turn is not $\beta\delta\alpha\gamma$ but $\beta\delta\alpha\gamma$.

Hence, these comparisons illustrate that the specific clustering performed in the current analyses highlighted a new main cluster that was not observed previously: the type IV$_1$ $\beta$-turn. Additionally, it showed the specificity of the type IV$_1$ and IV$_4$ $\beta$-turns in regards to their fine description. The type IV$_1$ $\beta$-turn was the only one to have been clearly characterized previously by both studies$^{42,51}$.

Koch and Klebe (KK) used a sophisticated approach to unify the assignment of turns of different lengths$^{52}$. This approach is not easily comparable to others because: (i) it is not based on the classical assignment rules and (ii) all the turns have been re-assigned. Hence, for $\beta$-turns, other features were used in the training in addition to the values of the dihedral angles $(\varphi, \psi)$ of the central residue. Classical and new $\beta$-turns were compared to the final definition of the 24 open KK $\beta$-turns (7 were considered to be non-turn-like structures) and 18 reverse KK $\beta$-turns presented in Supplemental Data S14 and S16 of ref. 52. Owing to the particular learning method, type I, II and II’ $\beta$-turns had no direct equivalent in the KK $\beta$-turns, whereas type I, IV$_1$, and IV$_4$ $\beta$-turns were associated with the KK type I $\beta$-turn (18% of the true turns). Type VIII $\beta$-turns were associated with the KK type VIII$^{33}$$\beta$-turn (6.5% of the true turns). Interestingly, type IV$_2$ $\beta$-turns were not associated with any KK $\beta$-turn types.

Hence, this comparison between studies indicated some similarities because the major turn (type I $\beta$-turn) could not distinguish between the two less frequent types (types IV$_3$ and IV$_4$ $\beta$-turn), whereas type VIII $\beta$-turns were easily found by using this approach. Similarly to previous results, the type IV$_4$ $\beta$-turn remained specific to our clustering. However, differences between the studies should be taken into account, such as the different learning method used by Koch and Klebe, considered more angles than ours and their training was conducted on the complete set of turns and not just the type IV $\beta$-turns.

Comparison with protein blocks. Table 5 shows the over- and under-representation of protein blocks for all the $\beta$-turn types. Type IV$^{w}$ $\beta$-turns were characterized by a PB motif of $[efghijko]$ $[bhijkno]$ $[gbhijiknop]$ $[acgip]$. As expected, this signature was more ambiguous in regards to the well-defined types, which showed a range of only one to four PBs at each position. The IV$^{w}$ represented only half of the previous $\beta$-turn IV$^{w}$ types. The only exception was the newly over-represented PBs $n$ and $p$ at positions $i$ and $i+1$ as well as the reduced over-representation of PBs $n$ and $p$ at positions $i+1$ and $i+3$, whereas 28/32 over-representations remained the same.

The newly defined type IV $\beta$-turns had stronger PB motifs. They could be analysed not only in regards to $\beta$-turn IV$^{w}$ but also in regards to II and VIII for types IV$_1$ and IV$_2$.

For type IV$_1$, the PB motif is $[aegp]$ $[aeghp]$ $[hikp]$ $[aill]$ and has no direct contradiction with the classical behaviours of $\beta$-turn IV$^{w}$. However, the PB motif of type II $\beta$-turns was less ambiguous, with only two main PBs at each position $[cg]$ $[hao]$ $[ik]$ $[al]$. IV$_3$ $\beta$-turns were clearly different, with 8 over-represented PBs that were under-represented in type II $\beta$-turns (PBs $a$ and $p$ at position $i$, PBs $a$, $e$ and $g$ at position $i+1$, PBs $h$ and $p$ at position $i+2$ and PBs $i$ at position $i+3$). Similarly, in type IV$_2$ $\beta$-turns, the PB motif was $[fkl]$ $[bklno]$ $[bglp]$ $[cg]$ and was comparable to

| $\beta$-turn | Thornton (1990) | Efimov (1986) | Efimov (1986) | Crawford (1973) |
|-------------|-----------------|----------------|----------------|-----------------|
| $i+1$       | $i+2$           | $i+1$         | $i+2$         | $d<5.7\,\AA$   |
| $\beta$-turn IV$_1$ | $\beta_2$ $\alpha_3$ | $\beta_2$ $\alpha_3$ | 99.4 | 0.6 | 23 |
| $\beta$-turn IV$_2$ | $\alpha_3$ $\beta_3$ | $\gamma$ $\delta$ | 72 | 28 | 37 |
| $\beta$-turn IV$_3$ | $\alpha_3$ $\beta_3$ | $\gamma/\alpha$ $\alpha$ | 71 | 29 | 54 |
| $\beta$-turn IV$_4$ | $\alpha_3$ $\beta_3$ | $\gamma$ $\gamma$ | 65 | 35 | 28 |
| $\beta$-turn IV$^{w}$ | $\alpha_3$ $\beta_3$ | $\gamma$ $\gamma$ | 64 | 36 | 19 |

Table 4. Torsion angle regions taken from Wilmot and Thornton, and Efimov, with turns and half-turn proportions as defined by Efimov and distance in regards to Crawford.
4. Interestingly, the type IV1 and II β-turns were clearly distinct in terms of dihedral angle distributions but had very similar amino acid compositions. Indeed, they shared the same over- or under-represented amino acid trends in 80% of the cases; only one inversion of amino acid preference was observed for the type IV3 β-turns at position i + 2 (alanine), because erasing half of the occurrences did not change the general trend of the unassigned turns.

Concerning the new turns defined in the current analyses, the four important points are as follows:

1. Type IVori and IVmisc β-turns remained strongly linked, because erasing half of the occurrences did not change the general trend of the unassigned turns.
2. IV3 and IV4 were clearly distinct in terms of dihedral angle distributions but had very similar amino acid compositions. Indeed, they shared the same over- or under-represented amino acid trends in 80% of the cases; only one inversion of amino acid preference was observed for the type IV3 β-turns at position i + 2 (alanine), because erasing half of the occurrences did not change the general trend of the unassigned turns.
3. The type VIII and IV2 β-turns were structurally close, with high sequence similarity. We found only one inversion between these types at position i + 2 for the valine residue.
4. Interestingly, the type IV3 and IV4 β-turns were close structurally but had strongly divergent sequences. At position i, no common amino acid over- or under-representation was observed. In the Ramachandran plot’s αi region, glycine represented 88% of the residues, whereas in γi, it was only 38% (with N 17%, D 9%, K 5%, E and R 4%, respectively). Interestingly, the type IV3 encompassed mainly the non-glycine residues at positions i and i + 1, and PBs b and p at position i + 2.

PB motifs of type IV3 and IV4 were clearly distinct in terms of dihedral angle distributions but had very similar amino acid compositions. Indeed, they shared the same over- or under-represented amino acid trends in 80% of the cases; only one inversion of amino acid preference was observed for the type IV3 β-turns at position i + 2 (alanine), because erasing half of the occurrences did not change the general trend of the unassigned turns.

Concerning the new turns defined in the current analyses, the four important points are as follows:

1. Type IVori and IVmisc β-turns remained strongly linked, because erasing half of the occurrences did not change the general trend of the unassigned turns.
2. IV3 and IV4 were clearly distinct in terms of dihedral angle distributions but had very similar amino acid compositions. Indeed, they shared the same over- or under-represented amino acid trends in 80% of the cases; only one inversion of amino acid preference was observed for the type IV3 β-turns at position i + 2 (alanine), because erasing half of the occurrences did not change the general trend of the unassigned turns.
3. The type VIII and IV2 β-turns were structurally close, with high sequence similarity. We found only one inversion between these types at position i + 2 for the valine residue.
4. Interestingly, the type IV3 and IV4 β-turns were close structurally but had strongly divergent sequences. At position i, no common amino acid over- or under-representation was observed. In the Ramachandran plot’s αi region, glycine represented 88% of the residues, whereas in γi, it was only 38% (with N 17%, D 9%, K 5%, E and R 4%, respectively). Interestingly, the type IV3 encompassed mainly the non-glycine residues at i + 2 (see Table 4). Moreover, proline and glycine residues were under-represented at position i + 2 of type II, although they were over-represented in type VIII β-turns. Additionally, the i + 2 positions of both types had more divergent residues. Figure 5 shows a Sammon map projection of all the β-turns. It emphasizes these relationships and highlights the strong differences between types IV1 and II, with the distance being quite substantial.

The type IV3 β-turn amino acid composition was similar to that of the two other new β-turn types, IV3 and IV4 (see Supplementary Information 12 and 13).

### Table 5. Protein blocks’ Z-scores of β-turn types.

| β-turn | i  | i+1 | i+2 | i+3 |
|--------|----|-----|-----|-----|
| β-turn I | (+) | FJKL | KLN | BGLOP | CGOP |
| β-turn II | (+) | αBCDEFGHIJKLMNOP | ABCDEFGHIJKLMNOP | ACDEFGHIJKLMNOP | ABDEFGHIJKLMNOP |
| β-turn III | (-) | EG | HO | JK | AL |
| β-turn IV | (-) | αBCDFIJKLMNOP | ABCDEFGIJKLMNOP | ABCDEFGIJKLMNOP | ABCDEFGIJKLMNOP |
| β-turn V | (+) | AFKP | ABL | CDGL | bCFK |
| β-turn VI | (-) | BCDHJLNO | BCDHJLNO | BCDHJLNO | BCDHJLNO |
| β-turn VII | (+) | EGHJNO | HO | IP | A |
| β-turn VIII | (-) | bCDHJLMP | bCDHJLMP | bCDHJLMP | bCDHJLMP |
| β-turn IX | (+) | abHO | J | ARLO | CGLP |
| β-turn X | (+) | DiM | abcDkMmp | cDkmp | adkm |
| β-turn XI | (-) | Cdp | EF | BHK | BIL |
| β-turn XII | (-) | bM | dkM | cdDm | cKm |
| β-turn XIII | (+) | Bi | aeFg | BK | gjo |
| β-turn XIV | (-) | m | m | m | m |
| β-turn XV | (+) | bCdi | aCD | Df | df |
| β-turn XVI | (-) | bM | DkM | kLM | cLM |
| β-turn XVII | (+) | EFGHJKO | BHJKLNO | ABGHJLNO | AGCGOP |
| β-turn XVIII | (-) | BCDM | CDFM | CDEFK | DFMK |
| β-turn XIX | (+) | AEGp | aEGHb | HJKP | AIL |
| β-turn XX | (-) | cDkMmp | bcDkMmp | bcDkMmp | cDkMmp |
| β-turn XXI | (+) | FJKL | BKLno | bGLP | cG |
| β-turn XXII | (+) | bcDefhmop | CDFghblop | acDefhkm | aDefhL |
| β-turn XXIII | (+) | Fjk | KL | BLMNop | cGMop |
| β-turn XXIV | (+) | bcDefhmop | abCDeFhlop | aCDeFhik | abDeFhik |
| β-turn XXV | (+) | FJKL | KLN | BGLOP | CGOP |
| β-turn XXVI | (+) | bDehmno | acDfihmp | acDefhkm | abDefhkm |
| β-turn XXVII | (+) | EFGHJKLNO | BHIJKLNO | ABHIJKLNO | AbCGP |
| β-turn XXVIII | (-) | bCDM | CDEFM | cDkM | DeM |

The type IVori β-turns but also had some differences compared with the type VIII β-turns. Hence, only half of the over-represented PBs in type VIII β-turns were found in type IV3 β-turns and 5 under-represented PBs were over-represented (PBs k, n and p at position i + 1, and PBs b and p at position i + 2).

PB motifs of type IV3 and IV4 β-turns were mainly associated with the most frequent β-turn, the type I β-turn, because their dihedral angles were in the same restricted area.

**Amino Acid Specificities of the new types.** β-turns have been widely analysed in terms of sequence–structure relationships, which have been incorporated in various prediction approaches. Table 6 shows the under- and over-represented amino acids in each type of turn. Some associations were expected because all of the different type VI β-turns were characterized by the proline at position i + 2.

Concerning the new turns defined in the current analyses, the four important points are as follows:

1. **Type IVori and IVmisc β-turns** remained strongly linked, because erasing half of the occurrences did not change the general trend of the unassigned turns.
2. **IV3 and IV4** were clearly distinct in terms of dihedral angle distributions but had very similar amino acid compositions. Indeed, they shared the same over- or under-represented amino acid trends in 80% of the cases; only one inversion of amino acid preference was observed for the type IV3 β-turns at position i + 2 (alanine), because erasing half of the occurrences did not change the general trend of the unassigned turns.
3. The type VIII and IV2 β-turns were structurally close, with high sequence similarity. We found only one inversion between these types at position i + 2 for the valine residue.
4. **Interestingly**, the type IV3 and II β-turns were close structurally but had strongly divergent sequences. At position i, no common amino acid over- or under-representation was observed. In the Ramachandran plot’s αi region, glycine represented 88% of the residues, whereas in γi, it was only 38% (with N 17%, D 9%, K 5%, E and R 4%, respectively). Interestingly, the type IV3 encompassed mainly the non-glycine residues at i + 2 (see Table 4). Moreover, proline and glycine residues were under-represented at position i + 3 of type II, although they were over-represented in type VIII β-turns. Additionally, the i + 2 positions of both types had more divergent residues. Figure 5 shows a Sammon map projection of all the β-turns. It emphasizes these relationships and highlights the strong differences between types IV1 and II, with the distance being quite substantial. The type IV3 β-turn amino acid composition was similar to that of the two other new β-turn types, IV3 and IV4 (see Supplementary Information 12 and 13).
Conclusions

β-turns are the most important secondary structures preceded by the α-helix and β-sheet. β-turns correspond to approximately 25 to 30% of all protein residues. The current classification of the different β-turns has remained unchanged for the past 30 years. In the 1980s and 1990s, different studies proposed extending the definition of turns, mainly on the basis of the division of a Ramachandran plot into 6 to 8 regions. These analyses of β-turns showed strong similarities with classical analyses and provided new definitions for the least frequently occurring turns. Two recent studies have expressed interest in redefining the definitions: (i) Koch and Klebe42 have used a very large modified Self-Organizing Map43,44 and (ii) George Rose’s group has defined 12 categories comprising different lengths. Nonetheless, these approaches were performed in a manner comparable to the secondary structure assignment that is still dominated by DSSP. Although different turn classifications have subsequently been proposed, none of them have been successfully used. The main idea in this study was not to redraw a novel classification but to extend the classical classification.

From an unsupervised classification, based exclusively on dihedral angles, four new types were defined. The two most frequently occurring, type IV1 and IV2, were similar to existing type II and VIII β-turns but had very distinct features. On the one hand, type IV1 and VIII β-turns were associated with stabilizing hydrogen bonds, unlike type VIII β-turns. On the other hand, type IV1 and II β-turns were very close in terms of dihedral angles but were distinct in terms of their amino acid content. Figure 5 clearly shows that type II β-turns were highly specific, whereas type IV1 β-turns had more classical characteristics, being closer to type I γ-turns but had minor differences. However, type IV2 β-turns were associated with stabilizing hydrogen bonds, unlike type VIII β-turns. On the other hand, type IV1 and II β-turns were highly similar in quality and sequence identity highlighted the high stability of the four main clusters (i.e., the new turns). For each simulation, the clusters were always found at similar frequencies and with similar dihedral values. However, the other clusters were substantially more variable. A simple analysis was also performed to evaluate the possibility of the presence of sub-clusters inside the different clusters by diminishing the authorized dihedral angle deviation.

### Table 6. Amino acid’s Z-score of β-turn types.

Colors underline the difference of new turns and the original type IVori (underlining new over- or under representation, in bold italics inversion of over- or under representation, see Methods section).

| β-turn  | i   | i+1 | i+2 | i+3 |
|---------|-----|-----|-----|-----|
| IV1     | (+) | cpG | PAN | whSNd |
| IV2     | (+) | iq | Pek | GN |
| IV3     | (+) | qP | Pe | GN |
| IV4     | (+) | st | G  | stN |
| IV5     | (+) | Vla | g  | ivlapsg |
| IV6*    | (+) | CPG | P  | ivlapstek |
| IV7     | (+) | Vc | g  | P |
| IV8     | (+) | Ve | g  | P |
| IV9     | (+) | Ve | g  | P |
| IV10    | (+) | Ve | g  | P |
| IV11    | (+) | Cpgn | P  | Ghtnd |
| IV12    | (+) | Cpgn | P  | Ghtnd |
| IV13    | (+) | Cpgn | P  | Ghtnd |

A classical question raised by any clustering methodology is the relevance of the results. Here, our results can be considered reliable, owing to their reproducibility and stability. The use of 10 different datasets ranging in quality and sequence identity highlighted the high stability of the four main clusters (i.e., the new turns). For each simulation, the clusters were always found at similar frequencies and with similar dihedral values. However, the other clusters were substantially more variable. A simple analysis was also performed to evaluate the possibility of the presence of sub-clusters inside the different clusters by diminishing the authorized dihedral angle deviation.
allowed during the training. Similarly, the centre of the four main clusters always appeared, thus supporting their stability.

Comparisons with the previous alternative classification proposed by Efimov and Thornton’s group emphasized the uniqueness of the approach. Notably, the most frequent new turn (type IV\(1\) β-turn) was not highlighted, although it is the 5th most occurring turn (including type IV\(\text{misc}\) β-turns). Only the type IV\(2\) β-turns were previously included.

This extended classification is relevant because it does not modify the currently accepted β-turn types, is highly stable (in regards to amino acid redundancy and the quality of protein resolution), and proposes new ways to analyse the architecture and dynamics of the protein or peptide structure of β-turns. Hence, we envision two potential applications of this classification system. The first one addresses molecular dynamics simulations in which researchers follow the dynamic evolution of type VIII β-turns. The change from type VIII to a type IV (i.e., IV\(1\)) during the simulations is very different when the turn is in fact a type IV\(2\) or IV\(\text{misc}\). The former case (type IV\(2\) β-turn) is a simple extension of this conformation, whereas the latter (type IV\(\text{misc}\) β-turn) is really a different independent conformation. The second example involves an analysis of conformational characteristics of asparaginyl residues in proteins. Interestingly, many are associated with turn conformations. With this new classification, only 16.5% (see Supplementary Information 14) were associated with miscellaneous turns (e.g., IV\(\text{misc}\)); thus, this classification provides a better description of local protein conformations and resolves the spectrum of IV\(\text{misc}\) turns to a greater extent.

An interesting point is that turns are often observed as tandem repeats, sometimes leading to long series of γ\(\beta\), β\(\gamma\), β\(\beta\) or γ\(\gamma\) turns. It is also notable that γ and β turns are associated with the same residues. In future work, we plan to investigate the succession of turns, particularly the ones mentioned in this study.

**References**

1. Pauling, L., Corey, R. B. & Branson, H. R. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci USA* **37**, 205–211 (1951).
2. Pauling, L. & Corey, R. B. The pleated sheet, a new layer configuration of polypeptide chains. *Proc Natl Acad Sci USA* **37**, 251–256 (1951).
3. Eisenberg, D. The discovery of the alpha-helix and beta-sheet, the principal structural features of proteins. *Proc Natl Acad Sci USA* **100**, 11207–11210 (2003).
4. Fourrier, L., Benros, C. & de Brevern, A. G. Use of a structural alphabet for analysis of short loops connecting repetitive structures. *BMC Bioinformatics* **5**, 58 (2004).
5. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).
6. Fodje, M. N. & Al-Karadaghi, S. Occurrence, conformational features and amino acid propensities for the pi-helix. *Protein Eng* **15**, 353–358 (2002).
7. Martin, J. et al. Protein secondary structure assignment revisited: a detailed analysis of different assignment methods. *BMC Structural Biology* **5**, 17 (2005).
8. Heinig, M. & Frishman, D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res* **32**, W500–502, doi: 10.1093/nar/gkh429 (2004).
9. Offmann, B., Tyagi, M. & de Brevern, A. G. Local Protein Structures. *Current Bioinformatics* **3**, 165–202 (2007).
10. Klose, D. P., Wallace, B. A. & Janes, R. W. 2Struc: the secondary structure server. *Bioinformatics* **26**, 2624–2625, doi: 10.1093/bioinformatics/btq480 (2010).
11. Calligari, P. A. & Kneller, G. R. ScrewFit: combining localization and description of protein secondary structure. *Acta Crystallogr D Biol Crystallogr* **68**, 1690–1693, doi: 10.1107/S0907444912039029 (2012).
12. Tyagi, M., Bornot, A., Offmann, B. & de Brevern, A. G. Analysis of loop boundaries using different local structure assignment methods. *Protein Sci* **18**, 1869–1881, doi: 10.1002/pro.198 (2009).
13. Krues, E., Thumfort, P., Tang, C. & Wingreen, N. S. Gibbs sampling and helix-cap motifs. *Nucleic Acids Res* **33**, 5343–5353, doi: 33/16/534366 (2005).
61. Tyagi, M., Bornot, A., Offmann, B. & de Brevern, A. G. Protein short loop prediction in terms of a structural alphabet.

63. Joseph, A. P.

58. Perskie, L. L. & Rose, G. D. Physical-chemical determinants of coil conformations in globular proteins. J Mol Biol 229, 472–493 (1993).

62. de Brevern, A. G., Etchebest, C. & Hazout, S. Bayesian probabilistic approach for predicting backbone structures in terms of protein structure units. J Mol Biol 315, 315–331 (2002).

56. Meissner, M., Koch, O., Klebe, G. & Schneider, G. Prediction of turn types in protein structure by machine-learning classifiers. Protein Sci 34, 434–447 (1995).

60. Wang, G. & Dunbrack, R. L. Jr. PISCES: a protein sequence culling server. Bioinformatics 17, 849–854 (2001).

50. Gordeev, A. B., Kargatov, A. M. & Efimov, A. V. PCBOST: Protein classification based on structural trees. Bioinformatics 21, 2100–2107 (2005).

49. Efimov, A. V. A structural tree for proteins containing 3beta-corners. Bioinformatics 21, 2100–2107 (2005).

28. Bornot, A. & de Brevern, A. G. Protein beta-turn assignments.

25. Chellgren, B. W. & Creamer, T. P. Short sequences of non-proline residues can adopt the polyproline II helical conformation. Protein Sci 3, 593–621 (1994).

19. Pauling, L. & Corey, R. B. The structure of fibrous proteins of the collagen-gelatin group. Proc Natl Acad Sci USA 37, 272–281 (1951).

20. Cowan, P. M., McGavin, S. & North, A. C. The polypeptide chain configuration of collagen. Nature 176, 1062–1064 (1955).

27. Fuchs, P. F. & Alix, A. J. High accuracy prediction of beta-turns and their types using propensities and multiple alignments. Proteins 59, 828–839 (2005).

58. Perskie, L. L. & Rose, G. D. Physical-chemical determinants of coil conformations in globular proteins. J Mol Biol 229, 472–493 (1993).

62. de Brevern, A. G., Etchebest, C. & Hazout, S. Bayesian probabilistic approach for predicting backbone structures in terms of protein structure units. J Mol Biol 315, 315–331 (2002).

56. Meissner, M., Koch, O., Klebe, G. & Schneider, G. Prediction of turn types in protein structure by machine-learning classifiers. Protein Sci 34, 434–447 (1995).

60. Wang, G. & Dunbrack, R. L. Jr. PISCES: a protein sequence culling server. Bioinformatics 17, 849–854 (2001).

50. Gordeev, A. B., Kargatov, A. M. & Efimov, A. V. PCBOST: Protein classification based on structural trees. Bioinformatics 21, 2100–2107 (2005).

49. Efimov, A. V. A structural tree for proteins containing 3beta-corners. Bioinformatics 21, 2100–2107 (2005).

28. Bornot, A. & de Brevern, A. G. Protein beta-turn assignments.

25. Chellgren, B. W. & Creamer, T. P. Short sequences of non-proline residues can adopt the polyproline II helical conformation. Protein Sci 3, 593–621 (1994).

19. Pauling, L. & Corey, R. B. The structure of fibrous proteins of the collagen-gelatin group. Proc Natl Acad Sci USA 37, 272–281 (1951).

20. Cowan, P. M., McGavin, S. & North, A. C. The polypeptide chain configuration of collagen. Nature 176, 1062–1064 (1955).

27. Fuchs, P. F. & Alix, A. J. High accuracy prediction of beta-turns and their types using propensities and multiple alignments. Proteins 59, 828–839 (2005).

58. Perskie, L. L. & Rose, G. D. Physical-chemical determinants of coil conformations in globular proteins. J Mol Biol 229, 472–493 (1993).

62. de Brevern, A. G., Etchebest, C. & Hazout, S. Bayesian probabilistic approach for predicting backbone structures in terms of protein structure units. J Mol Biol 315, 315–331 (2002).

56. Meissner, M., Koch, O., Klebe, G. & Schneider, G. Prediction of turn types in protein structure by machine-learning classifiers. Protein Sci 34, 434–447 (1995).

60. Wang, G. & Dunbrack, R. L. Jr. PISCES: a protein sequence culling server. Bioinformatics 17, 849–854 (2001).

50. Gordeev, A. B., Kargatov, A. M. & Efimov, A. V. PCBOST: Protein classification based on structural trees. Bioinformatics 21, 2100–2107 (2005).

49. Efimov, A. V. A structural tree for proteins containing 3beta-corners. Bioinformatics 21, 2100–2107 (2005).

28. Bornot, A. & de Brevern, A. G. Protein beta-turn assignments.

25. Chellgren, B. W. & Creamer, T. P. Short sequences of non-proline residues can adopt the polyproline II helical conformation. Protein Sci 3, 593–621 (1994).

19. Pauling, L. & Corey, R. B. The structure of fibrous proteins of the collagen-gelatin group. Proc Natl Acad Sci USA 37, 272–281 (1951).

20. Cowan, P. M., McGavin, S. & North, A. C. The polypeptide chain configuration of collagen. Nature 176, 1062–1064 (1955).

27. Fuchs, P. F. & Alix, A. J. High accuracy prediction of beta-turns and their types using propensities and multiple alignments. Proteins 59, 828–839 (2005).

58. Perskie, L. L. & Rose, G. D. Physical-chemical determinants of coil conformations in globular proteins. J Mol Biol 229, 472–493 (1993).

62. de Brevern, A. G., Etchebest, C. & Hazout, S. Bayesian probabilistic approach for predicting backbone structures in terms of protein structure units. J Mol Biol 315, 315–331 (2002).

56. Meissner, M., Koch, O., Klebe, G. & Schneider, G. Prediction of turn types in protein structure by machine-learning classifiers. Protein Sci 34, 434–447 (1995).

60. Wang, G. & Dunbrack, R. L. Jr. PISCES: a protein sequence culling server. Bioinformatics 17, 849–854 (2001).

50. Gordeev, A. B., Kargatov, A. M. & Efimov, A. V. PCBOST: Protein classification based on structural trees. Bioinformatics 21, 2100–2107 (2005).

49. Efimov, A. V. A structural tree for proteins containing 3beta-corners. Bioinformatics 21, 2100–2107 (2005).

28. Bornot, A. & de Brevern, A. G. Protein beta-turn assignments.
INTS (SESAME Grant).
by the programme "Investissements d'avenir" of the French National Research Agency, reference ANR-11-
by grants from the Laboratory of Excellence GR-Ex, reference ANR-11-LABX-0051. The labex GR-Ex is funded
University of Paris Diderot – Paris 7, French National Institute for Blood Transfusion (INTS), French Institute for
J.-C. Gelly, and especially T.J. Narwani. This work was supported by grants from the French Ministry of Research,
Venkatachalam, P. Balaram, N. Srinivasan and R. Sowdhamini and also to my colleagues C. Etchebest, P.F. J. Fuchs,
I would like to dedicate this research to Indian protein pioneers G.N. Ramachandran, C. Ramakrishnan, C.M.
95. Hayashi, I. & Ikura, M. Crystal structure of the amino-terminal microtubule-binding domain of end-binding protein 1 (EB1).
94. Schafer, K.
92. Zuo, Y., Wang, Y. & Malhotra, A. Crystal structure of Escherichia coli RNase D, an exoribonuclease involved in structured RNA
86. Becker, A. & Kabsch, W. X-ray structure of pyruvate formate-lyase in complex with pyruvate and CoA. How the enzyme uses the
85. de Sanctis, D.
88. Levy, C. W.
91. Hisano, T.
89. Grabarse, W.
83. Guruprasad, K., Prasad, M. S. & Kumar, G. R. Analysis of gammabeta, betagamma, gammagamma, betabeta multiple turns in
81. Srinivasan, N., Anuradha, V. S., Ramakrishnan, C., Sowdhamini, R. & Balaram, P. Conformational characteristics of asparaginyl
76. Sammon, J. A nonlinear mapping for data structure analysis.
75. Singh, H., Singh, S. & Raghava, G. P. In silico platform for predicting and initiating beta-turns in a protein at desired locations.
Proteins 83, 910–921, doi: 10.1002/prot.24783 (2015).
Sammon, J. A nonlinear mapping for data structure analysis. IEEE Transactions on Computers 18, 401–409. (1969).
Guruprasad, K. & Rajkumar, S. Beta-and gamma-turns in proteins revisited: a new set of amino acid turn-type dependent positional
preferences and potentials. J Bioess 25, 145–156 (2000).
87. Schuchhardt, J., Schneider, G., Reichelt, J., Schomburg, D. & Wrede, P. Local structural motifs of protein backbones are classified by
80. Ihaka, R. & Gentleman, R. R: A Language for Data Analysis and Graphics.
79. Singh, P. & Pramanik, S. C. Spectroscopic investigation of the gamma-turn and its application in the prediction of natural
44, 112–122 (1994).
71. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations.
70. Ihaka, R. & Gentleman, R. R: A Language for Data Analysis and Graphics.
69. Esque, J., Urbain, A., Etchebest, C. & de Brevern, A.G. Sequence-structure relationship study in all-alpha transmembrane proteins
47–50 (1996).
69. Schuchhardt, J., Schneider, G., Reichelt, J., Schomburg, D. & Wrede, P. Local structural motifs of protein backbones are classified by
self-organizing neural networks. Protein Eng 9, 833–842 (1996).
68. de Brevern, A. G. & Hazout, S. Hybrid protein model for optimally defining 3D protein structure fragments. Bioinformatics 19,
345–353 (2003).
67. Poulain, P. PBxplore: A program to explore protein structures with Protein Blocks. Technical report. (2016) Available at: https://
github.com/pierrepo/PBxplore. (Accessed: 21st June 2016).
66. Poulain, P. PBxplore: A program to explore protein structures with Protein Blocks. Technical report. (2016) Available at: https://
github.com/pierrepo/PBxplore. (Accessed: 21st June 2016).
65. Tyagi, M. et al. Protein Block Expert (PBE): a web-based protein structure analysis server using a structural alphabet. Nucleic Acids
Res 34, W119–123 (2006).
64. Deleser, J.-C., Deleers, G., Vermeiren, C., Balaram, P. & Schepens, J. Interpretation of gamma-turns using an unsupervised learning approach. Amino Acids 47, 2303–2322, doi: 10.1007/s00726-015-2010-5.1007/s00726-015-2010-5 (2015).
63. Jha, A. & Gentleman, R. R: A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics 5,
299–314 (1996).
62. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. J Mol Biol 7,
95–99 (1963).
61. Venkatachalam, P., Balaram, P., Srinivasan, N. & Sowdhamini, R. Conformational characteristics of asparaginyl residues in proteins.
Int J Pept Protein Res 44, 112–122 (1994).
60. Dobhek, H., Svrčítkych, V., Liss, J. & Meyer, O. Carbon monoxide induced decomposition of the active site [Ni-4Fe-5S] cluster of
doxygenase. J Am Chem Soc 126, 5382–5387, doi: 10.1021/ja047776v (2004).
59. Levy, C. W. et al. Insights into enzyme evolution revealed by the structure of methylaspartate ammonia lyase. Structure 10, 105–113
(2002).
58. Burmeister, W. P., Guilligay, D., Cusack, S., Wedell, G. & Armborg, N. Crystal structure of species D adenosine fiber knobs and their
sialic acid binding sites. J Virol 78, 7727–7736, doi: 10.1128/JVI.78.14.7727-7736.2004 (2004).
57. Grabarse, W. et al. On the mechanism of biological methane formation: structural evidence for conformational changes in methyl-
coenzyme M reductase upon substrate binding. J Mol Biol 309, 315–330, doi: 10.1006/jmbi.2001.4647 (2001).
56. Hisano, T. et al. Crystal structure of the (R)-specific enoyl-CoA hydratase from Aeromonas caivia involved in polyhydroxylalkanoate
biosynthesis. J Biol Chem 278, 617–624, doi: 10.1074/jbc.M205484200 (2003).
55. Zuo, Y., Wang, Y. & Malhotra, A. Crystal structure of Escherichia coli RlNase D, an oxoribonuclease involved in structured RNA
processing. Structure 13, 973–984, doi: 10.1016/S0969-2126(04)00415-5 (2005).
54. Kwak, B. Y. et al. Structure and mechanism of CTP-phosphocholine cytidylyltransferase (LicC) from Streptococcus pneumoniae. J
Biol Chem 277, 4343–4350, doi: 10.1074/jbc.M109163200 (2002).
53. Schaefer, K. et al. X-ray structures of the maltose-maldextrin-binding protein of the thermostabilic bacterial Amyllobacillus
caditerrariae provide insight into acid stability of proteins. J Mol Biol 335, 261–274 (2004).
52. Hayashi, I. & Iwata, M. Crystal structure of the amino-terminal microtubule-binding domain of end-binding protein 1 (EB1). J Biol
Chem 278, 36430–36434, doi: 10.1074/jbc.M305773200 (2003).
51. Wise, E. L., Graham, D. E., White, R. H. & Rayment, I. The structural determination of phosphosulfatase synthase from
Methanococcus jannaschii at 1.7–A resolution: an enolate that is not an enolate. J Biol Chem 278, 45858–45863, doi: 10.1074/jbc.
M304786200 (2003).
Acknowledgements
I thank the editor and anonymous reviewers for their constructive comments, which helped me improve the
manuscript. This work came from various trips and discussions I had during recent years in Bangalore, India, and
I would like to dedicate this research to Indian protein pioneers G.N. Ramachandran, C. Ramakrishnan, C.M.
Venkatachalam, P. Balaram, N. Srinivasan and R. Sowdhamini and also to my colleagues C. Etchebest, P.F.J. Fuchs,
J.-C. Gelly, and especially T.J. Narwani. This work was supported by grants from the French Ministry of Research,
University of Paris Diderot – Paris 7, French National Institute for Blood Transfusion (INTS), French Institute for
Health and Medical Research (INSERM). Adb also acknowledges the Indo-French Centre for the Promotion of
Advanced Research/CEFIPRA for collaborative grants (numbers 3903-E and 5302-2). This study was supported by
grants from the Laboratory of Excellence GR-Ex, reference ANR-11-LABX-0051. The labex GR-Ex is funded by
the programme "Investissements d’avenir" of the French National Research Agency, reference ANR-11-
IDEX-0005-02. Calculations were performed on an SGI cluster granted by Conseil Régional Ile de France and
INTS (SESAME Grant).
Author Contributions
A.G.d.B. designed and performed experiments, analysed data and wrote the paper.

Additional Information
Supplementary information accompanies this paper at http://www.nature.com/srep
Competing financial interests: The author declares no competing financial interests.
How to cite this article: de Brevern, A. G. Extension of the classical classification of β-turns. Sci. Rep. 6, 33191; doi: 10.1038/srep33191 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2016