ISSN 0006-3509, Biophysics, 2023, Vol. 68, No. 1, pp. 1–5. © Pleiades Publishing, Inc., 2023.
Russian Text © The Author(s), 2023, published in Biofizika, 2023, Vol. 68, No. 1, pp. 5–10.

MOLECULAR BIOPHYSICS

Conformational Aspects in the Formation of Structures of the Backbone of Polypeptide Chains in Proteins: the Relationship between Conformational Stability/Lability and β-Turns

I. Yu. Torshin[a,*], I. V. Filatov[b], A. V. Batyanovskii[c], K. V. Smirnov[d], A. A. Anashkina[d], N. G. Esipova[d], and V. G. Tumanyan[d]

[a] *Federal Research Center "Computer Science and Control," Moscow, 119333 Russia*
[b] *Moscow Institute of Physics and Technology, Dolgoprudny, Moscow oblast, 141700 Russia*
[c] *Institute of Biophysics and Cell Engineering, National Academy of Sciences of Belarus, Minsk, 220072 Belarus*
[d] *Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, 119991 Russia*
**e-mail: tiy135@yahoo.com*

Received November 14, 2022; revised December 11, 2022; accepted December 14, 2022

Abstract—To assess the nature of the relationship between the integral conformational stability of tetrapeptides and the main types of β-turns (which are also tetrapeptides), spectrum diagrams, the asymmetry of the distribution of conformationally stable and unstable tetrapeptides have been calculated. It has been shown that β-turns of types I', II, and II' consist mainly of conformationally labile peptides; this is consistent with the context-predetermined nature of their structure. Since, as we have shown earlier, in this case the conformation is imposed by external conditions (specifically, the closure of the cycle), the prevalence of conformation-labile peptides facilitates the formation of the structure due to external factors. The type I β-turn is an exception, since peptides with different conformational lability are distributed fairly even in it. It can be assumed that the formation of the type I β-turn is not contextually determined.

Keywords: conformational stable/labile segments of protein, local protein structure, conformational analysis, β-turns, statistical analysis, molecular dynamics

DOI: 10.1134/S0006350923010177

INTRODUCTION

According to general ideas, β-turns are considered as a typical structural element of a protein that is of interest both in itself and as an integral part of a certain level of the hierarchy of the protein structure. This hierarchy, generally speaking, cannot but reflect the process of protein folding into a globule and globule complexes. The significance of the problem of the β-turns is emphasized by the fact that in some proteins their proportion reaches 80%, that is, they occur more often than α-spirals and β-strands. According to the estimate made in [1], every fourth residue is part of a beta-turn β-turn. Often, the β-turns are components of a more complex structural element, primarily the so-called β-hairpins [2, 3].

The β-turn is characterized by a well-defined local structure (in other words, a certain type of secondary structure) of a certain size. Unlike the α-helix, the β-structure, or the PPII left helix (the study of which raises the question of their length and the criteria used to determine the length), the length of the β-turn is strictly four amino-acid residues. It is also important that the β-turn is a small structural motif, whose conformation is determined by a small number of degrees of freedom, which makes it possible to use iterative algorithms. Among the structural functions of the β-turn, the most obvious one is associated with a change in the course of the polypeptide chain to the reverse, which is reflected in the alternative name of the β-turn, "reverse turn." This kind of change in the course of the polypeptide chain contributes to its compactification and the formation of a protein globule.

Thus, the proportion of the β-turns in proteins is large and their structural and functional significance is quite obvious. However, there are still notable gaps and contradictions in the existing ideas about the β-turns. In particular, the annotation of these types of local structure and the methods of their prediction based on the amino-acid sequence need further improvement. In fact, although there has been remarkable progress in the prediction of the β-turns, it was not associated with deepening our understanding of this type of structure, but with the use of more effective methods of data mining and so-called machine learning (the methods of supporting vectors, convolutional neural networks, and others [4, 5]). The

further increase in the quality of predicting is limited not only by the difficulties in the initial annotation of the β-turns, but also because two different tasks actually have to be solved: finding the place of the β-turn in the polypeptide chain and determining the type of turn. It should be added that a purely bioinformatic approach, in which the similarity of amino-acid sequences is investigated, has a limited scope of application. Further deepening of the understanding of the problem of the β-turns should be sought by analyzing the physicochemical properties of these local structures.

One of the most important issues both for the whole protein and for the β-turns, is the search for relationships between the structure of the turns, the amino-acid sequences of the actual β-turns ("text") and the sequences of the flanking regions that play the role of a "context." In this study, the original problem of the relationship between the main structural features of a particular type of the β-turn and the conformational mobility/lability of its constituent segments has been posed and received a feasible solution. Such a relationship can make it possible to predict a β-turn by assessing the conformational stability/lability of its segments.

ABOUT THE CONTEXT OF β-TURNS IN PROTEIN STRUCTURES

The problems of the relationship between the structure and sequence of amino acids, the stabilization of the structure, the role of "text" and "context" in the case of the β-turns are reduced to solving specific optimization tasks according to geometric and energy criteria. It should be emphasized that the formation of all types of β-turns (and not only turns of type II') is associated with overcoming stereochemical obstacles. The presence of such stereochemical difficulties becomes obvious, in particular, from the consideration of Ramachandran maps for β-turns: the points on the maps for residues $i + 1$ and/or $i + 2$ in the turns are concentrated in the forbidden regions of the Ramachandran map (see, for example, [6]). A more or less "free" conformational map is characterized only by a type I β-turn. The question arises: how does the β-turn acquire a conformation that appears energetically "tense"? The second question is: since an energetically "stressed" structure has already been formed, how is the "unfavorable" energy inherent in this structure compensated?

To answer the first question, our team conducted an investigation by involving geometric representations of a decrease in the number of independent conformational parameters in the cycle formed by an oligopeptide [7]. Due to the closure of the cycle (through the hydrogen bond between the first and last residues of the β-turn), the number of independent conformational parameters decreases so much that the determination of conformation can be reduced to a purely geometric calculation. Such a geometric analysis was carried out for the main types of β-turns using independent methods, the original distance geometry procedure, and the conformation enumeration procedure [8]. It was demonstrated that the number of geometrically possible conformations in this case is reduced to two: the first of the solutions coincides with the data of X-ray diffraction analysis, and the second, although geometrically correct, falls on absolutely forbidden regions of the Ramachandran map (that is, corresponds to extremely pronounced steric difficulties). Thus, due to the closure of the hydrogen bond, a kind of "topological lock" is formed and this context factor leads to the formation of the β-turn conformation.

As for the answer to the second question, it is obvious that even with a fixed geometry of the β-turn, the energy of the corresponding local structure cannot be too "unfavorable," otherwise a cardinal restructuring will occur and a significantly different structure will arise. The presence of conformationally unstable segments would allow the unfavorable energy to "dissipate" over many degrees of freedom, and thus the "extra" stress would be removed. The inclusion of conformationally labile segments increases the number of degrees of freedom and the cycle effect is not observed. Thus, the segments with forced (imposed) conformation, taken by themselves, *a priori* would have to be conformationally labile.

THE CONFORMATIONAL STABILITY OF β-TURNS AND CONFIGURATIONS OF TETRAPEPTIDES

In connection with the above, we came to the need to evaluate the conformational stability of tetrapeptides forming the β-turns using the estimates of conformational lability/stability of tetrapeptides that we have developed earlier [9]. We recall that there is interest in the β-turns as special fragments of the secondary structure since, on the one hand, the conformation of the β-turn can be imposed by the context in the protein structure. On the other hand, the available algorithms for predicting the β-turns make it possible to successfully carry out such a prediction (with an accuracy of recognition of 75–82%) when considering a fairly small context in sequences around the β-turns (5–7 residues). Therefore, it is of interest to study exactly how the presence of the structure of the β-turns correlates with the conformational stability/lability of their constituent tetrapeptides ("texts") and flanking regions ("contexts") and whether there is such a relationship at all.

As shown in our previous study on conformational stability/lability measures s_1–s_5 [9], 122 100 different tetrapeptides were found in files from the PDB data bank, which were found at least five times in various structures. With a rarer occurrence of peptides (less than five cases across the entire PDB data bank), the use of the developed conformational lability measures

Table 1. Frequency of occurrence (in %) of various amino acids in conformationally labile motifs

Selection	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
Conformationally labile for types I', II, and II'	3.1	3.4	6.9	3.8	3.6	24.5	4.6	2.5	3.9	2.5	3.3	6.2	4.9	4.1	3.8	5.3	4.2	2.5	3.1	4.0
Average of the sample from the PDB	5.9	2.8	5.6	5.7	5.0	5.9	4.1	5.5	5.4	6.0	3.3	5.1	5.2	4.9	5.4	5.6	5.5	5.8	2.8	4.8

s_1–s_5 based on the collection of conformation statistics is not fully justified statistically.

The presence in the PDB of only 122100 out of ~160000 combinatorially possible tetrapeptides indicates, most likely, the peculiarities of the amino-acid sequences of crystallized proteins, which make up 90% of the data volume in the PDB (the remaining 10% of the structures were determined using NMR or electron microscopy). In confirmation of this, it can be noted that the frequencies of the same tri- and tetrapeptides differ significantly in the PDB and UNIPROT databases [10].

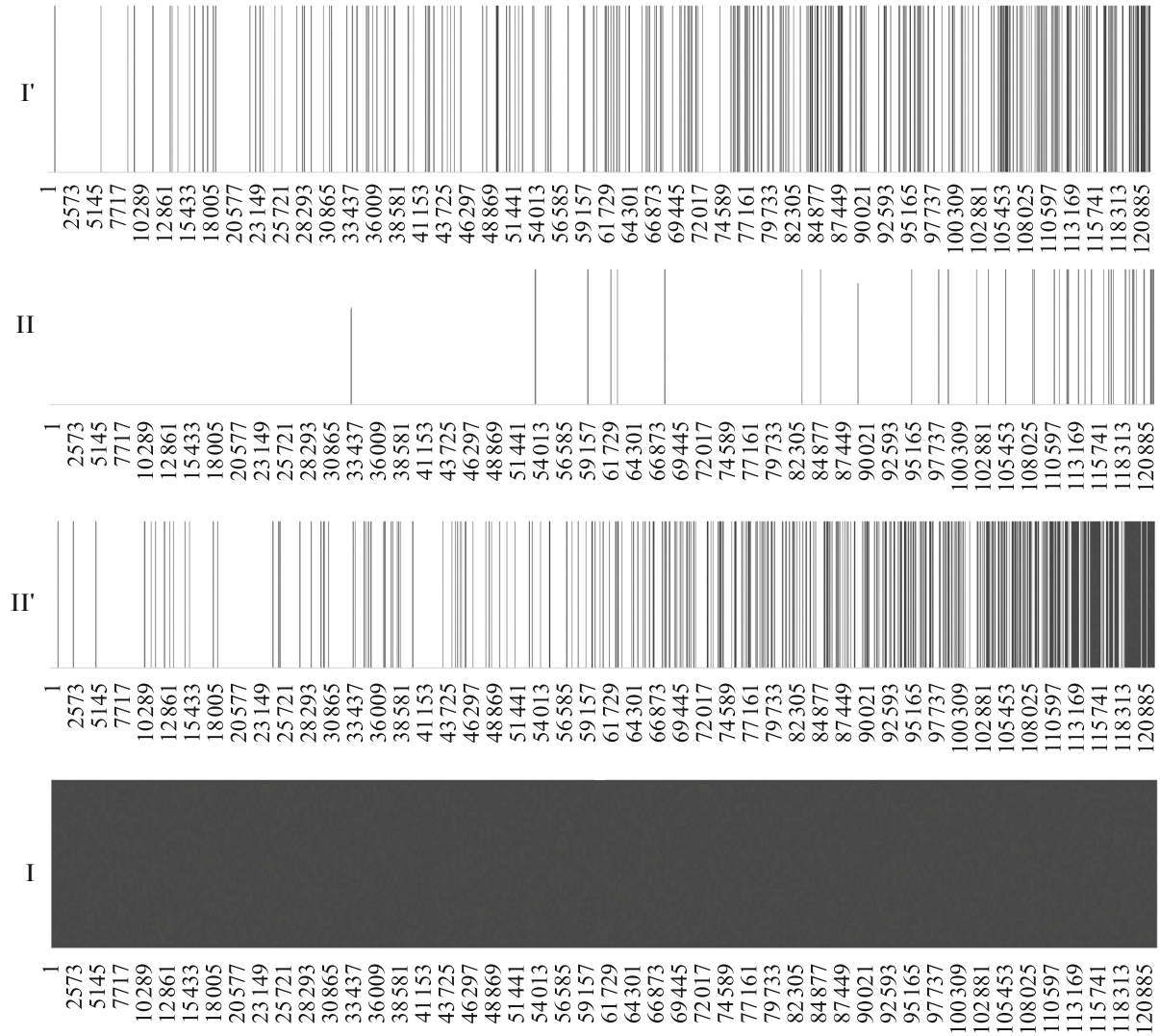

Fig. 1. Spectrum diagrams of the b_i values of the β-turns. The tetrapeptide stability rank is along the abscissa axis was calculated as the ordinal number of the b_i value in the list of tetrapeptides ordered in descending order of b_i values. A higher frequency of occurrence (higher line density) of conformationally labile peptides for three types of the turns (I', II, and II') is obvious. The spectrum diagram for the type I turn is obviously a uniform distribution of lines over all values of the ranks.

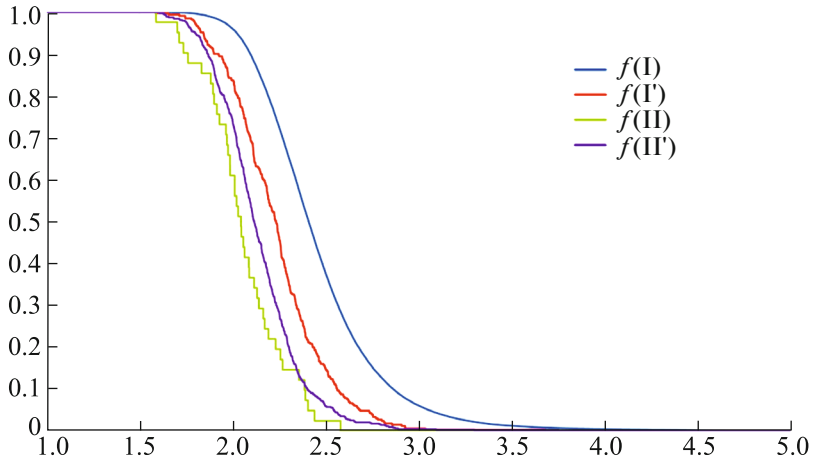

Fig. 2. Empirical distribution functions of the values of the integral assessment of conformational stability b_i.

Conformational lability measures s_1–s_5 were calculated for each of the tetrapeptides found in the PDB files [9]. Next, for each ith tetrapeptide, an integral estimate of the conformational stability b_i was calculated using the formula $b_i = 5 - \sum_{k=1,5} \omega_k s_k(i)$, where the weights of ω_k for each of the measures of conformational lability s_k were determined so that the value b_i of an arbitrary tetrapeptide lay in the range of values from 1 to 5 (1 is the minimum conformational stability, 5 is the maximum conformational stability).

In the general list of 122100 tetrapeptides, tetrapeptides were identified, for which for the second and third residues the predominant conformation (that is, the conformation that occurred more often than any other) was the β-turn of one of the four types (I, I', II, II'). The definitions of β-turns corresponded to the criteria from [6]. The following are the results of the analysis of tetrapeptides, the predominant conformation of which (more than 50% of the conformations found in the PDB for this peptide) was one of the four types of the turns (I, I', II, I, and I'). Only 64769 such tetrapetids were found.

Among these tetrapeptides with a predominant conformation in the form of one of the types of β-turns, peptides with the type I β-turn were most often found, corresponding (in most cases) to α-helices (in a broad sense, including 3/10 helices) or their "caping" fragments (63770 tetrapeptides). Tetrapeptides for which the most common conformation is the β-turn of the other three types, were much less common: 331 tetrapeptides were of type I', 41 tetrapeptides were of type II, and 627 tetrapeptides were of type II'. It is noteworthy that a greater number of different amino-acid sequences of tetrapeptides corresponds to the turn II', the most stressed from a stereochemical point of view.

It should be emphasized that the above data are new, unique, and have no world analogues, since they relate not to the frequency of occurrence of conformations (which is usually presented in the available literature), but to tetrapeptides (segments of four amino acids in length with a certain sequence), whose conformations are mainly associated with the above types of the turns; that is, these data are related to the "text" itself, and not to the "context" of the β-turn.

To analyze the relationship between the conformational stability of tetrapeptides and the β-turns of four types (I, I', II, and II'), the following approaches were used: (1) spectrum diagrams of b_i values; (2) empirical distribution functions of values of the integral assessment of conformational stability b_i; and (3) analysis of the distribution of tetrapeptides, mainly corresponding to the β-turns between the upper and lower parts of the total list of 122100 tetrapeptides, ordered in descending order of b_i values. All three independent methods of analysis showed similar results.

Spectrum diagrams make it possible to visually assess the differences in the frequency of occurrence of peptides of a given type (type of β-turn) at different values of b_i. The abscissa axis of the spectrum diagram shows the stability rank of the tetrapeptide calculated as the ordinal number of the b_i value in the list of tetrapeptides ordered in descending order of b_i values. Visual analysis of spectrum diagrams of b_i values showed that conformationally labile peptides are much more common for three types of the turns, namely, I', II, and especially II' (Fig. 1).

We consider statistical estimates of the distribution of peptides with a predominant conformation in the form of β-turns according to the b_i values. The analysis of conformationally labile peptides, which were much more common for three types of the turns (I', II, and II'), showed that these peptides were characterized by a pronounced excess of glycine and a reduced occurrence of alanine, glutamate, phenylalanine, isoleucine, lysine, leucine, arginine, and valine (Table 1). Compared with the occurrence of amino acids on

average in the PDB data bank, only the differences for glycine, alanine, isoleucine, leucine, and valine were reliable according to the χ^2 criterion. Conformationally stable peptides in this subsample did not show significant differences in amino-acid composition.

The analysis of the empirical distribution functions of the b_i values (Fig. 2) showed obvious shifts in the curves of the empirical distribution functions of the β-turns of types I', II, and II' towards lower b_i values, that is, towards more conformationally labile peptides. For example, the average b_i value for the type I turn was 2.4 ± 0.2 (mean ± standard deviation), and for the turns of types I', II', and II it was 2.22 ± 0.2, 2.1 ± 0.15, and 2.05 ± 0.12 respectively.

Analysis of the distribution of tetrapeptides corresponding to the β-turns between the upper and lower parts of the total list of 122 100 tetrapeptides, ordered in descending order of b_i values, confirmed the results described above. For the type I turns, the frequency of occurrence in the upper and lower halves of the list was almost the same (32955 peptides in the upper part and 30785 in the lower, that is, the frequencies were 0.548 and 0.512, respectively). In the case of three other types of the β-turns, statistically significant differences in the frequencies of occurrence with a predominance of conformationally labile peptides were noted: for type I' there were 74 and 256 tetrapeptides, that is, the frequencies differed by more than three times (0.0012 and 0.0043); for type II there were 3 and 37 tetrapeptides, the frequencies differed by an order of magnitude (5×10^{-5} and 6×10^{-4}); and for type II' there were 65 and 555 tetrapeptides, the frequencies differed by almost an order of magnitude (0.0011 and 0.0092). These differences for the β-turns of types I', II, and II' were significant according to the χ^2 criterion ($P < 0.008$ in all three cases).

Thus, the results of the analysis showed that among the β-turns of types I', II, and II' conformationally labile peptides are significantly more common. It is obvious that the proposed indicators of conformational lability of peptides can be used as additional independent information in computational schemes for the recognition of the β-turns of these types by amino-acid sequences of proteins. The data we obtained indicate a fundamental difference in the structure formation of the type I and other types of the β-turns. Unlike the other three types of the β-turns, the turns of the I type, apparently, were the least context-dependent.

CONCLUSIONS

The main result of this study is the finding of the unique properties of the β-turns of types I', II, and II' in the aspect of the amino-acid "text" of these local protein structures. In contrast to them, the type I β-turn does not exhibit pronounced peculiarities of the amino-acid composition, including in the sense of enrichment with conformation-labile sequences. The increased conformational lability of tetrapeptides forming the β-turns of I', II, and II' types contributes to the mitigation of stereochemical difficulties characteristic of the conformations of these local structures observed in the experiment.

FUNDING

The work was carried out with the financial support of the Russian Foundation for Basic Research, project no. 20-04-01085.

COMPLIANCE WITH ETHICAL STANDARDS

The authors declare that they have no conflicts of interest. This article does not contain any studies involving animals or human participants performed by any of the authors.

REFERENCES

1. W. Kabsch and C. Sander, Biopolymers **22** (12), 2577 (1983).
2. X. de la Cruz, E. G. Hutchinson, A. Shepherd, and J. M. Thornton, Proc. Natl. Acad. Sci. U. S. A. **99** (17), 11157 (2002).
3. M. Kumar, M. Bhasin, N. K. Natt, and G. P. S. Raghava, Nucleic Acids Res. **33**, W154 (2005).
4. M. K. Elbashir, J. Wang, F. X. Wu, et al., Proteome Sci. **11** (Suppl. 1), S5 (2013).
5. C. Fang, Y. Shang and D. Xu, Proteins **88** (1), 143 (2020).
6. A. G. de Brevern, Sci. Rep. **6**, 33191 (2016).
7. L. A. Uroshlev, I. Yu. Torshin, A. V. Batyanovsky, et al., Biophysics (Moscow) **60** (1), 1 (2015).
8. L. A. Uroshlev, I. Yu. Torshin, A. V. Batyanovsky, et al., Biophysics (Moscow) **64** (2), 195 (2019).
9. I. Yu. Torshin, A. V. Batyanovsky, L. A. Uroshlev, et al., Biophysics (Moscow) **64** (2), 182 (2019).
10. Yu. I. Zhuravlev, K. V. Rudakov, and I. Yu. Torshin, Tr. Fiz.-Tekh. Inst. **3** (4), 45 (2011).

Translated by E. Puchkov