HAL
open science

Protein beta-turn assignments.

Aurélie Bornot, Alexandre de Brevern

To cite this version:

Aurélie Bornot, Alexandre de Brevern. Protein beta-turn assignments.. Bioinformation, 2006, 1 (5), pp.153-5. inserm-00133658

HAL Id: inserm-00133658

https://inserm.hal.science/inserm-00133658v1

Submitted on 27 Feb 2007

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.

Bioinformation by Biomedical Informatics Publishing Group *open access*

www.bioinformation.net **Views & Challenges**

Protein beta-turn assignments

Aurélie Bornot & Alexandre G. de Brevern*

[1] Equipe de Bioinformatique Génomique et Moléculaire (EBGM), INSERM U726, Université Denis DIDEROT - Paris 7, case 7113, 2, place Jussieu, 75251 Paris, France, * Corresponding author: mailing address: Dr. de Brevern A.G., Equipe de Bioinformatique Génomique et Moléculaire (EBGM), INSERM U726, Université Denis DIDEROT - Paris 7, case 7113, 2, place Jussieu, 75251 Paris, France, E-mail : debrevern@ebgm.jussieu.fr, Tel: (33) 1 44 27 77 31, Fax: (33) 1 43 26 38 30

(Received March 27, 2006; Revised May 24, 2006; Accepted May 24, 2006; published online June 08, 2006)

Abstract:

A classical way to analyze protein 3D structures or models is to investigate their secondary structures. Their predictions are also widely used as a help to build new 3D models. Thus, hundreds of prediction methods have been proposed. Nonetheless before predicting, secondary structure assignment is required even if not trivial. Therefore numerous but diverging assignment methods have been developed. β-turns constitute the third most important secondary structures. However, no analysis to compare the β-turn distributions according to different secondary structure assignment methods has ever been done. We propose in this paper to analyze and evaluate the results of such a comparison. We highlight some important divergence that could have important consequence for the analysis and prediction of β-turns.

Background:

The description of protein structures in terms of secondary structures is widely used for analysis or prediction purposes. The secondary structures are classically described as composed of two repetitive states, the α-helix **[1]** and the β-sheet **[2]** states. All residues not associated to these states are assigned to the coil state, an undefined state. Numerous research teams have developed their own secondary structure assignment methods (SSAMs) using different criteria to describe the repetitive structures. DSSP remains the most widely-used program for secondary structure assignment. It is based on the detection of hydrogen-bonds defined by an electrostatic criterion. Secondary structure elements are then assigned according to characteristic hydrogen-bond patterns **[3]**. STRIDE is directly related to DSSP as it also uses hydrogen-bond patterns, even if their definitions are slightly different **[4]**. In addition, STRIDE takes into account (Φ/Ψ) angles to assign secondary structures. SECSTR belongs to the same family of methods **[5]**. XTLSSTR uses distances and angles calculated from the backbone geometry an is concerned with amide-amide interactions **[6]**. PSEA only considers Cα atoms. It is based on distance and angle criteria **[7]**. DEFINE relies on Cα coordinates only and compares Cα distances with distances in idealized secondary structure segments **[8]**. KAKSI is a recent approach based on distance between Cα atoms and dihedral angles **[9]**. SEGNO uses also the Φ and Ψ dihedral angles coupled with other angles to assign the secondary structures **[10]**. Nonetheless, only half of the residues are concerned with α-helices and β-strands. So, a more precise description of protein structures requires assignment of other local protein structures. β-turns are the most interesting local protein structures alongside the α-helices and the β-strands. They are constituted of 4 consecutive residues with a distance between Cα of first and fourth residues that has to be smaller than 7 Å. This restrictive distance implies a particular geometry to the backbone which turns back on itself **[11]**. As they orient α-helices and β-strands, they play a major role for the final protein topology. As an additional requirement, the central residues have to be non-helical in order to distinguish them from α-helices **[12]**. A serious issue raised by the variety of SSAMs is that they often yield diverging results about the extents of repetitive secondary structures **[9, 13, 14]**. Numerous analyses and prediction methods have been performed on the β-turns, but none comparison of β-turns assignment has been performed. In the present paper, we analyze the distribution of β-turns assignment according to different SSAMs.

Description:

Classically, the comparison of SSAMs only focuses on α-helix, β-strand and coil states **[9, 13, 14]**. Here, we have added the assignment of β-turns and compared their corresponding distribution. A high quality non-redundant set of 887 protein structures was selected from the PDB database according to the following criteria: X-ray structures with 1.6 Å or better resolution, and, no more than 20% pair wise sequence identity. In a first step, the secondary structure assignment was done with DSSP **[3]**, STRIDE **[4]**, SECSTR **[5]**, XTLSSTR **[6]**, and PSEA **[7]**. DEFINE **[8]**, KAKSI **[9]** and SEGNO **[10]** methods. Some methods assigned other states, e.g. turn using distance or hydrogen bond criteria between residues *i* and *i*+3, bend using kappa angle between residues *i*-2, *i* and *i*+2, polyproline II which is a helix with dihedral angle values in β-sheet region of Ramachandran map or β-bridge, single pair beta-sheet hydrogen bond formation. So, the description was reduced as follows: α corresponds to α-, 3_{10}- and π-helix, β corresponds to β-sheet and β-strand, and, *coil* encompasses all the rest. In a second step, β-turns were assigned following classical rules **[12]**, i.e. distance between residues *i* and *i*+3 less than 7 Å and the central residues of turns must be non-helical. Table 1 summarizes all the results of this analysis.

Repetitive structures corresponding to ~ 60% of the residues for all the SSAMs ranging between 58.05% and 61.51% (cf. Table 1, col. 2 and 3). DEFINE is clearly different with a mean value of repetitive structures equals to 54.24%. It corroborates previous observations that have already noted a high divergence of this method with every other SSAMS **[14]**. SSAMS that assigns less α-helix (mean value = 36.74%) assigns more β-strand (mean value = 22.25%). Analysis of coil frequency gives two major clusters above and below 17%. Coil frequencies are higher for STRIDE, DSSP, SEGNO, SECSTR, and XTLSSTR (between 19 and 21%, *i.e.*), while they are clearly lower for KAKSI, PSEA and DEFINE (~15%). DSSP and STRIDE turns frequencies (in parenthesis in Table 1, col. 5) are very close to the ones we determined with applications of classical rules.

Bioinformation, an open access forum
© Bornot & de Brevern, 2006, under the open access policy
The author(s) are responsible for the information published in this forum

Table 1: Distribution of secondary structure states (left) and confusion matrix for turn states assignments (right).

	α	β	coil[a]	turns[b,c]	DSSP	STRIDE	XTLSSTR	PSEA	DEFINE	SECSTR	KAKSI	SEGNO
DSSP	37.42	21.61	19.78	21.19[b] (20.53)[c]	--	89.03	76.39	85.48	59.30	84.28	74.70	87.55
STRIDE	38.88	22.16	19.06	19.90 (20.24)	94.49	--	79.33	88.53	59.78	85.12	79.07	91.22
XTLSSTR	41.04	19.57	19.57	19.82 (11.39)	81.32	79.55	--	83.76	58.46	74.50	77.22	85.53
PSEA	34.04	24.01	15.70	26.25	69.27	67.53	63.81	--	59.75	65.77	73.26	76.16
DEFINE	28.36	25.92	14.76	30.96	40.80	39.87	38.71	50.52	--	38.56	46.38	44.51
SECSTR	38.74	20.33	21.24	19.69	90.73	86.20	75.17	87.50	60.11	--	77.14	87.03
KAKSI	39.49	22.02	15.53	22.97	71.78	72.40	68.53	88.12	59.40	67.68	--	77.40
SEGNO	35.94	22.41	19.62	22.02	84.11	82.77	77.30	89.83	59.29	77.64	76.74	--

[a] coil state frequency corresponds to residues not associated to α-helix, β-strand or turns. [b] turn state frequency corresponds to residues assigned as β-turn and not associated to α-helix or β-strand (our assignment). [c] number in parenthesis are the frequency of turns originally given by the corresponding methods (original assignment method). For DSSP, it corresponds to turn and bent state.

Bioinformation, an open access forum
© Bornot & de Brevern, 2006, under the open access policy
The author(s) are responsible for the information published in this forum

For XTLSSTR, it is very different (+8.43%). Analysis of turn frequencies gives two major clusters. The first ones are associated to a frequency of turn residues near 20% (STRIDE, DSSP, SEGNO, SECSTR, XTLSSTR and KAKSI), the second ones are associated to a higher frequency (> 25%%, *i.e.* PSEA and DEFINE). This higher frequency is at the detriment of α-helix assignment. We compute a confusion matrix of β-turn assignment between each pairs of methods. It is defined as the number of times a residue assigned as turn by a $SSAM_i$ is also assigned in turn by $SSAM_j$. The turn confusion matrix (Table 1, right part) is asymmetric as the turn frequencies are different for each SSAMs. Table 1 highlights again the particularity of DEFINE assignment, its confusion values are near 50%. For all the other SSAMs, the values range from 63.81% to 94.49%.

The SSAMs based on hydrogen bonds (DSSP, STRIDE and SECSTR) show similar tendencies with confusion rates ranging from 84.28% to 94.49%. Even if the definitions are different for SEGNO, its results are quite similar to SSAMs based on hydrogen bonds (confusion values superior to 80%). The others SSAMs are clearly distinct. For instance, KAKSI turns is similar to PSEA turns with a high confusion value of 88.12% while it drops to ~70% for the SSAMs based on hydrogen bonds and only 68.53% for XTLSSTR.

Conclusion:

This analysis shows that β-turn frequencies are as stable as other repetitive secondary structures depending on the used SSAM. For residues non-assigned in repetitive structures, 20% are in β-turn. The use of β-turn is so quite interesting because less than 1/5 of the amino are left associated to a non-defined state. Nonetheless, this study shows also that using different SSAMs can give very different β-turn assignments. In fact, these divergences are directly related to the strong discrepancies in assignment of helix and sheet ends, as the turn assignments are performed in a second step. This problem can greatly influence sequence – structure analysis of β-turns and also could have repercussion on prediction methods (e.g. **[15]**).

In future work, we would like to study thoroughly the different beta-turn types between different SSAMs, examine the local environment of misassignments and consequences on the sequence–structure relationships.

References:

[1] L. Pauling & R.B. Corey. *Proc Natl Acad Sci U S A* 37: 235-240 (1951). [PMID: 14834145]
[2] L. Pauling & R.B. Corey. *Proc Natl Acad Sci U S A* 37: 251-256 (1951). [PMID: 14834147]
[3] W. Kabsch & C. Sander. *Biopolymers* 22: 2577-2637 (1983). [PMID: 6667333]
[4] D. Frishman & P. Argos. *Proteins* 23: 566-579 (1995). [PMID: 8749853]
[5] M.N. Fodje & S. Al-Karadaghi. *Protein Eng* 15: 353-358 (2002). [PMID: 12034854]
[6] S. M. King & W.C. Johnson. *Proteins* 35: 313-320 (1999). [PMID: 10328266]
[7] G. Labesse *et al*., *Comput Appl Biosci* 13: 291-295 (1997). [PMID: 9183534]
[8] F.M. Richards & C.E. Kundrot. *Proteins* 3: 71-84 (1988). [PMID: 3399495]
[9] J. Martin *et al*., *BMC Structural Biology* 5: 17 (2005). [PMID: 16164759]
[10] M.V. Cubellis *et al*., *BMC Bioinformatics* 6 Suppl 4: S8 (2005). [PMID: 16351757]
[11] C.M. Venkatachalam. *Biopolymers* 6: 1425-1436 (1968). [PMID: 5685102]
[12] E.G. Hutchinson & J.M. Thornton. *Protein Sci* 3: 2207-2216 (1994). [PMID: 7756980]
[13] N. Colloc'h *et al*., *Protein Eng* 6: 377-382 (1993). [PMID: 8332595]
[14] L. Fourrier *et a*l., *BMC Bioinformatics* 5: 58 (2004). [PMID: 15140270]
[15] P.F. Fuchs & A.J. Alix. *Proteins* 59: 828-39 (2005). [PMID: 15822097]

Edited by B. Offmann
Citation: Bornot & de Brevern, Bioinformation 1(5): 143-155 (2006)

Bioinformation, an open access forum
© Bornot & de Brevern, 2006, under the open access policy
The author(s) are responsible for the information published in this forum