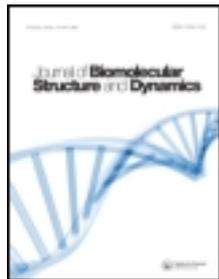


This article was downloaded by: [MSPU], [Alexey Gordeev]

On: 09 February 2013, At: 11:50

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Journal of Biomolecular Structure and Dynamics

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/tbsd20>

A novel structural tree for wrap-proteins, a subclass of ($\alpha+\beta$)-proteins

Eugenia A. Boshkova^a, Alexey B. Gordeev^a & Alexander V. Efimov^a

^a Institute of Protein Research, Russian Academy of Sciences, Pushchino, Moscow Region, 142290, Russian Federation

Version of record first published: 05 Feb 2013.

To cite this article: Eugenia A. Boshkova, Alexey B. Gordeev & Alexander V. Efimov (2013): A novel structural tree for wrap-proteins, a subclass of ($\alpha+\beta$)-proteins, Journal of Biomolecular Structure and Dynamics, DOI:10.1080/07391102.2012.760107

To link to this article: <http://dx.doi.org/10.1080/07391102.2012.760107>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

A novel structural tree for wrap-proteins, a subclass of ($\alpha+\beta$)-proteins

Eugenia A. Boshkova, Alexey B. Gordeev and Alexander V. Efimov*

Institute of Protein Research, Russian Academy of Sciences, Pushchino, Moscow Region 142290, Russian Federation

Communicated by Maxim Frank-Kamenetskii

(Received 2 November 2012; final version received 15 December 2012)

In this paper, a novel structural subclass of ($\alpha+\beta$)-proteins is presented. A characteristic feature of these proteins and domains is that they consist of strongly twisted and coiled β -sheets wrapped around one or two α -helices, so they are referred to here as wrap-proteins. It is shown that overall folds of the wrap-proteins can be obtained by stepwise addition of α -helices and/or β -strands to the strongly twisted and coiled β -hairpin taken as the starting structure in modeling. As a result of modeling, a structural tree for the wrap-proteins was constructed that includes 201 folds of which 49 occur in known nonhomologous proteins.

Keywords: classification; coiled β -hairpin; modeling; protein folding; unique fold

1. Introduction

The ($\alpha+\beta$)-proteins consist of a mixture of all- α and all- β regions which are segregated along the polypeptide chain (Levitt & Chothia, 1976). This structural class of proteins is rather large and can be subdivided into subclasses or superfamilies (Murzin, Brenner, Hubbard & Chothia, 1995; Orengo et al., 1997). In protein classification based on structural trees (PCBOST) (Gordeev, Kargatov & Efimov, 2010), there are also several subclasses of ($\alpha+\beta$)-proteins, e.g. proteins containing abCd-units (the abCd-unit has the overall fold which is very similar to that of the abcd-unit (Efimov, 1982) but contains helix C instead of strand c) and motifs analogous to the abCd-unit (Gordeev and Efimov, 2009, 2012) as well as four large groups of ($\alpha+\beta$)-proteins containing S-like β -sheets. For these subclasses, structural trees have been constructed and are available at <http://strees.protres.ru/>.

In this paper, one more structural subclass of ($\alpha+\beta$)-proteins is considered. A characteristic feature of proteins and domains of this subclass is that they consist of strongly twisted and coiled β -sheets wrapped around one or two α -helices, so they are referred to here as wrap-proteins. An analysis shows that overall folds of the wrap-proteins can be obtained by stepwise addition of secondary structural elements to the strongly twisted and coiled β -hairpin taking into account a set of simple rules described previously (Efimov, 1997).

2. Materials and methods

The database for the wrap-proteins was compiled using the Protein Data Bank (Bernstein et al., 1977) (<http://www.rcsb.org/pdb>) and SCOP (Murzin et al., 1995) (<http://scop.mrc-lmb.cam.ac.uk/scop/>). For example, proteins having cystatin-like, peptide deformylase and FKBP-like folds according to SCOP (Murzin et al., 1995) were included in this subclass. Proteins were manually selected using the RasMol molecular graphics program (Sayle and Milner-White, 1995). The secondary structure assignment was done with the RasMol. Possible homologies were revealed by protein-protein BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Tatusova and Madden, 1999). In total, our database includes 900 PDB entries of wrap-proteins and among them 136 are nonhomologous.

Theoretical modeling of protein folds and folding pathways using structural trees is of particular value in solving the protein folding problem since there are no experimental methods to observe protein folding in real time. In modeling, the structural motif having a unique overall fold and handedness is taken as the starting structure or the root structure of the tree. The larger structures are obtained by stepwise addition of α -helices and/or β -strands to the growing structure taking into account a restricted set of the rules inferred from the known principles of protein structure (Efimov, 1997):

*Corresponding author. Email: efimov@protres.ru

- (1) Overall folds of protein molecules and intermediate structures are taken into account and details of the structures are ignored.
- (2) A structural motif occurring in all the proteins of a given superfamily and having a unique overall fold, and unique handedness is taken as a starting structure in modeling or as a root structure of a tree.
- (3) The larger intermediate and protein structures are obtained by stepwise addition of α -helices and/or β -strands to a growing structure so that a structure obtained at the preceding step is maintained (it can be slightly modified). At each step, the α -helix or β -strand nearest to the growing structure along the chain is the first to be attached to the growing structure.
- (4) The obtained structures should be compact; α -helices and β -sheets should be packed in accordance with the rules that govern their packing (Chothia, Levitt, & Richardson, 1977; Efimov, 1977, 1979).
- (5) The α -helices and β -strands cannot be packed into one layer because of dehydration of the free NH and CO groups of the β -strands; thus, an α -helix should be packed into the α -helical layer and a β -strand into the β -layer of a growing structure.
- (6) Crossing of connections (Lim, Mazanov, & Efimov, 1978) and formation of knots (Richardson, 1977) are prohibited.
- (7) All the structural motifs (not only root motifs) or folding units should have corresponding handedness and overall folds. For example, all the β - α - β -units should form right-handed superhelices (Rao and Rossmann, 1973; Sternberg and Thornton, 1976).

A general scheme that represents the starting structural motif (the root motif), all the intermediate and completed structures connected by lines showing allowed pathways of structure growth, is referred to as the structural tree.

3. Results

An inspection of our database shows that all wrap-proteins contain strongly twisted and coiled β -hairpins. The strongly twisted and coiled β -hairpin can be represented as a double-helical structure in which the strands are twisted and coiled in a right-handed sense. A distinctive feature of this structure is that it is always formed by the right-turned β -hairpin in which the second β -strand runs on the right relative to the first one when viewed from the concave side of the superhelix (Efimov, 1991) (Figure 1). A more detailed stereochemical analysis shows that glycines located on the concave surface and

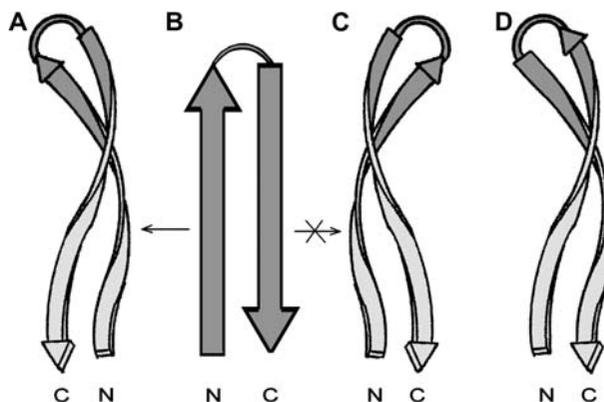


Figure 1. A β -hairpin can be strongly twisted and coiled to form the right-handed double-helical structure shown in A, but cannot be transformed into the superhelices shown in C and D (Efimov, 1991).

prolines located on the convex surface of the double-helical structure have to facilitate its formation (Efimov, 1991). Thus, the strongly twisted and coiled β -hairpin has the unique fold and handedness itself and can be taken as the starting structure in modeling of overall folds of the wrap-proteins.

Figure 2 represents a fragment of the structural tree constructed for the wrap-proteins. The number of possible overall folds that can be obtained from the root twisted and coiled β -hairpin is limited since the rules drastically reduce the number of allowed pathways of growth of the starting and intermediate structures. As seen, addition of an α -helix or a β -strand to the root β -hairpin at the first step can be done in different ways and results in formation of the folds shown in the bottom row of Figure 2. At the second step, addition of an α -helix or a β -strand to the structures of the bottom row can also be done in different ways etc. It should be noted that α -helices are packed into the concave surface of the strongly twisted and coiled β -hairpin or β -sheet. In these cases, a larger surface area is buried from water molecules (and, consequently, this is more favorable) than in the case of the α -helix packing onto the convex surface of the opposite β -sheet side. The complete structural tree of wrap-proteins includes 201 folds of which 49 occur in known proteins. The complete version of the tree is available at http://strees.protres.ru/7_2.xml as an extension of structural classification of proteins PCBOST (Gordeev et al., 2010). To date, our database includes 313 wrap-proteins (900 PDB-entries) and among them 136 are nonhomologous.

4. Discussion

In this paper, we present a novel subclass of (α + β)-proteins referred to as the wrap-proteins and the structural tree for them. It is shown that overall folds of the

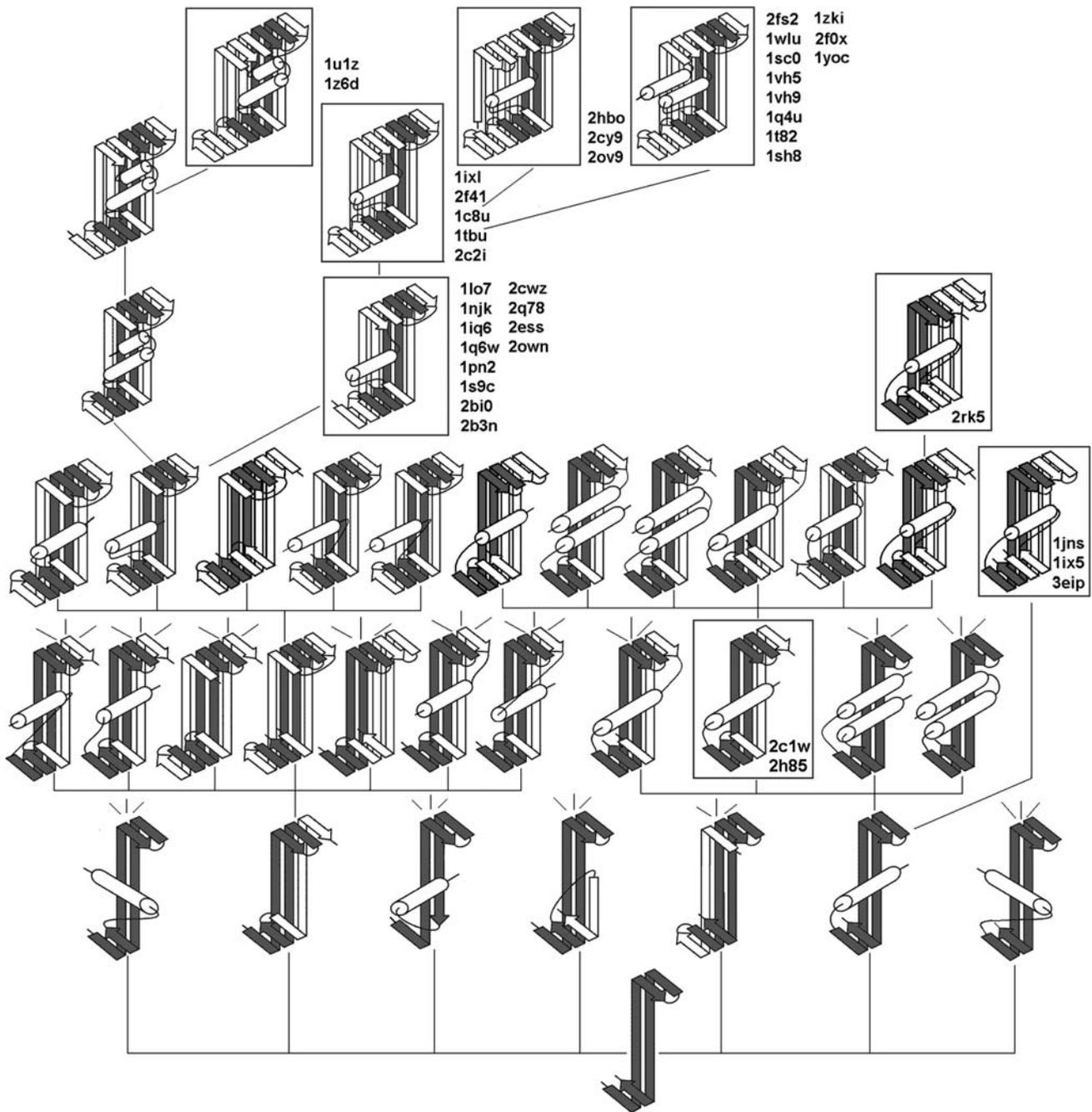


Figure 2. A fragment of the structural tree for the wrap-proteins, β -strands are shown with arrows directed from the N to C-ends and α -helices as cylinders. All structures are similarly oriented and are viewed from the concave surface of the root β -hairpin. The folds actually found in known proteins are framed.

wrap-proteins can be obtained by stepwise addition of α -helices and/or β -strands to the strongly twisted and coiled β -hairpin taken the starting structure in modeling. The structural tree also shows possible folding pathways of the wrap-proteins that lead to both known and yet unknown protein structures.

Although this approach does not predict a protein structure from its amino acid sequence, it does predict a

set of possible folds for a given structural class. A comparison of the first versions of structural trees constructed 15 years ago (Efimov, 1997) and updated structural trees (<http://strees.protres.ru/>) confirms this conclusion. Many overall folds included in the first structural trees (actually predicted) were not found in known proteins that time, but found in protein structures determined recently. Thus, the structural trees are good tool for the searching of all

possible folds, for the modeling of folding pathways of proteins and their structures, as well as for protein structure comparison and classification.

Acknowledgments

This work was supported by the Russian Foundation for Basic Research (Project No. 10-04-00727).

References

- Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Brice, M. D., Rodgers, J. R., ... Tasumi, M. (1977). The protein data bank: A computer-based archival file for macromolecular structures. *Journal of Molecular Biology*, *112*, 535–542.
- Chothia, C., Levitt, M., & Richardson, D. (1977). Structure of proteins: Packing of alpha-helices and pleated sheets. *Proceedings of the National Academy of Sciences USA*, *74*, 4130–4134.
- Efimov, A. V. (1977). Stereochemistry of the packing of alpha-spirals and beta-structure into a compact globule. *Doklady Akademii Nauk SSSR*, *235*, 699–702.
- Efimov, A. V. (1979). Packing of alpha-helices in globular proteins: Layer-structure of globin hydrophobic cores. *Journal of Molecular Biology*, *134*, 23–40.
- Efimov, A. V. (1982). Super-secondary structure of β -proteins. *Molecular Biology (Moscow)*, *16*, 799–806.
- Efimov, A. V. (1991). Structure of coiled β - β -hairpins and β - β -corners. *FEBS Letters*, *284*, 288–292.
- Efimov, A. V. (1997). Structural trees for protein superfamilies. *Proteins*, *28*, 241–260.
- Gordeev, A. B., & Efimov, A. V. (2009). Novel structural tree of (α + β)-proteins containing abCd-units. *Molecular Biology (Moscow)*, *43*, 480–484.
- Gordeev, A. B., & Efimov, A. V. (2012). Modeling of folds and folding pathways for some protein families of (α + β)- and (α / β)-classes. *Journal of Biomolecular Structure & Dynamics*, *31*, 4–16.
- Gordeev, A. B., Kargatov, A. M., & Efimov, A. V. (2010). PCBOST: Protein classification based on structural trees. *Biochemical and Biophysical Research Communications*, *397*, 470–471.
- Levitt, M., & Chothia, C. (1976). Structural patterns in globular proteins. *Nature*, *261*, 552–558.
- Lim, V. I., Mazanov, A. L., & Efimov, A. V. (1978). Stereochemical theory of the 3-dimensional structure of globular proteins: I Highly helical intermediate structures. *Molecular Biology (Moscow)*, *12*, 206–213.
- Murzin, A. G., Brenner, S. E., Hubbard, T., & Chothia, C. (1995). SCOP: A structural classification of proteins database for the investigation of sequences and structures. *Journal of Molecular Biology*, *247*, 536–540.
- Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B., & Thornton, J. M. (1997). CATH: A hierarchic classification of protein domain structures. *Structure*, *5*, 1093–1108.
- Rao, S. T., & Rossmann, M. G. (1973). Comparison of super-secondary structures in proteins. *Journal of Molecular Biology*, *76*, 241–256.
- Richardson, J. S. (1977). β -sheet topology and relatedness of proteins. *Nature*, *268*, 495–500.
- Sayle, R. A., & Milner-White, E. J. (1995). RasMol: Biomolecular graphics for all. *Trends in Biochemical Sciences*, *20*, 374–376.
- Sternberg, M. J. E., & Thornton, J. M. (1976). On the conformation of proteins: The handedness of the β -strand- α -helix- β -strand unit. *Journal of Molecular Biology*, *105*, 357–382.
- Tatusova, T. A., & Madden, T. L. (1999). Blast 2 sequences: A new tool for comparing protein and nucleotide sequences. *FEMS Microbiology Letters*, *174*, 247–250.