TWO-RESIDUE PERIODICITIES IN PROTEIN STRUCTURES : RESULTS FROM A SYSTEMATIC SEARCH IN 4-DIMENSIONAL RAMACHANDRAN SPACE

Ioannis Riziotis, Nicholas M. Glykos

Democritus University of Thrace, Dept. of Molecular Biology and Genetics, Alexandroupolis, Greece

INTRODUCTION

Secondary structure motifs in proteins are constructed by consecutive residues that share similar φ/ψ angle values. Representing the dihedral angle values in 2-dimensional *Ramachandran* space, regions that correspond to favoured residues of each secondary structure motif are formed. In our research, we raise the question of whether there are motifs in protein structure, which are defined by consecutive transitions between distinctive dihedral angle values or distinctive regions in the 2D *Ramachandran* space. For example, a set of continuous transitions between β-sheet and α_L-helix φ/ψ angle values, as shown in ***Fig. 1***. We developed a probabilistic algorithm to search for such patterns in a large sample of protein molecules submitted in the *Protein Data Bank*. Our current work is focused on tracing and describing motifs in protein structure that follow the pattern of transitions between two favoured φ/ψ pair angle value clusters in sets of five consecutive residues. Computational methods are used to cull these motifs from raw entries in the *Protein Data Bank* and clustering algorithms to produce representative structures. The main algorithm searches for these patterns in the 4-dimensional *Ramachandran* space (φ/ψ angle values for pairs of residues).

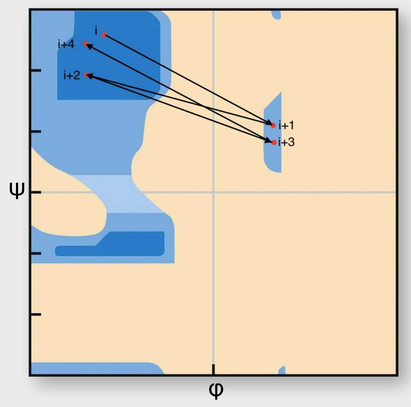

Fig. 1: *Our main hypothetic motif shown in Ramachandran plot. A transition pattern similar to the one of reverse turns, but more extended, is searched in the Protein Data Bank via a probabilistic algorithm.*

RESULTS

- Two representative examples of results (***Figs. 4, 5)*** are shown as *Ramachandran plots*. The transitions between distinctive regions are apparent.

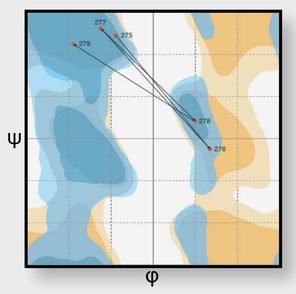

Fig. 4: *2DOV, residues A275-279 shown as Ramachandran plot. An example of transitions between β-sheet and α_L-helix regions*

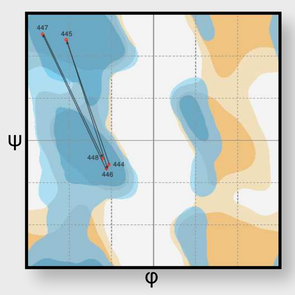

Fig. 5: *2ZF5, residues Y444-448 shown as Ramachandran plot. An example of transitions between a-helix and β-sheet regions*

- Hits with score >100.0 were used for dihedral clustering. ~6000 structures remained after omitting structures from homopolymers molecules and structures containing Glycine, Proline or Unknown residues.
- 6 dominant clusters were produced from dihedral clustering. ***Figs. 6-11*** show the superimposed structures of the clusters 1 (1675 structures), 6 (1411 structures) and 2 (137 structures) and the *Ramachandran plot* of each cluster.

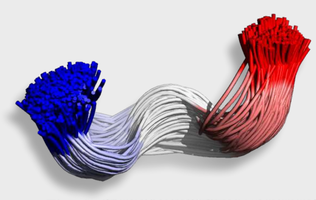

Fig. 6: *Cluster 1 (1675 structures)*

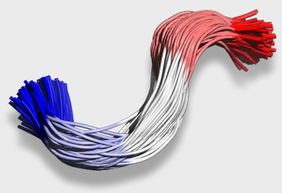

Fig. 8: *Cluster 6 (1411 structures)*

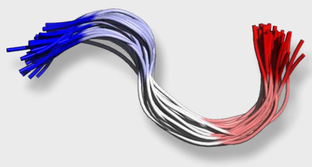

Fig. 10: *Cluster 2 (137 structures)*

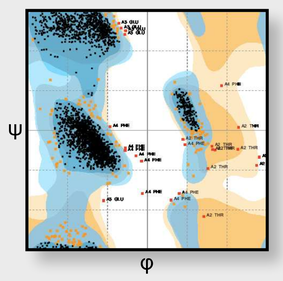

Fig. 7: *Cluster 1 Ramachandran Plot*

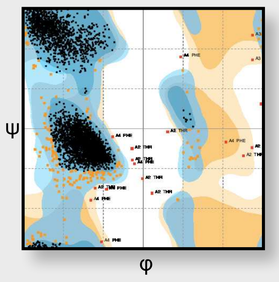

Fig. 9: *Cluster 6 Ramachandran Plot*

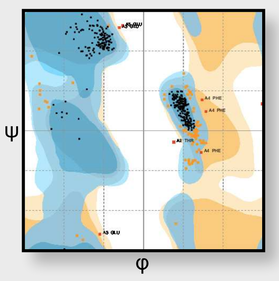

Fig. 11: *Cluster 2 Ramachandran Plot*

- ***Fig. 12*** shows example of 5 members of cluster 6 that seem to form a pattern, in superposition, with residues added in the N-terminus and C-terminus of the pentapeptide. The example is not a product of systematic research. More evidence is needed to conclude.

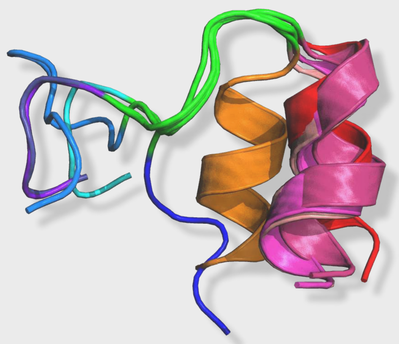

Fig. 12: *Superposition of 5 members of cluster 6 (shown in green) . N-terminus added residues are shown in shades of blue and C-terminus added residues are shown in shades of red*

METHODS

- 27118 protein entries downloaded from *PDB* using a *PISCES* culling list with the criteria of 3.0Å resolution and 80% identity cut-off.
- Extraction of *φ/ψ* angles of all residues except glycine and proline using *PROCHECK*.
- Calculation of euclidian distances of consecutive residue pairs in the 2D *Ramachandran* space, to define the standard deviation of two consecutive residues belonging in the same *Ramachandran* favoured region, by studying the distances histogram (***Fig. 2***).

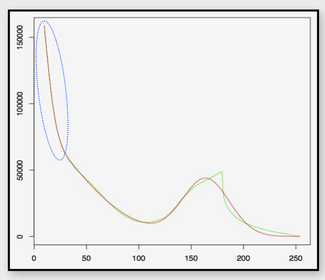

Fig. 2: *The [i - i+1] residue distances histogram (green line) and the fitted curve (red line). This histogram was used to define a dihedral angle value cluster in the 2D Ramachandran space, as a Gaussian distribution shown in the blue circle.*

- Use of the previous SD for developing a probabilistic algorithm to search the whole sample for two-residue periodicities, and score groups of five residues that match the hypothetic motif (***Fig.1***). The score given in each structure is the log-odds sum of the probability to content the hypothetic motif. For the conversion of euclidian *Ramachandran* distances to probabilities, the complementary Gauss error function was used (***Fig. 3***). *ANSI C* was used for implementing the main algorithm, *Perl* for various computational scripts and *R* for statistical analysis and clustering.

$$p = erfc\left(\frac{d}{2SD\sqrt{2}}\right) = 1 - \frac{2}{\sqrt{\pi}}\int_0^{\left(\frac{d}{2SD\sqrt{2}}\right)} e^{t^2}\,dt$$

Fig. 3: *The Gauss complementary error function.*
p : probability
d : euclidian distance
SD : standard deviation

- Cartesian cluster analysis (RMSD) of the potential high-scored results to produce representative structures using the R package with 1Å RMSD cut-off .

CONCLUSIONS

To conclude, we should review our results in comparison with the initial hypothesis:

- These early results show the potential existence of secondary structure motifs characterised by periodicities in the level of two residues.
- The periodicities seem to match our hypothetic model of transitions between two distinctive regions in the *Ramachandran* Plot.
- The conformation of the peptides found are comparable with β-turns. However, it should be annotated that Glycines and Prolines are totally omitted from the algorithmic search, to reduce noise and false positive results, such as type II and III reverse turns which firmly contain Glycine in their sequence.
- Cluster 2 contains a small number of structures as compared with clusters 1 and 6 but represents an interesting pattern made of transitions between b-sheet and α_L-helix.
- Continuous transitions between β-sheet and α-helix are shown in the *Ramachandran* plot of cluster 1, along with some residues in the α_L-helix region. This could be interpreted as error occurred from the cartesian clustering using the backbone of the peptides and not only the C_a atoms. The same applies to cluster 2 and 6 in a lesser extent.

FUTURE WORK

Our future intentions are mainly the characterisation of the patterns found in this part of the research, in terms of structure and function. We aim to answer the question of whether these motifs are conserved in primary structure level and if so, what is their functional role, if this exists. Furthermore, we are still searching for multi-residue periodicities by modifying the criteria used by the algorithm. For example, some early trials show the probable occurrence of three-residue periodicities. Some of the immediate plans also contain the count of Glycines and Prolines in the search to see the variation of our results.

REFERENCES

1. Lehninger, Albert L., David L. Nelson, and Michael M. Cox. Lehninger Principles of Biochemistry. New York: Worth Publishers, 2000.
2. G. Wang and R. L. Dunbrack, Jr. PISCES: a protein sequence culling server. Bioinformatics, 19:1589-1591, 2003
3. R. A. Laskowski, M. W. Macarthur, D. S. Moss, J. M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures J. Appl. Cryst., Vol. 26, pp. 283-291, 1993
4. S.C. Lovell, I.W. Davis, W.B. Arendall III, P.I.W. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson and D.C. Richardson (2002) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. Proteins: Structure, Function & Genetics. 50: 437-450.
5. The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

CONTACT INFO

Ioannis Riziotis, ioanrizi@mbg.duth.gr

Structural and Computational Biology Laboratory, NM Glykos' Group, Dept. of Molecular Biology and Genetics, DUTH

PDF QR Code: