DOI 10.18699/SBB-2022-150

Hierarchical structure of protein sequences – new insight of protein organization

Anashkina A.A.[1], Nekrasov A.N.[2]

[1] *Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*

[2] *Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, RAS, Moscow, Russia*

Key words: protein structure, hierarchy, protein sequences, ANIS method, protein design

Motivation and Aim: It should be noted that natural polypeptide chains (NPCs) have a number of properties that are unusual for polymers. In particular, NPCs are able to independently form stable and unique structures of spatial organization. This property allows "nature" to form many classes of "molecular machines" from NPCs, which ensure the flow of all physiological processes in living systems. Classical ideas about protein sequences are based on the idea that amino acid residues are the monomeric unit of NPCs. Such representations are natural and are associated with the chemical structure of polypeptide chains in which amino acid residues of different chemical type are linked by identical peptide groups. Initially, it was assumed that the properties of NPCs are determined by the patterns of distribution of amino acid residues along the polymer chain. However, repeated attempts to analyze the location of residues in the NPCs did not reveal significant patterns in the location. This suggested that there are other elementary structural units of NPCs. The aim of our work was to find the elementary units of the NPCs, allowing us to see the structural organization of the protein sequence.
Methods and Algorithms: To achieve the set task, the information entropy of a set of non-homologous natural protein sequences was studied. The study of the entropy characteristics of the NPCs showed that for groups of 5 successively located residues, a reduced level of informational entropy is observed and, therefore, blocks of precisely this size must be considered as elementary structural units of the NPCs. This approximation made it possible to develop a method that reveals the hierarchical structure of the NPCs – the Information Structure Analysis method (ANIS method) [10]. We assumed that the revealed hierarchical structures are associated both with the molecular evolution of NPCs and their functional activity.
Results: Further studies showed that the use of the ANIS method made it possible to solve a number of important practical problems in various fields [2] – the design of protein molecules, studies of intermolecular interactions, and the study of the mechanisms of protein functioning. In addition, it was shown on model structural blocks that, with certain combinations of amino acid residues, they can only be located in the predominant conformational given topology [9].
To date, the ANIS method has been successfully applied to solve a number of practical problems in the design of protein molecules. Thus, a truncated form of human 1-CYS peroxiredoxin was obtained, which retained about 95 % of the catalytic activity of the wild form of the protein [6]. In a modified form of human Interleukin-13, the ability to give a signal for the development of an allergic reaction was removed [7]. The specificity of oligopeptidase from Serratia proteamaculans was changed and the enzyme acquired the ability to cleave high molecular weight substrates [5]. A highly efficient hydrolase has been made from the bacteriophage φKZ polypeptide [3]. Minimal sites that retain

the functional activity of native proteins have been identified in human tumor necrosis factor [11] and human heat shock protein [12].

In addition, in the study of protein-protein interactions of hydrolase-inhibitor complexes, it was shown that the interface providing their interaction consists of almost 85 % pairs of sites with a low and high density of elements of the information structure of the lower level of the hierarchy [8]. Based on the revealed regularities in the formation of interchain interaction interfaces, a method for identifying immunogenic sites in proteins was proposed and an immune response was obtained for protein epitopes of Aspergillus fumigatus Asp f 2 and Asp f 3 [1]. Models for the functioning of a number of hydrolytic enzymes and heme-containing proteins have been proposed [4].

The totality of the presented theoretical and experimental data allows us to speak about the birth of a new paradigm of the structural organization of proteins, which is based on the entropy characteristics of the NPCs. This opens up fundamentally new opportunities not only for studying and modifying known NPCs, but also for "designing" fundamentally new "molecular machines" in terms of organization and function.

Acknowledgements: This research has been supported by Russian Foundation for Basic Research Grant No. 20-04-01085.

References

1. Alekseeva L. et al. Cryptic B-cell epitope identification through informational analysis of protein sequences. *Vaccine*. 2007;25(14):2688-2697.
2. Anashkina A.A. et al. Entropy analysis of protein sequences reveals a hierarchical organization. *Entropy*. 2021;23(12):1647.
3. Briers Y. et al. The structural peptidoglycan hydrolase gp181 of bacteriophage phiKZ. *Biochem Biophys Res Commun*. 2008;374(4):747-751.
4. Chertkova R.V. et al. New insight into the mechanism of mitochondrial cytochrome C function. *PLoS One*. 2017;12(5):e0178280.
5. Mikhailova A.G. et al. Truncated variants of Serratia proteamaculans oligopeptidase B having different activities. *Biochemistry (Mosc.)*. 2015;80(10):1331-1343.
6. Nekrasov A.N. et al. The novel approach to the protein design: active truncated forms of human 1-CYS peroxiredoxin. *J Biomol Struct Dyn*. 2007;24(5):455-462.
7. Nekrasov A.N. et al. Design of a novel interleukin-13 antagonist from analysis of informational structure. *Biochemistry (Mosc.).* 2009;74(4):399-405.
8. Nekrasov A.N. et al. Structural features of the interfaces in enzyme-inhibitor complexes. *J Biomol Struct Dyn*. 2010;28(1):85-96.
9. Nekrasov A.N. et al. A minimum set of stable blocks for rational design of polypeptide chains. *Biochimie.* 2019;160:88-92.
10. Nekrasov A.N. et al. Hierarchical structure of protein sequence. *Int J Mol Sci*. 2021;22(15):8339.
11. Shingarova L.N. et al. Expression and properties of human TNF peptide fragments. *Russ J Bioorganic Chem*. 2010;36(3):301-309.
12. Vlaskin et al. Activating effect of 70 kDa heat shock protein on human NK cells. *Immunologia (Russian).* 2007;28(2):74-79.