POXVIRAL CHEMOKINE-BINDING PROTEINS: THEORETICAL STUDY OF STRUCTURE AND FUNCTION EVOLUTION

K.V. Gunbin[2],

Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia
genkvg@gmail.com

D.V. Antonets,

State Research Center of Virology and Biotechnology "Vector", Novosibirsk, Russia
antonec@yandex.ru

T.S. Nepomnyashchikh

State Research Center of Virology and Biotechnology "Vector", Novosibirsk, Russia

Poxviridae is a family of large enveloped double-stranded DNA viruses replicating exclusively in the cytoplasm of infected cells. Their genomes vary in size from 130 to 360 kb and contain over 200 open reading frames. The majority of poxviral genes encode different proteins determining the host range, tissue tropism and immunomodulatory proteins, subverting anti-viral immune responses of the host. In particular, poxviruses produce special secreted proteins that bind cytokines and chemokines and block their activity. An important feature of these proteins is their pleiotropic mode of action – many of them are able to bind and inhibit more than one particular chemokine or even chemokines of distinct types (CC, XC, CXC or CX3C). The prototypic member of the family – a cowpox virus CC-chemokine binding protein vCCI bind a wide spectrum of CC-chemokine with nanomolar affinity thus preventing their interaction with cognate receptors on leukocytes blocking their activation and influx to viral replication site. More recently, it was another evasion strategy described that targets chemokine gradient formation by preventing chemokines from binding to glycosaminoglycans (GAGs) of intracellular matrix and cell surface. Consequently, a chemokine concentration gradient is not established near the infection site and leukocytes are not recruited. Such strategy is used by myxoma virus M-T7 protein and by vaccinia virus A41 protein. It's of great interest that despite low sequence identity poxviral chemokine-binding proteins share remarkably similar tertiary structures. Structural analysis revealed the recurrent presence of a conserved β-sandwich fold in these proteins that has not been observed in any eukaryotic or prokaryotic protein. This

domain was recently named PIE (Poxvirus Immune Evasion) domain [1]. PIE proteins are encoded by a plethora of genes in poxviral genomes, e.g., cowpox virus encodes 10 different PIEs: vCCI, A41, C8, M2, T4 (CPVX203), and some SECRET-domain containing proteins CrmB, CrmD, SCP-1, SCP-2, and SCP-3. The PIE subfamilies differ primarily in the number, size, and location of structural embellishments to the β-sandwich core that confer unique functional specificities. Therefore, an interesting question is the evolutionary relationships of PIE-containing proteins and their broad functions. With the help of phylogenetic analysis and molecular modeling here we tried to get some insights about the evolution of this protein superfamily and also about the changes of their molecular functions.

The protein sample was collected with DELTA-BLAST using 6 query proteins from [2], having experimentally reconstructed tertiary structures (2ffk, 2vga, 3on9, 4hkj, 4zk9, 5d28). These 6 proteins covered the maximum number of subfamilies of the chemokine-binding proteins type II protein family (CBPtII). This allowed us to further investigate both the evolution of the CBPtII primary protein sequences, and the evolution of their secondary and tertiary (3D) structures. For the last purpose, for each protein in the sample, using the SCRATCH-1D 1.1 system, the 8-element secondary structure and the degree of contact of each amino acid residue with the solvent were predicted. The obtained protein sample was aligned using PROMALS3D web-server, using data on secondary and tertiary protein structures. It is important to note that analysis of proteins using data on secondary and tertiary structures significantly improves the quality of the resulting alignment, in particular alignment of evolutionarily distant subfamilies of proteins. Further, the alignment of the primary sequences of proteins was translated into alignment of 8 elements of secondary structures and 20 classes of residue solvent accessibility values. The number of solvent accessibility classes is equal to the number of canonical amino acids that allows us to encode these classes into symbols of amino acids and then work with them just like with amino acids. The evolutionary variability of protein primary sequences and residue solvent accessibility sequences are studied by reconstruction of rate matrices of substitutions observed in CBPtII protein family, for this purpose the MODELESTIMATOR software was used [3]. The resulting rate matrices were used in the phylogenetic analysis. For the phylogenetic analysis of the 8-element secondary structures, the GTR evolutionary model was used. The phylogenetic analysis was based on the

maximum likelihood method implemented in the RAXML 8. The statistical significance of protein clusters (based on primary and secondary sequences, and on the residue solvent accessibility values) was investigated using a 50% jackknife procedure [4, 5], which is based on 300 pseudo-replicas. The final consensus trees were built using the SUMMTREES program of the DENDROPY 4.2 library based on a 50% consensus. To compare topologies and clusters obtained on the basis of primary and secondary sequences, as well as on the residue solvent accessibility values, the tanglegrams were reconstructed using DENDROSCOPE 3.4.0. To investigate the rate of the CBPtII protein surface evolution (obtained on the basis of residue solvent accessibility values), we reconstructed the phylogeny of Neuraminidase proteins of influenza A virus (using the above described approach, on the basis of residue solvent accessibility values) that were extracted from the IRD database. We used Neuraminidase proteins of influenza A virus taking into consideration numerous demonstrations of Neuraminidase adaptive evolution associated with a sharp change in protein surface.

The study of the topology of CBPtII proteins phylogenetic trees obtained on the basis of protein primary sequences allowed us to identify 7 main clusters: cluster 1, vCCI orthologs and vCCI-like proteins with long loop and strong negative charge on both sides; cluster 2, parapoxviral GIF-proteins (GM-CSF/IL-2 binding proteins), they have got a profound electronegative charge at one side, while they also do have a profound positive charge at the opposite side, that might interact with GAGs; cluster 3, parapoxviral chemokine-binding proteins, they are share amino acid and functional identity with the poxvirus type II CC-chemokine-binding (CBP-II) proteins and A41L chemokine binding virulence factor, they do have the negative charge at the most area, but they also have a patch of positive charge; cluster 4, A41 and A41-like proteins, that were shown to bind GAGs and to compete with chemokines for GAGs; cluster 5, CPXV218 SECRET-containing protein of cowpox virus; cluster 6, endoplasmic reticulum (ER)-localized apoptosis regulators, M-T4 type of the myxoma virus that may have a dual function in protecting infected lymphocytes from apoptosis and in modulating the inflammatory response to virus and CPXV203 cowpox virus protein, which is known to block the normal trafficking of MHC I from the ER to the plasma membrane; cluster 7, two domains containing proteins, N-terminal domain binds TNF and is homologous to cellular TNF receptor, while C-terminal domain (this domain is also called SECRET) belongs to PIE family,

the function of this domain is binding a number of CC, XC and CXC chemokines. Using the computational approach described above, we provided the comprehensive evidence that the diversification of CBPtII proteins was mainly due to a change in its protein surface. However, the comparison of stemness (the total length of the internal branches of the phylogenetic tree) of CBPtII and Neuraminidase phylogenetic trees shows that the rapid evolution of the CBPtII protein surface at the origin of CBPtII protein clusters is characteristic not for all 7 clusters of CBPtII protein family. The entire cluster 3 and the half of cluster 5 (the proteins related to ER-localized apoptosis regulators of the M-T4 type of the myxoma virus) does not characterized by rapid evolution of surface properties. Additionally, these two groups of CBPtII proteins characterized by lowest changes in secondary structures, and cluster 3 shows lowest rate of primary structure evolution. These facts allows us to conclude that 1) these two groups of CBPtII proteins are closely related to the last common ancestor of CBPtII protein family and 2) the ancestral function of CBPtII proteins may be a chemokine-binding.

Acknowledgements: This work was supported with RFBR grant #15-04-08956-a.

1. Nelson CA et al. (2015) Structural Conservation and Functional Diversity of the Poxvirus Immune Evasion (PIE) Domain Superfamily. *Viruses.* **7**:4878-4898.

2. Felix J et al. (2016) Structural basis of GM-CSF and IL-2 sequestration by the viral decoy receptor GIF. *Nat Commun*. **7**:13228.

3. Arvestad L. (2006) Efficient methods for estimating amino acid replacement rates. J Mol Evol. **62**:663-673.

4. Freudensteina JV, Davis JI. (2010) Branch support via resampling: an empirical study. *Cladistics* **26**:643-656.

5. Mort ME, J et al. (2000) Comparison of three methods for estimating internal support on phylogenetic trees. *Syst Biol*. **49**:160-171.