# Influenza Virus Nuclear Export Protein (NEP) 3D Structure Reconstruction Probed as an Example for Improvement of Approach to the Protein 3D Structures Prediction

E.N. Bogacheva*, A.O. Golovko**, A.N. Bogachev*, A.A. Dolgov*, V.A. Radyukhin**

**N.N. Semenov Institute of Chemical Physics Russian Academy of Sciences,*

*ul. Kosygina, 4, Moscow, 119991 Russia, e-mail: ben@chph.ras.ru*

***M.V. Lomonosov Moscow State University, 1/40 Leninskie Gory, Moscow, 119991, Russia: e-mail: nastiagolovko@mail.ru*

Influenza virus A is a major pathogen causing seasonal epidemics and pandemics that are associated with significant mortality, so detailed data on the structure and functions of the virus is of great importance. Until now, the mechanism of the virus infection of cells is unknown in details, and many phases of this process are hypothetical. Nuclear export protein participates in several important processes during the development of the infection. It mediates nuclear export and regulates activity of the viral polymerase complex. Structure of the full-sized NEP is still unresolved.

NEP is a 14 kD protein, composed of 121 amino acids. The NEP protein has two domains: C-domain consists of two alpha-helices, and N domain (aa 1-53) with unknown structure. At present only a 3D structure model for the N-terminal domain (the disordered region) obtained by the predictive method on the I-Tasser server (http://zhanglab.ccmb.med.umich.edu/I- TASSER/) is available: according to this model the domain consists of two α-helices - N1 and N2 [1]. A high-resolution structure was obtained and deciphered from X-ray diffraction data for the C-terminal fragment of NEP (aa 59-116), the C domain is a hairpin of two α-helices C1 and C2 connected by a short linker [2].

This work is one in a series of studies devoted to the study of the structure of proteins of the influenza virus A [3-6]. An attempt to suggest structure of the full-size NEP by means of bioinformatics tools and represent its 3D model was made. We have predicted the protein secondary and tertiary structures, as well as intrinsic disorder. Comparison of the predicted protein structures allows us for more detailed understanding the role of the protein in the infection mechanism. The low accuracy of the prediction programs of secondary and 3D structure requires using a large set of prediction programs and their subsequent averaging. Significantly improve the quality of predictions is also depends on the solution of the inverse problem: prediction of unstructured fragments (loops) in proteins.

## BIOINFORMATICS TOOLS

The server of the Institute of Bioinformatics of Switzerland ExPASy (http: /expasy.org/) was used when searching for the methods of protein structure predicting. The secondary structure of the N-domain NEP is defined using the APSSP2, Jpred, Jnet, JUFO, PHD, PHDpsi, Porter, PROF_king, PSIpred, SABLEand

SAMT99sec. Modeling of the spatial structure was carried out by the methods of I-Tasser, Scratch, ProteinPredict, RaptorX, Phyre2, Protein Model Portal, Modeller 9.8. To refine the structures obtained, the methods are used to predict the disordered regions in the protein IUPred , DISOPRED , DisemBL, FoldUnfold, PrDoc, FoldIndex, MFDP, DISISOclust, IUPredL (type of prediction - long disorder) and IUPredS (short type of prediction).

## RESULTS and DISCUSSION

A careful analysis of the earlier proposed the 3D model for NEP [1] showed that most binding sites with ligands in this structure locate in hard-to-reach places (crevices and deep pockets), that results in improper functional representation and, from our point of view, reduces the value of the model. In turn, the authors of this model admit the ambiguity of the chosen model.

Therefore, we conducted our own simulation of the full-length protein structure using a whole set of methods for predicting the secondary and tertiary structure. It is known that the accuracy of prediction methods does not exceed 80%, and in this case modeling with various approaches and programs can significantly improve the accuracy of the final model. In our case, the number of predicted alpha helices in the protein N-domain ranged from 2 (I-TASSER [7]) to 5 (SCRATCH [8]). The predicted N-domain of NEP protein by SPP [7] (Fig. 1a) has only one complete helix 3-17. Dimensions of remaining helices have less than two turns, so the question of their existence is controversial (28-34, 36-42, 46-52, 55-61). Energetically nonequilibrium inputs and outputs to and from the helix lead to the conventions of a precise definition. Fig. 1b shows a diametrically opposite view of the structure of this domain, obtained by the RaptorX server. A bundle of three antiparallel helices was determined for N domain.

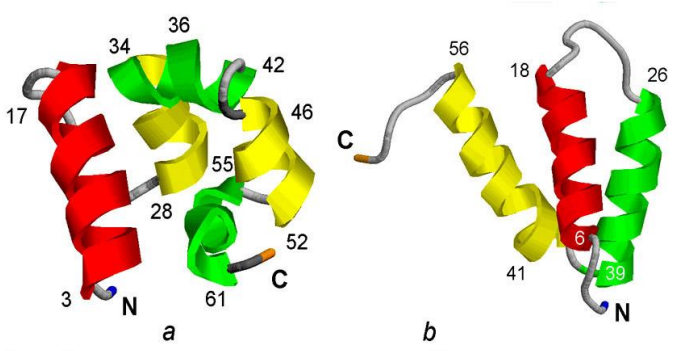


**Fig. 1.** The spatial structures of the N domain NEP protein, predicted using Scratch Protein Predict program (a) and Modeller 9.8 (b).

We looked for the best correlation across the polypeptide chain between two groups of mutually exclusive fragments of the molecule: the secondary structure (alpha helices in our case) and the disordered regions (loops). Fig. 2 shows the superposition of the disorder profile and the secondary structure elements predicted by the Modeller program along the polypeptide chain with the best correlation.

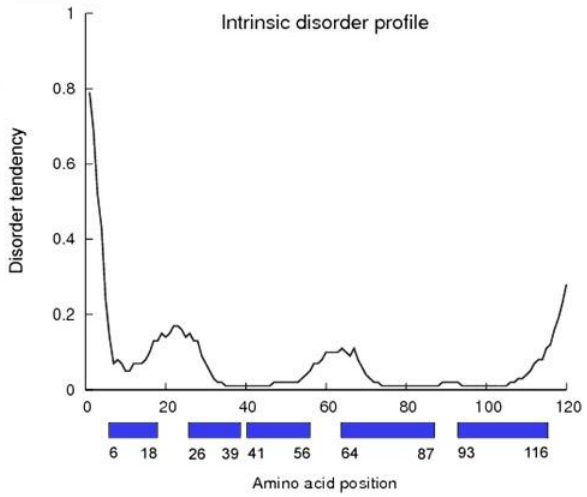


**Fig. 2.** Distribution of disordered regions along a polypeptide chain (disorder profile, IUPred program). Under the abscissa axis, the gray rectangles show alpha-helices predicted by the Modeller program.

Taking into consideration these factors, and also the X-ray data for the protein as well as the deposited NEP C-domain crystal structure (PDB ID: 1PD3) the most probable models for the NEP 3D structure were obtained using the Modeller 9.8 program [9]. In this scheme, the N-domain represents a triple-helical bundle of antiparallel helices. The simulation results are shown in Fig. 3.

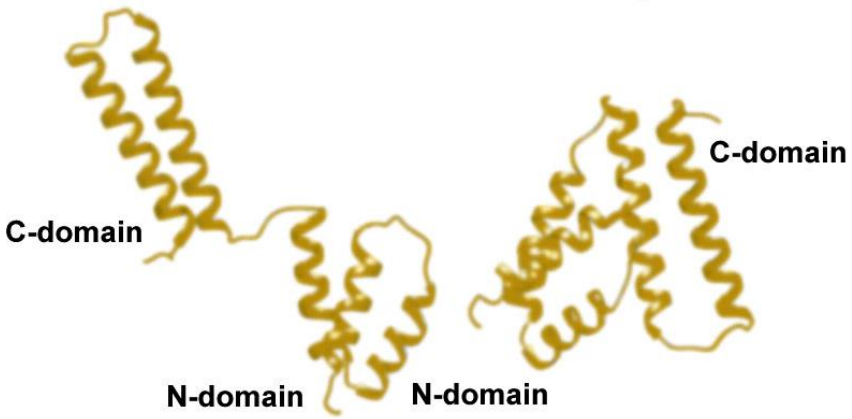


**Fig. 3.** Model of 3D structure of NEP protein: expanded (a) and compact conformation (b).

The 3D structures of NEP in Fig. 3 show the putative realization in at least two equiprobable, deployed and compact, conformations in this protein, both of them are conditioned by the presence of a flexible long loop of the amino acid sequence between the two domains.

The multifunctionality of this protein in the infected cell appears to be associated with its plastic 3D structure, which is provided at the expense of the unstructured regions contacting with various molecules partners. The 3D structure model obtained with the Modeller 9.8 program has the best correlation between the

methods of predicting the elements of the secondary structure and unordered regions.

Thus, the proposed three-dimensional protein structure is the first structural representation of the full-length NEP protein.

This work is supported in part by project of Russian Foundation for Basic Researches (№16-04-00563).

## REFERENCES


1. Darapaneni V., Prabhaker V.K., Kukol A. (2009) Large-scale analysis of influenza A virus sequences reveals potential drug target sites of non-structural proteins. *J. Gen. Virol.* **90**: 2124–2133.
2. Akarsu H., Burmeister W., Petosa C. (2003) Crystal structure of the M1 protein-binding domain of the in¯uenza A virus nuclear export protein (NEP/NS2). *EMBO J.* **22** (18): 4646–4655**.**
3. L. A. Baratova, E. N. Bogacheva, V. I. Gol'danskii, V. A. Kolb, A. S. Spirin, and A. V. Shishkov (2009) *Tritium Planigraphy of Biological Macromolecules.* Moscow: Nauka, 1999. 175.
4. Shishkov A.V., Bogacheva E.N., Dolgov A.A. et al. (2009) The *In Situ* Structural Characterization of the Influenza A Virus Matrix M1 Protein within a Virion. *Protein & Peptide Lett*. **16:** 1407-1413.
5. A. Shishkov, E. Bogacheva, A. Dolgov et al. (2011) Spatial structure peculiarities of influenza A virus matrix M1 protein in an acidic solution that simulates the internal lysosomal medium. *FEBS J.*, **278:** 4905-4916.
6. Shtykova E. V., Baratova L. A., Fedorova N. V. et al. (2013) Structural Analysis of Influenza A Virus Matrix Protein M1 and Its Self-Assemblies at Low pH. *PLoS ONE,* doi: 10.1371.
7. Brender, J. R., Shultis D., Khattak N.A., ZhangY. (2016) An Evolution Based Approach to de novo Protein Design. *Methods in Molecular Biology.* 1529: 243-264.
8. Cheng J., Randall A. Z., Sweredoski M. J., Baldi P. (2005) SCRATCH: a protein structure and structural feature prediction server. *Nucl. Acids Res.* 33, W72-W76. Web Server issue doi: 10.1093 / nar / gki396.
9. M. Kallerg, H. Wang, J. Peng, S. Wang, Z. Wang, H. Lu and J. Xu. (2012) Template-based protein modeling using the RaptorX web server. *Nature Protocols*. **7** (8): 1511-1522.