

Co-evolution analysis to predict protein-protein interactions within the Influenza virus envelope

Ramil R. MINTAEV

Department of Bioengineering and Bioinformatics, Moscow State University, Leninskie Gory 1-37, Moscow 119991, Russia, ramil.mintaev@gmail.com

Andrei V. ALEXEEVSKI

Department of Bioengineering and Bioinformatics & Belozersky Institute of Physico-Chemical Biology, Moscow State University, Leninskie Gory 1-37 & 1-40, Moscow 119991, Russia, aba@belozersky.msu.ru

Larisa V. KORDYUKOVA

Belozersky Institute of Physico-Chemical Biology, Moscow State University, Leninskie Gory 1-40, Moscow 119991, Russia, kord@belozersky.msu.ru

Influenza A virus belongs to Orthomyxoviridae family of enveloped viruses. Three transmembrane proteins are incorporated into the viral envelope: the surface glycoproteins hemagglutinin (HA) and neuraminidase (NA) that fulfill the virions' attachment and cell receptor destroying function, respectively, and a minor protein M2 facilitating virions budding and providing a proton input into the virion's interior. Major structural protein M1 underlies the viral lipid membrane. There are indirect experimental data confirming the existence of protein-protein interactions within the influenza virus envelope. Yet, there is no data available regarding the amino acids residues participating in those interactions.

To predict amino acid interactions for pairs of envelope proteins, we (1) calculated the so-called Direct Information (DI) based on an approach of [Marks et al., PLoS One 6(12) **2011**]. To distinguish true co-evolution couplings from the noisy set of correlations observed according to the multiple sequence alignment, the authors propose computing a set of direct residue couplings. Second, among those amino acid residue pairs possessing $DI > 0,1$ we have manually chosen those located within the topographically compatible protein domains. Next, we calculated amino acid pair frequencies for our candidates. Last, we checked whether the amino acid residues within a pair are compatible regarding their physicochemical properties.

We have found 2 amino acid (a.o.) pairs for the HA-M1 protein pair, 1 and 3 a.o. pairs for the HA-M2 and M1-M2 pairs, respectively (in total, 6 a.o. pairs, **Fig. 1**). No a.o. pairs were found for the HA-NA pair because of extreme conservation of the NA cytoplasmic tail.

Pair No.	1	2	3	4	5	6
Protein	Amino Acid Position					
HA	557	557	557			
M1	239	167		239	209	214
M2			54	54	77	77
DI	0.13	0.12	0.11	0.17	0.49	0.55
Amino acid pair frequency	N-A: 0.68 K-T: 0.23	N-T: 0.66 K-A: 0.23	N-R: 0.56 K-L: 0.23	A-R: 0.67 T-L: 0.2	A-R: 0.71 T-Q: 0.28	Q-R: 0.71 H-Q: 0.28

Fig. 1. Amino acid residues predicted to participate in the protein-protein interactions.

According to our predictions, a hypothetical model of protein-protein interactions within the viral envelope was proposed (**Fig.2**). Noteworthy, the interactions may not be realized simultaneously since there is different copying of the proteins within the virion.

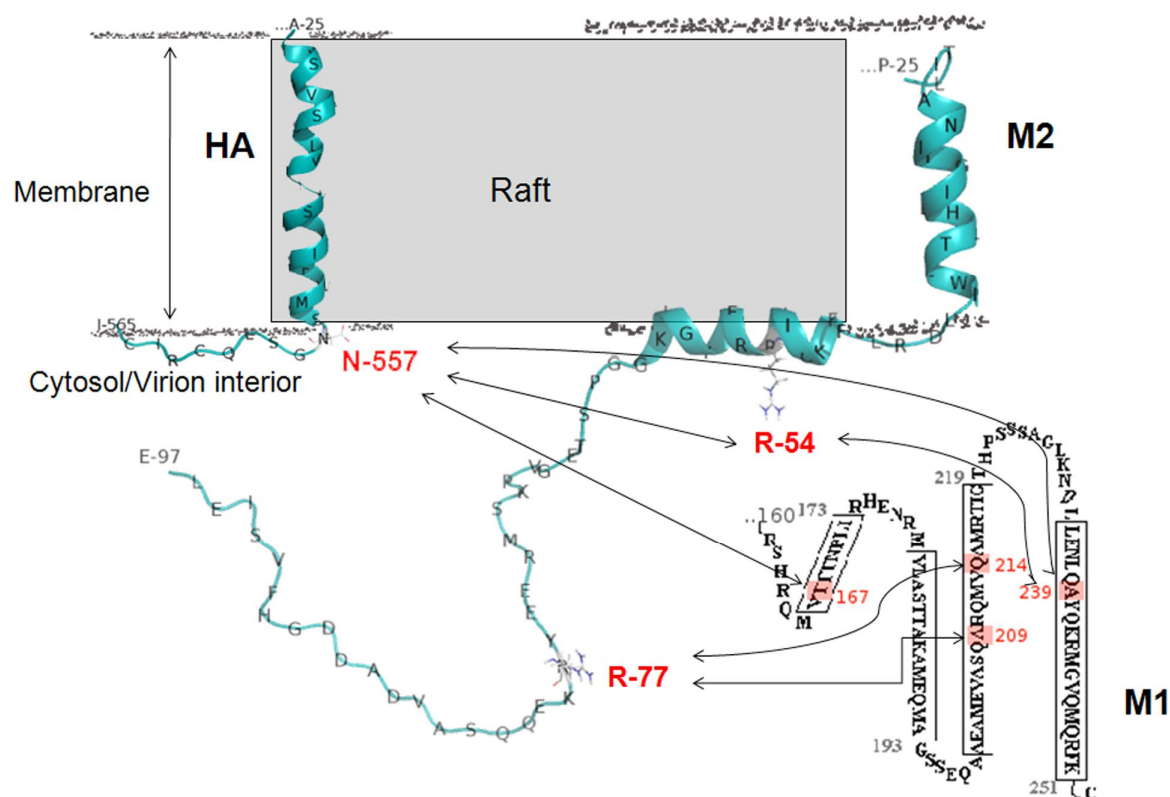


Fig.2. Hypothetical model of the protein-protein interactions within the viral envelope.

This work was supported by RFBR grants 11-04-91340-NNIO, 13-04-00290.