ISSN 0006-2979, Biochemistry (Moscow), 2018, Vol. 83, No. 11, pp. 1411-1421. © Pleiades Publishing, Ltd., 2018.
Original Russian Text © A. O. Golovko, O. N. Koroleva, A. P. Tolstova, N. V. Kuz'mina, E. V. Dubrovin, V. L. Drutsa, 2018, published in Biokhimiya, 2018, Vol. 83, No. 11, pp. 1746-1758.

Aggregation of Influenza A Virus Nuclear Export Protein

A. O. Golovko[1,a], O. N. Koroleva[2,b]*, A. P. Tolstova[3,c], N. V. Kuz'mina[4,5,d], E. V. Dubrovin[3,e], and V. L. Drutsa[6,f]

[1]*Lomonosov Moscow State University, Department of Bioengineering and Bioinformatics, 119991 Moscow, Russia*
[2]*Lomonosov Moscow State University, Department of Chemistry, 119991 Moscow, Russia*
[3]*Lomonosov Moscow State University, Department of Physics, 119991 Moscow, Russia*
[4]*Lomonosov Moscow State University, Department of Biology, 119991 Moscow, Russia*
[5]*Frumkin Institute of Physical Chemistry and Electrochemistry, Russian Academy of Sciences, 119071 Moscow, Russia*
[6]*Lomonosov Moscow State University, Belozersky Research Institute of Physico-Chemical Biology, 119991 Moscow, Russia*
[a]*e-mail: nastiagolovko@mail.ru*
[b]*e-mail: koroleva@genebee.msu.ru*
[c]*e-mail: tolstova@physics.msu.ru*
[d]*e-mail: kuzmina-natasha@inbox.ru*
[e]*e-mail: dubrovin@polly.phys.msu.ru*
[f]*e-mail: drutsa@genebee.msu.ru*

Received April 12, 2018
Revision received July 4, 2018

Abstract—Influenza A virus nuclear export protein (NEP) plays an important role in the viral life cycle. Recombinant NEP proteins containing $(His)_6$-tag at either *N*- or *C*-terminus were obtained by heterologous expression in *Escherichia coli* cells and their high propensity for aggregation was demonstrated. Dynamic light scattering technique was used to study the kinetics and properties of NEP aggregation in solutions under different conditions (pH, ionic strength, presence of low-molecular-weight additives and organic solvents). Using atomic force microscopy, the predominance of spherical aggregates in all examined NEP preparations was shown, with some amyloid-like structures being observed in the case of NEP-C protein. A number of structure prediction programs were used to identify aggregation-prone regions in the NEP structure. All-atom molecular dynamics simulations indicate a high rate of NEP molecule aggregation and reveal the regions preferentially involved in the intermolecular contacts that are located at the edges of the rod-like protein molecule. Our results suggest that NEP aggregation is determined by different types of interactions and represents an intrinsic property of the protein that appears to be necessary for its functioning *in vivo*.

DOI: 10.1134/S0006297918110111

Keywords: influenza A virus, nuclear export protein NEP, protein aggregation

Nuclear export protein (NEP) plays a key role in the life cycle of the influenza A virus. In particular, it regulates the activity of the viral polymerase complex [1], serves as a mediator between the complex of the viral ribonucleoprotein with the M1 matrix protein and nuclear export machinery [2, 3], and facilitates alterations of membrane curvature during virion budding [4].

Two domains have been identified in the NEP structure (121 a.a.): a weakly structured *N*-terminal domain and a highly structured *C*-terminal domain. The *C*-terminal domain represents an antiparallel amphiphilic hairpin of two α-helices connected with a short linker and is capable of crystallizing as a dimer [5]. Negatively charged glutamic acid residues are predominately concentrated on the hydrophilic side of the hairpin. The only hydrophobic Trp78 residue in this area plays an important role in the NEP interaction with the matrix protein M1 [5]. The *N*-terminal domain of the protein has a flexible conformation [6]; however, molecular modeling data indicate that it can form a hairpin from two antiparallel α-helices [7].

Previously, we have constructed plasmid vectors for efficient expression of the NEP and its variants (A/Puerto

Abbreviations: a.a., amino acid; AFM, atomic force microscopy; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate; DLS, dynamic light scattering; GuHCl, guanidine hydrochloride; NEP, nuclear export protein; NEP-C and NEP-N, recombinant NEP proteins with $(His)_6$-tag at the *C*- and *N*-termini, respectively; PdI, polydispersity index; R_h, hydrodynamic radius.

* To whom correspondence should be addressed.

Rico/8/1934 strain (H1N1)) in *Escherichia coli* cells that carried the $(His)_6$-tag at either *C*- or *N*-terminus (NEP-C or NEP-N, respectively). The obtained recombinant proteins were purified from solubilized inclusion bodies [8]. It was found that NEP exhibits an extremely strong tendency for aggregation. Lommer et al. in [6] mentioned that NEP forms insoluble precipitate, although the nature of the latter has not been investigated.

Aggregation is an inherent property of many viral proteins that is often associated with protein biological function. In particular, at least two influenza A virus proteins are capable of forming functionally significant aggregates: the matrix protein M1 that polymerizes with the layer formation under the viral particle membrane [9-11] and the PB1-F2 protein, whose amyloid-like aggregates disrupt the integrity of the mitochondrial membrane during viral infection thereby inducing cell apoptosis [12]. Like NEP, these proteins are characterized by the presence of extended unstructured regions common for the majority of amyloidogenic proteins [13].

The objective of this study was to investigate the features of NEP aggregation *in vitro* under various conditions in an attempt to identify structural elements participating in the protein self-assembly. The revealed properties of NEP aggregation could possibly help to establish functional significance of protein aggregates in the viral infection of cells.

MATERIALS AND METHODS

Reagents. The following reagents were used in the study: Hepes (Merck, Germany), SDS (Sigma, USA), GuHCl (Fluka, Germany), urea, arginine (Sigma, USA), and CHAPS (Amresco, Canada).

Analysis of NEP, NEP-C, and NEP-N expression. NEP, NEP-C, and NEP-N were expressed in *E. coli* ER1821 cells transformed with the Eptrt-NEP, Eptrt-NEPh, or Eptrt-hNEP plasmids, respectively, in 50 ml of DYT medium (16 g tryptone, 10 g yeast extract, 5 g NaCl per 1 liter) containing ampicillin (100 μg/ml) as described earlier [8]. The proteins were purified by affinity chromatography on Ni-NTA-agarose (Invitrogen, Germany) as described in [8] using buffer A (20 mM Hepes-Na, pH 7.2, 0.5 M NaCl).

Protein preparations were analyzed by Laemmli's SDS-PAGE in 15% separating polyacrylamide gel under denaturing conditions [14]. The gels were stained with Coomassie G-250; PageRuler Prestained Protein Ladder (Thermo Scientific, USA) was used as protein molecular weight markers.

Ultrafiltration was carried out using Ultracel YM-30, YM-50, and YM-100 membrane filters (Millipore, USA) with molecular weight cut-off limits of 30, 50, and 100 kDa, respectively.

Dynamic light scattering (DLS). Particle size distribution in the protein preparations was estimated by the method of Jachimska et al. [15] with slight modifications.

Sample preparation. The kinetics of protein aggregate accumulation was investigated within the first 24 h of protein incubation at room temperature in the protein preparations purified by affinity chromatography on a Ni-NTA-agarose in buffer A containing 0.3 M imidazole. The same protein preparations were used in the studies of the effect of protein concentration (0.3-1.5 mg/ml) on the aggregation process (Table 1). In order to transfer the protein into another buffer, an aliquot of the original solution containing 0.6 mg of the protein was mixed with an equal volume of saturated $(NH_4)_2SO_4$ solution (room temperature). The formed precipitate was separated by centrifugation, dissolved in 400 μl of the target buffer (Table 2), and passed through a nitrocellulose filter with a pore diameter of 0.22 μm. The resulting protein solution was placed into a ZEN0112 plastic cuvette with an optical path length of 1 cm and incubated for 1 min at 25°C.

DLS measurements were carried out with a Zetasizer Nano ZS instrument (Malvern Instruments Ltd., UK) equipped with a 4-mW He-Ne-laser (wavelength, 633 nm) at 25°C. Scattered light was recorded at a fixed angle of 173°. The values of refractive index and viscosity were esti-

Table 1. Hydrodynamic radii of NEP-C and NEP-N aggregates

Incubation conditions		R_h, nm			PdI
		fraction I	fraction II	fraction III	
NEP-C					
50 mM Hepes (pH 7.2), 0.5 M NaCl, 0.3 M imidazole, 0.05% Triton X-100	0 h	5.2 ± 0.2	37 ± 6	2140 ± 90	0.402 ± 0.027
	24 h	4.2 ± 0.2	39 ± 2	2200 ± 150	0.43 ± 0.06
NEP-N					
50 mM Hepes (pH 7.2), 0.5 M NaCl, 0.3 M imidazole, 0.05% Triton X-100	0 h	5.4 ± 0.2	90 ± 10	2450 ± 180	0.32 ± 0.08
	24 h	6.0 ± 0.5	50 ± 20	2320 ± 120	0.42 ± 0.14

Table 2. Effect of medium factors on hydrodynamic radii of NEP-C aggregates

Incubation conditions	R_h, nm			PdI
	fraction I	fraction II	fraction III	
20 mM Hepes (pH 7.2), 0.5 M NaCl (protein concentration 0.3 mg/ml)	5.3 ± 0.7	34.2 ± 13.0	2500 ± 120	0.48 ± 0.19
20 mM Hepes (pH 7.2), 0.5 M NaCl (protein concentration 0.75 mg/ml)	5.2 ± 0.3	32 ± 4 180 ± 70	2560 ± 170	0.47 ± 0.19
20 mM Hepes (pH 7.2), 0.5 M NaCl (protein concentration 1.5 mg/ml)	4.6 ± 0.4	24 ± 5 200 ± 60	2480 ± 200	0.46 ± 0.14
Citric acid/0.1 M sodium citrate (pH 4.2)	–	710 ± 160	–	0.53 ± 0.08
Citric acid/0.1 M sodium citrate (pH 5.5)	7.8 ± 0.6	83 ± 6	2400 ± 200	0.56 ± 0.05
20 mM Hepes (pH 7.2)	4.6 ± 1.2 8.6 ± 0.8	97 ± 7	2300 ± 160	0.75 ± 0.17
Na_2CO_3/$NaHCO_3$ (pH 10.8)	3.8 ± 0.9 7.8 ± 1.3	104 ± 6	–	0.668 ± 0.005
20 mM Hepes (pH 7.2), 0.1 M NaCl	5.1 ± 1.1	107 ± 8	2470 ± 80	0.88 ± 0.18
20 mM Hepes (pH 7.2), 0.5 M NaCl	6.9 ± 0.9	118 ± 9	2270 ± 150	0.77 ± 0.16
20 mM Hepes (pH 7.2), 1 M NaCl	7.5 ± 0.6	54 ± 3	2200 ± 150	0.554 ± 0.014
4 M urea in water	3.4 ± 0.5 13 ± 5	160 ± 30	–	0.76 ± 0.17
6 M GuHCl in water	–	92 ± 3	2340 ± 170	0.344 ± 0.016
1% SDS in water	5.4 ± 0.5	81 ± 9	2408 ± 150	0.60 ± 0.12
1% SDS in water, 90°C, 10 min	2.9 ± 0.6 16 ± 4	110 ± 30	2330 ± 150	0.35 ± 0.13
5% SDS in water	2.0 ± 0.2	33.6 ± 2.8	2320 ± 320	0.63 ± 0.05
1% CHAPS in water	–	80 ± 5	–	0.276 ± 0.013
20 mM Hepes (pH 7.2), 25% acetonitrile	7.5 ± 0.9	50 ± 6	2490 ± 80	0.31 ± 0.06
20 mM Hepes (pH 7.2), 25% ethanol	20.0 ± 2.3	118 ± 7	2260 ± 90	0.276 ± 0.015
100 mM ArgH(H_2PO_4) in water	6.8 ± 0.6	100 ± 4	–	0.610 ± 0.004
20 mM Hepes (pH 7.2), 4 M urea, 25% acetonitrile (mixture 1)	17 ± 3	90 ± 6	–	0.265 ± 0.009
20 mM Hepes (pH 7.2), 4 M urea, 1 M NaCl (mixture 2)	8.0 ± 0.7	54.2 ± 2.5	2230 ± 150	0.55 ± 0.03
20 mM Hepes (pH 7.2), 1 M NaCl, 25% acetonitrile (mixture 3)	–	37 ± 11 400 ± 100	2580 ± 80	0.49 ± 0.08
20 mM Hepes (pH 7.2), 4 M urea, 1 M NaCl, 25% acetonitrile (mixture 4)	–	370 ± 90	2620 ± 140	0.30 ± 0.06

mated for each protein solution. Each measurement included six readings for 10 s; a total of 50 measurements were performed. The correlation functions were analyzed using the Dispersion Technology Software (DTS) ver. 5.10.

Atomic force microscopy (AFM). Nanoparticle morphology in the protein preparations was examined by AFM according to a standard protocol [16] with optimization of sample preparation conditions and the following analysis.

Prior to the AFM experiments, protein aliquots (concentration, 0.1-0.3 mg/ml) were diluted 10-40-fold with deionized water. Next, the protein solution (1 μl) was applied onto a freshly cleaved mica surface for 1-3 min. The surface was dried with an air flow, washed by applying 100 μl of water, and dried again with an air flow.

AFM examination was carried out with a Nanoscope IIIa multimode atomic force microscope (Digital Instruments, USA) in the intermittent contact mode in air. Silicon HA_NC cantilevers (spring constants, 11.5 and 5.3 N/m; typical resonance frequencies, 254 and 152 kHz; TipsNano, Russia) were used for scanning. The scanning rate was 2 Hz with a raster scan at 512 × 512 pixels. The images were processed with the Femtoscan program (Advanced Technologies Center, Russia).

Prediction of regions facilitating protein aggregation. The following computer programs were used to search the NEP, NEP-N, and NEP-C structures for the presence of regions facilitating protein aggregation: TANGO [17], Aggrescan [18], FoldAmyloid [19], Pasta 2.0 [20], Waltz [21], AmyloidMutants [22], FISH Amyloid [23], GAP [24], and Met Amyl [25].

Computer modeling. Reconstruction of the *N*-terminal domain structure was performed using the I-Tasser program for predicting 3D structures of proteins based on their amino acid sequences [26].

Computer modeling was performed for NEP, NEP-C, NEP-N, and *N*-terminal (1-60 a.a.) and *C*-terminal (60-121 a.a.) fragments of NEP by molecular dynamic methods using the Gromacs 5.0.4 program [27] and the AMBER99sb-ILDN force field [28]. Each system contained 50 protein molecules randomly distributed in a 27,000-nm^3 volume, so that the minimal distance between any two molecules was no less than 1.5 nm, which exceeded the cut-off radius of Coulomb and van-der Waals forces in modeling. Next, water was added to the system (the three-site water model TIP3P) together with Na^+ and Cl^- ions at a concentration of 100 mM. All systems were subjected to energy minimization, first, according to the fastest descent algorithm [29], and then with the conjugate gradient method [30] followed by equilibration of the system for 100 ps in the NVT and NPT ensembles (canonical and isobaric-isothermal, respectively). The time of productive modeling was 150 ns for NEP, NEP-C, and NEP-N and 70 and 80 ns for the *N*- and *C*-terminal protein fragments, respectively.

RESULTS AND DISCUSSION

Influenza A virus NEP rapidly aggregates in solutions. Earlier, we have demonstrated its tendency for aggregation using cross-linking with glutaraldehyde [8]. In order to investigate NEP aggregation in more detail, we produced NEP recombinant variants using the previously constructed system for heterologous expression in bacterial cells [8]. The recombinant proteins carried the $(His)_6$-tag at either *N*- or *C*-termini; their structure was similar to that of the native NEP. Analysis of freshly purified protein preparations by SDS-PAGE demonstrated their homogeneity (Fig. 1a). However, no low-molecular-weight components corresponding to the protein monomers were detected in the filtrates obtained by ultrafiltration of freshly purified NEP-C or NEP-N through the membrane filters with the cut-off limits of 30 and 50 kDa. Therefore, we suggested that the proteins form oligomeric particles in solution within the first minutes after their purification. Incubation of the protein preparations at 4°C for several days or months resulted in accumulation of high-molecular-weight aggregates that could be detected by SDS-PAGE as Coomassie blue-stained zones at the start of the gel or at the boundary between the concentrating and separating gels (Fig. 1b). According to the literature data [13, 31], such resistance to the denaturing conditions is typical of highly structured aggregates including fibrillar (amyloid) aggregates.

We also studied the process of aggregate formation in NEP-C and NEP-N solutions under various conditions using DLS that allows estimating not only the size of aggregates but also their relative content in the mixture. Both proteins formed aggregates in buffer A containing 0.3 M imidazole with three main fractions (Table 1): I – low-molecular-weight (R_h, 2-20 nm); II – medium-molecular-weight (R_h, 20-700 nm); and III – high-molecular-weight (R_h, 2000-3000 nm). All three fractions were characterized with high polydispersity index (PdI > 0.1), which indicates their heterogeneous character. Incubation of protein solutions for more than 20 days often resulted in the formation of insoluble precipitate.

Monitoring for 24 h of the light scattering by the aggregated particles in the solutions of purified NEP-N and NEP-C revealed a decrease in the amount of type I aggregates and increase in the amount of type II aggregates (Table 1 and Fig. 2); both processes were more pronounced for NEP-C. We suggested that type II aggregates most likely formed from type I aggregates. The mass fraction of type III aggregates was extremely low in both cases. It must be noted that the introduction of an additional procedure of protein precipitation with ammonium sulfate followed by dissolving the precipitate in different buffer systems did not significantly affect the size distribution of the aggregates; the presence of the above-mentioned three fractions was observed in virtually all cases.

Many factors can affect protein aggregation including protein concentration, solution pH and ionic strength, temperature, and presence of various low-molecular-weight additives [31, 32]. To identify conditions that prevent or, on the contrary, facilitate the self-assembly of NEP, we used NEP-C, as this protein demonstrated higher tendency for aggregation. It could also be purified by affinity chromatography in a homogenous state with a high yield. According to the literature data [5], as well as

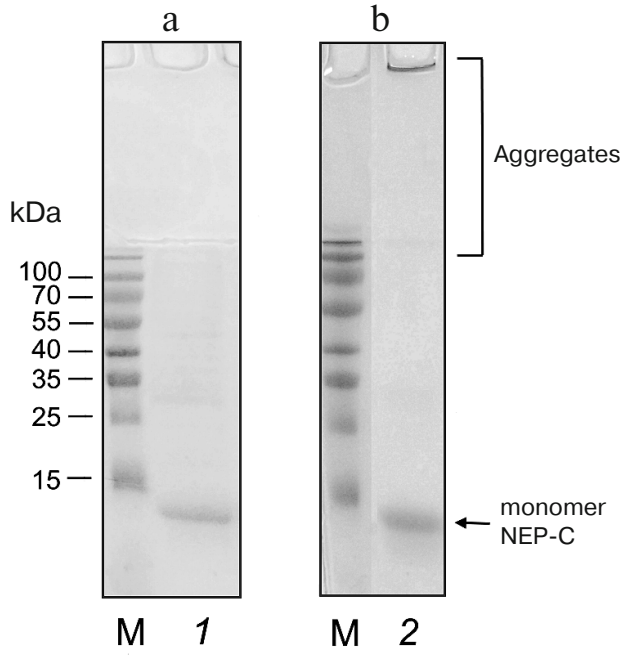

Fig. 1. SDS-PAGE of NEP-C preparations: *1*) freshly isolated; *2*) after incubation for 180 days at 25°C. M, protein molecular weight markers.

based on the results of computer modeling of native NEP and its $(His)_6$-tagged variants (see below), the *N*- and *C*-terminal NEP fragments do not affect the formation of the native protein 3D structure and mostly do not participate in the intermolecular interactions. Therefore, we suggested that NEP-C would serve as a good model that adequately represents the properties of native NEP.

Varying the NEP-C concentration in the range of 0.3-1.5 mg/ml did not noticeably affect the character and the rate of protein aggregation (Table 2). However, further increase in its concentration (>3 mg/ml) resulted in partial protein precipitation, which could have distorted quantitative characteristics of the aggregation process and resulted in the emergence of multiple scattering and fake signals in the DLS spectra. For this reason, in all further experiments on the effect of various factors on protein aggregation, the protein concentration in a solution was 1.5 mg/ml.

Studying the effect of pH on NEP-C aggregation revealed that at pH values close to the isoelectric point (p$I \approx 5.5$), the protein was relatively stable and did not precipitate (although many proteins exhibit this property [33]). The DLS profile changed only slightly at pH values above pI (5.5-10.8). The fractions I and II were predominant at these pH value, but disappeared from the solution at acidic pH (pH 4.2), when the higher-molecular-weight aggregates with $R_h \approx 700$-1000 nm emerged (Table 2 and Fig. 3a). This pH-dependence indicates involvement of electrostatic interactions (between the charged groups of amino acid residues) in the aggregate formation. However, most likely, electrostatic interactions are not the key type of interactions, because changes in the solution ionic strength within the 0-1 M NaCl range did not significantly affect the particle size distribution, although an increase in the NaCl concentration to 1 M was accompanied by a decrease in the aggregate size in fraction II and increase in the relative content of fraction I (Table 2).

Next, we investigated the effects of low-molecular-weight additives of various nature on protein aggregation. In the majority of experiments, pH was maintained at 7.2. The use of chaotropic agents such as 4 M urea and 6 M guanidine hydrochloride caused significant decrease in the fraction I content and slight increase in the fraction II content of (Table 2 and Fig. 3b). Furthermore, high-

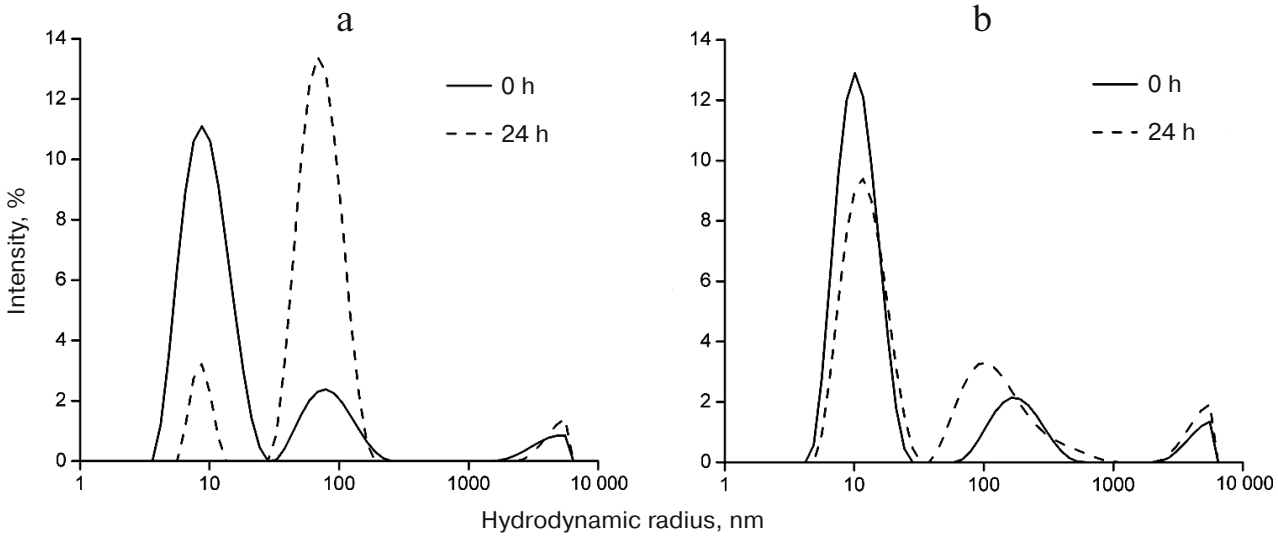

Fig. 2. DLS analysis of particle size distribution in the NEP-C (a) and NEP-N (b) solutions at different incubation times.

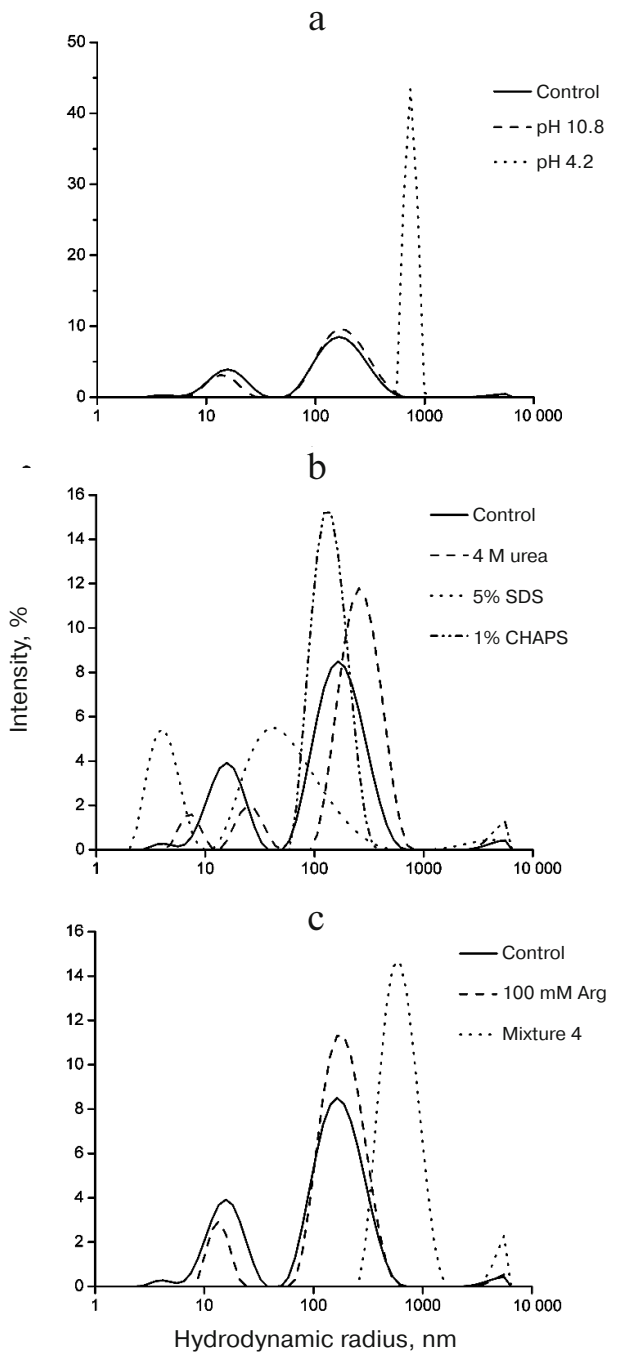

Fig. 3. DLS analysis of particle size distribution in the NEP-C solution at different pH values (a) and in the presence of low-molecular-weight additives (b) or their mixture (c). Control measurements were carried out in buffer A.

molecular-weight aggregates from fraction III were absent in the presence of 4 M urea. Much more pronounced effects were observed in the presence of detergents (Table 2 and Fig. 3b). For example, 5% SDS (ionic denaturing detergent) caused a significant decrease in both the size and the total number of particles in the mixture. In the case of zwitterionic non-denaturing detergent CHAPS, fractions I and III were completely absent, and the R_h of particles in fraction II decreases slightly. Taken together, these data indicate that hydrogen bonds and hydrophobic interaction play an important role in the formation of intermolecular contacts (especially, in fraction I). Similar results were obtained for the NEP-C solutions containing organic solvents such as acetonitrile (25%) and ethanol (25%) (Table 2) that also changed the DLS profiles most likely due to the disruption of hydrophobic interactions.

All further attempts to discover conditions that could prevent NEP aggregation were made taking into consideration the existence of multicenter intermolecular interactions of various nature in the aggregates. The amino acid arginine (L-Arg) in a form of its salts has been widely used for protein disaggregation [34, 35]. The exact mechanism of its action has not been yet elucidated; however, due to its structure, arginine can affect virtually all types of interactions and is often used as a stabilizer of protein solutions. However, we found that addition of arginine phosphate (10-100 mM) to the buffer solutions at different steps of protein purification and storage did not noticeably affect formation of fractions I and II, although resulted in the absence of the high-molecular-weight fraction III (Table 2 and Fig. 3c).

We also used various combinations of compounds investigated in this work for NEP-C disaggregation (Table 2). In all the cases, NEP formed the aggregates, although certain additivity in the actions of different factors was observed. For example, fraction I was absent in the case of mixture 4 (4 M urea + 1 M NaCl + 25% acetonitrile) (Fig. 3c), which suggests that the disaggregation conditions could be found by rational combination of different factors.

Based on the obtained results, we concluded that the most unstable fraction is fraction I that is likely composed of oligomeric precursors of larger aggregates. The high stability of fraction II indicates that the aggregates in this fraction are structured and not amorphous (amorphous aggregates are usually sensitive to the action of chaotropic agents and standard SDS-type detergents). Their resistance to disaggregating agents might be due to the fact that the interacting groups are hidden inside the highly organized structure. Since amyloid fibrils belong to this type of aggregates [13, 31], we believed it was important to investigate the particle morphology in the NEP solution.

Atomic force microscopy (AFM) was used to analyze the morphology of formed aggregates and to obtain additional data on the size of particles formed in the NEP, NEP-C, and NEP-N solutions (Fig. 4). The obtained AFM images of the particles in protein preparations show that NEP molecules adsorbed on the mica surface formed globules with a height of 1.2 ± 0.2 nm (n = 1000) (Fig. 4a, magnified image with several protein monomers is presented in the inset). The particles with a

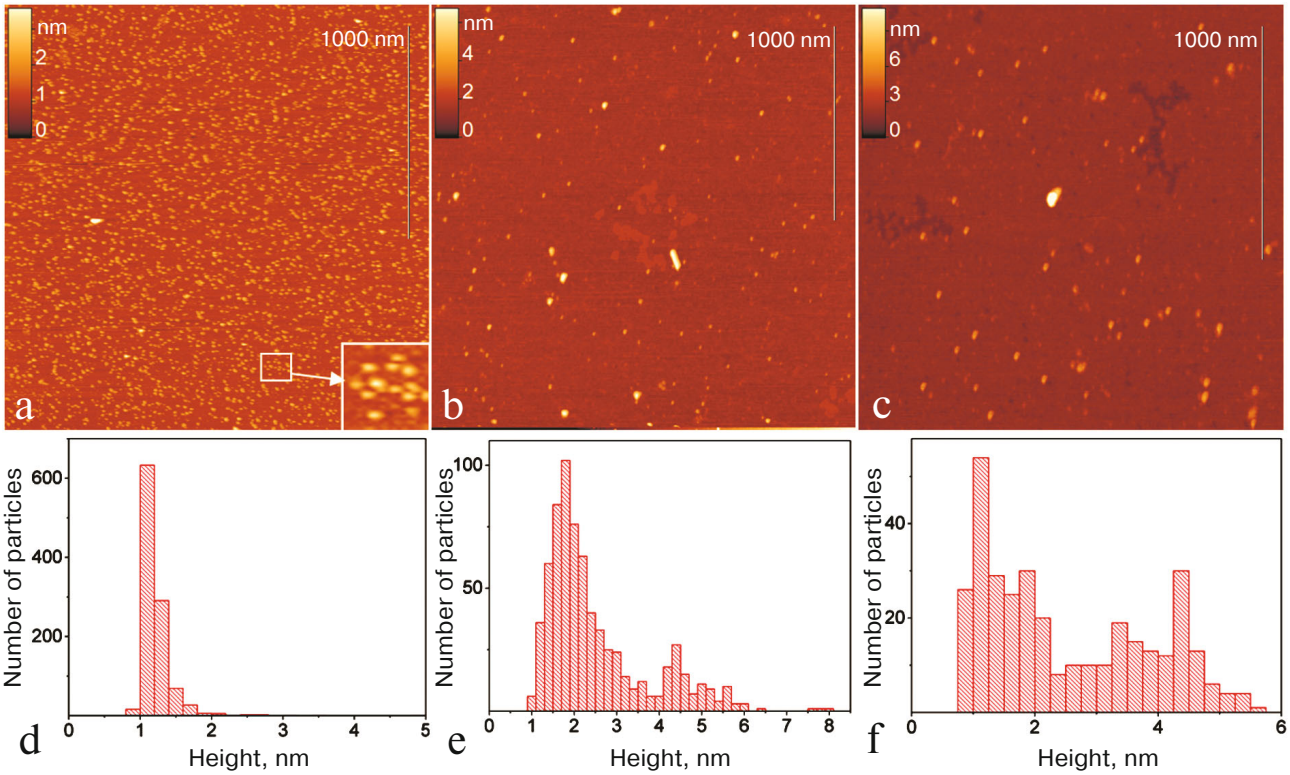

Fig. 4. AFM images of NEP (a), NEP-C (b), and NEP-N (c) preparations; (d-f) corresponding height distribution diagrams. Images size; 2×2 μm^2; inset size in image (a), 100×100 nm^2.

height up to 8 nm were also present (see histograms of height distribution in Fig. 4, d-f), that most likely represented protein aggregates. Based on geometric estimates, these aggregates consisted of tens and hundreds of protein molecules and could correspond to aggregates from fraction I with $R_h \approx 5$ nm found in the DLS experiments. A small fraction of elongated particles was found in the solutions; their number was higher in the NEP-C preparation (Fig. 4b). It can be assumed that these particles are the amyloid-type aggregates.

A number of computer programs allowing identification of the so-called amyloidogenic fragments in a protein structure were used in order to test the possibility of the amyloid fibril formation by NEP and to establish structural elements that can participate in this process (Fig. 5). The results of analysis revealed that fragments 4-16 and 29-43 in the *N*-terminal domain and fragments 67-82, 93-109, and 114-121 in the *C*-terminal domain could be such structural elements. It must be mentioned that the *C*-terminal domain containing clearly pronounced hydrophilic (55-75 and 80-95) and hydrophobic (98-112) regions is the most amyloidogenic.

In order to identify regions in the protein structure that could participate in aggregation, computer modeling of the aggregation process was performed for five protein structures (full-size NEP, NEP-C, and NEP-N, and *N*- and *C*-terminal fragments of NEP). NEP is represented in the protein 3D structure databases by only one structure (PDB ID: 1PD3 [5]; a.a. 63-116). The structure of the protein missing part was reconstructed using the I-Tasser program [26] that allows predicting 3D structure of a protein from its amino acid sequence. This portion of the NEP molecule consists of two parallel α-helices similar in structure to the *C*-terminal fragment. NEP molecule has an elongated rod-like shape formed by four α-helices (Fig. 6) similar to that described by Darapaneni et al. [7]. The protein size is $(5.2 \pm 0.3) \times (2.6 \pm 0.2)$ nm following 150 ns of modeling (number of measurements $n = 20$).

All protein variants except the *N*-terminal fragment demonstrated high aggregation rates already at the first modeling stages. Most molecules in these systems (~50 monomers per system) formed clusters (2-4 monomers each) after 50 ns with only insignificant number of single molecules. The next adhesion stage included cluster aggregation into larger conglomerates (only 3-5 per system) each containing 10-20 monomers. No clear regularities in the cluster formation were observed. In the case of the *N*-terminal fragment (a.a. 1-60), the total number of small clusters (only di- and trimers) was significantly larger (20 after 50 ns and 24 after 70 ns). These data provide an additional support to the suggestion that the *C*-terminal region offers the largest contribution to the aggregation properties of NEP.

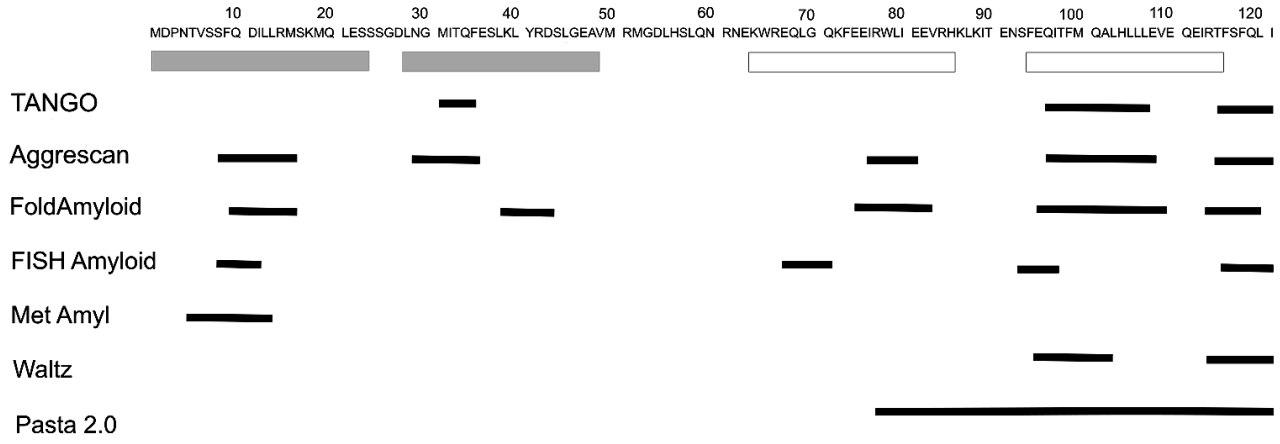

Fig. 5. NEP fragments facilitating protein aggregation (black rectangles) revealed with the help of computer programs (shown on the left). Grey and white rectangles designate regions involved in the formation of α-helices in the *N*- and *C*-domains of NEP, respectively.

In order to identify protein regions responsible for the aggregation, we built the histograms showing the frequency of amino acid residues in the contact sites. For this purpose, we developed an algorithm that used the coordinate file describing the final state of the protein ensemble to identify all amino acid residues from different molecules located at a distance less than 5 Å from each other (Fig. 7). The peaks in these histograms include amino acids most frequently found at the contact sites and correspond to regions 1-6, 50-70, 85-105, and 115-126, all located at the ends of the protein 3D model (Fig. 8). Residues 85-105 and 25-30 (peak is not visible on the histograms) are located at one end of the rod-like molecule, while residues 1-6, 50-70, and 115-126 – at the other end of the molecule. This suggests that the molecules in the clusters interact predominately via their ends. Different variants of such interactions are presented in Fig. 8. Amino acid residues forming ionic bonds (e.g., Asp, Glu, Lys) are prevalent in the 50-70 region; amino acid residues participating in hydrophobic interactions (mainly Leu and Phe) are found in the 85-105 region that also contains a certain number of charged residues (Lys, Glu); the 115-126 region contains hydrophobic amino acids (Phe, Leu).

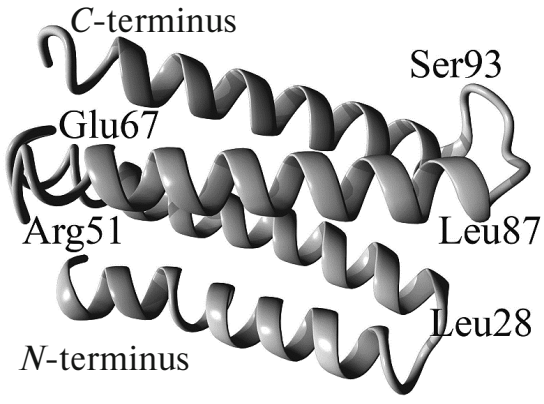

Fig. 6. 3D structure of NEP molecule reconstructed with the I-Tasser program.

It can be seen in Figs. 5 and 7 that the results of predictions of the amyloidogenic sites and regions responsible for protein aggregations correlate. These data are also in agreement with the abovementioned effects observed in the DLS experiments that indicated participation of electrostatic and hydrophobic interactions in the aggregate formation.

As was mentioned earlier, formation of NEP precipitates (aggregates) *in vitro* was observed during purification of the recombinant protein [6]. However, it was also reported that some NEP variants (e.g., in the case of influenza A/California/04/2009 virus) can form aggregates with predominant localization in the cell nucleus *in vivo* [36]. The biological significance of NEP aggregation has not been established. Taking into consideration that NEP is a multifunctional protein interacting with many viral and cellular components [37], it can be suggested that its aggregation could affect different stages of the virus life cycle. For example, it has been demonstrated in a number of studies that NEP regulates transcription and replication of viral DNA by modulating the activity of viral RNA polymerase [38-40]. Moreover, NEP activates RNA polymerase at low expression levels and inhibits it at high expression levels [41]. It is possible that regulation of RNA synthesis occurs via changes in the NEP concentration resulting from the protein ability for transition into its insoluble form. Aggregation might facilitate removal of the excessive NEP thereby creating a reserve of protein molecules that can serve as a source of soluble monomers if required.

It cannot be ruled out that formation of NEP aggregates in cells could occur as a result of protein interaction with various polymeric structures. In particular, the main

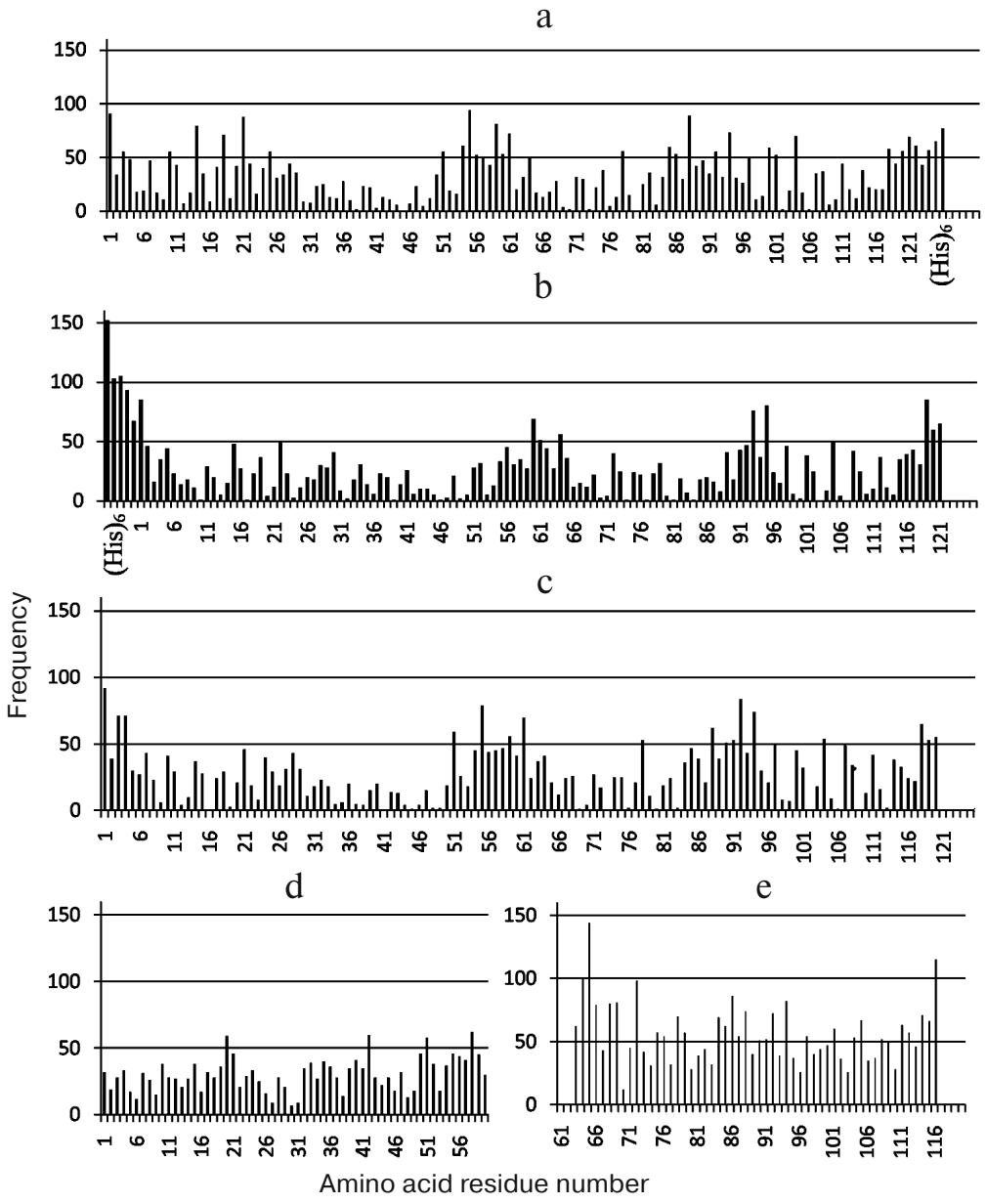

Fig. 7. Frequency of amino acid residues at the contact sites between molecules of the NEP variants: a) native NEP after 150 ns; b) NEP-N after 150 ns; c) NEP-C after 150 ns; d) *N*-terminal fragment of NEP (a.a. 1-60) after 70 ns; e) *C*-terminal fragment of NEP (a.a. 60-115) after 80 ns of modeling.

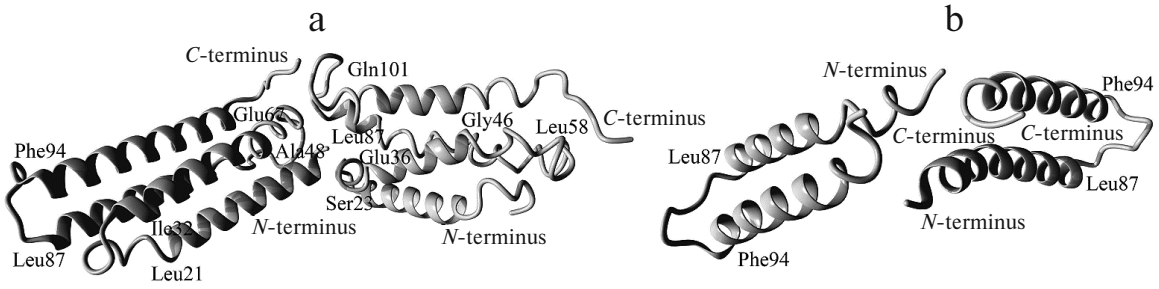

Fig. 8. Variants of interactions within the NEP-C cluster after 150 ns of modeling (a) and within the protein cluster of the NEP *C*-terminal fragment molecules after 80 ns of modeling (b).

function of NEP is the transport of viral ribonucleoprotein complex (vRNP) from the nucleus to the cytoplasm [37]. Some components of the export complex (e.g., M1 and NP proteins) [41, 42] are capable of oligo- and polymerization and might stimulate NEP aggregation by direct interaction with this protein. Moreover, transport of the vRNP complex to the cell membrane, where the assembly of the viral particles occurs, proceeds with the participation of microfilaments of the cell cytoskeleton that could also serve as a template for protein aggregation [41].

Another suggestion on the possible functional significance of NEP aggregation is related to the role of this protein in the process of virion budding [4]. It was established that NEP interacts with the cellular F_0F_1-ATPase, which is located near the lipid rafts and participates in the change of cell membrane curvature required for the following budding of the viral particle [4, 43]. It is likely that when the NEP concentration in the vicinity of the cell membrane is elevated, the protein can aggregate with the formation of rigid fibrillar amyloid-type structures that disrupt the membrane integrity and induce virion budding (as it has been shown for other amyloids) [44, 45]. However, all these hypotheses require experimental validation.

In conclusion, aggregation properties of recombinant NEP (both native and His_6-tagged) were studied using the DLS and AFM methods. It was shown that the protein can form structured aggregates under a wide range of *in vitro* condition. The effect of various external factors indicates that NEP aggregation is ensured by different types of molecular interaction, which was confirmed by modeling of aggregation process using molecular dynamics techniques. Aggregation is an inherent property of NEP that is likely required for its functioning *in vivo*.

Funding

This work was supported by the Russian Foundation for Basic Research (grants No. 16-04-00563 — preparation of proteins and study of protein aggregation using DLS method; mol_a No. 18-34-00623 — analysis of the morphology of nanoparticles in protein preparations using AFM method).

Acknowledgements

The authors are grateful to the Center for Collective Use of Supercomputers, Lomonosov Moscow State University, for providing equipment for this work.

Conflict of Interests

The authors declare no conflict of interests.

Compliance with Ethical Standards

This study does not include any experiments with the participation of humans or animals as objects of study.

REFERENCES

1. Manz, B., Schwemmle, M., and Brunotte, L. (2013) Adaptation of avian Influenza A virus polymerase in mammals to overcome the host species barrier, *J. Virol.*, **87**, 7200-7209.
2. Brunotte, L., Flies, J., and Bolte, H. (2014) The nuclear export protein of H5N1 influenza A viruses recruits matrix 1 (M1) protein to the viral ribonucleoprotein to mediate nuclear export, *J. Biol. Chem.*, **289**, 20067-20077.
3. Watanabe, K., Shimizu, T., Noda, S., Tsukahara, F., Maru, Y., and Kobayashi, N. (2014) Nuclear export of the influenza virus ribonucleoprotein complex: interaction of Hsc70 with viral proteins M1 and NS2, *FEBS Open Bio*, **4**, 683-688.
4. Gorai, T., Goto, H., Noda, T., Watanabe, T., Kozuka-Hata, H., Oyama, M., Takano, R., Neumann, G., Watanabe, S., and Kawaoka, Y. (2012) F_1F_0-ATPase, F-type proton-translocating ATPase, at the plasma membrane is critical for efficient influenza virus budding, *Proc. Natl. Acad. Sci. USA*, **109**, 4615-4620.
5. Akarsu, H., Burmeister, W. P., Petosa, C., Petit, I., Muller, C. W., Ruigrok, R. W., and Baudin, F. (2003) Crystal structure of the M1 protein-binding domain of the influenza A virus nuclear export protein (NEP/NS2), *EMBO J.*, **22**, 4646-4655.
6. Lommer, B. S., and Luo, M. (2002) Structural plasticity in influenza virus protein NS2 (NEP), *J. Biol. Chem.*, **277**, 7108-7117.
7. Darapaneni, V., Prabhaker, V. K., and 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.
8. Golovko, A. O., Koroleva, O. N., and Drutsa, V. L. (2017) Heterologous expression and isolation of influenza A virus nuclear export protein NEP, *Biochemistry (Moscow)*, **82**, 1529-1537.
9. Gomez-Puertas, P., Albo, C., Perez-Pastrana, E., Vivo, A., and Portela, A. (2000) Influenza virus matrix protein is the major driving force in virus budding, *J. Virol.*, **74**, 11538-11547.
10. Rossman, J. S., and Lamb, R. A. (2011) Influenza virus assembly and budding, *Virology*, **411**, 229-236.
11. Calder, L. J., Wasilewski, S., Berriman, J. A., and Rosenthal, P. B. (2010) Structural organization of a filamentous influenza A virus, *Proc. Natl. Acad. Sci. USA*, **107**, 10685-10690.
12. Vidic, J., Richard, C. A., Pechoux, C., Da Costa, B., Bertho, N., Mazerat, S., Delmas, B., and Chevalier, C. (2016) Amyloid assemblies of influenza A virus PB1-F2 protein damage membrane and induce cytotoxicity, *J. Biol. Chem.*, **291**, 739-751.
13 Uversky, V. N. (2008) Amyloidogenesis of natively unfolded proteins, *Curr. Alzheimer Res.*, **5**, 260-287.
14. Laemmli, U. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature*, **227**, 680-685.

15. Jachimska, B., Wasilewska, M., and Adamczyk, Z. (2008) Characterization of globular protein solutions by dynamic light scattering, electrophoretic mobility, and viscosity measurements, *Langmuir*, **24**, 6866-6872.
16. Muller, D. J., Janovjak, H., Lehto, T., Kuerschner, L., and Anderson, K. (2002) Observing structure, function and assembly of single proteins by AFM, *Prog. Biophys. Mol. Biol.*, **79**, 1-43.
17. Fernandez-Escamilla, A. M., Rousseau, F., Schymkowitz, J., and Serrano, L. (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins, *Nat. Biotechnol.*, **22**, 1302-1306.
18. Conchillo-Sole, O., de Groot, N. S., Aviles, F. X., Vendrell, J., Daura, X., and Ventura, S. (2007) AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides, *BMC Bioinformatics*, **8**, 65.
19. Garbuzynskiy, S. O., Lobanov, M. Y., and Galzitskaya, O. V. (2010) FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence, *Bioinformatics*, **26**, 326-332.
20. Walsh, I., Seno, F., Tosatto, S. C. E., and Trovato, A. (2014) PASTA2: an improved server for protein aggregation prediction, *Nucleic Acids Res.*, **42** (Web Server issue), W301-W307.
21. Maurer-Stroh, S., Debulpaep, M., Kuemmerer, N., Lopez de la Paz, M., Martins, I. C., Reumers, J., Morris, K. L., Copland, A., Serpell, L., Serrano, L., Schymkowitz, J. W., and Rousseau, F. (2010) Exploring the sequence determinants of amyloid structure using position-specific scoring matrices, *Nat. Methods*, **7**, 237-242.
22. O'Donnell, C. W., Waldispuhl, J., Lis, M., Halfmann, R., Devadas, S., Lindquist, S., and Berger, B. (2011) A method for probing the mutational landscape of amyloid structure, *Bioinformatics*, **27**, i34-i42.
23. Gasior, P., and Kotulska, M. (2014) FISH amyloid – a new method for finding amyloidogenic segments in proteins based on site-specific co-occurrence of amino acids, *BMC Bioinformatics*, **15**, 54.
24. Thangakani, A. M., Kumar, S., Nagarajan, R., Velmurugan, D., and Gromiha, M. M. (2014) GAP: towards almost hundred percent prediction of β-strand mediated aggregating peptides with distinct morphologies, *Bioinformatics*, **30**, 1983-1990.
25. Emily, M., Talvas, A., and Delamarche, C. (2013) MetAmyl: a METa-predictor for AMYLoid proteins, *PLoS One*, **8**, e79722.
26. Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015) The I-TASSER suite: protein structure and function prediction, *Nat. Methods*, **12**, 7-8.
27. Abraham, M. J., Murtola, T., Schulz, R., Pall, S., Smith, J. C., Hess, B., and Lindahl, E. (2015) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX*, **1**, 19-25.
28. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins*, **65**, 712-725.
29. Chaichian, M., and Demichev, A. (2001) *Path Integrals in Physics. Volume 1: Stochastic Process and Quantum Mechanics*, Taylor & Francis.
30. Zimmerman, K. (1991) All purpose molecular mechanics simulator and energy minimize, *J. Comp. Chem.*, **12**, 310-319.
31. Khan, M. V., Zakariya, S. M., and Khan, R. H. (2018) Protein folding, misfolding and aggregation: a tale of constructive to destructive assembly, *Int. J. Biol. Macromol.*, **112**, 217-229.
32. Jeong, J. S., Ansaloni, A., Mezzenga, R., Lashuel, H. A., and Dietler, G. (2013) Novel mechanistic insight into the molecular basis of amyloid polymorphism and secondary nucleation during amyloid formation, *J. Mol. Biol.*, **425**, 1765-1781.
33. Nehete, J. Y., Bhambar, R. S., Narkhede, M. R., and Gawali, S. R. (2013) Natural proteins: sources, isolation, characterization and applications, *Pharmacogn. Rev.*, **7**, 107-116.
34. Lange, C., and Rudolph, R. (2009) Suppression of protein aggregation by L-arginine, *Curr. Pharm. Biotechnol.*, **10**, 408-414.
35. Shukla, D., and Trout, B. L. (2010) Interaction of arginine with proteins and the mechanism by which it inhibits aggregation, *J. Phys. Chem. B*, **114**, 13426-13438.
36. Gao, S., Wang, S., Cao, S., Sun, L., Li, J., Bi, Y., Gao, G. F., and Liu, W. (2014) Characteristics of nucleocytoplasmic transport of H1N1 influenza A virus nuclear export protein, *J. Virol.*, **88**, 7455-7463.
37. Paterson, D., and Fodor, E. (2012) Emerging roles for the influenza A virus nuclear export protein (NEP), *PLoS Pathog.*, **8**, e1003019.
38. Robb, N. C., Smith, M., Vreede, F. T., and Fodor, E. (2009) NS2/NEP protein regulates transcription and replication of the influenza virus RNA genome, *J. Gen. Virol.*, **90**, 1398-1407.
39. Manz, B., Schwemmle, M., and Brunotte, L. (2013) Adaptation of avian influenza A virus polymerase in mammals to overcome the host species barrier, *J. Virol.*, **87**, 7200-7209.
40. Reuther, P., Giese, S., Gotz, V., Kilb, N., Manz, B., Brunotte, L., and Schwemmle, M. (2014) Adaptive mutations in the nuclear export protein of human-derived H5N1 strains facilitate a polymerase activity-enhancing conformation, *J. Virol.*, **88**, 263-271.
41. Smirnova, T. D., Danilenko, D. M., and Slita, A. V. (2013) Role of cellular cytoskeleton in influenza A infection cycle, *Tsitologiya*, **55**, 92-100.
42. Calder, L. J., Wasilewski, S., Berriman, J. A., and Rosenthal, P. B. (2010) Structural organization of a filamentous influenza A virus, *Proc. Natl. Acad. Sci. USA*, **107**, 10685-10690.
43. Martyna, A., and Rossman, J. (2014) Alterations of membrane curvature during influenza virus budding, *J. Biochem. Soc. Trans.*, **42**, 1425-1428.
44. Terakawa, M. S., Lin, Y., Kinoshita, M., Kanemura, S., Itoh, D., Sugiki, T., Okumura, M., Ramamoorthy, A., and Lee, Y. H. (2018) Impact of membrane curvature on amyloid aggregation, *Biochim. Biophys. Acta*, **1860**, 1741-1764.
45. Milanesi, L., Sheynis, T., Xue, W. F., Orlova, E. V., Hellewell, A. L., Jelinek, R., Hewitt, E. W., Radford, S. E., and Saibil, H. R. (2012) Direct three-dimensional visualization of membrane disruption by amyloid fibrils, *Proc. Natl. Acad. Sci. USA*, **109**, 20455-22460.