



Peptide Models of the Cytoplasmic Tail of Influenza A/H1N1 Virus Hemagglutinin Expand Understanding its pH-Dependent Modes of Interaction with Matrix Protein M1

Victor Vitoldovich Poboinev¹ · Vladislav Victorovich Khrustalev¹ · Anastasia Aleksandrovna Akunevich¹ · Nikolai Vladimirovich Shalygo¹ · Aleksander Nikolaevich Stojarov² · Tatyana Aleksandrovna Khrustaleva³ · Larisa Valentinovna Kordyukova⁴

Accepted: 27 February 2023 / Published online: 23 March 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Influenza A virus hemagglutinin (HA) is a major virus antigen. No cryo-electron microscopy or X-ray data can be obtained for the HA intraviral (cytoplasmic) domain (CT) post-translationally modified with long fatty acid residues bound to three highly conserved cysteines. We recently proposed a model of HA CT of Influenza A/H1N1 virus possessing an antiparallel beta structure based on the experimental secondary structure analysis of four 14–15 amino acid long synthetic peptides, corresponding to the HA CT sequence, with free or acetaminomethylated cysteines. To dispel doubts about possible non-specific “amyloid-like” aggregation of those synthetic peptides in phosphate buffer solution, we have determined the order of oligomers based on blue native gel electrophoresis, membrane filtration, fluorescence spectroscopy and molecular modeling approaches. We have found that unmodified peptides form only low molecular weight oligomers, while modified peptides form both oligomers of low order similar to those found for unmodified peptides and high order conglomerates, which however are not of beta-amyloid-like fold. This study confirms that the beta structure previously detected by circular dichroism spectroscopy analysis is more likely the result of intrinsic propensity of the HA CT amino acid sequence than the consequence of aggregation. The structures of low order oligomers of the synthetic peptides were used for *in silico* experiments on modeling of HA CT interactions with matrix protein M1 at physiological and acidic pH levels and revealed two different areas of binding. Finally, tripeptides capable of blocking interactions between HA CT and M1 were proposed.

Keywords Blue native gel electrophoresis · Cytoplasmic tail of hemagglutinin · Macromolecular docking · Tryptophan fluorescence · M1 protein · Tripeptides

1 Introduction

Influenza A virus hemagglutinin (HA) is a major virus antigen that forms trimers within a viral envelope. In the Protein Data Bank [1] there are no 3D-structures obtained for the cytoplasmic (intraviral) domain (the HA CT region) of any HA type/subtype using X-ray crystallography or cryo-electron microscopy analysis. This fragment is crucial at different stages of the virus life cycle. At the entry step, the HA CT participates in the disassembly of the viral envelope and the membrane fusion reaction when the pH level in the endolysosome decreases. During virus assembly at the plasma membrane it interacts with a membrane associated matrix protein M1 to form functionally competent virions. The relevant 3D models of the HA CT region may be further used in drug design studies.

✉ Victor Vitoldovich Poboinev
dremozzew@mail.ru

¹ Department of General Chemistry, Belarusian State Medical University, Dzerzhinskogo 83, Minsk, Belarus

² Department of Radiation Medicine and Ecology, Belarusian State Medical University, Dzerzhinskogo 83, Minsk, Belarus

³ Multidisciplinary Diagnostic Laboratory, Institute of Physiology of the National Academy of Sciences of Belarus, Minsk, Belarus

⁴ Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia

Previously we have offered a 3D model of the HA CT region for Influenza A/H1N1 virus in the form of a peptide [2]. The next exciting step of the study is the modeling of the HA CT interactions with matrix protein M1 at physiological pH of 7.4 in the cellular cytoplasm and at acidic pH of 5.0 in the lysosomes. However, to perform these *in silico* experiments one has to be sure in the relevance of the proposed peptide models for the native viral system.

The M1 is the most abundant protein in influenza virions acting as a hub for both integral proteins of the viral envelope and the nucleocapsid [3, 4]. The M1 protein contains an X-ray resolved global alpha-helical N-terminal (NM-) domain [5–7] and a structurally disordered C-terminal domain [8–10], which has not been crystallized. At neutral pH, the C-terminal domain binds to viral ribonucleoprotein segments (RNPs) and participates in a matrix layer formation [11]. In an acidic environment, it becomes more mobile, making possible the disintegration of the matrix layer and the transition of RNPs from the internal volume of the virion into the cytoplasm of the infected cell [11]. The contacts of the N-terminal domain of M1 with the inner surface of the lipid bilayer [12] and submembrane hemagglutinin domains, which are stable at neutral pH, should also be broken at low pH [4].

Within a virion, the C-terminal region of Influenza A virus hemagglutinin is associated with the lipid membrane via three fatty acid residues, two palmitates (C16:0) and a stearate (C18:0) bound via thioester bonds to specific cysteine residues located in the cytoplasmic tail and at the C-terminus of transmembrane domain, respectively [13]. It is almost impossible to model *in silico* such a complex structure of a lipopeptide with stearic and palmitic acid residues incorporated into the lipid bilayer [14]. That is why previously we have decided to use synthetic peptides corresponding to the last C-terminal 14 or 15 residues of Influenza A/H1N1 virus hemagglutinin to study their secondary structure in solution using circular dichroism (CD) and multi-bounce Horizontal Attenuated Total Reflectance Fourier Transformed Infrared (HATR-FTIR) spectroscopy analysis [2]. To prevent the possibility of disulfide bonds formation we substituted all three cysteine residues in two of those peptides by acetaminomethylcysteine residues [2].

Four peptides (WI14, FI15, WI14-ACM, and FI15-ACM) have been synthesized. The CD spectroscopy analysis of their saturated solutions at pH 7.4 revealed the formation of an antiparallel beta structure in all studied peptides [2]. At the same time, according to the results of fluorescence spectra analysis, the peptides are capable to form oligomers [2]. Various short hydrophobic peptides are known to form aggregates [15]. The subunits of such aggregates are usually organized as an intermolecular beta sheet [15]. Thus, there was a possibility that the studied short peptides corresponding to the HA CT sequence form beta structure in a buffer

solution, while the real cytoplasmic tail of hemagglutinin inside the virion does not. The presence of a beta structure in high molecular weight aggregates of those peptides may be considered non-specific and the models may be inappropriate. However, if the order of oligomers is low, the previously made statement about beta structure formation by the hemagglutinin cytoplasmic tail would not be compromised.

Thus, the first aim of the current study was to investigate the order of oligomers for peptides corresponding to the Influenza A/H1N1 hemagglutinin cytoplasmic tail. And the second aim was to use the obtained models of short peptides to investigate how exactly the HA CT may interact with N-terminal part of M1 protein at pH level of cytoplasm and pH level of lysosome.

The following methods were applied to check the molecular weight of model peptide oligomers in phosphate buffer solution and, correspondingly, the size/order of them: (1) blue native gel electrophoresis [16], (2) membrane centrifugal ultrafiltration [17] under the control of spectrofluorimetry [18], and (3) the protein-protein (actually, peptide-peptide) docking [19]. To elucidate whether aggregates of peptides are ‘amyloid-like’ or not, the classical test with specific dye (Congo Red) was performed [20].

The suitability of unmodified peptides WI14 and FI15 for further modeling experiments was easily confirmed, since they exist as low order oligomers at pH 7.4. Instead, the peptides with acetaminomethylcysteines were shown to form high molecular weight conglomerates along with the low order oligomers, which however were not of the beta-amyloid-like fold. The corresponding 3D-models of the HA C-terminal peptides have been used to predict the interaction modes of the HA CT region with the N-terminal domain of M1 crystallized at different pH levels. Finally, several tripeptide blockers of those interactions have been proposed.

2 Materials and Methods

2.1 Material

As the biochemical material for this study, four commercially synthesized peptides (Elabscience) were used as previously described [2]. The molecular mass of the FI15 peptide (NH₂-FWMCSNGSLQCRICI-COOH) is 1761.13 kDa, and purity level is equal to 95.84%. The molecular mass of the FI15-ACM peptide (NH₂-FWMC(ACM)SNGSLQC(ACM)RIC(ACM)I-COOH), containing acetaminomethylcysteine instead of cysteine, is 1974.43 kDa, and purity level is equal to 98.98%. The molecular mass of the WI14 peptide (NH₂-WMCSNGSLQCRICI-COOH) is 1613.36 kDa, and purity level is equal to 95.07%. The molecular mass of the WI14-ACM peptide (NH₂-WMC(ACM)SNGSLQC(ACM)RIC(ACM)I-COOH) is 1826.44 kDa, and purity level is

equal to 95.07%. The purity level of four synthesized peptides was determined by HPLC-MS analysis. The length of the intraviral (cytoplasmic) hemagglutinin domain of the influenza virus varies from 12 to 16 amino acid residues according to different Uniprot records and the tryptophan residue often indicates the border between cytoplasmic tail and transmembrane domain of the protein [2].

As the bioinformatics material, the model of the FI15 peptide and the refined homology-based model of the FI15 peptide were used as previously described [2]. The non-refined model of the FI15 peptide has two short beta-strands (Trp2-Met3 and Gln10-Cys11) and corresponds to the HA CT conformation at pH 7.4 according to CD and HATR-FTIR data [2]. The conformation of refined homology-based model possesses the minimal free energy and also has two beta-strands, but each encompasses four residues: Trp2-Ser5 and Ser8-Cys11. This form of the peptide model seems to be closer to that of the FI15 peptide at pH 5.0 [2].

2.2 Blue Native Gel Electrophoresis Followed by Silver Staining

As markers of molecular masses for native gel electrophoresis we used several oligomeric forms of cytochrome C (porcine, heart) (SERVA Electrophoresis, Germany) [21]. Saturated solutions of cytochrome C and WI14, FI15, WI14-ACM, and FI15-ACM synthetic peptides were obtained by dissolving of samples (200 mcg each) in 400 mcL of the running buffer that contained high concentration of Coomassie Brilliant Blue G-250 (0.25%) and keeping at 5 °C overnight. After that, each sample was centrifuged at 18,000 g for 10 min to remove the undissolved fraction. Volumes of the obtained saturated sample solutions used for the analysis were as follows: cytochrome C—10 mcL; WI14, FI15, WI14-ACM, and FI15-ACM—40 mcL. Each sample solution was mixed with 20 mcL of glycerol before loading onto the gel.

Electrophoresis was carried out in the Hoefer SE600 Standard Dual Cooled Vertical Electrophoresis Unit according to a modified protocol [16]. Separating gel (10%) has been casted by the way of the pouring together 0.5 M Bis-Tris buffer (pH 7.4, 11.8 mL), acrylamide-bisacrylamide (30% and 0.8%, correspondingly) stock solution (10 mL), deionized water (8.2 mL), 10% ammonium persulfate solution (150 mcL) and TEMED (30 mcL). Stacking gel (4%) was obtained by the way of mixing 0.5 M Bis-Tris buffer (pH 7.4, 1.56 mL), 30%/0.8% Acrylamide/Bis-acrylamide solution (0.78 mL), deionized water (3.66 mL), 10% ammonium persulfate solution (30 mcL) and TEMED (7.5 mcL). As the cathode buffer, 50 mM tricine, 15 mM Bis-Tris, and 0.02% Coomassie Brilliant Blue G-250 (pH 7.4) solutions were used. The anode buffer was a 50 mM Bis-Tris solution (pH 7.4).

Electrophoresis was started with the voltage equal to 120 V and it was increased to 150 V after the front had crossed the border between stacking and separating gels (after 45 min). The whole procedure lasted for 4 h in cooling conditions.

After that the gel was fixed for 30 min in a mixture of 70% ethanol, glacial acetic acid, and deionized water (4:1:5, v/v) and washed from nonspecifically bound Coomassie Brilliant Blue G-250 in a mixture of 70% ethanol, glacial acetic acid, and deionized water (1.36:1.00:7.64, v/v) three times within 48 h.

Silver staining was carried out according to the staining protocol [22]. The gel was washed twice with 20% ethanol for 10 min, and then by distilled water twice during 10 min. Next, the gel has been put into 0.02% solution of sodium thiosulfate for 1 min and washed three times by distilled water for 1 min. Then the gel has been put into 0.2% silver nitrate solution for 20 min and washed by distilled water trice during 10 min. After that the gel has been immersed in basic developing solution containing 18 g of sodium carbonate, 6 mL of 0.02% sodium thiosulfate solution, 288.6 mL of distilled water and 150 mcL of 40% formaldehyde solution. Once noticeable bands appeared, the reaction has been stopped by 5% acetic acid solution. This solution has been poured after 20 min, the gel has been washed in distilled water and scanned.

2.3 Membrane Filtration Followed by Fluorescence Spectroscopy

Membrane filtration of the WI14-ACM peptide was performed with Amicon Ultra devices (Merck, USA) [17]. The initial sample volume of the saturated WI14-ACM solution in 0.01 M phosphate buffer (pH 7.4) was equal to 400 mcL. Centrifugal ultrafiltration was carried out in three successive steps (12,000 g; 10 min each): initial saturated solution was filtered through the device with MWCO = 30 kDa, collected filtrate was passed through a device with MWCO = 10 kDa, and after that a new filtrate portion was passed through the device with MWCO = 3 kDa. The material retained by each of these three filters was washed away by centrifugation at 12,000 g for 10 min (during those procedures filters were reversed) in 400 mcL of 0.01 M phosphate buffer (pH 7.4).

The concentrations of peptides were determined by using absorption values at the wavelength of 280 nm, at which a single Trp residue in each peptide has a molar extinction coefficient equal to $5500 \text{ M}^{-1} \cdot \text{cm}^{-1}$. Each sample was diluted by 600 mcL of 0.01 M phosphate buffer (pH 7.4) and studied by spectrofluorimetry for the WI14-ACM presence ($\lambda_{\text{Ex}} = 280 \text{ nm}$; $\lambda_{\text{Em}} = 300\text{--}400 \text{ nm}$; step of 1 nm, and slit width of 2 nm). Each spectrum was recorded thrice and the resulted differential spectra for four samples of interest (heavier than 30 kDa; with molecular mass from 10 to

30 kDa; with molecular mass from 3 to 10 kDa; and lower than 3 kDa) minus the pure buffer are described in this study.

Membrane filtration of the saturated FI15 peptide solution in 0.01 M phosphate buffer (pH 7.4) was carried out by the Amicon Ultra device with MWCO = 10 kDa in the same manner as it was described above. However, because of the lower peptide concentration, slits of 5 nm were used for fluorescence spectroscopy. While the concentration of WI14-ACM is equal to 0.09 mg/mL, the concentration of FI15 is equal to 0.04 mg/mL.

So, solubility of the studied peptides in PBS is very low. Since we wanted to get closer to the conditions in the cell, we did not use other solvents than water, like trifluoroethanol, acetonitrile, or DMSO to increase the peptide's solubility because it could significantly change conformation of the peptides. Spectra of fluorescence were recorded using SOLAR CM2203 spectrofluorometer.

2.4 The Congo Red Spectroscopic Assay

The volume of the sample in cuvette was equal to 1 mL. A zero line of the phosphate buffer solution was recorded at 400 to 700 nm using SOLAR CM2203 spectrofluorometer at room temperature. Then 5 mL of the Congo Red solution (7 mg/mL in PBS) was added to the phosphate buffer, and the Congo Red absorbance spectrum was recorded at 400 to 650 nm. After that 10 mL of the saturated peptide solution was added to the cuvette, incubated for 30 min at room temperature and the "peptide + Congo Red" absorbance spectrum at 400 to 650 nm was recorded. To get the differential spectra of all peptides, we have subtracted the Congo Red absorbance spectra from the "peptide + Congo Red" absorbance spectra.

2.5 Determination of the Structure Stability of HA CT

Stability of the secondary structure of the HA CT region and its ability to undergo structural transitions was evaluated by the PentUnFOLD web server using all versions of that algorithm [23].

2.5.1 Protein-Protein Docking

The modeling of the FI15 oligomers structure was performed by Hex 8.0.0 program [19]. We used docking by shape, by electric charge, and by DARS (Decoys As the Reference State) [24] together. As an input, we have used the initial and the refined models of the FI15 peptide obtained previously [2]. Visualization of the obtained models was performed by RasMol [25]. Determination of surface accessibility for tryptophan residues in monomers, dimers, trimers, and tetramers of FI15 has been performed by the BIOVIA Discovery

Studio Visualizer [26]. Isoelectric points (pI) for WI14 and FI15 peptides have been calculated by the renewed Isoelectric Point Calculator 2.0 [27].

2.5.2 Quality Evaluation of the Obtained Models of Oligomers

In order to evaluate the quality of the constructed 3D models of FI15 trimer and tetramer at different pH levels, the following web servers were used: VADAR [28] and PROCHECK [29]. VADAR is able to evaluate > 30 of key structural parameters for the entire protein (peptide) and it is specifically designed for quantitative and qualitative assess of protein structures determined by X-ray crystallography, NMR spectroscopy, 3D-threading or homology modeling. The PROCHECK provides a detailed check of the stereochemistry of a protein structure.

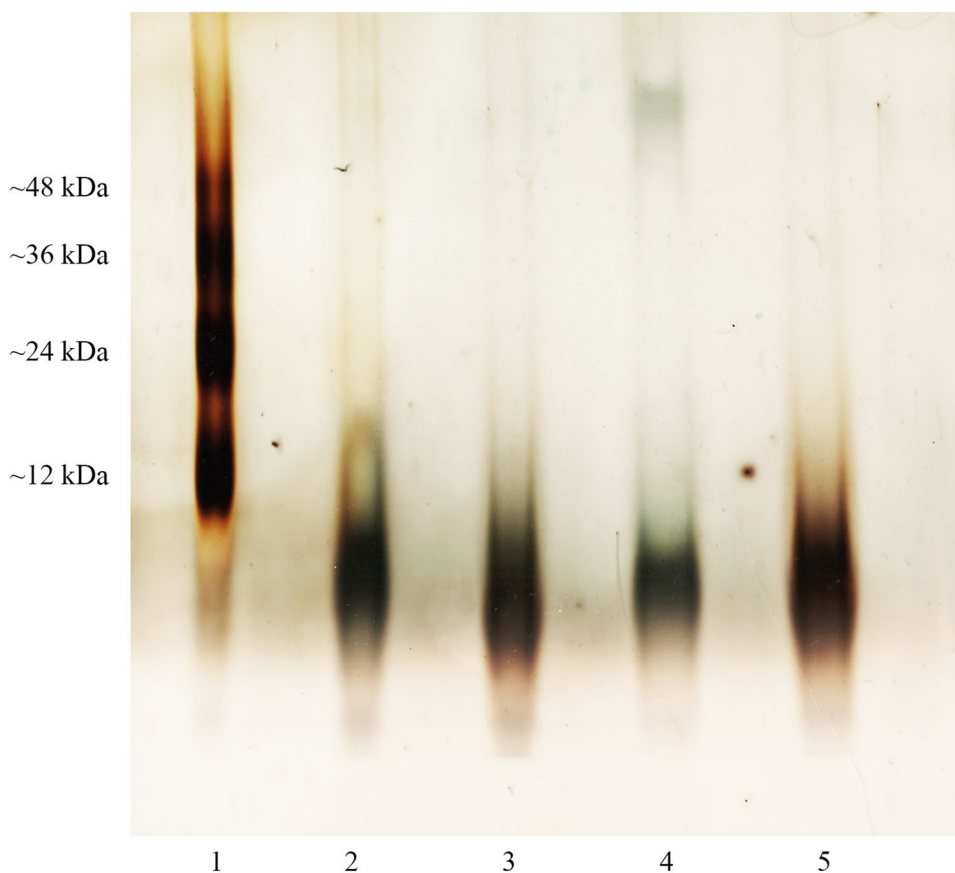
2.5.3 Modeling of M1 N-Terminal Domain at pH 7.4 and 5.0 and Its Interactions with the HA CT

In order to take into account possible changes occurring in conformation of both the M1 protein and the HA CT region during acidification of the medium, we have built two models of 3D structures of the M1 protein by the SWISS MODEL program using as a template two 3D structures of the M1 protein N-terminal domain at acidic and neutral pH levels (PDB ID: 4PUS [30] and PDB ID: 3MD2 [31]), respectively. The crystals for the first structure were formed at pH 4.7 and for the second structure they were formed at pH 7.0. Both structures were obtained for M1 from A/H1N1 Influenza virus strains. Molecular docking was carried out using the Hex 8.0.0 program, taking into account the shape, electrostatic interactions, and DARS. Two models of the FI15 peptide corresponding to the HA CT amino acid sequence were used as ligands: the first one at pH 7.4 and the second one at pH 5.0. While the first model (so called "non-refined" homology-based model, see more details in the "Material" sub-section above) contained short beta strands (2 amino acid residues each), the second model (refined homology-based model) contained longer beta strands (4 amino acid residues each) [2]. In order to make docking results more realistic we cut away hydrophobic side chains from these models of the HA CT region, since naturally they must be buried inside the lipid membrane (which was absent in our modeling system), and started their main chains from carboxylic carbon atom of Trp residue.

2.5.4 Modeling of Tripeptides Capable to Bind HA CT

All possible variants of tripeptides capable to bind HA CT have been modeled by the PepFOLD 3.5 algorithm [32]. We have modelled the acetylated and formylated tripeptides.

Fig. 1 Electrophoregram showing blue native gel electrophoresis and silver staining data obtained for cytochrome C and synthetic WI14, FI15, WI14-ACM, FI15-ACM peptides



- 1 - Cytochrome C: tetra-, tri-, di-, and monomers
- 2 - FI15-ACM
- 3 - FI15
- 4 - WI14-ACM
- 5 - WI14

Docking of tripeptides and the FI15 peptide models (the non-refined and the refined ones) was carried out using the Hex 8.0.0 program, again, at pH 7.4 and 5.0, respectively [19].

3 Results

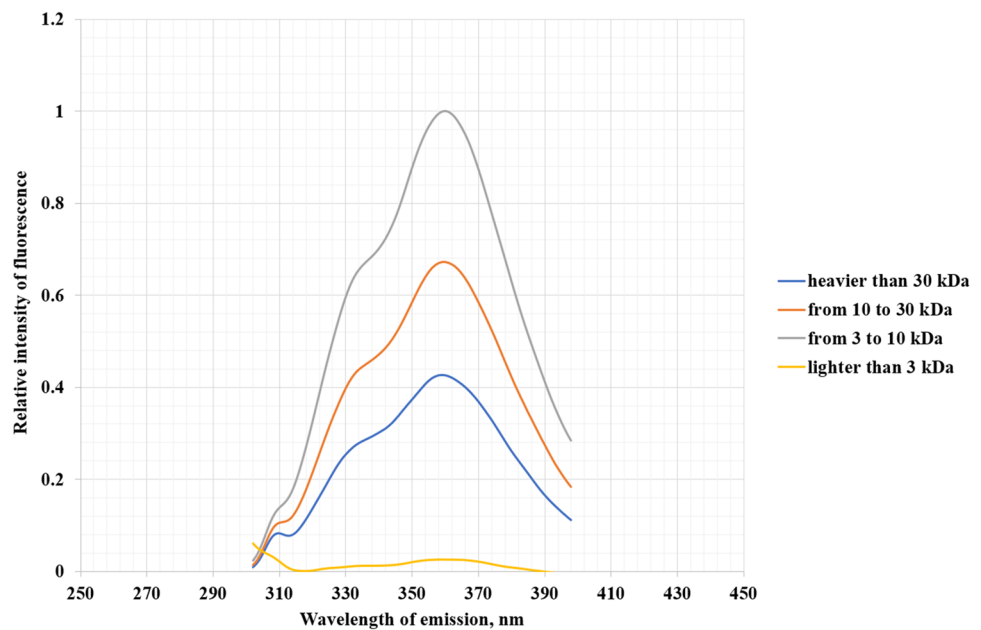
3.1 Blue Native Gel Electrophoresis

According to the Isoelectric Point Calculator 2.0, isoelectric points of the FI15 and WI14 peptides are equal to 8.71 and 9.08, respectively [27]. In this state, they will not move to the positively charged electrode during classical native gel electrophoresis, since the running buffer in this method has

a pH level equal to 8.8. To overcome this problem, blue native gel electrophoresis at pH 7.4 was carried out. The studied peptides were incubated with Coomassie Brilliant Blue G-250 overnight, and a running buffer contained that dye as well [16]. As a marker of molecular masses, we used cytochrome C protein that is known to form well characterized monomers (≈ 12 kDa), dimers (≈ 24 kDa), trimers (≈ 36 kDa), and tetramers (≈ 48 kDa) in a solution [21]. The pI value of cytochrome C is equal to 9.6, which is close to those of FI15 and WI14 peptides [33]. Incubation with Coomassie G-250 results in the binding of this negatively charged dye to proteins, but does not cause denaturation and dissociation of protein-protein complexes [16].

As one can see in Fig. 1, all four model synthetic peptides (WI14, FI15, WI14-ACM, and FI15-ACM) show a

Fig. 2 Fluorescence spectra of the WI14-ACM peptide solution in 0.01 M phosphate buffer (pH 7.4) fractionated by membrane filtration through filters with different molecular weight thresholds (3, 10, and 30 kDa)



major band with a molecular mass slightly lighter than that of the cytochrome C monomer (≈ 12 kDa). In addition, there is a heavier fraction of the FI15-ACM peptide with a molecular mass in the range from 12 to 24 kDa, since the corresponding band is located between the bands of cytochrome C monomers and dimers. Another acetaminomethylated peptide (WI14-ACM) showed an even heavier additional band with a molecular mass higher than the one of the cytochrome C tetramers (≈ 48 kDa), but the corresponding oligomers were still able to enter the 10% polyacrylamide gel and pass the upper segment of the path.

Thus, in a 0.01 M phosphate buffer solution (pH 7.4), FI15 and WI14 peptides form oligomers with a molecular mass lower than 12 kDa. Since the calculated molecular masses of the studied model peptides are equal to 1.76 kDa for FI15 and 1.61 kDa for WI14, we presume that the observed oligomers in solution are not larger than hexamers. Oligomers of the same order are also formed by the FI15-ACM and WI14-ACM peptides (their molecular masses are 1.97 and 1.83 kDa, respectively), which are obviously the main fraction. Yet, for some reason, presence of acetaminomethylcysteines led to the additional formation of a small portion of heavier oligomers.

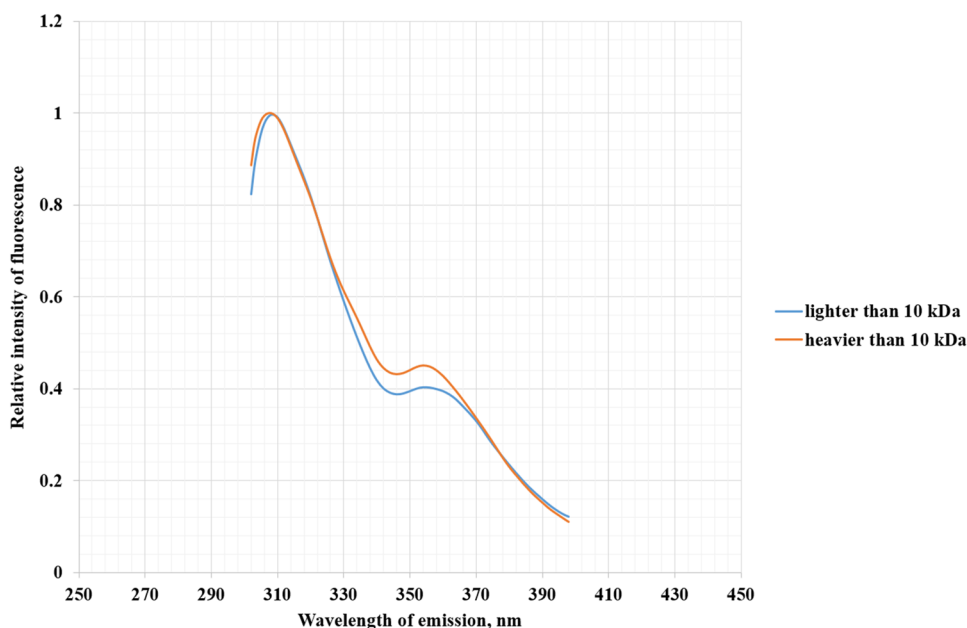
To elucidate further the oligomerization properties of our synthetic peptides, we applied to two of them, WI14-ACM and FI15, another approach – centrifugal ultrafiltration through a series of membrane filters with different molecular weight thresholds. Fluorescence spectra of all fractions were measured.

3.2 Membrane Filtration of the WI14-ACM Peptide

Since the WI14-ACM peptide exhibits the highest solubility (0.09 mg/mL) at pH 7.4 among all four model peptides and because this peptide forms high molecular mass oligomers (the order of such oligomers is the highest among all studied peptides), it was selected for the membrane filtration experiment. Sequential application of the filtration devices with MWCO of 30 kDa, 10 kDa, and 3 kDa let us conclude that the lower band seen at the electrophoregram in Fig. 1 corresponds to hexamers or lower order oligomers. Indeed, the strongest fluorescence signal at $\lambda_{Em} = 360$ nm has been detected in the fraction of WI14-ACM with molecular mass from 3 to 10 kDa (Fig. 2). This fraction corresponds to the complexes of up to six peptides since the molecular mass of WI14-ACM is 1.83 kDa. The signal for oligomers with molecular mass in the range from 10 to 30 kDa is lower (67% of the intensity of the one for the previous fraction), as well as the signal for conglomerates of a molecular mass heavier than 30 kDa (43% of the intensity for the first fraction). The signal for the fraction passed through the 3 kDa threshold filter (obviously the WI14-ACM monomers) is almost negligible (2.5% of the intensity of the first fraction). Taking into account the fact that the shape of fluorescence spectra is the same for all fractions (the maximum is at 360 nm), one can state that tryptophan residues are available for the water solvent even in high molecular mass oligomers [18].

Thus, we may conclude that the main fraction of the WI14-ACM peptide seen in the electrophoregram as a major elongated band with a molecular mass ≤ 12 kDa (Fig. 1) contains different oligomers, ranging from dimers to hexamers that is confirmed by the results of the membrane filtration

Fig. 3 Fluorescence spectra of the FI15 peptide solution in 0.01 M phosphate buffer (pH 7.4) fractionated by membrane filtration through a filter with a threshold of 10 kDa



experiments. The same is obviously true in the case of three other peptides, showing at the electrophoregram the main bands with approximately the same molecular mass as WI14-ACM (Fig. 1).

3.3 Membrane Filtration of the FI15 Peptide

Even though FI15 peptide has the lowest solubility (0.04 mg/mL) among the four model peptides, filtration of its saturated solution through the membrane with a MWCO = 10 kDa provided us with new data. The maximum of fluorescence for the FI15 tryptophan residue is at $\lambda_{Em} = 308$ nm [2], that is an evidence of its absolutely hydrophobic microenvironment. The filtration experiment showed us that the maximum of FI15 fluorescence remained at the same wavelength both in the fraction that did not pass through the membrane with a threshold of 10 kDa, and in the fraction that did pass through it (Fig. 3). Surprisingly, the amplitude of an additional fluorescence peak observed at $\lambda_{Em} = 360$ nm belonging to tryptophan residues exposed to the surrounding water molecules is even higher in large oligomers (those heavier than 10 kDa) than in small oligomers (those lighter than 10 kDa and composed of two to five peptide molecules). Obviously, oligomers of the FI15 peptide form different quaternary structures in water solution. However, in both “light” and “heavy” FI15 oligomers most tryptophan residues are completely surrounded by hydrophobic amino acid residues. These observations inspired us to carry out further modeling experiments using protein-protein docking tool Hex 8.0.0 to try to structurally substantiate the results obtained experimentally.

3.4 Congo Red Spectroscopy Assay

Additionally, we applied a standard Congo Red spectroscopy assay [20]. The absorbance spectra of the control Congo Red solution and solutions of the peptides incubated for 30 min with Congo Red demonstrated the same shape (Fig. 1S, A). The maximum of absorbance for Congo Red solution is at 493 nm, and the maxima of absorbance for WI14, WI14-ACM, FI15, and FI15-ACM peptides plus Congo Red are at 492 nm, meaning there are no essential maxima shifts between the pure Congo Red spectrum and that of the Congo Red in the presence of any dissolved peptide. On the differential spectra all peptides had the maximum of absorbance at 488 nm (Fig. 1S, B). In the presence of amyloid fibrils, a maximum of absorbance should be at 540 nm [34]. Thus, the assay with Congo Red did not reveal formation of any beta-amyloid structures in the saturated solutions of all studied peptides: WI14, WI14-ACM, FI15, FI15-ACM.

3.5 Modeling of FI15 Oligomers with at Least One Completely Buried Tryptophan Residue

As the input for the protein-protein docking, the non-refined and the refined homology-based models of the FI15 peptide were used [2]. In silico we created oligomers of different order to elucidate which variant corresponds to completely buried tryptophan residue experimentally detected as a fluorescent signal with maximum at 308 nm. According to the BIOVIA Discovery Studio Visualizer, the absolute solvent accessibility of tryptophan residue in FI15 monomer at pH 7.4 is equal to 199.531 (relative solvent accessibility is 82.403%). At pH 5.0 absolute solvent accessibility

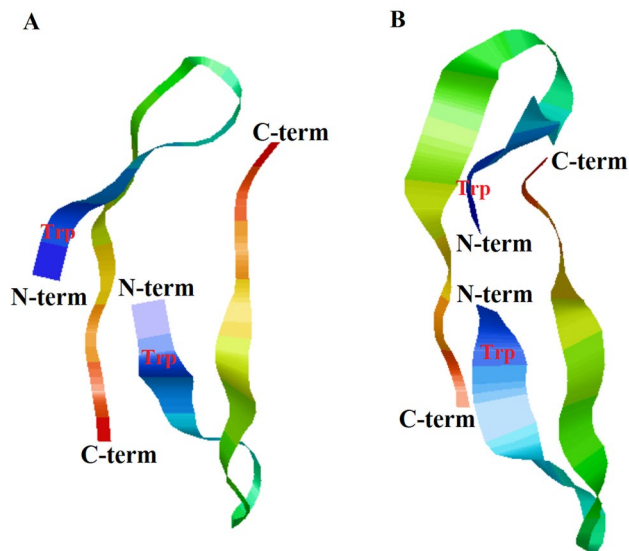


Fig. 4 FI15 dimer models with reduced surface accessibility of Trp residue at pH 7.4 (A) and pH 5.0 (B). Amino acid residues are colored according to the rainbow principle from purple/blue (N-terminus) to red (C-terminus). Tryptophan residues are indicated

of tryptophan residue in FI15 monomer is 165.309 and relative solvent accessibility is 68.269%. In the form of dimers at pH 7.4 accessibility of one of the tryptophan residues is decreased (from 199.531 to 86.44 or from 82.403 to 35.698%). Tryptophan residue from chain A is located between relatively hydrophobic fragments of both chains of a modeled dimer (Fig. 4A). At pH 5.0 accessibility of one of the tryptophan residues is also decreased from 165.309 to 58.369 (or from 68.269 to 24.105%), and that tryptophan residue contacts with a phenylalanine residue from another polypeptide chain. Noteworthy, in similar WI14 oligomers (those without initial phenylalanine residues) tryptophan is not buried [2]. As one can see in Fig. 4B, a hydrophobic C-tail of one of the modeled peptides contacts with N-terminal region of another peptide, a stalk tryptophan of the first peptide is situated between two relatively hydrophobic fragments and covered by a phenylalanine located at the N-terminus of the second peptide. However, the geometry of the complexes at two studied pH levels is different. One of the main differences is that at pH 5.0, phenylalanine from chain B is situated closer to the tryptophan from chain A, which causes a more pronounced decrease in the surface accessibility of the tryptophan residue to water solvent within the dimer. Actually, dimers described above were not the most energetically favorable ones (according to the energy calculated by Hex 8.0.0 for each model). Both models were selected by us from the range of 100 models according to the criterion of the lowest surface accessibility of one of the Trp residues.

The next docking step was performed for two FI15 dimers described above, and it was followed by the selection of models with lowest surface accessibility of one of the Trp residues. At pH 7.4 the lowest accessibility of one of the tryptophan residues is equal to 3.802 (1.57%). In addition to interactions already existing in a dimer, that tryptophan residue (Fig. 5A) is covered by hydrophobic C-tail of chain C and D. Tryptophan residue of the A chain is not available to water molecules (Fig. 5B). The “hydrophobic box” observed around Trp2-A is made of Leu9-A, Cys11-A, Ile13-B, Cys14-B, Ile15-B, Phe1-C, Arg12-C (aliphatic part of its side chain), Cys14-C, Cys11-D, Arg12-D (aliphatic part of its side chain) (designated orange in Fig. 5C). At pH 5.0 the lowest accessibility of one of the tryptophan residues lower, than at pH 7.4 and is equal to 2.013 (0.831%). In addition to interactions already existing in a dimer, tryptophan residue of the A chain is covered by both N-terminal phenylalanine of one of the two new polypeptide chains, and by its hydrophobic C-tail (Fig. 5D). Indeed, with the function of viewing the molecular surface in RasMol [25], it becomes obvious that tryptophan residue of the A chain is not available to water molecules (Fig. 5E). The hydrophobic box around Trp2 of chain A is made of amino acid residues from A, B and C chains (Phe1-C, Arg12-C (aliphatic part of its side chain), Ile13-C, Cys14-C, Ile15-C, Phe1-B, Met3-B, Cys11-B, Arg12-B (aliphatic part of its side chain), Leu9-A, Cys11-A) (Fig. 5F).

Thus, one of the chains within the tetramer is not “active” – it does not participate in the formation of a hydrophobic box around Trp2A at pH 5.0. That’s why we next tried to create a trimer model by performing docking between the dimer and a monomer. At pH 7.4 the solvent accessibility of tryptophan residue in the trimer (Fig. 6A, B) has increased somewhat compared to the corresponding value within the tetramer described above and reached 12.356 (5.103%). At pH 5.0 the best model of the trimer contains Trp2A residue with surface accessibility equal to 1.006 (0.416%), due to the C-terminus of the third polypeptide chain (Fig. 6C, D). Both models of trimers as models of dimers and tetramers were selected by us from the range of 100 models according to the criterion of the lowest surface accessibility of one of the Trp residues.

According to the PROCHECK server, Ramachandran plots of FI15 tetramer and trimer models at pH 5.0 are realistic: 83.3% of combinations of Psi and Phi dihedral angles are situated in the most favored regions and 16.7% are in additional allowed regions (Fig. 2S). The reliability of the FI15 tetramer model is also confirmed by the VADAR web server: 75% combinations of Psi and Phi dihedral angles are situated in core, 21% are in allowed area, and 3% are in generous area (in tetramer) vs. 75%, 20% and 4% (in trimer) (Fig. 3S). Based on VADAR web server, free energy of FI15 trimer and tetramer folding is -35.11 kJ/mol and

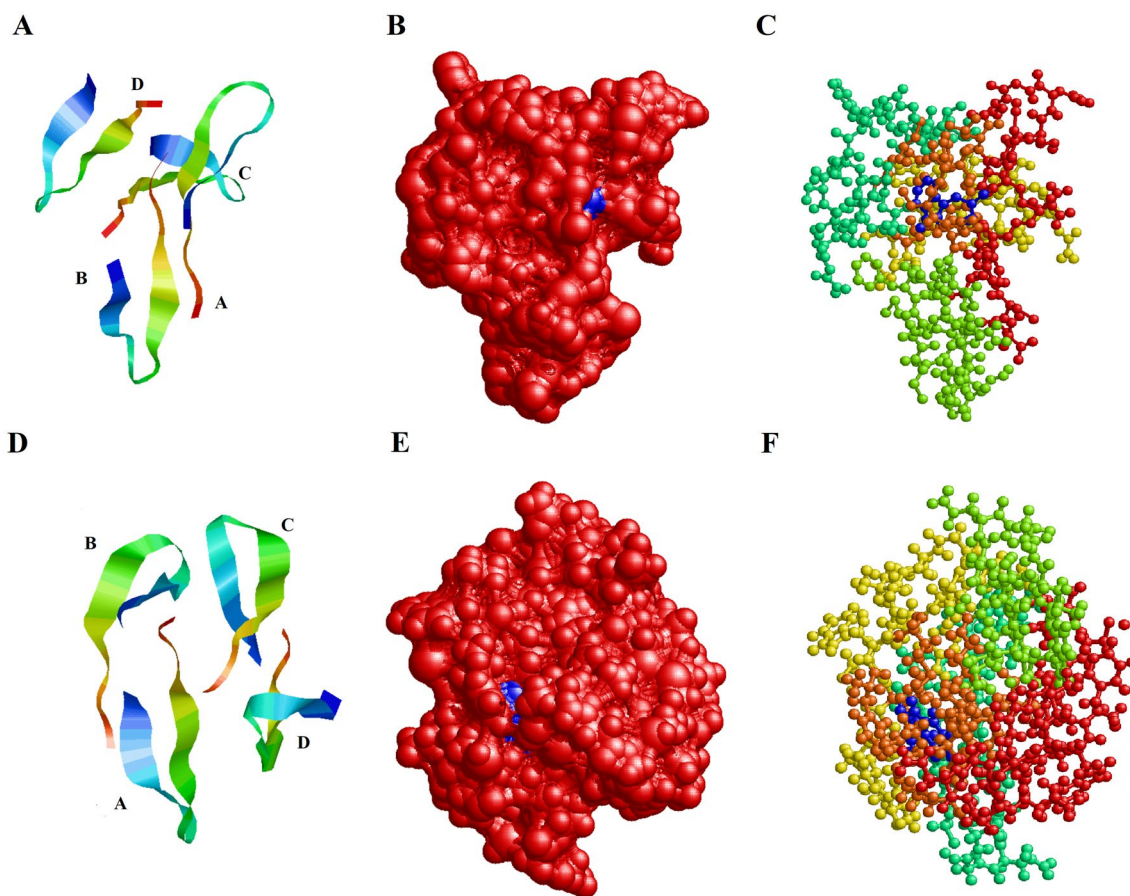


Fig. 5 FI15 tetramer models with a completely buried Trp residue of chain A within a tetramer at pH 7.4 (A, B, C) and pH 5.0 (D, E, F). **A, D:** Models of FI15 tetramer are represented as “Ribbons”. All chains are designated by letters. Amino acid residues are colored according to the rainbow principle from purple/blue (N-terminus) to red (C-terminus). **B, E:** Molecular surfaces of the FI15 tetramers: a buried tryptophan residue of the chain A is colored in blue, while

other amino acids residues are colored in red. **C, F:** Models are in “Ball & Stick” representation. Tryptophan residue of the chain A is colored in blue, hydrophobic amino acid residues around it are colored in orange, other residues are colored according to chains (chain A – in red, chain B – in green, chain C – in yellow, chain D – in azure)

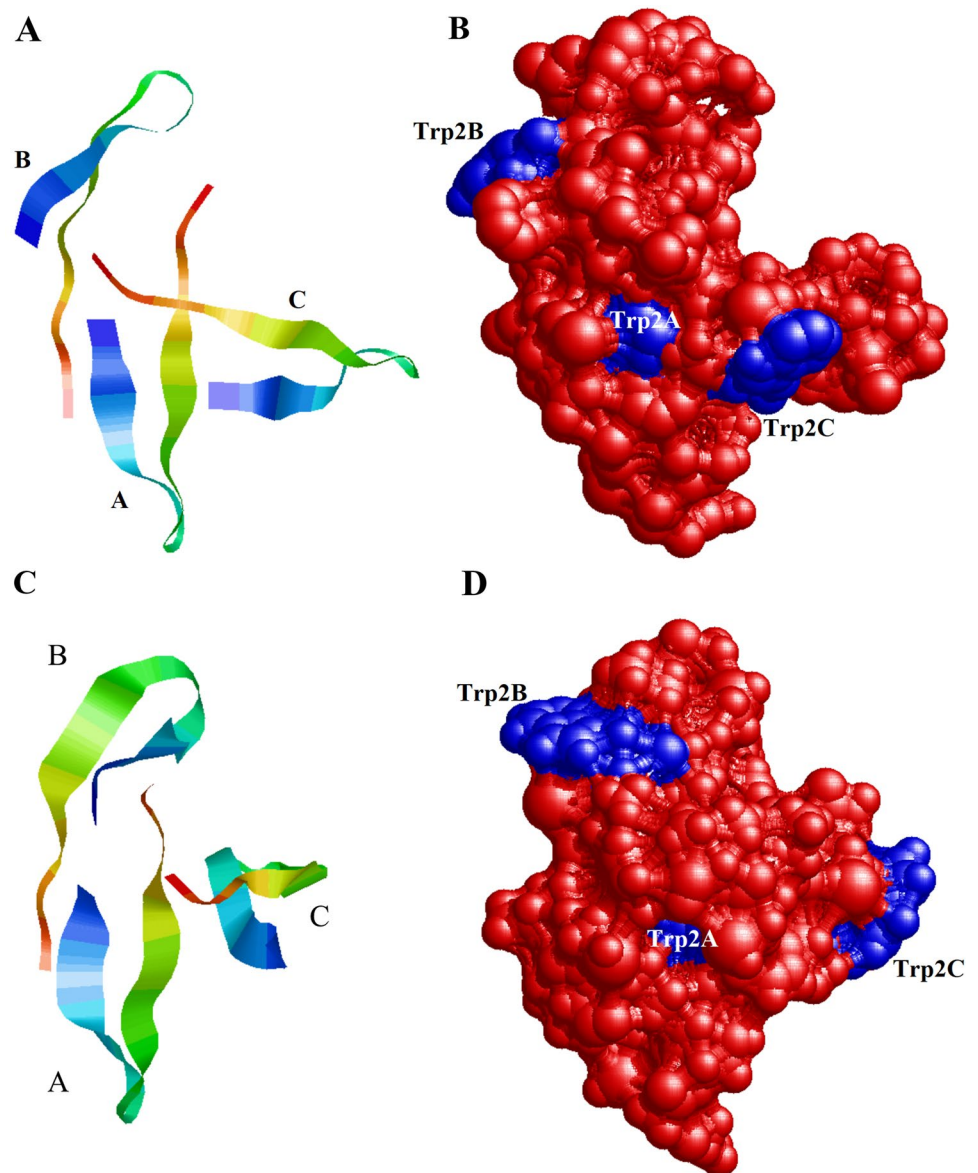
– 50.89 kJ/mol, respectively. At pH 7.4, 83.3% of combinations of Psi and Phi dihedral angles are situated in the most favored regions, 8.3% are in additional allowed regions and 8.3% are in generously allowed area (Fig. 4S). According to the VADAR web server 76% combinations of Psi and Phi dihedral angles are situated in core, 13% are in allowed area, 6% are in generous area and 3% are in disallowed area (in tetramer) vs. 82%, 8%, 6% and 2% (in trimer) (Fig. 5S). Based on VADAR web server, at pH 7.4 free energy of FI15 trimer and tetramer folding is -34.47 kJ/mol and -50.64 kJ/mol, respectively.

Thus, one of the tryptophan residues can be completely buried even in trimers (more likely, at acidic pH), and at a higher probability, in tetramers of the FI15 peptide. This explains the presence of the high fluorescence signal at

308 nm observed for the FI15 peptide solutions described above.

According to the previously published results, the reduction of disulfide bonds leads to the growth of the number of amino acid residues in beta structure for the FI15 peptide [2]. In our models with completely buried Trp2A residue, there are no disulfide bonds. So, those models are somehow correspond to the reduced state of FI15 in which Trp2A is as buried as in that with disulfide bonds. Taking into account all the data we obtained previously [2] and now we can hypothesize that disulfide bonds appear in penta- and hexamers of FI15. In addition, in penta- and hexamers there should be a parallel beta structure along with an antiparallel one [2]. Reduction of disulfide bonds should lead to structural change of those peptides in favor of the longer beta structure formation, getting it closer to the refined model of the FI15 peptide.

Fig. 6 FI15 trimer model with a completely buried Trp residue of chain A within a trimer at pH 7.4 (**A, B**) and pH 5.0 (**C, D**). **A, C**: FI15 trimer models are represented as ribbons with a completely buried Trp residue of chain A within a trimer. All chains are designated by letters. Amino acid residues are colored according to the rainbow principle from purple/blue (N-terminus) to red (C-terminus). **B, D**: The molecular surfaces of the FI15 trimers: tryptophan residues of all three chains are colored in blue, while other amino acid residues are colored in red



3.6 Determination of Structural Stability of the HA CT

We further addressed the *in silico* structural stability of the HA CT region based on the analysis of its amino acid sequence. For this, we applied a recently developed PentUnFOLD algorithm [23] that exists as three versions (1D, 2D, and 3D) to a model peptide FI15 (Table 1). At pH 7.4, a beta hairpin consisting of two beta strands (Trp2-Met3 and Gln10-Cys11) is determined by the DSSP and also by the 2D version of PentUnFOLD algorithm [23]. The first beta strand is more stable since Met3 is defined as stable beta strand residue (“ES” according to [23]) (Table 1). At the same time, the second beta strand is defined as nonstable (“EN”). But this beta strand should be prone to elongation, both from N- and C-termini, since Leu9, as

well as Arg12, are determined by the PentUnFOLD algorithm as “CE”, i.e. these amino acid residues are capable of structural shift from random coil to beta strand. According to the 3D version of the algorithm, all amino acid residues are capable of structural shift with the exception of Phe1 (Table 1). However, these transitions do not lead to a completely (intrinsically) disordered state. Only structural shifts from one element of the secondary structure to another, in particular, from a random coil to a beta strand and vice versa, are available for FI15, according to the PentUnFOLD predictions. The instability of the second beta strand (Leu9-Arg12) is partially confirmed by the 1D version of the PentUnFOLD algorithm. According to this version, Gln10 is defined as “EN”, while the remaining residues of this beta strand are classified as stable. The first beta strand is not determined by the probabilistic scales

Table 1 Stability of the secondary structure of FI15 model peptide determined by the PentUnFOLD algorithm.*

№ of amino acid residue	Amino acid residue	pH = 7.4			pH = 5.0		
		1D	2D	3D	1D	2D	3D
1	F			O			O
2	W			D			O
3	M	CE	ES	D	CE	ES	D
4	C	CH	CH	D	CH	ES	D
5	S			D			D
6	N			D			D
7	G			D			D
8	S			D			O
9	L	ES	CE	D	ES	ES	O
10	Q	EN	EN	D	EN	EN	D
11	C	ES	EN	D	ES	EN	D
12	R	ES	CE	D	ES	CE	D
13	I			D			D
14	C			D			D
15	I			D			D

*1D, 2D and 3D are three different versions of the PentUnFOLD algorithm [23]. «CE» is a fragment of random coil capable to structural transition to beta strand; «CH» is a fragment of random coil capable to structural transition to alpha helix; «ES» are residues of beta strands that have a low probability to turn into random coil; «EN» are residues of beta strands that have a high probability to turn into random coil; «D» is a disordered state; «O» is an ordered state [23]. The boundaries of beta strands defined by each version of the PentUnFOLD algorithm are highlighted in yellow

of the PentUnFOLD algorithm, however, the region of random coil corresponding to the first beta-strand (Trp2-Met3) is capable of structural transition to beta strand (Met3 is defined by the algorithm as “CE”) (Table 1).

At pH 5.0, the secondary structure of the FI15 peptide becomes more stable and ordered. The 2D version of the PentUnFOLD algorithm defines two longer beta strands (Trp2-Cys4 and Leu9-Cys11). At the same time, the amino acid residues Met3, Cys4 and Leu9 are stable (ES). Arg12, as at pH 7.4, belongs to random coil, but under certain conditions it is able to form the C-terminus of the second beta strand. According to the 3D version of the PentUnFOLD algorithm, Phe1, Trp2, Ser8, and Leu9 are defined as ordered elements of the secondary structure (O), while the rest are defined as D (disordered), implying they are able to change their secondary structure but not turn into the intrinsically disordered state (Table 1).

Thus, according to the PentUnFOLD calculations applied to the model peptide FI15, the structural organization of the HA CT region is flexible. The lengths of both beta-strands of FI15 (especially the C-terminal one) can extend during transition from neutral to acidic pH. Unlike the FI15's N-terminus, which is absent in the natural structure of HA CT, the elongation of the C-terminal beta-strand may be related to the physiological process of disassembly of the viral envelope with a decrease of pH level during the transmembrane pore formation and the entrance of the virus into the cytoplasm.

3.7 Modeling of the Interaction of Matrix Protein M1 with the HA CT

As was mentioned above, there is no X-ray resolved structure of the full-length M1. Only M1 N-terminal domain (two thirds of the whole amino acid sequence) was crystallized, and this part of the protein is supposed to interact with the viral membrane [5]. The N-terminal domain includes two sub-domains: N- (helices H1-H4) and M- (helices H6-H9) connected by a short helix H5, and sometimes is called an NM-domain [5, 11].

In order to predict the mechanism of the HA CT interaction with N-terminal domain of M1, protein-protein docking was carried out via the Hex 8.0.0 program between the 3D structures of the alpha-helical N-terminal domain of M1 taken at neutral and acidic pH levels and the FI15 peptide models mimicking the HA CT sequence at pH 7.4 and pH 5.0, respectively.

It is supposed that positively charged residues of M1 protein (Arg76 from loop 4; Arg77, Arg78 from helix H5; Arg101, Lys104 from helix H6, and Arg134 from helix H8) form electrostatic interactions with phosphatidylserine as a main negatively charged partner within the inner monolayer of the lipid membrane [12]. In addition, Gln75 and Gln81 obviously interact with phosphatidylcholine, the major phospholipid of the membrane [12]. Next to the amino acid residues mentioned above, there are also other positively charged amino acids, whose radicals are located on the same

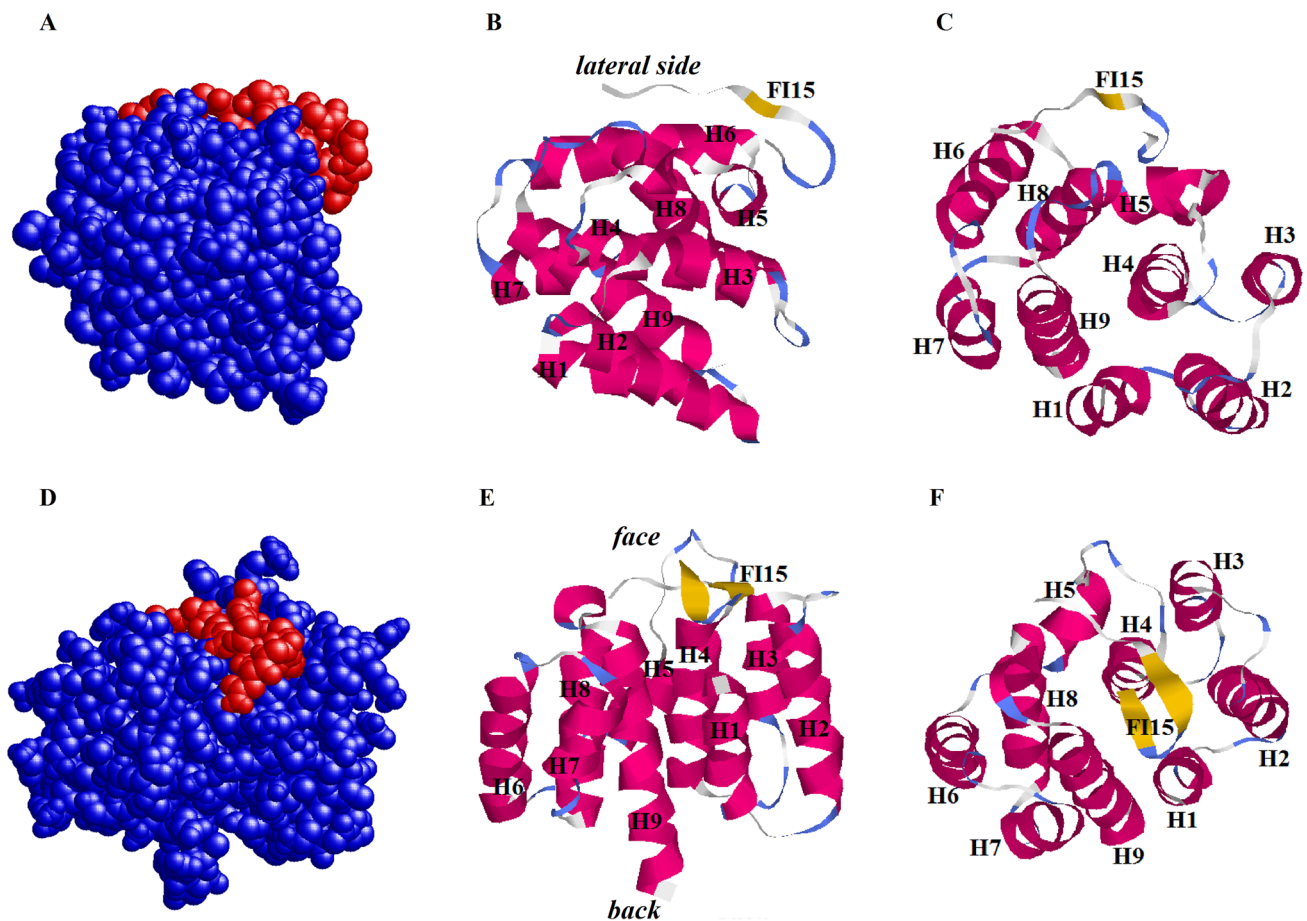


Fig. 7 The model of interactions between the FI15 peptide (mimicking the HA CT region) and M1 N-terminal domain at pH 7.4 (A–C) and pH 5.0 (D–F). In left panels (A, D) the FI15 atoms are shown in red, the M1 atoms are shown in blue. In middle and right panels (B,

C, E, F) alpha helices of M1 are shown in pink, a beta structure of the model FI15 peptide is shown in yellow. Nine alpha-helices of M1 N-terminal domain are labeled as H1–H9, and “lateral” side, “face” and “back” are indicated

surface: Lys47 from helix H3 and Arg105 at the end of helix H6. All these amino acid residues are located on the same “lateral” surface of the M1 protein’s N-terminal domain. The question arises, could this surface of the M1 N-domain serve as a site of interaction with the C-terminal beta structure of HA located directly under the lipid membrane?

Indeed, Hex 8.0.0 has found the most favorable position of the FI15 model peptide at the “lateral” surface of M1 near helices H5, H6, and H8 belonged to the M-subdomain of the N-terminal domain (along helix H6), when both HA CT and M1 models were taken at pH 7.4 (Fig. 7A–C). So, the HA CT seems to interact with the same lateral side of the N-terminal domain of M1 that interacts with lipid membrane itself.

However, in the case of FI15 peptide and M1 models taken at pH 5.0 their favorable mutual arrangement is quite different. The FI15 model peptide interacts with the “face” of the M1 N-terminal domain formed by the contacting ends of alpha-helices H1, H2, H4, H9 of M-subdomain as well

as loops connecting them (Fig. 7D–F). This kind of interaction would be impossible within a virion that is “resting” at neutral pH of the environment having tightly packed matrix layer in which the N-terminal domains of neighboring M1 molecules are well packed in the “face-to-back” manner [6, 7, 11, 31, 35].

The energy of binding for the best HA CT – M1 N-terminal domain complex at pH 7.4 is -504.6 kJ/mol, while at pH 5.0 that energy is essentially higher and equal to -403.01 kJ/mol. The control experiments showed that the best “acidic-like” complex at pH 7.4 and the best “physiological-like” complex at pH 5.0 have energies of binding even higher compared to the values indicated above, equal to just -388.22 kJ/mol, and -363.16 kJ/mol, respectively. We call the HA CT–M1 complex “acidic-like” if a 3D model of M1 N-terminal domain and a model of the HA CT, both built at pH 7.4, were arranged in the “acidic-like” manner (the HA CT interacts with the ends of alpha-helices of N-terminal domain of M1). Accordingly, the “physiological-like”

complex is a complex containing individual models of M1 N-terminal domain and the HA CT region built at pH 5.0 but arranged in relation to each other as at pH 7.4 (the HA CT interacts with the lateral side of M1 N-terminal domain along alpha-helices).

Thus, as a result of the *in silico* experiments, it was found that the models of the FI15 peptide corresponding to HA CT can bind to the same N-terminal domain of M1 regardless of the pH level, albeit to different surfaces of it: to the lateral side at neutral pH and to the ends of alpha-helices at acidic pH. The obtained energy values indicate that the HA CT – M1 contacts are more stable at neutral compared to acidic pH levels, which is reasonable from a virological point of view. In addition, when the pH level decreases, not only the HA CT region's structure and character of its interaction with the N-terminal domain of M1 changes, but there are also some local deviations in interactions of amino acid residues within the N-terminal domain itself. One can surmise that a complex between membrane and lateral side of M1 N-terminal domain may be strengthened by interactions with the cytoplasmic domain of hemagglutinin at pH 7.4, that is important for virus assembly. Subtle structural changes in both HA CT and M1 at pH 5.0 lead to preferable formation of another type of complex between them, which may play a functional role during the disassembly of the viral envelope. To check hypotheses about the functional roles of the HA CT – M1 complexes at various pH levels, one needs a kind of a blocker of interactions between the HA CT region and M1 N-terminal domain. In the next section we suggest several tripeptides that can be used in future experiments in cell culture to confirm the fact that HA CT and M1 do interact with each other.

3.8 Modeling of Tripeptides Capable to Bind HA CT

Dipeptides and tripeptides can penetrate into the cell by passive diffusion or by active transport with the help of oligopeptide transporter PepT1 [36]. Tripeptides have a larger interaction area, and, accordingly, should have a higher specificity than dipeptides. The competitive tripeptide, of course, will not be able to contact with hydrophobic surface of the cytoplasmic tail of hemagglutinin, which faces the membrane. That is why for the protein-protein (peptide-peptide) docking we have used slightly modified FI15 peptide models of the HA CT with removed side chains of hydrophobic residues and with removed hydrophobic residue Phe1.

On the hydrophilic surface of the HA CT of H1 subtype, besides the COOH-group of C-terminal isoleucine, there is only one positively charged residue – Arg12. For this reason, a distinctive feature of all possible tripeptides capable of binding the HA CT via electrostatic interactions should be the presence of a negatively charged residue, preferably aspartic acid, which demonstrates more pronounced

acidic properties compared to glutamic acid. To increase the specificity of binding with Arg12, aspartic acid must be at the C-terminus of the tripeptide, since in this case two negatively charged carboxyl groups will interact with the positively charged guanidine group of Arg12. Obviously, to keep strong interactions between the negatively charged tripeptide and Arg12, there should be no positively charged residues in the tripeptide. For the same reason, there should be no free N-terminal amino group. All these reasons have largely reduced a set of possible candidates: the composition of the tripeptide should be free of hydrophobic and positively charged residues (lysine and arginine), the C-terminus should contain aspartate, and the N-terminus should be modified, for example via formylation or acetylation.

The integral indicator reflecting the affinity of the ligand to a certain receptor is the free energy of binding: the less the energy value, the tighter the binding. The free energy of binding was determined for each formylated (Table 2) and acetylated (Table 3) candidate tripeptide in the “virtual reaction” with the model FI15 peptide mimicking its interaction with the HA CT. As it is seen from Tables 2 and 3, at pH 5.0, the most suitable tripeptide is NTD in both acetylated and formylated states (in tables 2 and 3 the lowest binding energy values among the studied tripeptides at pH 7.4 and at pH 5.0 in formylated and acetylated form are indicated in bold). However, the free binding energy of the FI15 model peptide with the formylated blocker tripeptide is slightly less than that of the acetylated one (– 227.78 vs. – 219.86 kJ/mol) meaning that the formylated tripeptide is a somewhat better candidate for blockers. At pH 7.4, the most suitable tripeptide in the formylated state is NQD, whose free binding energy with the model peptide FI15 is –215.31 kJ/mol, and in the acetylated state it is QGD, whose binding free energy is higher and equal to –210.16 kJ/mol. The binding energy of formylated NQD is lower since it has more options to form bonds with FI15.

Taken together, formylated NQD and NTD tripeptides could be used in future experiments regarding influenza virus envelope assembly or disassembly, respectively.

4 Discussion

Despite the fact that recently attention of the world community has been focused on coronaviruses, the influenza virus remains dangerous to the health of the human population around the world. WHO recommends countries to prepare for co-circulation of Influenza virus and SARS-CoV-2 [37]. Influenza type A virus is one of the serious pathogens that cannot be controlled by the current healthcare, even though vaccination against Influenza viruses is widespread around the globe [38, 39], and several specific antiviral drugs are known. The reasons why this virus

Table 2 The free binding energy of the formylated tripeptides with model FI15 peptide mimicking HA CT at pH 7.4 and 5.0

Tripeptide formylated	Binding energy pH 7.4. kJ/mol	Binding energy pH 5.0. kJ/mol
HSD	- 201.85	- 204.44
HTD	- 193.80	- 202.47
HHD	- 194.67	- 205.72
HQD	- 211.05	- 212.74
HND	- 187.86	- 209.83
HGD	- 203.72	- 194.52
NSD	- 197.85	- 211.09
NTD	- 200.49	- 227.78
NHD	- 194.38	- 196.45
NQD	- 215.31	- 207.46
NND	- 182.32	- 188.80
NGD	- 201.72	- 186.06
QSD	- 207.61	- 216.14
QTD	- 197.82	- 206.00
QHD	- 203.32	- 203.96
QQD	- 207.73	- 200.82
QND	- 207.05	- 203.16
QGD	- 213.12	- 201.40
GSD	- 201.34	- 204.60
GTD	- 196.65	- 199.83
GHD	- 196.58	- 199.76
GQD	- 190.37	- 199.49
GND	- 197.44	- 194.35
GGD	- 195.32	- 186.39
SSD	- 192.23	- 195.46
STD	- 183.19	- 185.51
SHD	- 186.03	- 201.67
SQD	- 187.89	- 197.97
SND	- 193.93	- 203.35
SGD	- 194.62	- 204.60
TSD	- 196.92	- 187.11
TTD	- 183.42	- 185.89
THD	- 195.41	- 210.98
TQD	- 184.92	- 192.68
TND	- 189.75	- 208.03
TGD	- 199.44	- 204.23

Table 3 The free binding energy of the acetylated tripeptides with model FI15 peptide mimicking HA CT at pH 7.4 and 5.0

Tripeptide acetylated	Binding energy pH 7.4. kJ/mol	Binding energy pH 5.0. kJ/mol
HSD	- 196.31	- 193.42
HTD	- 200.90	- 208.50
HHD	- 201.42	- 203.19
HQD	- 196.07	- 188.11
HND	- 190.81	- 198.61
HGD	- 196.76	- 209.24
NSD	- 193.03	- 211.07
NTD	- 196.69	- 219.86
NHD	- 194.51	- 188.75
NQD	- 195.39	- 189.18
NND	- 193.87	- 188.55
NGD	- 201.98	- 198.07
QSD	- 194.67	- 206.80
QTD	- 203.52	- 207.84
QHD	- 196.82	- 210.93
QQD	- 209.39	- 212.80
QND	- 189.20	- 199.16
QGD	- 210.16	- 212.26
GSD	- 190.26	- 203.12
GTD	- 199.12	- 198.21
GHD	- 193.56	- 212.84
GQD	- 195.56	- 201.29
GND	- 189.49	- 203.48
GGD	- 198.92	- 193.38
SSD	- 194.00	- 190.15
STD	- 188.91	- 180.08
SHD	- 184.62	- 202.02
SQD	- 200.43	- 196.43
SND	- 197.08	- 203.01
SGD	- 187.77	- 218.45
TSD	- 200.83	- 189.85
TTD	- 182.97	- 188.07
THD	- 186.36	- 211.99
TQD	- 199.41	- 183.84
TND	- 206.85	- 211.83
TGD	- 192.17	- 213.49

with the genome represented by eight separate minus-RNA molecules escapes immune responses and specific drugs are, first, in its ability to exchange genomic RNA molecules between different viruses (strains, subtypes) infected the same cell (antigenic shift) [40] and, second, in high propensity of its surface antigens, hemagglutinin and neuraminidase, to accumulate mutations (antigenic drift) [41]. An important strategy that can help drug designers is to focus on some very conserved regions of viral proteins that are responsible for important stages of the virus life

cycle. One of such conserved and poorly studied regions having highly complex architecture is the cytoplasmic tail of hemagglutinin.

In vivo, the cytoplasmic tail of hemagglutinin is obviously connected to the inner monolayer of the viral membrane by fatty acid residues covalently bound to its three C-terminal cysteine residues and, additionally, by side chains of its hydrophobic amino acid residues [42, 43]. As a first approximation to obtain a reliable model of such a complicated system we used four synthetic peptides

corresponding to the sequence of the inner domain of A/H1N1 Influenza hemagglutinin, with free or acetaminomethylated SH-groups. Several 3D-models of the HA CT region were proposed after experimental detection of beta structural elements in synthetic peptides in buffer solution, all possessing an antiparallel beta structure [2]. Now, to dispel doubts about a suitability of the obtained models for further *in silico* experiments, we have determined the order of oligomers of those four model peptides in buffer solution using an additional set of approaches, including membrane filtration through series of filters with discrete molecular weight thresholds coupled with fluorescence spectroscopy, a Congo Red assay for beta-amyloid formation and a wide spectrum of bioinformatics calculations.

We have come to the conclusion that all the studied peptides are not intrinsically disordered, but they can only undergo structural transitions from random coil to a beta strand or reverse transitions. In addition, we have found that WI14 and FI15 peptides form only oligomers of low order, while peptides with modified cysteines (WI14-ACM and FI15-ACM) form both oligomers of low order (similar to those found for the WI14 and FI15 peptides) and also high-order conglomerates, which, however, do not form beta-amyloid-like structures.

Having proved the reliability of the obtained models of submembrane domain of hemagglutinin, we have used them for macromolecular docking with M1 protein forming an endoskeleton beneath the virus membrane [44]. It is supposed that positively charged “lateral” surface of the M1 protein’s N-terminal domain interacts with the membrane. Using X-ray crystallography analysis it was revealed that at neutral pH 7.0 the N-terminal domains of two M1 protein molecules within a matrix sheet interact with each other by surfaces where the ends of the alpha-helices are located (in the “face-to-back” manner) [6, 7, 31]. However, at acidic pH they were shown to interact “side to side” (by lateral surfaces of long alpha-helices) [5, 11]. Moreover, in such dimers alpha-helices of the N- and M-subdomains of M1 are antiparallel to each other [5, 11].

At pH 7.4 the most favorable position of the HA CT model is that laying between helices H5, H6, and H8 along the lateral surface of M1’s M-subdomain (Fig. 7A–C). At acidic pH the favorable geometry of their complex is quite different. The HA CT is predicted to interact with the face of the same M-subdomain of M1, namely, with ends of alpha-helices H1, H2, H4, H9 and the loop connecting helices H4 and H5. That kind of complex may play a functional role during the disassembly of the viral envelope allowing transport of the viral RNPs into the cytoplasm. So, the inner domain of A/H1N1 Influenza virus hemagglutinin is yet another promising target for the development of specific antiviral medicines.

It is worth noting that the HA CT–M1 protein interactions have been developed and tuned in the process of molecular co-evolution of viral envelope proteins within every virus type/subtype/strain [45]. Now we have proposed two tripeptides as blockers of those interactions, which may be used for experimental tests in future. We would recommend the usage of formylated NQD tripeptide in viral envelope assembly experiments, while in viral envelope disassembly experiments we recommend the usage of formylated NTD tripeptide to confirm or deny the principles and importance of interactions of the HA CT region with N-terminal domain of M1 in the virus life cycle.

5 Conclusions

The synthetic peptides mimicking the cytoplasmic tail of Influenza A/H1 subtype virus hemagglutinin with three free cysteines (WI14 and FI15) dissolved in phosphate buffer solution at pH 7.4 form oligomers of a low order, starting from trimers up to hexamers. In contrast, model peptides with “blocked” sulfhydryl groups of cysteines (WI14-ACM and FI15-ACM) form higher molecular mass conglomerates in addition to low order oligomers. However, there are no amyloid-like fibrils in the saturated solutions of all studied peptides.

Several structures of FI15 trimers and tetramers with completely buried tryptophan residue emitting at 308 nm have been modeled. According to them, a single peptide molecule should form an asymmetric beta-hairpin to obey the requirement of the lack of the contact with water for at least one of tryptophan residues.

Macromolecular docking showed that the HA CT region is likely to interact with the lateral surface of M1 N-terminal domain which itself is supposed to interact with lipid membrane at pH 7.0 [12, 44]. At acidic pH 5.0, another complex becomes more favorable: the HA CT interacts with the “face” of M1 N-terminal domain formed by ends of alpha-helices. This observation let us hypothesize how the HA–M1 protein contacts that are stable at neutral pH level, may begin to re-organize when the pH level decreases. This fundamental issue of viral envelope disassembly is highly important for the enveloped virus families with pH-dependent life cycle.

Tripeptides that should be able to block interactions between the HA CT and M1 have been suggested in the final step of this study.

Supplementary Information The online version of this article contains supplementary material available <https://doi.org/10.1007/s10930-023-10101-z>.

Author Contributions VVP: conceptualization, methodology, investigation, formal analysis, writing—original draft preparation. VVK: methodology, investigation, formal analysis, writing—original draft preparation, supervision. AAA: investigation, editing. NVS: methodology, investigation, writing—review and editing. ANS: methodology, investigation, writing—review and editing. TAK: methodology, investigation, supervision, writing—review. LVK: methodology, investigation, supervision, writing—review and editing.

Funding This research was financially supported by the Russian Foundation for Basic Research, grant numbers 18-54-00019; 20-54-12007 (to L.V.K.) and by the Belarusian Republican Foundation for Fundamental Research, grant number B18R-113 (to V.V.K.).

Code Availability Not applicable.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical Approval Animals and human biomaterials were not used when writing this article.

References

- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242. <https://doi.org/10.1093/nar/28.1.235>
- Khrustalev VV, Kordyukova LV, Arutyunyan AM, Poboinev VV, Khrustaleva TA, Stojarov AN, Baratova LA, Sapon AS, Lugin VG (2020) The cytoplasmic tail of influenza A/H1N1 virus hemagglutinin is β -structural. *J Biomol Struct Dyn* 40:4642–4661. <https://doi.org/10.1080/07391102.2020.1860827>
- Shishkov AV, Goldanskii VI, Baratova LA, Fedorova NV, Ksenofontov AL, Zhirnov OP, Galkin AV (1999) The in situ spatial arrangement of the influenza A virus matrix protein M1 assessed by tritium bombardment. *PNAS* 96:7827–7830. <https://doi.org/10.1073/pnas.96.14.7827>
- Kordyukova LV, Shtykova EV, Baratova LA, Svergun DI, Batishchev OV (2019) Matrix proteins of enveloped viruses: a case study of Influenza A virus M1 protein. *J Biomol Struct Dyn* 37:671–690. <https://doi.org/10.1080/07391102.2018.1436089>
- Sha B, Luo M (1997) Structure of a bifunctional membrane-RNA binding protein, influenza virus matrix protein M1. *Nat Struct Biol* 4:239–244. <https://doi.org/10.1038/nsb0397-239>
- Harris A, Forouhar F, Qiu S, Sha B, Luo M (2001) The crystal structure of the influenza matrix protein M1 at neutral pH: M1-M1 protein interfaces can rotate in the oligomeric structures of M1. *Virology* 289:34–44. <https://doi.org/10.1006/viro.2001.1119>
- Arzt S, Baudin F, Barge A, Timmins P, Burmeister WP, Ruigrok RW (2001) Combined results from solution studies on intact influenza virus M1 protein and from a new crystal form of its N-terminal domain show that M1 is an elongated monomer. *Virology* 279:439–446. <https://doi.org/10.1006/viro.2000.0727>
- Ksenofontov AL, Dobrov EN, Fedorova NV, Radyukhin VA, Badun GA, Arutiunian AM, Bogacheva EN, Baratova LA (2011) Disordered regions in C-domain structure of influenza virus M1 protein. *Mol Biol (Mosk)* 45:689–696 Article in Russian
- Shishkov A, Bogacheva E, Fedorova N, Ksenofontov A, Badun G, Radyukhin V, Lukashina E, Serebryakova M, Dolgov A, Chulichkov A, Dobrov E, Baratova L (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. <https://doi.org/10.1111/j.1742-4658.2011.08392.x>
- Shtykova EV, Baratova LA, Fedorova NV, Radyukhin VA, Ksenofontov AL, Volkov VV, Shishkov AV, Dolgov AA, Shilova LA, Batishchev OV, Jeffries CM, Svergun DI (2013) Structural analysis of influenza A virus matrix protein M1 and its self-assemblies at low pH. *PLoS ONE* 8:e82431. <https://doi.org/10.1371/journal.pone.0082431>
- Shtykova EV, Dadinova LA, Fedorova NV, Golanikov AE, Bogacheva EN, Ksenofontov AL, Baratova LA, Shilova LA, Tashkin VY, Galimzyanov TR, Jeffries CM, Svergun DI, Batishchev OV (2017) Influenza virus matrix protein M1 preserves its conformation with pH, changing multimerization state at the priming stage due to electrostatics. *Sci Rep* 7:N16793. <https://doi.org/10.1038/s41598-017-16986-y>
- Hofer CT (2019) Structural determinants of the interaction between influenza A virus matrix protein M1 and lipid membranes. *Biochim Biophys Acta Biomembr* 1861:1123–1134. <https://doi.org/10.1016/j.bbmem.2019.03.013>
- Kordyukova LV, Serebryakova MV, Baratova LA, Veit M (2008) S acylation of the hemagglutinin of influenza viruses: mass spectrometry reveals site-specific attachment of stearic acid to a transmembrane cysteine. *J Virol* 82:9288–9292. <https://doi.org/10.1128/JVI.00704-08>
- Mineev KS, Lyukmanova EN, Krabben L, Serebryakova MV, Shulepko MA, Arseniev AS, Kordyukova LV, Veit M (2013) Structural investigation of influenza virus hemagglutinin membrane-anchoring peptide. *Protein Eng Des Sel* 26:547–552. <https://doi.org/10.1093/protein/gzt034>
- Nault L, Vendrely C, Bréchet Y, Bruckert F, Weidenhaupt M (2013) Peptides that form β -sheets on hydrophobic surfaces accelerate surface-induced insulin amyloid aggregation. *FEBS Lett* 587:1281–1286. <https://doi.org/10.1016/j.febslet.2012.11.036>
- Wittig I, Braun HP, Schägger H Blue native PAGE. *Nat Protoc* 1:418–28. <https://doi.org/10.1038/nprot.2006.62>
- Greening DW, Simpson RJ (2011) Low-molecular weight plasma proteome analysis using centrifugal ultrafiltration. *Methods Mol Biol* 728. https://doi.org/10.1007/978-1-61779-068-3_6. :109–24
- Khrustalev VV, Khrustaleva TA, Poboinev VV, Stojarov AN, Kordyukova LV, Akunevich AA (2021) Spectra of tryptophan fluorescence are the result of co-existence of certain most abundant stabilized excited state and certain most abundant destabilized excited state. *Spectrochim Acta A Mol Biomol Spectrosc* 257:119784. <https://doi.org/10.1016/j.saa.2021.119784>
- Ghoorah AW, Devignes MD, Smaïl-Tabbone M, Ritchie DW (2013) Protein docking using case-based reasoning. *Proteins* 81:2150–2158. <https://doi.org/10.1002/prot.24433>
- Yakupova EI, Bobyleva LG, Vikhlyantsev IM, Bobylev AG (2019) Congo Red and amyloids: history and relationship. *Biosci Rep* 39:BSR20181415. <https://doi.org/10.1042/BSR20181415>
- Hirota S, Hattori Y, Nagao S, Taketa M, Komori H, Kamikubo H, Wang Z, Takahashi I, Negi S, Sugiura Y, Kataoka M, Higuchi Y (2010) Cytochrome c polymerization by successive domain swapping at the C-terminal helix. *Proc Natl Acad Sci U S A* 107:12854–12859. <https://doi.org/10.1073/pnas.1001839107>
- Chevallet M, Luche S, Rabilloud T (2006) Silver staining of proteins in polyacrylamide gels. *Nat Protoc* 1:1852–1858. <https://doi.org/10.1038/nprot.2006.288>
- Poboinev VV, Khrustalev VV, Khrustaleva TA, Kasko TE, Popkov VD (2022) The PentUnFOLD algorithm as a tool to distinguish the dark and the light sides of the structural instability of proteins. *Amino Acids*. <https://doi.org/10.1007/s00726-022-03153-5>

24. Chuang GY, Kozakov D, Brenke R, Comeau SR, Vajda S (2008) DARS (Decoys as the reference state) potentials for protein-protein docking. *Biophys J* 95:4217–4227. <https://doi.org/10.1529/biophysj.108.135814>
25. Sayle RA, Milner-White EJ (1995) RASMOL: biomolecular graphics for all. *Trends Biochem Sci* 20:374–376. [https://doi.org/10.1016/s0968-0004\(00\)89080-5](https://doi.org/10.1016/s0968-0004(00)89080-5)
26. Dassault Systemes BIOVIA (2017) Discovery Studio modeling Environment, Release 2017. Dassault Systèmes, San Diego, CA, USA
27. Kozłowski LP (2021) IPC 2.0: prediction of isoelectric point and pKa dissociation constants. *Nucleic Acids Res* 49:285–292. <https://doi.org/10.1093/nar/gkab295>
28. Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD, Wishart DS (2003) VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res* 31:3316–3319. <https://doi.org/10.1093/nar/gkg565>
29. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK—a program to check the stereochemical quality of protein structures. *J App Cryst* 26:283–291. <https://doi.org/10.1107/S0021889892009944>
30. Safo MK, Musayev FN, Mosier PD, Zhou Q, Xie H, Desai UR (2014) Crystal structures of influenza a virus matrix protein M1: variations on a theme. *PLoS ONE* 9:e109510. <https://doi.org/10.1371/Journal.Pone.0109510>
31. Liu Q, Kloss B, Skowronski ES, Height J, Love J (2010) Not published.
32. Lamiable A, Thevenet P, Rey J, Vavrusa M, Derreumaux P, Tuffery P (2016) PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res* 44:449–454. <https://doi.org/10.1093/nar/gkw329>
33. Graf M, Galera García R, Wätzig H (2005) Protein adsorption in fused-silica and polyacrylamide-coated capillaries. *Electrophoresis* 26:2409–2417. <https://doi.org/10.1002/elps.200410360>
34. Groenning M (2010) Binding mode of thioflavin T and other molecular probes in the context of amyloid fibrils—current status. *J Chem Biol* 3:1–18. <https://doi.org/10.1007/s12154-009-0027-5>
35. Selzer L, Su Z, Pintilie GD, Chiu W, Kirkegaard K (2020) Full-length three-dimensional structure of the influenza a virus M1 protein and its organization into a matrix layer. *PLoS Biol* 18:e3000827. <https://doi.org/10.1371/journal.pbio.3000827>
36. Smith DE, Clemençon B, Hediger MA (2013) Proton-coupled oligopeptide transporter family SLC15: physiological, pharmacological and pathological implications. *Mol Aspects Med* 34:323–336. <https://doi.org/10.1016/j.mam.2012.11.003>
37. World Health Organization (2022) Influenza update № 416. <https://www.who.int/publications/m/item/influenza-update-n-416>. Accessed 19 April 2022
38. Klein NP, Fireman B, Goddard K, Zerbo O, Asher J, Zhou J, King J, Lewis N (2020) Vaccine effectiveness of cell-culture relative to egg-based inactivated influenza vaccine during the 2017–18 influenza season. *PLoS ONE* 15:e0229279. <https://doi.org/10.1371/journal.pone.0229279>
39. Panatto D, Lai PL, Mosca S, Lecini E, Orsi A, Signori A, Castaldi S, Pariani E, Pellegrinelli L, Galli C, Anselmi G, Icardi G, Team CI (2020) Influenza vaccination in Italian healthcare workers (2018–2019 season): strengths and weaknesses. Results of a cohort study in two large Italian hospitals. *Vaccines* 8:119. <https://doi.org/10.3390/vaccines8010119>
40. Moise L, Biron BM, Boyle CM, Yilmaz NK, Jang H, Schiffer C, Ross TM, Martin WD, De Groot AS (2018) T cell epitope engineering: an avian H7N9 influenza vaccine strategy for pandemic preparedness and response. *Hum Vaccin Immunother* 14:2203–2207. <https://doi.org/10.1080/21645515.2018.1495303>
41. Plant EP, Manukyan H, Sanchez JL, Laassri M, Ye Z (2020) Immune pressure on polymorphous influenza B populations results in diverse hemagglutinin escape mutants and lineage switching. *Vaccines* 8:125. <https://doi.org/10.3390/vaccines8010125>
42. Benton DJ, Nans A, Calder LJ, Turner J, Neu U, Lin YP, Ketelaars E, Kallewaard NL, Corti D, Lanzavecchia A, Gamblin SJ, Rosenthal PB, Skehel JJ (2018) Influenza hemagglutinin membrane anchor. *Proceedings of the National Academy of Sciences U S A* 115:10112–10117. <https://doi.org/10.1073/pnas.1810927115>
43. Brett K, Kordyukova LV, Serebryakova MV, Mintaev RR, Alexeevski AV, Veit M (2014) Site-specific S-acylation of influenza virus hemagglutinin: the location of the acylation site relative to the membrane border is the decisive factor for attachment of stearate. *J Biol Chem* 289:4978–4989. <https://doi.org/10.1074/jbc.M114.586180>
44. Peukes J, Xiong X, Erlendsson S, Qu K, Wan W, Calder LJ, Schraidt O, Kummer S, Freund SMV, Kräusslich HG, Briggs JAG (2020) The native structure of assembled influenza a virus M1 matrix protein. *Nature* 587:495–498. <https://doi.org/10.1038/s41586-020-2696-8>
45. Mintaev RR, Alexeevski AV, Kordyukova LV (2014) Co-evolution analysis to predict protein–protein interactions within influenza virus envelope. *J Bioinform Comput Biol* 12:1441008. <https://doi.org/10.1142/S021972001441008X>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.