

Article

Self-Assembly and Conformational Change in the Oligomeric Structure of the Ectodomain of the TBEV E Protein Studied via X-ray, Small-Angle X-ray Scattering, and Molecular Dynamics

Petr V. Konarev^{1,2}, Anna V. Vlaskina¹, Dmitry Korzhenevskiy³, Tatiana V. Rakitina^{1,4} , Dmitry Petrenko¹, Yulia Agapova¹, Yulia Kordonskaya¹  and Valeriya R. Samygina^{1,2,*} 

¹ NRC “Kurchatov Institute”, 123182 Moscow, Russia; taniarakitina@yahoo.com (T.V.R.)

² FRSC “Crystallography and Photonics” RAS, 119333 Moscow, Russia

³ Federal State Budget Institution “Federal Center for Brain Research and Neurotechnologies”, Federal Medical and Biological Agency of Russia, 117513 Moscow, Russia

⁴ Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, 117997 Moscow, Russia

* Correspondence: lera@crys.ras.ru

Abstract: The determination of the three-dimensional structures of viral proteins is a necessary step both for understanding the mechanisms of virus pathogenicity and for developing methods to combat viral infections. This study aimed to explore the folding and oligomeric state of the major component of the virion surface of the tick-borne encephalitis virus (TBEV), the ectodomain of the envelope E protein (ectoE), which was expressed in *E. coli* in a soluble form and purified from inclusion bodies as a mixture of dimeric and monomeric forms. The time-dependent assembly of monomers into dimers was detected using size-exclusion chromatography. An X-ray diffraction study of the ectoE crystals grown at pH 4.5 confirmed the dimeric folding of the recombinant protein typical for ectoE. The ability of ectoE dimers to self-assemble into tetramers was detected via small-angle X-ray scattering (SAXS) in combination with molecular dynamics. Such self-assembly occurred at protein concentrations above 4 mg/mL and depended on the pH of the solution. In contrast to stable, specific dimers, we observed that tetramers were stabilized with weak intermolecular contacts and were sensitive to environmental conditions. We discovered the ability of ectoE tetramers to change conformation under crystallization conditions. These results are important for understanding the crystallization process of viral proteins and may be of interest for the development of virus-like particles.

Keywords: ectodomain of TBEV protein E; self-assembly; SAXS; X-ray analysis; molecular dynamics



Citation: Konarev, P.V.; Vlaskina, A.V.; Korzhenevskiy, D.; Rakitina, T.V.; Petrenko, D.; Agapova, Y.; Kordonskaya, Y.; Samygina, V.R. Self-Assembly and Conformational Change in the Oligomeric Structure of the Ectodomain of the TBEV E Protein Studied via X-ray, Small-Angle X-ray Scattering, and Molecular Dynamics. *Crystals* **2023**, *13*, 1676. <https://doi.org/10.3390/cryst13121676>

Academic Editor: Abel Moreno

Received: 1 November 2023

Revised: 4 December 2023

Accepted: 8 December 2023

Published: 12 December 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Viruses are complex biological machines. To understand the mechanism of their functioning, information obtained using structural methods such as cryoEM, NMR, or X-ray crystallography is required. X-ray diffraction analysis is most often used to determine the structures of individual viral proteins, in particular, to study the structures of proteins that form the viral envelope or capsid. In the case of envelope proteins, recombinant soluble fragments, such as ectodomains or their individual domains, are often obtained. The viral envelope or capsid proteins are inherently prone to self-association. This property facilitates crystallization and helps in studying the structure-function relationships of viruses.

Flaviviruses are among the important emerging human viruses. The E protein is a membrane protein with a transmembrane helix located at the C-terminus, which is embedded in the membrane. The soluble part of the envelope E protein consists of three domains (Supplementary Figure S1). Protein E plays an important role in virion formation and cell infection [1]. The surface of the virion is composed of 180 copies of the E

protein ectodomain, which are arranged into head-to-tail dimers [2]. The ectodomain of the E protein on the virion surface retains the ability to undergo conformational rearrangements, which play an essential role in the functioning of flaviviruses [1]. Several crystal structures of full-length E protein ectodomains (residues 1–395) are known for different flaviviruses. They include Louping ill virus (PDB ID 6J5C) [3], West Nile virus (PDB ID 2I69, 2HG0) [4,5], Yellow fever virus (PDB ID 6IW4, 6IW5) [6], Dengue fever virus (PDB ID 1OAN, 1UZG, and 4UTC) [7–9], Zika virus (PDB ID 6JHM) [10], Usutu virus (PDB ID 6A0P) [11], and Japanese encephalitis virus (PDB ID 5MV1) [12]. The E protein ectodomains of the Louping ill virus and the Yellow Fever virus were expressed in *E. coli*. Other proteins were expressed in eukaryotic or insect expression systems. For TBEV, the structure of the full-length trypsinolysis-derived ectodomain (Neudoerfl strain, Central European subtype) was determined (PDB ID 1SVB) [13]. Recently, the structure of the recombinant full-length E protein ectodomain of the same strain in complex with a precursor fragment expressed in an insect expression system was determined (PDB ID 7QRE) [14].

The success of X-ray analysis depends on the quality of the protein sample. The preliminary characterization of proteins prior to crystallization experiments can increase the probability of obtaining diffraction-quality crystals [15]. The absence of irregular aggregates is a necessary baseline criterion that samples should meet. In some works, it was shown that when a precipitant used in crystallization is added to proteins, the oligomers formed in the solution are the structural units of the growing crystal [16,17]. There are also examples when a high protein concentration and a certain pH are sufficient for crystallization [18], suggesting the formation of oligomers that serve as ‘nucleation precursors’ for crystal formation [19].

In our work, we used small-angle X-ray scattering (SAXS) to verify the correct folding of the E protein ectodomain of TBEV (strain Sofjin) expressed in *E. coli*. The monodispersity of the ectoE and its oligomeric form were also checked prior to crystallization experiments. SAXS allowed the determination of not only the size but also the shape of the molecules. Although the biological form of ectoE is a dimer, we found that in the stabilization buffer at pH 8.0, the protein tended to form tetramers (dimer of dimers) in solution. To determine whether the detected tetramers were the building units of crystal packing, we compared structures formed in solution at pH 8.0 with clusters identified in ectoE structures obtained from crystals grown at two different pHs: 8.0 (strain Neudoerfl, pdb code 1SVB) and 4.5 (this work) and tested the stability of different types of predicted tetramers using molecular dynamics (MD).

2. Materials and Methods

2.1. E Protein Ectodomain Isolation and Purification, Crystallization, and Structure Refinement

The protein expressed in *Escherichia coli* was isolated and purified as described in [20]. After the refolding procedure, the target protein was purified via size-exclusion chromatography (SEC) on a Superdex 75 10/300 GL column (GE Healthcare, Uppsala, Sweden). The buffer for SEC was composed of 20 mM of Tris-Cl, 150 mM of NaCl, and pH 8.0. SEC was performed either immediately or 3–4 days after refolding. The experiments were carried out in duplicate or triplicate. Before and after SEC, the protein was stored at +4 °C.

The protein was used either immediately or was stored at +4 °C. Crystallization was performed as described in [19] at a protein concentration of 4 mg/mL. Crystals were grown in 1.6–1.8 M of Ammonium Sulphate, 2% MPD, 0.1 M of citric acid, and pH 4.5. The data set and the initial crystallographic model obtained in [20] were used to refine the structure. The structure was refined to a resolution of 3.2 Å using the Refmac [21] and Coot [22] programs. The diffraction data processing and refinement statistics are given in Table 1. Structure comparison and analysis were performed using the program Coot and CCP4 package [23].

Table 1. EctoE diffraction data and refinement statistics.

Protein	ectoE
Data collection	
Wavelength	0.98888
Resolution (Å)	50.0–3.20 (3.28–3.2) *
Space group	P4 ₁ 32
Unit cell parameters	
a = b = c (Å)	165.1
$\alpha = \beta = \gamma$ (°)	90
Completeness (%)	99.9 (100)
I/σ (I)	12.3 (2.43)
Multiplicity	3.8 (3.1)
R _{merge} (%)	15.8 (94.2)
CC _{1/2} (%)	99.9 (93.3)
No. unique reflections	50,107 (6384)
Refinement	
R _{work} /R _{free} (%)	20.2/28.6
No. atoms	
Protein	3028
Water	8
Average B factors (Å ²)	
Protein	121.9
Water	72.78
R.m.s. deviations	
Protein bond lengths (Å)	0.014
Protein bond angles (°)	2.12
Ramachandran analysis	-
Favored (%)	79.1
Allowed region (%)	12.5
Outliers (%)	9.5
PDB code	8R2L

* The data for the highest-resolution shell are shown in parentheses.

2.2. SAXS

SAXS measurements of two fractions of the TBEV protein E ectodomain, corresponding to monomer (fraction 1) and dimer (fraction 2) (according to SEC results), were performed at the BioMUR beamline of the Kurchatov Synchrotron Radiation Source (National Research Center “Kurchatov Institute”, Moscow, Russia) [24]. The concentration of fractions 1 and 2 was 2.0 mg/mL and 4.0 mg/mL, respectively. The samples were placed in quartz capillaries with a diameter of 1.5 mm. The sample-to-detector distance was 700 mm, which corresponds to a momentum transfer range of $0.15 \text{ nm}^{-1} < s < 3.8 \text{ nm}^{-1}$, where $s = (4\pi\sin\theta)/\lambda$, 2θ is the scattering angle, and $\lambda = 0.1445 \text{ nm}$, the X-ray wavelength. The signal was recorded using a two-dimensional pixel detector, PILATUS3 1 M (Dectris, Baden, Switzerland). The exposure time was 10 min. The measurements were performed at room temperature. The two-dimensional scattering pattern was averaged along the radial direction using the program Fit2D [25]. The subsequent SAXS data processing and analysis

were completed using the programs PRIMUS, GNOM, DAMMIN, CRY SOL, SASREF, OLIGOMER, and SUPALM from the package ATSAS [26]. The background scattering of the buffer was subtracted from the scattering curves of the sample solution using the program PRIMUS. The particle radius of gyration, R_g , was estimated within the Guinier approximation. The maximum particle size, D_{max} , and the distance distribution function $p(r)$ were computed using the indirect Fourier transform with the program GNOM. The excluded volume of the hydrated particle V_p was calculated using the Porod approximation. Three-dimensional *ab initio* shapes of ectoE from the two fractions were restored using the program DAMMIN. The scattering patterns from the monomeric model of the protein E ectodomain at pH 8.0 (PDB ID: 1SVB) and the compact/elongated tetramer models (derived from the crystallographic structures of TBEV at pH 4.5 (this work) and pH 8.0 (PDB ID: 1SVB) were calculated with the program CRY SOL. The best fits from their mixtures were obtained using the program OLIGOMER, which represents the scattering profile as a linear combination of curves computed from the compact and elongated tetramer models. Given the scattering curves of these components, the program OLIGOMER finds their volume fractions by solving a system of linear equations to minimize the discrepancy χ^2 between the experimental and calculated scattering curves. The discrepancy χ^2 is defined as follows:

$$\chi^2 = \frac{1}{N-1} \sum_{j=1}^N \left[\frac{I_{exp}(s_j) - cI_{calc}(s_j)}{\sigma(s_j)} \right]^2 \quad (1)$$

where N is the number of experimental points, c is a scaling factor, $I_{exp}(s_j)$ and $\sigma(s_j)$ are the experimental scattering intensities and associated errors, and $I_{calc}(s_j)$ is the calculated intensity from the oligomeric mixture.

Alternatively, the quaternary structure of fraction 2 of ectoE was restored via rigid body modeling using the program SASREF, which uses a simulated annealing protocol to generate an interconnected assembly of subunits without steric clashes that fit the scattering data. The relative positions and orientations of the two individual crystallographic dimers were refined by applying P2 symmetry. The models were superimposed on each other using the program SUPALM.

2.3. Molecular Modeling

The PROPKA server (Version 3.2 [27]) was used to determine the ionization states of amino acid residues at a pH of 4.5. Molecular dynamics (MD) simulations were performed using the GROMACS 2021 software package [28]. The Amber ff99SB-ILDN field [29] was selected due to the refined torsion potentials for certain atom groups. Each tetramer was placed in the center of a triclinic simulation box, and the dimensions of the box were set to ensure a minimum distance of 1 nm between the edge of the box and any protein atom. The simulation box was filled with an explicit solvent using the TIP4P-Ew [30] model. A negligible number (about 90) of chloride ions were added to neutralize the total charge of the box. Before performing productive MD calculations, the energy of the system in each box was minimized using the steepest descent method for 50,000 steps until the force on any atom dropped below 1000 kJ/(M·nm⁻²). Subsequently, the boxes underwent a 100 ps thermostat and a 100 ps barostat process in the NVT- and NPT-ensembles using the modified Berendsen (V-rescale) [31] and Parrinello–Raman [32] algorithms, respectively. Productive MD simulation was conducted using V-rescale and Parrinello–Raman barostats in the isothermal–isobaric ensemble. The standard leap-frog algorithm [33] with a time step of 2 fs was used for integration. Tetramer bond lengths were maintained using the LINCS algorithm, and long-range electrostatic interactions were processed with the PME summation method. The trajectories produced by each of the three tetramers lasted 100 ns at 23 °C.

The command “*gmx trjconv*” with the flag “*-pbc nojump*” was used to prepare the trajectories for analysis. For structural alignment, the atomic positions after the barostat step were assigned as the reference protein structure. To fit the tetramer trajectories to the reference ones, the command “*gmx trjconv*” with the flag “*-fit rot + trans*” was executed

to eliminate parallel transfers and rotations around the protein axis. The commands “*gmx rmsf*” and “*gmx rms*” were used to compute RMSF (root-mean-square fluctuation) and RMSD (root-mean-square deviation), respectively.

3. Results and Discussion

3.1. Size-Exclusion Chromatography

The E protein ectodomain was produced in *E. coli*, which is one of the most common and widely used protein expression systems for structural studies, and purified from inclusion bodies as described in [20]. However, this expression system does not always provide the necessary conditions for the correct folding of viral proteins. To preliminarily assess the stability and oligomeric state of recombinant ectoE in solution, we used SEC, which was performed immediately and also on the third and fourth days after protein refolding (Figure 1a). According to Figure 1a, both protein samples (freshly refolded and after 3 days of storage) were a mixture of dimers and monomers released in volumes corresponding to a protein with a molecular weight of 51 ± 1 and 36.5 ± 2.5 kDa. These values were close to the molecular weights calculated for ectoE dimers and monomers, respectively. After refolding, the monomer fraction predominated, while on the 3rd day, a shift in equilibrium towards the dimer was observed. After storage for four days, the protein became completely dimeric, with no evidence of the monomer (Figure 1a).

3.2. X-ray Analysis

Both monomeric and dimeric fractions gave crystals under the same crystallization conditions at pH 4.5 [20]. This agrees well with our observations from the SEC experiment, as crystals appeared at a higher concentration (4 mg/mL) compared to that used in the SEC, and the crystal growth took several days [20]. The overall 3.2 Å structure of the E protein ectodomain of the strain *Soffin*, obtained at pH 4.5 (pdb code 8R2L), is very similar to the other known structures of the full-length ectodomain of TBEV E protein [14] (Figure 1b). The root-mean-square deviations (RMSDs) of C α atoms between our structure and the superimposed ectodomain of TBEV E protein (strain Neudorfl) complexed with the pr protein (crystallized at a similar pH) [14] is 1.42 Å. The largest differences are observed in the pr protein binding region (Supplementary Figure S2). However, our structure has a more noticeable difference from the structure of the ectodomain of protein E of the Neudorfl strain obtained via thripsinolysis from TBEV virion and crystallized at pH 8.0 (pdb code 1SVB). The RMSD between structures at pH 4.5 and 8.0 is 1.59 Å (Figure 1c), with the largest difference observed in regions 145–163 of domain I and 202–213 of domain II; the tilt of domain II is also noticeable. Acidic pH triggers conformational changes in E proteins on the surface of viral particles. These rearrangements are an essential part of the virus invasion mechanism. It can explain observed differences.

A comparison of the X-ray dimer of the ectodomain of the E protein with the dimer in the asymmetric part of the cryoEM TBEV structure (Kuutsalo-14 strain, pdb code 7z51) reveals a more flattened shape (due to the tilt of domain II) and some difference in the C α -trace of regions in all three domains associated with crystal packing (see Supplementary Figure S3). Differences in the area of the conservative glycosylation site (Asn154) are mainly attributed to the expression system. The statistics of the diffraction data and structure refinement are summarized in Table 1. This structure, as well as the 1SVB structure, was used as a template for further modeling of E protein oligomers in solution.

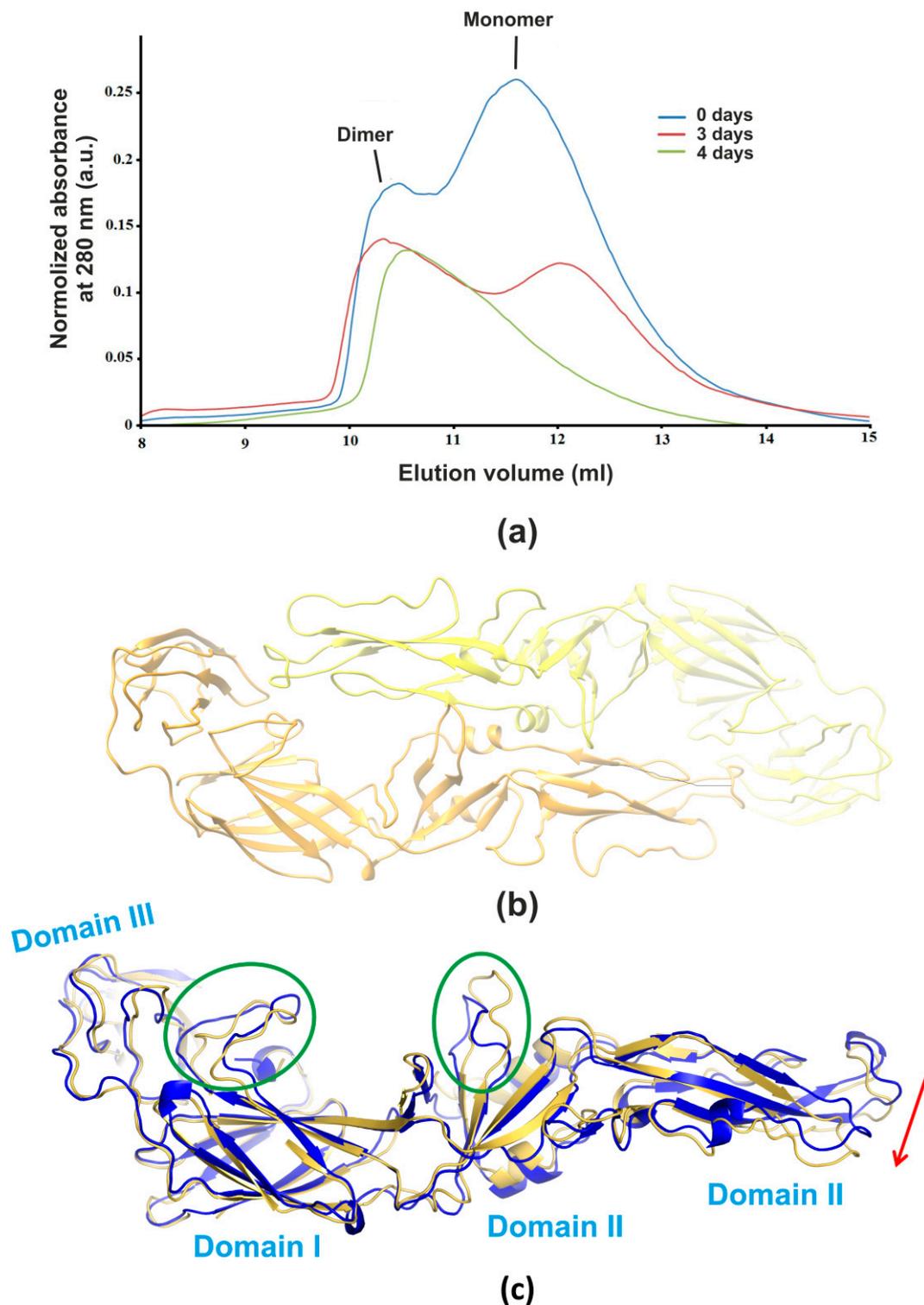


Figure 1. Dimer of the E protein ectodomain. (a) SEC chromatograms obtained on the 1st day and 3 and 4 days after the protein isolation from inclusion bodies. The protein (1 mg/mL) was loaded on a Superdex 75 10/300GL column equilibrated with a buffer of 20 mM of Tris-Cl, 150 mM of NaCl, and pH 8.0. (b) The X-ray structure in a cartoon representation. The molecule in the asymmetric unit is shown in yellow, and the molecule related by a twofold axis is shown in orange. (c) Superposition of dimers from the 8R2L (yellow) and 1SVB (blue) structures. The largest differences in α -trace (regions 145–161 and 202–212) are highlighted by green ovals. The direction of domain II tilt is indicated by the arrow.

3.3. SAXS

The ectoE was examined for monomer folding, sample quality, and the oligomeric state of the sample in solution via SAXS. First, fresh monomer and dimer fractions were studied at a concentration similar to that at which the crystallization was performed in the buffer used for SEC (pH 8.0). The overall structural parameters obtained from the SAXS data are summarized in Table 2. The experimental SAXS patterns are shown in Figure 2 (curve 1). The distance distribution function $p(r)$ displayed an asymmetric tail typical of elongated particles (Figure 2, inset).

Table 2. Overall structural parameters of the ectodomain of the TBEV envelope protein E derived from SAXS data. Here, R_g —the radius of gyration, D_{max} —the maximum particle size, V_p —the excluded volume of the hydrated particle, and MM_{exp} —the experimental molecular mass.

	R_g , nm	D_{max} , nm	V_p , nm ³	MM_{exp} , kDa
EctoE, fraction 1	3.95 ± 0.05	12.5 ± 0.5	67 ± 7	45 ± 5
EctoE, fraction 2	5.94 ± 0.15	20.0 ± 1.0	235 ± 15	160 ± 13

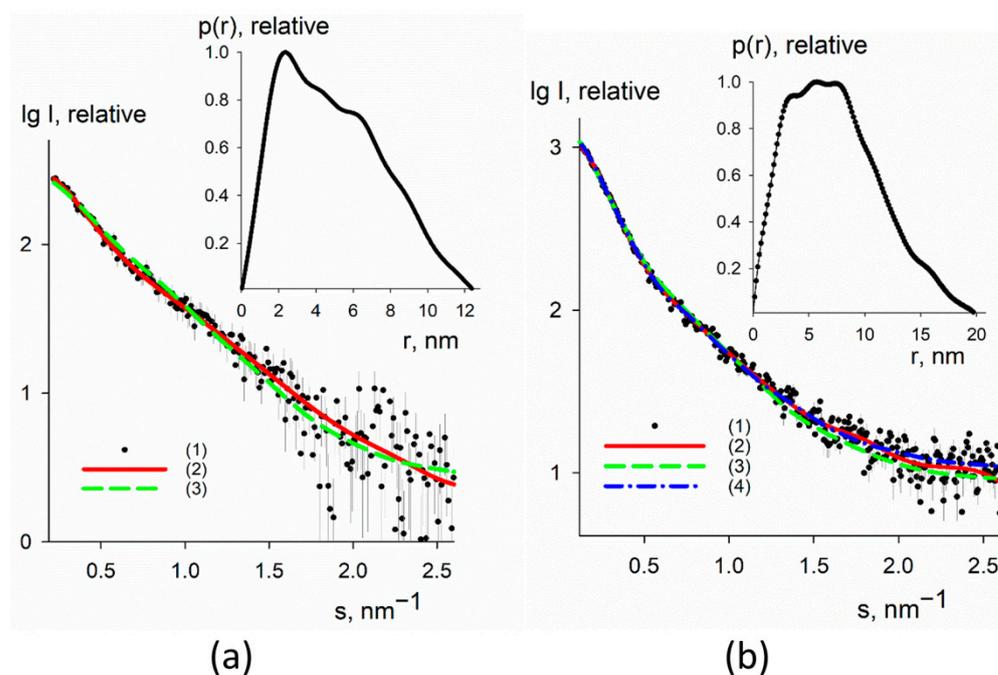


Figure 2. Experimental SAXS patterns of the ectodomain of the TBEV envelope protein E: (a) monomeric fraction and (b) tetrameric fraction. Experimental data (dots with error bars, curve 1), best fits from the *ab initio* shapes obtained with the program DAMMIN (solid red lines, curve 2), the calculated intensities from the crystallographic models (1SVB PDB for (a) and a mixture of compact and elongated tetramers for (b) at pH 8.0) (green dashed lines, curve 3), and the best fit from the rigid body tetramer model obtained with the program SASREF (blue dashed-dotted line, curve 4). The distance distribution functions $p(r)$ calculated with the program GNOM are shown in the inset.

The radius of gyration and the maximum particle size of fraction 1 ($R_g = 3.95 \pm 0.05$ nm and $D_{max} = 12.5 \pm 0.5$ nm) agree well with the theoretical values from the crystallographic model (PDB ID: 1SVB). The *ab initio* reconstructed shape fits the experimental SAXS data with $\chi^2 = 1.04$ (Figure 2, curve 2) and overlaps with the monomer model (Figure 3a). This fit indicates the correct folding of the E protein monomer and the biological relevance of the selected genetically engineered construct.

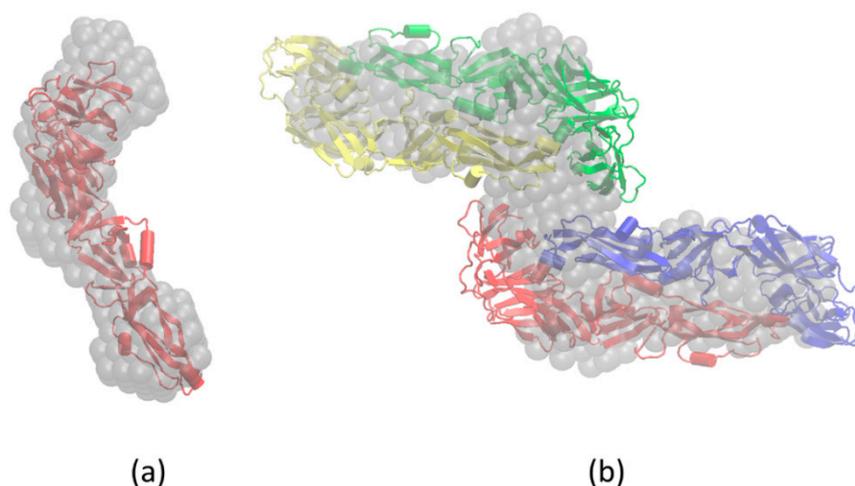


Figure 3. Ab initio and rigid body models of the ectodomain of the TBEV envelope protein E: (a) monomeric fraction and (b) tetrameric fraction. The *ab initio* shapes were restored with the program DAMMIN (gray beads) ($\chi^2 = 1.02$), and the rigid body model of the tetramer in cartoon representation was obtained using the program SASREF ($\chi^2 = 1.35$). Monomers are colored red, green, blue and yellow. The models were superimposed using the program SUPALM.

Meanwhile, the radius of gyration and the maximum particle size for fraction 2 of the E protein ectodomain ($R_g = 5.94 \pm 0.15$ nm and $D_{max} = 20.0 \pm 1.0$ nm) significantly exceed the theoretical values for the crystallographic models of the monomer and the dimer.

Different models were used to approximate this experimental curve. Figure 4 shows two variants of tetramers, elongated (ECT) or compact (CCT), that were generated from the crystallographic models at pH 8.0 (PDB ID 1SVB) and pH 4.5 (this work).

It can be seen that the experimental radius of gyration and the maximum particle size of fraction 2 are significantly larger than those of the compact tetramer models at both pHs ($R_g = 5.10$ nm and $D_{max} = 17.5$ nm, respectively), but significantly smaller than those of the elongated tetramer models ($R_g = 7.10$ nm and $D_{max} = 25.4$ nm, respectively). Meanwhile, the excluded volume of the hydrated particle and its molecular mass ($V_p = 235 \pm 15$ nm³, $MM_{exp} = 160 \pm 13$ kDa) correlate well with the theoretical molecular weight of the tetramer ($MM_{th} = 173.2$ kDa).

As expected, none of the crystallographic tetramer models individually were able to fit the data well (see Table 3). However, the mixture containing compact and elongated crystallographic tetramers (derived from the E protein ectodomain structure at pH 8.0) with volume fractions of species of 45% and 55%, respectively, yielded a reasonable fit to the data with $\chi^2 = 1.42$ (Figure 2b, curve 3).

Table 3. Overall structural parameters and the quality of the fit of the tetrameric crystallographic and rigid body models of ectoEto SAXS data. Here, R_g —the radius of gyration, D_{max} —the maximum particle size, and χ^2 —discrepancies between the calculated scattering curves and the experimental data. The best fits to the data from the mixture of crystallographic tetramers (pH 8.0) (obtained using the program OLIGOMER) and the rigid body model (obtained using the program SASREF) are shown in bold.

Models/(Programs)	R_g , nm	D_{max} , nm	χ^2
ECT, pH 8.0/ (CRYSOL)	7.1	25.4	4.62
CCT, pH 8.0/ (CRYSOL)	5.1	17.5	5.75
Mixture of crystallographic tetramers pH 8.0 (45% CCT + 55%ECT)/ (OLIGOMER)	--	--	1.42

Table 3. Cont.

Models/(Programs)	R_g , nm	D_{max} , nm	χ^2
ECT, pH 4.5/ (CRY SOL)	7.0	24.8	6.31
CCT, pH 4.5/ (CRY SOL)	4.5	15.0	17.97
Mixture of crystallographic tetramers pH 4.5 (25% CCT + 75% ECT)/ (OLIGOMER)	--	--	4.26
Rigid body model/(SASREF)	6.0	21.1	1.35

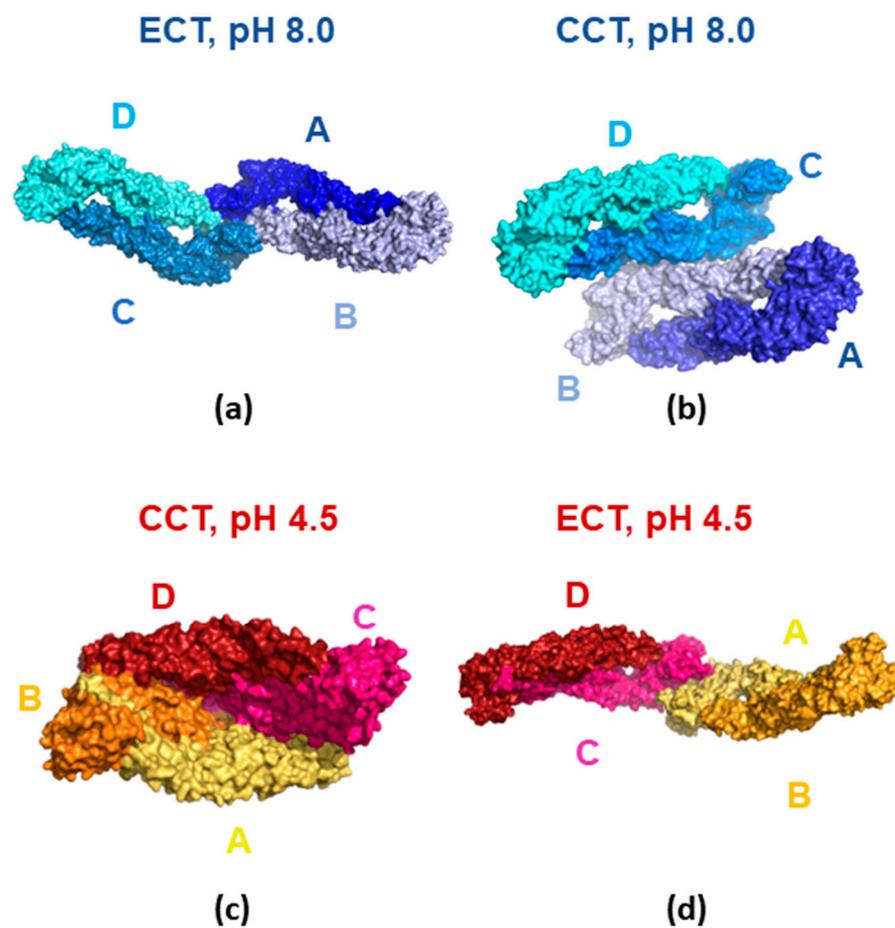


Figure 4. Types of tetramers in different crystal structures in surface representation (a,b) from the structure 1SVB at 8.0 pH; the Subunits A, B, C, and D of tetramers are colored bright blue, lilac, sky-blue, and cyan, respectively, (c,d) from X-ray structure at pH 4.5 (this work). Subunits A, B, C, and D of tetramers are colored yellow, orange, magenta, and brown–red, respectively. ECT is an elongated crystallographic tetramer, and CCT is a compact crystallographic tetramer.

At the same time, assuming the presence of a single metastable conformation of the tetramer in solution, one can alternatively apply *ab initio* and rigid body modeling approaches. Indeed, the restored *ab initio* shape and the rigid body model of the tetramer provided good fits to the experimental data with $\chi^2 = 1.02$ and 1.35, respectively (Figure 2b, curves 2 and 4). They also overlap well with each other (Figure 3b). The obtained tetramer has an intermediate conformation lying between the compact and elongated crystallographic structures.

Since both approaches provide almost the same fit qualities, it is impossible to reliably distinguish between them from each other without involving additional information.

Therefore, we assume that the system can adopt several tetrameric states depending on the physicochemical conditions of the environment, and it is easier for tetramer conformations with ‘loose’ interfaces between dimers (as is the case for the rigid body model).

Thus, according to the SAXS study, the ectodomain of the TBEV envelope protein E expressed in *E. coli* may be present in solution for a limited time as a monomer (at low concentration) or elongated tetramers (possibly with a small fraction of compact tetramers).

3.4. Molecular Modeling

To distinguish between two possible oligomeric models of ectoE (*ab initio* SAXS model or ECT/CCT mixture at pH 8.0), we employed molecular dynamics to test the stability of selected tetramers (*ab initio* tetramer, CCT, and ECT at pH 8.0) in solution at conditions close to the crystallization conditions.

The stability of molecules is routinely assessed using the RMSF graphs of C_{α} atoms. The RMSF captures, for each atom, the fluctuation around its average position and indicates the protein flexibility. The higher the RMSF value, the more mobile the atom is. The RMSF values of three tetramers are presented in Figure 5: A (ECT, pH 8.0), B (*ab initio* SAXS model), and C (CCT, pH 8.0). Tetramer A is found to be the most stable, as it is clearly seen that it has the lowest RMSF values (blue curve in Figure 5).

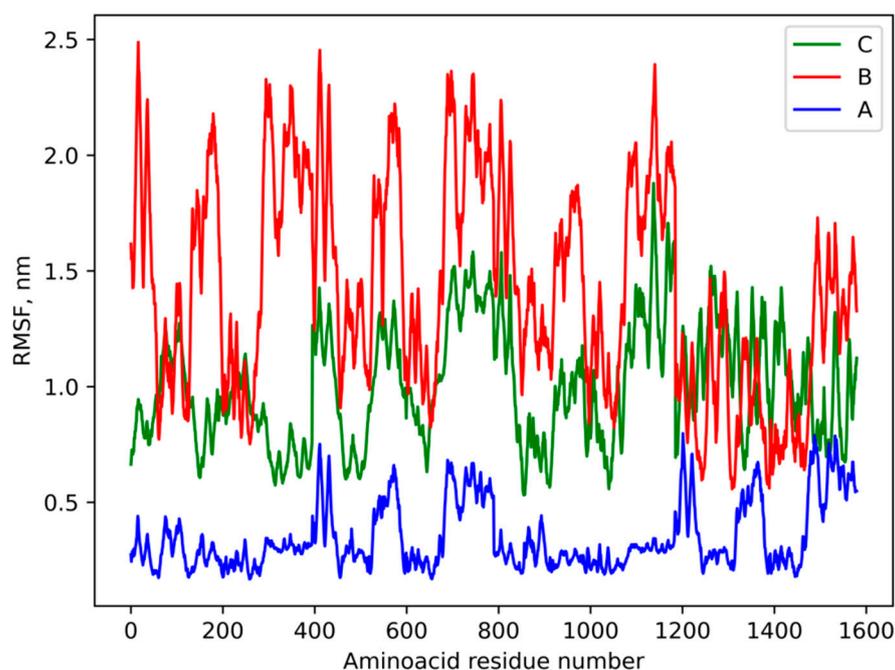


Figure 5. The RMSF of C_{α} atoms of three different tetramers, A–C, in water. Tetramer A (blue curve) is an elongated crystallographic tetramer from the structure 1SVB (Figure 4); tetramer B (red curve) is an *ab initio* SAXS model (Figure 3b); and tetramer C (green curve) is a compact crystallographic tetramer from the structure 1SVB (Figure 4).

We used the root-mean-square deviation (RMSD) as a standard measure to evaluate the average distance between coordinates and examined the change in the structure of tetramers during the simulations. The RMSDs of all C_{α} atoms are utilized to determine the deviations of all atoms simultaneously as a function of time, whereas the RMSFs represent the fluctuations around the average position of each atom during the dynamics. It can be seen in Figure 6 that the RMSF and RMSD data are consistent with each other, since tetramer A exhibits the structure most similar to the initial one (as in the crystal) throughout the simulation, while tetramers B and C immediately undergo significant transformations.

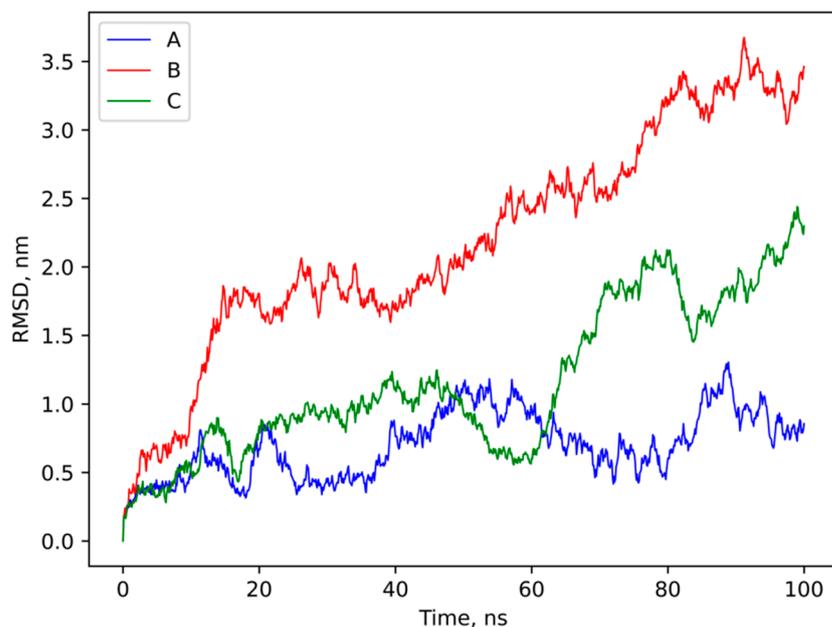


Figure 6. The RMSD of C_{α} atoms of three different tetramers A–C in water. The tetramer labeling and color scheme are the same as in Figure 5.

Moreover, the visual inspection (Figure 7) of the trajectories revealed that during 100 ns trajectories in water, the bonds between the subunits of tetramers B and C disappear and these molecules dissociate.

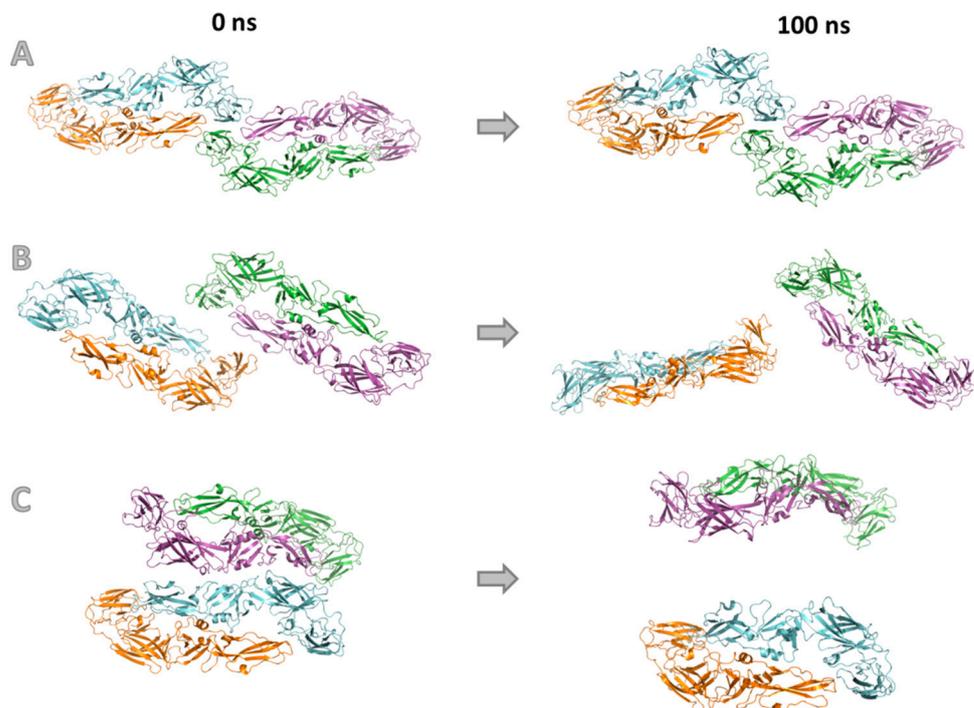


Figure 7. Tetramers (A–C) before (0 ns, on the left) and after (100 ns, on the right) MD simulations. The tetramer labeling scheme is the same as in Figure 5. Different monomers in tetramer are colored green, purple, blue and orange.

As a result, when the pH changed from 8.0 to 4.5, tetramer A (ECT, pH 8.0) proved to be the most stable. However, this most stable tetramer variant, which can exist in solution prior to crystal formation, is not the crystal packing unit of the X-ray structure at

crystallization pH. Together with the SAXS results showing that only a mixture of tetramer A and tetramer C describes protein clusters in solution, it can be assumed that the ab initio SAXS model is the most relevant. It is very likely that this tetramer with ‘loose’ interfaces between dimers is transient.

4. Conclusions

In general, it can be recommended to perform SAXS measurements prior to crystallization as a useful sample quality control, along with the SEC and the DLS techniques, which are commonly used. The DLS is based on the approximation of a spherical mathematical model and cannot tell anything about the shape of the molecules, while many proteins are not completely globular. In the case of E protein, we observed apparently specific oligomerization rather than nonspecific oligomerization, as one might think based on the large particle diameter. This was confirmed by the fact that these oligomers can yield high-quality diffracting crystals. We also found that this viral protein can self-associate in solution into an oligomer (tetramer) higher than the biologically functional dimer. Under certain conditions, this tetramer can be reconstructed through a rearrangement of the dimers. The rearrangement of subunits into an oligomer upon modification was found earlier in [34]. This finding is an interesting detail of the crystallization mechanism of the viral protein, as well as the crystal formation process, and may be useful for the development of virus-like particles (as a vaccine platform or the basis for a test system) based on self-association [35].

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/cryst13121676/s1>, Figure S1: Scheme of protein E organization. Figure S2: Superposition of ectoE and 7QRE. Figure S3: Influence of crystal packing on ectoE dimer conformation.

Author Contributions: Conceptualization, V.R.S.; methodology, V.R.S. and P.V.K.; validation, D.K.; formal analysis, V.R.S. and P.V.K.; investigation, T.V.R., A.V.V., Y.A., D.P. and D.K.; resources, V.R.S.; data curation, P.V.K. and V.R.S.; writing—original draft preparation, V.R.S., P.V.K. and Y.K.; writing—review and editing, V.R.S., P.V.K. and Y.K.; visualization, V.R.S., P.V.K. and Y.K.; supervision, V.R.S.; project administration, V.R.S.; funding acquisition, V.R.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NRC “Kurchatov Institute” in the part of molecular dynamics and was supported the Russian Ministry of Science and Education within the framework of the Federal Scientific and Technical Program for the Development of Synchrotron and Neutron Research and Research Infrastructure for 2019–2027 (Agreement No. 075-15-2021-1355 (12 October 2021)) in part of SEC, SAXS and XRD investigation.

Data Availability Statement: The data presented in this study are available. Structural data are deposited in the RCSB data bank with accession code 8R2L.

Acknowledgments: We acknowledge the European Synchrotron Radiation Facility for the provision of beam time. We also thank G.S. Peters for the support during SAXS measurements at beamline BioMUR.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kaufmann, B.; Rossmann, M.G. Molecular mechanisms involved in the early steps of flavivirus cell entry. *Microbes Infect.* **2011**, *13*, 1–9. [[CrossRef](#)] [[PubMed](#)]
2. Stiasny, K.; Kössl, C.; Lepault, J.; Rey, F.A.; Heinz, F.X. Characterization of a Structural Intermediate of Flavivirus Membrane Fusion. *PLoS Pathog.* **2007**, *3*, e20. [[CrossRef](#)] [[PubMed](#)]
3. Yang, X.; Qi, J.; Peng, R.; Dai, L.; Gould, E.A.; Gao, G.F.; Tien, P. Molecular Basis of a Protective/Neutralizing Monoclonal Antibody Targeting Envelope Proteins of both Tick-Borne Encephalitis Virus and Louping Ill Virus. *J. Virol.* **2019**, *93*, e02132–e2218. [[CrossRef](#)] [[PubMed](#)]
4. Kanai, R.; Kar, K.; Anthony, K.; Gould, L.H.; Ledizet, M.; Fikrig, E.; Marasco, W.A.; Koski, R.A.; Modis, Y. Crystal structure of West Nile Virus envelope glycoprotein reveals viral surface epitopes. *J. Virol.* **2006**, *80*, 11000–11008. [[CrossRef](#)] [[PubMed](#)]

5. Nybakken, G.E.; Nelson, C.A.; Chen, B.R.; Diamond, M.S.; Fremont, D.H. Crystal Structure of the West Nile Virus Envelope Glycoprotein. *J. Virol.* **2006**, *80*, 11467–11474. [[CrossRef](#)] [[PubMed](#)]
6. Lu, X.; Xiao, H.; Li, S.; Pang, X.; Song, J.; Liu, S.; Cheng, H.; Li, Y.; Wang, X.; Huang, C.; et al. Double Lock of a Human Neutralizing and Protective Monoclonal Antibody Targeting the Yellow Fever Virus Envelope. *Cell Rep.* **2019**, *26*, 438–446. [[CrossRef](#)]
7. Modis, Y.; Ogata, S.; Clements, D.; Harrison, S.C. A ligand-binding pocket in the dengue virus envelope glycoprotein. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 6986–6991. [[CrossRef](#)]
8. Modis, Y.; Ogata, S.; Clements, D.; Harrison, S.C. Surface Epitopes in the Crystal Structure of Dengue Virus Type 3 Envelope Glycoprotein. *J. Virol.* **2005**, *79*, 1223–1231. [[CrossRef](#)]
9. Rouvinski, A.; Guardado-Calvo, P.; Barba-Spaeth, G.; Duquerroy, S.; Vaney, M.-C.; Kikuti, C.M.; Sanchez, M.E.N.; Dejnirattisai, W.; Wongwiwat, W.; Haouz, A.; et al. Recognition determinants of broadly neutralizing human antibodies against dengue viruses. *Nature* **2015**, *520*, 109–113. [[CrossRef](#)]
10. Dai, L.; Song, J.; Lu, X.; Deng, Y.Q.; Musyoki, A.M.; Cheng, H.; Zhang, Y.; Yuan, Y.; Song, H.; Haywood, J.; et al. Structures of the Zika Virus Envelope Protein and Its Complex with a Flavivirus Broadly Protective Antibody. *Cell Host Microbe* **2016**, *19*, 696–704. [[CrossRef](#)]
11. Chen, Z.; Ye, F.; Lin, S.; Yang, F.; Cheng, Y.; Cao, Y.; Chen, Z.; Lu, G. Crystal structure of Usutu virus envelope protein in the pre-fusion state. *Virol. J.* **2018**, *15*, 183. [[CrossRef](#)] [[PubMed](#)]
12. Liu, X.; Zhao, X.; Na, R.; Li, L.; Warkentin, E.; Witt, J.; Lu, X.; Yu, Y.; Wei, Y.; Peng, G.; et al. The structure differences of Japanese encephalitis virus SA14 and SA14-14-2 E proteins elucidate the virulence attenuation mechanism. *Protein Cell* **2019**, *10*, 149–153. [[CrossRef](#)]
13. Zhang, Y.; Zhang, W.; Ogata, S.; Clements, D.; Strauss, J.H.; Baker, T.S.; Kuhn, R.J.; Rossmann, M.G. Conformational changes of the flavivirus E glycoprotein. *Structure* **2004**, *12*, 1607–1618. [[CrossRef](#)] [[PubMed](#)]
14. Vaney, M.C.; Dellarole, M.; Duquerroy, S.; Medits, I.; Tsouchnikas, G.; Rouvinski, A.; England, P.; Stiasny, K.; Heinz, F.X.; Rey, F.A. Evolution and activation mechanism of the flavivirus class II membrane-fusion machinery. *Nat. Commun.* **2022**, *13*, 3718. [[CrossRef](#)] [[PubMed](#)]
15. Hashizume, Y.; Inaka, K.; Furubayashi, N.; Kamo, M.; Takahashi, S.; Tanaka, H. Methods for Obtaining Better Diffractive Protein Crystals: From Sample Evaluation to Space Crystallization. *Crystals* **2020**, *10*, 78. [[CrossRef](#)]
16. Kordonskaya, Y.V.; Timofeev, V.I.; Dyakova, Y.A.; Marchenkova, M.A.; Pisarevsky, Y.V.; Kovalchuk, M.V. Free Energy Change during the Formation of Crystalline Contact between Lysozyme Monomers under Different Physical and Chemical Conditions. *Crystals* **2021**, *11*, 1121. [[CrossRef](#)]
17. Kovalchuk, M.V.; Blagov, A.E.; Dyakova, Y.A.; Gruzinov, A.Y.; Marchenkova, M.A.; Peters, G.S.; Pisarevsky, Y.V.; Timofeev, V.I.; Volkov, V.V. Investigation of the Initial Crystallization Stage in Lysozyme Solutions by Small-Angle X-ray Scattering. *Cryst. Growth Des.* **2016**, *16*, 1792–1797. [[CrossRef](#)]
18. Behlke, J.; Marg, A.; Paeschke, M.J. Protein crystallization with and without precipitants. *Appl. Cryst.* **1997**, *30*, 559–560. [[CrossRef](#)]
19. Vekilov, P.G.; Vorontsova, M.A. Nucleation precursors in protein crystallization. *Acta Crystallogr. F Struct. Biol. Commun.* **2014**, *70*, 271–282. [[CrossRef](#)]
20. Dubova, K.M.; Vlaskina, A.V.; Korzhenevskiy, D.A.; Agapova, Y.K.; Rakitina, T.V.; Samygina, V.R. Preliminary X-ray Diffraction Analysis of the Envelope (E) Protein of Far-Eastern Tick-Borne Encephalitis Virus Subtype (Sofjin Strain). *Crystallogr. Rep.* **2022**, *67*, 581–585. [[CrossRef](#)]
21. Murshudov, G.N.; Vagin, A.A.; Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Cryst.* **1997**, *D53*, 240–255. [[CrossRef](#)]
22. Emsley, P.; Lohkamp, B.; Scott, W.G.; Cowtan, K. Features and development of Coot. *Acta Cryst.* **2010**, *D66*, 486–501. [[CrossRef](#)]
23. Winn, M.D.; Ballard, C.C.; Cowtan, K.D.; Dodson, E.J.; Emsley, P.; Evans, P.R.; Keegan, R.M.; Krissinel, E.B.; Leslie, A.G.W.; McCoy, A.; et al. Overview of the CCP4 suite and current developments. *Acta Cryst.* **2011**, *D67*, 235–242. [[CrossRef](#)]
24. Peters, G.S.; Zakharchenko, O.A.; Konarev, P.V.; Karmazikov, Y.V.; Smirnov, M.A.; Zabelin, A.V.; Mukhamedzhanov, E.H.; Veligzhanin, A.A.; Blagov, A.E.; Kovalchuk, M.V. The small-angle X-ray scattering beamline BioMUR at the Kurchatov synchrotron radiation source. *Nucl. Instrum. Methods Phys. Res. Sect. A Accel. Spectrometers Detect. Assoc. Equip.* **2019**, *945*, 162616. [[CrossRef](#)]
25. Hammersley, A.P. FIT2D: A multi-purpose data reduction, analysis and visualization program. *J. Appl. Crystallogr.* **2016**, *49*, 646–652. [[CrossRef](#)]
26. Manalastas-Cantos, K.; Konarev, P.V.; Hajizadeh, N.R.; Kikhney, A.G.; Petoukhov, M.V.; Molodenskiy, D.S.; Panjkovich, A.; Mertens, H.D.T.; Gruzinov, A.; Borges, C.; et al. ATSAS 3.0: Expanded functionality and new tools for small-angle scattering data analysis. *J. Appl. Crystallogr.* **2021**, *54*, 343–355. [[CrossRef](#)]
27. Dolinsky, T.J.; Nielsen, J.E.; McCammon, J.A.; Baker, N.A. PDB2PQR: An Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations. *Nucleic Acids Res.* **2004**, *32*, W665–W667. [[CrossRef](#)]
28. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J.C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718. [[CrossRef](#)]
29. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J.L.; Dror, R.O.; Shaw, D.E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins Struct. Funct. Bioinform.* **2010**, *78*, 1950–1958. [[CrossRef](#)]
30. Horn, H.W.; Swope, W.C.; Pitner, J.W.; Madura, J.D.; Dick, T.J.; Hura, G.L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665–9678. [[CrossRef](#)]

31. Berendsen, H.J.C.; Postma, J.P.M.; Van Gunsteren, W.F.; Dinola, A.; Haak, J.R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690. [[CrossRef](#)]
32. Parrinello, M.; Rahman, A. Strain fluctuations and elastic constants. *J. Chem. Phys.* **1982**, *76*, 2662–2666. [[CrossRef](#)]
33. Van Gunsteren, W.F.; Berendsen, H.J.C. A Leap-Frog Algorithm for Stochastic Dynamics. *Mol. Simul.* **1998**, *1*, 173–185. [[CrossRef](#)]
34. Chiara, C.; Rees, M.; Menon, R.P.; Pauwels, K.; Lawrence, C.; Konarev, P.V.; Svergun, D.I.; Martin, S.R.; Chen, Y.W.; Pastore, A. Self-Assembly and Conformational Heterogeneity the AXH Domain of Ataxin-1: An Unusual Example of a Chameleon Fold. *Biophys. J.* **2013**, *104*, 1304–1313. [[CrossRef](#)]
35. Zhang, N.; Li, C.; Jiang, S.; Du, L. Recent Advances in the Development of Virus-Like Particle-Based Flavivirus Vaccines. *Vaccines* **2020**, *8*, 481. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.