



UNIVERSITATEA „POLITEHNICA” din BUCUREȘTI
ȘCOALA DOCTORALĂ CHIMIE APLICATĂ ȘI ȘTIINȚA MATERIALELOR

Nr. Decizie din

DOCTORAL THESIS SUMMARY

*EXPRESIA PROTEINEI DIN CAPSIDA ssCCMV ȘI SIMULARE DE DINAMICĂ
MOLECULARĂ A CAPSIDULUI ssCCMV*

*EXPRESSION OF THE ssCCMV CAPSID COAT PROTEIN AND MOLECULAR
DYNAMICS SIMULATIONS ON THE ssCCMV CAPSID*

Author: János SZÖVERFI

Ph.D. supervisor: Prof. dr. ing. Szabolcs LÁNYI

Keywords: heterologous expression, protein purification, molecular dynamics,
CCMV capsid, docking, mutation, protein dimers

COMISIA DE DOCTORAT

Președinte	Prof. dr. ing. Ileana RĂU	de la	Universitatea „Politehnica” din București
Conducător de doctorat	Prof. dr. ing. Szabolcs LÁNYI	de la	Universitatea „Politehnica” din București
Referent	Prof. dr. ing. Cristina COSTACHE	de la	Universitatea „Politehnica” din București
Referent	Prof. Habil. dr. ing. Csaba PAIZS	de la	Universitatea „Babes Bolyai”, Cluj Napoca
Referent	Prof. univ. dr. habil László SZILÁGYI	de la	Universitatea Sapiientia din Cluj-Napoca

BUCUREȘTI

2022

Contents

1. Introduction and objectives.....	5
3. Methods.....	7
3.1. Cloning, expression and purification of the CCMV coat protein	7
3.1.1. Bacterial transformation	7
3.1.2. Plasmid cloning by restriction enzyme digest (subcloning)	7
3.1.3. Shake flask test expression.....	7
3.1.4. Protein expression optimization.....	7
3.1.6. Agarose gel electrophoresis	7
3.1.7. SDS polyacrylamide gel electrophoresis	8
3.1.8. Protein purification	8
3.1.9. Digesting the pUBK_CCMV fusion protein	8
3.2. Generating dimer structures with ZDOCK	8
3.2.1. Energy minimization with CUDAGMIN	9
3.3. Studied systems.....	9
3.4. Molecular dynamics simulations	9
3.4.1. Long timescale MD	10
3.4.2. MD at different temperatures.....	10
3.4.3. Replica-exchange molecular dynamics simulation	10
3.4.4. Accelerated molecular dynamics simulations	10
3.5. Point mutations for T1 dimers.....	11
3.6. Analysis of trajectories	11
3.6.1. MM-GBSA analysis	11
4. Results.....	12
4.1. Cloning, expression and purification of the CCMV coat protein	12
4.1.1. Plasmid construction	12
4.1.2. Transformation and cloning	12
4.1.3. Plasmid digestion	12
4.1.4. Ligation	13
4.1.5. Shake flask test expression.....	13
4.1.6. Optimization of protein expression	14
4.1.7. Purification of the pUBK_CCMV	14
4.1.8. Deubiquitinating with ubiquitin C-terminal hydrolase 1 (YUH1).....	15
4.2. Energy minimization of dimer structures generated with ZDOCK	16

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

- 4.3. Comparing short timescale MD simulations for T1, T2 and T3 dimers 19
- 4.4. Dimers selected for dynamics investigation 21
- 4.5. Long timescale MD simulation in explicit solvent for T1, T2, T3 and TX protein dimers 21
- 4.6. Molecular dynamics simulations at different temperatures..... 23
- 4.7. Mutants of the T1 dimer 25
 - 4.7.1. Deletion mutant of T1 and T2 dimers 25
 - 4.7.2. Point mutations of T1 dimer..... 27
- 4.8. Replica exchange molecular dynamics simulations 29
- 4.9. Accelerated molecular dynamics on T1, T2, T3 for 500 ns in NPT ensemble 30
- 4.10. Pentamer of dimers 31
- 5. Conclusions..... 33
 - 5.1. Results and original contributions..... 33
 - 5.2. List of original publications 34
 - 5.2.1. Publications..... 34
 - 5.2.2. Conferences..... 35
 - 5.3. Perspectives for further developments 35
- Bibliography 36

Acknowledgements

First of all, I would like to thank Prof. Dr. Lányi Szabolcs for allowing me to participate in doctoral studies under his leadership and for his support, which allowed me to complete my doctoral thesis.

I would like to express my thanks to Prof. Dr. Albert Beáta and Dr. Orbán Csongor for introducing me into the world of biotechnology research and for always helping and supporting me when it was needed.

I would like to thank my PhD colleagues, especially Molnár-Nagy Katalin, Salamon Pál and Balázs-Bécsi Ildikó for helping me with my laboratory work and making the long evenings in the lab bearable with their good mood.

I would like to thank Dr. Fejér Szilárd for taking me into his team and selflessly assisting me throughout my doctoral studies. I spent many beautiful years in the Research Team of the Provitam Foundation and learned from him what a true leader is like.

Thank you to everyone who has helped me in any way, motivated me and kept me going so that I can get to the end of my studies.

Finally, my biggest thanks go to my wife, who has stood by me throughout my training years, supported me and endured my absence with patience. I am grateful to my son that during the breaks of stressful hours he was always able to make me feel better.

1. Introduction and objectives

Nowadays computational chemistry has a constantly growing importance in life sciences. Many phenomena of biological processes are widely examined *in vivo* and *in vitro*, but during the last few decades *in silico* research had a spectacular evolution.

Viruses are the simplest organisms, subsequently they are investigated for understanding fundamental properties and interactions of proteins, nucleic acids and other components [1][2][3]. By studying the formation of viruses, we can get valuable insight into self-assembly, a phenomenon met almost everywhere on different scales. The capsid of viruses is usually composed of coat proteins, multiplied to form a symmetrical shell around the genome. Using computational chemistry tools, we seek to understand how and why these proteins are arranged to form these capsids of well-defined structure.

The cowpea chlorotic mottle virus has been investigated for decades due to its ability to form fully functional empty capsids starting from monomers under various conditions [5]. Studies for understanding the process of self-assembly in detail were carried *in silico*, *in vitro* and *in vivo* with significant results, however many questions remain open.

In the current work a short presentation of the notion of self-assembly and the examined salt stable cowpea chlorotic mottle virus (ss-CCMV) is presented.

The research in this domain is represented both by *in vitro* and *in silico* experiments.

The first goal was the expression of the CCMV capsid protein *in vitro* and study the dimerization of the protein and the self-assembly of the empty capsid.

The ss-CCMV capsid protein will be expressed in *Escherichia coli* using a ubiquitous plasmid and purified for further investigations.

The designed gene is purchased from commercial agents and cloned into a ubiquitin-pET 19b based vector. The plasmid will be transformed in *E. coli* BL21(DE3) Rosetta host cell line for heterologous expression on small scale.

The built-in His-tag makes possible the purification of the proteins with batch nickel affinity chromatography and gel filtration column chromatography to achieve pure protein.

In the last part of *in vitro* experiments, the digestion of the fusion protein with ubiquitin C-terminal hydrolase 1 enzyme that will release the ss-CCMV capsid protein.

The principal aim of our research is to understand more deeply the processes and effects that affect the dimerization of the capsid proteins and the formation of the full capsid.

Previous *in silico* studies showed results concerning the first steps of dimer formation on atomistic levels. Various molecular dynamics simulation on different timescale and temperatures had results regarding the stability of the protein dimers [7].

We start from the experimental 3D structure of the capsid protein to simulate the formation of subunits. The interactions between the proteins are examined to predict the formation of the capsid. The three chains of the protein asymmetric unit (A, B, C) are separated and combined into possible dimers. The pairs of protein dimers are relaxed with the ClassicRelax protocol of Rosetta and a docking process is performed on the ZDOCK server to find 2000 best predicted dimer structures. Three types from the resulted dimers (BB, BC, CC) are subjected to subsequent energy minimization in the AMBER software package. The aim is to find the dimer with the lowest binding energy (most stable dimer) and compare this structure to the various experimentally known dimer interfaces.

In chapter 4.1. – 4.10. we study the self-assembling properties of the CCMV virus capsid proteins using state of the art computational methods. Dimers of the coat protein are investigated to understand the nature and the driving forces of self-assembly.

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

In the first phase of our work, we found correlation between the known binding interfaces of the protein dimers and the modelled ones and structural fits for all three types of interfaces. We can predict that the T1 CC dimers are the most important in the capsid formation process, result that is in good agreement with the experimentally determined pathway.

Molecular dynamics simulations are carried out for the structures with the lowest RMSD with respect to the original (crystallographically determined) interfaces on different time scales and temperature, to study the stability of the dimers and to determine the possibility of dissociation for the various dimers under certain conditions.

Studying the stability of proteins in different conditions provides a better understanding of their function and behavior. Self-assembly of the CCMV virus capsid can be modelled on realistic computational timescales if we consider the dissociation of protein complexes as a reverse process of the aggregation.

The aim of molecular dynamics studies presented in chapter 4.2.-4.6. is to find and understand the conditions that lead to a separation of the stable protein dimers. Results presented in chapter 4.1. showed that type 1 (T1) and type 2 (T2) interfaces are important during the capsid formation, while type 3 (T3) is an unstable dimer [6]. We performed dynamical simulation on T1, T2 and T3 protein dimers to get a better view on the interactions between subunits of the capsid. 7 mutated dimers of T1 were also used in the research.

The simulations performed were: long timescale molecular dynamics, classical molecular dynamics (cMD) on different temperatures, replica exchange molecular dynamics (REMD) and accelerated molecular dynamics (aMD).

One way to evaluate the stability of the dimer interface is to introduce point mutations and evaluate how the interface responds to them. We generated a large number of point mutants, and evaluated their relative stabilities with different methods.

3. Methods

This chapter describes the methods used throughout this work, the results of which are presented in Chapter 4.

3.1. Cloning, expression and purification of the CCMV coat protein

3.1.1. Bacterial transformation

The designed plasmids are introduced in bacterial cells during the transformation. The pUC57 plasmid containing the CCMV proteins DNA was transformed into *E. coli Top10* cloning cell line.

3.1.2. Plasmid cloning by restriction enzyme digest (subcloning)

The mixture for digestion was prepared similarly for the PUC57 and pUBK: 2 µl B buffer, 1,5 µl SaCII, 1,5 µl BamHI and 5 µl UP with 5 µl PUC57, respectively 10 µl pUBK in 20 µl final volume. After digestion both of the samples were purified by agarose gel electrophoresis. The ligation process was performed simultaneously for 2 hours on 25 °C and overnight on 16 °C. The next step is the transformation of the new plasmids into a TOP10 competent cell. 50 µl competent cell with 5 µl plasmid was submitted to heat shock at 42°C for 90 sec then shaken at 37°C. The desired colony of cells were isolated with Plasmid Miniprep Kit. A diagnostic digest was performed with SaCII and BamHI restriction endonucleases as described previously.

3.1.3. Shake flask test expression

The selected colonies were transformed into an *E. coli BL21(DE3) Rosetta* host cell line. Cells were grown on LB agar plates, containing kanamycin and a well isolated colony was inoculated in 10 ml of LB medium with 30 µg/ml kanamycin and incubated for 4 hours at 37 °C in a shaking incubator. The cell culture was transferred to 200 ml of LB medium and grown to OD₆₀₀=0.7. Expression was induced with 0.8 mM IPTG and incubated for 3 hours on 37 °C. Culture broth was centrifuged for 10 minutes on 20000 RPM and pellets were stored at -80 °C until further processing.

3.1.4. Protein expression optimization

To maximize the protein expression of the bacterial cells we planned an optimization process with changing the IPTG concentration and the temperature as follows:

- 0.1 mM IPTG at 18°C
- 1 mM IPTG at 18°C
- 0.5 mM IPTG at 37 °C
- 0.5 mM IPTG at 18°C
- 0.1 mM IPTG at 37 °C
- 1 mM IPTG at 37 °C

The experiment was conducted parallel for the samples selected from digestion results.

3.1.6. Agarose gel electrophoresis

Separation of genomic DNA, result of the restriction enzyme digest, was performed by agarose gel electrophoresis. The gel was prepared for 1% w/v of agarose and dissolved in TAE buffer. Redsafe was added to the boiled solution and poured in a cast.

5 µl of DNA samples were mixed with 1 µl of 6X loading dye and introduced into the wells of the gel. The electrophoresis was performed at constant 90 V and the gels were visualized under UV light.

3.1.7. SDS polyacrylamide gel electrophoresis

Usually, 25 µl of protein sample was mixed with 25 µl of 4x SDS sample buffer, heated at 98°C for 5 min and centrifuged for 10 min on 14000 rpm. 10 µl of sample was introduced in the pocket of the stacking gel. The analysis was performed with a 5% stacking gel and a 12% resolving gel at a constant voltage of 120 V on ice. After the electrophoresis the gel was treated with Coomassie Brilliant Blue dye for 40 min with constant shaking. In the end the gel was shaken with a destaining solution for 2-3 hours to visualize the bands on the gel.

3.1.8. Protein purification

4 ml of protein sample was loaded onto a HisTrap HP 5 mL column and an Akta Purifier system was used. As our protein contained a 10X His tag nickel affinity chromatography can be used. New or regenerated NTA-agarose beads were used on a column in a 1:10 NTA beads: protein ratio. The beads were introduced into the column and washed with 10 column volume (CV) washing buffer. Sample was added to the beads and rolled for 40 minutes on 4 °C. The flow through was saved and 1 ml washing buffer was added 8 times, followed by elution for 8 times. Fractions of 1 ml were saved and analyzed with SDS-PAGE.

3.1.9. Digesting the pUBK_CCMV fusion protein

YUH1 plasmid was transformed into *E. coli BL21(DE3) Rosetta* host cell line and inoculated in LB medium. The mixture was incubated on 37 °C, shaking on 250 RPM for 60 minutes, and grown on agar plates containing kanamycin on 37 °C overnight. Next day one isolated colony was inoculated in 7,5 ml LB medium, with 2,3 µl kanamycin stock (80 µg/ml) and incubated on 37 °C until OD₆₀₀=0.67. The mixture was induced with 200 µl of IPTG (1 mol/l). Culture broth was harvested by centrifugation at 4500 RPM 4 °C for 10 minutes. For purification pellet were suspended in 4X2 ml lysis buffer and sonicated for 3X20 sec to disrupt the cells. After 60 min of centrifugation on 20000 RPM and 4 °C the supernatant was collected and purified with batch nickel affinity chromatography. Fractions of 1 ml were collected and analyzed with SDS-PAGE.

YUH1 enzyme [8] cleave the ubiquitin from the purified protein, resulting the CCMV capsid protein monomer. Two fractions (CCMV1 and CCMV2) saved from the FPLC separation were concentrated with Amicon Ultra Centrifugal filter. 3,5 ml of both samples were centrifuged for 40 minutes on 4000 RPM and the concentrate was collected from the filter device sample reservoir. The concentration of the solutions was measured with Nanodrop Spectrophotometer at 260 nm. Solutions were mixed to an enzyme: protein ratio of 1:5 and digested for 60 minutes on room temperature. The mixtures were separated with batch nickel affinity chromatography as described previously.

3.2. Generating dimer structures with ZDOCK

The studied protein is a coat protein from the salt stable cowpea chlorotic mottle virus [9], resolution 2,7 Å. The protein contains 3 chains: A, B, C. We generated a protein-protein docking simulation on the ZDOCK server [10] for the possible permutations of chain B and chain C: chain B-chain B, chain B- chain C, chain C- chain C. Chain The input monomer structures were relaxed with the ClassicRelax protocol of Rosetta [11] to remove any clashes. The best 2000 of the predicted dimers from these input structures, based on the ZDOCK scoring function, were saved and processed for each pair.

3.2.1. Energy minimization with CUDAGMIN

The pdb files resulted from the docking process were prepared for energy minimization in the tleap software, from the AmberMD software package[12]. The force field used was ff03 [13], a modified version of the ff99 [14] .For solvation, we used the generalized Born solvation model (GB).The igb=2 option used by us is a modified version of the GB model[16]. The recommended radii set for the igb=2 model is mbondi2.

The salt concentration was set to 0,1 M. For initial runs, a cutoff of 12 Å was used. Subsequently, for CUDA runs (force field calculation optimized for graphical processing units - GPUs), no cutoffs were used for evaluating the non-bonded terms of the potential. The rGBmax value was set to 8.23 Å.The energy minimization of the models was performed using the GMIN software [17].

We used the L-BFGS minimizer part of GMIN, no basin-hopping steps were taken for the structures.

The energy evaluation for the system using an nVidia Tesla K40 GPU is about 100-200 times faster than the same evaluation using only a CPU core.The set of the dimers (2000 structures for BB, BC and CC) was minimized both in associated and dissociated form.The convergence criteria (SLOPPYCONV and TIGHTCONV) were set to 10^{-2} .The MAXERISE parameter is set to 10^{-4} .

The binding energy for each dimer is the difference between the potential energies for the associated and dissociated form of the dimers, resulted from the minimization run. The best alignment was searched based on three types of interfaces found in the icosahedral viral shell:

- type 1 (T1): CC, BA
- type 2 (T2): BC1, BC2, AA1, AA2
- type 3 (T3): AC, AB

We used the alpha carbons of residues 40 to 190 in each structure for RMSD calculation. We used the PERMOPT routine in GMIN for evaluating the backbone RMSDs.

3.3.Studied systems

Four structures were used to study the behavior of the CCMV dimers:

- The best structures from the global optimization were used for type 1 (T1), type 2 (T2) and type 3 (T3)
- The structure with the highest RMSD to interface T3, named TX
- Deletion mutants, where 10 residues from the N-terminals of both chains in T1 (T1DM) and T2 (T2DM) dimers were deleted
- five point-mutated mutants of T1 and two mutants based on the mutation sensitivity profile generated, with the highest $\Delta\Delta G$ and one with the lowest $\Delta\Delta G$
- a pentamer of dimers from the capsid of ss-CCMV

3.4.Molecular dynamics simulations

Molecular dynamics simulations were performed with the pmemd.cuda module of the Amber14 software package. The force field used was ff03 [13]. Simulations were carried out with implicit and explicit solvent as well. Explicit solvation was made with an octahedral box of TIP3P [18], where water molecules were added up to a distance of 8 Å from the protein.

The MD protocol was adapted from Mafucci *et al.* [19] and contains the following steps:

1. minimization of hydrogen atoms (1000 cycles of steepest descent and 5000 cycles of conjugated gradient)

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

2. minimization of water molecules (2000 cycles of steepest descent and 5000 cycles of conjugated gradient)
3. equilibration of the solvent box at 300 K by 100 ps of NVT and 100 ps of NPT simulation using a Langevin thermostat
4. minimization of the side chains and waters with backbone restraints of 25 kcal/mol
5. total minimization with backbone restraints of 10 kcal/mol (2500 cycles of steepest descent and 5000 cycles of conjugated gradient)
6. heating up the system to 300 K in 6 steps of 5 ps each (LT = 50 K), where backbone restraints were reduced from 10.0 kcal/mol to 5 kcal/mol
7. full equilibration in the NVT ensemble (100 ps, backbone restraints = 5.0 kcal/mol) and in the NPT ensemble (1 step of 200 ps, backbone restraints = 5 kcal/mol; 3 steps of 100 ps each, reducing the backbone restraints from 5.0 kcal/mol to 1.0 kcal/mol, and 1 step of 1 ns with 1.0 kcal/mol of backbone restraints)
8. production runs were conducted at 300K

An electrostatic cutoff of 8.0 Å, a Berendsen barostat, Particle Mesh Ewald summation (PME) for long-range electrostatic interactions, and the SHAKE algorithm were applied to all the calculations with explicit water.

3.4.1. Long timescale MD

T1, T2 and T3 dimers were simulated in explicit water for 2 μ s using the protocol from Chapter 3.4.

Simulation was performed on 300 K, temperature regulation was provided with the Langevin thermostat (NTT=3) with a collision frequency of GAMMA_LN=2. Periodic boundary conditions were applied using a cutoff with 8 Å to handle the long range nonbonded interactions. The timestep used was 2 fs (DT=0.002). The Particle Mesh Ewald method [20] was used and the SHAKE algorithm [21] was applied to constrain all bond lengths for hydrogen atoms.

3.4.2. MD at different temperatures

Four parallel runs in implicit solvent were performed for T1, T2, T3 and TX for 200 ns at 350 K. No cutoff was used (CUT=999), as there were no periodic boundary conditions. 600 ns of explicit solvent simulation were performed for T1, T2, T3 and TX at 350 K with the same parameters as listed in Chapter 3.4.

3.4.3. Replica-exchange molecular dynamics simulation

Temperature-based REMD simulations were used for the dimers mentioned in Section 3.1. at 8 different temperatures: 300.00 K, 306.65 K, 313.42 K, 320.28 K, 327.25 K, 334.34 K, 341.53 K, 348.81 K, 356.24 K. Temperatures were generated on the webservers <http://folding.bmc.uu.se/remd/> [22]

8 parallel runs were performed for 25 ns (200 ns in total).

3.4.4. Accelerated molecular dynamics simulations

aMD was performed on T1, T2, T3 dimers for 500 ns in NPT ensemble with the parameters described previously, with addition of aMD specific parameters listed below. Potential energy and dihedral energy values were extracted from the 2 μ s conventional molecular dynamics run.

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

Table 3-1 aMD specific simulation parameters for the tree type of dimers

Dimer	EthreshD (kcal/mol)	alphaD	EthreshP (kcal/mol)	alphaP
T1	3527	231	-145491	8186
T2	4457	184	-152923	799
T3	4438	184	-333461	799

3.5. Point mutations for T1 dimers

In silico mutations were performed on the interface residues of the T1 dimer with the MUTATE_MODEL routine of MODELLER [23]. The 178 mutants resulted were minimized in associated and dissociated forms to calculate the binding energies. Molecular dynamics simulations were performed for 100 ns on 300 K on 5 selected mutants with the protocol described at section 3.1.

3.6. Analysis of trajectories

Trajectory analysis and visualization were made with *cpptraj* software from the Amber14 package [24], VMD [25] and UCSF Chimera [26].

Cpptraj can read topology and coordinate files from different MD simulations and the following calculations were used: *rmsd*, *atomicfluct*, *molsurf* [27], *secstruct* [28]. Principal Component Analysis (PCA) was calculated for alpha-

3.6.1. MM-GBSA analysis

MM-GBSA is part of the Amber software package. Calculation was performed with the MMPBSA.py script [29]. Analysis was carried out for 10 ns MD production run. The protein dimers for the MM-GBSA run were selected from the 2000 structures minimized with GMIN, with an iRMSD (root-mean-square deviation of the interface residues motion) lower than 3 Å with respect to the original interface. Calculation was performed for 7 dimers of T1 and 5 dimers of T2. Since we did not find T3 interface dimer for the iRMSD criteria, we used the original T3 A-B dimer from the full capsid and the T3 dimer with the lowest iRMSD value.

4. Results

4.1. Cloning, expression and purification of the CCMV coat protein

4.1.1. Plasmid construction

The gene of the ss-CCMV protein, purchased from commercial agents, was cloned in pUC57 cloning vector with XbaI and BamHI restriction sites.

The gene was transformed in a pUBK vector (pUBK_CCMV).

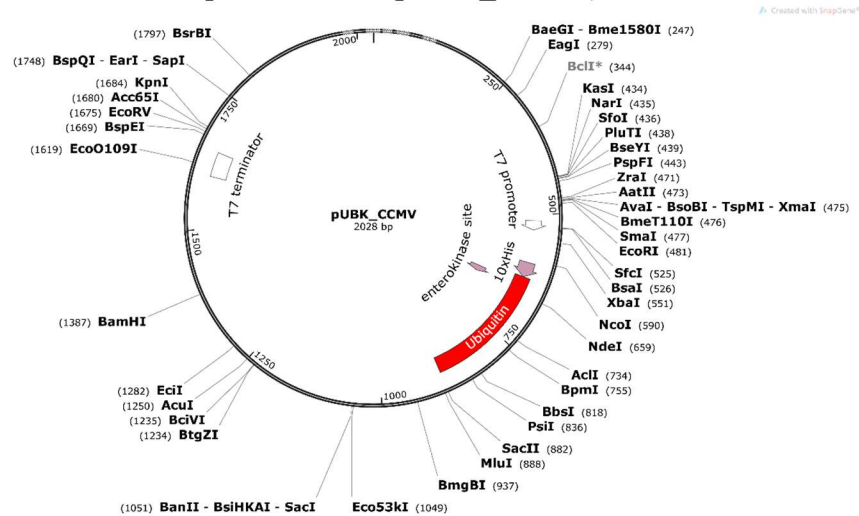


Fig. 4-1 Map of pUBK_CCMV plasmid

The plasmid (Fig. 4-1) contains an ubiquitin that assures the detection and increase the biological activity by affecting the solubility of the protein [30].

4.1.2. Transformation and cloning

The gene of ss-CCMV capsid protein was cloned in pUC57 by EcoRV cloning strategy (GenScript). The plasmid containing the CCMV proteins DNA was transformed into *E. coli Top10* cloning cell line.

4.1.3. Plasmid digestion

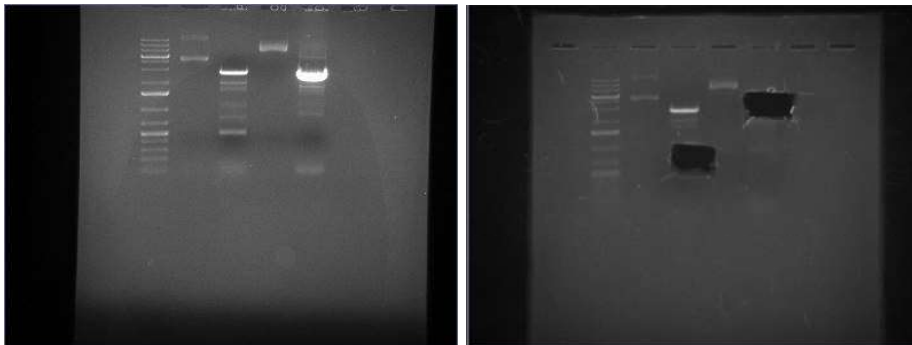


Fig. 4-2 Digestion of pUBK_CCMV plasmid

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

The DNA templates were digested with a double digestion reaction using two restriction endonucleases: BamHI and SacII at 37 °C for 60 minutes. After the reaction the enzymes were inactivated on 80 °C for 20 minutes and examined with 1,2% agarose gel electrophoresis (Fig, 4-2). The corresponding fragments were cut off and purified with *Gene Jet Gel Extraction Kit*. The concentration of the vector (pUBK) was 23,869 µg/ml and for the insert (ss-CCMV) 1.56 µg/ml.

4.1.4. Ligation

During the ligation reaction the vector and the insert prepared with digestion were combined. The reaction was promoted by T4 DNA ligase enzyme and performed for 2 hours on 25 °C and overnight on 16 °C. The product was transformed in *E. coli Top10* cell line and grown on agar plates containing ampicillin on 37 °C overnight. 4 well isolated colonies were inoculated in LB medium and incubated overnight on 37 °C in a shaking incubator. Recombinant plasmids were identified with double digestion reactions as described previously in 13 identical mixtures. The results of the digestion are shown on Fig. 4-3

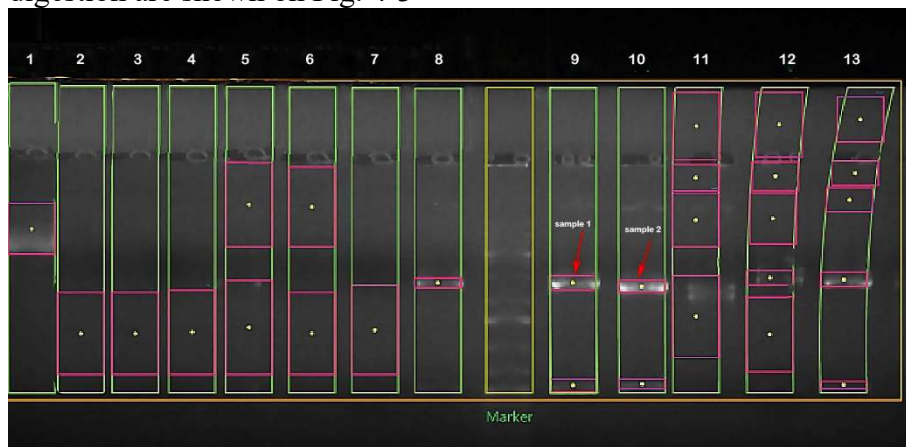


Fig. 4-3 Double digestion reactions for 13 samples of plasmid

Two samples (lane 9 and 10) were selected for further processing.

To confirm the selection of the samples gene sequencing was performed. Sample 1 showed an increased similarity to the theoretical one, thus this stock was used for expression.

4.1.5. Shake flask test expression

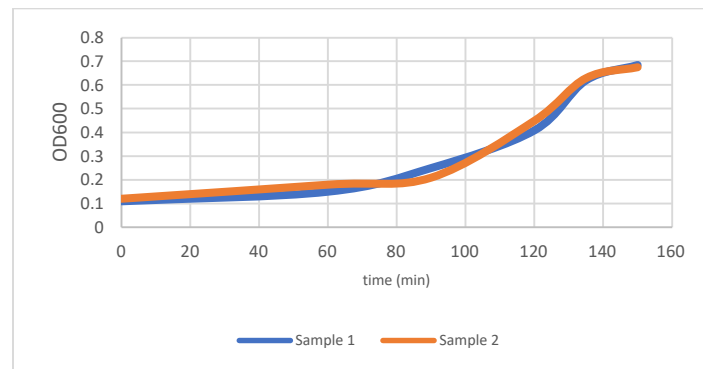


Fig. 4-4 Growth of *E. coli* during the test expression

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

The selected cell stocks were used for protein expression. First, the plasmids were transformed in *Escherichia coli* BL21(DE3) host cell line. After the OD600 reached 0.7 (Fig. 4-4) the cell culture was induced with 0.5 mM IPTG and incubated for 3 hours at 37 °C. The cell suspension was centrifuged for 10 minutes on 20000 rpm and 4 °C. The pellet was stored on -80 °C.

4.1.6. Optimization of protein expression

After shake flask test expression, we planned an optimization to maximize the expression yield. After test expression the SDS-PAGE results showed that the optimal conditions are 0.1 nM IPTG concentration and 37 °C incubation temperature (Fig. 4-5 lane 9)

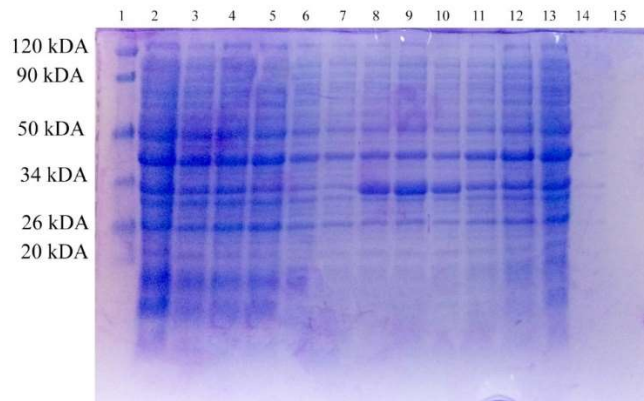


Fig. 4-5 SDS-PAGE for optimization of pUBK_CCMV expression

4.1.7. Purification of the pUBK_CCMV

Purification of pUBK_CCMV was carried out on an AKTA FPLC SYSTEM (Amersham Biosciences, Uppsala, Sweden). Fractions of 1,5 ml were saved and fractions belonging to the major peaks showed on the chromatogram (Fig. 4-6) were collected separately: green (CCMV I) and red (CCMV II). The fractions containing the proteins were dialyzed and stored for further analysis.

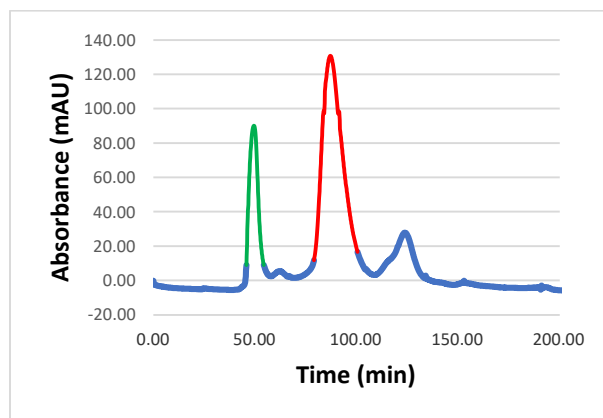


Fig. 4-6 Chromatogram for the purification of pUBK_CCMV protein expression

Results of the FPLC purification were verified by SDS-PAGE (Fig. 4-7). The proteins have been separated properly; however other proteins were bound beside the His-tagged CCMV protein.

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

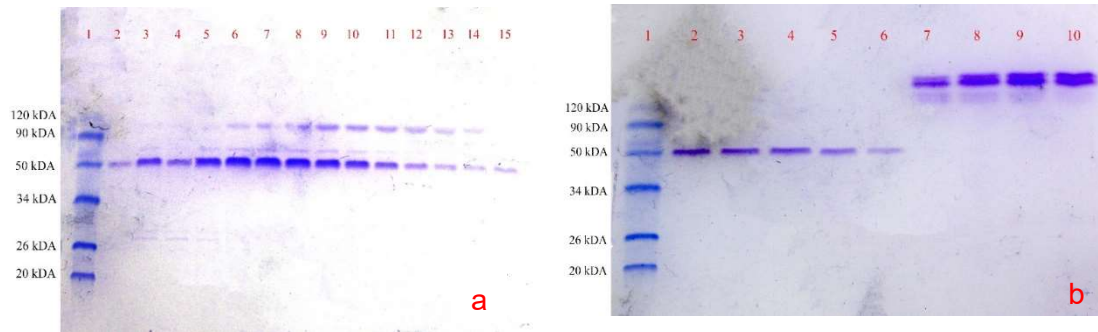


Fig. 4-7 SDS-PAGE verification after the purification of pUBK_CCMV on FPLC

4.1.8. Deubiquitinating with ubiquitin C-terminal hydrolase 1 (YUH1)

YUH1 protein was used to cleave the ubiquitin from the CCMV capsid protein.

Both of the collected fractions from FPLC purification were digested with YUH1. 8 ml of the CCMV1 solution was concentrated and resulted 2,2 ml with 550 $\mu\text{g/ml}$ concentration. Solutions were mixed to a final protein: enzyme ratio of 1:5 and after 1 hour of incubation was purified with Ni-NTA affinity chromatography.

After running an SDS_PAGE analysis the CCMV coat protein was clearly detected in the overflow fraction (Fig. 4-8, lane 2).

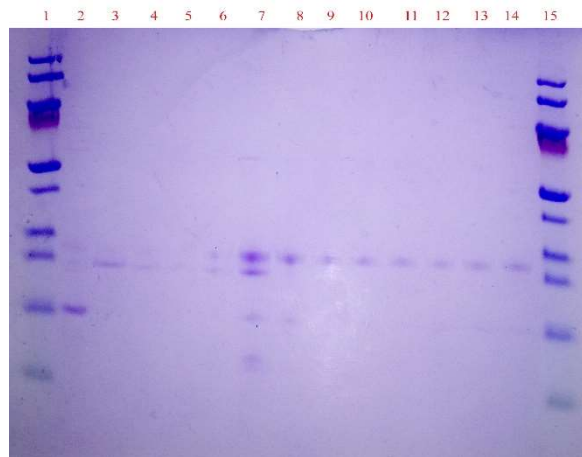


Fig. 4-8 Digestion of the pUBK_CCMV with YUH1

CCMV capsid protein can be produced with the plasmid construction presented, by heterologous expression. As this method is common and efficient, high yield of protein can be achieved.

4.2. Energy minimization of dimer structures generated with ZDOCK

The asymmetric unit in our reference crystal structure (PDB ID 1za7) contains three chains (Fig. 4-9), with almost identical heavy atom conformation. In the complete virus capsid with icosahedral symmetry, one can observe 8 possible interfaces between these chains (AA1, AA2, AB, AC, BA, BC1, BC2, CC). However, these can be classified into three types of interfaces, which we call T1, T2 and T3, respectively.

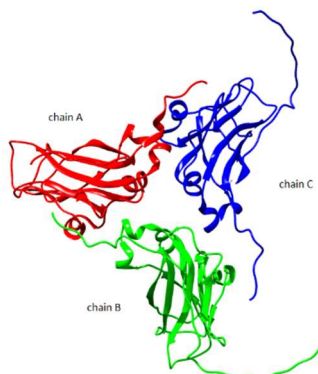


Fig. 4-9 The crystal structure of the asymmetric unit of ss CCMV(PDB ID 1za7)

For generating the dimer structures with the ZDOCK server, we input the coordinates of the B or C chain as the ‘substrate’, and the same chains for the ‘ligand’. ZDOCK runs were hence carried out on the following pairs: BB, BC, and CC. We left out the coordinates of chain A, since the crystal structure does not contain residues 26-40 from this structure (this belonging to the flexible N-terminal tail of the protein).

The ZDOCK run saved the best 2000 structures based on the docking score, and these structures were processed further for energy minimization with AMBER.

The energy minimization for the three sets of 2000 dimer structures was performed with the normal and CUDA accelerated versions of the GMIN software package, interfaced to the AMBER software suite [17].

We compare below the results of the four runs of minimization for each of the dimers (each pair in an associated and dissociated form of the ZDOCK structures).

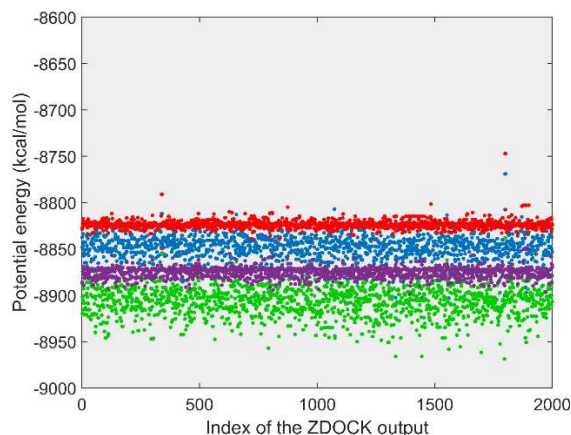


Fig. 4-10 Results of the energy minimization for BB dimer in associated form with GMIN (blue), CUDAGMIN (green) and in dissociated form with GMIN (red), CUDAGMIN (purple)

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

The third and fourth run was performed on graphics cards both in associated and dissociated forms. We can observe that the dissociated dimers have the highest energy values due to the lack of interactions between the chains, while in the associated runs clashes between the chains of the dimer can occur (Fig. 4-11). The binding energy was calculated from the difference of the energy values from associated and dissociated forms of the dimers.

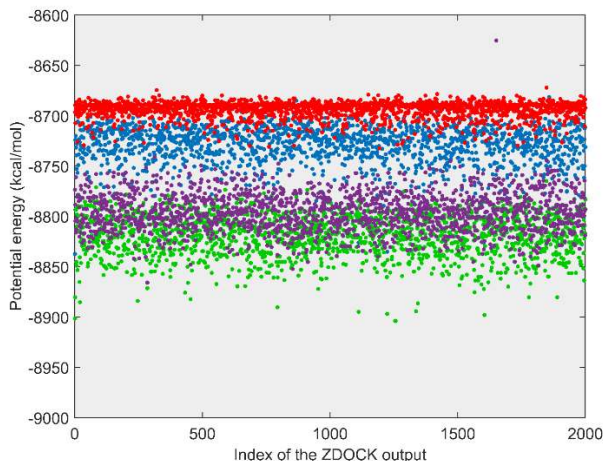


Fig. 4-11 Results of the energy minimization for BC dimer in associated form with GMIN (blue), CUDAGMIN (green) and in dissociated form with GMIN (red), CUDAGMIN (purple)

The energy minima of the BC dimers (Fig. 4-11) are generally higher than the those of BB and CC dimers. The energy values in the case of the BC dimer also present a larger scattering, due to the higher flexibility of the N-terminal tail of the protein. The conformational changes subsequently present a higher variety on a larger scale.

The reason of a larger variation in energies in case of the associated configurations is that in the associated form a large number of extra interactions are possible between the two chains, which are of course inexistent in the dissociated form.

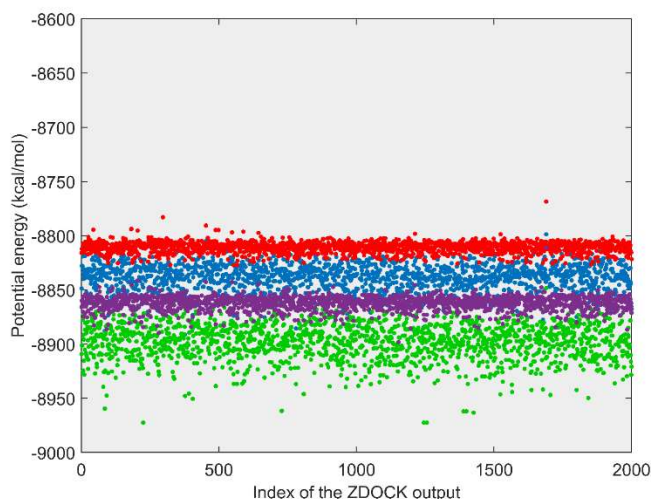


Fig. 4-12 Results of the energy minimization for CC dimer in associated form with GMIN (blue), CUDAGMIN (green) and in dissociated for with GMIN (red), CUDAGMIN (purple)

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

The potential energies calculated with AMBER, using the implicit solvent model, are not ‘pure’ potential energies, since the implicit solvent model contains free energy changes due to solvation/desolvation of surface atoms. Therefore, entropy changes associated with surface binding/unbinding of water are implicitly taken into account during evaluation of the Generalized Born energy term.

The calculated average energies for the associated and dissociated dimers are summarized in Table 4-1.

Table 4-1 Average energies of the BB, BC and CC dimers

	Average energy (kcal/mol)		Standard deviation	
	Associated	Dissociated	Associated	Dissociated
BB GMIN	-8847.675	-8823.670	44.803	4.593
BB CUDA	-8906.350	-8875.450	45.729	6.021
BC GMIN	-8710.890	-8693.810	186.979	8.566
BC CUDA	-8798.770	-8795.730	200.001	19.208
CC GMIN	-8833.065	-8810.840	102.098	4.845
CC CUDA	-8893.080	-8862.160	102.941	6.240

The calculated binding energy (Fig. 4-13) is represented as a function of the best alignment of the 6000 structures with the CC dimer (T1 interface). Here we can observe a funnel-like topology. We can observe that for several dimers a low RMSD value corresponds to low binding energy. The dimer with the lowest energy is similar with the CC type interface from the ss-CCMV coat protein, which probably has a major role in the first stages of capsid assembly.

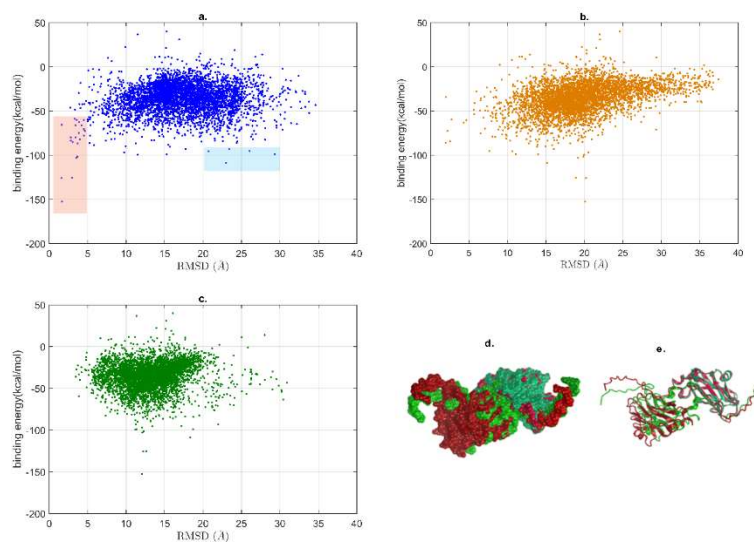


Fig. 4-13 Binding energies with respect to RMSD of the structures from the CC-T1 (a.), BC1-T2 (b.), AC-T3 (c.). Structure with the lowest binding energy superposed with the CC dimer of the virus capsid with surface (d.) and cartoon (e.) representations

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

We can conclude from the above three graphs that the T1 interface (CC and BA) is the most stable among all possible interfaces in the icosahedral capsid. Moreover, the structure with the smallest AMBER binding energy is one of T1 interface. This result suggests that global optimization with the AMBER potential is useful for blind prediction of the most stable interface between virus capsid proteins.

In Fig. 4-13 d. we represented the structure with the largest binding energy (red) aligned to the CC dimer of the virus capsid (green). The alignment was performed on residues 40 to 190 of one chain.

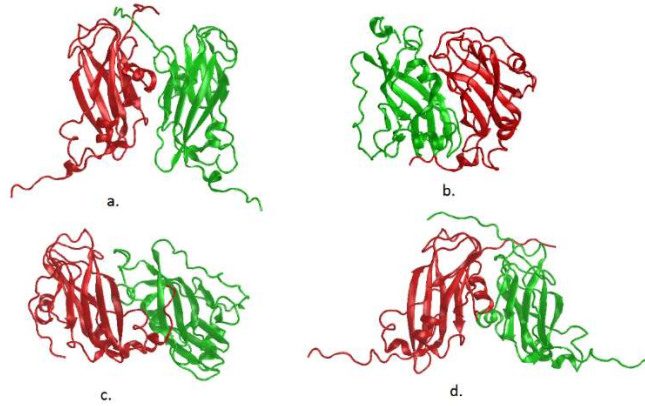


Fig. 4-14 Symmetrical structures: a. BB dimer with close to C2 symmetry; b. CC dimer with C1 symmetry c. CC dimer with C2 symmetry d. BB dimer with C2 symmetry

We find 4 other structures low in binding energy, which do not align well with the three types of interfaces (Fig. 4-14)

4.3. Comparing short timescale MD simulations for T1, T2 and T3 dimers

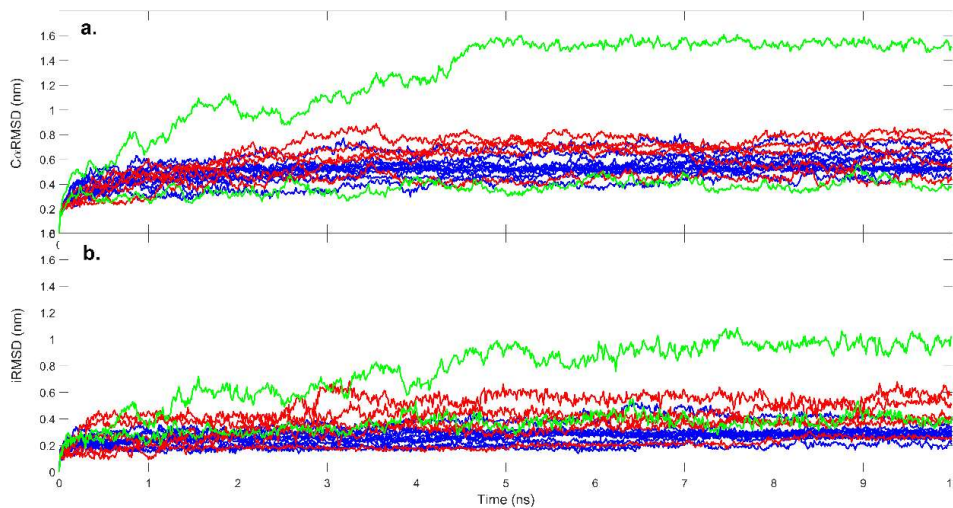


Fig. 4-15 Ca RMSD (a) and iRMSD (b) for T1 dimers (blue), T2 dimers (red) and T3 dimers (green)

We selected dimers from the 6000 ZDock structures with an iRMSD lower than 0.3 nm to the respective interface from the ss-CCMV capsid as starting structures for our MD simulations. Since we did not find T3 dimers corresponding to the criteria we selected two structures: the original AB dimer, taken from the capsid, and the structure with the lowest iRMSD from the ZDOCK results.

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

10 ns MD simulation in implicit solvent for the different interface type dimers were started. The change of iRMSD over the simulation is presented on Fig. 4-15. We can see a clear separation for the three types of dimer interfaces (Fig. 4-15).

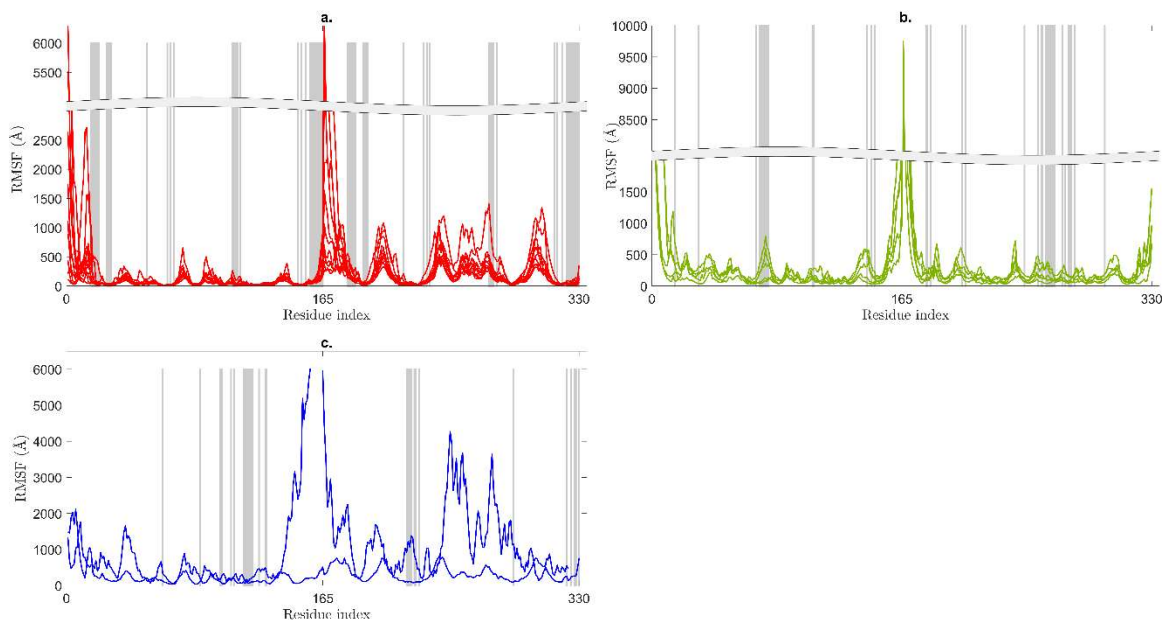


Fig. 4-16 RMSF for T1 (a.), T2 (b.) and T3 (c.) dimers with interface residues highlighted in gray.

Residue fluctuation results (Fig. 4-17) are consistent with the iRMSD calculations. Here it is interesting to look at the change in the residue interfaces (gray regions). For the T1 and T2 we cannot observe a significant fluctuation of the interface residues, while for T3 we can note that the highest motion is exactly for the interface residues.

MM-GBSA calculations were also performed on the 14 structures. The Results of the calculations are presented in Table 4-2, together with the binding energies calculated from the Amber minimization runs and the iRMSD. The average association energy is -185.7 ± 10.4 kcal/mol for T1, -108.2 ± 8.9 kcal/mol for T2 and -76.5 ± 6.7 kcal/mol for T3.

Results show that the T2 interface is about 40% less stable than the T1 interface. The iRMSD of the T3 interface increased higher than 1 nm during the 10 ns MD run. This suggests that the T3 interface does not play a role in the protein association.

Table 4-2 MM-GBSA binding energies

Minimum label	$-\Delta G_{\text{total}}$ (kcal/mol)	stdev	binding energy	Interface type	iRMS (nm)
BC1	-206.18	10.85	-152.52	1	0.167
BC2	-194.65	10.19	-125.72	1	0.156
BC410	-211.71	10.66	-125.44	1	0.299
BC407	-181.08	9.77	-83.42	1	0.279
BC10	-139.44	11.91	-79.93	1	0.28
BC3	-186.38	10.22	-65.47	1	0.161
BC73	-180.24	11.83	-46.16	1	0.275
BC11	-118.96	8.39	-85.90	2	0.195

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

BC63	-101.88	8.95	-84.33	2	0.252
BC61	-96.53	8.08	-72.41	2	0.220
BC1273	-111.58	8.59	-66.71	2	0.199
BC187	-112.16	10.15	-59.62	2	0.273
BC1000	-79.79	7.09	-45.40	3	0.716
BB321	-73.18	6.24	-36.13	3	0.341

Low binding energy values with low iRMSD are considered stable protein dimers. As we can see, among the T1 dimers three structures (BC1, BC2, BC410 – numbering according to the ZDOCK structure index) are below -100 kcal/mol and iRMSD values below 0.4 nm, with respect to the original protein dimer. T2 is represented with two good structures (BC11, BC63) with low binding energy values and iRMSD under 0.3 nm.

Based on these results three structures were selected to perform molecular dynamics simulations and investigate in detail the stability of them (T1 – BC1, T2 – BC11, T3 – BB321)

4.4. Dimers selected for dynamics investigation

Based on the energy minimization we selected three representative dimer structures with the lowest RMSD to the tree type of interface (T1, T2, T3). We also chose a structure with the highest RMSD to interface T3, named TX, to observe the dynamical behavior of a dimer with low similarity to all interfaces

4.5. Long timescale MD simulation in explicit solvent for T1, T2, T3 and TX protein dimers

Long timescale simulations make possible a deeper insight regarding the behavior of the dimers. Here we present the results for 2 μ s long simulations started from the various optimized configurations (T1, T2, T3 and TX) in explicit water. The interface C_{α} RMSD was calculated for all trajectories with reference to the first frame. Calculations were made for the backbone of the proteins and for the binding interface residues (Fig. 4-17).

The average C_{α} RMSD for the T1 dimer is 0.94 nm, higher than that for T2 (0.83 nm). The structure of T1 is more flexible, allowing an accentuated motion of the residues. The RMSD values for T3 are the largest with a higher fluctuation (1.08 nm). If we take a look at the interface RMSD values (Fig. 4-17, b.) we can see a clear separation of the three types of CCMV dimers.

The iRMSD value increases slower in case of the T2 dimer at the beginning of the simulation, but has an overall average value higher by 0.3 ns for the interface residues (1.14 nm) than in case of T1. Also, the value is not converged, showing an overall increasing trend. This result indicates that the T2 interface is less stable than T1, this being in good agreement with our previous study [6].

The interface of the T3 dimer is smaller than that of T1 and T2 and makes possible bigger motions of the chains. Therefore, the iRMSD values are between T1 and T2 with an average of 0.87 nm with a bigger fluctuation of the values.

Per residue fluctuations were calculated to monitor the motion of the residues (Fig. 4-17 c). There are some common motions of the protein parts in all three proteins. The 7-strand β -barrels are stable subunits of the protein monomer, as we can observe in the RMSF plot. The terminal tails of the monomers have large motions in every case, especially the N-terminal tail of T2 dimer. In Fig. 4-17 c. the interface residues of the dimers are represented with bars with the same color than for the respective protein.

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

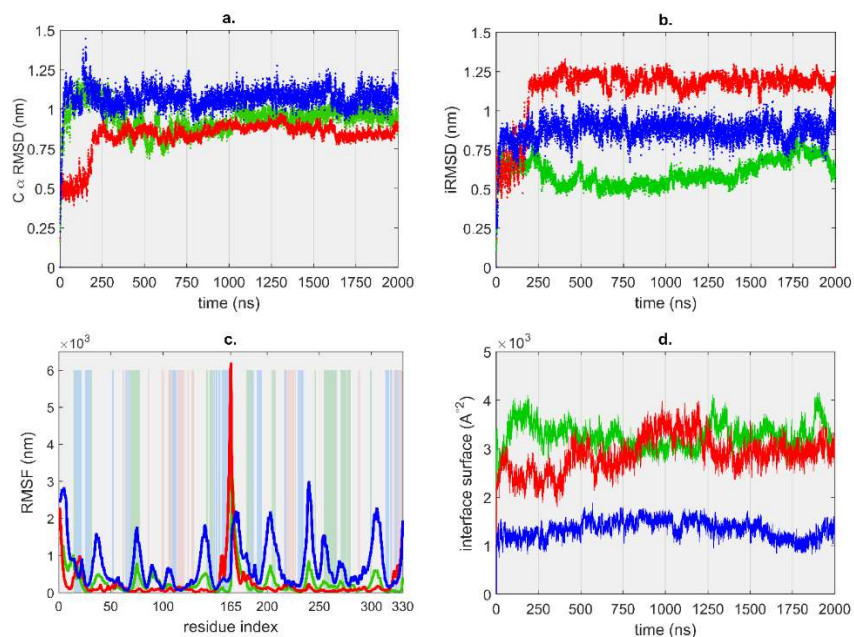


Fig. 4-17 C α RMSD (a.), iRMSD (b.), residue fluctuation (c.) and interface surface (d.) for T1 (green), T2 (red) and T3 (blue)

We calculated the interface surface area as the difference between the SASA and the sum of the standalone dimer's surfaces. The interface surface of T3 is the smallest and remains relatively constant during the simulation with minor temporary shifts ($1339.2 \pm 182.2 \text{ \AA}^2$). The T1 interface surface grows from the initial 14% of the whole surface to 19%. T2 presents a major growth of interface surface after 1000 ns of simulations, while the SASA is decreasing. The large fluctuations of values shows that the PPI of each dimer is changing permanently.

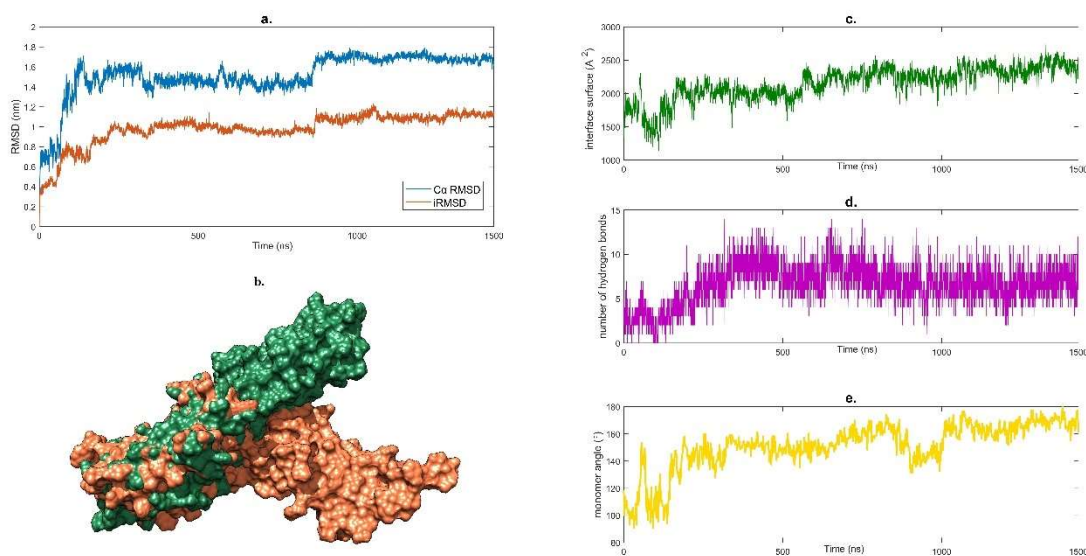


Fig. 4-18 Analysis of the 1.5 μ s NPT, explicit solvent simulation for the TX interface: a. C α RMSD and iRMSD; b. initial (red) and final (green) structures of TX superimposed with surface representation; c. change in the interface surface; d. number of hydrogen bonds; e. variation of the angle between the monomers of the protein.

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

1.5 μ s of constant pressure, explicit solvent simulation was performed on TX. Exploring the dynamics of a dimer with different structure than for the three types of dimers can lead to additional information about the behavior of the proteins under non-favorable contact conditions.

During the simulation sweeping changes occurred in the structure of TX dimer. Initially the chains of the protein were attached to each other through the N-terminal tail in a V-like shape (Fig. 4-18 b). In the early phase of the simulation the dimers are forming a rod-like structure. Due to this motion of the second chain the RMSD values are high, with an average of 1.51 ± 0.2 nm for the C_{α} RMSD and 0.98 ± 0.15 nm for the iRMSD (Fig. 4-18 a). The surface of the PPI is increasing throughout the simulation, suggesting an increasing stability of the dimer. During the simulation the number of hydrogen bonds is also growing with large fluctuations.

4.6. Molecular dynamics simulations at different temperatures

MD simulations were carried out on higher temperatures for T1, T2, T3 and TX dimers to study the conditions for the dissociation of dimers. The principal aim was to achieve a dissociation of the dimers for a better mechanistic understanding of how a capsid gets destabilized.

Four parallel simulations with different initial velocities were carried out in implicit solvent for T1, T2, T3 and TX on 350 K temperature for 200 ns. The four runs of the T1 dimer show high similarity to each other, with an average of 1.1 nm C_{α} RMSD and 0.67 nm iRMSD values (Fig. 4-19 a, b). The dimer is stable at 350 K, as shown also by the interface surface data (Fig. 4-19 c.). Changes in the interface are negligible for T1 runs on higher temperatures. Fluctuations in the values are due to the relative motions of the chains, but higher conformational variation cannot be observed.

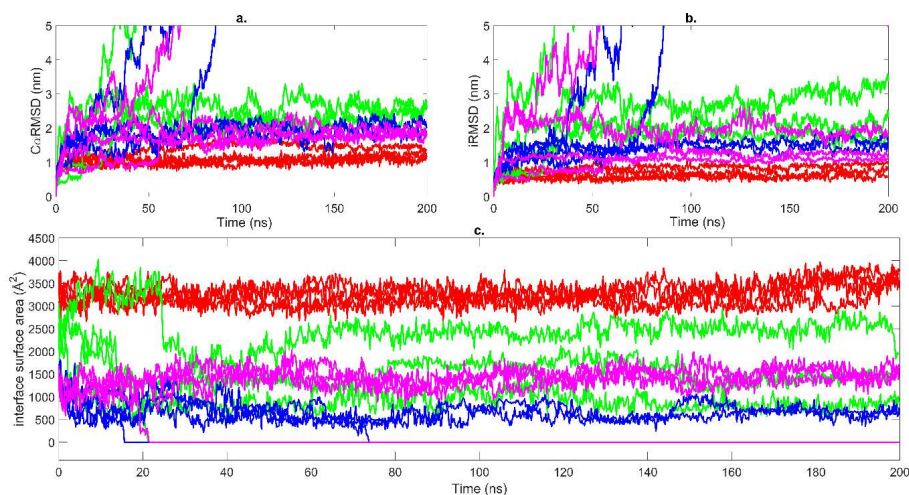


Fig. 4-19 C_{α} RMSD (a), iRMSD (b) and interface surface (c) for T1 (red), T2 (green), T3 (blue) and TX (purple) on 350 K, implicit solvent

The T2 dimer behaves differently to the T1. In three runs out of four the dimer approaches a state near dissociation; however, the N-terminal tails of the dimers remain attached and prevent the complete separation of the chains. An accentuated loss in the interface surface is also observed. In the fourth simulation the dissociation of the dimer is complete with a gradual loss of interactions between the N-terminal tails of the monomers.

The T3 structure also dissociates in two simulations, while in the others reaches a near-dissociation state. The interface is much smaller than for the other two types, hence the movement of the interface residues is accentuated.

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

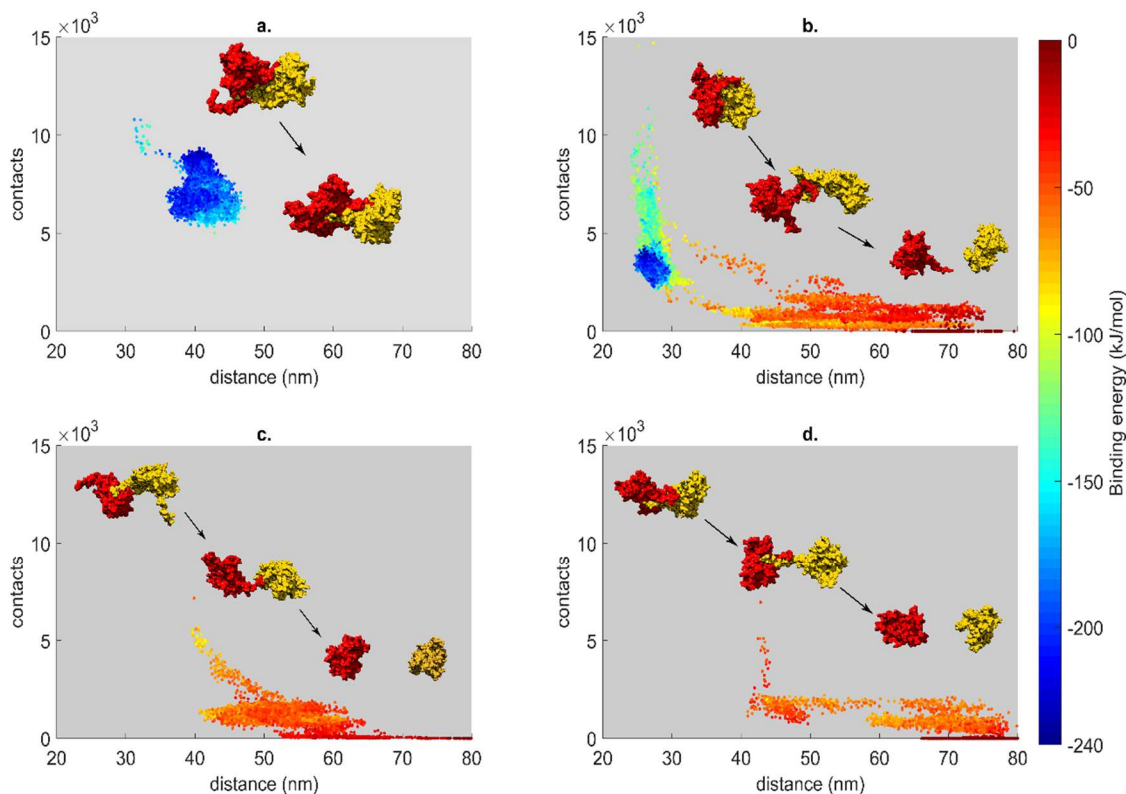


Fig. 4-20 Binding energy landscape as a function of monomer contacts and center of mass distances with initial to final structures presented, of the implicit water MD simulations for T1 (a), T2 (b), T3 (c) and TX (d) on 350K

Binding energy landscapes were created (Fig. 4-20) for the 200 ns simulations to observe the changes in the number of contacts on the PPI, distance between the chains of protein and the correlation with the binding energies.

T1 structures along the trajectories have low binding energy values, as no drastic changes can be detected. In the beginning the T2 dimer has a binding energy of -103 kJ/mol, which increases in time with the loss of contacts. The binding energy decreases exponentially to 0 as the monomers are moving away from each other to a total loss of contacts, reaching a dissociated state. T3 and TX have lower starting binding energies with fewer contacts. The landscapes are similar to that for T2 as they reach the dissociation, the binding energy and the number of contacts are becoming 0. After we found that in certain conditions the dissociation of the dimers is happening, we performed MD simulations in explicit solvent as well at 350 K, at constant pressure for 600 ns with the four dimers. The trends of C_{α} RMSD values (Fig. 4-21 a) for T1 (0.99 ± 0.15 nm) and T2 (0.78 ± 0.12 nm) are similar than for the 300 K simulations with higher fluctuations. In case of T3 (1.57 ± 0.33 nm) and TX (1.52 ± 0.42 nm) we see a major increase in values that can be explained with the structural changes of the dimers. Monomers in T1 and T2 are attached through a larger binding interface, thus the overall motion of the protein is reduced even on higher temperatures. T3 and TX are mostly bound by the tail regions of the monomers with a smaller interface, a larger motion of the protein is possible. The residues on the interface between the monomers show a different behavior (Fig. 4-21 b).

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

PCA values were linked to every atom and visualized with the NMWiz plugin in VMD [31]. Relative motions of the atoms are shown as arrows proportional with their values (Fig. 4-21 d-g).

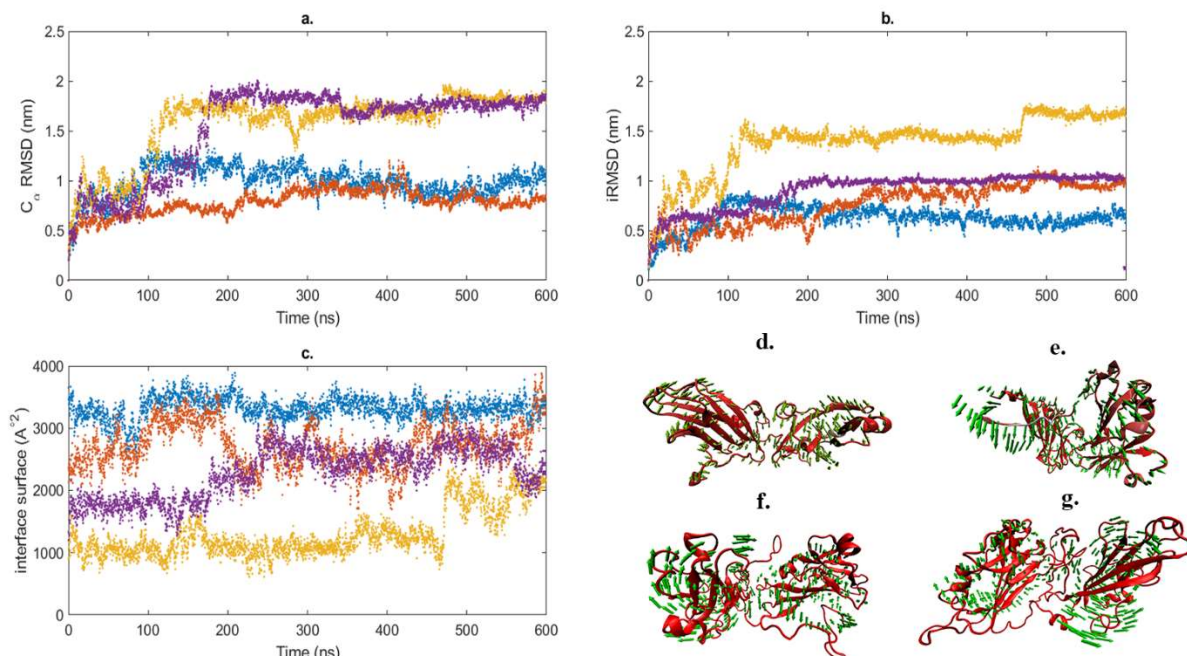


Fig. 4-21 C_{α} RMSD (a), iRMSD(b) and interface surface (c) for T1 (blue), T2 (red), T3 (yellow) and TX (purple) in explicit solvent, NPT ensemble, for 600 ns. Normal mode analysis for T1 (d), T2 (e), T3 (f) and TX (g).

The fluctuation of residues on 350 K in T1 shows a relatively balanced motion, higher values appear for irregular structures. The orientation of the arrows suggests a closing motion of the chains.

The accentuated motion of the N-terminal residues gives the biggest part of the overall fluctuation of the protein in T2 (Fig. 4-21e). A collective twisting motion of the residues in the first chain can be observed for T3 dimer (Fig. 4-21 f.), while the N-terminus of the second chain is moving away from the body of the protein. Similar motions are observed in case of TX (Fig. 4-21 g), however, with lesser impact on the interface between the monomers.

In conclusion, explicit solvent simulations at higher temperature tend to destabilize the tertiary structure of the dimers, while those in implicit solvent can destabilize the protein-protein interface.

4.7. Mutants of the T1 dimer

4.7.1. Deletion mutant of T1 and T2 dimers

During the high temperature MD runs, the T2 dimer reached a near dissociation state, however, the N-terminal tails of the dimers remained attached and prevented the complete separation of the chains. To overcome this effect, we deleted 10 residues from the N-terminals of both chains in T1 and T2 dimers and submitted these mutant proteins for further simulations.

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

NPT ensemble MD simulation was carried out for deletion mutant of T1 (T1DM) and T2 (T2DM) in explicit solvent with similar parameters as in section 4.5. During the 200 ns explicit solvent simulation on 350 K both of the dimers remained attached. The movement of the residues is accentuated, due to the increased temperature. C_{α} RMSD values remain below 0.72 ± 0.12 nm for T1DM (Fig. 4-22 a), while for the wild type in the same conditions these were around 0.99 ± 0.15 nm.

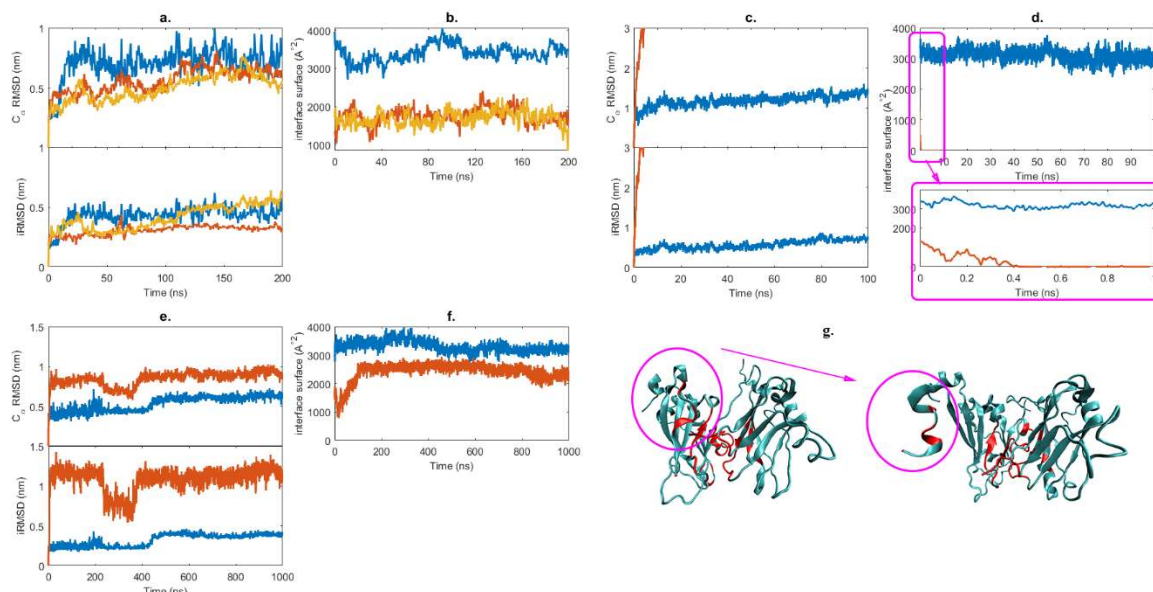


Fig. 4-22 C_{α} RMSD, iRMSD (a) and interface surface (b) for 200 ns NPT, explicit solvent simulation on 350K for T1DM (blue), T2DM (red), respectively T2DM on 400K (yellow)

C_{α} RMSD, iRMSD (c) and interface surface (d) for 100 ns implicit solvent simulation on 350K for T1DM (blue) and T2DM (red)

C_{α} RMSD, iRMSD (e) and interface surface (f) for 1 μ s implicit solvent simulation on 300K for T1DM (blue) and T2DM (red)

Similar differences can be observed for T2DM with an average of 0.61 ± 0.11 nm C_{α} RMSD on 350K and 0.60 ± 0.10 nm on 400K (400K data not presented); however, a convergence was not observed. The interface residues are behaving in same manner both at 350K and 400K. On this timescale no large difference can be observed in the movement of the dimers.

T1DM have a larger movement in 350K implicit solvent, but remains relative stable during the simulation (1.18 ± 0.15 nm C_{α} RMSD and 0.57 ± 0.11 nm iRMSD). The T2DM achieve a dissociation already within 1 ns of simulation (Fig. 4-22c). The lack of the N-terminal could cause the rapid dissociation, as in case of wild type T2 dimer the dissociation occurred after 20 ns.

We performed 1 μ s implicit simulation on 300 K to observe the overall stability of the deletion mutants (Fig. 4-22 e-f). Smaller values for T1DM on 300 K can indicate that terminal residues are important in preserving the compactness of the dimer. In case of T2DM we can detect anomalous behavior, as the C_{α} RMSD is lower than the iRMSD

These mutations are forced interventions in the structure of the proteins, but the results are demonstrating that the terminal residues are participating in the stabilization of the dimers, especially for the T1 dimer.

4.7.2. Point mutations of T1 dimer

Single point mutations on the interface of the T1 dimer were generated with replacing the original residues, both in chain B and C, based on properties as follows:

- Polar to nonpolar
- Nonpolar to polar
- Acidic to basic
- Basic to acidic
- Aromatic to polar

The resulted 178 mutant structures were minimized with CUDAGMIN with no cutoff for nonbonded interactions, and 10^{-4} RMS force convergence criterion with $igb=2$, $saltcon=0.1$. The binding energy values for all mutants, calculated from the difference of the energy values from associated and dissociated forms of the dimers, minimized to an RMS force of 10^{-4} kcal/mol, are shown in Table 4-3. Several mutant structures have low binding energy, 9 of them are below -150 kcal/mol. The results show that point mutations have the potential to significantly affect the dimerization.

Based on the calculated binding energy we chose five mutants with high energy and two mutants based on the mutation sensitivity profile generated, with the highest $\Delta\Delta G$ and one with the lowest $\Delta\Delta G$.

Table 4-3 List of selected mutations

Mutant	Mutated residue	Mutation	Binding energy (kcal/mol)	$\Delta\Delta G_{pred}$ (kcal/mol)
G16T	Glycine	Threonine	-68.5482	0.590324
S26A	Serine	Alanine	-92.6142	-0.40726
G64S	Glycine	Serine	-98.7593	0.64485
A111S	Alanine	Serine	-95.5852	0.241031
S160P	Serine	Proline	-93.2639	-0.13606
E151C	Glutamic acid	Cysteine		-2.77259
R65D	Arginine	Aspartic acid		2.330864

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

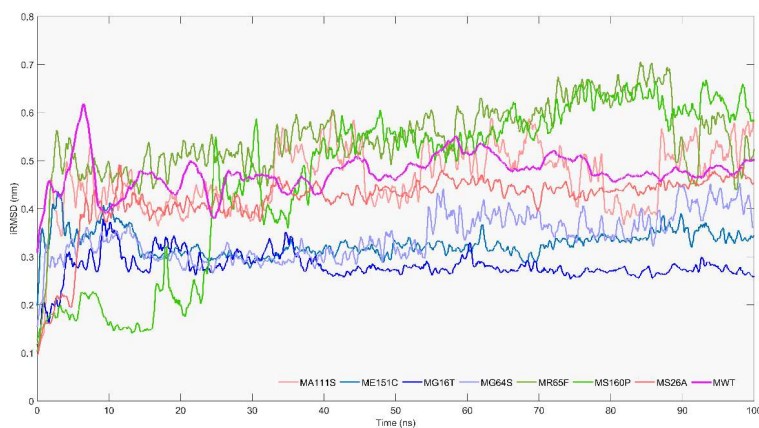


Fig. 4-23 Moving average of *i*RMSD for wild type (purple) and mutants for the 100 ns MD simulations on 300 K

MD runs were carried out for 100 ns on 300 K, in order to see the short timescale fluctuations due to the introduced mutations.

As all mutations were made on the PPI, we analyzed the behavior of the interface. The *i*RMSD of the wild type was compared with the 7 mutants (Fig. 4-23). S26A and A111S show a similar behavior to the wild type. Major differences are observed in the interface motions in both directions. G64S, G16T and E151C have lower values with around 0.2 nm, while R65D and S160P have higher values. The decrease in *i*RMSD values suggests a stabilization of the dimer, however we gained further insight into the changes caused by the mutations by a deeper analysis.

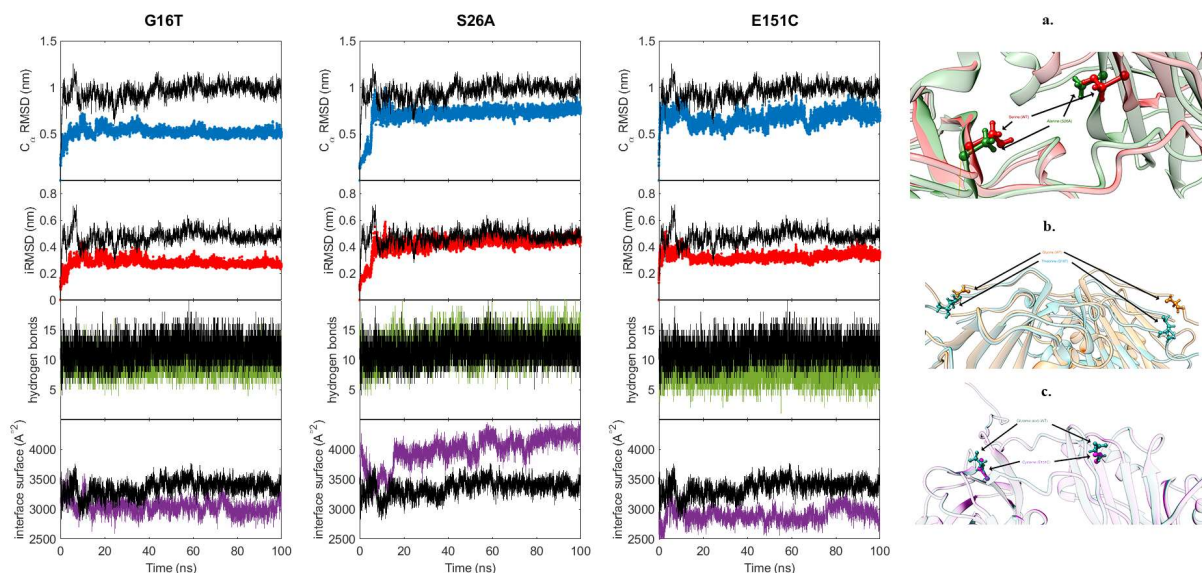


Fig. 4-24 C_{α} RMSD (blue), *i*RMSD (red), number of hydrogen bonds (green) and interface surface (purple) of G16T, S26A and E151C with respect to the WT (black). Mutated residues of G16T (a.), S26A (b.) and E151C (c.)

We selected three point mutated dimers and analyzed the 100 ns, implicit solvent MD simulations (Fig. 4-24).

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

In G16T a glycine was replaced with threonine at residue 16 and 181, parts of the N-terminal regions of the monomers (Fig. 4-24 a). Both C_{α} RMSD (0.52 ± 0.05 nm) and iRMSD (0.28 ± 0.03 nm) have decreased compared to WT (C_{α} RMSD = 0.92 ± 0.09 nm, iRMSD = 0.47 ± 0.05 nm).

In the second selected point mutant, the serine located on the 26th position is mutated to alanine (S26A). Bigger changes can be observed in the C_{α} RMSD values and interface surface. The fluctuations of interface residues are similar to those of the WT.

The E151C mutation stabilize the protein, based on the MAESTRO prediction. The results of the simulation show a similar behavior. The C_{α} RMSD and iRMSD is lower than for the WT with bigger fluctuations.

From our data gathered so far, we can conclude that further calculations are needed in order to forecast the consequence of the point mutations on the wild type CCMV protein dimer.

4.8. Replica exchange molecular dynamics simulations

Implicit solvent tREMD simulations were performed on T1, F161P and T2DM to overcome the free energy barriers, making possible a more accurate view on the changes during the simulations. Each simulation was performed for 25 ns, 200 ns in total for each structure. The temperatures used for simulations were 300.00 K, 306.65 K, 313.42 K, 320.28 K, 327.25 K, 334.34 K, 341.53 K, 348.81 K, 356.24 K. Replicas with the lowest and highest temperature distribution were selected from the 8 REMD production runs.

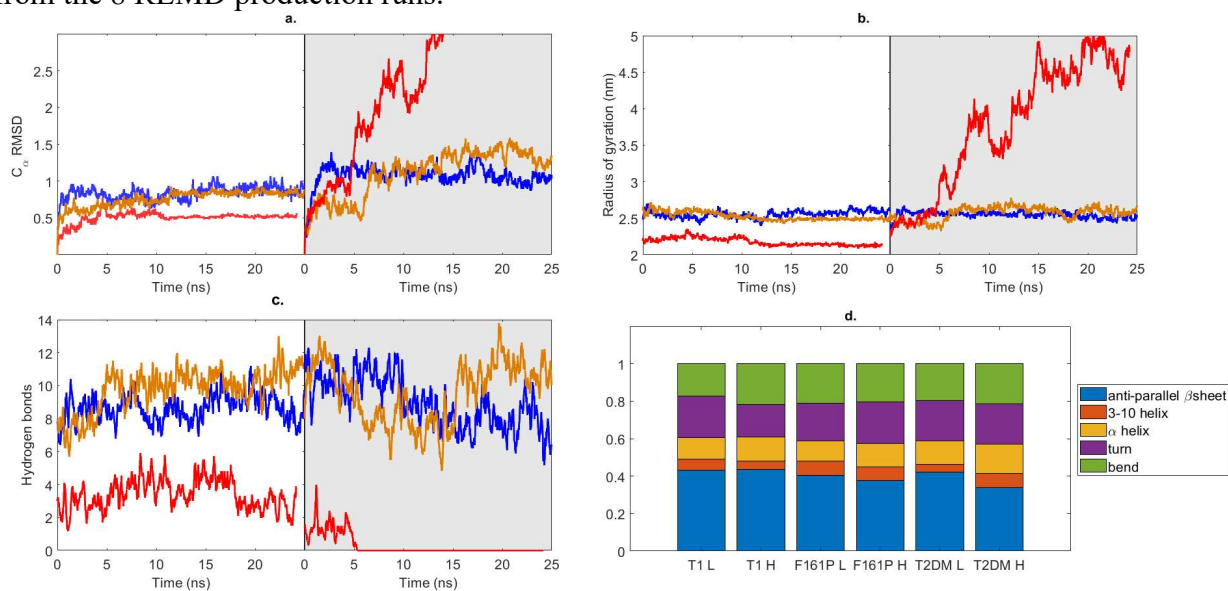


Fig. 4-25 Variation of C_{α} RMSD (a.), RoG (b.) and hydrogen bonds (c.) during the 25 ns REMD simulation for the low temperature distribution system (white background) and high temperature distribution system (gray background) of T1 (blue), F161P(yellow) and T2DM(red); Overall secondary structure composition of the investigated systems (d.)

The T1 dimer is stable in both cases with a minor fluctuation of the RoG (Fig. 4-25 b.). A major change is observed in case of the deletion mutant. On a lower temperature the RoG is smaller than in case of T1 and F161P because the lack of N-terminal residues. At higher temperatures the large increase of the RoG for the T2 deletion mutant is because of the dissociation. A slight change can be observed in the case of the F161P. On lower temperatures the stability of the dimer is increasing with the decrease of the RoG, while at higher temperatures a slow increase can be observed.

4.9. Accelerated molecular dynamics on T1, T2, T3 for 500 ns in NPT ensemble

Accelerated molecular dynamics (aMD) is an improved version of the conventional MD, where the potential is modified to reduce the height of local barriers, as the simulation can reach similar results in shorter time.

aMD was performed on T1, T2 and T3 dimers for 500 ns on 300 K and explicit solvent. The results are roughly similar to the 2 μ s cMD simulations. The average C $_{\alpha}$ RMSD of T1 was 0.92(\pm 0.15) nm. We can see an increase both in the C $_{\alpha}$ RMSD and iRMSD values after 400 ns, reaching the values of the cMD simulation. We cannot see this change in the interface surface; thus, the change is not taking place on the surface between the monomers. The change is caused by the opening movement of the monomers, the angle between the monomers is growing.

Lower RMSD values for T2 dimer is due to the attaching of the terminal tails to the body of the protein in the first phase of the simulation, hereby, the large movement of the tail regions is prohibited. T3 is behaving like in the cMD simulation, with larger conformational changes in the beginning of the simulation and fluctuation of the interface surface, caused by the loose structure of the dimer.

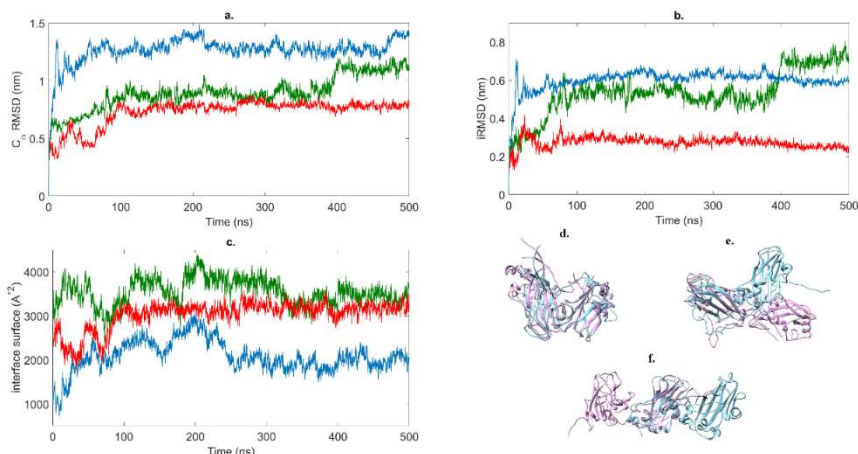


Fig. 4-26 C RMSD (a.), iRMSD (b.) and interface surface (c.) for aMD simulations for T1 (green), T2 (red) and T3(blue). Initial and final structures of the 500 ns simulations superposed are represented for T1 (d.), T2(e.) and T3 (f.)

After PCA calculations the first two eigenvectors were projected to generate a free energy landscape (Fig. 4-27), where the free energy inspects the direction of the fluctuation. The deep blue color indicates stable conformational states of molecules. Representative protein structures of the basins are also presented. The detailed analysis of cMD simulations is discussed in chapter 4.5.

T1 dimer has a single basin with a stable configuration (structure 1), as the simulation converged in the first phase. During the 500 ns aMD two energy minima can be observed (structure 2 and 3) in other locations than for the cMD. The change is due to the bending movement discussed above. However, the protein explored a larger conformational space during the aMD simulation.

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

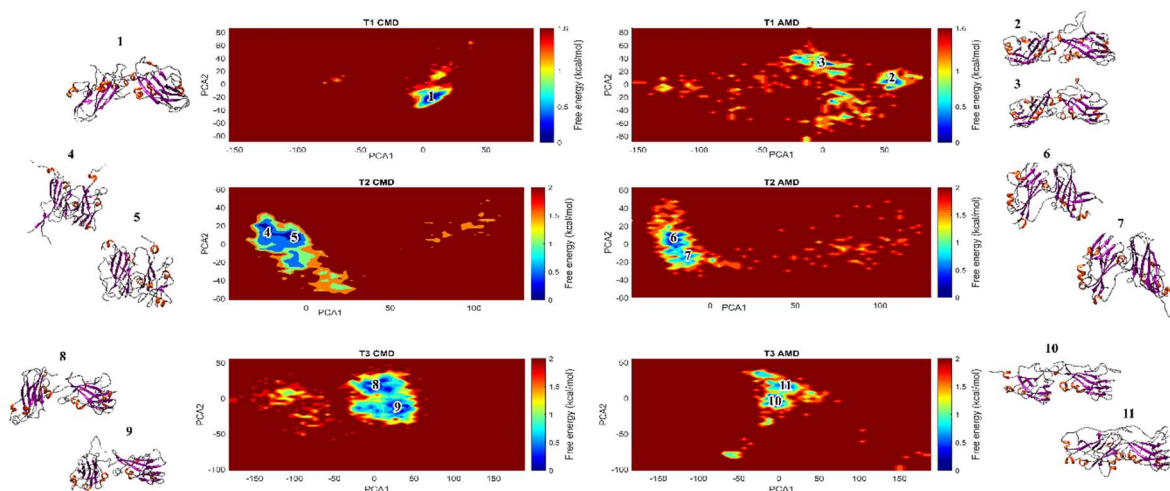


Fig. 4-27 Comparison of free energy landscapes for cMD (first column) and aMD (second column) simulations with the representative structures

T2 dimers have their minima in the same locations, the corresponding protein structures are also highly similar (structures 4 to 7). One can observe a different behavior in the case of T3 dimer. The protein has a larger mobility during the conventional simulation. The basins are found in the same places with low barriers between them, thus the transitions are likely to occur easily. The T3 dimer is relative unstable with less interactions between the monomers causing the flexibility of the protein.

4.10. Pentamer of dimers

A pentamer of dimers (PD) was selected from the capsid of ss-CCMV (pdb ID 1za7). The inner part of the PD consists of five C chains connected with the outer chains through T1 interfaces, while T2 interfaces are present between them (Fig. 4-28 c).

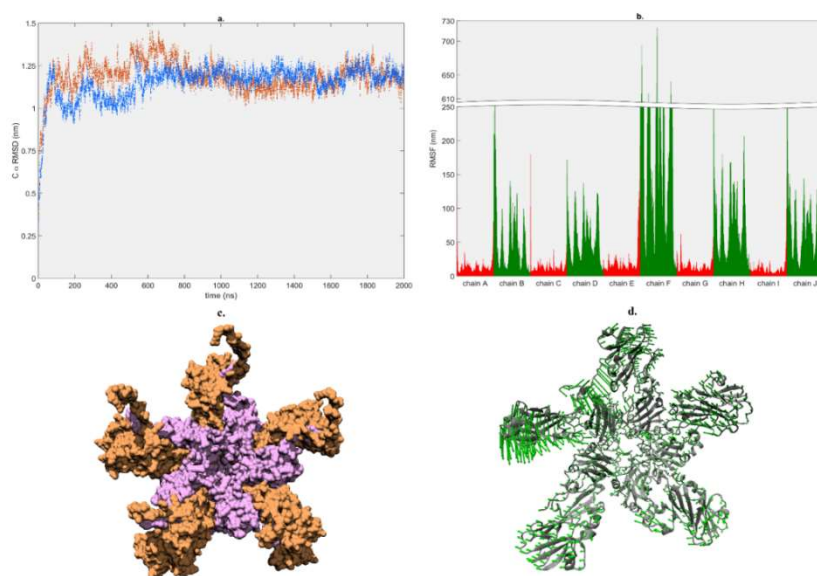


Fig. 4-28 C α RMSD of the 2 parallel runs of PD; b. Residue fluctuation of the PD, colored by chains: inner chains - red, outer chains – green; c. Structure of PD with surface representation; d. Motion of the residues during the simulation based on PCA calculation

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

Long timescale all-atom simulations were carried out in explicit water for the PD and analyzed for a better view of conformational changes in the CCMV capsid. Two simulations of NPT molecular dynamics were conducted for 2 microseconds in explicit water with different initial velocities. For a better view of the changes during the simulation we separated the dimers by the interface type, therefore we obtained 5 dimers of T1 interface and 5 dimers of T2 interface.

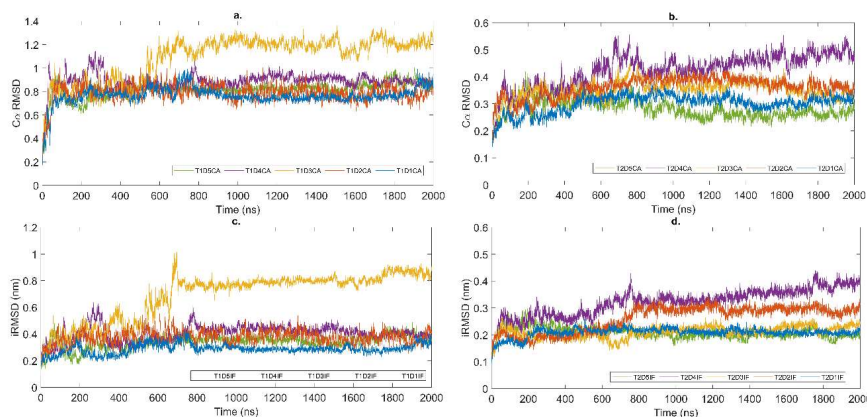


Fig. 4-29 C_{α} RMSD for T1 (a) and T2 (b) dimers. iRMSD for T1 (c) and T2 (d) dimers

Among the T1 dimers the third (T1D3) shows a different behavior from the others. After 500 ns simulation an increase of 0.4 nm both for C_{α} RMSD and iRMSD can be observed. The outer monomer moves upward from the plane of the PD. The shift is similar to that seen with a standalone T1 dimer. The N-terminal regions of the inner monomers are arranged in a specific conformation, forming H-bonds with the tail regions of two neighboring chains on the right side. The arrangement is conserved during the simulation, thus the role of the N-terminal region in the pentamer formation and stabilization is emphasized.

Taking into consideration the results, we can state that the T1 D3 dimer from the decamer is behaving more like the standalone T1 dimer.

The changes in the number of hydrogen bonds on the interfaces during the simulation was calculated. A small decrease in the number of hydrogen bonds can be observed for the two dimers (T1 D1, T1 D3). The other T1 dimers had similar values as for the standalone dimers.

T2 D3 and T2 D4 behave similarly to the standalone dimer. Although the movement of T2 dimers in the decamer is prohibited, the analogous behavior to the standalone dimer can lead to the conclusion that the interface between the monomers determines considerably the behavior of the decamer.

5. Conclusions

5.1. Results and original contributions

In the experimental part of the work, we studied the expression of the CCMV capsid protein in *E. coli* host cell.

The plasmid containing the DNA of the protein was acquired from commercial resource and transformed in *E. coli BL21(DE3) Rosetta* cells. Previously the ss-CCMV protein's DNA was cloned into a vector containing a His-tag and a ubiquitin fusion partner.

Cloning of the plasmid and expressing the pUBK_CCMV was successful, confirmed by sequencing. Expression was optimized with a series of experiments with changing the expression temperature and the concentration of inducer. The protein was purified with Ni-affinity chromatography and cleaved from the ubiquitin with the YUH1 enzyme. The expression system presented can be used to the expression of ss-CCMV's coat protein with high yields.

In our work we studied certain aspects of the self-assembly of the ss-CCMV virus capsid, through modelling and investigating the interactions between the capsid protein chains to form capsomers, constituents of the virus capsid. We examined the correlation between the known interfaces and the modelled protein dimers and found structural fits for all three types of interfaces. However, based on the binding energy and RMSD information, we can predict that the CC dimers (T1) are the most important in the formation of the capsid. This conclusion is confirmed by the original assembly pathway, which proceeds through dimerization and cooperative addition of dimers. Predictions made by the PISA webserver also show this interface to be the most important for forming the biological unit. It is also important to note that it is possible to construct the whole capsid by adding T1 dimers into a pentamer of dimers in a subsequent fashion. The pentamer of dimers is the experimentally determined nucleus for CCMV assembly [33].

The best few structures generated from ZDOCK and organized in order of the best docking score correspond quite well with our simulated data. For one particular dimer (BC), the first structures from the ZDOCK output are among the dimers with the lowest binding energies calculated with AMBER. This leads us to the conclusion that the AMBER force field can be used for protein-protein interface modelling with good predictive power. However, sampling of initial configurations is an important issue. The possible conformational flexibility of the protein also has to be taken into account; therefore, it is advisable to start ZDOCK runs from slightly different protein conformations. Any prediction is as good as the underlying sampling, therefore it is very important to start from as many different dimer configurations as possible, if we want to make blind docking.

After the docking studies, we investigated the behavior of the best dimer structure by carrying out MD simulations in various circumstances.

Studying in detail the interactions on protein complex binding interfaces can help to understand the overall process of capsid formation. Simulation of protein complex formation is difficult and time consuming, therefore we chose to analyze the reverse process, dissociation. Temperature is one factor that influences the stability of protein complexes. We performed calculations at different temperatures and different timescales.

Short timescale MD simulations performed on the structures with the best alignment to the original interfaces led to the selection of three dimers for further investigations.

During short simulations for the T1 ss-CCMV capsid protein dimer on higher temperatures the protein started to unfold, and we did not observe dissociation on this timescale. We predicted the

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

stability of the secondary structures, β -sheets arranged in a β -barrels seem to be more stable than the α -helices.

2 μ s long classical MD runs in explicit solvent proved the increased stability of T1 compared to the T2 dimer. Although the latter structure did not dissociate, the interface was continuously changing throughout the simulation. T3 was proved to be the least stable among the different interface types.

The dimers were simulated at different temperatures (300 and 350 K). In case of the T1 configuration only small temperature-dependent changes were noticed, on high temperatures a slow unfolding process has begun. In contrast to that, T2, as expected from its lower stability, reached a near-dissociated state on 350 K, only the N-terminal tails of the monomer chains prevented the separation. To reduce this effect, we deleted ten N-terminal residues from both chains and the repeated simulation resulted a complete dissociation of the protein chains.

Replica exchange simulations carried out in implicit solvent at a temperature range of 300 to 356 K support our previous results summarized below.

The T1 dimer is stable over longer simulations and on higher temperatures as well, unfolding is preferred instead of dissociation when increasing the temperature, showing the extremely high stability of the T1 interface. T2 is less stable and loses its original interface at higher temperatures but the total dissociation is obstructed by the N-terminal tail in some cases. The deletion mutant, originating from T2 dimer, shows a rapid dissociation on higher temperature. However, total dissociation was observed for the original T2 configuration as well in some high temperature implicit solvent MD runs. Fast dissociation was also observed for the T3 interface under those conditions. The flexible N-terminal tail can play a role in dimer formation by acting as an anchor, possibly facilitating initial contact between two monomers.

Point mutations were generated for the interface residues and the mutants were minimized. Four random structures were selected and MD simulations were performed to predict any changes affecting the stability of proteins. Significant changes cannot be observed and further calculations are needed to find a mutation that has a bigger impact on the protein stability.

After we gained a deeper insight in the behavior of the protein dimers, we chose a pentamer of dimers from the original ss-CCMV capsid, as this is the next level in the self-assembly process. The pentamer of dimers, containing T1 and T2 interfaces was stable during the 2 μ s, explicit solvent simulation. The separated dimers presented different compartment during the simulation, some of them showed similarity to the simulation of standalone dimers.

5.2. List of original publications

5.2.1. Publications

1. Z. Antal, **J. Szoverfi**, and S. N. Fejer, Predicting the Initial Steps of Salt-Stable Cowpea Chlorotic Mottle Virus Capsid Assembly with Atomistic Force Fields, *J. Chem. Inf. Model.*, vol. 57, no. 4, 2017, doi: 10.1021/acs.jcim.7b00078, **IF₂₀₁₇= 3.804**
2. **J. Szövérfi**, Cs. K. Orbán, B. Albert, K. Nagy, P. Salamon, Sz. Lányi, In Vitro Study Of The CCMV Capsid Protein: Cloning, Expression, And Purification, *U.P.B. Sci. Bull., Series B*, Vol. 83, Iss. 1, 2021, pp. 135-142
3. **Szoverfi, J.**, Fejer, S.N. Dynamic stability of salt stable cowpea chlorotic mottle virus capsid protein dimers and pentamers of dimers. *Sci Rep* 12, 14251 (2022). <https://doi.org/10.1038/s41598-022-18019-9>, **IF₂₀₂₁= 4.996**

5.2.2. Conferences

1. Molecular modeling in chemistry and biochemistry MOLMOD 2016, November 2016, Cluj Napoca, Romania “Modelling the dimerization of the CCMV capsid protein”, oral presentation
2. 22nd International Conference on Chemistry, November 2016, Timisoara, Romania, “Modelling the dimerization of the CCMV capsid protein”, oral presentation
3. 20th Romanian International Conference on Chemistry and Chemical Engineering, September 2017, Poiana Brasov, Romania, “Modelling the thermal stability of wild-type and mutant dimers of the CCMV capsid protein”, poster
4. 23rd International Conference on Chemistry, October, 2017, Deva, Romania, “Modelling the thermal stability of wild-type and mutant dimers of the CCMV capsid protein”, oral presentation
5. 24th International Conference on Chemistry, October, 2018, Sovata, Romania, “Cloning, Heterologous Expression and Molecular Dynamics Simulation of the CCMV Capsid Protein”, oral presentation
6. Molecular modeling in chemistry and biochemistry MOLMOD 2018, October 2019, Cluj Napoca, Romania, “Molecular Dynamics Studies of CCMV Capsid Protein oligomers”, oral presentation

5.3. Perspectives for further developments

The self-assembly of CCMV is widely studied both *in vitro* and *in silico* from the 60's, however many questions remained unanswered. The empty capsids are utilized as nanocarriers, thus numerous applications are possible.

To continue the investigation of the capsid formation a potential energy surface can be generated for two rigid protein units. An analytic function can be fitted to the potential energy surface.

With the coarse-graining of structures of the stable dimer interfaces, the simulate of the self-assembly of coarse-grained protein units can be simplified, achieving the simulation of the whole capsid.

Regarding the experimental part mutants can be generated, which disrupt the best interface, and a comparison of the assembly kinetics for the wild-type and the mutant protein can be made.

The digestion of the proteins and the purification process can be optimized further, for a higher yield of the CCMV capsid protein.

Large-scale expression of proteins can be performed in bioreactor.

In vitro capsid formation can be investigated and detected with fluorescence polarization assay.

Bibliography

- [1] A. Zeltins, "Construction and characterization of virus-like particles: A review," *Mol. Biotechnol.*, vol. 53, no. 1, pp. 92–107, 2013, doi: 10.1007/s12033-012-9598-4.
- [2] R. F. Bruinsma, G. J. L. Wuite, and W. H. Roos, "Physics of viral dynamics," *Nature Reviews Physics*, vol. 3, no. 2. Springer Nature, pp. 76–91, Feb. 01, 2021, doi: 10.1038/s42254-020-00267-1.
- [3] E. C. Dykeman, P. G. Stockley, and R. Twarock, "Building a viral capsid in the presence of genomic RNA," *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, vol. 87, no. 2, Feb. 2013, doi: 10.1103/PhysRevE.87.022717.
- [4] E. R. May, K. Arora, R. V. Mannige, H. D. Nguyen, and C. L. Brooks, "Multiscale Modeling of Virus Structure, Assembly, and Dynamics," pp. 167–189, 2012, doi: 10.1007/978-1-4614-2146-7_7.
- [5] L. Lavelle, J. P. Michel, and M. Gingery, "The disassembly, reassembly and stability of CCMV protein capsids," *J. Virol. Methods*, vol. 146, no. 1–2, pp. 311–316, 2007, doi: 10.1016/j.jviromet.2007.07.020.
- [6] Z. Antal, J. Szoverfi, and S. N. Fejer, "Predicting the Initial Steps of Salt-Stable Cowpea Chlorotic Mottle Virus Capsid Assembly with Atomistic Force Fields," *J. Chem. Inf. Model.*, vol. 57, no. 4, 2017, doi: 10.1021/acs.jcim.7b00078.
- [7] D. Law-Hine *et al.*, "Reconstruction of the disassembly pathway of an icosahedral viral capsid and shape determination of two successive intermediates," *J. Phys. Chem. Lett.*, vol. 6, no. 13, pp. 3471–3476, 2015, doi: 10.1021/acs.jpcllett.5b01478.
- [8] H. A. Yu *et al.*, "Characterization of ubiquitin C-terminal hydrolase 1 (YUH1) from *Saccharomyces cerevisiae* expressed in recombinant *Escherichia coli*," *Protein Expr. Purif.*, vol. 56, no. 1, pp. 20–26, 2007, doi: 10.1016/j.pep.2007.07.005.
- [9] "RCSB PDB - 1ZA7: The crystal structure of salt stable cowpea chlorotic mottle virus at 2.7 angstroms resolution." <https://www.rcsb.org/structure/1ZA7> (accessed May 29, 2021).
- [10] "ZDOCK Server: An automatic protein docking server." <https://zdock.umassmed.edu/> (accessed May 29, 2021).
- [11] S. Chaudhury, M. Berrondo, B. D. Weitzner, P. Muthu, H. Bergman, and J. J. Gray, "Benchmarking and analysis of protein docking performance in Rosetta v3.2," *PLoS One*, vol. 6, no. 8, p. 22477, 2011, doi: 10.1371/journal.pone.0022477.
- [12] C. Schafmeister, W. S. Ross, and V. Romanovski, "LEaP," *Univ. California, San Fr.*, 1995.
- [13] Y. Duan *et al.*, "A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations," *J. Comput. Chem.*, vol. 24, no. 16, pp. 1999–2012, 2003, doi: 10.1002/jcc.10349.
- [14] "How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? - Wang - 2000 - Journal of Computational Chemistry - Wiley Online Library." <https://onlinelibrary.wiley.com/doi/10.1002/1096-987X%28200009%2921%3A12%3C1049%3A%3AAID-JCC3%3E3.0.CO%3B2-F> (accessed May 29, 2021).
- [15] P. Alam and N. Van Hieu, "Protein unfolding versus β -sheet separation in spider silk nanocrystals," *Adv. Nat. Sci. Nanosci. Nanotechnol.*, vol. 5, no. 1, p., 2014, doi: 10.1088/2043-6262/5/1/015015.
- [16] A. Onufriev, D. Bashford, and D. A. Case, "Modification of the generalized born model suitable for macromolecules," *J. Phys. Chem. B*, vol. 104, no. 15, pp. 3712–3720, Apr. 2000, doi: 10.1021/jp994072s.
- [17] D. J. Wales and J. P. K. Doye, "Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms," *J. Phys. Chem. A*, vol. 101, no. 28, pp. 5111–5116, Jul. 1997, doi: 10.1021/jp970984n.
- [18] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein,

Expression of the ssCCMV capsid coat protein and molecular dynamics simulations on the ssCCMV capsid

- "Comparison of simple potential functions for simulating liquid water," *J. Chem. Phys.*, vol. 79, no. 2, p. 926, 1983, doi: 10.1063/1.445869.
- [19] I. Maffucci and A. Contini, "Improved Computation of Protein-Protein Relative Binding Energies with the Nwat-MMGBSA Method," *J. Chem. Inf. Model.*, vol. 56, no. 9, pp. 1692–1704, 2016, doi: 10.1021/acs.jcim.6b00196.
- [20] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems," *J. Chem. Phys.*, vol. 98, no. 12, pp. 10089–10092, 1993, doi: 10.1063/1.464397.
- [21] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes," *J. Comput. Phys.*, vol. 23, no. 3, pp. 327–341, 1977, doi: 10.1016/0021-9991(77)90098-5.
- [22] A. Patriksson, D. van der Spoel, H. J. C. Berendsen, B. M. Rode, R. Huber, and H. J. C. Berendsen, "A temperature predictor for parallel tempering simulations," *Phys. Chem. Chem. Phys.*, vol. 10, no. 15, p. 2073, Apr. 2008, doi: 10.1039/b716554d.
- [23] E. Feyfant, A. Sali, and A. Fiser, "Modeling mutations in protein structures.," *Protein Sci.*, vol. 16, no. 9, pp. 2030–41, 2007, doi: 10.1110/ps.072855507.
- [24] D. R. Roe and T. E. Cheatham, "PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data," *J. Chem. Theory Comput.*, vol. 9, no. 7, pp. 3084–3095, 2013, doi: 10.1021/ct400341p.
- [25] W. Humphrey, A. Dalke, and K. Schulten, "VMD: Visual molecular dynamics," *J. Mol. Graph.*, vol. 14, no. 1, pp. 33–38, 1996, doi: 10.1016/0263-7855(96)00018-5.
- [26] E. F. Pettersen *et al.*, "UCSF Chimera—A Visualization System for Exploratory Research and Analysis," *J. Comput. Chem.*, vol. 25, pp. 1605–1612, 2004, doi: 10.1002/jcc.20084.
- [27] M. L. Connolly, "Analytical molecular surface calculation," *J. Appl. Crystallogr.*, vol. 16, pp. 548–558, 1983, doi: 10.1107/S0021889883010985.
- [28] W. Kabsch and C. Sander, "Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features," *Biopolymers*, vol. 22, no. 12, pp. 2577–2637, 1983, doi: 10.1002/bip.360221211.
- [29] B. R. Miller, T. D. McGee, J. M. Swails, N. Homeyer, H. Gohlke, and A. E. Roitberg, "MMPBSA.py: An efficient program for end-state free energy calculations," *J. Chem. Theory Comput.*, vol. 8, no. 9, pp. 3314–3321, 2012, doi: 10.1021/ct300418h.
- [30] A.-M. Catanzariti, T. A. Soboleva, D. A. Jans, P. G. Board, and R. T. Baker, "An efficient system for high-level expression and easy purification of authentic recombinant proteins.," *Protein Sci.*, vol. 13, no. 5, pp. 1331–9, May 2004, doi: 10.1110/ps.04618904.
- [31] A. Bakan, L. M. Meireles, and I. Bahar, "ProDy: Protein dynamics inferred from theory and experiments," *Bioinformatics*, vol. 27, no. 11, pp. 1575–1577, 2011, doi: 10.1093/bioinformatics/btr168.
- [32] J. A. McCammon, B. R. Gelin, and M. Karplus, "Dynamics of folded proteins.," *Nature*, vol. 267, no. 5612, pp. 585–90, Jun. 1977, Accessed: Jan. 19, 2017. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/301613>.
- [33] A. Zlotnick, R. Aldrich, J. M. Johnson, P. Ceres, and M. J. Young, "Mechanism of capsid assembly for an icosahedral plant virus," *Virology*, vol. 277, no. 2, pp. 450–456, 2000, doi: 10.1006/viro.2000.0619.