MOLECULAR DYNAMICS SIMULATIONS OF ECHOVIRUS 1

Emmi Pohjolainen

Master's thesis in natural sciences
University of Jyväskylä, Department of physics
June 23, 2014
Supervisors: Prof. Hannu Häkkinen, Dr. Gerrit Groenhof

Abstract

Covalently attaching thiol-functionalized gold nanoclusters to virus surfaces could provide new visualization method for tracking viruses with TEM [1]. This type of imaging could provide better resolution images and shed light, e.g., on virus infection pathways [1]. The controllable binding of these nanoclusters would be essential, and this requires atomic scale information on potential binding sites. In addition atomic scale structural information on virus-nanocluster complexes and their dynamics are needed. Both types of information are potentially obtainable from an all-atom simulation. The aim of this work was to construct a valid model system and set up simulation for full Echovirus 1 capsid in salt solution. A model system of the full Echovirus 1 in salt solution is proposed and simulated up to 200 ns. Also comparative systems with different ligands are proposed and simulated. Based on the simulation the validity of the model is assessed and established in terms of convergence of energy terms, structure and relations to experiments. More detailed look is given considering behavior of potential binding sites, hydrophobic pocket and different pocket factors and cysteine residues. The recognized points of interest to be studied in further simulations are discussed. The proposed model then lays foundation for more extensive studies of this virus with molecular dynamics simulations.

Tiivistelmä

Kokeellisesti on havaittu, että tietyllä tavalla funktionalisoidut kultananoklusterit voivat kiinnittyä kovalenttisesti virusten pinnoille. Virusten leimaaminen kultananoklustereilla voi mahdollistaa niiden paremman kuvantamisen TEM:llä, jolloin voitaisiin saada tarkempaa tietoa esimerkiksi virusten infektiomekanismeista. Jotta kultaklusteri -leimaamista voitaisiin parhaiten hyödyntää kuvantamisessa, on klusterien sitoutumispaikkoja voitava kontrolloida tarkasti. Tähän puolestaan tarvitaan atomitason tietoa mahdollisten sitoutumispaikkojen rakenteista ja dynamiikasta viruksen pinnalla. Tämäntyyppistä atomitason tietoa voidaan saada molekyylidynamiikka simulaatiosta.

Kokonaisten virusten simuloiminen voi olla haastavaa. Erityisesti on kiinnitettävä huomiota mallin fysikaaliseen mielekkyyteen. Toisin sanoen, simulaatioiden tuloksia tulee aina verrata kokeellisiin faktoihin. Lisäksi on huomioitava, että tällaisten simulaatioiden aikaskaala, maksimissaan n. 1 μs, on hyvin lyhyt verrattuna kokeellisten havaintojen aikaskaalaan. Näin ollen on odotettavissa, että ns. perinteisessä simulaatiossa ei nähdä esimerkiksi laajoja konformaatiomuutoksia. Tietyissä tapauksissa kokeellista dataa voidaan kuitenkin hyödyntää niin, että simulaatiota voidaan sopivasti stimuloimalla ohjata haluttuun suuntaan.

Tämän työn tavoitteena oli valmistella ja toteuttaa molekyylidynamiikka simulaatio kokonaiselle Echovirus 1 systeemille. Erityisesti työ keskittyi mallin koostamiseen ja rakennetun mallin sopivuuden vahventamiseen käytettäväksi myöhemmissä simulaatioissa. Tähän liittyen energiatermien ja rakenteen konvergoitumista tarkkailtiin simulaatioissa. Lisäksi tarkasteltiin yleisiä dynaamisia piirteitä simulaatioissa, ja vertailtiin niitä kokeellisiin tietoihin. Lisäksi simulaatioiden varsinaisten tulosten analysoinnissa keskityttiin tarkemmin kultaklusterien mahdollisten sitoutumispaikkojen dynamiikkaan ja rakenteisiin. Lisäksi simulaatioiden perusteella pyrittiin löytämään kiinnostavia kohteita, joita voitaisiin tutkia myöhemmissä simulaatioissa eri menetelmin.

Ajettujen simulaatioiden perusteella todettiin, että ottaen huomioon systeemin koon ja aikaskaalan mallinnettu systeemi on fysikaalisesti mielekäs ja vakaa käytettäväksi myöhemmissä simulaatioissa ja muissa analyysimenetelmissä. Lisäksi todettiin, että sopivasti hyödyntämällä erilai-

sia analyysimenetelmiä, simulaatioista voidaan saada hyödyllistä tietoa liittyen viruksen rakenteeseen ja dynamiikkaan. Tässä työssä on siten luotu perustaa molekyylidynamiikka simulaatioiden hyödyntämiselle tässä virustutkimuksessa.

Acknowledgments

I would like to thank my supervisor Professor Hannu Häkkinen for the great opportunity of working with an interesting, multidisciplinary project and group and instructions throughout the work.

Also I would like to express my gratitude to my instructor Dr. Gerrit Groenhof for the extensive instructions, guidance and discussions throughout the work.

A special thanks also to Dr. Sami Malola for the instructions during this work.

There were also many more people contributing to this work in different ways, so in general I would like to thank everyone involved.

Contents

1 Introduction

1.1 Molecular dynamics simulations in biology

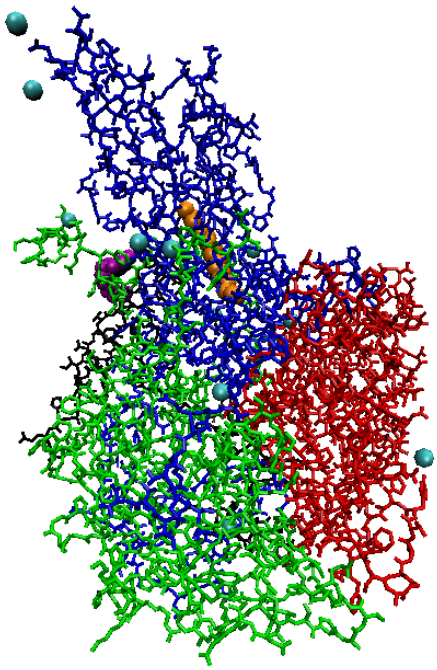
In theoretical studies of biological systems, classical molecular dynamics simulation is often the method of choice. This is because such a classical simulation can at best reach up to order of milliseconds of simulation time or handle systems containing millions of atoms, often a minimum required in solvated biological systems. Such a simulation can greatly complement experimental results by revealing atomistic details not available by experiments. While providing the details of dynamics at the atomistic level, there are challenges that limit the abilities of simulations in studying biological systems. [2–4]

One clear limitation is the time scale, which in practice, e.g., for a full virus system is limited to in order of hundreds of nanoseconds or up to microsecond. In a such complex biological system with traditional simulation this time scale can not cover for example opening of a virus. Therefore, making a connection between experiment and simulation is not necessary unambiguous. Even more so, large structural changes need somehow be triggered in the simulation in order for them to occur. This requires experimental data. In some cases proper experimental data can be utilized in driving the simulation to a certain direction by steering or targeting or by starting a simulation from a specific conformation. In that sense the time scale problem can be overcome and for example virus opening could possibly be studied.

Another important question with a large simulation is whether the model being used is actually physically meaningful. In addition and also related to this is the question whether the system is able to equilibrate enough in the simulation, that is, is the convergence sufficient and the model stable to be further analyzed or simulated. One additional challenge is brought by the sheer size of the system and, as a result, the magnitude of data it contains for each atom in the format of trajectory and energetics. Proper handling of data with statistical analysis and furthermore finding the points of interest is therefore not always trivial. In any case, a simulation always needs to be closely related to experiments to establish its validity.

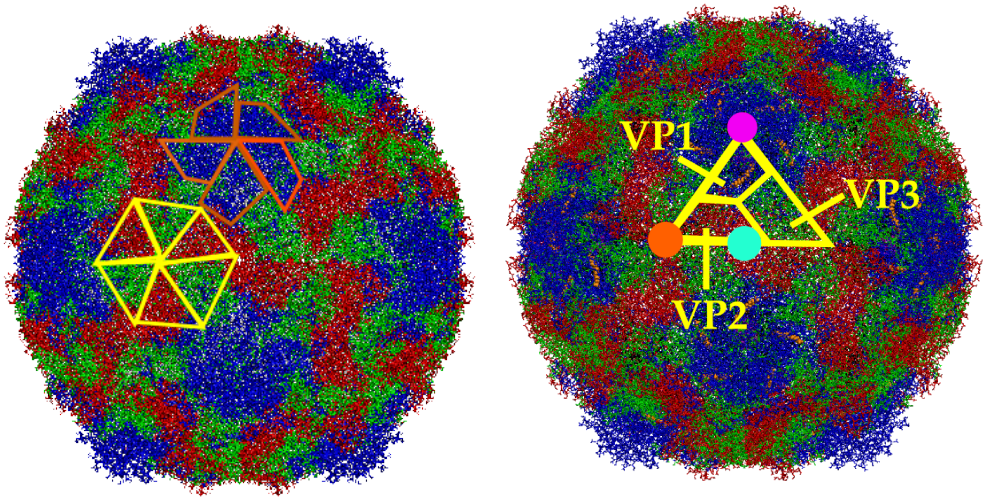
1.2 Echovirus 1

Echovirus 1 (EV1) belongs to Enterovirus genus, and it consists of the protein capsid and a single-stranded, negatively charged RNA genome. The capsid is composed of 60 protomers, each protomer containing 4 proteins: VP1-VP4. 60 protomers arrange to form an icosahedral capsid. The structure of the capsid is presented in figure 1. [5–7]



(a)

Figure 1: General structure of EV1 capsid with symmetry axes and surface shapes highlighted. (a) shows virus proteins VP1-VP4 (blue, red, green and black respectively), pocket factor (orange spheres), myristate group attached to VP4 (purple spheres) and crystallographic waters present in the structure (cyan spheres) in one protomer.



(b) (c)

Figure 1: General structure of EV1 capsid with symmetry axes and surface shapes highlighted. In (b) certain shapes on capsid surface are highlighted: pentamer (formed by VP1) and propeller (formed by VP2, VP3). (c) shows the arranging on VPs on capsid surface and also 5-, 3- and 2-fold symmetry axis are marked with magenta, orange and cyan points, respectively.

Proteins VP1-VP3 of one protomer form a triangular structure on the capsid surface. With 60 protomers they form a shell of approximate thickness of 3 nm and the different proteins are arranged in a way that VP1s form star-shaped structures around 5-fold symmetry axis; VP2s and VP3s in turn alternate around 3-fold symmetry axis forming propeller shaped structures. Between star-shaped mesas and propellers is formed the so called canyon, which contains residues important in binding to cell receptors. [5–7]

VP1 in addition contains a hydrophobic pocket. This pocket is occupied by a pocket factor, palmitic acid (C16 fatty acid) in EV1 [8]. The presence of the pocket factor is thought to be crucial in terms of capsid stability. It is also thought to play an important role in genome release: the presence of the pocket factor may prevent the capsid from opening and in order for genome to be released the pocket factor must be externalized. Another mechanism for pocket factor function could be that it causes certain con-

figurational changes enabling the genome release. In general, the pocket is an important drug target and it has been shown in many Picornaviruses that molecules that replace the pocket factor affect the infectivity. This infectivity could be limited by for example stabilizing the structure so that the genome cannot be released. [5, 9, 10]

In general, the attaching to receptors launches conformational changes in the capsid that eventually include externalization of VP4 and N-terminal of VP1. These changes could possibly also be induced by heating or decrease in pH. These are both irreversible changes that occur upon infection. Other structural changes include virus breathing motions, and possible partial externalization of some other parts of the capsid as well. The exit routes for RNA, VP4 and N-terminal of VP1 were previously proposed to be through 5-fold axis channel, whereas recent studies suggest exit close to 2-fold symmetry axis, where two VP2 helices would be torn apart during expansion. [11]

The release of the pocket factor is thought to destabilize the capsid causing externalization of VP4, N-terminal of VP1 and genome. Therefore, replacing a pocket factor with a proper molecule could aid the regulation of infectivity. Such a molecule needs to bind with high affinity, so that it will not be released upon receptor binding. The working mechanism of the pocket factor replacing molecule could then be then that since it won't be displaced, the capsid is stabilized so that the uncoating cannot occur. Other possible mechanism is that the binding of the pocket factor or its replacement induces conformational changes to receptor binding regions, thus inhibiting the actual attachment to the cell. In a way the pocket factor then both stabilizes the capsid before cell attaching and destabilizes the capsid after the attachment. In any case pocket factor release is thought be required for the uncoating. The empty particles without the pocket factor could also exhibit differences in the shape of the pocket. This possible closure of the empty pocket could initiate the uncoating via connection to the pentamer interface. [11–13]

VP4 is located on the inner surface of the capsid and interacts with the negatively charged genome. It is attached to myristate group (myristic acid is a C14 fatty acid) from its N-terminal. Presence of myristate group is thought to aid the genome encapsidation, pentamer formation and general capsid stability. Also it may be important in channel formation, through which parts of the virus externalize upon infection. [7, 9, 10]

In virus assembly, also genome may have significant role; however also empty capsids are known to be able to assemble. For example empty EV1 capsid can exist, but some of its properties, e.g., ability to attach cell receptors, could be limited. In general, empty particles can be expanded compared to full viruses: this could result in altered surface properties and interactions that hold the capsid together. [11]

Structural details and relations to experiments that are most interesting in the scope of this work are based on the observation that certain thiol-functionalized gold nanoclusters are able to attach to the EV1 capsid surface. The sites for attaching are some of the cysteine residues (sulfurs) close to the capsid surface. These gold clusters can then provide visualization method for TEM imaging of viruses; this could reveal interesting details on virus infection pathways and mechanisms. [1]

However utilization of gold clusters in imaging requires control over the specific positions where these clusters attach. Therefore it would be of interest to study which factors could affect the mechanism of this attachment and where potential binding sites are located. In addition, it would be of interest to study general structural characteristics and dynamics of the virus-gold cluster complexes. Both of these points would require atomic scale information on virus structure and its dynamics, making molecular dynamics simulation a suitable research tool to address these issues.

In addition to cysteine residues, a potential point for gold cluster attachment is the hydrophobic pocket. The natural pocket factor could be replaced by some other molecule containing a proper linker that in turn could be attached to the gold cluster. This would require the preferable binding of this new molecule compared to the natural pocket factor.

One drug molecule known to be effective towards some Enteroviruses in replacing the natural pocket factor and thus preventing the virus from opening is Pleconaril. Therefore, a potential molecule for replacing the natural pocket factor could be some derivative of Pleconaril molecule. One such molecule designed and synthesized at Jyväskylä University is Kirtan1. In addition to Pleconaril-like - parts it contains a long, aliphatic linker tail.

The interest in gold cluster attachment is its potential application as a visualization agent. In ideal case the gold clusters would be able to be con-

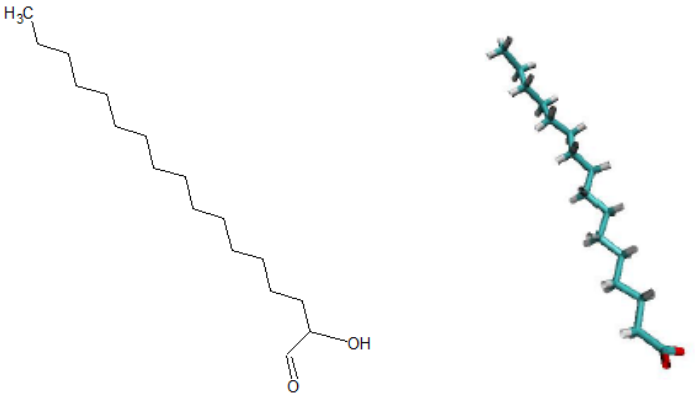
trollably attached to specific parts on virus capsid. The proper labeling with these gold clusters could enhance for example visualization of virus opening and movements in cells.

1.3 Goals

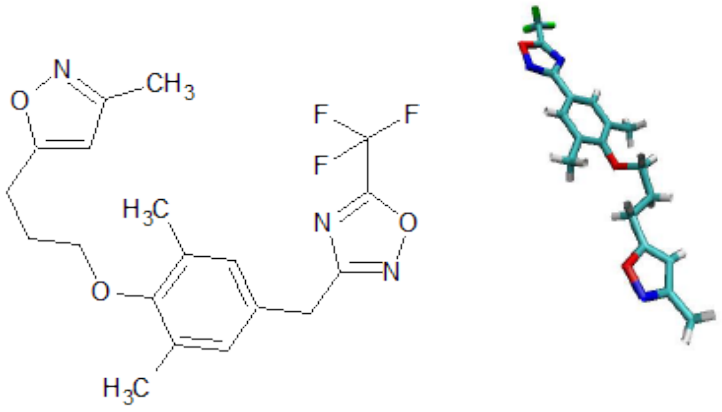
The previous work [14] mainly considered constructing and setting up a virus simulation (EV1) to establish that the model and simulation were in general working but not yet necessarily considering actual physical properties of the system. This model also lacked some potentially crucial parts of the system, that is ligands (pocket factor palmitic acid, covalently attached myristate group) and crystallographic waters present in the crystal structure. These crystallographic waters or some of them are thought represent bound metal ions since they exhibit occupancies larger than one and are located in negatively charged protein cages [8].

In this work the focus is more on discussing further the validity of the model in terms of energy convergence, structural changes etc., based now on significantly longer simulation times of hundreds of nanoseconds.

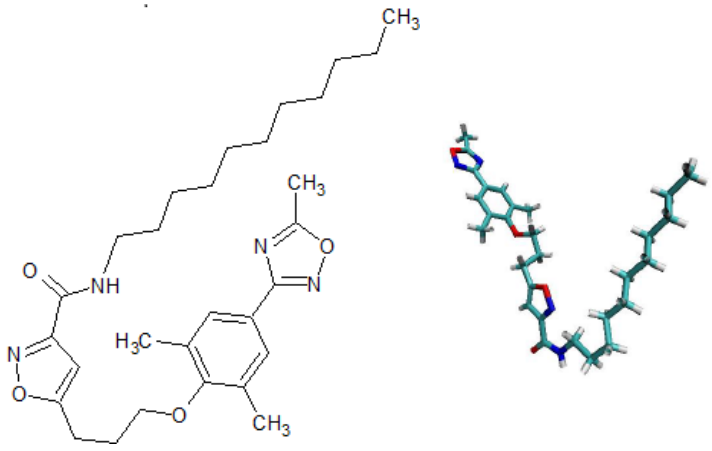
Since the hydrophobic pocket is one especial point of interest in the simulation, it was crucial that also the ligands would be included to the simulation. For comparison also the inclusion of the replacing molecules, Pleconaril and Kirtan1, are of interest. The structures of these pocket factor ligands are presented in figure 2. The structure of the additional ligand attached to VP4, myristic acid or C14 fatty acid is similar to that of palmitic acid (C16 fatty acid) and is not separately presented.



(a) Palmitic acid



(b) Pleconaril



(c) Kirtan1

Figure 2: Structures of different pocket factors.

With these aims in mind the goals for this work are to:
1) Construct systems with different pocket factors (palmitic acid, Kirtan1, Pleconaril) and the covalently attached ligand (myristate group) included and simulate these systems.
2) Extend all simulations to longer time scales in order to assess the usability of these models via observing the equilibration of energy and structure, i.e., the stability of the system.
3) With relation to 2) search for possible connections between simulation and experiment for further model establishment.
4) with relation to 2) address the issue of which properties could be studied with these simulations in more detail (and how) in future, i.e., which are the points of interest and how they are accessible.

The aims related more closely to the actual physical properties of these simulations at this point will be focused on the potential binding sites for gold clusters, that is
5) Behavior and interactions of different pocket factors and the hydrophobic pocket.
6) Dynamics of cysteine residues.

The sort of collective goal of this work is then to assess the usability of this model and molecular dynamics simulations generally in this virus study.

2 Theoretical background

2.1 Molecular dynamics simulations

Classical molecular dynamics (MD) simulation in its simplest form concerns the solving Newton's equation of motion for a set of atoms at each time step. In an all-atom simulation the equation of motion is solved for each atom at every time step. As a result trajectory for each atom is obtained. [15, 16]

2.1.1 MD and statistical mechanics

From statistical mechanical point of view MD corresponds to microcanonical ensemble with energy and number of atoms fixed (NVE). The ensemble can be changed by modifying the equations of motion. This can be performed so that the system is effectively coupled to a thermostat of certain temperature (NVT ensemble) and furthermore to barostat of certain pressure (NPT ensemble). [15, 16]

As an example of thermostat working principle is the velocity rescaled temperature coupling that was used in this work. The idea is that system's temperature T deviation from the temperature of the thermostat T_0 is corrected with time constant τ by: $\frac{dT}{dt} = \frac{T_0 - T}{\tau}$. That is, the strength of the coupling is controlled via τ since the temperature deviation decays exponentially with τ. This coupling is called Berendsen coupling. Berendsen coupling does not however generate correct ensemble, since kinetic energy fluctuations are suppressed. This can be corrected by velocity rescaling, where the kinetic energy is modified with certain stochastic term. [16]

As an example of barostat working principle is Berendsen pressure coupling used in this work. The idea is similar to Berendsen thermostat, that is, the pressure deviation is corrected with: $\frac{dP}{dt} = \frac{P_0 - P}{\tau_p}$ by rescaling box vectors and coordinates with certain matrix. As a result the system pressure averages to the pressure of the barostat, but the fluctuations are large. This is not in fact an exact NPT ensemble. [16]

In addition to trajectory data, then also thermodynamical quantities are obtainable from an equilibrium simulation. This is based on statistical mechanics and Ergodic hypothesis, according to which quantities from equilibrium simulation over sufficient time are related to thermodynamical ensemble averages, such as energy, temperature and pressure. [15, 16]

2.1.2 Description of interactions in MD simulation: Force field

In the actual simulation atoms are treated as point masses with point charges and they interact with each other via different forces. How these interactions are described depends on the mathematical form of the chosen force field, i.e., potential U that appears in the Newton's equation: $F_i = m_i a =$

$-\nabla_i U$. [15, 16]

The general form of a classical force field contains descriptions for bonded and non-bonded interactions. Bonded interactions are presented by harmonic potentials including bond stretching, bending and torsion and they exist between atoms that are bonded to each other. Non-bonded interactions include classical electrostatics or Coulomb interactions and potential describing the short-range repulsion and dispersion between atoms. The latter potential is often described with Lennard-Jones potential. [15, 16]

Each of these potentials contain parameters for different atoms, that are force-field specific. Therefore, it is preferable that the force field is chosen according to the type of the system. That is, if system being studied contains proteins one should use force field that is parameterized for similar systems. [15, 16]

In practice the most problematic interaction is the Coulomb interaction due to its long-ranged nature. That is, its explicit treatment in large system is impossible. Generally this is handled so that only short-ranged part defined by some cut-off distance is explicitly calculated and the long-range part is evaluated in reciprocal space by faster algorithms using Fourier transform. [15, 16]

2.1.3 Treatment of boundaries

In a way related to the treatment of Coulomb interactions is the treatment of system boundaries. Rigid or vacuum boundaries may bring up unphysical surface effects and therefore periodic boundaries are often used. That is, the system box is replicated several times in different directions to effectively fill up the whole space. In handling the periodicity the size of the simulation box and cut-offs need to be selected so that the system is not able to interact with itself. Periodicity may still cause some unphysical effects in the presence of solvent, but in a large system these effects should be negligible. [15, 16]

2.1.4 Classical MD and inclusion of quantum mechanics

Since the force field is phenomenological, the quantum effects are only effectively included in the formulation and parameters of the force field. This results in the fact that classical MD simulations are ground state processes, so that chemical reactions (bond breaking or formation) and processes involving different electronic states cannot be described by these force fields. [16]

In addition to force field specific potentials often some constraint potentials are added in the simulation. If the bonds were treated as classical harmonic oscillators they would exhibit high frequency motions which in turn means that for describing these motions short enough time step needs to be used in simulation. The purpose of the constraints is to remove these highest frequency motions by using a specific constraint algorithm. When the highest frequencies are removed, i.e., the bond is constrained the bond's behavior resembles more ground state quantum harmonic oscillator, which describes better the quantum mechanical nature of bond vibrations than the classical oscillator. So also in this respect the quantum effects of bond vibrations can better be included in the classical simulation. In practice, by using constraints the time step can be increased. An example of a constraint algorithm is the one used in this work, LINCS. With LINCS constraining occurs in two steps: 1) unconstrained update and 2) bond length correction. [16]

2.1.5 Integration of Newton's equation of motion

The equations of motion need to be integrated numerous times during a simulation. For this task certain integrator algorithm is used. These include for example leap frog and velocity Verlet integrators. Leap frog integrator as an example uses position r and force F at time t ,velocities at time $\frac{1}{2}t - \Delta t$, (where Δt is the timestep) in two relations to update position and velocity at each time step. Temperature and pressure coupling and constraints can be implemented via modifying or extending the equations of motion. [16]

The timestep needs to be selected so that it properly handles the dynamics of the system and does not bring up unphysical instabilities. That is, if

the time step is too large the system breaks down. On the other hand, in order the obtained trajectory data set to be statistically meaningful, the simulation must be able to proceed for a sufficient amount of time. In practice the timestep is selected between 1-5 fs. [15, 16]

2.1.6 General MD algorithm

Simplified flow of any MD algorithm follows the general route:
1) Input: initial conditions. System coordinates, velocities and description of interactions, i.e., the potential function are given as an input.
2) Force computation. Force F_i on every atom i is computed from the potential function U describing all the interactions and possible constraints with $F_i = -\frac{\partial V}{\partial r_i}$. Non-bonded forces are calculated for atom pairs and bonded interactions for a group of atoms depending on the potential form. In addition potential energy, kinetic energy and pressure tensor are calculated.
3) Update. Position and velocity for each atom is updated by calculating from the Newton's equation of motion using the selected integrator.
4) Output: current conditions. Current positions and velocities obtained from step 3) along with energy terms can be written to an output (trajectory). steps 2)-4) are repeated in the simulation for numerous steps to obtain sufficiently long trajectory. [16]

2.2 Water model

With proteins the simulations are performed in solvent, often water. Water can be treated implicitly as dielectric continuum or explicitly as molecules. In this work explicit water model was used. [17]

Different explicit water models exist, and they differ in terms of number of interaction sites and geometry. Thus they also produce different thermodynamical quantities. Often the geometry of water molecules is presented as rigid asymmetric rotor with six degrees of freedom; that is three centers of mass and three angles. To correctly reproduce all experimental properties of water at different temperatures and pressures still remains a challenge. [17, 18]

For this work one of the most often used water models, TIP3P [19] model was chosen. It contains three interaction sites with different partial charges interacting via Coulomb potential, and one Lennard-Jones 6-12 term describing interactions between two oxygen atoms. [18, 19]

This potential form contains parameters, that have been revised in the course of time. These improvements are made so that the model would agree as well as possible with experimental results obtained for water. Such properties for example include thermodynamics, diffusion coefficients, radial distribution functions, and hydrogen bond and energy distribution. In general the agreement of these properties are tested at 25 Celsius and 1 atm pressure. [17, 19]

2.3 Convergence and equilibrium

The error in molecular dynamics simulation comes from two sources: from the inaccurate model (force field inaccuracies) and inadequate statistical sampling. For the sampling to be sufficient, the system should converge, i.e., reach equilibrium state in the simulation. If the simulation is insufficiently sampled, the analysis is also meaningless. [20]

Said another way it means that if the system is statistically fully sampled it should be able to visit all the possible spatial configurations. The sampling needs to be sufficient in all these configurations. However there is in principle no way of knowing for sure whether some region in configuration space is being missed in a simulation. All this constitutes to the lack of ergodicity. [20]

The main question related to the statistical meaningfulness of the simulation results could then be summarized as: When is the system sufficiently equilibrated? The so called global sampling quality can be estimated by several different methods, both qualitative and quantitative. However there is no in a way universal or standardized method to decide whether this equilibrium is being reached. Several widely used methods for assessing the reaching of equilibrium include, convergence of root mean square deviations (RMSD), assessing of hydrogen bonds, intramolecular interaction energies, torsion and angle transitions and different cluster counting methods and principal components analysis. [20, 21]

Based on qualitative methods one cannot conclude that the simulation has equilibrated, but one can conclude if the simulation has not yet equilibrated. One general way of assessing this is based on root mean square deviations. If the longer time scale fluctuations do not reach system specific plateau value in the simulation, the system has not converged. RMSD does not contain direct information on which states have been sampled. [20]

Another qualitative method is cluster analysis. This analysis is based on pairwise distances within the structure; the system is divided to smaller structures or clusters based on some distance criteria. In short it could be assessed that if the rate of finding new clusters in the simulation becomes small, the simulation has converged at least to some extent. However cluster counting does not provide absolute determination of convergence either. [20]

Another popular method is the principal components analysis. The simplified idea is that from the vast amount of structural data larger scale characteristic motions or normal modes could be extracted and used as an alternative basis set. The fluctuations of the system can be projected onto this new basis set to reveal number of substates and transitions between them. The result for this would be that in a converged simulation there is a large number of transitions between these substates. [20]

Since the simulation in general starts by setting up a model system based on static X-ray crystallography structure of the protein, certain part in the beginning of a simulation always needs to be considered adjusting. That is, during the first few hundreds of picoseconds unphysical interactions caused by simulation setup are being removed. Further on, the spatial arrangement of the protein changes until some stable arrangement is reached and maintained. This is in general called convergence or equilibration of the system. [21,22]

Especially when the system is large and complex, such as a full virus, it is questionable whether equilibrium can be reached in a limited simulation time of hundreds of nanoseconds. And, if and when full equilibrium is not reached, it becomes questionable whether the statistical results obtained based on such simulation are quantitatively trustworthy and which qualities can be analyzed from such data. [20]

One problem in sampling of large complex biomolecular systems comes

from the fact that there are processes that occur at very different time scales. These processes are not necessarily independent of each other, but the fast-scale motions may be coupled to slower motions. This coupling is not necessarily seen in shorter trajectory. This then also complicates the sampling of individual parts of the system and assessing the convergence of single observables. In practice this means that only assessing longer trajectories may reveal insufficient sampling in shorter trajectories. Therefore, analyzing several time scales in simulation would be useful. [20]

Study of conformational entropies for different proteins show no convergence in simulations even up to 1 ms of simulation time. This means that in practice one cannot reach full equilibrium in a simulation, in the strict sense (that is, system properties such as entropy are not yet time-independent). In order for the entropy to converge, the protein would namely in practice have to be able to unfold during a simulation. This would take from milliseconds to hours, time scale certainly not accessible for an MD simulation. [22]

Full equilibrium in this strictest sense is not however always necessary in analyzing such a simulation, and throughout time molecular dynamics simulations have proved to be useful in studying proteins. This is because different properties exhibit different sensitivity to this lack of full equilibrium, that is, some properties become time-independent (equilibrated) faster than others. Therefore, excluding the initial equilibration, all simulation time is equally important and should be carefully used in analysis. [22]

2.4 Gromacs

Gromacs or GROningen MAchine for Chemical Simulations [23] is a free software molecular dynamics package designed especially to be used in simulating large biomolecular systems. It provides high performance with with efficient algorithmic optimizations and parallelizability. Gromacs supports several different force fields, water models and algorithms most often used in the field of molecular dynamics. Gromacs contains several tools for setting-up, minimizing, running and analyzing a molecular dynamics simulation. [16]

2.4.1 Simulating and analysis

There are some general procedures that take place in setting up, running and analyzing simulation performed with Gromacs. Here is presented a certain generalized workflow of a protein simulation, also utilized in this work. [16, 24]

Model system construction
The model set up generally in case of proteins starts from obtaining the crystal structure of the protein, often from Protein Databank. This structure basically contains coordinates for all the protein atoms, excluding hydrogens. If the structure is missing whole residues from crucial positions, these need to be modeled in manually or using tools designed for this. The structure may also contain molecules that are not considered as natural amino acids, and parameters for these are not necessarily available in the force field that is being used. Then one needs to consider the parameterization of these molecules, using procedures that are recommended for the specific force field that is being used. Parameters of different force fields should not be mixed, since the parameterization procedure is strongly force field specific, and the philosophy behind the parameter derivation may be completely different. [16, 24]

The starting structure for simulation needs to include hydrogens, that can be added with several different tools and programs. The adding of hydrogens in case of large protein structure is not necessarily unambiguous, since protonation states of some residues can vary greatly depending on the environment. Often it is useful to consider at least the protonation states of histidines in more detail, and not only trust the default values given by some program. [16, 24]

Running in Gromacs
In order for Gromacs to actually run the simulation, in addition to the starting structure (coordinates), Gromacs topology file of specific form is required. The topology file will contain information on atom types, charges, bond lengths, angles etc. All these are defined in the framework of specific force field, which is selected before constructing the topology. Also water model is selected at this point. The topology and addition of hydrogens can be performed using Gromacs tool pdb2gmx. [16, 24]

In addition to structure (coordinates) and the topology file, a certain pa-

rameter file specifying all the selections of the actual run (such as integrator, temperature and pressure coupling etc.) is required. These three files are preprocessed with Gromacs program grompp to produce a run input file, that is used as input for the simulation program mdrun. [16,24]

Energy minimization
The first actual simulation step is energy minimization. The first energy minimization is often performed for the obtained structure in vacuum: this should reveal whether there are severe overlaps or strains in the starting structure, that are unable to sufficiently relax during the minimization. The idea of energy minimization is to find a local energy minimum close to the starting structure, for example by moving locally down the steepest potential gradient, until the required force tolerance or number of steps is reached. This method of minimization is called steepest descents. The first energy minimization is then used for assuring that the structure is sufficiently stable. [16,24]

Simulation box
The simulation box is set by placing the protein in a periodic box. It is useful to select shape of the box so that the packing is optimal, that is, the number of solvent molecules that are needed to fill the box is as small as possible, e.g., rhombic dodecahedron box is often used, also in this work. Simultaneously one needs to consider the that the size of the box is sufficient, so that the protein does not interact with its periodic images improperly. [16,24]

Solvent addition and relaxation
The box is then filled with water of previously specified model with Gromacs tool genbox. Often with proteins also ions (,e.g., NaCl) are added to certain concentration so that the charge of the system will be simultaneously neutralized. Ions are added with Gromacs tool genion. The solvated system need to be again energy minimized, since the adding of solvent may generate unfavorable interactions and relaxation is therefore needed. [16,24]

The solvent is often further relaxed by running position restrained simulation. This means that the protein is kept still by defining restraints and only the solvent is allowed to move and therefore adapt further. [16,24]

Unrestrained simulations

After this adaptation, position restraints are removed. The first unrestrained simulation is often the temperature coupling; that is, creating NVT ensemble. This is run long enough so that the temperature converges. After NVT converged, NPT simulation is generated, i.e., pressure coupling is turned on. At least up to this point, simulations are considered purely equilibration. The total length of these simulations may vary according to the system, but for large protein is typically in order of hundreds of picoseconds. [16,24]

Actual production simulations are started after successful pressure coupling. The time scale of a production run is manifold larger than equilibration procedures, often in the order of hundreds of nanoseconds. [16,24]

Quality assurance and analysis
The production run produces trajectory, that is used in analysis. Analysis often contain quality assurance, in addition to finding answers to actual research questions. Quality assurance for example contains convergence checks of energy terms and structural terms. Also it is necessary make sure that an atom has not interacted with itself in the simulation, that is, each atom stays within long enough distance from its own periodic image. By quality assurance one makes sure that simulations have run properly, and that simulation data may be used for analysis of actual physical quantities. Gromacs provides several different analysis tools. [16,24]

2.5 AMBER force fields

As noted previously, error in molecular dynamics simulation comes from the inefficient sampling and inaccuracies in the model, i.e., force field. As the sampling has been enhanced with the increasing computer power and new simulation protocols, the better accuracy of the force field has become more and more important. After all, the physical meaningfulness of a simulation is first and foremost dependent on the force field parameters. The longer and longer simulations of different proteins and other systems have enabled bringing out these deficiencies in force fields, resulting in several improvements. [25]

Force field contains parameters (equation 1) for different atom types, for example different parameters for carbonyl and hydroxyl oxygen. That is,

atom types describe certain chemical environment. How the parameters are derived depends on the force field: Parameters can be fitted based on experimental data and/or quantum mechanical calculations. [18]

AMBER (=Assisted Model Building and Energy Refinement) force fields are designed for simulating proteins and nucleic acids. The original parameters were namely obtained by calibrating against vast amount of protein fragment data obtained both from experimental data and quantum mechanical calculations. The parameters have been improved several times in the course of time using more accurate quantum mechanical calculations on several different model systems. Improvements are made in order for the simulation to better correspond experimental, especially NMR, results. [18,26]

Mathematical form of AMBER corresponds to the one of general classical force field: it describes bonds, angles, dihedrals, Van der Waals and electrostatic interactions with two-body potentials. Parameters therefore include equilibrium bond parameters, partial charges, force constants and torsion parameters. Latest improvements often concern the torsion parameters. The functional form of AMBER force field is presented in equation 1. [27]

$$\begin{aligned} E_{tot} = & \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 \\ & + \sum_{dihedrals} \frac{V_n}{2}(1 + \cos(n\phi - \gamma)) \\ & + \sum_{i<j} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{4\pi\epsilon_0 R_{ij}} \right) \end{aligned} \tag{1}$$

Here r_{eq} and θ_{eq} are equilibrium parameters, K_r, K_θ, V_n are force constants. n is multiplicity and γ a phase angle for dihedral angle parameters. A, B, q are parameters of the nonbonded, Lennard-Jones and Coulomb, potentials. The Lennard-Jones potential can also be presented as $V_{LJ} = 4\epsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right)$, where ϵ is the depth of the potential well and σ is the distance at which the inter-particle potential reaches 0. [28]

The parameters in the first AMBER force fields were obtained from experimental data and gas phase simulations of several protein fragments. Later on as liquid state simulations became available, improvements to Van der Waals interactions, i.e, Lennard-Jones parameters as well as to partial charges could be done. [26]

Still today determining partial charges in large protein system is not unambiguous, since the charge of an atom can greatly vary depending on protein conformation. This is fundamentally due to electronic polarization, that is not explicitly included in AMBER force fields. The polarization is on average included in the sense that fixed partial charges are assigned based on fittings to several different conformations. [26]

The lack of describing the polarization and also the non-local electronic structure, leads to problems also in assigning torsion parameters. This could lead in overly stabilizing some parts of proteins, and therefore comparisons between converged protein simulations and experimental NMR data are compared in order to further improve torsion parameters. Also currently torsions potentials can be optimized to even better correspond to more and more accurate quantum mechanical calculations. [25, 26, 29]

The importance and also the difficulty in parameterization of the torsion parameters comes from the fact that torsion potential in a way accounts for the higher order potential terms that are not actually present in the force field formulation. Especially the main chain torsion potentials then need to be presented with high accuracy, for otherwise the balance between different conformations or secondary structure elements may distort. This is because these dihedral parameters are shared by all amino acids, and therefore their cumulative effect can be drastic. [26, 29]

2.5.1 Parameters

The original parameters of AMBER (developed 1984,1986) were obtained from different quantum mechanical calculations and experiments. Original atomic charges were obtained from ESP fittings with STO-3G level of theory. Van der Waals parameters in turn were obtained from fittings to amide crystal data along with liquid state simulations. [26]

Equilibrium bond and angle parameters and force constants were obtained from crystal structures and optimized to correspond to normal mode frequencies of several peptide fragments. Torsion force constants were obtained by adjusting the model to match energy barriers obtained from quantum mechanical calculations and experiments. In these torsional barriers, nonbonded interactions have an effect; that is, torsion force constants depend on Van der Waals and Coulomb parameters. This constitutes to the

fact that especially torsional force constants are difficult to transfer from one force field to another. [26]

Later as liquid state simulations and larger basis sets became available, also improvements to parameters were done. Now atomic charges could be derived with HF 6-31G* level of theory, that was known to properly overestimate bond dipoles to account for the exaggerated polarity of the current solvent models, such as TIP3P. Especially with implementation of RESP charge model, charge models could be further improved since the charges of buried atoms could be better defined. [26]

In addition to charges, improvements could be done to Van der Waals parameters with reference to liquid state simulations and experiments. Then, whereas other elements had fixed Van Der Waals parameters, different hydrogens were assigned different parameters according to the electronegativity of the atom they are bonded to. [26]

The last parameters to be improved are the torsional parameters, for they are coupled to the nonbonded parameters. Inclusion of torsional parameters is therefore also difficult, and often strictly empirical approaches are used so that the selected parameters would properly reproduce the total barrier profile obtained from experiment or calculation. Errors in torsional parameters have resulted in overly stabilizing some areas of the protein, and several attempts of corrections have been done to these parameters. In finding these errors, qualitative comparisons of simulation to NMR data of conformational populations has been useful. [26]

In the first force fields, torsion parameters were derived based on energy barrier profiles of certain dipeptide rotamers; later tetra peptides were used. Improvements in ff99SB and before concentrated in improving main chain torsion parameters of ff94. ff99SB gave significantly better results in respect to comparisons of experimental (NMR) and PDB survey data. [29]

ff99SB-ILDN (also used in this work) in turn provided improvements in side chain torsion parameters; parameters that had not been revised since the initial introduction. These improvements resulted in considerably better correspondence with NMR data than with previous force fields. The effects of side chain torsions are expected to be especially important in long simulations where buried parts of the protein are able to rotate. This affects the stability of protein folds and loops. [25]

Torsional parameters remain the most difficult ones to parameterize, but there is still room for improvement of these parameters. More general and practical description of parameter derivation in AMBER is presented in the following sections.

2.6 Parameterization of organic molecules for AMBER

As previously noted, the force field contains parameters for certain atom types, often for those of natural amino acids and possibly nucleic acids. Since the philosophy behind the force field parameters differs from force field to force field, parameters should not be mixed. That is, if some molecule is parameterized for some other force field than ,e.g., AMBER those parameters are not necessarily at all meaningful if used in AMBER force field.

2.6.1 Antechamber

Antechamber [30] is a program containing several sub-programs distributed with AmberTools [31,32] that can be used for parameterizing different organic molecules for AMBER. Atom types that are used by Antechamber are from general AMBER force field (GAFF) [28]. GAFF is compatible with other AMBER force fields, but the atom types in GAFF are more general than in protein-oriented AMBER force fields. Ideally, with atom types in GAFF, one could parameterize any organic molecule for AMBER. Therefore it is especially useful in drug design. [28,30]

In addition to atom parameters contained in GAFF, Antechamber uses HF/6-31G* RESP [33] or AM1-BCC methods [34,35] for deriving partial charges.

GAFF

The functional form of GAFF is similar to the AMBER force fields presented in equation 1.

Equilibrium bond r_{eq} and angle θ_{eq} parameters are obtained as mean val-

ues from experimental crystal structure data (X-ray and neutron diffraction) in addition to high level *ab initio* calculations (MP2/6-31G*). [28]

The related force constants, K_r, K_θ are estimated with different empirical rules. These are based on relation with calculated and empirical equilibrium parameters and linear least squares fitting. Bond force constants are optimized to reproduce experimental and *ab initio* vibrational frequencies. The consistency with AMBER force field was tested to be good. It is notable that the bond angle force constants in AMBER (and GAFF) are very simple, and hence the consistency is not as good with angle force constants as with bond force constants. [28]

Unlike bond lengths and angles, torsion parameters are coupled to the nonbonded energy terms and are therefore parameterized last. That is, they basically cover the polarization, charge transfer and many-body effects that cannot be presented with simple functional form. The actual torsion parameters are derived based on energy differences of two different conformations and rotational profiles, obtained experimentally or from *ab initio* calculations. That is, the torsional angle vs. conformation energy is scanned at high level *ab initio* calculation; the force constant V_i is then fitted to reproduce this *ab initio* rotational profile as well as possible. In addition, validation against NMR experiments in comparison to MD simulations is often done. [28]

Van der Waals (Lennard-Jones) parameters are the same as in AMBER force fields. This is because these parameters (internuclear separation at potential minimum, the depth of the potential well) (except for different hydrogens) are element, not atom type, specific. [28]

Charge derivation

Charges in GAFF are obtained with HF/6-31G* RESP [33] charge or AM1-BCC [34,35] derivation. AM1-BCC is a cheaper alternative to RESP method, and therefore often used in handling large amount of molecules. The charges derived with AM1-BCC are compatible with RESP charges, in the sense that the electrostatic potentials should match. [34, 35]

Atomic charge is a crucial concept in all chemistry. In field of molecular mechanics, the physical properties of the system greatly depend on charge. Therefore, atomic charges should be constructed in a way that preserves the quantitative description of the chemistry of the system. [33]

There are several methods ranging from quantum mechanical calculations to empirical models of deriving charges. One method is electrostatic potential (ESP) fit. The idea is that atom-centered point charges are derived using least squares fitting algorithm so that the electrostatic potential of the molecule would be reproduced. The electrostatic potential is calculated at each grid point from the quantum mechanical wavefunction, using big enough (6-31G*) basis set. Use of this basis set results in enlarged dipole moments, but this actually desirable since it in a way accounts implicitly for the polarization effects present in solvent; since in the most used solvent models (SPC, TIP3P) polarity is exaggerated. (Polarization is not explicitly included in AMBER force fields). That is, charges are already pre-polarized. [33,35,36]

ESP fit charges tend to reproduce interaction energies well, and they can be calculated straightforwardly. Weaknesses that appear with ESP fittings include conformation dependency, large charges and that charges are not easily transferable between functional groups in different molecules. These inaccuracies are fundamentally due to the worse determined charges of buried atoms. [33,36]

These factors can however be reduced by introducing restraints in form of penalty function of hyperbolic form to the fitting procedure, resulting in restraint electrostatic potential (RESP) fit. In general this procedure results in reducing the overall magnitude of especially the buried charges. [33]

Another factor taken into account in the RESP procedure is the appearance of symmetrical charges. That is, e.g., the charges on methyl hydrogens should be the same. This can be achieved by two stage approach: in the first part there is no forced charge symmetry applied on any atoms. In the second part refit of charges is done on symmetrical atoms, now forcing the symmetry. Charges of other atoms are constrained to the ones obtained in the first part. If the symmetries were forced already in the first part, it would affect the charges in the regions of polar heteroatoms. Therefore, the two stage approach provides a compromise. [33,36]

In general RESP model reproduces well solvation free energies and intramolecular conformational and important interaction energies. In addition charges exhibit smaller fluctuations between functional groups and charges are consistent with chemical intuition while simultaneously reflecting the effects of the chemical environment. The charges derived are

however still to some extent conformation dependent. [33, 36]

Since RESP charge derivation is computationally relatively heavy, an approach to emulate RESP charge derivation has been developed: AM1-BCC charge model. The idea is that the formal charge and electron delocalization are captured by AM1 population charges. To these charges the so called bond charge corrections (BCCs) are applied to produce AM1-BCC charges. The BCCs were obtained by fitting to the training set (2700 molecules) of RESP charges. With these parameters charges for vast number of organic molecules can be derived. The strength of the method is that no actual *ab initio* calculations need to be performed, but still high-quality atomic charges can be obtained to be used in solution phase simulations of organic molecules. [34, 35]

Workflow

The general workflow of Antechamber is the following; utilizing the tutorial in [37]: As an input coordinates of the molecule is required in form of e.g. pdb-file. In the first part point charges are calculated according to selected charge model (e.g. AM1-BCC). In addition bond types and atom types are assigned. As an output from the first part a file containing atom types, charges, coordinates and bonding information is obtained (mol2-file).

In the second part, parameters based on GAFF parameters need to be properly assigned. Antechamber provides a tool (parmchk) for checking if some of the needed parameters are missing. For the missing parameters some reasonable value may be automatically suggested by Antechamber or in some cases parameters need to be manually put in. In any case the missing parameters and related suggestions need to be assessed. As a result a residue topology file is obtained.

With program tLeap the actual parameters are gathered and can be saved in AMBER format. Conversion to Gromacs topology format can be done for example by using ACPYPE code [38]. Also then the meaningfulness of the results should be checked.

In summary then, Antechamber assigns the given molecule the required force field parameters, presented in equation 1 and a residue topology that contains information on atom connectivity, atom types and point charges.

Though the use of Antechamber in parameterizing a molecule is straightforward and automated, it is not a black-box method and the meaningfulness of the obtained (and possibly missing) parameters need to be assessed.

2.7 Theoretical virus studies

Crystal structures of several non-enveloped, icosahedral viruses are available. Taking also into account their relatively small size, they are ideal candidates for theoretical virus studies. Theoretical studies or molecular dynamics simulations of viruses generally aim to relate structural details of an all-atom simulation of the capsid to some virus functions. [2–4]

Bearing in mind the time-scale limitations of the simulations of such large systems, compromises and certain 'tricks' need to be sometimes applied. These include for example utilizing some experimental data or certain constraints to drive the simulation to a desired direction, that is, performing steered molecular dynamics simulations: these can for example address the mechanical properties of the capsid by forced deformation by pulling or pushing [39–41]. Of course, similar studies considering smaller mechanistic details of other than viruses also exist. Also different types of free energy calculations are used especially in studies of binding affinities of drug molecules and other ligands to different types of biomolecules, e.g., [42–50].

Interesting virus properties that have been addressed with simulations include for example (among several others) studies of capsid stabilities in different conditions [51–53], capsid self-assembly [54], capsid uncoating and genome release [55], effects of antiviral compounds or other molecules [56], effects of mutations [57], conformational transitions at different stages [58], etc. The numerous virus studies incorporating molecular dynamics simulations alongside experimental data show that in spite of the limitations of simulations, it is a valid tool in studying several interesting virus properties in detail that are not always available for experiments. Emphasis is however on the inclusion of experimental data, without which meaningfulness of the simulations are questionable.

2.8 Free energy calculations

Free energy fundamentally describes the system and is a crucial concept in understanding molecular processes. For biomolecules free energy calculations can provide valuable information related to, e.g, folding and unfolding of proteins, ligand binding, elastic properties, conformation transitions, and transport through protein channels. The free energy calculation however in general requires extensive sampling of the configuration space, hence the calculations are expensive for large systems. [59,60]

There exist also other than MD methods, such as non-dynamical and *ab initio* methods, that can be applied in free energy calculations. However with MD methods also the specifics of interactions can be explained. The strength of MD method lies in the fact that they take into account the dynamical behavior of the system instead of the mere structure, but are also computationally feasible. However MD methods always incorporate approximations to different extents, causing drawbacks in these methods. [61,62]

In the field of MD methods, for calculating free energies several techniques for equilibrium and non-equilibrium processes exist. Often the so called path-based free energy methods are applied and these have become relatively straightforward methods for studying small (<50 atoms), rigid ligands. In studies of protein-ligand interactions non-equilibrium fast growth methods based on Jarzynski equality potentially provide more efficient methods for calculating free energies compared to traditional equilibrium methods, such as umbrella sampling. Both types of methods utilize steered molecular dynamics simulations in sampling and free energies are studied via calculation of potential of mean force for, e.g., pulling the ligand out from its binding site. The PMF curve then describes the binding affinity between the protein and ligand. PMF is the potential obtained from the mean force of several different system configurations. [16,59,61,63]

Steered molecular dynamics (SMD) simulation in general enables studying conformational changes and unbinding of ligands in time scales of molecular dynamics simulations. The idea in short is that external forces are applied to the system and the response is analyzed. SMD simulations have several applications in ligand binding pathway studies and in studies of elastic properties of proteins. They can shed light on also binding

mechanisms in addition to mere numbers. SMD simulations work alongside experiments, suggesting possible mechanisms and interpretations related to experiments and also possibly suggest new experiments. [64]

In general ligand receptor interactions govern several biological processes. Therefore studying these interactions provides means to, for example, designing and testing interactions and effects of drug molecules in the system. [64]

The SMD simulation for protein-ligand complex then basically proceeds by unbinding the ligand from the protein by applying external forces. The external forces can for example restrain the ligand to a specific point and this restraint point can be shifted during the simulation so that the ligand eventually unbinds. That is, the ligand is pulled away from its binding site and the force experienced by the ligand then depends on the rate of the pulling and the force constant of the pulling force. Attention needs to be paid on selecting the path, i.e., direction of the force. The forces and ligand positions are then followed during the simulation to obtain information on the structure-function relationships of the ligand-protein complex. The mere forces however only give qualitative information, since these describe only the forces required to unbind the ligand and are not directly related to the actual binding affinity of the ligand. These rupture forces can however be utilized in comparing relative binding affinities of different ligands using simplified approximations of $\Delta G_{bind} \approx -F_{max} x_{eff}$. For quantitative information on binding affinity, free energies need to be calculated, that is, the free energy profile of the ligand along certain reaction coordinate needs to be obtained. This is a quantity called potential of mean force (PMF). [62, 64]

As mentioned earlier, different techniques for calculating the PMF curve exist; two most used methods in field of protein-ligand interaction studies include umbrella sampling (equilibrium technique) and Jarzynski equality or fast growth method (non-equilibrium technique). The difference is that equilibrium techniques require reversibility, in practice, simulations are slow and expensive. [63, 64]

The idea of an umbrella sampling simulation is that a number of configurations (called umbrella windows) are generated along a reaction coordinate so that the ligand's center of mass distance to the reference group (protein) is increased in each configuration. The ligand's position in the

sampling window is maintained by a biasing umbrella potential. In the simulations each of these windows is then equilibrated so that eventually numerous configurations along the reaction coordinate are obtained. The potentials obtained from these windows are unbiased and as a result PMF can be calculated for each window. PMF along the whole reaction coordinate is obtained from PMFs of adjacent windows. Often together with umbrella sampling the so called weighted histogram analysis is applied in handling the results to obtain continuous PMF curve. In general umbrella sampling method is computationally heavy: since it is an equilibrium approach, the system needs to be equilibrated in each sampling window before moving to the next one. [57, 63]

On the other hand, non-equilibrium method, Jarzynski equality or fast growth method does not require any other equilibration but one in the initial state. Utilizing Jarzynski equality in free energy calculation involves several steered molecular dynamics (SMD) or pulling simulations started from different initial, equilibrated configurations. From these simulations work functions along reaction coordinate are obtained, and the results are analyzed with Boltzmann averaging, that is a much simpler method than weighted histogram analysis. With complex biomolecular systems however Jarzynski method requires up to order of magnitude more data than umbrella sampling to produce reliable PMF curve. [63]

Basically, Jarzynski equality relates the work done in a nonequilibrium process to the change in equilibrium free energies. The Jarzynski relation is independent of the speed or reversibility of the process. Said another way, Jarzynski equality relates the ensemble average of an exponential of the total work performed on the system during non-equilibrium transition from one state to another to the free energy difference between these two states. [60, 63]

Jarzynski equality is derived from Clausius inequality. If a control parameter coupled to the system's Hamiltonian is switched, work is performed on the system. The average of this work is larger than or equal to the free energy difference between equilibrium states of the system corresponding to the initial and final value of the control parameter. The equality holds if the switching is done reversibly. This is called Clausius inequality, which can be turned to Jarzynski equality: $\left\langle e^{-\beta W} \right\rangle = e^{-\beta \Delta F}$ which does not require reversible switching. [65]

Then in practice, using Jarzynski equality the free energy difference can be obtained as $\Delta G_{AB} = -k_B T \log \left\langle e^{-\frac{W_{AB}}{k_B T}} \right\rangle$, where W_{AB} is the work performed for forcing the system along one path from initial state A to final state B. The part inside brackets means averaging over several trajectories with different initial states A. [16]

In Gromacs potential of mean force can be calculated in several different ways. One option is to use pull-code, that is, a steered molecular dynamics simulation. The pull code applies proper force constraints according to the selected pulling potential between two groups, one of which is pulled along a specified direction with specified velocity. The pull-code can be utilized with both umbrella sampling or Jarzynski equality techniques. Gromacs also provides a ready tool for performing weighted histogram analysis for umbrella sampling simulations. [16]

3 Computational methods

3.1 Overview of the previous work

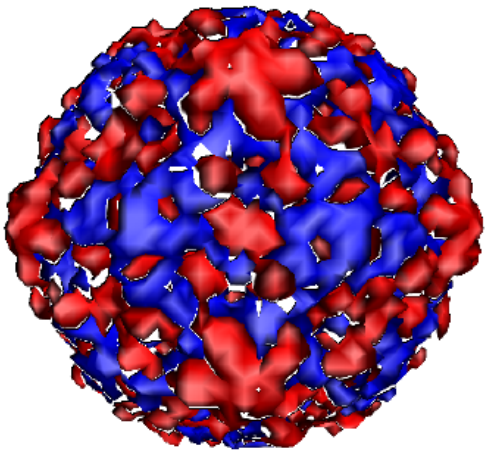
3.1.1 System description and construction

The capsid was constructed from the Protein databank 1EV1 protomer structure for Echovirus 1 [8], that contained proteins VP1-VP4 along with two different fatty acid ligands and crystal waters. The structure does not contain genome structure. Total of 14 residues were missing; first seven residues of VP2 and residues 16-22 of VP4. Since the residues missing from VP4 cause the chain to be broken, these known residues (mentioned in the crystal structure) were modeled in using Modeller 9.11 loop builder and also manual aligning to obtain stable structure. The ligands and crystal waters in the crystal structure were omitted from the full structure at this point. The full structure was constructed with VMD [66] mono2poly script [67], based on icosahedral capsid symmetry (PDB BIOMT records). Gromacs (4.5) [23] simulation package was used in all the simulations.

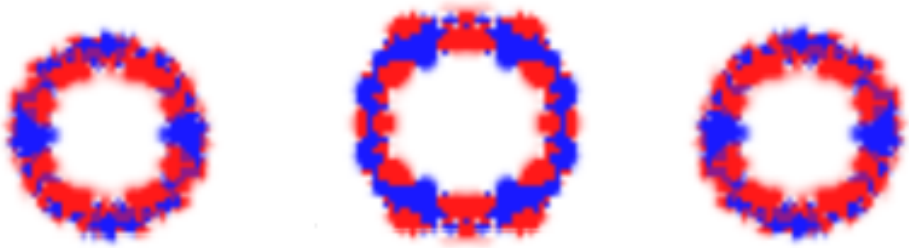
In constructing Gromacs topology selections for force field (AMBER99SB-ILDN [25]) and water model (TIP3P [19]) were done. Protonations for all

residues except for histidines were kept as default values (pH 7). Histidines were protonated based on visualizing the possibilities of hydrogen bond formations in one protomer; based on this histidines with indices 244, 109, 118, 145 and 216 were doubly protonated, i.e., positively charged. With these selections the total charge of the capsid was -420 e.

The large negative charge at this point raised suspicions on the stability of the capsid, since also the genome would be negative. For better visualization of the distribution of the potential on the capsid, APBS calculation [68] was performed utilizing VMD. An APBS calculation provides a way to evaluate the electrostatic properties of a biomolecule based on solving the Poisson-Boltzmann equation that is a relation between the electrostatic potential and dielectric properties of the solvent and solute, partial charges of the solute atoms, ionic strength of the solution and accessibility of the ions to solute interior [68]. This calculation revealed that there are areas of positive potential extending towards the interior of the capsid, i.e., towards the genome (figure 3). This suggests that in spite of the negative overall charge the capsid could still be stable since it contains areas of positive potential that could interact with the genome.



(a) Potential surface



(b) Volume slices of the capsid interior

Figure 3: (a) Potential surface of the Echovirus 1 capsid from APBS calculation. (b) Volume slices of the potential surface showing areas of positive potential extending towards the interior of the capsid. Red and blue colors correspond to negative and positive potentials, respectively.

The capsid was put to a dodecahedric periodic box that was filled with water. Ions (NaCl) were added to neutralize the system with concentration of 0.15 M. The system contained then approximately 3.2 million atoms.

3.1.2 Simulations

The setting up (energy minimization, temperature and pressure coupling and equilibration) and the first short production run of the first version of the virus system were performed on Jyväskylä University's Electra cluster (performance 0.5 ns/day with 64 cores). In these first simulations [14] it was established that in terms of simulation stability there were no severe problems with the system set up. That is, at least in the very short simulation (10 ns) the system started to show signs of at least partial convergence and the structure retained its shape. It was concluded that there were no severe unphysical instabilities in the model; this suggested that the model was physically meaningful enough to be used in further simulations.

3.2 Gromacs on HERMIT

Having sufficiently established the meaningfulness of the used model system, it was transferred to HLRS HERMIT machine. Simulating was tested with several different settings to find the optimal setup for running this particular simulation. With the available resources the optimal performance was found to be approximately 19 ns/day with 4096 cores.

In comparison to the previous work some of the running settings were in addition changed for better performance based on testing. These included changing to newer Gromacs [23] version (4.6.4), changing cut-off scheme to Verlet and constraining hydrogen bonds instead of all bonds. In addition dynamic load balancing was set to 0.7.

As the simulation was run further the performance occasionally decreased significantly below 19 ns/day with 4096 cores. Therefore to optimize both the performance and time in queue the runs were eventually performed with 2048 cores, then reaching performance of approximately 11 ns/day.

3.3 Ligand additions

The original crystal structure of EV1 [8] contained ligands and crystal waters, which were omitted from the first structure. Since these ligands are of

interest and possibly crucial to the structure, also system containing these ligands and also crystal waters was now constructed.

Parameterization in general requires first of all definition of an atom type for each atom in the molecule to be parameterized. Information for each atom type is presented in the force field: this information contains the atomic number and mass and Van der Waals, i.e., Lennard-Jones parameters for the atom. In AMBER the Lennard-Jones parameters are σ and ϵ, which correspond to depth of the potential well and distance for which inter-particle potential reaches 0, respectively. Bonded parameters, i.e., equilibrium bond lengths, angles, dihedrals, functions describing them and related force constants are defined between atom types present in the force field. The actual amino acid or molecule is described in a separate file, and it contains the list of atom names and atom type and partial charge for each of these atoms. Also atoms that are bonded to each other in the residue and improper dihedrals are defined here.

The ligands present in the crystal structure were palmitic acid (pocket factor; C16 fatty acid, figure 2) located in the hydrophobic pocket formed by VP1 and myristate group (myristic acid is a C14 fatty acid) covalently attached to the VP4 amino head. The force field does not readily contain parameters for these molecules. Since both ligands are however standard fatty acids, atom types and parameters for both were directly taken from the force field parameters utilizing those of a similar residue, glutamic acid. Partial charges were derived using RESP procedure used with AMBER force fields. Charges were derived assuming that palmitic acid is negatively charged, i.e., deprotonated inside the hydrophobic pocket. This assumption was merely done since in this simulation similar residues, glutamic and aspartic acid, were also by default deprotonated and negatively charged. Since myristic acid is covalently attached to the VP4 protein, i.e., exists as a neutral myristate group in the structure, also the charge derivation for the myristate group was done taking into account this covalent bonding to glycine residue of VP4 N-terminal. That is, the charge derivation was done for a structure where myristate group is attached to glycine residue.

For the charge derivation procedure the structures were first optimized and ESP calculation was performed with Gaussian [69] at HF/6-31F* level of theory. The ESP output was converted to proper format with Antechamber [30] so that input files for AmberTools [31] Respgen program in order

to perform two stage RESP [33] fitting were obtained. The fitting for both structures was done using Respgen program within Antechamber. For palmitic acid automated setups were used so that in the first part charges of all atoms were allowed to vary however requiring that the two carboxyl oxygens have equal charges. In the second part, all other charges were allowed to vary but degenerate hydrogens were restricted to have equal charges. For myristate group attached to glycine residue in the first part the charges of the glycine residue were constrained to the values readily available in the force field and the charges on the myristate group were allowed to vary. In the second part glycine charges were again constrained and charges on myristate part were allowed to vary but degenerate hydrogens were restricted to have equal charges. The obtained charges for both structures were compared to those of similar residue parts readily available in the force field and no drastic discrepancies were observed. In addition proper bonded parameters were modified to enable covalent attaching of myristate group to the VP4 N-terminal.

To establish the validity of the used parameters, a small model system containing several palmitic acid molecules in a water box was constructed. In a short simulation these molecules formed aggregates. Although simulation was only run very short, and no micelle formation was yet observed, the parameters should be meaningful enough.

Full icosahedral virus capsid was constructed from 1/60 protomer using VMD mono2poly algorithm (similarly to the previous system), containing now a total of 60 palmitic acid molecules, one in each hydrophobic pocket and total of 60 myristate groups each attached to different VP4.

In addition also crystal waters were now added to the system. Since crystal waters are thought to model bound metal ions [8], their presence in the simulation could be important. However since in [8] it was not indicated which ions these represent, these were modeled in as water molecules. However in constructing the full capsid from 1/60 protomers there were severe overlaps of certain crystal waters at very constricted spaces that caused problems in the energy minimizations. The crystal waters that formed problems were cramped at specific sites in groups of five; therefore the full system was constructed in a way that these sites only contained one water molecule instead of five. However in a long enough simulation these sites should become reoccupied if necessary, thus leaving some of the crystal waters out of the structure should not create drastic problems.

The rest of the crystal waters were left where they were after construction of the full capsid.

3.3.1 Replacement of the pocket factor

Replacing of the natural pocket factor (palmitic acid), i.e., docking of different molecules Pleconaril (known drug) and Kirtan1 was done elsewhere [70].

From the obtained protomer structures (1/60) with these new pocket factors docked, two new full systems were constructed (containing still myristate group attached to VP4 and crystal waters in place as described before). Now two comparative systems were obtained with one containing a total of 60 Pleconaril molecules, one in each hydrophobic pocket and the other correspondingly one Kirtan1 molecule in each hydrophobic pocket.

Since Pleconaril and Kirtan1 were more complicated structures (figure 2), yet still containing standard organic groups, the force field parameters were derived using a recommended procedure with AmberTools [31] package Antechamber tools [30]: Then the parameters come from GAFF and for charge derivation the AM1-BCC method was used (compatible with RESP charges). The Gromacs topologies were obtained using ACPYPE code (emulating amb2gmx) [38]. Antechamber uses parameters from GAFF [28]; these parameters and atom types are therefore not as such present in the used AMBER99SB-ILDN force field. Therefore the force field libraries needed were modified to contain the missing atom types and their parameters; these atoms were named N4-N6; C5-C9, H7-H9 and O3-O4. This procedure is potentially questionable, since in strict sense GAFF and AMBER99SB-ILDN are different force fields and parameters are not necessarily mixable. In comparing these derived bonded and non bonded parameters to those present in AMBER99SB-ILDN no discrepancies were observed and due to compatibility of these force fields this parameterization should be usable.

To be noted is that dihedral angles, propers and impropers (impropers keep planar groups planar), are described with different functions in Gromacs within AMBER99SB-ILDN force field Gromacs parameters and in topology produced by ACPYPE from Antechamber derived parameters. In AMBER99SB-ILDN the proper dihedrals are presented with Gromacs

function 9 and impropers with function 4 (periodic improper dihedral type). In ACPYPE topology the functions are in turn 1 and 3 (Rycakert-Bellemans dihedrals), respectively. Due to difference in functions then also parameters differ from each other. Basic bond and angle functions are the same in both (1) and thus also parameters are comparable and noted to be similar.

For these molecules no separate test runs were performed to test the parameters. These may be necessary in the future to establish the meaningfulness of the parameters.

3.4 Simulations: equilibration, product run, analysis

The full structures of viruses containing different sets of ligands were constructed from 1/60 protomers using VMD mono2poly algorithm that utilized symmetry operations described in the crystal structure. In addition to original crystal structure the protomer used in the construction contained missing loop from VP4, modeled to the structure in previous work (using Modeller 9.11 loop builder). The overlapping crystal waters were removed properly from the full structures.

Each system was (similar to original system) put to periodic, dodecahedral box with distance between solute and the box set to 1.2 nm. Water was then added to the box, followed by addition of ions (NaCl) at 0.15 M concentration. Ions were added so that they replaced water molecules, but not the crystal waters.

The whole system was first in short energy minimized to assure the quality of the produced structure. After this, different from treatment of the original system, short simulation with position restraints applied to protein with 200 K temperature coupling was performed for further equilibration. This additional step was performed since the immediate coupling to the 300 K temperature was noted to result in breaking of the system crashing of the simulation. After this, the system was coupled to 300 K temperature, and after that to 1 bar pressure. The production runs were started after pressure coupling. The production runs for all systems were performed for 24 hours at a time (that is, approximately 11 ns) and the simulations were continued from the checkpoint files written by Gromacs, so

that longer continuous trajectories could be obtained.

In the run parameters md integrator (uses leap-frog algorithm) was used with 2 fs timestep. As the cut-off scheme Verlet cutoff was selected with grid-type neighbor searching. Periodic boundaries were applied in all directions. For electrostatic calculations PME(=particle mesh ewald) was used with short-range cut-off of 1 nm and 0.12 grid spacing. For Van der Waals interactions cut-off scheme was used with dispersion corrections for energy and pressure. For temperature coupling v-rescale thermostat with time constant of 0.1 ps and reference temperature of 300 K was used. Protein and non-protein parts were connected to separate baths. For pressure coupling isotropic Berendsen barostat with time constant 1.0 ps and reference pressure of 1 bar were used. No velocities were generated in the run. The simulations were continued from previous simulations. LINCS constraint algorithm was used for constraining hydrogen bonds. Unless otherwise indicated, all the analysis were performed with Gromacs tools. VMD and its plugins were utilized in visualizing the structures and some of the analysis.

4 Results and discussion

The original system containing no ligands was simulated up to 207 ns. This will be called basic simulation in the following sections. The system with natural ligands (palmitic acid and myristate group) was simulated up to 114 ns; system with Pleconaril in the hydrophobic pockets (also myristate groups attached to VP4) up to 45 ns and system with Kirtan1 in the hydrophobic pockets (and myristate groups attached to VP4) up to 57 ns. These systems will be named based on the pocket factor in the following sections. In general the simulation times are short with respect to virus functions. Therefore, it is expected that no drastic structural changes should occur during these simulations. However if the systems were stable and physically meaningful it is expected to see convergence in energetic and structural terms at least to some extent already during these simulation times.

The construction of the capsid starting from the innermost parts is visualized in more detail in figure 4.

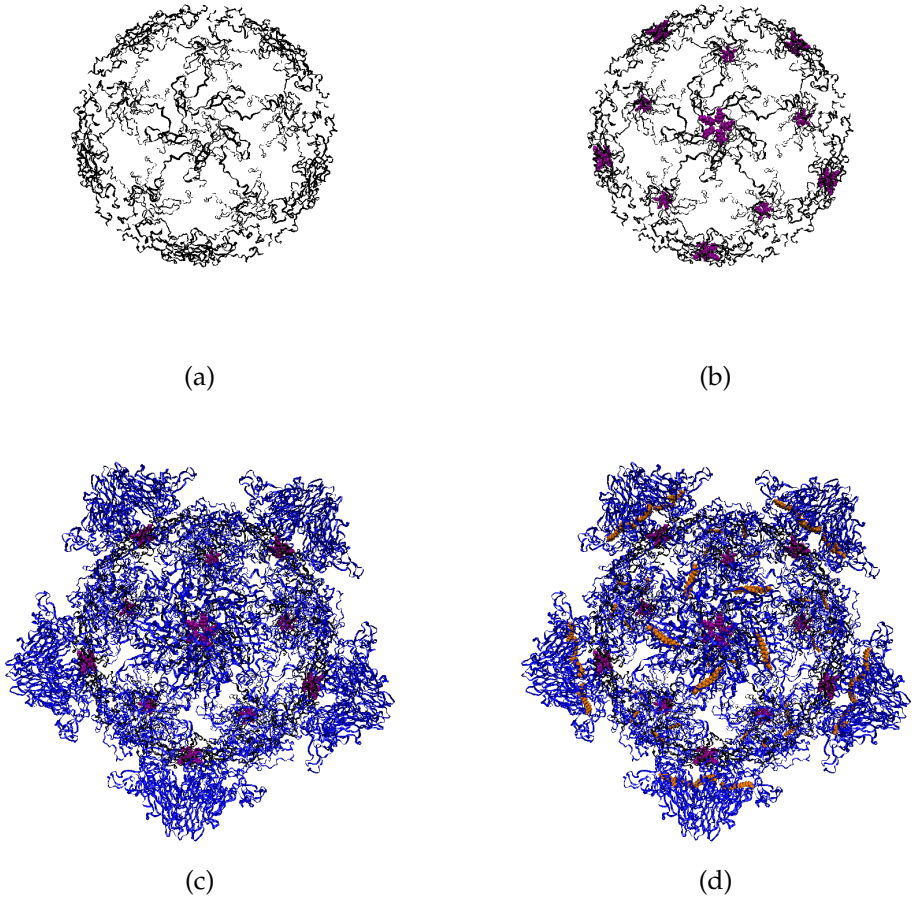
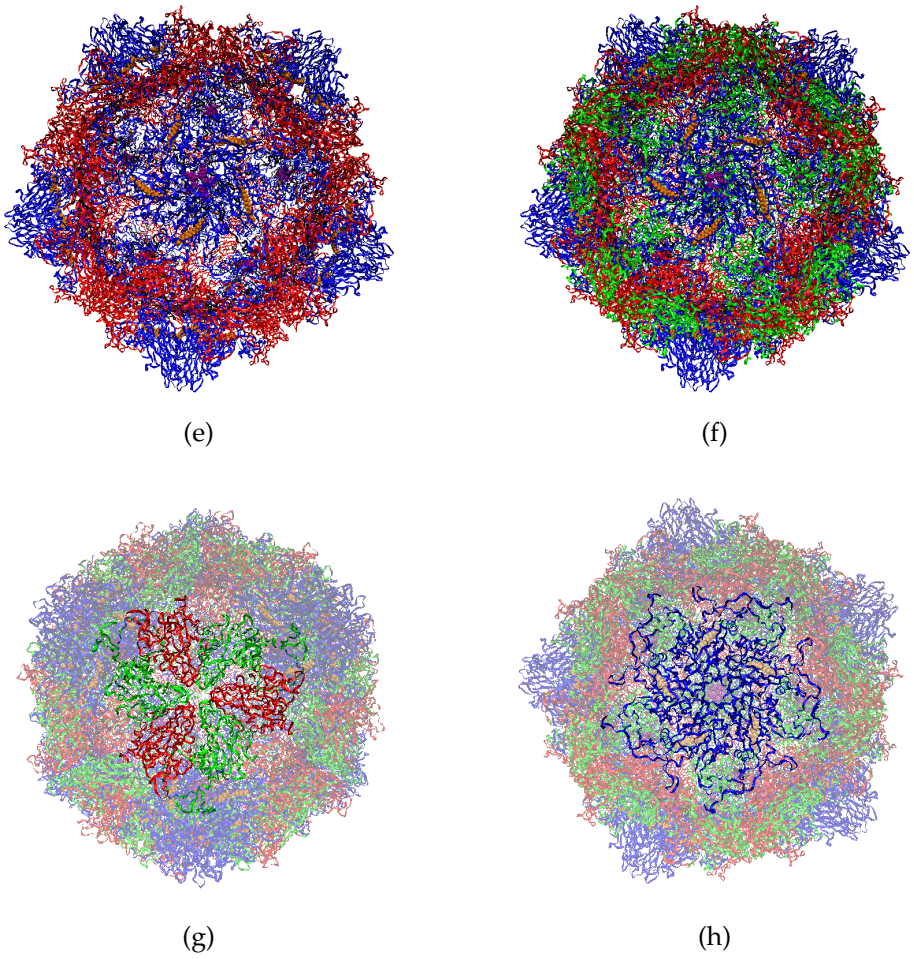


Figure 4: Construction of the capsid. (a) shows the inner protein VP4 (black). (b) shows myristate groups (purple) bonded to N-terminal of VP4. (c) shows VP1 (blue). (d) shows palmitic acids in the hydrophobic pockets of VP1 (orange spheres).

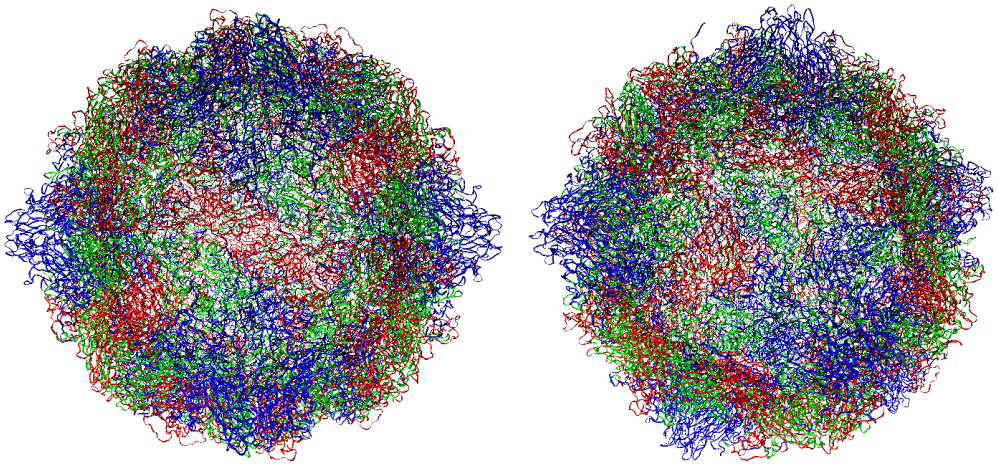


(e) (f)

(g) (h)

Figure 4: Construction of the capsid continued. (e) shows VP2 (red) and (f) VP3 (green). (g) highlights the propeller structure formed by VP2 and VP3 and (h) highlights the pentamer structure formed by VP1.

Figure 5 shows snapshots of the capsid along the basic trajectory. It can be seen that the structure moves in the simulation but no major distortions of the capsid shape can be clearly pointed from the snapshots.

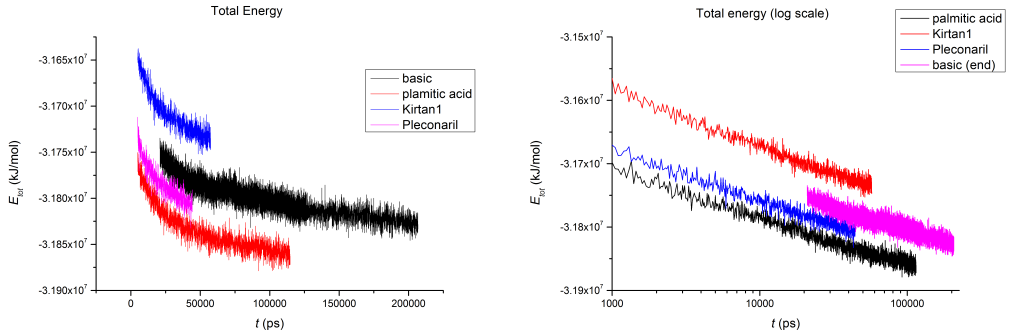


(a) 0 ns (b) 200 ns

Figure 5: Snapshots of the capsid along the trajectory of the basic simulation.

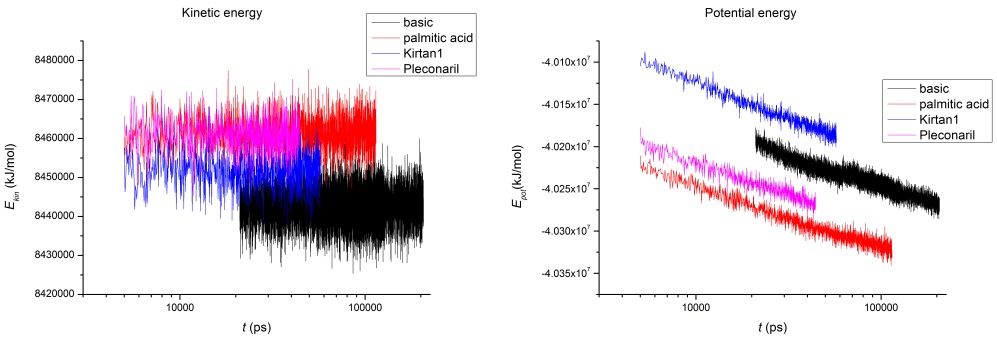
4.1 Model assessment

Evolving of different energetic terms in the simulations are presented in figures 6-9.



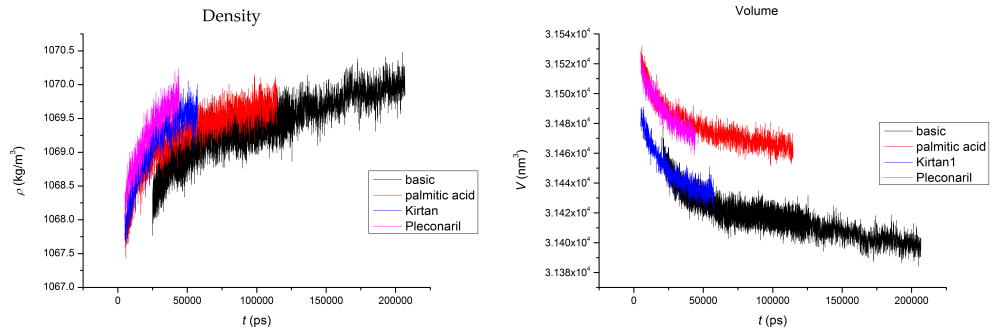
(a) (b)

Figure 6: Total energy (potential and kinetic energies) in different systems on (a) linear and (b) logarithmic scales.



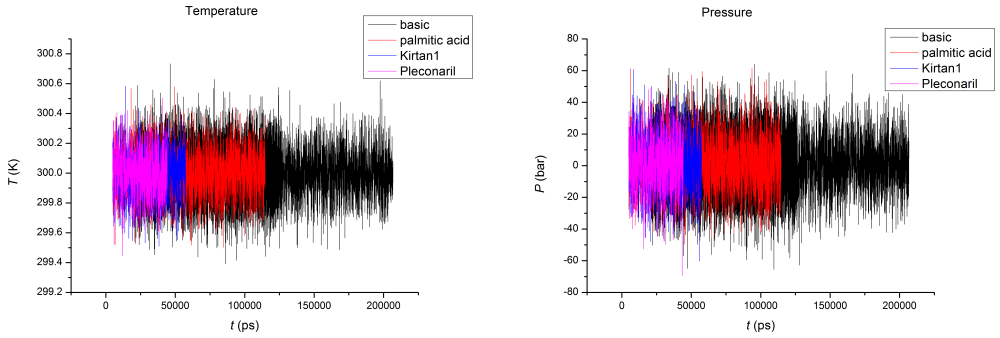
(a) (b)

Figure 7: Total energy (potential and kinetic energies) in different systems on (a) linear and (b) logarithmic scales.



(a) Density (b) Volume

Figure 8: (a) Density and (b) volume in different systems.



(a) Temperature

(b) Pressure

Figure 9: (a) Temperature and (b) pressure in different systems.

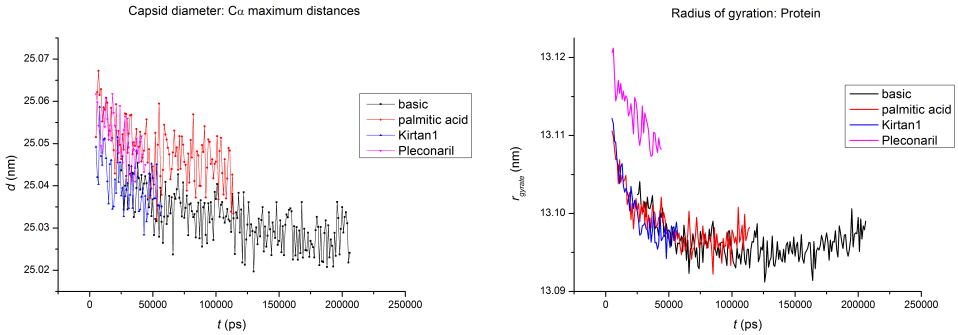
From figure 6(a) it can be seen that trend in the evolving of total energy is in general decreasing and negative for all the systems, showing that in very general level the simulations are stable and there are no severe unphysical large forces or instabilities evolving right from the start of the simulation. That is, each system is evolving towards an energy minimum. For the two longest simulations, basic and palmitic acid systems, the total energy also starts to show signs of convergence whereas the shorter simulations of Pleconaril and Kirtan1 systems still seem to exhibit decreasing trend. The convergence or its missing can be emphasized by plotting the energy on logarithmic scale (figure 6(b)): It can be seen that even the longest simulation still exhibits decreasing trend in total energy. Also it can be seen that the decrease in total energy on logarithmic time scale for all the systems appears almost linear. This could indicate some systematic long-term gradual drift in energy of the system. Whether this behavior proceeds to be linearly decreasing or converging cannot be yet assessed from these simulations. Looking more closely to the kinetic and potential energies separately, 7, it can be seen that on logarithmic scale kinetic energy does not show any drifting behavior, whereas potential energy does.

Evolving of density and volume in figure 8 show increasing and decreasing trends, respectively. The changes in both quantities are relatively small at the end parts of each simulation. However even for the longest simulation still trend of increase/decrease in these quantities can be observed. From figure 9(a) it can be seen that temperature shows only small fluctuations averaging to the set reference value of 300 K. Pressure (figure 9(b))

expectedly considering the properties of Berendsen coupling, shows large fluctuations but averages to the set reference of 1 bar.

The energetic terms, energies, volume and density, start to show signs of convergence towards the end of the longest simulations at least on linear time scale. On logarithmic scale total and potential energies show linearly decreasing trend. However the relative changes in all energy terms throughout the simulations are small. Hence for purposes of this work these simulations are sufficiently converged in terms of energetic quantities. The indications of differences in actual quantities between different systems can not be interpreted at this point.

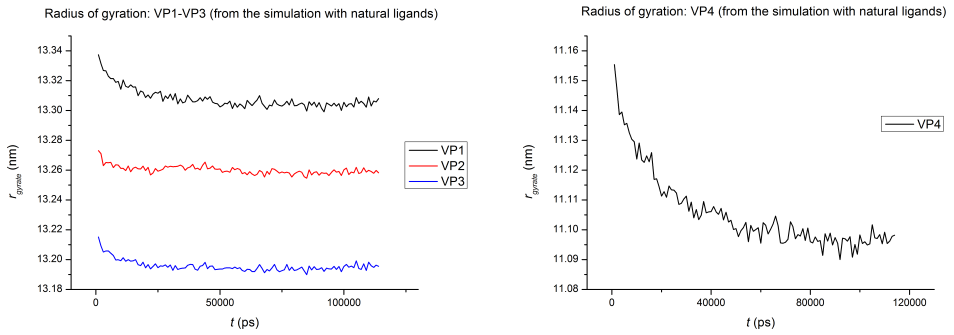
The convergence of structural terms is studied via evolving of capsid diameter, radius and root mean square deviations and root mean square fluctuations, figures 10-12. The diameter is defined at the calculated time points as the maximum distance of two C-alpha atoms. The radius is calculated as the radius of gyration of the protein and different proteins chains, VP1-VP4, with Gromacs tool g_gyrate. Time evolving of root mean square deviations for different system parts are presented for the basic simulation. These are calculated with Gromacs tool g_rms and the least squares fit is done on the initial structure, i.e., 0 ns structure. Time averaged, residue based root mean square fluctuations were calculated from the basic simulation of 207 ns for different protein chains VP1-VP4. Again least squares fit was done on the initial structure from the start, i.e., 0 ns of the simulation.



(a) Diameter

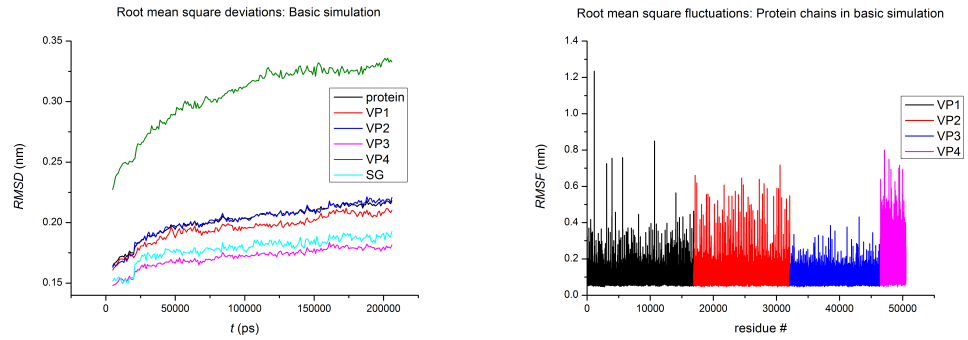
(b) Radius

Figure 10: (a) Capsid diameter and (b) Capsid radius (of gyration) in different systems.



(a) (b)

Figure 11: Radius of gyration of different protein chains (a) VP1-VP3 and (b) VP4 in the simulation with natural ligands.



(a) RMSD (b) RMSF

Figure 12: (a) Root mean square deviations of different system parts and (b) root mean square fluctuations of different protein chains in the basic simulation. SG denotes the cysteine sulfurs.

From figure 10(a) it can be seen that the fluctuations in diameters are small, in order of 0.2 Å. The general trend in diameters is decreasing. The diameters in general then show convergence in all systems. The differences in diameters between the systems are also small, less than 0.5 Å. In general it seems that the palmitic acid and Pleconaril systems exhibit slightly larger diameters than basic and Kirtan1 simulations. These differences could possibly indicate some effects on the structure caused by either missing

of the pocket factor or different structure of the pocket factor. However since the differences are very small, these possible indications are difficult to interpret.

The capsid radii from figure 10(b) show differences between simulated systems more clearly, since now also protein side chains are included in the size analysis. In general from the start of the simulation the trend in radii for all systems is decreasing. The longest, basic simulation seems to start to show increasing trend after more plateau region. However again the fluctuations and changes are very small, even this larger scale change is in order of 0.1 Å. The changing trend could possibly be related to the normal breathing motions of the virus. Radius related observations show also more clearly differences, although small, between the systems. It appears that the Pleconaril system exhibits larger radius compared to the other systems. This is in agreement with the observations made with trends of diameters as well. Again with the very small small differences this indication is not unambiguous to interpret. One possibility could be that the nature of the pocket factor does somehow affect the structural behavior of the capsid.

Figure 11 illustrates the differences in VP positions in the simulation with natural ligands (palmitic acid, myristate group). It can be seen, that during the 100 ns simulation the general positions of different VPs expectedly exhibit very minor changes. Figure 11 also illustrates the relative positions of different VPs; VP4 as the innermost protein exhibits the radius in order of 11 nm, whereas the surface protein chain VP1-VP3 exhibit radii in order of 13 nm.

Root mean square deviations in figure 12(a) illustrate that there are differences in behavior of between different system parts. In addition, the trend in these deviations is expectedly increasing, showing that the structure does fluctuate and evolves in the simulation. The deviations are modest, showing no alarming indications of, e.g., breaking of the structure. Also signs of convergence can be detected, indicating that the system structure is stabilizing. System parts exhibit differences in deviations, so that VP4 deviates clearly the most. Between other parts the differences are modest, so that in terms of these deviations VP3 seems to be the most stable. Deviations of protein as whole, VP1 and VP2 exhibit deviations close to each other. The larger deviations of the VP4 could be rationalized with the fact that majority of this protein is able to externalize upon binding to cell

receptors [10]. However, these deviations may also indicate the destabilization of VP4 structure since the simulation model does not include the genome, with which VP4 interacts in the actual, full virus. The root mean square fluctuations extracted from the basic simulation (figure 12(b)) in turn indicate that, in agreement with root mean square deviations, VP3 residues exhibit smallest fluctuations. The largest fluctuations on average are exhibited by VP4, again in agreement with the root mean square deviations. Also it is seen that VP1 contains highly mobile residues; this could possibly be rationalized with the fact that also the N-terminal of VP1 is able to externalize upon virus infection.

Summarizing observations from the structural terms, it could be said that expectedly drastic structural changes do no occur during these simulations. In this respect, the models are stable. The structural terms however show some differences between system parts and also between different systems. Due to smallness of these differences their actual indications are difficult to interpret. It could be interpreted, that the nature or complete missing of the pocket factor and its effects on the structural behavior of the capsid can be observed in the simulation. In addition, the differences in mobilities of different system parts observed in the simulation could be interpreted as being in agreement with some experimental observations.

4.2 General capsid dynamics

In spite of the limited time scale of the simulations, some results concerning the dynamical behavior of the capsid are extracted to more detailed extent than presented in considering the convergence and equilibration. However again large structural changes are not expected to occur during these simulations. Especially this section focuses on emphasizing differences between the model systems and parts within systems. In addition, focus is aimed at studying the dynamics of cysteine residues' sulfur atoms, since these are of interest in terms of the gold cluster attachment.

Related to these general dynamics, average root mean square deviations of different system parts and different systems are presented in figure 13. In addition, root mean square deviations of different system parts for the simulation with natural ligands are presented in figure 14 to be compared to the one in basic simulation (figure 12(a)). To further indicate the most

mobile parts of the capsid in the simulation with natural ligands root mean square fluctuations were converted to B-factors or temperature factors to visualize the most mobile parts better. The B-factor indicates the displacement of an atom from its mean value, i.e., its flexibility. The structural change of aggregation of myristate groups (attached to VP4) is visualized. Also solvent accessible surface area of the protein and hydrogen bonds within protein were calculated with Gromacs tools g_sas and g_hbond, respectively. In addition the diffusion of ions in the simulation relative to the capsid were observed: the amount of ions inside the capsid, by considering the capsid radius of 16 nm (found out to be the maximum radius in the previous work), were plotted.

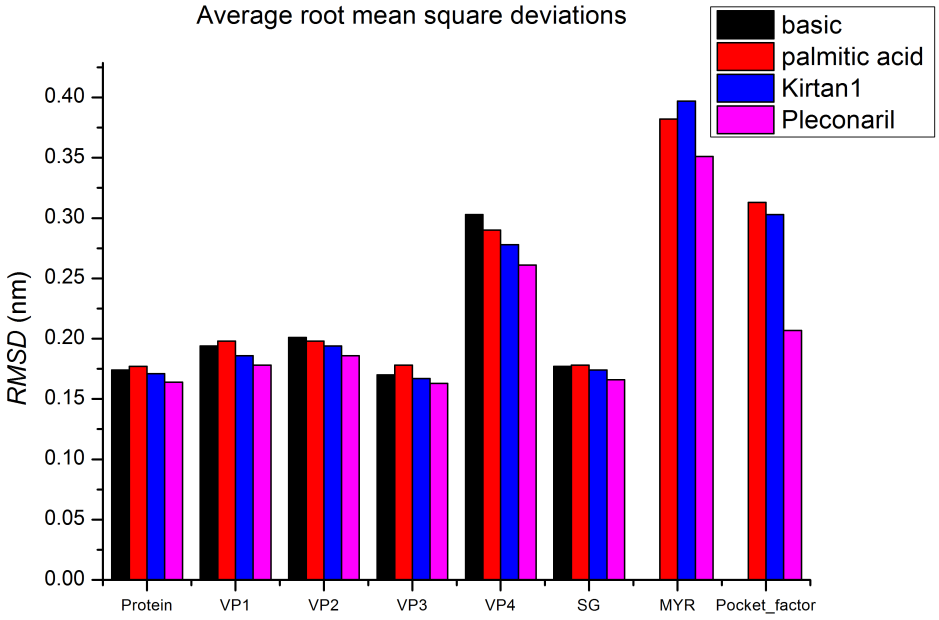


Figure 13: Average root mean square deviations of different parts of different systems. Averaged over simulation time.

From figure 13 it can be seen that expectedly the mobile ligands, i.e., myristate groups and pocket factor exhibit the largest on-average deviations. From the protein chains VP4 appears the most mobile, whereas VP3 appears the most immobile. The differences between systems are generally

speaking small and likely related to the differences in simulation times over which root mean square deviations were averaged. However the clearest, larger difference is observed with the pocket factors, so that Pleconaril shows clearly smaller deviations, suggesting its immobility related to palmitic acid and Kirtan1 molecules. This is to be expected due to rigid, ring-like structure of Pleconaril and Kirtan1 molecules. These differences are discussed in later section considering the hydrophobic pocket.

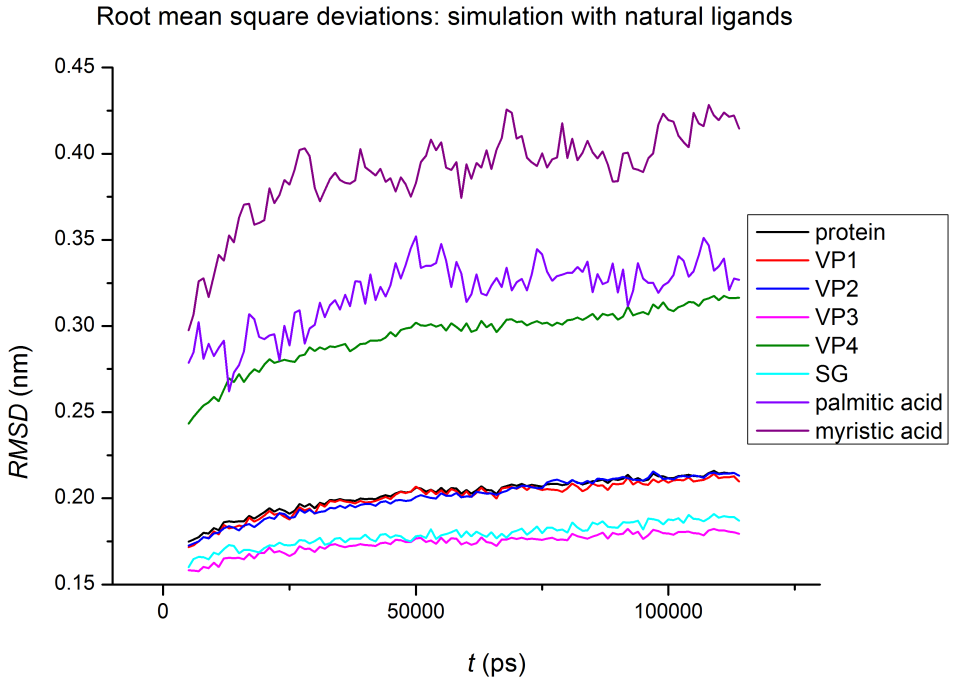
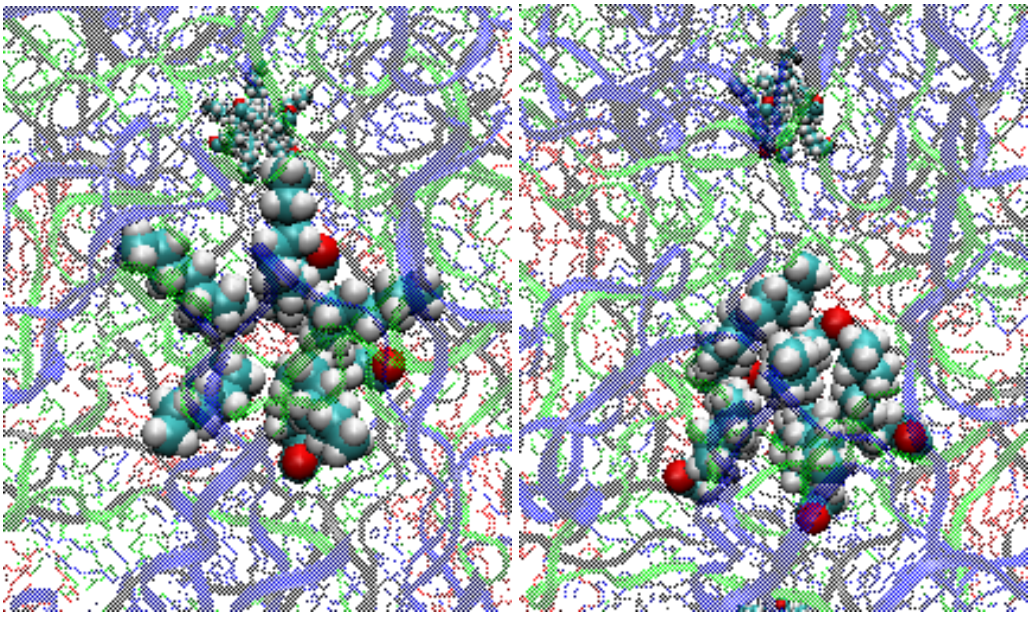


Figure 14: Root mean square deviations of different system parts in simulation with natural ligands.

Figure 14 shows root mean square deviations of lipid simulation comparable to those observed with the basic simulation, figure 12. In this sense, the missing of the pocket factor does not seem to affect the mobilities of system parts. Expectedly ligands, myristic and palmitic acids, exhibit larger deviations than other parts of the system. Myristate group seems more mobile than palmitic acid, possibly related it to having more space around it to move in.



(a) 0 ns, virus surface (b) 100 ns, virus surface

Figure 15: Aggregation of myristate group bundle at (a) 0 ns and (b) 100ns on the capsid surface in the simulation with natural ligands.

Figure 15 illustrates the moving of one bundle formed by five myristate groups attached to the VP4. It can be seen that in the beginning ((a)) there is a hole in the center of this bundle. In the simulation however these acids aggregate together in a way blocking this hole ((b)). This can be an effect of parameterization combined again with the fact that the model is missing the negatively charged genome from inside the VP4: the interactions of myristate groups and the genome could be crucial. It is therefore impossible to say, whether this aggregating indicates anything physical or is it in fact an unphysical occurrence.

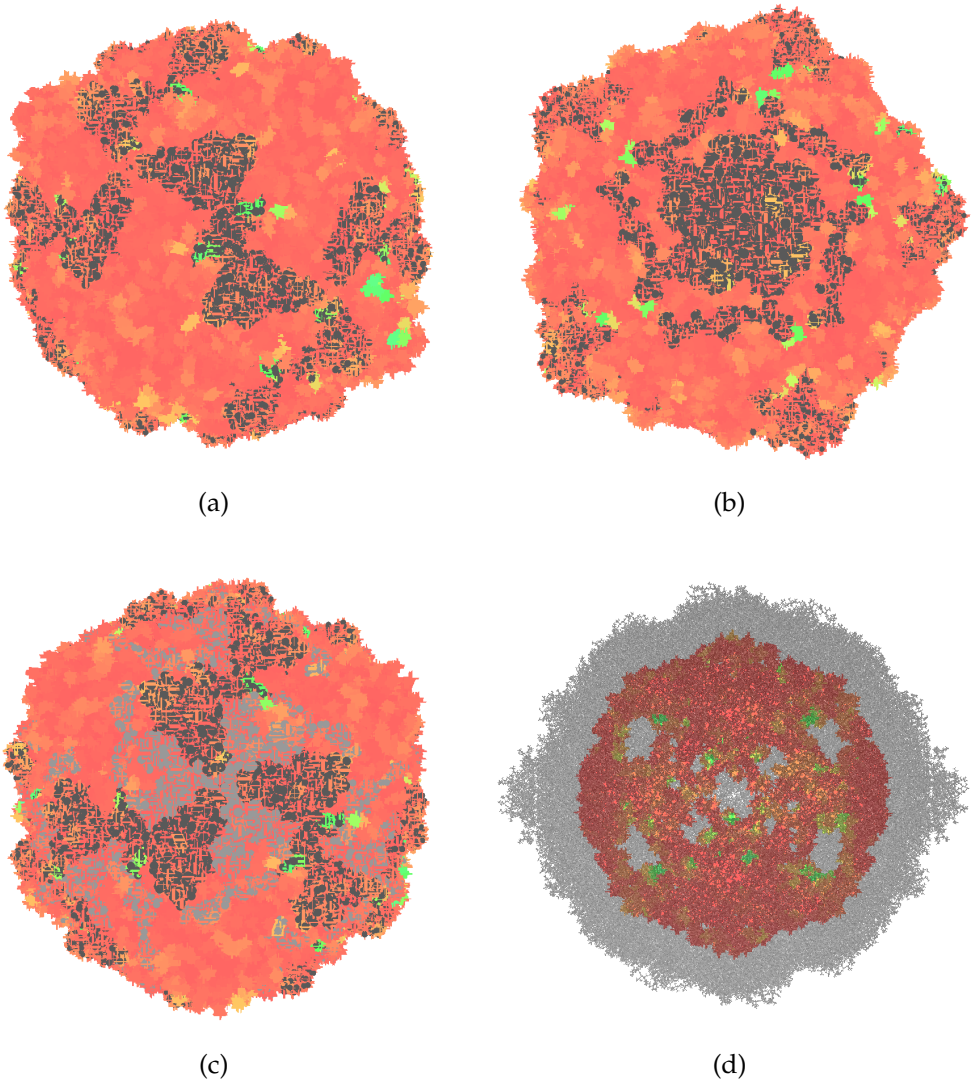


Figure 16: Mobilities of different protein parts from the simulation with natural ligands. Blue, green and yellow areas denote the more mobile parts, red the least mobile parts. Gray areas highlight different formations or symmetry axis on capsid surface: (a) highlights the two fold symmetry axis, (b) the pentamers, (c) propellers, and (d) the inner protein VP4.

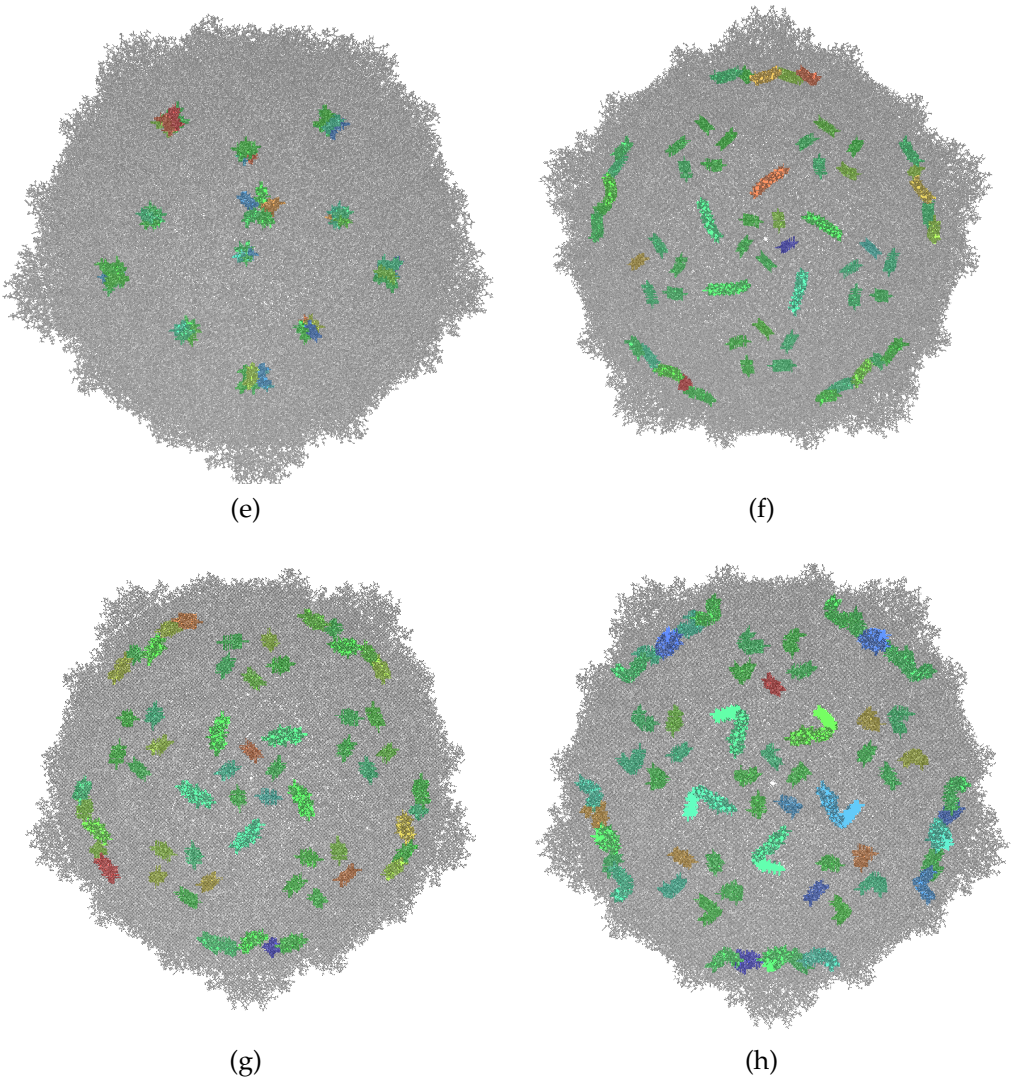


Figure 16: Mobilities of different protein parts. Blue, green and yellow areas denote the most mobile parts, red the least mobile parts. (e) shows the mobilities of myristate groups, (f) palmitic acids, (g) Pleconarils and (h) Kirtan1.

In figure 16 mobilities or flexibilities of different system parts are illustrated. With different capsid surface formations highlighted, it can be seen where the most mobile areas (colored green) are located. These parts could possibly indicate some weaker points in the capsid. It can be seen that

there are mobile areas located symmetrically close to the 2-fold symmetry axis (figure 16(a)) and also symmetrically around the pentamer formation (figure 16(b)). The higher flexibility of these positions could be rationalized with the experimental suggestion that the genome exits through the channel formed close to 2-fold symmetry axis, and the fact that the N-terminal of VP1 externalizes upon infection. No clear pattern is seen with respect to the propeller structure (figure 16(c)). Considering VP4 (figure 16(d)) it can be seen that it also contains areas of higher flexibility, possibly indicating weaker points in its structure. Considering ligands (figures 16(e)-(h)), it can merely be illustrated that there are differences in ligand mobilities within the same system based on their positions, possibly due to different entropic effects.

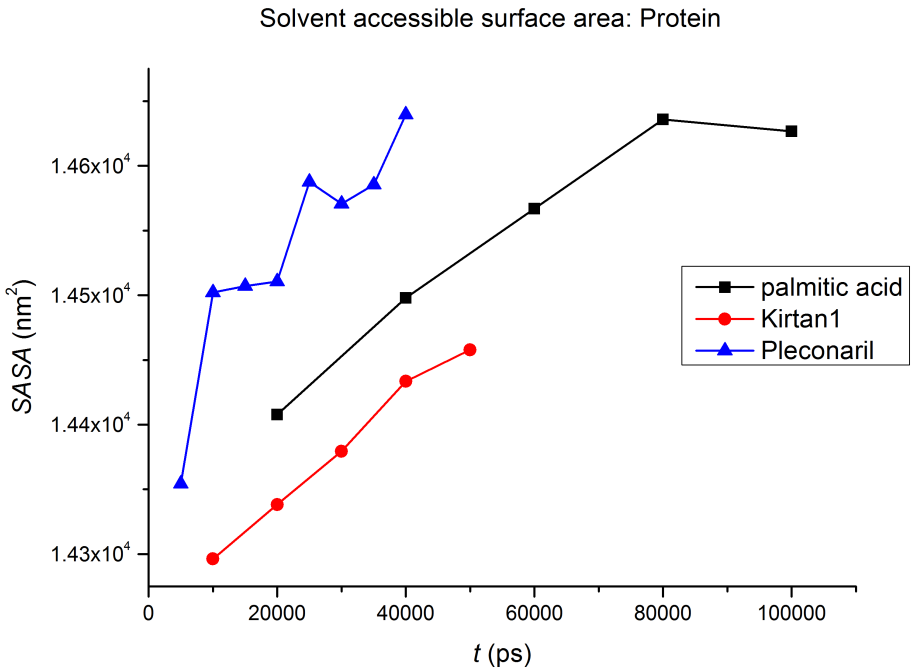


Figure 17: Solvent accessible surface area of protein in each system.

Figure 17 shows the evolving of protein solvent accessible surface area in palmitic acid, Kirtan1 and Pleconaril simulations. The trend in all is increasing and the differences relatively small considering the units of measure and the system sizes. It appears that the longest simulation with

natural ligands starts to show either signs of convergence or possible decrease at the very end of the simulation. Again due to the small differences and lack of convergence it is difficult to estimate whether these differences have any physical meaning related to, e.g., the nature of the pocket factor.

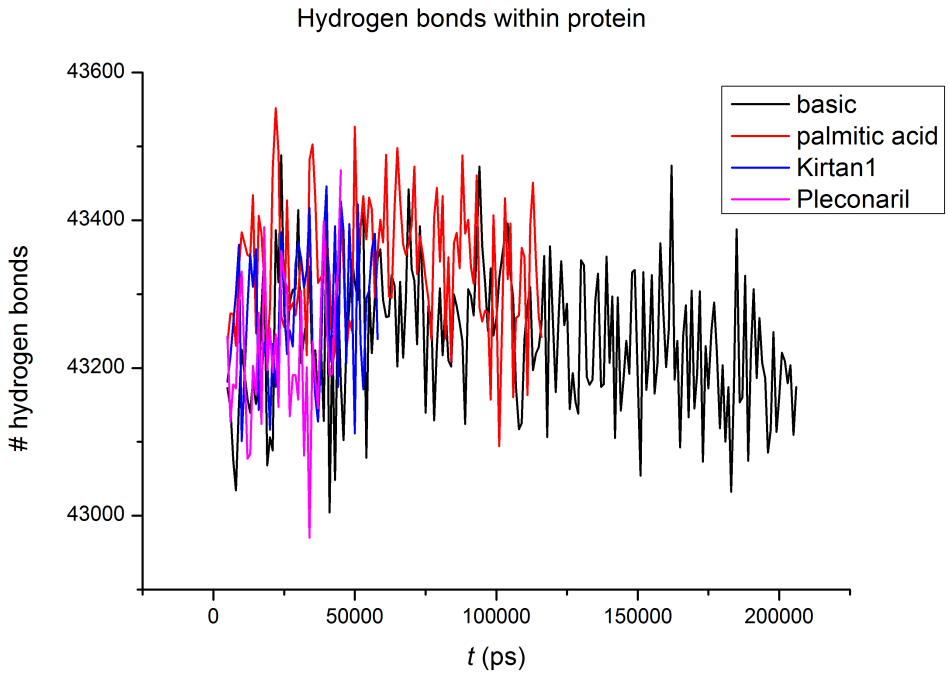
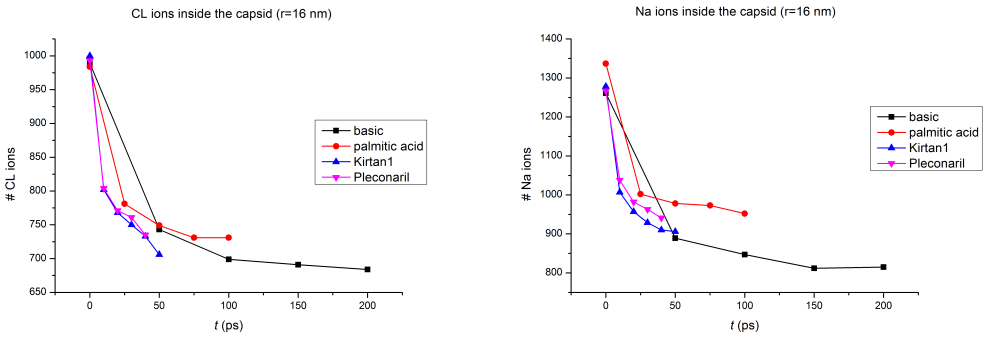


Figure 18: Protein internal hydrogen bonds in each simulation.

Internal hydrogen bonds, i.e., those within the protein are presented in figure 18 for each system. As a general note it can be said that no clear differences nor trends between and in different systems can be observed. On average it could be said that Pleconaril and Kirtan1 simulations may contain smaller amount of these hydrogen bonds, possibly indicating that the internal structure of the protein somehow changes with changing of the pocket factor from the natural ligand to Pleconaril or Kirtan1.



(a) (b)

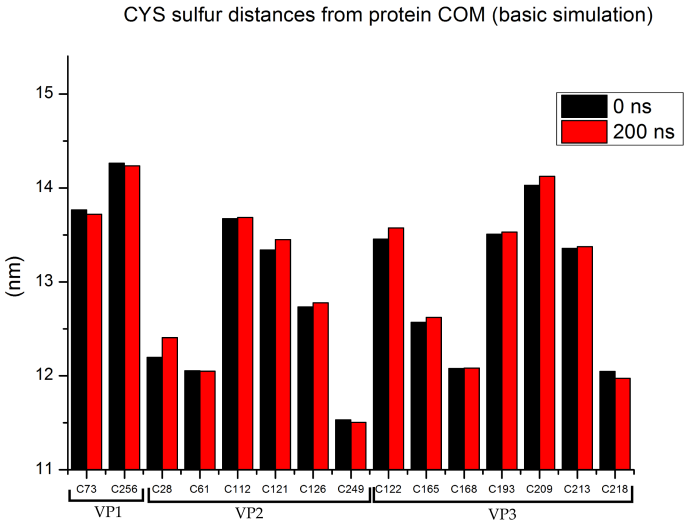
Figure 19: Number of different ion species ((a) Cl, (b) Na) inside the capsid in each simulation.

Also ion diffusion in the simulation was observed, the results are presented in figure 19. The observations indicate that both ion species are diffusing out of the capsid in all simulations. This in turn indicates that there are holes or channels present in the structure large enough to allow this diffusion. The exit routes of ions were not studied in more detailed in this work, but the general observation with Cl ions was that they tend to exit through the propeller centers formed by VP2 and VP3. The ability of ions to exit the capsid indicate that the system is able to adapt to the arbitrary ion concentration and positioning selected in the construction of the simulation.

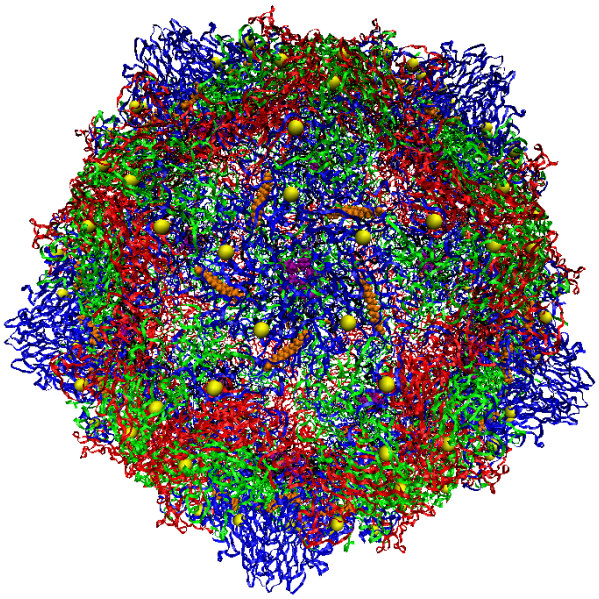
4.2.1 Cysteine sulfur dynamics

Cysteine sulfurs, especially those closest to the virus surface, are of interest due to their potential in attaching to the functionalized gold nanoclusters. However considering the crystal structure of the virus, these sulfurs are generally too buried to be able to attach the gold cluster without a proper, sufficiently long linker molecule. However experimentally the attaching is observed, so it is of interest to see whether the dynamics of these sulfurs in the simulation indicate somehow of their coming closer to the capsid surface. In addition, the presence of disulfide bonds are known to stabilize the structures of virus capsids in general. The crystal structure of EV1

did not indicate the positions disulfide bridges, so these were not set in the simulation either. However the possible positions of these bridges are discussed.



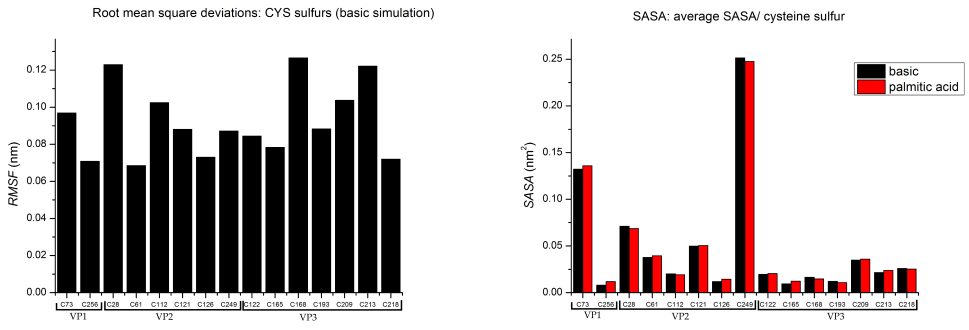
(a)



(b)

Figure 20: (a) Cysteine sulfur average distances from protein center of mass and its time evolving in basic simulation. (b) highlights the positions (yellow spheres) of the sulfurs closest to the surface (those of cysteines 73, 256 and 209)

Figure 20(a) shows, on average, the distance of cysteine sulfurs (SG) from the protein center of mass. Each residue, e.g. C73, is repeated 60 times in the complete structure, meaning that there are 60 sulfurs related to C73 residues. In general then, it can be seen that the sulfurs closest to the surface, i.e., furthest from the protein center of mass, are those related to cysteines C73, C256 and C209 (position illustrated in figure 20(b)). In the same figure also their time evolving is depicted. During the 200 ns basic simulation, it can be seen that in general the movements of sulfurs are relatively small. The largest general movements are observed with C28 sulfurs, that come closer to the surface by approximately amount of 2-3 Å. Considering the sulfurs closer to the surface to begin with (i.e. the most potential binding sites) it can be observed that bot C73 and C256 exhibit slight burial. However, C209 sulfurs exhibit small (approx. 1 Å) tendency to come closer to the surface. Based on the short simulation and small differences, it cannot clearly be pointed out which of the cysteines would provide the most potential binding sites. The slight externalization of C209 sulfurs could however suggest that these could become the most potential ones for the attaching. To see whether the externalizing trend continues, longer simulation would be needed.



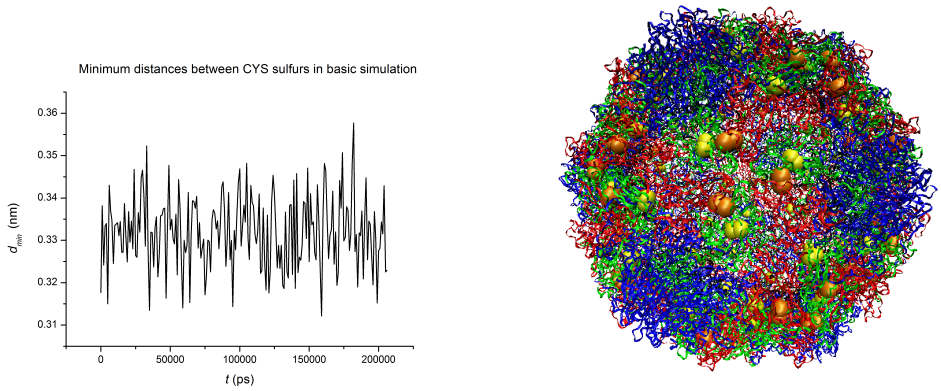
(a) RMSF

(b) SASA

Figure 21: (a) Root mean square fluctuations of and (b) solvent accessible surface areas of cysteine sulfurs in basic simulation.

Figure 21(a) in turn shows the time averaged root mean square fluctuations of each cysteine sulfurs. The differences are small, and the results do not indicate higher mobilities considering the sulfurs closest to the surface. Also solvent accessible surface areas for different sulfurs are presented in

figure 21(b). In spite of showing large differences, the areas are generally very small. However considering the sulfurs closest to the surface, C73 seem to exhibit largest solvent accessible surface area. This could indicate its potential as a binding site over C256 and C209. The differences in SASAs are very small between the two systems (basic, palmitic acid). Clearly the largest solvent accessible surface area is exhibited by sulfurs in CYS249. From figure 20(a) it can be however be seen that these residues is located far from the surface of the protein, most probably indicating its accessibility for the solvent from the inside of the capsid.



(a) Minimum distances

(b) Potential disulfide bridge positions

Figure 22: (a) Minimum distances between cysteine sulfurs in the basic simulation and (b) highlighted positions of the closest cysteine sulfur pairs, i.e., potential positions of disulfide bridges. Orange spheres are C121-C122 and yellow spheres C168-C218 pairs.

Figure 22(a) shows the minimum distances between two cysteine sulfurs during the basic simulation. In the simulation formations of disulfide bridges were not separately defined (since these bridges were not indicated in the crystal structure), meaning that in the simulation two sulfurs cannot come too close, for example the distance of disulfide bridge, to each other due to the repulsion resulting from their Van der Waals parameters. It is observed that the minimum distances in the simulation are in all cases between C168-C218 (both in VP3) or C121-C122 (C121 in VP2, C122 in VP3) sulfur pairs. These are highlighted in figure 22(b) to indicate possible positions of disulfide bridges in the structure. The possible bridge

between C121-C122 could for example strengthen the structure of the propeller. In future simulations these disulfide bridges could be imposed to the structure to see their possible effects on dynamics.

4.3 Hydrophobic pocket

Of a special interest in the simulation is the hydrophobic pocket and the pocket factor, due to its potential to be the binding site for the functionalized gold cluster. In order for the pocket to be a usable binding site, the natural pocket factor, palmitic acid, would be needed to be replaced by some other, proper molecule to which the gold cluster could be linked to. A known drug molecule, Pleconaril, is able to replace palmitic acid in the pocket. Kirtan1 is in turn a Pleconaril derivative, that provides a potential molecule for linking to the gold cluster. Interactions of these different pocket factors: palmitic acid, Kirtan1 and Pleconaril were studied in the simulations.

The structures of these pocket factors are presented in figure 2. Palmitic acid is a standard, C16 fatty acid whose parameters were taken from the existing force field parameters for similar residues. Pleconaril and Kirtan1 are more complicated molecules with aromatic rings, and their parameters were derived using Antechamber. The docking of the Kirtan1 and Pleconaril to hydrophobic pocket was done elsewhere [70]. The Pleconaril is located inside the pocket as a whole, whereas the Kirtan1 is positioned so that the long hydrocarbon tail sticks out from the pocket mouth [70]. The positions were found to be stable in short simulations inside one asymmetric unit (1/60) of the virus capsid [70]. In the simulations of this work each of the 60 pocket of a full virus is occupied by the pocket factor, that is, the results are also collectively for 60 molecules unless otherwise noted.

Figure 23 shows the radius of gyration for different pocket factors. Based on this, it can be estimated that the palmitic acids on average are located deeper in the pocket, i.e., closer to the protein center of mass compared to Pleconaril and Kirtan1. The Kirtan1 molecules expectedly exhibit larger radius due to the externalized hydrocarbon tails. The changes in these radii for each pocket factor during simulations are small, indicating that at least for these time scales they do not significantly move with respect to the protein center of mass.

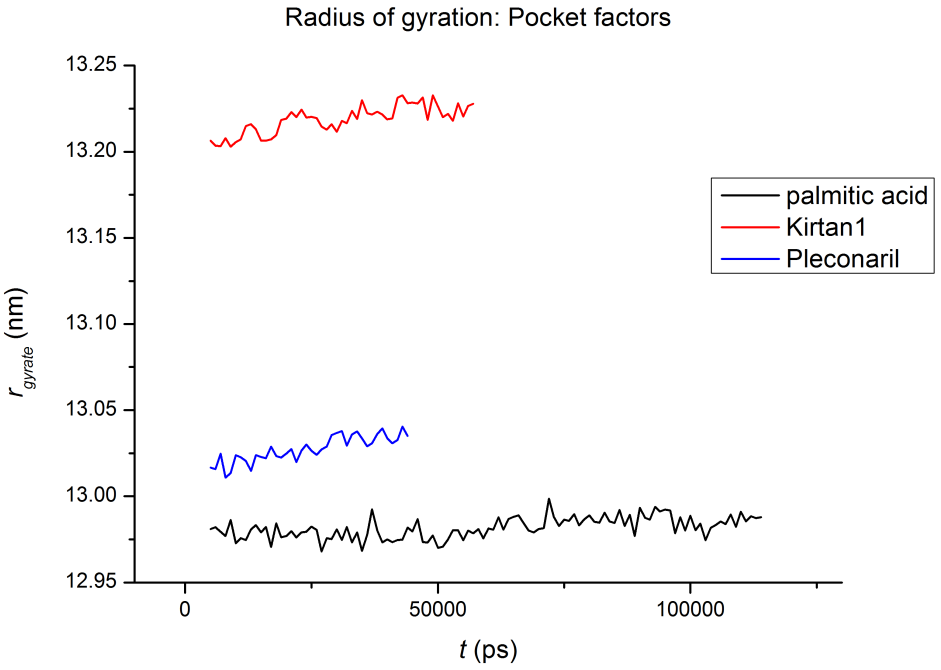


Figure 23: Radii of gyration of different pocket factors.

To study the interactions governing the binding to the pocket, hydrogen bonds between each pocket factor and protein were studied, and results are presented in figure 24. It can be seen that in terms of hydrogen bonds the natural ligand binds more favorably than Pleconaril or Kirtan1. However, since it is known that Pleconaril can replace palmitic acid, there need to be other interactions contributing to the binding affinity. These most likely include aromatic interactions.

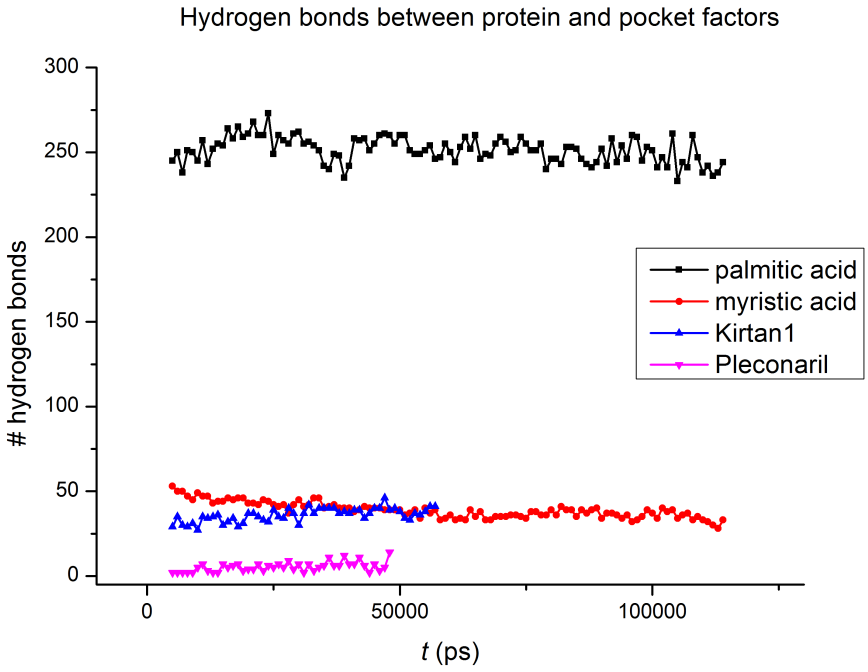
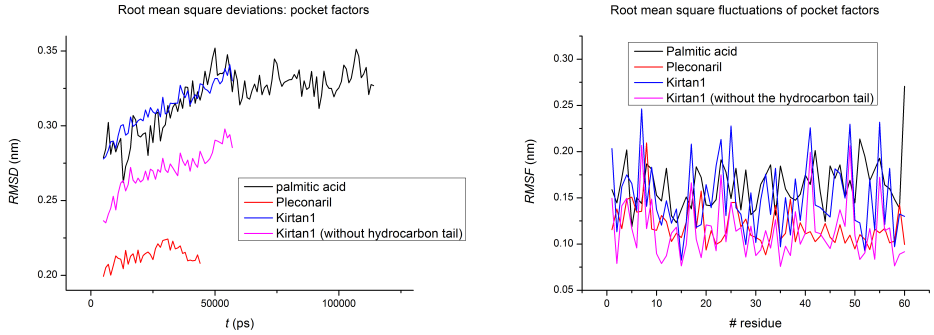


Figure 24: Hydrogen bonds between pocket factor and protein in each system. Results are for all the pocket factors (60) in each system.

To better study mobilities of different pocket factors, root mean square deviations and fluctuations are presented in figure 25. The least squares fitting for each frame was done on the initial protein structure at 0 ns. To better compare the pocket part of Kirtan1, also results omitting the hydrocarbon tail are presented. In terms of root mean square deviations (figure 25(a)), it appears that Pleconaril expectedly considering its rigid ring-like structure is the most immobile one. Expectedly palmitic acid is the most mobile or flexible one. Kirtan1 without the highly mobile hydrocarbon tail (also containing rigid ring structures) positions between these, exhibiting then larger deviation than Pleconaril. This could indicate that Kirtan1 does not stay as stably in the pocket as Pleconaril. The deviations in the dynamics of Kirtan1 also show relatively steeply increasing trend, possibly indicating its change of position in the pocket. Whether Kirtan1 actually starts to externalize from the pocket, would require more simulation time. General larger scale trend in RMSD for all the molecules is increasing, that is, fluctuation compared to the initial structure. This increase is steeper for

Kirtan1 and palmitic acid compared to Pleconaril. The longest simulation with palmitic acid however starts to show converging larger scale trend in RMSD, possibly suggesting the stabilization of pocket factors positions. In terms of root mean square fluctuations (figure 25(b)), in agreement to deviations, Pleconaril exhibits the smallest fluctuations, whereas palmitic acid exhibits the largest. Again Kirtan1 without the hydrocarbon tail places between these two.

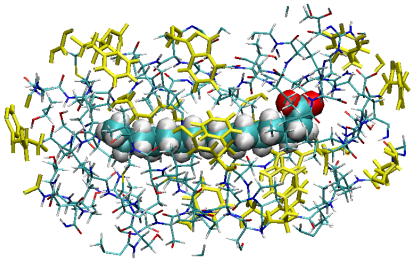


(a) RMSD

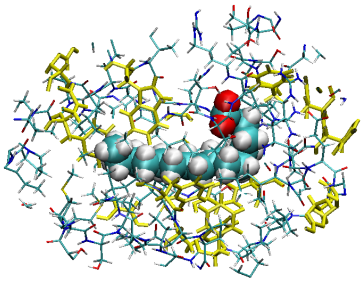
(b) RMSF

Figure 25: (a) Root mean square deviations and (b) fluctuations of different pocket factors.

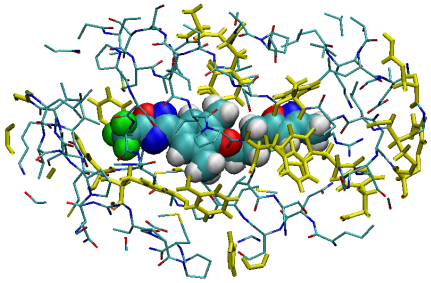
To visualize the movements of different pocket factors, snapshots at the start and the end of each simulation of one pocket and pocket factor, figure 26, and five pockets and pocket factors around the pentamer on the virus surface, figure 27, are presented. In figure 26, the aromatic residues close to the pocket factor (within 5 Å) are highlighted in yellow.



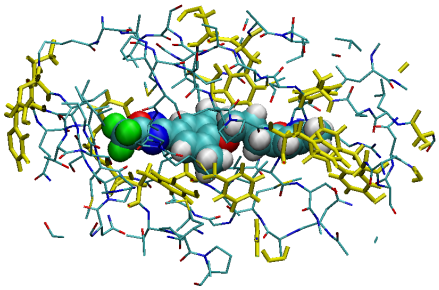
(a) 0 ns, PLT



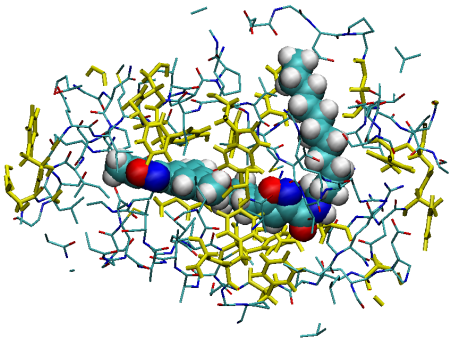
(b) 100 ns, PLT



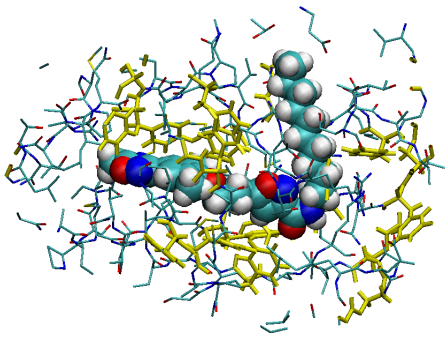
(c) 0 ns, PLR



(d) 44 ns, PLR



(e) 0 ns, PLE

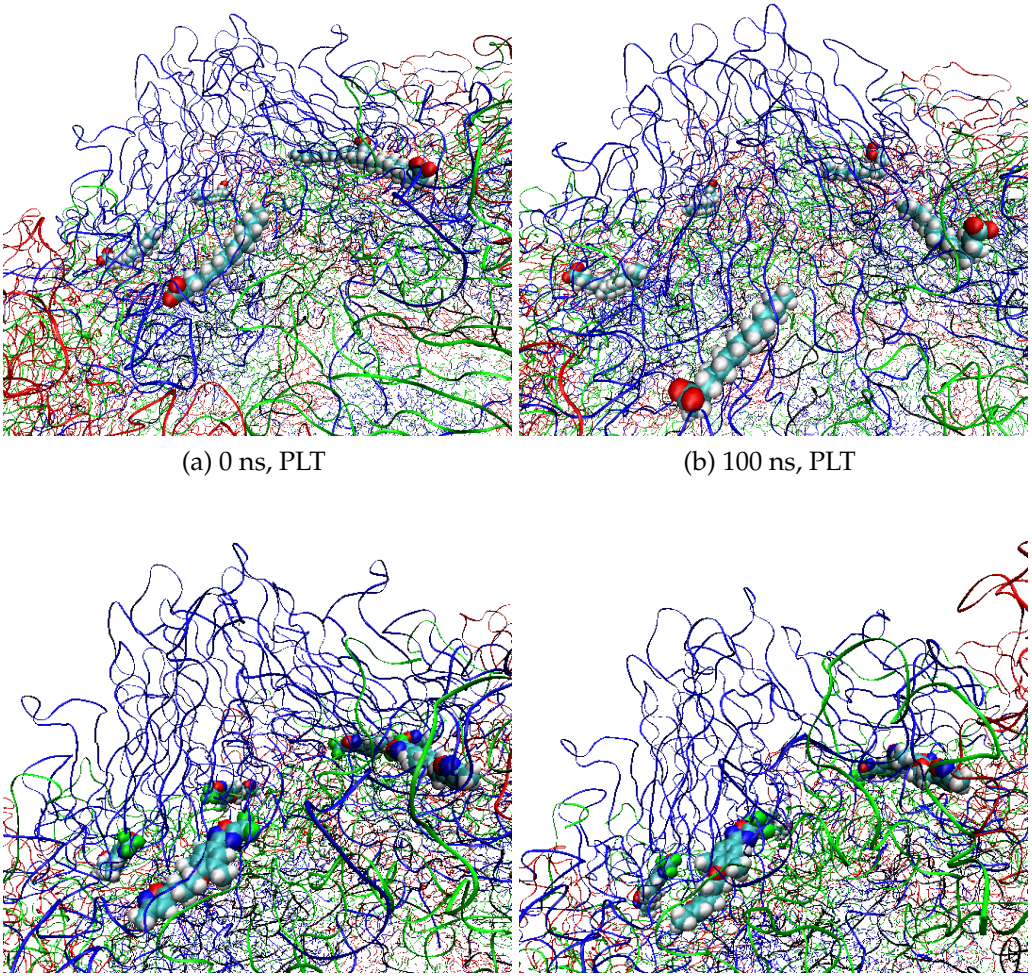


(f) 57 ns, PLE

Figure 26: Snapshots of pocket factors at the beginning and the end of each trajectory. The pocket factors are presented as Van der Waals spheres with cyan=C, blue=N, red=O, white=H, green=F. Aromatic residues close to the pocket factor are highlighted in yellow. PLT, PLR and PLR denote palmitic acid, Pleconaril and Kirtan1, respectively.

From figure 26 it can be seen that there are several aromatic residues lying

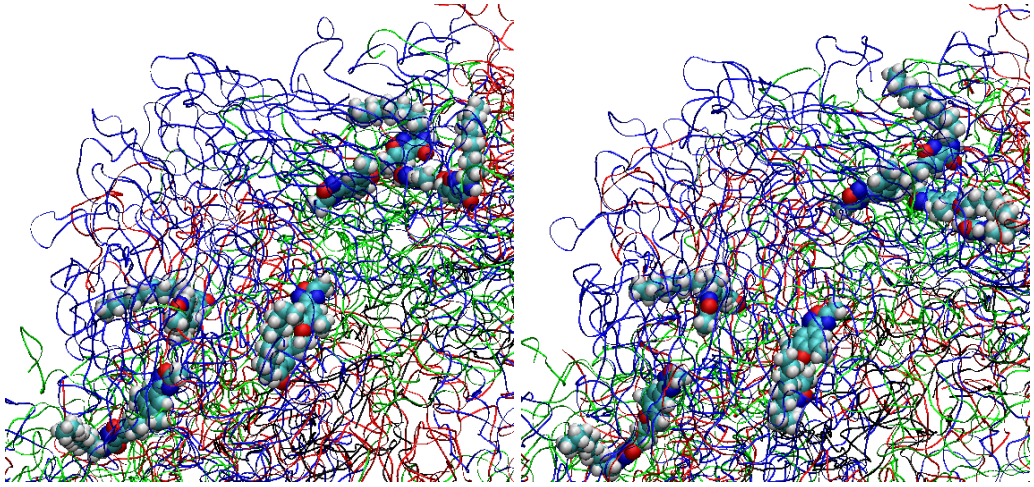
close to the pocket factor, which indicates that aromatic interactions most probably play a role in binding of Pleconaril and Kirtan1 in the pocket. From the motions of palmitic acid (figure 26(a)-(b)), it can be seen that structure twists in the simulation so that the acid head turns outwards from the virus surface. The same trend can be seen in figure 27(a)-(b). This may indicate the beginning externalization of the palmitic acid, and suggest that the pocket factor does not necessarily stay stably in the pocket in the simulation. On the other hand, with shorter time scale simulation, Pleconaril in turn shows no clear, visual signs of movements outwards from the virus surface, in figures 26(c)-(d) and 27(c)-(d). This could suggest its more stable staying in the pocket compared to palmitic acid. This would be in qualitative agreement with the higher binding affinity of Pleconaril over palmitic acid. Visualizing the motions of Kirtan1 (figure 26(e)-(f), figure 26(e)-(f)), it seems that similarly to Pleconaril the rigid, pocket part stays stably in the pocket in the simulation. In some cases the hydrocarbon tail expectedly shows higher mobility and tendency to reach out outwards from the pocket.



(a) 0 ns, PLT (b) 100 ns, PLT

(c) 0 ns, PLR (d) 45 ns, PLR

Figure 27: Snapshots of bundles pocket factors on the virus surface at the beginning and the end of each trajectory. PLT, PLR and PLE denote palmitic acid, Pleconaril and Kirtan1, respectively.



(e) 0 ns, PLE (f) 57 ns, PLE

Figure 27: Snapshots of bundles pocket factors on the virus surface at the beginning and the end of each trajectory. PLT, PLR and PLE denote palmitic acid, Pleconaril and Kirtan1, respectively.

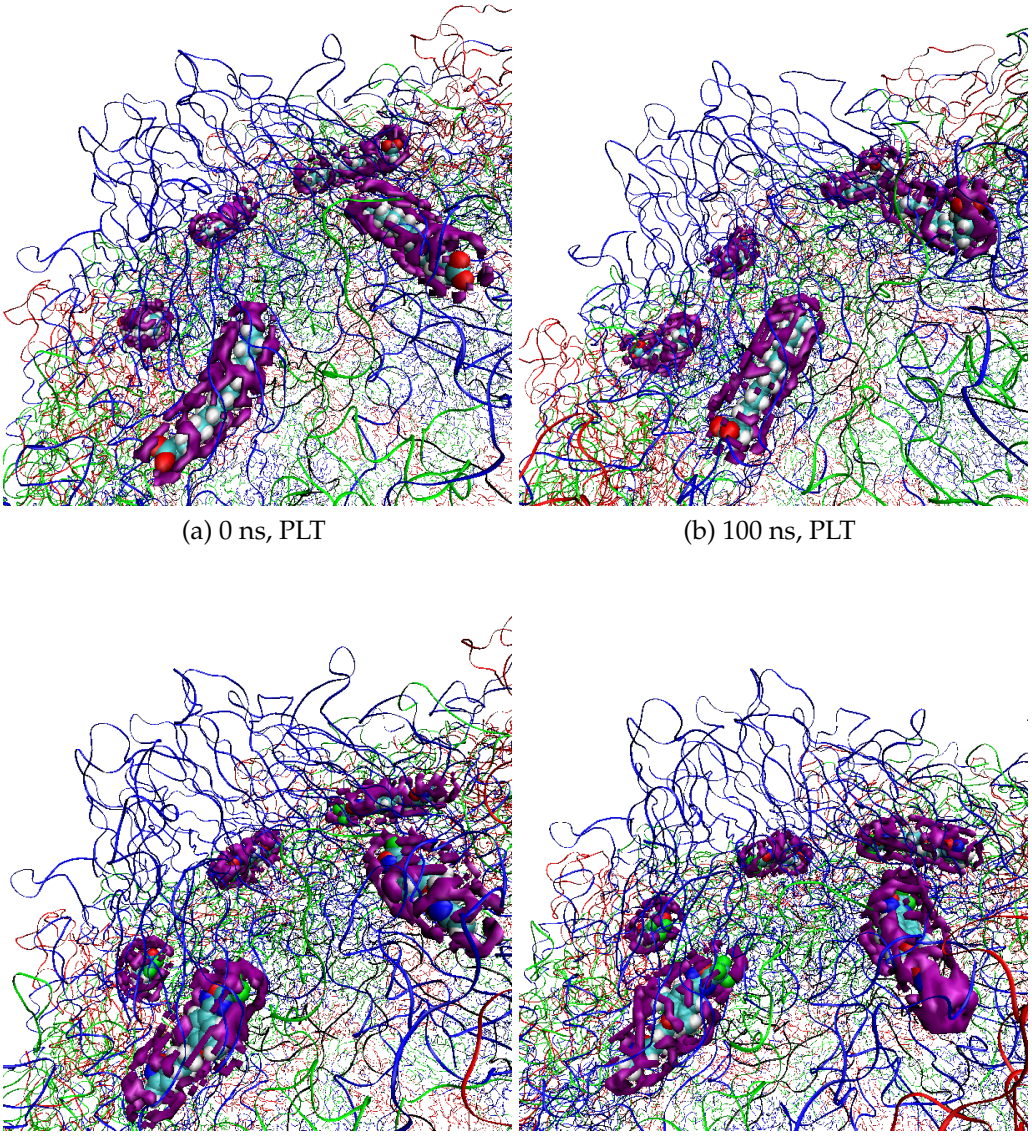
Table 1 in turn lists some results indicating the empty space or cavity around the pocket factors. Also solvent accessible areas are included. These were calculated with VMD VolArea package. The trends in the cavity volumes and areas vary from system to system. For palmitic acid the area around the pocket factors decreases significantly between 0 ns and 50 ns, and more moderately between 50 ns and 100 ns. Considering one single pocket factor, the cavity increases first from 0 ns to 50 ns, but moderately decreases from 50 ns to 100 ns. Considering larger area around the pocket factor, the cavity first clearly decreases from 0 ns to 50 ns, and exhibits small increase from 50 ns to 100 ns. The SASA of palmitic acids clearly decreases from 0 ns to 50 ns and moderately increases from 50 ns to 100 ns. For Pleconaril system the trend in cavity around the pocket factor first increases from 0 ns to 20 ns, and then slightly decreases from 20 ns to 40 ns. For single Pleconaril the trend in cavity is throughout increasing. Considering the larger area around the pocket factors, the cavity clearly decreases from 0 ns to 40 ns. The area in turn decreases only slightly. For Kirtan1, considered without the hydrocarbon tail, the trend in cavity volume from 0 ns to 25 ns is moderately increasing, and slightly decreases from 25 ns to 50 ns. Similar trend is observed for single Kirtan1 molecule.

The trend in larger area around the Kirtan1 molecules is throughout decreasing. Area in turn first increases from 0 ns to 25 and then decreases again from 25 ns to 50 ns.

Based on so few points general trends of behavior of the pocket around the different pocket factors cannot be extracted. Expectedly the most drastic changes both in cavity volumes and areas appear in the beginning of the simulation, but as the simulation proceeds the differences become smaller, but fluctuations appear. In general it appears that on average the cavity is largest around Kirtan1 molecule and smallest around palmitic acid molecule. Same is true for the solvent accessible surface area. These are both in agreement with the sizes of the molecules. Figure 28 illustrates the shape of the cavity around one pocket factor in each simulation. No significant variations are observed.

Table 1: Changes in the environment of the pocket factor during each simulation. PLT, PLR, PLE and pf denote palmitic acid, Pleconaril, Kirtan1 and pocket factor, respectively.

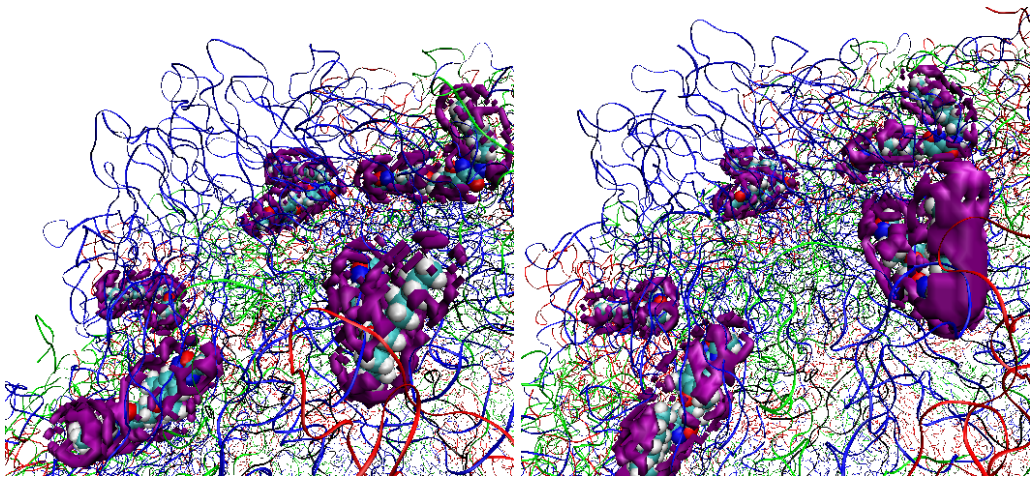
System	time (ns)	Cavity around pf (Å^3)	Cavity around one pf (Å^3)	Cavity around pf and 5 Å environment (Å^3)	Area (Å^2)
PLT	0	11915	215	27142	35518.143
PLT	50	11652	225	25244	34369.714
PLT	100	11611	199	25329	34543.138
PLR	0	11971	204	27025	35778.875
PLR	20	12191	218	-	-
PLR	40	12037	242	25396	35814.222
PLE	0	12425	206	27321	37380.138
PLE	25	12611	255	26674	37659.220
PLE	50	12587	239	26156	37339.121



(a) 0 ns, PLT (b) 100 ns, PLT

(c) 0 ns, PLR (d) 45 ns, PLR

Figure 28: Snapshots of pocket and pocket factor with the surrounding cavity highlighted in purple. From Kirtan1 molecule the hydrocarbon tail is omitted. PLT, PLR and PLE denote palmitic acid, Pleconaril and Kirtan1, respectively.



(e) 0 ns, PLE (f) 57 ns, PLE

Figure 28: Snapshots of pocket and pocket factor with the surrounding cavity highlighted in purple. From Kirtan1 molecule the hydrocarbon tail is omitted. PLT, PLR and PLE denote palmitic acid, Pleconaril and Kirtan1, respectively.

5 Summary and outlook

In this work a complex model system of the full Echovirus 1 for molecular dynamics simulations was successfully set up and simulated up to timescale of 200 ns. In addition to the virus capsid constructed of standard amino acids, different important ligand molecules were parameterized for the AMBER99SB-ILDN force field and included in the model and simulations. Although the simulations performed here do not reveal drastic conformational changes or clearly interpretable results of dynamics of the virus, this model and simulations are stable and sufficiently equilibrated to work as a starting point for future simulations and analysis methods, more focused on the effects of gold clusters. These include for example evaluation of binding affinities of different pocket factors. Also parameterization and inclusion of the actual gold cluster to the simulation is of interest.

The simulated systems expectedly do not show drastic structural changes

during the simulation times. During the simulations systems in general start to show signs of convergence both in energetic and structural terms. Both these observations suggest that our model is stable in the simulation, and therefore also meaningful to be simulated further to obtain more detailed data on structure and dynamics of this system.

In general the interpretation of small changes and variations in the structural terms is far from unambiguous with this short simulation times, due to both insufficient equilibration and also the fact that the simulation time is very short in the life of a virus. That is, it is difficult to interpret whether the differences or dynamics are actually physically meaningful. However, e.g., the observed differences in mobilities of different protein chains could be interpreted as being in agreement of experimental data stating that VP1 and VP4 are externalized upon virus infection. The high frequency changes in the radius of gyration can be of course related to the movements of the most mobile residues. However with the longest simulation also a starting, larger scale trend of the radius increase is observed; this could be related to the breathing motion of the capsid. How this proceeds in time could be studied in a longer simulation.

The observed diffusion of ions out of the capsid at least gives information that the system is able to adapt to the arbitrarily chosen ion concentration. That is, there should not form large osmotic pressure inside the capsid. This also suggests that the capsid structure contains channels or holes through which ions can exit. More careful and detailed tracking of these exit routes could reveal positions of such channels.

The pocket factor interactions studied here do not reveal anything unexpected. In terms of hydrogen bonds palmitic acid binds more favorably than Pleconaril or Kirtan1. However aromatic interactions are expected to be essential in binding of these molecules. The traditional simulation can not reveal much about these interactions. By tracking the movements of the different pocket factors during our short simulations show that the pocket factors stay in place in the pocket, and do not at least clearly start to externalize. However again the short simulation time does not necessarily capture the larger scale motions of these pocket factors. In order to better study the interactions and shed light on different binding of these ligands in future free energy calculations could be performed. In spite of missing experimental data, at least quantitative information could be extracted to compare which of the ligands potentially bonds the most favorably. This

could pave the way to also study the interactions in more detail.

Other aspects that could be of interest in further simulations include more detailed structural changes relating especially to hydrophobic pocket and the canyon area (cell receptor area) and possibly studying effects of imposing potential disulfide bridges that are actually missing form the crystal structure. In addition it would be of interest to recognize structural differences possibly related to the presence of the different pocket factors; this could suggest the function mechanism and the role of the pocket factor. Also potentially force measurements of different capsid parts in simulations could give interesting quantitative information on elastic properties of the capsid.

The observations made in this work in general emphasize further the role of experiments alongside simulations. Especially the time scale difference of simulation and experiment is a challenge. Looking even further to the possible future of these simulations, it may be that steered molecular dynamics simulation would best provide information from this type of simulations. That is, simulations could utilize experimental data to an extent that a simulation could actually be steered in the direction suggested by some experiment. In a way then, the time scale problem could be overcome if the system can be stimulated to undergo some conformational change.

In general simulating full virus systems is challenging. Along with difficulties in equilibration and the time scale issue, also other practical limitations have become clear. First, the current model is missing the genome. Although the capsid were able to exist without the genome, its missing can affect the structure, for example the size of the virus along with the structure of the surface. In addition a questionable issue is the ability of viruses to vary in time. The question then becomes of whether the model we are simulating based on the crystal structure from beyond 10 years has anything to do with the possibly mutated viruses existing today.

The simulations and the model proposed in this work however suggest that the system is stable and physically meaningful to be used in further simulations in studying this virus. The simulations run so far have also helped to recognize the points of interest and possibilities to be studied in further simulations, and also then promote that molecular dynamics and extensions to traditional simulations can be a valuable tool in virus studies

alongside experiments. The model proposed and simulations performed in this work then lay foundation of studying this and possibly also other virus systems with further molecular dynamics simulations and also extended methods such as free energy calculations.

References

[1] Varpu Marjomäki, Tanja Lahtinen, Mari Martikainen, Jaakko Koivisto, Sami Malola, Kirsi Salorinne, Mika Pettersson, and Hannu Häkkinen. Site-specific targeting of enterovirus capsid by functionalized monodisperse gold nanoclusters. *PNAS*, 111(4):1277–1281, 2014.

[2] Mark J. Foster. Molecular Modelling in Structural Biology. *Micron*, 33(4):365–384, 2002.

[3] Donald K Phelps, Brent Speelman, and Carol Beth Post. Theoretical studies of viral capsid proteins. *Current Opinion in Structural Biology*, 10(2):170–173, 2000.

[4] Daniel George Angelescu and Per Linse. Modelling of icosahedral viruses. *Current Opinion in Colloid & Interface Science*, 13(6):389–394, 2008.

[5] M. Luo. Assembly of viruses: Nonenveloped particles. In M.H.V. Van Regenmortel and B.W.J. Mahy, editors, *Encyclopedia Of Virology*. Academic Press, third edition, 2008.

[6] T. Hyypiä. Echoviruses. In M.H.V. Van Regenmortel and B.W.J. Mahy, editors, *Encyclopedia Of Virology*. Academic Press, third edition, 2008.

[7] K.J. Ertel and B.L. Semler. Picornaviruses: Molecular Biology. In M.H.V. Van Regenmortel and B.W.J. Mahy, editors, *Encyclopedia Of Virology*. Academic Press, third edition, 2008.

[8] Filman D.J., Wien M.W., J.A. Cunningham, J.M. Bergelson, and J.M. Hogle. Structure determination of echovirus 1. *Acta Cryst. D*, D54:1261–1272, 1998. PDB ID: 1EV1.

[9] M.D. Ryan and L.E. Hughes. Enteroviruses of Animals. In M.H.V. Van Regenmortel and B.W.J. Mahy, editors, *Encyclopedia Of Virology*. Academic Press, third edition, 2008.

[10] Antero Airaksinen. *The VP1 intracapsid hook and uncoating of enteroviruses*. PhD thesis, University of Helsinki, 2000.

[11] Xiangxi Wang, Wei Peng, Jingshan Ren, Zhongyu Hu, Jiwei Xu, Zhiyong Lou, Xumei Li, Weidong Yin, Xinliang Shen, Claudine Porta, Thomas S Walter, Gwyndaf Evans, Danny Axford, Robin Owen, David J Rowlands, Junzhi Wang, David I Stuart, Elizabeth E Fry, and Zihe Rao. A sensor-adaptor mechanism for enterovirus uncoating from structures of EV71. *Nature Structural & Molecular Biology*, 19:424–429, 2012.

[12] Pavel Plevka, Rushika Perera, Moh Lan Yap, Jane Cardosa, Richard J. Kuhn, and Michael G. Rossmann. Structure of human enterovirus 71 in complex with a capsid-binding inhibitor. *PNAS*, 110(14):5463–5467, 2013.

[13] Jodi K Muckelbauer, Marcia Kremer, Iwona Minor, Guy Diana, Frank J Dutko, James Groarke, Daniel C Pevear, and Michael G Rossmann. The structure of coxsackievirus B3 at 3.5 Å resolution. *Structure*, 3(7):653 – 667, 1995.

[14] Emmi Pohjolainen. Modeling of Structure and Dynamics of Enteroviruses With Molecular Dynamics Simulations. Research training thesis, University of Jyväskylä, Department of physics, 2013.

[15] D.C. Rapaport. *The Art Of Molecular Dynamics Simulation*. Cambridge University Press, 2004.

[16] D. van der Spoel, E. Lindahl, B. Hess, A.R. van Buuren, E. Apol, P.J. Meulenhoff, D.P. Tieleman, A.L.T.M. Sijbers, K.A. Feenstra, R. van Drunen, and H.J.C. Berendsen. *Gromacs user manual version 4.5.6*, `www.gromacs.org`, 2010.

[17] William L. Jorgensen and Julian Tirado-Rives. Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *PNAS*, 102(19):6665–6670, 2005.

[18] Alan Hinchlife. *Molecular Modelling for Beginners*. Wiley, 2003.

[19] William L. Jorgensen, Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.

[20] Alan Grossfield and Daniel M. Zuckerman. Chapter 2: Quantifying Uncertainty and Sampling Quality in Biomolecular Simulations. volume 5 of *Annual Reports in Computational Chemistry*, pages 23 – 48. Elsevier, 2009.

[21] Bernhard Knapp, Sophie Frantal, M. Cibena, Wolfgang Schreiner, and P. Bauer. Is an Intuitive Convergence Definition of Molecular Dynamics Simulations Solely Based on the Root Mean Square Deviation Possible? *Journal of Computational Biology*, pages 997–1005, 2011.

[22] Samuel Genheden and Ulf Ryde. Will molecular dynamics simulations of proteins ever reach equilibrium? *Phys. Chem. Chem. Phys.*, 14:8662–8677, 2012.

[23] Berk Hess, Carsten Kutzner, David van der Spoel, and Erik Lindahl. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008.

[24] Utrecht University NMR Spectroscopy Research Group, Bijvoet Center from Biomolecular Research. Molecular Dynamics simulation tutorial. `http://nmr.chem.uu.nl/~tsjerk/course/molmod/analysis.html`, 2008.

[25] Kresten Lindorff-Larsen, Stefano Piana, Kim Palmo, Paul Maragakis, John l. Klepeis, Ron O. Dror, and David E. Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proceedings of the National Academy of Sciences of the United States of America*, 102:1950–1958, 2010.

[26] Jay W. Ponder and David A. Case. Force fields for protein simulations. *Advances in protein chemistry*, 66:27–85, 2003.

[27] Yong Duan, Chun Wu, Shibasish Chowdhury, Mathew C. Lee, Guoming Xiong, Wei Zhang, Rong Yang, Piotr Cieplak, Ray Luo, Taisung Lee, James Caldwell, Junmei Wang, and Peter A. Kollman. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry*, 24(16):1999–2012, 2003.

[28] J. Wang, Wolf R. M., J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and testing of a general AMBER force field. *Journal of Computational Chemistry*, 25:1157–1174, 2004.

[29] Viktor Hornak, Robert Abel, Asim Okur, Bentley Strockbine, Adrian Roitberg, and Carlos Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics*, 65(3):712–725, 2006.

[30] J. Wang, W. Wang, Kollman P. A., and D. A. Case. Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling*, 25:247–260, 2006.

[31] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. A. Kollman. AMBER 12, 2012.

[32] D. A. Case, T. A. Darden, T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M. J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P. A Kollman. *AmberTools12 Reference Manual*, 2012.

[33] Christopher I. Bayly, Piotr Cieplak, Wendy Cornell, and Peter A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry*, 97(40):10269–10280, 1993.

[34] Araz Jakalian, Bruce L. Bush, David B. Jack, and Christopher I. Bayly. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *Journal of Computational Chemistry*, 21(2):132–146, 2000.

[35] Araz Jakalian, David B. Jack, and Christopher I. Bayly. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of Computational Chemistry*, 23(16):1623–1641, 2002.

[36] Wendy Cornell. Charge fitting philosophy. http://ambermd.org/doc6/html/AMBER-sh-19.4.html.

[37] Sishi Tang Ross Walker. Antechamber Tutorial. http://ambermd.org/tutorials/basic/tutorial4b/.

[38] Alan W Sousa da Silva and Wim F Vranken. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Research Notes*, 5(367), 2012.

[39] Gerard Adriaan Vliegenthart and Gerhard Gompper. Mechanical Deformation of Spherical Viruses with Icosahedral Symmetry. *Biophysical Journal*, 91:834–841, 2006.

[40] Mareike Zink and Helmut Grubmüller. Mechanical properties of the icosahedral shell of southern bean mosaic virus: A molecular dynamics study. *Biophysical Journal*, 96:1350–1363, 2009.

[41] Mareike Zink and Helmut Grubmüller. Primary changes of the mechanical properties of southern bean mosaic virus upon calcium removal. *Biophysical Journal*, 98:687–695, 2010.

[42] Olivier Michielin and Martin Karplus. Binding Free Energy Differences in a TCR-peptide- MHCComplex Induced by a Peptide Mutation: A Simulation Analysis. *J. Mol. Biol.*, 324:547–569, 2002.

[43] Yeng-Tseng Wang, Zhi-Yuan Su, Jun-Min Liao, and Cheng-Lung Chen. Potential of mean force for Syrian hamster prion epitope protein–Monoclonal fab 3f4 antibody interaction studies. *European Journal of Medicinal Chemistry*, 44:3504–3508, 2009.

[44] Yeng-Tseng Wang, Zhi-Yuan Su, and Cheng-Lung Chen. Potential of mean force of the hepatitis C virus core protein–monoclonal 19D9D6 antibody interaction. *Biophysical Chemistry*, 145:86–90, 2009.

[45] Deqiang Zhang, Justin Gullingsrud, and J. Andrew McCammon. Potentials of Mean Force for Acetylcholine Unbinding from the Alpha7 Nicotinic Acetylcholine Receptor Ligand-Binding Domain. *J. Am. Chem. Soc.*, 128:3019–3026, 2006.

[46] Helmut Grubmüller, Berthold Heymann, and Paul Tava. Ligand Binding: Molecular Mechanics Calculation of the Streptavidin-Biotin Rupture Force. *Science*, 271, 1996.

[47] H.Ariel Alvarez, Andrés N. McCarthy, and J.Raúl Grigera. A Molecular Dynamics Approach to Ligand-Receptor Interaction in the Aspirin-Human Serum Albumin Complex. *Journal of Biophysics*, 2012.

[48] Rommie Amaro and Zaida Luthey-Schulten. Molecular dynamics simulations of substrate channeling through an alpha-beta barrel protein. *Chemical Physics*, 307:147–155, 2004.

[49] Deqiang Zhang, Justin Gullingsrud, and J. Andrew McCammon. Potentials of Mean Force for Acetylcholine Unbinding from the Alpha7 Nicotinic Acetylcholine Receptor Ligand Binding Domain. *J. Am. Chem. Soc.*, 128(9):3019–3026, 2006.

[50] Hyung-June Woo and Benoît Roux. Calculation of absolute protein–ligand binding free energy from computer simulations. *PNAS*, 102(19):6825–6830, 2005.

[51] Alicia Claudia Lorenzo and Ernesto Raúl Caffarena. Elastic properties, young's modulus determination and structural stability of the tropocollagen molecule: a computational study by steered molecular dynamics. *Journal of Biomechanics*, 38(7):1527–1533, 2005.

[52] Sanjay Kumar and Mai Suan Li. Biomolecules under mechanical force. *Physics Reports*, 486(1-2):1–74, 2010.

[53] Anton Arkhipov, Wouter H. Roos, Gijs J.L. Wuite, and Klaus Schulten. Elucidating the mechanism behind irreversible deformation of viral capsids. *Biophysical Journal*, 97(7):2061–2069, 2009.

[54] D.C. Rapaport. Molecular dynamics simulation of reversibly self-assembling shells in solution using trapezoidal particles. *Physical review E*, 86, 2012.

[55] Eric D. Horowitz, K. Shefaet Rahman, David J. Dismuke Brian D. Bower, Michael R. Falvo, Jack D. Griffith, Stephen C. Harvey, and Aravind Asokana. Biophysical and ultrastructural characterization of adeno-associated virus capsid uncoating and genome release. *Journal of Virology*, 87(6):2994–3002, 2013.

[56] Donald K. Phelps, Peter J. Rossky, and Carol B. Post. Influence of an antiviral compound on the temperature dependence of viral protein flexibility and packing: a molecular dynamics study. *Journal of Molecular Biology*, 276(2):331–337, 1998.

[57] Justin A. Lemkul and David R. Bevan. Assessing the Stability of Alzheimer's Amyloid Protofibrils Using Molecular Dynamics. *The Journal of Physical Chemistry B*, 114(4):1652–1660, 2010. PMID: 20055378.

[58] APeter J. Ortoleva Yinglong Miao. Viral structural transition mechanisms revealed by multiscale molecular dynamics/order parameter extrapolation simulation. *Biopolymers*, 93(1), 2009.

[59] T. Baştuğ and S. Kuyucak. Application of Jarzynski's equality in simple versus complex systems. *Chemical Physics Letters*, 436:383–387, March 2007.

[60] Sanghyun Park, Fatemeh Khalili-Araghi, Emad Tajkhorshid, and Klaus Schulten. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *The Journal of Chemical Physics*, 119(6):3559–3566, 2003.

[61] Po-Chia Chen and Serdar Kuyucak. Accurate Determination of the Binding Free Energy for KcsA-Charybdotoxin Complex from the Potential of Mean Force Calculations with Restraints. *Biophysical Journal*, 100(10):2466 – 2474, 2011.

[62] Mai Suan Li and Binh Khanh Mai. Steered Molecular Dynamics-6 Promising Tool for Drug Design. *Current Bioinformatics*, 342(7):342–351, 2012.

[63] Turgut Baştuğ, Po-Chia Chen, Swarna M. Patra, and Serdar Kuyucak. Potential of mean force calculations of ligand binding to ion channels from Jarzynski's equality and umbrella sampling. *The Journal of Chemical Physics*, 128(15):–, 2008.

[64] Sergei Izrailev, Sergey Stepaniants, Barry Isralewitz, Dorina Kosztin, Hui Lu, Ferenc Molnar, Willy Wriggers, and Klaus Schulten. Steered Molecular Dynamics. In Peter Deuflhard, Jan Hermans, Benedict Leimkuhler, Alan E. Mark, Sebastian Reich, and Robert D. Skeel, editors, *Computational Molecular Dynamics: Challenges, Methods, Ideas*, volume 4 of *Lecture Notes in Computational Science and Engineering*, pages 39–65. Springer Berlin Heidelberg, 1999.

[65] Wolfgang Lechner and Christoph Dellago. On the efficiency of path sampling methods for the calculation of free energies from non-

equilibrium simulations, journal=Journal of Statistical Mechanics: Theory and Experiment. 2007(04):P04001, 2007.

[66] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[67] mono2poly script for VMD. `http://www.ks.uiuc.edu/Research/vmd/script_library/scripts/mono2poly/`.

[68] Baker NA, Sept D, Joseph S, Holst MJ, and McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci.*, 98:10037–10041, 2001.

[69] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox. Gaussian09 Revision D.01. Gaussian Inc. Wallingford CT 2009.

[70] Anna Ruokonen. KIRTAN 1:n epäkovalenttinen sitoutuminen echovirus 1:n hydrofobiseen taskuun. Research Project, University of Jyväskylä, Department of Chemistry, 2014.