University of South Bohemia in České Budějovice
Faculty of Science

Modelling of tick-borne encephalitis virus (TBEV) proteins and their complexes

Bachelor thesis

Marco Halbeisen

Supervisor: Mgr. Roman Tůma, Ph.D.

Co-supervisor: RNDr. Zdeněk Franta, Ph.D.

České Budějovice 2022

Halbeisen M., 2021: Modelling of tick-borne encephalitis virus (TBEV) protein and their complexes. Bc. Thesis, in English 57 p., Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation:

In this bachelor thesis the focus was on the study of non-structural proteins of the tick-borne encephalitis virus involved in the replicatory process of the virus. Non-structural protein number 3 (NS3) and number 5 (NS5) contain a helicase and an RNA-dependent RNA polymerase domain, respectively, and play a crucial role in the life cycle of the virus. Therefore, they are of interest to develop a potential antiviral drug inhibiting the replication of the pathogen. It has been found that these two proteins form a dimer during the replication of the viral RNA. The aim of the thesis was to identify the exact interaction of the two proteins with existing hydrogen-deuterium mass spectroscopy data as well as modelling the RNA into the complexes. The research was conducted by using computational biology tools including homology modelling using MODELLER, manual modelling in PyMOL and Chimera as well as Molecular Dynamics Simulations in GROMACS.

Declaration:

I hereby declare that I have worked on my bachelor's thesis independently and used only the sources listed in the bibliography. I hereby declare that, in accordance with Article 47b of Act No. 111/1998 in the valid wording, I agree with the publication of my bachelor thesis, in full to be kept in the Faculty of Science archive, in electronic form in publicly accessible part of the STAG database operated by the University of South Bohemia in České Budějovice accessible through its web pages. Further, I agree to the electronic publication of the comments of my supervisor and thesis opponents and the record of the proceedings and results of the thesis defence in accordance with aforementioned Act No. 111/1998. I also agree to the comparison of the text of my thesis with the Theses.cz thesis database operated by the National Registry of University Theses and a plagiarism detection system.

České Budějovice, 11.05.2022

……

Marco Halbeisen

Acknowledgment:

I would like to express my sincere gratitude to both of my supervisors Mgr. Roman Tůma, Ph.D. and RNDr. Zdeněk Franta, Ph.D. for their support during this work. Their effort to assist me in my professional development while working on my thesis went well beyond their duties. Mgr. Roman Tůma, Ph.D was like a mentor trying to introduce me into the field of science and often taking several hours a day explaining concepts to me. He made me feel like having my private teacher. I could ask him anything at any time.

Furthermore, both supervisors enabled me to pursue further projects in their lab which will open up a variety of opportunities in the future I would not have otherwise. Here, I would like to especially thank RNDr. Zdeněk Franta, Ph.D and Paulina Duhita Anindita Ph.D who sacrificed a lot of time explaining and demonstrating different techniques in the lab despite their own high workload.

In addition, I would like to thank the whole lab team supporting me with useful information in the lab as well as creating a uniquely friendly atmosphere which made going to work feel like meeting friends.

Finally, I would like to mention various cycling trips with Roman, Joel, Tom and Zdenek. Thank you for this amazing time I'll never forget!

Table of contents:

1. Introduction

1.1 Tick-borne encephalitis virus

In recent years, the attention of the tick-borne encephalitis virus (TBEV) and its associated research became the most important among tick-borne diseases in Europe, Asia as well as in the Far East. This thesis focusses exclusively on the European strain, however, due to their similarities the findings might be applicable to all strains. TBEV is part of the *Flaviviridae* family. They are characterized by a single strand positive RNA. The total length of the genome is around 11kb. The entire genome is transcribed and translated into a single polyprotein and further post-translationally cleaved into 7 non-structural (NS) and 3 structural proteins. The virus is transmitted by a variety of different ticks, with Ixodes Ricinus being the most common host in central Europe. An infection of the European strain usually manifests itself in two distinct clinical phases. The primary one includes flu-like symptoms, whereas in the second TBE can lead to a variety of life-threatening disease by breaching the blood brain barrier (BBB).[1]

The development of a vaccine against tick-borne encephalitis (TBE) has already started in the 1930s. There exists a variety of different vaccines developed by different countries. They are formaldehyde-inactivated vaccines with efficient cross-protection against various TBEV strains. Despite the potential treatment, TBE is the most common tick-transmitted disease in Europe. This can be explained by the TBEV vaccine coverage which - in most European countries - is only between 10-20% of all population with some exceptions like Austria having a coverage of around 88%. Furthermore, global warming enhances the risk of the virus to establish in colder regions as well.[2] Up to this date no antiviral has proven to be effective. The potential of nucleoside triphosphate (NTP) analogs interacting with the RNA dependent RNA polymerase domain and S-Adenosylmethionine (SAM) binding to the Mtase domain of the non-structural protein 5 proved to have an inhibiting effect on the virus replication but so far no candidate has made it through the clinical trials.[3,4] Further research must be conducted in order to find a specific anti-viral drug.[5]

1.2 Replication of TBEV

Most of the life cycle of TBEV takes place in the tick, however, there is little data available about the exact steps inside the tick. This data is mostly derived from mammalian cell studies. The life cycle of TBEV involves a complex assembly of the viral proteins accompanied by various maturation steps which are depicted in Figure 1. [6]

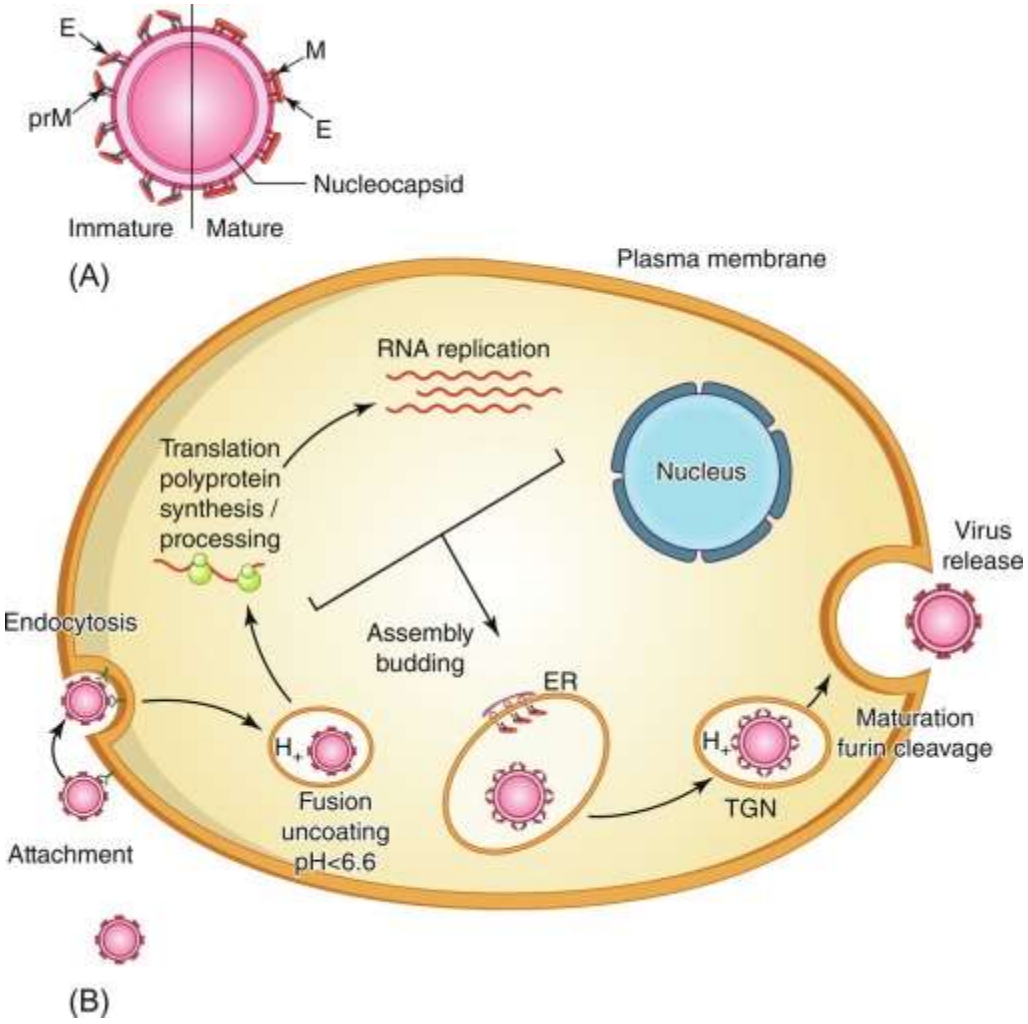

Figure 1: Figure A shows the structural organization of the virus. Figure B is a schematic representation of the replication cycle of TBEV within a mammalian cell. It shows the endocytosis, fusion uncoating involving the acidic pH, RNA release, translation and replication with the assistance of non-structural protein at the endoplasmic reticulum membrane. Finally, the viral assembly is followed by the transport through the trans-Golgi network (TGN) leading to the virus maturation and release. [7]

1.2.1 Entry

The entry of the flavivirus is triggered by receptor mediated clathrin dependent endocytosis on the mammalian membrane via the heparan sulfate receptor. Furthermore, micropinocytosis constitutes another potential way for the virus to enter the host cell. The virion is engulfed by an endocytic vesicle called the endosome which lowers the pH and by this triggers alteration of the viral particles. The formation of a pre-fusion trimer of envelope (E) proteins conformationally changes into a hairpin bringing the endosome membrane and the viral proteins into direct contact creating the post-fusion trimer. This hemi-fusion intermediate releases the nucleocapsid (NC) proteins into the cytosol. Additionally, it leads to the disintegration and release of the viral RNA. the uncoating events that are poorly understood. [8]

1.2.2 Translation and replication

The translation of the released RNA takes place at the endoplasmic reticulum (ER) membrane by host cell ribosomes. The RNA genome consists of a 5' cap and 5' untranslated region (UTR) followed by a single open reading frame (ORF) and ending with a 3' UTR. These 5' and 3' untranslated regions form conserved secondary structures needed for 5' and 3' long range interactions and subsequent replication. The translation product is a single polyprotein consisting of capsid (C) - precursor membrane protein (prM) – envelope protein (E) - NS1 - NS2A - NS2B - NS3 - NS4A - NS4B - NS5 which is cleaved by both host and viral proteases. Most proteins are non-structural (NS) proteins constituting the replication complex. They are either integral membrane proteins or have membrane anchors. Additionally, they form complexes with each other to ensure the correct transcription, translation, maturation and assembly of the viral proteins which is visualized in Figure 4. [9] But the composition, structure and conformational changes of these complexes remain elusive.

1.4 Role of NS3 and NS5 in replication

The *flaviviridae* require a minus as well as a plus strand RNA synthesis. Furthermore a 5' capping and methylation is crucial for the genome expression. The replication complex consists both of host and viral proteins which are embedded on the cytosolic part of the ER (Figure1). Two fundamental components of this machinery are the non-structural proteins number 3 and 5. They are of particular interest possessing the enzymatic capability to replicate the viral genome.[10] While NS3 consists of a C terminal helicase/RTPase domain (NS3H) and a N terminal serine protease domain. NS2B is necessary for the NS3 protease activity acting as a

cofactor required for cleavage between the NS proteins.[11] The NS3H domain serves three main purposes including the helicase activity which unwinds and separates double stranded RNA (replication intermediate), an RNA stimulated nucleoside triphosphate hydrolase (ATPase) and 5' RNA triphosphate activities which are involved in the 5' cap formation. [12] NS3H can be divided further into 3 subdomains consisting of seven conserved motifs situated in subdomain one and two. The interface of the three subdomains forms the RNA binding groove. The entry site from which the ssRNA is fed into the binding tunnel is at the interface of subdomain 2 and 3. The RNA form contacts with alpha helices 2, 7(subdomain 2) and alpha helix 6 (subdomain 3). The (+) ssRNA in contrast to the (-) ssRNA moves along the outside of the NS3H near the junction of subdomain 2 and 3. [13] The coupling between the RNA binding site which is approximately 30Å away from the NTPase/RTPase site is still poorly understood. Furthermore, lack of static structures to convey information about the motion involved in translocation of the RNA creates a need for approaches giving information on the dynamics.[14] Furthermore, residues G198, K199 and T200 of the Walker A motif in addition to residues of the Walker B motif located in the NTPase/RTPase are crucial for the binding of Mg^{2+} ion necessary for the hydrolysis to occur. [15] A visual representation of the apo structure of the NS3H (PDB: 7AYA, 1,83A) is depicted in Figure 2.

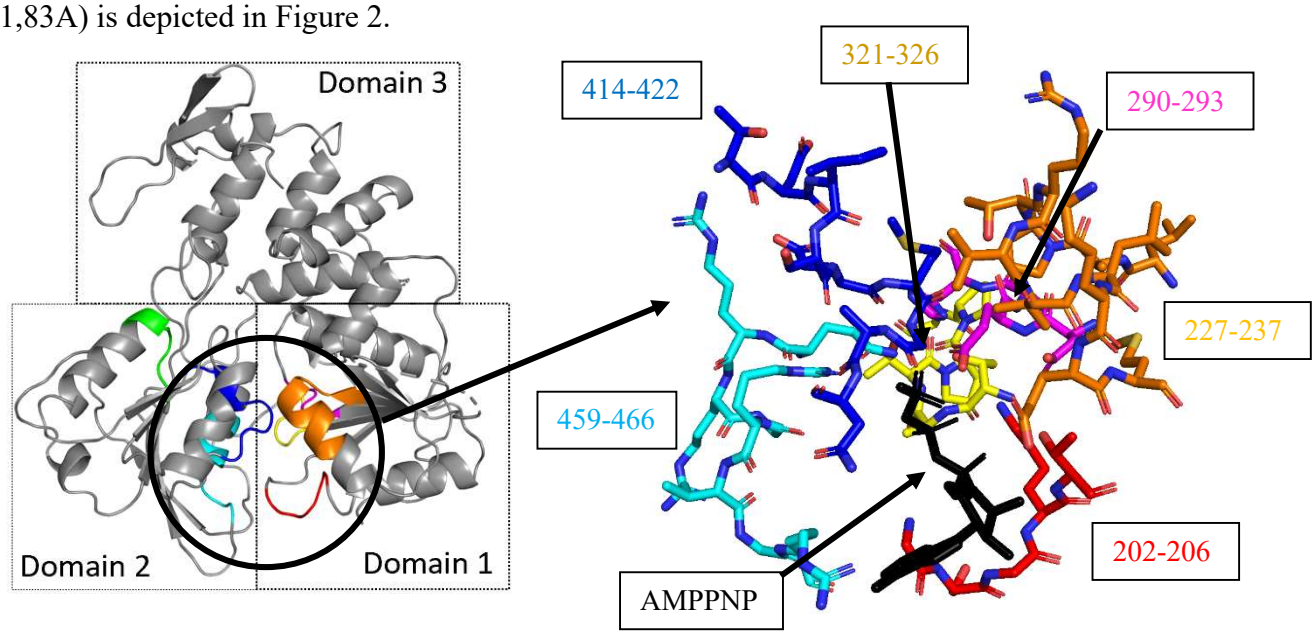

Figure 2: Showing the Subdomains 1,2 and 3 of the NS3H of TBEV. The division into the different motives 1-6 and 1a is colored in red, magenta, yellow, green, dark blue, orange and light blue, respectively. In the close look-up of the ATPase domain AMPPNP represented in black is bound to the ATPase site (PDB: 7BM0). The surrounding residues were hidden for visual simplicity. The residues were labelled with their residue number according to the apo structure (BDB 7AYA) using the corresponding color they are represented in.

NS5 on the other hand is the largest non-structural protein of encephalitis virus with a size of 103 KDa. It consists of two domains namely the N-terminal methyltransferase (MTase) and the C terminal RNA dependent RNA polymerase (RdRp) domain. The latter is depicted in Figure 3. The RdRp is capable of both (-) and (+) strand synthesis. [16] The MTase has a methyltransferase activity, however, it also exhibits a RNA guanyl transferase activity, both involved in 5' RNA capping and cap methylation. [17] The addition of a type one cap at its 5' end (m7GpppNm-RNA). The (-) strand which is only found as a double stranded RNA (dsRNA) in infected cells serves as a template for the (+) strand RNA genomic RNA. The dsRNA intermediate undergoes complementary strand synthesis by the RdRp immediately after being unwound by NS3H resulting in another dsRNA and an additional (+) ssRNA. [18] The 5' triphosphate end of the (+) ssRNA is converted into 5' diphosphate by the RNA triphosphatase activity of NS3H. Secondly, the guanyl transferase (GTase) activity of the NS5 MTase domain transfers GMP from GTP to the 5' diphosphate RNA.[19] Thirdly, a methyl group is first attached to the N7 position of the previously attached guanosine followed by an additional methyl group added to the 2'-OH of the ribose. This is again mediated by the NS5 MTase activity of NS5. The cofactor supplying the methyl group in both reactions is S-adenosyl-L-methionine (AdoMet).[20] In contrast to NS3 there is no crystal-structure available for the European strand of NS5 RdRp, however, a crystal structure with a resolution of a 3.17 Å of the RdRp and MTase of the eastern strand of TBEV has recently been published.[21]

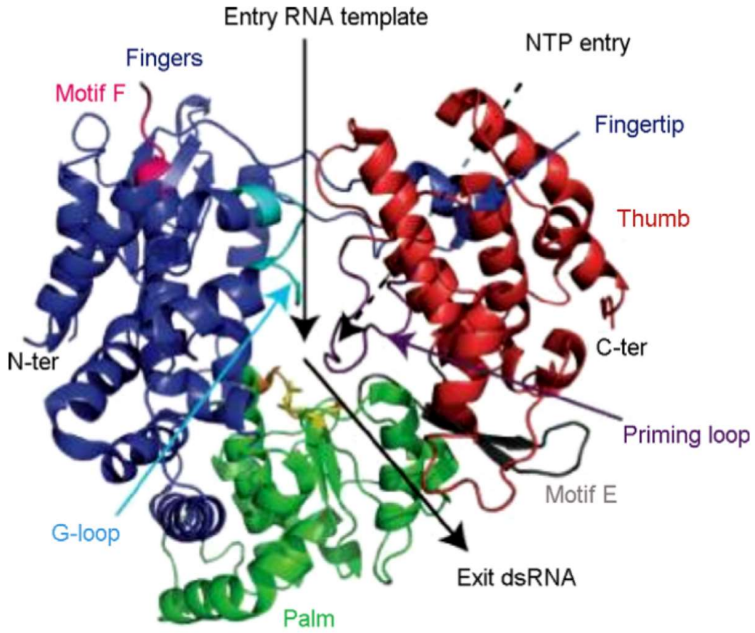

Figure 3: Showing a representation of a RdRp domain of the NS5 of flaviviruses. It is divided into three domains: The palm, the fingers and the thumb, colored green, blue and red, respectively. The entry and exit of the RNA as well as the NTP are binding site are indicated by a black arrow. Furthermore, the G-loop, the priming site as well as motive E and F are labelled.[22]

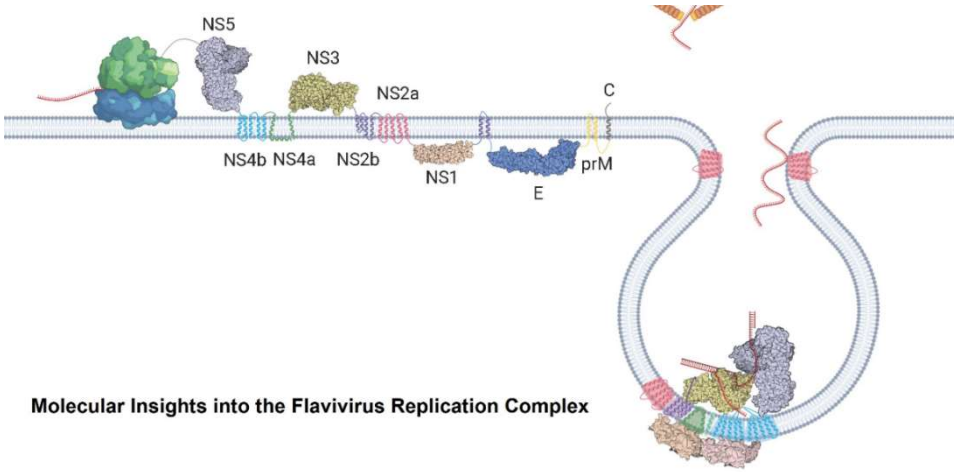

Figure 4: On the left side the polyprotein translation by a ribosome (green/blue) and its non-cleaved products along the membrane. The invagination of the membrane of the endoplasmic reticulum and the resulting interaction between the NS3 (green) and NS5 (grey) and the RNA is shown on the bottom right. They furthermore interact with nonstructural proteins NS4b (light blue) and NS2b (purple)[23]

1.5 Protein-protein interactions between NS3 and NS5

Both capping and RNA synthesis of (+) ssRNA requires NS3 and NS5 which are supported by other NS proteins. NS2B, NS4A and NS4B are cofactors by interacting via their hydrophilic regions with NS3 and also contain transmembrane regions to anchor NS3 into the ER membrane. [24] However, there is no reported interaction between any NS protein and the NS5 except for NS3, thus making it the only known anchor of the polymerase to the membrane. [10]

The complex formation of NS3 and NS5 has been validated using multiple techniques in vitro, including ELISA, surface plasmon resonance and pull-down experiment, as well as in vivo by employing co-immunoprecipitations.[25–28] This complex formation is independent of all other NS proteins. [25] Furthermore, the C-terminus of RdRp seems to be essential alongside with the helicase subdomain 3 for the complex formation. In addition, the NS5 thumb subdomain and particularly Lys-330 has shown to critical in stabilizing the interaction. [29] Nevertheless, the mentioned studies above do not address the role of RNA, if any, in this interaction. Likewise, the interplay and regulation of activities within the complex remains elusive due to the lack of structural information. Therefore, further research including full-length NS3 and NS5 proteins must be conducted to fully understand the interaction in positive strand synthesis. [13]

2. Aim of the study

The aim of this thesis was to further elucidate the working principle of the replication complex of TBEV focusing on the NS3 and NS5. The primary goal was to build a homology model for the European strand of the RdRp NS5. The result of this first aim was then taken and equilibrated using molecular dynamics (MD) to serve as a starting point for the second aim.

The interaction between NS3 and NS5 in TBEV is poorly understood. Using hydrogen deuterium exchange (HDX) coupled mass spectrometry (MS) data combined with a high ambiguity driven protein-protein docking approach, the interaction of those two proteins can be predicted and the model critically tested.

The third aim was to get a better understanding of the coupling between the RNA binding groove and the ATPase site of NS3 helicase. To pursuit these different components were modelled into the protein manually using PyMOL based on existing data of related flaviviruses as well as recently resolved TBEV NS3H X-ray structures provided by Dr. Anindita P., PDB codes: 7BLV, 7BM0, 7AY4The changes in the protein and ligands in a Newtonian based forcefield can be modelled using molecular dynamics (MD) simulations. Using the simulation, a working hypothesis about the mechanism of ATPase driven RNA translocation was made and tested experimentally by site directed mutagenesis.

3. Material and methods

This section covers both the computational methods' physical background used for homology modelling, docking of proteins and molecular dynamics simulations and the experimental procedure performed to test the hypothesis.

3.1 Computational Biology

In recent years in silico methods have become a valuable tool for drug design and the study of molecular mechanisms. While it is far from replacing experimental biology it still can serve as a valuable tool to hint potential mechanisms of actions happening at a molecular level. These mechanistic hypotheses can further be validated experimentally. The variety of computational programs gives further insight into the molecular world is growing day by day.

However, most theoretical approaches are mostly based on energy calculations of the system in question. While there exist simulation programs which aim to predict the interaction between molecules either by quantum mechanics or classical molecular mechanics based on Newtonian physics. Furthermore, machine learning approaches for the prediction of protein folding, protein-protein or protein-ligand interaction based on experimental data like Alphafold or Rosetta are getting more reliable year by year. [30,31]

3.1.1 Homology Modelling

A variety of homology-based protein modelling approaches exist both online and available for download. MODELLER is a comparative protein structure modelling which can predict the unresolved tertiary structure of a proteins by using phylogenetically related solved structures.[32] The prediction process is based on four main steps starting with the fold assignment. In this step related structures on the protein data base are scanned for similarity between the target and one or more template structures. Secondly, the template gets aligned with the target sequence. This is followed by the model building process based on the chosen template. Finally, the quality of the created model is evaluated by predicting the model errors.[33]

The fold assignment is done by the *SALIGN* command in MODELLER. The test set consists of 200 structures using pairwise alignments which have a sequence identity above 40%.[34] The alignment between the target and the template sequences is calculated by creating a scoring

matrix which gets optimized in multiple steps. A gap penalty as well as a pairwise matrix corresponding to the substitution are applied to score the alignment by adding both components. This process is commonly based on a dynamic programming approach. [35] MODELLER uses the BLOSUM62 substitution matrix by default. The cutoff for the E-values, which is a measure of the significance of the alignment based on sequence identity and statistical measures, for sequences is smaller or equal to 0.01. [33]

The template selection uses the first six lines of the alignment's output which share the highest sequence identity with the sequence of interest. The structural and sequence similarities are evaluated by the command *alignment.compare_structures ()* which assesses the alignment of the six sequences. The result is the basis of a multiple structure alignment executed by *malign3d(),* afterwards, the RMS and DRMS are calculated based on this structural alignment using *compare_structures().* Furthermore, pairwise sequence distances are written down and used as an input matrix to the *dendrogram ()* command calculating the clustering tree from the pairwise distances of the input matrix to visualize the differences between the six potential templates from the alignment output. Based on factors like the template structure's resolution and structural divergence among the templates, the most appropriate template can be chosen.[33]

The alignment with the chosen target sequence is usually performed by *align2d ()* taking both structural and sequence information into account. Additionally, a gap penalty function places gaps in solvent-exposed or loops. The output is in PIR format which is used for the following model-building stage while the secondly created PAP alignment file is for visual inspection.[36]

The 3D model is constructed automatically by MODELLER using its auto model class. The provided script will create a specified number of homology models. In this process the PIR format of the target-template is used where the function knows defines the template in the alignment file. The default modeler command to create homology models is called *make* creating output files in PDB format as well as a LOG file, which, besides reporting errors, also provides information on the restrains within the created structure. [33]

The lowest discrete optimized protein energy (DOPE) score corresponds to the best model since all the created models are calculated in reference to the same template structure. To assess the quality of the model on residue basis an evaluation with the DOPE potential can be performed. The energy profile can be visualized by graphic programs like GNUPLOT.[37]

Four potential candidates for homology modelling were available on the protein data base (PDB). For each of them 100 models were generated and the one with the lowest DOPE score was chosen. The plotted energy profiles between best the models of the templates were compared. To further reduce the DOPE the subdomain 1 and 2 of the YFV RdRp (PDB:6QSN) and the third subdomain of the JEV RdRp (PDB:5TFR) were selected in PyMOL and the remaining parts were deleted. They shared common residues which connected their subdomain 2 and 3 which enabled to manually combine the chosen autonomously folding units with each other using the edit function in PyMOL Throughout this manual modelling inter-residue clashes were introduced which led to sharp peaks in the Energy diagram. YASARA Energy Minimization Server was used to relax the hybrid model and eliminate the steric clashes. The relaxed model was then further equilibrated using molecular dynamics with explicit solvent for 500ns and is referred to as hybrid model. Meanwhile a crystal structure of the eastern strand RdRp of TBEV was published (resolution of 3.17 Å) [21] Homology models were then created for this newly resolved structure incorporating the slightly varying sequence in comparison to the Central European strand of TBEV. The resulting homology model was equilibrated again using GROMACS for 500 ns.

3.1.2 High ambiguity driven protein-protein docking

High ambiguity driven protein-protein docking (HADDOCK) is a software available online under the HADDOCK web portal. It predicts plausible protein-protein interactions as well as protein-ligand interactions using any experimental input giving information about the binding residues between two molecules such as NMR chemical shift perturbation (CPS) data, cryo-electron microscopy density maps or hydrogen deuterium exchange (HDX) coupled mass spectrometry (MS) data and cross-linking. This data-driven approach performs significantly better than the purely ab-initio docking programs. The interface provides opportunity to define up to 500 parameters including distance restraints (e.g. distance specific cross-linking). According to the defined input the pre-defined residues are forced to be at a specific distance to each other while maintaining contact by electrostatic interactions and hydrogen bonding. [38] The apo NS3 x-ray structure (PDB:7AYA, 1,83A) was combined with a TBEV NS5 homology model based the recently resolved NS5 RdRp of the eastern strand (PDB:7D6N, 3,17A) using MODELLER. The intermolecular interactions in this complex were previously determined by hydrogen deuterium exchange (HDX) by Dr. Filip Dycka in our research group and are summarized in Table 4. This data was used for the HADDOCK web server using standard

parameters which are summarized in the Appendix in Table 7.

3.1.3 Molecular Dynamics Simulation

The study of many-particle systems can be performed numerically by molecular dynamics (MD) simulation in programs such as GROMACS. This thesis focuses only on Newtonian based molecular dynamics simulations (i.e., classical MD). Classical MD poses several advantages over quantum mechanical, ab-initio calculations when it comes to larger, multicomponent systems, for instance, the complete knowledge of every particle of the system at any given time during the simulation. Furthermore, Newtonian based simulation can deal with non-equilibrium states and complex situations beyond the thermal equilibrium. However, this advantage can also mean a big disadvantage as the knowledge of every trajectory at any given time is hardly ever needed. This creates an enormous amount of data which can be seen as redundant. Moreover, MD simulations have an incomplete ergodicity meaning the sampling of configurations corresponding to higher energy states is less likely to be covered in the simulation time in comparison to low energy states. Therefore, the simulation outcome heavily depends on the configuration of the initial state i.e. the input structure leading to biased simulations results.[39] Furthermore, the most obvious disadvantage is the incorrect description of systems if quantum mechanical behavior is of importance i.e. where electron and proton transfer takes place like in bond formation and breaking .[40]

3.1.3.1 Classical Molecular Dynamics Simulations

The first and probably most important simplification is made by the assumption that atoms and ions behave based on the classical model and the quantum description can be neglected. One exception is the quantum mechanical description in the potential of partial charges which are based on quantum mechanical electron densities which are defined in the input files. This can be justified by comparing the De Broglie wavelength and compare it with the inter-particle distance. Where $\lambda = \frac{h}{p}$ describes the wavelength lambda with Plank's constant divided by the momentum of the particle which has a kinetic energy that can be described by $E_{kin} = \frac{p^2}{2m}$ and can be approximately described by $\frac{3}{2} * K_b * T$, where K_b is the Boltzmann constant and T describes the temperature in Kelvin and m the mass of the particle. Solving the equation for

the momentum and plugging this into the wavelength results in $\lambda = \frac{h}{\sqrt{3*m*K_b*T}}$. If for instance argon gas is taken which has a mass of approximately 40 a.u. and the temperature for a typical biological system is assumed to be 298,15 K solving the equation results in lambda being ~0.1Å. Comparing this result to typical distances of approximately 2-3Å it may thus be concluded that the motion of atoms in a simulated system containing atoms and ions is indeed classical. There do exist some exceptions to this rule like the behavior of protons at a specific pH which will not be dealt with in this work. This, however, does not apply to the interaction between molecules and atoms in terms of their potentials and forces as a result of the electronic structure. To be able to solve the electronic structure quantum physics is required. In classical molecular dynamics these interactions are approximated by classical model potentials based on empirical potentials using quantum mechanical calculations as an input. These potentials are approximated by the forcefield applicable for the system of interest. It results in a purely classical all atom many-particle computation.[41]

3.1.3.2 Energy calculations

The kinetic energy of the system is described by the sum of kinetic energies of N particles at their position $\underset{r_i}{\rightarrow}$ and their corresponding velocity $\underset{v_i}{\rightarrow} = \frac{\vec{p_i}}{m}$. The resulting kinetic energy can be described as $E_{kin} = \sum_{i=1}^{N} \frac{\vec{p_i^2}}{2m_i}$. In contrast, the potential energy is composed of the interaction between particles and the external potential $E_{pot} = v_{int}(\underset{r_i}{\rightarrow}, \ldots \underset{r_n}{\rightarrow}) + \sum_{i=1}^{N} u_{ext} * \underset{r_i}{\rightarrow}$. The forces acting can be summarized by $F_i = -\frac{d}{d\underset{r_i}{\rightarrow}} v_{int}(\underset{r_i}{\rightarrow}, \ldots \underset{r_N}{\rightarrow}) - \frac{d}{d\underset{r_i}{\rightarrow}} u_{ext} * \underset{r_i}{\rightarrow}$. The electrostatic potential is during the run is initially calculated by summing the charges per slice, followed by integrating twice of the charge distribution. The pair potential is calculated by using the Lennard-Jones potential in which the Van-der-Waals attraction describes the induced dipole-dipole interaction to be proportional to 1/r^6 and the short- range repulsion created by overlapping electron clouds to 1/r^12.

3.1.3.3 Time integration – Verlet algorithm

The Verlet algorithm can be used to simplify the integration of the Newtonian equation of motion and is employed in molecular dynamics programs such as GROMACS. The algorithm uses positions and accelerations from a previous time point (t_{n-1}) to calculate the new position

at time (tn+1) which speeds up the computation. This algorithm does not use explicit velocities. Hence this simplified integration can be used to calculate the trajectories of particles in the simulation with less computational effort and modest storage requirements while maintaining moderate accuracy of position. However, the accuracy of the velocities is low with an error of (O(τ^2)), but explicit velocities are not of much interest in MD interpretation.[42]

3.1.3.4 Geometry and boundary conditions

Macroscopic systems with an extremely high number of particles use periodic boundary condition to approximate large systems by the use of small simulations cells of defined size to include the influence of bulk solution. This box is repeated infinitely many times into all adjacent directions eliminating the boundaries of the system necessary to avoid an artifact in an isolated cluster. Therefore, each particle would interact with all the boxes added which emulates infinite repetition and effectively accounts for long-range interactions (e.g. electrostatics). There exist several possible shapes of the unit cell that are implemented within GROMACS. In this thesis only the standard cubic box was chosen which describes the area getting filled by the solvent molecules. The cut-off radius chosen by GORMACS to truncate the non-bonded interaction does not exceed half of the shortest box vector which is defined as: $R_c < ½ \min(\|a\|,\|b\|,\|c\|)$. This course of action avoids more than one image being within the cut-off distance of the force. In solution an additional restriction is applied where the length of each box vector exceeds the dimension of the macromolecule in the direction of this box's edge and additionally two times the cut-off radius R_c. [43]

3.1.3.5 Input and conditions

The protein and - if present - ligand were converted from a PDB to GRO file GROMACS using the command: *pdb2gmx -i X.pdb -p X.top -o X.gro -ignh* creating the topology file in the process. In the case of nucleic acid ligands, one has to first use a gaff file and compile the TOP (topology) file as well as the GRO file manually. The cubic box was created by calculating the appropriate parameters for the enzyme's size which is here referred to as X. Next, the defined box was solvated by using TIP3P water molecules using a padding water layer thickness of 1 nm. This approximation defines a rigid water molecule consisting of the three atoms with charges assigned to it according to the Lennard-Jones parameters. This model does not fully represent the interaction between the free electron pairs and other molecules by hydrogen

bonding resulting in a decreased viscosity of the solvent.[44] This can be considered an advantage since it drives the processes in the simulation happening more quickly than in real time, provided we are not interested in kinetics.[45] Furthermore, Na^+ or Cl^- ions were added to achieve net zero charge of the whole system. With these molecules and atoms added the system is first relaxed by energy minimization to eliminate e.g. steric clashes that would otherwise generate excess potential energy during MD and destabilize the trajectory. The minimization is done by GROMACS molecular mechanics module. This geometric optimization is further improved by short equilibrations including the solvent and ions surrounding the protein in order to avoid the collapse of the system due to the solvent being only relaxed within itself but not with the protein and ligands. At the same time, the temperature at which the simulation will be run at is introduced by a Berendsen-thermostat until the corresponding kinetic energies are reached (NVT simulation run, duration 100 ps), secondly, pressure is introduced until proper density is reached (NPT simulation run, 100 ps). This sets the stage for starting the actual molecular dynamics simulation. By choosing the required parameters in the md.mdp input file the run can be initiated using *gmx mdrun -deffn md_X.tpr* . [46]

Table 1: Conditions used in Simulations

Forcefield	Box shape	Solvent	Charge neutralization
GROMACS AMBER99SB-ILDN, nucleic AMBER94	Cubic box with at least 1 nm space to the edge	TIP3P 3-point solvent	Na (+) and Cl (-) ions for electroneutrality

For all 25 simulations performed the same conditions were chosen which are summarized in Table 5 in de appendix.

3.1.3.5 Analysis

Before starting with any analysis, we must correct for periodic boundary condition artefacts, e.g. simulated molecules reaching to adjacent boxes.

The root-mean standard deviation (RMSD) of every separately defined element can be calculated from the XTC output and TPR trajectory files. This is a measure of the RMSD to the initial state of the atoms after when initiating the MD simulation. To identify the principle states the molecules fluctuate in between during the simulation the package Bio3D in R-script can be employed to perform principal component analysis. [47] Furthermore, root-mean square fluctuation of residues can be quantified in relation to their average position.

3.2 Site directed mutagenesis

To confirm the results obtained in silico experimentally, site directed mutagenesis was performed on the NS3H. Residues which are involved in the phosphate release mechanism were identified by analyzing the data of the molecular dynamics simulations. Residues D290, Q459 and R463 are located directly in the ATP binding site and have previously been shown to play a crucial role in the hydrolysis of ATP. The primers were already available and they were mutated to serve as a negative control. The residues D296, R274, L539 and D543 are part of the RNA binding groove and were suspected to play a role in the translocation of the RNA which, as the molecular dynamics simulations suggest, is a key part of hydrolysis and subsequent phosphate release. The residue which the RNA interacted with leading to a phosphate release were compared to those interactions present, where no release of the inorganic phosphate was observed after exchanging the ATP with ADP and Pi. The mutation of those was suspected to slow down the translocation of the RNA, and hence, decrease the velocity of the phosphate release. Furthermore, also P229 was chosen to be mutated sitting in between the RNAse and the ATPase site and was suspected to be involved in the coupling between the two active sites. Hydrogen bonding between the residue and the RNA was only observed in 5 frames corresponding to the high RMSD structure. Furthermore, the residue was oriented in different direcitons due to the sliding movement of the RNA in the binding groove. All residues were mutated to the amino acid alanine. The analysis was done manually frame by frame using PyMOL. The chosen amino acids were subsequently mutated. Their effect on the rate of hydrolysis of the enzyme were measured by appropriate assays.

3.2.1 Primer design

The primers for the site directed were designed using the software provided by the New England Biolabs (https://international.neb.com/neb-primer-design-tools/neb-primer-design-tools). The amino acids to be changed were identified in the sequence and specified in the web tool. To ensure that no interaction between the residue and the ligand takes place anymore they were mutated into alanine. The codon in question was replaced by GCT. This codon was chosen since it is the most frequently used triplet to code for alanine in E-coli bacteria. The resulting primers were then ordered online at Generi Biotech and are summarized in table 2 below.

Table 2: The first column provides information on the single letter amino acid code of the wild type followed by the residue number of the amino acid of interest and the amino acid it will be changed

into which in this case is always alanine (A). Furthermore, the primer pairs are denoted forward and reverse. The second column provides the codons used flanking the triplet of interest which itself is intentionally written in lowercase letters. The third column shows the optimal annealing temperature of the primers proposed by the web tool while the final column summarizes the used consensus temperature for each primer.

Residue number	Primer sequence	Calculated Annealing Temperature Primers [°C]	Consensus Annealing Temperature Primers [°C]
D290A forward	GGCAATCATGgctGAGGCCCACT	63	62
D290A reverse	ACCTCCCAATTTTGTCTCC	61	
Q459A froward	TTCCGCTGCCgctCGGCGCGGAA	74, 72	72
Q459A reverse	GCCGTGGTCACTCGCCTG	72	
R463A forward	ACGGCGCGGAgctGTTGGTCGGC	74	72
R463A reverse	TGGGCAGCGGAAGCCGTG	72	
P229A forward	GGTGTTGGCTgctACTCGTGTGG	66	66
P229A reverse	AATGTTCTCAGGCGCCTG	65	
D296A forward	CCACTGGACTgctCCTCACAGCATA	63	64
D296A reverse	GCCTCATCCATGATTGCC	64	
R274A forward	CTATGTCAACgctCGGCTATTACCACAG	58	59
R274 reverse	GTTGCGTGACACATCACA	62	
L539A	CCGACATCTTgctACTCATTGTGAC	56	57

forward			
L539A reverse	AAGTGCTTTCTTTTCTCTTC	57	
D543A forward	CACTCATTGTgctTTTACACCGTGG	58	59
D543 reverse	AGAAGATGTCGGAAGTGC	61	

3.2.2 Mutagenesis and Transformation

The mutagenesis and transformation were performed following the manual of the Q5 Site-Directed Mutagenesis Kit provided by the New England Biolabs. To amplify the plasmid while introducing the desired mutation 1.5 µL of the forward and reverse primer were added having a concentration of 10 µM. Furthermore, 12.5 µL Q5 Hot Start High-Fidelity 2X Master Mix together with the template DNA which in this case was the plasmid vector pET-19b containing the NS3H were pipetted to the solution. The concentration of the template DNA was 1-25 ng. Nuclease free water was pipetted to the mixture to end up with a total volume of 25 µL. To take the different annealing temperatures of the primers into account a gradient polymerase chain reaction (PCR) was performed in a BiometraTone thermocycler from AnalyticJena.

A volume of 1uL of the PCR product was subsequently added to 5 µL 2X KLD reaction buffer, 1µL 10X KLD enzyme mix and 3 µL of nuclease-free water. The reagents were mixed by pipetting up and down and incubated for 5 minutes at room temperature allowing for the phosphorylation, ligation and template removal consisting of kinase, ligase and Dpln.

The transformation was performed using thawed NEB 5-alpha competent E. coli cells. 5 µL of the previous product were added to the cells and flicked 4-5 times and put on ice for 30 minutes. Afterwards, they were heat shocked for 30 seconds at 42°C and placed on ice for 5 minutes. To help the cells grow 250 µL of SOC media were added to the mixture and the shaker Eppendorf Innova S44i was set up at 37°C for 60 minutes at 250 rpm. The NEB 5-alpha were now ready to be plated onto a petri dish containing ampicillin to select the cells containing the pET-19b plasmid which contains an antibiotic resistance gene. The petri dishes were put into an incubator over night at 37°C.

3.2.2 Miniprep and Sequencing

To ensure that the mutagenesis was successful the plasmids were isolated using the NucleoSpin Plasmid for DNA purification. Eight colonies per mutant were picked from the previously incubated petri dishes using a standard pipette tip. Each of them was incubated overnight using 3ml SOC Outgrowth Medium B9020S with ampicillin added to it. The grown cells were pelleted in a standard benchtop centrifuge at 11,000 * g and the supernatant discarded. Secondly, 250 µL of Lysis Buffer 1 was added to the pellets and the cells were resuspended by vortexing. Subsequently, 250 µL of Buffer 2 were added and incubated for 5 minutes at room temperature. Thereafter, 300 µL of Buffer 3 was added and the solution was mixed by inverting the tube 6-8 times. To clarify the lysate the solution was centrifuged for 10 minutes at 11,000 * g at room temperature. In order to bind the DNA NucleoSpin Plasmid/Plasmid (NoLid) Columns were places into a Collection Tubes (2 mL) and the supernatant from the previous steps was decanted onto the columns. The construct was centrifuged at 11,000 * g for 1 minute and the flow through was discarded. To wash the silica membrane and remove any unwanted contaminants 500uL of an additional washing buffer were employed as well as 600 µL of a ethanol supplemented buffer. Both washing steps were performed by centrifuging the tubes at the same speed and duration as before. To dry the silica membrane an additional centrifugation step must be performed at 11,000 * g for two minutes. To elute the DNA the NucleoSpin Columns were placed into a standard 1.5 mL microcentrifuge tube and 50 µL of nuclease free water was added onto the column and incubated for one minute before elution. The tubes were centrifuged at 11,000 * g for 1 minute. Finally, the concentration of the miniprep product was measured using absorbance at 260 nm and nanophotometer from IMPLEN.

Standard Sanger sequencing was used to confirm the presence of the mutation in the sequence. The samples were sent to SEQme adding 500 ng of the plasmid and either 25 pmol of T7 promotor or T7 terminator primers depending on where the mutation was located within the sequence. The total volume required from the company if primers were added by the customer was 10 µL.

3.2.3 Maxiprep and NEB 5-alpha storage

Maxipreps of the plasmids carrying the site directed mutant genes were performed using the NucleoBond Xtra Midi isolation kit. The procedure is almost identical to the previously described isolation but with bigger volumes and driven by gravity rather than by centrifugation.

At the beginning 250 mL of E-coli cell culture of each mutant were prepared in an Erlenmeyer flask containing penicillin and incubated at 37°C at 180 rpm overnight. The entire volume of the flask was transferred into a bottle to be centrifuged at 4200 rpm for 20 minutes. The supernatant was poured into the previously used flask. 25 mL of this supernatant were used to resuspend the pellets and subsequently, pipetted into a 50 mL Falcon tube. The tubes were centrifuged once again at the previously set speed and duration. The supernatant was decanted once more leaving only the cell pellets in the Falcon tube. Furthermore, 10mL of resuspension buffer were added to the pellets and tube was vortexed until all the cells were dissolved in the buffer. Afterwards, 10ml of cell lysis buffer were added to break the cells apart. The tube was inverted 5 times to make sure everything got in contact with the buffer and incubated for 5 minutes at room temperature. Meanwhile, 10mL of equilibration buffer were added to the NucleoBond Xtra Column. To neutralize the previously added buffer 10mL of the provided neutralization buffer were added until the blue color turned completely white. This caused the solution to form a slimy, viscous lysate. Now that the lysate was neutralized, and the column equilibrated the solution could be loaded into the filter. The column was washed two times with 10 mL containing the filter and 15ml after the filter had been removed, respectively. The elution of the DNA was performed using 10ml elution buffer. The eluent was collected in a 15 mL Falcon tube. 5 mL of isopropanol at room temperature were added to precipitate the DNA. The flask was then centrifuged at 4200 rpm for 15-30 minutes. The plasmid formed a precipitate. The supernatant was carefully removed by decanting and pipetting. The pellet was transferred into a 2ml tube and and equal volume of 70% ethanol was added. The small microcentrifuge tube was centrifuged once more for 5 minutes at 15000 rpm and the supernatant was removed one more time. The pellet was left to dry at room temperature until all the ethanol had evaporated. Finally, the pellet was resuspended in nuclease free water and the concentration determined in the same manner as done for the miniprep.

To store the transformed NEB 5-alpha cells containing pET-19b plasmid with the desired mutation ROTI Store-Cryo vials were employed. Again, 3 mL of media with antibiotics with the colony containing the desired plasmid were incubated in a shaker overnight. Subsequently, 0.5mL of the media was transferred into the cryo-vials and incubated for 30 seconds while the tube was inverted several times. The liquid was removed to solely leave the blue beads in the vial. The vials were labelled and put into the -80°C freezer for long term storage.

3.2.4 Transformation pET19b into E. coli (DE) Codon plus RIPL

The transformation of the pET-19b plasmid into the expression competent E. coli (DE) Codon plus RIPL cells could be done starting with the maxiprep which had been prepared for each mutant previously. As in the previously performed transformation 20-50 ng of DNA was required. The cells were thawed, and the plasmid calculated amount was added to it and mixed by flicking the tube several times and put on ice for 30 minutes. Thereafter, the cells were heat shocked in water at 42°C for 30 seconds and immediately placed on ice for 5 minutes. 250uL of SOC media was added into the mixture and the microtubes were incubated at 37°C for an hour at 250 rpm. Finally, the cells were plated again, however, this time on a petri dish containing both ampicillin and chloramphenicol. The chloramphenicol resistance gene was contained within the Codon plus cells; therefore, the combination of both antibiotics ensured a successful transformation. The plated cells were incubated overnight at 37°C and could be stored for further use at 4°C.

3.2.5 Pilot Expression for pET19B-NS3 Helicase mutants

The optimal expression temperature and time for each of the mutants could be determined performing a pilot expression. The competent E. coli BL21 (DE) Codon plus RIPL cells containing the desired plasmid were picked from the petri dish and incubated by adding 5ml of Luria-Bertani (LB) media together with both ampicillin and chloramphenicol at 37°C at 200rpm overnight. For each mutant four 25mL culture tubes were prepared and 10mL LB media containing both antibiotics added. To this 500 µL of the overnight culture was added. The tubes were labelled 18°C+, 18°C, 30°C+ and 30°C- referring to the used temperature and induction, respectively. All the tubes were first placed into a shaker set up for 37°C with 220 rpm until they reached an optical density (OD) of 0.5 which roughly corresponded to a 600 nM concentration. At this time point the tubes were placed into the fridge to stop the cell proliferation in the exponential phase. Afterwards, in the tubes labelled with a '+' protein production was induced by adding 10 µL per 10 mL of the galactose analog Isopropyl- β -D-1-thiogalactopyranoside (IPTG). The tubes were either incubated at 18°C or 30°C and 1 mL aliquots were taken at 0, 2 and 6 hours for both temperatures and additionally at 4 hours for 30°C and overnight for the expression at 18°C. These aliquots were centrifuged for 5 minutes at 4200 rpm and the supernatant was removed by pipetting. The remaining pellets were stored in the freezer at -20°C until all timepoints were collected.

To check whether the expressed protein had been folded properly to stay in solution a soluble and insoluble isolation was performed. The collected pellets were thawed and resuspended in 150 µL lysis buffer. Liquid nitrogen was employed to freeze the resuspended cells. A heating block was prepared at 42°C to thaw the cells. This cycle was repeated three times. The lysed cells were then centrifuged down at 14000rpm for 10 minutes at 4°C. The supernatant was transferred into new tubes and 50 µL of 4 times SDS-PAGE sample buffer was added. The remaining pellet was dissolved in 200 µL 1 times SDS-PAGE sample buffer which had been diluted by phosphate buffered saline (PBS). The tubes were heated on the heating block at 95°C for 5 minutes. At this point all proteins were denatured and prepared to be loaded onto the SDS-PAGE. The resulting gel indicated whether the protein remained in the soluble or insoluble fraction as well as the rate of production at the different temperature and timepoints could be judged semi-quantitatively.

3.2.6 Large-scale protein expression

Once the optimal temperature for each of the mutants had been determined the large-scale protein expression was performed. The transformed E. coli BL21 (DE) Codon plus RIPL cell containing the desired mutant was picked from the petri dish and added to 100 mL of LB media in an Erlenmeyer flask and grown at 180 rpm at 37°C overnight. The next day this volume was transferred into 700 mL media to end up with a total of 800mL. This was performed twice for a total of 1.6 L per mutant divided into two bottles. The bottles were incubated again at 37°C at 180 rpm until they reached an OD of 0.5. Once they reached the value the temperature got adjusted to the optimal expression temperature and a volume of 800 µL of IPTG was added to initiate the expression. After incubating them for the previously determined duration the volume was transferred into bottles and centrifuged for 30 minutes at 4200 rpm. The supernatant was removed in a similar manner as in the maxiprep and resuspended in a volume of 25 mL. This was transferred into a 50 mL Falcon tube for each of the two bottles. The tubes were centrifuged again for the same duration and speed. The supernatant was removed, and the pellets could be stored in a -80°C freezer until the purification was performed.

3.2.7 Protein purification

For purification, a series of chromatographic techniques was employed after the cells had been lysed. The pellets were thawed on ice for 30 minutes and 10 mL lysis buffer, DNAse, and RNAse were added. To dissolve the cells, they were pipetted up and down until everything was dissolved. Afterwards, a LM20 French Press from Microfluidizer was employed at 20000 psi to lyse the cells. This was repeated for three times and collected in an additional Falcon tube. The volume was then centrifuged using a CP90WX Ultracentrifuge from Hitachi Koki Himac. The rotor is specified to hold the desired volume and the centrifuge was set for 25000 rpm. The tubes had to be carefully weighed using an analytic scale to avoid disbalance between metal tubes holding the samples. The vacuum was established, and the samples were spun for an hour. The protein of interest should now be in solution. Therefore, the supernatant was transferred into a Falcon tube for further purification.

The first chromatographic technique chosen to separate the Helicase from any unwanted proteins is affinity chromatography. The purification was performed using the ÄKTA Pure chromatography system. The column provided by the company for purifying the protein with an attached histidine tag was HisTrapTM HP with a column volume of 5ml and column dimensions of 1.6 x 2.5 cm and a column hardware pressure limit of 0.5MPa. The matrix is made of highly cross-linked spherical agarose (6%). The metal ion capacity is ~15umol Ni2+/mL medium. The dynamic binding capacity is at least 40 mg of histidine tagged protein/mL medium. The chosen flow rate for binding is 2 mL/min while washing was performed at 5 mL/min. Two buffers were prepared with the binding buffer containing 20mM Na-HEPES, 500 mM NaCl at a pH of 7. The elution buffer contained 20mM of Na-HEPES with 500mM NaCl and 1M imidazole at a pH of 7.

After the column was washed with five volumes of water and binding buffer and afterwards eluted with 100% elution buffer to make sure no contaminations from previous runs were bound to the column. Furthermore, 5 more column volumes of binding buffer were run to equilibrate the column. The proteins were loaded onto the column and 5% of the elution buffer was added to wash the column. The concentration of the elution buffer was subsequently increased to 60% elution buffer to elute the protein of interest. All fractions containing proteins which lead to increases in the measured UV absorbance were collected. The concentration was then further increased to 100% to elute everything bound to the column and to prepare the matrix for further use.

The second technique employed for further purification of the desired enzyme was ion

exchange chromatography. The column chosen for this chromatography was a HiTrapTM 5 mL column with column dimensions of 1.6 x 2.5 cm with a column hardware pressure limit of 0.5MPa. The total ionic capacity was 0.15 – 0.20 mmol (H+/mL medium) while the dynamic binding capacity was defined as 50mg HAS/mL medium (20mM Tris-HCl). The mean particle size is defined as 34um and the bead structure is formed of 6% highly cross linked spherical agarose. The maximum flow rate is givens as 20mL/min while the recommended flow rate is defined as 5 mL/min, although, for the helicase to bind a flow rate of 1-1.5 mL/min was chosen. Again, two buffers were employed for the binding and elution of the protein. The binding buffer consisted of 20 mM Na-HEPES and 150 mM of NaCl with a pH of 7.0 while the elution buffer contained 20 mM Na-HEPES with 2 M NaCl at the same pH-value.

The column was washed with water and binding buffer, eluted, and again equilibrated with 5 column volumes as described in the affinity chromatography. To determine the exact ionic strength of the elution buffer needed for to elute each of the mutants from the matrix a gradient elution was performed. The proteins were loaded, and the system equilibrated with loading buffer. Thereafter, unwanted proteins were eluted with 20% elution buffer. At 60% of elution buffer and 40% binding buffer the desired enzyme was displaced by the elution buffer and collected for further purification. Each of the flow throughs and the elution were run on the gel for control.

The last chromatographic technique for purification of the helicase was size exclusion. In contrast to two other techniques there is no real chemical interaction between the column matrix and the amino acids. The exclusion is solely based on size. For this purpose a Superdex 200 10/300GL was chosen with a total volume of 24 mL with column dimensions of 30 cm and an inner diameter of 10mm. The matrix consists of an agarose-dextrane composite with an average particle size of 13um and separates globular proteins in the range of 10-600 kDa. The flow rate is between 0.25 and 0.75 mL/min with 1.5 Mpa. The column was washed with water and subsequently buffer (20mM Na-HEPES and 150 mM of NaCl with a pH of 7.0) and 500 µL of the previously purified and concentrated sample were loaded onto the column and pushed through the column with buffer. After approximately 14 minutes the desired protein was collected by fractionation in a standard 96 well plate. The concentrations of the fractions containing the peaks were subsequently labeled and stored in the fridge. An aliquot of all flow throughs as well as the elution were collected and run on the gel.

3.2.8 Enzymatic phosphate release assays

The EnzChek Phosphate Assay Kit (E-6646) supplied by Thermo Fisher was employed for sensitive and fast spectrophotometric quantitative detection of inorganic phosphate release in solution suitable for catalyzation of ATP by ATPases and GTPases. A 96 well plate suitable for the plate reader was loaded as follows: 20 µL of 6.25 mM Helicase were mixed with 10 uL of 1 times reaction buffer (containing 0.05 M Tris-HCl, 1mM MgCl, 0.1mM of sodium azide at a pH of 7.5), 40uL of 200mM 2-amino-6mercapto-7-methylpurine riboside (MSEG), 2 µL of 1 U/mL purine nucleoside phosphorylase (PNP), 20uL of 10mM ATP, 10 mL of 0.5 mg/mL Poly-A and the remaining volume to 200 mL was filled with dH2O. For each of the mutants, the wild type and the blank three wells were filled. The absorbance was measured at 360nm which was produced when the inorganic phosphate reacted with the MSEG catalyzed by the enzyme PNP to form ribose 1-phosphate and 2-amino-6-mercapto-7-methylpurine with the latter having a maximal absorption at 360nm. The prepared wells were read by a plate reader at 30°C for every 30 seconds.

In the described manner two measurements with two repeats were performed. The first assay measurement was performed using 24 wells including blank, wild type and the six mutants with each of them being measured in three separate wells. The ATP concentration was held constant at the previously mentioned concentration. This measurement was performed to compare the relative velocity at a given substrate concentration.

The second experiment included the wild and the mutation introduced at residue number $_{Pro}229_{Ala}$. This mutant was chosen based on the first experiment. Varying ATP concentrations were chosen to assess the K_m of the mutation in comparison to the wild type.

The ATP was diluted to create 4 mM, 2 mM, 1 mM, 0.5 mM, 0.25 mM, 0.125 mM, 62.5 µM, 32 µM, 16 µM, 8 µM solutions. This was prepared by creating a 140 µL 4 mM stock and transferring 70 µL of the stock to a microcentrifuge tube and adding 70 µL of ddH2O and repeating the process until arriving at 8 µM.

4. Results

4.1 Homology Model NS5 RdRp

The DOPE score of the NS5 RdRp hybrid model was compared to the DOPE score of the NS5 RdRp which is depicted in Figure 4. While the energy averaged over all residues does not differ significantly, there are more fluctuations present in the DOPE of the hybrid model without energy minimization. After Energy minimization using the YASARA webtool the DOPE score spikes around residues 90, 120 and 420 disappear. Especially the residues 1-300 show a completely different Energy profile which in accordance when comparing the structures visually. Nevertheless, they model shows high similarity of the DOPE profile compared to the model based on the recently resolved crystal structure. The DOPE score of these models is plotted in Figure 5. The comparison of the initial models of the YFV and JEV are visualized in Figure 16 as well as their DOPE in Figure 17 in the appendix. The comparison of the DOPE score of the individual models was clearly lowered by creating a hybrid and removing steric clashes creating an overall improved model. A visual comparison of the hybrid model and the models in cartoon representation is visualized in Figure 6.

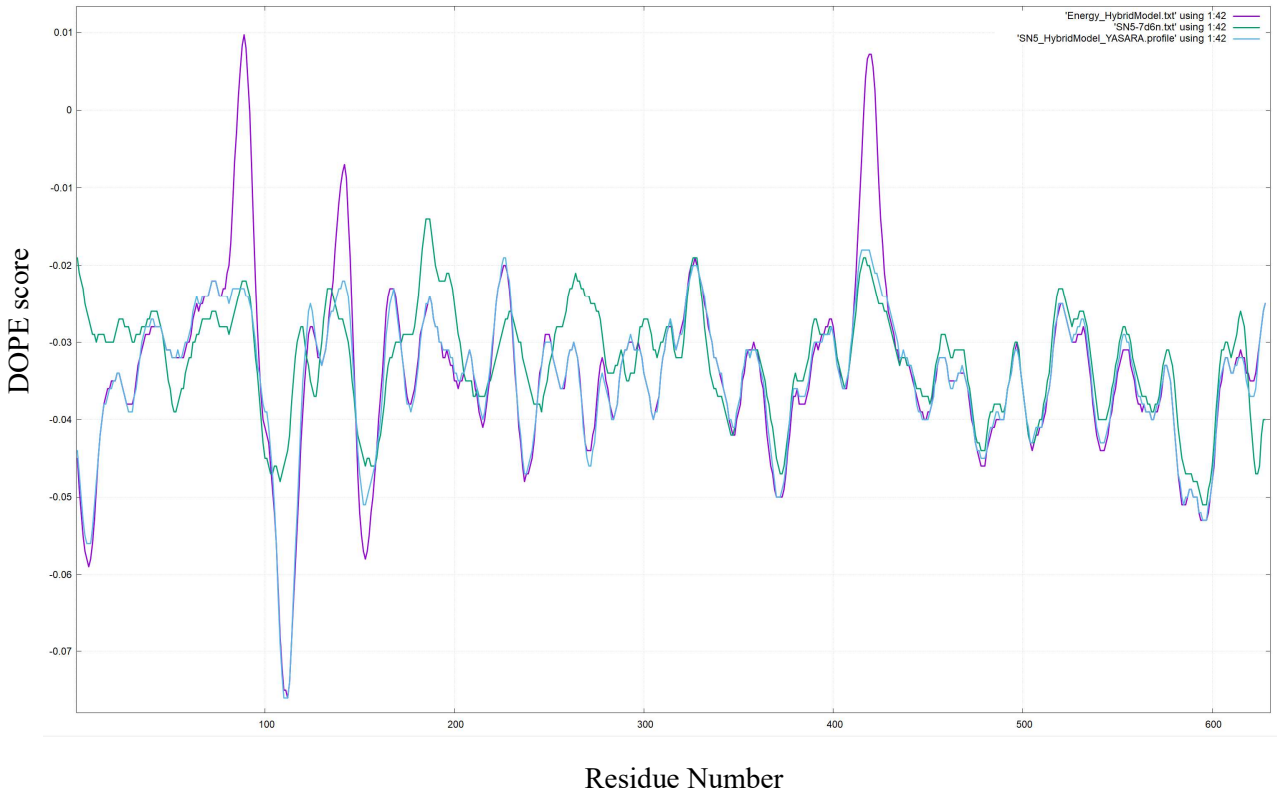

Figure 5: DOPE score of the NS5 RdRp of the hybrid model without energy minimization is represented in purple, the relaxed model is represented in blue, and the model based on the x-ray structure of the Eastern strand is visualized in green as a function of the residue number of the NS5 RdRp.

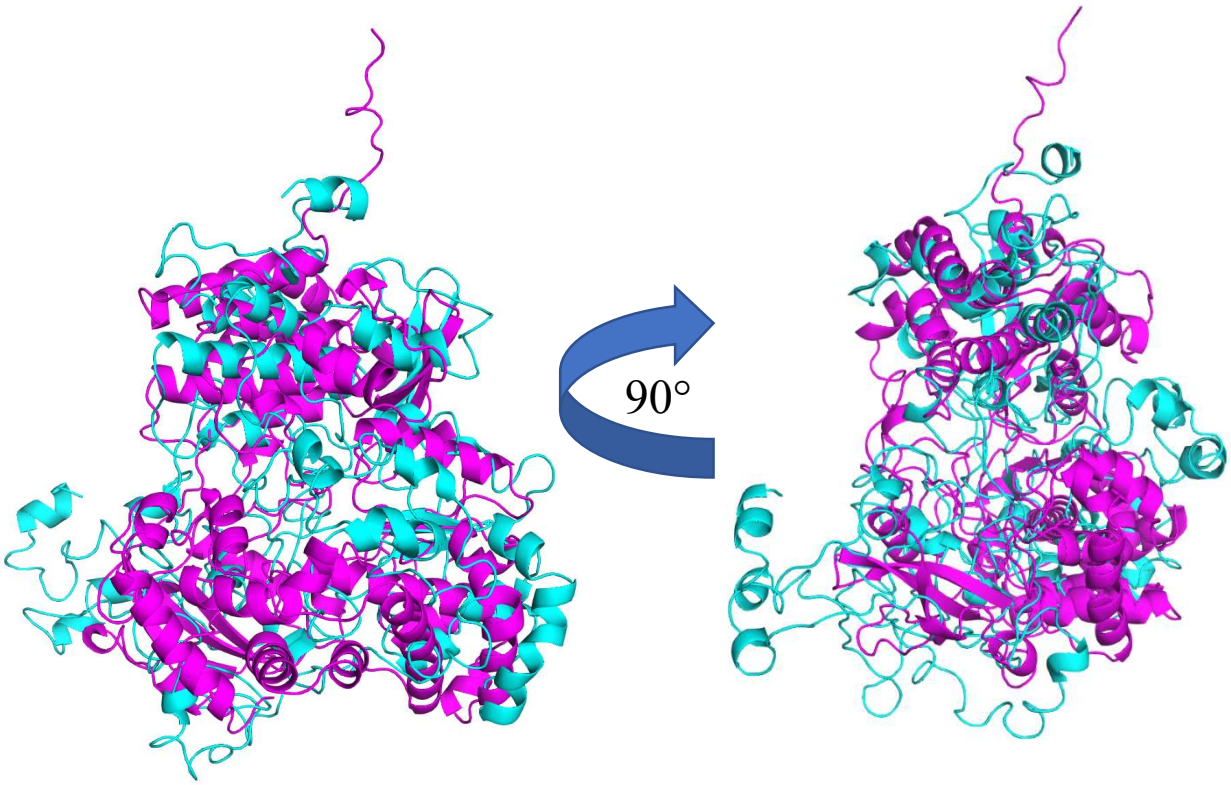

Figure 6: Cartoon representation of the hybrid model depicted in cyan, and the model based on PDB 7n6b. The models were compared using the *align* function in PyMOL. The difference in the DOPE score visualized in Figure 5 of residues 1-300 are seen in the lower part of the comparison where the formed β-sheets of the are not present in the x-ray-based model.

4.2 Docking NS3 and NS5

The strengths and weaknesses of each of the best three clusters generated by HADDOCK are in the Appendix in Figure 18-25. HADDOCK combines interface RMSD, desolvation energy, Van der Waals energy, electrostatic energy and restraint energy and identifies the best cluster with all the given parameters are given equal weight. The model with the lowest HADDOCK score, and hence, the lowest Energy is visualized in Figure 7. An interaction cutoff of 5 Å for weak electrostatic interactions, 3 Å for strong electrostatic interactions and 2 Å for hydrogen bonds between donor H and acceptor or less than 3.2 Å between donor and acceptor atom was chosen. The interactions predicted were compared to the crosslinking results obtained by RNDr. Zdeněk Franta, Ph.D in Table 3. The residues predicted by HADDOCK for the best three models for are not in accordance with the experimental data. The experimentally determined residues are on average more than 50 Å away from the NS3H RdRp interaction site predicted by the webserver. The orientation of the remaining 7 clusters showed an identical orientation of the two proteins relative to each other.

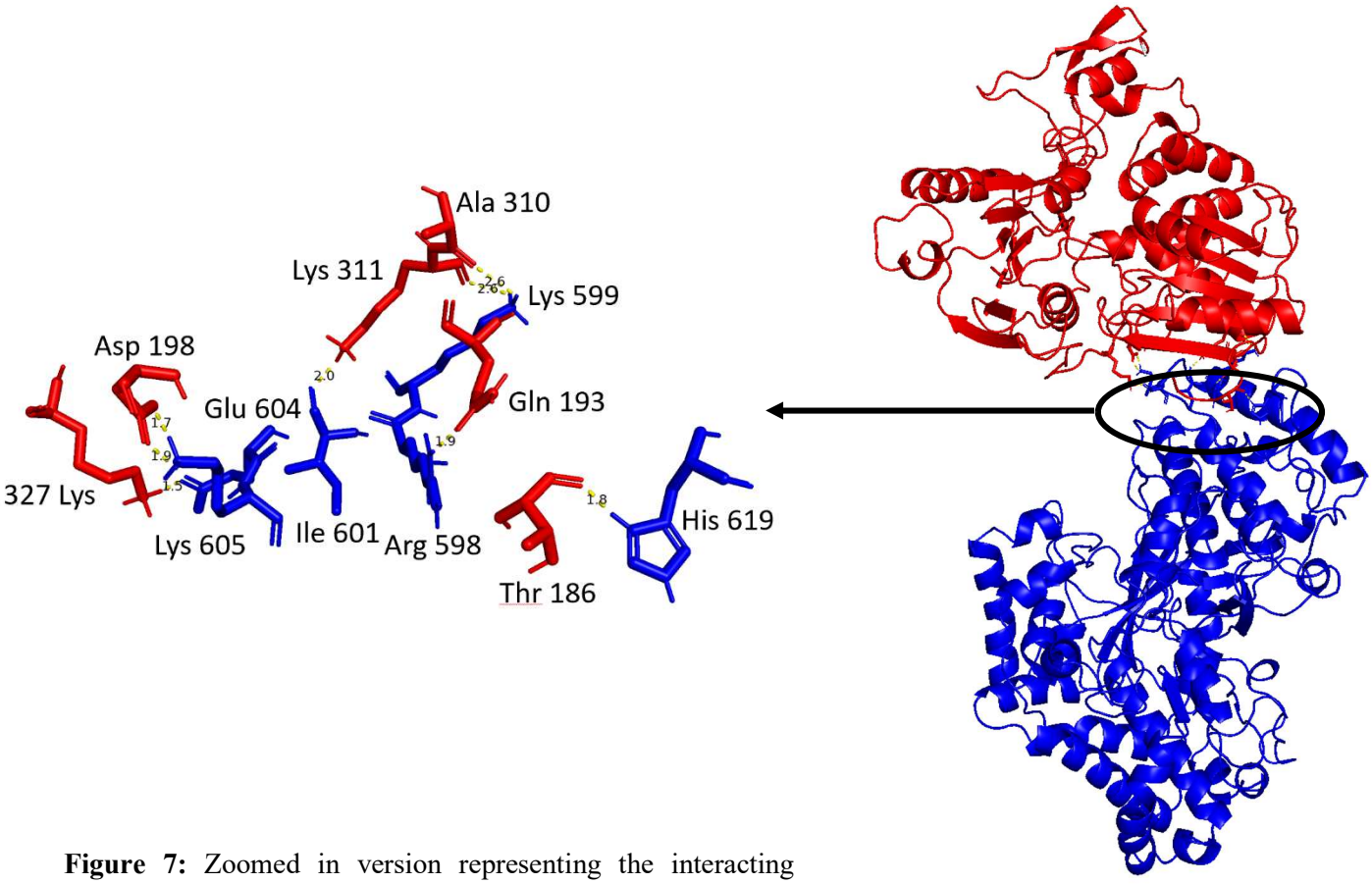

Figure 7: Zoomed in version representing the interacting residues between the NS3H and the NS5 RdRp. It shows the RdRp represented in blue binding with its C-terminus which is located at thumb to the Subdomain 1 of the NS3H. The hydrogen bonds are visualized with yellow dotted lines and connect the residues shown as sticks. The rest of the two enzymes is hidden in the stick representation for a visual simplicity.

Table 3: Comparison of the HADDOCK predication with the crosslinking results of our research Group.

Crosslinking Results			
Enzyme	Enzyme	Residue (NS3)	Residue (NS5)
NS3	NS5	246	314
NS3	NS5	369	314
NS3	NS5	295	311
HADDOCK Cluster 1			
APO (PDB: 7AY4)	NS5 Model (MODELLER from PDB 7D6n and MD)	Residue (NS3)	Residue (NS5)
7AY4	NS5 Model	186	619
7AY4	NS5 Model	311	609
7AY4	NS5 Model	311	599
7AY4	NS5 Model	327	604
7AY4	NS5 Model	198	604
7AY4	NS5 Model	311	601
HADDOCK Cluster 2			
7AY4	NS5 Model	186	595
7AY4	NS5 Model	187	595
7AY4	NS5 Model	198	607
7AY4	NS5 Model	311	603
7AY4	NS5 Model	311	604
7AY4	NS5 Model	325	605
7AY4	NS5 Model	198	608
HADDOCK Cluster 3			
7AY4	NS5 Model	312	605
7AY4	NS5 Model	313	602
7AY4	NS5 Model	313	598
7AY4	NS5 Model	314	604
7AY4	NS5 Model	314	603
7AY4	NS5 Model	327	107
7AY4	NS5 Model	501	107

4.3 Molecular Dynamics Simulations and Phosphate release in NS3H

To elucidate the coupling between the ATPase activity and the mechanical motion of RNA along the RNA binding groove the NTPase cycle of the NS3H the publication "Allostery in the dengue virus NS3 helicase: Insights into the NTPase cycle from molecular simulations" by Davidson B. Russell et. Al. was taken as an inspiration. In the publication the authors investigate the allosteric effects of RNA and NTPase substrates on the protein structure using a multiscale computational analysis. Furthermore, they present pathways of short-range, residue-residue interactions connecting the two active sites. [48] All simulations of the NS3H with the used ions and ligands are summarized in Table 2 in the appendix. Furthermore, a chronological scheme of the simulated configurations is shown in Figure 8. The insertion of ssRNA based on an available Dengue structure (PDB: 2JLU) led to repositioning of the P-loop as well as conformational changes in Motif VI and alpha helix 7 which had previously been observed in a crystal structure by Dahai Luo, et. al..[49] ATP was introduced as a substrate using

the crystal structure containing the AMPPNP as a template for positioning. The simulation led to two main conformations being identified. The ATP stabilized its position after roughly 550 ns while the RNA fluctuated between two principal conformations corresponding to a sliding along the binding groove depicted in the Root mean square deviation (RMSD) of the RNA relative to the protein depicted in Figure 10A. Furthermore, a cartoon comparison of the two states is visualized in Figure 9. These two structures were taken for further investigation and are referred to as highRMSD (forward movement of the RNA) and lowRMSD (backwards movement of the RNA). In both confirmations ATP was replaced with ADP and Pi. The single negatively charged phosphate used by this group did not leave the Helicase of the Dengue virus, this was confirmed with the TBEV helicase in simulation 6 and 7 for both positions of the RNA, respectively. At physiological conditions in the cytosol, where NS3H is catalytically active, the pH is around 7.4. The phosphate is already negatively charged when being attack from the hydroxyl ion generated from the catalytic water carrying an additional negative charge. Furthermore, the pKa of the second hydrogen of H_3PO_4 is at 7.21 making it reasonable for the inorganic phosphate to be double negatively charged. Hence, the Pi was replaced by HPO_4^{2-} using a newly created force field (courtesy of Dr. David Reha). This led to the phosphate being released in multiple repeats using the highRMSD ranging from 400-1500us while the lowRMSD showed no release for any of the repeats. Control simulations numbered 4 and 5 were performed using ADP and ATP without RNA. [48] The binding interactions between the RNA and the residues situated in the binding groove in the low- and highRMSD were identified for 100 frames each. Only those amino acids which formed hydrogen bonds in the latter structures were chosen to be mutated. Additionally, three already available primers for residues in the ATPase domain were available of which only residue 290 was successfully mutated alongside those in the RNAse domain and to which should inhibit the ATPase activity even in the presence of a poly-A RNA strand, and therefore, can be used as a control. The primers used are summarized in Table 3 in the appendix.

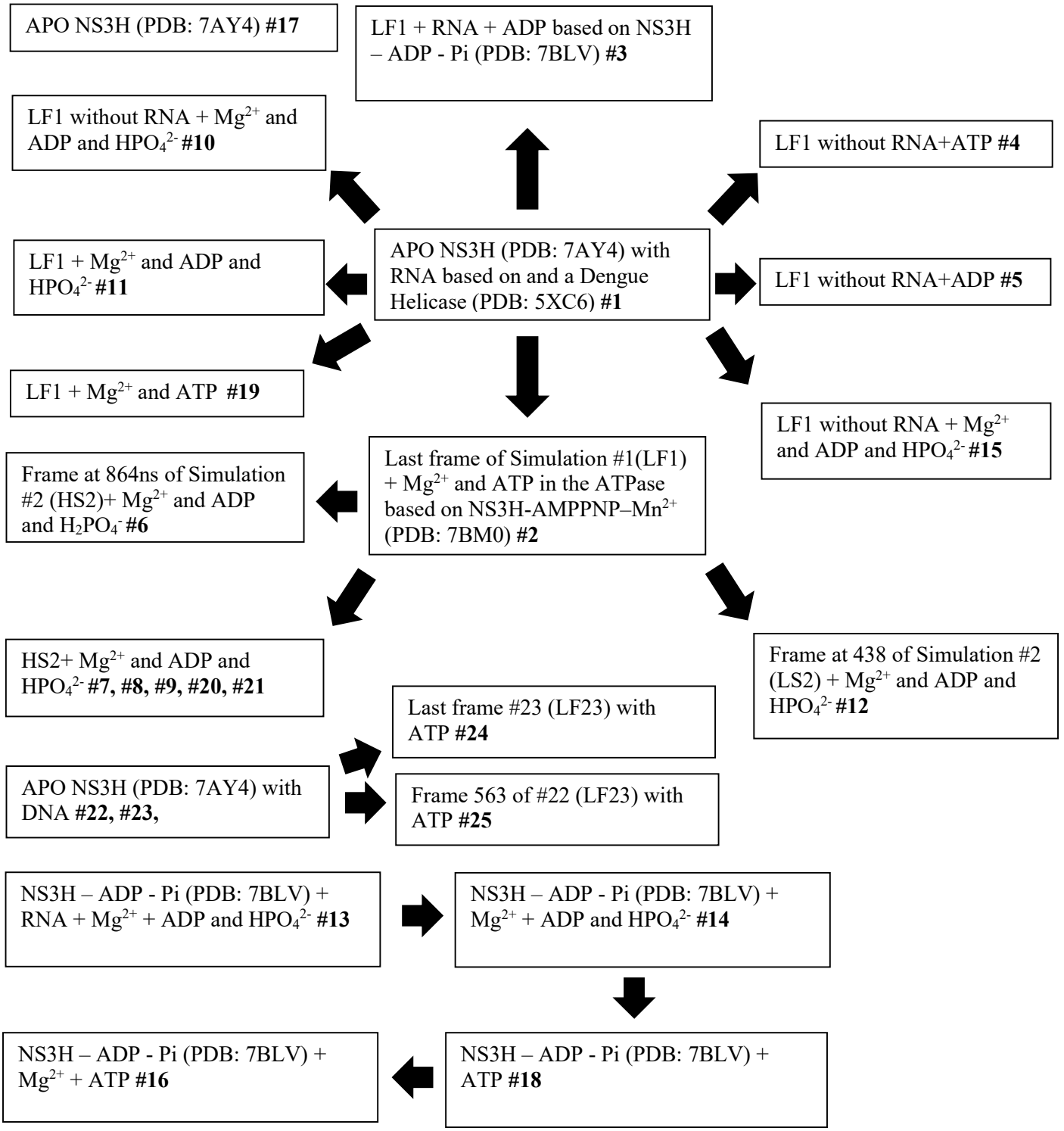

Figure 8: Flow chart of the simulations of NS3H using various starting structure for protein and positioning of the ligands which are referred to with corresponding four letter their PDB code. The numbering used corresponds to the chronological order of the simulations.

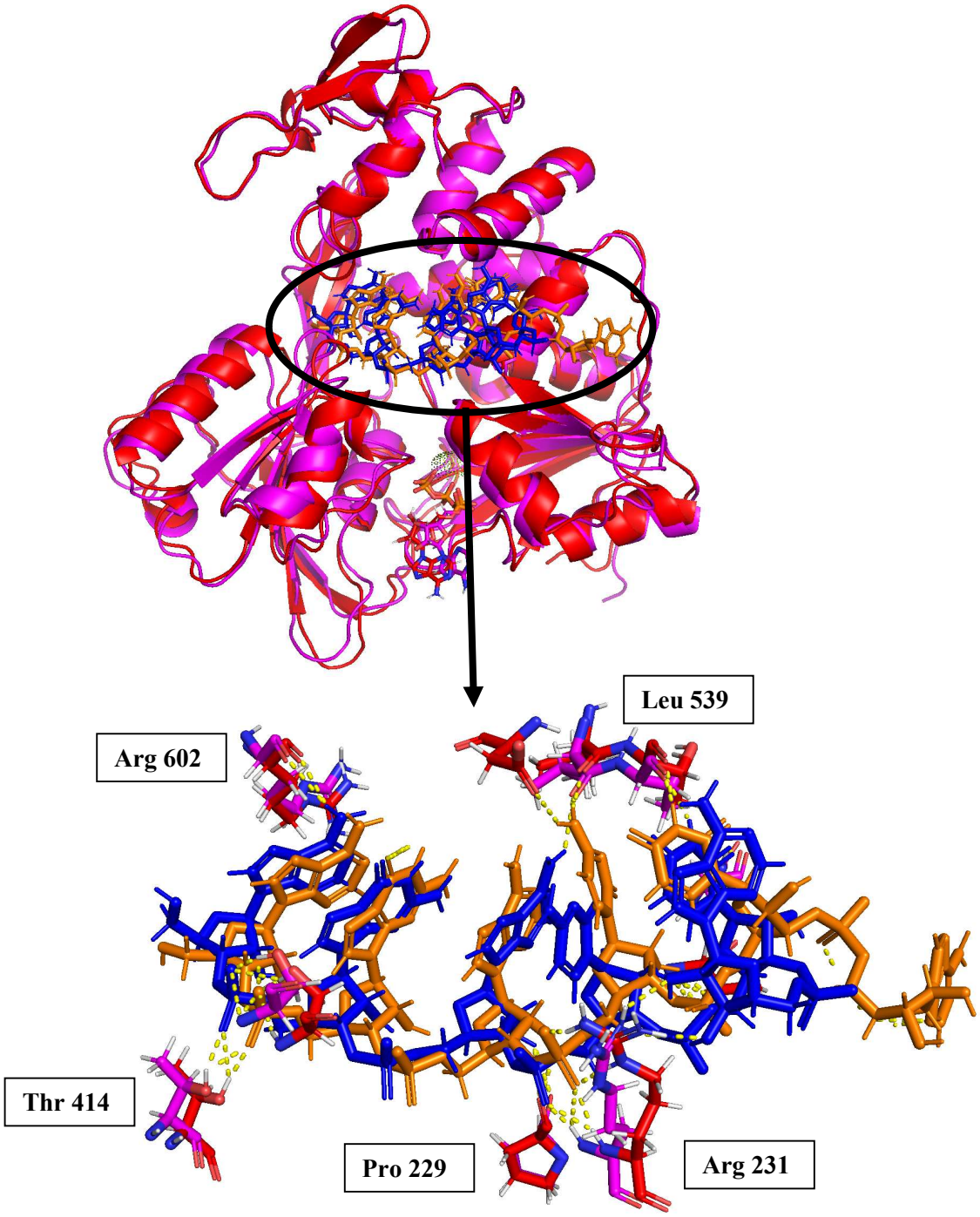

Figure 9: Cartoon representation of the high and the low RMSD structures indicating to the sliding of the RNA along the RNA binding grove. The red structure is representing the structure where the RNA already moved forwards inside the groove while the magenta-colored structure represents the NS3H with initial position of the RNA. The RNAs are for visualization colored orange and blue, respectively, with their common contacts shown in the zoomed in depiction. Both enzymes have ATP with magnesium bound in their ATPase site. The position of the ATP changes with the RNA movement. Additionally, the residue P229 is visualized which is exclusively bound to the structure corresponding to the high RMSD.

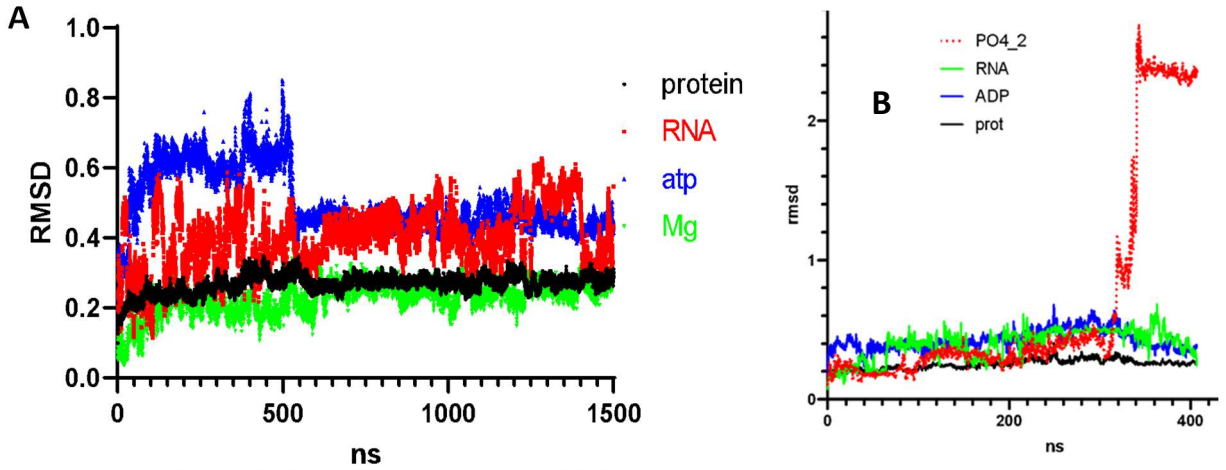

Figure 10A: RMSD of simulation 2 for ATP, RNA Magnesium and the protein backbone.
B: Showing the RMSD during the phosphate release depicting ADP, RNA inorganic phosphate and the protein backbone.

4.3.1 Phosphate release pathway:

Phosphate release followed a reproducible pathway with four main steps, represented in Figure 11. The initial state of the gamma phosphate is close to the beta phosphate. After 300-1500 ns of simulation time using the previously stated input parameters the phosphate is stabilized by the helix 7 to the left and the ADP to the right which corresponds to position 2 in Figure 11. At this point the phosphate release is initiated an no fluctuation back into positions further back in the ATPase site is observed in all repeats. After that the phosphate is released either to the bottom as depicted in the in Figure 11 or does leave directly through the helical gate formed by helix 7 and 2. In both observed release pathways an opening of the helical gate is observed. The final position of the phosphate was different in different simulations but on the outside of the protein.

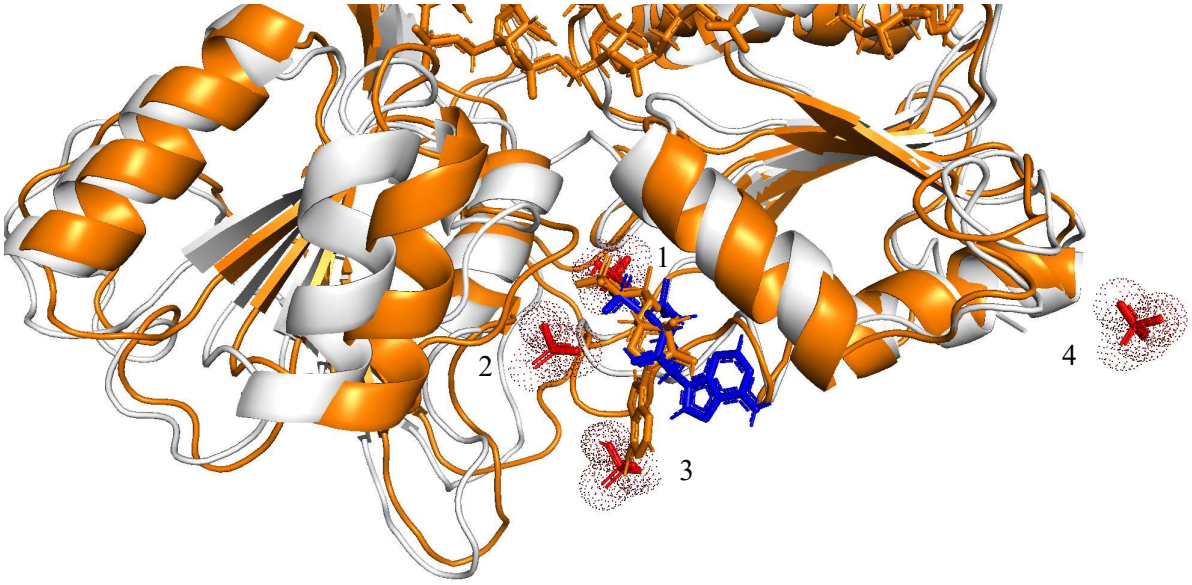

Figure 11: Phosphate release pathway of the doubly negative charge inorganic phosphate. The initial structure is colored in orange with the corresponding phosphate numbered one while the structure upon the release indicated by the opening of the helical gate is represented in the white structure with the corresponding phosphate number three. The inorganic phosphates are numberd chronologically. The amount of time simulated for the phosphate to end up in its final position was 900ns.

4.3 Mechanistic Hypothesis and SDM results

The selection of the residues to be mutated were based on the procedure described in 3.2. The mutations for residue D290A, P229A, D296A, R274A, L539A and D543A were successful. However, when sequencing the isolated plasmids containing the mutations from the transfected NEB 5-alpha cells there were no positive results achieved for the primers Q459A and R463A. This might be due to the high GC content which could have influenced the correct annealing of the primers. Hence, the following procedure was only performed on the six remaining plasmids. The pilot expression after transfecting the expression cells resulted in a uniform picture favoring 30°C and a time of 4-6 hours for all mutants. There were only subtle differences in the expression levels. After the large-scale expression at this predetermined time and temperature differences in the yield became apparent. Especially for helicases with the mutation P229A and L539A a significantly lower yield was achieved which added up on their already suppressed expression compared to the other four mutants. This trend continued for the IEC, however, there was no significant loss in concentration when performing the SEC. The SDS-PAGE results are depicted in Figure 12.

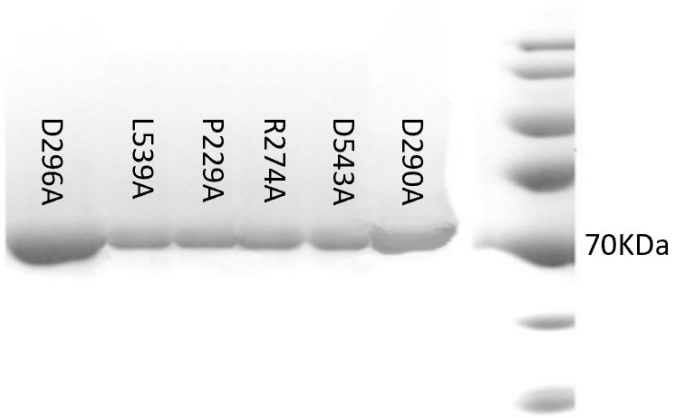

Figure 12: The SDS-PAGE of all the purified helicase mutants with a protein latter on the far right.

The results of the phosphate release assays are depicted in Figure 13 below. The measurements include a control, the wild type and the six mutants. To each enzyme three wells were assigned. The results are averaged over two of the three wells due to outliers. The corresponding velocity from the measurement values was subsequently calculated. The protein concentration was determined by using the absorbance at 280 nm from the detection device of the size exclusion chromatographic system and cross-checked with the absorbance using a nanophotometer at 280 nm. The differences in concentration were accounted in the calculations.

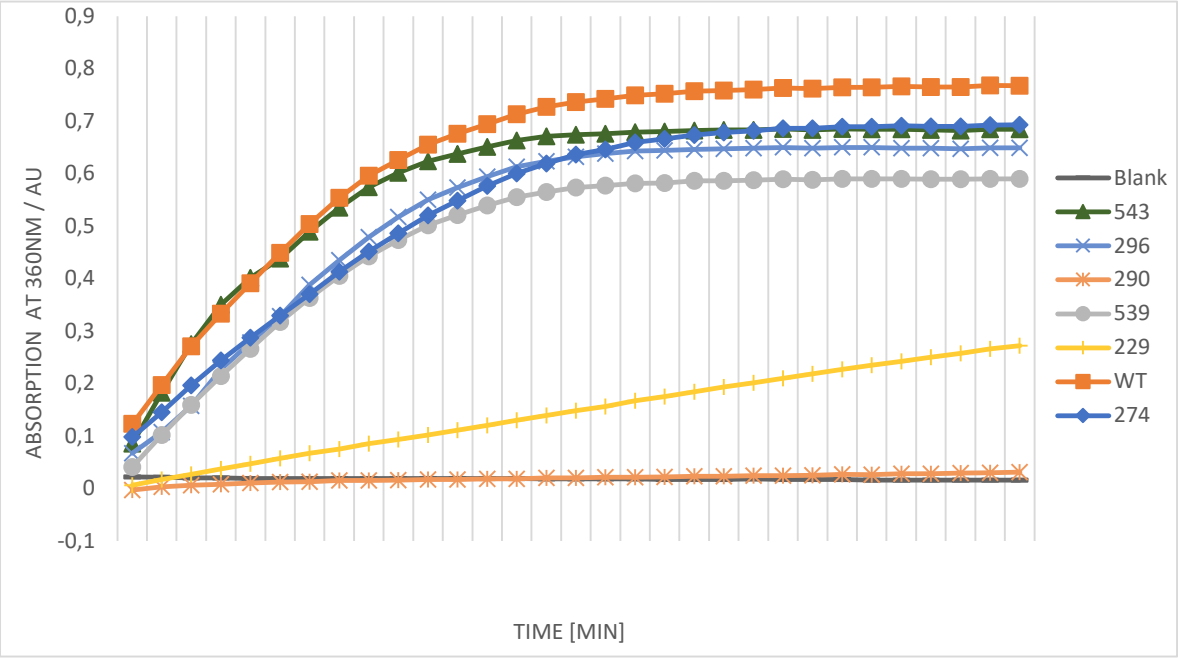

Figure 13: Depiction of the absorption measured in Au versus the time in minutes for each of the six mutants, the wild type, and the control.

To determine the velocity the slope of the linear increase of absorption (i.e. the first ten seconds) of the corresponding curves were measured with units of [Au/sec] and converted to [µM/min] which corresponds to the conversion of substrate which in this case is ATP hydrolyzed into phosphate and adenosine di phosphate (ADP) per minute by dividing it by the characteristic extinction coefficient of MESG at 360nm. The results are summarized in table 4 below.

Table 4: Calculation of the velocity at 0,06µM enzyme concentration and 10mM ATP.

Enzyme	Regression [au/min]	Velocity [M/min]	Velocity [µM/min]
WT	0,0537	6,3929E-06	6,39285714
290	0,0009	1,0714E-07	0,10714286
229	0,0089	1,0595E-06	1,05952381
539	0,0421	5,0119E-06	5,01190476
274	0,0402	4,7857E-06	4,78571429
296	0,0464	5,5238E-06	5,52380952
543	0,052	6,1905E-06	6,19047619

The different mutants show decreased ATPase velocity compared to the wild type. The velocity for varying ATP concentration were measured for the mutation P229A and compared to the WT. The corresponding results are summarized in Table 5.

Table 5: Calculation of ATPase velocities determined at increasing ATP concentration with 0,06 µM NS3H WT and P229A respectively.

Enzyme	Regression [au/min]	Velocity [M/min]	Velocity [µM/min]
WT 8µM	0.0016	1.9048E-07	0.1904762
16µM	0.0019	2.2619E-07	0.2261905
32µM	0.0031	3.6905E-07	0.3690476
64µM	0.0044	5.2381E-07	0.5238095
128µM	0.0104	1.2381E-06	1.2380952
256µM	0.0157	1.869E-06	1.8690476
0.5mM	0.0233	2.7738E-06	2.7738095
1mM	0.0271	3.2262E-06	3.2261905
2mM	0.0253	3.0119E-06	3.0119048
4mM	0.0231	2.7586 E-06	2.7579663
229 8µM	0.0009	1.0714E-07	0.10714286
16µM	0.0009	1.0714E-07	0.10714286
32µM	0.0009	1.0714E-07	0.10714286
64µM	0.0009	1.0714E-07	0.10714286
128µM	0.0009	1.0714E-07	0.10714286
256µM	0.0009	1.0714E-07	0.10714286
0.5mM	0.0009	1.0714E-07	0.10714286
1mM	0.0008	9.5238E-08	0.0952381
2mM	0.0009	1.0714E-07	0.10714286
4mM	0.0006	7.1429E-08	0.07142857

The graphical representation below depicting the data of Table 5. There is a significant decrease in the velocity above an ATP concentration of 1 mM (Figure 14). A much lower velocity for the mutant P229A was measured as to be expected from the previous results. However, the mutant showed no ATPase activity in second ATPase assay using varying ATP concentrations.

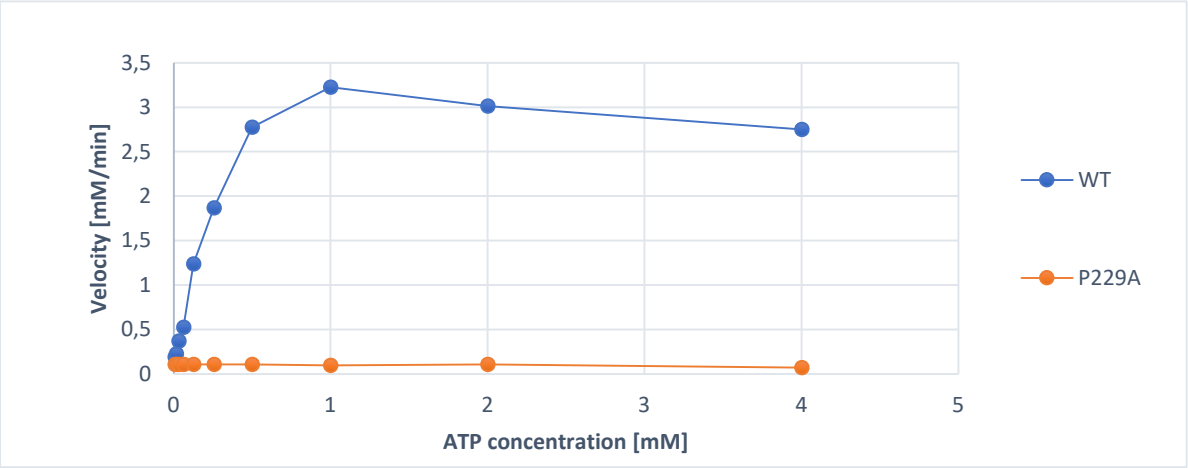

Figure 14: Velocity in mM/min plotted against the corresponding ATP concentration.

Based on the linear portion of the plot above both the K_m and the Vmax were determined for both the WT and the mutant P229A using a Lineweaver-Burk-plot. The Vmax for the NS3H WT was found to be 3.22 mM/min ATP concentration. The K_m calculated to be 57.53 µM respectively. The mutation P229A showed no ATPase activity.

5. Discussion

The initial model created of the NS5 RdRp shows the strengths and weaknesses of homology modelling. It produces reasonable approximations for proteins where the closely related enzymes assume a similar tertiary structure. If the overall structure does differ significantly due to the SNIPs introduced, the approach is not flexible enough.[50] When comparing the hybrid model with the x-ray based homology model which shares a higher sequence similarity used as a modelling template this limitation is observed. The thumb in the NS5 is much more compressed and consists of multiple beta sheets according to the newly resolved structure. The homology modelling software MODELLER as well as other modelling interfaces such as Swiss-Model were not able to predict this. The available x-ray structures of the YFV and the JEV show a less compressed thumb with alpha helices present. MODELLER tried to hold on to these tertiary structures being not flexible enough to form the beta strands based on the sequence alone. This is caused by the limited understanding of beta-sheet structure compared to that of alpha helices. [51] Furthermore, the fluctuations in the thumb subdomain during the simulation of the hybrid model confirms this hypothesis.

The interaction between the homology model of the NS3H and the NS5 RdRp predicted by HADDOCK is not in accordance with the experimental data. The contacts which were experimentally determined could be used to assist the HADDOCK run with additional parameters to improve the results and end up with an orientation of the NS3H relative to the NS5 which is in accordance with findings of related falviviridae.[29,48] The best model judging by the HADDOCK score which is depicted in Figure 7 does not resemble the complex predicted for the dengue virus.[13] Furthermore, orientation of the NS3H and the RdRp does not allow RNA to be fed through in a continuous manner as the previous models suggests.[10] For further analysis more data on the complex formation has to be gathered and a new model has to be constructed. Furthermore, a series of molecular dynamics simulations must be employed to evaluate the stability of the hydrogen bonds between the enzymes as well as the modelling of the RNA into the complex.

The investigating the NS3H NTPase cycle using molecular dynamics simulations did supplement the currently available static structure of the NS3H bound to various ligands.[52] The allosteric coupling between the NTPase and the RNAse site was investigated using

GROMACS leading to a hypothesis which was could be further tested by site directed mutagenesis. Furthermore, the release of the inorganic phosphate which is the rate limiting step in the NTPase cycle could be shown in multiple simulations enhancing the current understanding of the NTPase cycle of the NS3H. This is visualized schematically in Figure 15.

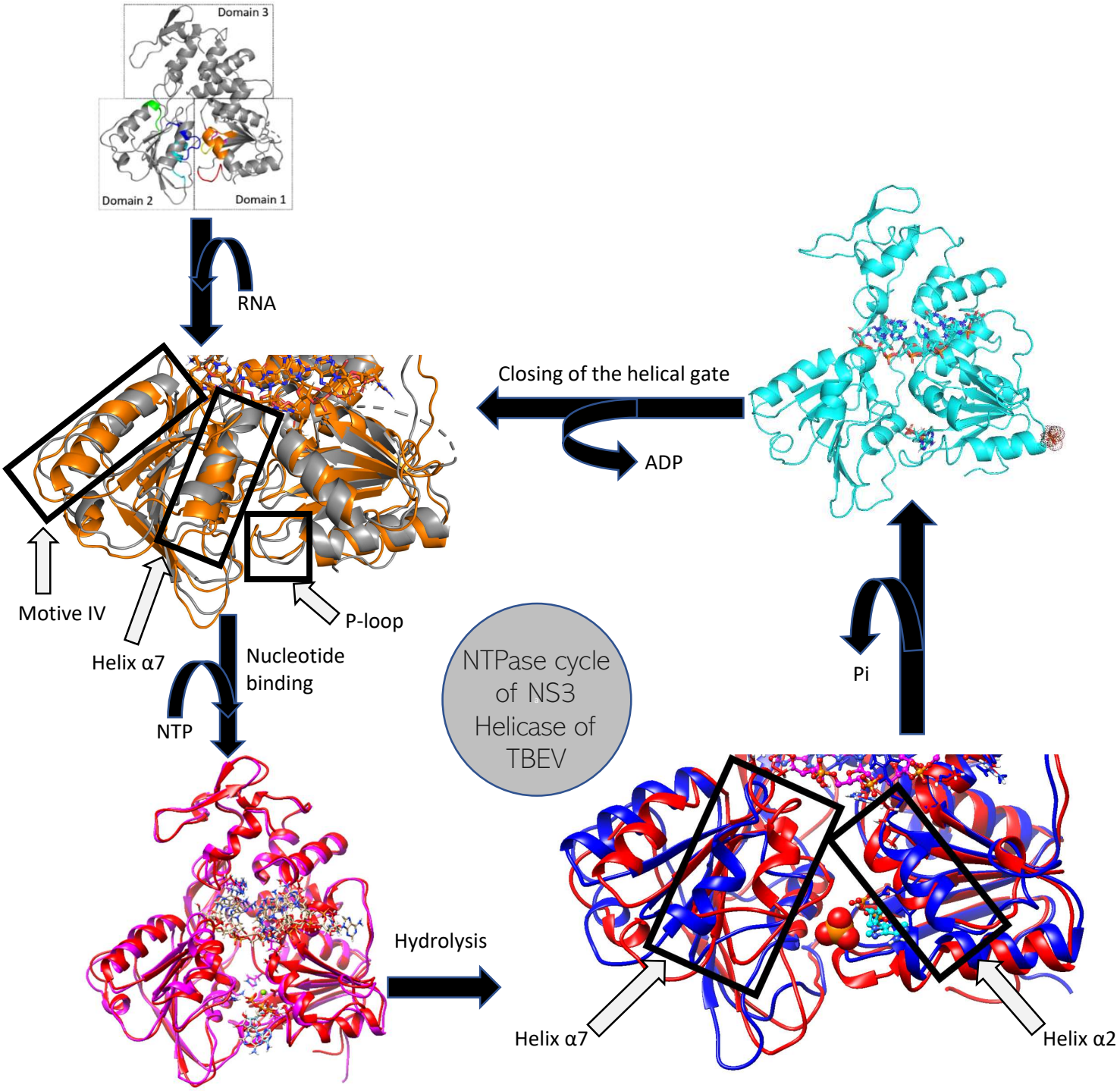

Figure 15: The apo NS3 is depicted in gray on the top of the figure representing the starting structure. The RNA bound NS3H is depicted in orange with its structural changes in motive 4, alpha helix 7 and the p-loop highlighted. The simulation including ATP in the active site resulted in subtle opening of the helical gate and fluctuations between two principle states corresponding to a sliding movement of the

RNA inside the binding groove were observed which are represented by the colors red (forward) and magenta (backwards) corresponding to a high and low RMSD of the RNA. When replacing the ATP with ADP and HPO_4^{2-} the release of the inorganic phosphate was observed for various repeats for the NS3H with the high RMSD structure. Upon the release a further opening of the helical gate was observed which is depicted in the blue frame on the bottom right compared to the starting position in red (high RMSD). The final structure depicted in cyan shows the inorganic phosphate already on the outside of the NS3H. The release of ADP from the active site is accompanied by a closing of the helical gate to end up with the RNA bound NS3H depicted in orange to complete the NTPase cycle of the helicase.

The mutation 290 shows no ATPase activity which can be expected since this residue is situated in the ATPase domain and serves a crucial part in the hydrolysis. This constitures a negative control. The velocities for mutations D296A, R274A, L539A and D543A were only slightly decreased compared to the WT leading to the conclusion that they are involved in moving the RNA along the binding groove, however, the presence of the remaining residues in contact are sufficient for the sliding of the substrate. More work has to be done in order to see whether double mutants would further decrease the enzyme velocity.[53]

While the mutant P229A did show a highly decrease ATPase activity using a 1 mM ATP concentration no activity wasobserved in the phosphate release assay with varying ATP concentrations. The positive results of the first assay was most likely caused by an experimental error. The complete inactivation of the ATPase activity could either be caused by a folding issue which must be further investigated by trying to crystallize the mutant and analyzing its x-ray diffraction pattern or due to the RNA binding. While the residue is almost never in contact with the RNA itself it does occasionally bind to it, and moreover, change its position upon the movement of the substrate along the binding groove. [54]

Furthermore, the inorganic phosphate release might be the power stroke required to move the RNA in the unwinding process by loosening its contacts as proposed in HCV by Qixin Wang, et al.. [55] For which both the release pathway as well as the confirmational changes in the protein backbone enhance the understanding of the mechanistic nature of the NS3H of TBEV.. The values for the WT are in accordance with the results obtained for the Dengue Virus by Sampath A. et al., whereas for the mutant P229A no corresponding measurement in related viruses was found. [56]

6. Conclusion

A combined approach of using both computational and experimental methods to investigate the non-structural proteins of the TBEV was used. Throughout this work the strength and weaknesses of in silico methods became apparent.

The homology modelling using MODELLER did produce a reasonable model if the template structure shared a similar tertiary structure, but the modelling failed to be flexible enough where this was not the case. The incorporation of artificial intelligence and machine learning to predict protein folding could help to overcome this.

The docking of the NS3H and the homology model was not in accordance with the HDX data even though they were in the input of the prediction. Both the position of the NS3H and the orientation does not allow the RNA to be fed through continuously which most likely indicated that the web-tool predicted it wrong.

The molecular dynamics simulations gave valuable insight into the translocation of the RNA along the binding groove as well as the phosphate release mechanism of the TBEV accompanied by the opening of the helical gate during the exit.

Furthermore, they laid the foundations for the site directed mutagenesis which revealed insights into the coupling between the RNAse and the ATPase domain of the viral NS3H. While the mutations P229A showed no ATPase activity further research has to be done to evaluate whether this introduced mutation does disrupt the folding of the protein or simply disrupts the coupling between the two enzymatically active sites of the NS3H.

7. Literature

(1) Lenz, N.; Engler, O.; Grandgirard, D.; Leib, S. L.; Ackermann-Gäumann, R. Evaluation of Antivirals against Tick-Borne Encephalitis Virus in Organotypic Brain Slices of Rat Cerebellum. *PLoS ONE* **2018**, *13* (10), e0205294. https://doi.org/10.1371/journal.pone.0205294.

(2) Gray, J. S.; Dautel, H.; Estrada-Peña, A.; Kahl, O.; Lindgren, E. Effects of Climate Change on Ticks and Tick-Borne Diseases in Europe. *Interdisciplinary Perspectives on Infectious Diseases* **2009**, *2009*, 1–12. https://doi.org/10.1155/2009/593232.

(3) Gejji, V.; Svoboda, P.; Stefanik, M.; Wang, H.; Salat, J.; Eyer, L.; Ruzek, D.; Fernando, S. An RNA-Dependent RNA Polymerase Inhibitor for Tick-Borne Encephalitis Virus. *Virology* **2020**, *546*, 13–19. https://doi.org/10.1016/j.virol.2020.03.006.

(4) Galabov, A. S.; Mukova, L.; Abashev, Y. P.; Wassilewa, L.; Tzvetkov, P.; Minkov, V.; Barinskiy, I. F.; Ouzounov, S.; Sidzhakova, D. Cycluridine: A Novel Antiviral Effective against Flaviviruses. *Antivir Chem Chemother* **2017**, *25* (2), 58–67. https://doi.org/10.1177/2040206617723442.

(5) Lani, R.; Moghaddam, E.; Haghani, A.; Chang, L.-Y.; AbuBakar, S.; Zandi, K. Tick-Borne Viruses: A Review from the Perspective of Therapeutic Approaches. *Ticks and Tick-borne Diseases* **2014**, *5* (5), 457–465. https://doi.org/10.1016/j.ttbdis.2014.04.001.

(6) Füzik, T.; Formanová, P.; Růžek, D.; Yoshii, K.; Niedrig, M.; Plevka, P. Structure of Tick-Borne Encephalitis Virus and Its Neutralization by a Monoclonal Antibody. *Nat Commun* **2018**, *9* (1), 436. https://doi.org/10.1038/s41467-018-02882-0.

(7) Hayes, E. B. Flaviviruses. In *Principles and Practice of Pediatric Infectious Disease*; Elsevier, 2008; pp 1082–1087. https://doi.org/10.1016/B978-0-7020-3468-8.50224-8.

(8) Pulkkinen, L.; Butcher, S.; Anastasina, M. Tick-Borne Encephalitis Virus: A Structural View. *Viruses* **2018**, *10* (7), 350. https://doi.org/10.3390/v10070350.

(9) *Fields Virology*, 5th ed.; Fields, B. N., Knipe, D. M., Howley, P. M., Eds.; Wolters Kluwer Health/Lippincott Williams & Wilkins: Philadelphia, 2007.

(10) Klema, V.; Padmanabhan, R.; Choi, K. Flaviviral Replication Complex: Coordination between RNA Synthesis and 5'-RNA Capping. *Viruses* **2015**, *7* (8), 4640–4656. https://doi.org/10.3390/v7082837.

(11) Falgout, B.; Pethel, M.; Zhang, Y. M.; Lai, C. J. Both Nonstructural Proteins NS2B and NS3 Are Required for the Proteolytic Processing of Dengue Virus Nonstructural Proteins. *J Virol* **1991**, *65* (5), 2467–2475. https://doi.org/10.1128/jvi.65.5.2467-2475.1991.

(12) Li, H.; Clum, S.; You, S.; Ebner, K. E.; Padmanabhan, R. The Serine Protease and RNA-Stimulated Nucleoside Triphosphatase and RNA Helicase Functional Domains of Dengue Virus Type 2 NS3 Converge within a Region of 20 Amino Acids. *J Virol* **1999**, *73* (4), 3108–3116. https://doi.org/10.1128/JVI.73.4.3108-3116.1999.

(13) Brand, C.; Bisaillon, M.; Geiss, B. J. Organization of the *Flavivirus* RNA Replicase Complex: *Flavivirus* RNA Replicase Complex. *WIREs RNA* **2017**, *8* (6), e1437. https://doi.org/10.1002/wrna.1437.

(14) Li, K.; Phoo, W. W.; Luo, D. Functional Interplay among the Flavivirus NS3 Protease, Helicase, and Cofactors. *Virol. Sin.* **2014**, *29* (2), 74–85. https://doi.org/10.1007/s12250-014-3438-6.

(15) Assenberg, R.; Mastrangelo, E.; Walter, T. S.; Verma, A.; Milani, M.; Owens, R. J.; Stuart, D. I.; Grimes, J. M.; Mancini, E. J. Crystal Structure of a Novel Conformational State of the Flavivirus NS3 Protein: Implications for Polyprotein Processing and Viral Replication. *J Virol* **2009**, *83* (24), 12895–12906. https://doi.org/10.1128/JVI.00942-09.

(16) Chu, P. W. G.; Westaway, E. G. Replication Strategy of Kunjin Virus: Evidence for Recycling Role of Replicative Form RNA as Template in Semiconservative and Asymmetric Replication. *Virology* **1985**, *140* (1), 68–79. https://doi.org/10.1016/0042-6822(85)90446-5.

(17) Issur, M.; Geiss, B. J.; Bougie, I.; Picard-Jean, F.; Despins, S.; Mayette, J.; Hobdey, S. E.; Bisaillon, M. The Flavivirus NS5 Protein Is a True RNA Guanylyltransferase That Catalyzes a Two-Step Reaction to Form the RNA Cap Structure. *RNA* **2009**, *15* (12), 2340–2350.

https://doi.org/10.1261/rna.1609709.
(18) Wengler, G.; Wengler, G.; Gross, H. J. Studies on Virus-Specific Nucleic Acids Synthesized in Vertebrate and Mosquito Cells Infected with Flaviviruses. *Virology* **1978**, *89* (2), 423–437. https://doi.org/10.1016/0042-6822(78)90185-X.
(19) Henderson, B. R.; Saeedi, B. J.; Campagnola, G.; Geiss, B. J. Analysis of RNA Binding by the Dengue Virus NS5 RNA Capping Enzyme. *PLoS ONE* **2011**, *6* (10), e25795. https://doi.org/10.1371/journal.pone.0025795.
(20) Ray, D.; Shah, A.; Tilgner, M.; Guo, Y.; Zhao, Y.; Dong, H.; Deas, T. S.; Zhou, Y.; Li, H.; Shi, P.-Y. West Nile Virus 5′-Cap Structure Is Formed by Sequential Guanine N-7 and Ribose 2′-O Methylations by Nonstructural Protein 5. *J Virol* **2006**, *80* (17), 8362–8370. https://doi.org/10.1128/JVI.00814-06.
(21) Yang, J.; Jing, X.; Yi, W.; Li, X.-D.; Yao, C.; Zhang, B.; Zheng, Z.; Wang, H.; Gong, P. Crystal Structure of a Tick-Borne Flavivirus RNA-Dependent RNA Polymerase Suggests a Host Adaptation Hotspot in RNA Viruses. *Nucleic Acids Research* **2021**, *49* (3), 1567–1580. https://doi.org/10.1093/nar/gkaa1250.
(22) Bollati, M.; Alvarez, K.; Assenberg, R.; Baronti, C.; Canard, B.; Cook, S.; Coutard, B.; Decroly, E.; de Lamballerie, X.; Gould, E. A.; Grard, G.; Grimes, J. M.; Hilgenfeld, R.; Jansson, A. M.; Malet, H.; Mancini, E. J.; Mastrangelo, E.; Mattevi, A.; Milani, M.; Moureau, G.; Neyts, J.; Owens, R. J.; Ren, J.; Selisko, B.; Speroni, S.; Steuber, H.; Stuart, D. I.; Unge, T.; Bolognesi, M. Structure and Functionality in Flavivirus NS-Proteins: Perspectives for Drug Design. *Antiviral Research* **2010**, *87* (2), 125–148. https://doi.org/10.1016/j.antiviral.2009.11.009.
(23) van den Elsen, K.; Quek, J. P.; Luo, D. Molecular Insights into the Flavivirus Replication Complex. *Viruses* **2021**, *13* (6), 956. https://doi.org/10.3390/v13060956.
(24) Li, Y.; Li, Q.; Wong, Y. L.; Liew, L. S. Y.; Kang, C. Membrane Topology of NS2B of Dengue Virus Revealed by NMR Spectroscopy. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2015**, *1848* (10), 2244–2252. https://doi.org/10.1016/j.bbamem.2015.06.010.
(25) Kapoor, M.; Zhang, L.; Ramachandra, M.; Kusukawa, J.; Ebner, K. E.; Padmanabhan, R. Association between NS3 and NS5 Proteins of Dengue Virus Type 2 in the Putative RNA Replicase Is Linked to Differential Phosphorylation of NS5. *Journal of Biological Chemistry* **1995**, *270* (32), 19100–19106. https://doi.org/10.1074/jbc.270.32.19100.
(26) Moreland, N. J.; Tay, M. Y. F.; Lim, E.; Rathore, A. P. S.; Lim, A. P. C.; Hanson, B. J.; Vasudevan, S. G. Monoclonal Antibodies against Dengue NS2B and NS3 Proteins for the Study of Protein Interactions in the Flaviviral Replication Complex. *Journal of Virological Methods* **2012**, *179* (1), 97–103. https://doi.org/10.1016/j.jviromet.2011.10.006.
(27) Cui, T.; Sugrue, R. J.; Xu, Q.; Lee, A. K. W.; Chan, Y.-C.; Fu, J. Recombinant Dengue Virus Type 1 NS3 Protein Exhibits Specific Viral RNA Binding and NTPase Activity Regulated by the NS5 Protein. *Virology* **1998**, *246* (2), 409–417. https://doi.org/10.1006/viro.1998.9213.
(28) Zou, G.; Chen, Y.-L.; Dong, H.; Lim, C. C.; Yap, L. J.; Yau, Y. H.; Shochat, S. G.; Lescar, J.; Shi, P.-Y. Functional Analysis of Two Cavities in Flavivirus NS5 Polymerase. *Journal of Biological Chemistry* **2011**, *286* (16), 14362–14372. https://doi.org/10.1074/jbc.M110.214189.
(29) Tay, M. Y. F.; Saw, W. G.; Zhao, Y.; Chan, K. W. K.; Singh, D.; Chong, Y.; Forwood, J. K.; Ooi, E. E.; Grüber, G.; Lescar, J.; Luo, D.; Vasudevan, S. G. The C-Terminal 50 Amino Acid Residues of Dengue NS3 Protein Are Important for NS3-NS5 Interaction and Viral Replication. *Journal of Biological Chemistry* **2015**, *290* (4), 2379–2394. https://doi.org/10.1074/jbc.M114.607341.
(30) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.
(31) Das, R.; Baker, D. Macromolecular Modeling with Rosetta. *Annu. Rev. Biochem.* **2008**, *77* (1), 363–382. https://doi.org/10.1146/annurev.biochem.77.062906.171838.

(32) Šali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology* **1993**, *234* (3), 779–815. https://doi.org/10.1006/jmbi.1993.1626.
(33) Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics* **2016**, *54* (1). https://doi.org/10.1002/cpbi.3.
(34) Marti-Renom, M. A. Alignment of Protein Sequences by Their Profiles. *Protein Science* **2004**, *13* (4), 1071–1087. https://doi.org/10.1110/ps.03379804.
(35) Needleman, S. B.; Wunsch, C. D. A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins. *Journal of Molecular Biology* **1970**, *48* (3), 443–453. https://doi.org/10.1016/0022-2836(70)90057-4.
(36) Madhusudhan, M. S.; Marti-Renom, M. A.; Sanchez, R.; Sali, A. Variable Gap Penalty for Protein Sequence–Structure Alignment. *Protein Engineering, Design and Selection* **2006**, *19* (3), 129–133. https://doi.org/10.1093/protein/gzj005.
(37) Shen, M.; Sali, A. Statistical Potential for Assessment and Prediction of Protein Structures. *Protein Sci.* **2006**, *15* (11), 2507–2524. https://doi.org/10.1110/ps.062416606.
(38) van Zundert, G. C. P.; Rodrigues, J. P. G. L. M.; Trellet, M.; Schmitz, C.; Kastritis, P. L.; Karaca, E.; Melquiond, A. S. J.; van Dijk, M.; de Vries, S. J.; Bonvin, A. M. J. J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *Journal of Molecular Biology* **2016**, *428* (4), 720–725. https://doi.org/10.1016/j.jmb.2015.09.014.
(39) Skeel, R. D. What Makes Molecular Dynamics Work? *SIAM J. Sci. Comput.* **2009**, *31* (2), 1363–1378. https://doi.org/10.1137/070683660.
(40) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Computational science series; Academic Press: San Diego, 2002.
(41) González, M. A. Force Fields and Molecular Dynamics Simulations. *JDN* **2011**, *12*, 169–200. https://doi.org/10.1051/sfn/201112009.
(42) Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159* (1), 98–103. https://doi.org/10.1103/PhysRev.159.98.
(43) Hess, B., K., C. ,. van der Spoel, D. and Lindahl, E. J. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular SimulationChem. *Theory Comput.* **2008**, *4*, 435–447.
(44) González, M. A.; Abascal, J. L. F. The Shear Viscosity of Rigid Water Models. *The Journal of Chemical Physics* **2010**, *132* (9), 096101. https://doi.org/10.1063/1.3330544.
(45) Sheridan, S.; Gräter, F.; Daday, C. How Fast Is Too Fast in Force-Probe Molecular Dynamics Simulations? *J. Phys. Chem. B* **2019**, *123* (17), 3658–3664. https://doi.org/10.1021/acs.jpcb.9b01251.
(46) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: A Package for Molecular Simulation and Trajectory Analysis. *J Mol Model* **2001**, *7* (8), 306–317. https://doi.org/10.1007/s008940100045.
(47) Grant, B. J.; Rodrigues, A. P. C.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. D. Bio3d: An R Package for the Comparative Analysis of Protein Structures. *Bioinformatics* **2006**, *22* (21), 2695–2696. https://doi.org/10.1093/bioinformatics/btl461.
(48) Davidson, R. B.; Hendrix, J.; Geiss, B. J.; McCullagh, M. Allostery in the Dengue Virus NS3 Helicase: Insights into the NTPase Cycle from Molecular Simulations. *PLoS Comput Biol* **2018**, *14* (4), e1006103. https://doi.org/10.1371/journal.pcbi.1006103.
(49) Luo, D.; Xu, T.; Watson, R. P.; Scherer-Becker, D.; Sampath, A.; Jahnke, W.; Yeong, S. S.; Wang, C. H.; Lim, S. P.; Strongin, A.; Vasudevan, S. G.; Lescar, J. Insights into RNA Unwinding and ATP Hydrolysis by the Flavivirus NS3 Protein. *EMBO J* **2008**, *27* (23), 3209–3219. https://doi.org/10.1038/emboj.2008.232.
(50) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Research* **2018**, *46* (W1), W296–W303. https://doi.org/10.1093/nar/gky427.
(51) Hughes, R. M.; Waters, M. L. Model Systems for β-Hairpins and β-Sheets. *Current Opinion in Structural Biology* **2006**, *16* (4), 514–524. https://doi.org/10.1016/j.sbi.2006.06.008.

(52) Larsen, A. H.; Wang, Y.; Bottaro, S.; Grudinin, S.; Arleth, L.; Lindorff-Larsen, K. Combining Molecular Dynamics Simulations with Small-Angle X-Ray and Neutron Scattering Data to Study Multi-Domain Proteins in Solution. *PLoS Comput Biol* **2020**, *16* (4), e1007870. https://doi.org/10.1371/journal.pcbi.1007870.
(53) Mildvan, A. S.; Weber, D. J.; Kuliopulos, A. Quantitative Interpretations of Double Mutations of Enzymes. *Archives of Biochemistry and Biophysics* **1992**, *294* (2), 327–340. https://doi.org/10.1016/0003-9861(92)90692-P.
(54) Kumar, A.; Biswas, P. Effect of Site-directed Point Mutations on Protein Misfolding: A Simulation Study. *Proteins* **2019**, *87* (9), 760–773. https://doi.org/10.1002/prot.25702.
(55) Wang, Q.; Arnold, J. J.; Uchida, A.; Raney, K. D.; Cameron, C. E. Phosphate Release Contributes to the Rate-Limiting Step for Unwinding by an RNA Helicase. *Nucleic Acids Research* **2010**, *38* (4), 1312–1324. https://doi.org/10.1093/nar/gkp1118.
(56) Sampath, A.; Xu, T.; Chao, A.; Luo, D.; Lescar, J.; Vasudevan, S. G. Structure-Based Mutational Analysis of the NS3 Helicase from Dengue Virus. *J Virol* **2006**, *80* (13), 6686–6690. https://doi.org/10.1128/JVI.02215-05.

8. Appendix

Figure 16: Visual representation of the homology model using YFV RdRp (PDB:6QSN), the JEV RdRp (PDB:5TFR) and the hybrid of the two models as a template coloured blue red and blue, respectively.

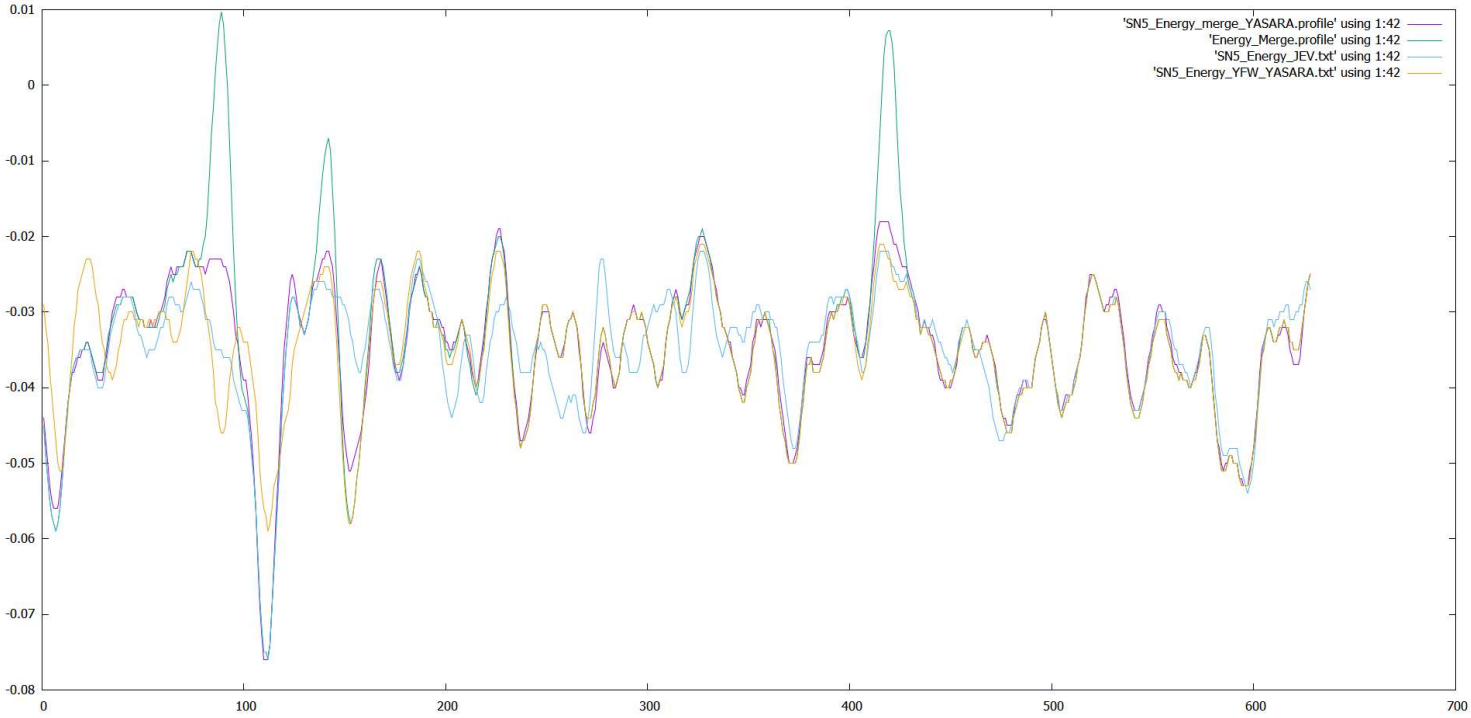

Figure 17: Comparison of the DOPE of the homology model of JEV, YFW, the initial hybrid and the hybrid without steric clashes due to the manual modelling in PyMOL represented in blue, yellow, green and magenta respectively.

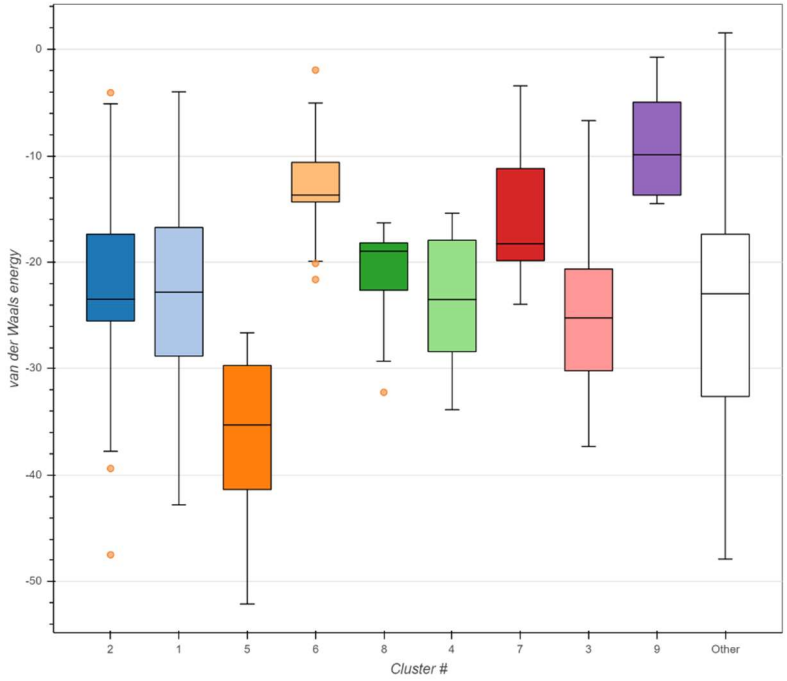

Figure 18: Representation of the boxplots on the van der Waals Energy of the individual clusters. The clusters 1-10 are represented by dark blue, light blue, orange, yellow, dark green, red, pink, purple and white for all following Figures, respectively.

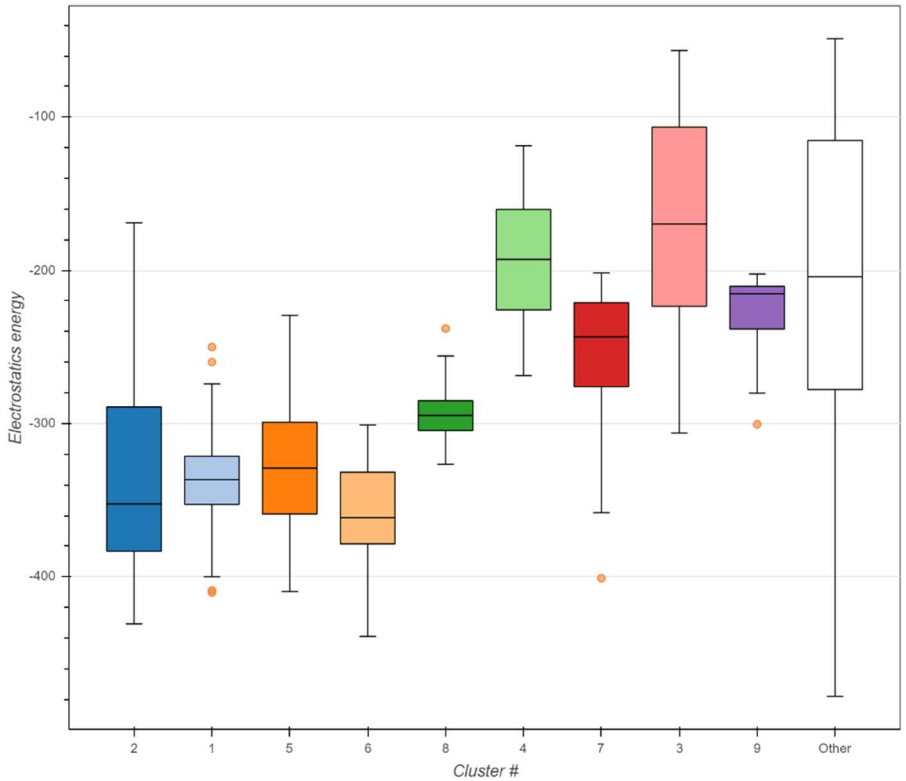

Figure 19: Representation of the boxplots on the electrostatic energy of the clusters.

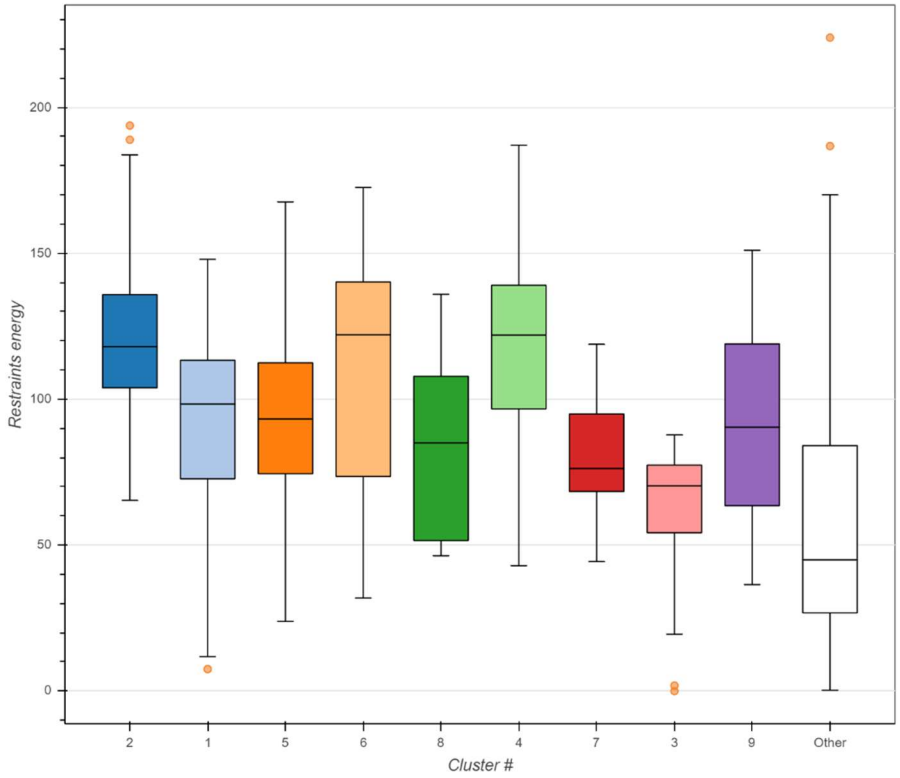

Figure 20: Representation of the boxplots on the restraint energy of the clusters.

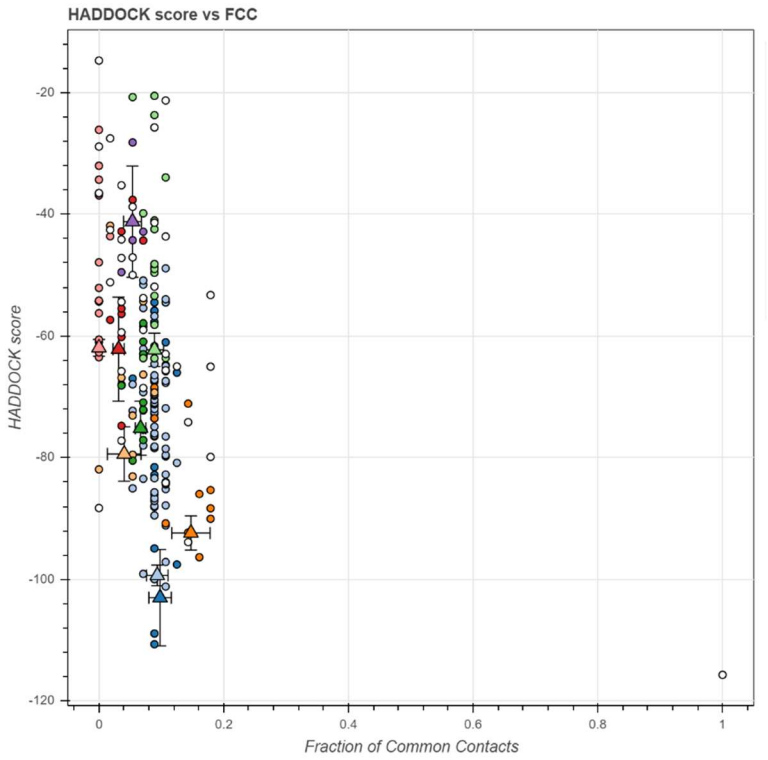

Figure 21: HADDOCK score plotted against the Fraction of Common Contacts (FCC).

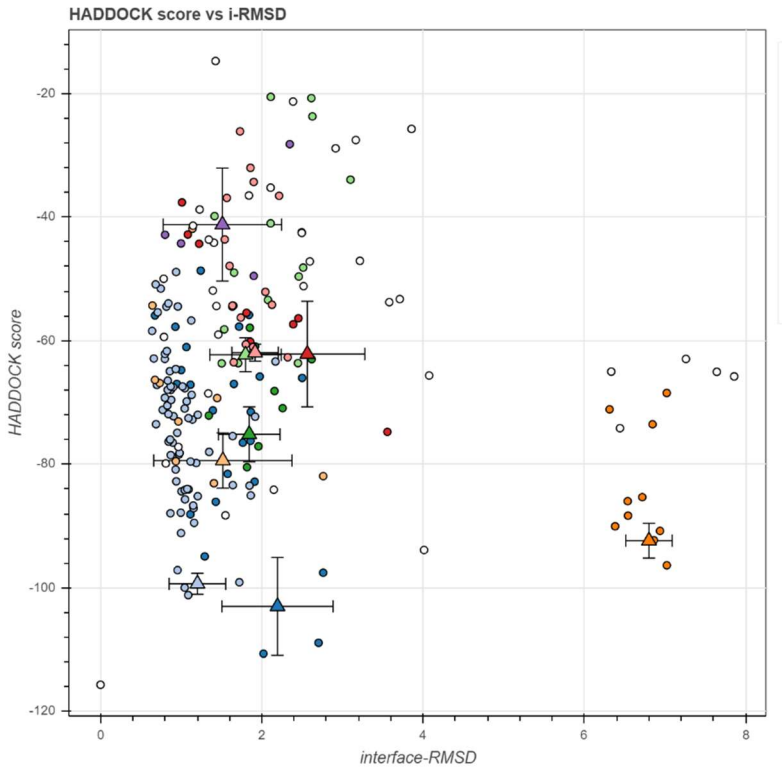

Figure 22: HADDOCK score plotted against the Fraction of Common Contacts (FCC).

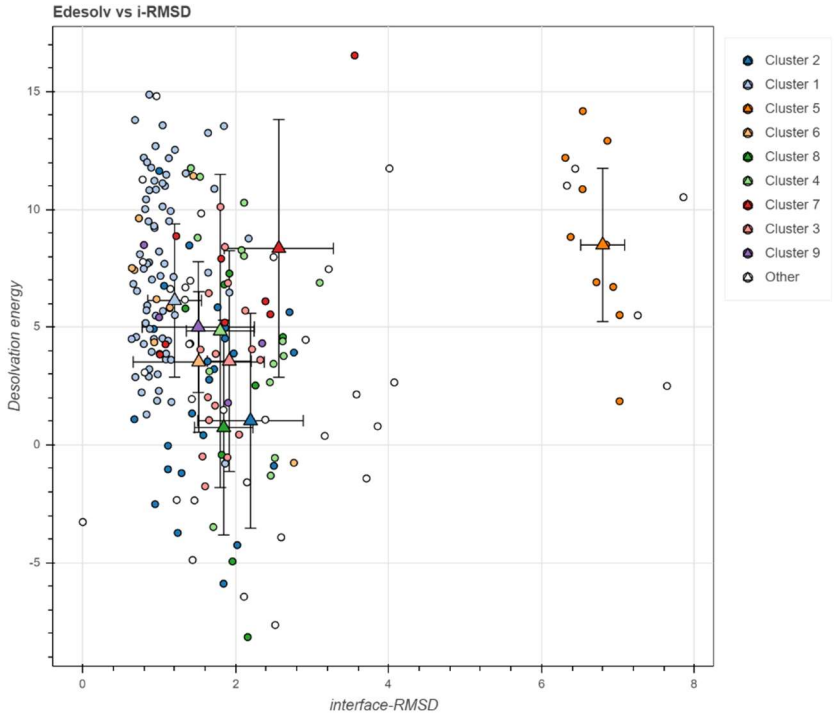

Figure 23: Desolvation Energy plotted against the interface-Root mean square deviation (RMSD).

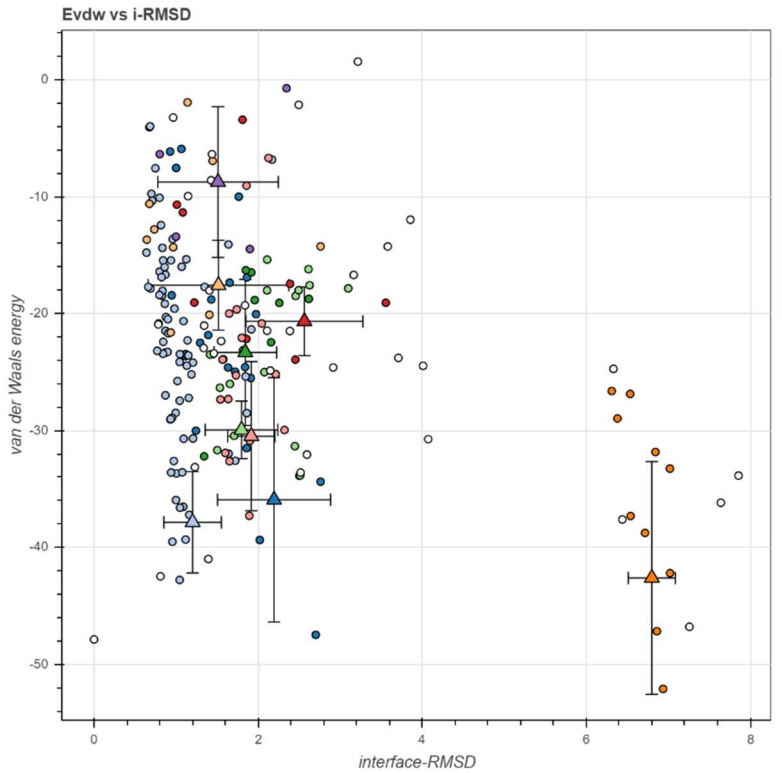

Figure 24: Van der Waals Energy plotted against the interface-Root mean square deviation (RMSD).

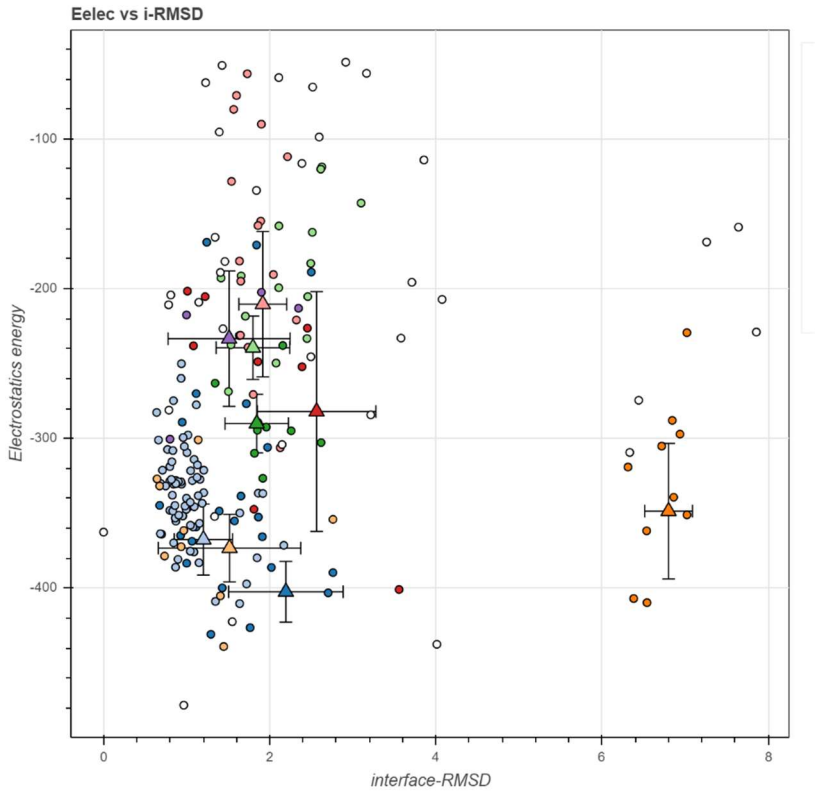

Figure 25: Electrostatic energy plotted against the interface-Root mean square deviation (RMSD).

Table 6: Inputs used for the molecular dynamics simulations in GROMACS summarized in the table above.

Simulations	Protein	Inorganic Ion	Ligand 1	Ligand 2	Ligand 3	Duration [ns]
1	PDB: 7AY4	-	ssRNA (AGACUA)	-	-	500
2	Last frame of Simulation 1 (LFS1)	Magnesium ion (Mg 2+) in the ATPase site	ssRNA (AGACUA)	ATP	-	1500
3	LFS1	Mg 2+	ssRNA (AGACUA)	ADP	-	1000
4	LFS1	Mg 2+	-	ATP		1000
5	LFS1	Mg 2+	-	ADP		1000
6	Frame at 864ns of Simulation 2 (HS2)	Mg 2+	ssRNA (AGACUA)	ADP	PO4 (-)	1500
7	HS2	Mg 2+	ssRNA (AGACUA)	ADP	PO4 (2-)	600
8	HS2	Mg 2+	ssRNA (AGACUA)	ADP	PO4 (2-)	1500
9	HS2	Mg 2+	ssRNA (AGACUA)	ADP	PO4 (2-)	800
10	LFS1	Mg 2+	-	ADP	PO4 (2-)	1500
11	LFS1	Mg 2+	ssRNA (AGACUA)	ADP	PO4 (2-)	1500
12	Frame at 438ns of Simulation 2 (HS2)	Mg 2+	ssRNA (AGACUA)	ADP	PO4 (2-)	1500
13	PDB: 2BLV*	Mg 2+	ssRNA (AGACUA)	ADP	PO4 (2-)	1500
14	PDB: 2BLV*	Mg 2+	-	ADP	PO4 (2-)	1500
15	LFS1	Mg 2+	ssRNA (AGACUA)	ADP	PO4 (2-)	1500
16	LSF1	Mg 2+	-	ADP	PO4 (2-)	1500
17	PDB: 7AY4	-	-	-	-	500
18	PDB: 2BLV*	Mg 2+	-	ATP	-	1000
19	LFS1	Mg 2+	ssRNA (AGACUA)	ATP		
20	HS2	Mg 2+	ssRNA (AGACUA)	ADP	PO4 (2-)	1500

21	HS2	Mg 2+	ssRNA (AGACUA)	ADP	PO4 (2-)	1500
22	PDB: 7AY4	-	DNA (AGACTA)	-	-	1500
23	PDB: 7AY4	-	DNA (AGACTA)	-	-	800
24	Last frame of Simulation 23 (LFS3)	Mg 2+	DNA (AGACTA)	ATP	-	2000
25	Frame 563 of Simulation 22	Mg 2+	DNA (AGACTA)	ATP	-	1500

Table 7: Summary of parameters used in the HADDOCK webserver. All distances are given in Ångström while energies are given in Joules.

Parameter	Setting
Distance Restrains:	
DNA/RNA restrains	*If one of your molecules is DNA/RNA, restraints are automatically created to preserve its structure. Uncheck this option if you are docking with unstructured DNA/RNA.*
Removal of non-polar hydrogens	*HADDOCK deletes by default all hydrogens except those bonded to a polar atom (N, O). Uncheck this option if you have NOEs or other specific restraints to non-polar hydrogens.*
Force constant for centre of mass contact restrains	1.0
Force constant for surface contact restrains	1.0
Random exclusion of a fraction of the ambiguous restrains (AIRs)	On
Number of partitions for random exclusion	2.0000
Gyration radius	17.78
Sampling Parameters:	
Number of structures for rigid body docking	1000
Number of trials for rigid body docking (RBD)	5
Sample 180° rotated solutions during rigid body EM	On
Number of structures for semi-flexible refinement	200
Preform final refinement	On
Solvent of use	Water
Number of structures to analyse	200
Number of structures for final refinement	200
Solvate Docking Model	Off
Cut-off of flexible segments of neighbouring residues of contacts	5.0
Clustering parameters:	

Clustering method	Fraction of Common Contacts (FCC)
RMSD cut-off for clustering	0.6
Minimum cluster size	4
Dihedral and hydrogen bond restrains:	
Automatically define backbone dihedral restraints from structure	Off
Error bound for the dihedral restraints	10
Energy and interaction parameters	
Non-bonded parameters	OPLSX
Include electrostatic during RBD	On
Distance dependent (rdie) dielectric	On
Scaling of intermolecular interactions for RB EM	1.0
Rigid body dynamic	0.001
SA with flexible side chains	0.001 (initial) 1.0 (final)
SA with flexible backbone and side chains	0.05 (initial) 1.0 (final)
Analysis parameters:	
Limited analysis of results	Cluster
Cutoff distance proton-acceptor to define hydrogen bonding	2.5
Cutoff distance carbon-carbon to define hydrogen bonding	3.9