The University of South Bohemia in České Budějovice
Faculty of Science

Study of Viral Protein Complexes using Mass Spectrometry

Bachelor thesis

Markus Mairhofer

Advisor: Ph.D. Dyčka Filip, Mgr.

České Budějovice 2024

Mairhofer M., 2024: Study of viral protein complexes using Mass Spectrometry, Bc. Thesis, in English 47 p. Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic.

Annotation

This thesis studied the tick-borne encephalitis virus non-structural protein domains NS3 helicase and NS5 RdRp and their protein-protein interaction regions using Mass Spectrometry techniques. Cross-linking Mass Spectrometry and Hydrogen-Deuterium Exchange Mass Spectrometry were employed to investigate regions and peptides involved in NS3-NS5 interaction. The obtained data can be used to further investigate complex formation and methods of shutting down the replication complex of the virus.

Declaration

I declare that I am the author of this qualification thesis and that in writing it I have used the sources and literature displayed in the list of used sources only.

Linz, 08.05.2024.

……..

Mairhofer Markus

Acknowledgements

I hereby thank my supervisor Ph.D. Dyčka Filip, Mgr. for the opportunity to work on this bachelor thesis and for the guidance throughout. I also want to thank Zdeněk Franta, Petra Havlíčková and Paulina Duhita Anindita for the preparation of the proteins for our experiments.

Lastly, I want to thank my family, friends and colleagues who always supported me on my academic journey.

Abstract

The tick-borne encephalitis virus (TBEV) is a Flavivirus commonly found in Europe, Siberia, and the Far-East. Infection can cause mild symptoms to fatal encephalitis and there is no anti-viral treatment available. Therefore, it is imperative to better understand the viral life-cycle and mechanisms of replication. Similar to other Flavivirus, TBEV employs a replication complex formed by the viral non-structural proteins NS3 and NS5 to replicate and synthesis new viral RNA genome. Understanding the formation of the replication complex could lead to the development of anti-viral treatment aiming to shut down said complex. Therefore, in this thesis work the TBEV NS3 helicase domain and NS5 RdRp domain were studied using Cross-linking Mass Spectrometry (CLMS) and Hydrogen-Deuterium Exchange Mass Spectrometry (HDX) to gain insight into the regions involved in protein-protein interaction. CLMS uses a cross-linking agent which is able to bi-covalently bind to two different Lysine residues that are in proximity to each other (based on the agent's specific maximum length) revealing possible protein-protein interaction regions. Through this we were able to identify possible interaction regions between the Finger and Thumb domain of the TBEV NS5 RdRp and the Domain III of the NS3 helicase. Furthermore, through HDX we found that the deuterium uptake was lower in the complex state for the peptides (residues 76-85 in the Finger domain; residues 486-495, 500-509 and 598-613 in the Thumb domain) close to the cross-linked residues (Lys-90, -88 and -612) on the RdRp than in the single protein state, which suggests that these peptides were possibly involved in protein-protein interaction such as hydrogen bonding.

Table of contents

1. Introduction

1.1. TBEV

The tick-borne encephalitis virus (TBEV) is a virus of the genus Flavivirus, belonging to the Flaviviridae family. [1] [2] [3] Viral infection can cause mild fever to fatal encephalitis. TBEV is split into the European, Siberian and Far-Eastern subtype. [4] [5] Infection with the European subtype most often results in a biphasic course, with a viremic phase that can last up to 8 days, whereby, individuals with encephalomyelitis are likely to skip this phase, followed by an asymptomatic interval of 1-33 days, before reaching the second phase of meningoencephalitis. [6] The other two subtypes normally occur monophasic, bypassing the asymptomatic stage. [4] Tick-borne encephalitis (TBE) can also take on a chronic form, often linked to the Siberian subtype, which can occur as a gradually progressive form or be latent for years. [7] The main method of transmission of TBEV to humans is through the bites of infected ticks, although it has been shown that infection can also occur from the consumption of unpasteurized milk and cheese of domesticated animals. [8] So far, there are no specific antiviral treatment options for a TBEV infection, only preventative measures, such as the immunization via vaccination are effective. [9]

1.1.1. Virion structure and host cell entry

The mature spherical, 50 nano-meter wide flavivirus virion particle consists of a viral genome-capsid(C) complex core surrounded by a lipid bilayer and a smooth surfaced outer shell. The outer shell is composed of membrane-anchored glycoproteins E and M (**Figure 1**). [10]

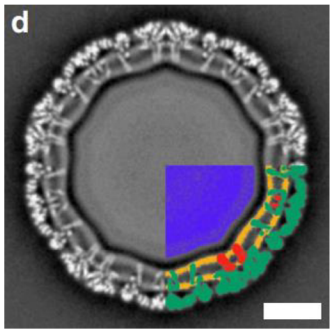


Figure 1 - Depiction of the central slice of a TBEV virion electron density map. The bottom-right quadrant shows the virus membrane (orange), the M-proteins (red) and the E-proteins (green), as well as the nucleocapsid core (blue). [11]

The envelope (outer shell) of flavivirus show a so-called herringbone pattern incorporated by three heterotetramers each consisting of two E-proteins and two M-proteins. These heterotetramers are therefore considered to be the building blocks of the mature virion (**Figure 2**). [11]

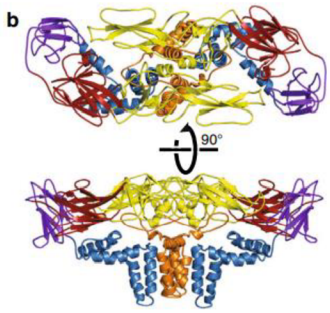


Figure 2 - Cartoon depiction of a heterotetramer. The two M-proteins are coloured in orange and the domain I of the E-proteins in red, the domain II in yellow and the domain III in violet. [11]

In order to enter a target cell, the ectodomains of the protein E of flavivirus first need to interact with host-cell receptors or other molecules on the cell membrane, with one of them being heparan sulfate, which can act as an attachment receptor. [12] In the case of TBEV a cellular entry factor has been identified in the T-cell immunoglobulin and mucin domain I (TIM-1). [13] Upon successful attachment the virus undergoes receptor-mediated clathrin-dependent endocytosis. [14] Hereby, the virus gets recruited to clathrin-coated pits that then pinch off from the membrane to form coated vesicles on the inside of the cell. [15] After endocytic uptake the viral membrane undergoes fusion with a cellular membrane dependent on the protonation of viral E-protein residue/s by the acidic pH in the endosomes. The process of fusion can however, also be initiated at the host-cell plasma membrane through cellular receptors. [16] After the fusion, the nucleocapsid is able to diffuse into the cytoplasm of the host-cell [17], where it is uncoated, releasing the the viral RNA genome into the cytoplasm. [18]

1.1.2. Genome translation and replication

The single-stranded, positive-sense RNA genome of TBEV/flavivirus has a length of around 11 kilo-bases and contains only one open reading frame (ORF), sandwiched between 5'- and 3'-untranslated regions (UTRs). Both UTRs are responsible for the coordination of RNA synthesis and the 3'-UTR is also able to interact with viral and host proteins to regulate RNA genome translation, as well as, play a role in the formation of vesicle packets. The ORF is translated into a single, large polyprotein by the host cell ribosomes. [19] This polyprotein, made up of circa 3400 amino acids, then gets cleaved by cellular and viral protease, leaving behind three structural proteins (capsid-C, pre-membrane/membrane-prM/M and envelope-E) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) (**Figure 3**). [20] [21] [22] [23]

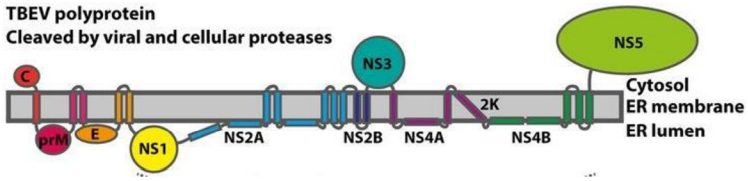


Figure 3 - Schematic overview of the proposed arrangement of the polyprotein composed of structural and non-structural viral proteins on the host cell endoplasmic reticulum. [23]

The genome replication takes place in the cytoplasm, in virus-invoked replication compartments at the host cells endoplasmic reticulum, which are comprised of vesicles and an enclosed extravesicular area. These compartments are able to provide an ideal environment for viral genome replication and protecting the replication machinery as well as, the newly synthesized RNA from the scrutiny of the host cells immune system (**Figure 4**). [24]

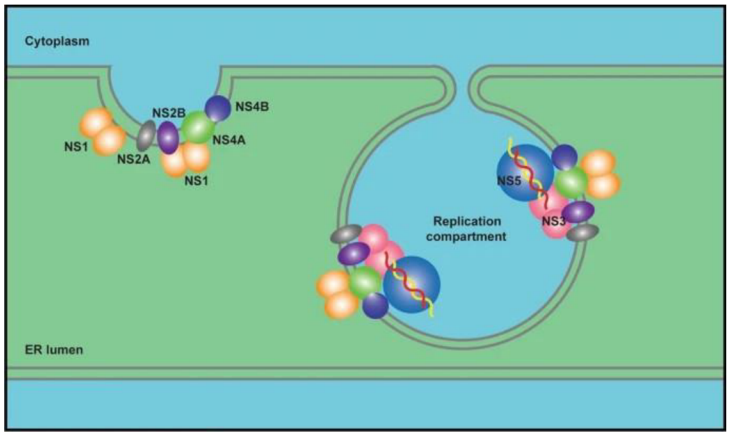


Figure 4 - Schematic overview of a replication compartment formed by viral membrane-active NS proteins. In these compartments the replication complex is established for the synthesis of new viral RNA. [25]

These replication compartments are likely to be established through the combined effort of different non-structural proteins, with the most probable suspects being NS1, NS2A and B, NS4A and NS4B. For successful virus replication, the replication complex (RC), consisting of the NS3 and NS5 protein, as well as, the viral RNA must be recruited to the replication compartment. This is accomplished by NS proteins who possess an ER anchor. [25] NS3 gets recruited by NS2B [26], however, the recruiter for NS5 has not yet been identified, although it is speculated that NS3 may play a role in NS5 recruitment. [27] The viral RNA is most likely recruited by NS5, as the binding of the 5'-UTR's stem-loop A (SLA) to NS5 was found to initiate the viral replication process. It has been suggested that NS5 binds SLA at the same time as recruiting the 3'-end of the RNA genome to the template-binding channel without overlap. Then, during the synthesis of the negative-strand RNA template, SLA stays bound until it is needed to finish the template. This guarantees the synthesis of complete genomes through intact viral RNA templates. [28]

After the replication process is finished, newly-synthesized viral RNA is encompassed by viral C-proteins to form nucleocapsids (NCs). This step likely occurs in proximity to the replication compartments. The freshly assembled NCs then bud into the ER lumen through the membrane to obtain a lipid envelope and the proteins prM and E, forming so-called immature particles. [29] These particles continue to travel through the Golgi apparatus and the trans-Golgi network where the lower pH incites conformational changes that lead to the formation of mature particles. The trimeric prM-E spikes on the surface of the particle rearrange to the heterotetramers associated with mature particles, exposing the pr peptides to cleavage by a host protease called furin. [30] The pr remains bound to the E-M-M-E heterotetramer until the now mature particle leaves the cell via endocytosis arriving in a neutral environment. The pr peptide is not able to dissociate under acidic conditions, which avoids premature membrane fusion in the trans-Golgi network (**Figure 5**). [31]

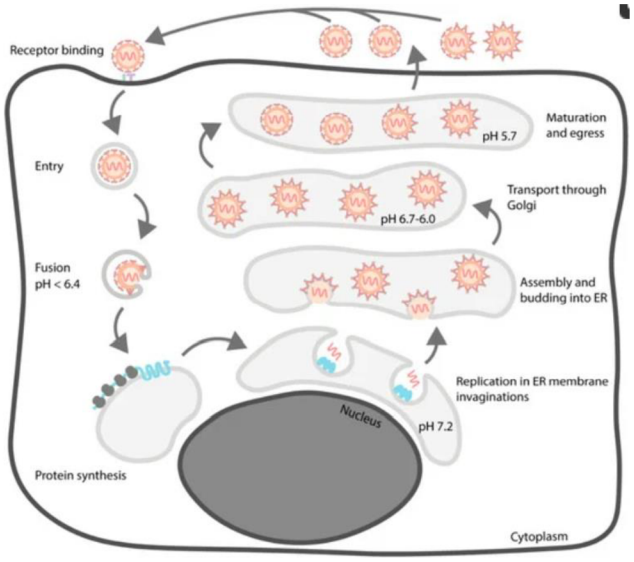


Figure 5 - Schematic overview of the life cycle of the tick-borne encephalitis virus. From host-cell entry to the replication at the endoplasmic reticulum and the budding and assembly of the viral particle at the ER lumen to the subsequent transportation out of the cell and maturation of the virion to its infectious form. [29]

1.1.3. NS5

NS5 is the largest non-structural protein with a length of ~900 amino acids. The protein consists of two domains, a C-terminal RNA-dependant RNA polymerase (RdRp) and a N-terminal methyl transferase (MTase) domain respectively. These two catalytic domains are connected via a ~10 residue long inter-domain region. [32] The MTase of NS5 is necessary for guanine N-7 and ribose 2'-OH methylation in the RNA capping process, which in turn provides better stability for and more efficient translation of the viral RNA. [33] On the other hand, the RdRp of NS5 is critical for viral genome replication. As previously mentioned, the RdRp is able to bind the SLA at the 5' end at the same time as initiating *de novo* RNA synthesis at the 3' end of the (+) strand RNA. After the RdRp generates a complete (-) strand, the now double-stranded RNA is used as a template for the synthesis of new (+) strand RNA genome. [34]

The structure of the NS5 RdRp consists of three domains, the palm, finger and thumb domains, which surround the polymerase active site (**Figure 6**). In contrast to the global structure, the catalytic module structure of RdRp is relatively conserved among different positive-strand RNA viruses. However, it was found that the regions between the catalytic motifs B and C in flavivirus and Flaviviridae as well as, other virus families show strong structural variability that is likely connected to their specific hosts due to in part being solvent accessible. [35]

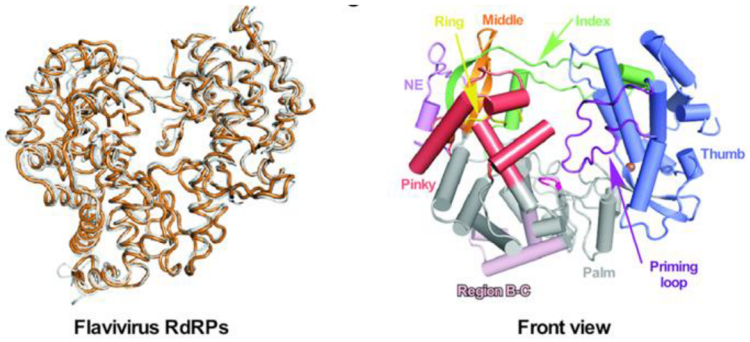


*Figure 6 - Cartoon depiction of the RdRps of different flavivirus (TBEV in dark orange) (**left**), next to the cartoon depiction of the TBEV NS5 RdRp crystal structure with its domains (**right**); The fingers are separated in pinky, ring, middle and index finger. The NE stands for N-terminal extension. [35]*

The TBEV RdRp also has an NTP interacting site, which is located between motif F (part of the finger domain) and the palm domain. Whereby, motifs A, C and E (part of the palm domain) constituted the so-called NTP-tunnel (**Figure 7**) and motifs E and F act as priming site, motif F as interrogating site and motifs C and E as catalytic site. [36]

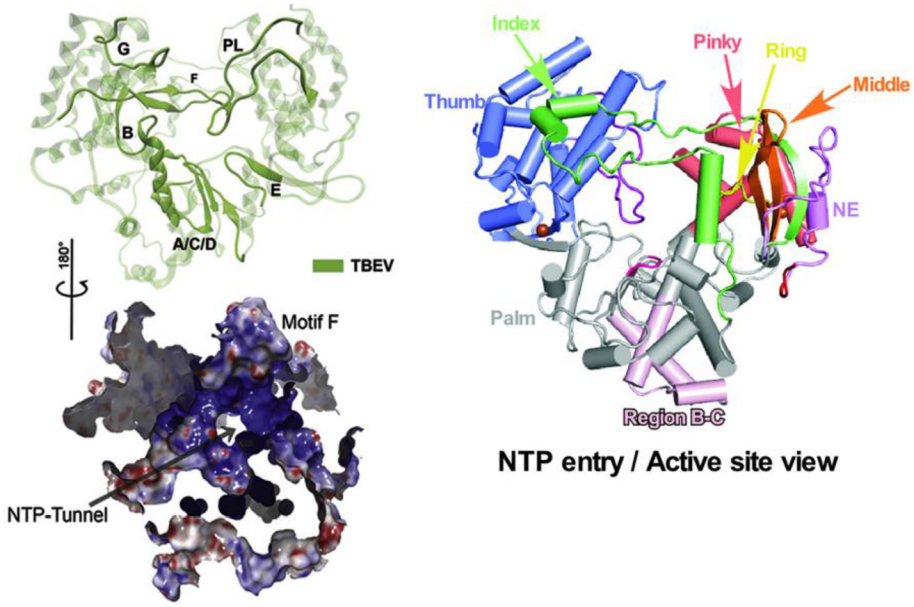


*Figure 7 - Cartoon depictions of the TBEV NTP relevant structural motifs (**top left**); Close-up of the NTP-tunnel and the potential surface charge distributions of the motifs involved (**bottom left**; white = neutral, blue = positive, red = negative) [36]; Cartoon depiction of the crystal structure of the TBEV RdRp with the view on the NTP entry/NTP-tunnel (**right**) [35].*

1.1.4. NS3

This non-structural protein possesses both a N-terminal protease and a C-terminal helicase domain. The role of the protease is the cleavage of the viral polyprotein; however, it first needs to be stabilized and activated by the hydrophilic region and the C-terminal region of NS2B to form the NS3-NS2B protease. [37] [27] The task of the C-terminal helicase domain (~434 amino acids long) is the unwinding of viral double-stranded RNA intermediate, which is supported by the energy provided from nucleotide triphosphate (NTP) hydrolysis through the NTPase of the helicase. Furthermore, the helicase has RNA 5'-triphosphatase activity, removing the terminal y-phosphate from the 5'-triphosphate end of the (+) strand RNA before the cap methylation of the MTase of NS5. [38]

The NS3 helicase consists of domain I, II and III, whereby, Motifs I, II and VI, which sit at the borders of domains I and II, take part in the binding of NTP to the NTP hydrolysis site that can be found between those two domains (**Figure 8**). On the other hand, the involvement of domains I, II and III is likely necessary for the binding of viral RNA, as evidenced by four specific RNA recognition residues, conserved in other flavivirus, that can be found as part of a positively charged channel between the three domains (**Figure 9**). [39]

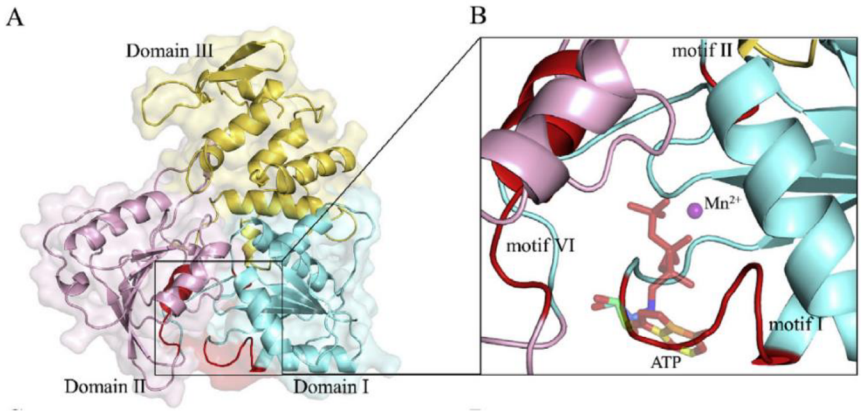


Figure 8 - Cartoon depiction of the NS3 helicase of TBEV with a close-up of the NTP binding site, which is able to bind divalent metal cations and ATP, and the Motifs I, II and VI involved in the binding process. [39]

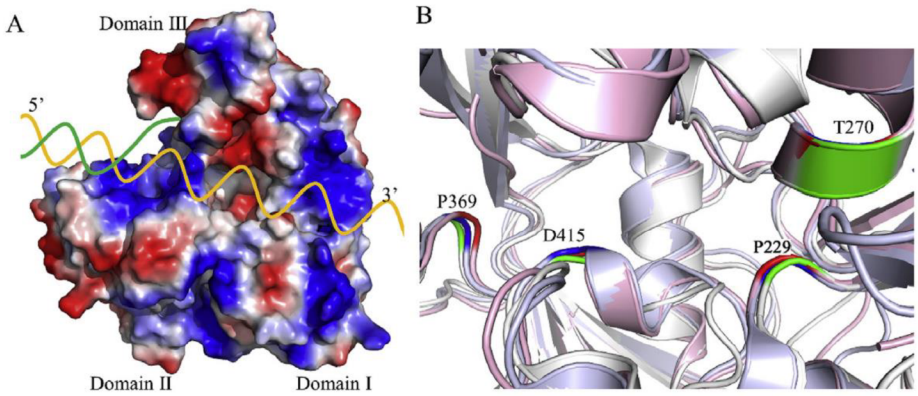


Figure 9 - On the left; surface charge distribution map of the NS3 helicase of TBEV (blue = positively charged; red = negatively charged) with double stranded RNA binding, whereby one strand (orange) moves along the channel and the other winds around the back the helicase (green). On the right: cartoon depiction of a close-up of the residues involved in RNA-recognition of different flavivirus (TBEV (red), ZIKV (green), and DENV-4 (blue)) [39]

1.1.5. NS3-NS5 interaction

The interaction between NS3 and NS5 is vital for successful replication of flavivirus. To summarize, the RdRp of NS5 synthesizes a (-) strand RNA from the genomic (+) strand RNA to generate a double stranded intermediate, which is then unwound by NS3, whereby, the (-) strand is used as template for the synthesis of multiple new (+) strand genome. [40] It has been found that the unwinding of dsRNA by the NS3 helicase is directly supported by interaction with NS5, stimulating the NTPase activity of NS3, most likely to couple the unwinding process with the synthesis of new viral RNA. [40] [41] Furthermore, two interaction regions have been identified in residues 566-585 of NS3 and residues 320-341 of NS5 of the dengue virus, whereby, the negative surface charge on NS3 and the positive one on NS5 in these regions indicates a charge-based interaction. The specific interaction points appear to be residue 570 of NS3 and residue 330 of NS5. The former is an asparagine residue in most flavivirus except in yellow fever, where it is replaced by a histidine. Similarly, residue 330 of NS5 is a lysine in most flavivirus but a tyrosine in the yellow fever virus. [42]

1.2. Cross-linking Mass Spectrometry (CLMS)

Cross-linking Mass Spectrometry (CLMS) is a complementary structural elucidation method, that can provide supporting information to traditional structural elucidation tools such as, X-ray crystallography, nuclear magnetic resonance (NMR) and cryogenic electron microscopy (Cryo-EM) as well as, identify structural information on protein complexes in their native environment (e.g. in the cell). [43] [44] The main attraction of CLMS is the specific distance restraint imposed by the crosslinking agent that is used, which provides region-level insight into the proximity of protein subunits. A cross-linking agent binds two residues that are close to each other covalently, whereby, there is an upper limit to the distance between these two residues since the length of the cross-linking agent cannot be exceeded. [45] The most frequently used cross-linking agents are homobifunctional primary amine selective and are able to cross-link two lysine residues or lysine residues to the amino terminus. [45] [46] The typical workflow of a CLMS measurements starts with the reaction of target proteins and protein-complexes with the chosen cross-linking agent, after which the proteins are digested by an enzyme such as trypsin to generate a mixture of peptides. These peptides are then separated by liquid chromatography and analysed via mass spectrometry. Lastly, the cross-linked peptides and their link positions are identified through a search of the MS database (**Figure 10**). [47]

Hereby, the software used to analyse the MS data can be dependent on the type of cross-linking reagent employed, however most software is able to work with all reagents. For example, software such as XlinkX or MeroX can be used for the analysis, whereby, cleavable reagents profit from MS spectra observable specific cross-link reporter ions, which facilitate the identification of intact cross-linked peptide masses. [48] Afterwards, the cross-link data can be viewed through visualization tools such as xiNET, which is able to portray the data as a node-link diagram that includes the protein sequences and linkages sites as well as, linked peptides in relation to the complete sequence (and therefore, in relation to domains etc.) and even cross-linking reaction products. All the software needs are the cross-link data, the protein sequence data and optional annotation data. [49] The identified cross-links can also be visualized in a 3D model of the proteins/protein-complexes. For this, software such as, pyMOL is used. PyMOL is able to depict the 3D structure of the target proteins in ribbons, cartoons, dots, surfaces, spheres, sticks and lines. Furthermore, pyMOL is built on the programming language Python, which makes it accessible for Python plugins. [50] The identified cross-link positions and the corresponding cross-links can then be manually added onto the 3D structure or a script can be written to automize the process. [51] A new plugin by the name of PyXlinkViewer is also able to automatically map all of the experimentally determined intra- and intermolecular cross-links onto the 3D model as well as, measure the distance between two cross-linked residues to determine if the threshold set (dictated by the length of the cross-linking agent) is violated or satisfied. [52]

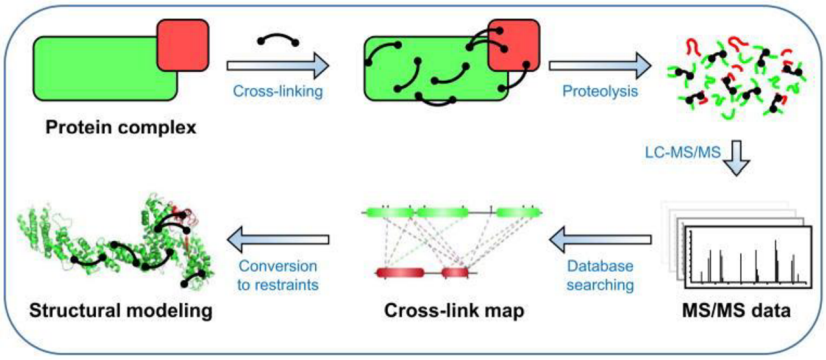


Figure 10 - Illustration of the workflow of an CLMS measurement. (Proteolysis = enzymatic digestion of proteins; LC-MS = liquid chromatography-mass spectrometry; Conversion to restraints = structure elucidation through the distance restraints of the cross-linker) [47]

1.3. Hydrogen-Deuterium Exchange Mass Spectrometry (HDX MS)

The underlying principle of Hydrogen-Deuterium Exchange (HDX) is the exchange of backbone amide hydrogens of amino acids with deuterium (D). This exchange reaction is catalyzed by OD^- and D_3O^+ ions as well as D_2O, the mathematical relationship of which is described in **equation 1**. Beside the pH, the exchange rate is also dependant on temperature, which is tied to the activation energies of exchange reactions and ionization of D_2O. [53]

1 - General equation for the 'chemical' exchange rate constant k_{ch} of a backbone amide, dependent on the intrinsic rate constants (k_{int}) of the base-, acid-, water-catalyzed exchange reactions. [57]

$$k_{ch} = k_{int,H}[D_3O^+] + k_{int,OH}[OD^-] + k_{H2O}[D_2O]$$

Furthermore, the neighbouring peptide side chains and therefore, the amino acid sequence influence the exchange rate through inductive or steric blocking effects. [53] Additionally, amides within a protein structure are either protected or unprotected depending on the state of the protein. If the protein changes conformation or unfolds, affected amides are solvent accessible and can be subjected to exchange at a specific rate (k_{ch}), which is described by **equation 2**. Accessibility is mostly dependant on hydrogen bonding and the degree of solvent occlusion (for e.g., in hydrophobic regions). [54]

2 - Equation for the exchange of a backbone amide in a protein structure. cl = closed confirmation of protein, amide not accessible for exchange; op = open confirmation of protein, amide accessible for exchange; k_{cl} = rate constant of closing; k_{op} = rate constant of opening. [57]

$$NH_{cl} \overset{k_{cl}\ k_{op}}{\longleftrightarrow} NH_{op} \overset{k_{ch}}{\longrightarrow} ND_{op} \overset{k_{op}\ k_{cl}}{\longleftrightarrow} ND_{cl}$$

Assuming the protein native state is the closed state, the exchange rate in the closed state (k_{ex}) can be simplified as the product of the frequency of opening and the fraction of exchanged hydrogens at a specific position during the opening, as seen in **equation 3**. [55]

3 – [55]

$$k_{ex} = \frac{k_{op}k_{ch}}{k_{cl} + k_{ch}}$$

With this in mind, there are generally to regimes to consider when it comes to protein dynamics. The first (EX1) occurs when $k_{ch} >> k_{cl}$, this means that k_{ex} is equal to k_{op}, therefore, with each opening comes a complete exchange reaction of all affected amides. The second regime (EX2) takes place when $k_{cl} >> k_{ch}$, meaning that $k_{ex} = K_{op}*k_{ch}$, with $K_{op} = k_{op}/k_{cl}$. Hereby, measuring k_{ex} and k_{ch} is directly tied to the free energy of opening of the protein as seen in **equation 4**. [56]

4 - Equation for the free energy of opening, whereby, Pf is the protection factor of a backbone amide resulting from dividing the exchange rate in the closed state k_{ex} by the intrinsic exchange rate k_{ch}. [56]

$$\Delta\Delta G_{op} = -RTln\left(\frac{k_{ex}}{k_{ch}}\right) = -RTln(Pf)$$

It has been found that all proteins fall into the category of EX2 kinetics, and it is rare that a protein undergoes EX1 kinetics. However, by increasing the pH or using a denaturant, it is possible to cause EX1 kinetics to occur. Furthermore, proteins can exhibit both EX1 and EX2 kinetics in the same region and the identification of which regime is present in which region can be done by the characteristic isotope patterns in a mass spectra. [57]

Coupled with Mass Spectrometry, HDX can provide information on the dynamics and confirmation of proteins by measuring the deuterium uptake at various time intervals. Hereby, the uptake can be measured globally or more popularly, by quenching the exchange reaction and proteolytically cleaving the protein before MS analysis. [58] Since deuterium is the heavier hydrogen isotope, the mass of measured peptide fragments will increase over time and the mean mass shift is indicative of the percentage of deuterium uptake at any given time. This can be used for e.g., to identify ligand-binding sites, as the peptides involved will not be available for exchange due to already partaking in hydrogen bonding with the ligand. [59]

The typical workflow of HDX MS experiments starts with the incubation of the previously purified target protein/s in deuterated buffer solution (usually 10-fold dilution of the sample) to start the exchange reactions. The exchange is then quenched with cold trifluoroacetic acid or formic acid solution at around 2.5 pH. Following this, the sample is digested, usually directly on a liquid chromatography pepsin column, and washed with acetonitrile in a trap column (desalting) to then be hydrophobically separated and eluted with a solvent gradient on a C18 column. After the elution an electrospray ionizes the peptides and they are analysed via time-of-flight or orbitrap mass analysers. The exchange can then be analysed and visualized with the corresponding software. [60]

2. Objective

The aim of this thesis work was the structural analysis of the two non-structural protein domains NS3 helicase and NS5 RdRp of the tick-borne encephalitis virus and especially of the replication complex that is formed by these two proteins using Cross-linking Mass Spectrometry and Hydrogen-Deuterium Exchange Mass Spectrometry. The formation of this complex is vital for the replication of viral positive-sense single-stranded RNA genome and structural elucidation could proof useful in the design of currently unavailable antiviral drugs. Through CLMS possible interaction region/points between the two proteins can be identified by inter-molecular cross-links, whereby, the distance restraint imposed by the cross-linking agent is indicative of the proximity of the involved residues. Similarly, through HDX the deuterium uptake of the two proteins individually can be compared to the uptake of the protein complex and therefore, possible interaction regions can be identified based on the difference in deuterium uptake for those regions.

3. Materials and Methods

3.1. Chemicals

High-quality solvents for liquid chromatography/mass spectrometry were from VWR Chemicals (Czech Republic). PierceTM trypsin MS Grade, cross-linking agent bis(sulfosuccinimidyl)suberate (BS3) were purchased from Thermo Fisher Scientific (USA). EmporeTM Solid Phase Extraction Disks C18 used for preparation of Stage Tips were obtained from Merck (Germany). All other chemicals, including deuterium oxide (99.9%), were of analytical purity grade from Merck (Germany). Pepsin and HPLC columns, syringes and other material for HDX MS were from Waters (USA). CaptiveSpray and other accessories for timsTOF instrument were from Bruker (USA) and nanoLC columns were from Thermo Fisher Scientific (USA).

3.2. Protein preparation

The two proteins NS5 RdRp and NS3 helicase used for CLMS and HDX were prepared by Zdeněk Franta, Petra Havlíčková [61] and Paulina Duhita Anindita [38].

3.3. Cross-link Mass Spectrometry

For the results depicted in **Table 2** the proteins were cross-linked using 1.3 mg of fresh bis(sulfosuccinimidyl)suberate (BS3, Thermo Fisher Scientific) diluted in 227.1 μL HEPES buffer to a concentration of 10 mM BS3. The buffer used for the incubation was composed of 20 mM HEPES and 150 mM NaCl. The mixtures of NS5 RdRp and NS3 helicase were prepared to the ratios of 1:1, 1:4 and 4:1 of NS3:NS5 with a total protein concentration of 4 μM and BS3 concentration of 800 μM. The measurements done on the proteins individually had the same total protein concentration. Before the BS3 was added the proteins were mixed together on ice and left to sit for 30 minutes. After that, the incubation with the cross-link buffer took another 30 minutes at room temperature before being quenched by 100 mM ammonium bicarbonate (Sigma). The proteins were then reduced with 10 mM dithiothreitol (DTT) for 40 minutes at 56 °C and subsequently alkylated using 55 mM iodoacetamide (IAM) in the absence of light. The alkylation was stopped by incubation in 50 mM DTT for 15 minutes.

To digest the proteins, Trypsin was dissolved in buffer and 80 μL (0.2 μg Trypsin) were added to the sample for an overnight digestion. Following this, the peptides were purified using C18 Stage Tips ([62]) and the sample was dissolved in 20 μL of 3% acetonitrile, 0.1% formic acid, of which 2 μL were used for timsTOF Mass Spectrometry analysis. The separation of the peptides was done via HPLC, whereby, the sample was loaded onto a AcclaimTM PepMapTM 100 C18 trapping column (300 μm i.d., 5 mm length, 5 μm particle size and 100 Å pore size; Thermo Fisher Scientific). After 2 minutes the peptides were eluted onto a AcclaimTM PepMapTM 100 C18 analytical column (75 μm i.d., 150 mm length, particle size of 2 μm and 100 Å pore size; Thermo Fisher Scientific) and separated by linear gradient of 5-35% acetonitrile/0.1% formic acid at a flow rate of 0.3 μL/min for 48 minutes. The temperature of the column oven was set to 35 °C and the results were collected via PASEF scan mode with positive polarity. The ionization was done by Electrospray using a CaptiveSpray (Bruker Daltonics, Bremen, Germany) with a capillary voltage of 1500 V, as well as a flow rate of dry gas at 3 L/min and a dry temperature of 180 °C. The ions were gathered for 100 ms and 10 PASEF MS/MS scans were collected via topN acquisition cycle with an ion mobility range (1/K0) of 0.6-1.6 Vs/cm2. The mass spectra were obtained in a m/z range of 100-1700. In order to filter out the single charged ions of low m/z a polygon filter was used. The collision energies were ramped up from 20-59 eV in 5 steps of identical width between 0.6 and 1.6 Vs/cm2 of 1/K0 values. Together with the protein sequences in form of FASTA files, the raw data was loaded up in the MaxQuant software v2.0.3.0 ([63]). The default settings for the BS3 cross-linker were used for the peptide search. The precursor precision was set to 5 ppm and fragment ion precision to 10 ppm. The false discovery rate (FDR) cut off was set to 1%. A decoy database was established as a reversed sequence. Cysteine carbamidomethylation and methionine oxidation were used as static and variable modifications.

3.4. Hydrogen-Deuterium Exchange Mass Spectrometry

As labelling/equilibration buffer 20 mM Tris and 150 mM NaCl were used (pH/pD 7). Proteins NS3 helicase and NS5 RdRp were kept separately or mixed together in equilibration buffer at 4 °C. For labelling, 4 µL of protein solution was added to 56 µL of labelling buffer and incubated at 20 °C. Afterwards the labelling reaction was quenched with 1.6 M GHCl in 0.8% FA buffer at 0 °C. Each sample was injected into the HDX manager system (Waters, MA, USA) for an online digestion by Enzymate™ BEH Pepsin Column (Waters) connected to an Acquity UPLC M-Class system (Waters) with a constant flow rate of 100 µL/min of a 0.2% formic acid mobile phase. The resulting peptides were trapped on an Acquity BEH C18 Vanguard Column (1.7 µm; Waters) for 2 minutes before being eluted onto an Acquity BEH C18 analytical column (1.7 µm, 1 mm i.d., 100 mm length; Waters). The peptides were separated in a linear gradient of 8-35% acetonitrile/0.2% formic acid for 7 minutes with a flow rate of 40 µL/min. The deuterium labelling and injection into the UPLC system was automized by a PAL RTC system (LEAP technologies, USA). During the separation step the columns were kept in HDX manager chambers at 0 °C and lastly, the peptides were analyzed by a Synapt G2-Si mass spectrometer (Waters). For ionization ESI was used with a voltage of 3 kV for the capillary and 40 V for the sampling cone as well as, a source offset of 80 V. The source temperature was 100 °C and the desolvation temperature 250 °C. The cone gas flow was set to 50 L/h, the desolvation gas flow to 600 L/h and the pressure of the nebulizer gas to 6.5 bar. The measurements were performed in the MS^E or MS mode in positive polarity and the MS^E data was obtained in low- and high-energy modes whereas, the collision energy was increasing from 20 up to 40 V for the high energy mode. The ProteinLynx Global Server software (PLGS) V. 3.0.3 (Waters) was used for MS database searching against an in-house database containing the protein sequences as well as, common contaminants (Max Planck Institute of Biochemistry, Martinsried, Germany). The threshold for low and high energy was set to 200 and 40 counts respectively and the parameters for the database search were non-specific (enzyme specificity), methionine oxidation (variable modification), 4% (false discovery rate), minimum 1 (fragment ion matches per peptide), minimum 7 (fragment ion matches per protein), 3 (peptide matches per protein). The software DynamX V. 3.0 (Waters) was used for the HDX data analysis. Only peptides with a minimum intensity of 500, minimum sequence length of 5 and minimum number of products of 2 were processed. Manual correction of the DynamX data had to be made. The deuterium uptake plots were visualized in DynamX and the relative fractional uptake at all labelling times was exported from DynamX and loaded into the pyMOL software V. 2.5.2 for 3D visualization.

3.5. Structure prediction of the NS5 RdRp; Schematic representation and 3D visualization

The structure of the NS5 RdRp protein was predicted using the software pyhre2 ([64]) with the default and normal mode settings as well as, the addition of His-Tag sequence at the start of the protein sequence. The NS3 helicase structure was taken from the website: https://www.rcsb.org/structure/7OJ4. The schematic representation of the CLMS results was done with the help of the software Xinet ([49]) and similarly to the HDX results the 3D visualization was done in pyMOL. For the schematic representation the CLMS results were gathered in a csv file and shortened to only include Protein1, Protein2, Link-Pos.1, Link-Pos.2 and Score. The csv file was then uploaded to Xinet together with the FASTA file containing both protein sequences.

Lastly, a pyMOL-Python script was used to visualize the cross-links on the 3D structure in pyMOL.

4. Results

4.1. Cross-link Mass Spectrometry

The Protein NS3 helicase (7oj4) as well as the protein NS5 RdRp (tbevRdRP8AA) were incubated with BS3, a crosslinker, after which they were digested by Trypsin and measured via Mass Spectrometry. The raw data were then analysed by the software MaxQuant ([63]), which identified the peptides and positions of the amino acid residues that were involved in cross-linking and assigned them a Score. A high Score is hereby indicative of a low probability of the cross-link being a false-positive. Furthermore, the distance between two cross-linked residues was calculated using the software pyMOL. BS3 alone has a length of 11.4 Å and together with the lysine residues it is attached to, a maximum length of 25 Å should be expected (the distance is calculated from the peptide bone alpha carbon instead of the lysine ends). However, a threshold of 30 Å was set to allow for more tolerance. The variability in the length of cross-links is a result of the flexibility of the cross-linker. Despite that, the distance of cross-links has to be 30 Å or less, since it is not possible for a cross-link to exceed the length of BS3 and both lysine residues combined.

Lastly, it has to be noted that the sequence numbering for pyMOL is different than the numbering for the actual FASTA sequences (for example, Ala-33 in FASTA is Ala-182 in pyMOL) as can be seen in **Figure 20**, **Figure 21** and **Figure 22**, **Figure 23**. Therefore, Link-Pos. refers to the protein sequence used in the MS experiments (**Figure 21** and **Figure 23**).

The following results were obtained over several measurements, each using multiple mixtures of NS3 and NS5 to different ratios, as well as measuring NS3 and NS5 separately. The results of the individual measurements were then put together in a single table (**Table 2**). Hereby, it was decided to not include mono-links (cross-linker is attached to only one residue), since they are irrelevant to the objective.

Overall, 138 unique cross-links were able to be identified, 108 of which are within the defined threshold of 30 Å (Xinet schematic representation: **Figure 11** and **Figure 12**). Notably, the number of intra-molecular cross-links located on the NS3 helicase are significantly higher than those on the NS5 RdRp, with 78 to 56. In total 18 of the unique cross-links are above the threshold which could meant that either the model which was used in pyMOL differs from the real structure of the proteins, especially that of NS5 RdRp, or the protein in question formed a multimer or oligomer with a second molecule of its kind, leading to an inter-molecular cross-link where a real distance cannot be calculated. Furthermore, 6 unique cross-links, namely Link-Positions 96-96 (Lys-244 in pyMOL), 457-457 (Lys-605 in pyMOL), 225-225 (Lys-373 in pyMOL), 451-451 (Lys-599 in pyMOL), 87-87 (Lys-235 in pyMOL) and again 87-87 (Lys-235 in pyMOL) on the NS3 helicase, possess a distance of 0, which indicates that they are self-linked, possibly due to the formation of multimers or oligomers of the NS3 helicase. However, from these results a definitive conclusion cannot be drawn. Lastly, 4 of the cross-links outside of the threshold are inter-molecular cross-links, namely Link-Positions 457-90 (Lys-605 NS3- Lys-93 NS5 in pyMOL), 457-88 (Lys-605 NS3- Lys-91 NS5 in pyMOL), 612-383 (Lys-615 NS5- Lys-531 NS3 in pyMOL) and 90-457 (Lys-93 NS5- Lys-605 NS3 in pyMOL), with an unknown distance. The determination of the length of inter-molecular cross-links was not possible due to a lack of understanding of the interactions between the two analysed proteins. It has to be noted that the inter-molecular cross-link between Lys-612 (Lys-615 in pyMOL) of the NS5 RdRp and Lys-383 (Lys-531 in pyMOL) of the NS3 helicase is not part of the schematic nor 3D representation, since this cross-link was only identified in one experiment.

The most prevalent Link-Position is 457 (NS3 helicase) with 23 involvements in cross-linking, 3 of which are inter-molecular cross-links. For the NS5 RdRp the most common Link-Position is 189, which participated in 11 cross-links, none of which are inter-molecular cross-links.

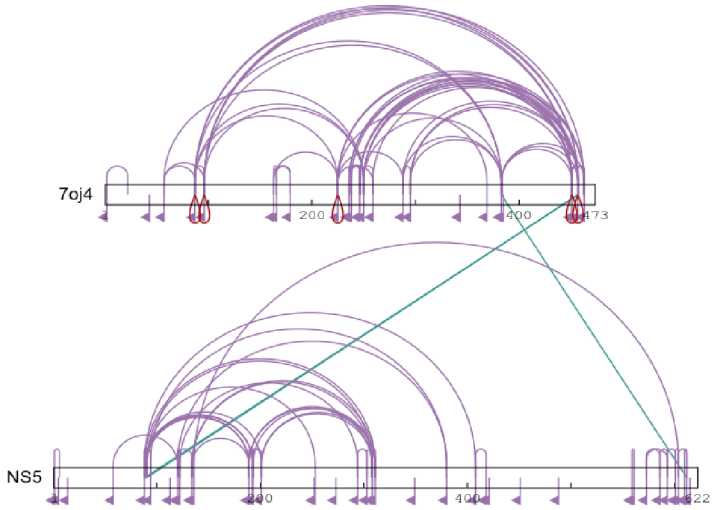


*Figure 11 - Schematic overview of the results from **Table 2** using the software xiNET [49]. The bars represent the protein sequences, the curved lines (arcs) the intra-protein cross-links, the small loops in red the self-linked cross-links and the straight lines in between the two proteins the inter-protein cross-links.*

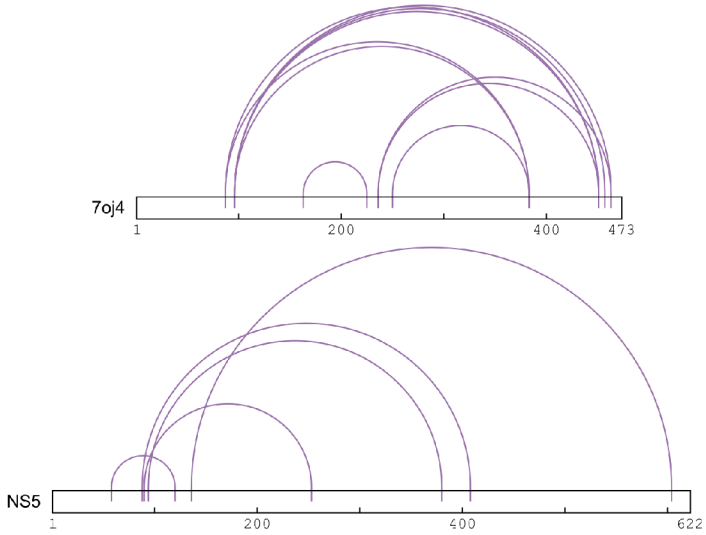


Figure 12 - Schematic overview of the cross-links from Table 1 that do not satisfy the threshold of 30 Å using xiNET [49].

To visualize the data gathered in **Table 2** the software pyMOL was used, which is able to depict the 3D structure of the NS3 helicase and NS5 RdRp with their respective cross-links as well as, calculate the distance between cross-linked peptides with respect to the 3D models used (**Figure 13**). The structure of the NS3 helicase is known, however, that of NS5 RdRp was predicted using phyre2 ([64]). Regarding the 3D model, the green coloured cross-links have a distance below the threshold of 30 Å, whereby red means that the threshold was crossed. This could be either due to the formation of multimers or oligomers as discussed above, or the real structure is slightly different than that of the model. A yellow colour indicates mono-links (cross-linker is linked to only one residue) and a violet colouration is for crosslinks on the same residue. Lastly, magenta-coloured cross-links represent the inter-molecular cross-links.

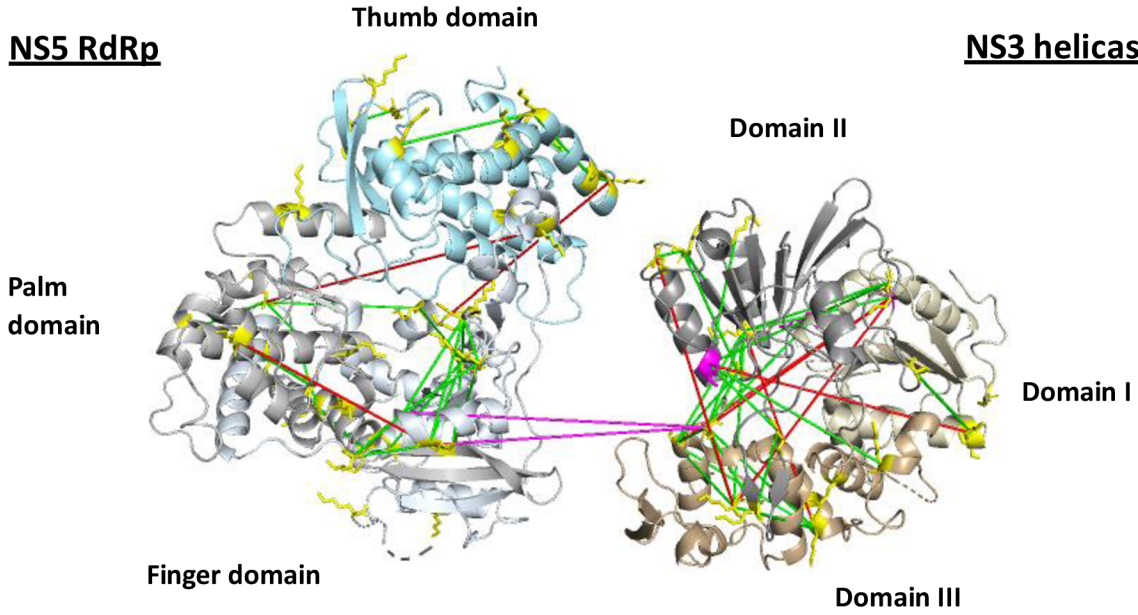
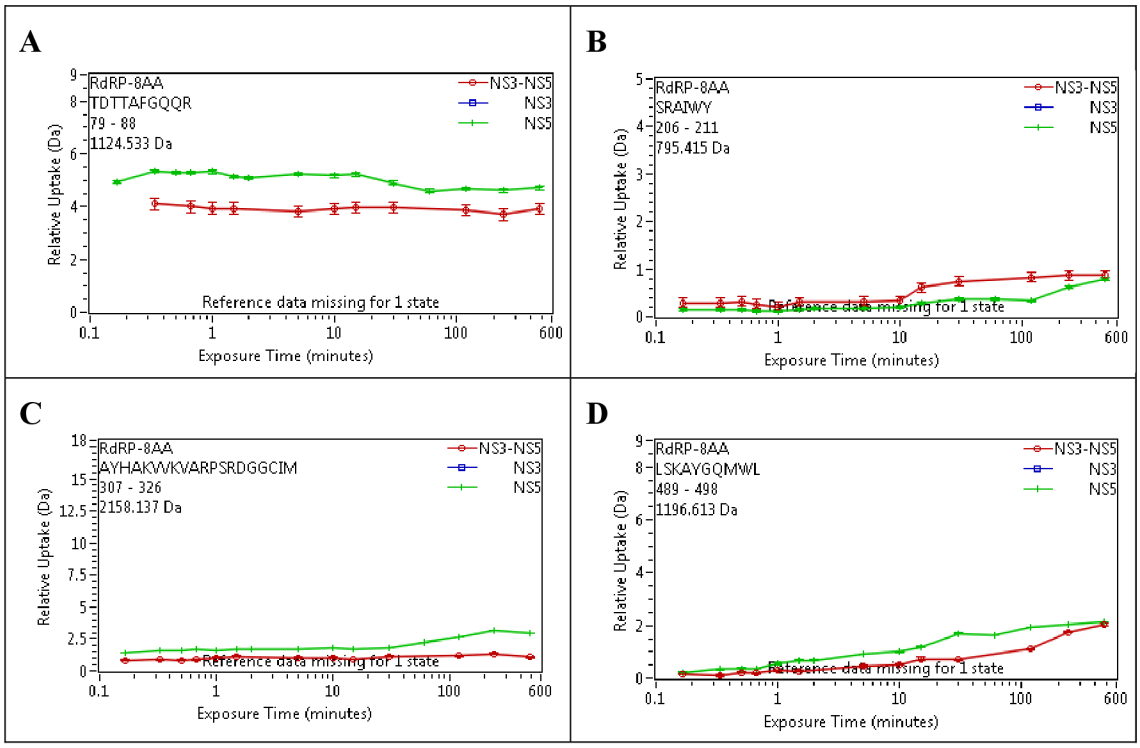


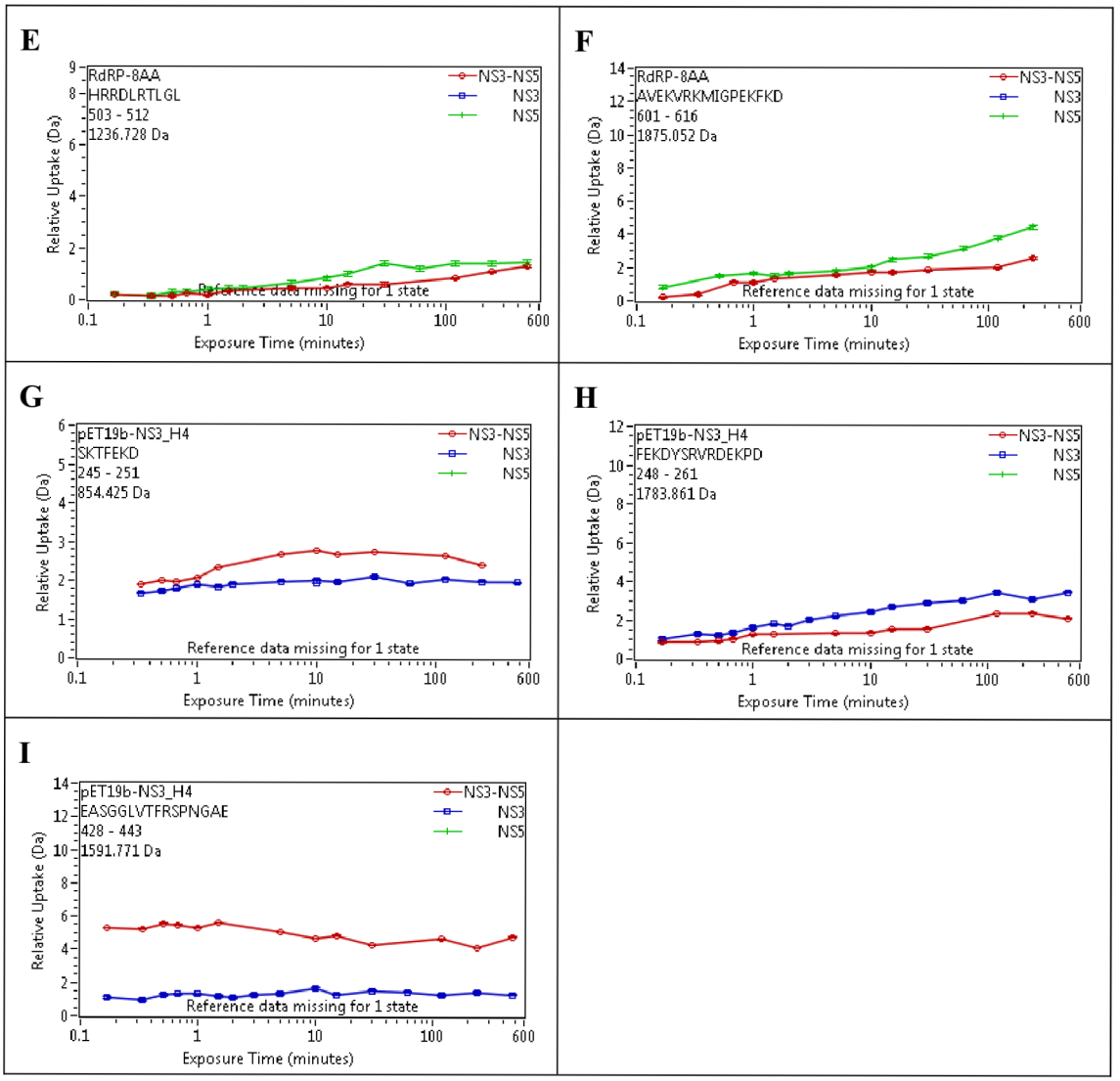
*Figure 13 - 3D model of the NS5 RdRp (left) and the NS3 helicase (right) including all of the cross-links from **Table 2**.*

4.2. Hydrogen-Deuterium Exchange Mass Spectrometry

The proteins NS3 helicase and NS5 RdRp were incubated separately and together in a deuterium labelling buffer for different time intervals ranging from 10 seconds to 8 hours. This was done in order to compare the deuterium uptake of the proteins alone and when they are in complex. After the incubation the deuterium exchange was stopped with a quenching buffer and the proteins were loaded into the HDX System, where they were then digested and the resulting peptides were separated by HPLC, analysed by Electrospray ionization MS/MS and the raw data was processed in DynamX (Waters), which also generated the deuterium uptake plots depicting the relative deuterium uptake for different states over time (for e.g., complex state and single protein state) and the peptides with the most significant changes in relative deuterium uptake were selected as seen in **Table 1**. For a 3D visualization of the deuterium uptake the DynamX data for different labelling times was exported to pyMOL as seen in **Figure 14** for the labelling time of 2 hours.

Table 1 - Uptake plots of the peptides that showed a significant difference in deuterium uptake between the single protein and complex state.





It has to be noted that, the amino acid numbering for the NS5 RdRp results on the uptake plots is identical to the numbering in pyMOL, however the numbering for the NS3 helicase results on the uptake plots is differently from the one in pyMOL.

For the residues 79-88 (**Table 1A**) found in the Finger domain of the NS5 RdRp the relative deuterium uptake was significantly lower for the complex state than for the single protein state for every labelling time as well as, the relatively high deuterium uptake reached its maximum within the first few seconds. For the residues 206-211 (**Table 1B**) found in the Finger domain the relative deuterium uptake was slightly higher for the complex state within the first 10 minutes of labelling and significantly higher until 8 hours of labelling time, where the relative deuterium uptake was the same for both states. For residues 489-498 (**Table 1D**) and 503-512 (**Table 1E**) found in the Thumb domain the opposite occurred, with the relative deuterium uptake being significantly higher for the single protein state between 10 minutes and 2 hours of labelling time and the same after 8 hours of labelling. For the residues 307-326 (**Table 1C**) and 601-616 (**Table 1F**) found in the Palm and Thumb domain of the NS5 RdRp respectively the relative deuterium uptake was slightly lower for the complex state for approximately 30 minutes of labelling time, however, after that the difference between single protein state and complex state uptake increased significantly with the largest difference at and after 1 hour and 30 minutes of labelling.

For the residues 245-251 (393-399 in pyMOL) (**Table 1G**) found in domain II of the NS3 helicase the relative deuterium uptake was slightly higher in the complex state than in the single protein state for approximately 1 minute of labelling time and then significantly higher from 1 minute to approximately 1 hour and 30 minutes. However, for the residues 248-261 (396-409 in pyMOL) (**Table 1H**) the single protein state saw slightly higher relative uptake than the complex state for 1 minute and 30 seconds of labelling and significantly higher uptake after that, with the largest difference at approximately 40 minutes of labelling time. Lastly, for the residues 428-443 (576-591 in pyMOL) (**Table 1I**) found in domain III the relative deuterium uptake was significantly lower for the single protein state than for the complex state for every labelling time and the relatively high deuterium uptake for the complex state reached its maximum within the first few seconds.

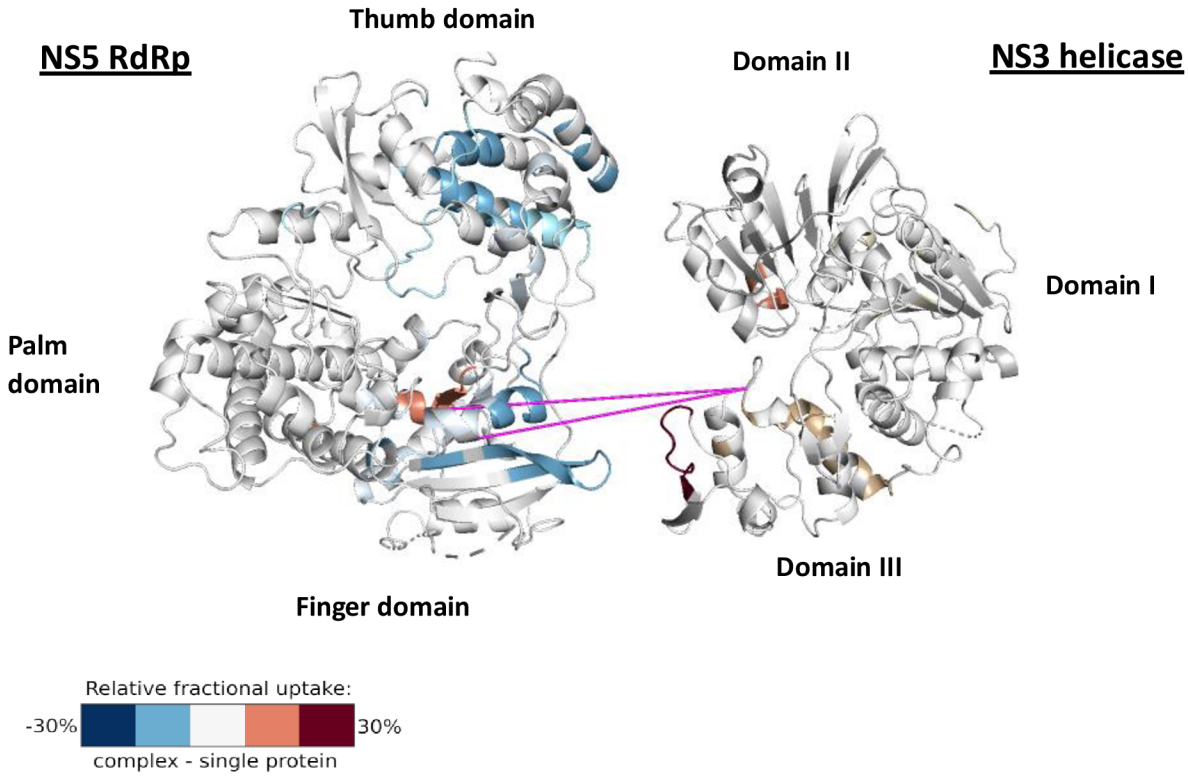


Figure 14 - 3D model of the proteins NS5 RdRp (left) and NS3 helicase (right) including the inter-molecular cross-links (magenta sticks) and the deuterium uptake for 2 hours of labelling time as a cartoon depiction.

5. Discussion

5.1. Cross-linking Mass Spectrometry

At the time of the Cross-linking Mass Spectrometry measurements a crystal structure of the TBEV NS5 RdRp was not yet available and therefore, for the 3D model a phyre2 predicted structure was used [64]. However, as of 2021 the crystal structure has been solved [35]. Compared to the crystal structure the predicted structure of TBEV NS5 RdRp looks to be relatively accurate (**Figure 15**). The intra-molecular cross-links of the NS5 RdRp that exceeded the set threshold of 30 Å (coloured in red) indicate that the distance between the cross-linked residues in the predicted model does not match the real structure.

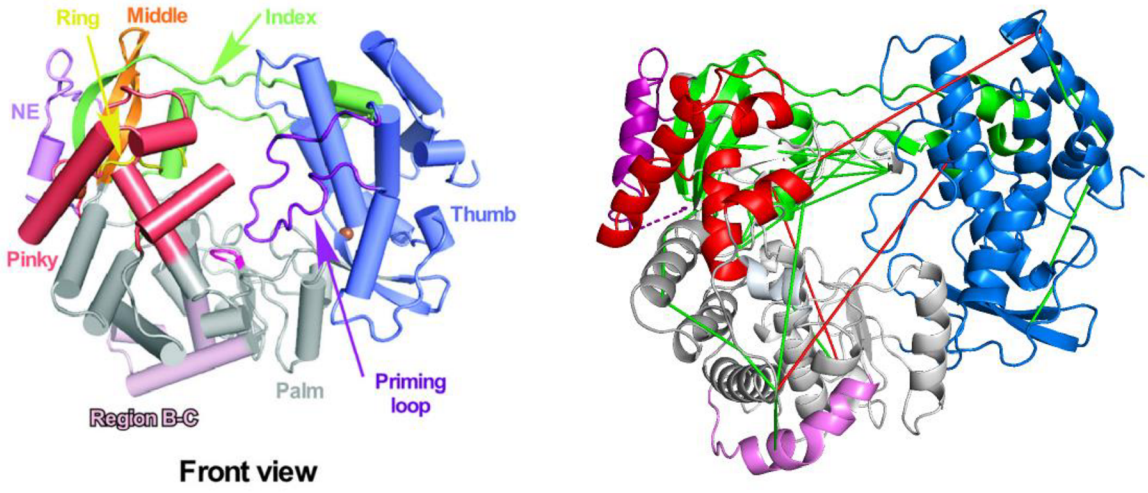


Figure 15 - Comparison of the crystal structure of TBEV NS5 RdRp (left) [35] to the phyre2 predicted 3D model (right); The Pinky finger domain is coloured in red, the NE domain in magenta, the Ring finger domain in yellow, the Index finger domain in green, the Thumb domain and the Priming Loop in blue, the Palm domain in grey and lastly, the B-C region in pink on the predicted model; including the threshold satisfied (green sticks) and unsatisfied (red sticks) intra-molecular cross-links.

The main objective of this thesis work was to provide information on the structure of the NS5 RdRp and NS3 helicase complex that is formed in order to efficiently replicate the viral positive-sence single stranded RNA genome. Therefore, the inter-molecular cross-links detected in the cross-linking experiments are of most significance, since they provide information on the regions of both proteins that are within interaction range of each other.

With this in mind, the inter-molecular cross-linked residues identified in this thesis work Lys-90, -88 and -612 (Lys-93, -91 and -615 in pyMOL) for the NS5 RdRp and Lys-457 and -383 (Lys-605 and -531 in pyMOL) for the NS3 helicase can be found in the Finger domain and the Thumb domain of the NS5 RdRp and domain III of the NS3 helicase respectively (**Figure 16** and **Figure 17**). These results show that the Finger domain and the Thumb domain are close to the domain III of the helicase (known to be involved in the interaction between the two proteins), presumably, while in complex with the RdRp. A previous study on another flavivirus, namely the dengue virus (DENV), revealed possible interaction residues (Asn-570 on the DENV NS3 helicase domain III and Lys-330 on the DENV NS5 RdRp Thumb domain) the former of which was suggested to be highly conserved among flavivirus within the possible interaction region (residues 320-341 on the DENV NS5 RdRp and residues 566-585 on the DENV NS3 helicase) [42]. In our case the DENV helicase residue Asn-570 corresponds to the residue Asn-572 (in pyMOL) and the residues 566-585 correspond to residues 568-587 (in pyMOL) (as seen by the sequence alignment found in the supplementary information of a previous study [39]). The DENV RdRp residue Lys-330 corresponds to the residue Trp-65 (in pyMOL) and residues 320-341 correspond to residues 55-76 (in pyMOL) (according to another previous study [65]). Furthermore, the amino acids surrounding DENV Lys-330 are conserved in TBEV (DENV: INGVVKLLTK(330)PWD; TBEV: INGVVKLLSW(65)PWN). The inter-molecular cross-linked residues Lys-93 and -91 (in pyMOL) are in the same domain as the residue Trp-65 (Finger domain of the TBEV NS5 RdRp), whereby, Trp-65 is directly between the Finger and Thumb domain, which was suggested to be involved in NS3-NS5 interaction, as seen in **Figure 16**. The two lysines each formed an inter-molcular corss-link with the Lys-605 (in pyMOL) of the TBEV NS3 helicase, which can be found in domain III close to the residue Asn-572, which was also suggested to be involved in NS3-NS5 interaction (**Figure 16**). Furthermore, the inter-molecular cross-linked residue Lys-615 (in pyMOL) of the TBEV RdRp formed its cross-link with the Lys-531 (in pyMOL) of the TBEV helicase, the former is located in the Thumb domain and the latter in domain III, which would put the residue Asn-572 between Lys-531 and Lys-605 of the NS3 helicase (**Figure 17**). Although it was not possible to determine the exact length of the inter-molecular cross-links it can be assumed that they are below 30 Å if the identified cross-links were not false-positives.

This would suggest that the Thumb domain of the TBEV RdRp and the domain III of the TBEV helicase are closer than shown in the 3D model (for reference the **Figure 17** depicts the back view with the positions of the cross-linked residues marked in red; The 3D model was created by merging the two proteins together in pyMOL and it doesn't fit the real interaction model which will be done by molecular dynamics in the future).

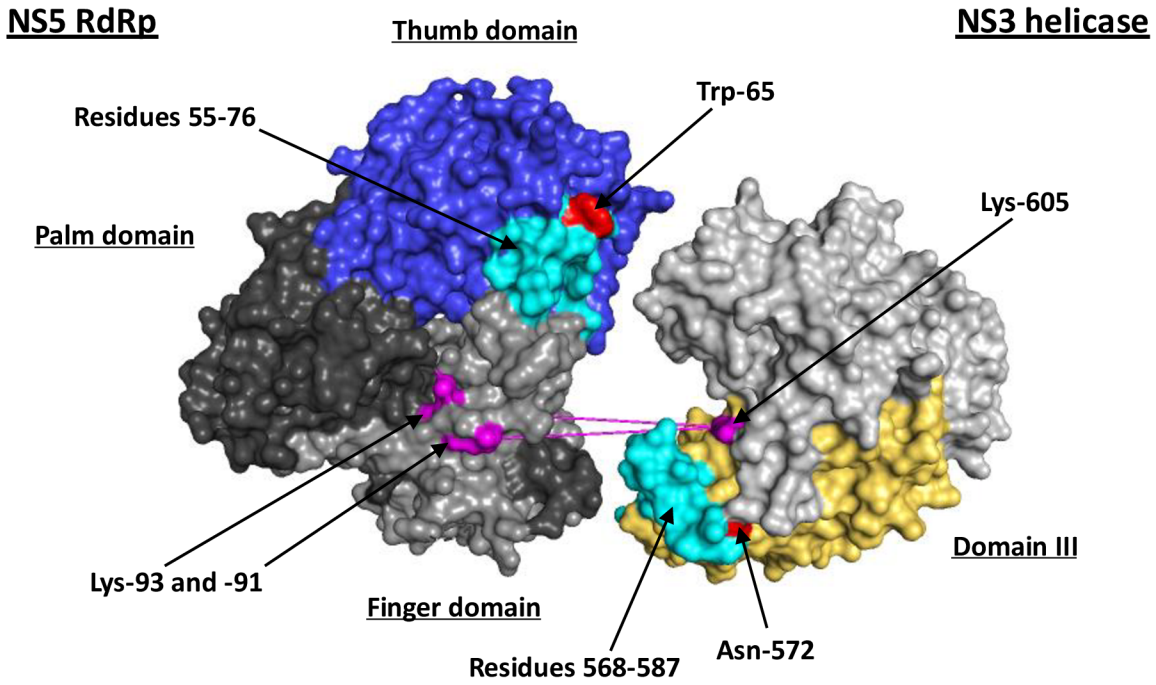


Figure 16 - 3D model of the NS5 RdRp (left) and the NS3 helicase (right) as surface depiction. The Palm domain of the NS5 RdRp is coloured in dark grey, the Finger domain in light grey, the residues Trp-65 and Asn-572 in red, the residues 55-76 and 568-587 in cyan and the residues Lys-93 and -91 in magenta; The domain III of the NS3 helicase is coloured in yellow-orange, the residues 566-585 in red and the residue Lys-605 in magenta; Lastly, the two inter-molecular cross-links are depicted as magenta sticks.

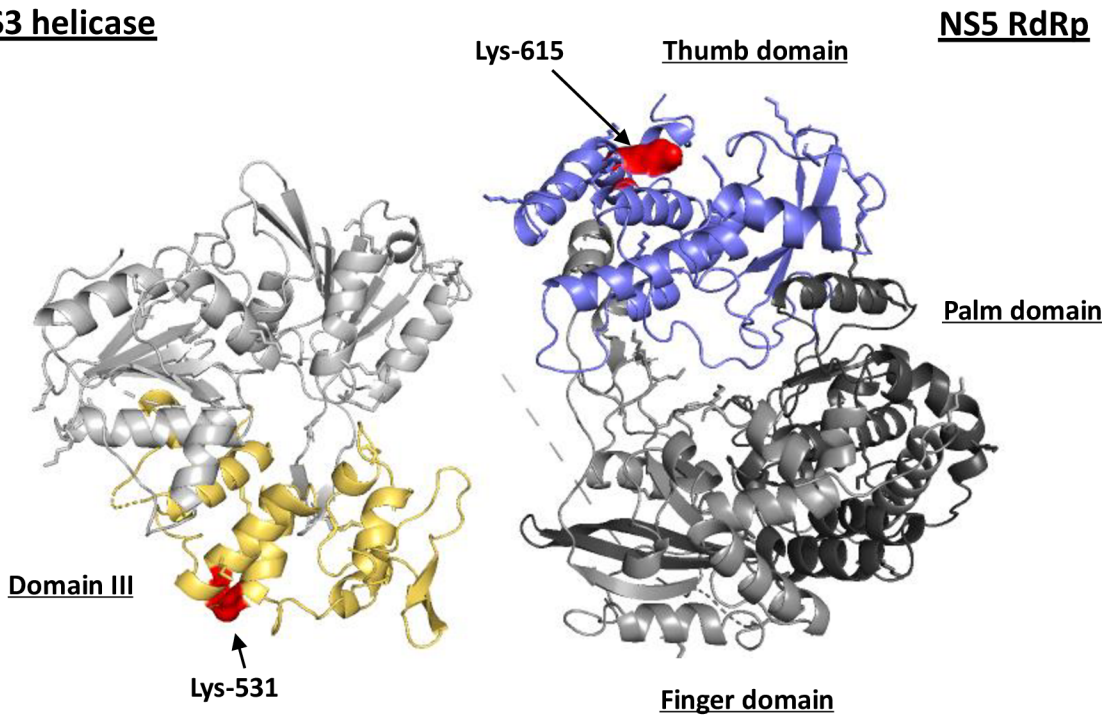


Figure 17 - 3D model of the NS3 helicase (left) and NS5 RdRp (right) as a cartoon depiction. The intermolecular cross-linked residues Lys-531 and Lys-615 are coloured in red, the domain III in yellow-orange, the Thumb domain in blue, the Palm domain in black and the Finger domain in grey.

Another interesting group of cross-links are the self-linked cross-links. As mentioned in the Results section these cross-links are linked to the same residue on each end of the cross-linking agent. This is not possible on the same molecule and therefore, it has been suggested that these might be an indicator for the formation of multimers/oligomers. All of the self-linked cross-links were found on the NS3 helicase and none on the NS5 RdRp (**Figure 11**).

Previous research on the oligomerization of the NS3 helicase of West Nile Virus (WNV) has shown that the helicase has a low tendency to form dimers, trimers and above, however, the whole NS3 protein does form such multimers and oligomers even in higher orders, which is also important in turn for the helicase activity. [66] A more recent study on the crystal structure of the NS3 helicase of TBEV has also shown that the helicase has no stable oligomer in the crystal formation, indicating that the natural state of the helicase is monomeric. [39] While WNV and TBEV are two different viruses they both belong to the flaviviruses, which have a highly conserved helicase sequence. [66]

It has to be mentioned that the WNV NS3 helicase was studied without the presence of the NS5 RdRp and the sequence of the TBEV helicase and the WNV helicase does show some significant differences, with only a sequence identity of 49% according to BLAST (**Figure 24**). [67] [68] Therefore, the observed self-linked cross-links on the TBEV helicase could be the result of oligomerization or similar interactions. Although a Score of above 40 for all self-linked cross-links would indicate that they are not false-positives it cannot be ruled out that at least a few of them are.

Similarly, for the intra-molecular cross-links above the set threshold of 30 Å, it is possible that the formation of multimers or oligomers lead to a greater calculated distance, whereas in reality the residues are closer together but on two different molecules of the same protein. However, it is likely that the 3D model is different than the real structure in solution for these implicated peptides and regions, which is more likely for the NS5 RdRp model since it is not based on the crystal structure, but also the model of the NS3 helicase, which is based on the known structure, could be slightly different than the structure in solution during the experiment. Again, false-positives cannot be ruled out although the Score of all cross-links that fall under this category is above 40 as well.

5.2. Hydrogen-Deuterium Exchange Mass Spectrometry

The hydrogen-deuterium exchange rate and maximum uptake of every peptide is highly dependent on the accessibility of the back-bone amides to the solvent. Hereby, hydrogens that are involved in hydrogen bonding for example are more resistant to deuterium exchange. Therefore, it is expected that peptides involved in NS5-NS3 interaction experience a lower relative deuterium uptake when in the complex state as opposed to the single protein state. For the NS5 RdRp following peptides showed significantly lower deuterium uptake in the complex state: residues 79-88 (TDTTAFGQQR) (**Table 1A**), 307-326 (AYHAKVVKVARPSRDGGCIM) (**Table 1C**), 489-498 (LKSAYGQMWL) (**Table 1D**), 503-512 (HRDLRTLGL) (**Table 1E**) and 601-616 (AVEKVRKMIGPEKFKD) (**Table 1F**). The first peptide can be found in the Finger domain, the second in the Palm domain and the last three in the Thumb domain. Crucially, the peptide consisting of residues 79-88 is on the same α-helix as Lys-91 and -93 that are involved in inter-molecular cross-linking, thus complementing the findings of the CLMS experiment and identifying a possible interaction region on the NS5 RdRp.

Furthermore, the peptide containing residues 307-326 (**Table 1C**) found in the Palm domain is spatially fairly close to the residues 79-88 as seen in **Figure 18** (here the 3D model should be accurate), therefore, it is plausible that this peptide is also similarly involved in NS3-NS5 interaction. The peptide consisting of residues 601-616 (**Table 1F**) found on the Thumb domain includes the inter-molecularly cross-linked residue Lys-615. Together with the other two peptides found in the Thumb domain (residues 489-498 (**Table 1D**) and residues 503-512 (**Table 1E**)) the interaction region of the Thumb domain could have been correctly identified. Lastly, for the NS5 RdRp the peptide containing residues 206-211 (**Table 1B**) had a higher deuterium uptake in the complex state as opposed to the single protein state which would suggest that upon complex formation the RdRp structure opens up at this position, whereby, the peptide is a part of a beta sheet and α-helix confirmation as seen in **Figure 19** (marked in red).

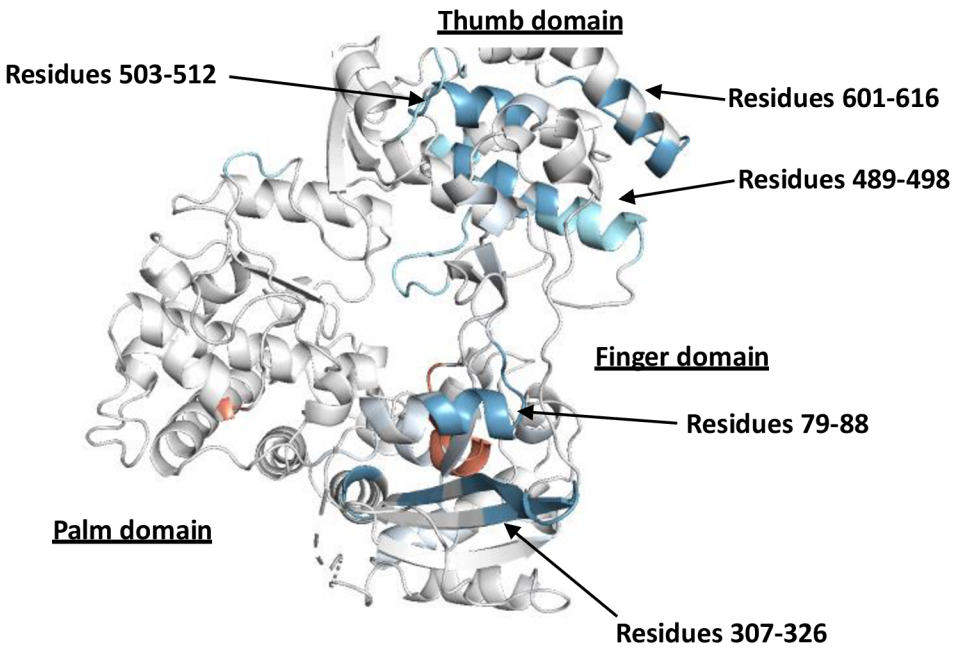


Figure 18 - 3D model of the NS5 RdRp front view for a deuterium labelling time of 2 hours as a cartoon depiction. The possible interaction regions (peptides with lower relative deuterium uptake in the NS3-NS5 complex state) are marked in blue.

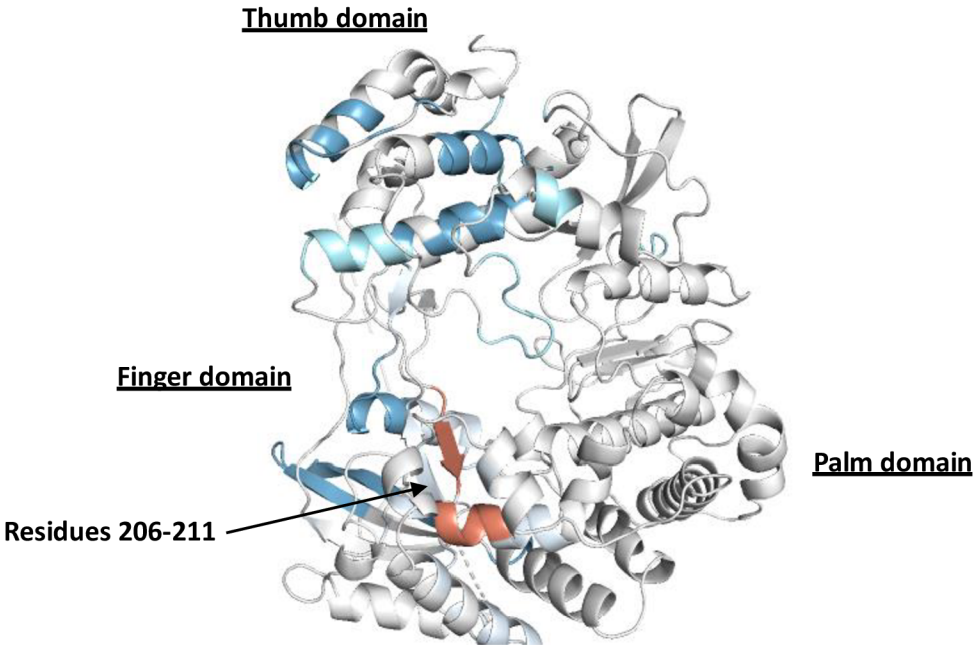


Figure 19 - 3D model of the NS5 RdRp back view for a deuterium labelling time of 2 hours, as a cartoon depiction. The peptide containing residues 206-211 (which saw higher deuterium uptake in the complex state) was marked in red.

For the NS3 helicase only one peptide was identified that had a significantly lower deuterium uptake in the complex state: 248-261 (396-409 in pyMOL) (FEKDYSRVRDEKPD) (**Table 1H**). Unfortunately, this peptide is not marked in blue on the 3D model and is also neither close to the inter-molecular cross-linked residue 457 (605 in pyMOL) nor is it close or contains any lysine involved in self-linked cross-linking. It is highly likely that the NS3 helicase concentration in the HDX sample containing both NS3 helicase and NS5 RdRp was significantly larger than assumed and therefore, not every helicase was in complex with RdRp, meaning there was an excess of single protein state NS3 helicase which greatly influenced the results.

6. Conclusion

The aim of this thesis was the study of NS3-NS5 interaction and identification of possible interaction regions via the analysis of NS3 helicase and NS5 RdRp by two different Mass Spectrometry techniques. Through the Cross-linking Mass Spectrometry experiments 3 unique inter-molecular cross-links were identified which showed that the NS3-NS5 interaction region is most likely located in the Finger and Thumb domain of the NS5 RdRp and the domain III of the NS3 helicase. Together with the results of the Hydrogen-Deuterium Exchange Mass Spectrometry experiments the interaction region could be narrowed down to residues 76-85 (79-88 in pyMOL) of the Finger domain and residues 486-613 (489-616 in pyMOL) of the Thumb domain on the NS5 RdRp (**Figure 18**). Unfortunately, the results of the HDX experiments were inconclusive for the NS3 helicase. It was intended to have the concentration of NS5 RdRp and NS3 helicase at a 1:1 ratio, such that every helicase would be in complex with every RdRp protein. Unfortunately, it seems that the concentration of NS3 helicase was too high, therefore, for further experiments the protein concentrations should be closer scrutinized beforehand. However, through the CLMS results alone and a comparison to a previous study on the dengue virus, which revealed a possible interaction region and interaction point on the DENV NS3 helicase, the domain III of the TBEV NS3 helicase was identified to be involved in NS3-NS5 interaction. [42] This interaction region could possibly be narrowed down to residues 420-439 (568-587 in pyMOL) and residue Asn-424 (Asn-572 in pyMOL) on the domain III of the NS3 helicase (**Figure 16**).

Lastly, the 6 unique self-linked cross-links discovered on the NS3 helicase point towards the formation of oligomers. The availability of complex-free NS3 helicase proteins due to the suspectedly high concentration of helicase could have been a factor in the formation of oligomers, although from these CLMS results a definitive conclusion cannot be drawn. It could also be the case that NS3 helicase is able to form oligomers while in complex with NS5 RdRp, therefore, further research on the subject matter is needed. For example, the obtained data from CLMS and HDX can be used in a Molecular Dynamics simulation to study NS3 helicase and NS5 RdRp in complex.

[1] K. L. Mansfield, N. Johnson, L. P. Phipps, J. R. Stephenson, A. R. Fooks, and T. Solomon, 'Tick-borne encephalitis virus – a review of an emerging zoonosis', *J. Gen. Virol.*, vol. 90, no. 8, pp. 1781–1794, Aug. 2009, doi: 10.1099/vir.0.011437-0.
[2] G. Grard *et al.*, 'Genetic characterization of tick-borne flaviviruses: New insights into evolution, pathogenetic determinants and taxonomy', *Virology*, vol. 361, no. 1, pp. 80–92, Apr. 2007, doi: 10.1016/j.virol.2006.09.015.
[3] E. Martello *et al.*, 'Systematic review on the non-vectorial transmission of Tick-borne encephalitis virus (TBEv)', *Ticks Tick-Borne Dis.*, vol. 13, no. 6, p. 102028, Nov. 2022, doi: 10.1016/j.ttbdis.2022.102028.
[4] K. Yoshii, 'Epidemiology and pathological mechanisms of tick-borne encephalitis', *J. Vet. Med. Sci.*, vol. 81, no. 3, pp. 343–347, 2019, doi: 10.1292/jvms.18-0373.
[5] P. Bogovic, 'Tick-borne encephalitis: A review of epidemiology, clinical characteristics, and management', *World J. Clin. Cases*, vol. 3, no. 5, p. 430, 2015, doi: 10.12998/wjcc.v3.i5.430.
[6] M. Haglund and G. Günther, 'Tick-borne encephalitis—pathogenesis, clinical course and long-term follow-up', *Vaccine*, vol. 21, pp. S11–S18, Apr. 2003, doi: 10.1016/S0264-410X(02)00811-3.
[7] D. Ruzek *et al.*, 'Tick-borne encephalitis in Europe and Russia: Review of pathogenesis, clinical features, therapy, and vaccines', *Antiviral Res.*, vol. 164, pp. 23–51, Apr. 2019, doi: 10.1016/j.antiviral.2019.01.014.
[8] K. M. Paulsen *et al.*, 'Tick-borne encephalitis virus in cows and unpasteurized cow milk from Norway', *Zoonoses Public Health*, vol. 66, no. 2, pp. 216–222, Mar. 2019, doi: 10.1111/zph.12554.
[9] E. Pustijanac, M. Buršić, J. Talapko, I. Škrlec, T. Meštrović, and D. Lišnjić, 'Tick-Borne Encephalitis Virus: A Comprehensive Review of Transmission, Pathogenesis, Epidemiology, Clinical Manifestations, Diagnosis, and Prevention', *Microorganisms*, vol. 11, no. 7, p. 1634, Jun. 2023, doi: 10.3390/microorganisms11071634.
[10] B. Kaufmann and M. G. Rossmann, 'Molecular mechanisms involved in the early steps of flavivirus cell entry', *Microbes Infect.*, vol. 13, no. 1, pp. 1–9, Jan. 2011, doi: 10.1016/j.micinf.2010.09.005.
[11] T. Füzik, P. Formanová, D. Růžek, K. Yoshii, M. Niedrig, and P. Plevka, 'Structure of tick-borne encephalitis virus and its neutralization by a monoclonal antibody', *Nat. Commun.*, vol. 9, no. 1, p. 436, Jan. 2018, doi: 10.1038/s41467-018-02882-0.
[12] H. Kroschewski, S. L. Allison, F. X. Heinz, and C. W. Mandl, 'Role of heparan sulfate for attachment and entry of tick-borne encephalitis virus', *Virology*, vol. 308, no. 1, pp. 92–100, Mar. 2003, doi: 10.1016/S0042-6822(02)00097-1.
[13] X. Zhang *et al.*, 'T-Cell Immunoglobulin and Mucin Domain 1 (TIM-1) Is a Functional Entry Factor for Tick-Borne Encephalitis Virus', *mBio*, vol. 13, no. 1, pp. e02860-21, Feb. 2022, doi: 10.1128/mbio.02860-21.
[14] M. Perera-Lecoin, L. Meertens, X. Carnec, and A. Amara, 'Flavivirus Entry Receptors: An Update', *Viruses*, vol. 6, no. 1, pp. 69–88, Dec. 2013, doi: 10.3390/v6010069.
[15] E. Veiga and P. Cossart, 'The role of clathrin-dependent endocytosis in bacterial internalization', *Trends Cell Biol.*, vol. 16, no. 10, pp. 499–504, Oct. 2006, doi: 10.1016/j.tcb.2006.08.005.
[16] R. Fritz, K. Stiasny, and F. X. Heinz, 'Identification of specific histidines as pH sensors in flavivirus membrane fusion', *J. Cell Biol.*, vol. 183, no. 2, pp. 353–361, Oct. 2008, doi: 10.1083/jcb.200806081.
[17] X. Zhang *et al.*, 'Structure of Acidic pH Dengue Virus Showing the Fusogenic Glycoprotein Trimers', *J. Virol.*, vol. 89, no. 1, pp. 743–750, Jan. 2015, doi: 10.1128/JVI.02411-14.
[18] G. Gerold, J. Bruening, B. Weigel, and T. Pietschmann, 'Protein Interactions during the Flavivirus and Hepacivirus Life Cycle', *Mol. Cell. Proteomics*, vol. 16, no. 4, pp. S75–S91, Apr. 2017, doi: 10.1074/mcp.R116.065649.
[19] K. Van Den Elsen, J. P. Quek, and D. Luo, 'Molecular Insights into the Flavivirus Replication Complex', *Viruses*, vol. 13, no. 6, p. 956, May 2021, doi: 10.3390/v13060956.

[20] M. Sakai, K. Yoshii, Y. Sunden, K. Yokozawa, M. Hirano, and H. Kariwa, 'Variable region of the 3' UTR is a critical virulence factor in the Far-Eastern subtype of tick-borne encephalitis virus in a mouse model', *J. Gen. Virol.*, vol. 95, no. 4, pp. 823–835, Apr. 2014, doi: 10.1099/vir.0.060046-0.
[21] D. Ott *et al.*, 'Tick-borne encephalitis virus (TBEV) prevalence in field-collected ticks (Ixodes ricinus) and phylogenetic, structural and virulence analysis in a TBE high-risk endemic area in southwestern Germany', *Parasit. Vectors*, vol. 13, no. 1, p. 303, Dec. 2020, doi: 10.1186/s13071-020-04146-7.
[22] Y. Tang, Y. Diao, X. Gao, C. Yu, L. Chen, and D. Zhang, 'Analysis of the Complete Genome of Tembusu Virus, a Flavivirus Isolated from Ducks in China: Analysis of the Complete Genome of Tembusu Virus', *Transbound. Emerg. Dis.*, vol. 59, no. 4, pp. 336–343, Aug. 2012, doi: 10.1111/j.1865-1682.2011.01275.x.
[23] W.-L. Yau, V. Nguyen-Dinh, E. Larsson, R. Lindqvist, A. K. Överby, and R. Lundmark, 'Model System for the Formation of Tick-Borne Encephalitis Virus Replication Compartments without Viral RNA Replication', *J. Virol.*, vol. 93, no. 18, pp. e00292-19, Sep. 2019, doi: 10.1128/JVI.00292-19.
[24] L. Miorin *et al.*, 'Three-Dimensional Architecture of Tick-Borne Encephalitis Virus Replication Sites and Trafficking of the Replicated RNA', *J. Virol.*, vol. 87, no. 11, pp. 6469–6481, Jun. 2013, doi: 10.1128/JVI.03456-12.
[25] Y. Ci and L. Shi, 'Compartmentalized replication organelle of flavivirus at the ER and the factors involved', *Cell. Mol. Life Sci.*, vol. 78, no. 11, pp. 4939–4954, Jun. 2021, doi: 10.1007/s00018-021-03834-6.
[26] H. Xing *et al.*, 'Zika NS2B is a crucial factor recruiting NS3 to the ER and activating its protease activity', *Virus Res.*, vol. 275, p. 197793, Jan. 2020, doi: 10.1016/j.virusres.2019.197793.
[27] C. Brand, M. Bisaillon, and B. J. Geiss, 'Organization of the *Flavivirus* RNA replicase complex', *WIREs RNA*, vol. 8, no. 6, p. e1437, Nov. 2017, doi: 10.1002/wrna.1437.
[28] E. Lee *et al.*, 'Structures of flavivirus RNA promoters suggest two binding modes with NS5 polymerase', *Nat. Commun.*, vol. 12, no. 1, p. 2530, May 2021, doi: 10.1038/s41467-021-22846-1.
[29] L. Pulkkinen, S. Butcher, and M. Anastasina, 'Tick-Borne Encephalitis Virus: A Structural View', *Viruses*, vol. 10, no. 7, p. 350, Jun. 2018, doi: 10.3390/v10070350.
[30] V. A. Kostyuchenko, Q. Zhang, J. L. Tan, T.-S. Ng, and S.-M. Lok, 'Immature and Mature Dengue Serotype 1 Virus Structures Provide Insight into the Maturation Process', *J. Virol.*, vol. 87, no. 13, pp. 7700–7707, Jul. 2013, doi: 10.1128/JVI.00197-13.
[31] L. Li *et al.*, 'The Flavivirus Precursor Membrane-Envelope Protein Complex: Structure and Maturation', *Science*, vol. 319, no. 5871, pp. 1830–1834, Mar. 2008, doi: 10.1126/science.1153263.
[32] Y. Zhao *et al.*, 'A Crystal Structure of the Dengue Virus NS5 Protein Reveals a Novel Inter-domain Interface Essential for Protein Flexibility and Virus Replication', *PLOS Pathog.*, vol. 11, no. 3, p. e1004682, Mar. 2015, doi: 10.1371/journal.ppat.1004682.
[33] Y. Zhou *et al.*, 'Structure and Function of Flavivirus NS5 Methyltransferase', *J. Virol.*, vol. 81, no. 8, pp. 3891–3903, Apr. 2007, doi: 10.1128/JVI.02704-06.
[34] H. Malet *et al.*, 'The flavivirus polymerase as a target for drug discovery', *Antiviral Res.*, vol. 80, no. 1, pp. 23–35, Oct. 2008, doi: 10.1016/j.antiviral.2008.06.007.
[35] J. Yang *et al.*, 'Crystal structure of a tick-borne flavivirus RNA-dependent RNA polymerase suggests a host adaptation hotspot in RNA viruses', *Nucleic Acids Res.*, vol. 49, no. 3, pp. 1567–1580, Feb. 2021, doi: 10.1093/nar/gkaa1250.
[36] J. J. Valdés, P. T. Butterill, and D. Růžek, 'Flaviviridae viruses use a common molecular mechanism to escape nucleoside analogue inhibitors', *Biochem. Biophys. Res. Commun.*, vol. 492, no. 4, pp. 652–658, Oct. 2017, doi: 10.1016/j.bbrc.2017.03.068.
[37] D. Akaberi *et al.*, 'Targeting the NS2B-NS3 protease of tick-borne encephalitis virus with pan-flaviviral protease inhibitors', *Antiviral Res.*, vol. 190, p. 105074, Jun. 2021, doi: 10.1016/j.antiviral.2021.105074.

[38] P. D. Anindita, M. Halbeisen, D. Řeha, R. Tuma, and Z. Franta, 'Mechanistic insight into the RNA-stimulated ATPase activity of tick-borne encephalitis virus helicase', *J. Biol. Chem.*, vol. 298, no. 10, p. 102383, Oct. 2022, doi: 10.1016/j.jbc.2022.102383.
[39] C. Chen *et al.*, 'Crystal structure of the NS3 helicase of tick-borne encephalitis virus', *Biochem. Biophys. Res. Commun.*, vol. 528, no. 3, pp. 601–606, Jul. 2020, doi: 10.1016/j.bbrc.2020.05.138.
[40] S. Xu *et al.*, 'Zika virus NS3 is a canonical RNA helicase stimulated by NS5 RNA polymerase', *Nucleic Acids Res.*, vol. 47, no. 16, pp. 8693–8707, Sep. 2019, doi: 10.1093/nar/gkz650.
[41] C. Yon *et al.*, 'Modulation of the Nucleoside Triphosphatase/RNA Helicase and 5'-RNA Triphosphatase Activities of Dengue Virus Type 2 Nonstructural Protein 3 (NS3) by Interaction with NS5, the RNA-dependent RNA Polymerase', *J. Biol. Chem.*, vol. 280, no. 29, pp. 27412–27419, Jul. 2005, doi: 10.1074/jbc.M501393200.
[42] M. Y. F. Tay *et al.*, 'The C-terminal 50 Amino Acid Residues of Dengue NS3 Protein Are Important for NS3-NS5 Interaction and Viral Replication', *J. Biol. Chem.*, vol. 290, no. 4, pp. 2379–2394, Jan. 2015, doi: 10.1074/jbc.M114.607341.
[43] F. J. O'Reilly and J. Rappsilber, 'Cross-linking mass spectrometry: methods and applications in structural, molecular and systems biology', *Nat. Struct. Mol. Biol.*, vol. 25, no. 11, pp. 1000–1008, Nov. 2018, doi: 10.1038/s41594-018-0147-0.
[44] J. D. Chavez and J. E. Bruce, 'Chemical cross-linking with mass spectrometry: a tool for systems structural biology', *Curr. Opin. Chem. Biol.*, vol. 48, pp. 8–18, Feb. 2019, doi: 10.1016/j.cbpa.2018.08.006.
[45] A. Leitner, M. Faini, F. Stengel, and R. Aebersold, 'Crosslinking and Mass Spectrometry: An Integrated Technology to Understand the Structure and Function of Molecular Machines', *Trends Biochem. Sci.*, vol. 41, no. 1, pp. 20–32, Jan. 2016, doi: 10.1016/j.tibs.2015.10.008.
[46] P. Novák and G. H. Kruppa, 'Intra-Molecular Cross-Linking of Acidic Residues for Protein Structure Studies', *Eur. J. Mass Spectrom.*, vol. 14, no. 6, pp. 355–365, Dec. 2008, doi: 10.1255/ejms.963.
[47] C. Yu and L. Huang, 'Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology', *Anal. Chem.*, vol. 90, no. 1, pp. 144–165, Jan. 2018, doi: 10.1021/acs.analchem.7b04431.
[48] C. Iacobucci, M. Götze, and A. Sinz, 'Cross-linking/mass spectrometry to get a closer view on protein interaction networks', *Curr. Opin. Biotechnol.*, vol. 63, pp. 48–53, Jun. 2020, doi: 10.1016/j.copbio.2019.12.009.
[49] C. W. Combe, L. Fischer, and J. Rappsilber, 'xiNET: Cross-link Network Maps With Residue Resolution', *Mol. Cell. Proteomics*, vol. 14, no. 4, pp. 1137–1147, Apr. 2015, doi: 10.1074/mcp.O114.042259.
[50] S. Yuan, H. C. S. Chan, and Z. Hu, 'Using PyMOL as a platform for computational drug design', *WIREs Comput. Mol. Sci.*, vol. 7, no. 2, p. e1298, Mar. 2017, doi: 10.1002/wcms.1298.
[51] M. O. Debelyy, P. Waridel, M. Quadroni, R. Schneiter, and A. Conzelmann, 'Chemical crosslinking and mass spectrometry to elucidate the topology of integral membrane proteins', *PLOS ONE*, vol. 12, no. 10, p. e0186840, Oct. 2017, doi: 10.1371/journal.pone.0186840.
[52] B. Schiffrin, S. E. Radford, D. J. Brockwell, and A. N. Calabrese, 'PyXLinkViewer : A flexible tool for visualization of protein chemical crosslinking data within the PyMOL molecular graphics system', *Protein Sci.*, vol. 29, no. 8, pp. 1851–1857, Aug. 2020, doi: 10.1002/pro.3902.
[53] Y. Bai, J. S. Milne, L. Mayne, and S. W. Englander, 'Primary structure effects on peptide group hydrogen exchange', *Proteins Struct. Funct. Bioinforma.*, vol. 17, no. 1, pp. 75–86, Sep. 1993, doi: 10.1002/prot.340170110.
[54] E. I. James, T. A. Murphree, C. Vorauer, J. R. Engen, and M. Guttman, 'Advances in Hydrogen/Deuterium Exchange Mass Spectrometry and the Pursuit of Challenging Biological Systems', *Chem. Rev.*, vol. 122, no. 8, pp. 7562–7623, Apr. 2022, doi: 10.1021/acs.chemrev.1c00279.

[55] H. Xiao *et al.*, 'Mapping protein energy landscapes with amide hydrogen exchange and mass spectrometry: I. A generalized model for a two-state protein and comparison with experiment', *Protein Sci.*, vol. 14, no. 2, pp. 543–557, Feb. 2005, doi: 10.1110/ps.041001705.
[56] J. Zheng, T. Strutzenberg, B. D. Pascal, and P. R. Griffin, 'Protein dynamics and conformational changes explored by hydrogen/deuterium exchange mass spectrometry', *Curr. Opin. Struct. Biol.*, vol. 58, pp. 305–313, Oct. 2019, doi: 10.1016/j.sbi.2019.06.007.
[57] T. E. Wales and J. R. Engen, 'Hydrogen exchange mass spectrometry for the analysis of protein dynamics', *Mass Spectrom. Rev.*, vol. 25, no. 1, pp. 158–170, Jan. 2006, doi: 10.1002/mas.20064.
[58] D. Narang, C. Lento, and D. J. Wilson, 'HDX-MS: An Analytical Tool to Capture Protein Motion in Action', *Biomedicines*, vol. 8, no. 7, p. 224, Jul. 2020, doi: 10.3390/biomedicines8070224.
[59] V. Vinciauskaite and G. R. Masson, 'Fundamentals of HDX-MS', *Essays Biochem.*, vol. 67, no. 2, pp. 301–314, Mar. 2023, doi: 10.1042/EBC20220111.
[60] I. Oganesyan, C. Lento, and D. J. Wilson, 'Contemporary hydrogen deuterium exchange mass spectrometry', *Methods*, vol. 144, pp. 27–42, Jul. 2018, doi: 10.1016/j.ymeth.2018.04.023.
[61] P. Havlickova, J. Crossley, Z. Gardian, F. Dycka, I. Kuta Smatanova, and Z. Franta, 'Structural and functional studies of TBEV non-structural protein 5', *Acta Crystallogr. Sect. Found. Adv.*, vol. 78, no. a2, pp. e339–e339, Aug. 2022, doi: 10.1107/S2053273322093925.
[62] J. Rappsilber, M. Mann, and Y. Ishihama, 'Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips', *Nat. Protoc.*, vol. 2, no. 8, pp. 1896–1906, Aug. 2007, doi: 10.1038/nprot.2007.261.
[63] J. Cox and M. Mann, 'MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification', *Nat. Biotechnol.*, vol. 26, no. 12, pp. 1367–1372, Dec. 2008, doi: 10.1038/nbt.1511.
[64] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, and M. J. E. Sternberg, 'The Phyre2 web portal for protein modeling, prediction and analysis', *Nat. Protoc.*, vol. 10, no. 6, pp. 845–858, Jun. 2015, doi: 10.1038/nprot.2015.053.
[65] G. Zou *et al.*, 'Functional Analysis of Two Cavities in Flavivirus NS5 Polymerase', *J. Biol. Chem.*, vol. 286, no. 16, pp. 14362–14372, Apr. 2011, doi: 10.1074/jbc.M110.214189.
[66] M. J. Feito, J. Gómez-Gutiérrez, S. Ayora, J. C. Alonso, D. Peterson, and F. Gavilanes, 'Insights into the oligomerization state–helicase activity relationship of West Nile virus NS3 NTPase/helicase', *Virus Res.*, vol. 135, no. 1, pp. 166–174, Jul. 2008, doi: 10.1016/j.virusres.2008.03.008.
[67] S. Altschul, 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', *Nucleic Acids Res.*, vol. 25, no. 17, pp. 3389–3402, Sep. 1997, doi: 10.1093/nar/25.17.3389.
[68] S. F. Altschul *et al.*, 'Protein database searches using compositionally adjusted substitution matrices', *FEBS J.*, vol. 272, no. 20, pp. 5101–5109, Oct. 2005, doi: 10.1111/j.1742-4658.2005.04945.x.

7. Appendix

Figure 20 - Sequence and residue numbering of TBEV NS3 helicase in pdb structure which was loaded into pyMOL for protein structure visualization.

Figure 22 - Sequence and residue numbering of TBEV NS5 RdRp in pdb structure which was loaded into pyMOL for protein structure visualization

```
MGHHHHHHHHHHSSGHIDDDDKHMEKSRPNLPQAVVGTGWTSKGQITVLDMHPGS    55
GKTHRVLPELIRQCIDRRLRTLVLAPTRVVLKEMERALNGKRVRFHSPAVSDQQA   110
GGAIVDVMCHATYVNRRLLPQGRQNWEVAIMDEAHWTDPHSIAARGHLYTLAKEN   165
KCALVLMTATPPGKSEPFPESNGAITSEERQIPDGEWRDGFDWITEYEGRTAWFV   220
PSIAKGGAIARTLRQKGKSVICLNSKTFEKDYSRVRDEKPDFVVTTDISEMGANL   275
DVSRVIDGRTNIKPEEVDGKVELTGTRRVTTASAAQRRGRVGRQDGRTDEYIYSG   330
QCDDDDSGLVQWKEAQILLDNITTLRGPVATFYGPEQDKMPEVAGHFRLTEEKRK   385
HFRHLLTHCDFTPWLAWHVAANVSSVTDRSWTWEGPEANAVDEASGDLVTFRSPN   440
GAERTLRPVWKDARMFKEGRDIKEFVAYASGRR                          473
```

Figure 21 – Sequence of TBEV NS3 helicase. This sequence was used for peptide identification in all Mass Spectrometry experiments.

Figure 23 - Sequence of TBEV NS5 RdRp. This sequence was used for peptide identification in all Mass Spectrometry experiments.

Table 2 - Cross-linked peptides between NS5 RdRp (NS5) and NS3 helicase (7oj4), collected over several measurements of different mixtures to various ratios of NS5/NS3. A distance of "-" means that the distance could not be calculated since the complex structure is not known and a distance of "0" means that the residue is linked to itself; A distance of " " means that the two cross-linked residues belong to the same peptide but are in different positions.

Protein1	Protein2	LinkPos1	LinkPos2	Pep. Seq. 1	Pep. Seq. 2	Distance / Å	Score
NS5	NS5	593	601	AEWAKNIWGAVEKVR	AEWAKNIWGAVEKVR	15	79.485
NS5	NS5	601	604	AEWAKNIWGAVEKVR	KMIGPEK	5	153.21
7oj4	7oj4	96	96	ALNGKR	ALNGKR	0	59.818
7oj4	7oj4	96	457	ALNGKR	MFKEGR	42	70.919
NS5	NS5	586	593	AQDMLCSSLVGRKER	AEWAKNIWGAVEK	10	115.88
NS5	NS5	308	311	AYHAKVVKVAR	AYHAKVVKVAR	10	85.777
NS5	NS5	189	308	CAHCVYNMMGKR	AYHAKVVK	20	46.964
NS5	NS5	189	90	CAHCVYNMMGKR	DKVDTK	14	42.809
NS5	NS5	189	134	CAHCVYNMMGKR	EEFIAKVK	23	189.5
NS5	NS5	189	193	CAHCVYNMMGKR	EKKLGEFGVAK	11	117.67
NS5	NS5	189	193	CAHCVYNMMGKR	KLGEFGVAK	11	70.98
NS5	NS5	189	201	CAHCVYNMMGKR	LGEFGVAKGSR	11	104.44
NS5	NS5	189	134	CAHCVYNMMGKR	MCSREEFIAKVK	23	123.53
NS5	NS5	189	88	CAHCVYNMMGKR	VFKDKVDTK	16	201.44
NS5	NS5	189	311	CAHCVYNMMGKR	VVKVAR	15	85.288
NS5	NS5	189	311	CAHCVYNMMGKR	VVKVARPSR	15	46.318
7oj4	7oj4	179	163	CALVLMTATPPGKSEPFPESNGAITSER	GHLYTLAKENK	21	206.33

7oj4	7oj4	463	457	DIKEFVAYASGR	MFKEGR	11	151.56
7oj4	7oj4	463	246	DIKEFVAYASGR	SVICLNSKTFEK	15	73.21
7oj4	7oj4	463	250	DIKEFVAYASGR	TFEKDYSR	21	150.59
7oj4	7oj4	463	451	DIKEFVAYASGR	TLRPVWKDAR	16	91.401
7oj4	7oj4	463	96	DIKEFVAYASGRR	ALNGKR	44	45.863
7oj4	7oj4	463	383	DIKEFVAYASGRR	LTEEKR	26	40.801
7oj4	7oj4	463	457	DIKEFVAYASGRR	MFKEGR	11	198.87
7oj4	7oj4	463	246	DIKEFVAYASGRR	SVICLNSKTFEK	15	106.57
7oj4	7oj4	463	250	DIKEFVAYASGRR	TFEKDYSR	21	173.58
7oj4	7oj4	463	451	DIKEFVAYASGRR	TLRPVWKDAR	16	221.24
7oj4	7oj4	463	236	DIKEFVAYASGRR	TLRQKGK	33	44.601
NS5	NS5	90	94	DKVDTKAQEPQP GTR	DKVDTKAQEPQP GTR	10	118.26
NS5	NS5	573	593	DVPYLPKAQDML CSSLVGR	AEWAKNIWGAVE K	19	198.39
NS5	NS5	573	586	DVPYLPKAQDML CSSLVGR	KERAEWAK	16	45.546
NS5	NS5	573	593	DVPYLPKAQDML CSSLVGRK	AEWAKNIWGAVE K	19	136.89
NS5	NS5	573	593	DVPYLPKAQDML CSSLVGRK	AEWAKNIWGAVE KVR	19	42.828
NS5	NS5	134	308	EEFIAKVK	AYHAKVVK	25	135.55
NS5	NS5	134	120	EEFIAKVK	LAQKSKPR	25	92.925
NS5	NS5	134	311	EEFIAKVK	VVKVAR	27	70.1
NS5	NS5	192	193	EKKLGEFGVAK	EKKLGEFGVAK	4	175.93
NS5	NS5	193	88	EKKLGEFGVAK	VFKDKVDTK	19	67.605
7oj4	7oj4	166	179	ENKCALVLMTATP PGKSEPFPESNGAI TSEER	ENKCALVLMTATP PGKSEPFPESNGA ITSEER	27	20.075
NS5	NS5	408	88	FGKALYFLNDMAK	VFKDKVDTK	31	116.29
NS5	NS5	408	418	FGKALYFLNDMAK TR	FGKALYFLNDMA KTR	14	13.298
NS5	NS5	612	604	FKDYLSCMDR	KMIGPEK	14	66.285
7oj4	7oj4	2	22	GHHHHHHHHHH SSGHIDDDDKHM EK	GHHHHHHHHHH SSGHIDDDDKHM EK		89.879
7oj4	7oj4	163	166	GHLYTLAKENKCA LVLMTATPPGK	GHLYTLAKENKCA LVLMTATPPGK	8	163.51
7oj4	7oj4	238	457	GKSVICLNSK	MFKEGR	27	119.01
7oj4	7oj4	238	250	GKSVICLNSK	TFEKDYSR	22	140.84
7oj4	7oj4	246	463	GKSVICLNSKTFEK	DIKEFVAYASGRR	15	112.54
7oj4	7oj4	238	246	GKSVICLNSKTFEK	GKSVICLNSKTFEK	22	44.135
7oj4	7oj4	369	225	GPVATFYGPEQDK MPEVAGHFR	TAWFVPSIAKGGA IAR	28	64.096
7oj4	7oj4	57	96	GQITVLDMHPGS GKTHR	ALNGKR	17	159.26
7oj4	7oj4	57	250	GQITVLDMHPGS GKTHR	TFEKDYSR	24	57.989
7oj4	7oj4	57	87	GQITVLDMHPGS GKTHR	VVLKEMER	15	200.99

NS5	NS5	193	90	KLGEFGVAK	VFKDKVDTK	16	53.001
NS5	NS5	193	201	KLGEFGVAKGSR	KLGEFGVAKGSR	7	77.287
NS5	NS5	253	90	KLSTLNGGLFYADDTAGWDTK	VFKDKVDTK	35	47.728
NS5	NS5	120	134	LAQKSKPR	EEFIAKVK	25	63.727
NS5	NS5	120	122	LAQKSKPR	LAQKSKPR	7	88.421
NS5	NS5	201	308	LGEFGVAKGSR	AYHAKVVK	27	78.486
NS5	NS5	201	90	LGEFGVAKGSR	VFKDKVDTK	13	149.09
NS5	NS5	201	311	LGEFGVAKGSR	VVKVARPSR	22	112.42
7oj4	7oj4	383	96	LTEEKR	ALNGKR	40	45.68
7oj4	7oj4	383	457	LTEEKR	MFKEGR	27	78.934
7oj4	7oj4	383	457	LTEEKRK	MFKEGR	27	59.542
NS5	NS5	88	94	MAMTDTTAFGQQRVFKDK	VDTKAQEPQPGTR	12	184.06
NS5	NS5	134	308	MCSREEFIAKVK	AYHAKVVK	25	126.63
NS5	NS5	134	189	MCSREEFIAKVK	CAHCVYNMMGKR	23	141
NS5	NS5	134	120	MCSREEFIAKVK	LAQKSKPR	25	92.977
7oj4	7oj4	457	96	MFKEGR	ALNGKR	42	94.302
7oj4	7oj4	457	383	MFKEGR	LTEEKR	27	85.731
7oj4	7oj4	457	457	MFKEGR	MFKEGR	0	64.866
7oj4	NS5	457	90	MFKEGRDIK	DKVDTK	-	47.302
7oj4	NS5	457	88	MFKEGRDIK	VFKDKVDTK	-	42.317
NS5	7oj4	612	383	MIGPEKFKDYLSCMDR	LTEEKR	-	42.338
NS5	NS5	610	612	MIGPEKFKDYLSCMDR	MIGPEKFKDYLSCMDR	7	168.79
NS5	NS5	601	604	NIWGAVEKVR	KMIGPEK	5	178.15
7oj4	7oj4	295	451	PEEVDGKVELTGTRR	TLRPVWKDAR	5	57.989
7oj4	7oj4	451	457	PVWKDAR	MFKEGR	12	65.246
7oj4	7oj4	236	238	QKGKSVICLNSK	QKGKSVICLNSK	6	166.68
NS5	NS5	303	308	QLATTIMQKAYHAKVVK	QLATTIMQKAYHAKVVK	7	83.31
7oj4	7oj4	246	463	SVICLNSKTFEK	DIKEFVAYASGR	15	153.82
7oj4	7oj4	246	457	SVICLNSKTFEK	MFKEGR	11	80.96
7oj4	7oj4	250	463	SVICLNSKTFEKDYSR	DIKEFVAYASGR	21	111.58
7oj4	7oj4	246	463	SVICLNSKTFEKDYSR	DIKEFVAYASGRR	15	337.62
7oj4	7oj4	246	457	SVICLNSKTFEKDYSR	MFKEGR	11	213.24
7oj4	7oj4	246	250	SVICLNSKTFEKDYSR	SVICLNSKTFEKDYSR	8	195.27
7oj4	7oj4	246	451	SVICLNSKTFEKDYSR	TLRPVWKDAR	21	109.05
NS5	NS5	58	120	TAPTGSAASLINGVVKLLSWPWNAR	AVNDWILERLAQKSK	52	55.364
7oj4	7oj4	225	163	TAWFVPSIAKGGAIAR	GHLYTLAKENK	42	90.385

7oj4	7oj4	225	457	TAWFVPSIAKGGAIAR	MFKEGR	13	168.32
7oj4	7oj4	225	225	TAWFVPSIAKGGAIAR	TAWFVPSIAKGGAIAR	0	56.147
7oj4	7oj4	225	451	TAWFVPSIAKGGAIAR	TLRPVWKDAR	23	70.414
7oj4	7oj4	225	87	TAWFVPSIAKGGAIAR	VVLKEMER	30	69.935
7oj4	7oj4	225	451	TAWFVPSIAKGGAIARTLR	PVWKDAR	23	66.325
7oj4	7oj4	250	96	TFEKDYSR	ALNGKR	27	106.62
7oj4	7oj4	250	383	TFEKDYSR	LTEEKR	33	43.897
7oj4	7oj4	250	457	TFEKDYSR	MFKEGR	19	118.28
7oj4	7oj4	250	87	TFEKDYSR	VVLKEMER	13	234.47
7oj4	7oj4	451	96	TLRPVWKDAR	ALNGKR	47	53.967
7oj4	7oj4	451	383	TLRPVWKDAR	LTEEKR	20	78.342
7oj4	7oj4	451	383	TLRPVWKDAR	LTEEKRK	20	48.467
7oj4	7oj4	451	457	TLRPVWKDAR	MFKEGR	12	244.73
7oj4	7oj4	451	457	TLRPVWKDAR	MFKEGRDIK	12	92.439
7oj4	7oj4	451	250	TLRPVWKDAR	TFEKDYSR	29	117.93
7oj4	7oj4	451	451	TLRPVWKDAR	TLRPVWKDAR	0	67.93
7oj4	7oj4	451	236	TLRPVWKDAR	TLRQKGK	39	109.6
7oj4	7oj4	451	87	TLRPVWKDAR	VVLKEMER	35	56.482
7oj4	7oj4	236	451	TLRQKGK	PVWKDAR	39	57.348
7oj4	7oj4	87	250	TLVLAPTRVVLKEMER	TFEKDYSR	13	195.77
7oj4	7oj4	87	87	TLVLAPTRVVLKEMER	VVLKEMER	0	72.184
7oj4	7oj4	288	457	TNIKPEEVDGK	MFKEGR	14	90.731
7oj4	7oj4	295	383	TNIKPEEVDGKVELTGTR	LTEEKR	21	117.74
7oj4	7oj4	295	457	TNIKPEEVDGKVELTGTR	MFKEGR	15	112.32
7oj4	7oj4	288	225	TNIKPEEVDGKVELTGTR	TAWFVPSIAKGGAIAR	14	96.982
7oj4	7oj4	295	451	TNIKPEEVDGKVELTGTR	TLRPVWKDAR	5	128.4
7oj4	7oj4	288	295	TNIKPEEVDGKVELTGTR	TNIKPEEVDGKVELTGTR	13	132.78
NS5	NS5	94	380	VDTKAQEPQPGTR	VERWLKEHGEER	34	57.009
NS5	NS5	90	308	VFKDKVDTK	AYHAKVVK	18	85.554
NS5	NS5	90	193	VFKDKVDTK	KLGEFGVAK	16	52.823
NS5	7oj4	90	457	VFKDKVDTK	MFKEGRDIK	-	41.644
NS5	NS5	88	90	VFKDKVDTK	VFKDKVDTK	6	162.41
NS5	NS5	88	311	VFKDKVDTK	VVKVARPSR	15	53.3
7oj4	7oj4	288	383	VIDGRTNIKPEEVDGK	LTEEKR	29	73.208
NS5	NS5	136	604	VKSNAALGAWSDEQNR	KMIGPEK	36	153.23

NS5	NS5	136	604	VKSNAALGAWSD EQNRWASAR	KMIGPEK	36	69.152
NS5	NS5	1	6	VLAEDKVKEQDV QER	VLAEDKVKEQDV QER		63.401
7oj4	7oj4	259	457	VRDEKPDFVVTTD ISEMGANLDVSR	MFKEGR	28	181.37
7oj4	7oj4	259	225	VRDEKPDFVVTTD ISEMGANLDVSR	TAWFVPSIAKGGA IAR	22	116.73
7oj4	7oj4	259	250	VRDEKPDFVVTTD ISEMGANLDVSR	TFEKDYSR	16	224.79
NS5	NS5	311	193	VVKVARPSR	KLGEFGVAK	25	46.156
7oj4	7oj4	87	96	VVLKEMER	ALNGKR	15	163.66
7oj4	7oj4	87	383	VVLKEMER	LTEEKR	34	81.625
7oj4	7oj4	87	457	VVLKEMER	MFKEGR	29	80.763
7oj4	7oj4	87	250	VVLKEMER	TFEKDYSR	13	225.72
7oj4	7oj4	87	87	VVLKEMER	VVLKEMER	0	78.334
NS5	NS5	559	561	VWILDNPFMQNK GKVMEWR	VWILDNPFMQNK GKVMEWR	6	85.77
NS5	NS5	380	120	WLKEHGEER	LAQKSKPR	17	90.412
NS5	NS5	294	122	YMEGEHKQLATTI MQK	LAQKSKPR	13	223.06
NS5	NS5	294	303	YMEGEHKQLATTI MQKAYHAK	YMEGEHKQLATTI MQKAYHAK	13	139.96

Score	Expect	Method	Identities	Positives	Gaps
422 bits(1086)	3e-147	Compositional matrix adjust.	212/435(49%)	287/435(65%)	5/435(1%)

```
Query  186  KKQITVLDLHPGAGKTRKILPQIIKEAINKRLRTAVLAPTRVVAAEMSEALRGLPIRYQT  245
            K QITVLD+HPG+GKT ++LP++I++ I++RLRT VLAPTRVV  EM  AL G  +R+ +
Sbjct  43   KGQITVLDMHPGSGKTHRVLPELIRQCIDRRLRTLVLAPTRVVLKEMERALNGKRVRFHS  102

Query  246  SAVHREHSGNEIVDVMCHATLTHRLMSPHRVPNYNLFIMDEAHFTDPASIAARGYIATKV  305
             AV  + +G  IVDVMCHAT  +R + P    N+ + IMDEAH+TDP SIAARG++ T
Sbjct  103  PAVSDQQAGGAIVDVMCHATYVNRRLLPQGRQNWEVAIMDEAHWTDPHSIAARGHLYTLA  162

Query  306  ELGEAAAIFMTATPPGTSDPFPESNAPISDMQTEIPDRAWNTGYEWITEYVGKTVWFVPS  365
            +  + A + MTATPPG S+PFPESN  I+  + +IPD  W  G++WITEY G+T WFVPS
Sbjct  163  KENKCALVLMTATPPGKSEPFPESNGAITSEERQIPDGEWRDGFDWITEYEGRTAWFVPS  222

Query  366  VKMGNEIALCLQRAGKKVIQLNRKSYETEYPKCKNDDWDFVITTDISEMGANFKASRVID  425
            +  G  IA  L++ GK VI LN K++E +Y + +++  DFV+TTDISEMGAN   SRVID
Sbjct  223  IAKGGAIARTLRQKGKSVICLNSKTFEKDYSRVRDEKPDFVVTTDISEMGANLDVSRVID  282

Query  426  SRKSVKPTIIEEGDGRVILGEPSAITAASAAQRRGRIGRNPSQVGDEYCYGGHTNEDDSN  485
             R ++KP   EE DG+V L     +T ASAAQRRGR+GR   +  DEY Y G  ++DDS
Sbjct  283  GRTNIKP---EEVDGKVELTGTRRVTTASAAQRRGRVGRQDGRT-DEYIYSGQCDDDDSG  338

Query  486  FAHWTEARIMLDNINMPNGLVAQLYQPEREKVYTMDGEYRLRGEERKNFLEFLRTADLPV  545
               W EA+I+LDNI    G VA  Y PE++K+  + G +RL  E+RK+F   L   D
Sbjct  339  LVQWKEAQILLDNITTLRGPVATFYGPEQDKMPEVAGHFRLTEEKRKHFRHLLTHCDFTP  398

Query  546  WLAYKVAAAGISYHDRKWCFDGPRTNTILEDNNE-VEVITKLGERKILRPRWADARVYSD  604
            WLA+ VAA   S  DR W ++GP  N + E + + V   +  G  + LRP W DAR++ +
Sbjct  399  WLAWHVAANVSSVTDRSWTWEGPEANAVDEASGDLVTFRSPNGAERTLRPVWKDARMFKE  458

Query  605  HQALKSFKDFASGKR  619
             + +K F  +ASG+R
Sbjct  459  GRDIKEFVAYASGRR  473
```

Figure 24 - Sequence alignment of West Nile Virus NS3 polyprotein (Query) and TBEV NS3 helicase (Sbjct). Done with BLAST.