

**Comparison of sequences and structures between NS3-NS2B of various Flaviviruses**

*Submitted to Worcester Polytechnic Institute for partial fulfillment  
of the requirements for the Degree of Bachelor of Science*

By: Thai Thuy Anh Le

May-6-2021

**Advisors:**

Dr. Patricia Z. Musacchio (WPI)

Dr. George A. Kaminski (WPI)

**Sponsors:**

Dr. Celia A. Schiffer (UMass Medical School)

Dr. Nese Kurt Yilmaz (UMass Medical School)

**This report represents the work of one or more WPI undergraduate students submitted to the faculty as evidence of completion of a degree requirement. WPI routinely publishes these reports on the web without editorial or peer review**

## Table of Content

Acknowledgement.....	3
Abstract.....	4
I. Introduction.....	5
1.1 Background of flavivirus and disease.....	5
a)    Epidemiology.....	5
b)    Flavivirus’s structure.....	12
c)    Flavivirus’s life cycle and cell entry.....	13
d)    Genome.....	14
1.2 Structural proteins of Flavivirus.....	16
1.3 . Non-structural 3 (NS3) Protease-Helicase.....	18
a) Structures.....	18
b) Function.....	19
II. Methodology.....	20
2.1    Sequence Alignment.....	20
2.2    Structural Comparison.....	21
III. Results and Discussions.....	21
IV. Appendix.....	24
V. Reference.....	79

### Acknowledgement

I would like to thank you to Dr. Celia A. Schiffer and Dr. Nese Kurt Yilmaz from University of Massachusetts Medical School for guiding me and providing me an invaluable opportunity to conduct my research in collaboration with the Schiffer lab. This opportunity has shaped my approach in science, both in the present and in the future. I would like to thank you Dr. Patricia Z. Musacchio of Worcester Polytechnic Institute, who has been acting as my mentor for this project, and has giving me wonderful advice in conducting research, despite the current pandemic situation. Additionally, I want to thank you Dr. George A. Kaminski of Worcester Polytechnic Institute and Winnie W. Mkandawire from University of Massachusetts Medical School for advising me in using protein structural and sequences alignment software and websites, all of which play a crucial role in helping me conducting this research. . I also would like to thank you my roommate, Vien Phuong Thuy Le, who helped me optimize my workflow with a C/C++ based program that counts and labels number for the residues in the CLUSTAL W files. For this project, molecular graphics and analyses were performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311. Multiple sequences alignment was performed by web-based protein sequence alignment software T-Coffee and MUSCLES, developed by Di Tommaso's group.

## Abstract

Flavivirus is a genus of virus that belong to the *Flaviviridae* family, a family of positive single-stranded, enveloped RNA viruses that transmit to humans through arthropods as their vectors. There has been recorded of more than 70 flaviviruses, with some of them causing potentially fatal human illness: yellow fever virus (YFV), West Nile virus (WNV), Dengue virus (DENV), Zika virus (ZIKV), and Murray Valley encephalitis virus (MVEV).

In flavivirus, non-structural (NS) proteins have been found to play a vital role in the proliferation and transmission of flavivirus. The complex NS2B-NS3 protein is associated with the endoplasmic reticulum membrane and possesses dual functions for processing RNA and polyprotein. This article is a comparison in structures and sequences between various NS3 proteins among various human-borne flaviviruses and one swine-borne pestivirus, utilizing 3D structural modeling programs USCF Chimera, and global sequence alignment programs T-COFFEE Espresso, and MUSCLES.

Based on the structural and sequence analysis data, it was observed that the helicase domains of the Flavivirus species are the most conserved domains in terms of structures and sequences, while the protease regions display more variations and diversities. Interestingly, despite being far from relation on the phylogenetic tree, DENV4 helicase domains share a high score percent identity and low RMSD values when compared to Kunjin virus and MVEV (with DENV belongs to the *Aedes* mosquito-borne group, and MVEV, Kunjin belong to the *Culex* mosquito-borne group).

## Introduction

### 1.1 Background of flavivirus and disease

#### a) Epidemiology

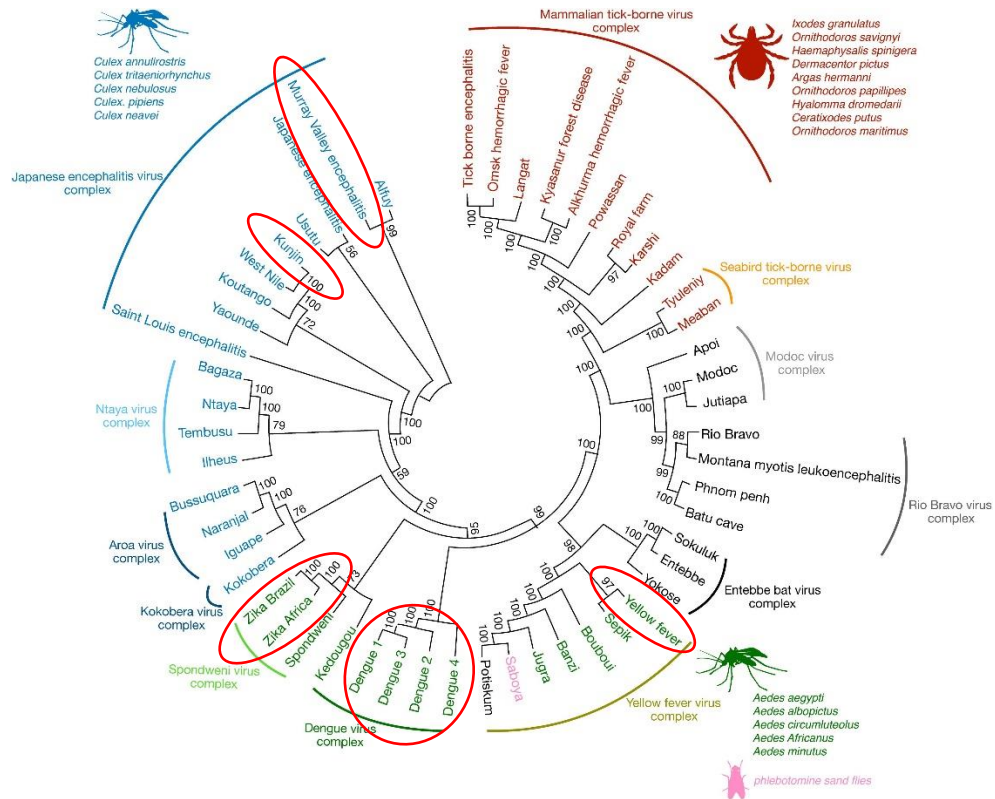


Figure 1. Phylogenetic tree of Flaviridae family members. Adapted from Rathore, Abhay P. S. and St. John, Ashley L., 2020. The

Flavivirus species that are in the scope of this study are circled in red.

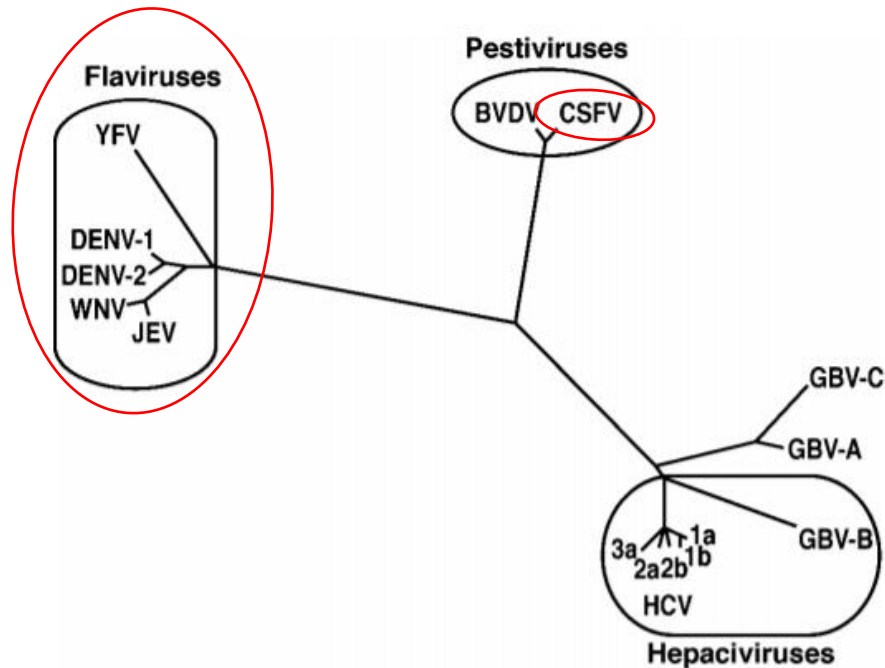


Figure 2. Phylogenetic tree of Flaviviridae family members. Pestivirus genus' relation to Flavivirus genus. Adapted from Lindenbach et. al, 2007 (original source: Simmonds, 1995).

The Flaviviridae is a large family of viral pathogens responsible for causing severe disease and mortality in humans and animals. The family consists of four genera: Flavivirus, Pestivirus, Pegivirus and Hepacivirus (Hayes, 2012; Luo, 2008 and Rathore et. al, 2020). The Pestivirus genus contains 11 species, many of which, such as Classical Swine Fever (CSF), can infect species in the family *Bovidae* (e.g.: cattle, sheep, and goats) and the family *Suidae* (species of swine) (Lindenbach et. al, 2007, ICTV, n.d.). Diseases associated with this genus include hemorrhagic, fever, abortion, and fatal mucosal disease. The Flavivirus genus, which is the largest of the three, contains more than 70 species including Dengue Virus (DV), Murray Valley encephalitis (MVEV), West Nile Virus (WNV), Yellow Fever Virus (YFV), and Zika Virus (ZIKV). These diseases are known to cause systemic febrile syndromes and hemorrhagic fever. The encephalitis ones can cause neurotropic, neurologic inflammation, which can lead to

permanent brain damages in some cases. Other flaviviruses, while are not known to pose a risk to human health yet can elicit cross-reactive antibodies (Geok Saw et al., 2019).

- West Nile virus (WNV):

West Nile virus was discovered in 1937 in Uganda, Eastern Africa. Additionally, it also has a subtype, known as Kunjin virus. WNV is transmitted mostly by *Culex* mosquitoes (Hayes, 2012). Infection can also occur through contact with human milk, infected blood, and organ transplant. WNV epidemic frequently happens in Africa, Asia, Australia, Europe, North and Latin America, and the Caribbean (Hayes, 2012 and CDC, 2020). WNV infections are usually asymptomatic. However, symptoms can manifest after 2-14 days of incubation (less than 1% of infected patients have neurotropic symptoms). Typically, patients develop headaches, fever, myalgia, weakness, abdominal pain, nausea and vomiting, diarrhea. Severe symptoms include malaise, neck stiffness, movement disorders like tremors, seizures, acute flaccid paralysis (with or without meningitis). The infection rate has been recorded to be less than 3% in the U.S, but this number can be higher in other places. From 1999-2005, there were more than 19000 cases of WNV in the US (CDC, 2020). As for prevention and treatment, several vaccines are under development, but none are available so far. Ribavirin and interferon (antiviral drugs) are used in clinical settings, but the results are inconclusive, so most treatments are intensive-supportive care.

- Zika Virus (ZIKV):

Zika virus was identified in 1947. Its name was based on the Zika Forest in Uganda (CDC, 2019). The virus is mostly transmitted by *Aedes* mosquitoes. In human, infection human can occur as perinatal transmission (from mother to fetus) or from sexual contact. (CDC, 2019).



ZIKV epidemic is located in tropical Africa, Southeast Asia, Pacific Island; however, there was a large-scale outbreak in the Americas in 2015-2016 (CDC, 2019). Like many other flaviviruses, ZIKV infection can be asymptomatic (Hayes, 2012; Geok Saw et al., 2019). If infected, incubation can take place from 3-14 days. Once symptoms manifest, patients will develop rash, fever, headache, joint pain, muscle pain and red eye. Modern-day Zika has developed more pathogenicity comparing to the newly discovered ones, and has known to cause encephalitis, acute disseminated encephalomyelitis, and Guillain-Barre syndrome as well (Hayes, 2012). In pregnant women, Zika can cause birth defects such as microcephaly, miscarriage, stillbirths, and other birth defects. According to CDC in 2018, 90% of cases of ZIKV infection happened in Africa. Despite a long-known history, no vaccines for ZIKV have been developed so far. Supportive treatments are required, such as replenishing fluids and electrolytes, using non-steroid anti-inflammatory drugs, and other pain medications. Interferons are utilized for treating ZIKV, however, there has been reported interferon-resistant ZIKV strain.

- Yellow Fever (YFV):

Yellow Fever virus was among the first of the Flavivirus genus to be identified and studied and was the basis for the name “*Flavivirus*” (with *flavus* means 'yellow' in Latin) (Mitchel, 1744). YFV transmission mostly occur through *Aedes* and *Haemogogus* mosquitoes' bites. Direct human transmission has not been recorded, but transmission of “the attenuated vaccine strain from a mother to her fetus has been described” (Hayes, 2012). YFV epidemic has been found in Africa, and South America, as well as in the U.S. Once infected, the incubation period ranges from 3-6 days. Patients will develop the acute onset of headache, fever, chills, and myalgia, along with photophobia, back pain, anorexia and vomiting, restlessness. They will also have viremia. In rare cases, patient might acquire myocardial dysfunction. Although central

nervous system invasion is rare, once happens, can cause convulsions, coma and delirium. As for prevention and treatment, infected patients will need intensive supportive care. In severe cases, such as with neutropenia patients, prophylactic antibiotics will be utilized in case of secondary bacterial infection. Interferon can be used as a preventive medication but *not as a treatment* given 24 hours after infection (Hayes, 2012). Immunoglobulin with neutralizing antibodies against YFV is beneficial as a therapeutic drug but required given in the early clinical course.

- Murray Valley encephalitis (MVEV):

Murray Valley virus was first recognized in 1917, isolated from a human brain tissue in 1951 (Rust, 2014). MVEV transmission mostly occur by *Culex annulirostris*, *Aedes normanesis* mosquitoes and waterbird (which is believed to be a part of the cycle) (Floridis et. al, 2018; Hayes, 2012). MVEV encephalitis epidemic are recorded to be in Australia and Papua New Guinea. Enzootic infection also found in the Top End, Katherine, and Barkly regions of Northern Territory and Northern Western Australia (Floridis et. al, 2018). Interestingly, MVEV has same antigenic complex with JEV and WNV (Floridis et. al, 2018; Hayes, 2012; Knox et al., 2012).

Typical symptoms of MVEV infection are relentless progression to death, flaccid paralysis, cranial nerve palsy, tremor and encephalitis. While the majority of infection is asymptomatic, between 1 in 150, and 1 in 1000 infections end up with symptomatic results. Fatality rate is 15-30%, 30-50% survivors suffer long-term sequelae. Sentinel chicken flocks are used for flavivirus surveillance, acting as an early warning system to detect MVEV activity in the Northern Territory of Australia. Similar to many flavivirus encephalitis and neurological illness, there are no clinically proven therapeutic options for MVEV so far. In Knox et al., 2012, where the researchers discussed the case of a 69-year-old man living in Murray River town of Mildura presented to the hospital for MVEV infection, the patient received intensive, supportive care,

with acyclovir and benzylpenicillin for his secondary bacterial infection. He was also treated with corticosteroid to modulate immune system responds, and more broad-spectrum antibiotics have been added to his treatment, but to no avail (the patient was announced brain death on Day 10 of hospitalization). Ribavirin and Interferon alfa-2a have shown inconsistent results in clinical, placebo-controlled studies against JEV (which is a flavivirus with same antigenic complex to MVEV and WNV), and immunoglobulin therapy is also unclear in its therapeutic role (in 1917, a 12-year-old patient with symptoms suggestive of MVEV infection, received immune serum and recovered).

- Dengue virus (DENV):

Dengue virus is first discovered in 1943 by Ren Kimura and Susumu Hotta, by isolating blood samples of patients during the 1943 epidemic in Nagasaki, Japan ( Beasley et al., 2008). The serotype that the scientists discovered at the time was proven later to be DENV-1. In fact, it was found out later that there are four known subtypes of DENV: DENV-1, 2, 3, and 4. Each serotype/subtype produces type-specific immunity that cannot be crossed-immune, meaning one person could acquire 4 different dengue antibodies in their lifetime 40% of the world's population, roughly 3 billion people, “live in areas with a risk of dengue.” (CDC, 2020; Hayes, 2012). DENV infection is frequently transmitted through the bites of *Aedes* mosquitoes, such as *Aedes aegypti* and *Aedes albopictus*. The disease is epidemic in Tropical Asia and Americas. Interestingly, there are observation of enzootic cycles of dengue infection in West Africa and Malaysia, although they do not directly influence the infection rate in human (Hayes, 2012). When infected, patients will acquire the following symptoms: dengue hemorrhagic fever (DHF), dengue fever, and dengue shock syndrome (DSS). Incubation period can be from 4-7 days (or as long as 2-15 days). Most infections are asymptomatic or mild illness; however, classic dengue

fever is an acute illness. Most severe manifestation include increased vascular permeability, increasing hematocrit, development of pleural effusions, and positive tourniquet-test. Progression to decreased circulatory volume and shock can occur suddenly. Fatality rate of DSS in untreated patient is about 50%. However, mortality rate is below 1% for early diagnosis. Similar to other flavivirus epidemic, no specific treatment nor vaccine for DENV is found yet. Patient with classic dengue fever should be managed with rest, closed observation, and adequate hydration. It is crucial not to use Aspirin during the course of treatment since Aspirin can interference with platelet function. Vaccines are under development, but not available currently.

- Classical swine fever virus (CSF):

Classical swine fever outbreak was first recorded in 1833, in Ohio of the U.S. (Moennig et al., 2008). CSF spreads through direct contact with cattle, sheep, or goats (Risatti et al., 2020). Wild boars and species within the *Suidae* family, vehicles and equipment, and even asymptomatic infected pigs can also carry potential risks of infection. CSF is considered an endemic in certain countries of Central and South America, in the Caribbean basin, and in many pork-producing countries in Asia. Currently, there is no record or research demonstrating CSF infection in human. In swine, symptoms manifest as fever, hemorrhages, ataxia, and purple discoloration of the skin. Fever over 41°C (105.8°F) is usually seen and continues until the final stages of the disease, accompanied by diarrhea and vomit. Incubation ranges from 3-7 days, with death occurring within 10 days after infection (in highly virulent and fatal CSF variants). Most of the outbreaks (95%) occurred in backyard piggeries. The principal causes of transmission of CSF were the introduction of infected pigs (38%), movements of people (37%) and unknown origin (13%). The epidemiological relationships with 15 affected farms explained 31 outbreaks. The

overall attack and mortality rates were 39% and 32%, respectively. The main clinical signs were high fever (67%), incoordination of movements (54%), and prostration (52%).

To prevent CSF outbreak, CSF live-attenuated vaccines are considered to be effective choice, being able to induce protection shortly after vaccination (within 3 days) (Risatti et. al, 2020). Countries free of CSF, however, forbid the use of prophylactic vaccination and exert strict control of movement of animals, including domestic pigs. In these countries, herds affected by an outbreak of CSF are quarantined and exposed animals eliminated. Pre-emptive slaughter of pigs is sometimes applied within established quarantine zones. Oral vaccination of wild boars has been used successfully within the European Union using a live-attenuated vaccine delivered via baits.

#### *b) Flavivirus's structure*

Flaviviruses have small size (~ 50nm), with an electron dense core approximately 30nm, and a lipid envelope (Lindenbach et. al, 2007; Creative Diagnostics, n.d.). Its densities vary between 1.19 to 1.23g/cm<sup>3</sup>, depending on the host's physiology, which determines the lipid's features.

Like many human host viruses, flaviviruses are composed of:

- Genomic, single stranded, (+) sense RNA (~11kb): codes for 3 structural proteins: capsid (C protein), membrane (M, which is expressed as prM, the precursor to M and envelope (E protein) and 7 nonstructural proteins: NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5.
- Protein capsid (C protein): uniformity with an icosahedral capsid and closefitting, spiked envelope. The size is about 30nm. The whole particle is 45nm in diameter.

- Membrane protein (M protein): produced when the virus particle matures and exists as prM (precursor M) in immature virion.
- Envelop protein (E protein): its glycoprotein mediates fusion and binding for virus entry; also functions as the antigenic marker of the virus.
- Lipid envelope: composed of a lipid bilayer with two or more species of envelope (E) glycoproteins surrounding a nucleocapsid.

c) *Flavivirus's life cycle and cell entry*

The life cycle of flavivirus upon infecting human's immune cells, particularly the dendritic cells, happens in the following order:

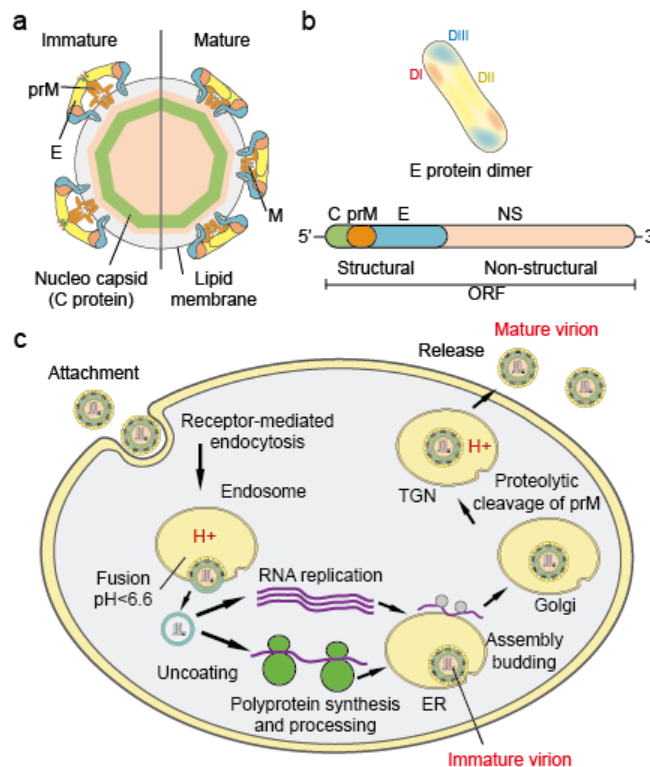


Figure 3. The overall structure and life cycle of a typical Flavivirus viral particle. Adapted from Creative Diagnostics.

First, the virus attaches to the cell surface through receptor-mediated endocytosis. Viral RNA genome enters the host cell's cytoplasm. After that, viral RNA replicates in the rough

endoplasmic reticulum (ER) and the double membrane vesicle packets (VP) derived from the Golgi-apparatus. The non-structural protein and double stranded RNA are then concentrated in the VP to prepare for viral RNA synthesis. The newly synthesized RNA is exported to the intermembrane space of the VP and subsequently exits into cytoplasm. Then, viral particles assembled in the rough ER. The newly synthesized genome is packaged by a viral capsid protein (C) which surrounded by a lipid bilayer embedding E (envelope) and prM (precursor of membrane) proteins. The immature virions are spiky, measuring 60 nm in diameter. The immature virions then bud off from the lumen of the rough ER and transported to the trans-Golgi network (TGN) for maturation process to take place. After maturation, the viral particles buds outside the host's cells, taking up a part of the lipid membrane of the host. From here, the flavivirus particles flow in the circulatory system of the patient and capable of infecting new host or new vectors.

#### d) Genome

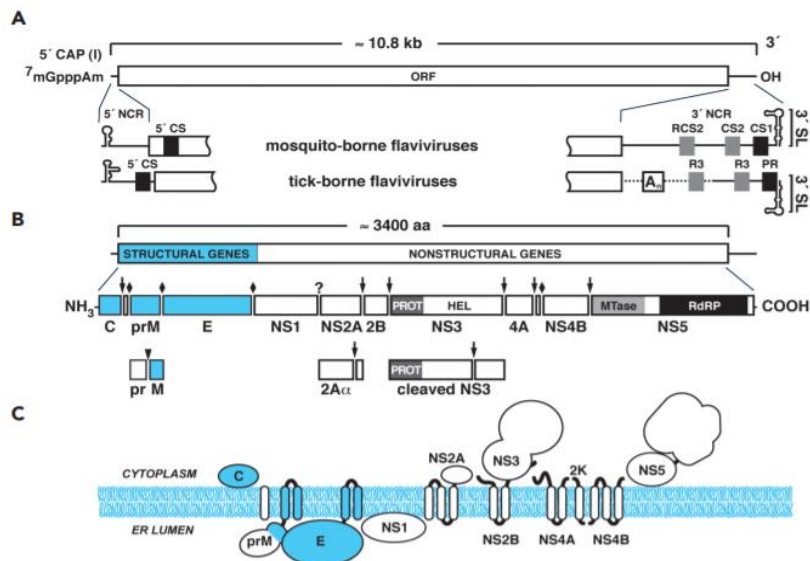


Figure 4. A: Genomic structure of a typical Flavivirus viral particle. B: Diagram and cleavage sites of a single, expressed polypeptide. C. Sites of expressions of the proteins after cleavage in mature Flavivirus particle. Adapted from Lindenbach et. al, 2007.

Flavivirus possesses a single stranded, (+) sense RNA that encodes for one long polypeptide, one that can be cleaved into The genome of flavivirus is constructed using cDNA clones. It is found that the genome of flavivirus is a single stranded, (+) sense RNA (~11kb), with sedimentation at 42S, and a 5' type I cap, m<sup>7</sup>GpppAmpN<sub>2</sub>. Unlike messenger RNA (mRNA), this genome lacks 3' polyadenylate tail. Their genomes have one single open reading frame, with 5' and 3' noncoding regions (NCR) of ~100 nucleotide (nt), 400 and 700 nt, respectively.

The 5' NCR vary between different species of flavivirus; however, common secondary structures can be found within this region. These structures can alter the translation of the genome: for example, some morpholinos (nucleotide analogs that bind to the short sequences in the transcription start sites, or the pre-mRNA splice sites) can prevent DENV RNA translation, and thus prevent the virus from replication. If there is mutation in this region, or in the complementary region of the negative strand, then flavivirus cannot replicate.

The 3' NCR also exhibits varieties between species of flavivirus. Still, this region has some conservation among each species, such as the existence of a long 3' stem-loop (3' SL) with essential virus-specific and host-specific functional regions within 3' SL, found in WNV and DENV-2 mutational analysis. 3' SL can interact with other proteins that share similar functions, including NS3 and NS5.

The pathogenicity of flavivirus depends on its ability to translate the genome. To facilitate translation process, the virus has evolved beneficial traits. For example, its 5' NCR and 3'NCR contain specialized structures as mentioned above that target specific hosts (the difference



between tick-borne viruses and mosquito-borne viruses), thus enhancing the flaviviruses chance to infect. Overall, translation is dependent on the cap of the RNA and initiated by ribosomal scanning. The product of the translation is a single, long polypeptide that is co and post-translationally cleaved into at least 10 proteins. On the N-terminal side, one fourth of this long chain polypeptide encodes for structural proteins: C protein, E protein and pr-M protein. The non-structural proteins come next: NS1-NS2A-NS2B-NS3-NS4A-2K-NS4B-NS5. For cleaving, the host signal peptidases cleave between the C/prM, prM/E, E/NS1 and 2K-NS4B. The virus serine protease then proceeds to cleave NS2A/NS2B, NS2B/NS3, NS3/NS4A, NS4A/2K and NS4B/NS5. Currently, the enzyme that cleaves NS1-2A is unidentifiable.

### ***1.2. Structural proteins of Flavivirus***

- Protein capsid (C protein):

The capsid protein of flavivirus is a highly basic protein, with a weight approximately 11 kd. The charged residues are concentrated at the N- and C-termini, with an internal hydrophobic region locates in the middle that also meditates membrane binding and entry of the virus. Newly synthesized C protein (anchC) contains a C-terminal hydrophobic, which is a signal peptide for prM's ER translocation. After maturation, this hydrophobic region will be cleaved by viral serine protease. In mature virus, C protein has a conformation of a dimer, with each monomer containing four alpha helices. RNA binding and membrane interaction surfaces are heavily relied on the specific asymmetric distribution of positive charged and hydrophobic residues on C protein. Additionally, it is found that large portions deletion in C protein-coding sequences can be fatal to the virus replication, but second-sites changes that increase hydrophobicity of downstream sequences can save the virus. The mechanism that induces C protein dimers to

organize within nucleocapsids is not yet clear, but it is suspected that interactions with DNA or RNA trigger the process.

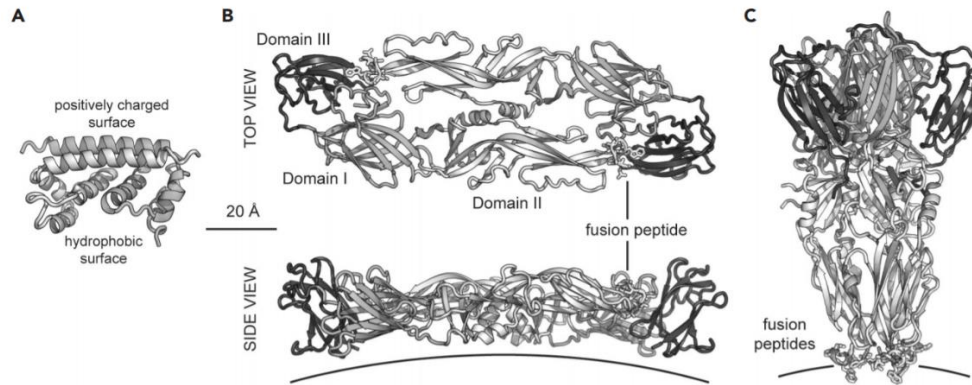


Figure 5. Structure of C and E proteins of an example Flavivirus.

A: Dengue virus 2 (DENV-2) capsid protein Dimer.

B: DENV-2 E glycoprotein dimer.

C: Tick-borne encephalitis virus (TBEV) E protein trimers.

Adapted from Lindenbach et. al, 2007.

- Membrane protein (M protein):

The glycoprotein precursor of M protein is translocated into the ER by the C-terminal hydrophobic region of C protein. The process is influenced by the cleavage of signal peptidase, caused by viral serine protease, thus generates mature C protein. The N-terminal region of prM contains the linked glycosylation that are all disulfide linked. In immature virus, prM is cleaved into pr and M fragments by the furin protease. In addition, prM protein also assists in folding of E protein. When prM cleavage is inhibited, the viral particle will be render noninfectious and immature.

- Envelop protein (E protein):

The E protein is ~ 53kDa and being one of the most major protein that is expressed on the outside membrane of the virus. Its glycoprotein meditates fusion and binding for virus entry and also functions as the antigenic marker of the virus. The E protein is a type I membrane protein, with 12 cysteines that form disulfide bonds, and can be N-glycosylated. The folding, stabilization and secretion of E is dependent on prM, and thus the complex prM-E, if disrupted, will yield noninfectious proteins.

### ***1.3.Non-structural 3 (NS3) Protease-Helicase***

#### *a) Structures*

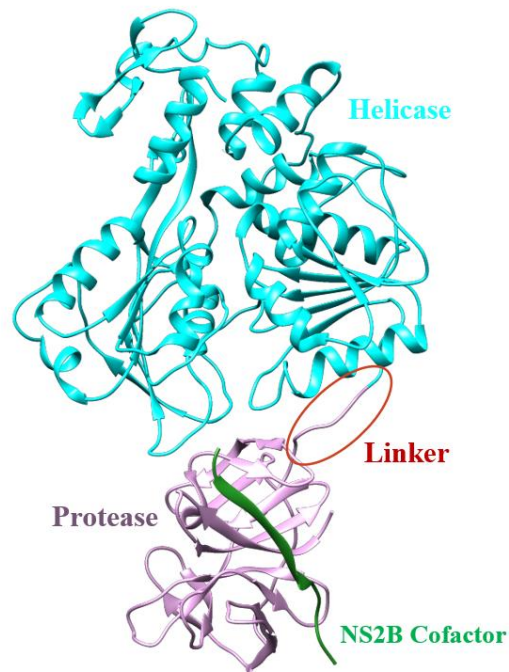


Figure 6. Example of an NS2B/NS3 protein. PDB ID: 2VBC. Image was generated by UCSF Chimera

NS3 protein is a large, ~70kDa multifunctional protein that is important for the virus polyprotein processing and RNA replication. The protein has two main domains: the helicase region (C terminal) and the protease region (N-terminal third of the protein) (Lindenbach et al., 2007). The protein consists of approximately 618 residues, and the two main domains are linked by a glycine-rich linker region. The protease region has a catalytic serine complex, the His-51, Asp-75, and Ser-135 (found in DENV4 NS3 protein); meanwhile, the hydrolysis/ RTPase/ NTPase catalytic cores of the helicase region rely on the motifs within its three subdomains (Luo et al., 2008).

Associating with NS3 protein is co-factor NS2B. NS2B is a small, ~14kDa protein, that forms a complex NS2B NS3 serine protease. This membrane-associated protein locates in within the fold of the protease domain, and its activity depends on a central peptide. Mutation causing lack of NS2B in structures can cause severe instability, such as autoproteolytic cleavage, transcleavage and insolubility in solvent when crystallized. Additionally, studies of WNV and DENV-2 crystal structures reveal that NS2B co-factor also lends the NS3 protease region a  $\beta$  strand to form the chymotrypsin-like fold.

#### *b) Function*

NS3 protein's main functions are polypeptide cleavage, RNA hydrogen bonds cleavage (helicase activity), and NTPase/RTPase activity. On the helicase domain, the structure consists of three subdomains, with two subdomains conserved among the Flavivirus NS3 helicase family members. These two subdomains are involved in hydrolysis of NTP, in addition to a unique C-terminal subdomain that may also involve in virus-specific RNA and protein recognition. The helicase region's catalytic cores for NTPase/RTPase and hydrolysis activity depend on a set of conserved motifs (motifs I, II, and VI) within its two N-terminal Rec-A-like subdomains (Luo et.

al, 2008, Luo et. al, 2015). In DENV-2 (see Appendix) and TBEV, the protease region is shorter comparing to other flavivirus NS3 protease, and thus the serine protease catalytic sites are located in the helicase domain instead. It was observed that Kunjin virus with large in-frame deletion mutations in the helicase region can still be compensated in trans, while deletion mutation in serine protease domain results in loss-of-function completely (Jones et al., 2005., Khromykh et. al, 2000). The linker region also plays an important role in the functionality of the domains. Its interdomain flexibility allows the protease region and the helicase region to be able to approach substrate effectively; deletion or insertion mutation in the linker will significantly reduce the ATPase/hydrolysis and proteolysis activity (Luo et al., 2010). The serine protease catalytic triad locates in the central left of NS3 protease region; this triad can recognize the positively charged residues Arginine/Lysine for initiation of cleaving.

## II. Methodology

### a) Sequence Alignment

For sequence alignment, FASTA (“fast-all”) files of the following Flavivirus NS2B/NS3 proteins were obtained from RCSB Protein Data Bank:

Virus names	Subspecies/Serotype	Protease (PDB ID)	Helicase (PDB ID)	Helicase-Protease (PDB ID)
<b>Zika (ZIKV)</b>		5GJ4	5JMT	
<b>Dengue (DENV)</b>	DENV1	3L6P		
	DENV2	4M9T	2BHR (RNA helicase). 2BMF (RNA helicase at 2.4Å resolution)	
	DENV3	(not found yet)	(not found yet)	
	DENV4			2VBC (NS3 protease-helicase). 2WHX (2nd conformation of 2VBC).

				2WZQ (Insertion mutation E173GP174 of 2VBC)
West Nile (WNV)	Kunjin virus		2QEQ	
Yellow Fever (YFV)		6URV	1YKS	

Table 1. Summary of PDB ID of NS2B/NS3 protease-helicase of various Flaviviruses

Multiple sequences alignment was conducted through T-Coffee and MUSCLES. The algorithm parameter is PAM 350 matrix, and the final results were exported in ClustalW format. The sequences were aligned in 10-residue-segments for homologies comparison (See Appendix).

#### *b) Structural Comparison*

3D crystallized structures of the aforementioned proteins in Table 1 were obtained from RCSB Protein Data Bank as PDB files. Structural analysis, including RMSD values, percent identity, and superimpositions snapshots, were performed in UCSF Chimera. For proteins that are closely related in the phylogenetic tree (DENV serotypes, MVEV, ZIKV, YFV, Kunjin), Needleman-Wunsch algorithm and BLOSUM-45 matrix were applied. For proteins that are far away in the phylogenetic tree (CSF), Smith-Waterman algorithm and BLOSUM-65 matrix were applied (See Appendix).

### **III. Results and Discussions**

- NS3 Protein (Helicase-Protease)

Based on the RMSD evaluation, DENV4 serotypes share the lowest RMSD among each other (RMSD across all 580 pairs: 0.876Å), following by MVEV-DENV4 with RMSD across all 409 pairs at 1.138Å; 5WX1- MVEV with RMSD value across all 231 pairs at 2.481Å and 5WX1-DENV4 with RMSD value across all 5 pairs at 3.252Å. The percent identity goes from

the highest, which is the DENV4 serotype: 99.51% for 2WZQ-2WHX, 99.51% for 2WHX-2VBC, and 92.56% for 2WZQ-2VBC. MVEV-DENV4 ranks second with the percent identities: 42.88% (2WV9- 2VBC), 42.72% (2WV9-2WHX), 42.49% (2WV9-2WZQ). The lowest ones are CSF-MVEV at 5.65%, and CSF-DENV4 with the percent identity: 0.32% (5WX1-2WHX), 0.32% (5WX1-2VBC), 0.48% (5WX1-2WZQ)

- NS3 Protein (Helicase region)

Regarding the RMSD values, ZIKV-DENV4 has the lowest value across all 423 pairs at 0.985Å, followed by Kunjin virus-DENV4 with RMSD values across all 403 pairs at 1.343Å, DENV2 (2BHR)-DENV4 with RMSD values across all 431 pairs at 1.345Å, YFV-DENV4 with RMSD values across all 399 pairs at 1.361Å, and DENV2 (2BMF)-DENV4 with RMSD values 442 pairs at 2.989Å. The percent identity goes from the highest, which is among DENV2 crystallizations for 94.46% (2BMF- 2BHR); DENV2-DENV4 for 77.83% (2BMF- 2WHX), 77.38% (2BMF-2VBC), 77.16% (2BMF -2WZQ), 76.72% (2BHR-2WHX), 76.50% (2BHR-2VBC), 76.27% (2BHR vs 2WZQ). Followed are Kunjin virus- DENV4 for 61.52% (2QEQ-2WHX), 61.06% (2QEQ-2VBC), 61.06% (2QEQ - 2WZQ); YFV-DENV4 for 44.09% (1YKS-2WHX), 43.86% (1YKS- 2VBC), 43.18% (1YKS-2WZQ); ZIKV-DENV4 for 66.67% (5JMT-2VBC), 66.44% (5JMT-2WHX), and 66.44% (5JMT-2WZQ).

- NS3 Protein (Protease region)

With the RMSD obtained, the ranking from the lowest to the highest RMSD values goes from YFV-DENV4 for value across all 150 pairs is 2.322Å; DENV1-DENV4 with RMSD value across all 21 pairs at 2.393Å; DENV2- DENV4 with RMSD value across all 21 pairs at 2.432Å, and ZIKV-DENV4 with RMSD value across all 13 pairs at 2.816Å. Regarding the percent

identity, ranking from the highest to the lowest values are YFV-DENV4 at 47.95% (6URV-2VBC), 45.61% (6URV-2WZQ); ZIKV-DENV4 at 45.20% (5GJ4-2WHX); DENV1- DENV4 at 39.83% (3L6P -2VBC), 38.56% (3L6P- 2WZQ); DENV2-DENV4 at 38.06% (4M9T-2VBC), 37.25% (4M9T vs 2WZQ), 2.02% (4M9T-2WHX); YFV-DENV4 at 1.75% (6URV-2WHX), DENV1-DENV4 at 1.27% (3L6P-2WHX); ZIKV-DENV4 at 1.13% (5GJ4-2VBC) and 0.00% (5GJ4-2WZQ).

In addition, it was found that 2WHX and 2VBC are different at residues 13 (A13 in 2WHX, T13 in 2VBC and 2WZQ), 19 (T19 in 2WHX, S19 in 2VBC and 2WZQ), and 62 (T62 in 2WHX, S62 in 2VBC and 2WZQ). This might be one of the reasons that 2WHX has a protease region that is 161 degree turned from the protease region of 2VBC (Luo et al 2010). Another reason given in the Luo et. al study is that these 2 conformations are in an equilibrium in terms of energy, hence the conformation can be adapted interchangeably depending on the state of energy at the time of crystallization.

The information obtained from structural and sequence alignments have confirmed the relationship in the phylogenetic trees among the Flavivirus species, and as well as with the Pestivirus; CSF has a low percent identity and a high RMSD value when being compared to DENV4 and MVEV- two species from the Flavivirus genus. Interestingly, despite being far from relation on the phylogenetic tree, DENV4 helicase domains shares a high score percent identity and low RMSD values when compared to Kunjin virus and MVEV (with DENV belongs to the *Aedes* mosquito-borne group, and MVEV, Kunjin belong to the *Culex* mosquito group). The helicase domains of the Flavivirus species are indeed the most conserved domains in terms of structures (motif I, II, and VI) and sequences, with scoring of percent identity that can be more than 70%. Meanwhile, the protease domains are more prone to point mutation and variability,



suggesting that each virus species has evolved to possess diverse versions of protease regions to increase pathogenicity (to accommodate different hosts), yet still maintain the same helicase region for the purpose of replication. Both these domains are very ideal targets for making antiviral drugs.


#### IV. Appendix

##### Annotation:

The parameter for comparison is the “Point Accepted Mutation” (PAM) 350 matrix, which means the software will align segments with the least amount of point mutation that occurs in the compared sequences (in this case, about 350 mutations per 100 amino acids is the expected range). Scoring is assigned based on the physiological characteristics of the amino acid (hydrophobic, hydrophilic, aromatic, etc) and the amounts of mutations per limited range. Scoring can be positive or negative values. If the scoring is larger than zero/positive, the compared sequences are considered to be closely related. If the scoring is smaller than zero/negative, it is assumed that the referred sequences and the compared sequences are not closely related. The scoring can be used to assess the closeness in terms of functionality and phylogenetic trees.

For annotation purpose, the following symbols and markings are:

$\alpha 1$   
 Alpha helix

$\beta 1$   
 Beta strand

\* (asterisk) are positions that have a single, fully conserved residue. No highlight.

⋮ (colon) are positions that have conservation between groups of strongly similar properties – with scoring  $> 0.5$  in the Gonnet PAM 350 matrix. **Highlight in yellow.**

⋮ (period) are positions that have conservation between groups of weakly similar properties- with scoring  $\leq 0.5$  in the Gonnet PAM 350 matrix. **Highlight in magenta.**

Blank spaces are positions that do not fall in any scenarios of the above, which can be a non-conserved mutation. **Highlight in orange.**

⊖ are positions that lacking an amino acid, either due to lacking a corresponding structure (when comparing an NS3 protease-helicase against an NS3 protease fragment), deletion mutation, or an insertion mutation has changed the order of the aligned sequences (when comparing to other protein sequences). **Highlight in red.**

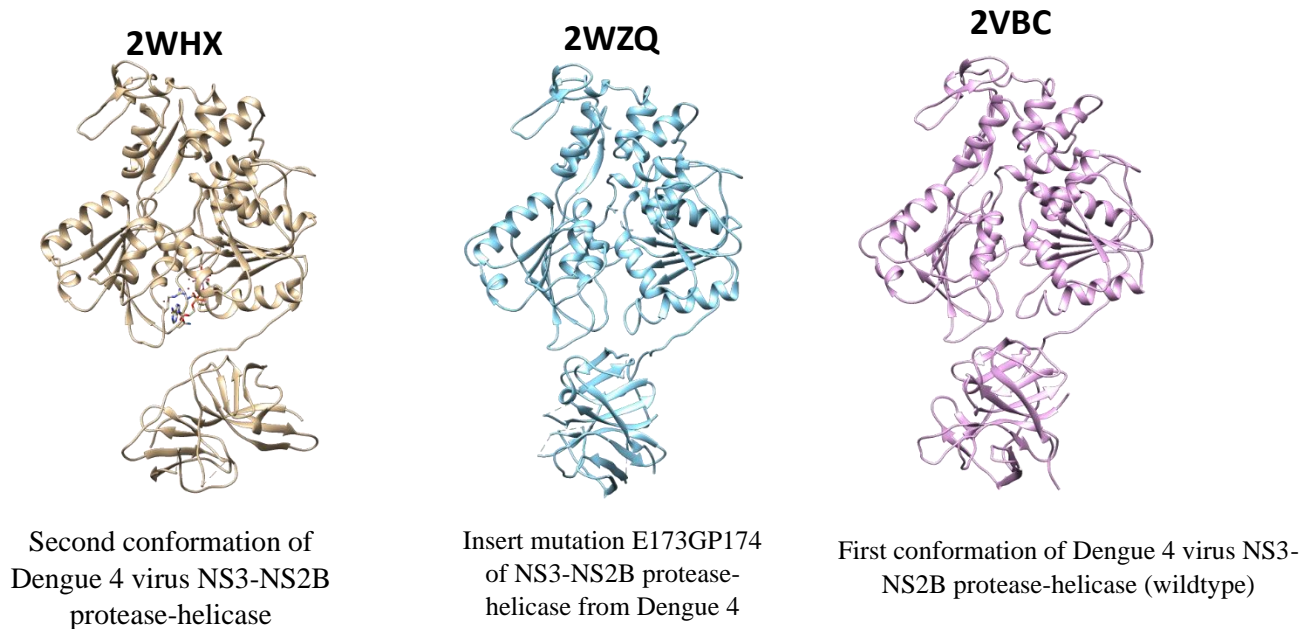


Figure 7. DENV 4 NS3 protein in superimposed position, border scale 0.8. Image was generated by UCSF Chimera

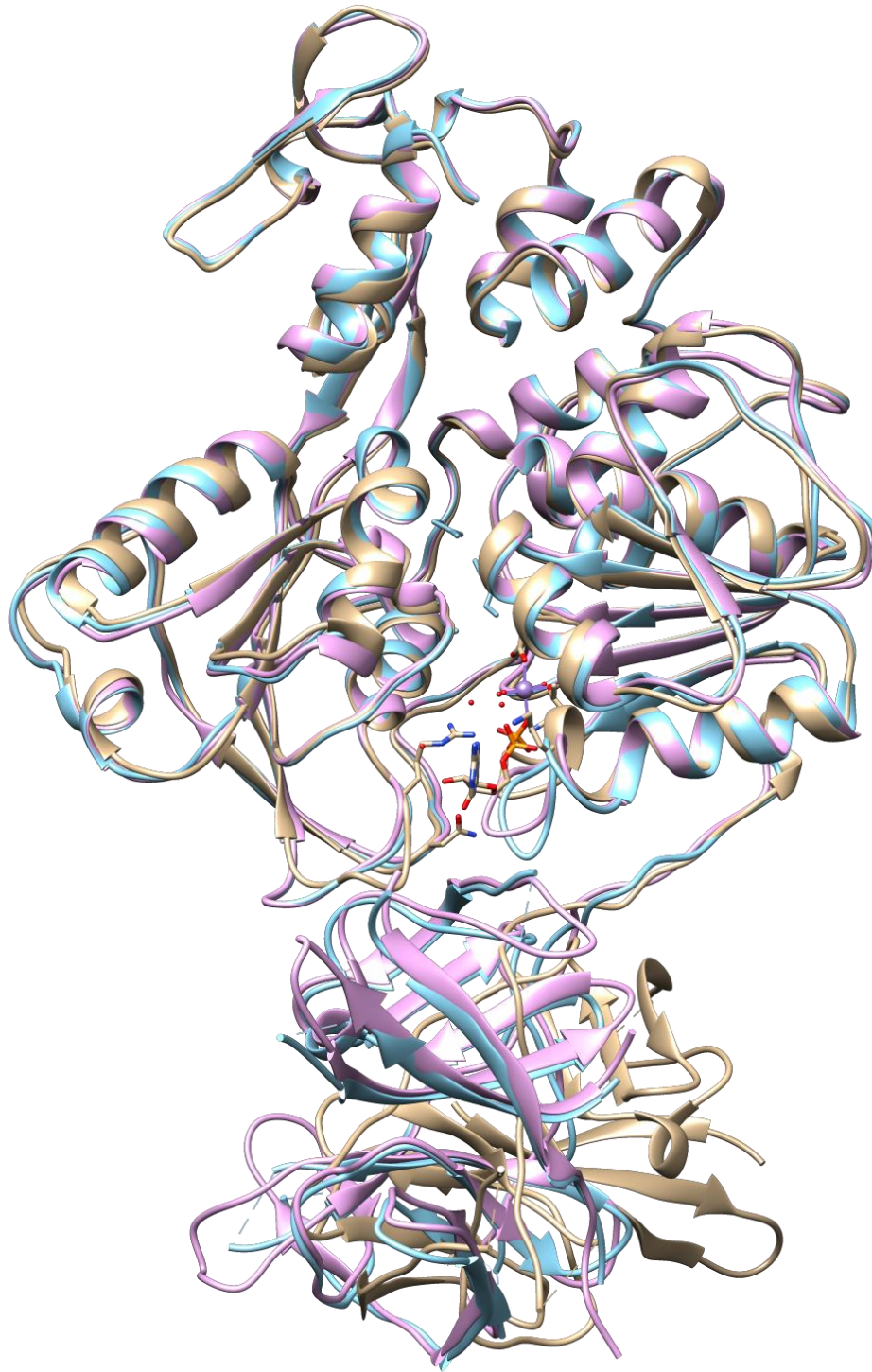


Figure 8. DENV 4 NS3 protein in superimposed position, border scale 0.0. Image was generated by UCSF Chimera

***Percent identity: 99.51% (2WZQ vs 2WHX)***

**Percent identity: 92.56% (2WZQ vs 2VBC)**

**Percent identity: 99.51% (2WHX vs 2VBC)**

**RMSD:  $\alpha$  Carbon pairs [182-438], 0.970 (2WHX vs 2VBC)**

**$\alpha$  Carbon pairs [82-619], 0.972 (2WZQ vs 2WHX)**

**$\alpha$  Carbon pairs [20-619] 0.876 (2WZQ vs 2VBC).**

**Across all 580 pairs: 0.876**

## NS2B CO FACTOR

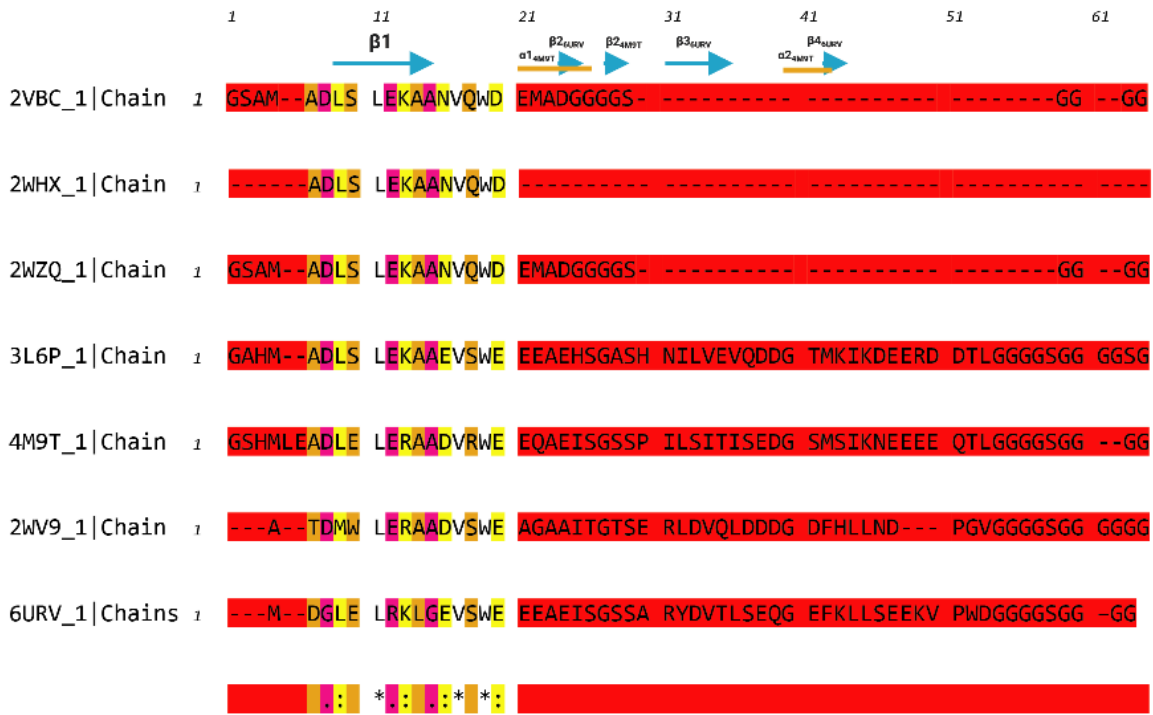
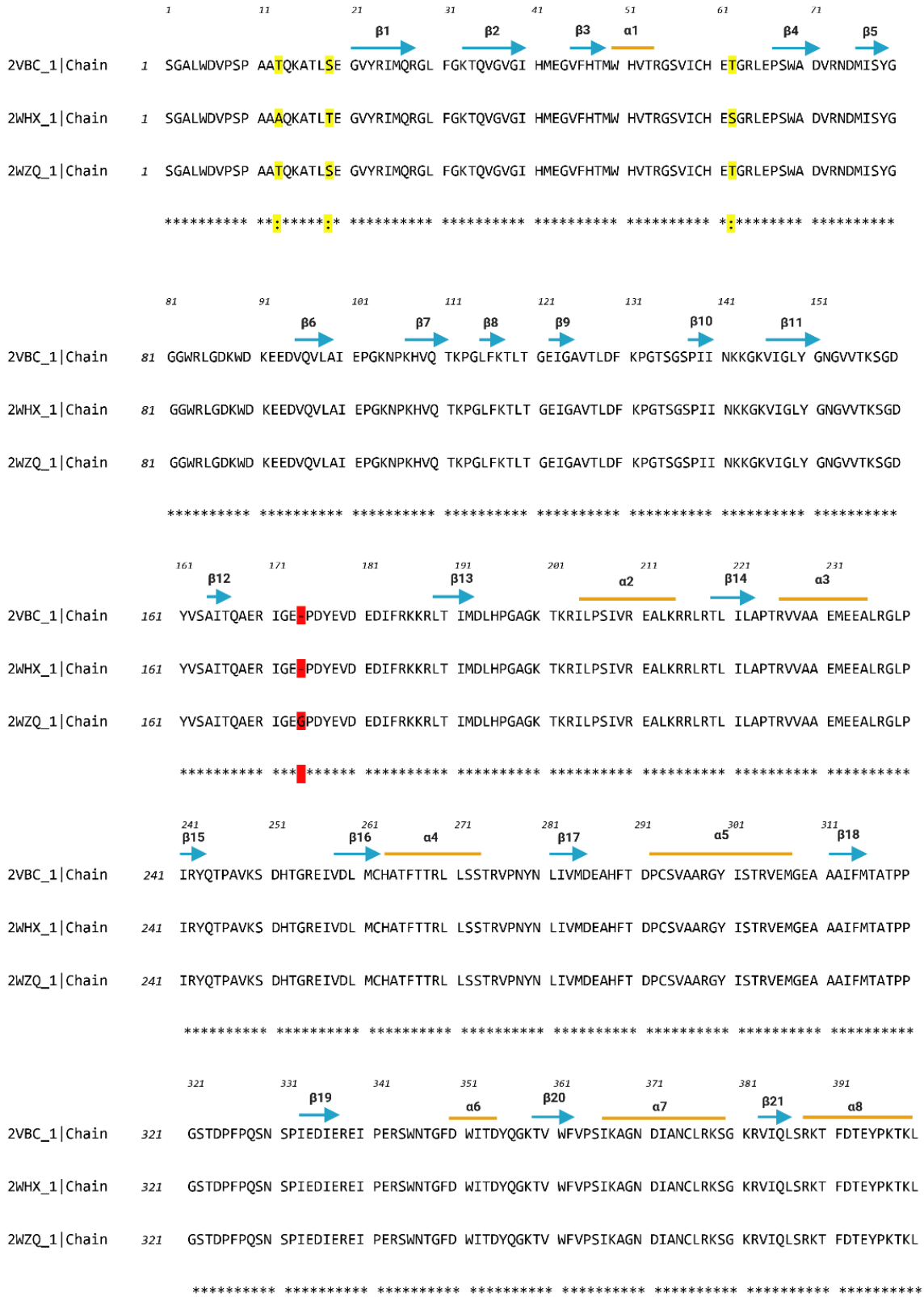


Figure 9. Sequence alignment of NS2B co-factor of various flaviviruses (Table 1). Alignment done by T-Coffee Expresso.



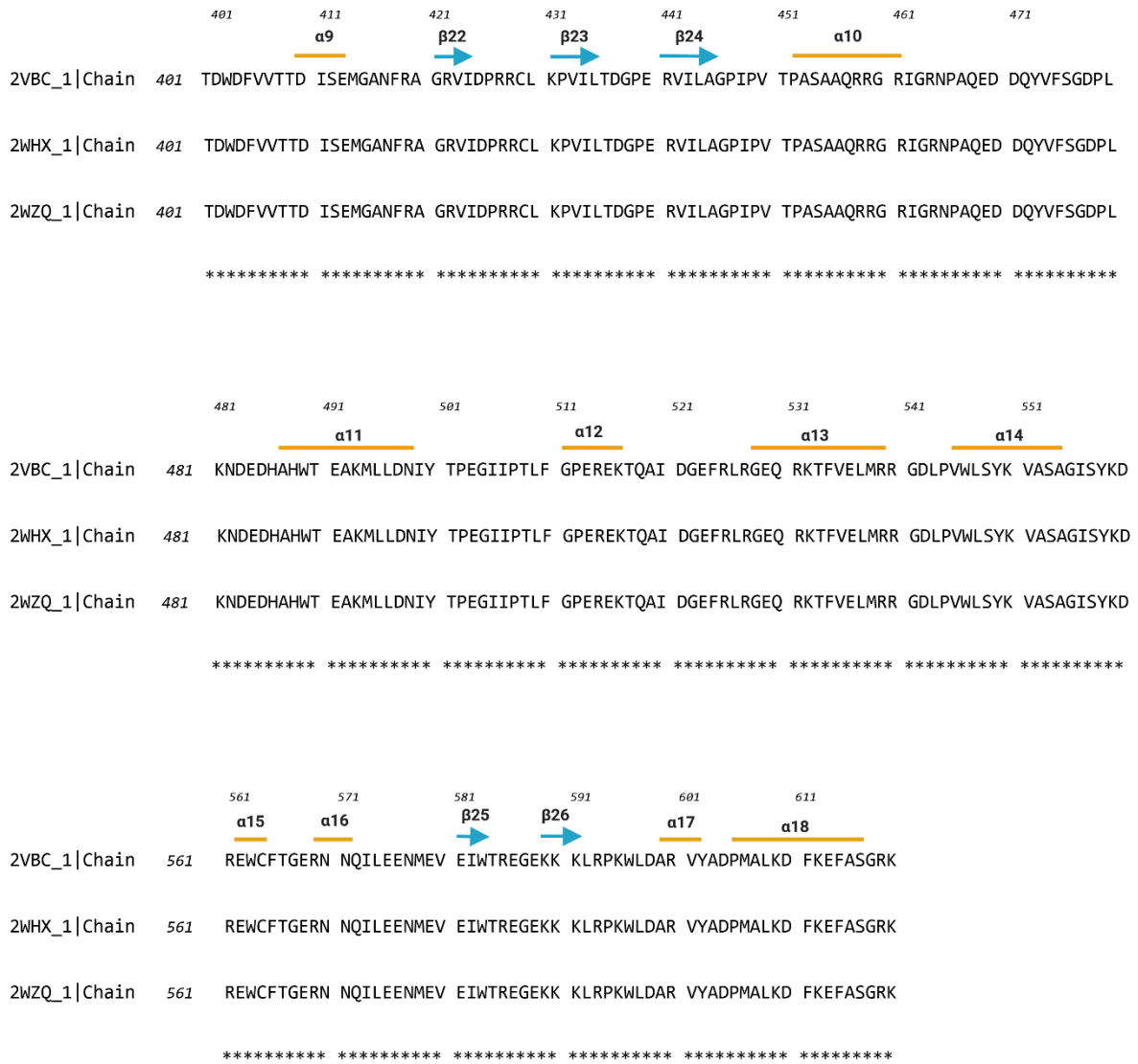


Figure 10. Sequence alignment of DENV 4 NS3 protein (Chain A). Alignment done by T-Coffee Expresso.

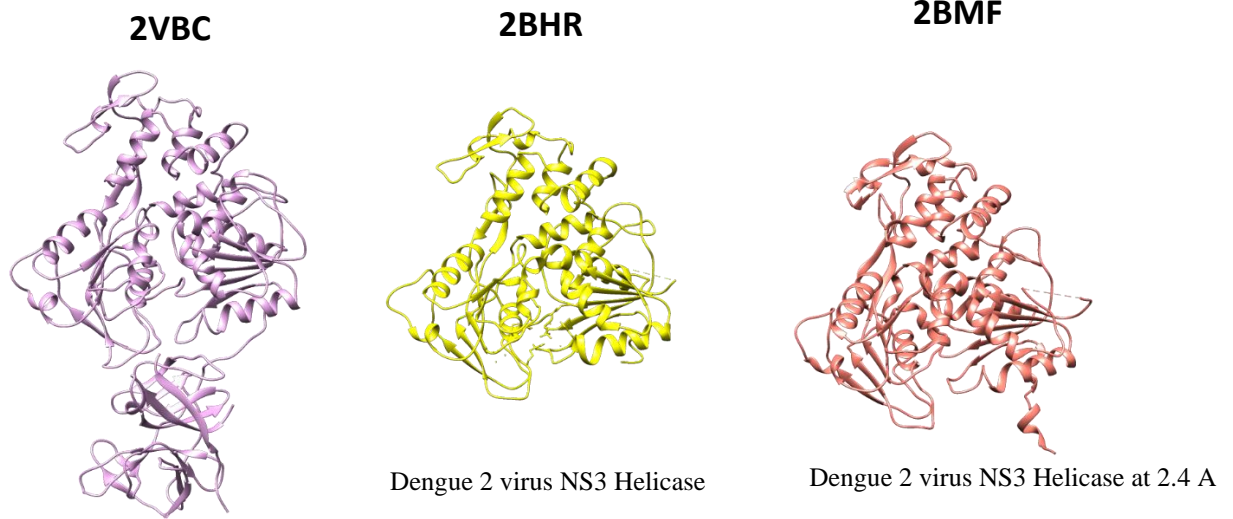


Figure 11. DENV 4 NS3 protein and DENV 2 helicases in superimposed position, border scale 0.8. Image was generated by UCSF Chimera.

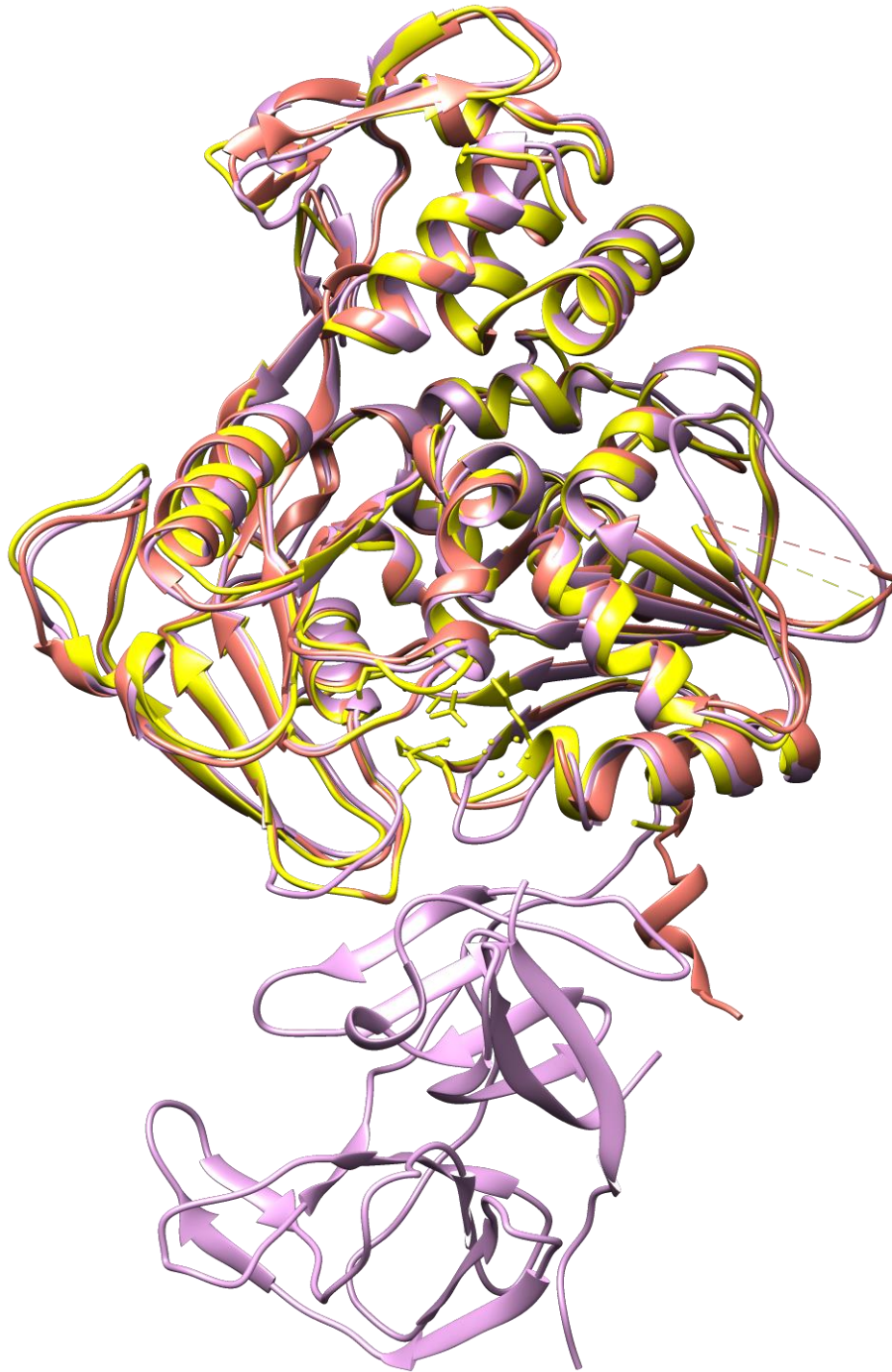


Figure 12. DENV 4 NS3 protein and DENV 2 helicases in superimposed position, border scale 0.0. Image was generated by

UCSF Chimera.

***Percent identity: 76.72% (2BHR vs 2WHX)***



*Percent identity: 76.50% (2BHR vs 2VBC)*

*Percent identity: 76.27% (2BHR vs 2WZQ)*

*Percent identity: 77.83% (2BMF vs 2WHX)*

*Percent identity: 77.38% (2BMF vs 2VBC)*

*Percent identity: 77.16% (2BMF vs 2WZQ)*

*Percent identity: 94.46% (2BMF vs 2BHR)*

*RMSD:  $\alpha$  Carbon pairs [182-438], 1.144 (2BHR vs 2VBC)*

*$\alpha$  Carbon pairs [182-438], 1.031(2BHR vs 2WHX)*

*$\alpha$  Carbon pairs [182-438], 1.163 (2BHR vs 2WZQ).*

*Across all 431 pairs: 1.345*

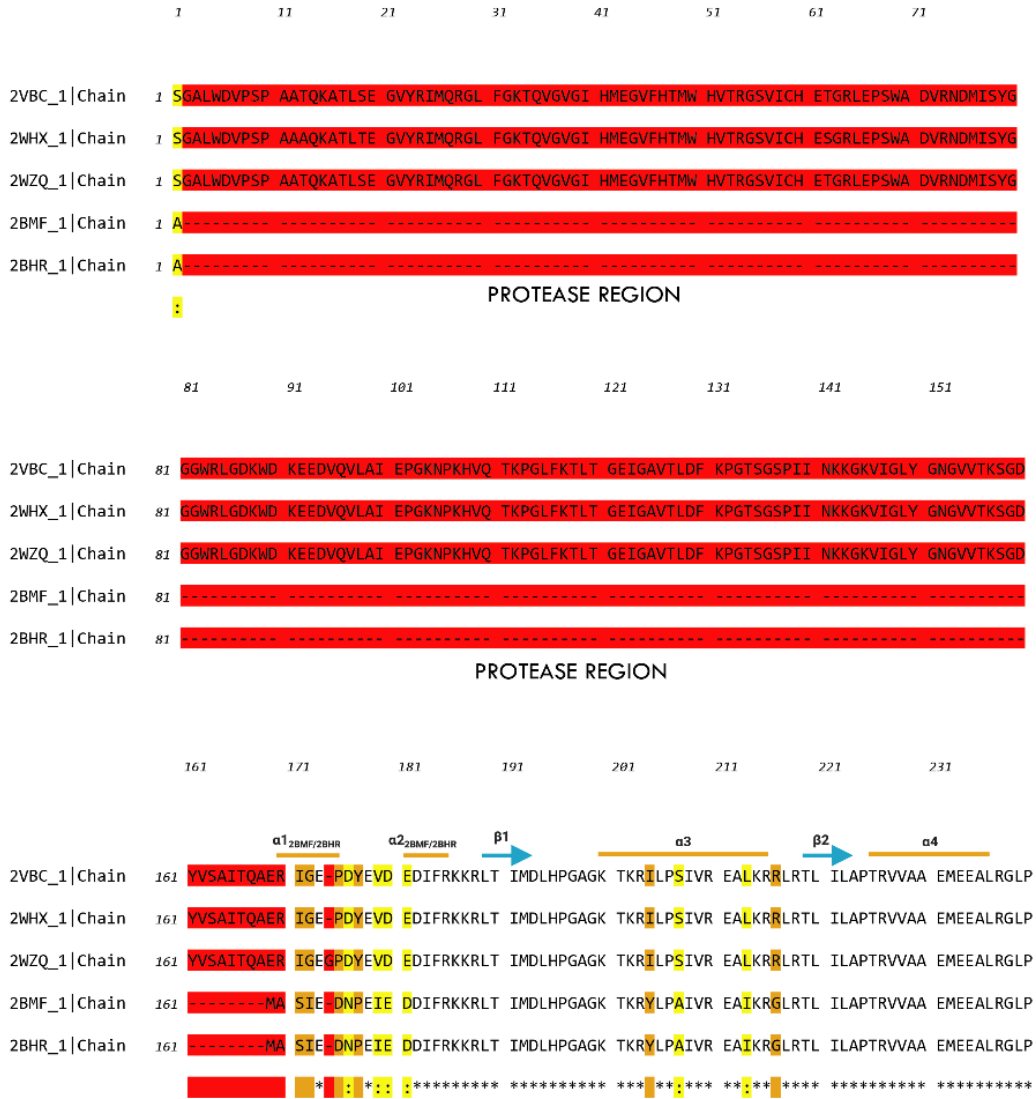
*RMSD:  $\alpha$  Carbon pairs [182-438], 1.208 (2BMF vs 2VBC)*

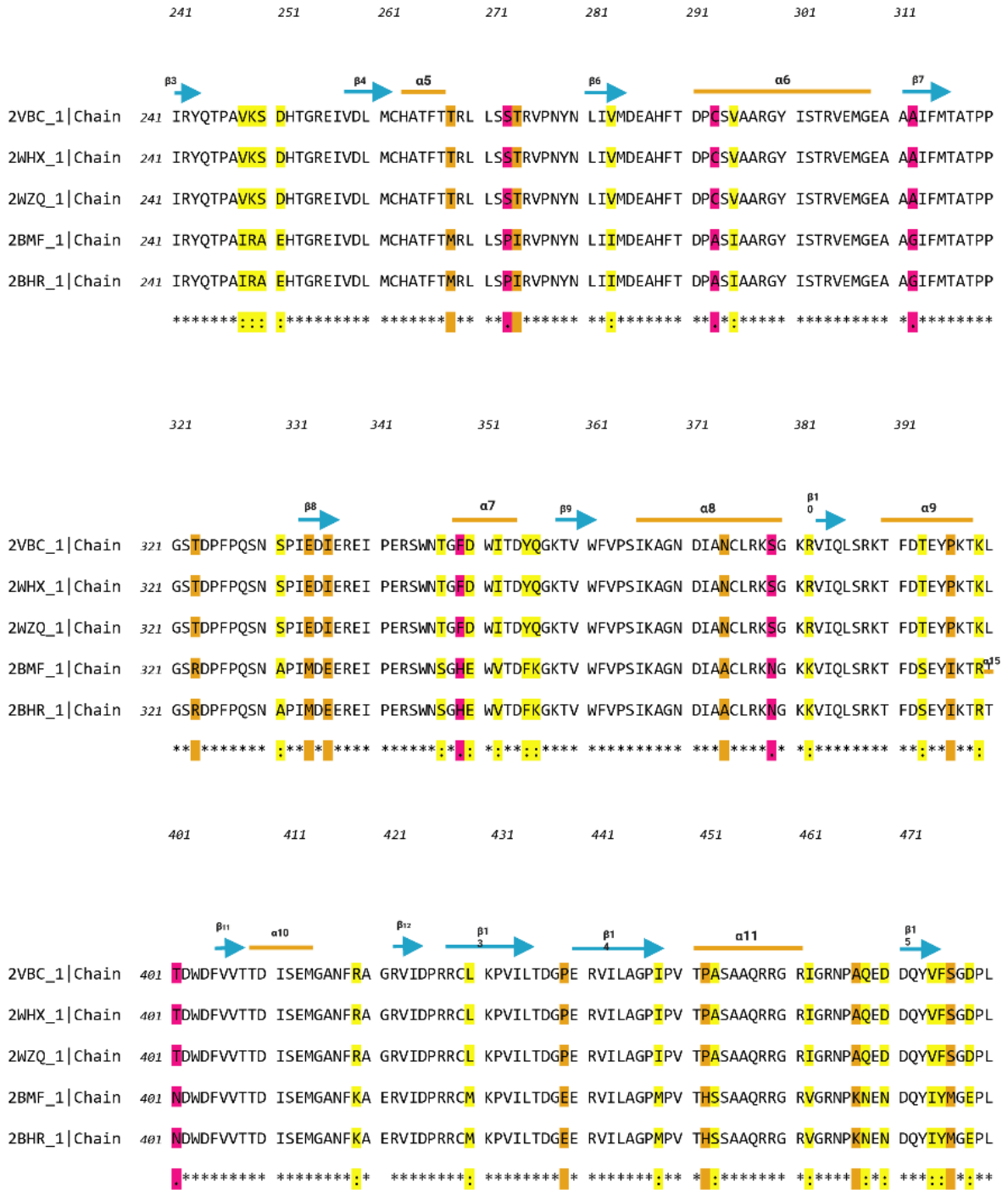
*$\alpha$  Carbon pairs [182-438], 1.357 (2BMF vs 2WHX)*

*$\alpha$  Carbon pairs [182-438], 1.193 (2BMF vs 2WZQ).*

*Across all 442 pairs: 2.989*

*RMSD:  $\alpha$  Carbon pairs [182-438], 1.285 (2BHR vs 2BMF)*







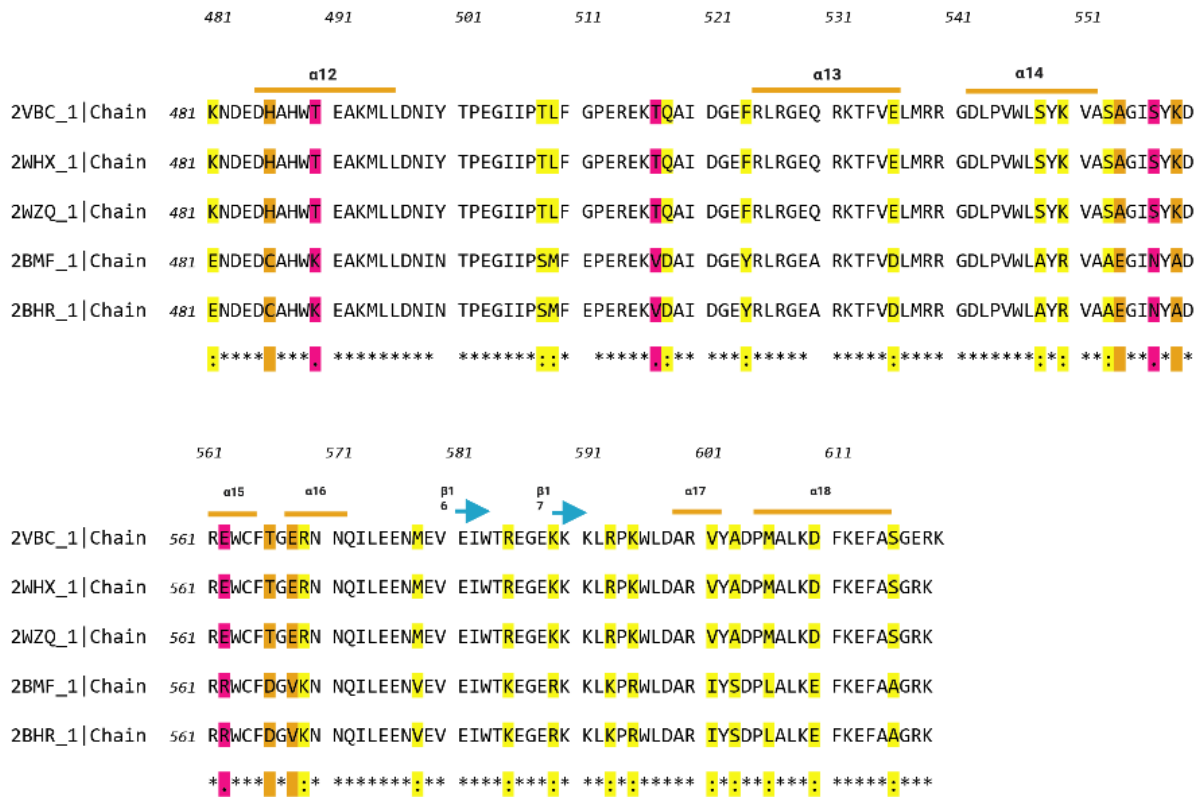


Figure 13. Sequence alignment of DENV 4 NS3 protein (Chain A) and DENV 2 helicases (Chain A). Alignment done by T-Coffee Expresso.

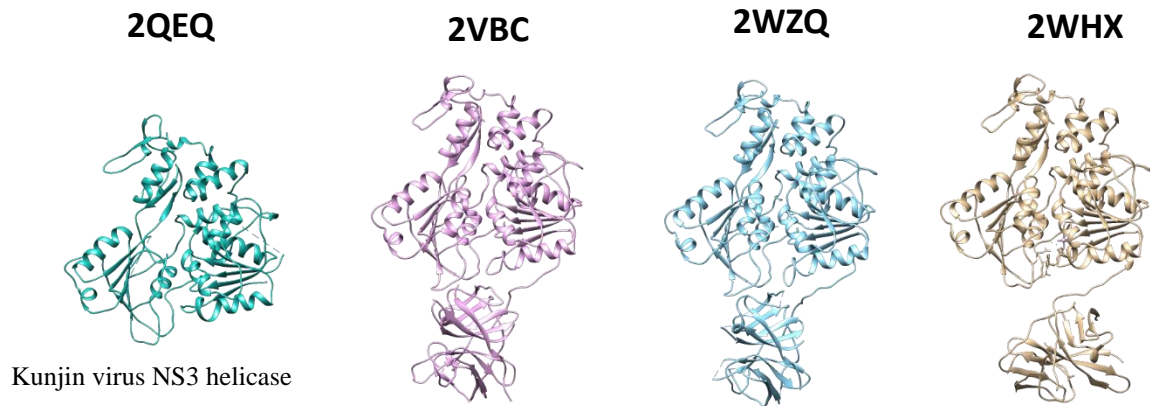


Figure 14. DENV 4 NS3 protein and Kunjin virus NS3 helicase in superimposed position, border scale 0.8. Image was generated by UCSF Chimera.

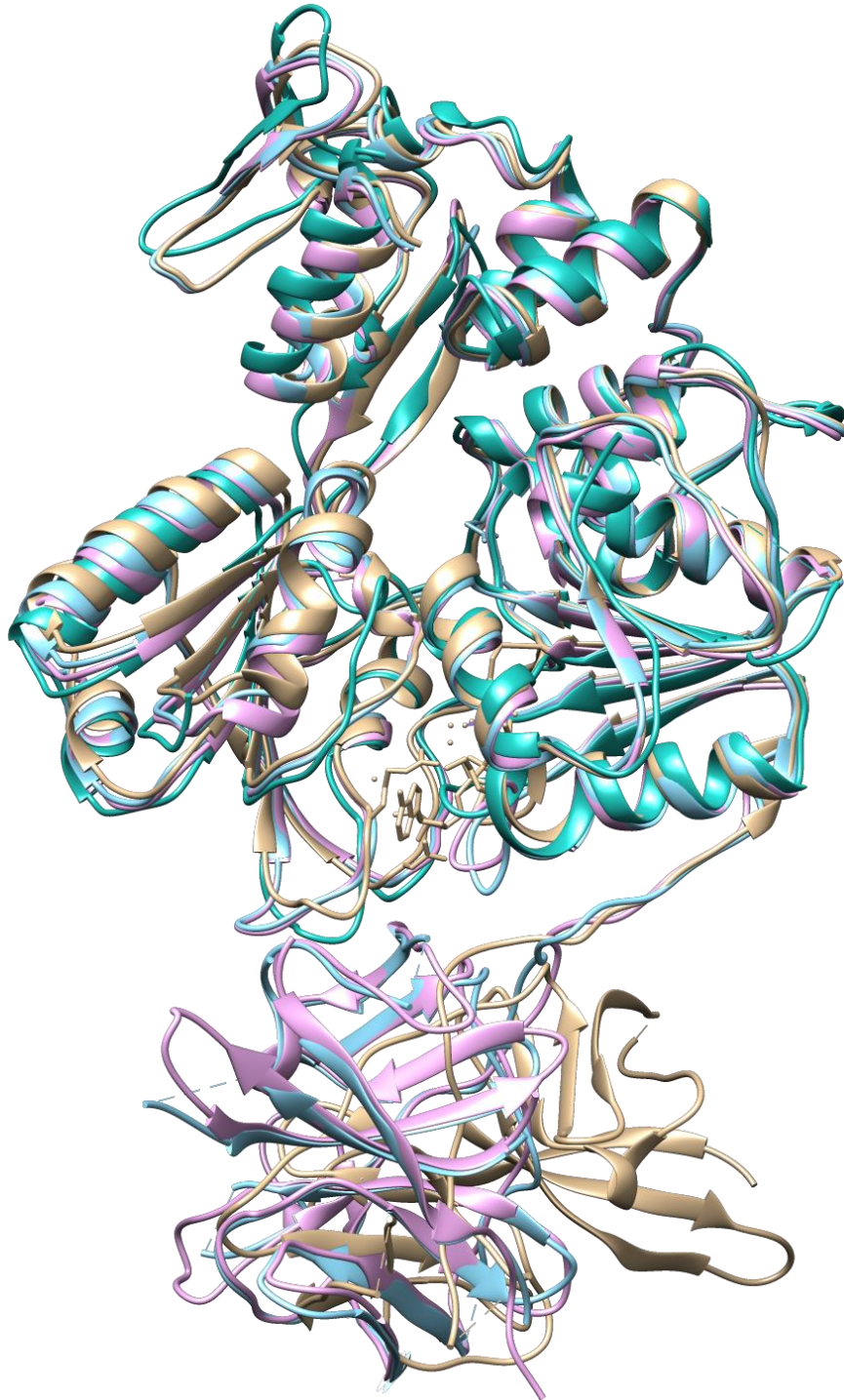


Figure 15. DENV 4 NS3 protein and Kunjin virus NS3 helicase in superimposed position, border scale 0.0. Image was generated by UCSF Chimera.

***Percent identity: 61.52% (2QEQ vs 2WHX)***

*Percent identity: 61.06% (2QEQ vs 2VBC)*

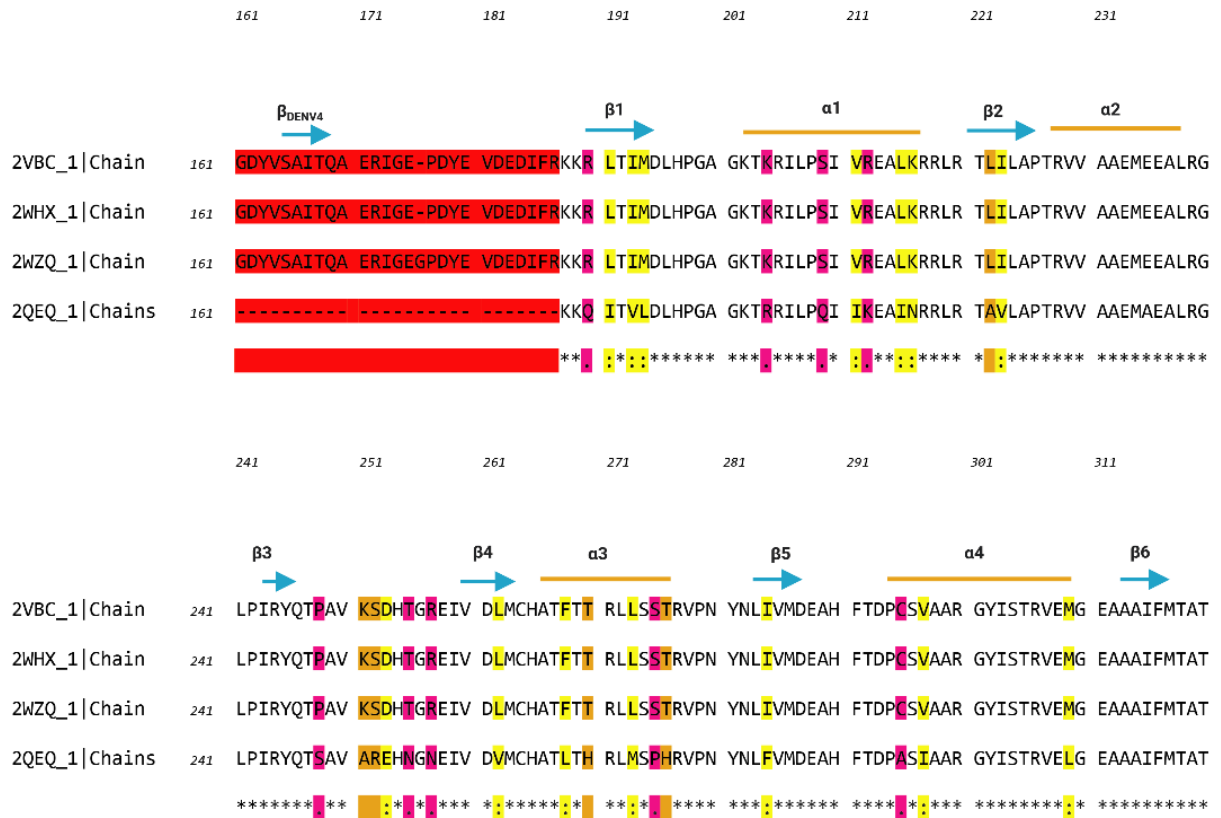
*Percent identity: 61.06% (2QEQ vs 2WZQ)*

*RMSD:  $\alpha$  Carbon pairs [376-839], 1.507 (2QEQ vs 2VBC)*

*$\alpha$  Carbon pairs [376-839], 1.915 (2QEQ vs 2WHX)*

*$\alpha$  Carbon pairs [376-839], 1.525 (2QEQ vs 2WZQ).*

*Across all 403 pairs: 1.343*







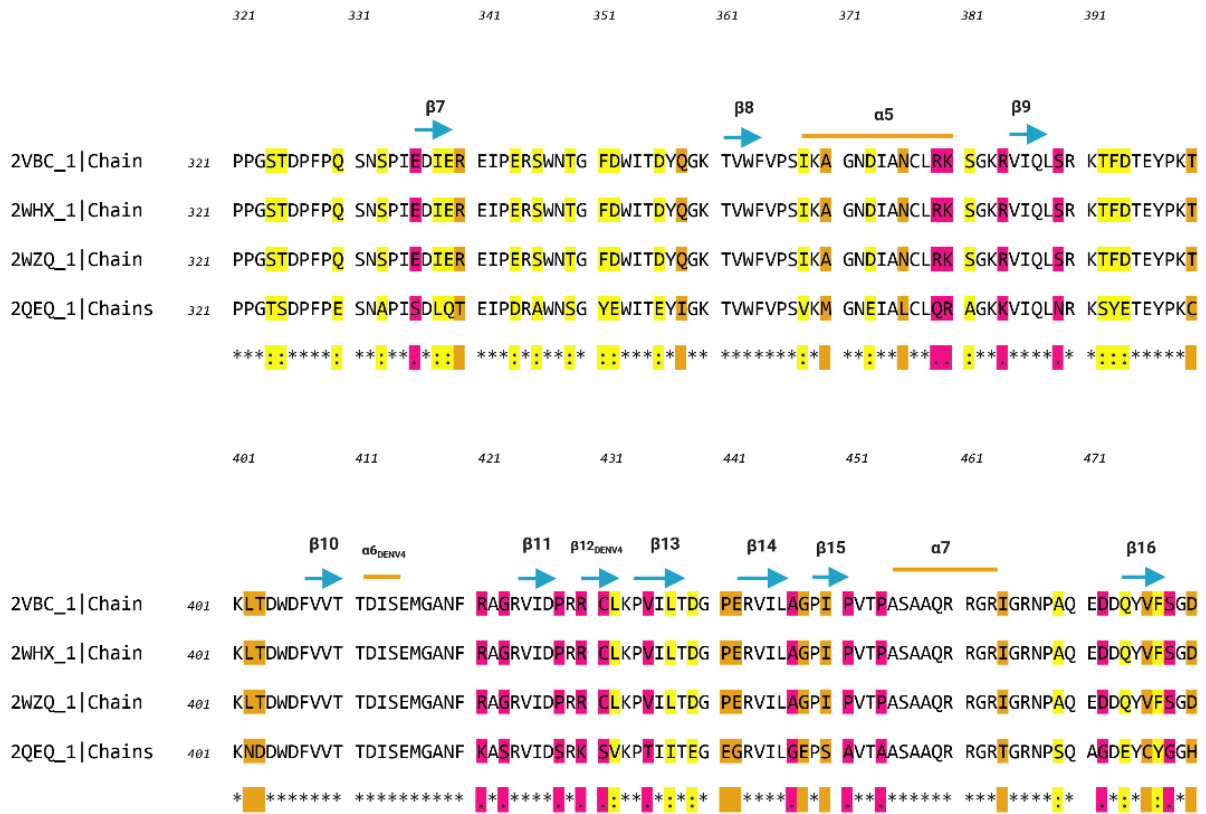


Figure 16. Sequence alignment of DENV 4 NS3 protein (Chain A) and Kunjin virus NS3 helicase (Chain A). Alignment done by T-Coffee Expresso.

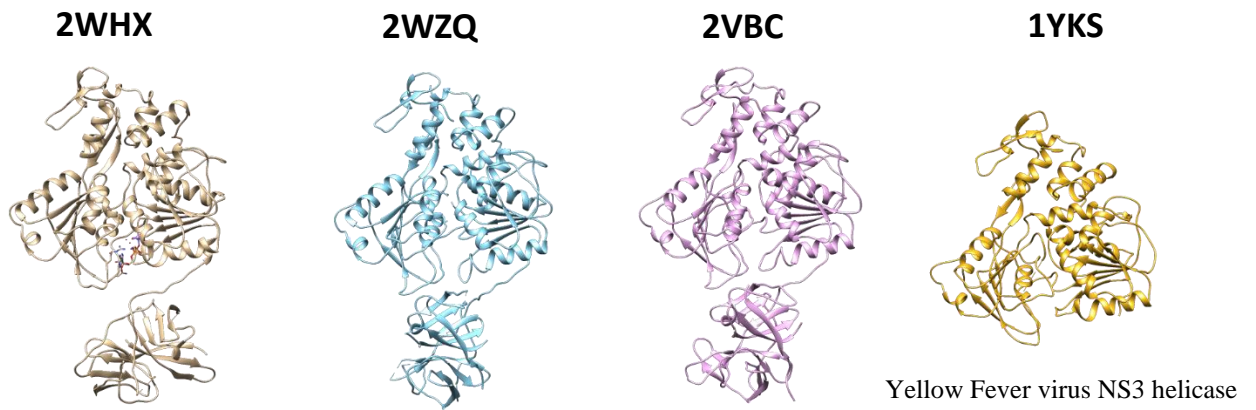


Figure 17. DENV 4 NS3 protein and Yellow Fever virus NS3 helicase in superimposed position, border scale 0.8. Image was generated by UCSF Chimera.

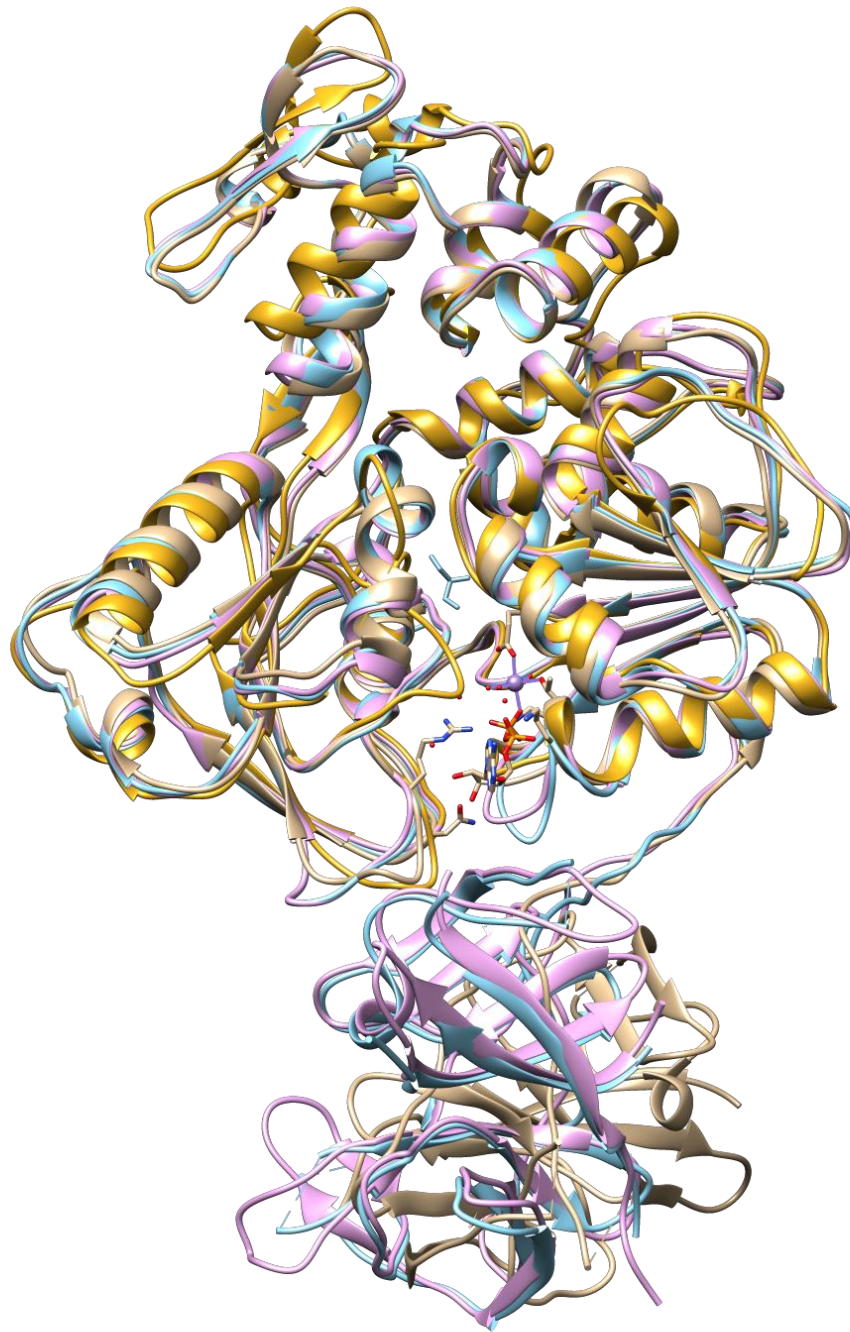


Figure 18. DENV 4 NS3 protein and Yellow Fever virus NS3 helicase in superimposed position, border scale 0.0. Image was generated by UCSF Chimera.

***Percent identity: 44.09% (1YKS vs 2WHX)***

*Percent identity: 43.86% (1YKS vs 2VBC)*

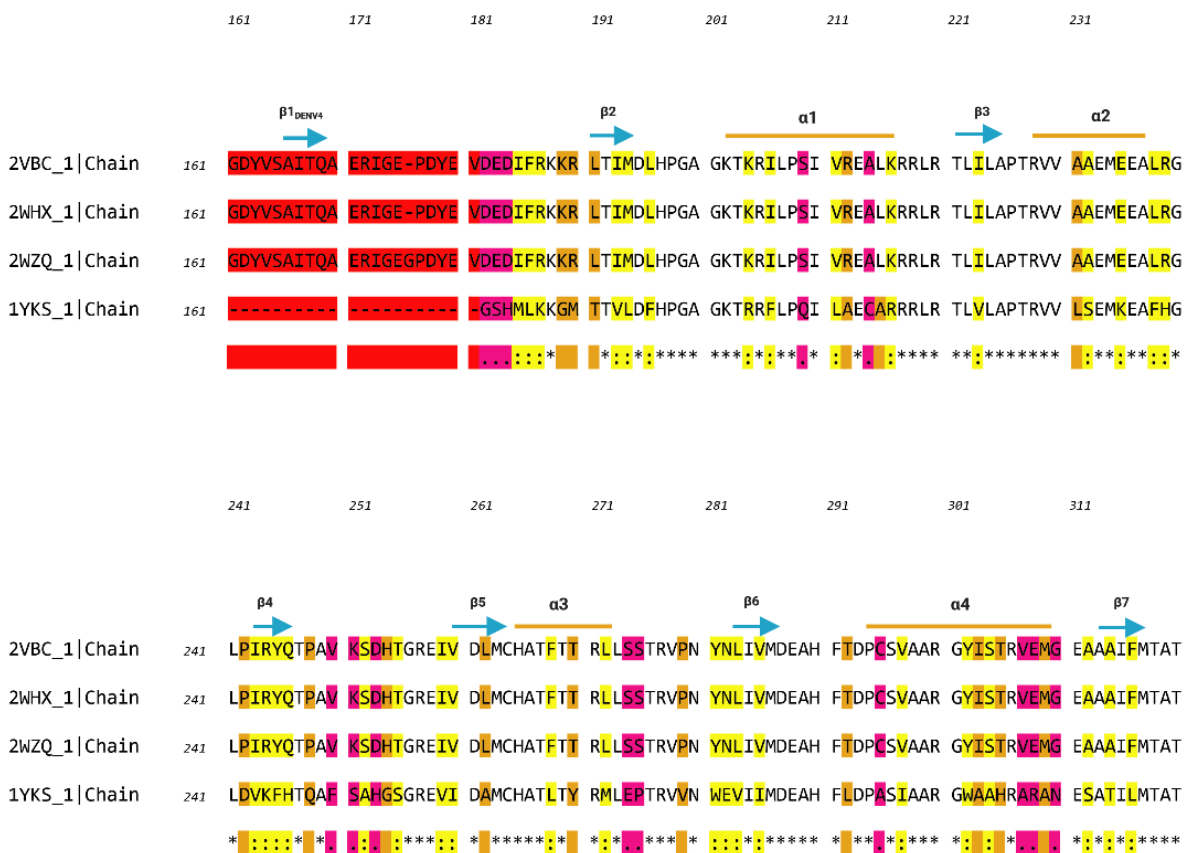
*Percent identity: 43.18% (1YKS vs 2WZQ)*

*RMSD:  $\alpha$  Carbon pairs [373-851], 1.695 (1YKS vs 2VBC)*

*$\alpha$  Carbon pairs [373-851], 1.341 (1YKS vs 2WHX)*

*$\alpha$  Carbon pairs [373-851], 1.722 (1YKS vs 2WZQ).*

*Across all 399 pairs: 1.361*





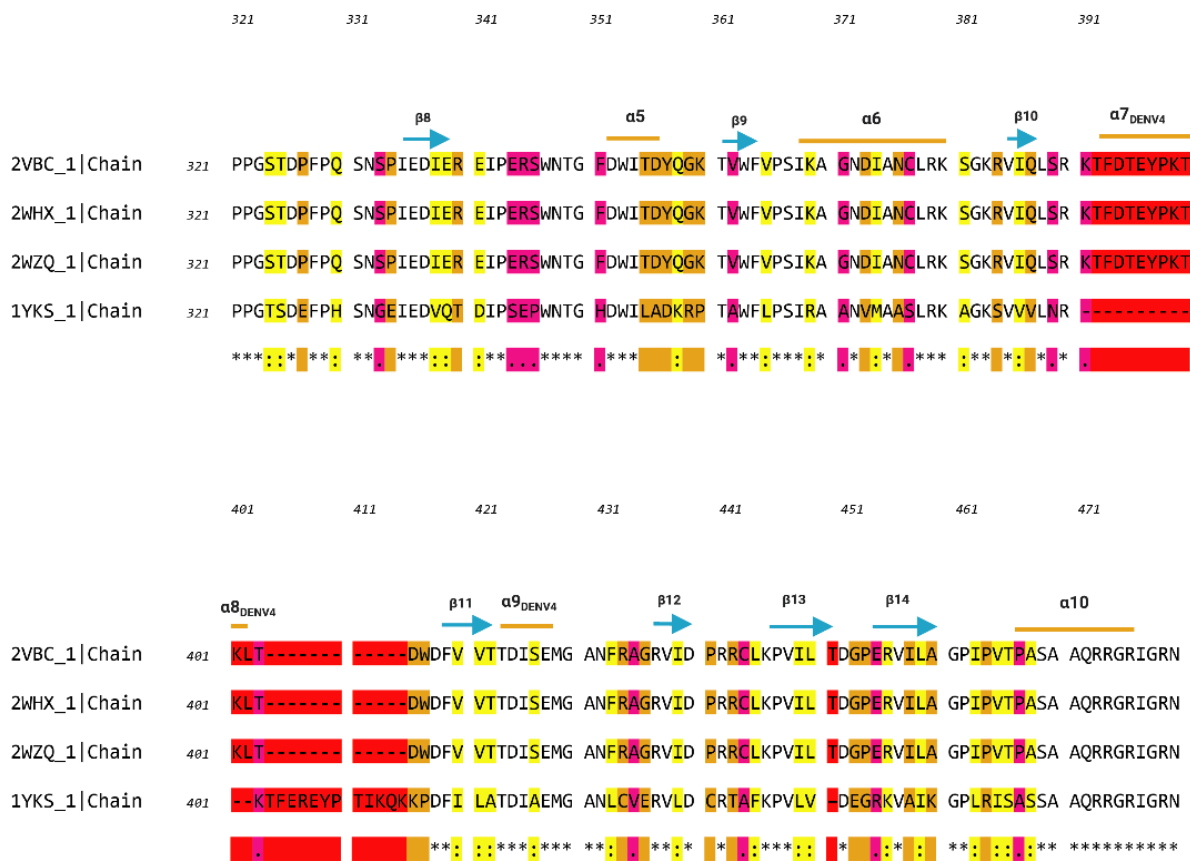


Figure 19. Sequence alignment of DENV 4 NS3 protein (Chain A) and Yellow Fever virus NS3 helicase (Chain A). Alignment done by T-Coffee Expresso.

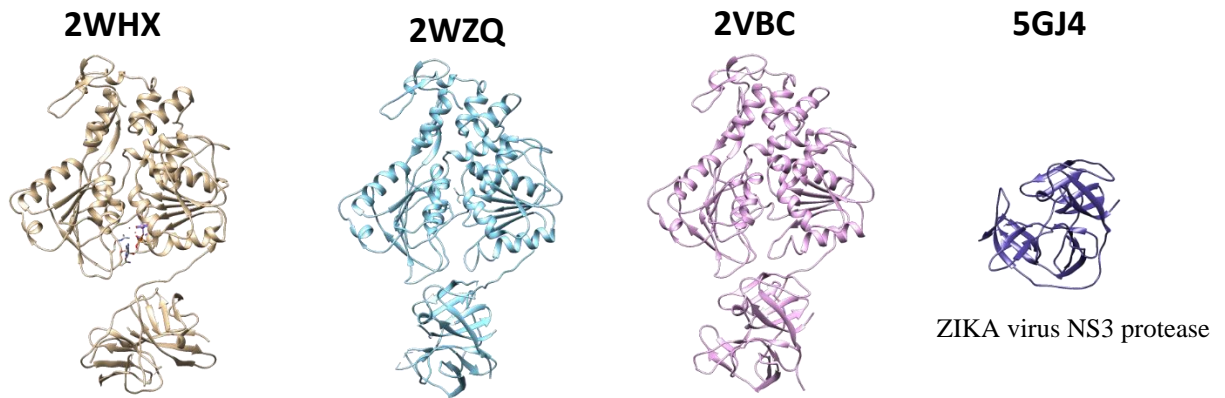


Figure 20. DENV 4 NS3 protein and ZIKA virus NS3 protease in superimposed position, border scale 0.8. Image was generated by UCSF Chimera.



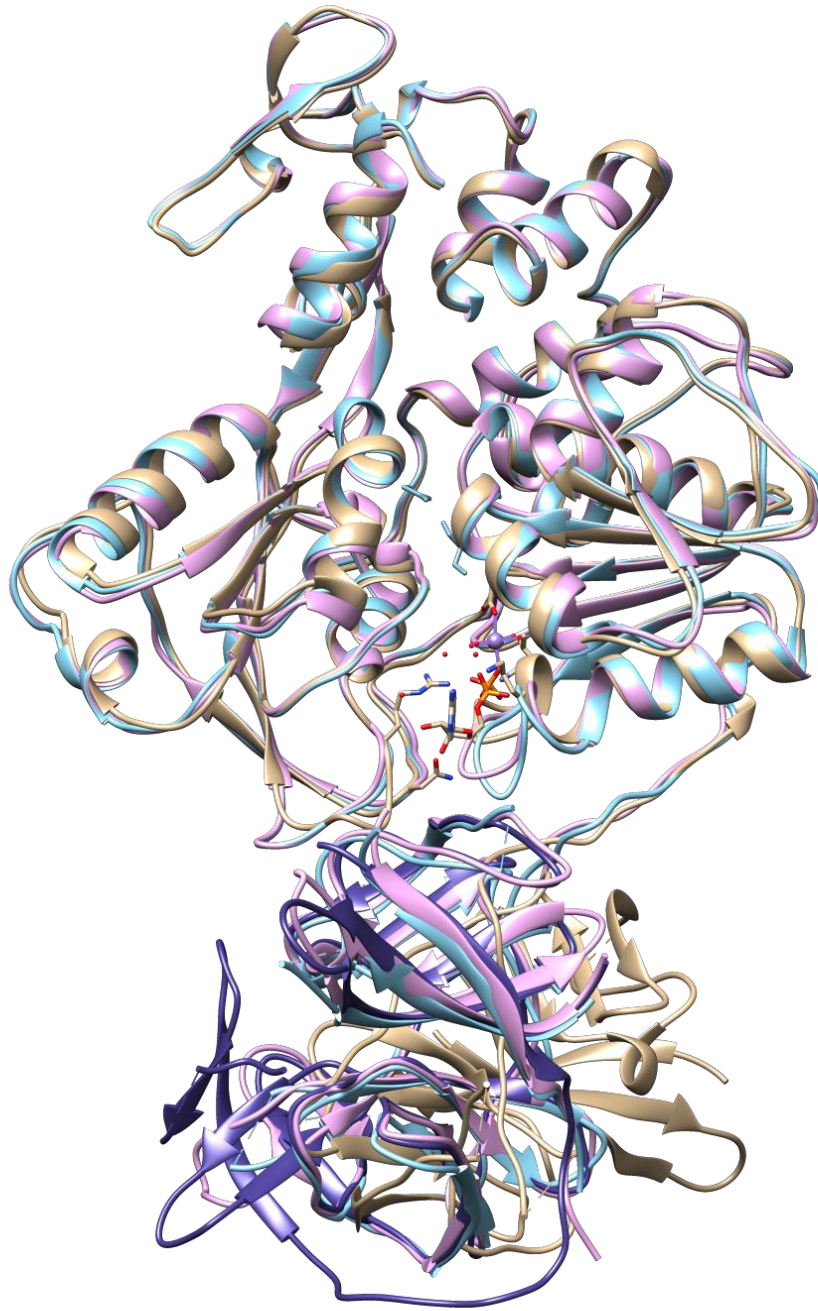


Figure 21. DENV 4 NS3 protein and ZIKA virus NS3 protease in superimposed position, border scale 0.0. Image was generated by UCSF Chimera.

***Percent identity: 45.20% (5GJ4 vs 2WHX)***

*Percent identity: 1.13% (5GJ4 vs 2VBC)*

*Percent identity: 0.00% (5GJ4 vs 2WZQ)*

*RMSD:  $\alpha$  Carbon pairs [208-865], 3.410 (5GJ4 vs 2VBC)*

*$\alpha$  Carbon pairs [208-865], 0.473 (5GJ4 vs 2WHX)*

*$\alpha$  Carbon pairs [208-865], 3.369 (5GJ4 vs 2WZQ).*

*Across all 13 pairs: 2.816*

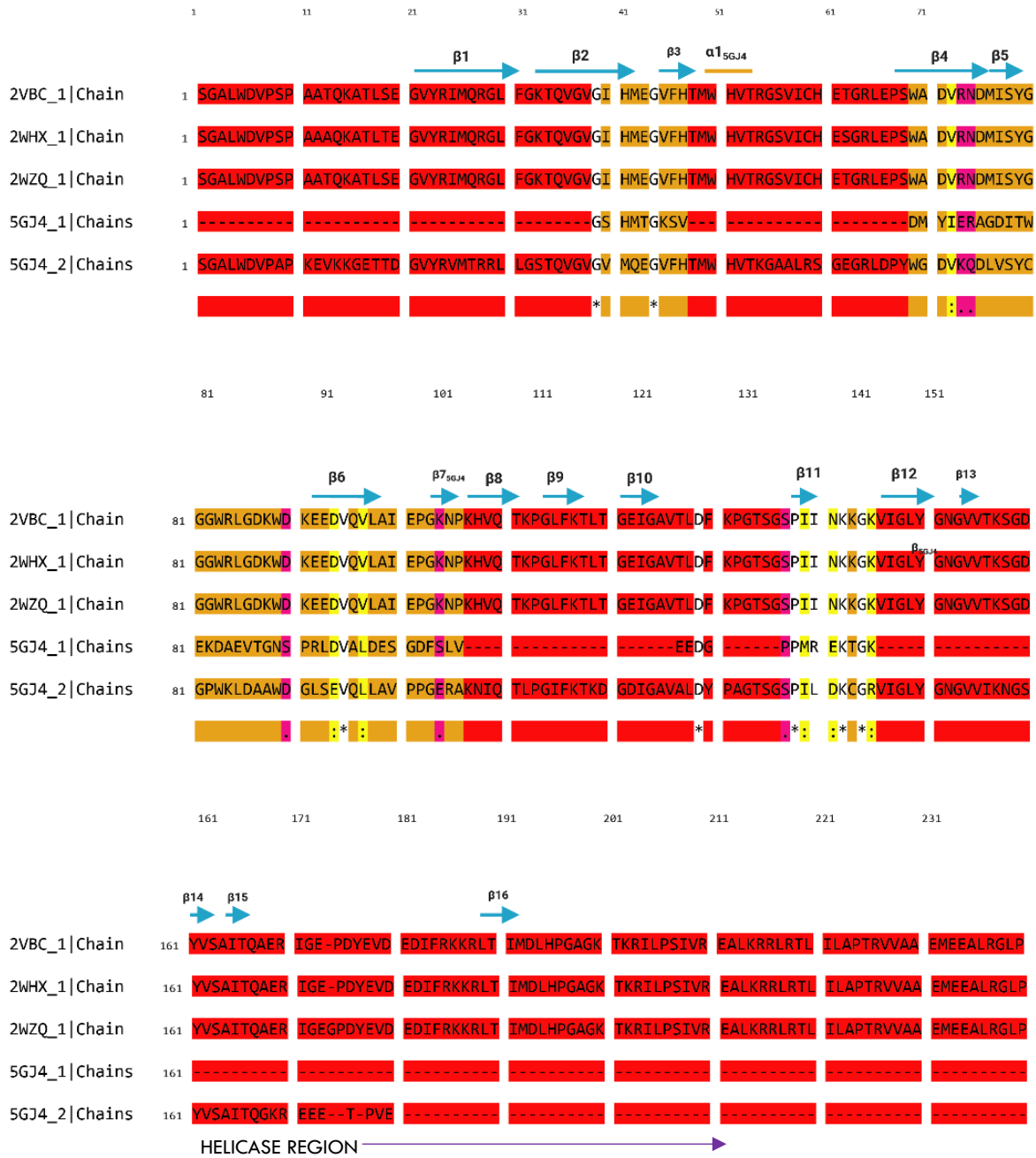


Figure 22. Sequence alignment of DENV 4 NS3 protein (Chain A) and ZIKA virus NS3 protease (Chain A and B). Alignment done by T-Coffee Expresso.

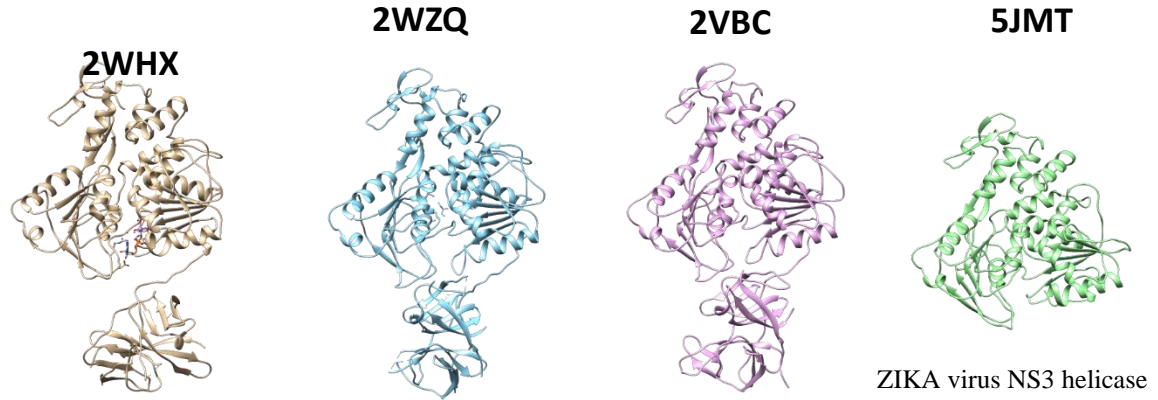


Figure 23. DENV 4 NS3 protein and ZIKA virus NS3 helicase in superimposed position, border scale 0.8. Image was generated by UCSF Chimera.

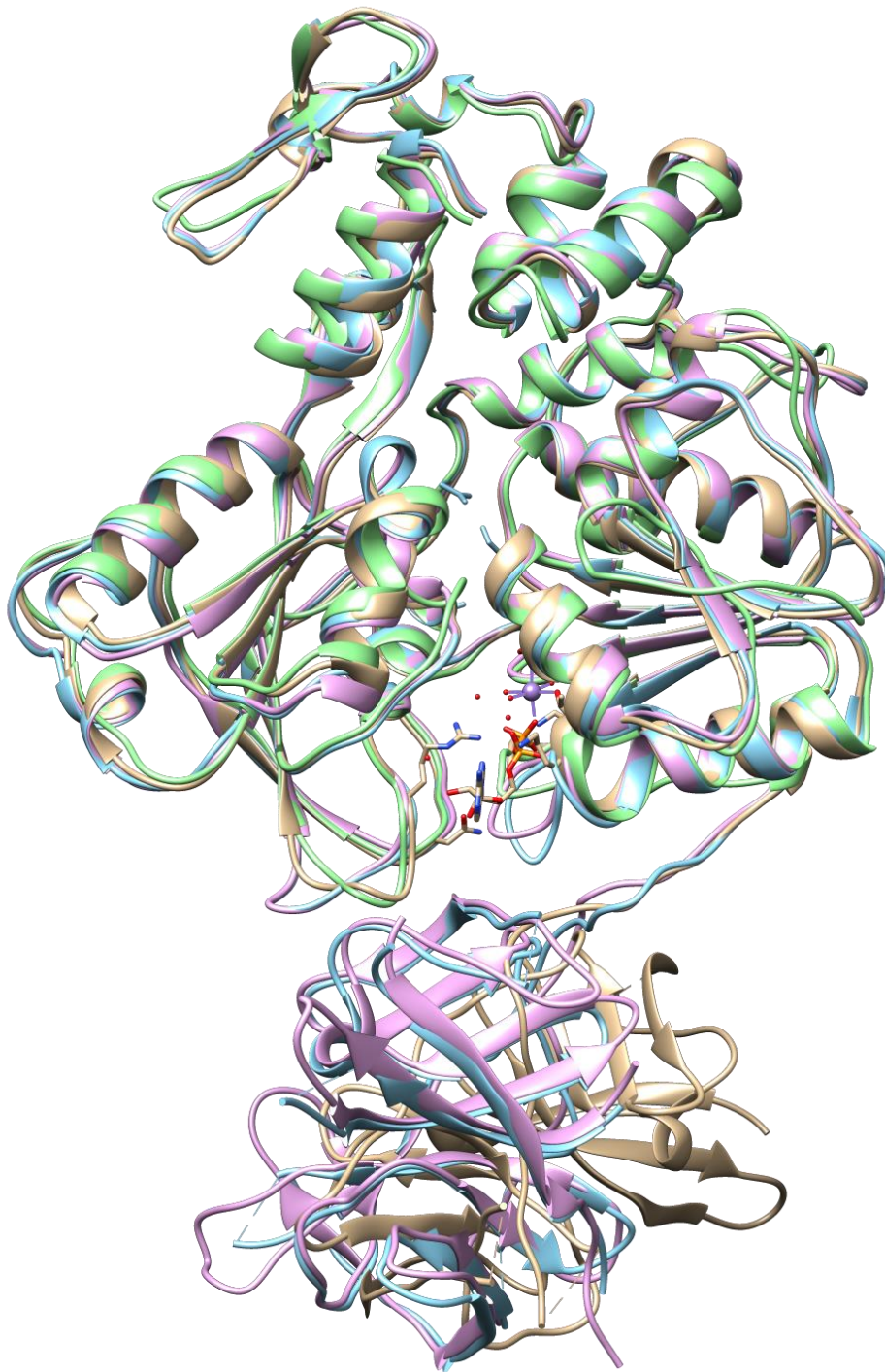


Figure 24. DENV 4 NS3 protein and ZIKA virus NS3 helicase in superimposed position, border scale 0.0. Image was generated by UCSF Chimera.

**Percent identity: 66.44% (5JMT vs 2WHX)**

**Percent identity: 66.67% (5JMT vs 2VBC)**

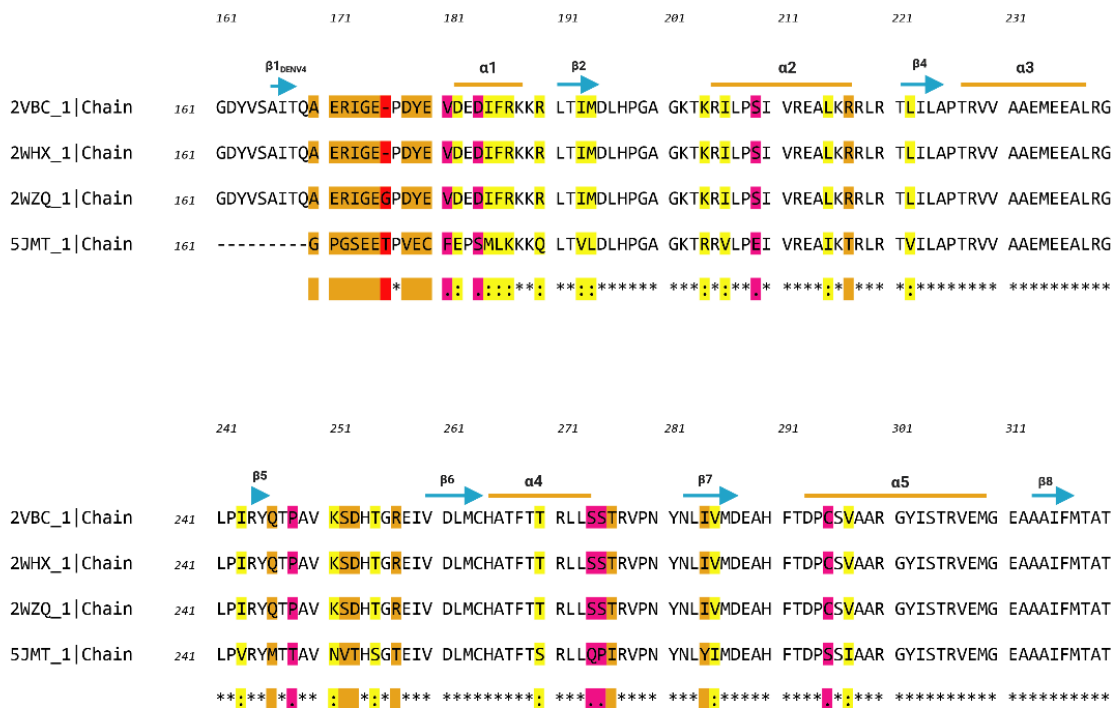
**Percent identity: 66.44% (5JMT vs 2WZQ)**

**RMSD:  $\alpha$  Carbon pairs [382-838], 1.167 (5JMT vs 2VBC)**

**$\alpha$  Carbon pairs [382-838], 1.193 (5JMT vs 2WHX)**

**$\alpha$  Carbon pairs [382-838], 1.220 (5JMT vs 2WZQ).**

**Across all 423 pairs: 0.985**



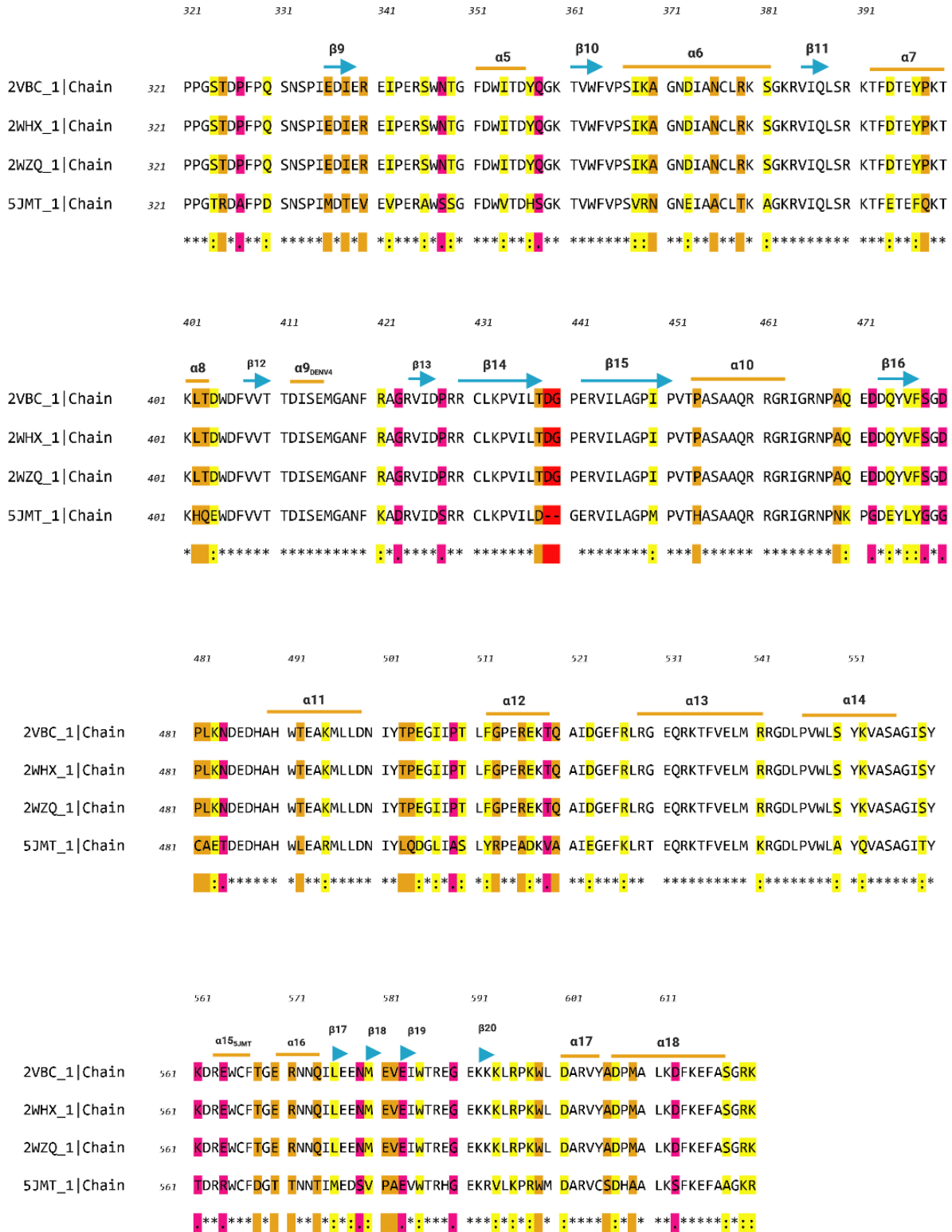


Figure 25. Sequence alignment of DENV 4 NS3 protein (Chain A) and ZIKA virus NS3 helicase (Chain A). Alignment done by T-Coffee Expresso.

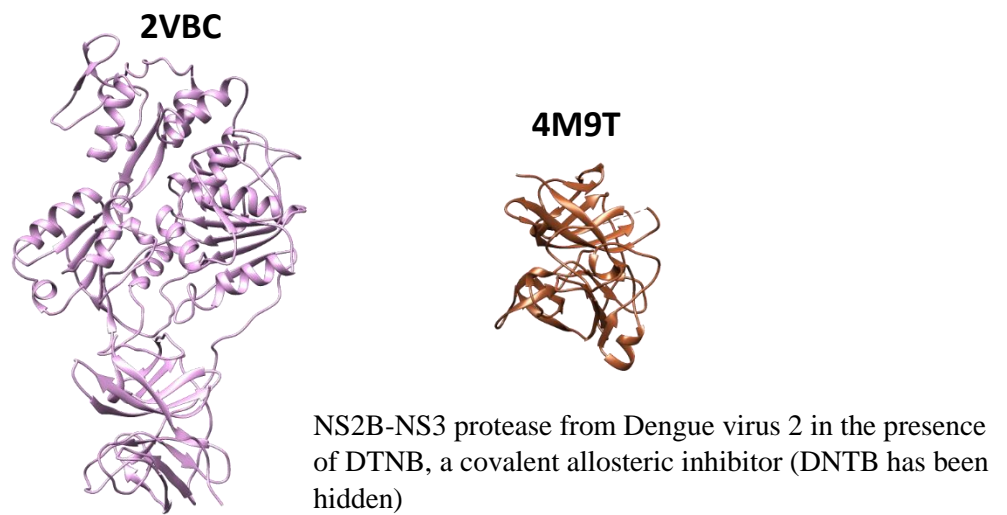


Figure 26. DENV 4 NS3 protein and DENV 2 NS3 protease in superimposed position, border scale 0.8. Image was generated by UCSF Chimera.



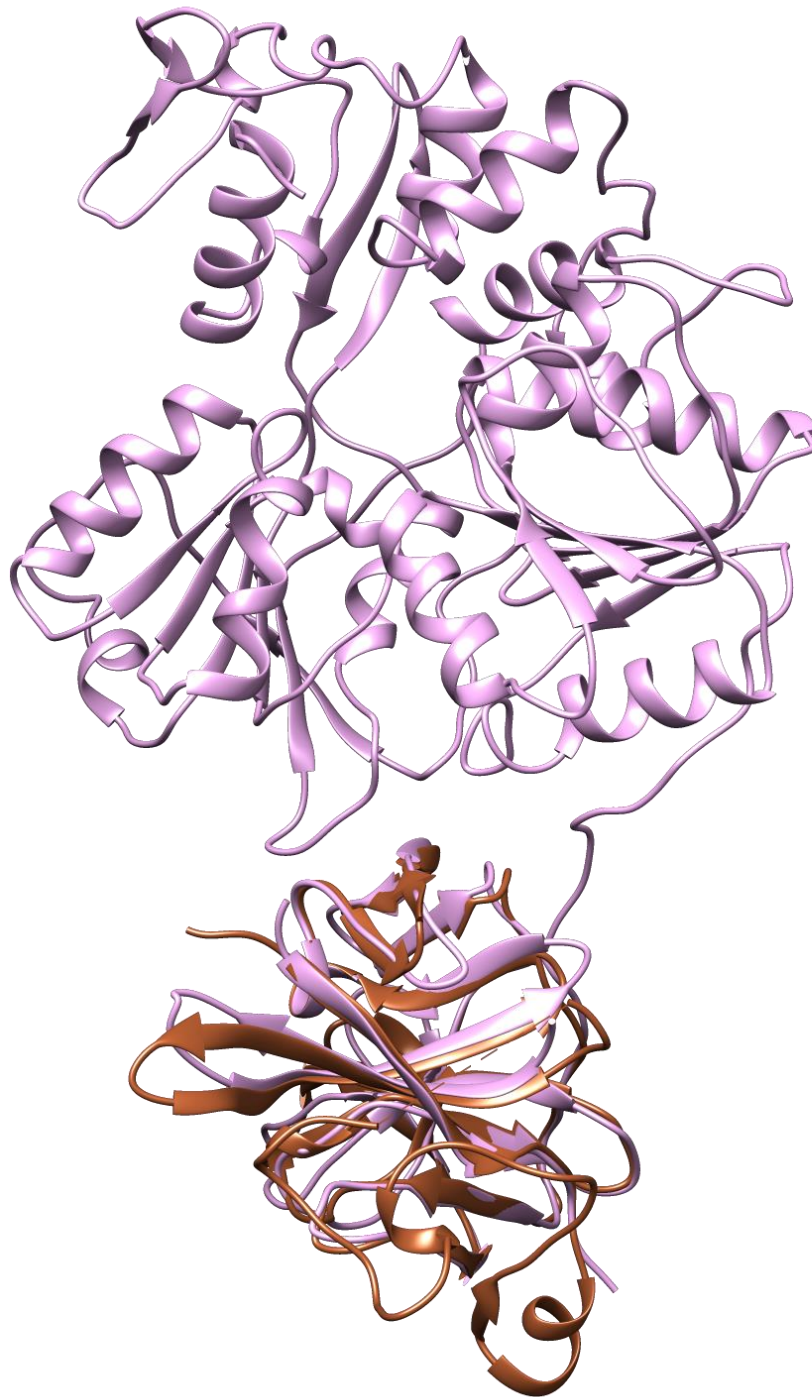


Figure 27. DENV 4 NS3 protein and DENV 2 NS3 protease in superimposed position, border scale 0.0. Image was generated by

UCSF Chimera.

***Percent identity: 2.02% (4M9T vs 2WHX)***

Percent identity: 38.06% (4M9T vs 2VBC)

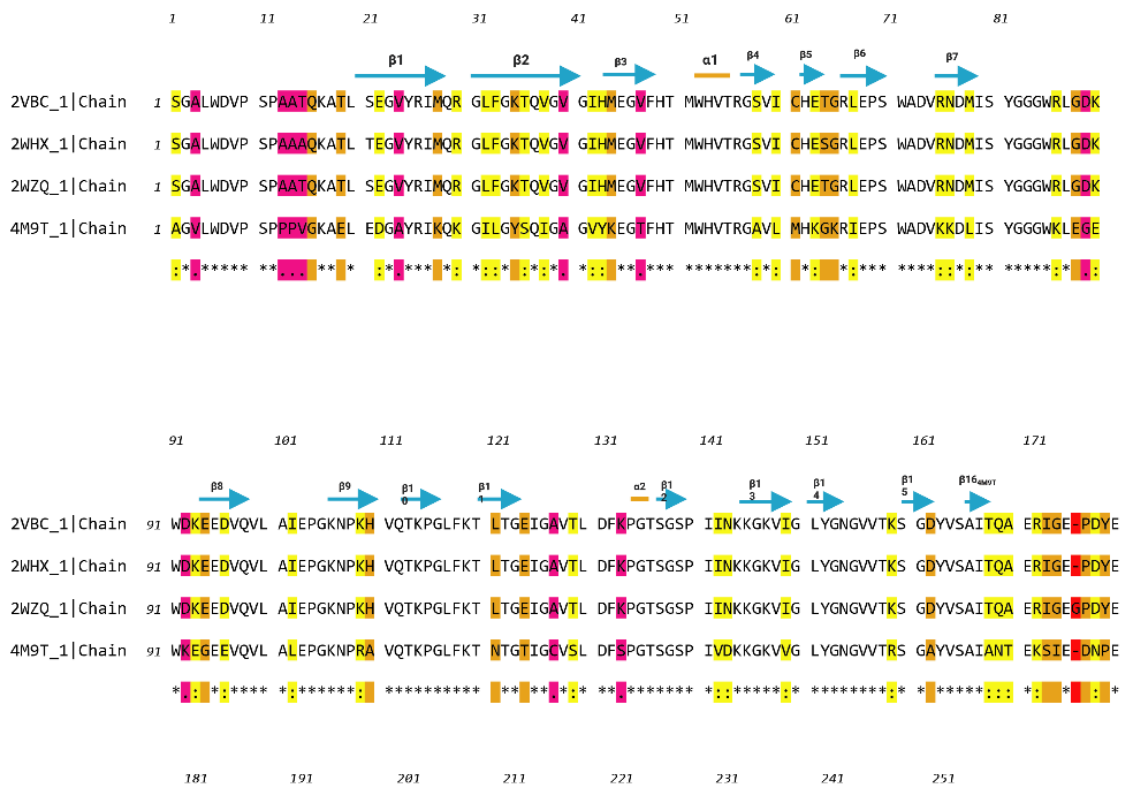
Percent identity: 37.25% (4M9T vs 2WZQ)

RMSD:  $\alpha$  Carbon pairs [217-915], 0.823 (4M9T vs 2VBC)

$\alpha$  Carbon pairs [217-915], 3.183 (4M9T vs 2WHX)

$\alpha$  Carbon pairs [217-915], 1.263 (4M9T vs 2WZQ).

Across all 21 pairs: 2.432



#### HELICASE REGION

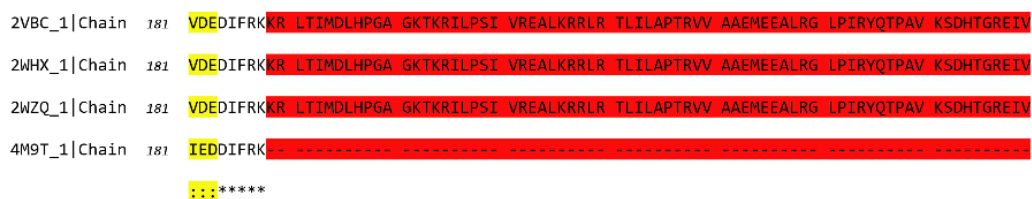


Figure 28. Sequence alignment of DENV 4 NS3 protein (Chain A) and DENV 2 NS3 protease (Chain A). Alignment done by T-Coffee Expresso.

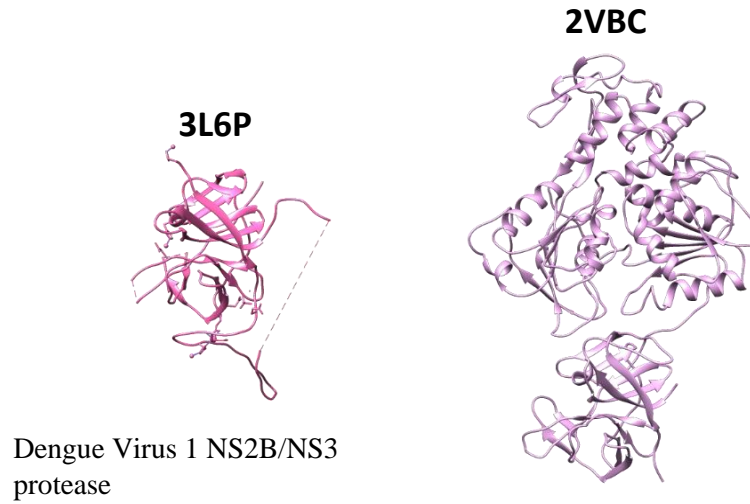


Figure 29. DENV 4 NS3 protein and DENV 1 NS3 protease in superimposed position, border scale 0.8. Image was generated by UCSF Chimera.

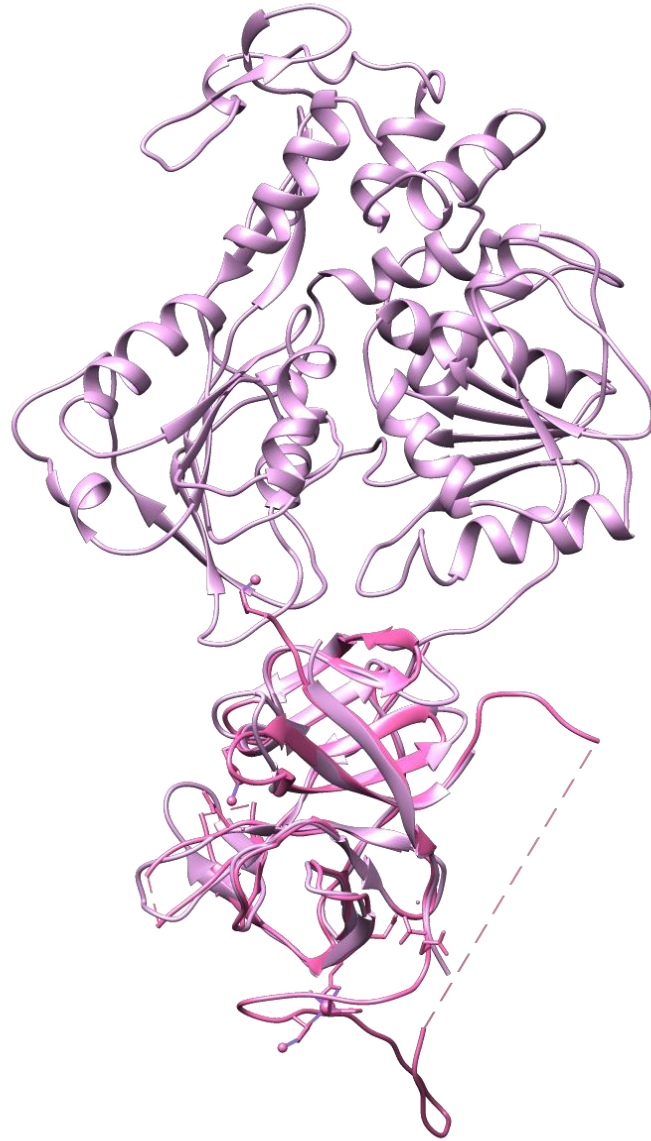


Figure 30. DENV 4 NS3 protein and DENV 1 NS3 protease in superimposed position, border scale 0.0. Image was generated by

UCSF Chimera.

***Percent identity: 1.27% (3L6P vs 2WHX)***



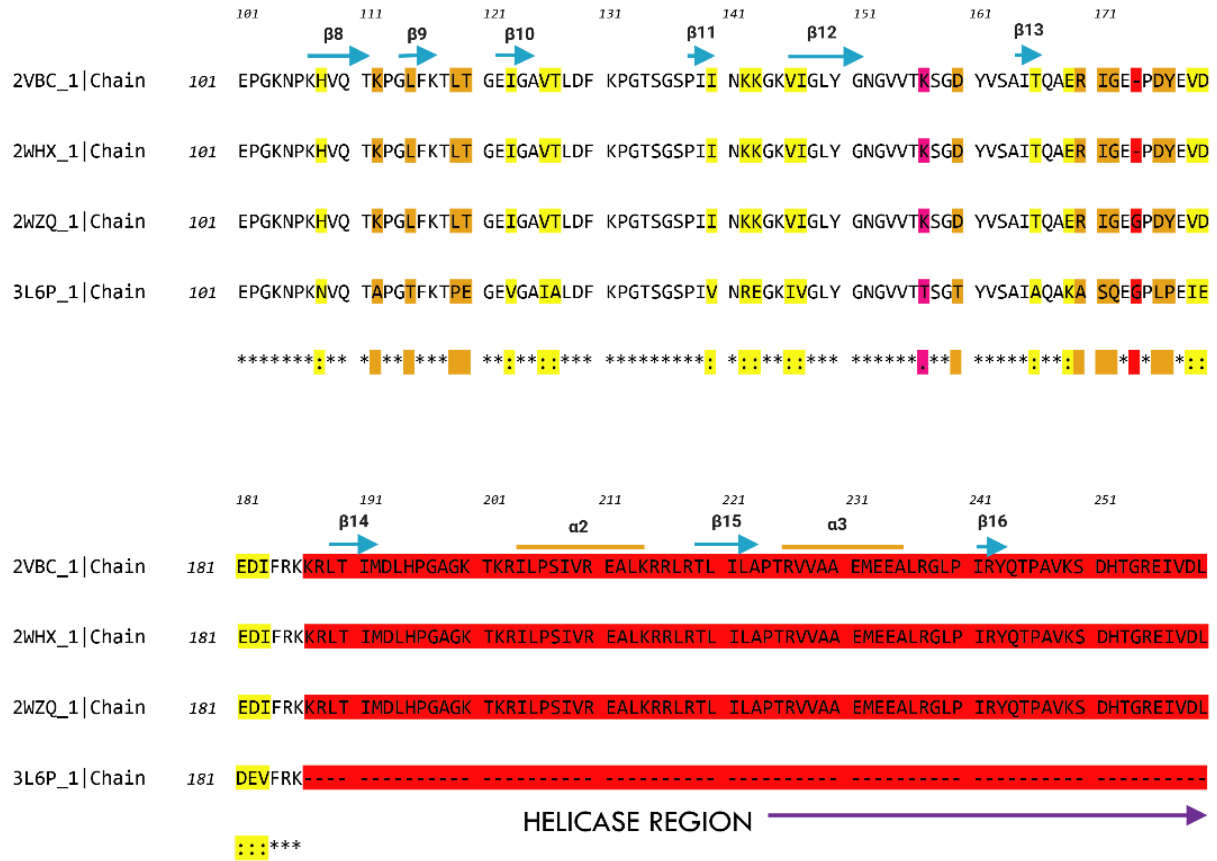


Figure 31. Sequence alignment of DENV 4 NS3 protein (Chain A) and DENV 1 NS3 protease (Chain A). Alignment done by T-Coffee Expresso.

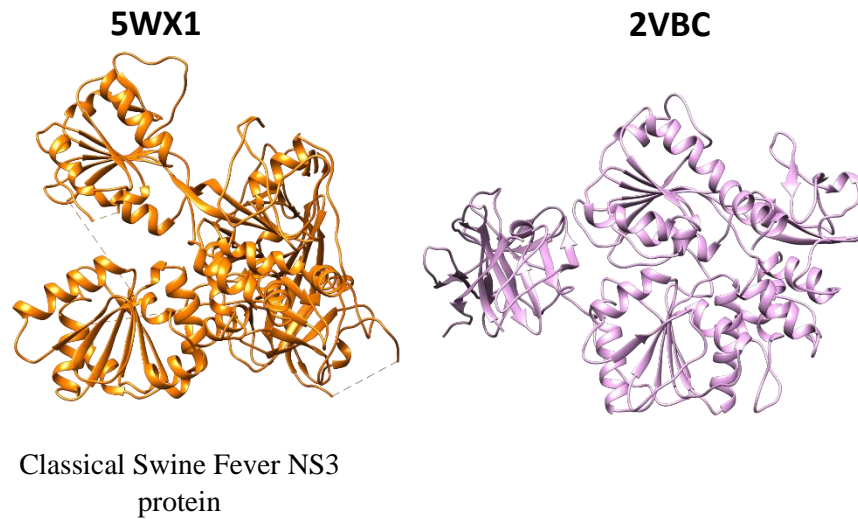


Figure 32. DENV 4 NS3 protein and CSF NS3 protein in superimposed position, border scale 0.8. Image was generated by UCSF Chimera.

***Percent identity: 0.32% (5WX1 vs 2WHX)***

***Percent identity: 0.32% (5WX1 vs 2VBC)***

***Percent identity: 0.48% (5WX1 vs 2WZQ)***

***RMSD:  $\alpha$  Carbon pairs [382-1504], 4.630 (5WX1 vs 2VBC)***

***$\alpha$  Carbon pairs [382-1504], 0.972 (5WX1 vs 2WHX)***

***$\alpha$  Carbon pairs [382-1504], 3.125 (5WX1 vs 2WZQ).***

***Across all 5 pairs: 3.252***

***Overall percent identity: 5.83%***

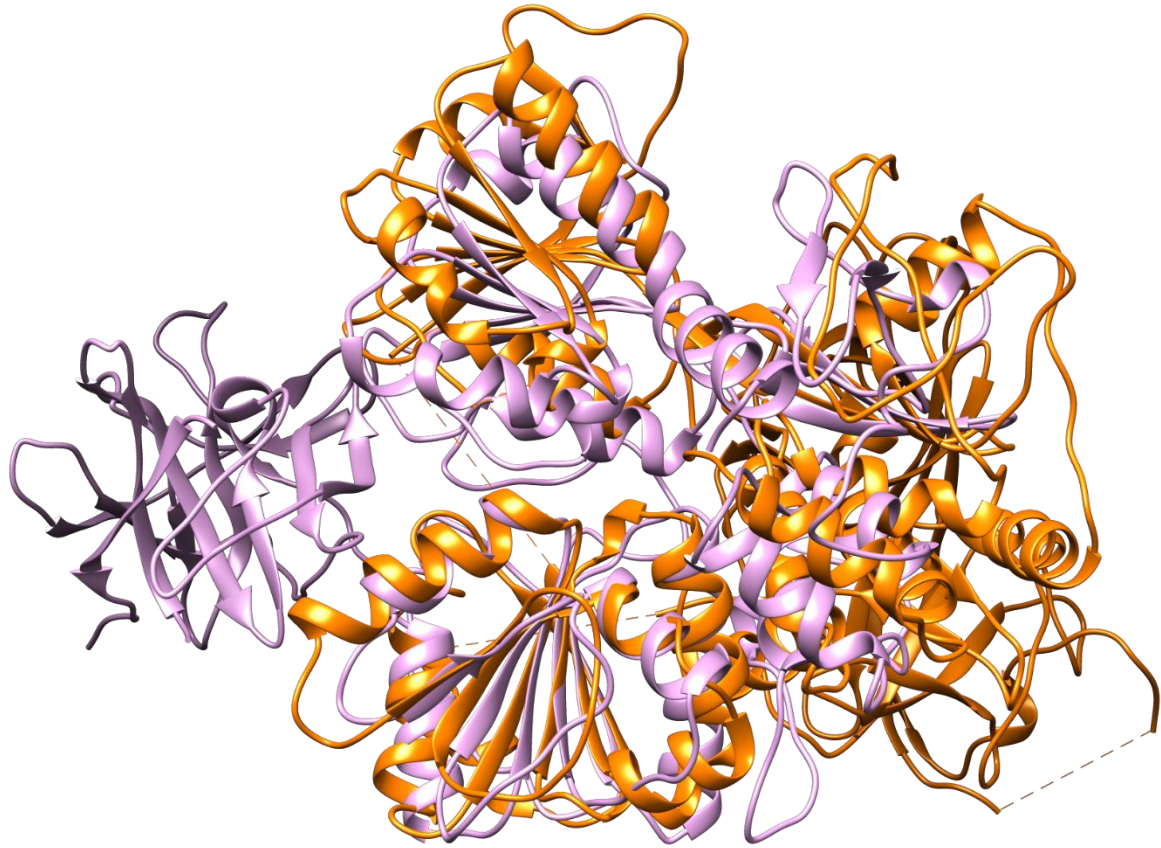


Figure 33. DENV 4 NS3 protein and CSF NS3 protein in superimposed position, border scale 0.0, horizontal orientation. Image

was generated by UCSF Chimera.



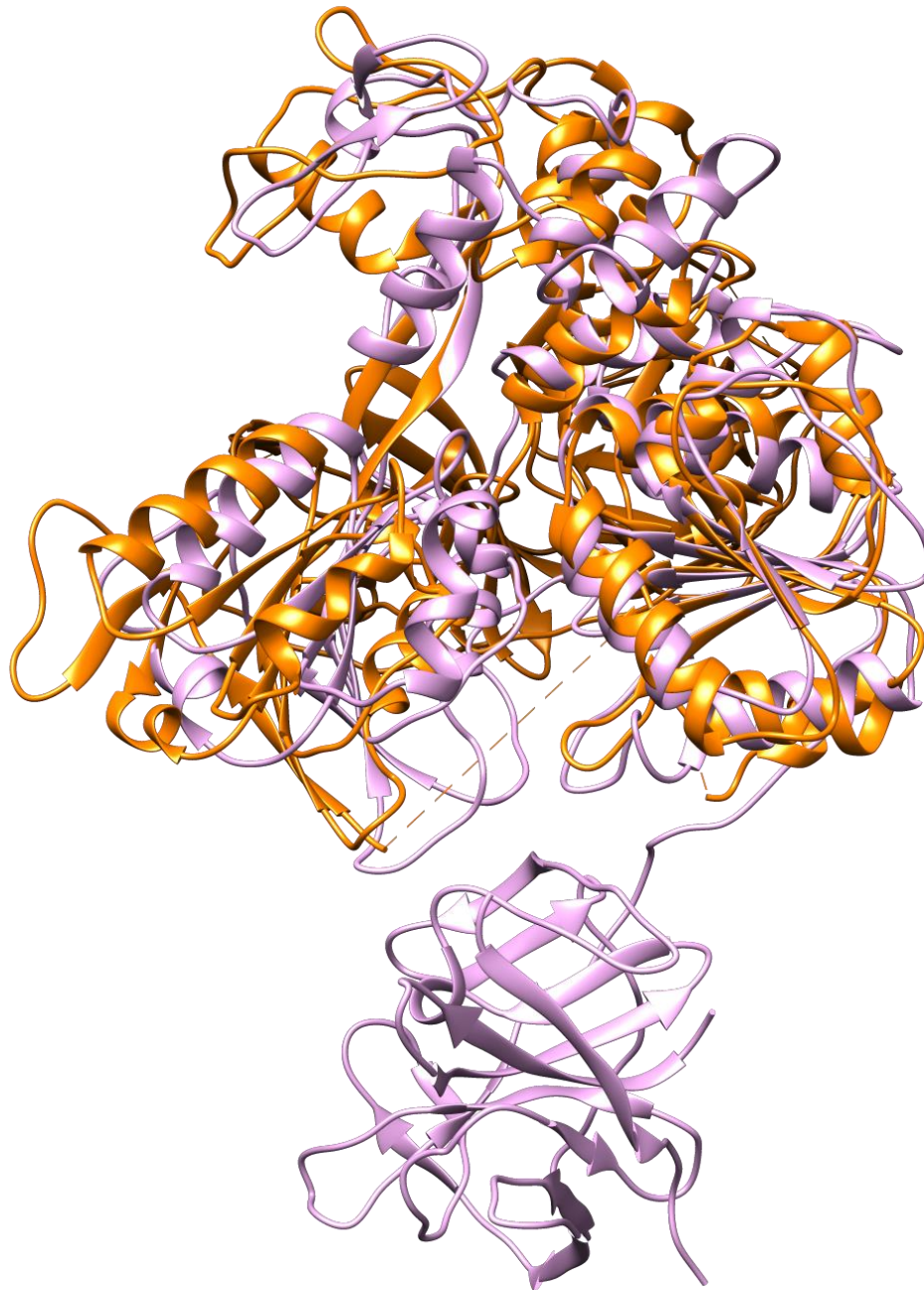


Figure 34. DENV 4 NS3 protein and CSF NS3 protein in superimposed position, border scale 0.0, vertical orientation. Image was generated by UCSF Chimera.



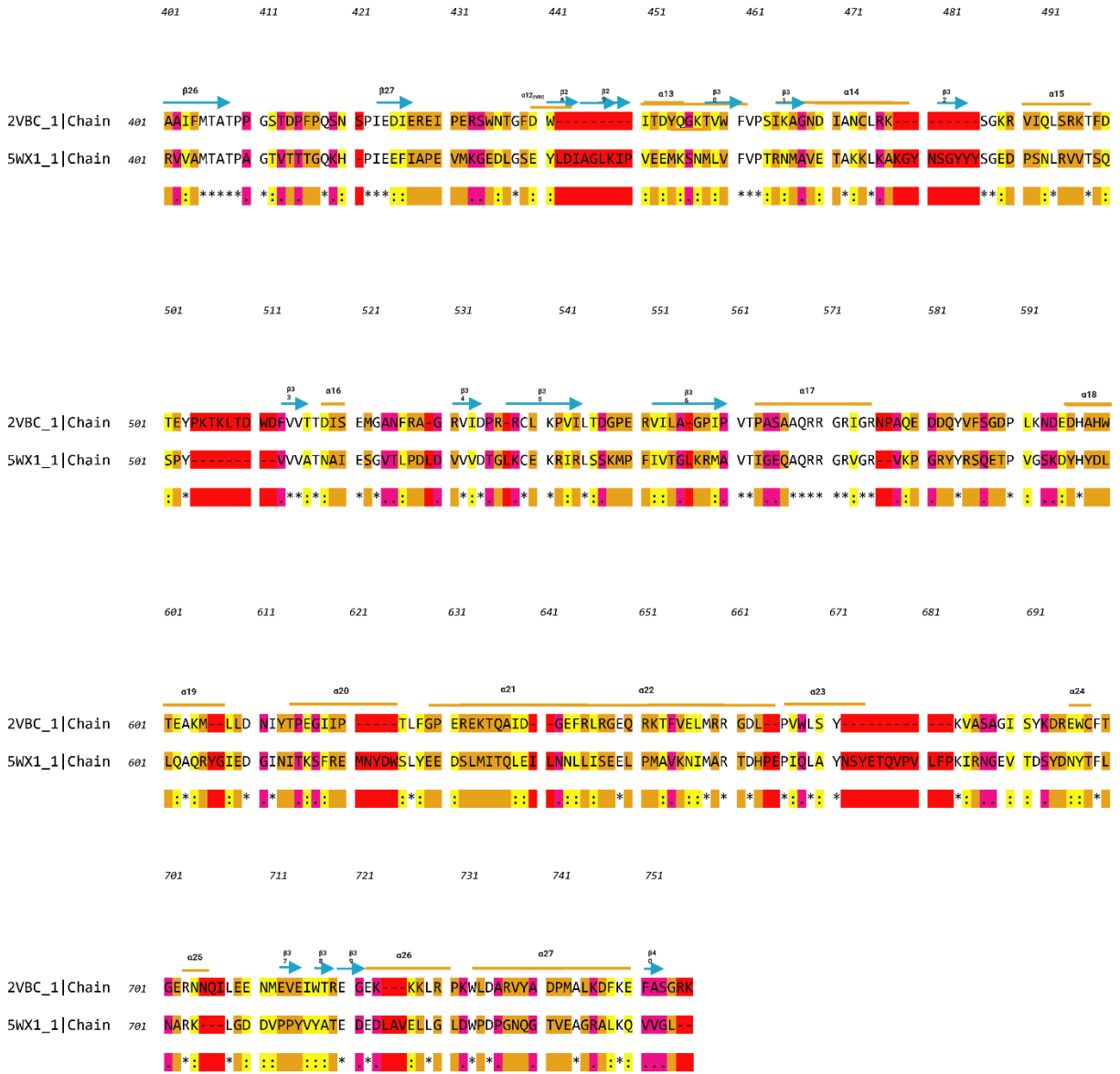


Figure 35. Sequence alignment of DENV 4 NS3 protein (Chain A) and CSF NS3 protein (Chain A). Alignment done by T-Coffee Expresso.

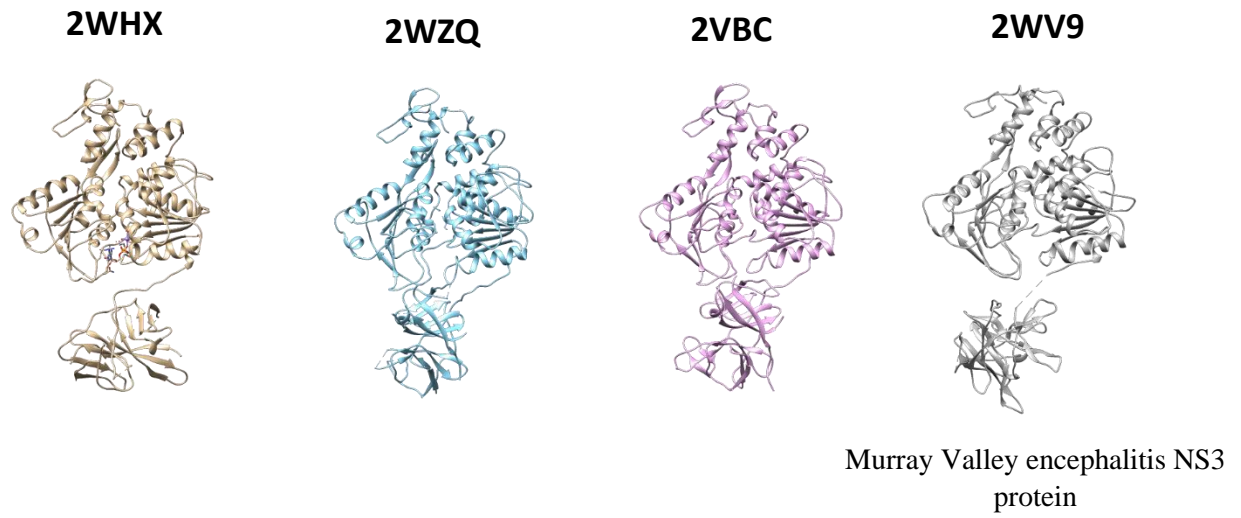


Figure 36. DENV 4 NS3 protein and MVEV NS3 protein in superimposed position, border scale 0.8. Image was generated by UCSF Chimera.

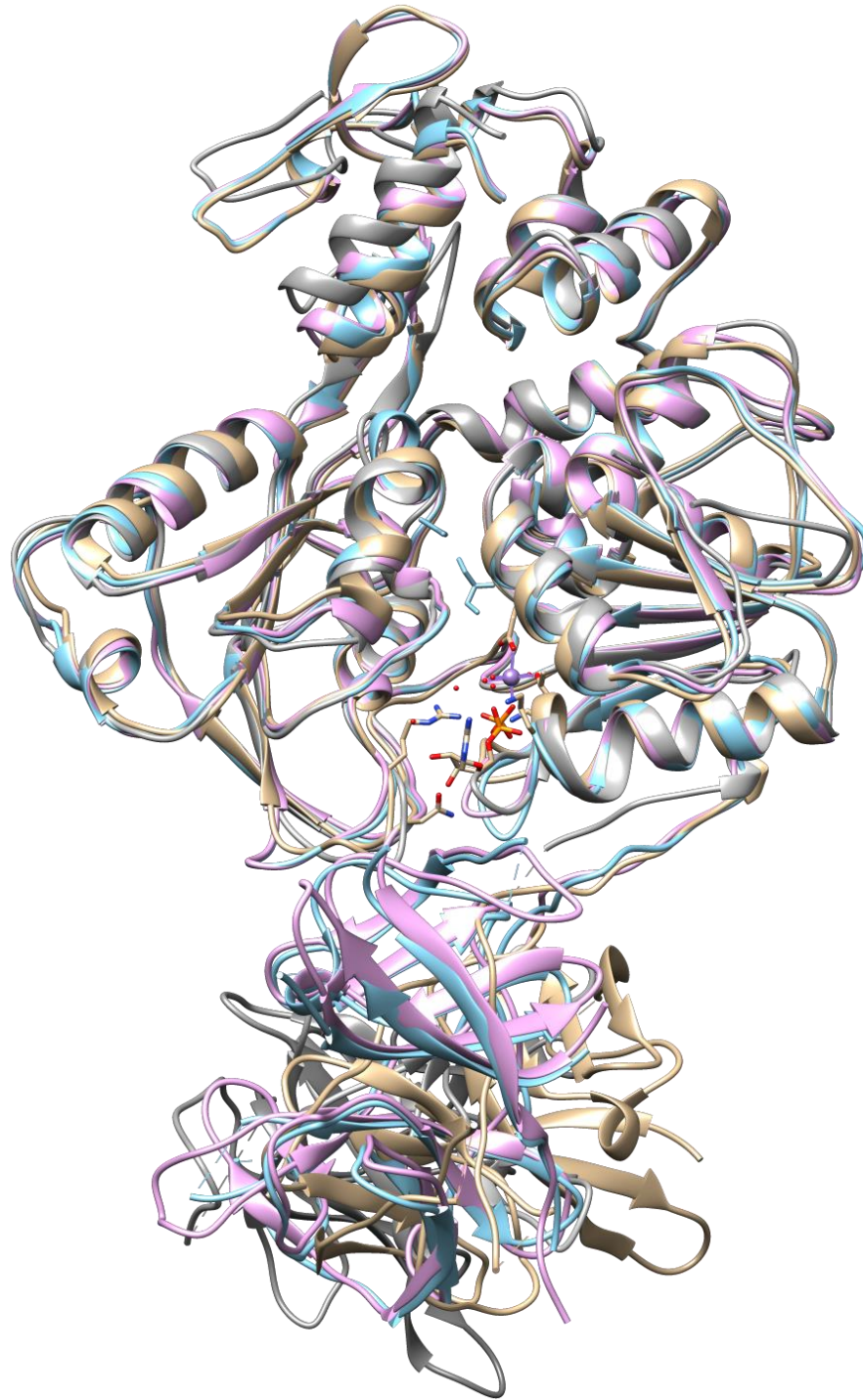


Figure 37. DENV 4 NS3 protein and MVEV NS3 protein in superimposed position, border scale 0.0. Image was generated by

UCSF Chimera.

***Percent identity: 42.72% (2WV9 vs 2WHX)***

**Percent identity: 42.88% (2WV9 vs 2VBC)**

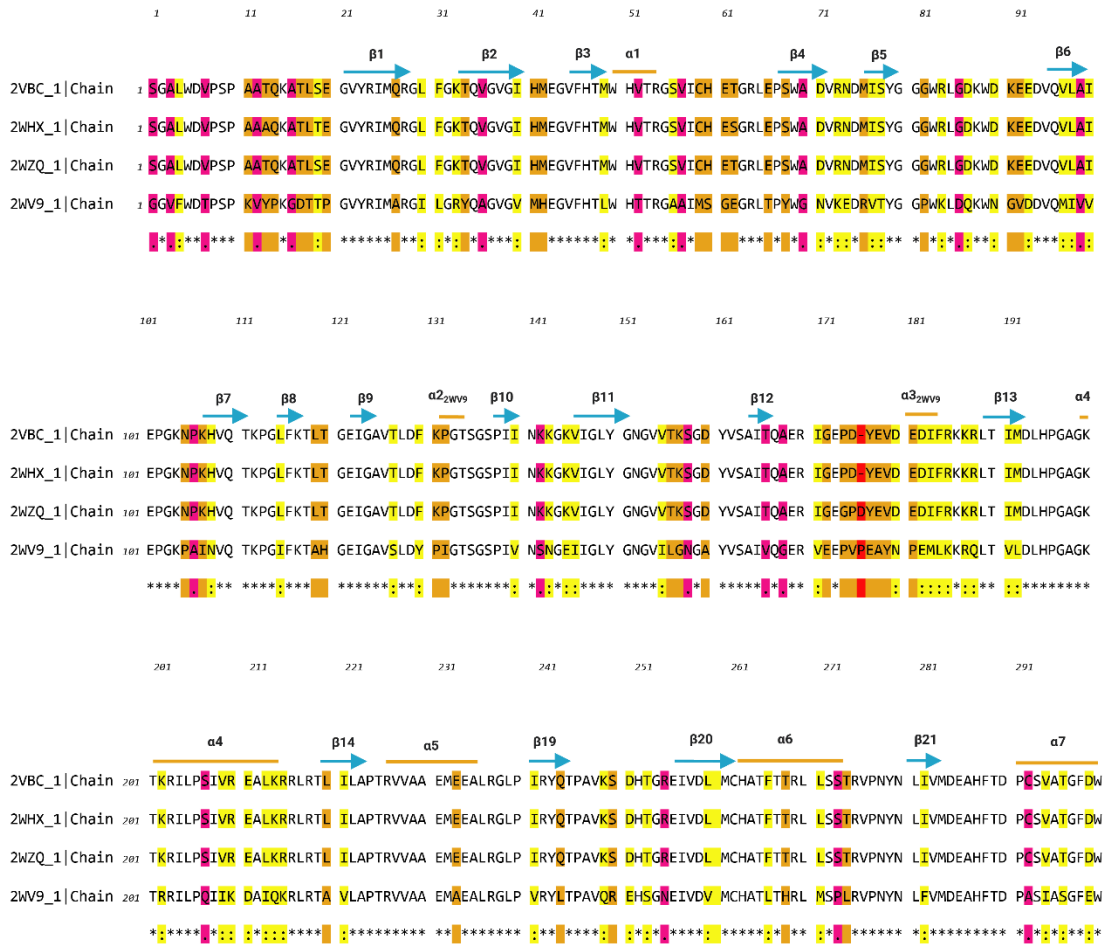
**Percent identity: 42.49% (2WV9 vs 2WZQ)**

**RMSD:  $\alpha$  Carbon pairs [601-1075], 1.360 (2WV9 vs 2VBC)**

**$\alpha$  Carbon pairs [601-1075], 1.609 (2WV9 vs 2WHX)**

**$\alpha$  Carbon pairs [601-1075], 1.368 (2WV9 vs 2WZQ).**

**Across all 409 pairs: 1.138**





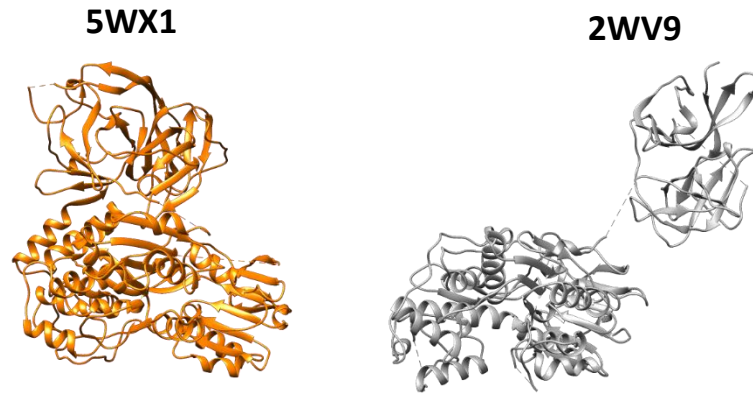


Figure 39. CSF NS3 protein and MVEV NS3 protein in superimposed position, border scale 0.8. Image was generated by UCSF Chimera.



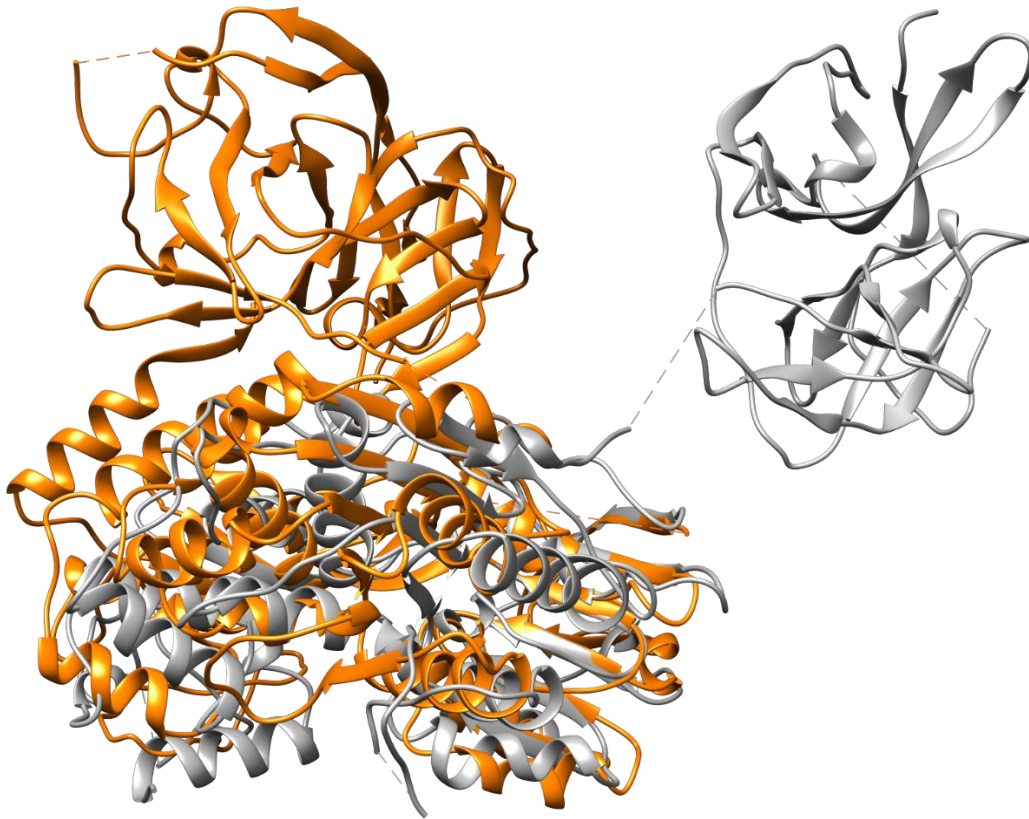


Figure 40. CSF NS3 protein and MVEV NS3 protein in superimposed position, border scale 0.0. Image was generated by UCSF Chimera.

***Percent identity: 5.65% (2WV9 vs 5WX1)***

***RMSD:  $\alpha$  Carbon pairs [485-1175], 0.970 (2WHX vs 2VBC)***

***Across all 231 pairs: 2.481***



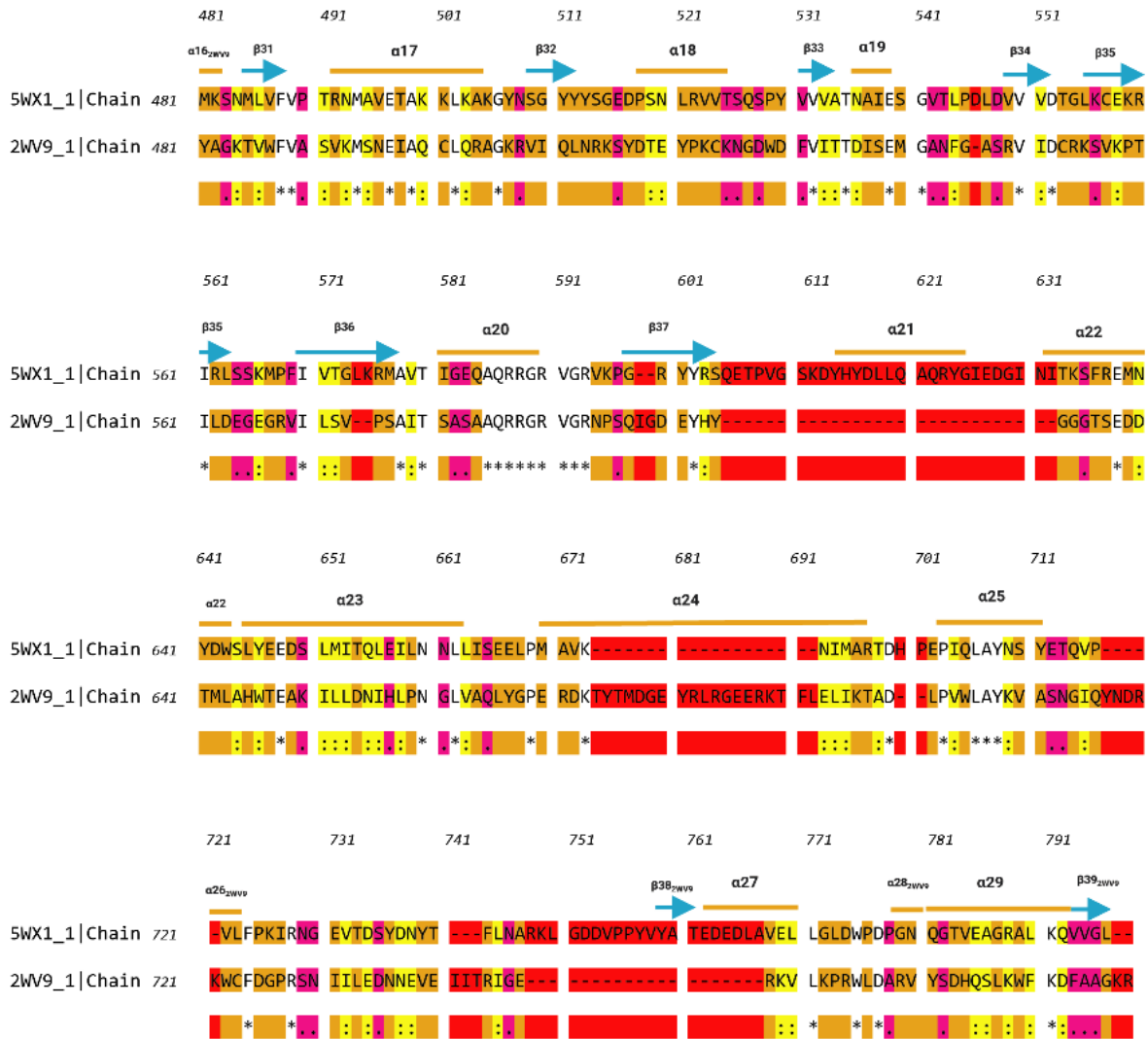


Figure 41. Sequence alignment of CSF NS3 protein (Chain A) and MVEV NS3 protein (Chain A). Alignment done by T-Coffee Expresso.

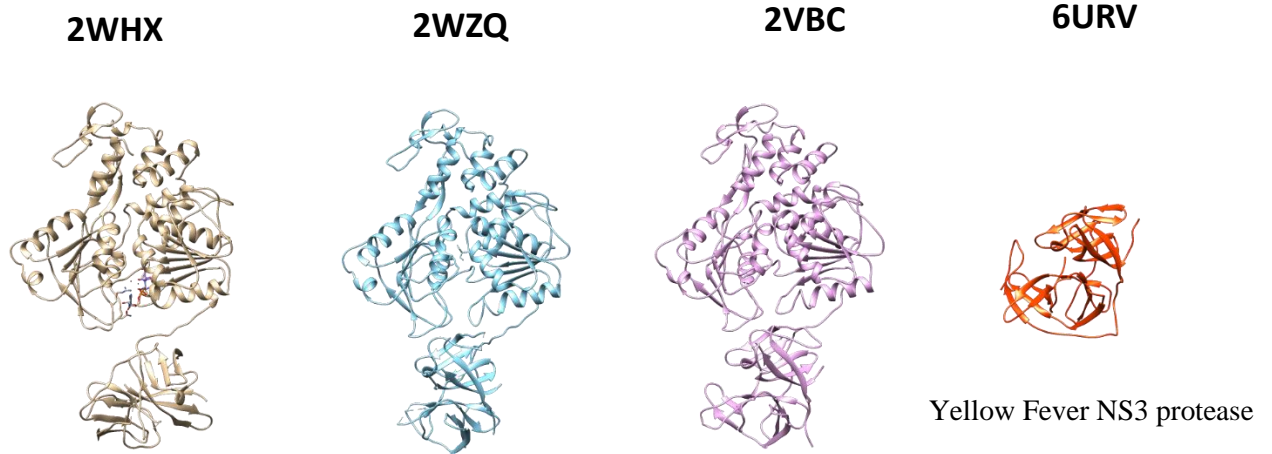


Figure 42. DENV 4 NS3 protein and Yellow Fever virus NS3 protease in superimposed position, border scale 0.8.

Image was generated by UCSF Chimera.

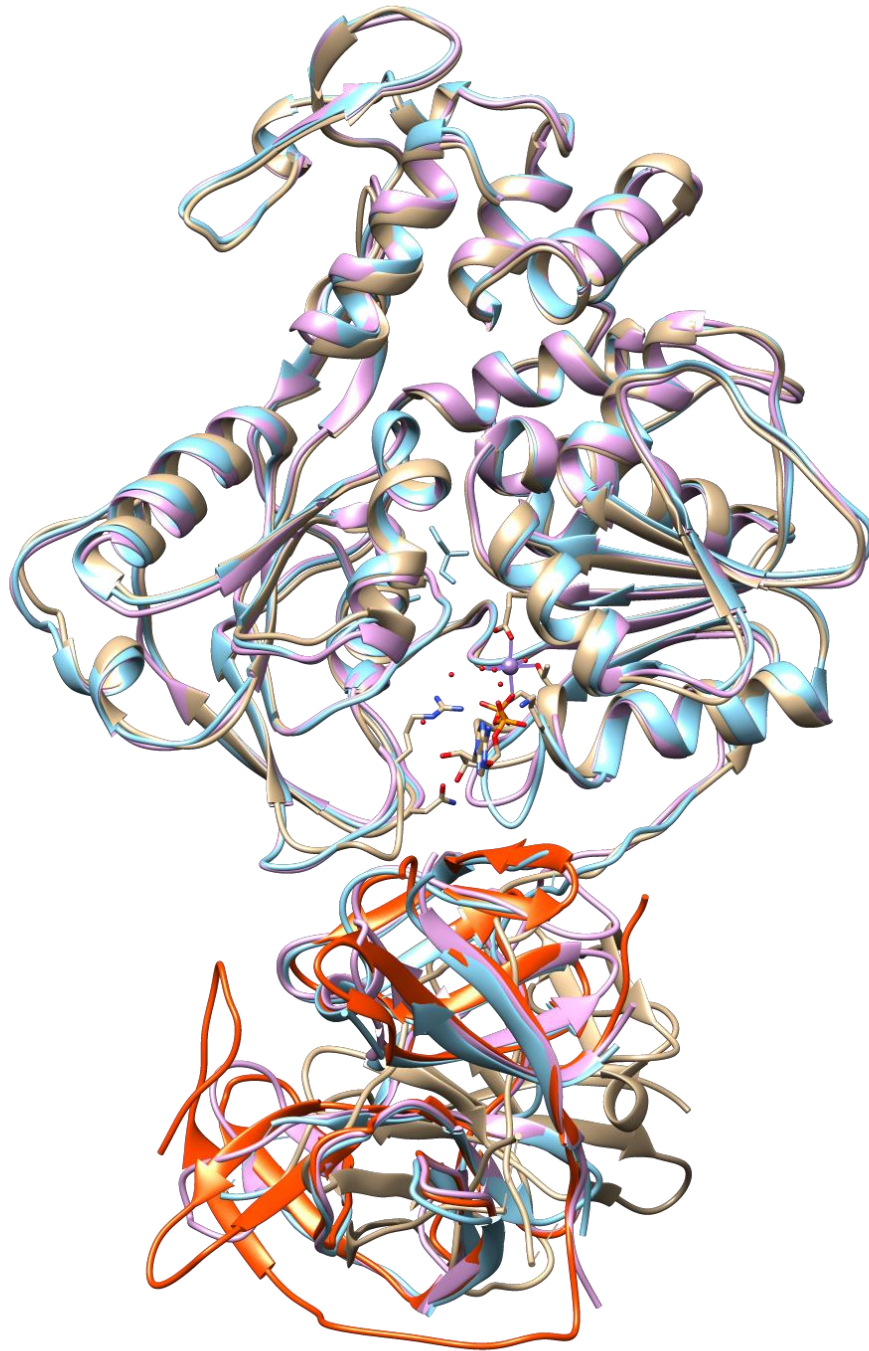


Figure 43. DENV 4 NS3 protein and Yellow Fever virus NS3 protease in superimposed position, border scale 0.0. Image was generated by UCSF Chimera.

***Percent identity: 1.75% (6URV vs 2WHX)***

*Percent identity: 45.61% (6URV vs 2WZQ)*

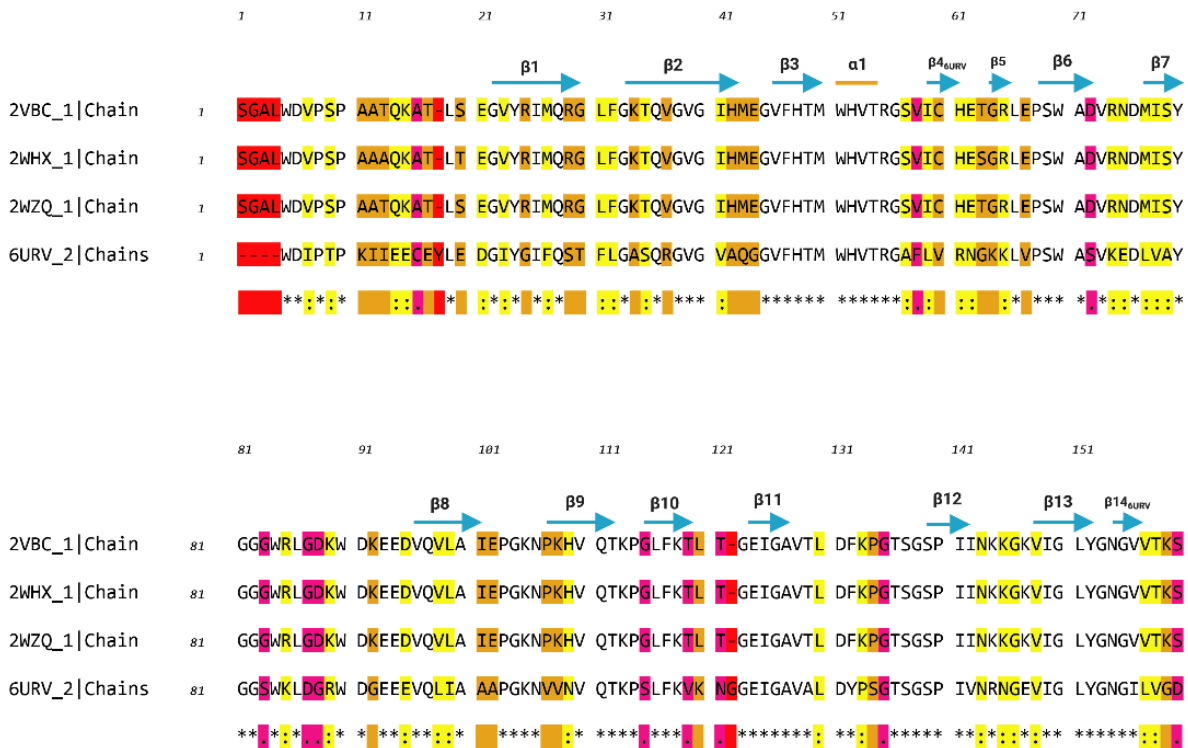
*Percent identity: 47.95% (6URV vs 2VBC)*

*RMSD:  $\alpha$  Carbon pairs [152-838], 0.733 (6URV vs 2VBC)*

*$\alpha$  Carbon pairs [152-838], 3.036 (6URV vs 2WHX)*

*$\alpha$  Carbon pairs [152-838], 0.959 (6URV vs 2WZQ).*

*Across all 150 pairs: 2.322*



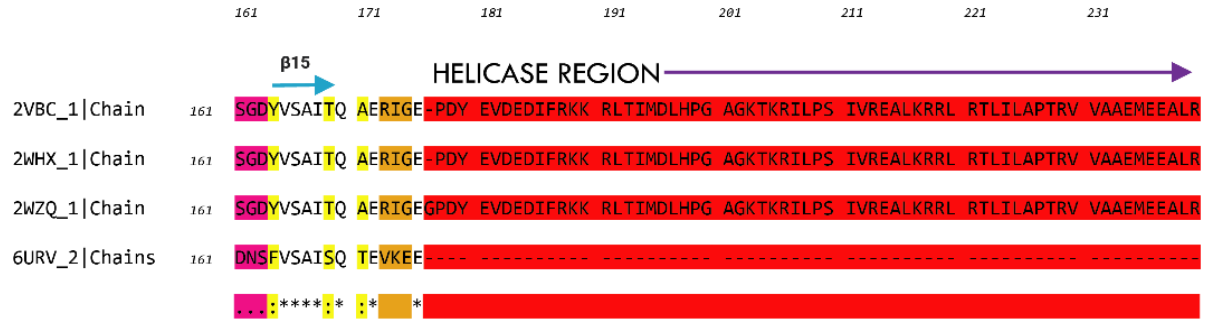


Figure 44. Sequence alignment of DENV 4 NS3 protein (Chain A) and Yellow Fever virus NS3 protease (Chain B). Alignment done by T-Coffee Expresso.

## V. Reference

- 1) Luo, D., G. Vasudevan, S., and Lescar, J. (2015). The flavivirus NS2B–NS3 protease–helicase as a target for antiviral drug development. *Antiviral Research*, 118, 148-158. ISSN 0166-3542. Doi: <https://doi.org/10.1016/j.antiviral.2015.03.014>.
- 2) Luo, D., Wei, N., Doan, D.N., Paradkar, Prasad N., Chong, Y., Davidson, Andrew D., Kotaka, M., Lescar, J., and Vasudevan, S.G. (2010). Flexibility between the protease and helicase domains of the dengue virus NS3 protein conferred by the linker region and its functional implications. *J Biol Chem*, 285(24), 18817-18827. Doi:10.1074/jbc.M109.090936.
- 3) Geok Saw, W., Pan, A., Subramanian, M.S., Manimekalai, A.G., and Grüber, G. (2019). Structure and flexibility of non-structural proteins 3 and -5 of Dengue- and Zika viruses in solution. *Progress in Biophysics and Molecular Biology*, 143, 67-77. ISSN 0079-6107 Doi: <https://doi.org/10.1016/j.pbiomolbio.2018.08.008>.
- 4) Luo, D., Xu, T., Hunke, C., Grüber, G., Vasudevan, S.G., and Lescar, J. (2008). Crystal structure of the NS3 protease-helicase from dengue virus. *J Virol*, 82(1), 173-183. Doi:10.1128/JVI.01788-07.
- 5) Li, K., Phoo, W. and Luo, D. (2014). Functional interplay among the flavivirus NS3 protease, helicase, and cofactors. *Virologica Sinica*, 29(2), 74-85. Doi:10.1007/s12250-014-3438-6.
- 6) Hayes, Edward B. (2012). Flaviviruses. *Principles and Practice of Pediatric Infectious Diseases*. (4<sup>th</sup> ed.), chapter 218. <https://www.sciencedirect.com/book/9781437727029/principles-and-practice-of-pediatric-infectious-diseases>



- 7) Floridis, J., McGuinness, S.L., Kurucz, N., Burrow, J.N., Baird, R., and Francis, J.R. (2018). Murray Valley Encephalitis Virus: An Ongoing Cause of Encephalitis in Australia's North. *Trop Med Infect Dis*, 3(2), 49. Doi:10.3390/tropicalmed3020049.
- 8) Knox, J., Cowan, R.U., Doyle, J.S., Ligtermoet, M.K., Archer, J.S., Burrow, J.N., Tong S.Y., Currie, B.J., Mackenzie, J.S., Smith, D.W., Catton, M., Moran, R.J., Aboltins, C.A., Richards, J.S. (2012). Murray Valley encephalitis: a review of clinical features, diagnosis, and treatment. *Med J Aust*, 196(5), 322-6. Doi: 10.5694/mja11.11026. PMID: 22432670.
- 9) Lindenbach, B.D., Thiel, H.J., and Rice, C.M. (2007). Flaviviridae: The Viruses and Their Replication. *Fields Virology* (Knipe, D.M. and Howley, P.M., 5<sup>th</sup> ed). Lippincott-Raven Publishers.
- 10) Barrett, A. D.T. and Higgs, S. (2007). Yellow Fever: A Disease that Has Yet to be Conquered. *Annual Review of Entomology*, 52(1), 209-229. Doi: 10.1146/annurev.ento.52.110405.091454
- 11) Risatti, G. R., and Borca, M. (2020, May) *Classical Swine Fever*. Merck Manual.[https://www.merckvetmanual.com/generalized-conditions/classical-swine-fever/classical-swine-fever#:~:text=Classical%20swine%20fever%20\(CSF\)%20is,abonortion%20may%20also%20be%20observed](https://www.merckvetmanual.com/generalized-conditions/classical-swine-fever/classical-swine-fever#:~:text=Classical%20swine%20fever%20(CSF)%20is,abonortion%20may%20also%20be%20observed). Retrieved on March 8<sup>th</sup> 2021.
- 12) Pineda, P., Deluque, A., Peña, M., Diaz, O. L., Allepuz, A., & Casal, J. (2020). Descriptive epidemiology of classical swine fever outbreaks in the period 2013-2018 in Colombia. *PLoS one*, 15(6), e0234490. <https://doi.org/10.1371/journal.pone.0234490>
- 13) Rathore, Abhay P. S. and St. John, Ashley L. (2020). Cross-Reactive Immunity Among Flaviviruses. *Front. Immunol*, 11(334). Doi: 10.3389/fimmu.2020.00334

- 14) Center for Disease control and Prevention (2020, June). *West Nile*. Retrieved on April 17<sup>th</sup>, 2021. <https://www.cdc.gov/westnile/index.html>
- 15) Center for Disease control and Prevention (2019, October). *Zika*. Retrieved on April 17<sup>th</sup>, 2021. <https://www.cdc.gov/zika/about/overview.html#:~:text=Zika%20virus%20was%20first%20discovered,probably%20occurred%20in%20many%20locations.>
- 16) Center for Disease control and Prevention (2020, July). *Dengue*. Retrieved on April 17<sup>th</sup>, 2021. <https://www.cdc.gov/dengue/index.html>
- 17) Mitchell, J. (1744). Account of the Yellow fever which prevailed in Virginia in the years 1737, 1741, and 1742, in a letter to the late Cadwallader Colden, Esq. of New York, from the late John Mitchell, M.D.F.R.S. of Virginia. *American Medical and Philosophical Register*, 4, 181-215. Reprinted in 1814. [https://books.google.com/books?id=\\_EZJAAAAYAAJ&pg=PA181#v=onepage&q&f=false](https://books.google.com/books?id=_EZJAAAAYAAJ&pg=PA181#v=onepage&q&f=false)
- 18) Rust, R.S. (2014). Arbovirus Encephalitis. *Encyclopedia of the Neurological Sciences* (Aminoff, M.J., Daroff, R.B., 2<sup>nd</sup> ed.). Academic Press. ISBN 9780123851581. Doi: <https://doi.org/10.1016/B978-0-12-385157-4.00371-7>. (<https://www.sciencedirect.com/science/article/pii/B9780123851574003717>)
- 19) Mancini, E. J., Assenberg, R., Verma, A., Walter, T. S., Tuma, R., Grimes, J. M., Owens, R. J., & Stuart, D. I. (2007). Structure of the Murray Valley encephalitis virus RNA helicase at 1.9 Angstrom resolution. *Protein science*, 16(10), 2294–2300. Doi: <https://doi.org/10.1110/ps.072843107>

- 20) Beasley, D. W. C. and Barrett, A. D. T. (2008). The Infectious Agent. *In Dengue: Tropical Medicine: Science and Practice* (Pasvol, G. and Hoffman, S. L., 5<sup>th</sup> ed.), 29-74. London: Imperial College Press.
- 21) Moennig, V. and Greiser-Wilke, I. (2008). Classical Swine Fever Virus. *Encyclopedia of Virology* (Mahy, B. W. J., and Van Regenmortel, M. H.V., 3<sup>rd</sup> ed.), 525-532. Academic Press.
- 22) Creative Diagnostics. (n.d). *Flavivirus*. <https://www.creative-diagnostics.com/Flavivirus.htm#:~:text=Flaviviruses%20show%20morphological%20uniformity%20with,about%2010%20kb%20in%20size>.
- 23) Di Tommaso, P., Moretti, S., Xenarios, I., Orobittg, M., Montanyola, A., Chang, J. M., Taly, J. F., & Notredame, C. (2011). T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic acids research*, 39(Web Server issue), W13–W17. Doi: <https://doi.org/10.1093/nar/gkr245>
- 24) Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera-a visualization system for exploratory research and analysis. *J Comput Chem*, 25(13), 1605-12.
- 25) Simmonds, P. (1995). Variability of hepatitis C virus. *Hepatology*, 21(2), 570–583.
- 26) ICTV. (n.d). "*Virus Taxonomy: 2018b Release*. <https://talk.ictvonline.org/taxonomy/>. Retrieved April 25<sup>th</sup> , 2021.
- 27) 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., and Lescar, J. (2008). Insights into

RNA unwinding and ATP hydrolysis by the flavivirus NS3 protein. *EMBO J.*, 27(23), 3209-3219. Doi: 10.1038/emboj.2008.232

28) Jones, C.T., Patkar, C.G., and Kuhn, R.J. (2005). Construction and applications of yellow fever virus replicons. *Virology*, 331(2), 247–259.

29) Khromykh, A.A., Sedlak, P.L., and Westaway, E.G. (2000). Cis- and trans-acting elements in flavivirus RNA replication. *J Virol*, 74(7), 3253–3263.