

TABLE OF CONTENTS

Chapter 1: Introduction	3
1.1 Dengue: A Clinical Approach	3
1.2 Dengue: Prevention and Treatment.....	7
1.3 Dengue Virus: genome and Life Cycle.....	8
1.4 Viral Proteins.....	10
1.4.1 M	10
1.4.2 E.....	10
1.4.3 Capsid.....	11
1.4.4 NS1	11
1.4.5 NS2a	12
1.4.6 NS2b	12
1.4.7 NS3	12
1.4.8 NS4a	12
1.4.9 NS4b	13
1.4.10 NS5	13
1.5 Protein-Protein Interactions (PPI).....	13
1.5.1 In-Vitro based approach.....	14
1.5.2 In-Vivo based approach.....	15
1.5.3 In-Silico Based Approach.....	16
1.5.4 Structure-based Prediction Approach	16
1.5.5 Sequence-based Prediction Approach	16
1.5.6 Domain-Pairs-based Approach.....	17
1.5.7 Hot-spot-based Approach	17
1.6 Project Outline	17
1.6.1 Hypothesis and aim.....	18
1.6.2 Objectives.....	18
Chapter 2: Methods	20
2.1 Resources	20
2.1.1 Databases.....	20
2.1.2 Software and tools	22

2.2 Workflow.....	23
2.2.1 The Structural Approach	23
2.2.2 The Genomics Approach	27
2.2.3 Python Script	27
Chapter 3: Results and discussion: Data Collection and Processing	30
3.1 Firth Method: Structural Approach.....	30
3.1.1 Step 1: Obtaining a Protein-Protein Structure List.....	30
3.1.2 Step 2: Filtering and organizing the PDB ID Lists.....	31
3.1.3 Step 3: Acquiring Domains	32
3.1.4 Step 4: Merging Similar Domains and Proteins	32
3.2 Second Method: Sequence Approach.....	33
3.2.1 Step 1: Retrieving Dengue Protein Sequence.....	33
3.2.2 Step 2: Modelling Domains	34
3.2.3 Step 3: Proteins-Domain Association	35
3.2.4 Step 4: Filtering Data through PDBSum	36
3.2.5 Step 5: Merging IRIndex Database	36
3.2.6 Step 6: Classification of Data.....	36
Chapter 4: Results and Disucssion: Functional Candidates.....	38
4.1 Sample Selection	38
4.2 P06935: West Nile virus Polyprotein.....	39
4.2.1 Introduction to West Nile Virus	39
4.2.2 P06935: Interaction information.....	40
4.2.3 Analysis: Sequence Alignment	40
4.2.4 Analysis: Structural Alignment	40
4.2.5 Hotspot Analysis.....	46
Chapter 5: Conclusion	53
5.1 Summary	53
5.2 Drawbacks	53
5.3 Future Research	54
Chapter 6: Biography	56

CHAPTER 1: INTRODUCTION

1.1 Dengue: A Clinical Approach

Dengue Fever is a deadly disease caused by the dengue fever virus. It is one of the most widely spread mosquito-borne diseases (WHO and HealthMap.org, 1997). According to the World Health Organization (WHO, 2012), around 40 % of the world population is at risk of contracting dengue, especially in the tropical and subtropical regions (WHO and HealthMap.org, 1997) (Fig 1.1). It is estimated that the rate of dengue infection might be as high as 100 million cases annually and about 500,000 per year require hospitalization (Halstead et al., 2007; WHO, 2012).

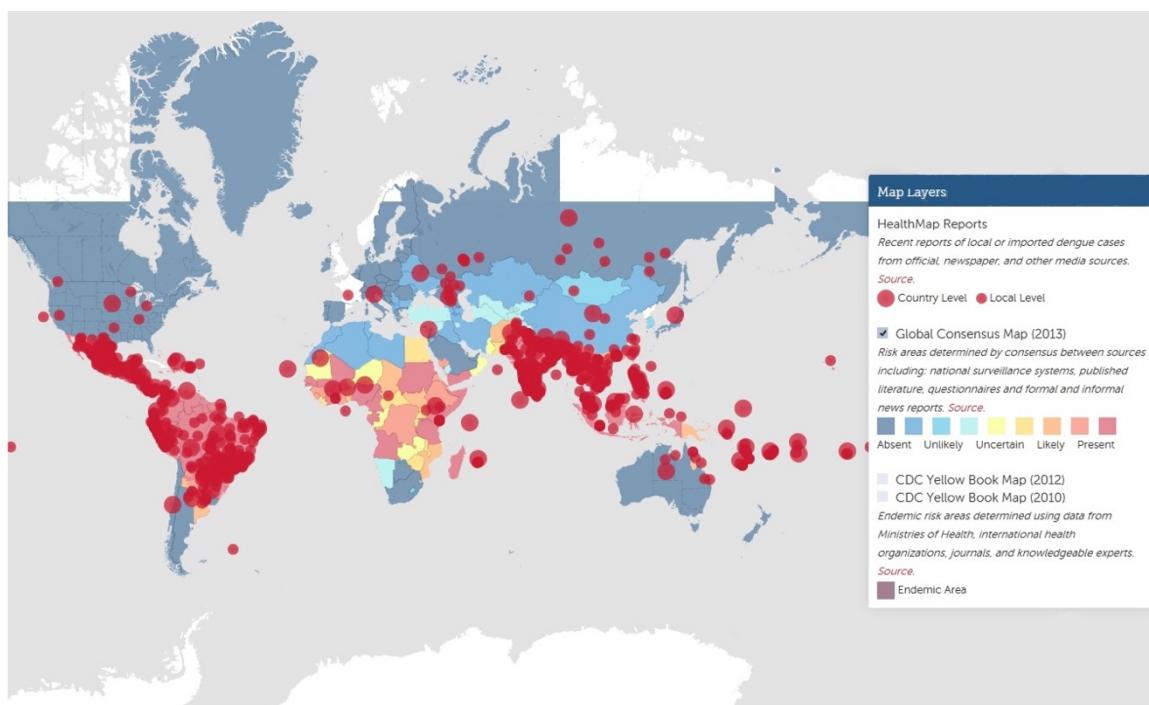


Figure 1-1 Dengue Wold Map (August 2017) (WHO and HealthMap.org, 1997)

This map reveals the indices of dengue infection in the world map from February 2017 to August 2017. Infection is mainly located on the equatorial line.

Concern is increasing respect to an increasing spread of this disease, as a result of the virus re-introduction into the mosquito-infected regions, caused by global warming (Centers and Prevention, 2010; Dorji et al., 2009; Kovats et al., 2001; La Ruche et al., 2010; McMichael et al., 2006; Pandey et al., 2008).

Re-introduction of the virus into mosquito-infected regions produced predicted increase of 44% and 53% increase risk of dengue infection by 2055 and 2085 (Hales et al., 2002). In 2002, the WHO increased their commitment to fight against dengue, mostly evidenced in the revision of the International Health Regulation, published in 2005 World Health Assembly WHA58.3 (World, 2005).

The symptoms of this disease include high fever, body aches and rashes WHO, 2009. Most hospital patients are children and teenagers, who usually present serious and complicated symptoms of dengue (WHO and HealthMap.org, 1997). Usually, full recovery takes a week; however, some cases may develop haemorrhagic manifestations leading to the most severe form of the disease, called Dengue Haemorrhagic Fever (DHF) (Rigau-Pérez et al., 1998) (Fig 1.2). The most severe complication of DHF is the Dengue Shock Syndrome (DSS), consisting of hypovolemic shock caused by excessive plasma leakage from the blood vessels, as an over-response to the infection (Rigau-Pérez et al., 1998) (Gluber, 1998).

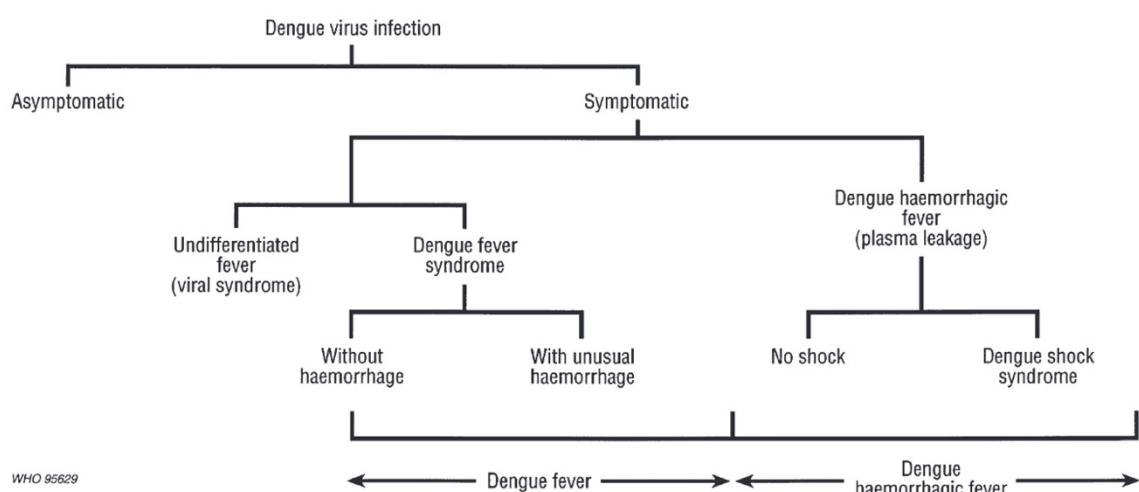


Figure 1-2 Manifestations of Dengue Infection (adapted from WHO, 2009)

This graph describes the clinical manifestations of the Dengue Virus Infection. Dengue Infection can be symptomatic or asymptomatic. When Dengue infection is symptomatic, it can be divided into dengue fever and dengue haemorrhagic fever. Dengue fever may be undifferentiated fever or dengue fever syndrome, either without haemorrhage or with unusual haemorrhage. Dengue haemorrhagic fever (with plasma leakage) can be divided into shock and dengue shock syndrome.

Dengue Infection is described as a systemic and dynamic disease. The symptoms presented vary greatly, from severe to non-severe clinical manifestations. After the initial incubation period, the disease develops three identifiable phases Lum et al., 2014; Rigau-Pérez et al., 1998]:

- Febrile Phase

- Critical Phase
- Recovery Phase

An array of symptoms characterizes the Febrile Phase. The most common are: high fever, facial flushing, skin erythema, widespread body pain, myalgia, arthralgia, headache, sore throat, injected pharynx, conjunctival infection, anorexia, nausea, and vomiting. Non-common symptoms include mild haemorrhagic manifestations like: petechiae, mucosal membrane bleeding, massive vaginal bleeding, and gastrointestinal bleeding (Kalayanarooj et al., 1997; Balmaseda et al., 2006; Lum et al., 2008; Ismail, 2012) (Table 1.1). This phase extended from 2 to 7 days. The symptoms are indistinguishable between severe and non-severe dengue cases. For this reason, monitoring the patients respect to indicators of progress toward the Critical Phase constitutes a priority (Fig 1.3).

Dengue Fever: Symptoms	Dengue Haemorrhagic Fever: Symptoms
Severe headache	Fever
Sever eye pain	Positive Tourniquet test
Joint pain	Hepatomegaly
Muscle and/or bone pain	Thrombocytopenia
Rash	Dengue Shock Syndrome
Middle bleed Manifestation	
Low white cell count	(b) Symptoms of Dengue Haemorrhagic Fever
(a) Symptoms of Dengue Fever	

Table 1-1: Clinical Symptoms of Dengue fever (Gubler, 1997)

(a) These are the symptoms that integrate the clinical diagnosis of Dengue Fever. (b) These are the symptoms that integrate the clinical diagnosis of Dengue Haemorrhagic Fever.

The positive tourniquet test is a clinical test that helps distinguish between severe and non-severe dengue cases. The first abnormality present in the full blood count is a decrease in the total white cell count (indicating high probability of dengue).

The beginning of the Critical Phase is marked by increase capillary permeability in parallel with increasing haematocrit levels (Lum et al., 2014). This usually occurs between the days 3 to 7, and the Critical Period of plasma leakage usually lasts from 24 to 48 hours (Lum et al., 2014) (Fig 1.3). The distinction between severe and non-severe malaria becomes apparent: if the patient does not experience

an increase in the capillary permeability, the condition will improve, while an increase in the capillary permeability will further aggravate the condition. The usual diagnostic tools used are abdominal ultrasound and chest x-ray.

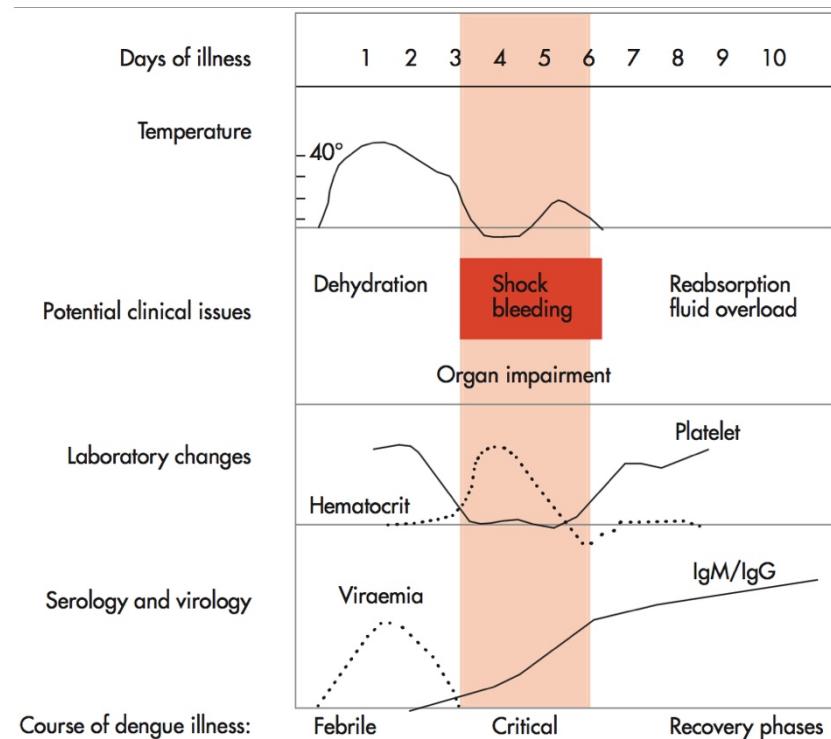


Figure 1-3 Dengue Disease Progression (Yip, 1980)

Yip, 1980 imaged the 3 phases in time progression: Febrile, Critical and Recovery. The Febrile Phase (first 3 days) shows temperature increase up to 40 degrees, with a steady decrease by the end of the 3rd day. There are also signs of dehydration, decrease of platelets, and increase of viremia. The Critical Phase (between the days 3 – 6), is characterized by temperature decrease, shock bleeding (including organ impairment), haematocrit increase and constant decrease of platelets. Also, viremia disappears.

Dengue Shock presents distinctive warning signs, such as subnormal body temperature, progressive organ impairment, metabolic acidosis, and intravascular coagulation (Lum et al., 2014). This leads to the Dengue Shock Syndrome (DSS). The severity of DSS depends on the severity of the haemorrhage caused by the previous symptoms, which causes haematocrit decrease (Lum et al., 2014). This ultimately increases the severity of the DSS. Patients are also prone to develop severe organ impairment like hepatitis, encephalitis, myocarditis and/or severe bleeding without obvious plasma leakage or shock (Lum et al., 2014).

The Recovery Phase starts after the patient survives the 24-48 hours of the Critical Phase (Fig 1.3).

After surviving the Critical Phase, gradual re-absorption of extravascular compartment fluid takes place (Lum et al., 2014). During the Recovery Phase, the patient's well-being improves, appetite returns, gastrointestinal symptoms reduce, haemodynamic status stabilizes and diuresis ensues (Lum et al., 2014). The haematocrit also stabilizes, usually have blood cell count stability precedes platelet recovery (Lum et al., 2014).

Some patients may develop a rash, or have bradycardia and electrocardiography alterations.

Fluid therapy is recommended with caution, since excess is associated with pulmonary oedema or congestive heart failure.

Severe dengue is characterized by one or more of the following features:

- Plasma leakage that may lead to shock.
- Severe bleeding,
- Severe organ impairment.

As dengue vascular permeability progresses, hypovolaemia worsens and results in shock. This usually takes place in 4 to 5 days of illness.

1.2 Dengue: Prevention and Treatment

The first generation of vaccines was developed with live attenuated virus LAV (Bhamarapravati and Suttee, 2000) or inactivated virus (Lyons et al., 2007). LAV has not been successful due to insufficient immunogenicity, excess reactogenicity and an imbalanced response to each dengue serotype (Edelman, 2007).

Current efforts are being placed on creating a vaccine based on recombinant chimeric virus. The four serotypes of dengue have distinctively antigenic profiles, with an amino acid homology that vary between 65% and 70%. This effort is being made to create a balanced immunological response against all four serotypes, as well as to reduce dengue antibody-dependant enhancement (ADE) of infection (Dejnirattisai et al., 2010).

At this moment, the only dengue vaccine available and approved by the WHO is Dengvaxia by Sanofi Pasteur. The WHO recommends that the candidates for vaccinations meet the following criteria (World, 2017):

- Individuals should go through an immunological screening.
- Individuals should have seroprevalence of 50 or higher.
- Administration of 3-dose series given at 0/6/12 months schedule.
- Individuals should be 9 years old or higher

1.3 Dengue Virus: genome and Life Cycle

Dengue virus belongs to the genus Flavivirus of the Flaviviridae family; other members include the West Nile Virus (WNV), Yellow Fever Virus (YFV) and Japanese Encephalitis Virus (JEV) (Harris et al., 2008). The Flaviviridae family is arthropod-borne, meaning that it requires an insect as a host to complete its life cycle. The dengue virus rarely causes neurotropic disease, unlike other flaviviruses such as JEV and WNV. Dengue Virus and YFV requires a human host to complete their life cycle (Rigau-Pérez et al., 1998).

The genome of flavivirus consists of one molecule of positive single-stranded RNA (Rice et al., 1985). The genome is encapsulated in a viral capsid protein shell, which is developed by a membrane derived from a host cell. This single poly-protein encodes 10 proteins (Iglesias and Gamarnik, 2011; Rice et al., 1985).

Structural Proteins:

- Precursor of Membrane Protein (PrM)
- Envelop Protein (E)

Capsid

- Non-Structural (NS) Proteins:
- Non-Structural Protein 1 (NS1)
- Non-Structural Protein 2A (NS2A)

- Non-Structural Protein 2B (NS2B)
- Non-Structural Protein 3 (NS3)
- Non-Structural Protein 4A (NS4A)
- Non-Structural Protein 4B (NS4B)
- Non-Structural Protein 5 (NS5)

The known molecular process of infection proceeds as follows (Rodenhuis-Zybert et al., 2010):

- Dengue binds to receptors on the host cell receptors at the surface to induce endocytosis of the virion.
- Ph. decreases causing release of viral RNA in the cytoplasm via structural changes
- RNA is translated into a polyprotein and is inserted in the host endoplasmic reticulum membrane
- The Polyprotein is cleaved by host proteases furin and signalase and by viral protease NS2B/NS3 complex, to form the ten individual proteins.
- Viral RNA is transferred to replication complexes containing NS5 (viral RNA-dependent RNA polymerase). This takes place in vesicles packets; induced by viral infection.
- Viral RNA is coated with capsid forming a nucleocapsid which buds into the ER lumen and acquires a virus envelope from the ER membrane.
- The virus envelope also contains PrM and E proteins embedded during dengue polyprotein translation at the ER membrane.
- The PrM protein of virions in the lumen of the ER and Golgi apparatus are further cleaved by host protease
- Then the host cell excretes the virus via the secretion pathway.

1.4 Viral Proteins

1.4.1 M

M Protein precursor (PrM) is a 26 kDa glycoprotein located in the ER lumen and anchored to the ER membrane during synthesis (Lindenbach, 2005). PrM binds to the envelop protein, in order to prevent premature fusion of the Golgi membrane (Yu et al., 2009).

In vitro studies suggest that a mixture between mature and immature disseminating virions is found in human cell cultures and infected mosquitoes. This can prevent the elicitation of the immune system, since the uncleaved PrM of an immature virion elicits non-neutralizing antibodies.

The ectodomain of M interacts with Tctex-1, a dynein light chain that functions in cargo binding (Brault et al., 2011). Silencing studies have revealed that silencing of tctex-1 reduces dengue replication in cell culture (Brault et al., 2011).

1.4.2 E

The envelope Glycoprotein or E protein is a major protein on the surface of flavivirus virions of approximately 53kD (Schlender et al., 1996). It is synthesized as a Type I membrane protein, which holds 12 conserved cysteines that form disulphide bonds and in some species E is N-glycosylated. Various essential process, including folding, stabilization in low pH and secretion is dependent on co-expression with prM; This is because E is a class II fusion protein that serves as a mediator for both receptor binding and membrane fusion.

The Envelope protein is composed of three domains:

- D1.- It forms an eight-stranded B-barrel. Alongside DIII, forms a hydrophobic pocket that buries DII fusion peptide.
- DII.- It is a long finger-like domain that projects alongside the virus surface. The fusion peptide is located within this domain.

- DIII.- It maintains an immunoglobulin-like fold. Alongside DI, it forms a hydrophobic pocket that buries DII fusion peptide. Another role is also involved in this process.

1.4.3 CAPSID

Capsid is a building block for the shell of a nucleocapsid (Kuhn et al., 2002). It is a 11kDa protein, 100 amino-acid-long protein that contains several basic amino acids, which give it a positive charge (isoelectric point = 12.5) (Kuhn et al., 2002). After translation, it has a transmembrane domain at its C-terminus, which allows it to anchor to the ER, while the N-terminus is exposed to the cytoplasm (Lindenbach, 2005). Maturation occurs when dengue serine protease NS2B/NS3 cleaves the transmembrane domain (Shiryaev et al., 2006). After cleavage, it naturally forms homodimers that become part of the icosahedral nucleocapsid shell (Kuhn et al., 2002; Shiryaev et al., 2006).

It also contains potential nuclear localization signals (NLS) and it has been known to be located in the nucleus or nucleolus in a several cell lines (Chen et al., 2002; Sangiambut et al., 2008). Nuclear localized capsid interacts with death-domain associated protein (DAXX). It is a multi-functional protein known to have a role in apoptosis (or programmed cell death). Nuclear Capsid can sensitize cells to undergo apoptosis (Nagila et al., 2011). Other potential roles of nuclear and nucleolar localized capsid are still unclear.

1.4.4 NS1

NS1 is a 46 kDA glycoprotein and its function is still unclear. Lindenbach, 2005 showed immature NS1, as part of the dengue polyprotein residue. NS1 is cleaved by the host signals and it can form homodimers associated to the membrane (Winkler et al., 1988; Winkler et al., 1989); although the mechanism remains uncertain. Mutagenesis studies have shown the importance of NS1 for an efficient RNA replication. It also seems to play a role in immunosuppression, by suppressing through binding to C1, C4 and C4BP (Avirutnan et al., 2011; Avirutnan et al., 2010; Somnuke et al., 2011).

1.4.5 NS2A

NS2A function has not been extensively studied. Possible functions of NS2A include Virion assembly and inhibition of interferon signalling. A study performed by (Munoz-Jordan et al., 2003) showed that NS2a could suppress interferon-beta-stimulated gene expression.

1.4.6 NS2B

The main known function of NS2B is as a cofactor for NS3, a viral serine protease (Falgout et al., 1991).

1.4.7 NS3

NS3, a 70kDa cytoplasmic protein (Lindenbach, 2005), has two function domains. One is a N-Terminal serine protease domain and a C-terminal RNA helicase Domain (Lindenbach, 2005). The N-Terminal Serine protease is responsible for cleaving the main dengue polyproteins into the individual proteins that it encodes; this process takes place in the cytoplasmic side of the ER membrane.

The C terminal RNA helicase domain is required for RNA viral replication (Sampath and Padmanabhan, 2009). It has RNA triphosphatase activity, which plays a role in 5'capping of the viral RNA (Bartelma and Padmanabhan, 2002; Chernov et al., 2008; Lindenbach, 2005; Locatelli et al., 2001).

Another possible function is to induce apoptosis. WNV NS3 shows a direct interaction and cleavage of caspase 8, which induces apoptosis. This possible function could account for the tissue damage, present in patients with dengue.

1.4.8 NS4A

There is not concrete evidence of the role of NS4A, a 16kDa transmembrane protein. Miller et al., 2007, reported that it might play a role in ER membrane rearrangement, since Huh-7/T7 cells which expressed it show rates of membrane rearrangement. Another possible function was shown by Munoz-Jordan et al., 2003, in which it was shown that expression of NS4A caused suppression of the Interferon-beta-stimulated gene expression.

1.4.9 NS4B

NS4B is a strong interferon inhibitor, stronger THAN NS2A AND NS4A. (Kelley et al., 2011; Kelley et al., 2012).

1.4.10 NS5

NS5 is a multi-functional protein with a N-terminal with methyltransferase domain (MTAse) and C-terminal RNA-Dependent RNA polymerase (RdRp)(Lindenbach, 2005). N-terminal functions for RNA capping (Egloff, 2002) while the C-terminal is required for RNA replication (Ackermann and Padmanabhan, 2001).

NS5 and NS3 form a complex which is essential for RNA replication, since RdRp function and the helicase activity of NS3 are both required for RNA replication.

NS5 also plays a role in interfering Interferon pathway by binding to STAT2 and promoting its degradation (Ashour et al., 2009; Schulze zur Wiesch et al., 2005).

1.5 Protein-Protein Interactions (PPI)

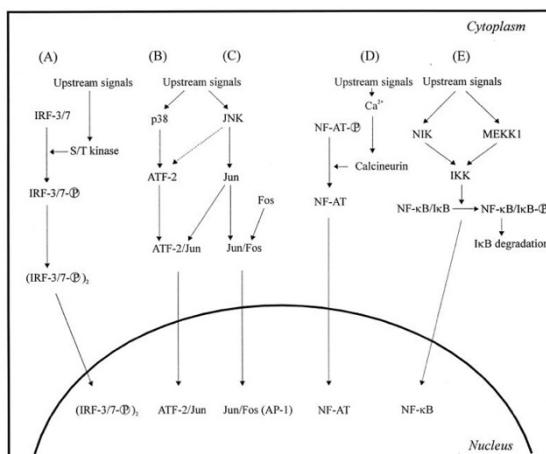


Figure 1-4 Immunological pathways activated by dengue infection (Mogensen and Paludan, 2001b)

One or more of the signal transduction cascades leads to the expression of Virus induced expression of Cytokines and Chemokines. These later leads to the activation of IRF-3/7(A), ATF-2/Jun(B), AP-1(C), NF-AT(D), and NF- κ B(E).

Protein-Protein Interactions are building blocks of molecular pathways that run all physiological mechanism from a cell to an entire body.

One such mechanism that is dependent on protein-protein interactions are the immunological responses triggered by viral infection. Figure 1.4 shows the most important signal transduction pathways activated by viruses. Interferon (IFN) Regulatory Factor 3 (IRF-3) and IRF-7 are virus-activated transcription factors that play an important role in the IFN-alfa/beta expression. Mitogen-Activated Protein (MAP) Kinases p38 and Jun N-terminal Kinase (JNK) are also activated by viruses; further activation takes place with the Activating Transcription Factor 2 (ATF-2) and Jun. Jun can form homodimers and heterodimer with ATF-2 and Fos (responsible for). Another transcription factor activated by virus infection is the nuclear factor of activated T cells (NF-AT). Activation of NF-AT is a hallmark of most infections, including viral infections. Upon infection, MAP Kinase Kinase Kinase (MAP3K) are activated, promoting the activation of a large kinase, able to phosphorylate I κ B in two specific amino-terminal serine residues. I κ B is phosphorylated by I κ B Kinase Alfa (IKK alpha) and IKK Beta. Phosphorylated I κ B is a degradation that leads to migration of NF-AT to the nucleus, activating transcription (Mogensen and Paludan, 2001).

There are 3 main ways of studying Protein-Protein Interactions and identifying them. The first one involves the in-vitro based approach; the second one, the in-vivo based approach, and the third one, the in-silico based approach.

1.5.1 IN-VITRO BASED APPROACH

One in-vitro analysis technique is TAP Tagging. The first attempt in a high-throughput manner was for analysing the yeast interactome. The method used by Gavin et al., 2002 is based on a double tagging of the protein involved on its chromosomal locus, with later identification through SDS-PAGE, followed by mass spectrometry analysis (a two-step identification process). This allows us to identify the collaborator of the protein of interest. The strength of this method permits the identification of a wide variety of protein complexes by testing their activity.

Another in-vitro method is affinity chromatography (Meng et al., 2014). This is a highly responsive method that can detect weakest interaction in proteins. However, this method can generate false positive results. For this reason, it should be accompanied by other methods, in order to cross-check and verify the results.

Co-immunoprecipitation is another method that confirms interaction, using whole cell extract (Xiong et al., 2014). This method places the targets in a complex mixture, with all possible elements required for a successful interaction, leaving the proteins in their native form. Eukaryotic cell lines can be used, allowing the study of interactions that cannot be reproduced in a prokaryotic model.

Another method, which is becoming increasingly popular due to its ability for detecting proteins, monitoring the expression levels and probing protein interactions and functions, is Protein microarray (Chen et al., 2014). This method consists of a piece of glass, where affixed protein molecules are located in an ordered manner, in separate locations. The purpose of these methods is to carry out efficient and sensitive high-throughput protein analysis.

X-ray crystallography and Nuclear Magnetic Resonance (NMR) are methods that enable visualization of protein structure. X-Ray crystallography allows to visualize proteins at the atomic level. It can show how protein interacts and the changes in conformational changes. NMR is based on magnetically active nuclei oriented by a strong magnetic field with unique characteristics, such as absorption of electromagnetic radiation and characteristic frequencies.

1.5.2 IN-VIVO BASED APPROACH

Yeast Two Hybrid (Y2H) is an in-vivo method. It requires two protein domains. A DNA binding domain to aid in DNA binding and an Activation domain that allows transcription of DNA (Cagney et al., 2000). It allows direct recognition of PPI. However, it can lead to false positive interactions and false negative interactions; false negative reactions caused by the inability of proteins to fold and function in a prokaryotic environment (for eukaryotic proteins). For case of proteins that require post-translational modifications, these cannot be detected. Efforts have been undertaken to reduce the errors caused by Y2H (James et al., 1996). Other techniques that are widely used are: Bioluminescence Resonance Energy Transfer (BRET), Fluorescence Resonance Energy Transfers (FRET) (Piston and Kremers, 2007).

1.5.3 IN-SILICO BASED APPROACH

Data generated in in-vivo and in-vitro approaches have allowed the development of tools that aid in the detection of protein-protein-interactions. The data created from the previous 2 approaches lacks the information of possible PPIs. In order to understand the context of potential interactions, there are methods able to analyse a wide range of possible interactions between proteins, as necessary (Valencia and Pazos, 2002).

In response to this, in-silico based approaches were developed to support interactions that were explored experimentally. These include structure-based prediction approach, Sequence-based Prediction Approach, Domain-Pairs-based Approach, In-silico Two Hybrid, among others.

1.5.4 STRUCTURE-BASED PREDICTION APPROACH

Structure based prediction approaches are aimed to predict protein-protein interactions, if two proteins have similar structure. If two proteins, A, and B, are known interactors, then there might be two other proteins, A' and B', with similar structures. Due to their similarity in structure, it can be possible that A' and B interact and that A and B' interact. The Multimeric Threading Approach has been used by Lu et al., 2002, in order to generate over 2800 protein-protein interactions in yeast, and over 1100 have been confirmed in DIP.

There is another algorithm, which is used by Hosur et al., 2012: the Coev2Net algorithm. This is a three-step process that involves: prediction of binding interfaces, evaluation of compatibility of the interface with an interface co-evolution based model, and an evaluation of the confidence score for interaction.

1.5.5 SEQUENCE-BASED PREDICTION APPROACH

The principle of the sequence-based prediction approach states that an interaction found in one species can infer interaction in another species. Hosur et al., 2011 developed a new algorithm that used a threading-based approach, which takes sequences as input. Interface Weighted RAPtor (IWARP), which

predicts whether two proteins interact by combining a novel linear programming approach (Hosur et al., 2011).

1.5.6 DOMAIN-PAIRS-BASED APPROACH

Domains are distinct, compact, and stable protein structural units that fold independently. There also tend to be distinct regions of protein sequence, with high conservation tendency throughout evolution. Since they are structural and functional units, they play a role in the development of class prediction, location prediction, and membrane-type prediction.

There is evidence that domain-domain interactions are more reliable than their corresponding PPIs (Memiševic et al., 2013). Due to this the use of domains as predictors of PPIs is consider a reliable tool.

1.5.7 HOT-SPOT-BASED APPROACH

Hotspots are a group of residues that contribute to the majority of the binding free energy. Classical recognition of Hotspot is done by “testing”, which mutates the residues to an alanine, and if the energy difference is equal or more than 2kCal/mol, it is called a Hotspot.

As mentioned before, not all amino acids contribute equally to the binding in PPI. Bogan and Thorn, 1998 stated that there are 3 amino acids that seem to heavily involve: Try, Arg and Trp. They also determined that the surrounded amino acids played an important role, such as preventing water binding.

1.6 Project Outline

The methods have been unsuccessful, therefore we proceeded to use a new method, molecular mimicry. Molecular mimicry are similarities shared between macromolecules found on pathogens and in host tissues; these similarities can be structural, functional, or immunological.

In order to identify possible new functions from modelled or identified cases of molecular mimicry the role of such macromolecules requires to be known; both clinical and molecular information on dengue were merged to provide a clear understanding of the possible findings.

1.6.1 HYPOTHESIS AND AIM

Dengue is a neglected disease, which affects populations with low resources; so, there is a need to find an inexpensive method to provide new possible functions or molecular pathways for a better understanding of how dengue affects the host during infection. This could bring new possibilities for drug targeting as well.

The Computational Method brings a powerful and inexpensive way to bring new potential targets and interactions which have been previously unknown. A repository of possible new interactions could benefit and guide researches like the PrePPI. This repository stores new possible functions between yeast and human proteins.

No analysis of mimicry found through in-silico methods have been done on the literature. Previous work has only focused on whole sequence similarity or Genome-Wide studies.

Sequence-only based analysis has several drawbacks. The first one is that they do not detect remote homology. Detection is rare if the sequences diverge greatly. Also, a whole-sequence similarity approach could not detect homology between protein domains. Respect to mimicry, it is known that for dengue mimicry, protein sequence with host proteins has almost no evolutionary linkage. Meanwhile, structure mimicry is the primary way for mimic protein to function through co-evolution between the invader and the host.

The purpose of this project is to search for mimicry through the use of combination with in-silico methods (a combination of domain homology, Hotspot analysis with focus on combined sequence/structure similarity methods), in order to find new possible interactions for known dengue proteins.

1.6.2 OBJECTIVES

The objectives of this project are the following:

- To collect data about dengue protein structures from publicly available databases.
- To associate domain homology found in dengue protein with proteins from other species.

- To associate new interactions from publicly available databases by using sequence homology/structural homology and Hotspot homology.

CHAPTER 2: METHODS

2.1 Resources

2.1.1 DATABASES

Various databases have been used to develop the domain pipeline. Some databases have been consulted to obtain specific information (e.g. Obtain the UniProtCode). In most cases, specific files were downloaded from these databases and used as input files in the pipeline. All databases used, their application in the project and relevant references are summarized in Table 2.1.

Databases used in the project	
Database	Application
RCSB (Parasuraman, 2012)	Search known protein structures between Human protein (H1) and other organisms (Ox). It was also used to obtain the UniProt identifier of the proteins in the structure, as well as to determine whether the structure was between two proteins or an antigen and a protein.
PDBe (Velankar et al., 2010)	Search of known protein structures between Human Protein (H1) and another Human Protein (H2).
UniProt (Bairoch et al., 2005)	Determine whether the organism of the protein within the PDB for example: is it Human or not? It was also used to obtain the known protein sequences of the polyprotein of dengue virus serotypes.
CATH 3D (Orengo et al., 1997)	Search the known domains in individual proteins. It was also used to model possible domains for protein sequence of the polyprotein of dengue virus serotypes.

Table 2-1: Databases used in the project and their application (with relevant references)

This project made use of 4 main sources of data. In order to access the amino acid sequence of proteins, UniProt was used to obtain and identify the information required. In order to access the known structures of the identify proteins, both RCSB and PDBe was used; although they hold the same data, both were used. In order to obtain information about the domains modelled or identify on proteins CATH 3D was used.

A Research Collaboratory for Structural Bioinformatics (RCSB) of the Protein Data Bank was established in 1971 by Walter Hamilton (Parasuraman, 2012). It contained 7 structures. Currently, it is the unique worldwide repository for structure data. RCSB was used to obtain the protein data list, which contained the following information:

- TAXONOMY is Dengue virus group and TAXONOMY is just Homo sapiens (human)
- TAXONOMY is Viruses and TAXONOMY is Homo Sapiens (human)
- TAXONOMY is Archaea and TAXONOMY is Homo Sapiens (human)
- TAXONOMY is Bacteria (eubacteria) and TAXONOMY is Homo Sapiens (human)
- TAXONOMY is Eukaryota (eukaryotes) and TAXONOMY is Homo Sapiens (human)

The Protein Data Bank in Europe (PDBe) is the European resource for collection, organization and dissemination of data on biological macromolecular structures (Velankar et al., 2010). It belongs to the worldwide Protein Data Bank group, which in collaboration with RSCB PDB and EMDataBank, maintains and provides access to global repositories of macromolecular structures. This database was used to acquire the list of known Human-Human protein complexes.

The Universal Protein Resource (UniProt) is a comprehensive worldwide resource for protein sequence and annotation data (Bairoch et al., 2005). UniProt is product of the collaboration between the European Bioinformatics Institute (EMBL-EBI), the SIB Swiss Institute of Bioinformatics and the Protein Information Resources (PIR). The individual information (in text format) was used in a script, in order to determine if the protein belonged to the genus Homo Sapiens. It was also used to obtain the polyprotein protein sequence of the 4 dengue serotypes. Class, Architecture, Topology, Homologous Superfamily, Sequence Family, Orthologous Family, Like Domain, Identical Domain and Domain Counter (CATH SOLID) Database, is a hierarchical domain classification of protein structures in the Protein Data Bank (Orengo et al., 1997). It provides information on the structure and function of that protein, including known domains obtained with structural data or by modelling. CATH 3D was used to obtain known domains on the protein-protein interaction list, obtained through RSCB and PDBe. It was also used to analyse the protein sequence of dengue polyprotein in order to obtain possible domains.

2.1.2 SOFTWARE AND TOOLS

For this project, we used Python as programming language, with BioPython packages for analysis. We used ArchSchema for related Pfam domain architectures. For the visualization and further analysis of the sequence and structural superimposition Jalview was used. HotSpot Wizard was used for modelling hotspots of the proteins selected. MacPymol was used to structural superimposition (Table 2.2).

Tools used for the project

Tool	Application
Python (G., 1995)	Programming Language.
BioPython (Cock et al., 2009)	Python Tools for Computational Biology.
ArchSchema Database (Orengo et al., 1997)	Domain Architecture Mapping.
JalView (Waterhouse et al., 2009)	Visualisation of sequence alignments.
HotSpot Wizard (Pavelka et al., 2009)	Automatic Identification of Hotspot Sites Tool.
MacPymol (Schrodinger, 2015)	Visualisation of Molecular Models; distance measurements.

Table 2-2: Tools used for the project and their application (with relevant references).

Python is the programming language chosen for this project, due to great array of packages for analysing the data required as well as requiring limited amount of time for knowing it. ArchSchema was used to map the domain architecture of the targets identify. Jalview was used due to its ability to create several analyses within the program as well as been able to easily export the results from each process. HotSpot Wizard, it is a great tool for identification of hotspots. MacPymol, versatility and resources allow for easy visualization of the structural elements of the PDB files which contain the structural information of the target proteins.

2.1.2.1 Implementation

This part provides a detail of the mental framework used for data collection and data analysis for this Thesis. The First Section describes the Structural Approach used for this Thesis (Top-down prices). It describes how data was collected, stored, and filtered, as well as analysed. Section Two describes the genomic approach used for this Thesis (Bottom-up process). It describes the process through which the

protein sequence for dengue polyprotein of all serotypes was used to obtain possible domains and further association of these domains to possible functions.

2.2 Workflow

In this Thesis, 2 workflows were used. The first workflow searches domain similarities from Human-Human protein structures, Human-Bacteria protein structures, Human-Virus protein structures, Human-Archaea protein structures and Human Eukaryotes protein structures. An additional list was produced to have only Human-Dengue structures.

The second workflow is aimed to search domain similarities and associated functions in selected domain-modelled dengue-only structures, obtained from RCSB/PDBe.

2.2.1 THE STRUCTURAL APPROACH

The first part of the workflow is shown on Fig 2.1.

In Step 1, protein structure based on the established parameters is acquired. The parameters are: Human-Human, Human-Bacteria, Human-Virus, Human- Dengue and Human Eukaryote structures.

In Step 2, the data acquired in Step 1 is filtered, so that synthetic molecules and DNA bound molecules are filtered out from the results. At this moment, there is a data differentiation between antigen bound structures and protein-protein bound structures. The last process consists of extraction of the Unique UniProt codes found in the structures.

In Step 3, the data extracted and a series of data dictionaries are created. The following are the main parameters:

- Protein-Protein
- Protein-Antigen

The following sub-parameters were also used:

- Human-Human: PDB, UniProt and Uniprot, Domain

- Human-Bacteria: PDB, UniProt and Uniprot, Domain
- Human-Virus: PDB, UniProt and Uniprot, Domain
- Human-Bacteria: PDB, UniProt and Uniprot, Domain
- Human-Eukaryote: PDB, UniProt and Uniprot, Domain
- Human-Dengue: PDB, UniProt and Uniprot, Domain

The next section of data analysis is set forth on Fig 2.2

Step 1: The dictionaries created at the end of Fig 2.1 were used to do find domains with similar domains in other dictionaries.

Step 2: The data collected by the process will be filter by similar domains with dengue- human structures and stores in a list.

Step 3: Samples from similar domains associated with dengue will be taken for structural comparison analysis.

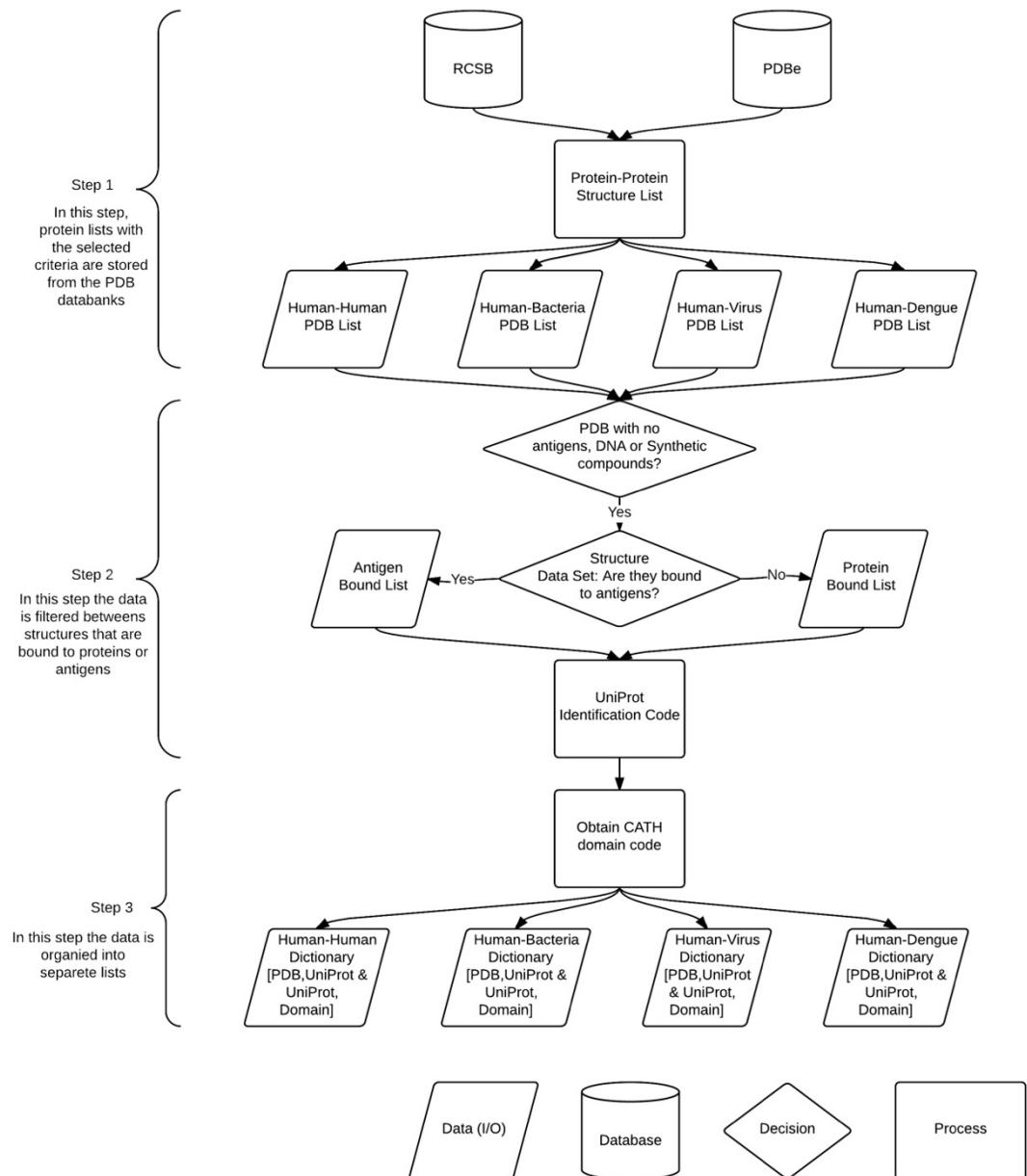


Figure 2-1 This is the first workflow used for the Structural approach

Step 1) required obtaining the information of the known structures from both Databases RCSB and PDBe for quality control and data verification; later the data collected was organized and stored in Lists which represents four desired protein interactions based on Human and Organism's Kingdom (Human-Human, Human-Bacteria, Human-Virus and, Human Dengue). Step 2) Starts by filtering the data collected in Step 1, to eliminate DNA-bound proteins, antigens and/or synthetic compounds. If the data collected has none of those characteristics, it is stored separately in different file; if bound to antigens, it is saved in a separate list. Then the UniProt IDs are obtained from the gathered protein-protein structures, as well as the CATH domain from the individual proteins and stored using the same organization categories as Step 1).

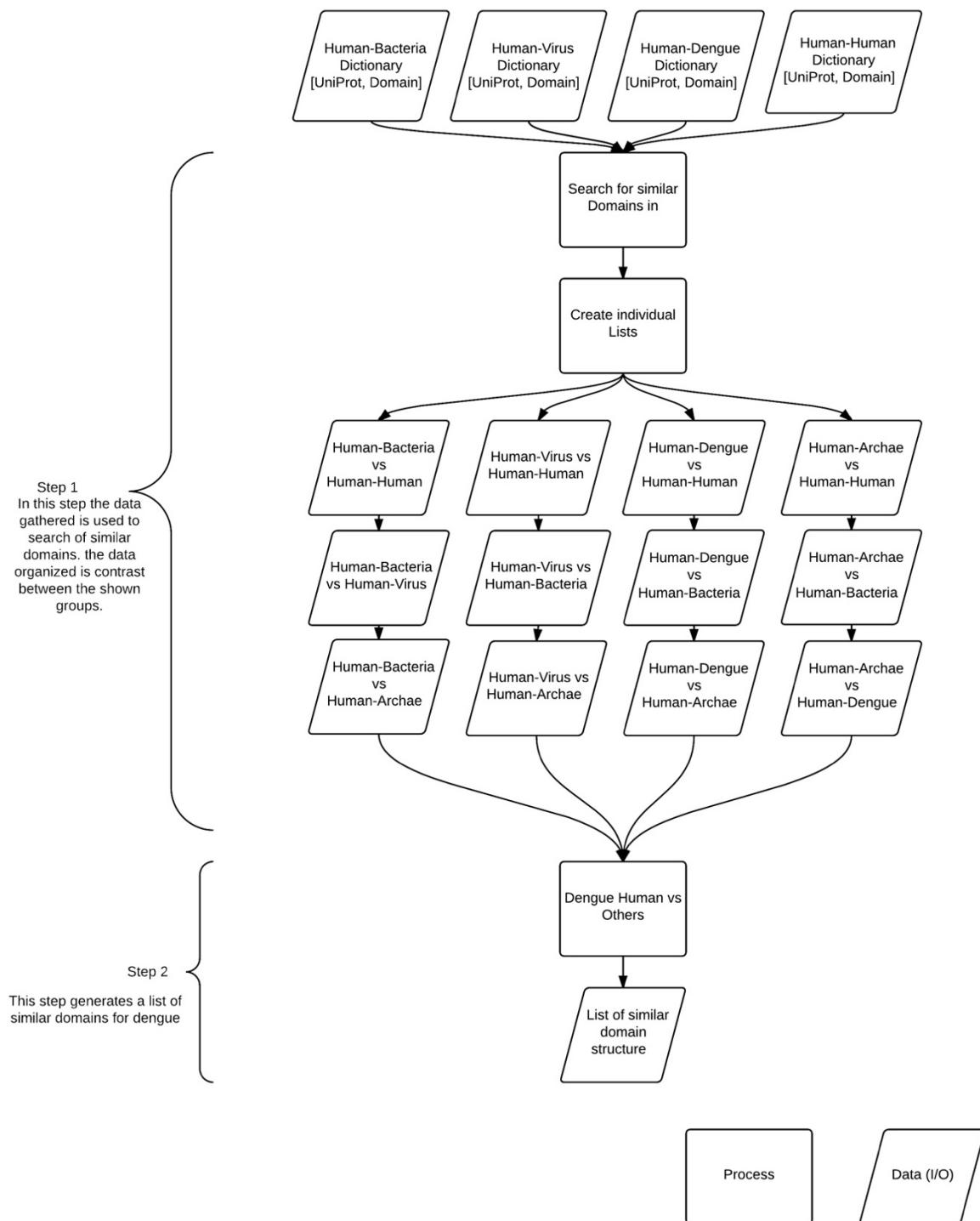


Figure 2-2 This is the second workflow used for the Sequence approach

Step 1) The data stored in the first workflow of the sequence approach would be used to tools to model possible new domains from each one of the proteins identified. Step 2) These data would be organized in different lists that represent any connection between Human vs Kingdom/Human/Dengue and Human vs Kingdom/Human/Dengue in order to establish a possible new role of a protein due to its identify interaction with another protein from another kingdom member. This data will later be organized in structured lists, so they can later be use for further analysis.

2.2.2 THE GENOMICS APPROACH

The first part of the workflow is shown in Fig 2.3:

- In Step 1, Uniprot database is used to extract the liner sequence of individual dengue proteins.
- In Step 2, the protein sequences are used to model domains using CATH Gene 3D.
- In Step 3, the domains modelled are store individually.

The second part of the workflow is shown in Fig 2.4:

- In Step 1, the PDB codes associated with each domain super-family were stored in lists.
- In Step 2, the PDB codes were filtered and only those found in PDB Sum were stored.
- In Step 3, UniProt code associated with the PDB codes was used to search for interaction in the IRIIndex Database.
- In Step 4, each protein was separated into Human, Dengue, Archaea, Bacteria, Virus and Non-human Eukaryote groups.

2.2.3 PYTHON SCRIPT

All python scripts created for this project are accessible at (Alarcon Aulestia, 2017)

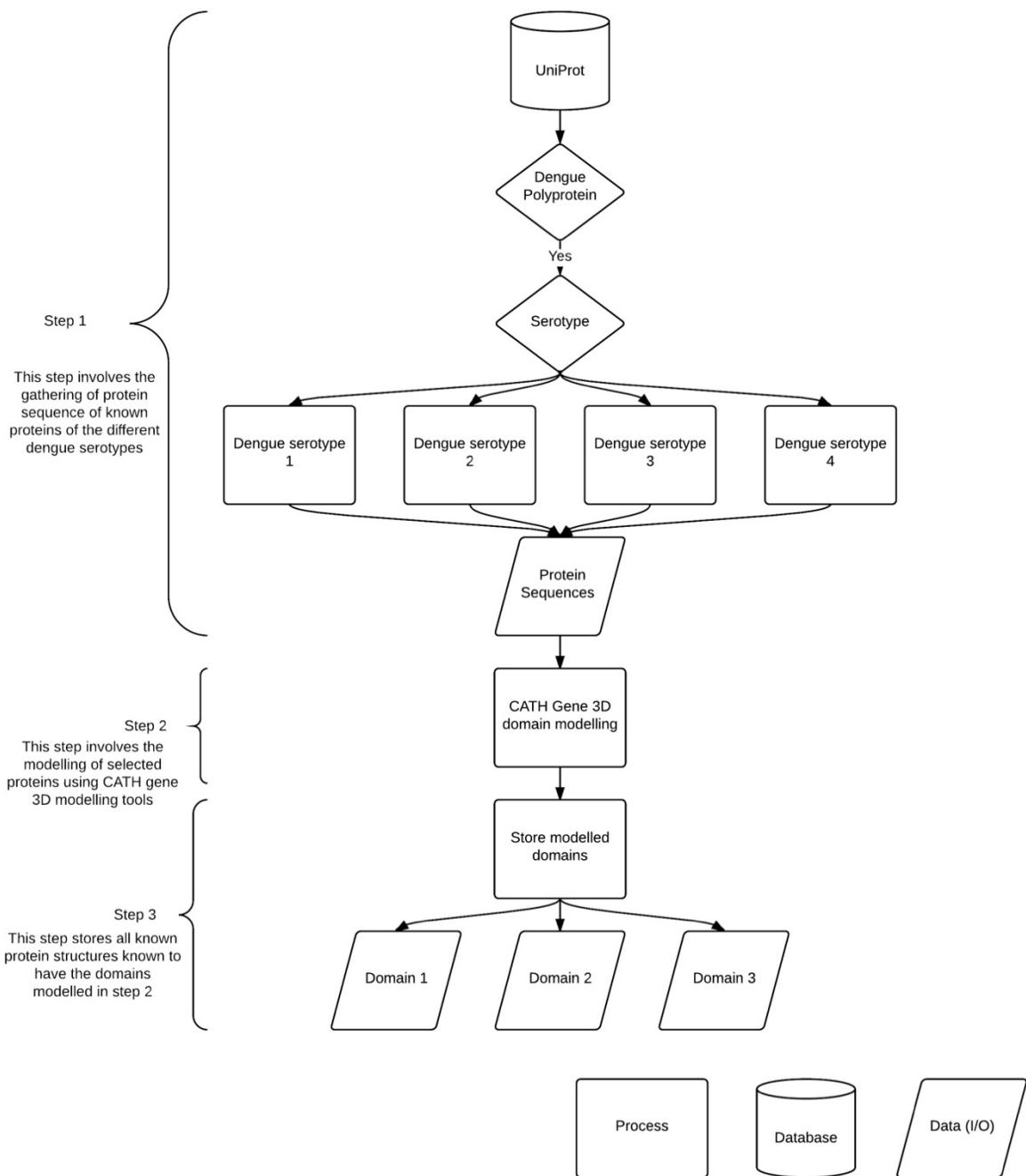


Figure 2-3 This is the first workflow used for the Sequence approach

The first phase of the genome approach starts by obtaining the polyprotein linear sequence of all four serotypes of dengue. Later the protein sequences are going to be process through CATH Gene 3D and later store by the domains identified.

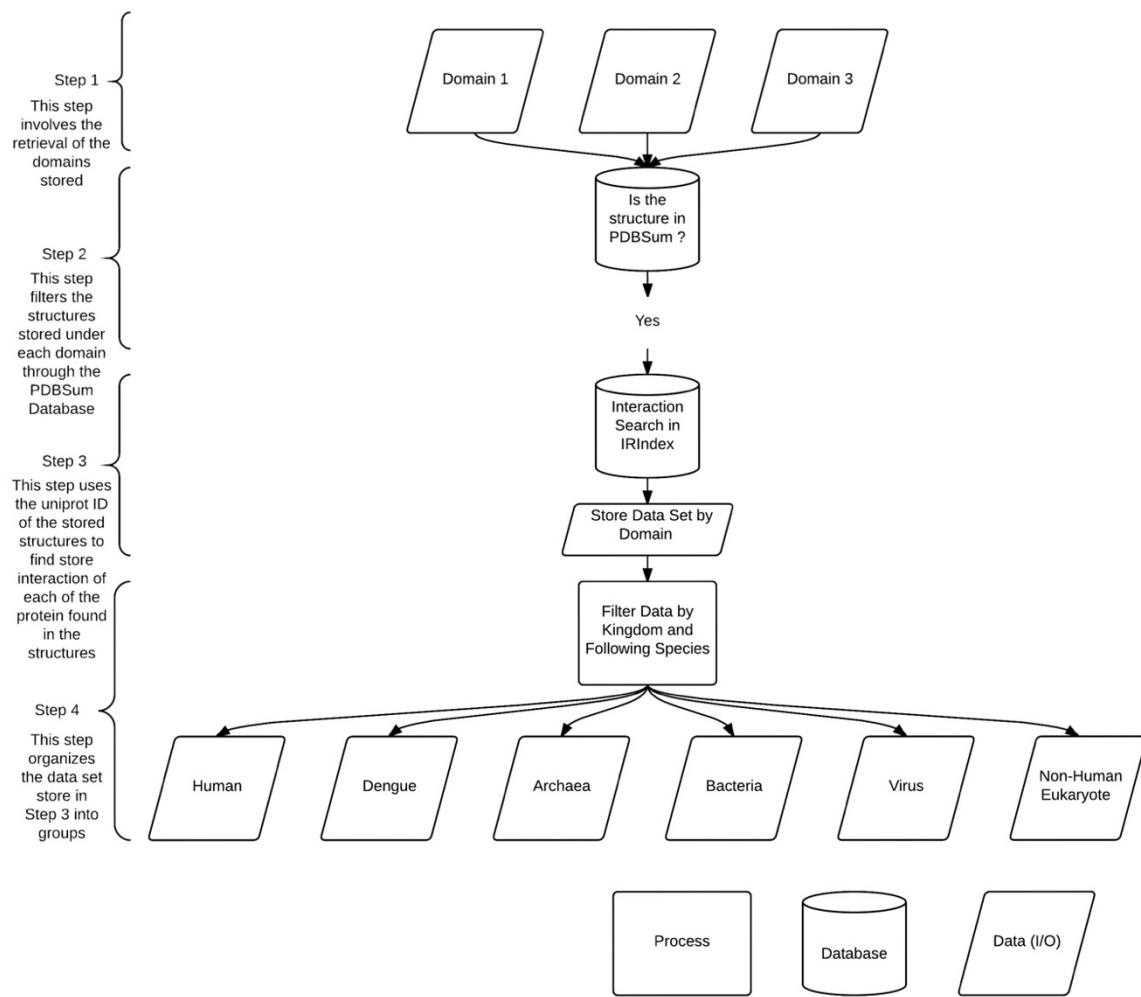


Figure 2-4 This is the second workflow used for the Sequence approach

The data in the list created from the first phase of the Genomic Approach is going to be further process by identifying whether there are known structures for those proteins with those domains (by the use of the PDBSum Database), and whatever there is an known interaction (by the use of IRIndex Database) . Later the proteins are organized into different lists, Human, Dengue, Archaea, Bacteria, Virus and Non-Human Eukaryote.

CHAPTER 3: RESULTS AND DISCUSSION: DATA COLLECTION AND PROCESSING

3.1 Firth Method: Structural Approach

3.1.1 STEP 1: OBTAINING A PROTEIN-PROTEIN STRUCTURE LIST

In this first step, the list of protein structures with the following criteria was created:

- Human-Dengue binding structures
- Human-Bacteria binding structures
- Human-Virus Binding Structures
- Human-Archaea Binding Structures
- Human-Human Binding Structures
- Human-Eukaryote Binding Structures

We have used the RCSB and PDBe databases for retrieving the Protein Data Bank code (PDB Codes). We decided to start the method by acquiring the already known structures with Human proteins interacting with proteins of other kingdoms. This was done because structural similarities yield a higher probability that two proteins share similar function.

The RCSB Data Bank holds 113816 biological macromolecular structures and alongside Protein Data Bank in Europe (PDBe), it belongs to the worldwide Protein Databank (WWPDB). The initial data was gathered from RCSB and PDBe.

The data gathered from RCSB and PDBe was saved in the order files mention in the Method Section, they hold a list of PDB IDs of all the protein codes that meet the criteria. These files contains PDB ID, Entity ID, Organism Scientific Name, Tax ID, Organism Synonyms, Rank (genus, subfamily, family, order, phylum, super-kingdom), Genus, Experimental Data Available, Title, Entry Authors, Sample Preparation

Method, Detector, Detector Type, Synchrotron Beamline, Beam Source Name, Diffraction Protocol, Structure Determination Method, Refinement Software, Space Group, Synchrotron Site, Structure Solution Software, Data Reduction Software, Data Scaling Software, Crystallization and Reservoir.

Search Results of Human Bound Structures	
Group	Structures Found
Archae	17
Bacteria	397
Virus	721
Human	9627
Dengue	14

Table 3-1: Search Results of Human Bound Structures

The results from the Initial Human Bound Structures are 17 Human-Archaea bound proteins, 397 Human-Bacteria bound proteins, 721 Human-virus bound proteins, 9627 Human-Human bound proteins and 14 Human-Dengue bound proteins

The PDB IDs were filtered, so that there were no repetitions. After applying it only 10776 were obtained, from which, 17 were Archaea, 397 were Bacteria, 721 were Virus, 9627 were Human and 14 were Dengue. (Table 3.1)

3.1.2 STEP 2: FILTERING AND ORGANIZING THE PDB ID LISTS

Archae		
Structures with non- Immunoglobulin	Structures with Immunoglobulin	Total proteins found
1	16	17
Bacteria		
Structures with non- Immunoglobulin	Structures with Immunoglobulin	Total proteins found
11	16	27
Virus		
Structures with non- Immunoglobulin	Structures with Immunoglobulin	Total proteins found
289	432	721
Human		
Structures with non- Immunoglobulin	Structures with Immunoglobulin	Total proteins found
7757	2070	9827
Dengue		
Structures with non- Immunoglobulin	Structures with Immunoglobulin	Total proteins found
0	14	14

Table 3-2: Results of structures for Archae, Bacteria, Virus Human and Dengue

After filtering the data collected previously to identify if the protein-protein structure were between non-Immunoglobin and Immunoglobin, the results are: Archae, 1 non-immunoglobin, 16 immunoglobin; Bacteria, 11 non-immunglobin, 16 immunoglobin; Virus, 289 non-immunoglobin, 432 immunoglobin; Human, 7757 non-immunoglobin, 2070 immunoglobin; dengue 0 non-immunoglobin, 14 immunoglobin.

The lists were divided into 2 different dictionaries, those that were bound to antigens and those that were not. The results showed that of the 17 Human-Archaea structures, 1 is bound to an antigen; of the 397

bacteria structures, 82 were bound to antigens; of the 721 Human-virus structures, 432 were bound to antigens; of the 9827 Human-Human structures, 2070 were bound to antigens, and from the 14 dengue structures, 14 were bound to antigens Table 3.2.

The lists were separated into kingdoms, in order to organize and analyse the data individually. Also, for further comparison analysis, it is more convenient to have the data independently stored.

Proteins with associated domains		
Group	PPI	PAI
Archae	14	2
Bacteria	310	79
Dengue	0	16
Virus	334	280
Human	3358	973

Table 3-3: Protein structures with associated domains found for each individual

The results for the identified associated domains between PPI list and PAI list are as follow: Archae, 14 PPI, 2 PAI; Bacteria, 310 PPI, 79 PAI; Dengue, 0 PPI, 16 PAI; Virus, 334 PPI, PAI 280; Human, 3358 PPI, 973 PAI.

3.1.3 STEP 3: ACQUIRING DOMAINS

The CATH domains were associated to targets in all the lists using the file containing the database of ArchSchema. These file was downloaded from the CATH Gene 3D website, and the file contains 37719526 domains associations totally. After data filtration, there were 14 Archaea Protein-Protein interaction structures (PPis) out of the 17 which had associated domains in the ArchSchema file and 1 Protein Antigen interaction structure (PAIs) out of the 2. Bacteria had 310 PPis with associated domains out of 397 and 79 PAIs out of 82. Dengue had 0 PPis out of 0 and 16 out 14 PAIs. Virus had 334 PPis out of 721 and 280 out of 432 PAIs. Human had 3358 out of 9827 PPis and 973 out of 2070 PAIs (Table 3.3).

The increase in number of PAIs with a domain association, in comparison with the number of PAIs originally stored, could be explained by more than one domain associated with that PAIs, or a repeat domain in a different section of the protein sequence.

3.1.4 STEP 4: MERGING SIMILAR DOMAINS AND PROTEINS

This Step consists of finding similar domain and proteins among the Human-Other (Bacteria, Virus, Dengue, Archaea) and the Human-Human lists.

In an association of proteins, there is a Protein A and a Protein B that interact with each other. Protein A interacts with Protein B through Domain A and Domain B. Same thing occurs with Other A and Human A', which interact through Domain A and Domain B. So, if a Human A' is found in the Human A Interaction and Human B Interaction, it possibly means that Other can interact with the same protein as Human A.

Respect to the final results from the Structural Approach, no Dengue proteins were stored in the PPI. All Dengue-Human structures stored in PDB are bound to antigens. Due to the nature of Antigen binding, the domains and interactions in these structures are not to be used for modelling possible new functions.

Lamentably the results obtain through this method did not give any usable results for further analysis for molecular mimicry. Due to this failure in finding suitable interactions the second method, a more sequence approach, was used.

3.2 Second Method: Sequence Approach

3.2.1 STEP 1: RETRIEVING DENGUE PROTEIN SEQUENCE

Due to lack of data obtained from the Structural Approach, the Sequence Approach was used.

As mentioned before, RCSB is a Protein Data Bank that has stored 113816 biological macromolecular structures. Specific dengue proteins were selected with the search under the parameter 'Dengue'. These were the selected protein structures:

- 3l6p: Crystal Structure of Dengue Virus 1 NS2B/NS3 protease
- 1l9k: Crystal Structure of Dengue Virus 2 methyltransferase
- 1tg8: The structure of Dengue virus 2 E glycoprotein
- 1oan: Crystal Structure if Dengue 2 Virus Envelop Protein
- 1uzg: Crystal Structure of the Dengue virus 3 Envelop Protein

- 3we1: Crystal structure of Dengue 4 Envelope protein domain III near-atomic resolution cryo-EM structure of Dengue serotype 4 virus
- 4cct: Dengue 1 cryo-EM reconstruction
- 4al8: Structure of Dengue virus DIII in complex with Fab 2H12
- 2bhr: Dengue Virus RNA Helicase
- 2fom: Dengue Virus NS2B/NS3 Protease

The selected group of protein structures used is not PH dependent and belong to several stereotypes of dengue. Therefore, the different stereotypes of Dengue will be represented in the sample.

Dengue Domains Modelled		
PDB	Domain	Times
3l6p	2.40.10.120	1
	2.40.10.10	1
1l9k	3.40.50.150	1
1oanA/B	2.60.98.10	3
	3.30.387.10	1
	3.30.67.10	1
	2.60.40.350	1
	2.60.98.10	3
1uzgA/B	3.30.387.10	1
	3.30.67.10	1
	2.60.40.350	1
	2.60.40.350	3
3we1B	2.60.40.350	3
4cbfA/B	2.60.98.10	4
	3.30.387.10	1
	3.30.67.10	1
	2.60.40.350	1
4al8L/H/C	2.60.40.10	2
	2.60.40.350	1
2bhrA/B	3.40.50.300	2
2fomB	2.40.10.120	1
	2.40.10.10	1

Table 3-4: Dengue Domains Modelled

The chosen structures, 3l6p, 1l9k, 1oanA/B, 1uzgA/B, 3we1B, 4cbfA/B, 4al8L/H/C, 2bhrA/B and 2fomB had their domains modelled through GENE 3D CATH. 3l6p had 2 domains, of which both only were repeated once. 1l9k had 1 domain with only one repetitions. 1oanA/B had 4 individual domains identified with a total of 6 domains found. 1uzgA/B had 4 individual domains identified with a total of 6 domains found. 3we1B had 1 domain identified, however it was repeated three times. 4cbfA/B had 4 individual domains identified with a total of 7 domains found. 4al8L/H/C had 2 individual domains identified and 3 domains found. 4bhrA/B had only one domain identified, and it repeated twice. 2fomB had 2 single domains identified with a single repetition found.

3.2.2 STEP 2: MODELLING DOMAINS

GENE 3D CATH website offers modelling of domains and uses an array of options. The PFam option allows you to copy the protein sequence of the desired protein, and the outcome will be the possible domains. It also includes the hit sequence. All 11 proteins were used in the modelling procedure, and a total of 9 domains were modelled (Table 3.4).

3.2.3 STEP 3: PROTEINS-DOMAIN ASSOCIATION

Using the data file from Gene 3D Cath database, all associated proteins with each of the domains were extracted. A total of 2210360 structures were found for the 9 domains, from which 3.30.67.10 had 11146 structures, 2.60.98.10 and 13863 structures; 2.60.40.10 had 193912 structures; 2.40.10.120 had 18048 structures; 3.30.387.10 had 10861 structures; 3.40.50.300 had 1563155 structures; 2.40.10.10 had 71722 structures; 2.60.40.350 had 11960 structures; 3.40.50.150 had 315693 structures (Table 3.5)

Structures Associated with the modelled domains	
Domain	Proteins Stored
3.30.67.10	11146
2.60.98.10	13863
2.60.40.10	193912
2.40.10.120	18048
3.30.387.10	10861
2.40.50.300	1563155
2.40.10.10	71722
2.60.40.350	11960
3.40.50.150	315693

Table 3-5: Number of structures associated with the modelled domains=

The nine identified domains from 9 selected dengue proteins had stored the following amount of protein structures associated with them: 3.30.67.10, 11146 proteins; 2.60.98.10, 13863 proteins; 2.60.40.10, 193912 proteins; 2.40.10.120, 18048 proteins; 3.30.387.10, 10861 proteins; 2.40.50.300, 1563155 proteins; 2.40.10.10, 71722 proteins; 2.60.40.350, 11960 proteins; 3.40.50.150, 315693 proteins.

Proteins with known interactions	
Domain	Proteins Stored
3.30.67.10	19
2.60.98.10	19
2.60.40.10	2470
2.40.10.120	192
3.30.387.10	19
3.40.50.300	2511
2.40.10.10	1508
2.60.40.350	19
3.40.50.150	184

Table 3-6: Proteins with known interactions

The data obtained in Table 3.5 was later further processed to discern whether they had known interactions already identified: 3.30.67.10, 19 interactions; 2.60.98.10, 19 interactions; 2.60.40.10, 2470 interactions; 2.40.10.120, 192 interactions; 3.30.387.10, 19 interactions; 2.40.50.300, 2511 interactions; 2.40.10.10, 1508 interactions; 2.60.40.350, 19 interactions; 3.40.50.150, 184 interactions.

3.2.4 STEP 4: FILTERING DATA THROUGH PDBSUM

PDBsum is a pictorial database that gives an overview of the contents of 3D structures stored in the Protein Data Bank. To assure that the proteins used for analysis have a structure associated with them, the proteins data extracted was filtered through ‘seqdata.dat’ from the protein sequence annotations file of PDBSum. After the data was filtered, the results were the same as in Step 3.

3.2.5 STEP 5: MERGING IRINDEX DATABASE

IRefIndex is a database that provides an index of protein interactions available through databases like Bind, BioGRID, CORUM, DIP, HPRD, InnateDB, IntAct, MatrixDB, MINT, MPact, MPIDB and MPPI. ‘All.mitab.04072015.txt’ file from IRefIndex was downloaded and it holds a total of 1425070 interactions store in it.

The data stored in Step 4 was used to look for proteins that have interactions in the IRefIndex Database. After filtering the data through the IRefIndex Database, domain 3.30.67.10 had 19 proteins with interactions; domain 2.60.98.10 had 19 proteins with interactions; domain 2.60.40.10 had 2470 proteins with interactions; domain 2.40.10.120 had 192 proteins with interaction; domain 3.30.387.10 had 19 proteins with interaction; domain 3.40.50.300 had 2511 proteins with interaction; domain 2.40.10.10 had 1508 proteins with interaction; domain 2.60.40.350 had 19 proteins with interaction; domain 3.40.50.150 had 184 proteins with interaction (Table 3.6).

3.2.6 STEP 6: CLASSIFICATION OF DATA

The data group obtained was filtered and organized by using a script that stored the data into the following groups:

- Human
- Virus
- Bacteria
- Non-Human Eukaryotes
- Archaea

The script also avoided repeated proteins to be present in the organization. If the protein was repeated, but had a different domain associated with it, the domain was stored under the same protein.

The results showed a total of 825; human had a total of 417 proteins; virus had a total of 18 proteins; bacteria had a total of 128 proteins; non-human eukaryote had a total of 254 proteins, and archaea had a total of 8 proteins (Table 3.7).

CHAPTER 4: RESULTS AND DISCUSSION: FUNCTIONAL CANDIDATES

4.1 Sample Selection

The result of the sequence approach yields a total of 825 proteins. These proteins were registered in tables with the following data:

- UniProtM: UniProt Code from the obtained protein.
- UniProtI: UniProt Code of the interactor protein. obtained from the `verb|All.mitab.04072015.txt|` file.
- Domains: Found for the protein.
- pdbM: PDB of the UniProt (Main).
- pdbI: PDB of the Interactor.
- Complex: Complex code for this particular interaction.
- IntType: Interaction Type.
- SpeM: Species of UniProt (Main).
- TaxIDM: Taxological ID of UniProtM.
- Spel: Species of UniProtI.
- TaxIDI: Taxological ID of UniProtI.
- BioRoleM: Biological role of UniProtM.
- BioRoleI: UniProtI's Biological Role.
- Namel: UniProtI's Name
- PUBMED: Pubmed ID of paper describing the interaction.

The data was organized in this manner, in order to discern possible samples for further analysis.

Starting with the Interaction Type, there are 103 interaction types stored in the database and used by iRefIndex Database (in the 'All.mitab.04072015.txt')(Razick et al., 2008); among them: Association,

Physical Association, Direct Interaction, Covalent Bonding, Disulphide Bond, Transglutamination Reaction, Enzymatic Reaction, Acetylation Reaction, ADP Ribosylation Reaction, Amidation Reaction. Of all interaction types, these are the ones selected to filter the sample data:

- Association
- Physical Association
- Direct Interaction
- Covalent Bonding

Another factor was the SpeInt or Species of the Interactor. In this particular case, we were looking for samples that had a known interaction with Human Proteins.

Another factor of selection was the BioRoleInt or the Biological Role of the Interactor. This means, Interaction in known molecular pathways or responsible or known to play a role in viral infection.

Based on these criteria, the following proteins were selected:

P06935: West Nile virus Polyprotein

4.2 P06935: West Nile virus Polyprotein

4.2.1 INTRODUCTION TO WEST NILE VIRUS

West Nile Virus (WNV) has a bird-mosquito-bird transmission cycle. The main mosquito species to be transmitted is the Culex mosquito. West Nile Virus is part of the 70 viruses that belong to the family Flaviviridae. It is member of the Japanese encephalitis sero - complex. It also has 5 phylogenetic lineages, being lineage 1 and 2 associated with outbreaks in human.

The WNV is a positive single-stranded RNA virus, with no 3'Poly(A) tract. The genome encodes 10 mature proteins; 3 structural proteins, Capsid, Membrane, enveloped and encoded within the 5' section of the genomic ORF. Other 7 non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS5) are encoded in the 3' section.

4.2.2 P06935: INTERACTION INFORMATION

WNV Poly protein (P06935) interacts with Integrin beta-3 (P05106). According to Sharma-Walia et al., 2004, Integrin Alpha-3 Beta-1 interacts with Envelop Protein Domain III. This is through Physical Interaction.

There are 2 known structures of the WNV E protein domain III: Crystal Structure Analysis of the West Nile virus envelope (E) protein domain III (2p5p), NMR Structure of Domain III of the West Nile Virus Envelope Protein, Strain 385-99 (1s6n). 1s6N was selected for further analysis (Fig 4.2). Modelling by Gene 3D Cath webtool suggested that the domain 2.60.40.350 was found in 1s6n; this domain occupies almost the entirety of the structure.^{ooo}

Integrin beta 3 are known to be responsible for entry of several type of viruses, such as Hanta viruses (Gavrilovskaya et al., 1998), Western Nile Virus (Chu and Ng, 2004), parechovirus 1 (Joki-Korpela et al., 2001), cytomegalovirus (Feire et al., 2004), herpesvirus (Nemerow and Cheresh, 2002).

4.2.3 ANALYSIS: SEQUENCE ALIGNMENT

The first analysis carried out was ClustalWS alignment. The proteins were aligned in the sequence in which the domain 2.50.40.350 is found in 1s6n. As seen from Fig 4.2, there is a high incidence of conservation in the amino acid sequence. This is supported by the conservation scores of the amino acids and the consensus, as shown in Fig 4.1.

This shows that the sequence of the proteins is quite similar, giving a high probably of same domains, function, and structure.

4.2.4 ANALYSIS: STRUCTURAL ALIGNMENT

The next step is aligning the structures. Individual structures were acquired from PDB, from the structures that were modelled with 2.60.40.350 domain:

- 1oan (Fig 4.3)
- 1uzg (Fig 4.4)

- 3we1 (Fig 4.5)
- 4al8 (Fig 4.6)
- 4cbf DF chain (Fig 4.7)

After acquiring the individual structures for each of the proteins, they were structurally superimposed by using MacPymol. For the 1s6n:1oan superimposition, 105 atoms were aligned, and a Root Mean Square Deviation (RMSD) value of 2.069 Å was calculated for 87 of 87 atoms (Fig 4.8, Fig 4.12A). For 1s6n:1uzg super imposed structure, 105 atoms were aligned, and a RMSD value of 2.238 Å was calculated for 89 of 89 atoms (Fig 4.9, Fig. 4.12B). For 1s6n:3we1 superimposition, 96 atoms were aligned, and a RMSD value of 2.309 Å was calculated for 87 of 87 atoms (Fig 4.10, Fig 4.12C). For 1s6n:4al8 structure superimposition, 98 atoms were aligned, a RMSD value of 2.632 Å was calculated for 93 of 93 atoms (Fig 4.11, Fig 4.12D). For 1s6n:4cct structural superimposition, 104 atoms were aligned, a RMSD value of 2.603 Å was calculated for 90 of 90 atoms (Fig 4.12E). For 1s6n:4cbf F Chain superimposition, 22 atoms were aligned, a RMSD value of 9.246 Å was calculated for 22 of 22 atoms (Fig4.12F). For 1s6n:4cbfA chain structural superimposition, 107 atoms were aligned, a RMSD Value of 2.433 Å was calculated for 91 of 91 atoms (Fig 4.12G).

Maiorov and Crippen, 1994a stated that when analysing the fold or structure of super imposed structure, the value of RMSD is an important value. If RMSD value of superimposition between 2 structures is close to 0. This means that both structures are very identical, if not similar. It is not clear which value could be considered the minimum required for identifying two identical proteins (Maiorov and Crippen, 1994a; Maiorov and Crippen, 1994b).

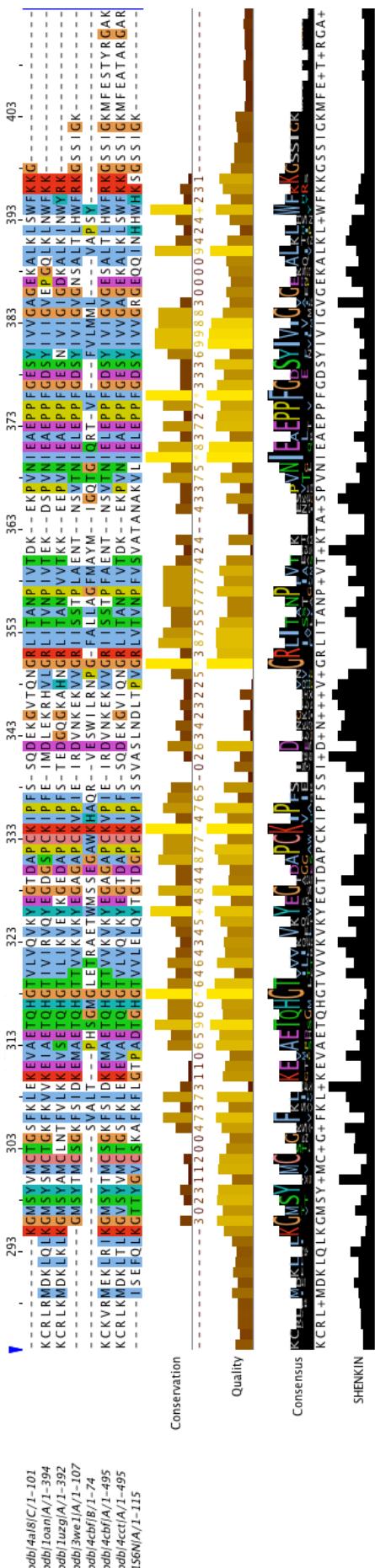


Figure 4-1 Alignment

This image shows the alignment done through JalView of the obtain structures, 4a18/C, 1oan/A, 1juze/A, 3we1/A, 4bcf/B, 4ctt/A and 1s6n/A.

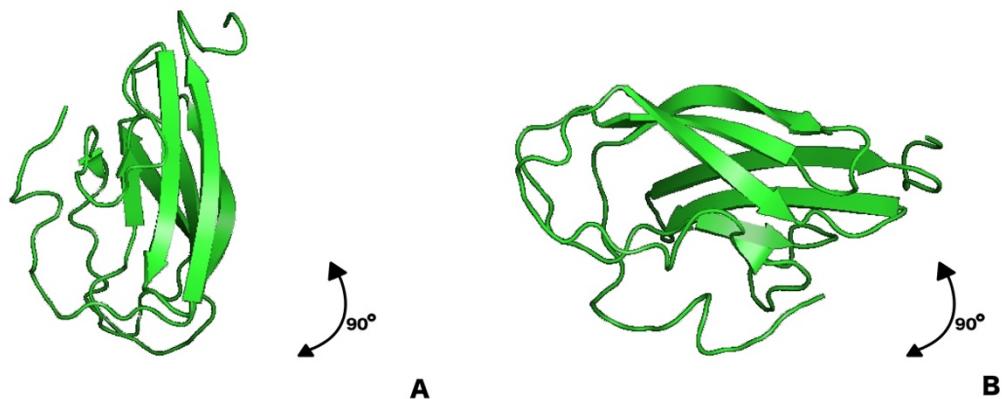


Figure 4-2 Structure of NMR Structure of Domain III of the West Nile Virus Envelope Protein, Strain 385-99 (1s6n)

A) Frontal view corresponding to 1s6n B) Rotate view corresponding to 1s6n

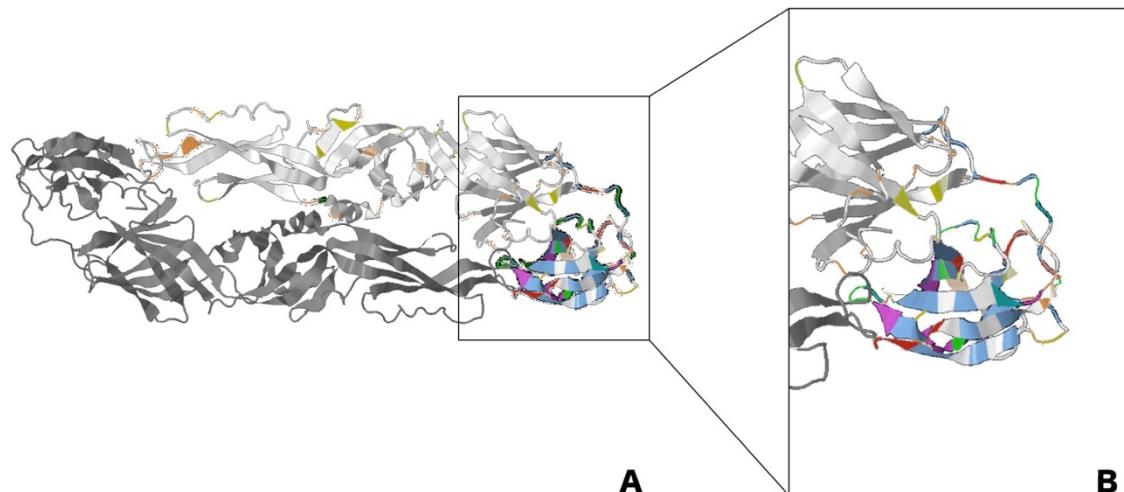


Figure 4-3 Structure of protein 1oan: Crystal Structure of the Dengue Type 2 Virus Envelope Protein

A) Corresponds to a frontal side view of 1oan. B) It is the region that is similar to the modelled 1s6n.

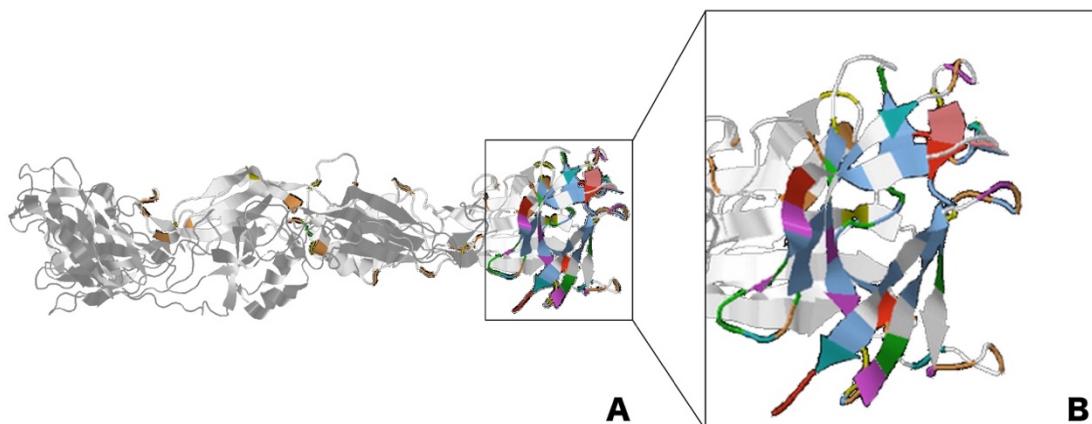


Figure 4-4 Structure of protein 1uzg: Crystal Structure of the Dengue Type 3 Virus Envelope Protein

A) These a frontal view of the protein structure 1uzg. B) This is the area modelled to correspond to the structure of 1s6n

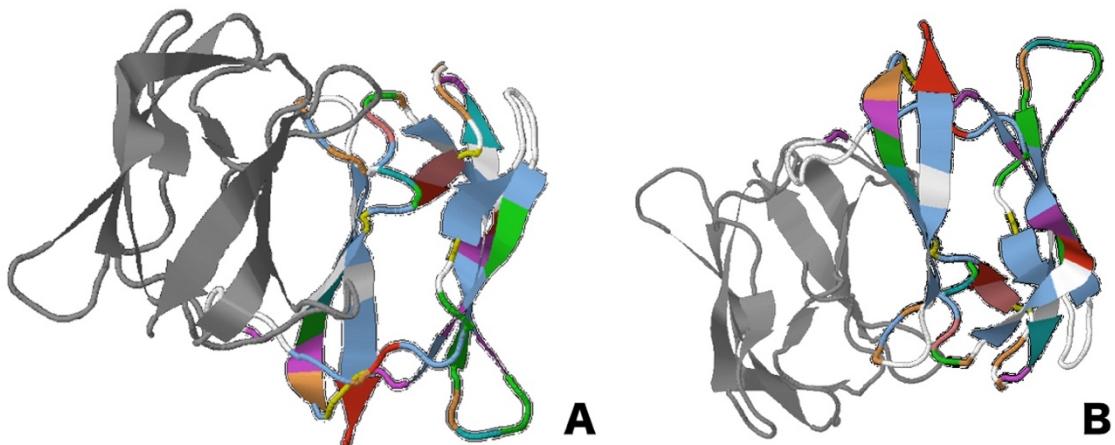


Figure 4-5 Structure of protein 3we1: Crystal structure of Dengue 4 Envelope protein domain III

A) The left side corresponding to 3we1 B) The side corresponding to a 180 degree rotation in the y axis

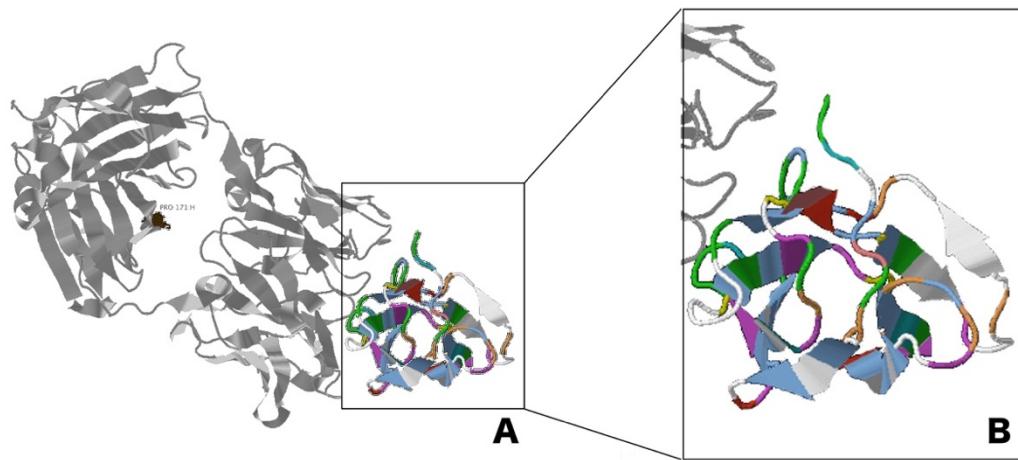


Figure 4-6 Structure of protein 4al8: Structure of Dengue virus DIII in complex with Fab 2H12

A) the left side corresponding to 4al8 B) This is the area modelled to correspond to the structure of 1s6n

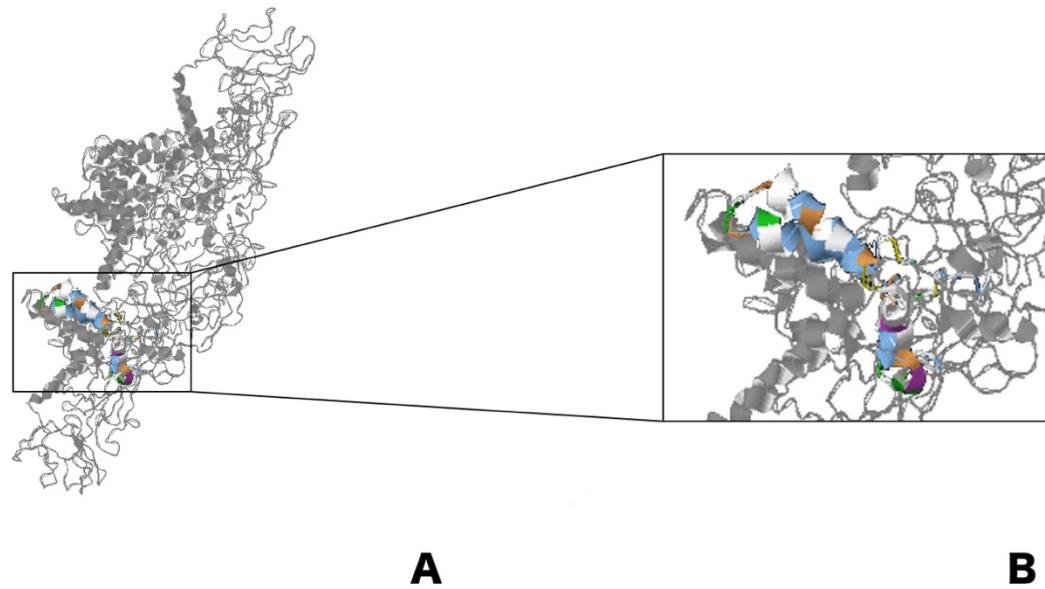


Figure 4-7 Structure of protein 4cbf DF chain: Near-atomic resolution cryo-EM structure of Dengue serotype 4 virus

A) This is 4cbf DF chain frontal view. B) This is the area modelled to correspond to the structure of 1s6n

From the values gathered, excluding the RMSD 9.246 Å for 4cbf F Chain, because the 22-atom alignment is too low in comparison to the average of over 100 atoms aligned in the other superimpositions, the RMSD mean value is 2.38067 (SD 0.21816). These values are close to 0 indicate that the structure of domain 2.60.40.350 of 2s6n and modelled structures in the Dengue proteins are very similar.

4.2.5 HOTSPOT ANALYSIS

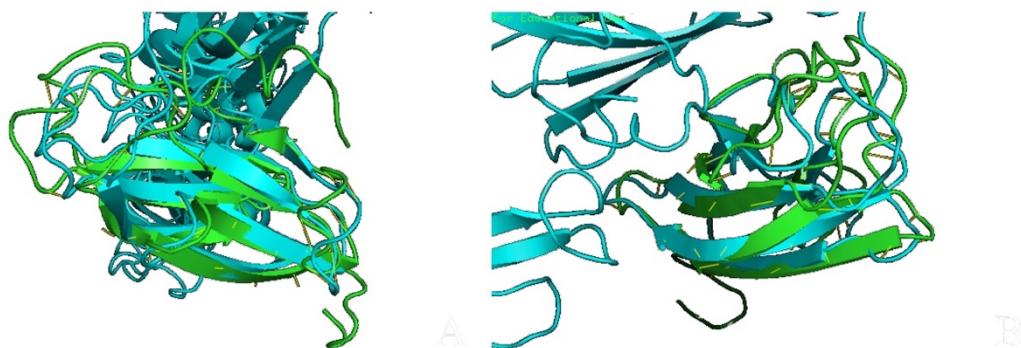


Figure 4-8 Structure superposition between 1s6n (green) and 1oan (turquoise)

A) The superimposition of both structures is shown from a frontal orientation. B) The superimposition of both structures is shown from a downward orientation.

Hotspot analysis was done through Predict Protein (reference). It uses PROFsis, which is a machine learning-based method that identifies residues from protein sequence. It uses predicted structural features with evolutionary information to get the strongest possible prediction.

By using the sequence alignment, the hotspots were highlighted in black squares in Fig. 4.13. Looking at these results, there are 3 sets of hotspot analyses that are conserved throughout all structures: The Glutamic Acid (E), Threonine (T), Glutamine (Q) group, the Phenylalanine (F), Threonine (T), Glutamine (Q) group and Aspartic Acid (D), Threonine (T), Glycine (G) group; the TDG hotspot column for 1S6N and

the Serine (S), Leucine (L), Asparagine (N), Aspartic Acid (D), Leucine (L), Threonine (T) and Proline (P) hotspot column.

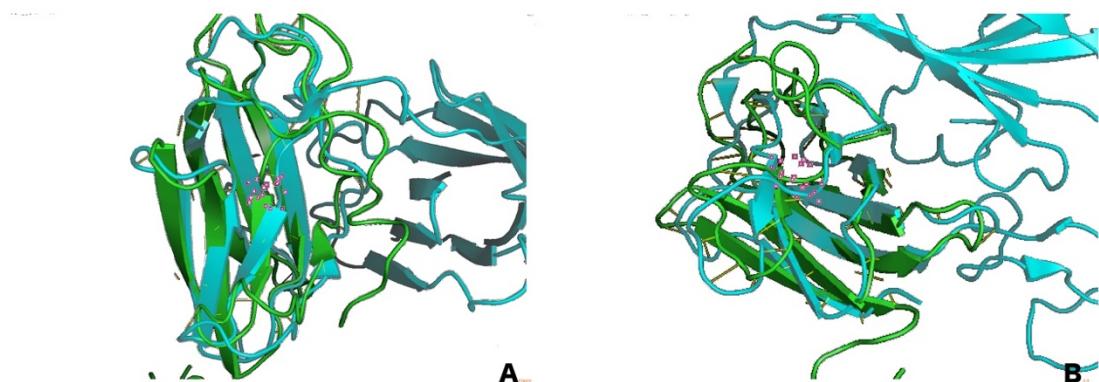


Figure 4-9 Structure superposition between 1s6n (green) and 1uzg (turquoise)

A) The superimposition of both structures is shown from a frontal orientation. B) The superimposition of both structures is shown from a downward orientation.

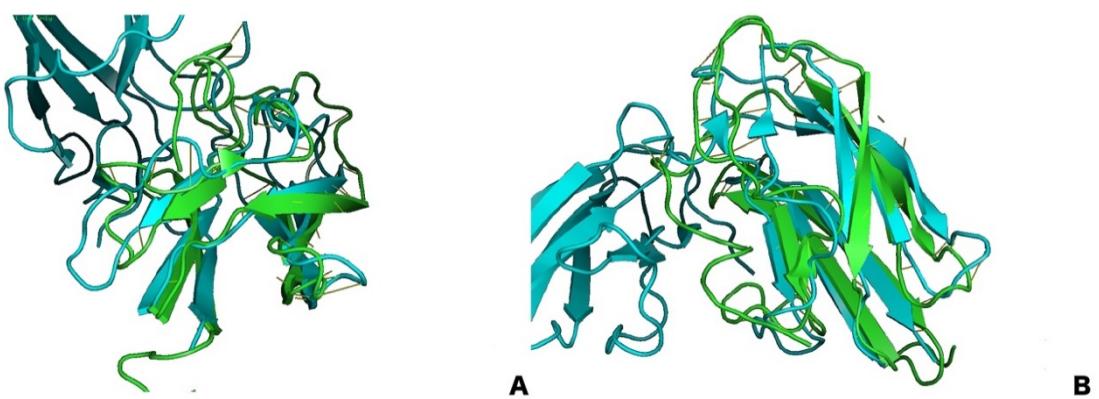


Figure 4-10 Structure superposition between 1s6n (green) and 3we1 (turquoise)

A) The superimposition of both structures is shown from a frontal orientation. B) The superimposition of both structures is shown from a downward orientation.

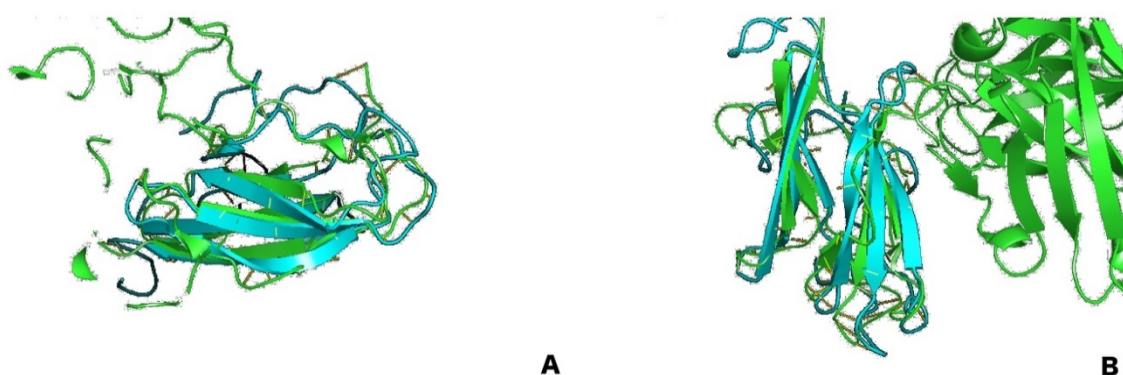


Figure 4-11 Structure superposition between 1s6n (green) and 4al8 (turquoise)

A) The superimposition of both structures is shown from a frontal orientation. B) The superimposition of both structures is shown from a downward orientation.

Two other groups that are not highlighted constantly are the Threonine (T), Aspartic Acid (D)and Glycine (G) columns. Although there are found hotspots for most proteins, the Tyrosine (Y) and Methionine (M)/Valine(V) residues are conserved throughout the sequences. The other group is the F residue in position 376, which is highly conserved and not highlighted as a hotspot by the analysis.

A	/1S6N 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96 101 106 ---GWC SKAFKFLGTPADTGHTVWLEQYGTGDPCKWPISSV ASLNDLT
B	/1S6N 11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96 101 106 ---YG CSKAFKFLGTPADTGHTVWLEQYGTGDPCKWPISSVASLNDLT PVGRLV
C	/1S6N 11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96 101 106 I SEFQLKGTTYGWCSKAFKFLGTPADTGHTVWLEQYGTGDPCKWPISSVASLNDLT PVGRLV
D	/1S6N 6 11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96 101 106 I SEFQLKGTTYGWCSKAFKFLGTPADTGHTVWLEQYGTGDPCKWPISSVASLNDLT PVGRLV
E	/1S6N 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96 101 106 -TTYGWC SKAFKFLGTPADTGHTVWLEQYGTGDPCKWPISSV ASLNDLT
F	/1S6N 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96 101 106 111 -ADTGHTWLEQ---YTG DGPCKWPISSV ASLNDLT
G	/1S6N 11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96 101 106 111 -Y GWCSKAFKFLGTPADTGHTVWLEQYGTGDPCKWPISSVASLNDLT PVGRLV

Figure 4-2 Structural Alignment of West Nile Virus 1s6N vs Dengue Proteins

(A) In here an amino acid chain alignment between 1s6N and s1OAN is shown. (B) In here an amino acid chain alignment between 1s6N and 1UZG. (C) In this section, the linear amino acid alignment between 1s6N and 3WE1. (D) Shows linear amino acid alignment between 1s6N and 4AL8 (E) In this section the linear amino acid alignment between 1s6N and 4Cct is shown. (F) In this section, the linear amino acid alignment between 1s6N and 4cbf (G) Shows linear amino acid alignment between 1s6N and 4CBF.

Sequence alignments cannot provide an accurate picture of the possible folding and structures overlapping. The results from the hotspot analysis were also overlapped with the super-imposition results shown in the previous section.

In Fig 4.14 we see a reconfirmation of the 3 main groups previously found in the sequence alignment and hotspot overlap. The Aspartic Acid (D), Threonine (T), Glycine (G) group is found in all structures. The second group is the Serine (S), Leucine (L), Asparagine (N), Aspartic Acid (D), Leucine (L), Threonine (T) and Proline (P) group, with exception of B and D (Fig 4.14) All others have some overlap in hotspots. The TDG, on the other hand, only has D, B (Fig 4.14) for the TDG group.

The conserved groups are not highlighted as hotspot. The phenylalanine (F) group (position 88) is structurally aligned and conserved. This is also the case for the Threonine (T), Glycine (G), Valine (V) group, which is a conserved group.

The superimposition for the 4cbf F chain structure was excluded because of its structure, which is smaller than others, and the alignment of Amino Acids was lower than the group average of over 100 residues aligned.

According to these results, the hotspot areas, or the critical area for binding- binding interactions are conserved at the structural level, as well as the sequence level.

These results indicate that the possible interaction by 1S6N with Integrin Beta-3 is possible in all sampled dengue proteins. This is furthered supported by (reference). (reference) indicated that Dengue infection produces an over expression of Integrin Beta-3 protein.

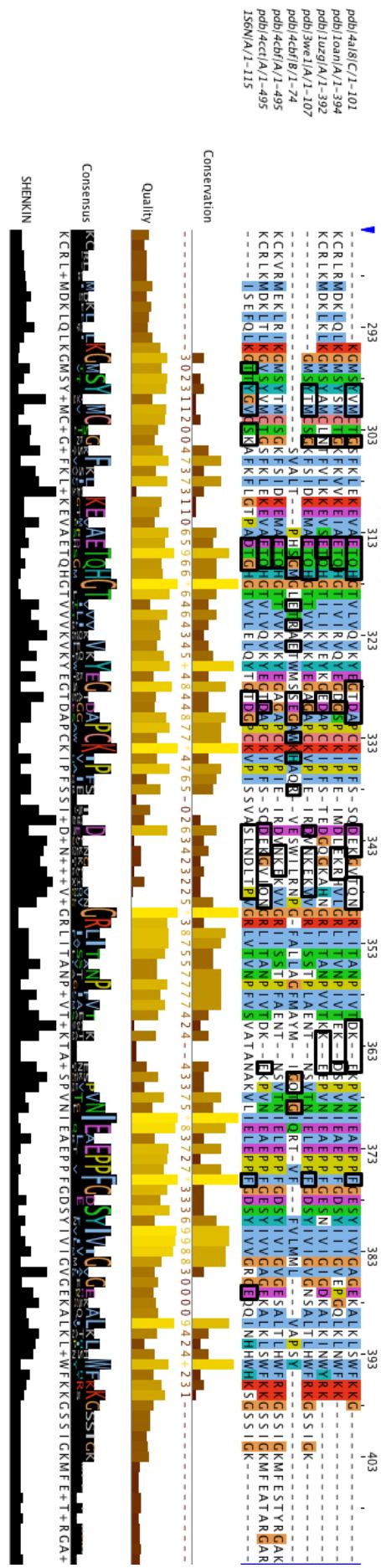


Figure 4-3 Alignment showing Results from HotSpot Analysis (Superimposition with Fig. 4.1)

In here all hotspots by analysis through PROFilts in Predict Protein for 4al8, 1oan, 1uzg, 2we1, 4bcl, 4cct, and 1s6n are highlighted in black squares.

CHAPTER 5: CONCLUSION

5.1 Summary

The initial structural approach in this piece of work did not yield results for possible new functions in Dengue-Human associated proteins. Mainly due to protein-protein interaction structures stored in PDB are paired with human antigen proteins.

The results gathered from the sequence approached yielded more interesting results. A final count of 825 proteins interaction from 9 modelled domains give rise to a possible array of new unknown interactions discovered.

One example is the interaction found between Envelop protein domain 3 of WNV (1S6N structure) and Human Integrin protein beta-3. Further Sequence Analysis shows that the sequence among all proteins sharing the same domain 2.60.40.350 is very similar, thus increasing the possibilities of a conserved function among them.

Structural superimposition scores (RMSD) also show the similarity of the structural domain 2.60.40.350 in all modelled Dengue proteins and 1S6N. Hotspot analysis also revealed that both in the linear sequence and the structural super imposition, there is a high conservation of hotspot.

All these results point to a possible interaction between these Dengue proteins and human integrin beta-3; the same as the WNV enveloped protein.

5.2 Drawbacks

Due to time constrains, further analysis of the samples used or acquired was not done. From the 825 associations found, only 2 were used for the analysis.

The first drawback is the selection criteria. Concerning the protein chosen, the selection criteria was not an abstract criterion such as statically significant samples, but rather a subjective criterion similar to close lineage for the case of Western Nile Virus protein P06935. It was based on lineage and its

interaction with human protein studied, like Integrin beta-3 (P05106). This thesis is focused on searching for mimicry of already known and established functions.

The second drawback was the use of data obtained from the Structural Approach. Although no association was obtained from the Method 1, the data set could be used for further analysis aimed to look for new potential associations, probably not found through the sequencing method. It could also support the strength of a possible mimic function, which could be occurring at the molecular level instead of being a possible model.

As described in the Introduction, there are more specific, but computational demanding methods for establishing protein-protein interactions. More accurate results could be produced with the use of these computationally demanded processes. At the same time, the use of tools such as (sample of tools), which modelled a possible interaction between two proteins, could further increase the probability that these two proteins interact with each other.

5.3 Future Research

Zhang et al., 2012 created a repository of interaction between yeast and human proteins, by using similar strategies to the one used in this project.

The research group took a series of steps, which included: sequence alignment, structural alignment, template creation (through superimposition of similar models found in PDB), calculated scores through Bayesian Statistics to create high-confidence interactions. Some of these interactions have been proven experimentally by independent research groups (Zhang et al., 2012).

Using similar methods, which include domain modelling and the methods used in this project, a repository for Dengue domains could be created with high-confidence interactions. These could give way to new molecular pathways previously unknown, or possible new targets for Dengue antimicrobial drugs.

A new possible way for further application of the methods used here consists of searching for new possible interactions for other pathogens which need for further understanding of their molecular

mechanism. Also, as the databases are updated, more possible interactions could be further modelled and studied.

CHAPTER 6: BIOGRAPHY

- Centers, for Disease Control and (CDC Prevention (2010), ‘Locally acquired Dengue--Key West, Florida, 2009-2010.’, *MMWR Morb Mortal Wkly Rep*, 59 (19), 577-81.
- Ackermann, Matt and R. Padmanabhan (2001), ‘De NovoSynthesis of RNA by the Dengue Virus RNA-dependent RNA Polymerase Exhibits Temperature Dependence at the Initiation but Not Elongation Phase’, *Journal of Biological Chemistry*, 276 (43), 39926-37.
- Alarcon Aulestia, Luis Alberto (2017), *MScProject-GWD Molecular Mimicry of Host Pathogen Interaction*, (GitHub).
- Ashour, J., et al. (2009), ‘NS5 of Dengue Virus Mediates STAT2 Binding and Degradation’, *Journal of Virology*, 83 (11), 5408-18.
- Avirutnan, P., et al. (2011), ‘Binding of Flavivirus Nonstructural Protein NS1 to C4b Binding Protein Modulates Complement Activation’, *The Journal of Immunology*, 187 (1), 424-33.
- Avirutnan, Panisadee, et al. (2010), ‘Antagonism of the complement component C4 by flavivirus nonstructural protein NS1’, *The Journal of Experimental Medicine*, 207 (4), 793-806.
- Bairoch, A., et al. (2005), ‘The Universal Protein Resource (UniProt).’, *Nucleic Acids Res*, 33
- Balmaseda, Angel, et al. (2006), *Assessment Of The World Health Organization Scheme For Classification Of Dengue Severity In Nicaragua*, (73;) 1059-62.
- Bartelma, Greg and R. Padmanabhan (2002), ‘Expression, Purification, and Characterization of the RNA 5'-Triphosphatase Activity of Dengue Virus Type 2 Nonstructural Protein 3’, *Virology*, 299 (1), 122-32.
- Bhamarapravati, N and Y Suttee (2000), ‘Live attenuated tetravalent dengue vaccine’, *Vaccine*, 18 44-47.
- Bogan, Andrew A and Kurt S Thorn (1998), ‘Anatomy of hot spots in protein interfaces’, *Journal of Molecular Biology*, 280 (1), 1-9.
- Brault, Jean-Baptiste, et al. (2011), ‘The interaction of flavivirus M protein with light chain Tctex-1 of human dynein plays a role in late stages of virus replication’, *Virology*, 417 (2), 369-78.
- Cagney, Gerard, Peter Uetz, and Stanley Fields (2000), ‘High-throughput screening for protein-protein interactions using two-hybrid assay’, *Methods in Enzymology*, 3-14.
- Chen, Feng, et al. (2014), ‘Up-Regulating CYP3A4 Expression in C3A Cells by Transfection with a Novel Chimeric Regulator of hPXR-p53-AD’, *PLoS ONE*, 9 (5), e95752.
- Chen, Shiow-Yi, et al. (2002), ‘Mechanisms for Inhibition of Hepatitis B Virus Gene Expression and Replication by Hepatitis C Virus Core Protein’, *Journal of Biological Chemistry*, 278 (1), 591-607.
- Chernov, Andrei V., et al. (2008), ‘The Two-component NS2B-NS3 Proteinase Represses DNA Unwinding Activity of the West Nile Virus NS3 Helicase’, *Journal of Biological Chemistry*, 283 (25), 17270-78.
- Chu, Justin Jang-hann and Mah-Lee Ng (2004), ‘Interaction of West Nile Virus with avß3Integrin Mediates Virus Entry into Cells’, *Journal of Biological Chemistry*, 279 (52), 54533-41.
- Cock, P. J. A., et al. (2009), ‘Biopython: freely available Python tools for computational molecular biology and bioinformatics’, *Bioinformatics*, 25 (11), 1422-23.
- Dejnirattisai, W., et al. (2010), ‘Cross-Reacting Antibodies Enhance Dengue Virus Infection in Humans’, *Science*, 328 (5979), 745-48.
- Dorji, Tandin, et al. (2009), ‘Diversity and Origin of Dengue Virus Serotypes 1, 2, and 3,

- Bhutan', *Emerging Infectious Diseases*, 15 (10), 1630-32.
- Edelman, R. (2007), 'Dengue Vaccines Approach the Finish Line', *Clinical Infectious Diseases*, 45 (Supplement 1), S56-60.
- Egloff, M.-P. (2002), 'An RNA cap (nucleoside-2'-O)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization', *The EMBO Journal*, 21 (11), 2757-68.
- Falgout, B., et al. (1991), 'Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins.', *J Virol*, 65 (5), 2467-75.
- Feire, A. L., H. Koss, and T. Compton (2004), 'Cellular integrins function as entry receptors for human cytomegalovirus via a highly conserved disintegrin-like domain', *Proceedings of the National Academy of Sciences*, 101 (43), 15470-75.
- G., van Rossum (1995), 'CMIFed: A Presentation Environment for Portable Hypermedia Documents', *Centrum voor Wiskunde en Informatica (CWI)*,
- Gavin, Anne-Claude, et al. (2002), 'Functional organization of the yeast proteome by systematic analysis of protein complexes', *Nature*, 415 (6868), 141-47.
- Gavrilovskaya, Irina N., et al. (1998), 'β3 integrins mediate the cellular entry of hantaviruses that cause respiratory failure', *Proceedings of the National Academy of Sciences*, 95 (12), 7074-79.
- Gubler, Duane J. (1997), 'Dengue and dengue hemorrhagic fever', *Seminars in Pediatric Infectious Diseases*, 8 (1), 3-9.
- Hales, Simon, et al. (2002), 'Potential effect of population and climate changes on global distribution of dengue fever: an empirical model', *The Lancet*, 360 (9336), 830-34.
- Halstead, Scott B, Jose A Suaya, and Donald S Shepard (2007), 'The burden of dengue infection', *The Lancet*, 369 (9571), 1410-11.
- Harris, Eva, et al. (2008), 'Molecular Biology of Flaviviruses', *Novartis Foundation Symposia*, 23-40.
- Hosur, Raghavendra, et al. (2011), 'iWRAP: An Interface Threading Approach with Application to Prediction of Cancer-Related Protein-Protein Interactions', *Journal of Molecular Biology*, 405 (5), 1295-310.
- Hosur, Raghavendra, et al. (2012), 'A computational framework for boosting confidence in high-throughput protein-protein interaction datasets', *Genome Biology*, 13 (8), R76.
- Iglesias, Nestor G. and Andrea V. Gamarnik (2011), 'Dynamic RNA structures in the dengue virus genome', *RNA Biology*, 8 (2), 249-57.
- Ismail, Noor (2012), *Early predictors of dengue infection in adults (EPOD) A Malaysian outpatient experience*, 36: 105-115.
- James, P, J Halladay, and EA Craig (1996), 'Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast', *Genetics*,
- Joki-Korpela, P., et al. (2001), 'Entry of Human Parechovirus 1', *Journal of Virology*, 75 (4), 1958-67.
- Kalayanarooj, Siripen, et al. (1997), 'Early Clinical and Laboratory Indicators of Acute Dengue Illness The Journal of infectious diseases', 176 313-21.
- Kelley, James F., et al. (2011), 'Maturation of dengue virus nonstructural protein 4B in monocytes enhances production of dengue hemorrhagic fever-associated chemokines and cytokines', *Virology*, 418 (1), 27-39.
- Kelley, James F., Pakieli H. Kaufusi, and Vivek R. Nerurkar (2012), 'Dengue hemorrhagic fever-associated immunomediators induced via maturation of dengue virus nonstructural 4B protein in monocytes modulate endothelial cell adhesion molecules and human microvascular endothelial cells permeability', *Virology*, 422 (2), 326-37.
- Kovats, R. S., et al. (2001), 'Early effects of climate change: do they include changes in vector-borne disease?', *Philosophical Transactions of the Royal Society B: Biological*

- Sciences*, 356 (1411), 1057-68.
- Kuhn, Richard J., et al. (2002), ‘Structure of Dengue Virus’, *Cell*, 108 (5), 717-25.
- La Ruche, G, et al. (2010), ‘First two autochthonous dengue virus infections in metropolitan France, September 2010.’, *Euro Surveill*, 15 (39), 19676.
- Lindenbach, B. D. (2005), ‘Complete Replication of Hepatitis C Virus in Cell Culture’, *Science*, 309 (5734), 623-26.
- Locatelli, Giada A, et al. (2001), ‘Hepatitis C virus NS3 NTPase/Helicase: different stereoselectivity in nucleoside triphosphate utilisation suggests that NTPase and helicase activities are coupled by a nucleotide-dependent rate limiting step’, *Journal of Molecular Biology*, 313 (4), 683-94.
- Lu, Long, Hui Lu, and Jeffrey Skolnick (2002), ‘MULTIPROSPECTOR: An algorithm for the prediction of protein-protein interactions by multimeric threading’, *Proteins: Structure, Function, and Genetics*, 49 (3), 350-64.
- Lum, LCS, CJ Ng, and EM Khoo (2014), ‘Managing dengue fever in primary care: A practical approach’, *Malays Fam Physician Malaysian Family Physician : the Official Journal of the Academy of Family Physicians of Malaysia*, 9 (2), 2-10.
- Lum, Lucy, et al. (2008), ‘Quality of life of dengue patients The American journal of tropical medicine and hygiene’, 78 862-67.
- Lyons, Arthur, et al. (2007), ‘A Phase 2 study of a purified, inactivated virus vaccine to prevent Japanese encephalitis’, *Vaccine*, 25 (17), 3445-53.
- Maiorov, V. N. and G. M. Crippen (1994a), ‘Learning about protein folding via potential functions’, *Proteins: Structure, Function, and Genetics*, 20 (2), 167-73.
- Maiorov, Vladimir N. and Gordon M. Crippen (1994b), ‘Significance of Root-Mean-Square Deviation in Comparing Three-dimensional Structures of Globular Proteins’, *Journal of Molecular Biology*, 235 (2), 625-34.
- McMichael, Anthony J, Rosalie E Woodruff, and Simon Hales (2006), ‘Climate change and human health: present and future risks’, *The Lancet*, 367 (9513), 859-69.
- Memiševic, Vesna, Anders Wallqvist, and Jaques Reifman (2013), ‘Reconstituting protein interaction networks using parameter-dependent domain-domain interactions’, *BMC Bioinformatics*, 14 (1), 154.
- Meng, Qingshi, Liqun Rao, and Yinghong Pan (2014), ‘Enrichment and analysis of rice seedling ubiquitin-related proteins using four UBA domains (GST-qUBAs)’, *Plant Science*, 229 172-80.
- Miller, Sven, et al. (2007), ‘The Non-structural Protein 4A of Dengue Virus Is an Integral Membrane Protein Inducing Membrane Alterations in a 2K-regulated Manner’, *Journal of Biological Chemistry*, 282 (12), 8873-82.
- Mogensen, T. H. and S. R. Paludan (2001), ‘Molecular Pathways in Virus-Induced Cytokine Production’, *Microbiology and Molecular Biology Reviews*, 65 (1), 131-50.
- Munoz-Jordan, J. L., et al. (2003), ‘Inhibition of interferon signaling by dengue virus’, *Proceedings of the National Academy of Sciences*, 100 (24), 14333-38.
- Nagila, Amar, et al. (2011), ‘Role of CD137 signaling in dengue virus-mediated apoptosis’, *Biochemical and Biophysical Research Communications*, 410 (3), 428-33.
- Nemerow, Glen R. and David A. Cheresh (2002), ‘Herpesvirus hijacks an integrin’, *Nature Cell Biology*, 4 E69-71.
- Orengo, CA, et al. (1997), ‘CATH – a hierachic classification of protein domain structures’, *Structure*, 5 (8), 1093-109.
- Pandey, Basu Dev, et al. (2008), ‘Dengue Virus, Nepal’, *Emerging Infectious Diseases*, 14 (3), 514-15.
- Parasuraman, S (2012), ‘Protein data bank’, *Journal of Pharmacology and Pharmacotherapeutics*, 3 (4), 351.

- Pavelka, Antonín, Eva Sebestova, and Jiří Damborský (2009), ‘HotSpot Wizard: A Web Server for Identification of Hot Spots in Protein Engineering Nucleic acids research’, 37 W376-83.
- Piston, David W. and Gert-Jan Kremers (2007), ‘Fluorescent protein FRET: the good, the bad and the ugly’, *Trends in Biochemical Sciences*, 32 (9), 407-14.
- Razick, Sabry, George Magklaras, and Ian M Donaldson (2008), ‘iRefIndex: A consolidated protein interaction database with provenance’, *BMC Bioinformatics*, 9 (1), 405.
- Rice, C., et al. (1985), ‘Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution’, *Science*, 229 (4715), 726-33.
- Rigau-Pérez, José G, et al. (1998), ‘Dengue and dengue haemorrhagic fever’, *The Lancet*, 352 (9132), 971-77.
- Rodenhuis-Zybert, Izabela A., Jan Wilschut, and Jolanda M. Smit (2010), ‘Dengue virus life cycle: viral and host factors modulating infectivity’, *Cellular and Molecular Life Sciences*, 67 (16), 2773-86.
- Sangiambut, S., et al. (2008), ‘Multiple regions in dengue virus capsid protein contribute to nuclear localization during virus infection’, *Journal of General Virology*, 89 (5), 1254-64.
- Schlender, Jorg, et al. (1996), ‘Interaction of measles virus glycoproteins with the surface of uninfected peripheral blood lymphocytes induces immunosuppression in vitro’, *Proceedings of the National Academy of Sciences*, 93 (23), 13194-99.
- Schrodinger, LLC (2015), ‘The PyMOL Molecular Graphics System, Version~1.8’,
- Schulze zur Wiesch, J., et al. (2005), ‘Broad Repertoire of the CD4+ Th Cell Response in Spontaneously Controlled Hepatitis C Virus Infection Includes Dominant and Highly Promiscuous Epitopes’, *The Journal of Immunology*, 175 (6), 3603-13.
- Sharma-Walia, N., et al. (2004), ‘Kaposi’s Sarcoma-Associated Herpesvirus/Human Herpesvirus 8 Envelope Glycoprotein gB Induces the Integrin-Dependent Focal Adhesion Kinase-Src-Phosphatidylinositol 3-Kinase-Rho GTPase Signal Pathways and Cytoskeletal Rearrangements’, *Journal of Virology*, 78 (8), 4207-23.
- Shiryaev, SergeyA., et al. (2006), ‘Cleavage targets and the D-arginine-based inhibitors of the West Nile virus NS3 processing proteinase’, *Biochemical Journal*, 393 (2), 503-11.
- Somnuke, Pawit, et al. (2011), ‘N-linked glycosylation of dengue virus NS1 protein modulates secretion, cell-surface expression, hexamer stability, and interactions with human complement’, *Virology*, 413 (2), 253-64.
- Valencia, A and F Pazos (2002), ‘Computational methods for the prediction of protein interactions.’, *Curr Opin Struct Biol*, 12 (3), 368-73.
- Velankar, S., et al. (2010), ‘PDBe: Protein Data Bank in Europe’, *Nucleic Acids Research*, 39 (Database), D402-10.
- Waterhouse, A. M., et al. (2009), ‘Jalview Version 2--a multiple sequence alignment editor and analysis workbench’, *Bioinformatics*, 25 (9), 1189-91.
- WHO and HealthMap.org (1997), *Health Map: Flu Map, Contagious Disease Surveillance, Virus Awareness 2014*, (World Health Organization).
- WHO (2009), *Dengue : guidelines for diagnosis treatment prevention and control*, (World Health Organization).
- (2012), *Global strategy for dengue prevention and control, 2012-2020*, (World Health Organization).
- Winkler, Gunther, et al. (1988), ‘Evidence that the mature form of the flavivirus nonstructural protein NS1 is a dimer’, *Virology*, 162 (1), 187-96.
- Winkler, Gunther, et al. (1989), ‘Newly synthesized dengue-2 virus nonstructural protein NS1 is a soluble protein but becomes partially hydrophobic and membrane-associated after dimerization’, *Virology*, 171 (1), 302-5.

- World, Health Assembly (2005), *Revision of the International Health Regulations*, ((WHA58.3. 2005 May 23); World Health Organization) 159.
- World, Health Organization (2017), ‘Dengue vaccine: WHO position paper, July 2016 – recommendations’, *Vaccine*, 35 (9), 1200-1.
- Xiong, Xiangyang, et al. (2014), ‘Heat shock protein 90 β stabilizes focal adhesion kinase and enhances cell migration and invasion in breast cancer cells’, *Experimental Cell Research*, 326 (1), 78-89.
- Yu, I.-M., et al. (2009), ‘Association of the pr Peptides with Dengue Virus at Acidic pH Blocks Membrane Fusion’, *Journal of Virology*, 83 (23), 12101-7.
- Zhang, Qiangfeng Cliff, et al. (2012), ‘Structure-based prediction of protein–protein interactions on a genome-wide scale’, *Nature*, 490 (7421), 556-60.