CLONING, EXPRESSION, PURIFICATION OF HPgV (GBV-C VIRUS) NS5B PROTEIN – RNA DEPENDENT RNA POLYMERASE FROM INCLUSION BODIES

_____

A University Thesis Presented to the Faculty

of

California State University, East Bay

_____

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Biological Sciences

_____

By

Madhumitha Seshadri

December 2018

ABSTRACT

The GB Virus type – C (GBV-C), formerly known as Hepatitis G Virus/HGV is a member of the Flaviviridae family that has recently been classified under the genus – Pegivirus. It has a positive sense single stranded RNA genome. Among the members of Flaviviridae, most are pathogenic like that of Hepatitis C virus (HCV), West Nile virus, Yellow fever virus and Dengue virus; GBV-C is found to be a non-pathogenic virus. GBV-C is found to infect peripheral blood mononuclear cells (PBMCs), $CD4^+$ and B-cells. The viral RNA dependent RNA polymerase (RdRp) encoded by the NS5B gene, is responsible for replication of the GBV-C genome. Recently, there have been studies showing that co-infection of HIV-1 with GBV-C has shown to slow the replication of HIV-1 and delays the progression to AIDS (Acquired Immune Deficiency Syndrome). However, the mechanism by which GBV-C aids in slowing the replication of HIV-1 is uncertain. Understanding the mechanism of viral replication aids in providing more insight on the viral life cycle and its role in slowing down the HIV-1 replication during co-infection. As mentioned above, most of the members of Flaviviridae are highly pathogenic and understanding the structure and mechanism of the GBV-C polymerase will be useful to study drug target mechanisms and inhibition of replication of other pathogenic members of the family as polymerases have found to be structurally similar. The GBV-C genome possesses nucleic acid sequence similarity with the Hepatitis- C virus (HCV), thus making it a potential model to be used for therapeutic studies since it is a non-pathogenic model. In this Project, the coding sequence of NS5B gene that codes for GBV-C RNA dependent RNA polymerase (RdRp) will be cloned into an expression vector fused with metal binding tag that will allow the purification of the polymerase. The results of this research will potentially aid in obtaining the crystal structure of the protein that can be used for crystallization studies using X-ray diffraction method. Further, the crystallized structure will be used to determine the 3-D structure of the protein.

The structure of protein will be helpful in providing a potential non-pathogenic model for therapeutic studies on pathogenic members of the Flaviviridae family by giving more insight on developing inhibitors against drug targets, vaccine development and other clinical research as well as in obtaining a coherent picture of the replication of GBV-C and its role in co-infection with HIV-1.

CLONING, EXPRESSION, PURIFICATION OF HPgV (GBV-C VIRUS) NS5B PROTEIN – RNA DEPENDENT RNA POLYMERASE FROM INCLUSION BODIES

By

Madhumitha Seshadri

Approved:	Date:
Professor Christoph Baysdorfer	11/26/18
Professor Claudia Udhe-Stone	2018 Nov 29
Professor Ann McPartland	11/26/2018

ACKNOWLEDGEMENTS

I would like to express my heartfelt thanks and gratitude to Dr.Curr, for welcoming me into his lab and supporting me in my research. Throughout the project, Dr.Curr had been very kind, understanding and patient in guiding me. Without Dr.Curr, it would have been impossible for me to have completed this project. It is an honor to have worked in his lab.

I also extend my humble gratitude to Professor Stone, Professor Baysdorfer and Professor McPartland for their continuous support during this project. I am thankful to all my professors for being an inspiration to me.

I would like to thank the Center for Student Research for the financial support by providing the student research fellowship.

I also thank my fellow graduate students for their support throughout my project.

Finally, I thank my parents, my husband and my daughter for all the support and understanding during the course of my project. Without their support and co-operation, I would never be able to pursue my interest and passion in science.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Section 1: Discovery

The GB virus C (GBV-C) is an enveloped, circular virus with a positive sense single- stranded RNA genome. The virus was discovered in late 1995, as a new member of Flaviviridae family, which was thought to be a novel etiologic agent that was responsible for non-A to non-E hepatitis. (3-8).

Earlier in 1967, Deinhardt, F. et al. studied on disease transmission of human viral hepatitis to marmoset monkeys, to develop a new experimental model for this disease. Serum from a 34-year-old surgeon initialed G.B., who had developed acute hepatitis after unknowingly exposing himself to hepatitis serum, was inoculated into marmosets and later all of those developed hepatitis (3). The above "GB agent" that was passaged in primates was then passaged into tamarins and the "GB agent" was found to contain two unique flavivirus-like genomes characterized as GBV-A and GBV-B (4,5). The genomes were cloned and sequenced and the "GB agent" causing the hepatitis was represented as two new genera in the Flaviviridae family. Later, a clone from non -A -non -B hepatitis causing agent was isolated from genome of blood borne virus and was designated as Hepatitis C Virus (HCV) (6).

In late 1995, two independent research groups simultaneously identified a potentially new hepatitis causing virus. One of the groups identified the potential agent that was responsible for non-A to non-E hepatitis and tentatively named it as GBV-C (8). Another group isolated a closely related RNA virus from plasma of a patient with chronic hepatitis and designated it as HGV, found to be distantly related to HCV, GBV-A and GBV-B. Later, the studies on GBV-C and HGV showed that they were isolates of same virus (9, 10). Multiple studies have showed that GBV-C does not cause any disease in humans and replicates poorly in hepatocytes (11, 12) and hence the terminology HGV is no longer accurate. Due to its significance in sequence similarity with HCV and potential HIV-1 therapeutics, it is importance to study the GBV - C viral replication, infection and

transmission.

Section 2: Infection and transmission

GBV-C, which was initially thought to be a hepatotropic virus (13) was later shown that it is a virus replicating in Peripheral Blood Mononuclear Cells (PBMC) (14, 15) and their RNA transcripts from full length cDNA are shown to be infectious in primary CD4-positive T cells (16).

So far, parental transmission and sexual transmission of GBV-C have been determined and is evident that the prevalence and incidence of the infection of GBV-C is higher in risk groups that are prone to sexually transmitted or blood borne infections. (17, 18).

It is considered to be unnecessary to screen blood donors for GBV-C and to eliminate them from being a blood donor, as GBV-C is prevalent in major donor population and it is not pathogenic to humans. (19, 48).

It was found that at the time of blood donation, about 1-4% of healthy blood donors in most of the developed countries were found to contain GBV-C viremia and about 13% were found to have anti-E2 (anti-second envelope glycoprotein) antibodies, which shows the presence of prior infection. (20, 48). The prevalence of GBV-C RNA was found to be higher in patients infected with HIV-1. (21). The GBV-C viremia is found to be about 20% among population with chronic HCV infection and 20-40% among individuals who are HIV-1 positive (22).

In vitro PBMC based cultures demonstrated different replication phenotypes of GBV-C in plasma of individuals with HIV coinfection (15).

Section 3: GBV-C interaction with HIV

So far, several studies have made associations between GBV-C infection and an increase in survival rate of patients infected with HIV-1. (23, 24, 49, 50, 51, 52). GBV-C

shares common routes of transmission with HCV and HIV-1, resulting in co- and triple-infection in humans (25).

Nattermann *et al.,* demonstrated that interaction between GBV-C E2 and CS81 might contribute in delayed progression towards HIV infection in patients who are co infected with HGV (26).

Furthermore, a study about interaction of GBV-C NS3 protease has revealed some insights on inhibition of HIV-1 replication, without decreasing HIV receptor expression (27). Studies on inhibition of HIV-1 replication in $CD4^+$ Jurkat T cells by GBV-C NS5A phosphoprotein has been made to propose a novel therapeutic approach for anti-HIV therapy (28).

Many studies have also shown that GBV-C co infection with HIV has shown increase in $CD4^+$ T cell counts and lowering of HIV viral load and delay in progression of HIV infection to AIDS (24, 29,30,31). Upregulation of IL-16 (Interleukin) mediated by GBV-C has allowed for HIV-1 suppression in patients co- infected with GBV-C (32).

There is no clear evidence and no lucid picture on how the coinfection of GBV-C and HIV results in delay in progression of HIV-1 to AIDS. For such studies, understanding the replication of GBV-C will play a major role in understanding the interaction between GBV-C and HIV-1. To understand the GBV-C replication process, it is important to get detailed insight about polymerase structure and function. A detailed analysis of active site of polymerase will provide more information needed for development of drugs, not just against GBV-C but against other members of Flaviviridae family. For these reasons, determining the 3-D structure of protein polymerase coded by NS5B gene becomes significant.

Section 4: The Flaviviridae Family

The Flaviviridae family of viruses have been associated with various significant human infections. The Flaviviridae viruses are classified into three genera: Pestivirus,

Flavivirus and Hepaciviruses, recently a fourth genus was proposed as Pegivirus, which consists of GBV-A, GBV-C and newly isolated GBV- D.

Pestiviruses are widespread cattle pathogens present worldwide (Cytopathic Bovine Viral Diarrhea viruses (33). They include the Classical Swine Fever Virus (CSFV) and Bovine Viral Diarrheal Virus (BVDV). The Flavivirus genus consists of over 60 species most of which are transmitted by either the mosquito or the tick vector. (34, 35).

The Flaviviridae represents some of the deadliest human pathogens, such as, West Nile virus, Dengue virus, Yellow fever virus and Japanese encephalitis virus. Most of these arthropod-borne flaviviruses cause severe hemorrhagic fevers in humans (36).

The Hepacivirus genus represents the HCV as well as the GBV-B. About 3% of world's population is infected with HCV and it is of a major clinical significance as it leads to chronic liver infection (37).

Recently, a new genus was proposed as Pegivirus which represents the viruses GBV-A, GBV-C and GBV-D (Isolated from bat), based on their phylogenetic relationships, structure of genome and non-pathogenic nature (20).

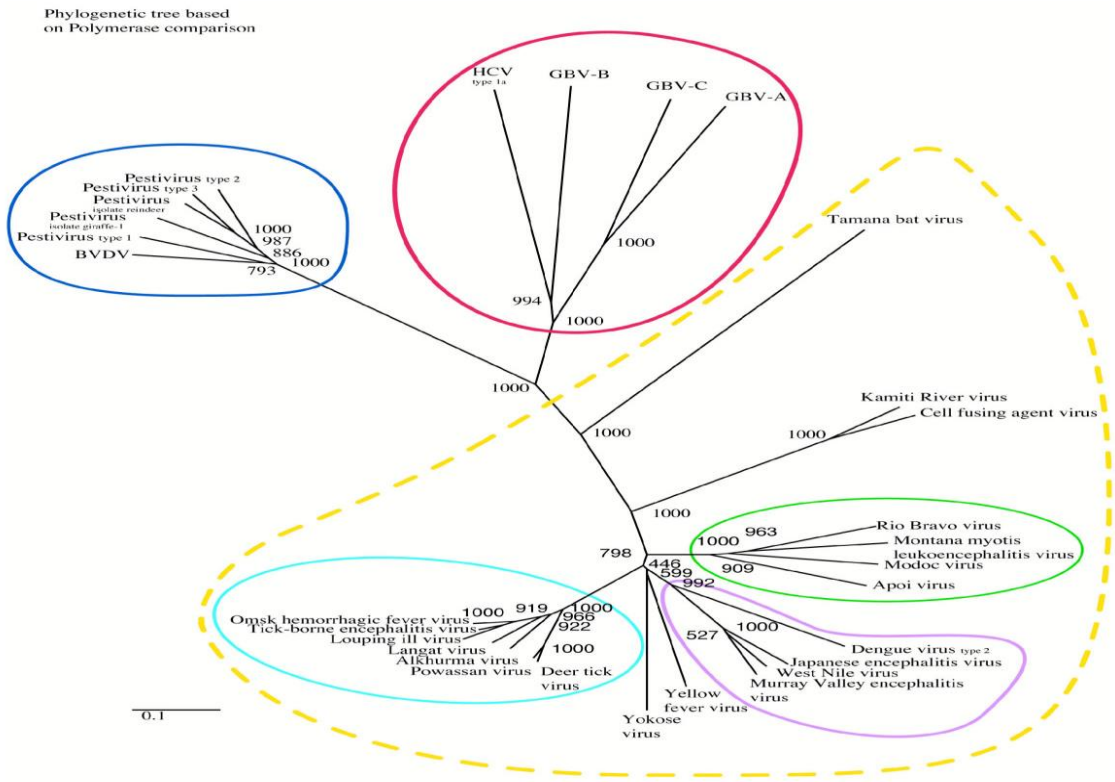


Figure 1. The Phylogenetic tree of Flaviviridae. Number at the nodes indicate the statistical support of the branching order by bootstrap criteria. The bar at the bottom of the phylogram indicates the evolutionary distance, to which the branch lengths are scaled based on the estimate divergence. The dashed yellow line indicates the flagitious genus, the blue line indicates the pest virus genus and the red line indicates hepacivirus genus. (2). In the above figure, Ferron et al have shown a NS5B-based phylogenetic analysis showing that RNA dependent RNA polymerases of GB viruses are closest to HCV in *Flaviviridae*. (2)

Section 5: Genome and Replication

The Flaviviridae family are enveloped, positive sense single stranded RNA viruses. Which means they have their genome directly utilized as if it were mRNA. The Flaviviridae genome has one single open reading frame (ORF) and on both ends of this ORF non-structural regions of approximately 100 nucleotides exist (38). 5’ non-coding

regions are not highly conserved throughout different Flaviviridae species and 3’ non-coding regions tend to be highly variable between the viruses that are transmitted by mosquitoes and the viruses that are transmitted by ticks. The genus Flavivirus has a type I cap structure at the 5' untranslated region (UTR) and a highly structured 3' UTR, and the remaining three groups exhibit translational control by means of an internal ribosomal entry site (IRES) in the 5'- UTR and a much shorter less-structured 3'- UTR (39).

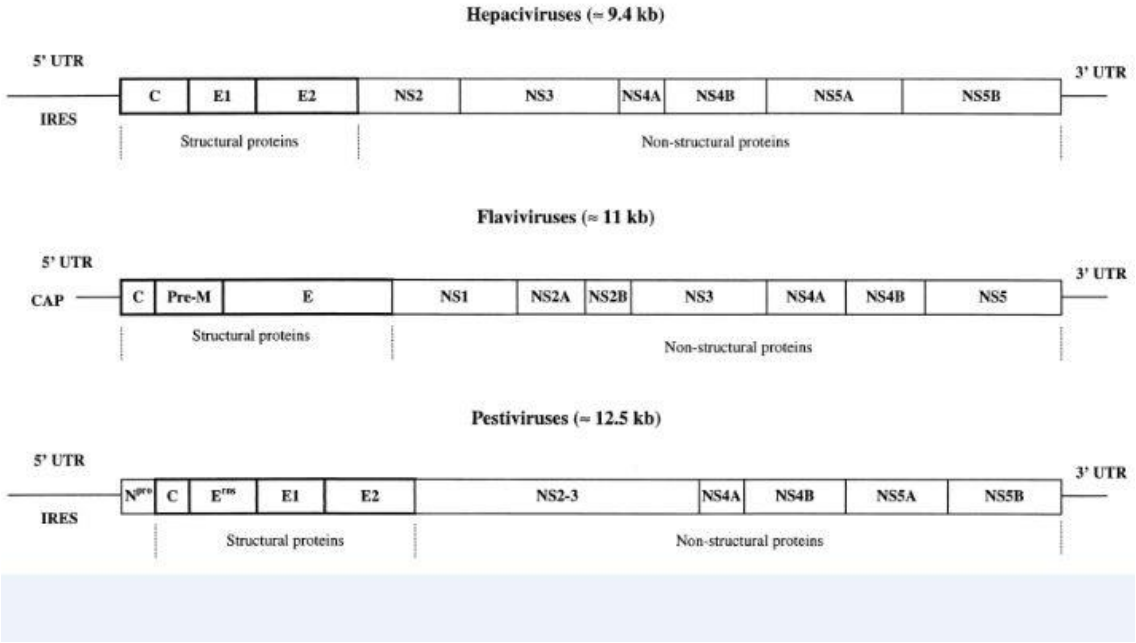


Figure 2. Genomic organization of members of the *Flaviviridae*. The viral genome consists of a single-stranded RNA molecule of positive polarity which is capped in flaviviruses and contains an IRES in hepaciviruses and pestiviruses. UTR are present at the 5' and 3' ends of the genome. Boxes indicate mature proteins generated by proteolytic processing. (40)

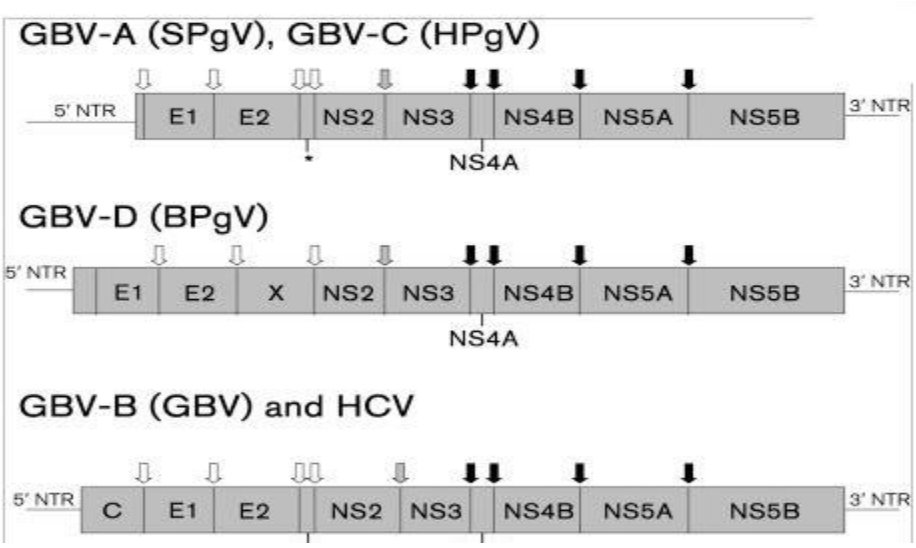


Figure 3. Genome organization of the GB viruses and HCV. All four viruses contain a positive-polarity ssRNA genome with a 5' NTR and 3' NTR. The genome encodes a polyprotein that is co- and post-translationally cleaved into individual viral proteins. The structural proteins include core (C) and envelope glycoproteins (E1 and E2), and the NS proteins include NS2–NS5B. (20)

The structural proteins in GB virus and HCV carry the enzymatic activity needed for production of new virions. The structural proteins E1 and E2 are found on surface of Flaviviruses and are cleaved by host protease (41). E2 is the prevalent antigen that produces antibodies in host cell that is infected by GBV-C although the receptors for GBV-C remains unknown. (42). The non-structural proteins NS2, NS3, NS4A, NS5A and NS5B are necessary for replication and infectivity of the virus. The NS2 and NS3 proteins carryout the autocatalytic activity and allow the auto-cleavage of NS2/NS3 (41). The rest of the NS proteins are cleaved by the protease formed by the NS3-NS4A complex where NS3 acts as a chemotropism- like serine protease and NS4A as its cofactor. (43). NS5A is thought to be involved in interferon sensitivity and interacts with the double stranded RNA protein kinase (PKR) (44). The NS5B is found at the carboxy-terminal of the polyprotein. NS5B is the RNA dependent RNA polymerase essential for the virus infectivity and replication of its genome. It is encoded by a 1.7Kb sequence at

the 3`-terminal end of the virus genome and is translated into a predicted 62.29 KDa functional protein (16).

Section 6: GBV-C as a Potential Model for Therapeutics

GBV-C has been proposed as a potential model to study the Hepatitis C Virus (HCV) due to its phylogenetic relation with HCV. The presence of a shared protease substrate makes GBV-C, a useful model to test anti-viral compounds against HCV (45). By determining the 3-D structure of GBV-C NS5B, GBV-C can be made a potential model for study of hepatitis due to its similarity in nucleic acid sequence.

This GBV-C isolate was found to contain amino acid substitutions in both NS5A and NS5B coding regions similar to that of changes identified in HCV NS5A and NS5B, which was found to be involved in interferon sensitivity and enhanced RNA replication (46). Thus, understanding such substitutions in GBV-C proteins will aid in knowing its role in viral replication and evasion of host immune response. Compared to Bovine Viral Diarrheal virus (BVDV), GBV-C is closer to HCV in terms of model polymerase for anti-viral drug screening and therapeutics (47).

Section 7: Transition site inhibitors as potential inhibitors of RNA dependent RNA polymerases in Flaviviridae family:

Structure of RNA Dependent RNA Polymerases in Flaviviridae Family

The RNA Dependent RNA Polymerases in the Flaviviridae family are found to have the shape of right hand with fingers, palm and thumb domains and the inner surface formed by three domains is found to form a central template- binding channel (Figure 4) (53). The RdRps have unique finger domain which has several strands of long inserts called “fingertips”, forming an entrance to the template binding channel. Among the polymerases, the catalytic domain which is the palm domain is found to be highly conserved, structurally (53).

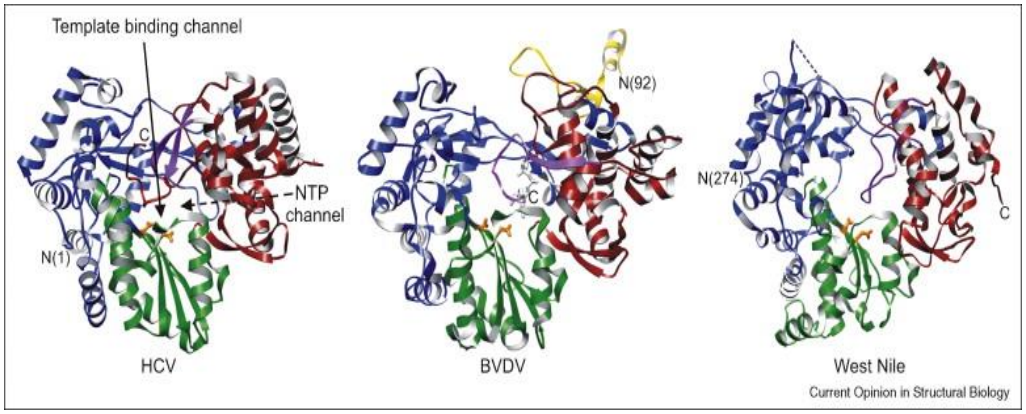


Figure 4: Structures of *Flaviviridae* RNA polymerases. RdRp domains of HCV, BVDV, and West Nile virus are shown. The fingers, palm, and thumb domains are shown in blue, green, and red, respectively. The C-terminal motif is colored in purple. The N-termini and C-termini are labeled, and the N-terminal residue for each structure is indicated in parenthesis. The directions that the template and NTP would take to enter the active site are indicated by arrows (53).

Catalytic Motifs of RdRps in Flaviviridae Family

The palm domain in all the viral RNA dependent RNA polymerases harbors the motifs A-E with motifs A-C bearing most conserved features. The motifs A and C, each containing aspartic acid residues are universally conserved in all single -subunit nucleic acid polymerases and they play key role in two-metal ion catalysis. Motif B has highly conserved serine that plays central role in 2`hydroxyl group of NTP ribose recognition. The glycine reside in motif B is conserved in all RdRps. The catalytic motifs F/G in the polymerases are shared by all viral RdRps with conservations in sequence and/or structures (54).

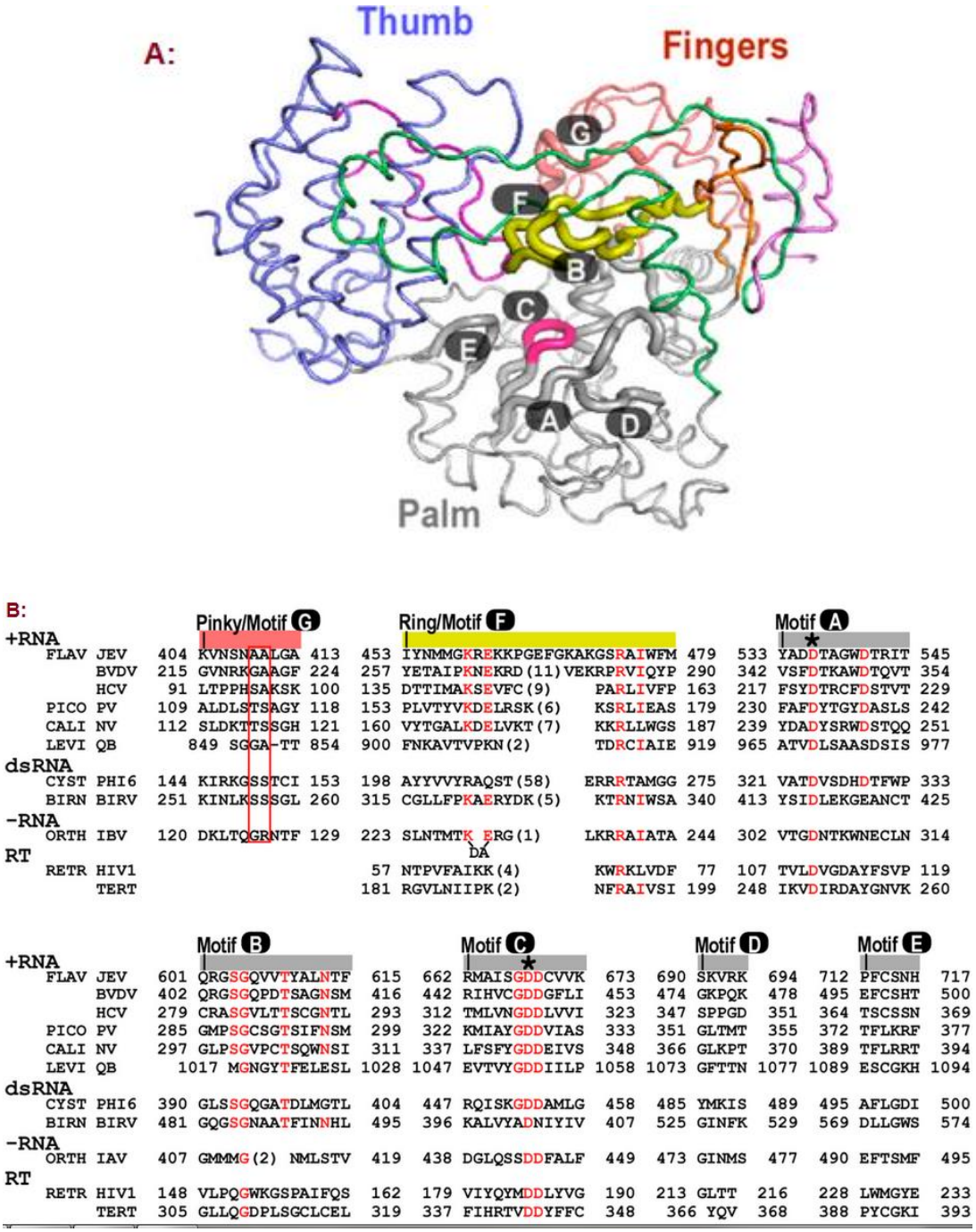
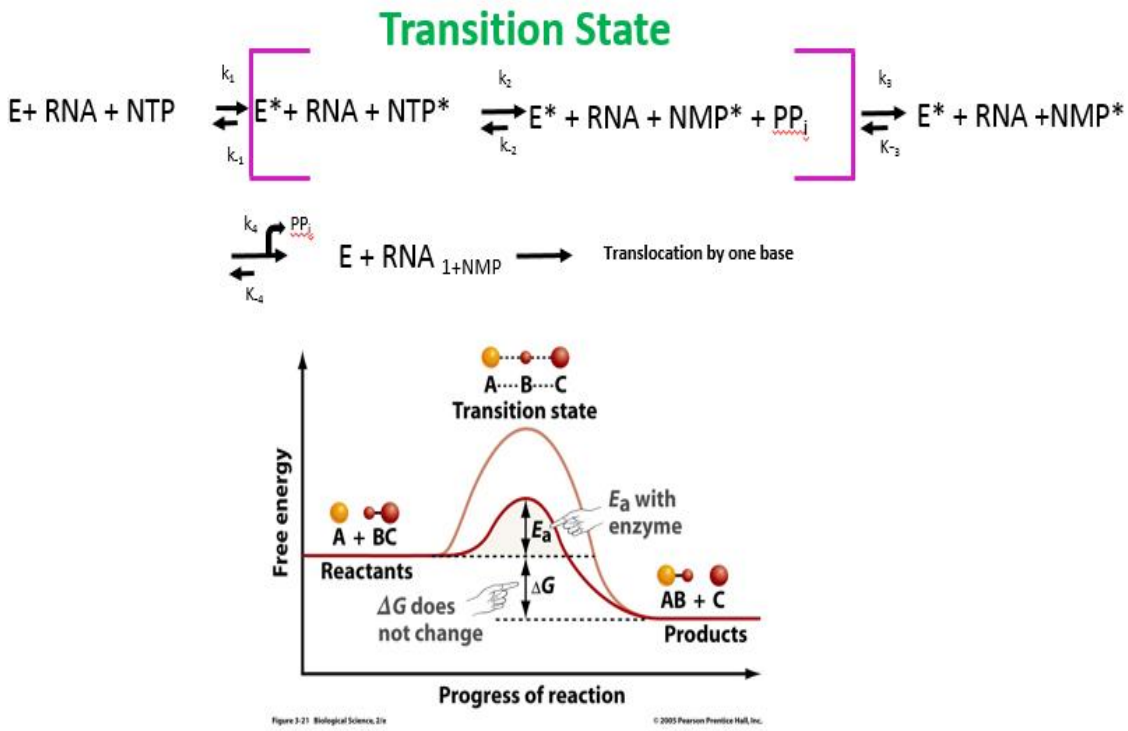


Figure 5: (Figure 5.A): Spatial organization of the JEV RdRP catalytic motifs A–G. **(Figure 5. B):** A structure-based sequence alignment depicting the conservation of RdRP motifs (Protein Data Bank (PDB) entries: 4K6M, 1S4F, 1NB4, 3OL6, 3BSO, 3AVT, 1HI0, 2PGG, 4WRT, 1RTD, and 3DU6). (54)

Transition state inhibitors targeting NS5B:

Methylthioadenosine nucleosidase are the enzymes, catalyzing the hydrolytic deadenylation reaction of 5'-methylthioadenosine and S-adenosylhomocysteine. This is a very important reaction in many bacterial species. There is a stereoisomeric arrangement occurring at the anomeric carbon that changes the distance of the nitrogen during the transition stage, thereby accounting for an early and late stage transition state. Guitierrez et al., has created two transition analogs that work on at the picomolar level in *E. coli* Methylthioadenosine nucleosidases (Figure 6) (55).



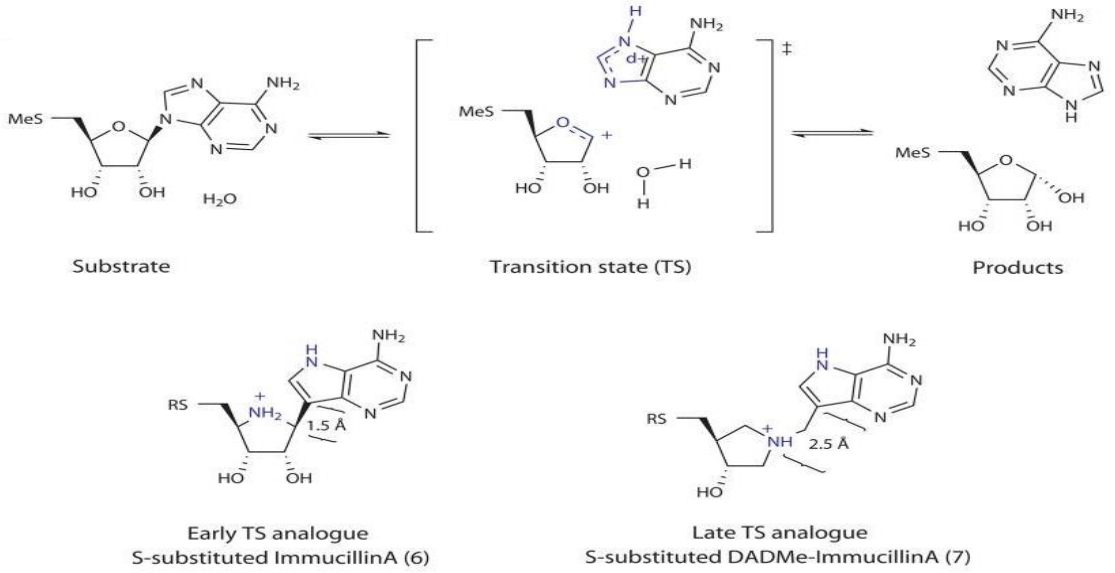


Figure 6. Transition state in an enzymatic reaction and transition state inhibitors (55)

Transition state inhibitors can be used as a potential therapeutic drug to prevent replication of viruses. The structural similarity and the identity in the catalytic domains among the RNA Dependent RNA Polymerases of Flaviviridae help to develop potential transition-state inhibitors that can act against most of the RNA Dependent RNA Polymerases in the Flaviviridae family. The 3D structure determination of GBV-C NS5B will aid in using GBV-C NS5B as a potential model for therapeutic purposes to develop antiviral products such as transition state inhibitors.

Section 8: Inclusion Bodies

Inclusion bodies are naturally occurring structures in most bacterial species. They usually act as storage units for such compound such as glycogen and Polyhydroxyalkanoates. However, as recombinant proteins are over-expressed in bacterial expression systems, the sheer amount of protein can allow for proteins to aggregate forming inclusion bodies. This aggregation makes them insoluble within the cell and purification of functional protein is no longer obtainable. Caroline Roth, a former student in the lab, was able to express NS5B within bacterial cells, but could not prevent

the formation of inclusion bodies. A potential reason was the inadvertent inclusion of a Cysteine at the N-terminus of the protein, as well as biochemical properties that occurred during the over-expression process. The focus of this research was to clone GBV-C, express the protein in bacterial cells and remove the protein from inclusion bodies.

CHAPTER 2: AIMS

AIM #1: Cloning NS5B into an expression vector (pFN6K)

AIM #2: Expression of the tagged NS5B protein

AIM #3: Purification of the tagged NS5B protein

CHAPTER 3: MATERIALS AND METHODS

Section 1: GBV-C genome: The Iowan strain of the GBV-C genome (Reagent pAF121950) is present as cDNA on pCR2.1-TOPO plasmid, as previously described and was obtained through the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH: pAF121950 from Dr. Jinhua Xiang and Dr. Jack Stapleton (16), Appendix 1.

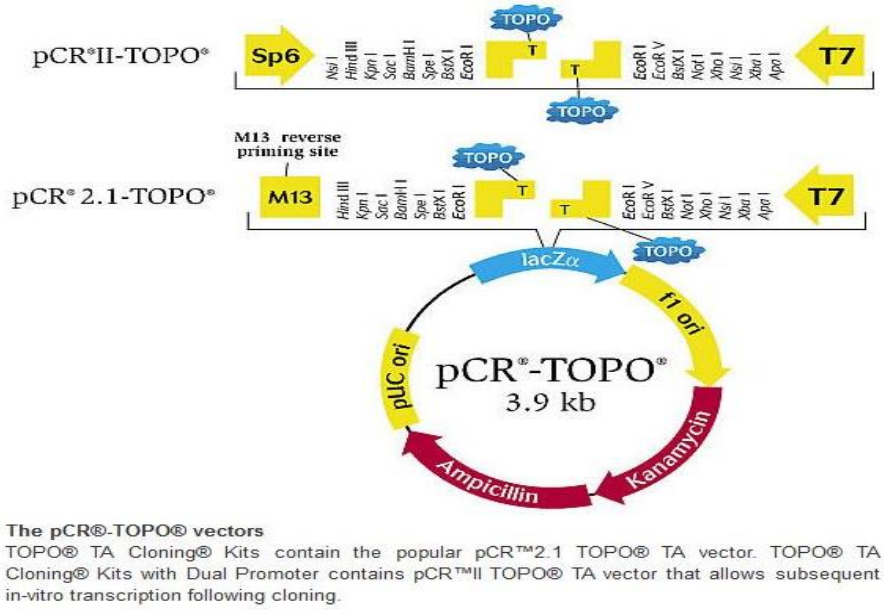


Figure 7: pCR2.1 TOPO (Invitrogen, Carlsbad, CA) vector map with the GBV-C cDNA genome.

The complete construct is 13,229bp (9,329bp for GBV-C's genome and 3,900 for pCR2.1 TOPO vector).

Section 2: NS5B coding sequence and PCR amplification of NS5B coding sequence

The GBV-C genome of 9.39kbp (GenBank accession # AF121950), is present as cDNA on the pCR2.1-Topo vector (Life technologies, Carlsbad, CA {Fig 6}). This construct was previously made by Xiang, J. et al. (Full-Length GB Virus C (Hepatitis G Virus) RNA Transcripts Are Infectious in Primary CD4-Positive T Cells, Jinhua Xiang,2000). First, the DH5α E. *coli* strains were transformed with pCR2.1, carrying the

GBV-C genome using Transform Aid Kit (Fermentas, Glen Burnie, MD). Transformed bacteria was selected by adding ampicillin(100μg/ml) to the LB-agar plates. Single colony will be picked and amplified in 4ml liquid LB with ampicillin(100μg/ml). The vector will be purified with Qiagen Miniprep Kit (Qiagen, Germantown, MD).

Section 3: Predicting the coding sequence for NS5B

The sequence of the GBV- C genome was retrieved form NCBI database (Hepatitis G virus strain Iowan, complete genome GenBank: AF121950.1). Since we know the genomic organization is conserved throughout the *Flaviviridae* family and that the NS5B is present in the upstream of the 3'UTR region of GBV-C that has been published (16), the genome was analyzed using GenBank (GenBank: AAD31765.1) for conserved domains with that of HCV-NS5B and the predicted GBV-C NS5B sequence was retrieved. The ExPASy (Swiss Institute of Bioinformatics (SIB)) translation tool was used to identify the correct open reading frame (ORF). The protein found to contain 563 amino acid residues. Also, the molecular weight of protein was estimated using ExPASy Compute pI/Mw (Swiss Institute of Bioinformatics (SIB)).

The translated coding sequence was run through BLASTp (National Institute of Health) which is a Basic Local Alignment Search Tool (BLAST) that the amino-acid sequences of proteins using algorithms. To amplify the NS5B coding sequence for cloning, primers were designed as per manufacturer protocol, with following integrated sites:

Forward Primer with SgfI site and start codon:

5'-GCAAGCGATCGCCATGTCCTTCTCTTACATTTGGTCT-3' (Tm = 65.9^{o}c)

Reverse primer with PmeI site:

5'-GAATGTTTAAACCCCGAAGAGGGCTACGATGAGCA-3' (Tm=64.9^{o}c)

The PCR ran for 30 cycles (1- 98°C for 1 min, 2-98°C for 20 sec, 3-55°C for 30 sec, 4-72°C for 1 min, 5- repeat step 2 to 4 29 times, 6- 72°C for 10 min) with 1X PCR

master mix 2X (containing Phusion DNA polymerase, dNTPS and $MgCl_2$, DMSO from NEB, Ipswich, MA), 0.5µM of forward and reverse primer and 172.1 ng of pCR2.1-TOPO to a 50µL reaction volume. After amplification of NS5B coding sequence with the specified primers, PCR products were run on a 1% agarose gel to quickly verify the correct length of the products.

Section 4: Cloning of NS5B Coding Region into Expression Vector – pFN6K

TheNS5B coding sequence was cloned into the expression vector pFN6K (Promega, Madison, WI, {Fig 5}), which allows the tagging of the protein at the N-terminal with a HQ tag for facilitated purification using metal affinity chromatography.

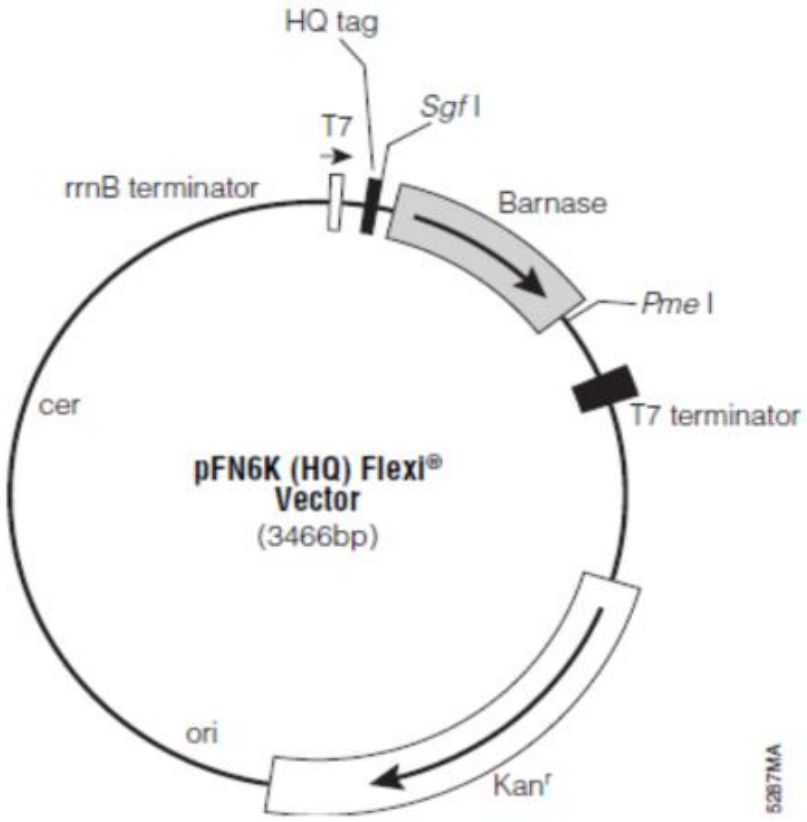


Figure 8. pFN6K expression vector: The pFN6K was purchased from Promega (Madison, WI).

The pFN6K Flexi vector contains a T7 promoter for bacterial expression of a protein coding region. The vector appends an N-terminal “HQ” tag coding region, which can be used to purify the expressed protein. The vector contains the lethal barnase gene for positive selection of the insert, a Kanamycin-resistance gene for selection of the plasmid and unique SgfI and PmeI sites, which allow easy insertion or transfer of the sequence of interest.

Section 4.1: PCR Product Purification

Gel extraction (QIAquick Gel Extraction Kit, Qiagen, Germantown, MD) was carried out to purify the PCR product of the amplified NS5B coding region along with the restriction sites. Once the agarose gel was run with samples from PCR of NS5B amplification, the bands containing the NS5B coding region in the appropriate molecular weight was cut from the gel and was subjected to gel extraction. The product thus obtained was run in agarose gel electrophoresis to check for clean band in the appropriate molecular weight.

Section 4.2: Agarose gel electrophoresis

One percent agarose gels were run to verify products obtained by plasmid purification using Spin Miniprep kit (Qiagen, Germantown, MD), PCR amplification of NS5B coding region, and restriction digestion of vector and digested coding region form the plasmid. To make 1% agarose gels, 0.5g of agarose powder was mixed with 50ml 1XTAE, with the addition of 2µl of 10mg/ml of Ethidium Bromide dye. The samples were added with appropriate amount of 6X loading dye from Invitrogen (Carlsbad, CA). Gels were run at for 50 – 60 minutes at 90 volts and Quick-Load® Purple 2-Log DNA Ladder (0.1 - 10.0 kb) (NEB, Ipswich, MA) was used to measure the appropriate size of the band.

Section 4.3: Restriction digest of the PCR product and the Vector

The PCR product was cut with SgfI and PmeI (Promega, Madison, WI) at 37°C for

60 min and restriction reaction was stopped by placing it at 65°C for 20 min. The cut PCR product will be cleaned of the small fragments created by treatment with restriction enzymes by using a PureLink™ PCR Purification Kit (Invitrogen, Carlsbad, CA).

The flexi pFN6K vector contains the lethal Barnase gene (335bp), which will be cut out by SgfI and PmeI digestion at 37°C for 30 min and reaction will be stopped at 65°C for 20 min (Promega, Madison, WI).

Section 4.4: Ligation of the Purified PCR Product and the Digested Vector

The digested vector and the purified PCR products will be mixed together at a ratio of 1:2 and ligated with T4 DNA ligase (New England Biolabs (NEB), Ipswich, MA) at 4^{o}C overnight, as well as at room temperature for two hours. After ligation the final product should be approximately 4.8kbp (3085bp vector + 1.7kb NS5B coding sequence), with the coding sequence in frame with the HQ tag, the expected size will be verified on a 1% agarose gel.

Section 5: Transformation of pFC7K Expression Vector, Carrying the NS5B Coding Region into BL21(DE3) E. *coli* cells for Expression

After Ligation, the correct construct will be transformed into competent BL21(DE3) *E. coli* cells using the TransformAid Kit (Fermentas, Glen Burnie, MD). The transformed bacteria will be selected on LB agar plates with Kanamycin (25µg/ml).

Section 5.1: Chemically competent bacterial cells

High efficiency chemically competent BL21(DE3) E. *coli* cells suitable for T7 protein expression were used. (catalog # c2566h, NEB, Ipswich, MA). The T7 express competent E. *coli* cells were streaked on an agar plate containing LB (Luria-Bertani) media and incubated at 37 ^{o}C, overnight. Once the colonies appear, single colony was inoculated into 2ml of C-medium (TransformAid, Thermo Fisher Scientific, Waltham, MA) and incubated at 37 ^{o}C, overnight in a shaker. The overnight bacterial culture was

ready for transformation of purified ligated product (NS5B coding region + pFN6K expression vector) into the T7 express competent E. *coli* cells.

Section 5.2: Bacterial Transformation

The tubes containing the TransformAid C-medium (1.5 ml of culture per transformation) were pre-warmed at 37 ^{o}C for 20 minutes. The agar plates containing the LB media (Luria – Bertani), supplemented with the antibiotic Kanamycin (25µg/ml), were pre-warmed at 37 ^{o}C for 20 minutes before plating.

1/10th volume of overnight culture containing the T7 express competent E. *coli* cells were added to the pre-warmed C-medium (1.5ml) and incubated at 37 ^{o}C shaker for 20 minutes. The transformation was carried as per the manufacturer's instructions in the TransformAid Bacterial Transformation Kit (Thermo Fisher Scientific, Waltham, MA) in the ratio of 10ng of ligated product (NS5B coding region +pFN6K vector) to a vial of 200µl of T7 express E. *coli* competent cells. The pre -warmed plates containing the LB media (Luria – Bertani) supplemented with antibiotic kanamycin (25µg/ml) cells were incubated with the cells resuspended with the ligation mixture at 37 ^{o}C, overnight.

Following a 24 hours incubation, the plates were checked for colony growth. Only bacteria that possess the transformed pFN6K vector plasmid would be able to survive the antibiotic - Kanamycin selective pressure. If the pFN6K vector does not have proper ligated NS5B coding region, the coding region for Barnase would prove lethal for the bacteria and result in bacterial death.

Section 6: Expression of GBV-C RNA polymerase (NS5B)

Section 6.1: Cell culture for protein induction

Single bacterial colony from transformed agar plates were inoculated into 5ml LB (Luria-Bertani) broth, supplemented with antibiotic- Kanamycin (25µg/ml) and incubated in 37 ^{o}C shaker, overnight. This culture was used for starter culture preparation. Two

different clones (clone A and clone B) from the transformed agar plates were used for protein induction.

Section 6.2: Starter culture preparation

One-tenth of the culture of transformed BL21(DE3) E. *coli* cell culture was inoculated each in two different 50 ml conical tubes containing 10ml LB broth media, supplemented with the Kanamycin (25µg/ml) and incubated at 37^{o}C overnight, in a shaker.

Section 6.3: Induction of NS5B Expression in Recombinant Clones

Clones (clone A and clone B) were grown in 50ml LB (Luria – Bertani) broth, supplemented with Kanamycin (25µg/ml) both for induction and for control (non - induced). Optical density (OD_{600}) of the bacterial cultures were measured at 600nm using spectrophotometer. Cultures were induced at different OD_{600} of 0.6, 0.8 and 1; to check the optimal OD_{600} level for protein expression. Other parameters such as duration of induction (2 hours, 4 hours, 6 hours, overnight) and induction temperature (37 ^{o}C and 30 ^{o}C) were carried out at different conditions to check for optimal protein expression. Different IPTG (Isopropyl β-D-1-thiogalactopyranoside) concentrations (40µM – 400µM from 100mM stock) were used for induction. After the induction, the cells were collected and centrifuged at 4 ^{o}C for 20 minutes at 6000 rpm.

The cell pellet was re-suspended in ice-cold PBS (Phosphate-buffered saline) and cells were centrifuged again and cell pellets collected for lysis. For every 10ml cell culture pellets, the following were added for lysis:

- 2ml fast-break cell lysis reagent (Promega, Madison, WI)
- 1µl DNase
- 10µl Lysozyme
- 500µl PMSF (Thermo Fisher Scientific, Waltham, MA)
- 500 µl EDTA-free Protease Inhibitor Cocktail (sigma-Aldrich, St. Louis, MO)

The resuspended cells were let to stand at room temperature for 20 minutes. Then they were sonicated with the Sonic Dismembrator Model 100 from (Thermo Fisher Scientific, Waltham, MA) on ice with a pulse mode on level 6, alternating between 10 seconds ON and 30 seconds OFF for a total of 5 cycles. The samples were then centrifuged for 30 minutes at 10000 rpm at 4 ^{o}C. Both pellets and supernatant were analyzed for protein expression using 4-12% Bis-Tris Plus SDS-PAGE (sodium dodecyl sulfate - polyacrylamide gel electrophoresis) gel. Both pellets and supernatant of non-Induced samples were analyzed.

Section 6.4: SDS-PAGE

The bolt® system (Invitrogen, Carlsbad, CA) was used for protein electrophoresis. The 4-12% Bis-Tris Plus Gels were selected. These gels are designed for optimal separation of a wide range of protein sizes under denaturing conditions. Both pellets and supernatant samples were treated with sample reducing agent and LDS sample buffer (Invitrogen, Carlsbad, CA) and then heated at 70°C for 15 minutes. Spectra multicolor broad range protein ladder (Thermo Fisher Scientific, Waltham, MA). Coomassie blue stain was used to stain the PAGE gels.

Section 7: Purification of the GBV-C RNA Polymerase (NS5B) by Immobilized Metal Affinity Chromatography (IMAC)

The expressed NS5B protein was purified using HisLink™ protein purification resin (Promega, Madison, WI). The resin contained a high level of tetradentate-chelated nickel for purifying expressed polyhistidine- or HQ-tagged proteins from bacteria. The resin was used in gravity – flow chromatography column where the cleared lysate was added over a HisLink™ column containing the resins and this method is usually sufficient for complete capture and efficient elution of polyhistidine- or HQ-tagged proteins.

First the column was added with sufficient volume of resins and it was equilibrated with 5 column volumes of binding buffer containing 100mM HEPES (pH

7.5), 10mM imidazole and 500mM NaCl. The clear lysate was then added to the resin and the rate of flow through was maintained as 1-2ml / minute. The column was washed with wash buffer containing 100mM HEPES (pH 7.5), 20mM imidazole and 500mM NaCl. Then the elution was carried out with various imidazole concentrations (100mM, 200mM, 500mM and 800m M) along with 100mM HEPES (pH 7.5), in order to release any NS5B protein bound to the column. The eluted fractions were checked on SDS-PAGE gel.

Section 7.1: DNA Purity and Concentration Assessment

The NanoDrop2000 from (ThermoFisher Scientific, (Waltham, MA) was used to quantify and assess purity of DNA.

Section 8: Homology Modeling of the GBV-C Virus RNA Polymerase

The NS5B protein sequence was translated from the nucleotide sequence of the Iowan genome (GenBank accession # AF121950). Homology modeling of the protein was conducted using Swiss model workspace. Swiss-Model is fully automated protein structure homology-modeling server. It assists and guides in building protein homology improved hierarchical approach for template selection. HCV polymerase (4kb7.1. A) was chosen as template as it has maximum sequence identity, compared to other templates.

Section 8.1: SAVES (The Structure Analysis and Verification Server)

SAVES, a structural analysis and verification server, was used to analyze protein structure using the following tools:

- ERRAT: ERRAT analyzes the statistics of non-bonded interactions between different atom types and plots the value of the error function versus position of a 9-residue sliding window, calculated by a comparison with statistics from highly refined structures.

- VERIFY_3D: VERIFY 3D determines the compatibility of an atomic model (3D)

with its own amino acid sequence (1D) by assigning a structural class based on its location and environment (alpha, beta, loop, polar, non-polar etc.) and comparing the results to good structures.

- PROCHECK: Checks the stereochemical quality of a protein structure by analyzing residue-by-residue geometry and overall structure geometry. It uses Ramachandran plot assessment which helps to visualize backbone dihedral angles ψ against φ of amino acid residues in protein structure and identify sterically allowed regions for these angles.

The predicted model provides a significant structure for GBV-C NS5B which can be further used in docking studies. For example, Ferron et al has worked on docking GBV-C with GTP Based on similar studies in HCV and BVDV (Bovine Viral Diarrhea Virus). Their docking results proposed that for the GTP binding site, motif E is a signature sequence and GTP is required for firmly holding the initiation complex (2).

Section 8.2: Superimposition of modeled NS5B from GB Virus- C with known structure of HCV NS5B using SuperPose version 1.0 (http://wishart.biology.ualberta.ca/superpose/))

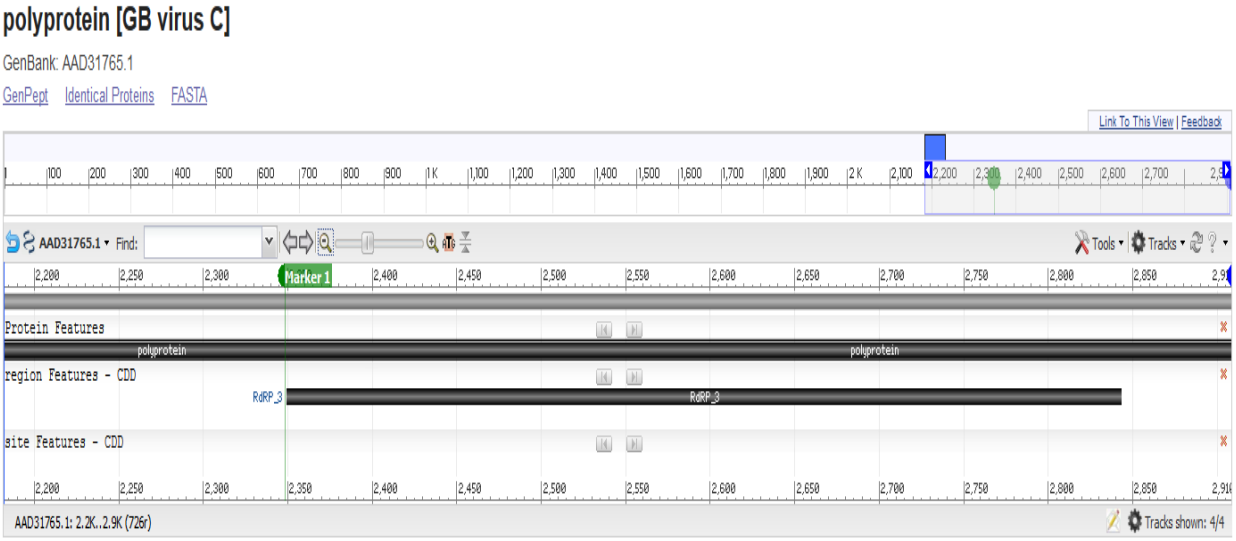
SuperPose is a server for superimposition of two or more structures that generates sequence alignment, structural alignments and RMSD values. The modeled structure of GBV-C NS5B was superimposed with known HCV NS5B structure taken from PDB (Protein Data Bank - https://www.rcsb.org/) (PDB ID: 4kb7.1. A). The RMSD (Root Mean Square Deviation) value gives quantitative measure of similarities in atomic co-ordinates between two superimposed structures. It is expressed in Angstrom (Å). The similar structures generally have lower RMSD value between 1 - 3 Å.

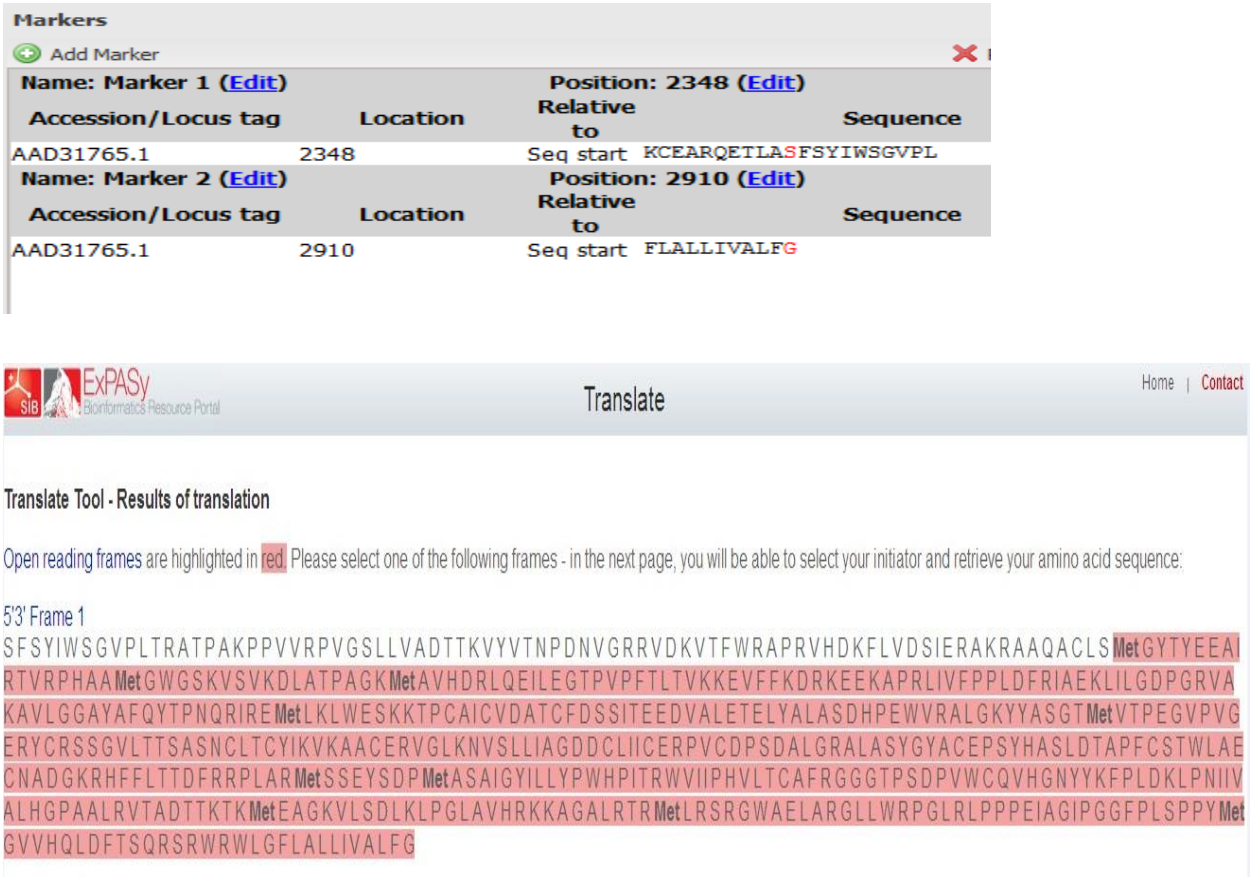
CHAPTER 4: RESULTS

Section 1: Finding NS5B Coding Sequence and Translation of Sequence

The NS5B coding region in the genome of GBV-C was found using NCBI-Gene, as mentioned in materials and methods. The NS5B genome was analyzed using NCBI-GenBank for conserved domains and by using HCV-NS5B and that of known published 3'UTR sequence of GBV-C, the NS5B gene sequence (Appendix 2) was predicted for GBV-C (Figure 6) and was found to be 1692bp long. The GBV-C NS5B gene sequence was then translated into protein sequence using ExPASy (Swiss Institute of Bioinformatics (SIB)) translate tool by finding the right ORF (Figure 6) and the protein sequence was found to be 563 amino acids long (Appendix 3). The molecular weight of protein was estimated using ExPASy Compute pI/Mw (Swiss Institute of Bioinformatics (SIB)) to be 62.29 KDa (Kilodaltons).

The BLASTp was used to verify the protein sequence identity with known sequences in the database and the results matched with various Blast hits with sequence identity of 100% and 99% and an E-value of 0 with the sequences of GB Virus-C polyprotein and putative NS5B of GB Virus-C (Appendix 4, 5 and 6).





Markers

Add Marker

Name: Marker 1 (Edit)		Position: 2348 (Edit)	
Accession/Locus tag	Location	Relative to	Sequence
AAD31765.1	2348	Seq start	KCEARQETLASFSYIWSGVPL
Name: Marker 2 (Edit)		**Position: 2910 (Edit)**	
Accession/Locus tag	Location	Relative to	Sequence
AAD31765.1	2910	Seq start	FLALLIVALFG

Figure 9: Sequence retrieval using NCBI Gene and protein translation using Expasy translate.

Section 2: Plasmid extraction of cDNA containing GBV-C genome; PCR amplification, cloning and transformation of NS5B coding region into expression vector

The pCR2. TOPO plasmid was extracted from the DH5-Alpha *E.coli* cells using mini-prep and it was run in 1% agarose gel. The lane showed the plasmid at appropriate band range (13Kb) (Figure 7).

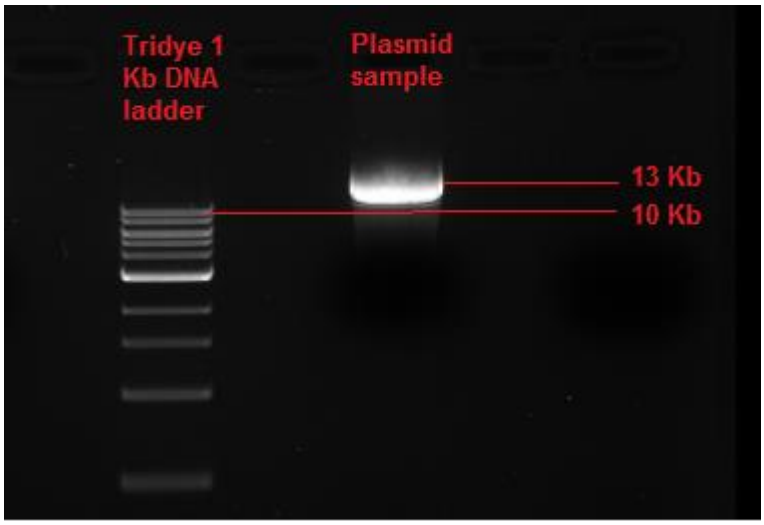


Figure 10: Agarose gel electrophoresis of mini-prep purified pCR2. TOPO carrying the whole GBV-C genome (~13 Kb). (Lanes showing DNA ladder and plasmid sample).

The NS5B coding region was amplified using PCR along with the incorporated restriction sites and the PCR product was run in 1% agarose gel for checking the band range of NS5B gene. The NS5b gene was found in appropriate band range (1.7 Kb) (Figure 8).

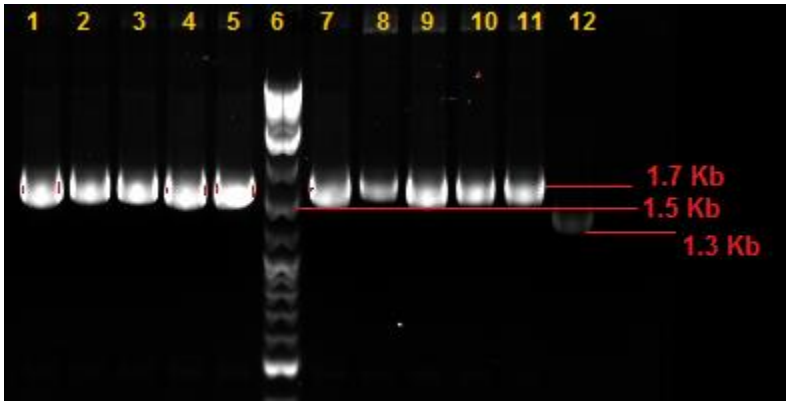


Figure 11: Agarose gel electrophoresis showing amplified NS5B gene with sgfI/pmeI primer pair. (Lanes 1-5 and 7-11 are PCR samples; Lane 6 is DNA ladder (TriDye™ 2-Log DNA Ladder); Lane 12 is Positive control of 1.3 Kb).

The PCR product that was purified by gel extraction was run in 1% agarose gel and the gel showed clear band of amplified NS5B gene in the range 1.7Kb (Figure 9).

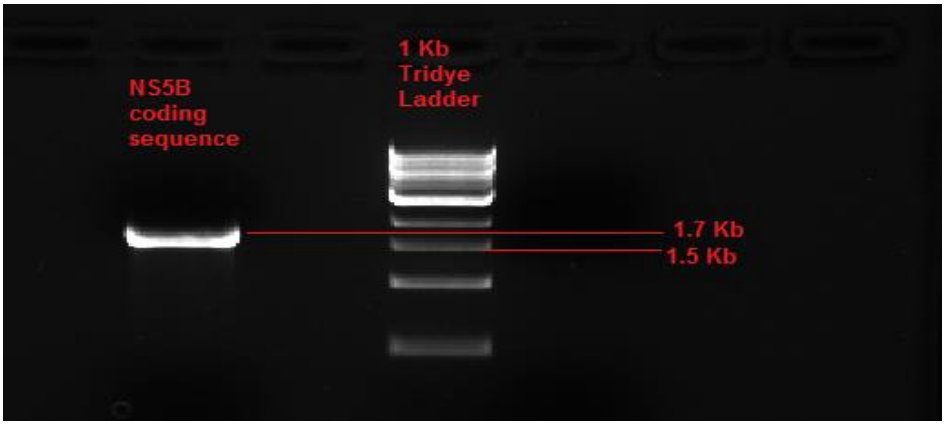


Figure 12: Agarose gel electrophoresis showing Gel extraction of Purified PCR product of amplified NS5B coding sequence. (Lanes showing NS5B coding region and DNA ladder).

After the NS5B gene was cloned into the expression vector and transformed into the *E.coli* BL21(DE3) cells and PCR amplification of NS5B was done to verify if NS5B is incorporated in to the vector present in the *E.coli* cells. The PCR amplified NS5B gene sequence was confirmed by running the samples in 1% agarose gels and the gel showed NS5B in appropriate band range of 1.7Kb (Figure 10).

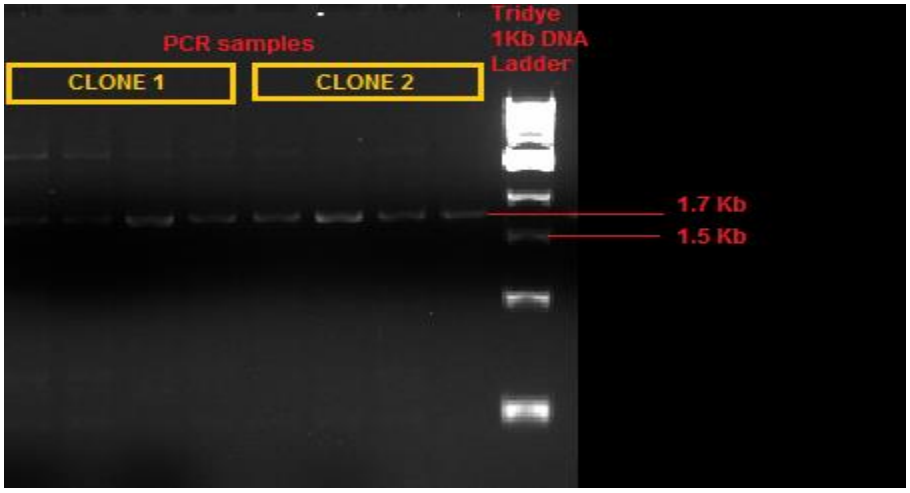


Figure 13: Agarose Gel electrophoresis showing PCR amplification of NS5B gene in plasmid present in the transformed clone to confirm presence of plasmid in the transformed clones. (Lanes 1-4 shows CLONE 1 sample; Lanes 5-8 shows CLONE 2 samples and Lane 9 is the DNA ladder).

Section 3: Expression and Purification of GB virus-C NS5B protein

Various expression parameters were tried to achieve optimal expression of GBV-C NS5B protein.

Protein Expression parameters:

A) OD_{600}: 0.6

Temperature (ºC)	IPTG Concentration(mM)	Induction duration (Hours)
37	**0.2**	**2,3,5,6**
	0.4	**2,3,5,6**
	0.6	**2,3,5,6**
	0.8	**2,3,5,6**
30	**0.4**	**8**
	1	**8**
25	**0.4**	**16.5**
	1	**16.5**

B) OD_{600}: 0.8

Temperature (°C)	IPTG Concentration(mM)	Induction duration (Hours)
37	**0.3**	**4**
30	**0.3**	**4**

C) OD_{600}: 1

Temperature (°C)	IPTG Concentration(mM)	Induction duration (Hours)
37	**0.5**	**1,2,4**
30	**0.3**	**4**
	0.4	**4**
	0.5	**3**
	0.6	**4**
	0.7	**4**

The optimal expression level was attained when cultures were induced at an OD_{600} of 0.8 for 4 hours of induction at 30°C with 400µM IPTG concentration. With the above parameters, the expression was found more in insoluble form in the pellets. Same induction method was carried out by changing the duration of induction to overnight at 25 °C with IPTG concentrations 400µM and 1mM and the expression was attained in the supernatant (Figure 11). The non- induced sample did not have any expression (Appendix 10 -Figure 20).

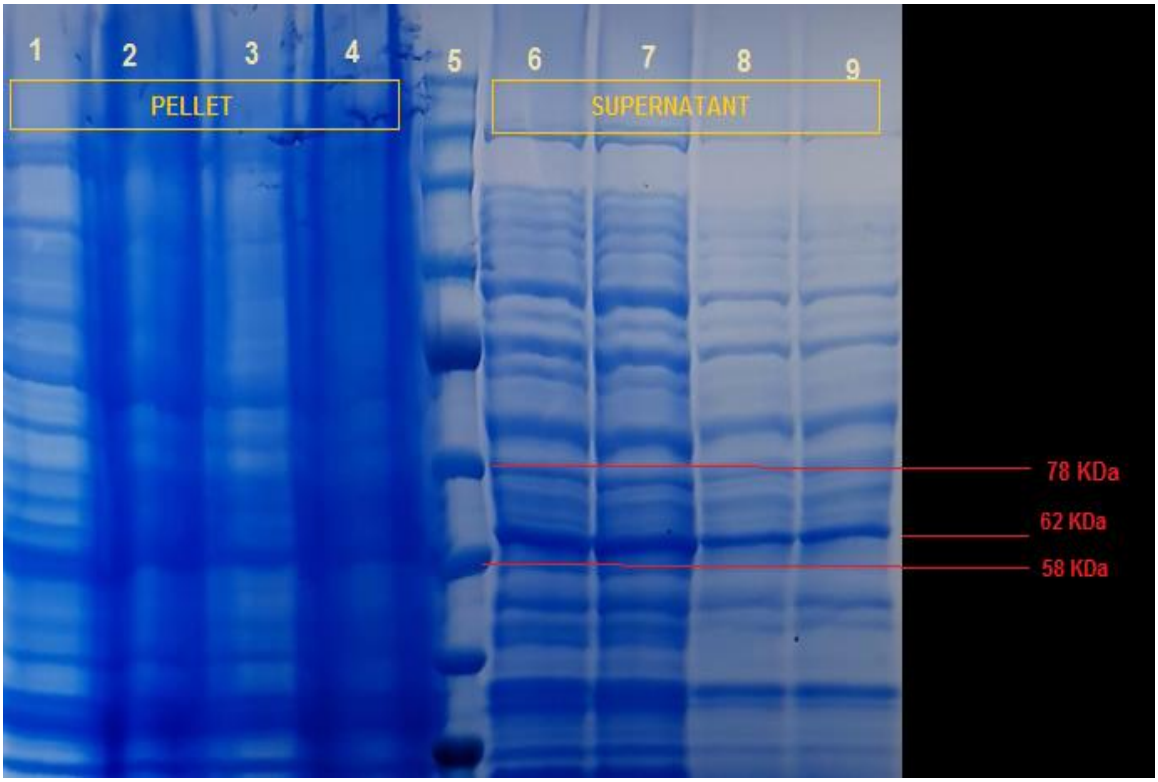


Figure 14: SDS-PAGE gel showing the expressed NS5B protein from GB Virus-C. Lanes 1 and 2 are pellets from clone 1; lanes 3 and 4 are pellets from clone 2; lane 5 is Protein standard; Lanes 6 and 7 are supernatant from clone 1; lanes 8 and 9 are supernatant from clone 2. Lanes 1,3,6 and 8 have IPTG concentration 400 µM. Lanes 2,4,7 and 8 have IPTG concentration 1mM.

The expressed protein from clone 1 was purified using IMAC column using His-Link resins and was eluted with 100mM Imidazole. The purified protein was obtained in second elution (Figure 12).

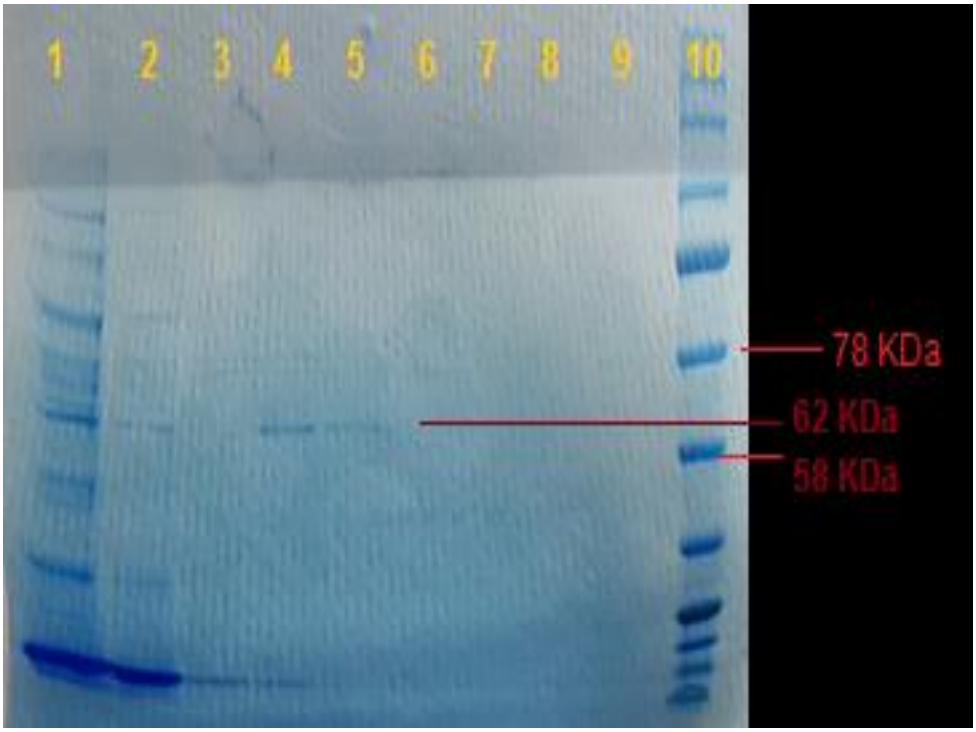


Figure 15: SDS-PAGE gel showing the purified NS5B protein from GB Virus-C using His-Link resin IMAC column. Lane 1 is cell lysate. Lane 2 is fraction from wash buffer. Lane 3 is fraction from first elution with 100mM Imidazole. Lane 4 is second elution with 100nM Imidazole which shows maximum fraction of purified protein. Lane 5-9 are fractions from following elution. Lane 10 is protein standard.

Section 4: Homology Modeling of the GBV-C Virus RNA Polymerase:

The homology modeling of translated protein sequence (Appendix 3) was done using Swiss Model Workspace with known HCV NS5B as template to predict the possible protein structure of GBV-C NS5B. The predicted structure was viewed in

Rasmol (http://www.openrasmol.org/) (Figure 13).

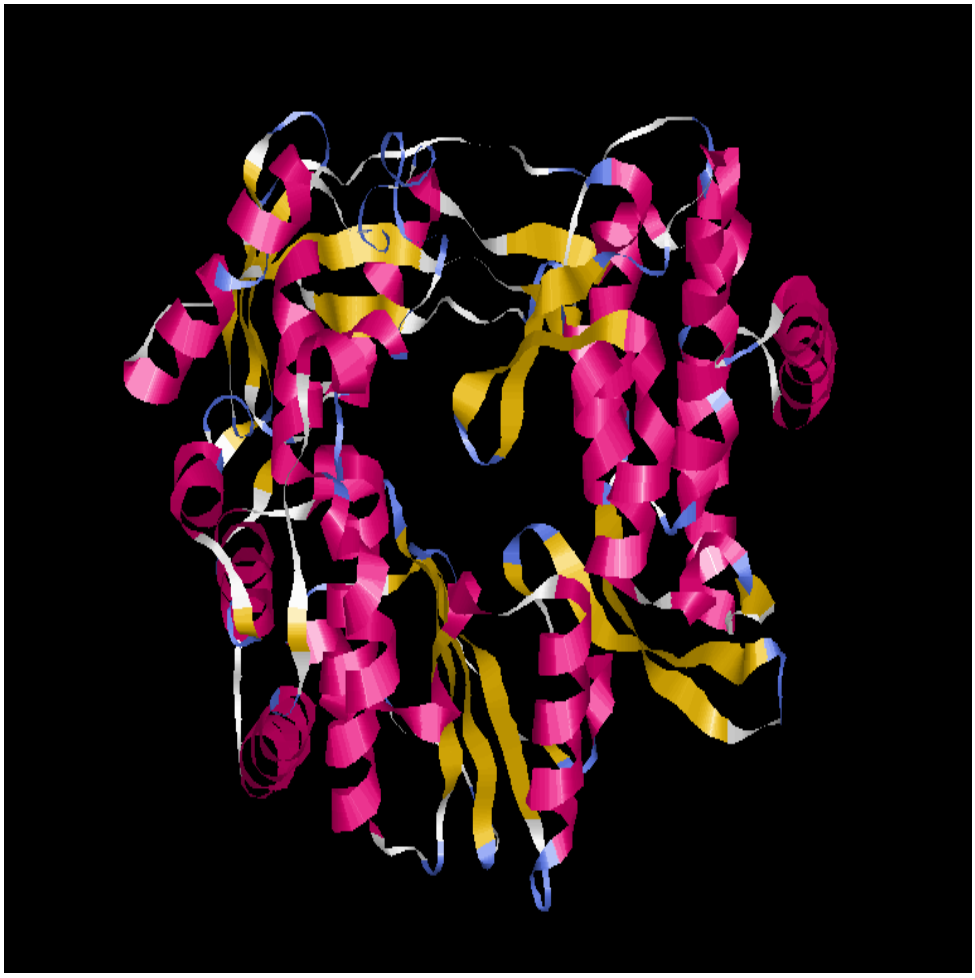


Figure 16: Modelled structure visualized using Rasmol viewer with Ribbons display colored based on structure. The Rasmol "structure color" scheme colors the molecule by protein secondary structure. Alpha helices are colored magenta, beta sheets are colored yellow, turns are colored pale blue and all other residues are colored white.

Section 5: Protein Structure Verification

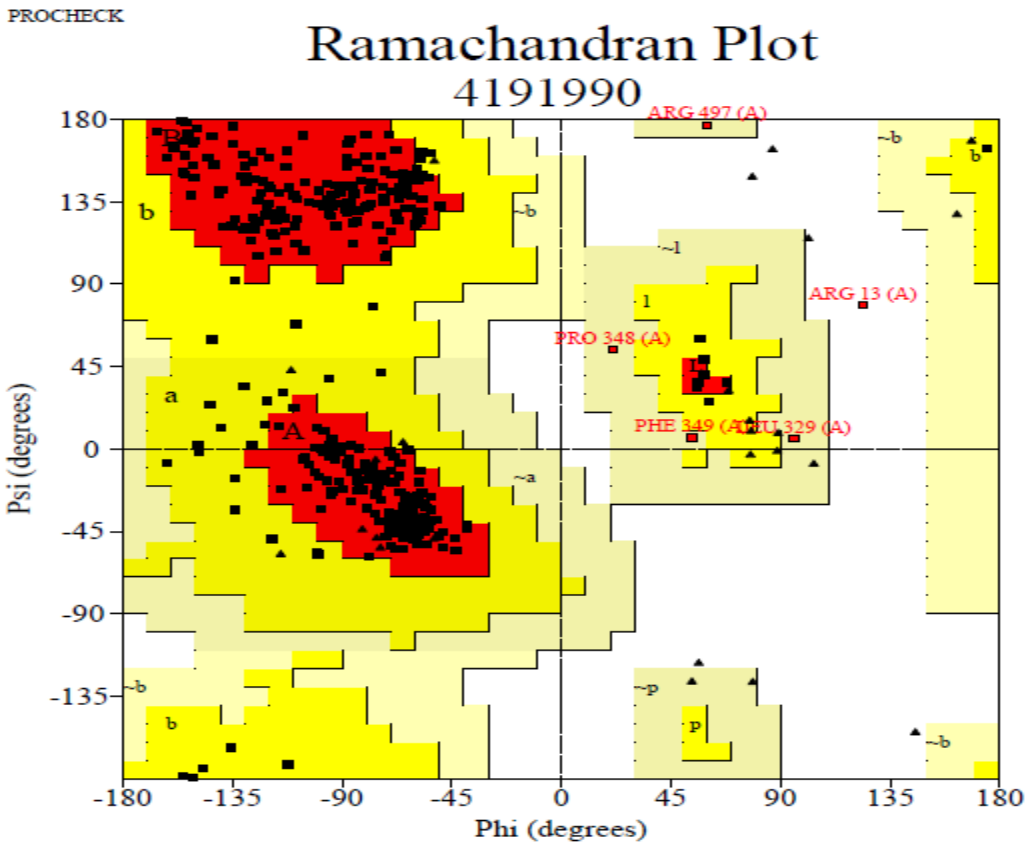
Section 5.1: SAVES result:

VERIFY3D: The result showed 93.80% of the residues had an averaged 3D-1D score >= 0.2 for VERIFY3D which is pass score. The result shows that the modeled protein structure is compatible with its sequence.

ERRAT: ERRAT has the Overall Quality Factor to be 91.1885. The preferred score for ERRAT is >95 %. The predicted model shows an ERRAT score in the higher range for a refined protein structure.

Section 5.2 Ramachandran Plot

The SAVES PROCHECK was used for Ramachandran plot assessment. According to Ramachandran plot assessment of a protein structure, a good quality model would be expected to have over 90% residues in favored region. The predicted model of NS5B has 92.1 % of residues in the favored region which shows a good quality structure for a protein (Figure 14).



Plot statistics

Residues in most favoured regions [A,B,L]	397	92.1%
Residues in additional allowed regions [a,b,l,p]	30	7.0%
Residues in generously allowed regions [~a,~b,~l,~p]	3	0.7%
Residues in disallowed regions	1	0.2%
Number of non-glycine and non-proline residues	431	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	32	
Number of proline residues	35	
Total number of residues	500	

Figure 17: Ramachandran plot for the modeled structure of NS5B, using SAVES PROCHECK.

Section 5.3: Superimposition of Modeled NS5B from GBV- C with the Known Structure of HCV NS5B Using SuperPose Software

The superimposition of the NS5B structure modeled from GBV-C (blue) and NS5B from known HCV structure (yellow), using Rasmol viewer shows clear similarity in structures (Figure 15). The RMSD value for the superimposed structure of modeled NS5B from GB Virus- C with known structure of HCV NS5B is 1.17 Å, showing that the two structures are similar in positions of atom co-ordinates.

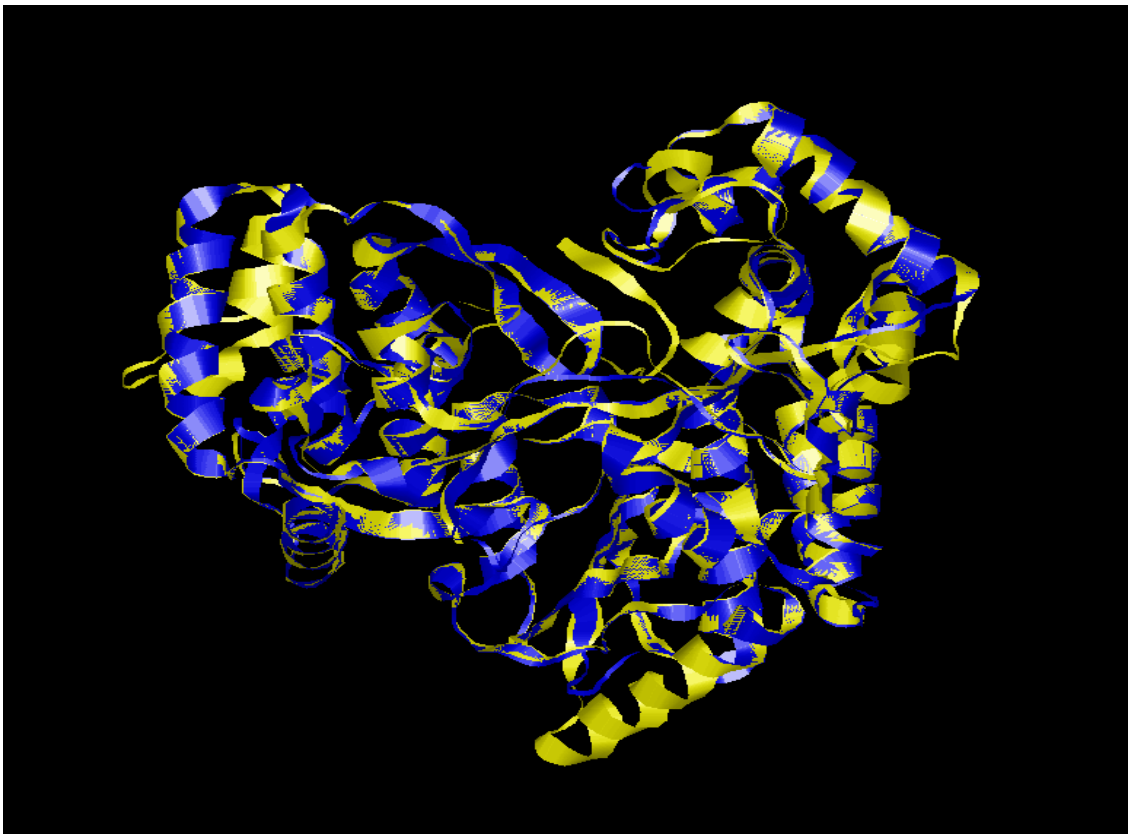


Figure 18: Superimposed structure of modeled GBV-C NS5B (blue color) with HCV NS5B (yellow color) viewed using Rasmol viewer (http://www.openrasmol.org/).

CHAPTER 5: CONCLUSION AND DISCUSSION

The coding sequence of NS5B (RNA dependent RNA polymerase) from GB virus- C was identified and cloned into a bacterial expression vector. The expression vector was transformed into BL21(DE3) cells for expression of protein. The protein was found to be expressed at the required band range of molecular weight. The expressed protein contained a polyhistidine tag and was purified using IMAC purification. The purified protein was obtained at appropriate band range of molecular weight.

The protein structure of NS5B (RNA dependent RNA polymerase) from GB virus-C was modelled using Swiss model workspace using the template nucleotide sequence of Iowan genome. The modelled structure was verified for energetically allowed regions of dihedral angles ψ against φ of amino acid residues in the backbone structure of protein. The result showed about 92.1% of residues in the favored region which is above the expected percentage of 90 for a good modeled protein structure. The superimposition of modeled NS5B structure from GBV-C with known NS5B structure from HCV showed greater similarity. This makes GBV-C NS5B as a potential model for HCV therapeutics.

The results of this research will aid in obtaining the crystal structure of the protein that can be used for crystallization studies using X-ray diffraction method. Further, the crystallized structure can be used to determine the 3-D structure of protein. The structure of protein is very important in providing a potential non-pathogenic model for therapeutic studies on pathogenic members of the Flaviviridae family, including hepatitis- C virus; by giving more insight on developing inhibitors against drug targets, vaccine development and other clinical research. The protein structure will also prove very helpful in obtaining a clear picture of the replication of GBV-C and its role in co-infection with HIV-1.

CHAPTER 6: REFERENCES

1. T. A. Steitz, t *et al,* A unified polymerase mechanism for nonhomologous DNA and RNA polymerases. Science 266, 2022-2025 (1994);

2. F. Ferron, C. Bussetta, *et al,* The modeled structure of the RNA dependent RNA polymerase of GBV-C virus suggests a role for motif E in Flaviviridae RNA polymerases. BMC Bioinformatics 6, 255 (2005);

3. F. Deinhardt, A. W. Holmes*et al,* Studies on the transmission of human viral hepatitis to marmoset monkeys. I. Transmission of disease, serial passages, and description of liver lesions. J Exp Med 125, 673-688 (1967);

4. A. S. Muerhoff, *et al,* Genomic organization of GB viruses A and B: two new members of the Flaviviridae associated with GB agent hepatitis. J Virol 69, 5621-5630 (1995);

5. J. N. Simons, *et al,* Identification of two flavivirus-like genomes in the GB hepatitis agent. Proc Natl Acad Sci U S A 92, 3401-3405 (1995);

6. Q. L. Choo, *et al,* Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science 244, 359-362 (1989);

7. J. Linnen, J. *et al,* Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. Science 271, 505-508 (1996);

8. J. N. Simons, *et al,* Isolation of novel virus-like sequences associated with human hepatitis. Nat Med 1, 564-569 (1995);

9. A. J. Zuckerman *et al,* Alphabet of hepatitis viruses. Lancet 347, 558-559 (1996);

10. L. Shao, *et al,* Sequence of hepatitis G virus genome isolated from a Japanese patient with non-A-E-hepatitis: amplification and cloning by long reverse transcription-PCR. Biochem Biophys Res Commun 228, 785-791 (1996);

11. H. J. Alter *et al,* The cloning and clinical implications of HGV and HGBV-C. N Engl J Med 334, 1536-1537 (1996);

12. D. Theodore*et al,* hepatitis G virus, or human orphan flavivirus? *Hepatology* **25**, 1285-1286 (1997);

13. Seipp S et al, Hepatotropism of GB virus C (GBV-C): GBV-C replication in human hepatocytes and cells of human hepatoma cell lines. (1999);

14. MARTA FOGEDA, et *al,* In Vitro Infection of Human Peripheral Blood Mononuclear Cells by GB Virus C/Hepatitis G Virus. (1999);

15. Cult *et al,* Clinical isolates of GB virus type C vary in their ability to persist and

replicate in peripheral blood mononuclear cell (2003);

16. Xiang J1 *et al,* Full-length GB virus C (Hepatitis G virus) RNA transcripts are infectious in primary CD4-positive T cells. (2000);

17. Scallan *et al,* Sexual transmission of GB virus C/hepatitis G virus. (1992);

18. Christensen *et al,* GB virus C epidemiology in Denmark: different routes of transmission in children and low- and high-risk adults. (2003);

19. M A Sathar *et al,* GB Virus C/Hepatitis G Virus (GBV-C/HGV): still looking for a disease (2000);

20. Jack T. Stapleton *et al,* The GB viruses: a review and proposed classification of GBV-A, GBV-C (HGV), and GBV-D in genus Pegivirus within the family Flaviviridae (2011);

21. H Keyvani *et al,* Prevalence of GBV-C RNA in HIV infected individuals in Tehran, Iran (2010);

22. Mohr EL *et al,* GB virus type C interactions with HIV: the role of envelope glycoproteins. (2009);

23. Polgreen PM *et al,* GB virus type C/hepatitis G virus: a non-pathogenic flavivirus associated with prolonged survival in HIV-infected individuals. (2003);

24. Jack T. Stapleton *et al,* GB virus type C/Hepatitis G virus. (2003);

25. Carolyn *et al,* Persistent GB Virus C Infection and Survival in HIV-Infected Men

26. Nattermann *et al,* Regulation of CC chemokine receptor 5 in hepatitis G virus infection. (2003);

27. Sarah L. *et al,* The GB Virus C (GBV-C) NS3 Serine Protease Inhibits HIV-

1 Replication in a CD4+ T Lymphocyte Cell Line without Decreasing HIV Receptor Expression

28. Jinhua Xiang* et al, An 85-aa segment of the GB virus type C NS5A

phosphoprotein inhibits HIV-1 replication in CD4 Jurkat T cells. (2006);

29. Heringlake S, *et al,* GB virus C/hepatitis G virus infection: a favorable prognostic factor in human immunodeficiency virus-infected patients? J. Infect. Dis. (1998);

30. Yeo AE, *et al,* Effect of hepatitis G virus infection on progression of HIV infection in patients with hemophilia. Multicenter Hemophilia Cohort Study. Ann. Intern. Med. (2000);

31. Tillmann HL, *et al,* Infection with GB virus C and reduced mortality among HIV-infected patients. N. Engl. J. Med. (2001);

32. Richard Allen White III *et al,* GBV-C acute infected cells reveal IL-16 cell mediated

downregulation of critical hosts proteins involved in HIV-1 replication (2009);

33. Ernst PETERHANS *et al,* Cytopathic bovine viral diarrhea viruses (BVDV):

emerging pestiviruses doomed to extinction. (2010);

34. Petersen LR *et al,* Shifting epidemiology of Flaviviridae. J Travel Med. (2005);

35. Hienz FX, Collett MS, Purcell RH, et al. Flaviviridae. In: van Ragenmorteil CM, Fauquet CM, Bishop DHK, *et al.,* eds. Taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses. San Diego, CA: Academic Press; (2000);

36. Christopher D. *et al,* Fatal Hemorrhagic Fever Caused by West Nile Virus in the United States (2006);

37. S Zaltron *et al,* Epidemiology of Hepatitis C Virus (HCV) Infection Theodore Sy,,,,,,Chronic HCV infection: epidemiological and clinical relevance (2012);

38. Arie J. Zuckerman *et al,* Medical Microbiology. 4th edition. Chapter 70Hepatitis Viruses (1996);

39. Thurner C *et al,* Conserved RNA secondary structures in Flaviviridae genomes. (2004);

40. Pieter Leyssen et al, Perspectives for the Treatment of Infections with Flaviviridae (2000);

41. Leary TP *et al,* Sequence and genomic organization of GBV-C: a novel member of the flaviviridae associated with human non-A-E hepatitis. (1996);

42. McLinden JH1 *et al,* Characterization of an immunodominant antigenic site on GB virus C glycoprotein E2 that is involved in cell binding. (2006);

43. Belyaev AS *et al,* Hepatitis G virus encodes protease activities which can affect processing of the virus putative nonstructural proteins. (1998);

44. Xiang J *et al,* GB virus type C NS5A sequence polymorphisms: association with interferon susceptibility and inhibition of PKR-mediated eIF2alpha phosphorylation.

, (2005);

45. Beames B1 *et al, GB* virus B as a model for hepatitis C virus. (2001);

46. Blight et al., 2000; Gale et al., 1998; Krieger et al. (2001);

47. François Ferron *et al,* The modeled structure of the RNA dependent RNA polymerase of GBV-C Virus suggests a role for motif E in Flaviviridae RNA polymerases. (2005);

48. Lefihre, et al, Prevalence of GB virus type /hepatitis G virus RNA and of anti-E2 in

individuals at high or low risk for blood-borne or sexually transmitted viruses: evidence of sexual and parenteral transmission. (1999);

49. Carolynne Schwarze-Zander, *et al*, GBV-C shares common routes of transmission with HCV and HIV, resulting in co- and triple-infection in humans. Role of GB virus C in modulating HIV disease. (2012);

50. Nirjal Bhattarai et *al,* GB virus C: the good boy virus? (2012);

51. Emma L. Mohr *et al,* GB Virus Type C Envelope Protein E2 Elicits Antibodies That

React with a Cellular Antigen on HIV-1 Particles and Neutralize Diverse HIV-1 Isolates.

(2010);

52. Jack T. Stapleton *et al,* GB Virus Type C: a Beneficial Infection? (2009);

53. Kyung H Choi *et al,* RNA-dependent RNA polymerases from Flaviviridae. (2009);

54. Jiqin Wu *et al,* A Structural Overview of RNA-Dependent RNA Polymerases from

the Flaviviridae Family. (2015);

55. Guitierrez et al, "Picomolar Inhibitors as Transition-State Probes of 5′-Methylthioadenosine Nucleosidases. (2007);

CHAPTER 7: APPENDIX

Appendix 1: GB virus- C Iowan strain:

>(AF121950) Deep sequenced Iowan strain NS5B 1689bp from 7,339 to 9027

ccgacgcctatctaagtagacgcaatgactcggcgccgacctggcgaccggccaaaaggtggtggatgggtgatgccagggt
tggtaggtcgtaaatcccggtcatcttggtagccactataggtgggtcttaagggaaggtcaagattcctcttgtgcctgtggcga
gacagcgcacggtccacaggtgttggccctaccggtgttaataagggcccgacgtcaggctcgtcgttaaaccgagcccgtca
cccacctgggcaaacgacgcccacgtacggtccacgtcgcccttcaatgtctctcttgaccaataggtttatccggcgagttgac
aaggaccagtgggggccgggggttatggggaaggaccccaaaccctgcccttcccggtgggccgggaaatgcatggggcc
acccagctccgcggcggcctgcagccggggtagcccaagaatccttcgggtgagggcgggtggcatttctcttttctataccat
catggcagtccttctgctccttctcgtggttgaggccggggccattctggccccggccacccacgcttgtcgagcgaatgggca
atatttcctcacaaattgctgtgccccggaagacatcgggttctgcctggaaggcggatgcctggtggccctggggtgcacggtt
tgcaccgaccgttgctggccactgtatcaggcgggtttggctgtgcggcctggcaagtccgcggcccagctcgttggggaact
ggggagcctgtacgggcccttgtcggtctcggcttacgtagccgggatcctgggtctgggcgaggtttactccggggtcctgac
agttggtgttgcgttgacgcgccgggtctacccgatgcccaacctgacgtgtgcagtagaatgtgagttaaagtgggaaagtga
gttttggagatggactgagcagttggcctccaattactggattctggaataccttggaaagtcccatttgacttttggagaggagtg
atgagcctgacccctctgttggtctgcgtggccgcattgctcttgctggagcaacggattgtcatggtcttcctgctggtgacgatg
gcggggatgtcgcaaggcgccccgcctccgttttggggtcccgccccttgactacgggttgacgtggcagtcatgctcctgc
agggctaacgggtcgcgtattcccactggggagagggtgtgggatcgagggaatgtcacgctcttgtgtgactgccccaacgg
cccctgggtctggctcccggccttctgccaggcggttggctggggcgaccccatcacccattggagccacggacaaaaccagt
ggcccctatcatgccccccaatatgtctatgggtctgtgtccgtaacgtgcgtgtggggttccgtgtcttggtttgcctcgaccggcg
gtcgtgattcgaagatcgatgtgtggagtttggtgccggttggatctgccagctgcaccatagccgctctagggtcatcggatcgc
gacacggtggttgagctctccgagtggggagtcccgtgcgtaacgtgtatcctggaccgtcggcctgcctcatgtggcacctgt
gtgcgggactgctggcccgaaaccgggtcggttagattccctttccatcggtgcggcacggggcctcggctgacaaaggactt
ggaagctgtgcccttcgtcaacaggacaactcccttcaccataaggggtcccctgggcaaccaggggagaggcaacccggtg
cggtcgcccctgggttttgggtcctacaccatgaccaagatccgggattccctgcatttggtgaaatgtcccacaccagccatag
agcctccaactggaacgttcgggttcttccccggagttccgccccttaacaactgcatgctgctaggcacggaagtgtctgaggc
attgggcggagctgggcttacgggggggttctacgagcctctggttcgcaggtgttcggagctgatgggacgccgaaatccgg
tttgcccggggtacgcatggctgtcctctggtagacctgacgggttcatacacgtccaggggcacctgcaggaggtggatgcg

ggcaacttcatccctcctccacgctggttgctcttggattttgtatttgtcctgctctatctgatgaagctggctgaggcacggctggt
cccgttgattctgctcctgctgtggtggtgggtgaaccagttggcagttctaggactgccggctgtggacgctgccgtggcgggt
gaagtttttgcgggccctgccttgtcatggtgtttgggccttcccactgtcagtatgatactaggtctagcaaacctggtgttgtatttt
cggtggatgggccctcagcgcctcatgttcctcgtgttgtggaagctcgctcggggagctttcccgctggcacttttgatggggat
ttcggcgacccgcgggcgcacctctgtgctcggggccgagttctgcttcgatgtcacattcgaggtggacacttcggtgttgggc
tgggtggtggccagcgtggtggcttgggccatagcgctcctgagctcaatgagcgcaggggggtggaagcacaaggccgtg
atctataggacgtggtgtaaagggtaccaggctgtgcgccagagggtggtgcggagccccctcggggaggggcgtcctacca
agcctctgacgttcgcctggtgcttggcctcatacatctggccggatgctgtgatgatggtggtggtggccttggtcctcctcttcg
gcctgttcgacgcactggactgggccctggaggagctcctggtctcccggccctcgttacggcgactggcacgggtggttgag
tgctgtgtgatggcgggcgagaaggccaccaccatccgactggtctccaagatgtgcgcaagaggggcctacctgtttgacca
catgggctctttctcgcgcgctgtcaaggagcgcttgttggaatgggacgcggctttggagcccttgtcattcactaggacggact
gtcgcatcatcagagatgccgcgaggaccctgtcctgcggacagtgcgtcatgggtttacccgtggtagcacggcgcggtgat
gaggttcttatcggcgtctttcaagatgtgaaccatttgcctcccgggtttgtcccgactgcaccagttgtcatccgtcggtgcgga
aagggcttcctgggggtcacgaaggcagccttgacaggtagggatcctgacttacatccagggaacgtcatggtgttggggac
ggctacgtcacgaagcatgggcacatgtctgaatggcctgctgttcacaactttccatggggcttcatcccgaaccatcgccacg
cccgtgggggcccttaatcccaggtggtggtcagccagtgatgacgtcacggtgtacccgcttccagatggggcaacttcgttg
acgccctgcacttgccaggcggagtcctgttgggttattagatccgacggggctttgtgccatggcttgagcaaggggacaag
gttgagctggatgtggccatggaggtctctgacttccgtggttcgtctggttcaccggtcctttgcgacgaagggcacgcagtag
gaatgctcgtgtcagtgctccactctggcggcagggttactgcggcgcgattcactaggccgtggactcaagtaccaacagatg
ccaagactaccacagaaccccctccggtgccggcaaaaggagttttcaaggaggccccgttgtttatgcctacgggggcggga
aagagcacccgcgtaccgttggagtacggcaacatgggccacaaggtcttgatcttgaacccgtcggtagctaccgtgagggc
catgggcccatacatggagcggctggcggggaaacaccccagtatttactgtggccatgacaccactgctttcacaaggatcac
tgactcgccccttacgtattccacttacggaaggtttttggccaaccctaggcagatgctgaggggtgtgtcggtggtcatttgtga
cgagtgccacagtcatgactcaactgtgttgttgggcattgggcgtgtcagggagctggcgcgaggatgtggagtgcaattggt
gctctacgccactgccacccctcccggatccccgatgacccagcatccatcaatcattgagacaaaactggacgtgggagagat
ccccttctatgggcatggcatacctcttgagcggatgcggaccggaaggcatctcgtattctgccactccaaggctgagtgcgag
cgcctggcgggccagttttcggctaggggggtaaatgccatcgcctattacaggggaaagacagctctatcatcaaagatgga
gacctggtggtgtgtgctacagacgcactatccactgggtacactgggaacttcgattctgtcaccgattgtgggttagtggtgga
ggaggtcgtcgaggtgacccttgatcccaccattaccatctccctgcgcacggtgcccgcatcggctgaactgtcgatgcagcg

gcgaggacgcacgggtaggggcaggtctgggcgctactactacgcgggggtcggcaaggcccctgctggtgtggtgcgctc
aggtcctgtctggtcggcggtggaagccggtgtgacctggtacggaatggaacctgacctgacagcaaacctactgagacttta
cgacgactgcccttacaccgcagccgtcgcagctgacattggggaagccgcggtgttcttttcggggcttgccccgttgaggat
gcatcccgatgttagctgggcaaaagttcgcggcgtcaactggcccctcctggtgggtgttcagcggaccatgtgccgggaaa
cactgtctcccggcccatcggatgacccccagtgggcaggtctgaagggcccgaatcctgtcccactcctgctgaggtggggc
aatgatttaccatctaaagtggccggccatcacatcgtggacgacctggtccgtaggctcggggtggcggagggttacgtccgc
tgcgatgcgggacccatcttgatggtgggcctcgctattgcgggggcatgatctatgcgtcatacaccgggtctctcgtggtgg
ttacagactgggatgtgaaggggggtggcagccccctttatcggcatggagaccaggccacgccccagccggttgtgcaggt
ccccccggtagaccatcggccgggggggagagtctgcgccatcggatgccaagacagtgacagatgcggtggcggccatcca
ggtggattgcgattggtcagtcatgaccctgtcgatcggggaagtgctgtccttggctcaggctaagacggccgaggcctacgc
agctaccaccaagtggcttgctggctgctacacggggacgcgggccgtccccactgtttcaattgttgacaagctcttcgccggg
ggctgggcggcggtggtgggccattgctacagtgtaatagctgcggcagtggcggcctacggggcttccaggagccctccatt
ggctgctgccgcttcctacctcatggggttgggcgtcggaggcaacgcacaaacccgcttagcctccgctctcctactaggggc
tgctgggaccgctctgggcacgcctgtcgtggggttaaccatggcgggcgcgttcatgggggtgctagcgtctccccctcctt
ggtcaccattttaccgggggccgtgggggggctgggagggcgtggtgaacgcggctagccttgtcttcgactttatggcgggga
aactatcatcagaagatctgtggtatgccatcccagtgctaaccagtccgggggcaggacttgcggggatcgccctcgggttgg
tgttgtactcagctaacaactctggcactaccacttggttgaaccgtctgctgactacattaccaaggtcctcatgcatccctgacag
ttactttcagcaggccgattactgtgacaaggtctcagctgtgctccgacgcttgagcctcactcgcaccgtggttgccctggtca
acagggagcctaaggtggatgaggttcaggtggggtacgtctgggacttgtgggagtggatcatgcgtcaagtgcgcatggtg
atggccagacttcgggccctctgccccgtggtgtcattacccttatggcactgcggggaggggtggtccggagaatggttgttg
gacggccatgttgagagtcgttgtctttgtggttgcgtgatcaccggtgatgttttgaatgggcaactcaaagaaccagtttactcta
ccaagctgtgcaggcattattggatggggacagtccctgtgaacatgctgggctatggcgagacatcgcctttgctcgcctcaga
caccccgaaggtggtaccattcgggacgtctgggtgggctgaggtggtggtgacccctacccacgttgtgatcaggcgaacat
ccgcctacaaactgctgcgccagcaaatcctgtcggctgctgttgctgagccctattacgtcgacggcataccggtctcatggga
cgcggacgcgcgagcgcctgccatggtctatggccctgggcaaagtgtcaccattgacggggaacgctacacccttccgcatc
aactgcggcttaggaatgtggcgccctctgaggtgtcatccgaggtgtccattgacattgggacggagactgaagactcagaac
tgactgaggccgacctgccgccggcggctgcagcccttcaggctatcgagaatgctgcgagaattcttgagcctcacatagatg
tcatcatggaggattgcagtacaccctctctttgtggaagtagccgagagatgcctgtgtggggagaagacataccccgcactcc
atcgccagcactaatctcggttactgagagcagcccagatgagaagaccccgtcggtgtcctcctcgcaggaggataccccgt

cctctgactcattcgaggtcatccaagagtccgagacagccgaaggggaggaaagcgtcttcaacgtggctctttccgtactaaa
agccttgtttccacagagcgatgccacaagaaagcttaccgttaagatgtcatgctgtgttgagaagagcgtaacacgcttcttttc
attgggattgacggtcgctgacgtggcaagcctgtgtgagatggaaatccagaaccatacagcctattgtgacaaggtgcgcac
tccgcttgaattgcaggttgggtgcttggtgggcaatgaacttacctttgaatgtgacaagtgtgaggctaggcaagagaccttgg
cttctttctcttacatttggtctggggtgccactgacgagggccactccggccaagccccctgtggtgaggccggttggctccttg
ctggtggccgacaccaccaaggtgtatgtcaccaacccggacaatgttgggagaagagttgacaaggttaccttctggcgtgcc
cctagggttcatgacaaattcctcgtggactccatagagcgcgctaagagggcagctcaagcctgcctaagcatgggttacactt
atgaggaggcaataaggactgtaaggccacatgctgccatgggctggggatctaaggtgtcggtcaaggacctcgccacccct
gcggggaagatggctgtccatgaccggctccaggagatacttgaagggacgccagtccccctttactcttactgtgaaaaaggag
gtgttcttcaaagaccgaaaggaggagaaggcccccccgcctcattgtgttcccccccctggacttccggatagctgaaaagctta
ttctgggagacccgggacgggtagccaaggcggtgttgggggggcttacgccttccagtacaccccaaatcagcgagttag
ggagatgctcaaactgtgggagtcaaagaaaacaccatgcgccatctgtgtggacgccacatgcttcgacagtagcataactga
agaggacgtggcgctggagacagagctttatgccctggcttcagaccatccagaatgggtgcgtgccctggggaaatactatg
cctctggcacaatggtaacccccgaggggtgccagtgggtgagaggtattgtagatcctcaggggtcttgaccaccagtgcg
agcaactgcttgacttgctatatcaaggtgaaagccgcctgtgagagggtggggctgaaaaatgtctcgctcctcatcgctggcg
atgactgtttgatcatatgcgaacggcctgtgtgcgatcctagcgacgctttgggcagagccctggcgagctacgggtacgcatg
cgagccttcgtatcatgcatcactggacacggccccttctgctccacttggctagctgagtgcaatgcagatgggaaacgccat
ttcttccttaccacggactttcggaggcccctcgctcgcatgtcgagcgagtacagcgacccaatggcttcggccatcggttacat
cctcctataccttggcatcccatcacacggtgggtcatcatccctcacgtgctcacctgcgcgtttaggggtggtggcacaccgt
ctgatcctgtgtggtgccaggtacatggtaattactacaagtttccactggacaaactgcctaacatcatcgtggccctccacgga
ccagcagcgtcgagggttaccgcagacacaaccaagacaaaaatggaggctggcaaggtgctgagcgacctcaagctccct
ggcctagcagtccaccggaagaaggccggggcattgcgaacgcgtatgctccggtcgcgcggttgggctgagttggctaggg
ggctgttgtggcatccaggcctgcggcttccccccccggagattgctggtatccccgggggtttccccctttcccccccctatatg
ggggtggttcatcaattggattttacaagccagaggagtcgctggcggtggttggggttcttagccctgctcatcgtagccctcttc
gggtgaactaaattcatctgttgcggcaaggtccggtgactgatcatcactggaggaggttcccgccctccccgccccaggggt
ctccccgctgggtaaaaagggcccggccttgggaggcatggtggttactaacccccctggcagggtcaaagcctgatggtgcta
atgcactgccacttcggtggcgggtcgctaccttatagcgtaatccgtgactacgggctgctcgcagagccctccccggatggg
gcacagtgcactgtgatctgaaggggtgcaccccggtaagagctcggcccaaag

Appendix 2: NS5B predicted gene sequence:

>NS5B PREDICTED GENE SEQUENCE

tccttctcttacatttggtctggggtgccactgacgagggccactccggccaagccccctgtggtgaggccggttggctccttgct
ggtggccgacaccaccaaggtgtatgtcaccaacccggacaatgttgggagaagagttgacaaggttaccttctggcgtgccc
ctagggttcatgacaaattcctcgtggactccatagagcgcgctaagagggcagctcaagcctgcctaagcatgggttacactta
tgaggaggcaataaggactgtaaggccacatgctgccatgggctggggatctaaggtgtcggtcaaggacctcgccacccctg
cggggaagatggctgtccatgaccggctccaggagatacttgaagggacgccagtccccctttactcttactgtgaaaaaggaag
tgttcttcaaagaccgaaaggaagagaaggcccccccgcctcattgtgttcccccccctggacttccggatagctgaaaagcttatt
ctgggagaccctggacgggtagccaaggcggtgttgggggggggcctacgccttccagtacaccccaaatcagcgaattaggg
agatgctcaaactgtgggaatcaaagaagacaccatgcgccatctgtgtggacgccacatgcttcgacagtagcataactgaag
aggacgtggcgctggagacagagctttatgccctggcttcagaccatccagaatgggtgcgtgccctggggaaatactatgcct
ctggcacaatggtaacccccgaggggtgccagtgggtgagaggtattgtagatcctcaggggtcttgaccaccagtgcgagc
aactgcttgacttgctatatcaaggtgaaagccgcctgtgagagggtggggctgaaaaatgtctcgctcctcatcgctggcgatg
actgtttgatcatatgcgaacggcctgtgtgcgatcctagcgacgctttgggcagagccctggcgagctacgggtacgcatgcg
agccttcgtatcatgcatcactggacacggcccccttctgctccacttggctagctgagtgcaatgcagatgggaaacgccatttc
ttcctgaccacggactttcggaggcccctcgctcgcatgtcgagcgagtacagtgacccaatggcttcggccatcggttacatcc
tcctataccttggcatcctatcacacggtgggtcatcatccctcacgtgctcacctgcgcgtttaggggtggtggcacaccgtct
gatcctgtgtggtgccaggtacatggtaattactacaagtttccactggacaaactgcctaacatcatcgtggccctccacggacc
agcagcgttgagggttaccgcagacacaactaagacaaaaatggaggctggcaaggtgctgagcgacctcaagctccctggc
ctagcagtccaccggaagaaggccggggcattgcgaacgcgtatgctccggtcgcgcggttgggctgagttggctaggggggc
tgttgtggcgtccaggcctgcggcttccccctccggagattgctggtatccccgggggtttccccctttcccccccctatatgggg
gtggttcatcaattggatttcacaagccagaggagtcgctggcggtggttggggttcttagccctgctcatcgtagccctcttcgg
gtga

Appendix 3: NS5B coding sequence translated by Expasy Translate:

Number of amino acids: 563 Molecular Weight: 62.22 KDa

>VIRT124745

SFSYIWSGVPLTRATPAKPPVVRPVGSLLVADTTKVYVTNPDNVGRRVDKVTFW
RAPRVHDKFLVDSIERAKRAAQACLSMGYTYEEAIRTVRPHAAMGWGSKVSVK
DLATPAGKMAVHDRLQEILEGTPVPFTLTVKKEVFFKDRKEEKAPRLIVFPPLDF
RIAEKLILGDPGRVAKAVLGGAYAFQYTPNQRVREMLKLWESKKTPCAICVDAT
CFDSSITEEDVALETELYALASDHPEWVRALGKYYASGTMVTPEGVPVGERYCR
SSGVLTTSASNCLTCYIKVKAACERVGLKNVSLLIAGDDCLIICERPVCDPSDALG
RALASYGYACEPSYHASLDTAPFCSTWLAECNADGKRHFFLTTDFRRPLARMSS
EYSDPMASAIGYILLYPWHPITRWVIIPHVLTCAFRGGGTPSDPVWCQVHGNYYK
FPLDKLPNIIVALHGPAASRVTADTTKTKMEAGKVLSDLKLPGLAVHRKKAGAL
RTRMLRSRGWAELARGLLWHPGLRLPPPEIAGIPGGFPLSPPYMGVVHQLDFTSQ
RSRWRWLGFLALLIVALFG

Appendix 4: Similarity of translated protein sequence with previously found sequences.

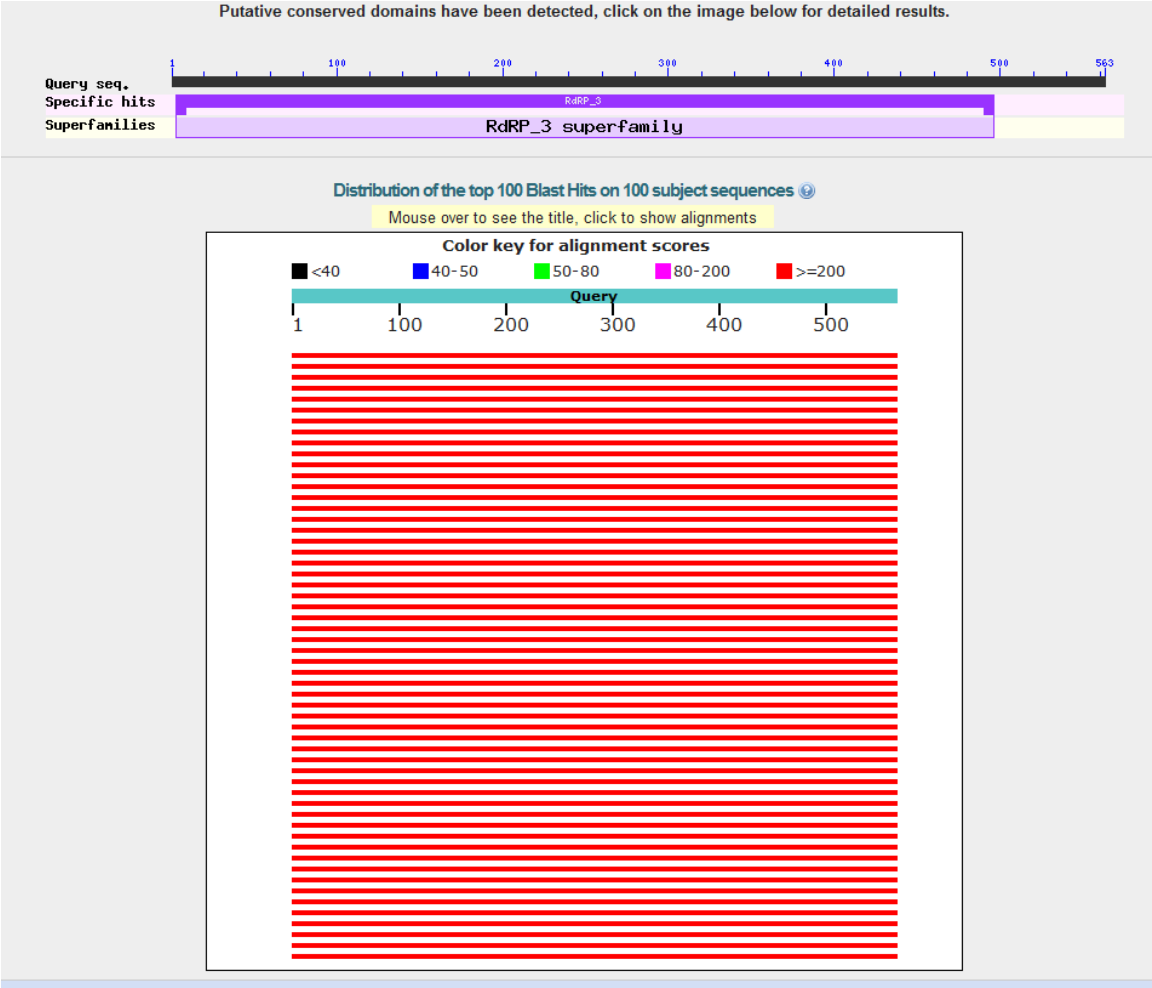


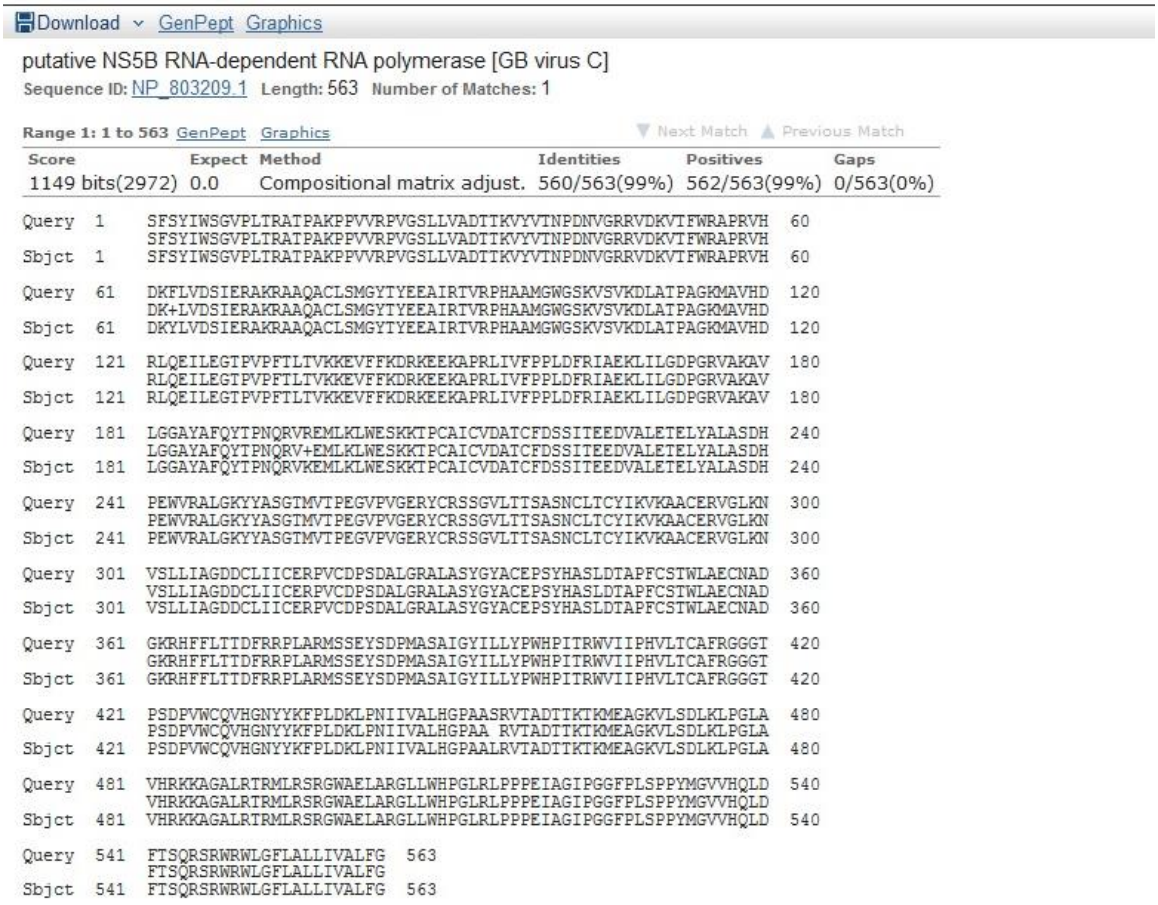
Figure 19: Protein BLAST result showing the hits of translated NS5B protein

Appendix 5:

	Max score	Total score	Query cover	E value	Ident	Accession
polyprotein precursor [GB virus C]	1154	1154	100%	0.0	99%	NP_043570.1
polyprotein [GB virus C]	1154	1154	100%	0.0	99%	BAA23103.1
polyprotein [GB virus C]	1153	1153	100%	0.0	99%	BAA19580.1
polyprotein [GB virus C]	1153	1153	100%	0.0	99%	AAC09229.1
polyprotein [Human pegivirus]	1152	1152	100%	0.0	99%	CVH74182.1
polyprotein [Human pegivirus]	1152	1152	100%	0.0	99%	CVH74189.1
unnamed protein product [GB virus C]	1152	1152	100%	0.0	99%	BAA13321.1
polyprotein [GB virus C]	1152	1152	100%	0.0	99%	AAD31765.1
polyprotein [Human pegivirus]	1151	1151	100%	0.0	99%	CVH74180.1
polyprotein [Pegivirus A]	1151	1151	100%	0.0	99%	CVH74176.1
polyprotein [Pegivirus A]	1151	1151	100%	0.0	99%	CVH74174.1
polyprotein [Human pegivirus]	1151	1151	100%	0.0	99%	CVH74178.1
polyprotein [Pegivirus A]	1151	1151	100%	0.0	99%	CVH74173.1
polyprotein [GB virus C]	1150	1150	100%	0.0	98%	BAA23723.1
putative NS5B RNA-dependent RNA polymerase [GB virus C]	1149	1149	100%	0.0	99%	NP_803209.1
polyprotein [GB virus C]	1149	1149	100%	0.0	99%	BAA22479.1
polyprotein [GB virus C]	1149	1149	100%	0.0	99%	BAA22476.1
polyprotein [GB virus C]	1148	1148	100%	0.0	99%	AHA61262.1
polyprotein [GB virus C]	1148	1148	100%	0.0	99%	AAA96965.1
polyprotein [GB virus C]	1147	1147	100%	0.0	99%	BAA22988.1
polyprotein [Human pegivirus]	1147	1147	100%	0.0	99%	CVH74184.1
polyprotein [Human pegivirus]	1147	1147	100%	0.0	99%	CVH74181.1

Figure 20: BLAST hits with sequence identity of 100% and 99% and with an E-value of 0.

Appendix 6:



Download GenPept Graphics

putative NS5B RNA-dependent RNA polymerase [GB virus C]

Sequence ID: NP_803209.1 Length: 563 Number of Matches: 1

Range 1: 1 to 563 GenPept Graphics Next Match Previous Match

Score	Expect	Method	Identities	Positives	Gaps
1149 bits(2972)	0.0	Compositional matrix adjust.	560/563(99%)	562/563(99%)	0/563(0%)

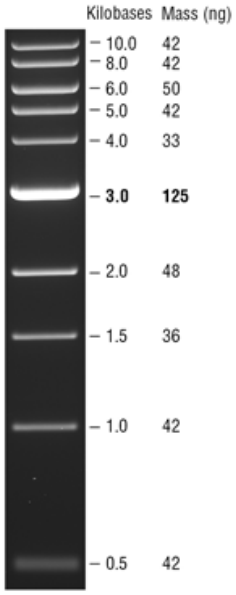
```
Query  1    SFSYIWSGVPLTRATPAKPPVVRPVGSLLVADTTKVYVTNPDNVGRRVDKVTFWRAPRVH  60
            SFSYIWSGVPLTRATPAKPPVVRPVGSLLVADTTKVYVTNPDNVGRRVDKVTFWRAPRVH
Sbjct  1    SFSYIWSGVPLTRATPAKPPVVRPVGSLLVADTTKVYVTNPDNVGRRVDKVTFWRAPRVH  60

Query  61   DKFLVDSIERAKRAAQACLSMGYTYEEAIRTVRPHAAMGWGSKVSVKDLATPAGKMAVHD  120
            DK+LVDSIERAKRAAQACLSMGYTYEEAIRTVRPHAAMGWGSKVSVKDLATPAGKMAVHD
Sbjct  61   DKYLVDSIERAKRAAQACLSMGYTYEEAIRTVRPHAAMGWGSKVSVKDLATPAGKMAVHD  120

Query  121  RLQEILEGTPVPFTLTVKKEVFFKDRKEEKAPRLIVFPPLDFRIAEKLILGDPGRVAKAV  180
            RLQEILEGTPVPFTLTVKKEVFFKDRKEEKAPRLIVFPPLDFRIAEKLILGDPGRVAKAV
Sbjct  121  RLQEILEGTPVPFTLTVKKEVFFKDRKEEKAPRLIVFPPLDFRIAEKLILGDPGRVAKAV  180

Query  181  LGGAYAFQYTPNQRVREMLKLWESKKTPCAICVDATCFDSSITEEDVALETELYALASDH  240
            LGGAYAFQYTPNQRV+EMLKLWESKKTPCAICVDATCFDSSITEEDVALETELYALASDH
Sbjct  181  LGGAYAFQYTPNQRVKEMLKLWESKKTPCAICVDATCFDSSITEEDVALETELYALASDH  240

Query  241  PEWVRALGKYYASGTMVTPEGVPVGERYCRSSGVLTTSASNCLTCYIKVKAACERVGLKN  300
            PEWVRALGKYYASGTMVTPEGVPVGERYCRSSGVLTTSASNCLTCYIKVKAACERVGLKN
Sbjct  241  PEWVRALGKYYASGTMVTPEGVPVGERYCRSSGVLTTSASNCLTCYIKVKAACERVGLKN  300

Query  301  VSLLIAGDDCLIICERPVCDPSDALGRALASYGYACEPSYHASLDTAPFCSTWLAECNAD  360
            VSLLIAGDDCLIICERPVCDPSDALGRALASYGYACEPSYHASLDTAPFCSTWLAECNAD
Sbjct  301  VSLLIAGDDCLIICERPVCDPSDALGRALASYGYACEPSYHASLDTAPFCSTWLAECNAD  360

Query  361  GKRHFFLTTDFRRPLARMSSEYSDPMASAIGYILLYPWHPITRWVIIPHVLTCAFRGGGT  420
            GKRHFFLTTDFRRPLARMSSEYSDPMASAIGYILLYPWHPITRWVIIPHVLTCAFRGGGT
Sbjct  361  GKRHFFLTTDFRRPLARMSSEYSDPMASAIGYILLYPWHPITRWVIIPHVLTCAFRGGGT  420

Query  421  PSDPVWCQVHGNYYKFPLDKLPNIIVALHGPAASRVTADTIKTKMEAGKVLSDLKLPGLA  480
            PSDPVWCQVHGNYYKFPLDKLPNIIVALHGPAA RVTADTIKTKMEAGKVLSDLKLPGLA
Sbjct  421  PSDPVWCQVHGNYYKFPLDKLPNIIVALHGPAALRVTADTIKTKMEAGKVLSDLKLPGLA  480

Query  481  VHRKKAGALRTRMLRSRGWAELARGLLWHPGLRLPPPEIAGIPGGFPLSPPYMGVVHQLD  540
            VHRKKAGALRTRMLRSRGWAELARGLLWHPGLRLPPPEIAGIPGGFPLSPPYMGVVHQLD
Sbjct  481  VHRKKAGALRTRMLRSRGWAELARGLLWHPGLRLPPPEIAGIPGGFPLSPPYMGVVHQLD  540

Query  541  FTSQRSRWRWLGFLALLIVALFG  563
            FTSQRSRWRWLGFLALLIVALFG
Sbjct  541  FTSQRSRWRWLGFLALLIVALFG  563
```

Figure 21: BLAST result showing alignment between translated protein sequence of GBV-C NS5B (by Expasy) and BLAST sequence identity with putative NS5B RNA-dependent RNA polymerase [GBV-C]

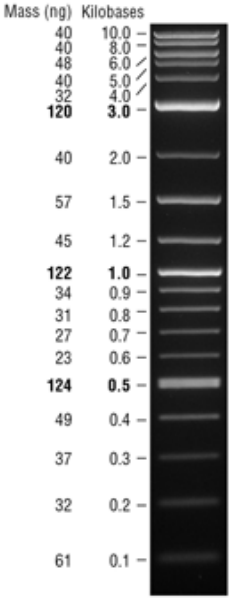
Appendix 7: Nanodrop readings of plasmid purified using Miniprep:

	SAMPLE
CONCENTRATION	110.8
A260	2.217
A280	1.151
260/280	1.93
260/230	2.48

Appendix 8: TriDye™ 1 kb DNA Ladder:



Appendix 9: TriDye™ 2-Log DNA Ladder:



Appendix 10: Non-Induced expression of GBV-C NS5B:

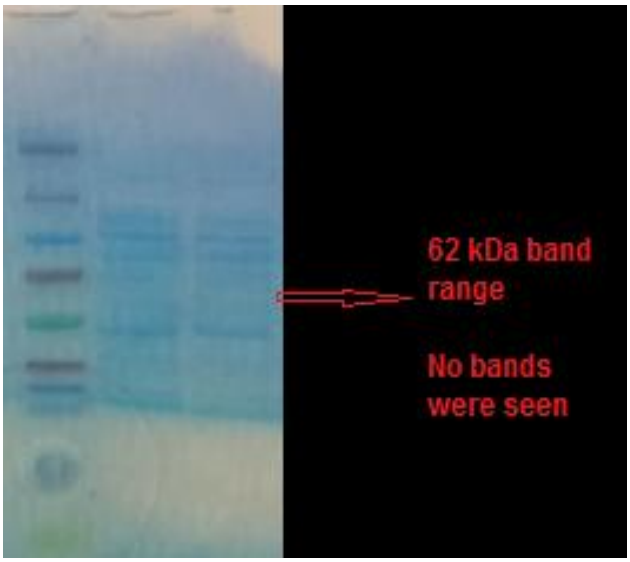
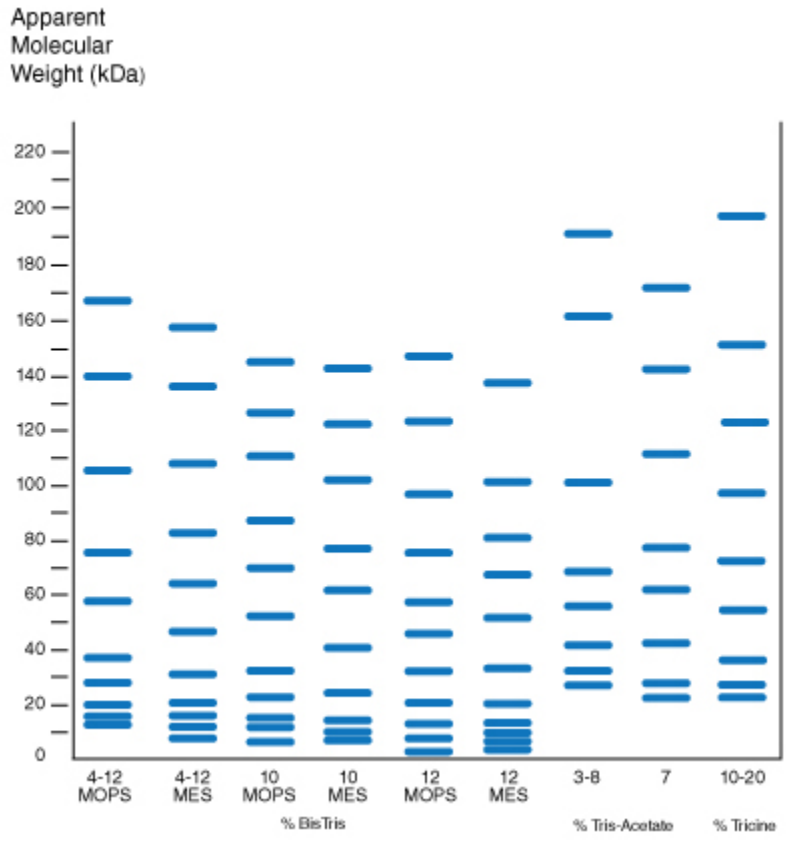


Figure 22: SDS Page showing Non-Induced expression of GBV-C NS5B. No expression was seen in the transformed BL21(DE3) cells when the expression was not induced with IPTG (0.4mM)

Appendix 11: Blue Prestained Protein Standard, Broad Range (11-190 kDa)

NEB (Ipswich, MA: In 4-12 % Bis-Tris gel with MOPS buffer



Appendix 12: The Rasmol results of the protein structure information:

```
RasMol> show information
Secondary Structure ... Calculated
Experiment Technique .. THEORETICAL MODEL (SWISS-MODEL SERVER)
Number of Groups ..... 500
Number of Atoms ...... 3883
Number of Bonds ...... 3983
Number of H-Bonds ..... 329
Number of Helices ..... 30
Number of Strands ..... 24
Number of Turns ...... 42
```