Functional and Phylogenetic Analysis of 5′ and 3′ Untranslated Regions (UTRs) of Hepatitis C Virus Genome (HCV) for the Potential Target Prediction of its Self- Regulation

By

Haleema Sadia

NUST00000277012

Department of Healthcare Biotechnology

Atta-ur-Rahman School of Applied Biosciences (ASAB)

National University of Sciences and Technology (NUST)

Islamabad, Pakistan

2021

Functional and Phylogenetic Analysis of 5′ and 3′ Untranslated Regions (UTRs) of Hepatitis C Virus Genome (HCV) for the Potential Target Prediction of its Self-Regulation

A thesis submitted in partial fulfillment of the requirement for the degree of Master of Sciences in Healthcare Biotechnology

By

Haleema Sadia

NUST00000277012

Supervised by: **Dr. Sobia Manzoor**

Co-supervisor: **Dr. Fazal Adnan**

Department of Healthcare Biotechnology

Atta-ur-Rahman School of Applied Biosciences (ASAB)

National University of Sciences and Technology (NUST)
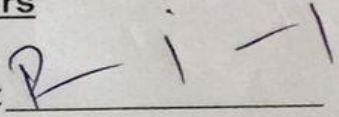
Islamabad, Pakistan

2021

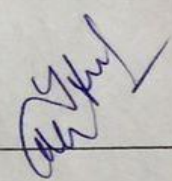# National University of Sciences & Technology

MS THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by:
(Student Name & Regn No.) <u>Haleema Sadia Reg No. NUST277012</u>

Titled: <u>Functional and phylogentic of 5' and 3' Untranslated regions (UTRs) of HCV genome for potential target prediction of its self regulation</u> be accepted in partial fulfillment of the requirements for the award of <u>MS Degree in Healthcare Biotechnology</u> degree with (__A__ grade).
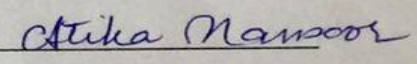
Examination Committee Members
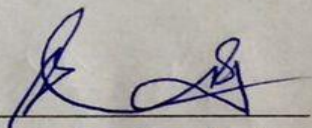
1. Name: <u>Dr. Rumeza Hanif</u>   Signature: _____

2. Name: <u>Dr. Aneela Javed</u>   Signature: _____

3. Name: <u>Dr. Atika Mansoor</u>   Signature: _Atika Mansoor_

Supervisor's name: <u>Dr. Sobia Manzoor</u>   Signature: _____
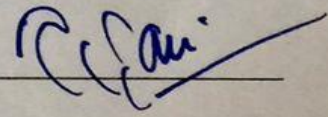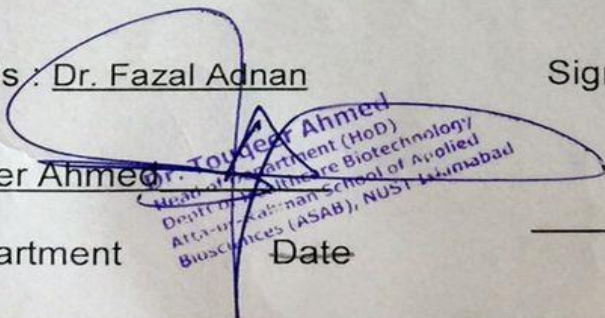   Date: _30.8.2021_

Co-Supervisor's : <u>Dr. Fazal Adnan</u>   Signature: _____

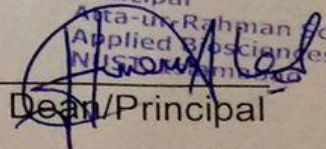<u>Dr. Touqeer Ahmed</u>
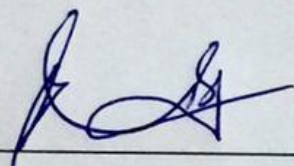
Head of Department   Date   30.8.2021   Date

COUNTERSINGED

Dr. Hussnain A. Janjua
Principal
Atta-ur-Rahman School of
Applied Biosciences (ASAB)
NUST Islamabad

Dean/Principal

Date: _30.8.2021_

THESIS ACCEPTANCE CERTIFICATE

Certified that final contents and form of MS/MPhill thesis entitled **"Functional and Phylogenetic Analysis of 5′ and 3′ Untranslated Regions (UTRs) of Hepatitis C Virus Genome (HCV) for the Potential Target Prediction of its Self- Regulation"** written by Ms. Haleema Sadia, (Registration No. 000000277012), of ASAB has been vetted by undersigned, found complete in all respects as per NUST Status/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial fulfillment for award of MS/Mphil degree. It is further certified that necessary amendments as pointed out by GEC members of the scholar have also been incorporated in the said tasks.

Signature: _____

Dr Sobia Manzoor, PhD
Tenured Associate Professor
Head of Department (HoD)
Deptt of Healthcare Biotechnology
Atta-ur-Rahman School of Applied
Biosciences (ASAB), NUST Islamabad

Supervisor: Dr. Sobia Manzoor

Date: 9.11.2021

Signature (HOD): _____

Dr. Touqeer Ahmed
r. of Department (HoD)
of Healthcare Biotechnology
School of
(ASAB), NUST Islamabad

Date: 9.11.2021

Dr. Hussnain...
Principal
Atta-ur-Rahman School of
Applied Biosciences (ASAB)
NUST, Islamabad

Signature (Dean/Principal): _____

Date: 10.11.2021

CERTIFICATE FOR PLAGIARISM

It is to confirm that MS thesis entitled "**Functional and Phylogenetic Analysis of 5′ and 3′ Untranslated Regions (UTRs) of Hepatitis C Virus Genome (HCV) for the Potential Target Prediction of its Self- Regulation**" by Ms. Haleema Sadia, Registration No. 00000277012 has been examined by me. I undertake that,

1. Thesis has significant new work/knowledge as compared to already present elsewhere. No sentence, table, equation, diagram, paragraph or section has been copied verbatim from previous work except when placed under quotation marks or duly referenced.

2. The work presented is original and own work if the author i.e. there is no plagiarism. No idea, results or words of others has been presented as author's own work.

3. There is no fabrication of data or results such that the research is not accurately represented in the records. The thesis has been checked using Turnitin (a copy of the originality report attached and found within the limits as per HEC plagiarism policy and instruction issued from time to time.

Dr Sobia Manzoor, PhD
Tenured Associate Professor
Head of Department (HoD)
Deptt of Healthcare Biotechnology
Atta-ur-Rahman School of Applied
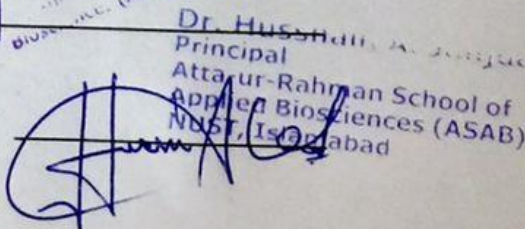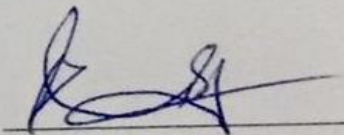Biosciences (ASAB), NUST Islamabad

9.11.2021

(Supervisor)

Dr. Sobia Manzoor
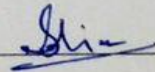
Tenured Associate Professor

Healthcare Biotechnology

ASAB, NUST

DECLARATION

I, Haleema Sadia, certify that the research work presented in this thesis titled **"Functional and Phylogenetic Analysis of 5′ and 3′ Untranslated Regions (UTRs) of Hepatitis C Virus Genome (HCV) for the Potential Target Prediction of its Self- Regulation"** is the result of my own work. Where information has been derived from other sources, it has been properly acknowledged/mentioned in the thesis. The work herein was carried out while I was a post-graduate student at Atta-ur-Rahman School of Applied Biosciences, NUST.

Haleema Sadia

NUST00000277012

DEDICATED TO

All my Teachers to date

For their role in making me who I am today, especially,

Dr. Sobia Manzoor

and

Dr. Fazal Adnan

The two teachers who mean the most to me
Forever thankful for their support & guidance

My beloved Family

For their support throughout the journey

&

Me, Myself & I

For the perseverance, compassion and
the resolution to never give up

ACKNOWLEDGEMENTS

Praise be to Allah Almighty who is the most beneficent, the most merciful, the ultimate source of all knowledge. The One who has been with me at every step of the way and never left me even at my lowest; The Creator, The Exalted, The Bestower who has blessed me immensely and kept me firm on this path of knowledge that I chose.

Firstly, I would like to express my thankfulness to my second home, my institution National University of Sciences and Technology (NUST) and my department ASAB, Principal ASAB and HOD Healthcare Biotechnology for providing a better working environment with required resources to conduct this research.

I would like to thank my supervisor **Dr. Sobia Manzoor** for the support, guidance and motivation throughout my research. It was a great learning experience working under her supervision, and in this lab. I am grateful to my GEC members Dr. Aneela Javed, Dr. Rumeza Hanif and external GEC member Dr. Atika Mansoorfor their input in this thesis. Foremost, I express my deepest gratitude to **Dr. Fazal Adnan**, for his invaluable assistance, encouragement and support throughout this year. I feel extremely lucky to have him as my mentor whose door was always open whenever I ran into trouble or had a long list of questions. For bearing with me patiently and being there for me, I will always be thankful to him.

My appreciation goes out to my fellows and friends specially **Moomal Masood** , whose unwavering support kept me going, Moomal Masood, Eeman Rehman who have been with me in this journey and the fellows from and I am extremely thankful to our lab assistant

Table of Contents

List of Acronyms

HCV	Hepatitis C virus
HCC	Hepatocellular carcinoma
UTRs	Untranslated regions
NcRNA	Non-coding RNA
ORF	Open reading frame
RdRp	RNA dependent RNA polymerase
NS	Non-structural
IRES	Internal ribosome entry site
%	Percentage
ALT	Alaninine amino transferase
NHANES	National Health and Nutrition Examination Survey
HIV	Human Immunodeficiency virus
NIH	National Institutes of Health
ELISA	Enzyme-linked immunosorbent assay
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
NICE	National Institute for Clinical Excellence
MRNAs	Messenger RNAs
EM	Electron microscopy

SL	Stem loops
HnRNPs	Heterogeneous nuclear ribonucleoproteins
NLs	Nuclear localization signals
ER	Endoplasmic reticulum
HVR1	Hypervariable region 1
HDL	High density lipoproteins
ARFP	Alternate reading frame protein
TM	Transmembrane regions
LEL	large extracellular loop
SR-B1	Scavenger receptor B type 1
ICAM-3	Intercellular adhesion molecule-3
ASGP-R	Asialoglycoprotein receptor
HSPG	Heparin sulfate proteoglycans
Eif	Eukaryotic initiation factors
E1 protein	Envelope glycoprotein 1
E2 protein	Envelope glycoprotein 2
NCBI	National centre of biotechnology information
C protein	Core protein
IFN	Interferon

List of Figures

4.7	Interaction of 5′ UTR HCV-3a with core, envelope glycoproteins(E1,E2),NS1,NS2,NS4A,NS4B,NS5A,NS5B	54
4.8	Interaction of 3′ UTR HCV-3a with core, envelope glycoproteins(E1,E2),NS1,NS2,NS4A,NS4B,NS5A,NS5B	56
4.9	Phylogenetic tree of 5′ UTRs of constructed through Mega X	60
4.10	Phylogenetic tree of 3′ UTRs of constructed through Mega X	61

List of Tables

Abstract

Hepatitis C virus (HCV) is an utmost communal well-being problem globally. Globally about 177.5 million individuals are tainted with HCV. Pakistan has second largest load of HCV with a country wide prevalence of 4.8 %. About 17 million mortal are tainted with HCV in Pakistan. Documenting the genetics of HCV has important implications for understanding disease severity as well in therapeutics. In this insilico study we conducted functional analysis of non-coding 5' and 3' untranslated regions (UTRs) sequence of various HCV genotypes by multiple sequence alignment. Alongside we formulated and compared the secondary structures of these UTRs in Hepatitis C virus (HCV) of different regions and genotypes. RNA-RNA interaction was used to compare 5' and 3' non-coding UTRs with rest of coding viral genome using the most prevalent HCV-3a as model. Secondary structures of both UTRs regions were analyzed to confirm the structural stability and its involvement in self-regulation of genes for survival and pathogenicity in the host. Lastly a phylogenetic analysis was constructed to determine homology and origin of HCV-3a virus in Pakistan. 5′ UTR sequence show high homology in Pakistan HCV-2a to HCV-3a sequence. The secondary structures of 5' and 3' UTRs of HCV-2a and HCV-3a of Pakistan showed significant structural differences and varying degree of stabilities in comparison to other countries i.e. India, Egypt, Us and Japan. In our study the 5' UTR of HCV-3a Pakistan has less minimum free energies as compare to 3' UTR of same genotypes. The interaction of non-coding 5′ and 3′ UTR with ORF of HCV-3a Pakistan strain show 6-7 binding sites with low hybridization energies <-20 kcal/mol at core, E1, E2, NS2, NS4A, NS4B and NS5A. Phylogenetic analysis show that 5′ UTR HCV-2a Pakistan is closely related to US HCV-2a and 5′ UTR of HCV-3a Pakistan is closely related to that of 5′UTR of HCV-3a of India.

Chapter 1

Introduction

Chapter 1: Introduction

1.1 Background and problem statement

Chronic hepatitis C virus (HCV) sepsis a notable public-health problem accountable for cirrhosis, liver damage and hepatocellular carcinoma (HCC) (Arshad & Ashfaq, 2017). Hepatitis C is a worldwide haleness issue infecting 177.5 million people and approximately the infection becomes chronic in 71.1 million peoples (Spearman et al., 2019). According to a recent estimate Pakistan has the second-largest load of hepatitis C, with a countrywide prevalence of 4.8% (Abbas & Abbas, 2020). In Pakistan roughly 8-10% individuals are HCV carriers and 17 million people are infected with HCV (Al Kanaani et al., 2018). Dominant liberated hazards for anti-HCV receptivity were male sex, blood transfusion, and intravenous drug use (Spearman et al., 2019).

The HCV virus is a blood-borne which is grouped into 7 genotypes and 67 subtypes all responsible for causing Hepatitis C infection (Smith et al., 2014). Hepatitis C virus genome encodes 10 constructional and non-constructional proteins and is 9.4 kb in length. It is bounded by duet cis reserving RNA elements eminent as 5' and 3' untranslated regions (UTRs). In upstream of relocation instigation codon 5' UTR is present which consists of 341 nucleotides (Forton et al., 2004). At the 5' UTR internal ribosome ingress site (IRES) is present which is a highly preserved site and the translation initiation takes place through this site which acts in a cap-independent mode. The 3' UTR which is present towards C-terminal consists of roughly 225 nucleotides and is shorter than 5' UTR. In three region the 3' UTR is assembled which ranges from 5' to 3' UTR consisting on of an inconsistent domain of about 30-40 nt , a highly preserved 3' X region that is comprised of triplicate stem loop composition SL3,SL2 and SL1 and a poly U/UC tract (Kolykhalov et al., 1996).

The 3' untranslated region present at periphery of viral genomes in entangled in viral duplication and translation (Sadia et al., 2013).
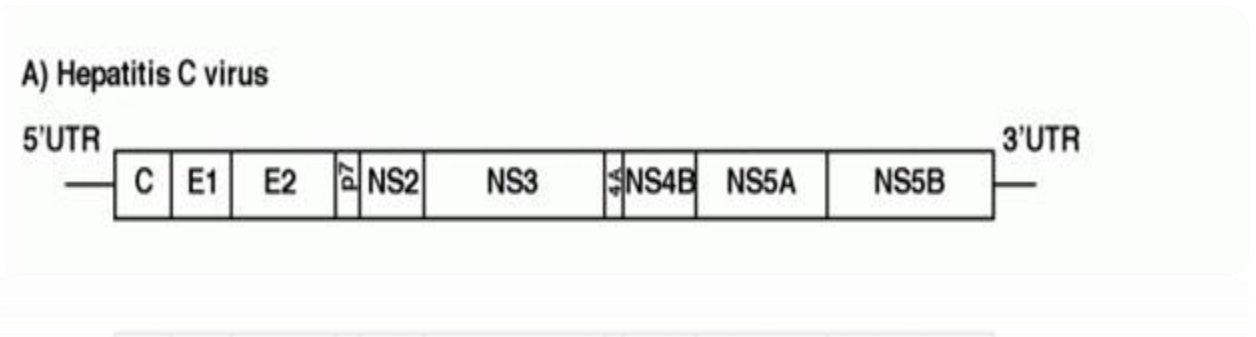
Fig 1.1: Genomic organization of hepatitis C virus. Adapted from (Vladimir *et al.*, 2018)

Currently antivirals are given against HCV infection but no vaccine has been developed yet. The existing antivirals and therapeutics are showing limitations due to lack of specificity, resistance or ineffectiveness due to frequent mutations in the HCV genome. The most important reason for in-effective treatment of HCV is presence of its unique characteristics "quasispecies" which create a major hurdle. A key feature of viral development is the genesis of viral proteins by ribosome of human compere. In HCV this operation is controlled by the viral 5' and 3' untranslated but accurate controlling process has not been configured. Therefore, in our study our focus would be to conduct various analyses on the HCV genome particularly the 5' and 3' non-coding UTR sequences. Our study will aim to bridge the literature gap by studying the conservation of UTR sequences, its secondary structure and stability, its interaction with viral proteins and lastly by conducting a phylogenetic analysis of viral UTRs across the globe. Therefore understanding the gene regulation mechanisms in the HCV genome in future can help provide alternative strategies to combat infection rate and avoid limitations of the existing therapeutics.

Chapter 2

Literature review

Chapter 2: Literature review

2.1 Hepatitis C

Hepatitis C is a liver disease bring about by HCV. HCV is affiliated to the Flaviviridae appendage and is an exclusive positive stranded RNA virus. HCV is an enveloped virus and has a small size (55-65nm). It's a blood borne virus and a significant source of persistent liver disease. It has been acknowledged as global health problem because of the development to hepatocellular cancer and cirrhosis(Modi & Liang, 2008). HCV replicates in liver hepatocytes and is target specific (Hussain, 2013).

2.1.1 Natural history

Hepatitis C virus is entangled in 40% of long-term liver disease and is common significant blood borne sepsis in the U.S. In 1989 Choo et al first secluded HCV from the individual fluid with non-A non-B hepatitis. HCV a recently discovered virus was found to be accountable for 90% of non-A, non –B hepatitis after discovery of HCV replication (Chen & Morgan, 2006).

2.1.2 Genomic organization of HCV and function

Hepatitis C virus is a participant to the appendage *Flaviviridae* and is a RNA virus. The Flaviviridae appendage is splitted into three subdivsions: pestivirus, hepacivirus and flavivirus.. HCV is a part of hepacivirus genus which comprise atleast 7 genotypes and various subtypes. The *Flavivirdae* genome range in size from 9.6 to twelve chiliad nucleotides and carries a positive strand RNA molecule with an 0RF comprise a precursor protein of three thousands amino acids or more (Lindenbach & Rice, 2005).
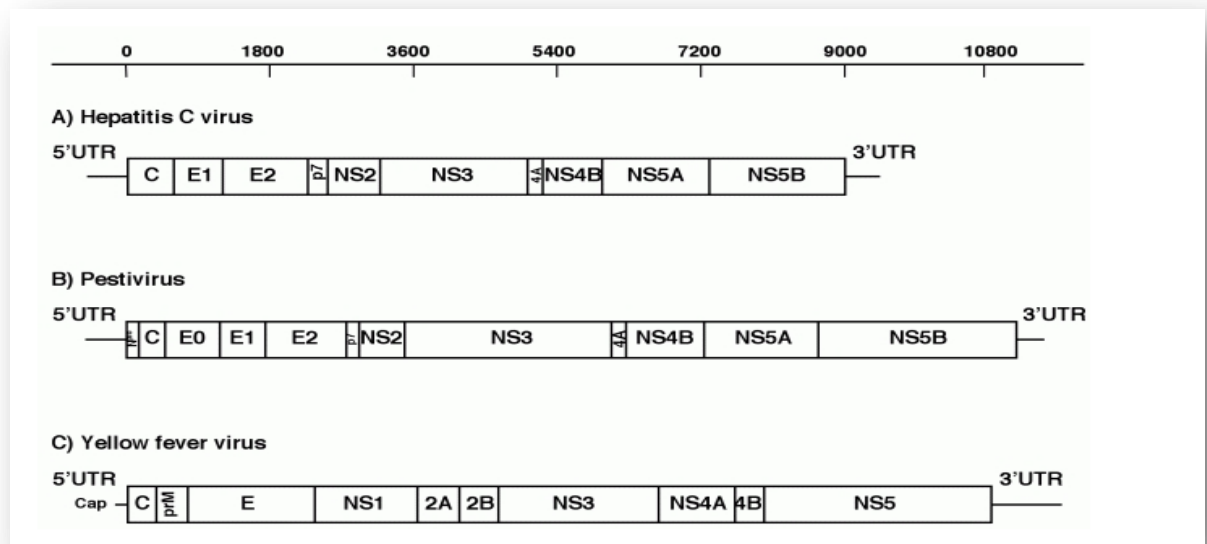
Fig 2.1: The arrangement of *Flaviviridae* genomes. Top HCV, middle pestivirus and at bottom yellow fever virus is being shown. Adapted from (Thurner et al., 2004)

Flaviviirdae seems to trigger receptor mediated endocytosis and hitch to single or additional host markers assembled as a marker composite. The nucleocapsid is exported to cytol after hitching of viral shell with cellular layers. After removal of capsid the process of translation starts in the cytol and it leads to genesis of polyprotein. By viral and host proteases the polyprotein then splits into ten non-constitutional and structural ligands. The viral duplication composite that is connected with cellular layers helps in duplication of viral genomes. Viral duplication takes place in cytol which leads to production of complete length negative-strand RNA interposed. From cytoplasmic vesicles progeny virions are assembled which are manufactured by sprouting through intracellular layers. In spite of the described likeness escorted by appendage of other *Flaviviridae* genera, HCV do show an amount of epidemiological, virological and pathophysiological differences. The translation process in Flavivirus is usually cap-dependent and the type 1 cap structure present in 5' UTR helps in moderation of the translation process. This cap independent structure is

present at 5' UTR present next to unchanged preserved AG region and a comparatively squat spraw; present towards the polyprotein coding region (Brinton & Dispoto, 1988). However if we compare the 5' UTR of HCV to GB viruses and pestivirus the translation process is entirely different. Because the 5' UTR of HCV consists of IRES which acts in a cap-independent mode for translation startup which facilitates undeviating hitching of molecular factors, ribosomal subdivsion and succeeding relocation (Thurner et al., 2004).

HCV has tapered tissue co-receptor and host distinctiveness. HCV is spread entirely through undeviating gore-to-gore association betwixt humans. Flaviviruses are mainly transmitted through ticks or mosquitos and possibly contaminate a wide span of vertebrate brute, with humans being a periphery host. No intimate pestivirus can debase mortal and no investigated insect transmitter existed. Contamination spread by flaviviruses are short termed and restricted in vertebrate brute; however HCV has an increased regulatory rate in mortal. Adapted and strong cellular and humoral immune responses have been exhibited to be intricate in pestivirus and flavivirus infection protection and recovery (T. Farci et al., 1996).

2.1.3 Routes of transmission

In 20th century later half, the extensive reachability of injectable therapies and extended illegal injection drug use were accountable for fast pop up of hepatitis C virus (HCV). Illegal injection drug use and iatrogenic exposures have been significant risk hazards for HCV transmission globally. Insecure therapeutic injection application seem to be accountable for most infections in developing countries (Feinstone et al., 2001). Backer testing has nearly abolished transmission related in sepsis in developed countries but sepsis transferred to patients by risky injection practices has turn out to be problematic. Major

hazard for HCV is injection drug use, incidence residue elevated between new injectors, and this conduct probably subscribe to reveal alliance among HCV-positive persons and histories of non-injection drug use, incarceration and tattooing. (M. J. Alter, 2011).

2.1.4 Acute Hepatitis C

The prevalence of severe hepatitis C changes from 1980 to 1995 from one lac eighty thousand cases to thirty thousand cases per year (Armstrong et al., 2006). Severe hepatitis C sepsis is uncommonly identified since the bulk of intensely contaminated patients have no signs. In the transmission environs, where severe outbreak of HCV sepsis existed, seventy to eighty percent cases were having no symptoms (McCaughan et al., 1992).

Around twenty to thirty percent of adults with acute HCV infection may flourish aloof manifestations. After exposure the symptomatic outset span from three to twelve weeks (H. J. Alter & Seeff, 2000).

2.1.5 Persistent hepatitis C

Chronic hepatitis C infection differs from acute in a way that in chronic HCV the HCV RNA remains in blood for over a period of six months. In only fifteen to twenty percent of patients HCV infection resolve automatically and HCV RNA set off undiscovered and ALT levels come to typical. But in about 75%-85% of infected individuals HCV virus stays longer than 6 months and they are unable to remove the virus from the body which leads to development of chronic hepatitis (Chen et al., 2006).

2.2 Hazards for developing chronic HCV sepsis

2.2.1 Age

The rate of HCV severity seems to be reduced in immature subjects. In a study conducted

by National Health and Nutrition Examination Survey (NHANES), the severity tariff was determined thirty percent in individuals below the age of 20 years (Vogt, 2000). Deep rooted chased research in youngster with after transmission hepatitis exhibit that only fifty five to sixty percent of youngster hold on HCV RNA positive in maturity (Sasaki et al., 1997).

2.2.2 Gender

Chronicity of HCV infections seems to be lowered in women especially younger women's. Confirmation of this emanate from retroactive inspection of duet huge eruption that women who are pregnant and has attained Rh immune globulin that has been contaminated with HCV has lower rate of HCV infection (Hillyer, 2000).

2.2.3 Race

Amid variable tribal and ethnic groups with HCV sepsis there is variability in tariff of severe HCV infection, enlargement of complications and retaliation to cure. For indefinite purposes, African Americans seem to have greater amount of severe HCV infection as compare to Hispanic whites and Caucasians (L. B. Seeff et al., 2001). The NHANES study also showed a greater tariff of severe infection between African Americans (86%) in comparison to Caucasians (68%). The NHANES study showed that in comparison to African Americans and Caucasians the African Americans should greater amount of infection i.e. eighty six percent. In addition to combination therapy of ribavirin and interferon the African American showed reduced rate of continuous viral response (SVR) (Nguyen et al., 2004).

2.2.4 Immune and jaundice response

The severity outlay of HCV sepsis is reduced in subjects who evolve jaundice through the short term outburst of HCV sepsis in comparison to which have no jaundice. The proficiency of immune retaliation performs a significant function in the progression of severe hepatitis C and results to development of liver fibrosis (Chen & Morgan, 2006).

2.3 Risk elements for advanced progression of hepatic fibrosis

2.3.1 Alcohol utilization

In subjects with persistent HCV infection alcohol usage seems to be a significant factor which results in development of fibrosis. Alcohol utilization results in development of continuous liver disease and there is powerful evidence which showed that higher levels of alcohol intake results in persistent HCV infection (Wiley et al., 19).

2.3.2 Age at sepsis time

Multiple studies have exhibited a dominant alliance between fibrosis tariff and age at infection hour. The level of fibrosis and inflammation also plays a significant role in further progression to HCV liver disease. The data shows that the development of HCV sepsis to liver fibrosis increases at an irregular rate so it got changes as the patient ages to an older age (Chen & Morgan, 2006).

2.3.2 Simultaneous infection with HIV and HBV

The dominant risk factors that result in liver fibrosis are simultaneous infection with HBV and HIV. In injection drug users and hemophiliacs simultaneous infection with HCV and HIV are significantly common. HIV low CD4 counts and seropositivity exhibit to speed up HCV liver fibrosis (Ragni & Belle, 2001). Vice versa HCV has been linked with rapid

development of HIV to acquired immunodeficiency syndrome (AIDS) (Lesens et al., 1999).

2.3.3 Comorbid situations

In the development of HCV infection comorbid situation and HCV sepsis plays a dominant function. Amantadine has thought to be linked with more hostile hepatic infection. Individuals with cellular immune impairment and humoral immunoglobulin deficiency have exhibited dominant increased rates of development to cirrhosis than immunocompetent patients (Berenguer et al., 2000).

2.3.4 Long term complications

2.3.4.1 HCC and cirrhosis

HCV development to chronic infection is frequently hushed and no sign and symptoms are seen until it progresses to later stage liver disease and then the complexities to HCC starts to develop. The characteristics of deteriorated hepatomegaly incorporate portal hypertensive gastropathy, upper gastrointestinal bleeding and hepatic encephalopathy (Fattovich et al., 1997). The accumulative likelihood of an event of healthcare fails five percent at 1 year and expands to 30% at ten years from verification of hepatomegaly (Hu & Tong, 1999). The chance for HCC was enhanced seventeen fold in HCV contaminated subjects in comparison to HCV negative controls in a meta examination of case control studies (Papatheodoridis et al., 2001).
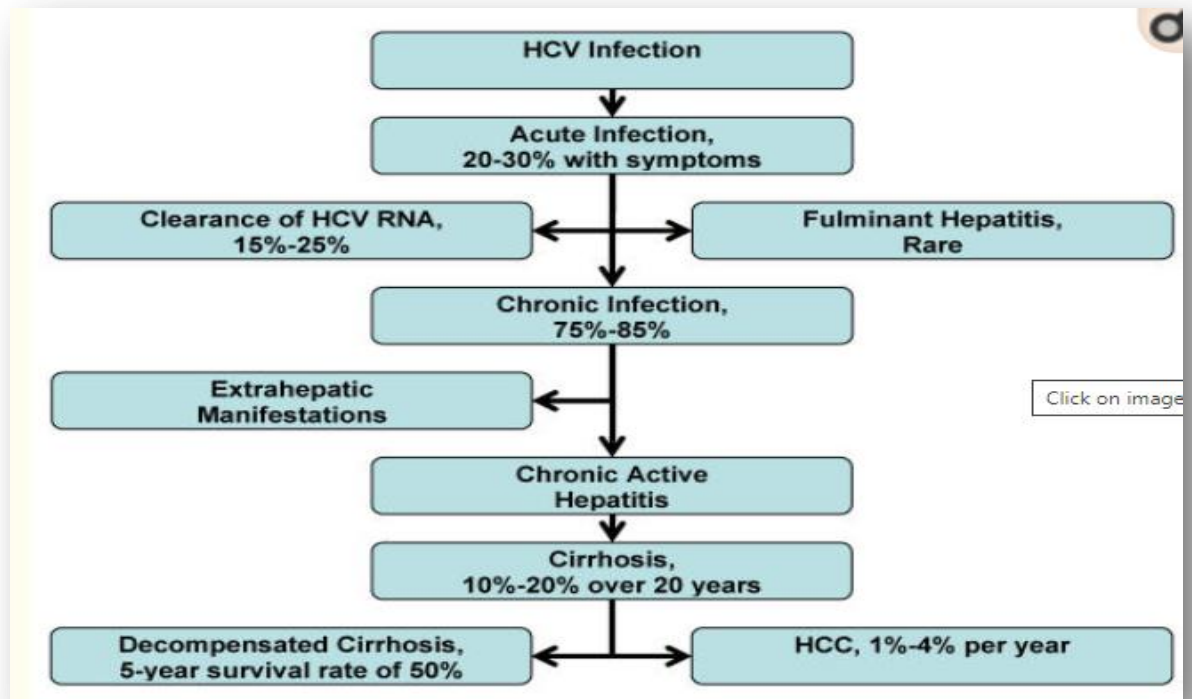
Fig 2.2: Stages of HCV infection from acute to chronic. Adapted from (Papatheodoridis *et al.,* 2001).

2.4 Symptoms of HCV:

Inflammation with the hepatitis C virus leads to hepatic in addition to extra hepatic affliction. Amidst a pupilage of individuals initial symptom is a critical ailment reminding extra manifestation of critical hepatitis. Symptoms last 2-12 weeks and the average incubation period are 7 weeks. Patient with persistent HCV are normally asymptomatic, however muscle pain, fatigue, starved and right portion ache still happen. Ailments associated with sepsis incorporate thyroditis, polyarteritis nodosa, Sjögren's syndrome, lichen planus and necessary mixed cryoglobulinemia (Lunel, 1994).

2.4.1 Outside liver symptoms

Chronic HCV infection has been linked with multiple outside liver symptoms. These manifestations may include dermatologic, hematologic, multiple organ system and rheumatologic systems (S C Gumber, n.d.). The most significant extrahepatic manifestation is cryoglobulinemia. In 50% of subjects with long-term HCV infection cryoglobulins are present. Medical symptoms may bloom in 25% to 30% of HCV patients with mixed cryoglobulinemia which ranges from skin rashes, Raynaud's phenomenon, purpura, arthralgias. Immune complex deposition led to the development of clinical symptoms (El-Serag et al., 2002).

Other outside liver symptoms constructed in patient with severe HCV infection are vitiligo and lichen planus. Existent is some evidence that propose a connection betwixt severe HCV sepsis and non-Hodgkin's lymphoma and Hodgkin's . Particularly there is ambiguity if these connected ailments are originated straightly from HCV sepsis or the cardinal resistant provoke engender by severe sepsis (El-Serag et al., 2002).

2.4.2 Diagnosis of HCV

ELISA tests should be used to examine the patients with surmise HCV sepsis for virus agglutinin (Goffin et al., 1994). If agglutinin is observed the subject is believed to be at danger regardless of undefined serum tests, viremia must be looked for by polymerase chain reaction (PCR). The finest compute of the expanse of sepsis is measured by liver biopsy, routine liver trials correspond badly with fibrosis and necroinflammatory outcome, hence it is also beneficial to eliminating other diagnosis like as alcohol induced liver ailment. Patients with long-term HCV complications should be monitored for HCV

therapy. The main purpose of HCV therapy is to attain a continuous virological retialition and avoiding regain of symptoms (Davis, 2002).

2.4.3 Standard treatment of Hepatitis C

Pegylated interferons has been given for treatments. The pegylated antivirals have different pharmacodynamics characteristics and are given once a week. The half-life of peg-aplha 2 is longer and is secreted through liver mainly it is given in a fixed dose while peg-alpha 2b is mainly secreted by kidney and dose is administered according to body-weight (Vince et al., 2009).

According to different genotype the response rate is different. Peginterferon/ribavirin therapy given to individuals with HCV genotype 1 has different response rate of 40-50% while same therapy given to people with genotype 2 and 3 the feedback rate varies to 80%. Twenty four weeks may be sufficient for genotype 2 and 3 and trials with Pega exhibit that these subjects need less ribavirin dose in comparison to genotype 1(Manns et al., 2001)..

2.4.4 Provisions

Interferons should be used in special care especially patients that face depression problems. As ribavirin has proved to be cancer causing in animals females of accouchement age should be careful on need of prophylactic (Leonard B. Seeff & Hoofnagle, 2002). Ribavirin can result to red blood lysis single therapy as an alternative solution when ribavirin is indicated in patient with HCV(Salazar, 1996).

2.5 HCV genome structure and organization

HCV consists of a single-stranded positive sense genomic RNA of about 9.4 kb in size and

contains a poly-A tail at 3' terminal. The sequence contains 5' untranslated region of about

341 bases and 3' UTR of 225 bases.
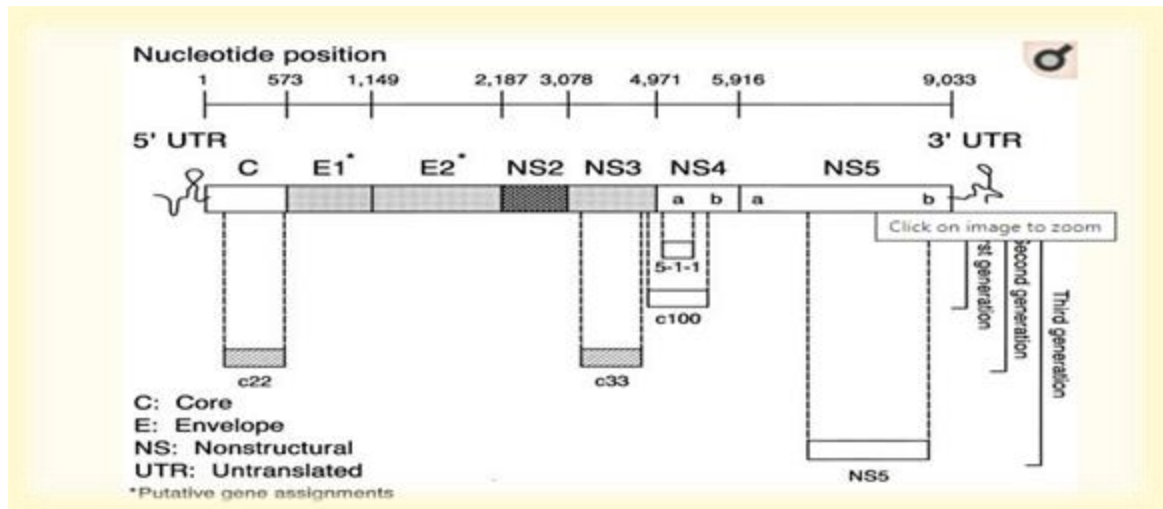
Fig 2.3: Genomic organization of HCV. Adapted from (Conne et al., 2000)

2.5.1 General functions of UTR

After transcription when genes are expressed to form proteins, UTRs are known to play

critical roles which include translation efficiency and shift of mRNAs out of the nucleus

provide stability and involved in subcellular localization (Bashirullah et al., 2001).

 UTRs also play various other functions such as encoding selenoproteins in an action

moderated by a homologous stem loop structure present at 3' terminal and is involved in

definite positioning of selenocysteine at stop codon of mRNA. The importance of UTRs in

controlling gene function is highlighted by locating the mutations that change the UTR and

result in severe immune response (Conne et al., 2000)..

2.5.1.1 HCV 5′ untranslated region

HCV 5' UTR is present at 5' terminal and consists of 341 nucleotides present before of translation initiation codon. It constitute for the most preserve region of genome showing sequence identity of sixty percent with GBV-B and fifty percent with pestivirus. The 5' UTR consists of four constitutional domain I, II ,III, IV. Domain two to four with some twelve to thirty nucleotides constitute the IRES which is involved in cap independent translation and contains the start codon. IRES has ability to startup the translation by directly binding to 40 s ribosomal subunits. Electron micrographs reveals that domain two to four form visual areas and pliable hook among domains two and four (Beales et al., 2001).

2.5.1.2 HCV 3′ untranslated region

The 3' UTR sequence is present at 3' terminals and consists of 225 nucleotides. It consists of three regions a changeable domain of thirty to forty nucleotides, a preserved nucleotide of about 98 nucleotides (3' X region) which is comprised of three stem loop structures known as SL1,SL2, SL3 and a poly U/UC region (Kolykhalov et al., 1996). The 3' UTR interconnect with these stable stem loop regions and also viral protein NS5B coding sequence. The 3' X region and some portion of poly U tract are involved in viral replication and residue sequence of 3' UTR is involved in enhancement of viral replication (Yi & Lemon, 2003).
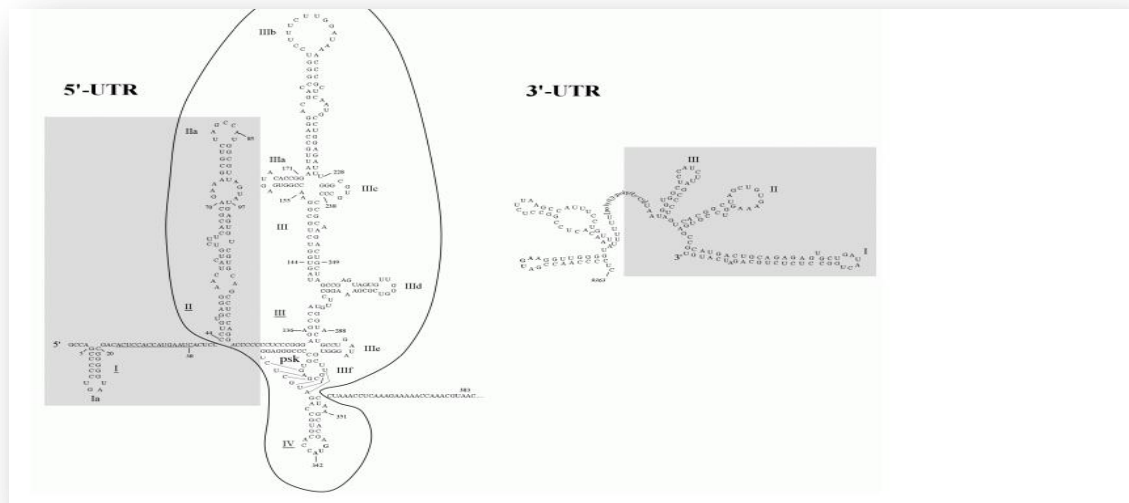
Fig 2.4: The structures of the 5′ UTR and 3′ UTR of HCV RNA. Adapted from (Rijnbrand and Lemon, 2000)

2.5.1.2.1 Regulation by UTRs

The UTRs are involved in regulation by different ways. With different RNA-hitching proteins nucleotides patterns present in 5' and 3' UTR interconnect with these binding proteins. On amalgam of primary and secondary structures the living pursuit of RNA level depends. The interchange that takes place betwixt sequence element that reside in UTR and non-coding elements exhibit to perform key functions (Sweeney et al., 1996)..

Different RNA-binding proetins present in cytol involved in protein formation also perfroms vital role in expanded diversity of statutory process, like as pre-mRNA meshing and 3' end converting within the nucleus (Xu et al., 2001). This practical linkage among translational affair in the cytoplasm and in nucleus may describe that nuclear processing of mRNA can play role in controlling the cytol fate (Kataoka et al., 2000)..

2.5.1.2.2 General structure characteristics of untranslated regions

Juxtaposition of numerous partial and completed genome sequences disclose some preserved features of UTRs structure. The mean size of 5' UTRs normally remain constant within the figure of 100-200 nucleotides while the mean size of 3' UTR is more fluctuating and it ranges from 800 nucleotides in brute and humans and 200 nucleotides in plants and fungi. The size of both 5' and 3' UTR is quite fluctuating in different species and is quietly noticeable and ranges from dozen to some thousand nucleotides (Pesole et al., 2001)..

More introns are present in 5' UTR as compare to 3' UTR in genomic region correlated to mRNA. 30% genes in metazoe has round about wholly 5' UTR regions and they consist of more introns as compare to 3' UTR although they are much length than 5' UTR. Alternatives UTRs can be generated through through use of various splice donor/ acceptor sites and various transcription initiator sites (Grabowski & Black, 2001).
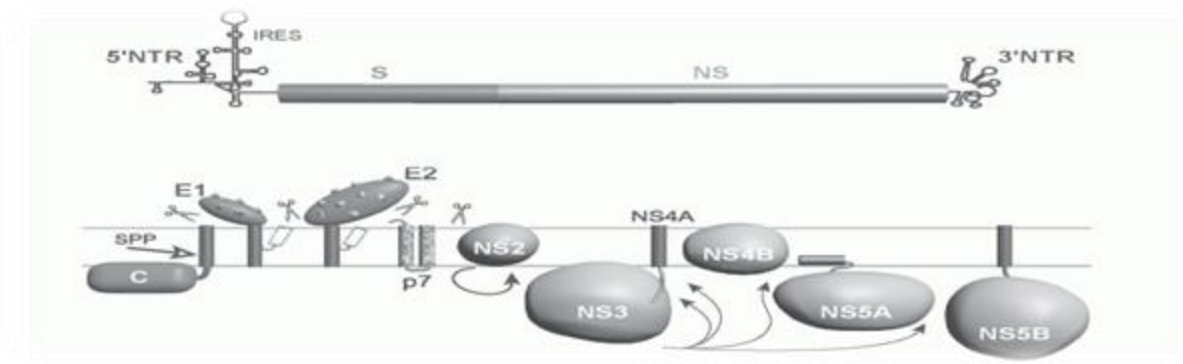

Fig 2.5: HCV genome compartmentalization (top) and polyprotein rectifying (bottom). Adapted from (Penin *et al.,* 2004)

2.5.2 Functions and characteristics of HCV proteins

Hanging on the genotype the ORF of HCV contains of 9024-9111 nucleotides. The ORF

contains 11 different structural and non-structural proteins. A small protein p7 functions

has not been known yet. The 3 structural proteins are present towards the N-terminal whole

non-constitutional (7 proteins) are present towards C-terminal.

HCV protein	Function	Apparent molecular weight (kDa)
Core	Nucleocapsid	23 (precursor) 21 (mature)
F/ARF[a]-protein	?	16–17
E1	Envelope Fusion domain?	33–35
E2	Envelope Receptor binding Fusion domain?	70–72
p7	Calcium ion channel (viroporin)	7
NS2	NS2-3 autoprotease	21–23
NS3	Component of NS2-3 and NS3-4A proteinases NTPase/helicase	69
NS4A	NS3-4A proteinase cofactor	6
NS4B	Membranous web induction	27
NS5A	RNA replication by formation of replication complexes	56 (basal form) 58 (hyperphosphorylated form)
NS5B	RNA-dependant RNA polymerase	68

Fig 2.6: HCV proteins and their functions in the viral life cycle. Adapted from
Bartenschlager et al., 2004.

2.5.2.1 Structural proteins

2.5.2.1.1 Core protein

The core protein of HCV is RNA-binding protein, is basic, and apparently constitute the

viral capsid. The size of core protein ranges from 23 kDa and consists of 191 aa. Core

protein exist in different sizes but 21 kDa protein structure is considered to be standard.

The core protein consists of three regions a N–terminal a C-terminal and signal for

envelope protein E1. Various positive charges are contained by domain D1 and is

significantly participate in nuclear localization and binding of RNA. For interconnection of core protein with lipid droplets and ER domain D2 is accountable (Schwer et al., 2004).

Core protein layer and unlayered structure appears to exist in compound or uncompounded forms. The HCV core protein can exist as NLPs when manifested in artificial systems (Baumert et al., 1998)..

Apart from its function in viral shell formation the core protein plays some other roles as well. It plays role in interconnecting with different host and viral proteins that may be significant in viral circuit ion. The HCV core protein has anti and pro apoptotic purpose restoring hepatocyte and involved in tissue damage development (Cho et al., 1998). HCV core protein also connect with other cellular protein like c-myc which is a proto oncogene and is involved in carcinogenesis which can leads to hepatocellular carcinoma (Moriya et al., 1997).

2.5.2.1.2 E1 and E2 envelope glycoproteins

For viral alloy and ingress E1 and E2 envelope glycoproteins are involved. They are necessary components which take parts in formation of viral shell. The molecular weight of E2 ranges between 70-72 kilo Dalton and is heavier than E1 whose weight ranges from 33-35 kDa. They do perform some other functions such as assembly, localization and membrane spanning etc. Multiple cysteine and proline residues are contained by sub domains of E1 and E2. E1 and E2 are extremely glycosylated.. Besides that, E2 consists of hypervalent regions with a amino acid sequences varying up to 80% between HCV subtypes and genotypes (Zibert et al., 1997).

In the initial steps of infection E2 plays a crucial role. Viral adjunct is believed to be commence through E2 interconnection with assorted element of recipient composite. Due

to positive region of HVR1 it can bind with negatively charged molecules present at cell surface which can help in attachment and determination of E1 and E2 to receptor surface (Voisset et al., 2005).

2.5.2.1.3 Frameshift protein

In the N-terminal of HCV polyprotein ribosomal shift take place which results in formation of frameshift protein. In severely infected patients, agglutinin to ligands from the F protein were discovered, proposing that the protein is manufactured in the course of sepsis (Walewski et al., 2001). Although the required protein formation procedure ruling the yield and frequency of the F protein through the course of the several stages of HCV sepsis are unknown so the function of F protein has not been elucidated yet (Baril & Brakier-Gingras, 2005)..

2.5.2.2 Nonstructural proteins

2.5.2.2.1 P7

The size of p7 protein is 63 amino acid and is classified to be an essential layer protein. It constitutes two membrane spanning estate containing alpha helix and fixes by cytol kink. In chimpanzees p7 appears to be crucial as liver exchanges between humans and chimpanzees can leads to deletions and mutations in its cytol hook that inhibited the infectivity. To viroporin appendage p7 belongs and it can act as calcium ion channel but these studies should be confirmed outside the cell (Sakai et al., 2003).

2.7.2.2.2 NS2

NS2 is a non glycan attached protein and its weight ranges from 21-23kDa. At amino acid position 839-883 and 928-960 it consists of two internal signal sequences that are accountable for ER membrane connection. NS2 is a momentary protein which unties its

proteolytic activity following self-dissolution from NS3 and is demean by the by protein kinase 2 e in phosphorylation-dependent mode. NS2 can interconnect to provider cell protein in addition to its proteolytic activity like influence demonstrator genes and liver defined pro apoptotic cell death supervised by non-liver and liver specified enhancers and promoters (Samina et al., 2010).

2.5.2.2.3 NS3-NS4A

NS3 is a multi-purpose protein which consist of alpha helicase domain in its C-terminal and the N-terminal consist of serine protease domain. NS3-4A also hold up other properties and it interconnect with recipient cell route that might perform a vital role in infection circuition. For anti-HCV treatments regimen NS3-NS4A is major target (Morillas et al., 2005).

2.5.2.2.6 NS4B

NS4B is derived from localization of ER and is an integral membrane protein of having size 261 aa. NS4B is believed to have for membrane spanning domains and constitute a N-terminal helix that are accountable for membrane linkage. It also plays vital function as a layer anchor for duplication composite. Addition properties include cellular synthesis hindrance, transformation and alteration of HCV NS5B RdRP activity and initiation of cytokines (Kadoya et al., 2005).

2.5.2.2.7 NS5A

NS5A is a phosphate covered zinc metallic protein and the weight ranges from 55-58 kDa. It plays a significant function in virus regulation and replication. The N-terminal domain of NS5A contains amino acid and a lipohlilic and hydrophilic properties which is required for replication composite joining. The NS5A protein contains three domains located at N-

terminus. Domain I present in N-terminus contains four cysteine leftovers and a zinc motif. Changes in zinc binding region leads to halt in HCV RNA replication due to mutations in NS5A region.

Procedure through which NS5A manages HCV duplication are not wholly clear. NS5A interconnects to C-terminal which then help in connection to lipid rafts derived from intracellular layers. An alternative study proposed that the amount of NS5A phosphorylation performs a significant function in viral circuition by modulating a switch from duplication to congregation, by which hyperactive phosphorylation shape purpose to preserve the duplication complex in a congregation incompetent state (Wang et al., 2005).

Various roles were allocated to NS5A grounded on its interconnection with host cell proteins. Such as, NS5A seems to play a part in interferon reluctance by inhibiting and irrevocable to protein kinase. NS5A besides prop up RNA to RNA replication restorative roles, and seems to play part in the cell growth control. Although, these inspection remains to be established *in vivo* (Tellinghuisen & Rice, 2002).

2.5.2.2.8 NS5B RNA-Dependent RNA Polymerase

The membrane anchored protein has different classes and NS5B is a part of tail-anchored proteins. The C-terminal of this protein forms an alpha helical membrane spanning domain which results in after translational targeting to the endoplasmic reticulum cytosolic side and the Spartan protein area is revealed at this site (Moradpour et al., 2005).

The glass composition of NS5B exposed that the RdRp has a rational "thumb,fingers and palm" composition set up by its N-terminal amino acids.. The RdRp is an alternative significant earmark for anti-HCV drugs progression

NS5B is part of layered proteins commonly known as tail-anchored proteins as they are present at 3' terminal. The C-terminal region of this protein manifests α-helical membrane-spanning preserve accountable for post-translational attacking to cytol of ER. The structural composition of NS5B exhibit that it do contain a traditional structure of thumb, palm and fingers. Interconnection uniting the thumb and finger sub zones leads to a wholly enclosed catalytic area that secure formation of sense and anti-sense RNAs (Morillas et al., 2005)..

Interconnection among NS5B and cellular elements has been revealed. The C-terminus and N-terminus interconnection of NS5B may performs a significant function in building of HCV replication complex (Saunier et al., 2003)..

2.6 Phylogenetic analysis:

Extensive literature and phylogenetic studies are available recently but they have used partial HCV genomic sequences of capsid, envelope, NS5B or other genes for locating viral factors accountable for transmission trends of the viruses. Several other studies have used partial 3' UTR sequences for conducting phylogenetic analysis. Some of the research has used the open reading frame or whole genomic sequences for elucidating the evolutionary relationships of HCV genotype 3a that is either confined to a specific region or country.

Therefore there is demand to govern whether fixating on specific non-coding 5' and 3' UTR regions of HCV or using entire genome sequences and partial ORF sequences is befitted for genotyping the hepatitis C virus isolate in Pakistan and worldwide.

Purpose of the study:

- The first objective of this study is to find best possible deletions and mutations that happened in the Hepatitis C 5′ UTRs of different origins and genotypes.

- The second aim is to study interaction of 5′ and 3′ UTR against the proteins of Hepatitis C virus at mRNA level

- The third aim is to get more insight to Hepatitis C virus lineage in Pakistan of 5′ and 3′ UTRs sequences especially Hepatitis C genotype 3a as it is most prevalent genotype in Pakistan.

Chapter 3

Methodology

Chapter 3: Methodology

3. Methodology

The below steps show step by step framework or methodology followed in the study focusing on the different bioinformatics tools used for the results

1. Sequence retrieval of HCV UTR through **NCBI viral genome database**

2. Functional analysis of the 5' UTR of different HCV strains by Multiple sequence alignment using **Clustal W in Mega X software**

3. Prediction and comparison of secondary structures of 5′ and 3′ UTR of HCV genotypes 2a-3a of various HCV strains using **RNAfold.**

4. Prediction of consensus secondary structures of set of HCV genotype HCV-2 and HCV-3 sequences using **RNAalifold.**

5. RNA-RNA interaction is used to compare 5′ and 3′ UTRs with rest of viral genome with most prevalent HCV genotype 3a of Pakistan as model using **IntaRNA.**

6. Phlylogentic tree was constructed to determine the homology and origin of HCV genotype 2a,3a in Pakistan using **Mega X.**

3.1 Sequence retrieval through NCBI

Complete sequence of the different Hepatitis C virus isolates from all over the world were

fragmented manually into individual 5′ and 3′ UTR sequences. Similarly the protein

sequences of ORF were also manually separated from the whole genome. The RNA

sequences of HCV were retrieved from NCBI viral genome database

HCV Virus genotype	Location	Accession Number	Year of Collection	UTR Sequence
2a	Pakistan	AB444580.1	2009	5′
2	Egypt	HQ228206	2011	5′
2a	US	NC_009823	2019	5′
2	Japan	D63821	2005	5′
3a	Pakistan	JN588558	2011	5′
3a	India	GQ275355	2011	5′
3a	US	D28917	2005	5′
3a	Pakistan	JN588558	2011	3′
3a	India	GQ275355	2011	3′
3a	US	D28917	2005	3′

Table 3.1 : Accession number of HCV isolates used in the study

For the interaction of HCV virus 5′ and 3′ UTR regions with mRNA of viral proteins we have used Hepatitis C virus genotype 3a as our model. Database of viral protein mRNA sequences are given in table. They are all from Pakistani HCV genotype 3a virus.

Viral protein	Accession number
Core protein	KY364192
Glycoprotein E1	ABY83293
Glycoprotein E2	EU399721
Viroporin p7 (NS1)	AM423050
RDRP (NS5B)	HM172568

Non-structural 4A (NS4A)	HQ822054
Non-structural 4B(NS4B)	GQ325251
NS2	FJ865505

Table 3.2 : Accession number of viral proteins used in study

In NCBI viral genome database the sequences were retrieved by addition of above mentioned accession number of HCV genomes and proteins. After retrieval of whole genome sequence they were fragmented manually into 5′ and 3′ UTR sequences. Sequences were then saved in FASTA format.

Steps:

1. Open NCBI viral genome database

2. Add accession number of HCV genome or desired proteins in the search tab then select nucleotide.

3. Click the search option

4. Whole genome sequence will be opened

5. Manually fragment it into 5′ and 3′ UTR sequence

6. Save file in FASTA format

3.2 Clustal W

Clustal W software was used for the Multiple Sequence Alignment and functional analysis of 5′ UTR region of the dengue viral database. From the result of this tool we can infer the homology and evolutionary relationships between the sequences studied. The input sequences are added to the tool and the output is clustal alignment format without residue or base numbering.

Steps:

1. Open Mega X

2. Go to Align option

3. Select Edit/Build alignment
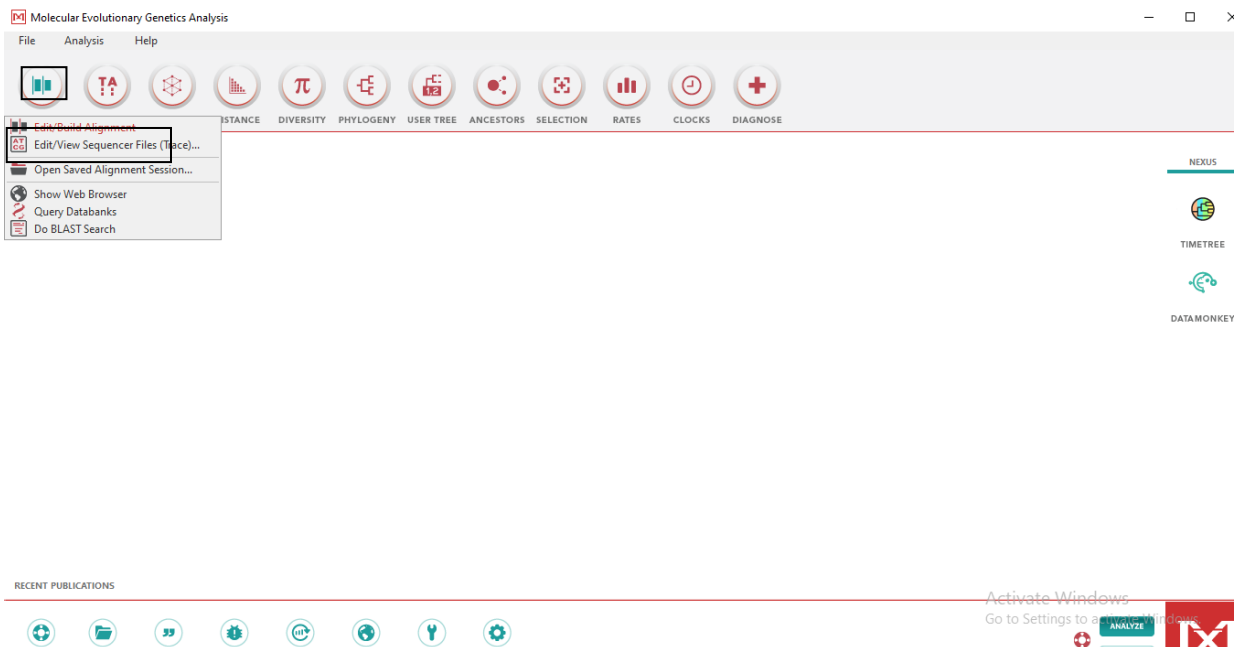
4.While selecting DNA/protein select DNA alignment
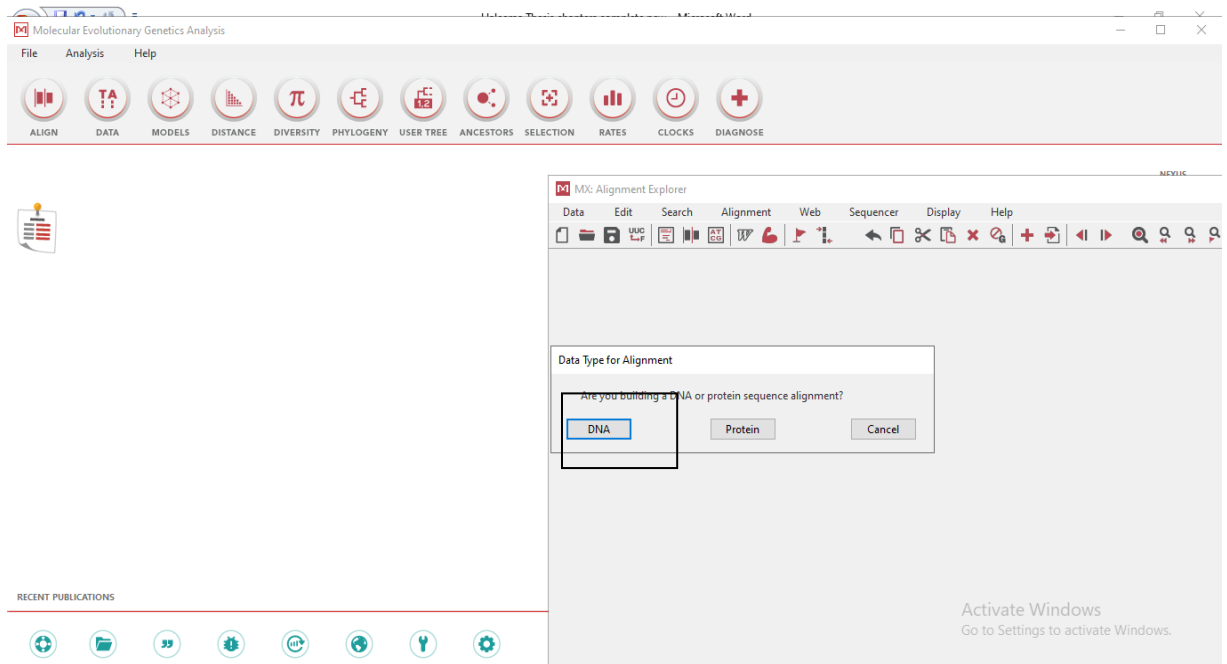
5.Copy and paste the UTR sequences retrieved from NCBI
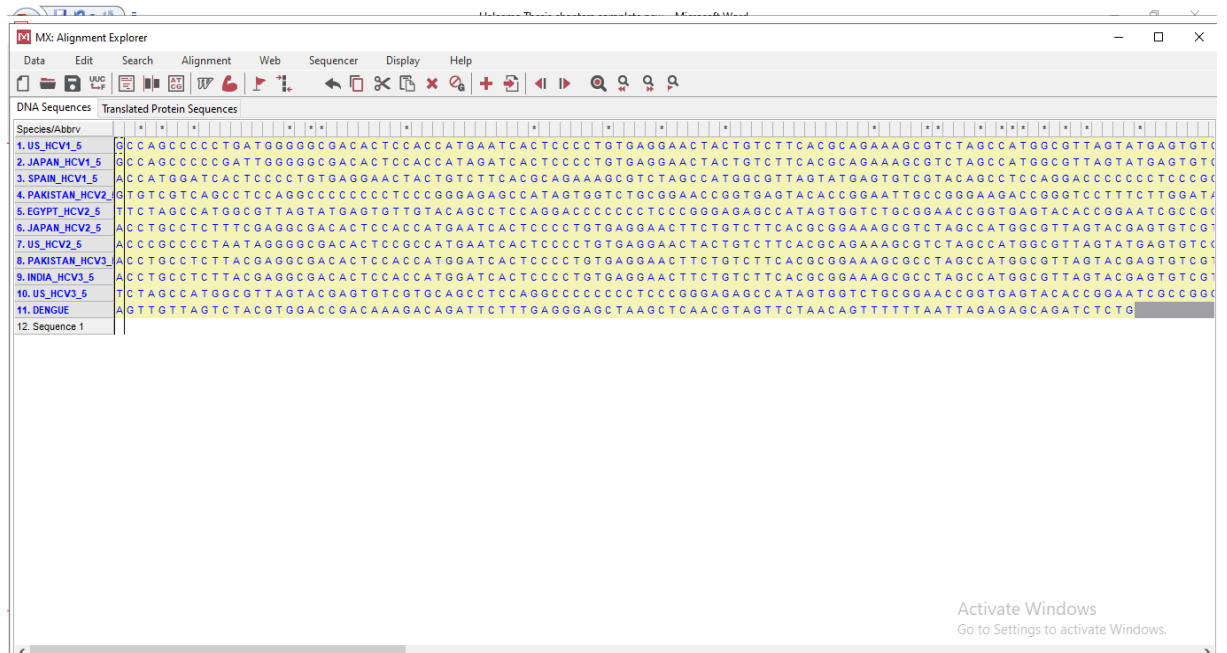
7. In alignment option select alignment by Clustal W.
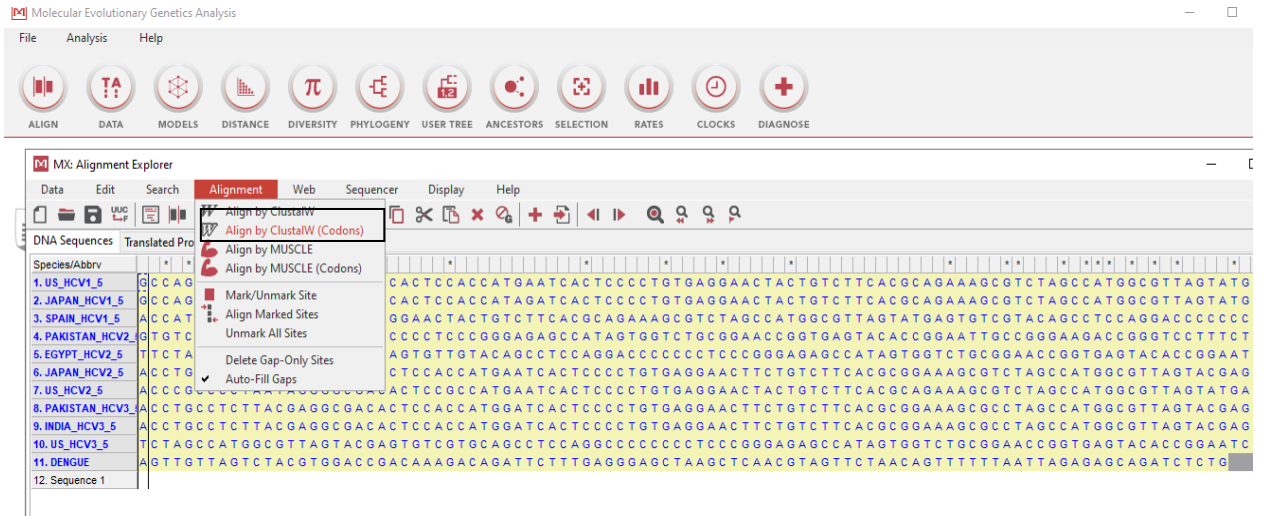
8. Select default parameters and press ok.
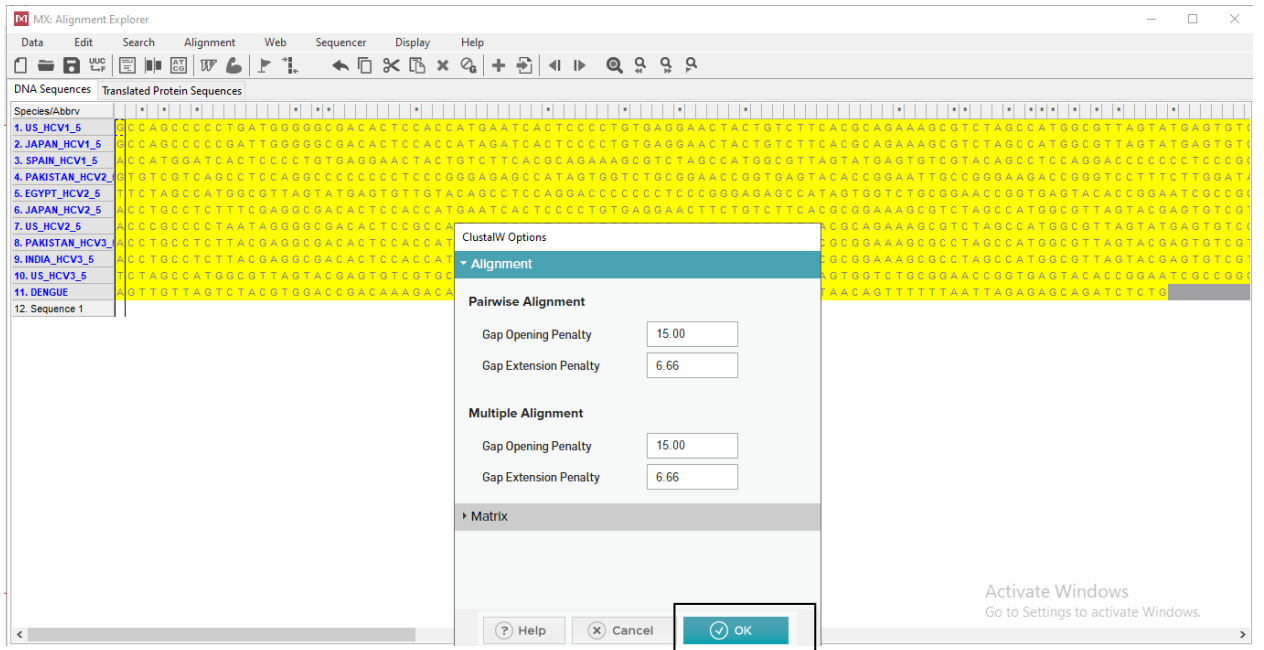
8. The sequences will be aligned as follows

3.3 RNA fold

The RNA fold is used for the prediction of secondary structures .5′ and 3′ UTRS are single

stranded RNA sequences an using this online web tool we compared the secondary

structures of HCV genotype 2a-3a. It displays interactive RNA secondary structure plot

with reliability annotation (partition function folding only) along with the mountain plot.

The settings are as follows

Steps:

1. Open RNAfold

2. Copy and paste the retrieved FASTA format UTR sequence

3. Among basic settings the folding algorithm was set to show minimum free energy

(MFE) and partition functions

   o Besides the minimum free energy (MFE) structures this also calculates the

   base pairing probability matrix and partition functions

4. Isolated base pairs: Not allowed

3.4 RNAalifold

The RNAalifold is a bioinformatics tool that will forecast an agreement of secondary

structures of a set of aligned sequences of single stranded RNA.

Steps

1. Sequences were first aligned using Clustal W and then saved in FASTA format before running in RNAalifold.
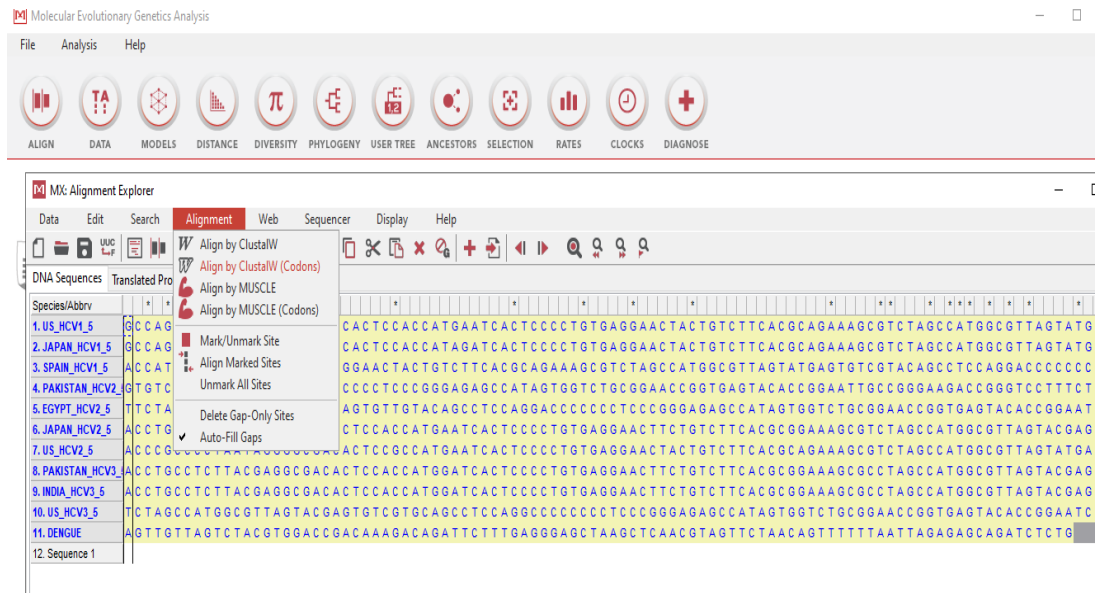
2. The output options are similar as the RNAfold tool in previous section. Base setting related to folding algorithms is also the same.

3. But RNAalifold has an additional setting:

   RIBOSUM scoring: Ribsoum scoring matrices are used which exhibit better gap character handling and the matrix is selected automatically according to the maximal and minimal pairwise identities of the sequences in the alignment file

3.5 IntaRNA

For the accurate and speedy predictions of interactions between two RNA sequences or molecules IntaRNA is program used.

Steps

1. Opening of IntaRNA

2. Both the target (mRNA) and query(UTR) sequences are added in FASTA or other supported format. The settings are as follows.

Output parameters

- o No of suboptimal interaction=0

Seed parameters

- o Minimum number of base pair in seed=7

- o Maximum numbers of mismatches in seed=0

Folding parameters

o   Folding window size target=150 nucleotides

o   Maximum base pair distance= 100 nucleotides

3.6 Mega X

This tool allows the interference of evolutionary relationships by using multiple aligned sequences and creating a phylogenetic tree that indicates the lineage and phylogeny of the various sequences fed to the software. 5′ UTR of dengue virus was used as an outgroup. The results were then displayed in a phylogenetic tree or cladogram.

Steps

1. Opening of Mega X

2.Sequences were first aligned through Clustal W

3.On phylogeny select tree construction through maximum likelihood

4.Bootsrap value was selected 100

Chapter 4

Results

Chapter 4: Results

4.1 Results

We conducted five different analyses in the study to fulfill the two main objectives of the

study. First to pinpoint the possible deletions/mutations sequences within the 5′ and 3′ UTR

sequences of HCV that lead to the generation of novel variations in the genome of the

virus. And second to study the evolutionary sequences of Hepatitis C viruses of various

genotypes across different countries.
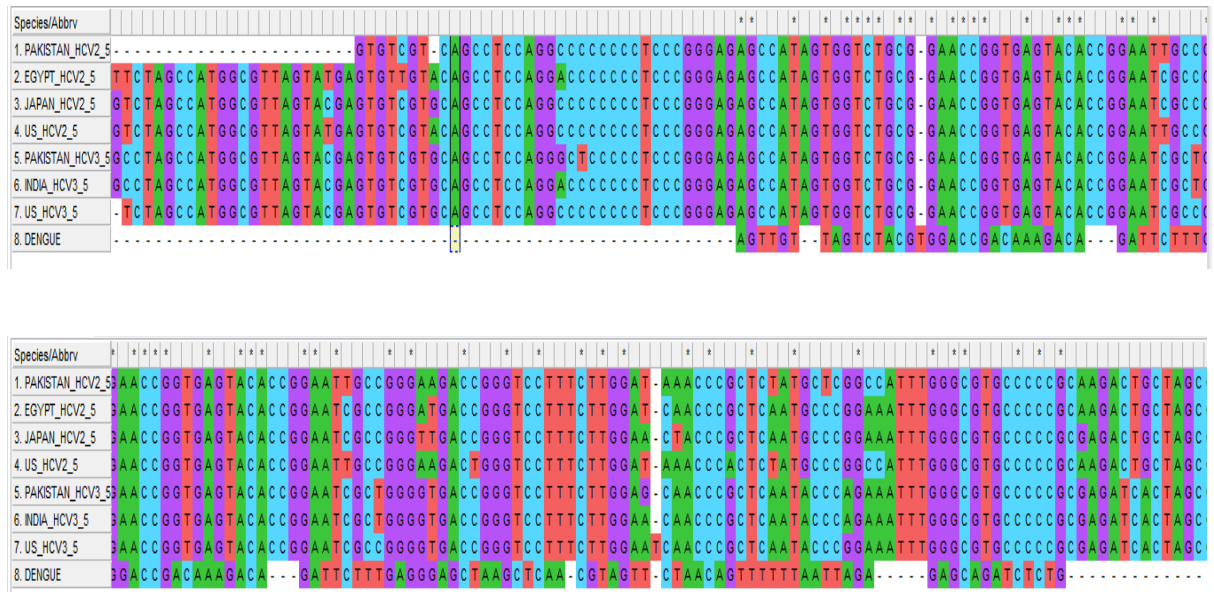
Figure 4.1: Multiple sequence alignment of 7 different 5′ UTR HCV isolates

For comparison of three or more biological sequences multiple sequence alignment is used.

We retrieved HCV genotypes 2 and 3 5′ UTR sequences from US, Japan, Spain, Pakistan,

India, Egypt. The results show highly conserved 5′ UTR sequences in the promoter

sequence among different HCV genotypes around the world..HCV-2a of Pakistan show

some deletions and nucleotide substitutions in 5′ UTR sequence while HCV-3a Pakistan

show high conservation in sequence when compared to other genotypes The steric sign shows the conserved sequences. Dengue is used as an outlier which is used to validate the results. The 5′ UTR sequence is mostly conserved that is why it is used for diagnostic purposes.

4.2 Secondary structures of HCV 5′ and 3′ UTR sequences drawn using RNA fold

Secondary structure prediction for RNA is essentially a collection of predicted base pairs following a set of rules and certain algorithms. Base pairs can be either G-C or A-U or the weaker G-U pair. Secondary structures can be used to regulate the replication of single stranded RNA viruses and also for further three dimensional modeling. The comparative approach requires RNA sequences to be similar in length so that they can be reliably aligned.
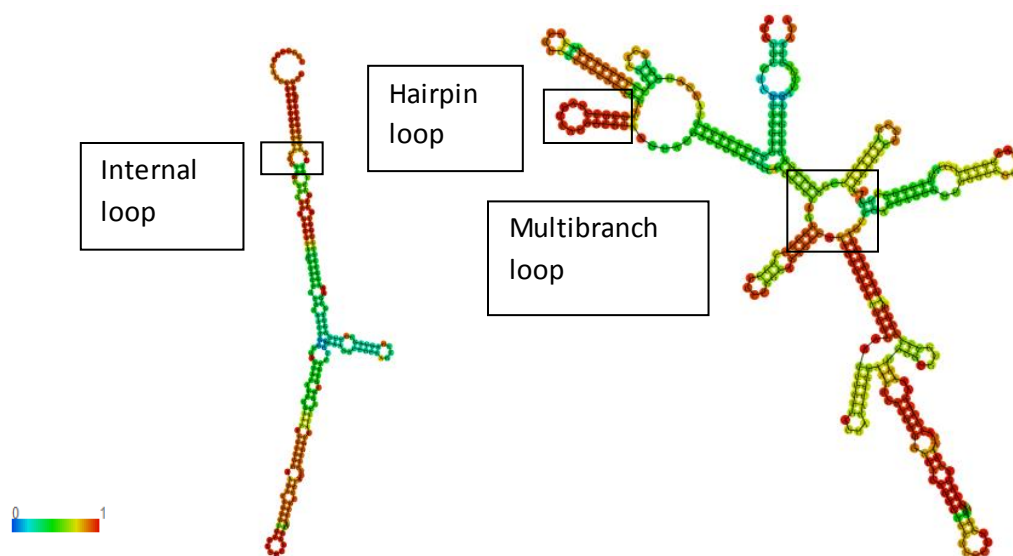
Fig 4.2a HCV-2a 5′ UTR Pakistan   **Fig 4.2b** HCV-3a 5′ UTR Pakistan

Figure 4.2(a,b) Comparison of secondary structures of 5′ UTRs HCV-2a and HCV-3a of Pakistan for stability evaluation

Figure 4.2 shows a comparison of HCV-2a and HCV-3a 5′ UTR sequences of Pakistan. The different color show coding of base-pair probabilities. Red color shows highest probability of base pairing as HCV-3a 5′ UTR is comparatively more stable and has low minimum free energy than HCV-2a 5′ UTR and blue color show least base pair probability. Low MFE shows less energy needed for folding and energetically stable structure. HCV-3a 5′ UTR has more stem loops structure as compared to HCV-2a 5′ UTR which also play parts in giving stability to structure.

Region	**Secondary structure of 5′ UTR**	**Minimum free energy(kcal/mol)**	**Hairpin loops**	**Internal loops**	**Multibranch loops**
Pakistan	HCV-2a	-92.50	2	13	0
Pakistan	HCV-3a	-140.80	9	7	2

Table 4.1: Comparison of secondary structures of 5′ UTRs HCV-2a and HCV-3a of Pakistan for stability evaluation
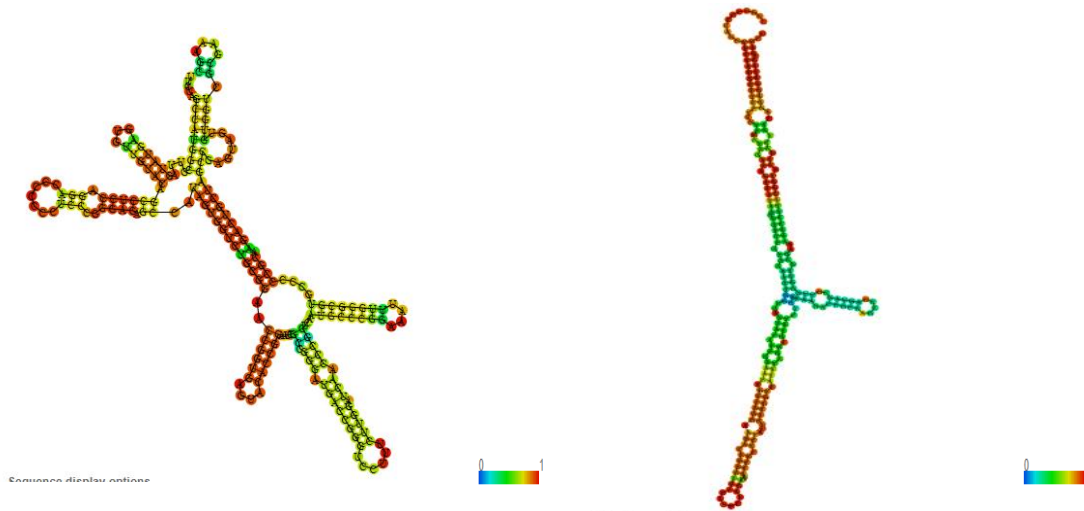
Fig 4.3a: 5' UTR of HCV-2 Egypt   **Fig 4.3b:** 5' UTR of HCV-2a Pakistan
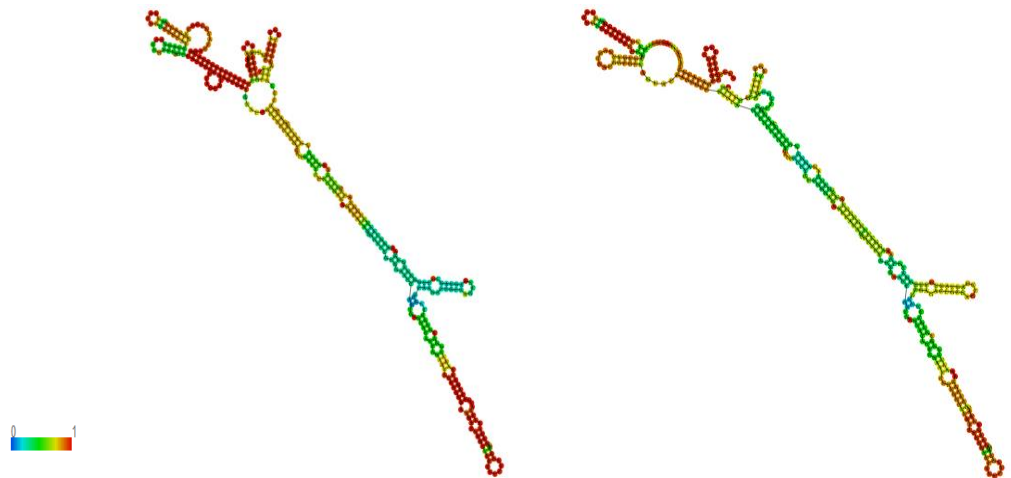
Fig 4.3c: 5′ UTR of HCV2-2a US   **Fig 4.3d:** 5′ UTR of HCV-2 Japan

Figure 4.3 (a,b,c,d): Comparison of secondary structures of 5′ UTRs HCV-2a of Pakistan with different countries for stability evaluation

Region	Secondary Structure of 5′ UTR	Minimum Free energy (Kcal/mol)	Hairpin loops	Internal loops	Multibranch loops
Pakistan	HCV-2a	**-92.50**	2	13	0
Egypt	HCV-2	**-72.40**	6	3	1
US	HCV-2a	**-141.20**	6	13	1
Japan	HCV-2	**-136.10**	6	12	1

Table 4.2: Secondary structures of HCV-2 5′ UTR of Pakistan, Egypt, US, Japan and their

minimum free energies

Table 4.2 shows a comparison HCV-2a 5′ UTR secondary structures of Pakistan, US and Japan and their minimum free energies. The most stable structure is HCV-2a 5′ UTR of US as it has more less MFE followed by Japan, Pakistan and Egypt
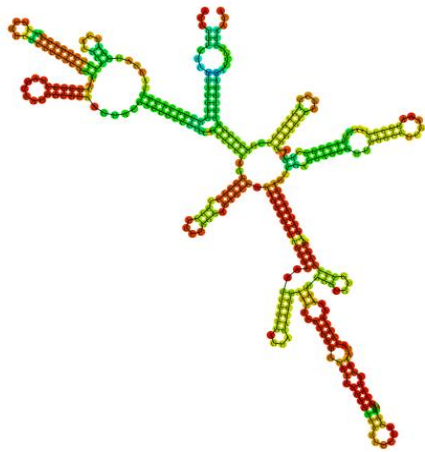
Fig 4.4a: 5′UTR of HCV-3a Pakistan   **Fig 4.4b:** 5′ UTR of HCV-3a India
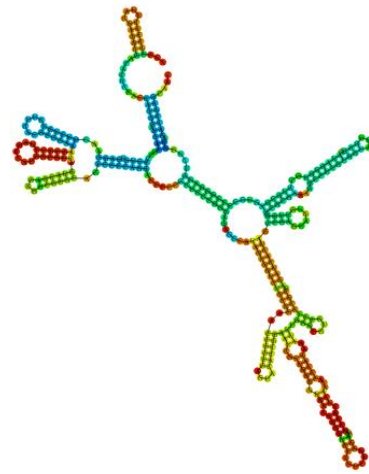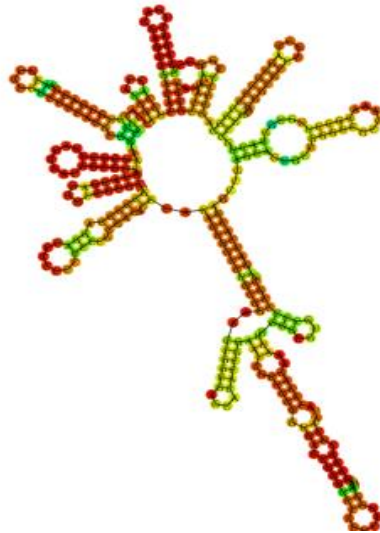
Fig 4.4c: 5′ UTR of HCV-3a US

Fig 4.4(a,b,c): Comparison of secondary structures of 5′ UTRs of HCV-3a of Pakistan with different countries for stability evaluation

Region	Secondary structure of 5′ UTR	Minimum free energy(kcal/mol)	Hairpin loop	Internal loop	Multi Loop
Pakistan	HCV-3a	**-140.80**	9	7	2
India	HCV-3a	**-135.60**	9	4	2
US	HCV-3a	**-144.75**	12	6	1

Table 4.3: Secondary structures of 5′ UTR of HCV-3a of Pakistan, HCV-3a India and

HCV-3a US and their minimum free energies

Table 4.3 shows a comparison of 5′ UTR of Pakistan, India and US and their minimum free energies. The most stable secondary structure is HCV-3a of US as it has less MFE as compared to other structures.
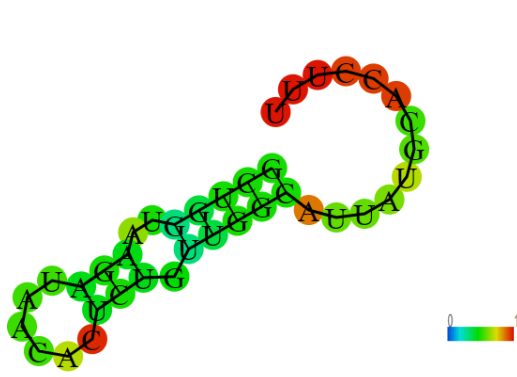
Fig 4.5a: HCV-3a 3′ UTR of Pakistan   **Fig 4.5b:** HCV-3a 3′ UTR of India
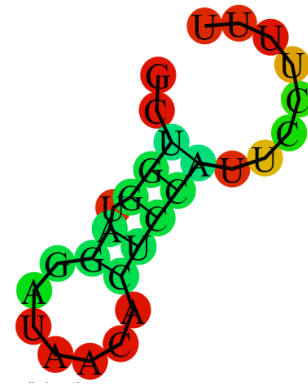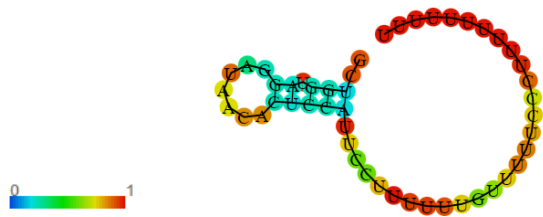
Fig 4.5c: HCV-3a 3′ UTR of US

Figure 4.5(a,b,c): Comparison of 3′ UTR of HCV-3a Pakistan, India and US

Region	Secondary Structure of 3′ UTR	Minimum Free energy (Kcal/mol)	Hairpin loop	Internal loop	Multi branch loop
Pakistan	HCV-3a	**-4.20**	1	1	0
India	HCV-3a	**-2.90**	1	0	0
US	HCV-3i	**-3.97**	1	0	0

Table 4.4: Comparison of secondary structures of 3′ UTRs of HCV-3a of Pakistan with

different countries for stability evaluation

The above table shows a comparison of secondary structures of HCV-3 Pakistan, India and US. These structures are relatively less complex as compared to 5' UTR sequences because the 3′ UTR is less lengthy as compared to 5′ UTR sequence. The most stable secondary structures of 3′ UTR is of Pakistan as it has low minimum free energy in comparison to other structures. The minimum free energy is higher as compared to 5′ UTR of HCV-3 because the 3′ UTR structure is relatively less complex as compare to 5′ UTR of HCV-3 of different countries.

4.3 Consensus 5′ UTR sequences of Hepatitis C virus genotypes HCV-2a and HCV-3a from different countries and Pakistan using RNAalifold

This tool predicts consensus secondary structures of the set of all RNA sequences aligned.

In the first case the sequences are 5′ UTRs sequences of HCV-2 and HCV-3 of US, India, Egypt, Pakistan are aligned with each other.
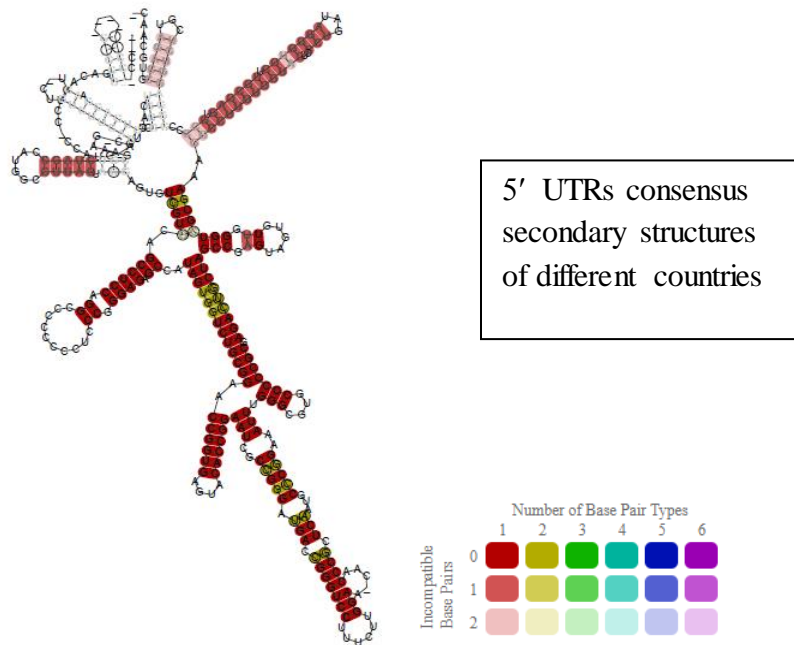
Figure 4.6: 5′ UTR consensus secondary structure of different countries

Figure shows the consensus sequence of UTRs. This sequence can be used for testing and these sequences show high conservation. According to color code the red color shows sequences which are conserved with each other while yellow color shows mutations.

4.4 Interaction of 5′ and 3′ UTRs with viral mRNAs protein sequence

IntaRNA generally shows interaction of non-coding RNA like 5′ and 3′ with RNA coding

proteins both structural and non-structural. We will be taking Pakistan Hepatitis C virus

genotype 3 as model because it is the most common type in the country.

- **Interaction between HCV -3a 5′ UTR and Core protein**

```
                         334                    342
                          |                      |
             5'-AUG..ACGAC          AGGUC..UGU-3'
   Target                  CCCCGGCGA
                           | | | | | | | | |
   Query                   GGGGUCGCU
             3'-ACA..CCAGU          AAGGC..CCA-5'
                          |                      |
                         180                    172
```

Fig 4.7a: Best interaction is nucleotide position 335-341 for core protein with 173-179 position for 5′ UTR

- **Interaction between HCV -3a 5′ UTR and E1 protein**

```
                         225                    235
                          |                      |
             5'-CUA..UUAGU          CGAUG..GCC-3'
   Target                  GGGCGCGCCCA
                           | | | | | | | | | | |
   Query                   CCCGUGCGGGU
             3'-ACA..GCGCC          UUAAA..CCA-5'
                          |                      |
                         236                    226
```

Fig 4.7b: Best interaction is nucleotide position 226-234 for E1 protein with 227-235 position for 5′ UTR

- **Interaction between HCV -3a 5′ UTR and E2 protein**

```
                82                                      107
                |                                        |
        5'-CAC..AGAAA      GCU       ACAC        GGGGC..GCA-3'
Target            CUGCA     GAUCA     CUAUGGCUC
                  | | | | |  | | | | |  | | | | | | | | |
Query             GGCGU     CUGGU     GAUACCGAG
        3'-ACA..GCCAA                            AGGGC..CCA-5'
                |                                        |
               151                                      133
```

Fig 4.7c: Best interaction is nucleotide position for 83-106 for E2 protein with 134-150 position for 5′ UTR

- **Interaction between HCV- 3a 5′ UTR and P7 protein**

```
                140                                     159
                |                                        |
        5'-GCC..GGGUC      GUC              GUCCU..GCU-3'
Target            UUUG    CCUAGCAUUGCUC
                  | | | |  | | | | | | | | | | | |
Query             AAGC    GGGUUGUGAUGAG
        3'-ACA..CCGGA      GCU              CCGAU..CCA-5'
                |                                        |
               273                                      254
```
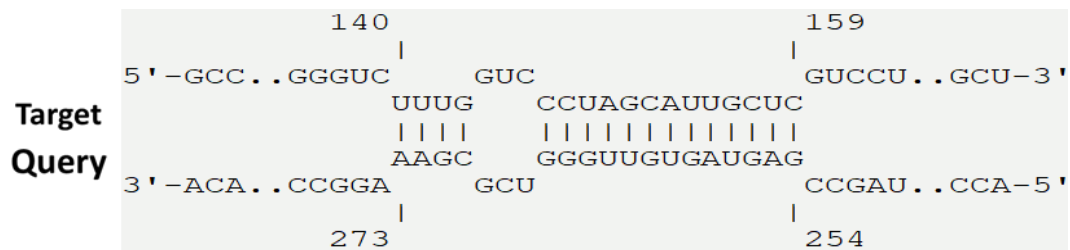
Fig 4.7d: Best interaction is for nucleotide position 141-158 position for P7 protein with 255-272 position for5' UTR

- **Interaction between HCV-3a 5'UTR and NS5B protein**

```
               18                                       47
               |                                         |
       5'-AUG..UGGGU      G      U   UCG         AGCCU..UGC-3'
Target           GUUGCUC    GAGGGG CC   CGG   CCCUAGC
                 | | | | | | |  | | | | | | | |  | | |  | | | | | | |
Query            CAACGAG    CUUUCC GG   GCC   GGGGUCG
       3'-ACA..UCGCC      GUU      U         AGU    CUAAG..CCA-5'
               |                                         |
              205                                       174
```
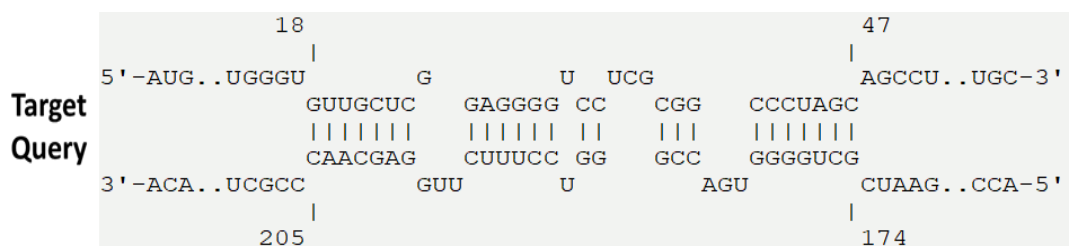
Fig 4.7e: Best interaction is for nucleotide position 19-48 for NS5B protein with 175-204 position for 5′ UTR.

- **Interaction between HCV- 3a 5′ UTR and NS4A protein**

```
                    458                                    481
                     |                                      |
          5'-UCA..CUUGG          AAAAUCAU          AACUC..UGU-3'
Target              UGGCUUUC              GGGGGGGG
                    | | | | | | | |       | | | | | | | |
Query               ACCGAGAG              CCCUCCCC
          3'-ACA..GUGAU           GG                 CUCGG..CCA-5'
                     |                                      |
                    138                                    121
```
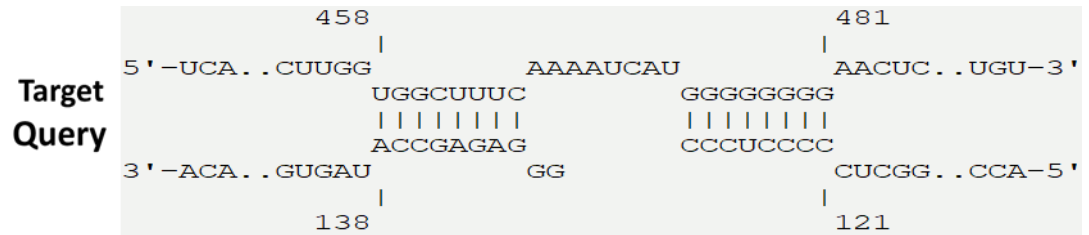
Fig 4.7f: Best interaction is for nucleotide position 459-480 for NS4A protein with 122-137 nucleotide position for 5′ UTR

- **Interaction between HCV -3a 5′ UTR and NS2 protein**

```
                  9                              31
                  |                              |
Target  5'-CUUUGGUC            UCC           UCCCC..UUA-3'
           CCUAGCAUUGCUCG          UCUUGC
           | | | | | | | | | | | | | | | |    | | | | | |
           GGGUUGUGAUGAGC       AGAGCG
Query 3'-ACA...GCGCU          CGAUCACU      CCCCC...CCA-5'
               |                     |
              266                   239
```
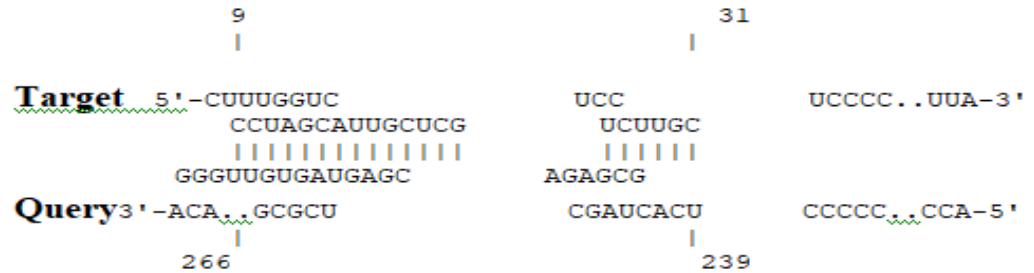
Fig 4.7g: Best interaction is for nucleotide position 10-30 for NS2 protein with 240-265 nucleotide position for 5′ UTR

- **Interaction between HCV-3a 3′ UTR and Core protein**

```
     550                          568
                |                           |
     5'-AUG..UCUCU     U       CCAU        CUUGU-3'
Target          UGC  UAAU       CCAACAG
                | | |  | | | |     | | | | | | |
Query           ACG  AUUA       GGUUGUC
         3'-UUUCC    U        C            UCACA..UCG-5'
                |                           |
               33                          18
```
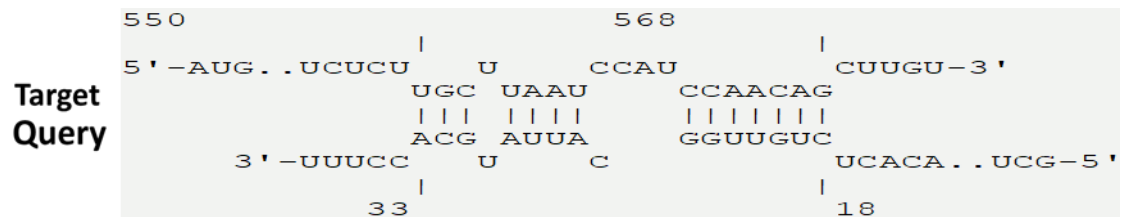
Fig 4.8a: Best interaction is nucleotide position 551-567 for core protein with 19-32 nucleotide position for **3′** UTR

- **Interaction between HCV-3a 3′ UTR and E1protein**

```
34                40
                  |              |
5'-CUA..AUGUC              ACGAC..GCC-3'
                  CUUACCA
                  |||||||
                  GAAUGGU
3'-UUU..CAAUA              CG-5'
                  |        |
                  9        3
```
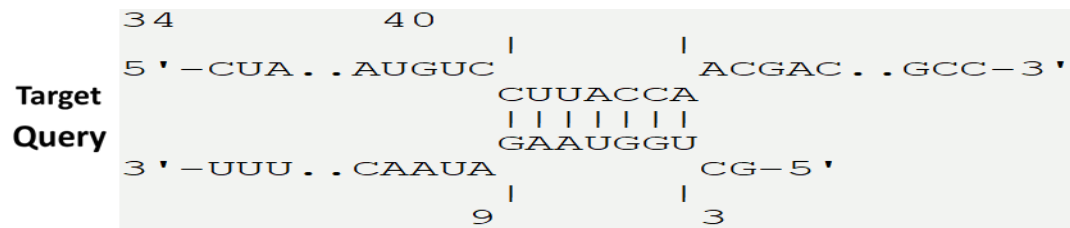
Fig 4.8b: Best interaction is nucleotide position 35-39 for EI protein with 4-8 nucleotide position for **3′** UTR

- **Interaction between HCV-3a 3′ UTR and E2 protein**

```
919                        934
        |                        |
5'-CAC..UGCCC                    UAAUA..GCA-3'
          GCAU    UGUCAACAGGGU
          ||||    ||||||||||||
          CGUA    ACGGUUGUCUCA
   3'-UUUCCA    UU              CAAUA..UCG-5'
          |                        |
          32                      15
```
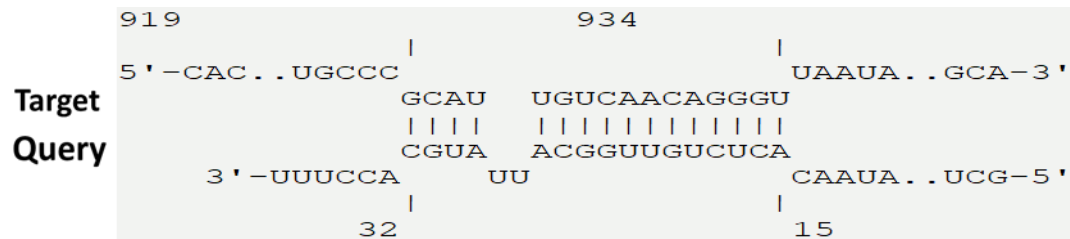
Fig 4.8c: Best interaction is nucleotide position 920-933 for E2 protein with 16-31 nucleotide position for **3′** UTR

- **Interaction between HCV-3a 3′ UTR and P7 protein**

```
            85          91
            |           |
5'-GCC..GCGCA        ACGUG..GCU-3'
            GCGUGGU
            |||||||
            CGUAUUA
   3'-UUUCCA        CGGUU..UCG-5'
            |           |
            32          26
```
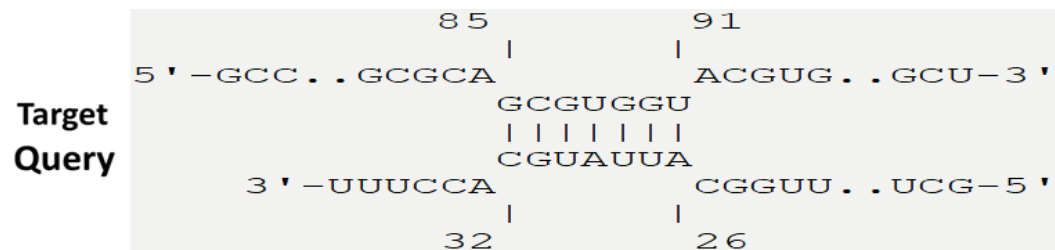
Fig 4.8d: Best interaction is nucleotide position 86-90 for P7 protein with 27-31 nucleotide position for 3′ UTR.

- **Interaction between HCV-3a 3′ UTR and RDRP protein**

```
                    345           351
                     |             |
        5'-AGU..UUCUC          ACCCG..UGG-3'
Target              GUAUGAU
                    | | | | | | |
Query               CGUAUUA
            3'-UUUCCA          CGGUU..UCG-5'
                     |             |
                    32            26
```
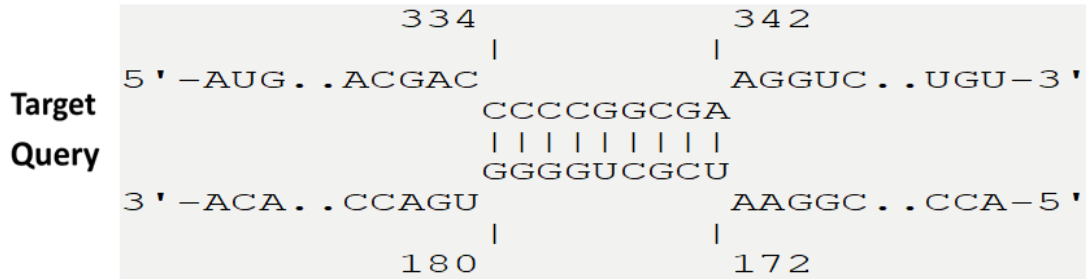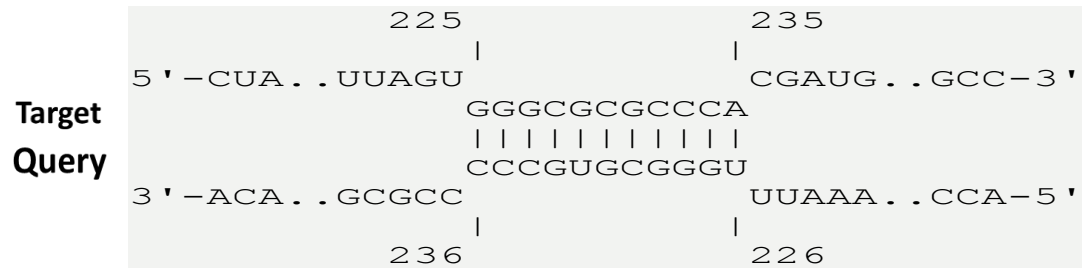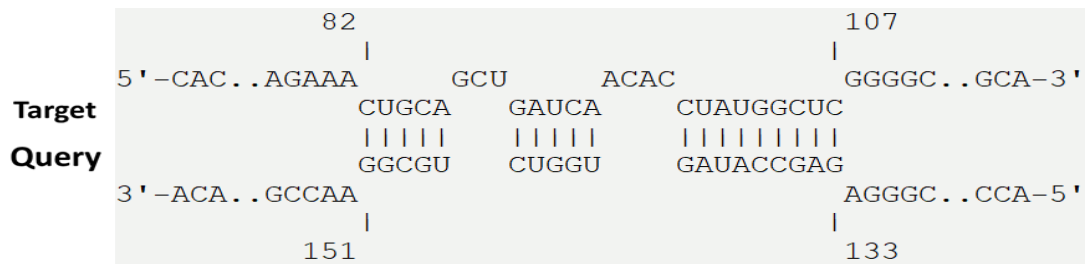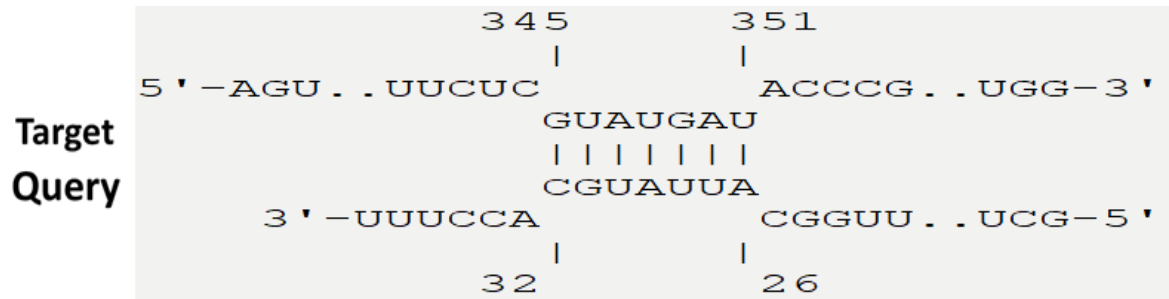
Fig 4.8f: Best interaction is nucleotide position for 346-350 for RdRp protein with 27-31 position for 3′ UTR

- **Interaction between HCV-3a 3′UTR and NS4A protein**

```
                  4                      19
                  |                       |
        5'-AUG          CCU     UGCUC..UGC-3'
Target       GCCAGCA     GGGUGU
             | | | | | | |   | | | | | |
Query        CGGUUGU     CUCACA
     3'-UUU..UAUUA                AUAGA..UCG-5'
             |                       |
             25                      13
```
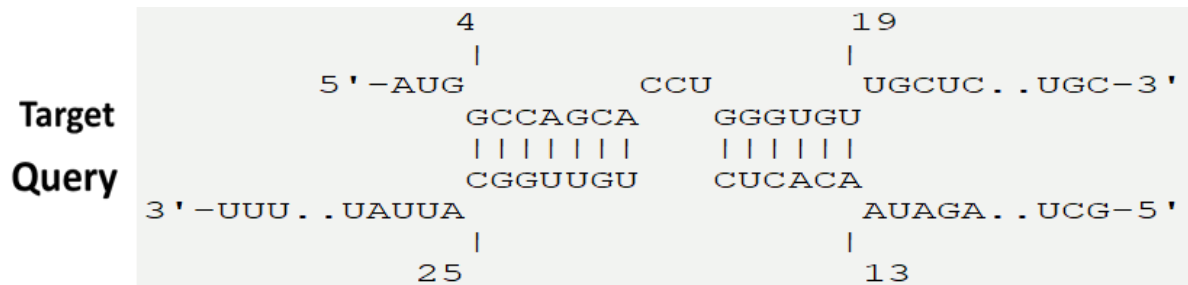
Fig 4.8g: Best nucleotide position is 5-18 nucleotide position for NS4A protein with 14-24 position for 3′ UTR

- **Interaction between HCV3 3′ UTR and NS4B protein**

```
   490                                 521
        |                                 |
5'-UCA..UCCCC        G   C  G   CA    GU    C  CAUAC..UGU-3'
Target         AAUGCCGA  GA GU GU   AUCU  UGCC GC
               | | | | | | | |  | |  | |  | |   | | | |   | | | |  | |
Query          UUACGGUU  CU CA CA   UAGA  AUGG CG
3'-UUUCCACGUA        GU        A            U-5'
        |                                 |
       28                                 1
```
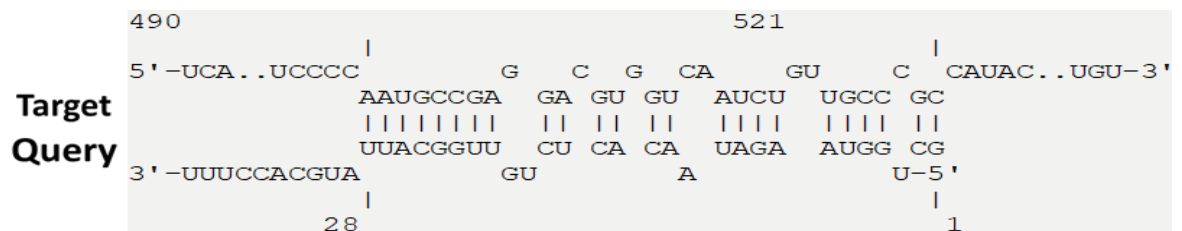
Fig 4.8h: Best nucleotide position is for 491-520 position for NS4B protein with 2-29 position for 3′ UTR

Hybridization Energy (kcal/mol)	Core Protein	E1	E2	Viroporin P7(NS1)	NS2	NS4A	NS4B	Ns5A	RdRp NS5B
5' UTR	-20.58	-23.94	-23.25	-20.58	-24.92	-35.37	-26.67	_	_
3' UTR	_	_	-21.16	_	_	_	_	_	_

Table 4.5: Hybridization energies of interaction of 3' and 5' UTRs with viral proteins

We have only included the results with hybridization energies <- 20kcal/mol which shows highly favorable interaction. Low hybridization energy <-20 means less energy needed for interaction + highly stable complex.

4.5 Phylogenetic tree

The sequences of both UTRs of Hepatitis C virus from all over the world were fed in Mega 6 software and phylogenetic tree was constructed
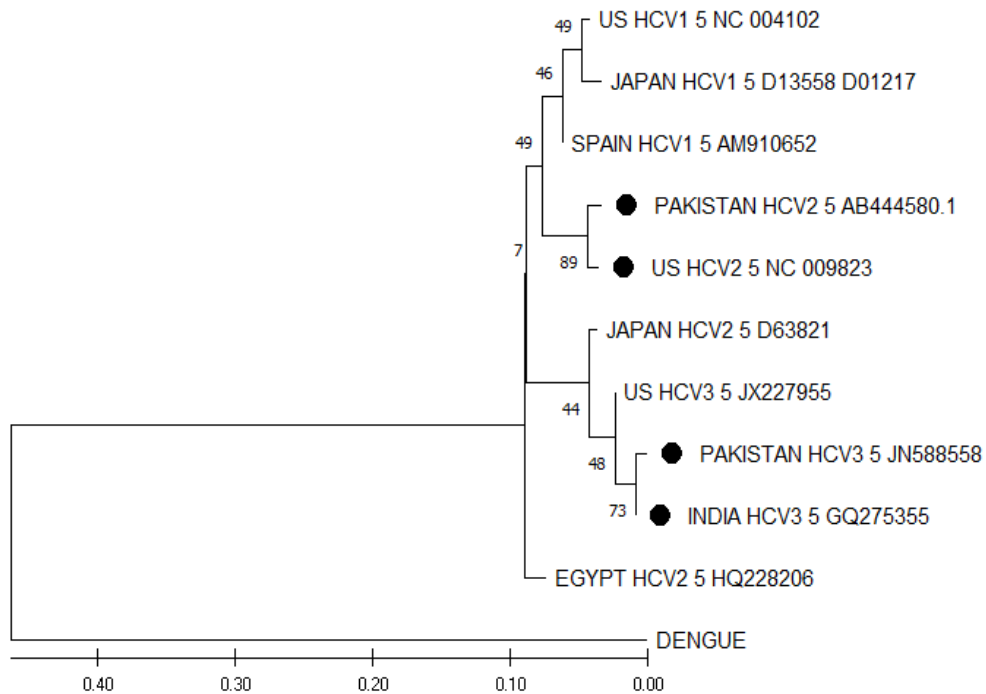
Fig 4.9: Phylogenetic tree of 5′ UTRs of constructed through Mega X

Phylogenetic tree was constructed using Mega X software. The results show that Pakistani HCV-2a 5′ UTR sequences is closely related to US HCV-2a 5′ UTR sequence. Pakistan HCV-3a 5' UTR sequence is closely related to India HCV-3a 5' UTR sequence.
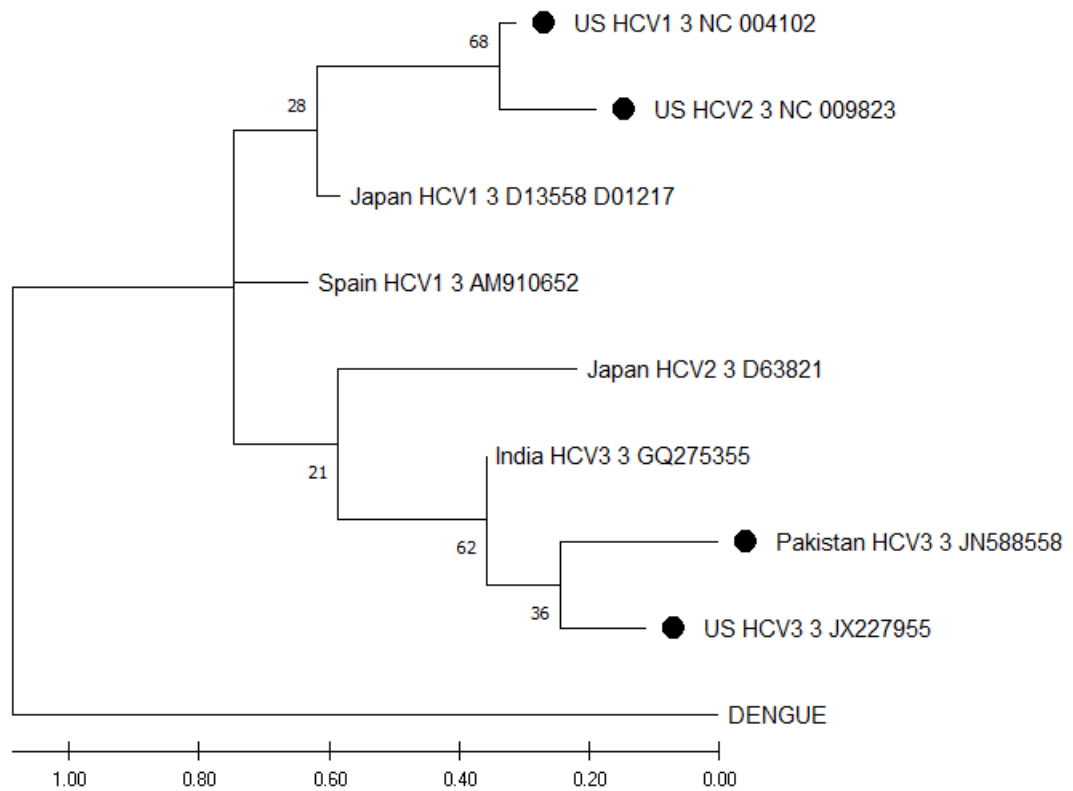
Fig 4.10: Phylogenetic tree of 3′ UTRs of constructed through Mega X

Phylogenetic tree was constructed using Mega X software. The results show that Pakistani HCV-3a 3' UTR sequence is closely related to US HCV-3i 3' UTR sequence.

Chapter 5

Discussion

Chapter 5: Discussion

Discussion

Persistent hepatitis C virus (HCV) sepsis is a dominant communal-health problem accountable for cirrhosis, liver damage and hepatocellular carcinoma (HCC) (Arshad & Ashfaq, 2017). Hepatitis C is a worldwide well-being problem infecting 177.5 million people and approximately 71.1 million subjects are contaminated severely with hepatitis C virus (HCV) (Spearman et al., 2019). According to a recent estimate Pakistan has the second-largest load of hepatitis C, with a countrywide prevalence of 4.8% (Abbas & Abbas, 2020).

The protein production of HCV starts from a preserved region known as 5' untranslated region that constitutes IRES which contain start codon and involved in initiation of translation. The region present at C-terminal known as 3' UTR is involved in duplication and translation (Sadia et al., 2013). The 5' and 3' UTRs are involved in translation and replication of HCV so in this study we conducted various analyses on these non-coding regions.

In this study the result of multiple sequence alignment of 5' UTR of HCV-2a and HCV-3a Pakistan compared with other countries show high conservation in 5' UTR of 324 to 341 nucleotides sequences. Our findings has been in accordance with former findings done by (Beales et al., 2015) in which he showed that sequencing data of HCV genome show that among HCV strains the 5' UTR of 324 to 341 nucleotides is the most unchanged region..

The secondary structures of 5' and 3' UTRs of HCV-2a and HCV-3a of Pakistan compared with other countries i.e. India, Egypt, US and Japan show significant structural differences and varying degree of stability due to hairpin and internal loop structures. There are no other reports on HCV UTRs secondary structures. But secondary structures of coding region of Dengue virus proposed by (Karen et al., 2016) showed conservation in hairpin structures. The difference is due to the fact that UTRs are non-coding regions and when they fold into complex secondary structures show varying degree of stability as UTRs sequence length vary from one genotype to other genotype which leads to different base pair attractions resulting in variant secondary structures and stabilities. In this study the secondary structures of 5' UTRs of HCV-2a and HCV-3a have less minimum free energies calculated as compared to 3' UTRs of the same genotypes. However secondary structures of 5' UTRs of Dengue virus has more minimum free energies (MFE) as compared to 3' UTRs of Dengue virus (Ng et al., 2017). So our result is not in accordance with the study conducted on Dengue virus by (Ng et al., 2017) because in dengue virus 5' region is shorter in comparison to 3' UTR region while in HCV 5' UTR region is longer as compared to 3' UTR and as MFE depends on sequence length so greater the sequence length the MFE calculated will be lower.

The interaction of 5' and 3' UTRs of HCV-3a Pakistan with viral protein mRNAs show promising results. UTRs when hybridize with mRNA of viral proteins will either provide stability to mRNA or destabilize it resulting in translational activation or translational suppression. In our insilico study we predicted that when 5' UTR interact with mRNA sequence of core protein it results in translational suppression. Our result is in line with the

in-vitro study conducted by (Shimoike et al., 2015) which stated that interconnection of core protein HCV with viral mRNA and UTRs leads to inhibition of its translation.

In this insilico study we predicted that when 5' UTR pf HCV-3a Pakistan interact with mRNA sequence of envelope glycoproteins E1 and E2 protein it results in translational activation. Our result is in line with the in-vitro study conducted by (Taylor et al., 2015) which stated that interconnection of HCV envelope protein with viral mRNA and UTRs leads to activation of its translation by inhibiting protein kinase (PKR).

In this insilico study we predicted that when 5' UTR pf HCV-3a Pakistan interact with mRNA sequence of non-structural (NS2 and NS5A) protein it results in translational activation. There are no other reports on interaction of these protein mRNA sequence with UTRs of hepatitis C virus. But a similar in-vitro study conducted by (Alvarez et al., 2015) on Dengue virus non-structural protein interaction with 3' UTR leads to translational activation.

The results of phylogenetic analysis show that Pakistani HCV-2a 5' UTR sequences is closely related to US HCV-2a 5' UTR sequence and 5' UTR of Pakistan HCV-3a sequence is closely related to India HCV-3a 5' UTR sequence. Our study has been in accordance with the previous results in which phylogenetic analysis of whole genome of HCV genotype 2a and 3a of Pakistani were found to be nearly related to many countries including India, Thailand, Vietnam, Argentina, USA, Brazil, Russia, Afghanistan and Mexico (Ghori et al., 2016). Pakistani sequences were set up to be nearly related to US because a number of populations were found to reside in US due to which they have shared the genotype sequences with each other. Genotype 3a is most common in Pakistan (Aleena et al., 2019) and it has been described in many researches that genotype 3a has has emerged

in Indian-subcontinent many years ago. So HCV genotype 3a of India and Pakistan were found to be closely related as India and Pakistan was Indian subcontinent for decades (Zehender et al., 2013).

Chapter 6

Conclusion and future prospects

6.1 Conclusion and future prospects

The result of our study suggested some possible deletion sequences as identified through multiple sequence alignment that will reduce the genesis of contagious hepatitis C virus. The finding of our study demonstrated that UTRs sequences show conservation in primary structures but secondary structures show varying degree of stabilities in some genotypes. The different secondary structure stabilities showed that in some genotypes the UTRs are more stable indicating that they could be more involved in development of cirrhosis. The result of interaction of viral protein mRNA sequences with UTRs demonstrated some potential target sites which could help the researchers in designing antivirals for hepatitis C virus in future.

Result of this study are useful to pinpoint the sequences where mutation/deletions can result in inhibition of viral replication and translation as shown by interaction with core, envelope, NS2, NS4A, NS4B and NS5A. Further experiments need to be conducted to confirm the predicted insilico results that will reduce viral pathogenicity and host cell cytopathicity. We hope that our findings of the 5' UTRs in HCV translation will facilitate in comprehending viral infection lifecycle.

Chapter 7

References

Chapter 7: References

References

Abbas, Z., & Abbas, M. (2020). The cost of eliminating hepatitis C in Pakistan. In *The Lancet Global Health* (Vol. 8, Issue 3, pp. e323–e324). https://doi.org/10.1016/S2214-109X(20)30036-X

Al Kanaani, Z., Mahmud, S., Kouyoumjian, S. P., & Abu-Raddad, L. J. (2018). The epidemiology of hepatitis C virus in Pakistan: Systematic review and meta-analyses. In *Royal Society Open Science* (Vol. 5, Issue 4). https://doi.org/10.1098/rsos.180257

Alter, H. J., & Seeff, L. B. (2000). Recovery, persistence, and sequelae in hepatitis C virus infection: A perspective on long-term outcome. In *Seminars in Liver Disease* (Vol. 20, Issue 1, pp. 17–35). https://doi.org/10.1055/s-2000-9505

Alter, M. J. (2011). HCV routes of transmission: What goes around comes around. *Seminars in Liver Disease*, *31*(4), 340–346. https://doi.org/10.1055/s-0031-1297923

Appel, N., Pietschmann, T., & Bartenschlager, R. (2005). Mutational Analysis of Hepatitis C Virus Nonstructural Protein 5A: Potential Role of Differential Phosphorylation in RNA Replication and Identification of a Genetically Flexible Domain. In *Journal of Virology* (Vol. 79, Issue 5, pp. 3187–3194). https://doi.org/10.1128/jvi.79.5.3187-3194.2005

Armstrong, G. L., Wasley, A., Simard, E. P., McQuillan, G. M., Kuhnert, W. L., & Alter, M. J. (2006). The prevalence of hepatitis C virus infection in the United States, 1999 through 2002. In *Annals of Internal Medicine* (Vol. 144, Issue 10, pp. 705–714). https://doi.org/10.7326/0003-4819-144-10-200605160-00004

Arshad, A., & Ashfaq, U. A. (2017). Epidemiology of hepatitis c infection in Pakistan:

Current estimate and major risk factors. *Critical Reviews in Eukaryotic Gene Expression*, *27*(1), 63–77. https://doi.org/10.1615/CritRevEukaryotGeneExpr.2017018953

Baril, M., & Brakier-Gingras, L. (2005). Translation of the F protein of hepatitis C virus is initiated at a non-AUG codon in a +1 reading frame relative to the polyprotein. In *Nucleic Acids Research* (Vol. 33, Issue 5, pp. 1474–1486). https://doi.org/10.1093/nar/gki292

Bashirullah, A., Cooperstock, R. L., & Lipshitz, H. D. (2001). Spatial and temporal control of RNA stability. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 98, Issue 13, pp. 7025–7028). https://doi.org/10.1073/pnas.111145698

Baumert, T. F., Ito, S., Wong, D. T., & Liang, T. J. (1998). Hepatitis C Virus Structural Proteins Assemble into Viruslike Particles in Insect Cells. In *Journal of Virology* (Vol. 72, Issue 5, pp. 3827–3836). https://doi.org/10.1128/jvi.72.5.3827-3836.1998

Beales, L. P., Rowlands, D. J., & Holzenburg, A. (2001). The internal ribosome entry site (IRES) of hepatitis C virus visualized by electron microscopy. In *Rna* (Vol. 7, Issue 5, pp. 661–670). https://doi.org/10.1017/S1355838201001406

Berenguer, M., Ferrell, L., Watson, J., Prieto, M., Kim, M., Rayón, M., Córdoba, J., Herola, A., Ascher, N., Mir, J., Berenguer, J., & Wright, T. L. (2000). HCV-related fibrosis progression following liver transplantation: Increase in recent years. In *Journal of Hepatology* (Vol. 32, Issue 4, pp. 673–684). https://doi.org/10.1016/S0168-8278(00)80231-7

Brinton, M. A., & Dispoto, J. H. (1988). Sequence and secondary structure analysis of the 5′-terminal region of flavivirus genome RNA. In *Virology* (Vol. 162, Issue 2, pp. 290–

299). https://doi.org/10.1016/0042-6822(88)90468-0

Bruix, J., Calvet, X., Costa, J., Ventura, M., Bruguera, M., Castillo, R., María Barrera, J., Ercilla, G., María Sanchez-Tapias, J., Vall, M., Bru, C., & Rodes, J. (1989). Prevalence of Antibodies To Hepatitis C Virus in Spanish Patients With Hepatocellular Carcinoma and Hepatic Cirrhosis. In *The Lancet* (Vol. 334, Issue 8670, pp. 1004–1006). https://doi.org/10.1016/S0140-6736(89)91015-5

Chen, S. L., & Morgan, T. R. (2006). The natural history of hepatitis C virus (HCV) infection. In *International Journal of Medical Sciences* (Vol. 3, Issue 2, pp. 47–52). https://doi.org/10.7150/ijms.3.47

Cho, H.-S., Ha, N.-C., Kang, L.-W., Chung, K. M., Back, S. H., Jang, S. K., & Oh, B.-H. (1998). Crystal Structure of RNA Helicase from Genotype 1b Hepatitis C Virus. In *Journal of Biological Chemistry* (Vol. 273, Issue 24, pp. 15045–15052). https://doi.org/10.1074/jbc.273.24.15045

Choo, Q. (n.d.). *Choo_ Isolation of a cDNA clone derived from a blood-born.*

Cocquerel, L., Wychowski, C., Minner, F., Penin, F., & Dubuisson, J. (2000). Charged Residues in the Transmembrane Domains of Hepatitis C Virus Glycoproteins Play a Major Role in the Processing, Subcellular Localization, and Assembly of These Envelope Proteins. In *Journal of Virology* (Vol. 74, Issue 8, pp. 3623–3633). https://doi.org/10.1128/jvi.74.8.3623-3633.2000

Conne, B., Stutz, A., & Vassalli, J. D. (2000). The 3' untranslated region of messenger RNA: A molecular "hotspot" for pathology? In *Nature Medicine* (Vol. 6, Issue 6, pp. 637–641). https://doi.org/10.1038/76211

Davis, G. L. (2002). Monitoring of viral levels during therapy of hepatitis C. In *Hepatology* (Vol. 36, Issue 5 I). https://doi.org/10.1053/jhep.2002.36798

El-Serag, H. B., Hampel, H., Yeh, C., & Rabeneck, L. (2002). Extrahepatic manifestations of hepatitis C among United States male veterans. In *Hepatology* (Vol. 36, Issue 6, pp. 1439–1445). https://doi.org/10.1053/jhep.2002.37191

Farci, P., Alter, H. J., Shimoda, A., Govindarajan, S., Cheung, L. C., Melpolder, J. C., Sacher, R. A., Shih, J. W., & Purcell, R. H. (1996). Hepatitis C Virus–Associated Fulminant Hepatic Failure. In *New England Journal of Medicine* (Vol. 335, Issue 9, pp. 631–634). https://doi.org/10.1056/nejm199608293350904

Farci, P., Alter, H. J., Wong, D., Miller, R. H., Shih, J. W., Jett, B., & Purcell, R. H. (1991). A Long-Term Study of Hepatitis C Virus Replication in Non-A, Non-B Hepatitis. In *New England Journal of Medicine* (Vol. 325, Issue 2, pp. 98–104). https://doi.org/10.1056/nejm199107113250205

Farci, T., Shimoda, A., Wong, D., Cabezon, T., De Gioannis, D., Strazzera, A., Shimizu, Y., Shapiro, M., Alter, H. J., & Purcell, R. H. (1996). Prevention of hepatitis C virus infection in chimpanzees by hyperimmune serum against the hypervariable region 1 of the envelope 2 protein. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 93, Issue 26, pp. 15394–15399). https://doi.org/10.1073/pnas.93.26.15394

Fattovich, G., Giustina, G., Degos, F., Tremolada, F., Diodati, G., Almasio, P., Nevens, F., Solinas, A., Mura, D., Brouwer, J. T., Thomas, H., Njapoum, C., Casarin, C., Bonetti, P., Fuschi, P., Basho, J., Tocco, A., Bhalla, A., Galassini, R., … Realdi, G. (1997). Morbidity and mortality in compensated cirrhosis type C: A retrospective follow-up study of 384 patients. In *Gastroenterology* (Vol. 112, Issue 2, pp. 463–472). https://doi.org/10.1053/gast.1997.v112.pm9024300

Feinstone, S. M., Kapikian, A. Z., Purcell, R. H., Alter, H. J., Holland, P. V., &

Zuckerman, J. (2001). Transfusion-associated hepatitis not due to viral hepatitis type A or B. In *Reviews in Medical Virology* (Vol. 11, Issue 1, pp. 3–9). https://doi.org/10.1002/rmv.304

Forton, D. M., Karayiannis, P., Mahmud, N., Taylor-Robinson, S. D., & Thomas, H. C. (2004). Identification of Unique Hepatitis C Virus Quasispecies in the Central Nervous System and Comparative Analysis of Internal Translational Efficiency of Brain, Liver, and Serum Variants. In *Journal of Virology* (Vol. 78, Issue 10, pp. 5170–5183). https://doi.org/10.1128/jvi.78.10.5170-5183.2004

Gao, L., Aizaki, H., He, J.-W., & Lai, M. M. C. (2004). Interactions between Viral Nonstructural Proteins and Host Protein hVAP-33 Mediate the Formation of Hepatitis C Virus RNA Replication Complex on Lipid Raft. In *Journal of Virology* (Vol. 78, Issue 7, pp. 3480–3488). https://doi.org/10.1128/jvi.78.7.3480-3488.2004

Goffin, E., Pirson, Y., Cornu, C., Jadoul, M., & van Ypersele de Strihou, C. (1994). Significance of NS3 and NS5 antigens in screening for HCV antibody. In *Lancet* (Vol. 343, Issue 8901, p. 854).

Grabowski, P. J., & Black, D. L. (2001). Alternative RNA splicing in the nervous system. In *Progress in Neurobiology* (Vol. 65, Issue 3, pp. 289–308). https://doi.org/10.1016/S0301-0082(01)00007-7

Hassan, M., Ghozlan, H., & Abdel-Kader, O. (2005). Activation of c-Jun NH2-terminal kinase (JNK) signaling pathway is essential for the stimulation of hepatitis C virus (HCV) non-structural protein 3 (NS3)-mediated cell growth. In *Virology* (Vol. 333, Issue 2, pp. 324–336). https://doi.org/10.1016/j.virol.2005.01.008

HCV Genome and Life Cycle - Hepatitis C Viruses - NCBI Bookshelf. (n.d.).

Hillyer, C. D. (2000). Clinical outcomes after hepatitis C infection from contaminated anti-

D immune globulin. In *Transfusion Medicine Reviews* (Vol. 14, Issue 1, pp. 85–86). https://doi.org/10.1016/s0887-7963(00)80122-4

Honda, M., Ping, L. H., Rijnbrand, R. C. A., Amphlett, E., Clarke, B., Rowlands, D., & Lemon, S. M. (1996). Structural requirements for initiation of translation by internal ribosome entry within genome-length hepatitis C virus RNA. In *Virology* (Vol. 222, Issue 1, pp. 31–42). https://doi.org/10.1006/viro.1996.0395

Hu, K. Q., & Tong, M. J. (1999). The long-term outcomes of patients with compensated hepatitis C virus- related cirrhosis and history of parenteral exposure in the United States. In *Hepatology* (Vol. 29, Issue 4, pp. 1311–1316). https://doi.org/10.1002/hep.510290424

Hussain, Z. (2013). Genomic Heterogeneity of Hepatitis Viruses (A-E): Role in Clinical Implications and Treatment. In *Practical Management of Chronic Viral Hepatitis*. https://doi.org/10.5772/55231

Kadoya, H., Nagano-Fujii, M., Deng, L., Nakazono, N., & Hotta, H. (2005). Nonstructural proteins 4A and 4B of hepatitis C virus transactivate the interleukin 8 promoter. In *Microbiology and Immunology* (Vol. 49, Issue 3, pp. 265–273). https://doi.org/10.1111/j.1348-0421.2005.tb03728.x

Kataoka, N., Yong, J., Kim, V. N., Velazquez, F., Perkinson, R. A., Wang, F., & Dreyfuss, G. (2000). Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. In *Molecular Cell* (Vol. 6, Issue 3, pp. 673–682). https://doi.org/10.1016/S1097-2765(00)00065-4

Kolykhalov, A. A., Feinstone, S. M., & Rice, C. M. (1996). Identification of a highly conserved sequence element at the 3' terminus of hepatitis C virus genome RNA. In *Journal of virology* (Vol. 70, Issue 6, pp. 3363–3371).

https://doi.org/10.1128/jvi.70.6.3363-3371.1996

Lerat, H., Shimizu, Y. K., & Lemon, S. M. (2000). Cell Type-Specific Enhancement of
Hepatitis C Virus Internal Ribosome Entry Site-Directed Translation due to 5′
Nontranslated Region Substitutions Selected during Passage of Virus in
Lymphoblastoid Cells. In *Journal of Virology* (Vol. 74, Issue 15, pp. 7024–7031).
https://doi.org/10.1128/jvi.74.15.7024-7031.2000

Lesens, O., Deschênes, M., Steben, M., Bélanger, G., & Tsoukas, C. M. (1999). Hepatitis C
virus is related to progressive liver disease in human immunodeficiency virus-positive
hemophiliacs and should be treated as an opportunistic infection. In *Journal of
Infectious Diseases* (Vol. 179, Issue 5, pp. 1254–1258).
https://doi.org/10.1086/314720

Levin, M. K., Gurjar, M., & Patel, S. S. (2005). A Brownian motor mechanism of
translocation and strand separation by hepatitis C virus helicase. In *Nature Structural
and Molecular Biology* (Vol. 12, Issue 5, pp. 429–435).
https://doi.org/10.1038/nsmb920

Lin, W., Kim, S. S., Yeung, E., Kamegaya, Y., Blackard, J. T., Kim, K. A., Holtzman, M.
J., & Chung, R. T. (2006). Hepatitis C Virus Core Protein Blocks Interferon Signaling
by Interaction with the STAT1 SH2 Domain. In *Journal of Virology* (Vol. 80, Issue
18, pp. 9226–9235). https://doi.org/10.1128/jvi.00459-06

Lindenbach, B. D., & Rice, C. M. (2005). Unravelling hepatitis C virus replication from
genome to function. In *Nature* (Vol. 436, Issue 7053, pp. 933–938).
https://doi.org/10.1038/nature04077

Lunel, F. (1994). Hepatitis C virus and autoimmunity: Fortuitous association or reality? In
Gastroenterology (Vol. 107, Issue 5, pp. 1550–1555). https://doi.org/10.1016/0016-

5085(94)90564-9

Manns, M. P., McHutchison, J. G., Gordon, S. C., Rustgi, V. K., Shiffman, M., Reindollar, R., Goodman, Z. D., Koury, K., Ling, M. H., & Albrecht, J. K. (2001). Peginterferon alfa-2b plus ribavirin compared with interferonalfa-2b plus ribavirin for initial treatment of chronic hepatitis C: A randomised trial. In *Lancet* (Vol. 358, Issue 9286, pp. 958–965). https://doi.org/10.1016/S0140-6736(01)06102-5

McCaughan, G. W., McGuinness, P. H., Bishop, G. A., Painter, D. M., Lien, A. S. M., Tulloch, R., Wylie, B. R., & Archer, G. T. (1992). Clinical assessment and incidence of hepatitis C RNA in 50 consecutive RIBA-positive volunteer blood donors. In *Medical Journal of Australia* (Vol. 157, Issue 4, pp. 231–233). https://doi.org/10.5694/j.1326-5377.1992.tb137124.x

McHutchison, J. G., & Poynard, T. (1999). Combination therapy with interferon plus ribavirin for the initial treatment of chronic hepatitis C. In *Seminars in Liver Disease* (Vol. 19, Issue 1 SUPPL. 1, pp. 57–66).

Miller, R. H., & Purcell, R. H. (1990). Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 87, Issue 6, pp. 2057–2061). https://doi.org/10.1073/pnas.87.6.2057

Modi, A. A., & Liang, T. J. (2008). Hepatitis C: A clinical review. In *Oral Diseases* (Vol. 14, Issue 1, pp. 10–14). https://doi.org/10.1111/j.1601-0825.2007.01419.x

Moradpour, D., Brass, V., & Penin, F. (2005). Function follows form: The structure of the N-terminal domain of HCV NS5A. In *Hepatology* (Vol. 42, Issue 3, pp. 732–735). https://doi.org/10.1002/hep.20851

Morillas, R. M., Montoliu, S., & Planas, R. (2005). Hepatitis C. In *FMC Formacion*

Medica Continuada en Atencion Primaria (Vol. 12, Issue 2, pp. 74–85). https://doi.org/10.1016/S1134-2072(05)71167-1

Moriya, K., Yotsuyanagi, H., Shintani, Y., Fujie, H., Ishibashi, K., Matsuura, Y., Miyamura, T., & Koike, K. (1997). Hepatitis C virus core protein induces hepatic steatosis in transgenic mice. In *Journal of General Virology* (Vol. 78, Issue 7, pp. 1527–1531). https://doi.org/10.1099/0022-1317-78-7-1527

Nguyen, M. H., Whittemore, A. S., Garcia, R. T., Tawfeek, S. A., Ning, J., Lam, S., Wright, T. L., & Keeffe, E. B. (2004). Role of ethnicity in risk for hepatocellular carcinoma in patients with chronic hepatitis C and cirrhosis. In *Clinical Gastroenterology and Hepatology* (Vol. 2, Issue 9, pp. 820–824). https://doi.org/10.1016/S1542-3565(04)00353-2

O., P., & S., H. (2016). Next big threat for pakistan hepatocellular carcinoma (HCC). *Journal of the Pakistan Medical Association*, *66*(6), 735–739.

Papatheodoridis, G. V., Papadimitropoulos, V. C., & Hadziyannis, S. J. (2001). Effect of interferon therapy on the development of hepatocellular carcinoma in patients with hepatitis C virus-related cirrhosis: A meta-analysis. In *Alimentary Pharmacology and Therapeutics* (Vol. 15, Issue 5, pp. 689–698). https://doi.org/10.1046/j.1365-2036.2001.00979.x

Pesole, G., Mignone, F., Gissi, C., Grillo, G., Licciulli, F., & Liuni, S. (2001). Structural and functional features of eukaryotic mRNA untranslated regions. In *Gene* (Vol. 276, Issues 1–2, pp. 73–81). https://doi.org/10.1016/S0378-1119(01)00674-6

Polyak, S. J., Khabar, K. S. A., Paschal, D. M., Ezelle, H. J., Duverlie, G., Barber, G. N., Levy, D. E., Mukaida, N., & Gretch, D. R. (2001). Hepatitis C Virus Nonstructural 5A Protein Induces Interleukin-8, Leading to Partial Inhibition of the Interferon-Induced

Antiviral Response. In *Journal of Virology* (Vol. 75, Issue 13, pp. 6095–6106). https://doi.org/10.1128/jvi.75.13.6095-6106.2001

Ragni, M. V., & Belle, S. H. (2001). Impact of human immunodeficiency virus infection on progression to end-stage liver disease in individuals with hemophilia and hepatitis C virus infection. In *Journal of Infectious Diseases* (Vol. 183, Issue 7, pp. 1112–1115). https://doi.org/10.1086/319273

Rosen, H. R., & Martin, P. (2000). Viral hepatitis in the liver transplant recipient. In *Infectious Disease Clinics of North America* (Vol. 14, Issue 3, pp. 761–784). https://doi.org/10.1016/S0891-5520(05)70130-6

S C Gumber, S. C. (n.d.). *Hepatitis C_ a multifaceted disease*.

Sakai, A., St. Claire, M., Faulk, K., Govindarajan, S., Emerson, S. U., Purcell, R. H., & Bukh, J. (2003). The p7 polypeptide of hepatitis C virus is critical for infectivity and contains functionally important genotype-specific sequences. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 100, Issue 20, pp. 11646–11651). https://doi.org/10.1073/pnas.1834545100

Salazar, A. E. (1996). Management of hepatitis C. In *Journal of the American Board of Family Practice* (Vol. 9, Issue 5, p. 391).

Sasaki, N., Matsui, A., Momoi, M., Tsuda, F., & Okamoto, H. (1997). Loss of circulating hepatitis C virus in children who developed a persistent carrier state after mother-to-baby transmission. In *Pediatric Research* (Vol. 42, Issue 3, pp. 263–267). https://doi.org/10.1203/00006450-199709000-00003

Schwer, B., Ren, S., Pietschmann, T., Kartenbeck, J., Kaehlcke, K., Bartenschlager, R., Yen, T. S. B., & Ott, M. (2004). Targeting of Hepatitis C Virus Core Protein to Mitochondria through a Novel C-Terminal Localization Motif. In *Journal of Virology*

(Vol. 78, Issue 15, pp. 7958–7968). https://doi.org/10.1128/jvi.78.15.7958-7968.2004

Seeff, L. B., Hollinger, F. B., Alter, H. J., Wright, E. C., Cain, C. M. B., Buskell, Z. J., Ishak, K. G., Iber, F. L., Toro, D., Samanta, A., Koretz, R. L., Perrillo, R. P., Kalloo, D., Desor, V., Schellhase, L., Orebaugh, C., Barton, L., Orrington, M., DeMedina, M., … Robinette, D. (2001). Long-term mortality and morbidity of transfusion-associated non-A, non-B, and type C hepatitis: A National Heart, Lung, and Blood Institute Collaborative Study. In *Hepatology* (Vol. 33, Issue 2, pp. 455–463). https://doi.org/10.1053/jhep.2001.21905

Seeff, Leonard B., & Hoofnagle, J. H. (2002). National Institutes of Health Consensus Development Conference: Management of Hepatitis C: 2002. In *Hepatology* (Vol. 36, Issue 5 I, pp. s1–s2). https://doi.org/10.1053/jhep.2002.36992

Shi, S. T., Lee, K.-J., Aizaki, H., Hwang, S. B., & Lai, M. M. C. (2003). Hepatitis C Virus RNA Replication Occurs on a Detergent-Resistant Membrane That Cofractionates with Caveolin-2. In *Journal of Virology* (Vol. 77, Issue 7, pp. 4160–4168). https://doi.org/10.1128/jvi.77.7.4160-4168.2003

Shih, C. M., Lo, S. J., Miyamura, T., Chen, S. Y., & Lee, Y. H. (1993). Suppression of hepatitis B virus expression and replication by hepatitis C virus core protein in HuH-7 cells. In *Journal of Virology* (Vol. 67, Issue 10, pp. 5823–5832). https://doi.org/10.1128/jvi.67.10.5823-5832.1993

Smith, D. B., Bukh, J., Kuiken, C., Muerhoff, A. S., Rice, C. M., Stapleton, J. T., & Simmonds, P. (2014). Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: Updated criteria and genotype assignment web resource. In *Hepatology* (Vol. 59, Issue 1, pp. 318–327). https://doi.org/10.1002/hep.26744

Spearman, C. W., Dusheiko, G. M., Hellard, M., & Sonderup, M. (2019). Hepatitis C. In

The Lancet. https://doi.org/10.1016/S0140-6736(19)32320-7

Sumpter, R., Loo, Y.-M., Foy, E., Li, K., Yoneyama, M., Fujita, T., Lemon, S. M., & Gale, M. (2005). Regulating Intracellular Antiviral Defense and Permissiveness to Hepatitis C Virus RNA Replication through a Cellular RNA Helicase, RIG-I. In *Journal of Virology* (Vol. 79, Issue 5, pp. 2689–2699). https://doi.org/10.1128/jvi.79.5.2689-2699.2005

Sweeney, R., Fan, Q., & Yao, M. C. (1996). Antisense ribosomes: rRNA as a vehicle for antisense RNAs. In *Proceedings of the National Academy of Sciences of the United States of America* (Vol. 93, Issue 16, pp. 8518–8523). https://doi.org/10.1073/pnas.93.16.8518

Taylor, D. R., Shi, S. T., Romano, P. R., Barber, G. N., & Lai, M. M. C. (1999). Inhibition of the interferon-inducible protein kinase PKR by HCV E2 protein. *Science*, *285*(5424), 107–110. https://doi.org/10.1126/science.285.5424.107

Tellinghuisen, T. L., Marcotrigiano, J., Gorbalenya, A. E., & Rice, C. M. (2004). The NS5A protein of hepatitis C virus is a zinc metalloprotein. In *Journal of Biological Chemistry* (Vol. 279, Issue 47, pp. 48576–48587). https://doi.org/10.1074/jbc.M407787200

Tellinghuisen, T. L., & Rice, C. M. (2002). Interaction between hepatitis C virus proteins and host cell factors. In *Current Opinion in Microbiology* (Vol. 5, Issue 4, pp. 419–427). https://doi.org/10.1016/S1369-5274(02)00341-7

Thurner, C., Witwer, C., Hofacker, I. L., & Stadler, P. F. (2004). Conserved RNA secondary structures in Flaviviridae genomes. In *Journal of General Virology* (Vol. 85, Issue 5, pp. 1113–1124). https://doi.org/10.1099/vir.0.19462-0

Timchenko, L. T. (1999). Myotonic Dystrophy: The Role of RNA CUG Triplet Repeats. In

The American Journal of Human Genetics (Vol. 64, Issue 2, pp. 360–364). https://doi.org/10.1086/302268

Tycowski, K. T., Guo, Y. E., Lee, N., Moss, W. N., Vallery, T. K., Xie, M., & Steitz, J. A. (2015). Viral noncoding RNAs: More surprises. In *Genes and Development* (Vol. 29, Issue 6, pp. 567–584). https://doi.org/10.1101/gad.259077.115

Umar, M., Hamama-tul-Bushra, Ahmad, M., Khurram, M., Usman, S., Arif, M., Adam, T., Minhas, Z., Arif, A., Naeem, A., Ejaz, K., Butt, Z., & Bilal, M. (2010). Hepatitis C in Pakistan: A review of available data. In *Hepatitis Monthly* (Vol. 10, Issue 3, pp. 205–214).

Villano, S. A., Vlahov, D., Nelson, K. E., Cohn, S., & Thomas, D. L. (1999). Persistence of viremia and the importance of long-term follow-up after acute hepatitis C infection. In *Hepatology* (Vol. 29, Issue 3, pp. 908–914). https://doi.org/10.1002/hep.510290311

Vince, A., Židovec Lepej, S., Kukelac, I., Čajić, V., Burek, V., Dušek, D., & Budimir, J. (2009). Diagnosis and treatment of hepatitis C. In *Infektološki Glasnik* (Vol. 29, Issue 2, pp. 49–56).

Vogt, M. (2000). Prevalence and clinical outcome of hepatitis C infection in children who underwent cardiac surgery before the implementation of blood-donor screening. In *Clinical pediatrics* (Vol. 39, Issue 4, p. 249). https://doi.org/10.1177/000992280003900411

Voisset, C., Callens, N., Blanchard, E., Op De Beeck, A., Dubuisson, J., & Vu-Dac, N. (2005). High density lipoproteins facilitate hepatitis C virus entry through the scavenger receptor class B type I. In *Journal of Biological Chemistry* (Vol. 280, Issue 9, pp. 7793–7799). https://doi.org/10.1074/jbc.M411600200

Walewski, J. L., Keller, T. R., Stump, D. D., & Branch, A. D. (2001). Evidence for a new

hepatitis C virus antigen encoded in an overlapping reading frame. In *Rna* (Vol. 7, Issue 5, pp. 710–721). https://doi.org/10.1017/S1355838201010111

WHO. (1999). Global surveillance and control of hepatitis C. In *Journal of Viral Hepatitis* (Vol. 6, pp. 35–47).

Wiley, T. E., Mccarthy, M., Breidi, L., Mccarthy, M., & Layden, T. J. (1998). Impact of alcohol on the histological and clinical progression of hepatitis C infection. In *Hepatology* (Vol. 28, Issue 3, pp. 805–809). https://doi.org/10.1002/hep.510280330

Xu, N., Chen, C.-Y. A., & Shyu, A.-B. (2001). Versatile Role for hnRNP D Isoforms in the Differential Regulation of Cytoplasmic mRNA Turnover. In *Molecular and Cellular Biology* (Vol. 21, Issue 20, pp. 6960–6971). https://doi.org/10.1128/mcb.21.20.6960-6971.2001

Yano, M., Kumada, H., Kage, M., Ikeda, K., Shimamatsu, K., Inoue, O., Hashimoto, E., Lefkowitch, J. H., Ludwig, J., & Okuda, K. (1996). The long-term pathological evolution of chronic hepatitis C. In *Hepatology* (Vol. 23, Issue 6, pp. 1334–1340). https://doi.org/10.1053/jhep.1996.v23.pm0008675148

Yasui, K., Wakita, T., Tsukiyama-Kohara, K., Funahashi, S.-I., Ichikawa, M., Kajita, T., Moradpour, D., Wands, J. R., & Kohara, M. (1998). The Native Form and Maturation Process of Hepatitis C Virus Core Protein. In *Journal of Virology* (Vol. 72, Issue 7, pp. 6048–6055). https://doi.org/10.1128/jvi.72.7.6048-6055.1998

Yi, M., & Lemon, S. M. (2003). Structure-function analysis of the 3′ stem-loop of hepatitis C virus genomic RNA and its role in viral RNA replication. In *Rna* (Vol. 9, Issue 3, pp. 331–345). https://doi.org/10.1261/rna.2144203

Chapter 8

Appendices

Appendix

Protein name	Protein Id
Core protein	AQU12741.1
Glycoprotein E1	EU399720.1
Glycoprotein E2	ABY83294.1
Viroporin p7	CAM24200.1
RdRp	ADM46203.1
NS4A	AEC48776.1
NS4B	ACQ45270.1

Table 1: Protein ids of HCV genotype 3a of Pakistan

turnitin

Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author:	Haleema Sadia
Assignment title:	Dr. Sobia
Submission title:	Functional and phylogenetic analysis of 5' and 3' untranslate...
File name:	Haleema_thesis_file__Final_and_final_19_10_2021.docx
File size:	5.3M
Page count:	83
Word count:	10,294
Character count:	54,970
Submission date:	22-Oct-2021 11:08AM (UTC+0500)
Submission ID:	1680837060

Functional and phylogenetic analysis of 5' and 3' untranslated
regions (UTRs) of Hepatitis C virus genome (HCV) for the
potential target prediction of its self- regulation

By

Haleema Sadia
NUST00000277012

Department of Healthcare Biotechnology
Atta-ur-Rahman School of Applied Biosciences (ASAB)
National University of Sciences and Technology (NUST)
Islamabad, Pakistan
2021

9.11.2021

Dr Sobia Manzoor, PhD
Tenured Associate Professor
Head of Department (HoD)
Deptt of Healthcare Biotechnology
Atta-ur-Rahman School of Applied
Biosciences (ASAB), NUST Islamabad