MOLECULAR CLONING and CHARACTERIZATION OF THE COMMON 1b SUBTYPE OF HCV FROM TURKEY

A THESIS SUBMITTED TO
THE DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS
AND
THE INSTITUTE OF ENGINEERING AND SCIENCE OF
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF MASTER OF SCIENCE

By
AGLI ÖZTAN

JULY, 1999

MOLECULAR CLONING and CHARACTERIZATION OF THE COMMON 1b SUBTYPE OF HCV FROM TURKEY

A THESIS SUBMITTED TO

THE DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS

AND

THE INSTITUTE OF ENGINEERING AND SCIENCE

OF BILKENT UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF MASTER OF SCIENCE

By

ASLI ÖZTAN

JULY, 1999

I certify that I read this thesis and that in my opinion it is fully adequate, in scope and in quality, as thesis for the degree of Master of Science.

Prof. Dr. Mehmet Öztürk

I certify that I read this thesis and that in my opinion it is fully adequate, in scope and in quality, as thesis for the degree of Master of Science.

Prof. Dr. Hikmet Akkız

I certify that I read this thesis and that in my opinion it is fully adequate, in scope and in quality, as thesis for the degree of Master of Science.

Assist. Prof. Işık Yulug

Approved for Institute of Engineering and Science.

Prof.Dr. Mehmet Baray,
Director of Institute of Engineering and Science

ABSTRACT

MOLECULAR CLONING and CHARACTERIZATION OF THE COMMON 1b SUBTYPE OF HCV FROM TURKEY

Aslı Öztan

MSc. in Molecular Biology and Genetics

Supervisor: Prof. Mehmet Öztürk

July 1999

Hepatitis C Virus is a major cause of acute and chronic hepatitis worldwide. 80-90% of Hepatitis C Virus infections become chronic and 75% of these cases lead to liver disease, including cirrhosis, liver failure and hepatocellular carcinoma. Hepatitis C Virus was first identified by molecular cloning of the viral genome in 1989. Hepatitis C Virus is an enveloped virus containing a positive stranded RNA genome with a size of around 9.5 kilobases. In terms of genomic organization, it was accepted as a member of Flaviviridae family as a new genus named Hepaciviruses. The single-stranded RNA genome encodes a single open reading frame, which is transcribed into a single polypeptide of 3010 or 3030 amino acids and cleaved into viral proteins Core, E1, E2/p7, NS2, NS3, NS4A, NS4B, NS5A, NS5B by host and viral proteases. Sequencing, serotyping and RFLP studies indicate that Hepatitis C Virus genome is highly variable. There are six distinct genotypes and at least 74 subtypes with different distributions between the geographic areas. Variability is not equally distributed throughout the genome. 5' UTR, some parts of the 3' UTR and capsid protein are the most conserved regions. The predominant genotype in the Turkish population was found to be 1b by sequencing of the 5'-UTR. In this study, the entire sequence encompassing the complete coding region and partial non coding regions of the genotype 1b obtained from a HCV-infected Turkish patient was cloned to investigate its evolutionary relationship with other genotypes and to study its overall genome organization,. In order to characterize the viral genome, viral RNA was extracted from the serum, cDNA was synthesized, the HCV genome was amplified by PCR in 7 overlapping fragments, PCR fragments were cloned into bacterial vectors and cloned inserts were sequenced by automated sequencing methodology. The partial sequence data covering 70% of the cloned HCV genome indicate that the Turkish 1b genotype displays high homology to other 1b genotypes, but differs from others by distinct amino acid changes. To our knowledge, this is the first report about the HCV genome structure from Turkey. The HCV subgenomic fragments obtained here will serve to further molecular and immunologic studies on this dominant form of HCV found in Turkish patients.

Key Words: Hepatitis C Virus; 1b; genome; cloning; sequencing; Turkish Isolate.

ÖZET

TÜRKİYE DE BASKIN OLARAK GÖRÜNEN HEPATİT C VİRÜSÜ 1b ALT TİPİNİN MOLEKÜLER KLONLANMASI VE KARAKTERİZASYONU

Aslı Öztan

Moleküler Biyoloji ve Genetik Yüksek Lisans

Tez Yöneticisi: Prof. Mehmet Öztük

Temmuz 1999

Hepatit C Virüsü dünyada akut ve kronik hepatitin en önemli nedenidir. Hepatit C Virüsü enfeksiyonlarinin yüzde 80 ile yüzde 90 gibi bir oranı kronikleşir ve buna bağlı olarak siroz, karaciğer yetmezliği ve karaciğer kanseri gibi hastalıklara sebep olur. Hepatit C Virüsü ilk defa 1989 yılında örtüşen bir çok kopya DNA'nın klonlanması ile bulunmuştur. Bu zarflı, pozitif tek iplikli RNA virüsü 9.5 kilobazlık bir genoma sahiptir. Genom organizasyonuna göre HCV Flaviviridae virus ailesinin altında Hepasivirusler diye adlandırılan yeni bir generada sınıflandırılmıştır. Virüsün tek iplikli RNA genomu 3010 ile 3030 amino asitten oluşan tek bir polypeptid sentezleyen açık okuma çerçevesi içerir ve viral ya da hücre proteazları tarafından Core, E1, E2/p7, NS2, NS3, NS4A, NS4B, NS5A ve NS5B virüs proteinlerine parçalanır. Dizi analizi, serotipleme ve RFLP metodları ile virüs genomunun çok sayıda değişime uğradığı anlaşılmıştır. Şimdiye kadar farklı coğrafi alanlara yayılmış 6 değişik genotip ile 74 civarında altip bulunmuştur. Virüs genomunun dizisinde gözlemlenen hızlı değişim genomun tüm bölgelerine eşit şekilde dağılmamıştır. 5'-UTR, 3'-UTR'ın bir kısmı ve kapsid proteini en çok korunan bölgelerdir. Türk popülasyonunda en çok rastlanan genotip 5' UTR dizi analizinin yapılması ile 1b olarak saptanmıştır. Bu çalışmada da ilk Türk HCV izolatının karakterize edilmesi ve virüsün evriminin incelenmesi amacı ile bir Türk hastadan izole edilen Hepatit C Virüsü genomu klonlanmış ve dizi analizi yapılmıştır. İzlenen yöntemler virüs RNA sının hastanın serumundan izole edilmesi, kopya DNA nın sentezlenmesi, genomun örtüşen 7 tane polimeraz zincir reaksiyonu ile çoğaltılması, bu fragmentların bakteri plasmidlerine klonlanması ve klonlanan bölgelerin otomatik dizi analizi yöntemi ile analiz edilmesidir. Klonlamış HCV genomunun %70 lik dizi verisine göre, Türk 1b genotipinin diğer 1b genotipleriyle homolojisi yüksek olmakla birlikte, bazi amino asit farklılıkları da belirlenmiştir. Elimizdeki verilere göre, bu çalışma Türkiye'den HCV genomu yapısıyla ilgili ilk rapordur. Burada elde edilen HCV alt genom parçaları ileride Türkiye'de baskın olan bu alt tür ile ilgili moleküler ve immunolojik çalışmalarda kullanılabilecektir.

Anahtar Kelimeler: Hepatit C Virüsü; 1b; genome; klonlama; sekanslama; Türkiye izolatı.

ACKNOWLEDGEMENTS

I would like to express my greatest gratitude to my thesis advisor Prof. Mehmet Öztürk, who had understood my passion in virology and gave me a chance to be a part of the virology project. It was a great opportunity for me to share his scientific vision during this study.

I would like to thank to my undergraduate advisor Prof. Aslıhan Tolun, for giving me the "first" chance and for her undoubtfull trust and encouragement throughout the years; Uğur Yavuzer, for all the discussions that enhanced my scientific way of thinking, for her trust in me more than I had in myself during the nightmares and above all for seeing rather than looking; Cengiz Yakıcıer for being a person as he is; Işık Yuluğ for sharing her scientific experiences, all the support during these two years and her endless tolerance to me; Rengül Çetin Atalay, who is not that much old at all, for all the chats in the lab and her kind personal interest.

"*Dün baştan başa onlarla açıklandı, yarın onlarla kurıldu baştanbaşa.* (A.Timuçin)"

Gülayşe, the blue color in my picture, things wouldn't be the same without your smile, encouragement and unquestionable belief. Thank you not only for all the sleepless nights, the difficult days and pains that you were with me but for all the successes and happiness that we could share with each other. **Bleda**, the whisperer of my conscience, "*Birbirinden en ayrı gemilerde bile bizim için yol aynı olurdu., nehirden yukarı- çünkü bizi ayni kaynak bekliyor* (R.M.Rilke)", for being a better thorn in the game than me. **Tolga Emre**, my favorite puzzle, you are the magenta in the picture, sometimes breaking things down, but surely one of the irreplaceable .

Hilal, the mother goddess Kybele, surviving against the most difficult experience of her legend throughout the centuries. "*Onlar savaşçılardır, acıyı ve sevinci direncin dupduru suyuyla yoğururlar.* (A.Timuçin)". Thank you for each **A**, **T**, **G** and **C** in this thesis and every second, minute and hour you had spend for me both professional and personal.

My special thanks go to Esra, the one and only virology group member, for sharing your deep knowledge on HCV and for being the guide of me in the virology world; Tuba for all the units of the enzymes she shared with me and her patience on answering my endless questions; Ayça, the ultimate support and ocean blue on my palette, for sharing the great joy of being on "çimenler" or "manzara" and all the great memories in Boğaziçi; Suha, the one whom I had sheltered not to scream all the way out; Tunç, the soul mate in a parallel universe, for being a part of my unconscious land and taking Aruoba and the dark side of the moon with you; Ersan, the "kırk yılda bir gibisin", for being the safest ground.

Finally, I would like to thank to my family Akın and Güler Öztan, for bringing me up as I am. They gave me the chance to explore my own limits and never loose their faith in me to date.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVATIONS

A	Adenine
Ala	(A) Alanine
Arg	(R) Arginine
Asn	(N) Asparagine
Asp	(D) Aspartic Acid
Asx	(B) Asparagine
bp	base pairs
C	Cytosine
cDNA	Complementary Deoxyribonucleic Acid
C-Terminus	Carboxyl-terminus
Cys	(C) Cysteine
ddH_2O	deionized distilled water
DNA	Deoxyribonucleic Acid
dNTP	deoxynucleotide triphosphate
ds	double strand
EDTA	diaminoethane tetra-acetic acid
EtBr	Ethidium Bromide
G	Guanine
Gln	(Q) Glutamine
Glu	(E) Glutamic Acid
Glx	(Z) Glutamic Acid
Gly	(G) Glycine
HBV	Hepatitis B Virus
HCC	Hepatocellular Carcinoma
HCV	Hepatitis C Virus
His	(H) Histidine
HIV	Human Immunodeficiency Virus
HVR	Hyper Variable Region
IFN-α	Interferon Alfa
Ile	(I) Isoleucine
IPTG	Isopropylthio-b-D-galactoside
IRES	Internal Ribosome Entry Site
ISDR	Interferon Sensitivity Determining Region
kb	kilobase(s)
LB	Luria-Bertoni Media
Leu	(L) Leucine
Lys	(K) Lysine
Met	(M) Methionine
NANBH	Non-A Non-B Hepatitis
NS	Non-structural Protein
N-Terminus	Amino-Terminus

NTP	Nucleoside triphosphate
PCR	Polymerase Chain Reaction
Phe	(F) Phenylalanine
Pro	(P) Prolin
RNA	Ribonucleic Acid

1. INTRODUCTION

Hepatitis C Virus is a major cause of non-A, non-B hepatitis worldwide. In 80-90% of the cases, the HCV infections become chronic hepatitis which may lead to cirrhosis, liver failure and hepatocellular carcinoma (Caselmann *et al.*, 1996). HCV infection is typically mild in its early stages and rarely recognized until it has caused significant damage to the liver. The cycle of disease from infection to significant liver damage can take 20 years or more. Known risk factors of the HCV are mainly transfusions of blood and blood products, intravenous drug usage, parental or sexual transmission and needle-stick accidents (Okamoto *et al.*, 1991a). Since 1990, after the introduction of blood tests for detection of antibodies against HCV proteins, the posttransfusion incidence of HCV has decreased in the developed countries. The risk of transfusion related hepatitis is in the range of 1 in 100,000 units transfused in these countries. In developing countries lack of detection tests for blood donors, multiple usage of syringes, low socioeconomic levels and lack of public education leads to a drastic increase in new HCV cases. In Turkey mandatory detection of antibodies against HCV in the blood donors' sera started in 1997. The data based on tests made on blood donors in Turkey showed that the 1,5% of the population is anti-HCV positive (Figure 1.1; Thomas *et al.*, 1994). Persistent HCV infection is common in transplant recipients. It was reported that HCV is the cause of chronic hepatitis in approximately 10% of all transplant recipients. Antiviral therapy with interferon-alpha is effective in only a minority of transplant patients, and since allografts

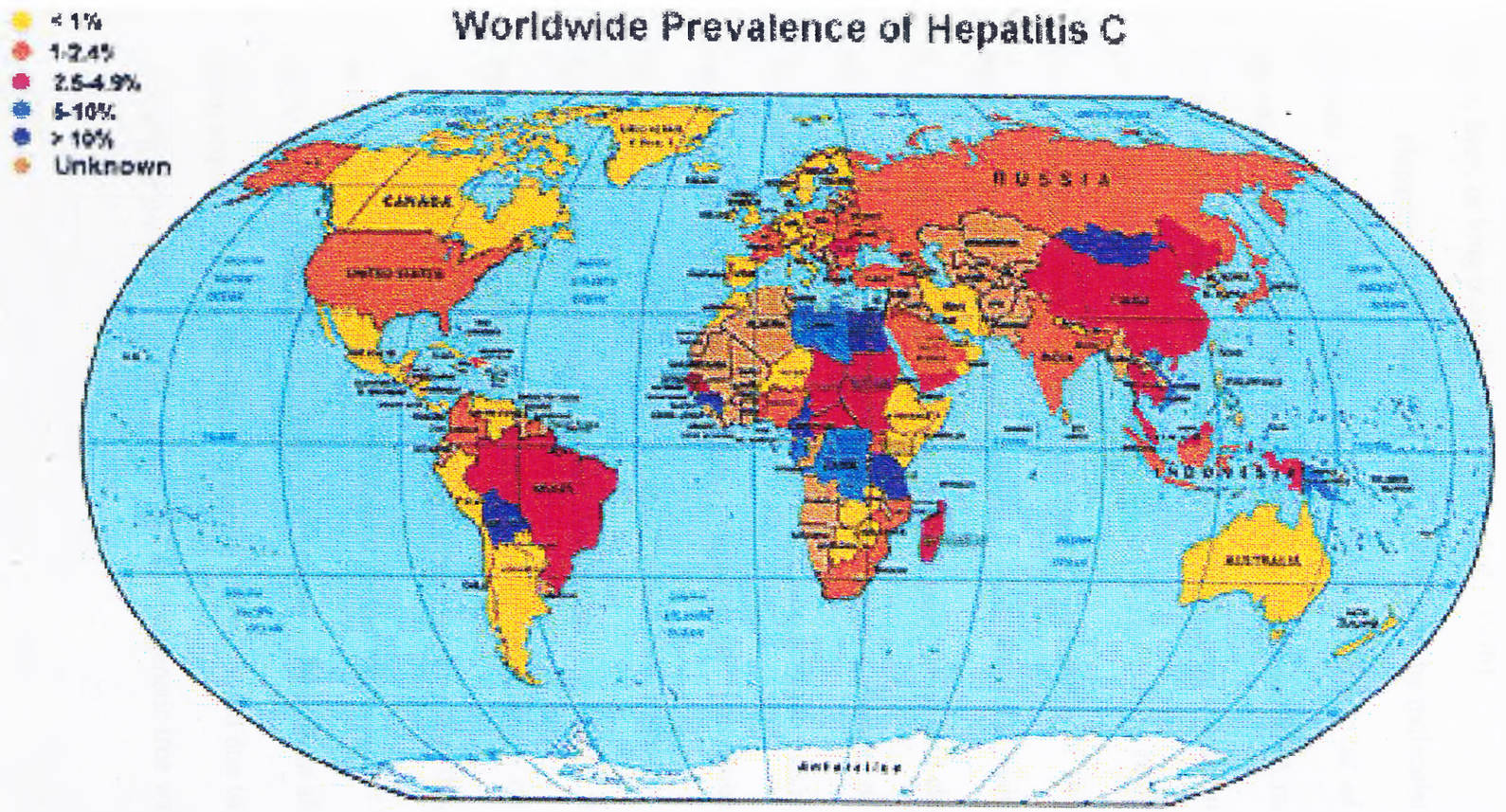


Figure1.1: Incidence of Hepatitis C Virus cases worldwide

from HCV infected donors are quite efficient in transmitting the virus, great attention is paid to the appropriate use of organs from HCV-positive donors. At present, these organs should be particularly targeted for patients in emergency need of lifesaving heart, liver, or lung transplants (Fishman *et al.*, 1996).

Hepatitis C virus was first identified in 1989 by molecular cloning of the viral genome in several overlapping cDNA fragments (Choo *et al.*, 1989). The virus genome is approximately 9.5 kb positive single stranded RNA molecule (Figure 1.2). RNA encodes a single polypeptide which is cleaved into structural and non-structural proteins by viral and host proteases (Figure 1.3). Genome replication takes place in the cytoplasm through an RNA intermediate. In terms of the presence of a lipid envelope, sensitivity to organic solvents, single stranded positive RNA genome, a long open reading frame, organization of the polyprotein and cleavage of polyprotein into several viral proteins, HCV was classified as a member of the Flaviviridae and has been placed in a new monotypic genus in this family, named hepaciviruses The nucleotide sequence is highly variable, the most divergent isolates sharing only 60% nucleotide sequence homology. Variability rate of the genome was estimated to be $1.44X10^{-3}$ base substitution/site/year. According to this data, theoretically all of the nucleotides may change during 700 years of infection in chimpanzees. However variability/stability of the genome varies by region within the genome. HVR1 and some regions of the 3'-UTR were found to be the most variable sites between various HCV strains (Simmonds P., 1998). Until today isolates from all over the world have been sequenced and a classification system was proposed due to the differences in the 5'-UTR of the different HCV groups and a phylogenetic tree was constructed where

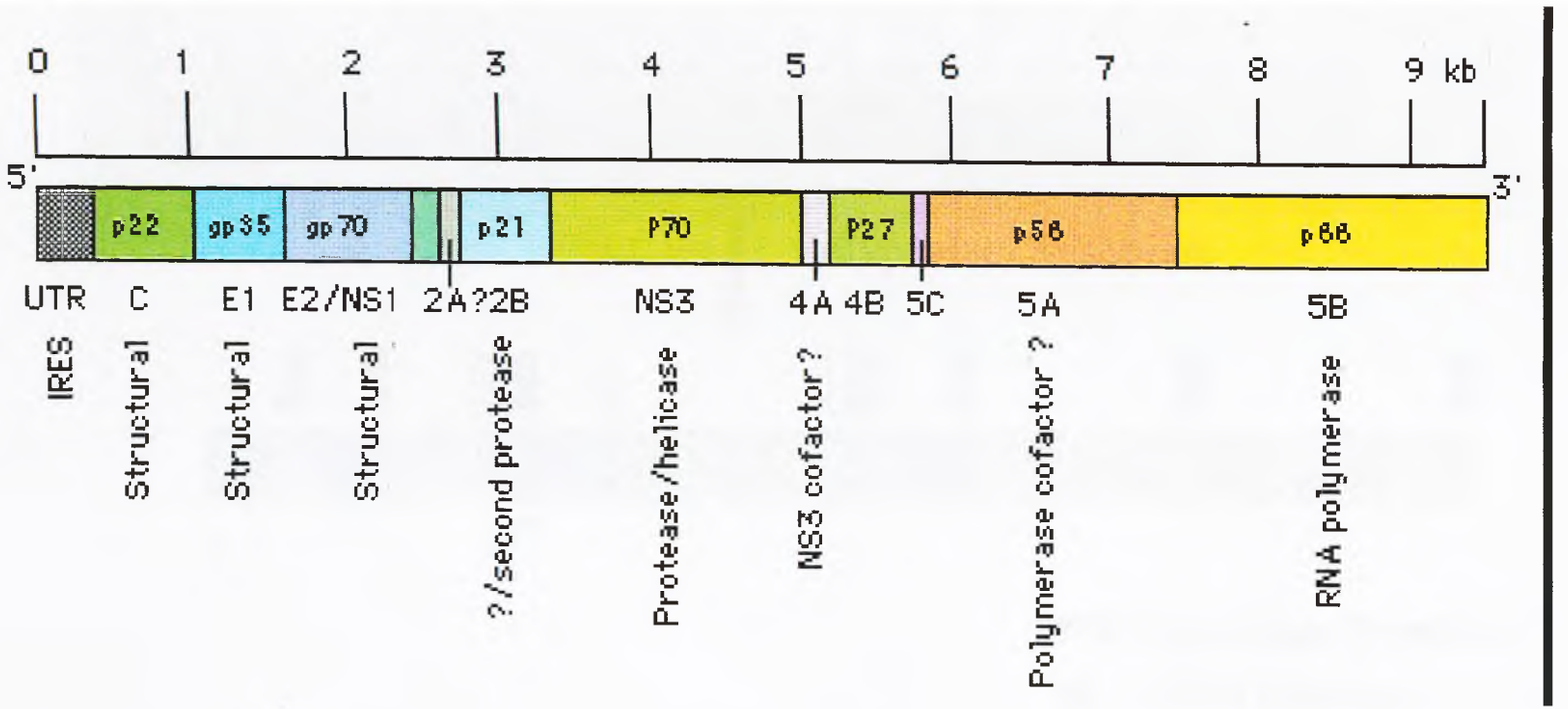


Figure 1.2: Schematic illustration of HCV genome and polyprotein.

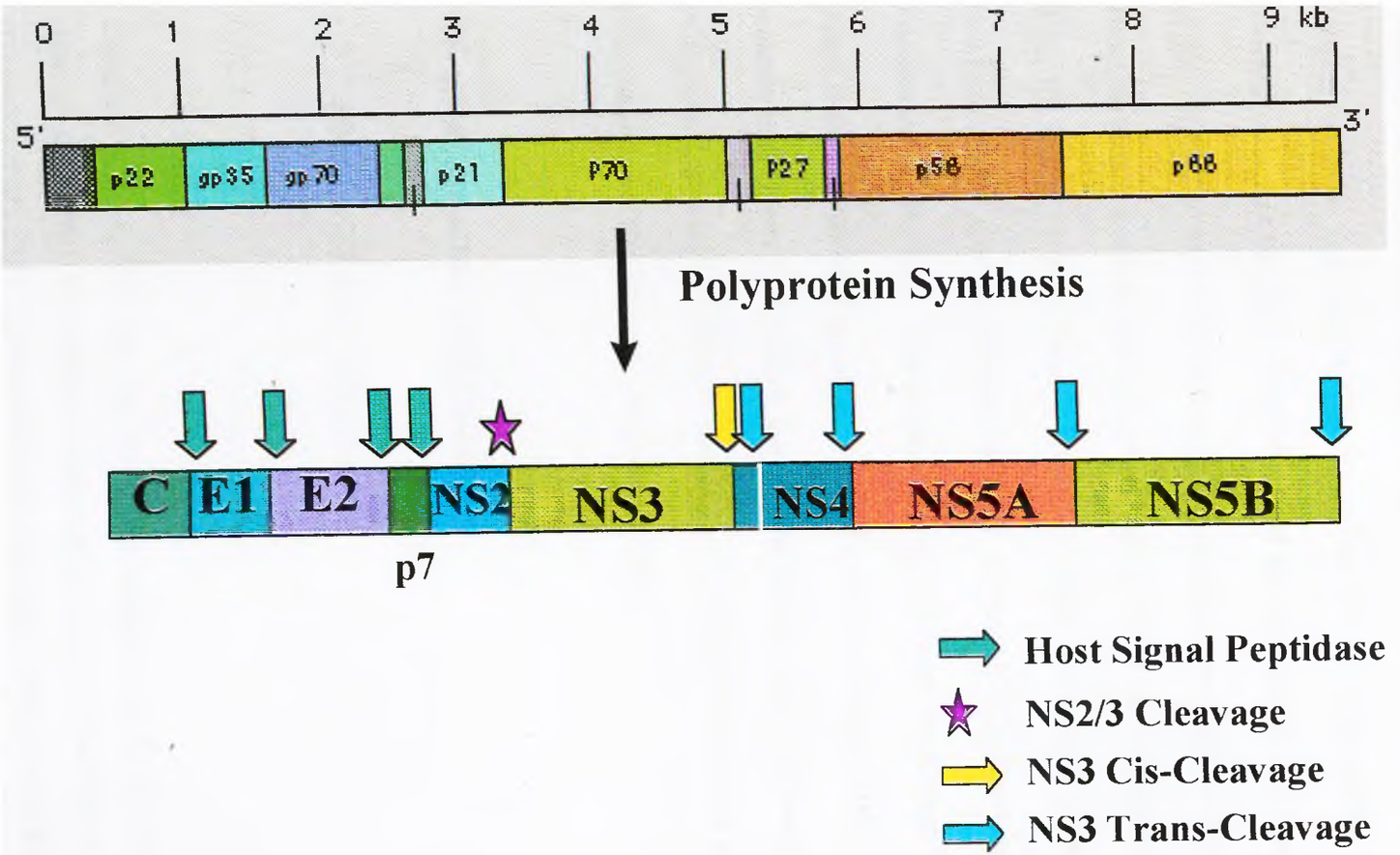


Figure 1.3: Schematic illustration of HCV polyprotein processing.

HCV isolates were grouped into 6 equally divergent genotypes and each into several subtypes. Main types are numbered 1 to 6 and subtypes have been lettered a, b and c in order of discovery (Simmonds *et al.*, 1993). Besides these clearly divergent isolates, there are different variants of the virus within the same individual reflecting the high variability rate of the viral genome. The quasispecies are not different enough to be grouped as a separate type or subtype but diverged enough to escape from the host's immune system. Some genotypes of HCV such as 1a, 1b, 2a and 2b have shown a broad worldwide distribution, whereas others have been seen in specific geographical areas. Genotype 1b seems to be the most common variant in Japan, 1a is commonly seen in Europe and United States, and type 4 is the predominant genotype in Middle East and Central Africa. Turkey is between the regions of Southern Europe and Middle East, but interestingly the predominant genotype of Turkish population was found to be 1b at a frequency higher than Japan (Yildiz *et al.*, unpublished data).

1.1 VIRAL GENOME ORGANIZATION

1.1.1 UNTRANSLATED REGIONS

1.1.1.1 5'-UTR

Complete 5' UTR spans the 341 nucleotide region at the 5' site of the genome (Lee *et al.*, 1992) but shorter sequences have been detected and reported (Okamoto *et al.*, 1990, Choo *et al.*, 1991, Kato *et al.*, 1990). 5'-UTR is the

most conserved region in the genome and used in genotyping studies because of the differences specific to the various types.

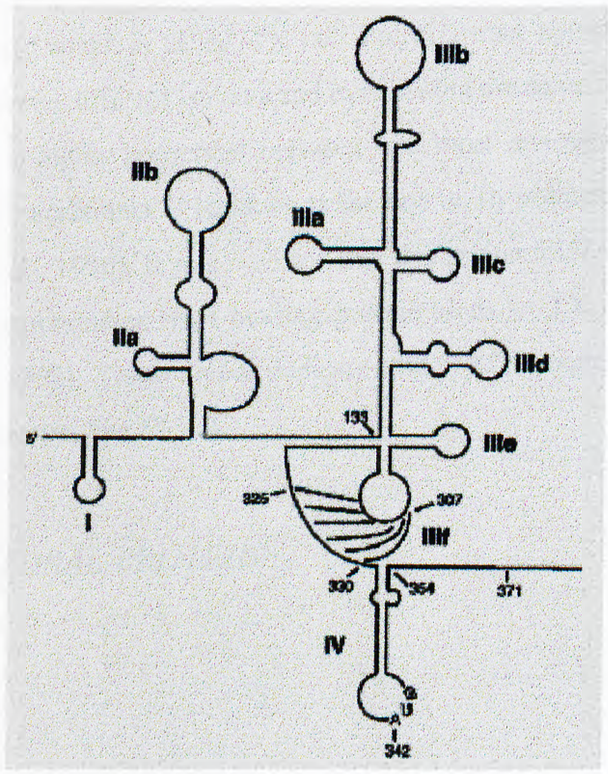


Figure 1.4: Model of secondary and tertiary structure of the HCV IRES proposed by Honda *et al.*, 1996.

A large conserved stem-loop structure (Figure 1.4) which acts as the Internal Ribosome Entrance Site (IRES; Wang *et al.*, 1993) is present at the proximal part. HCV viral genome translation is cap-independent and the IRES is involved in the regulation of replication and translation (Honda *et al.*, 1999). IRES provides a structure which is able to direct ribosomes to the AUG codon at position 342 for translation initiation. p57 polypyrimidine tract binding protein (PTB; Ali *et al.*, 1995) and 2 subunits of eIF3 were found to bind to the HCV IRES (Buratti *et al.*, 1998).

1.1.1.2 3'-UTR

3' end of the HCV genome has a tripartite structure with a conventional 3' end sequence, a poly(U) tract in some isolates and a highly conserved 98 nucleotide sequence which is called 3'X tail. Poly(U) tract appears to be highly heterogenous between different isolates and even within the same infected liver. In contrast 3'X tail is highly conserved between two most divergent genotypes 1a and 1b except two additional U residues at the end in 1b isolates (Tanaka *et al.*, 1996, Yamada *et al.*, 1996). It was reported that this region folds into a stem-loop structure and polypyrimidine tract binding protein binds to 3'X tail just like 5'-UTR (Ito *et al.*, 1998). These finding indicate that 3'-UTR region might also be important in replication like 5'UTR.

1.1.2 STRUCTURAL PROTEINS

1.1.2.1 CORE

The first structural protein is the HCV core protein (p22). It does not have any N-glycosylation sites. Amino acid sequence was found to be well conserved between the different HCV isolates (Bukh *et al.* 1994), which may reflect the importance of the function of this protein for the survival of the virus. Also because of this property core protein can be used for efficient detection of anti HCV antibodies (Doorn *et al* 1994).

The core protein is cleaved from the polyprotein by the cellular signal peptidase which recognize a specific sequence in the peptide chain. The protein then undergoes a complex process of proteolytic maturation and two main different products have been described: C173 (residues from 1 to 173) and C191 (residues from 1 to 191). Core protein products are localized in the cytoplasm but

when the expression of C173 takes place in the absence of C191, it is translocated to the nucleus (Santolini *et al* 1994).

HCV core protein is thought to be a multifunctional protein. It was shown that core protein binds to cellular membranes, RNA molecules and 60S ribosomal subunit (Santolini *et al.*, 1994). There is also a domain for binding to HCV E1 protein (Lo *et al.*, 1996). There are some data about the multimerization and dimerization ability of the core protein (Santolini *et al.* 1994). It can be detected in viral particles in patients serum by immunoelectron microscopy (Kaito *et al.*, 1994) and related to the capsid protein C of flaviviruses (Santolini *et al* 1994). All these data indicate that the core protein is the nucleocapsid protein of the virus and it is involved in the viral assembly. Besides having a role in viral encapsulation, it was shown to have a trans-acting regulatory role. It possesses a nuclear localization signal and as described previously, some forms of core protein were detected in the nuclei. HCV core protein is involved in the transcriptional repression of p53 (Ray *et al.* 1997) and repression $p21^{WAF1/Cip1/Sid1}$ promoter activities (Ray *et al* 1998), and also trans activates c-myc oncogene (Ray *et al.* 1996). Yeast two hybrid library screening studies have identified an interaction with the cytoplasmic tail of Lymphotoxin-β receptor (Matsumoto *et al.* 1997, Chen *et al* 1997) which is involved in host defense mechanism. Lymphotoxin-β receptors are the members of tumor necrosis factor receptors (TNFR). After the identification of Lymphotoxin-β receptor, it was shown that core protein also binds to the cytoplasmic domain of TNFR 1 which is another member of the TNFR family. This binding enhances TNF-induced apoptosis (Zhu *et al.*, 1998). In a recent study a transgenic mice model was used to show that the core protein induces hepatocellular carcinoma (Moriya *et al* 1998).

1.1.2.2 ENVELOPE PROTEINS

Other structural proteins of HCV are E1 and E2 glycoproteins. Molecular mass of E1 varies from 28 to 35 kDa depending on the strain and conditions of expression (Ralston *et al.*,1993). Between residues 350-390 a stretch of hydrophobic amino acids is present, which is thought to act as a transmembrane anchor (Heinz *et al.*, 1992). In an immunoprecipitation experiment, the E1 protein was found to be precipitated by an anti-core antibody in the presence but not in the absence of the core protein, indicating that the E1 protein can interact with the core protein. This interaction is independent of whether the E1 and the core genes are linked in *cis* or separated in different DNA constructs for expression. Deletion-mapping studies indicate that the carboxy-terminal sequences of both the core and the E1 proteins are important for their interaction. Since a small portion of E1 sequence is exposed to the cytosolic side of the endoplasmic reticulum, the interaction between the core and the E1 proteins most likely takes place in the endoplasmic reticulum membrane (Lo *et al.*, 1996). Other envelope protein E2, gp72 is more heavily glycosylated than E1. It is encoded by E2/p7 and has a protein backbone of 32 kDa (Bradley *et al.*, 1992). Full length gp72 is not secreted but remains membrane-associated and there is an ER retention signal in the C terminal 29 amino acids (Cocquerel *et al.*, 1998). The C terminal position of the E2 protein is not absolutely clear at present. The location of E2 was predicted to be between amino acids 384-729 (Takamizawa *et al*, 1991). It was reported that (Hijikata *et al.*, 1991) the full E2 region does not extend aa 740 and the N terminal position of NS2 lies about amino acid 810 (Matsuura *et al.*, 1994).

Therefore residues encoding the predicted structural protein p7 could not be detected and its expression status is not known. Comparison of the available E2/p7 sequences revealed the presence of hypervariable regions (Kato *et al.*, 1992) at the N terminal of the E2 protein. HVR1 is one of these regions and it is located at the downstream of the cleavage site between E1 and E2/p7, covering 30 N terminal residues of the E2/p7 protein (Weiner *et al.*, 1991). It lacks a conserved secondary structure and resembles the V3 loop of HIV gp120 (Weiner *et al.*, 1992). Specific antibody reactions were detected against peptides corresponding to linear epitopes in VR1, indicating that the N terminal part of E2 region encodes antigenically distant variants, subjected to immune selection (Weiner *et al.*, 1992; Lesniewski *et al.*, 1993). The observed hypervariability may result from sequential mutations leading to escape mutants (Kato *et al.*, 1993) and variability rate is related with the immune pressure. This hypothesis was substantiated by the lack of variability in the HVR1 in an agammaglobulinemic patient over a period of 2.5 years (Kumar *et al.*, 1994). Escape from the host immune system by means of mutations in HVR1 might be involved in the mechanism of persistent infection by HCV, which results in chronic hepatitis and hepatocellular carcinoma.

It was observed that E1 and E2 form heterodimers which indicates that envelope proteins are involved in virus morphogenesis at budding. Characterization of HCV glycoprotein complex formation indicates that a majority of these proteins are misfolded aggregates (Deleersnyder *et al.*,1997). Analysis of HCV glycoprotein assembly in viral and nonviral expression systems showed similar results, which indicates that tendency toward aggregation is not due to the expression systems used but it is an intrinsic property of HCV glycoproteins. It

was observed that formation of stable E1-E2 complexes is slow because of the slow folding of these proteins. In the absence of E2, E1 does not fold properly which suggests that E2 plays a chaperon-like role in the folding of E1 (Michalak *et al.*, 1997). In addition, it was shown that ER chaperons like BiP, calnexin and calreticulin interact with E1 and E2 (Choukhi *et al.*, 1998). In a relatively recent study (Pileri *et al.*, 1998) it was found that E2 binds human CD81, a tetraspanin expressed on various cell types including hepatocytes and B lymphocytes. Binding of E2 was mapped to the major extracellular loop of CD81. This observation indicates that E2 is involved in virus-host interaction and extremely important in understanding viral interactions, cell culture studies and future vaccination studies.

1.1.3 NON-STRUCTURAL PROTEINS

Non-structural proteins of the HCV are located at the C-terminus of the genome. They are involved in the viral life cycle in the host cell, rather than being a structural component of the mature virion. Cleavages in the nonstructural protein region are accomplished by two viral protease activities located within the NS2–NS3 region. Cleavage at the NS2–NS3 junction occurs through an intramolecular (cis) mechanism in a zinc-dependent reaction (Hijikata *et al.*,1993, Grakoui *et al.*,1993a). In contrast, cleavages mediated by the NS3 serine protease release NS3, NS4A, NS4B, NS5A and NS5B proteins (Grakòui A. *et al.*,1993b).

1.1.3.1 NON-STRUCTURAL PROTEIN 2 (NS2)

HCV NS-2 is a transmembrane protein with a C-terminus located into the lumen of the endoplasmic reticulum and N-terminus located in the cytosol. There is a proteolytic function mapped to the C-terminus of the protein and it is involved in the autocatalytic cleavage of NS2/NS3 region in combination with serine protease NS3. Proteolysis at the NS2–NS3 junction was proposed to be catalyzed by a zinc protease. This hypothesis was based on observations that the NS2–NS3 proteolytic activity in in-*vitro* transcription/translation assays was inhibited by metal chelators such as EDTA and phenanthroline, and stimulated by exogenous zinc, but there is no direct structural evidence to support it (Pieroni *et al.*, 1997). An alternative to the metalloprotease hypothesis involves cysteine-protease-mediated cleavage at the NS2–NS3 site. In a recent study a relation between NS2-3 autoproteolysis and phosphorylation of NS5A was reported. This was supported by the loss of phosphorylated form of NS5A upon partial or complete deletion of the NS2 region from the HCV polyprotein. Also deletion of the amino acid residues between 810 and 907 of NS2 protein was shown to abolish the phosphorylated form of NS5A and resulted in an impairment of the autocleavage at the junction of NS2/NS3. Site-directed mutagenesis experiments that disrupt the NS2-3 autoprotease activity and studies that show the expression of NS2-3 precursor is sufficient to restore phosphorylation in trans are the other supports of this suggestion. Based on these data, apart form the its protealytic role, NS2 protein was postulated to be involved in phosphorylation of the NS5A and become on of the most important proteins in HCV life cycle (Liu *et al.*, 1999).

1.1.3.2 NON-STRUCTURAL PROTEIN 3 (NS3)

HCV NS3 protein is a 70-kDa multifunctional enzyme with three known catalytic activities in two different domains. The serine protease function was mapped to the N-terminal of the protein, whereas nucleoside triphosphatase (NTPase) (Preugschat *et al.*, 1996) and RNA helicase functions (Kadaré *et al.*, 1997) reside in the remaining C-terminal region. The HCV NS3 helicase can unwind dsRNA as well as dsDNA and RNA-DNA heteroduplexes in the 3'-to-5' direction by using NTP or dNTP as the energy source. The helicase and NTPase activities are inhibited in the presence of monovalent cations and NTPase activity is enhanced by the presence of polynucleotides (Tai *et al.*, 1996). Despite the high degree of sequence variability in the genome of HCV, serine protease domain of NS3 protein is highly conserved and carries common serine protease domains, like His-1083, Asp-1107 and Ser-1165 residues that form the enzyme catalytic triad at the active site of the protein. Temporal hierarchy in NS3 mediated protein cleavage was shown by several transient protein expression studies in cultured mammalian cells and in vitro transcription/translation assays (Failla *et al.*, 1995; Hahm *et al.*, 1995; Lin and Rice, 1995). The cleavage between NS3 and NS4A is the first event in the cascade and occur in cis to produce a mature noncovalent NS3-NS4A complex. Then cleavage at the following junctions occur in trans. The difference between the cis- and trans-cleavage sites is the presence of threonine as the P1 residue in the NS3/NS4A junction, whereas cysteine in the other junctions (Grakoui *et al.*, 1993c). Binding of NS4A and presence of a single Zn^{+2} atom per molecule are the necessary factors for the NS3 proteinase function. Binding of

NS4A enhances the trans-cleavage of the protease by stabilizing the protein. There are several interaction studies between the cellular proteins and HCV NS3 protein. It was observed that NS3 is co-localized with wild type p53 but not with mutant p53 in the nucleus, whereas when expressed alone it is localized both in the cytoplasm and in the nucleus to the same extend (Ishido *et al.*, 1997). NS3 suppresses actionomycine D-induced apoptosis (Fujita *et al.*, 1996) through the decreased expression of p53 (Ishido *et al.*, 1997), but interestingly expression of $p21^{WAF1}$, a downstream element of p53 induced apoptosis, is not effected (Fujita *et al.*, 1996). NS3 was also shown to be involved in signal transduction pathway by inhibiting and sequestering Protein Kinase A (PKA) activity in a cyclic AMP (cAMP)- independent manner (Borowski *et al.*, 1997).

1.1.3.3 NON-STRUCTURAL PROTEIN 4 (NS4)

The HCV NS4 protein is cleaved into 2 regions NS4A and NS4B, by transacting NS3 serine protease. NS4A, as stated before, is the structural component of NS3 protease. A 14 amino acid hydrophobic region of NS4A has been identified as being sufficient for this protease cofactor function of the protein (Butkiewicz *et al.*, 1996). The β strand in the central domain of the NS4A intercalates into the N-terminal part of the NS3 proteinase and becomes an integral part of the protein. The N-terminal 30 residues of the NS3 proteinase form β strand with the neighboring molecules in the absence of NS4A (Kim *et al.*, 1996, Love at al., 1996). It was also observed that NS4A forms nonionic detergent-stable complex with the NS4B-NS5A protein complex through the central domain which may indicate that four nonstructural proteins might function

as a multisubunit protein complex. This protein complex might be involved in the replication of HCV genome in combination with viral RNA-dependent RNA polymerase and some possible cellular cofactors (Lin *et al.*, 1997). It was also shown that NS4A is involved in the phosphorylation of the central region of NS5A (Asabe *et al.*, 1997).

1.1.3.4 NON-STRUCTURAL PROTEIN 5 (NS5)

The HCV NS5 region encodes two proteins, NS5A and NS5B. NS5A is a phosphoprotein, which exists in differentially phosphorylated forms of 56 k-Da and 58 k-Da (Kaneko *et al.*, 1994). When only the NS5A region of the HCV polyprotein is expressed in cultured cells, expression of the 56-kDa product is higher than the 58-kDa product, whereas production of 58-kDa protein is increased in the presence of NS4A. Both forms of the HCV NS5A protein are located in the perinuclear membrane fraction (Tanji *et al.*, 1995). In a recent study, it was shown that NS4A-dependent hyperphosphorylation of NS5a is genotype specific. In the presence of NS4A it was observed that HCV-2a NS5A is phosphorylated but not hyperphosphorylated and it was not associated with NS4A, whereas HCV-1b NS5A is hyperphosphorylated and associated with NS4A. These results suggest that different phosphorylation status of HCV NS5A region might reflect different virological features of these two genotypes (Hirota *et al.*, 1999). It was reported that NS5A protein, without its 146 amino-terminal amino acids and fused to the DNA-binding domain of GAL4, strongly activates transcription in yeast and human hepatoma cells (Kato *et al.*, 1997). Another important characteristic of this protein is the presence of interferon sensitivity

determining region (ISDR). Like other viral transactivator proteins known to repress interferon-induced gene expression, the ISDR in NS5A overlaps one of the acidic amino acid regions, putative domains conferring this activity. Presently, the only therapeutic approach to hepatitis caused by HCV is long-term treatment with α-interferon, alone or in combination with ribavirin (Main *et al.*, 1995). Patient age, duration of infection, presence of cirrhosis before the treatment and genotype are the factors that influence the response to the treatment. Both 1a and 1b exhibit a high level of resistance to interferon. This resistance was correlated with mutations between amino acid positions 237-276 of NS5A protein (Fukuma *et al.*, 1998). ISDR was thought to be involved in IFN mechanism via PKR pathway. IFN-induced cellular antiviral response is mediated by Mx proteins, the 2'-5' oligoadenylate synthetase, RNAse L and PKR. Upon induction by IFN, these proteins block viral gene expression at multiple levels. PKR is known to phosphorylate the α subunit of the eukaryotic initiation factor 1 (eIF-2α) and leads to cessation of protein synthesis. Many viruses have evolved mechanisms to block PKR activity, such as the PKR kinase inhibitor activity of HIV tat protein. It was found that NS5A represses PKR activity through a direct interaction with the protein kinase catalytic domain upon binding to PKR (Gale *et al.*, 1998). Recently it was reported that NS5A protein represses transcription of the cell cycle regulatory gene p21WAF1, while it activates the human proliferating cell nuclear antigen gene in murine fibroblasts and human hepatoma cells. Furthermore, introduction of NS5A into murine fibroblasts (NIH3T3) promoted anchorage-independent growth and tumor formation in nude mice. According to these data, NS5A might exhibit a role in cell growth regulation in addition to the other activities (Ghosh *et al.*, 1999).

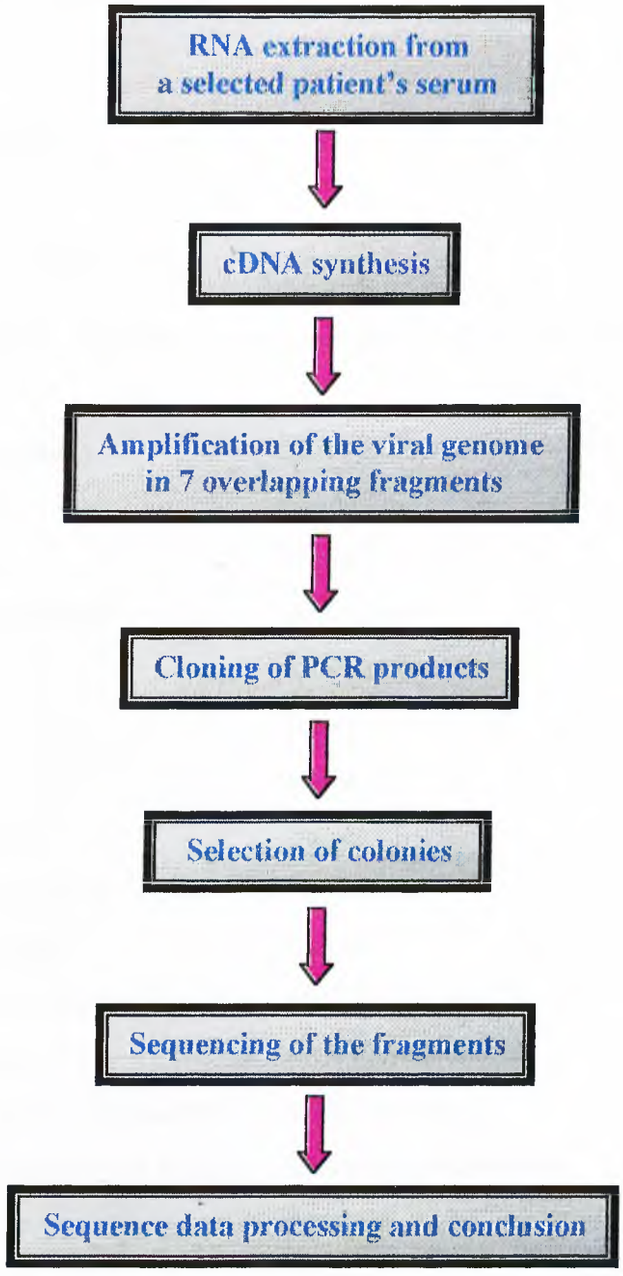
The amino acid sequence studies on NS5B have shown that there is a homology with the other known RNA-dependent RNA polymerase sequences (Miller and Purcell, 1990). A Gly-Asp-Asp sequence motif is characteristic of RNA-dependent RNA polymerases of positive strand RNA viruses and it is present between the residues 2737 and 2739 of the NS5B region of the HCV genome. According to this sequence data, NS5B was identified as the region that encodes viral polymerase and later this activity was demonstrated in vitro (Chung and Kaplan, 1992). In a cell-free system, the RNA-dependent RNA polymerase activity of the protein has been demonstrated and, by using the baculovirus expression system the functional and biochemical characteristics of the NS5B protein were identified (Behrens *et al.*, 1996). Another result of this study was about the unspecificity of the polymerase activity. RNA synthesis was observed with unrelated input RNA molecules which suggests that the NS5B protein is necessary but not sufficient for the replication of the HCV genome. Some other viral or cellular proteins are needed for the polymerase specificity. The localization of the NS5B protein was found to be at the perinuclear region in association with the nuclear membrane and the endoplasmic reticulum or Golgi complex (Hwang *et al.*, 1997). This finding suggests that HCV replication might take place in the membrane complex and this is consistent with the replication region of other RNA viruses. It has been suggested that, in some isolates, a secondary structure of the genomic RNA exists at the region encoding the C terminal part of the NS4 region. This structure formation might serve as an IRES which might be located upstream of a common inframe ATG codon and accelerate the replication rate of the viral genomic RNA (Okamoto *et al.*, 1992).

As a summary, chronic hepatitis C virus (HCV) infection occurs in about 3 percent of the world population and is a major cause of liver disease. HCV infection is also associated with cryoglobulinemia, a B lymphocyte proliferative disorder (Rice *et al.*, 1999). High incidence of the virus, direct involvement with several liver and some immunologic diseases, lack of a successful therapy or vaccine lead to an increasing interest in molecular biology studies on HCV. As an RNA virus, the mutation rate of the virus is high and the differences between the caused pathologies, immunity levels and the resistance of some of the variants to IFN therapy show that characterization of the variants is one of the most important steps in HCV studies.

1.2 AIM OF THE STUDY AND STRATEGY

As a rapidly evolving RNA virus, HCV contains several base changes throughout the genome. As summarized previously this variability is not evenly distributed between the regions. Based on the classification system proposed by Simmonds *et al.* in 1994, 6 main genotypes and several subtypes were defined. The characterization of a predominant genotype in a selected population is necessary due to the differences between the pathologies caused , immune responses created and responses to the IFN treatment between the genotypes. The analysis of the sequence data will provide information about the molecular biology of the virus, its persistence, treatment, evolution and possible vaccination studies. The recent data based on tests made on blood donors in Turkey showed that 1,5% of the population is anti-HCV positive. To date several clinical data about HCV status in Turkey were published but the molecular data of the HCV virus type in Turkish population was not stated.

In this study we aimed to characterize a HCV genome isolated representing the predominant genotype of Turkish population. Based on the data obtained from a previous study of genotyping of the HCV in the Turkish population, a patient with 1b subtype was selected. The viral genome was divided into 7 overlapping PCR fragments each corresponding to one or two viral protein and amplified. PCR products were cloned into bacterial vectors and sequenced. Sequencing data were compared with known 1b genomes and differences were discussed. The strategy of the project is summarized below.


RNA extraction from a selected patient's serum
cDNA synthesis
Amplification of the viral genome in 7 overlapping fragments
Cloning of PCR products
Selection of colonies
Sequencing of the fragments
Sequence data processing and conclusion

2. MATERIALS AND METHODS

2.1 MATERIALS

2.1.1 Chemicals and Enzymes

All enzymes or reagents were purchased from the sources listed below:

Restriction Enzymes	MBI Fermentas and Appligene
Ribonuclease A	Sigma
RNAse Inhibitor	Promega
Cloned *Pfu* Polymerase	Stratagene
dNTPs	MBI Fermentas
IPTG	Sigma
X-Gal	Sigma
Gel Extraction Kit	Qiagen
Plasmid Purification Kit	Qiagen
PCR Purification Kit	Qiagen
100 bp plus DNA Marker	MBI Fermentas
1 kb DNA Marker	MBI Fermentas
First Strand cDNA Synthesis Kit	MBI Fermentas
Expand High Fidelity PCR System	Boehringer Mannheim
pGEM-T Easy Vector System	Promega
Ampicilin	Sigma

All other chemicals were purchased from Sigma, Difco or Carlo-Erba. Oligodeoxynucleotides were synthesized on a Oligo 1000M DNA Synthesizer (Beckman) in the Molecular Biology and Genetics Department, Bilkent University and Iontek company, Bursa, Turkey.

2.1.2 Equipment

The list of equipment used and their manufacturers is given below:

Automatic Pipettes	
P2	Gilson
P10	Gilson
P20	Rainin, Pipet-plus
P200	Rainin, Pipet-plus
P1000	Rainin, Pipet-plus
GeneAmp PCR System 9600 and 2400	Perkin Elmer
Centrifuges	
Biofuge (pico)	Heraus Instruments
Avanti J-25 I	Beckman
GS-15 R	Beckman
Gel Tanks	
Horizontal Minicell	E-C Apparatus Corporation
pH Meter	E-C Apparatus Corporation
Spectrophotometer DU 640	Beckman
Power Supply PAC 300	BioRad
UV Transilluminator	BioRad GelDoc 2000

2.1.3 Plastic Disposables

The list of plastic disposables used and their manufacturers is given below:

Microcentrifuge Tubes	Costar
0.2 ul	
1.5 ul	
2.0 ul	
Petri Dishes	LP
Syringes	Fefarma

Filter	Costar
Test Tubes	Costar and Nunc
15 ml	
50 ml	
Tips	Costar and USA/Scientific Plastics
Filter Tips	Greiner

2.1.4 Bacterial Strain

The bacterial strain used in this study is *E.coli* HB101. The genotype of the strain is: *thi-1, hsdS20 (rB–, mB–), supE44, recA13, ara-14, leuB6, proA2, lacY1, galK2, rpsL20 (strr), xyl-5, mtl-1*

Bacteria were grown in liquid culture in Luria-Bertoni (LB) medium. For selective medium ampicillin was used in this study. For long term studies, *E.coli* strains were stored in glycerol in 1/1 ratio, at -70°C.

2.1.5 Solutions and Media

2.1.5.1 General Solutions

EDTA:	0.5 M (stock solution), pH 8.0
$CaCl_2$:	1M (stock solution), 50mM (working solution)
1X TAE:	40 mM Tris-Acetate, 1mM EDTA (pH 8.0)
1X TBE:	45mM Tris-borate, 1mM EDTA (pH 8.0)
EtBr:	10 mg/ml water (stock solution) 30 ng/ml (working solution)
1X Gel Loading Buffer:	0.25% bromophenol blue, 0.25% xylene cyanol, 50% glycerol, 1mM EDTA
X-Gal	50 mg/ml, disssolved in N,N´-dimethyl-formamide, stored at -20°C.
IPTG	100mM, filter-sterilized and stored at 4°C.

2.1.5.2 Solutions for RNA Extraction

Guanidium Thiocyanate	4M, prepared with RNase free ddH_2O and stored at 4°C.
N-Laurylsarosine	0.5%, prepared with RNase free ddH_2O and stored at 4°C.
Sodium Acetate	25mM (pH 7.0), prepared with RNase free ddH_2O and stored at 4°C.

2.1.5.3 Solutions for Plasmid DNA Isolation:

Solution I	50 mM Glucose, 25 mM Tris-Cl, pH 8.0, 10 M EDTA
Solution II:	0.2 N NaOH, 1% (wt/vol) SDS
Solution III:	3 M Potassium Acetate, pH 4.8
Phenol/Chloroform:	1/1 ratio of phenol and chloroform.

2.1.5.4 Microbiological Media and Antibiotic

LB:	1% g bacto-tryptone, 0.5% yeast extract, 1% NaCl.
LB-Agar:	1% g bacto-tryptone, 0.5% yeast extract, 1% NaCl, 1.5% Agar.
Glycerol Stock Solution:	65% glycerol, 0.1 M $MgSO_4$, 0.025 M Tris-Cl (pH 8.0).
Ampicillin:	100 mg/ml solution in ddH_2O, sterilized by filtration and stored at -20°C (stock solution). 100 ug/ml (working solution)

2.2 METHODS

2.2.1 Patient

A female, 59 years old, Turkish patient from the Gastroenterology Department of Çukurova University in Turkey, who was found to be anti-HCV antibody positive with second-generation ELISA test was chosen for complete genome sequencing of Hepatitis C Virus. HCV RNA detection with PCR amplification was positive and according to the sequencing results of the 5'-UTR, the genotype of the sample was previously identified as subtype 1b.

2.2.2 Viral RNA Extraction and cDNA synthesis

Total RNA was extracted from patient's sera by single step guanidium thicyanate method (Ausbel *et al.*, 1991) with minor modifications (Wilson *et al.*, 1995).

RNA was extracted from 300 ul of serum with the freshly made lysis buffer containing guanidine-HSCN and mercaptoethanol in the presence of an RNA carrier; RNA was then recovered by isopropanol precipitation and resuspended in 10 ul DEPC–ddH2O and denatured at 90 °C. First strand cDNA was synthesized by First Strand cDNA Synthesis kit (Cataloge# K1612, MBI Fermentas). cDNA synthesis was performed on 10 ul resuspended RNA in a 20 ul reaction volume by using 0.2 ug of random primer, 40 U of M-MuLV Reverse Transcriptase, 20 U of Ribonuclease Inhibitor (Promega), and 1mM (each) deoxyribonucleotide at 37 °C for 1 hr.

2.2.3 Oligonucleotide Synthesis

Two sets of primers were synthesized for this study. First set for amplification reactions and second set for the sequencing reactions. Sequences of the primers used in this study were given in Table 2.1 and Table 2.2.

Oligonucleotides used in PCR were synthesized by Lutfiye Mesci (Bilkent University, Department of Molecular Biology and Genetics), at Oligo 1000M

DNA Synthesizer by utilizing cyanoethyl-phosphoramidite chemistry. After synthesis oligonucleotides were cleaved and deprotected by UltraFast Cleavage and Deprotection Kit (Beckman) and dried by the SpeedVac (Beckman). The concentrations of the oligonucleotides were determined by using oligodeoxynucleotide quantification option of the spectrophotometer. Oligonucleotides for the internal regions used in sequencing reactions were synthesized by IONTEK (Bursa, Turkey).

Table 2.1: PCR Primers

Primer	Fragment	Region	Site	Primer Sequence
F1	Zeus	5'UTR+Core	36	**ATCACTCCCCTGTGAGGAAC**
CoreR			964	**(G/A)GAGCA(G/A)TCGTTCGTGACAT**
E1F	Charon	E1 + E 2	850	**CCCGGTTGCTCTTTCTCTATC**
E2R			2791	**ATGC(A/G)GCCATCTCCCGGTC**
NS2F	NS2	NS2	2738	**T(C/T)CT(A/G)(C/T)TG(G/T)C(G/A)TTACCACC**
NS2R			3449	**GT(T/C)TG(T/C)TG(G/A/T)G(A/C)GTAGGCCGT**
NS3F	NS3	NS3	3354	**CCGAAGGGGGA(A/G)GGAGAT**
NS3R			5326	**GCACCCA(G/A)GTGCT(A/C/G)GT(G/A)ACGAC**
NS4F	NS4	NS4A+NS4B	5283	**ATGCATGTCGGC(T/C)GACCT**
NS4R			6329	**TG(G/A)AGCCA(G/A)GTCTTGAAGTC**
NS5AF	NS5A	NS5A	6142	**TATGTGCCTGAGAGCGACG**
NS5AR			7695	**(A/G)CG(C/T)AGCAAAGAGTTGCTCA**
NS5BF	NS5B	NS5B	7543	**AGCGACGGGTC(T/C)TGGTCTAC**
NS5BR			9397	**CCTGGAGTG(G/T)TT(A/G)GCTCCC**

Table 2.2: Sequencing Primers

Primer	Fragment	Region	Site	Sequence
E2Fseq1	Charon	Envelope 2	1591	TGGCACATCACAGGACTG
NS3Fseq1	NS3	NS3	3771	CGGCGGGGCGACAGCAGGG
NS4Rseq1	NS4	NS4A + NS4B	5875	CACCTTCCCAAGGCCTAT
NS5Afseq1	NS5A	NS5A	6627	GTGGGGGATTTCCACTACG
NS5Arseq1	NS5A	NS5A	7294	TCTTCCTCCGTGGAGGT
NS5Bfseq1	NS5B	NS5B	8014	TGGCAAAAAATGAGGTTTTCTG
NS5BRseq1	NS5B	NS5B	8888	TAGGGCTTTCTCAAGTTGCT
T7	Vector	Forward Primer, Multiple Cloning Site	-	TAATACGACTCACTATAGGG
M13	Vector	Reverse Primer, Multiple Cloning Site	-	CAGGAAACAGCTATGAC

2.2.4 Polymerase Chain Reaction

HCV genome was amplified in 7 overlapping PCR fragments. Primer sets for each fragment were designed from the most conserved regions in the desired area. Alignments of the regions that the primers were designed from were given in Appendix A.

Due to the insufficient amount of template DNA, two PCR reactions with the same primer sets were set up for each fragment. Fragments smaller than 1 kb were amplified with Cloned *Pfu* DNA Polymerase (Catalog# 600153, Stratagene) in the first round in order to decrease the mutation rate and in the second round Expand High Fidelity PCR System (Catalog# 1732650, Boehringer Mannheim) was used to obtain PCR products with A nucleotide overhangs. Fragments larger than 1 kb were amplified with Expand High Fidelity System in both PCR cycles. Optimized PCR reaction conditions for fragments smaller than 1 kb and larger than 1 kb are given in Table 2.3 and 2.4 respectively. Amplifications of all fragments were performed as hot start PCR.

Table 2.3: PCR Conditions for Fragments Smaller than 1 kb.

Fragment	1st PCR Mix	2nd PCR Mix	PCR Conditions
Zeus	• ¼ of the total cDNA • 50 pmol/ul primer • 1X *Pfu* Buffer • 10 mM dNTP • 2.5 units *Pfu*	• 1/5 of the 1st PCR product • 50 pmol/ul primer • 5X Buffer with $MgCl_2$ • 10 mM dNTP • 2 units Expand High Fidelity PCR Enzyme	94 °C 3' 94 °C 30'' 60 °C 50'' 72 °C 1'] 30 cycles 72 °C 10'
NS2	• ¼ of the total cDNA • 100pmol/ul primer • 1X *Pfu* Buffer • 10 mM dNTP • 1% DMSO • 2,5 units *Pfu*	• 1/5 of the 1st PCR product • 100pmol/ul primer • 5X Buffer with $MgCl_2$ • 10 mM dNTP • 2 units Expand High Fidelity PCR Enzyme	94 °C 3' 94 °C 30' 60 °C 50' 72 °C 1'] 30 cycles 72 °C 10'

Table 2.4: PCR Conditions for Fragments Longer than 1 kb.

Fragment	1st PCR Mix	2nd PCR Mix	PCR Conditions
Charon NS3 NS4 NS5A NS5B	• ¼ of the total cDNA • 50 pmol/ul primer •5X Buffer with $MgCl_2$ • 10 mM dNTP • 2 units Expand High Fidelity PCR Enzyme	•1/5 of the 1st PCR product •50 pmol/ul primer •5X Buffer with $MgCl_2$ • 10 mM dNTP • 2 units Expand High Fidelity PCR Enzyme	94 °C 3' 94 °C 30'' 60 °C 50'' 72 °C 2'] 30 cycles 72 °C 10'

2.2.5 *Agarose Gel Electrophoresis of DNA Fragments*

DNA fragments were separated by gel electrophoresis using agarose of 1% w/v in 1X TAE or TBE buffer (Maniatis *et al.*, 1982). 0.2-0.4 ug EtBr/ml of gel was added to visualize the DNA bands. DNA samples were loaded into the wells with 1X loading buffer. Electrophoresis was performed by running the gel for 1 hour at 100 V. The DNA bands were visualized by illuminating with UV light at 302nm and the gels were photographed.

The size and concentration of the PCR fragments were checked in the agarose gel. Bands with the correct size were isolated from the agarose gel slices by using QIAquick Gel Extraction Kit (Catalog#: 28106, Qiagen Inc.), according to the manufacturers instructions.

2.2.6 *Electrophoresis Markers*

The length of DNA fragments were estimated by comparing them to known molecular weight standards, which had been run on the same gel. DNA markers used in this study were 1 kb DNA Ladder (Catalog# SM0311, MBI Fermentas) and 100 bp DNA Ladder Plus (Catalog# SM0321S, MBI Fermentas. Sizes of the fragments are given in Figure 2.1.

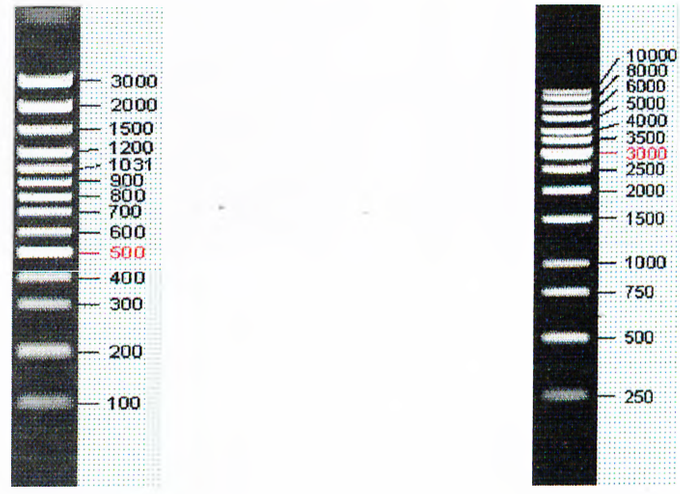


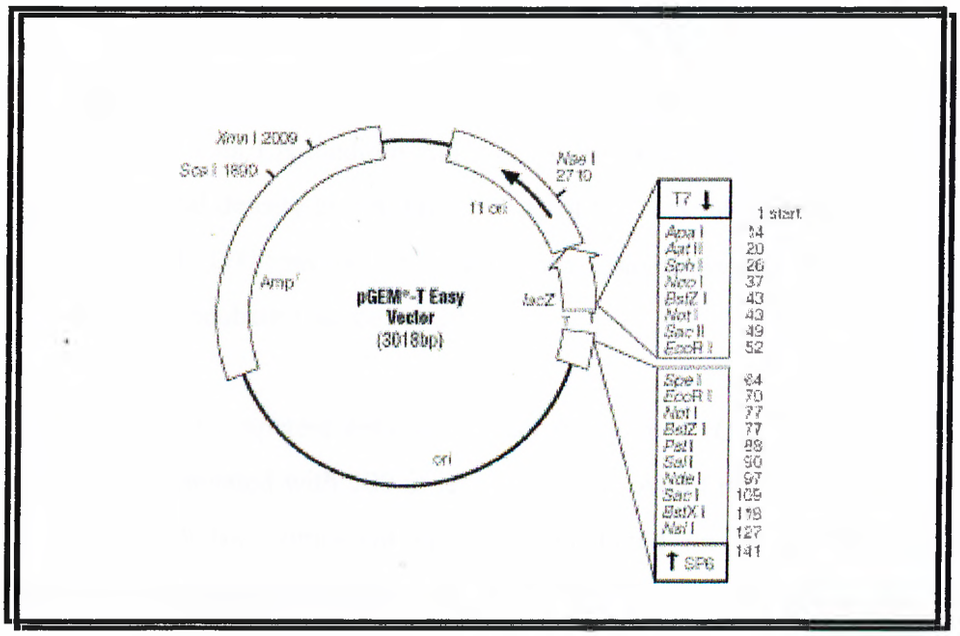
Figure 2.1: DNA size markers. Panel A; 100 bp DNA Ladder Plus and Panel B; 1kb DNA Ladder.

2.2.7 Cloning of PCR Products

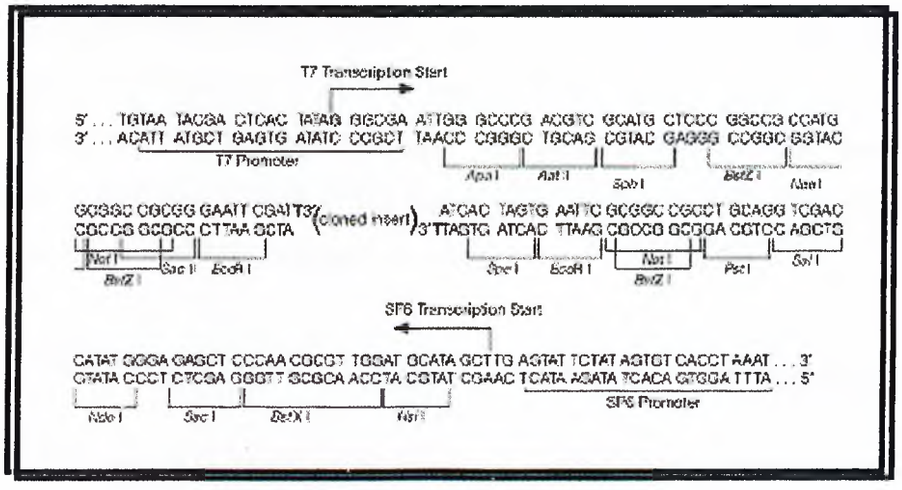
Primers for PCR reactions do not contain any linker sites. In order to clone the amplified fragments, pGEM®-T Easy Vector System (Cataloge # TM042, Promega) was used. The vector map is given in Figure 2.2. pGEM®-T Easy Vector is a linearized vector with T overhangs in the multiple cloning site. *BstZ*I, *Eco*RI, *Not*I enzyme sites are present at the two sides of the cloning site and they are used to check the presence of an insert with a single restriction enzyme reaction. T7 and M13 primer sequences, that are present at the two ends of the cloning site, were used for sequencing of the insert. Multiple cloning site of the vector is in the middle of the LacZ gene. Presence of an insert interrupts gene expression and colonies containing insert appears white whereas colonies containing religated vectors appear as blue, in the presence of X-Gal and IPTG. Cloning reactions were set up in 10 ul total volume and optimized conditions are given in Table 2.5.

Table 2.5: Cloning conditions of PCR products

	Reaction Mix			
PCR Products	**Insert**	**pGEM®-T Easy Vector**	**Ligase**	**LigaseBuffer**
ZEUS	60 ng	15 ng	0.5 u	1ul
Charon	20 ng	15 ng	0.5 u	1ul
NS2	60 ng	15 ng	0.5 u	1ul
NS3	20 ng	15 ng	0.5 u	1ul
NS4	20 ng	15 ng	0.5 u	1ul
NS5A	20 ng	15 ng	0.5 u	1ul
NS5B	20 ng	15 ng	0.5 u	1ul



A



B

Figure 2.2: A.) pGEM®-T Easy Vector Map and B.) Multiple Cloning Site

2.2.8 Transformation of E.Coli

Transformation of plasmid DNA into *E.coli* was achieved by using calcium chloride method. The following procedure is based on Ausubel *et al.* (1991).

Preparation of Competent Cells: 200 ul of HB101 overnight grown culture was inoculated into 5 ml LB medium and cells were grown at 37°C, shaking at 200 rpm to an optical density at 590 nm (OD_{590}) of 0.4. 1.5 ml of growing cells were centrifuged at 13,000 rpm for 1 min and gently resuspended in 500 ul ice-cold 50 mM $CaCl_2$ and incubated on ice for 30 min.

Transformation: Competent cells were centrifuged at 13.000 rpm for 1 min and pellets were suspended with 500 ul ice-cold $CaCl_2$. All of the cloning mix (10 ul) was mixed with the competent cells and incubated on ice for 30 min. The competent cells were heat shocked at 42°C for 90 seconds and then incubated on ice for 2 minutes. 1 ml of LB medium was added onto the competent cells and incubated at 39°C water bath for 1 hour to allow the expression of antibiotic resistance gene before plating onto the selective LB-Agar plates. After incubation cells were centrifuged for 2 minutes and pellet was resuspended in 100 ul left supernatant and plated onto LB-Agar plates containing 100ug/ml ampicillin, 50mg/ml X-Gal, and 100mM IPTG. Plates were incubated at 37°C overnight for the selection of antibiotic resistant transformants.

2.2.9 Small Scale Preparation of Plasmid DNA (Mini-Preperation)

The method is based on alkaline lysis method of the Birnboim and Doly (1979). Colonies that appear as white after overnight incubation at 37°C were selected and inoculated into 5 ml LB media containing ampicillin. After overnight incubation at the 37°C shaker, 1.5 ml of the bacterial culture was centrifuged at 13,000 rpm for 1 minute. After discarding the supernatant, cells were resuspended in 0.1 ml ice cold solution I and 0.2 ml freshly prepared solution II, 0.5 ml solution III and 0.5 ml phenol/chloroform were added. The tube was shacked vigorously by vortexing 2 minutes and centrifuged 5 minutes at 13,000 rpm. Upper phase that

contains the plasmid DNA was transferred into a clean tube and 1 ml ice cold absolute ethanol was added. The tube was centrifuged for 10 minutes at 13,000 rpm and supernatant was discarded. The pellet was washed with ice cold 70% ethanol, left at room temperature for 15-20 minutes to dry and resuspended in 20 ul ddH_2O containing 10 ug/ml RNase A. Samples were kept at -20°C for long-term storage.

2.2.10 Restriction Enzyme Digestion of DNA

Presence of the correct size inserts in the mini-preps of selected colonies, were checked with restriction enzyme digestions. Restriction enzyme digestions of DNA were carried out in a total volume of 10 ul with 5-10 units restriction enzyme with 1X recommended buffer. Digestion reactions of mini-preps were incubated for 1.5 hours at 37°C.

2.2.11 Medium and Maximum Scale Isolation (Midi-Preperation)

Cells were grown in 100 ml ampicillin containing Lb for overnight to achieve complete saturation. Plasmid DNA was isolated, by using QIAfilter plasmid midi kit (Cataloge# 10 12145, Qiagen inc.), according to the manufacturer's instructions. The DNA samples were stored at -20°C.

2.2.12 Quantification of DNA

Concentrations and purity of nucleic acids were determined by measuring absorbency at 260nm and 280nm in a spectrophotometer (Beckmann Instruments Inc., CA, USA). Nucleic acid samples displaying OD_{260} and OD_{280} values in the range of 1.8 to 2.0 are regarded as highly pure. A value of OD_{260}=1.0 corresponds to a concentration of approximately 50 ug/ul double stranded DNA (Maniatis *et al.*, 1982).

2.2.13 Automated Sequencing of DNA Fragments

The automated sequencing of the inserts were carried on an ABI Prism 310 Sequencer at Bilkent University, Department of Molecular Biology and Genetics. To increase the sequencing length, Big Dye™ Terminator Cycle Sequencing Ready Rection Kit (Catalog# 4363149, ABI Prism) in combination with long capillary of the 310 Sequencer were used. First sequencing reactions were carried out with M13 and T7 universal primers that are present at the two ends of the cloning site, on the vector sequence. For the middle sequences internal primers were designed and synthesized according to the outer sequencing data obtained from M13 and T7 reactions.

3. RESULTS

3.1 Introduction

Hepatitis C Virus is a major cause of non-A, non-B hepatitis worldwide. In a majority of patients the virus leads to the development of chronic hepatitis and later cirrhosis and hepatocellular carcinoma. Since 1989 several complete or nearly complete genomic sequences have been published (Choo *et al.*, 1989, 1992, Inchauspe *et al.*, 1991, Ogata *et al.*, 1991, Okamoto *et al.*, 1991b, Takamizawa *et al.*, 1991, Yeh *et al.*, 1997, Rispeter *et al.*, 1997, Aizaki *et al.*, 1998, Chamberlain *et al.*, 1997a, 1997b). Comparison of the nucleotide and polypeptide amino acid sequences of these isolates revealed considerable heterogeneity which leads to the classification of multiple genotypes. According to the results of the genotyping studies, the distribution of the genotypes of the virus was mapped. In a previous study the predominant genotype of Turkey was found to be 1b by sequencing the 5'-UTR region (Yıldız *et al.*, unpublished data). Based on the previous studies indicating the high variability rate of the viral genome, it was decided to clone and sequence the full length HCV genome representing the first Turkish isolate. The strategy to clone the viral genome was based on 1) Patient selection according to the subtype, age and interferon therapy, 2) RNA isolation, 3) cDNA synthesis according to previously optimized conditions (Yıldız *et al.*, unpublished data), 4) Amplification of the fragments by PCR, 5) Direct cloning of the PCR products and 6) Sequencing of the cloned regions.

Table 3.1: Isolates of Submitted HCV 1b Complete Genomes

Isolate	Definition	Subtype	Accession	Author
AF054248	HCV HC-J4, pCV-J4L2S	1b	AF054248	Yanagi,M *et al.*
AF054249	HCV HC-J4, pCV-J4L4S	1b	AF054249	Yanagi,M *et al.*
AF054247	HCV HC-J4, pCV-J4L6S	1b	AF054247	Yanagi,M *et al.*
HCU45476	HCV isolate HD-1	1b	U45476	Mueller,H.M.
HPCK1R1	HCV clone HCV-K1-R1	1b	D50480	Enomoto,N.
HPCK1R2	HCV clone HCV-K1-R2	1b	D50481	Enomoto,N.
HPCK1R3	HCV clone HCV-K1-R3	1b	D50482	Enomoto,N.
HPCK1S1	HCV clone HCV-K1-S1	1b	D50483	Enomoto,N.
HPCK1S2	HCV clone HCV-K1-S2	1b	D50485	Enomoto,N.
HPCK1S3	HCV clone HCV-K1-S3	1b	D50484	Enomoto,N.

3.2 Strategy for Amplification of the Viral Genome

Sequences of the previously published 1b isolates, given in Table 3.1 were aligned and analyzed. For the possible future expression studies, the viral genome was divided into 7 overlapping fragments each corresponding to one or two viral proteins, as shown in Figure 3.1. From the desired areas, most conserved regions were selected for primer design, and variability problem in some of the desired areas was solved by designing degenerate primers. Aligned sequences of the primer regions were given in Appendix.

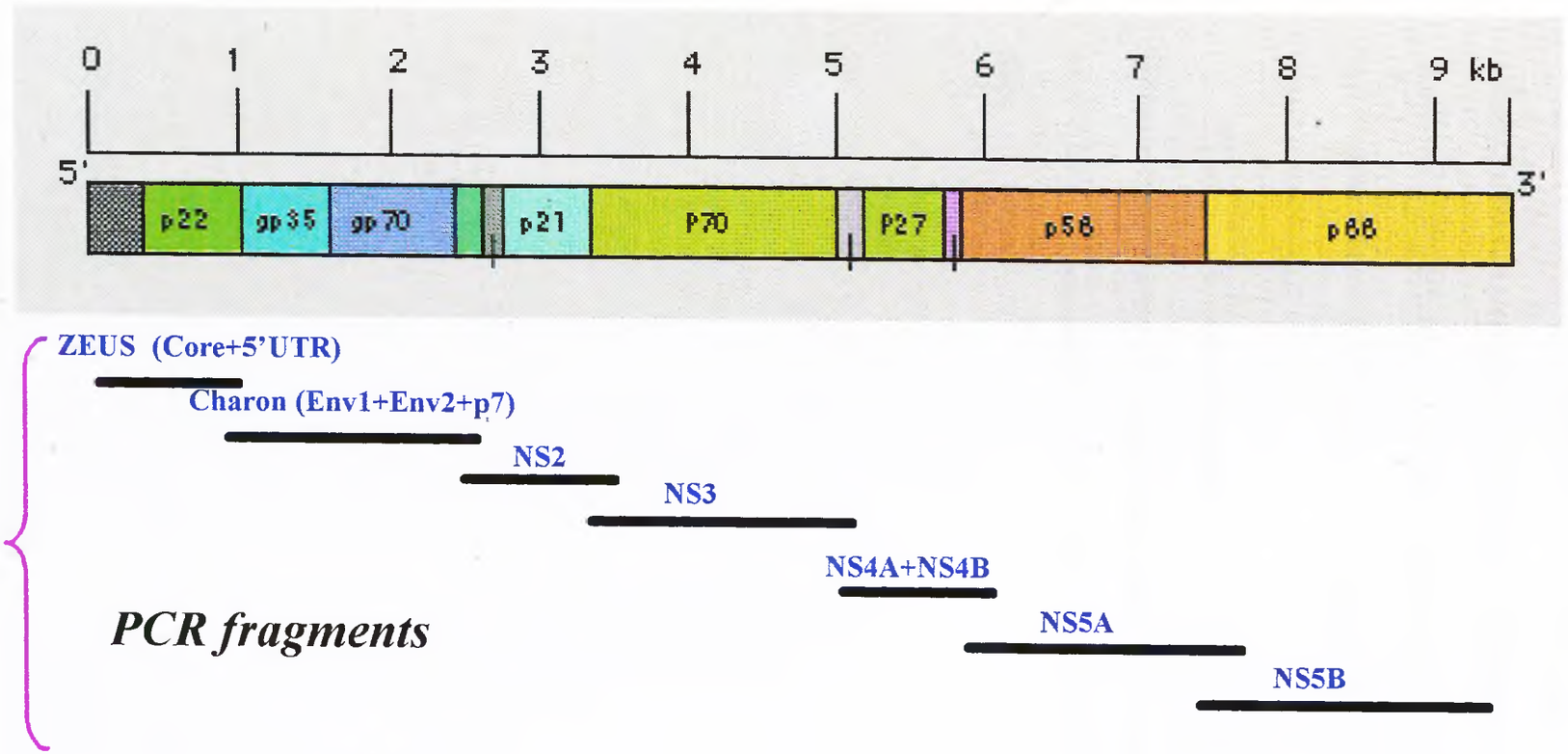


Figure 3.1: Amplification strategy of HCV genome

In Figures 3.2 and 3.3, agarose gel electrophoresis profiles of PCR products o the regions Zeus corresponding to 5'-UTR and Core (Panel A), Charor corresponding to Envelope 1 and Envelope 2 (Panel B), Non-Structural 2 (Panel A) Non-Structural 3 (Panel C), Non-Structural 4 including regions of Non-Structural 4A and Non-Structural 4B (Panel D) , Non-Structural 5A (Panel E), Non-Structural 5E (Panel F) were given.

There were unspecific bands in some of the PCR reactions. In order tc eliminate the possibility of cloning wrong insert, PCR products were run on agarose gel and bands corresponding to the expected sizes were isolated from the agorose gel slices.

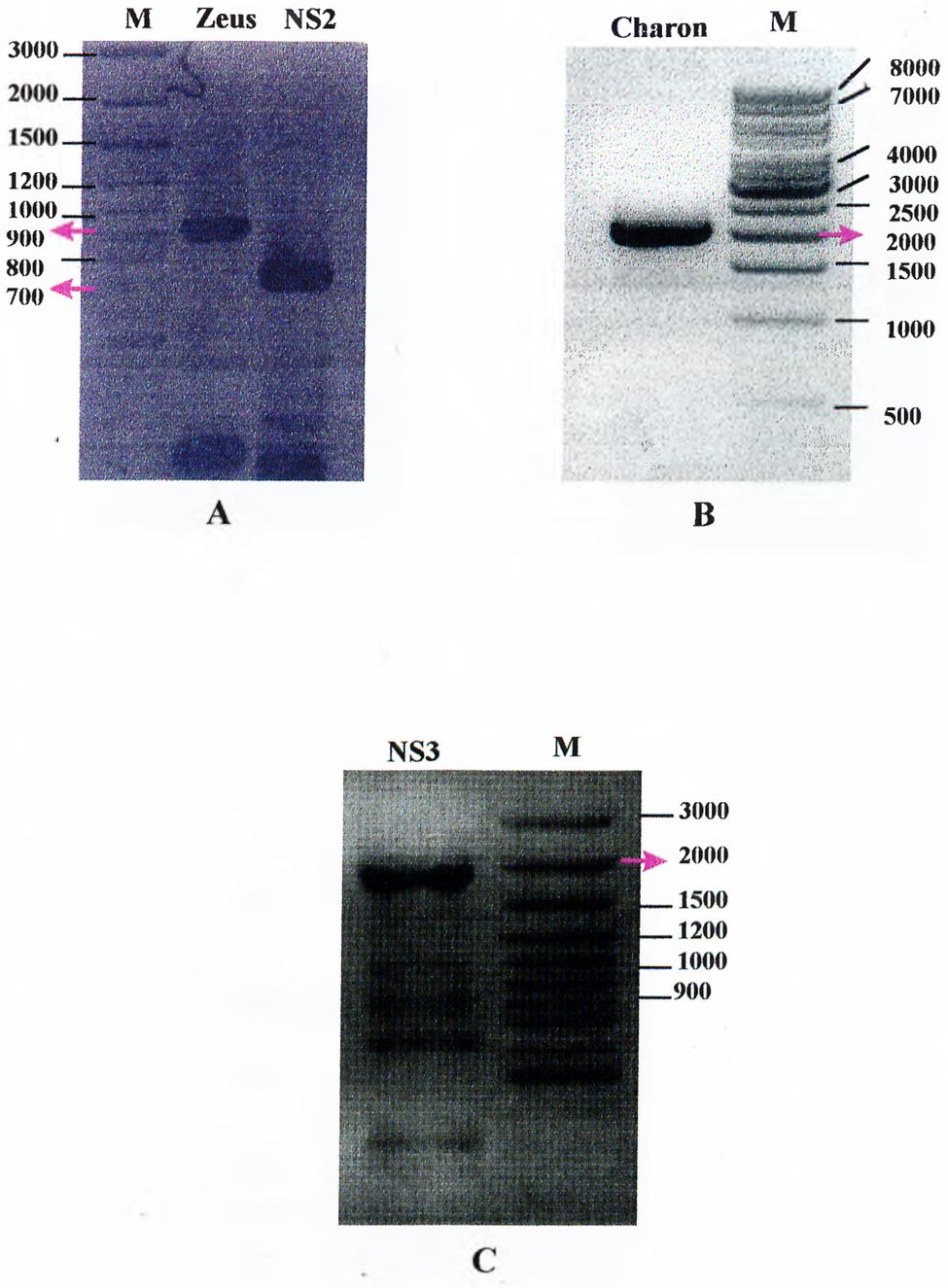
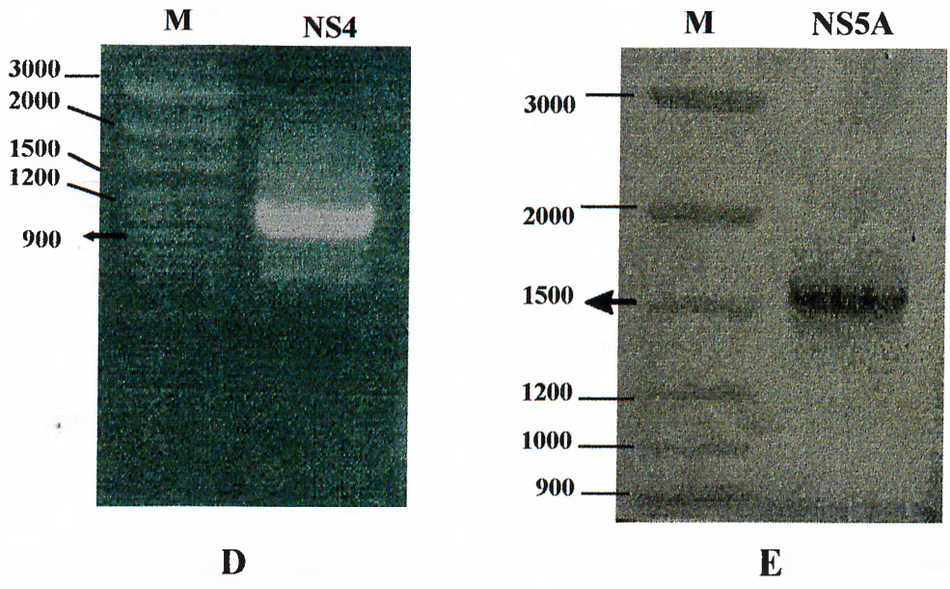


Figure 3.2: Agarose gel electrophoresis of PCR products. DNA size marker in Panel A and C is 100 bp DNA ladder Plus (MBI), DNA size marker in Panel B is 1 kb DNA ladder (MBI).



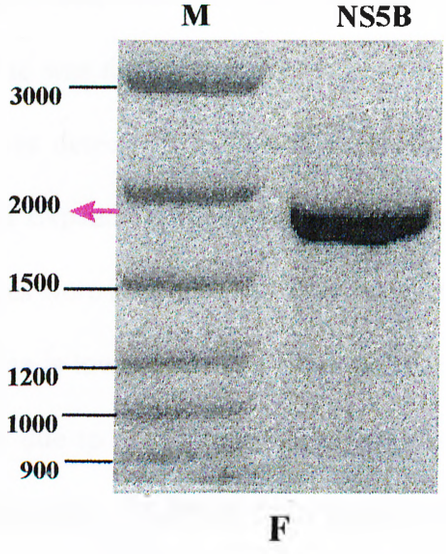


Figure 3.3: Agarose gel electrophoresis of PCR products. DNA size marker in Panel D, C and F is 100 bp DNA Ladder Plus (MBI).

3.3 Strategy for Cloning of the PCR Products

As the sequence of the viral isolate was not known exactly, instead of designing primers with restriction enzyme sites, purified PCR products having A overhangs, due to the activity of the *Taq* polymerase, were cloned into a pGEM®-T Easy Vector by Promega. Cloning reactions were optimized for each fragment according to the data given on Table 2.5 and transformed into HB101 using $CaCl_2$ methodology. Successful cloning of an insert in the pGEM®-T Easy Vectors interrupts the coding sequence of beta-galactosidase; recombinant clones can usually be identified by color screening on indicator plates containing X-Gal and IPTG. On average 20-30 colonies were selected for each cloning reaction and mini-preps of the colonies were screened with *Eco*RI digestion for all of the fragments except NS5A. An internal *Eco*RI site was detected in the NS5A fragment and the presence of the correct size insert was detected by *Not*I digestion for this fragment. During PCR amplification *Pfu* and Expand High Fidelity Enzyme System which contain *Taq* and *Pwo* DNA polymerases were used. *Pfu* and *Pwo* enzymes have proof reading activities and their mutation rate is lower than *Taq* DNA polymerase, but in order to check if the base changes are due to a PCR mistake or not, 2-4 colonies were selected for sequencing step of the study. 4 colonies from fragments Zeus, NS3, NS4 and NS5A, 3 colonies from fragments Charon and NS2, 2 colonies from fragment NS5A were positive and selected for large scale plasmid isolation. The clones carrying the right constructs were setup for 100mL cultures in LB supplemented with 100μg/mL ampicillin, and incubated overnight at 37°C. Plasmid isolation was then carried out using QIAfilter plasmid midi kit from Qiagen. Selected colonies were screened with *Eco*RI and *Not*I digestion once more for the correct size of the insert and results of

the midi-prep or mini-prep digestions are given on Figure 3.4 and 3.5. Size of the linearized vector without any insert is 3 kb. Insert sizes are 900 bp for Zeus (Panel A), 1950 bp for Charon (Panel B), 700 bp for NS2 (Panel C), 1980 bp for NS3 (Panel D), 1000 bp for NS4 (Panel E), 1500 bp for NS5A (Panel F) and 1800 bp for NS5B (Panel G). Extra band on Panel F is probably due to insufficient enzyme digestion and corresponds to linear vector with the insert.

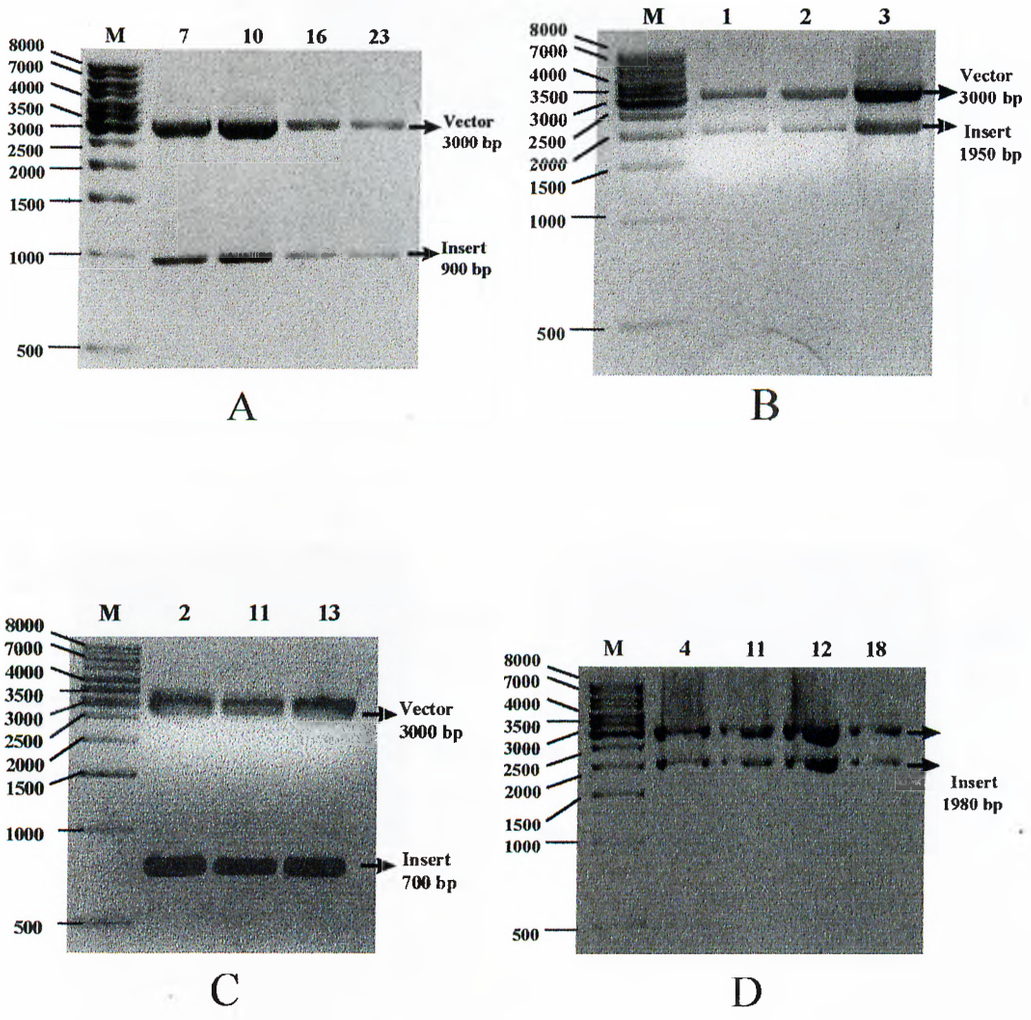
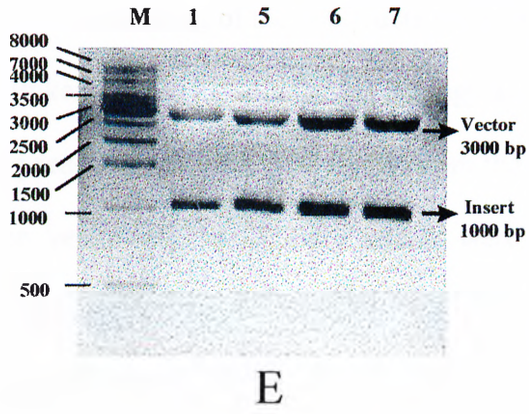
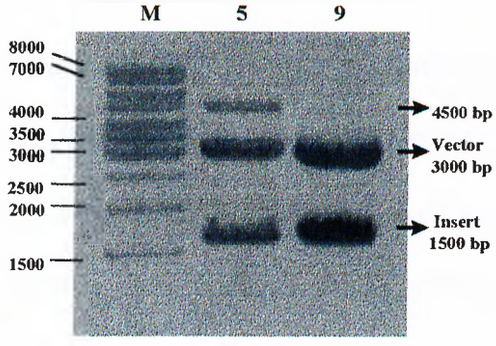


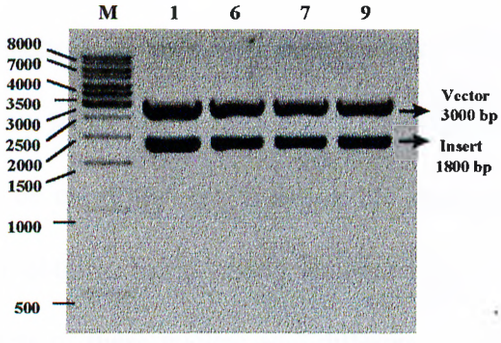
Figure 3.4: Agarose gel electrophoresis results of restriction enzyme digestion of selected colonies. DNA size marker is 1 kb DNA Ladder (MBI). Numbers on the lanes correspond to the colony numbers.



E



F



G

Figure 3.5: Agarose gel electrophoresis results of restriction enzyme digestion of selected colonies. DNA size marker is 1 kb DNA Ladder (MBI). Numbers on the lanes correspond to the colony numbers.

3.4 Sequencing of the Cloned Fragments

The automated sequencing of the constructs were carried out on an ABI Prism 310 Sequencer at Bilkent University, Molecular Biology and Genetics Department. In the first sequencing reaction M13 and T7 primers from the both sides of the multiple cloning site of the vector pGEM®-T Easy vector were used. Using Big Dye™ Terminator Cycle Sequencing Ready Reaction Kit and long capillary of the 310 Sequencer had enhanced the reading length and sequence quality. At the end of the first sequencing reaction fragments Zeus and NS2 were read completely. In order to read the sequences in the middle of long fragments internal primers were designed from the data obtained from the sequences read by M13 and T7 primers. At the end of these two sets of automated sequencing reactions partial sequence data was collected and analyzed by aligning with the previously submitted 1b sequences given on Table 3.1. Differences in the sequence data will be confirmed by repeating the sequencing reactions with further internal primers to obtain double strand sequencing and by sequencing the other selected clones. Effects of these nucleic acid differences on the amino acid sequence were analyzed. RNA and protein sequences of the Turkish isolate of HCV 1b, read by the end of the sequencing reactions, are given on Figure 3.6. Base changes that lead to amino acid changes are summarized below and at Table 3.2. Codon and base numbers are according to Turkish HCV 1b isolate.

3.4.1 ZEUS

Length of the fragment corresponding to 5'UTR and coding region of the Core protein is 881 bp. Sequence of the full length fragment was analyzed. In 5'-UTR there is an insertion of G residue at the site of 295th bp. and one base change from C to A at the site of 283th bp.

There are 10 base differences in the Core region that were not reported before and 2 of them results in amino acid differences. At codon 48, Ala which is a hydrophobic amino acid is changed into Pro, an amino acid with a cyclic side chain. Second amino acid shift is the base difference at codon 147. Leu, a hydrophobic amino acid, is changed into another a hydrophobic amino acid Val. Core is one of the most conserved regions in the sequenced part of the viral genome and both functional amino acid changes are specific to Turkish isolate.

3.4.2 CHARON

Total length of the fragment corresponding to the regions coding for E1, e2/p7 is 1854 bp. 5 sequencing reactions were set up for this fragment and only the reactions of M13 and T7 primers had worked. 248 bp from the 5' region and 509 bp form the 3' region of the fragment were analyzed. 23 base differences were observed, 16 of them are the base changes at conserved regions and 5 of them results in amino acid changes. At codon 78 hydrophobic amino acid Ala is present in previously submitted HCV 1b subtypes, whereas in Turkish HCV 1b isolate it is changed into acidic amino acid Asp. At codon 462 Asp or Asn is changed in to basic Histidine. These differences can be functional at protein level.

3.4.3 NS2

Length of the fragment coding NS2 protein is 651 bp. Full length of the region was sequenced by M13 and T7 primers. Sequence data for both strands of the region was available and possible changes was confirmed from both strands. 12 base differences were detected, 9 of them are in conserved regions and 8 of them leads to amino acid changes. Leu, a hydrophobic amino acid, became Ser which is a hydrophilic amino acid, at codon 37. Hydrophobic Val or Ile is present at codon 52 of other isolates, whereas it is changed into hydrophilic Thr in Turkish HCV 1b isolate. Another functional amino acid change from one aromatic amino acid Phe into a hydrophobic amino acid, Leu is observed at codon 129.

3.4.4 NS3

Fragment encoding NS3 protein is 1902 bp in length. 3 sequencing reactions with M13, T7 and an internal forward primer were set up for the NS3 fragment. A region corresponding to 1121 bp of the fragment; 843 bp from the 5' end and 278 bp form the 3' end, were analyzed. 26 base differences were observed, 20 of them take place at conserved residues and 14 of them results in 13 amino acid changes. 6 of theses amino acid changes can be classified as functional changes. Ile or Val at codon 71 are hydrophobic and changed into Thr, a hydrophilic molecule. Gly, a hydrophobic amino acid, at codons 239 and 237 is changed into basic amino acid Lys. An acidic amino acid, Asp at codon 276 is changed into an aromatic one, Tyr. At codon 590 cyclic amino acid, Pro became a basic amino acid, Arg.

3.4.5 NS4

The size of the NS4 fragment corresponding to the regions encoding NS4A and NS4B is 879 bp. Sequence data of the fragment except an 16 bp region was analyzed. 21 base changes, 16 at conserved residues, 5 at variable regions leading to amino acid changes, were detected. As a possible functional difference at codon 35, hydrophobic amino acid, Leu is changed into an aromatic amino acid, Phe.

3.4.6 NS5A

The sequence of the full length NS5A fragment, except a 10 bp region, was analyzed. 48 base changes were detected and 18 of them lead to 16 amino acid changes. 43 of the base changes are at conserved residues whereas 5 of them are at the variable sites. An amino acid change at codon 58 from Pro to Ser leads to a change from an amino acid with a cyclic side chain to a hydrophilic one and can be important for the protein structure. At codon 107 a hydrophilic amino acid Ser is changed into hydrophobic Thr. At codon 123 basic amino acid Arg is changed into Pro, a cyclic one and at codon 198, hydrophobic Val became hydrophilic Thr. Codons 204 and 255 encompass amino acid changes from hydrophilic Thr to hydrophobic Ile and vice versa respectively. At codon 309 aromatic Phe is changed into hydrophobic Thr. Other hydrophilic to hydrophobic amino acid changes were observed at codons 360 and 410, from Thr to Ala and Val to Ala respectively. Hydrophilic Ser is changed into cyclic amino acid Pro. As described before, ISDR is located in NS5A protein and changes in this region, that were not reported previously, might be important in future studies on IFN treatment. As shown in Figure 3.5, the green highlighted region which shows the ISDR location in Turkish HCV 1b isolate, encompasses 4 base changes and two of them effecting the same codon 255, lead to

a change from hydrophobic Ile to hydrophilic Thr. After confirming the base change, this functional amino acid difference should be analyzed in detail.

3.4.7 NS5B

The length of the fragment corresponding to the region encoding NS5B protein is 1784 bp. The sequence data of a 985 bp region, corresponding to 485 bp from the 5' end and 500 bp form the 3' end, was analyzed. There are 28 base changes and 22 of them are from conserved regions. 11 of them lead to 8 amino acid changes. At codon 132, 2 base changes lead to a change from hydrophilic Thr to hydrophobic Val. Acidic Glu was changed into Arg due to the difference in 3 of the bases corresponding to that region. Codon 146 encompasses a change from basic Lys to hydrophobic Ile. At codon 539, basic Arg is present instead of acidic Gln .

Figure 3.6 (See the end of the figure for the legend)

(From 36th bp of 5'UTR)
ATCACTCCCCTGTGAGGAACTACTGTCTTCACGCAGAAAGCGTCTAGCCATGG
CGTTAGTATGAGTGTCGTGCAGCCTCCAGGACCCCCCCTCCCGGGAGAgCCAT
agTggTCTGCGGAACCGGTGAGTACACCGGAATTGCCAGGACGACCGGgTCCTT
TCTTGGATCAACCCGCTCAATGCCTGGAGATttgggCGTGCCCCCGCGAGACTGC
TAGCCGAGTAGTGTTGGGTCGCGAAAGGCCTTGTGGTACTGCCTGATAGGGTC
TTGCGAGTGCCCAggGAGGTCTcGGTAgACCGTGCACC---(CORE)---aTGAGCAC
M S
GAATCCTaAACCTCAaaGAAAAACCAAACGTAACACCAACCGCCGGcCACAGGA
T N P K P Q R K T K R N T N R R P Q

CGTTAAGTTCCCGGGCGGTGGTCAGAtCGTTGGTGGAGTTtACtTGTTGCCgCGC
D V K F P G G G Q I V G G V Y L L P R

AgGGGcCCCAGgtTGGGTGtGCgCCCgACTAgGAAgACtTCCgAgCGgtcgAAACCTc
R G P R L G V R P T R K T S E R S K P

GTGgAAgGcGacaACCTATcCCcAaGGcTCGcCAgCCCgAgGGCAgGGCCTGGgCacA
R G R R Q P I P K A R Q P E G R A W A

gCCcgGGTACCCTTGGCCCCTCTATGGCAATGAGGGCTTGGGGTGGGCAGGAT
Q P G Y P W P L Y G N E G L G W A G

GGCTCCTGTCACCCCGTGGCTCTCGGCCTAGTTGGGGCCCCACGGACCCCCGG
W L L S P R G S R P S W G P T D P R

CGTAGGTCGCGTAACTTGGGTAAGGTCATCGATACCCTCACATGCGGCTTCGC
R R S R N L G K V I D T L T C G F

CGACCTCATGGGGTATATTCCGCTTGTCGGCGCCCCCGTGGGGGGCGCTGCCA
A D L M G Y I P L V G A P V G G A A

GGGCCCTGGCGCATGGCGTCCGGGTTCTGGAGGACGGCGTGAATTATGCAAC
R A L A H G V R V L E D G V N Y A

AGGGAATTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTTGCTATCTTG
T G N L P G C S F S I F L L A L L S

TCTGACCATCCCTGCTTCCGCT---(E1)---TATGAGGTGCGCAACGTGTCCGGGG
C L T I P A S A---------- Y E V R N V S G

gGTACCATGTCACGAACGACTGCTCCAACTCAAGCATaGTGTATGAGGCAGCG
A Y H V T N D C S N S S I V Y E A A

GACATGATCATGCACACCCCCGGGTGCGTGCCCTGCGTTCGGGAGAACAACTC
D M I M H T P G C V P C V R E N N

CTCCCGCTGCTGGGTgGCGCTCACTCCCACGCTtGCGGCTAGGAATGGCAGCGT
S S R C W V A L T P T L A A R N G S

Figure 3.6 Contd.

CCCCACTACGGCAATACGACGCCATGTCaACTTGCTCGTTGGGGCGGATACCT
V P T T A I R R H V N L L V G A D T

TCTGCTCCGC---(ENV1-2, 1098 bp)---TCAGCGCTGCATGCAATTGGACTCGGGG
F C S----------S A A C N W T R

AGAGCGTTGTCACTTGGAGGACAGGGACAGATCGGAGCTCAGCCCGCTGCTG
G E R C H L E D R D R S E L S P L L

CTGTCTACAACAGAGTGGCAAGTGTTGCCCTGTTCCTTCACCACCTTGCCCGC
L S T T E W Q V L P C S F T T L P

CCTGTCCACTGGTTTGATCCATCTACATCAGAACATCGTGGACGTGCAATACC
A L S T G L I H L H Q N I V D V Q Y

TGTACGGTATAGGGTCGGCGGTTGTCTCCTTTGCCATCAAATGGGAGTATATC
L Y G I G S A V V S F A I K W E Y I

CTGCTGCTCTTCCTCCTCCTGGCGGACGCGCGAGTCTGCGCTTGCCTGTGGAT
L L L F L L L A D A R V C A C L W

GATGCTGCTGATAGCCCAGGCTGAGGCCGCCCTAGAGAACCTGGTGGTCCTCA
M M L L I A Q A E A A L E N L V V L

ATGCGGCGTCCGTGGCCGGGGCACACGGCATTCTCTCCTTCCTTGTGTTCTTC
N A A S V A G A H G I L S F L V F F

TGTGCTGCCTGGTACATCAAGGGCAGGCTGGTGGGTGGGGCGGCATACGCCC
C A A W Y I K G R L V G G A A Y A

TGTACGGCGTGTGGCCGCTGCTCCTGCTCCTGCTGGCgTTACCACCGCGAGCA
L Y G V W P L L L L L L A L P P R A

TATGCC---(NS2)---ATGGACCGGGAGATGGCTGCATCGTGTGGAGGCGCGGTTT
Y A---------- M D R E M A A S C G G A V

TTGTAGGTCTGGCAATCTTgacCTTGTCACCGCACTATAAAGTGCTCCTAGCAA
F V G L A I L T L S P H Y K V L L A

AGCTCATATGGTGGTCACAGTACTTTATCACCAGGGCCGAGGCGCATGTGCAA
K L I W W S Q Y F I T R A E A H V Q

GTGTGGACCCCCCCCCTCAACGTTCGGGGGGGCCGCGATGCCATCATCCTCCT
V W T P P L N V R G G R D A I I L

CACTTGCGCGGTCCACCCAGAGCTAATCTTTGAAATCACCAAAATTTTGCTCG
L T C A V H P E L I F E I T K I L L

CCGTGCTCGGTCCGCTCATGGTGCTCCAGGCTGGCATAACCAgAGTGCCGTAC
A V L G P L M V L Q A G I T R V P Y

Figure 3.6 Contd.

Figure 3.6 Contd.

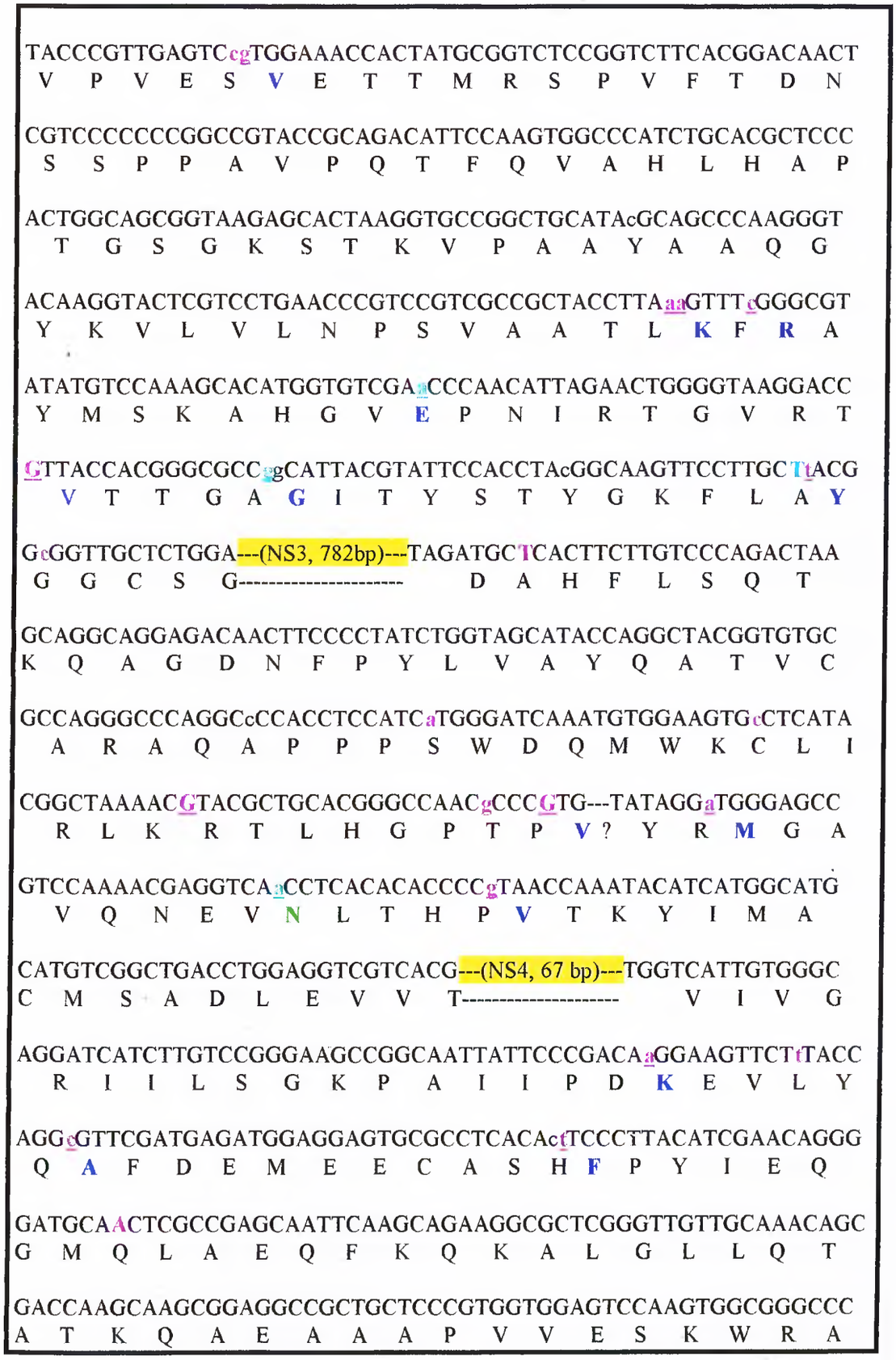


Figure 3.6 Contd.

```
TTGAGACCTTTTGGGCGAAGCACATGTGGAACTTCATtAGCGGGATACAGTACTT
  L   E   T   F   W   A   K   H   M   W   N   F   I   S   G   I   Q   Y

AGCAGGCTTGTCCACTCTGCCTGGGAACCCCGCAATAGCATCACTGATGGCATT
L   A   G   L   S   T   L   P   G   N   P   A   I   A   S   L   M   A

TACAGCCTCTATCACTAGCCCGCTCACCACTCAACACACCCTCCTGTTTAACATC
F   T   A   S   I   T   S   P   L   T   T   Q   H   T   L   L   F   N   I

TTGGGGGGATGGGTaGCCGCCCAACTCGCaCCCCCCAGCGCCGCTTC-(NS4, 16bp)
  L   G   G   W   V   A   A   Q   L   A   P   P   S   A   A ----------

GGCATCGCAGGTGCGGCTGTTGCCAGCATAGGCCTTGGGAAGGTGCTTGTGGAC
   G   I   A   G   A   A   V   A   S   I   G   L   G   K   V   L   V   D

ATCCTGGCGGGCTATGGAGCGGGGGTGGCAGGCGCGCTCGTGGCCTTCAAGGTC
   I   L   A   G   Y   G   A   G   V   A   G   A   L   V   A   F   K   V

ATGAGTGGCGAGATGCCCTCTACCGAGGATCTGGTCAACTTACTCCCAGCTATC
   M   S   G   E   M   P   S   T   E   D   L   V   N   L   L   P   A   I

CTCTCTCCTGGTGCCCTGGTCGTCGGAGTCGTGTGCGCAGCAATACTGCGGCGG
   L   S   P   G   A   L   V   V   G   V   V   C   A   A   I   L   R   R

CATGTGGGCCCAGGAGAGGGGGCGGTACAATGGATGAACCGGCTGATAGCGTT
   H   V   G   P   G   E   G   A   V   Q   W   M   N   R   L   I   A

CGCTCCACGGGGTAACCACGTTCCCCCCACGCACTAtGTGCCTGAGAGCGACGCT
F   A   P   R   G   N   H   V   S   P   T   H   Y   V   P   E   S   D   A

GCAGCGCGTGTCACACAGATCCTCTcTAGCCTTACCATCACTCAGCTGCTGAAGA
   A   A   R   V   T   Q   I   L   S   S   L   T   I   T   Q   L   L   K

GGCTCCACCAGTGGATTAATGAgGACTGCTCCACGCCATGT--------(NS5A)---------
  R   L   H   Q   W   I   N   E   D   C   S   T   P   C----------

tCCGGCTCGTGGCTAAGAGACATCTGGGACTGGATATGCACGGTGTTGTCTGACT
   S   G   S   W   L   R   D   I   W   D   W   I   C   T   V   L   S   D

TCAAGACCTGGCTCCAGTCCAAGCTTCTGCCGCGGTTACCGGGAGTCCCCTTCT
   F   K   T   W   L   Q   S   K   L   L   P   R   L   P   G   V   P   F

TCTCATGCCAACGTGGGTATAAGGGAGTCTGGCGCGGAGACGGCATCATGCACA
   F   S   C   Q   R   G   Y   K   G   V   W   R   G   D   G   I   M   H

CCACCTGTTCATGTGGGGCACAGATCACCGGACATGTCAAAAACGGTTCCATGA
   T   T   C   S   C   G   A   Q   I   T   G   H   V   K   N   G   S   M

GGATCGTTGGGCCAAAAACCTGCAGtaACACGTGGCATGGAACATTTCCcATCAA
   R   I   V   G   P   K   T   C   S   N   T   W   H   G   T   F   P   I
```

Figure 3.6 Contd.

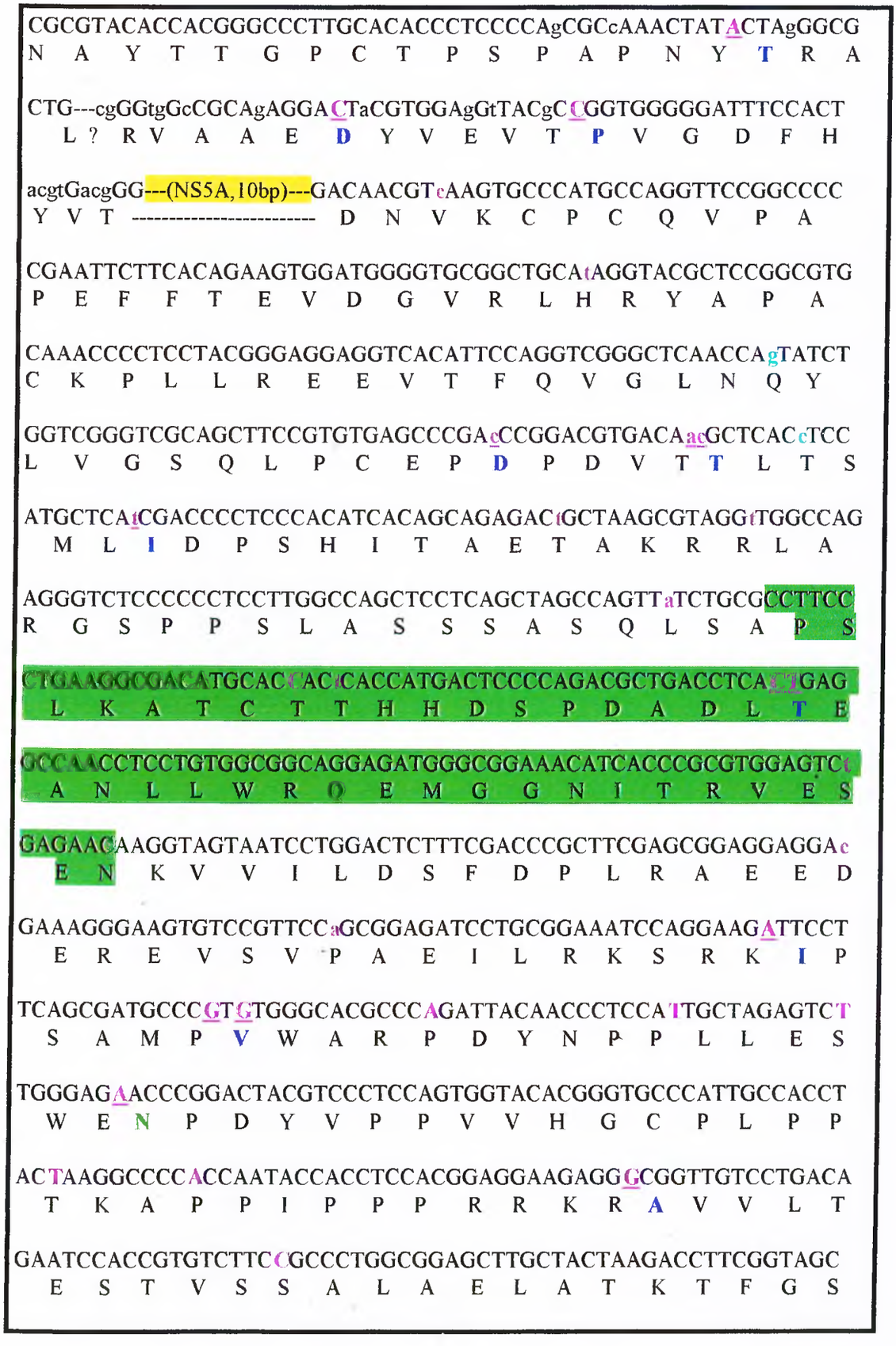


Figure 3.6 Contd.

Figure 3.6 Contd.

TCAAtagGGTGGCTTCATGCCTCAGGAAACTTGGGGTACCGCCCTTGCGAGTCTG
I N R V A S C L R K L G V P P L R V

GAGACATCGGGCCAGAAGTGTCCGCGCTAAGCTACTGTCTCAGGGGGGGAGGG
W R H R A R S V R A K L L S Q G G R

CTGCCACATGTGGTAGATACCTCTTCAACTGGGCAGTAAGGACCAAGCTCAAAC
A A T C G R Y L F N W A V R T K L K

TCACTCCAATTCCGGCTGCATCTCGGTTGgACTTG---AN-TGGTTTGTTGCTGGTT
L T P I P A A S R L D L-------- W F V A G

ACAGCGGGGGAGACATATATCACAGCCTGTCTCGTGCCCGACCCCGCTGGTTTA
Y S G G D I Y H S L S R A R P R W F

TGTTGTGCCTCTTCCTACTTTCCGTAGGGGTAGGCATCTATCTGCTCCCCAACCG
M L C L F L L S V G V G I Y L L P N

ATGA--- (3'UTR)--- ACGGGGAGCCAACCACTCCAGG
R *----------

Figure 3.6: RNA and amino acid sequences of the Turkish isolate of the HCV subtype 1b. Missing bases are shown with dashes. Numbers and fragment names in the brackets correspond to the missing bases in side those fragments.

Underlined bases correspond to the base changes that cause amino acid differences.

Green highlighted region : ISDR.

Yellow highlighted regions : Strating point of fragments.

X : Base changes on conserved regions, detected only in Turkish HCV isolate.

X : Base changes on variable sites.

Y : Amino acid changes on conserved regions.

Y : Amino acid changes on variable regions.

Region	Variable Base	Variable Codon	Amino Acid Change	HCV 1b Turkey	HCV 1b Isolates									
					AF05424 7	AF05424 8	AF05424 9	HCU 45476	HPCK1 R1	HPCK1R 2	HPCK1R 3	HPCK1S 1	HPCK1 S2	HPCK1 S3
5'-UTR	283	♦	♦	A	C	C	C	C	C	C	C	C	C	C
	295 (Insertion)	♦	♦	G	—	—	—	—	—	—	—	—	—	—
Core	450	48	A ► P	CCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG	GCG
	738	147	L ► V	GTG	CTG	CTG	CTG	CTG	TTG	TTG	TTG	TTG	TTG	CTG
E1 +E2	26	9	I/V/L ► A	GCG	ATA	ATA	ATA	GTG	CTG	GTG	GTG	CTG	GTG	GTG
	214	72	D ► N	AAC	GAC	GAC	GAC	GAT	GAT	GAT	GAT	GAT	GAC	GAT
	233	78	A ► D	GAT	GCT	GCT	GCT	GCT	GCT	GCT	GCC	GCT	GCT	GCC
	1349	450	N/T/K ► S	AGC	AAT	AAT	AAT	AAC	ACC	AAC	AAA	ACC	AAC	AAA
	1384	462	N/D ► H	CAC	AAC	AAC	AAC	GAC	GAT	GAC	AAC		GAC	AAC
NS2	110	37	L ► S	TCA	TTA	TTA	TTA	TTA	TTG	TTA	TTG	TTA	TTA	TTA
	142	48	M/L ► V	GTG	ATG	ATG	ATG	TTG	TTG	CTG	TTG	TTG	TTG	TTG
	155	52	V/I ► T	ACC	GTC	GTC	GTC	ATC	ATC	ATC	ATC	ATC	ATC	ATC
	257	85	I ► V	GTG	ATA	ATA	ATA	ATA	ATA	ATA	ATA	ATA	ATA	ATA
	356	119	V ► A	GCC	GTC	GTC	GTC	GTC	GTC	GTC	GTC	GTC	GTC	GTC
	384	129	F ► L	CTC	TTC	TTC	TTC	TTC	TTC	TTC	TTC	TTC	TTC	TTC
	418	140	V ► L	CTT	GTT	GTT	GTT	GTT	GTC	GTC	GTC	GTC	GTT	GTT
	497	166	V ► I	ATC	GTC	GTC	GTC	GTT	GTC	GTC	GTC	GTC	GTC	GTC
NS3	212	71	I/V ► T	ACC	ATC	ATC	ATC	GTC	ATC	ATC	ATC	ATC	ATC	ATC
	523	175	M ► V	GTG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG	ATG
	709	237	G ► K	AAG	GGG	GGG	GGG	GGC	GGG	GGC	GGA	GGG	GGC	GGA
	710	237	G ► K	AAG	GGG	GGG	GGG	GGC	GGG	GGC	GGA	GGG	GGC	GGA
	715	239	G ► K	CGG	GGG	GGG	GGG	GGG	GGG	GGG	GGG	GGG	GGG	GGG
	747	249	D ► E	GAA	GAC	GAC	GAC	GAC	GAT	GAC	GAC	GAT	GAC	GAC
	775	259	I ► V	GTT	ATT	ATT	ATT	ATC	ATC	ATC	ATC	ATC	ATC	ATC
	790	264	S/P ► G	GGC	TCC	TCC	TCC	CCT	CCC	CCT	CCC	CCC	CCT	CCT
	826	276	D ► Y	TAC	GAC	GAC	GAC	GAC	GAC	GAT	GAT	GAC	GAT	GAT
	1769	590	P ► R	CGT	CCT	CCT	CCT	CCT	CCT	CCT	CCT	CCT	CCT	CCT
	1792	598	L ► V	GTG	CTG	CTG	CTG	CTG	CTC	CTG	CTG	CTG	CTG	CTG
	1804	602	L ► M	ATG	CTA	CTA	CTA	CTA	CTA	CTA	CTA	CTA	CTG	CTA
	1829	610	I/T ► N	AAC	ATC	ATC	ATC	ACC	ACC	ATC	ACC	ACC	ATC	ATA
	1843	615	I ► V	GTA	ATA	ATA	ATA	ATC	ATA	ATA	ATA	ATA	ATA	ATA
NS4	123	18	R ► K	AAG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG
	140	24	E ► A	GCG	GAG	GAG	GAG	GAG	GAG	GAG	GAG	GAG	GAG	GAG
	172	35	L ► F	TTC	CTT	CTT	CTT	CTC	CTC	CTT	CTC	CTC	CTT	CTC
	548	154	G ► A	GCC	GGC	GGC	GGC	GGC	GGC	GGC	GGC	GGC	GGC	GGC
	799	245	S ► P	CCA	TCG	TCG	TCG	TTC	TCG	TCG	TCG	TCG	TCG	TCG
NS5A	22	8	V ► I	ATC	GTT	GTT	GTT	GTT	GTT	GTC	GTT	GTT	GTC	GTT
	172	58	P ► S	TCA	CCA	CCA	CCA	CCA	CCA	CCA	CCC	CCC	CCA	CCC
	319	107	S ► T	ACT	TCA	TCA	TCA	TCA	TCT	TCC	TCC	TCT	TCC	TCC
	351	117	E ► D	GAC	GAG	GAG	GAG	GAG	GAG	GAG	GAG	GAG	GAG	GAG
	368	123	R ► P	CCG	CGT	CGT	CGT	CGG	CGG	CGG	CGG	CGG	CGG	CGG
	579	193	E ► D	GAC	GAA	GAA	GAA	GAA	GAA	GAA	GAA	GAA	GAA	GAA
	592	198	V ► T	ACG	GTG	GTG	GTG	GTG	GTG	GTG	GTG	GTG	GTG	GTG
	593	198	V ► T	ACG	GTG	GTG	GTG	GTG	GTG	GTG	GTG	GTG	GTG	GTG
	611	204	T ► I	ATC	ACC	ACC	ACC	ACC	ACC	ACC	ACC	ACC	ACC	ACC
	764	255	I ► T	ACT	ATC	ATC	ATC	ATC	ATC	ATC	ATC	ATC	ATC	ATC
	765	255	I ► T	ACT	ATC	ATC	ATC	ATC	ATC	ATC	ATC	ATC	ATC	ATC
	925	309	F ► I	ATT	TTC	TTC	TTC	TTT	TTT	TTC	TTC	TTT	TTC	TTC
	943	315	I ► V	GTG	ATA	ATA	ATA	ATA	ATA	ATA	ATA	ATA	ATA	ATA
	945	315	I ► V	GTG	ATA	ATA	ATA	ATA	ATA	ATA	ATA	ATA	ATA	ATA
	991	331	D/A ► N	AAC	GAC	GAC	GAC	GAC	GAC	GAC	GAC	GAC	GAC	GCC
	1078	360	T ► A	GCG	ACG	ACG	ACG	ACG	ACG	ACG	ACG	ACG	ACG	ACG
	1201	401	S ► P	CCC	TCC	TCC	TCC	TCC	TCC	TCC	TCC	TCC	TCC	TCC
	1229	410	N ► A	GCT	GTT	GTT	GTT	GTT	GTT	GTT	GTT	GTT	GTT	GTT
NS5B	391	130	D ► E	GAG	GAC	GAC	GAC	GAC	GAT	GAC	GAT	GAT	GAC	GAT
	395	132	T ► V	GTC	ACC	ACC	ACC	ACC	ACC	ACC	ACC	ACC	ACC	ACC
	396	132	T ► V	GTC	ACC	ACC	ACC	ACC	ACC	ACC	ACC	ACC	ACC	ACC
	434	145	E ► R	CGA	GAG	GAG	GAG	GAG	GAG	GAG	GAG	GAG	GAG	GAG
	435	145	E ► R	CGA	GAG	GAG	GAG	GAG	GAG	GAG	GAG	GAG	GAG	GAG
	436	145	E ► R	CGA	GAG	GAG	GAG	GAG	GAG	GAG	GAG	GAG	GAG	GAG
	439	146	K ► I	ATT	AAG	AAG	AAG	AAA	AAA	AAG	AAA	AAA	AAG	AAA
	468	156	V ► G	GGG	GTA	GTA	GTA	GTG	GTG	GTG	GTA	GTG	GTG	GTT
	1290	430	A ► V	GTC	GCT	GCT	GCT	GCC	GCC	GCG	GCT	GCC	GCG	GCT
	1617	539	Q ► R	CGG	CAG	CAG	CAG	CAG	CAG	CAG	CAG	CAG	CAG	CAG
	1715	572	L ► F	TTC	CTC	CTC	CTC	CTC	CTC	CTC	CTC	CTC	CTC	CTC

Table 3.2: Summary of codon differences between the HCV 1b Turkish isolate and other characterized HCV 1b genomes. X: Differences on the conserved regions

4. DISCUSSION

The aim of this study was to characterize a Hepatitis C Virus genome isolated from a Turkish patient by cloning and sequencing the complete genome.

As an RNA virus, HCV genome is highly variable. This is due to the lack of proofreading mechanism of the RNA polymerase of the virus that is involved in the replication of the viral genome. The mutation rate throughout the genome differs between the regions. The most variable regions are the immunogenic regions so that the virus can escape from the immune pressure due to the differences in these sites. Soon after the first complete genome sequence of HCV was reported (Choo *et al.*, 1991) different isolates all around the world were submitted. Several subtypes of HCV have been found in different geographic areas (Bukh *et al.*, 1993). Today the most widely accepted classification of HCV genotypes is that put forward by Simmonds *et al.* in 1994, which allows the definition of six virus types and a great number of subtypes. The nucleotide sequence identity between types is around 66-69% and that between subtypes of the same type 77-80%, whereas an identity of 91-98% is found between sequences of the same virus subtype (Simmonds *et al.*, 1995). HCV subtype 1b is the predominant genotype found in Turkey. To investigate more fully its evolutionary relationship with other genotypes of HCV and to study its overall genome organization, the genome sequence, encompassing the partial coding region and non coding regions of the genotype 1b obtained from a HCV infected Turkish patient, was determined.

The most important parameter in this study is the reliability of reported base differences. In previously done cloning experiments, many non-functional isolates

were reported due to PCR mistakes or cloning artifacts. During PCR reactions, DNA polymerase enzymes are capable of incorporating wrong bases. Many commercially available enzymes have proofreading activity to reduce the PCR base mutation rate. In this study a combination of two enzymes *Taq* and *Pwo* DNA Polymerases were used as a part of a commercial PCR kit. *Pwo* has proof reading activity whereas *Taq* is necessary to increase the fragment length. Using an enzyme with the proofreading activity reduced the probability of a misincorporation considerably but did not eliminate the problem completely. Also in a study it was reported that during cloning reactions, nonfunctional genomes are selected over functional ones which indicates that during replication in the bacteria, mutated fragments are favored (Forns *et al.*, 1997). In order to eliminate this problem it was decided to choose more than one colony for each fragment. A possible base change that results in an amino acid change should be confirmed by sequencing the other colonies of that region. If same base change is reported in the sequences of other colonies, this increases the chance of a real base change rather than a PCR-based mutation or a cloning artifact. The resulting sequence represents the predominant genome of the virus that is present in the patient's serum.

In this study sequence of the 70% of the viral genome is presented. Blast search of the available sequences have shown that clones that were sequenced have the correct inserts and the viral genome is clearly subtype 1b. According to the sequencing results given in Figure 3.6, there are total of 168 base differences and 139 of them were on the conserved regions of other sequenced isolates. 63 of 168 base differences cause 56 amino acid changes in the protein sequence. There is only one insertion in the 5'-UTR which does not effect the reading frame of the polyprotein. Considering the sequenced portion of the genome, 5'-UTR and Core are regions

showing the highest homology whereas NS5A encounters most of the base changes. In the ISDR there is an amino acid change which may effect the protein structure but since the sequence data is not confirmed exactly, the reliability of the mutation is not known. Considering the high mutation rate of the viral genome, a high number of base differences were expected but these differences should not interfere with the protein function, either by being silent mutations or causing changes to similar amino acids. Approximately 64% of the base changes in the sequenced region of the HCV 1b Turkish isolate do not effect the amino acid sequence and approximately 43% of the amino acid changes are the variations between different types of amino acids. These data increase the possibility that the sequencing data is reliable and the fragments may represent a functional viral genome.

5. PERSPECTIVE

At the end of the first part of the project, 6652 bases of the 9.5 kb viral genome were analyzed. In order to complete the sequences of the missing regions of the genome, more internal primers should be designed and sequencing reactions will be repeated. According to the sequencing data of this part of the genome, first Turkish isolate of HCV 1b resembles previously submitted HCV 1b isolates but contains several amino acid changes on the conserved residues of the 10 clones given on Table 3.1. Sequencing the regions with the specific differences should be repeated with internal primers and base changes should be confirmed by reading the double strand of the variable regions in order to be sure that the differences are not the sequencing artifacts. Also in order to eliminate the PCR artifact probability other

selected colonies should be sequenced in order to see the same variations. After completion of the nucleic acid sequencing and confirmation of mutations on different clones, the exact structure of the Turkish 1b genotype will be known. Further studies will be needed to test whether the significant mutations are a common future of Turkish 1b genotype, rather than changes associated to the isolate that we studied.

If there are specific changes associated to Turkish 1b genotypes, such changes should be analyzed in terms of viral pathogenecity, immune response and viral therapy. As we obtained viral genome fragments each corresponding to a distinct protein-coding region of the virus, these fragments are valuable tools for further studies, including but not limited to the production of recombinant proteins as candidate antigens for immunodiagnosis and vaccine production. The same fragments can also be inserted into mammalian expression vectors to perform molecular studies related to different aspects of virus-associated cellular events, such as the elucidation of viral entry into the host cells, viral replication as well as phenotypic changes in host cells some of which may shed further light onto the suspected role of viral protein in malignant transformation of hepatocytes.

6. REFERENCES

Aizaki, H. , Aoki, Y., Harada, T., Ishii, K., Suzuki, T., Nagamori, S., Toda, G., Matsuura, Y., and Miyamura, T. (1998). Full-length complementary DNA of hepatitis C virus genome from an infectious blood sample. Hepatology. **27**(2):621-7.

Ali, N. and A. Siddiqui (1995). "Interaction of polypyrimidine tract-binding protein with the 5' noncoding region of the hepatitis C virus RNA genome and its functional requirement in internal initiation of translation." J Virol **69**(10): 6367-75.

Asabe, S. I., Y. Tanji, S. Satoh, T. Kaneko, K. Kimura and K. Shimotohno (1997). "The N-terminal region of hepatitis C virus-encoded NS5A is important for NS4A-dependent phosphorylation." J Virol **71**(1): 790-6.

Ausubel F. M. (1987). Current protocols in molecular biology. New York, Published by Greene Pub. Associates and Wiley-Interscience : J. Wiley.

Behrens, S. E., L. Tomei and R. De Francesco (1996). "Identification and properties of the RNA-dependent RNA polymerase of hepatitis C virus." Embo J **15**(1): 12-22.

Borowski, P., K. Oehlmann, M. Heiland and R. Laufs (1997). "Nonstructural protein 3 of hepatitis C virus blocks the distribution of the free catalytic subunit of cyclic AMP-dependent protein kinase." J Virol **71**(4): 2838-43.

Bradley, D. W., M. J. Beach and M. A. Purdy (1992). "Recent developments in the molecular cloning and characterization of hepatitis C and E viruses." Microb Pathog **12**(6): 391-8.

Bukh, J., Purcell, R. H., and Miller, R. H. (1993). At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide. Proc Natl Acad Sci U S A. **90**(17):8234-8.

Bukh, J., R. H. Purcell and R. H. Miller (1994). "Sequence analysis of the core gene of 14 hepatitis C virus genotypes." Proc Natl Acad Sci U S A **91**(17): 8239-43.

Buratti, E., S. Tisminetzky, M. Zotti and F. E. Baralle (1998). "Functional analysis of the interaction between HCV 5'UTR and putative subunits of eukaryotic translation initiation factor eIF3." Nucleic Acids Res **26**(13): 3179-87.

Butkiewicz, N. J., M. Wendel, R. Zhang, R. Jubin, J. Pichardo, E. B. Smith, A. M. Hart, R. Ingram, J. Durkin, P. W. Mui, M. G. Murray, L. Ramanathan and B. Dasmahapatra (1996). "Enhancement of hepatitis C virus NS3 proteinase activity by association with NS4A-specific synthetic peptides: identification of

sequence and critical residues of NS4A for the cofactor activity." Virology **225**(2): 328-38.

Caselmann, W. H. and M. Alt (1996). "Hepatitis C virus infection as a major risk factor for hepatocellular carcinoma." J Hepatol **24**(2): 61-6.

Chamberlain, R. W., Adams, N., Saeed, A. A., Simmonds, P., and Elliott, R. M. (1997). Complete nucleotide sequence of a type 4 hepatitis C virus variant, the predominant genotype in the Middle East. J Gen Virol. **78**(Pt 6):1341-7.

Chamberlain, R. W., Adams, N. J., Taylor, L. A., Simmonds, P., and Elliott, R. M. (1997). The complete coding sequence of hepatitis C virus genotype 5a, the predominant genotype in South Africa. Biochem Biophys Res Commun. **236**(1):44-9.

Chen, C. M., L. R. You, L. H. Hwang and Y. H. Lee (1997). "Direct interaction of hepatitis C virus core protein with the cellular lymphotoxin-beta receptor modulates the signal pathway of the lymphotoxin-beta receptor." J Virol **71**(12): 9417-26.

Choo, Q. L., G. Kuo, A. J. Weiner, L. R. Overby, D. W. Bradley and M. Houghton (1989). "Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome." Science **244**(4902): 359-62.

Choo, Q. L., K. H. Richman, J. H. Han, K. Berger, C. Lee, C. Dong, C. Gallegos, D. Coit, R. Medina-Selby, P. J. Barr and *et al.* (1991). "Genetic organization and diversity of the hepatitis C virus." Proc Natl Acad Sci U S A **88**(6): 2451-5.

Chouki, A., S. Ung, C. Wychowski and J. Dubuisson (1998). "Involvement of endoplasmic reticulum chaperones in the folding of hepatitis C virus glycoproteins." J Virol **72**(5): 3851-8.

Chung, R. T., Kaplan, L.M. (1992). "Isolation and characterization of an HCV-specific RNA-dependent RNA polymerase activity from extracts of infected liver tissue.In Houghton, M., Bonino, F., D'Aquino, M., (eds.)." "Hepatitis C Virus and Related Viruses. Molecular Virology and Pathogenesis.", First annual meeting Venice, July, 1992. **Abstract A15**: p21.

Cocquerel, L., J. C. Meunier, A. Pillez, C. Wychowski and J. Dubuisson (1998). "A retention signal necessary and sufficient for endoplasmic reticulum localization maps to the transmembrane domain of hepatitis C virus glycoprotein E2." J Virol **72**(3): 2183-91

Deleersnyder, V., A. Pillez, C. Wychowski, K. Blight, J. Xu, Y. S. Hahn, C. M. Rice and J. Dubuisson (1997). "Formation of native hepatitis C virus glycoprotein complexes." J Virol **71**(1): 697-704.

Failla, C., L. Tomei and R. De Francesco (1995). "An amino-terminal domain of the hepatitis C virus NS3 protease is essential for interaction with NS4A." J Virol **69**(3): 1769-77.

Fishman, J. A., R. H. Rubin, M. J. Koziel and B. J. Periera (1996). "Hepatitis C virus and organ transplantation." Transplantation **62**(2): 147-54.

Forns, X., J. Bukh, R. H. Purcell and S. U. Emerson (1997). "How Escherichia coli can bias the results of molecular cloning: preferential selection of defective genomes of hepatitis C virus during the cloning procedure." Proc Natl Acad Sci U S A **94**(25): 13909-14.

Fujita, T., S. Ishido, S. Muramatsu, M. Itoh and H. Hotta (1996). "Suppression of actinomycin D-induced apoptosis by the NS3 protein of hepatitis C virus." Biochem Biophys Res Commun **229**(3): 825-31.

Fukuma, T., N. Enomoto, F. Marumo and C. Sato (1998). "Mutations in the interferon-sensitivity determining region of hepatitis C virus and transcriptional activity of the nonstructural region 5A protein." Hepatology **28**(4): 1147-53.

Grakoui, A., D. W. McCourt, C. Wychowski, S. M. Feinstone and C. M. Rice (1993a). "A second hepatitis C virus-encoded proteinase." Proc Natl Acad Sci U S A **90**(22): 10583-7.

Grakoui, A., C. Wychowski, C. Lin, S. M. Feinstone and C. M. Rice (1993b). "Expression and identification of hepatitis C virus polyprotein cleavage products." J Virol **67**(3): 1385-95.

Grakoui, A., D. W. McCourt, C. Wychowski, S. M. Feinstone and C. M. Rice (1993c). "Characterization of the hepatitis C virus-encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites." J Virol **67**(5): 2832-43.

Hahm, B., D. S. Han, S. H. Back, O. K. Song, M. J. Cho, C. J. Kim, K. Shimotohno and S. K. Jang (1995). "NS3-4A of hepatitis C virus is a chymotrypsin-like protease." J Virol **69**(4): 2534-9.

Heinz, F. X. (1992). "Comparative molecular biology of flaviviruses and hepatitis C virus." Arch Virol Suppl **4**: 163-71.

Hijikata, M., N. Kato, Y. Ootsuyama, M. Nakagawa and K. Shimotohno (1991). "Gene mapping of the putative structural region of the hepatitis C virus genome by in vitro processing analysis." Proc Natl Acad Sci U S A **88**(13): 5547-51.

Hijikata, M., H. Mizushima, T. Akagi, S. Mori, N. Kakiuchi, N. Kato, T. Tanaka, K. Kimura and K. Shimotohno (1993). "Two distinct proteinase activities required for the processing of a putative nonstructural precursor protein of hepatitis C virus." J Virol **67**(8): 4665-75.

Hirota, M., S. Satoh, S. Asabe, M. Kohara, K. Tsukiyama-Kohara, N. Kato, M. Hijikata and K. Shimotohno (1999). "Phosphorylation of nonstructural 5A protein of hepatitis C virus: HCV group-specific hyperphosphorylation." Virology **257**(1): 130-7.

Honda, M., L. H. Ping, R. C. Rijnbrand, E. Amphlett, B. Clarke, D. Rowlands and S. M. Lemon (1996). "Structural requirements for initiation of translation by internal ribosome entry within genome-length hepatitis C virus RNA." Virology **222**(1): 31-42.

Honda, M., M. R. Beard, L. H. Ping and S. M. Lemon (1999). "A phylogenetically conserved stem-loop structure at the 5' border of the internal ribosome entry site of hepatitis C virus is required for cap-independent viral translation." J Virol **73**(2): 1165-74.

Hwang, S. B., K. J. Park, Y. S. Kim, Y. C. Sung and M. M. Lai (1997). "Hepatitis C virus NS5B protein is a membrane-associated phosphoprotein with a predominantly perinuclear localization." Virology **227**(2): 439-46.

Inchauspe, G., Zebedee, S., Lee, D. H., Sugitani, M., Nasoff, M., and Prince, A. M. (1991). Genomic structure of the human prototype strain H of hepatitis C virus: comparison with American and Japanese isolates. Proc Natl Acad Sci U S A. **88**(22):10292-6.

Ishido, S., S. Muramatsu, T. Fujita, Y. Iwanaga, W. Y. Tong, Y. Katayama, M. Itoh and H. Hotta (1997). "Wild-type, but not mutant-type, p53 enhances nuclear accumulation of the NS3 protein of hepatitis C virus." Biochem Biophys Res Commun **230**(2): 431-6.

Ito, T., S. M. Tahara and M. M. Lai (1998). "The 3'-untranslated region of hepatitis C virus RNA enhances translation from an internal ribosomal entry site." J Virol **72**(11): 8789-96.

Kadare, G. and A. L. Haenni (1997). "Virus-encoded RNA helicases." J Virol **71**(4): 2583-90.

Kaito, M., S. Watanabe, K. Tsukiyama-Kohara, K. Yamaguchi, Y. Kobayashi, M. Konishi, M. Yokoi, S. Ishida, S. Suzuki and M. Kohara (1994). "Hepatitis C virus particle detected by immunoelectron microscopic study." J Gen Virol **75**(Pt 7): 1755-60.

Kaneko, T., Y. Tanji, S. Satoh, M. Hijikata, S. Asabe, K. Kimura and K. Shimotohno (1994). "Production of two phosphoproteins from the NS5A region of the hepatitis C viral genome." Biochem Biophys Res Commun **205**(1): 320-6.

Kato, N., M. Hijikata, Y. Ootsuyama, M. Nakagawa, S. Ohkoshi and K. Shimotohno (1990). "Sequence diversity of hepatitis C viral genomes." Mol Biol Med **7**(6): 495-501.

Kato, N., Y. Ootsuyama, S. Ohkoshi, T. Nakazawa, H. Sekiya, M. Hijikata and K. Shimotohno (1992). "Characterization of hypervariable regions in the putative envelope protein of hepatitis C virus." Biochem Biophys Res Commun **189**(1): 119-27.

Kato, N., H. Sekiya, Y. Ootsuyama, T. Nakazawa, M. Hijikata, S. Ohkoshi and K. Shimotohno (1993). "Humoral immune response to hypervariable region 1 of the putative envelope glycoprotein (gp70) of hepatitis C virus." J Virol **67**(7): 3923-30.

Kato, N., K. H. Lan, S. K. Ono-Nita, Y. Shiratori and M. Omata (1997). "Hepatitis C virus nonstructural region 5A protein is a potent transcriptional activator." J Virol **71**(11): 8856-9.

Kim, J. L., K. A. Morgenstern, C. Lin, T. Fox, M. D. Dwyer, J. A. Landro, S. P. Chambers, W. Markland, C. A. Lepre, E. T. O'Malley, S. L. Harbeson, C. M. Rice, M. A. Murcko, P. R. Caron and J. A. Thomson (1996). "Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide [published erratum appears in Cell 1997 Apr 4;89(1):159]." Cell **87**(2): 343-55.

Kumar, U., J. Monjardino and H. C. Thomas (1994). "Hypervariable region of hepatitis C virus envelope glycoprotein (E2/NS1) in an agammaglobulinemic patient [see comments]." Gastroenterology **106**(4): 1072-5.

Lee, C. H., C. Cheng, J. Wang and L. Lumeng (1992). "Identification of hepatitis C viruses with a nonconserved sequence of the 5' untranslated region." J Clin Microbiol **30**(6): 1602-4.

Lesniewski, R. R., K. M. Boardway, J. M. Casey, S. M. Desai, S. G. Devare, T. K. Leung and I. K. Mushahwar (1993). "Hypervariable 5'-terminus of hepatitis C virus E2/NS1 encodes antigenically distinct variants." J Med Virol **40**(2): 150-6.

Lin, C. and C. M. Rice (1995). "The hepatitis C virus NS3 serine proteinase and NS4A cofactor: establishment of a cell-free trans-processing assay." Proc Natl Acad Sci U S A **92**(17): 7622-6.

Lin, C., J. W. Wu, K. Hsiao and M. S. Su (1997). "The hepatitis C virus NS4A protein: interactions with the NS4B and NS5A proteins." J Virol **71**(9): 6465-71.

Liu, Q., R. A. Bhat, A. M. Prince and P. Zhang (1999). "The hepatitis C virus NS2 protein generated by NS2-3 autocleavage is required for NS5A phosphorylation." Biochem Biophys Res Commun **254**(3): 572-7.

Lo, S. Y., M. J. Selby and J. H. Ou (1996). "Interaction between hepatitis C virus core protein and E1 envelope protein." J Virol **70**(8): 5177-82.

Love, R. A., H. E. Parge, J. A. Wickersham, Z. Hostomsky, N. Habuka, E. W. Moomaw, T. Adachi and Z. Hostomska (1996). "The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site." Cell **87**(2): 331-42.

Lohmann, V., K. r. F, J. Koch, U. Herian, L. Theilmann and R. Bartenschlager (1999). "Replication of Subgenomic Hepatitis C Virus RNAs in a Hepatoma Cell Line." Science **285**(5424): 110-113.

Maniatis, T., E. F. Fritsch and J. Sambrook (1982). Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory.

Matsuura, Y., T. Suzuki, R. Suzuki, M. Sato, H. Aizaki, I. Saito and T. Miyamura (1994). "Processing of E1 and E2 glycoproteins of hepatitis C virus expressed in mammalian and insect cells." Virology **205**(1): 141-50.

Matsumoto, M., T. Y. Hsieh, N. Zhu, T. VanArsdale, S. B. Hwang, K. S. Jeng, A. E. Gorbalenya, S. Y. Lo, J. H. Ou, C. F. Ware and M. M. Lai (1997). "Hepatitis C virus core protein interacts with the cytoplasmic tail of lymphotoxin-beta receptor." J Virol **71**(2): 1301-9.

Miller, R. H. and R. H. Purcell (1990). "Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups." Proc Natl Acad Sci U S A **87**(6): 2057-61.

Michalak, J. P., C. Wychowski, A. Choukhi, J. C. Meunier, S. Ung, C. M. Rice and J. Dubuisson (1997). "Characterization of truncated forms of hepatitis C virus glycoproteins." J Gen Virol **78**(Pt 9): 2299-306.

Moriya, K., H. Fujie, Y. Shintani, H. Yotsuyanagi, T. Tsutsumi, K. Ishibashi, Y. Matsuura, S. Kimura, T. Miyamura and K. Koike (1998). "The core protein of hepatitis C virus induces hepatocellular carcinoma in transgenic mice." Nat Med **4**(9): 1065-7.

Ogata, N., Alter, H. J., Miller, R. H., and Purcell, R. H. (1991). Nucleotide sequence and mutation rate of the H strain of hepatitis C virus. Proc Natl Acad Sci U S A. **88**(8):3392-6.

Okamoto, H., S. Okada, Y. Sugiyama, S. Yotsumoto, T. Tanaka, H. Yoshizawa, F. Tsuda, Y. Miyakawa and M. Mayumi (1990). "The 5'-terminal sequence of the hepatitis C virus genome." Jpn J Exp Med **60**(3): 167-77.

Okamoto, N., M. Mizokami, H. Kano, E. Orito and N. Yoshihara (1991a). "[Risk of hepatitis C virus infection by needlestick among medical employees]." Kansenshogaku Zasshi **65**(11): 1470-5

Okamoto, H., Okada, S., Sugiyama, Y., Kurai, K., Iizuka, H., Machida, A., Miyakawa, Y., and Mayumi, M. (1991b). Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions. J Gen Virol. **72**(Pt

11):2697-704.

Okamoto, H., M. Kojima, S. Okada, H. Yoshizawa, H. Iizuka, T. Tanaka, E. E. Muchmore, D. A. Peterson, Y. Ito and S. Mishiro (1992). "Genetic drift of hepatitis C virus during an 8.2-year infection in a chimpanzee: variability and stability." Virology **190**(2): 894-9.

Pieroni, L., E. Santolini, C. Fipaldini, L. Pacini, G. Migliaccio and N. La Monica (1997). "In vitro study of the NS2-3 protease of hepatitis C virus." J Virol **71**(9): 6373-80.

Pileri, P., Y. Uematsu, S. Campagnoli, G. Galli, F. Falugi, R. Petracca, A. J. Weiner, M. Houghton, D. Rosa, G. Grandi and S. Abrignani (1998). "Binding of hepatitis C virus to CD81." Science **282**(5390): 938-41.

Poch, O., I. Sauvaget, M. Delarue and N. Tordo (1989). "Identification of four conserved motifs among the RNA-dependent polymerase encoding elements." Embo J **8**(12): 3867-74.

Preugschat, F., D. R. Averett, B. E. Clarke and D. J. T. Porter (1996). "A steady-state and pre-steady-state kinetic analysis of the NTPase activity associated with the hepatitis C virus NS3 helicase domain." J Biol Chem **271**(40): 24449-57.

Ralston, R., K. Thudium, K. Berger, C. Kuo, B. Gervase, J. Hall, M. Selby, G. Kuo, M. Houghton and Q. L. Choo (1993). "Characterization of hepatitis C virus envelope glycoprotein complexes expressed by recombinant vaccinia viruses." J Virol **67**(11): 6753-61.

Ray, R. B., K. Meyer and R. Ray (1996). "Suppression of apoptotic cell death by hepatitis C virus core protein." Virology **226**(2): 176-82.

Ray, R. B., R. Steele, K. Meyer and R. Ray (1997). "Transcriptional repression of p53 promoter by hepatitis C virus core protein." J Biol Chem **272**(17): 10983-6.

Ray, R. B., R. Steele, K. Meyer and R. Ray (1998). "Hepatitis C virus core protein represses p21WAF1/Cip1/Sid1 promoter activity." Gene **208**(2): 331-6.

Rice, C. M. (1999). "Is CD81 the key to hepatitis C virus entry?" Hepatology **29**(3): 990-2.

Simmonds, P. (1998). "Variability of the hepatitis C virus genome." Curr Stud Hematol Blood Transfus **62**: 38-63.

Rispeter, K., Lu, M., Lechner, S., Zibert, A., and Roggendorf, M. (1997). Cloning and characterization of a complete open reading frame of the hepatitis C virus genome in only two cDNA fragments. J Gen Virol. **78**(Pt 11):2751-9.

Santolini, E., G. Migliaccio and N. La Monica (1994). "Biosynthesis and

biochemical properties of the hepatitis C virus core protein." J Virol **68**(6): 3631-41.

Simmonds, P., E. C. Holmes, T. A. Cha, S. W. Chan, F. McOmish, B. Irvine, E. Beall, P. L. Yap, J. Kolberg and M. S. Urdea (1993). "Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region." J Gen Virol **74**(Pt 11): 2391-9.

Simmonds, P., Alberti, A., Alter, H. J., Bonino, F., Bradley, D. W., Brechot, C., Brouwer, J. T., Chan, S. W., Chayama, K., Chen, D. S., and *et al.* (1994). A proposed system for the nomenclature of hepatitis C viral genotypes [letter]. Hepatology. **19**(5):1321-4.

Simmonds, P. (1995). Variability of hepatitis C virus. Hepatology. **21**(2):570-83

Spangberg, K. and S. Schwartz (1999). "Poly(C)-binding protein interacts with the hepatitis C virus 5' untranslated region." J Gen Virol **80**(Pt 6): 1371-6.

Takamizawa, A., C. Mori, I. Fuke, S. Manabe, S. Murakami, J. Fujita, E. Onishi, T. Andoh, I. Yoshida and H. Okayama (1991). "Structure and organization of the hepatitis C virus genome isolated from human carriers." J Virol **65**(3): 1105-13.

Taylor, D. R., S. T. Shi, P. R. Romano, G. N. Barber and M. M. Lai (1999). "Inhibition of the Interferon- Inducible Protein Kinase PKR by HCV E2 Protein." Science **285**(5424): 107-110

Yeh, C. T., Lu, S. C., Chu, C. M., and Liaw, Y. F. (1997). Molecular cloning of a defective hepatitis C virus genome from the ascitic fluid of a patient with hepatocellular carcinoma. J Gen Virol. **78**(Pt 11):2761-70

Thomas DL, Mahley RW, Badur S, Palaoglu E, Quinn TC (1994). "The epidemiology of hepatitis C in Turkey." Infection **22**(6):411-4

Weiner, A. J., M. J. Brauer, J. Rosenblatt, K. H. Richman, J. Tung, K. Crawford, F. Bonino, G. Saracco, Q. L. Choo, M. Houghton and *et al.* (1991). "Variable and hypervariable domains are found in the regions of HCV corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins." Virology **180**(2): 842-8.

Weiner, A. J., H. M. Geysen, C. Christopherson, J. E. Hall, T. J. Mason, G. Saracco, F. Bonino, K. Crawford, C. D. Marion, K. A. Crawford and *et al.* (1992). "Evidence for immune selection of hepatitis C virus (HCV) putative envelope glycoprotein variants: potential role in chronic HCV infections." Proc Natl Acad Sci U S A **89**(8): 3468-72.

7. APPENDIX

Alignments of PCR primer sequences of 10 HCV subtype 1b isolates.

FORWARD-1 PRIMER

```
AF054249      ATCACTCCCCTGTGAGGAACTACTG
AF054247      .........................
AF054248      .........................
HCU45476      .........................
HPCK1R1       .........................
HPCK1R2       .........................
HPCK1R3       .........................
HPCK1S1       .........................
HPCK1S2       .........................
HPCK1S3       .........................
```

CORE REVERSE PRIMER

```
AF054249      ATGTCACGAACGACTGCTCCAACT
AF054247      ........................
AF054248      ........................
HCU45476      ..................T.....
HPCK1R1       ............T...........
HPCK1R2       ..................T...G
HPCK1R3       ........................
HPCK1S1       ............T...........
HPCK1S2       ........................G
HPCK1S3       ........................
```

E1 FORWARD PRIMER

```
AF054249      CCCGGTTGCTCTTTCTCTATC
AF054247      .....................
AF054248      .....................
HCU45476      .....................
HPCK1R1       .....................
HPCK1R2       .....................
HPCK1R3       .....................
HPCK1S1       .....................
HPCK1S2       .....................
HPCK1S3       .....................
```

E2 REVERSE PRIMER

```
AF054249      GACCGGGAGATGGCTGCAT
AF054247      ...................
AF054248      ...................
HCU45476      ...................
HPCK1R1       ..............C....
HPCK1R2       ...................
HPCK1R3       ...................
HPCK1S1       ..............C....
HPCK1S2       ...................
HPCK1S3       ...................
```

NS2 FORWARD PRIMER

```
AF054249      TCCTACTGGCGTTACCACC
AF054247      ...................
AF054248      ...................
HCU45476      ....T...A..........
HPCK1R1       ...G...............
HPCK1R2       ...G...............
HPCK1R3       .T..G...T..........
HPCK1S1       ...G...............
HPCK1S2       ...G...............
HPCK1S3       .T..G...T..........
```

NS2 REVERSE PRIMER

```
AF054249      ACGGCCTACTCCCAACAAAC
AF054247      ....................
AF054248      ....................
HCU45476      .........G.T.....G..
HPCK1R1       ..................G..
HPCK1R2       ..............G..G..
HPCK1R3       ...........A.....G..
HPCK1S1       ..................G..
HPCK1S2       ..............G..G..
HPCK1S3       ..................G..
```

NS3 FORWARD PRIMER

```
AF054249      CCGAAGGGGGAAGGAGAT
AF054247      ..................
AF054248      ..................
HCU45476      ..........G.......
HPCK1R1       ..........G.......
HPCK1R2       ..................
HPCK1R3       ..........G.......
HPCK1S1       ..........G.......
HPCK1S2       ..................
HPCK1S3       ..................
```

NS3 REVERSE PRIMER

```
AF054249      GTCGTCACTAGCACCTGGGTGC
AF054247      ......................
AF054248      ......................
HCU45476      ........G.............
HPCK1R1       ........C.............
HPCK1R2       .....T..G.............
HPCK1R3       ........G.....T.......
HPCK1S1       ........C.............
HPCK1S2       ........G.............
HPCK1S3       ........G.....T.......
```

NS4 FORWARD PRIMER

```
AF054249      ATGCATGTCGGCTGACCT
AF054247      ..................
AF054248      ..................
HCU45476      ..........C.....
HPCK1R1       ..................
HPCK1R2       ..................
HPCK1R3       ..................
HPCK1S1       ..................
HPCK1S2       ..................
HPCK1S3       .........A........
```

NS4 REVERSE PRIMER

```
AF054249      GACTTCAAGACCTGGCTCCA
AF054247      ....................
AF054248      ....................
HCU45476      ....................
HPCK1R1       ................T...
HPCK1R2       ....................
HPCK1R3       ....................
HPCK1S1       ................T...
HPCK1S2       ...........T........
HPCK1S3       ....................
```

NS5A FORWARD PRIMER

```
AF054249      TATGTGCCTGAGAGCGACG
AF054247      ...................
AF054248      ...................
HCU45476      ...................
HPCK1R1       ...................
HPCK1R2       ...................
HPCK1R3       ...................
HPCK1S1       ...................
HPCK1S2       ...................
HPCK1S3       ...................
```

NS5A REVERSE PRIMER

```
AF054249      TGAGCAACTCTTTGCTGCGT
AF054247      ....................
AF054248      ....................
HCU45476      ................A..C
HPCK1R1       ....................
HPCK1R2       ...................C
HPCK1R3       ....................
HPCK1S1       ....................
HPCK1S2       ...................C
HPCK1S3       ....................
```

NS5B FORWARD PRIMER

```
AF054249      AGCGACGGGTCTTGGTCTAC
AF054247      ....................
AF054248      ....................
HCU45476      ....................
HPCK1R1       ....................
HPCK1R2       ....................
HPCK1R3       ....................
HPCK1S1       ....................
HPCK1S2       ....................
HPCK1S3       ..........C.........
```

NS5B REVERSE PRIMER

```
AF054249      CGGAGCTAACCACTCCAGG
AF054247      ....................
AF054248      ....................
HCU45476      .........A..........
HPCK1R1       .........A..........
HPCK1R2       ......C.............
HPCK1R3       .........A..........
HPCK1S1       .........A..........
HPCK1S2       .........A..........
HPCK1S3       ....................
```