CHARLES UNIVERISTY IN PRAGUE

Faculty of Science

Department of Genetics and Microbiology

Mgr. Anas Khawaja

A study of the HCV IRES variability: An experimental approach coupled with design of a large-scale mutation database

Studie rozmanitosti HCV IRES: propojení experimentálního přístupu s přípravou a hodnocením rozsáhlé databáze mutací

Ph.D. Thesis

Advisor: RNDr. Martin Pospíšek, Ph.D.

Prague, 2016

Declaration:

I declare that I wrote this work on my own and all sources and literature are properly cited. I also declare, that this work or its major part was not previously used for obtaining of the same or any other academic degree.

Anas Khawaja.

Prague, May 2016.

Acknowledgements

The journey of countless miles begins with a single step and just like every road has an end, alien as it may sound, PhD too is not endless. The people who supported and encouraged me throughout my study are to be blamed for my continuous suffering.

First and foremost, I would like to express my enormous gratitude to my supervisor Martin Pospisek for his support, guidance and constructive critique. I would also like to acknowledge my co-supervisor Vaclav Vopalensky for offering his advice, expertise and encouragement of my effort over the years which at times must have seemed a lost cause.

I am grateful to all my lab colleagues especially Sara, Vero, Michal, Pepa and Silvia for assisting me in several ways and helping me to stay in touch with reality.

I would always remember the endless ‘chilling’ sessions I have had with my now friends-for-life, Humi, Hamzah, Arshad and Bilal. It is really sad that ‘chilling’ is coming to an end since everyone is leaving but I shall always cherish our time that we have spent in Prague.

I would also like to extend my gratitude to my family for their constant support and love. To my late uncle, Khawaja Ahmad Nadeem who always believed in me and is a source of inspiration. I am also thankful to my best-friend Ali Khawaja for helping me survive all the stress in difficult times.

My loving brother, Usama, and sisters, Zunera and Hira, nephews and nieces: none of you left my side and your faith in me kept me going. A big thank you to my Mom, Shagufta Tenvir and Dad, Imtiaz Khawaja for their support and especially patience for the amount of years this thesis consumed.

Abstract

Translation initiation in the hepatitis C virus (HCV) occurs through a cap-independent mechanism that involves an internal ribosome entry site (IRES) capable of interaction with and utilization of the eukaryotic translational machinery. We focused on the structural configuration of the different HCV-IRES domains and the impact of IRES primary sequence variations on secondary structure conservation and function. For this purpose we introduced into our laboratory, methods such as denaturing gradient and temperature gradient gel electrophoresis for screening the degree of heterogeneity and total amount of HCV-IRES variability accumulated in HCV infected patients over a period of time. The selected samples showed variable migration pattern of the HCV-IRES (from all the patients) visualized in DGGE and TGGE, were sequenced and evaluated for translation efficiency using flow cytometry. In some cases, we discovered that multiple mutations, even those scattered across different domains of HCV-IRES, led to restoration of the HCV-IRES translational activity, although the individual occurrences of these mutations were found to be deleterious. We propose that such observation may be attributed to probable long-range inter- and/or intra-domain functional interactions. We established a large-scale HCV-IRES variation database (HCVIVdb; www.hcvivbd.org) comprises ~1900 mutations acquired from majority of the HCV-IRES mutation-linked studies. The HCVIVdb contributes extensively by providing comprehensive HCV-IRES mutation dataset that can be utilized in overall evaluation of the functional role(s) played by structural IRES elements in modulation of HCV-IRES translation initiation. The design of the knowledge base web interface and advanced search tools is conducive to perform and visualize multiple analyses by comparisons of the collated HCVIVdb data, leading to new findings. We also identified ~20 novel HCV-IRES mutations and determined their influence on HCV-IRES translation efficiency by flow cytometry. Further elucidation of these novel mutations in the context of probable HCV-IRES structure conformations and contacts with translation machinery was demonstrated using the available mutation (HCVIVdb), structure and biochemical data. For validation of our HCVIVdb dataset we used multiple sequence alignment of the HCV genome data from other resources. These validations showed a positive correlation and signified the conservation of specific nucleotides and hypervariability of others.

The extended data of the nucleotide variability also provides wider knowledge about the evolutionary advantage and preservation of specific bases at each HCV-IRES nucleotide position. The structural conformation, sequence preservation and variability, and translational machinery association with the HCV-IRES regions are also thoroughly discussed, along with other factors that can affect and influence the formation of translation initiation complexes.

Moreover, our results indicate interplay between codon bias and mRNA secondary structure as determinants of translation efficiency of the HCV-core RNA during HCV-IRES mediated translation initiation.

Keywords: Hepatitis C virus, HCV, internal ribosome entry site, IRES, database, translation efficiency, structure, quasispecies, codon bias, core.

Abstrakt

Iniciace translace u viru hepatitidy C (HCV) probíhá mechanismem nezávislým na čepičce, který využívá vnitřního vazebného místo pro ribozom (IRES), které interaguje s eukaryotním translačním aparátem a je schopné jej využívat. V našem výzkumu jsme se zaměřili na strukturní konformaci jednotlivých domén HCV-IRES a na dopad primární sekvence IRES na konzervovanost a funkci této sekundární struktury. Za tímto účelem byl proveden *screening* pacientů infikovaných virem hepatitidy C pro zjištění míry heterogenity a celkové variability HCV-IRES vzniklé u pacientů v průběhu jejich nemoci pomocí optimalizování metod, jako jsou gelové elektroforézy s denaturačním nebo teplotním gradientem. Účinnost translace mutovaných HCV-IRES sekvencí byla měřena pomocí průtokové cytometrie. V některých případech jsme zjistili, že několikanásobné mutace, včetně těch, které jsou rozesety mezi různé domény HCV-IRES, vedly k obnově translační aktivity, ačkoli tyto mutace účinnost translace snižovaly, pokud byly analyzovány jednotlivě. Navrhujeme, že tato pozorování by mohla být vysvětlena potenciálními mezi- a/nebo vnitro-doménovými funkčními interakcemi působícími na velkou vzdálenost. Také jsme nalezli nové mutace, jejichž vliv na účinnost translace HCV-IRES byl experimentálně ověřen. Vytvoření rozsáhlé databáze variací HCV-IRES (*HCV-IRES variation database*, HCVIVdb) dále přispívá k pochopení dopadu mutací na syntézu virových proteinů, a to na úrovni domén, subdomén nebo jednotlivých nukleotidů. Návrh vyhledávacích nástrojů a webového rozhraní k přístupu k databázi byl vytvořen s cílem umožnit a vizualizovat mnohonásobné analýzy. Námi získaná data byla srovnána s HCV sekvencemi pocházejících z jiných zdrojů. Toto srovnání vykazuje pozitivní korelaci a signifikantní konzervovanost určitých nukleotidů a hypervariabilitu nukleotidů jiných. Pro přesné fungování vyžaduje HCV-IRES specifickou interakci svých domén s podjednotkami ribozomu a s částí eukaryotických translačních iniciačních faktorů (eIF). Strukturní konformace, konzervovanost či variabilita sekvence a propojení translační mašinérie s oblastmi HCV-IRES jsou také podrobně probírány, spolu s dalšími faktory, které mohou ovlivňovat sestavování translačních iniciačních komplexů a působit na ně.

Klíčová slova: Virus hepatitidy C, HCV, vnitřní vazebné místo pro ribozom, IRES, databáze, účinnost translace, struktura, kvazidruhy, eukaryotické iniciační faktory.

Table of Contents

1 Abbreviations

3D	three-dimensional
Å	angstrom (10^{-10} m)
(-)	Watson-Crick base-pair
(•)	non-canonical base-pair
aa	amino acid
ATP	adenosine triphosphate
β-hairpin	beta hairpin
bp	base pair
BRC	bioinformatics resource center
CD	circular dichroism
cDNA	complementary DNA
CrPV	cricket paralysis virus
Cryo-EM	cryo electron microscopy
CSFV	classic swine fever virus
CTL	cytotoxic T-cell
Da	dalton
DAA	direct-acting antivirals
DEPC	diethylpyrocarbonate
DDBJ	DNA Data Bank of Japan
ddH_2O	deionised distilled water
ddNTPs	dideoxyribonucleotide triphosphates
DGGE	denaturing gradient gel electrophoresis
DMEM	Dulbecco´s modified Eagle's medium
DMS	dimethyl sulfate
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTPs	deoxyribonucleotide triphosphates
DTT	dithiotreitol
E1	envelop gene 1
E2	envelop gene 2
EDTA	ethylenediaminetetraacetic acid
EGFP	enhanced green fluorescent protein
eIFs	eukaryotic translation initiation factors
eIF2-TC	eIF2-GTP-Met $tRNA_i$, the ternary complex
EM	electron microscopy
EMBL	European Molecular Biology Laboratory
ES7	expansion segment 7
EtBr	ethidium bromide

FACS	fluorescence-activated cell sorting
FBS	fetal bovine serum
FCS	fetal calf serum
FRET	Förster resonance energy transfer
GDP	guanosine diphosphate
GFP	green fluorescent protein
GTP	guanosine triphosphate
h	helice
HCC	hepatocellular carcinoma
HBV	hepatitis B virus
HCV	hepatitis C virus
HCVIVdb	hepatitis C virus IRES variation database
HEV	hepatitis E virus
HIw	high with loops
HIwo	high without loops
HRP	horseradish peroxidase
HGMD	Human Gene Mutation Database
HVDB	hepatitis virus database
HVR	hypervariable region
IFN-α	interferon alpha
IMEx	The International Molecular Exchange consortium
INSD	International Nucleotide Sequence Database
IP	immunoprecipitation
IRES	internal ribosomal entry site
ISDR	interferon sensitivity determining region
ISG	interferon-stimulated genes
jIIIabc	junction IIIabc
kb	kilobase
kDa	kilodalton
Low	low with loops
Lowo	low without loops
Met-tRNA$_i$	methionyl transfer RNA
mRNA	messenger RNA
miRNA	microRNA
MS	mass spectrometry
MSA	multiple sequence alignment
NCR	non-coding region
NGS	next-generation sequencing
NIAID	National Institute of Allergy and Infectious Diseases
NMR	Nuclear Magnetic Resonance
nt.	nucleotide
NVR	non-sustained viral response
ORF	open reading frame

PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
P_i	inorganic phosphate
PKR	protein kinase R
rpS	ribosomal proteins
RdRp	RNA dependent RNA polymerase
RNA	ribonucleic acid
RNase	ribonuclease
rRNA	ribosomal RNA
S	Svedberg, sedimentation (10^{-13} seconds or 100 femtosecond).
SAXS	small-angle x-ray scattering data
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SHAPE	selective 2'-OH acylation analysed by primer extension
siRNA	small interfering RNA
SL	stem-loop
SNP	single nucleotide polymorphism
ssDNA	single strand deoxyribonucleic acid
ssRNA	single strand ribonucleic acid
SVR	sustained viral response
TAE	Tris-acetate-EDTA
tAI	transfer RNA adaptation index
tRNA	transfer RNA
TBS	Tris-buffered saline
TEMED	tetramethylethylenediamine
TGGE	temperature gradient gel electrophoresis
UTR	untranslated region
UV	ultraviolet
ViPR	Virus Pathogen Database and Analysis Resource
VSV	vesicular stomatitis virus
WT	wild type

2 Introduction

The protein synthesis of hepatitis C virus (HCV) genome is accomplished via an internal ribosome entry site (IRES) element that provides an alternative mechanism for translation initiation. The recognition of 5' cap or the scanning procedure is substituted with direct recruitment and positioning of the small ribosome subunit to the initiation codon without further requirement of all the eukaryotic initiation factors (eIFs) but only a small subset. The complex independently folded domains' (I-IV) architecture of the HCV-IRES establishes high-affinity interaction(s) with the 40S subunit promoting the stable binding of eIFs in order to form active 80S complex that drives the viral mRNA translation. The appropriate functionality of viral protein synthesis is highly dependent on the overall domain organization (I-IV) of HCV-IRES in the context of its fairly conserved sequence and structure. The ~341 nucleotide spanning IRES is largely susceptible to variations across various conserved RNA motifs and elements, observed through extensive mutagenesis, and culminate in the disruption of translation.

The focal point of our study was to investigate and establish, on a large-scale, relationship between sequence changes in overall HCV-IRES and the functional output as a direct result of these mutations. The comprehensive evaluation of the observed phenotype(s) was concluded through the comparative analyses with our own results and published mutation and biochemical data. In the process, techniques were developed to screen large number of HCV-IRES samples from the patients and differentiate them on the basis of genetic heterogeneity making it a very accessible and useful tool to characterize the genetic variants of virus population(s). The influence of novel mutations on HCV-IRES function was established and intriguing phenomenon such as long-range inter- and intra-domain interaction was detected and analyzed from the patients. In addition, we also investigated into the HCV patients, treated or non-treated with standard combination therapy of interferon (IFN) and ribavirin to determine any link between HCV-IRES nucleotide changes as a result of therapy and the respective viral response outcome. A large-scale knowledge-base was also developed constituting the variations and other information, gathered from bulk of the HCV-IRES mutation-associated studies,

assembled and organized into various classifications. The objective was to set up an analytical tool equipped with simple / readily accessible web interface and extensive search engine that provides in-depth information on all the available specialized HCV-IRES variation data. The data from HCV-IRES variation database (HCVIVdb) was validated by retrieving information from other resources and comparing these HCV genome sequence entries with our dataset to look for correlation. The correlation between the datasets also lead to other interesting finds on part of the HCV-IRES conservation and hypervariability of various nucleotides. These results, in general, offer detailed insights into the functional dynamics of HCV-IRES ranging from an overall capacity of domains and subdomains to individual nucleotides.

Furthermore, we synthesized HCV-core constructs that varied in their codon usage and maintenance of mRNA structure. We found that both the codon bias and mRNA structure interplay as determinants of protein expression levels of truncated HCV-core RNA.

3 Literature Review

Understanding the potential of HCV-IRES domains to modulate translation initiation via their structure and function

Hepatitis C virus (HCV) is a blood-borne pathogen with an estimated global prevalence of 130-200 million cases of chronic infection (Gravitz, 2011), increasing these patients' risk of developing liver cirrhosis and hepatocellular carcinoma. The standard chemotherapeutic treatment for HCV is the combination of pegylated interferon alpha (IFN-α) with guanosine analogue ribavirin, which is linked to side effects and can often lead to a relapse in many patients (Feld and Hoofnagle, 2005, Dienstag and McHutchison, 2006). Only recently, other treatments based on the inhibition of either the NS5B viral polymerase (sofosbuvir) (Sofia et al., 2010, Lawitz et al., 2013, Mizokami et al., 2015) or the NS3/4A viral protease (simeprevir) (Fried et al., 2013, Gutierrez et al., 2015), boceprevir (Malcolm et al., 2005, Nagpal et al., 2015) and telaprevir (Perni et al., 2006, Neukam et al., 2015) have been approved.

The viral genome comprises a 9.6 kb long, single-stranded, positive sense RNA molecule (Choo et al., 1991, Takamizawa et al., 1991) that possesses a highly conserved 5' untranslated region (5' UTR), a single open reading frame (ORF) and a conserved 3' UTR. The ORF encodes a single large polyprotein of >3000 amino acids, which is co- and post-translationally cleaved by cellular and viral proteases to yield the mature viral structural (core, E1, and E2) and non-structural (p7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B) proteins. Translation of the polyprotein is driven by an internal ribosomal entry site (IRES) that occupies most of the 5' UTR of the viral RNA (Honda et al., 1996a). Translation initiation mediated by the HCV-IRES is distinct from canonical eukaryotic mRNA translation initiation in that it does not require the presence of the m^7G cap moiety at the 5' mRNA end along with all of the eukaryotic translation initiation factors (eIFs) that are normally responsible for the recruitment of a ribosome to mRNA and its scanning for the correct AUG initiation codon. Instead, the HCV-IRES directly recruits and positions the small ribosomal

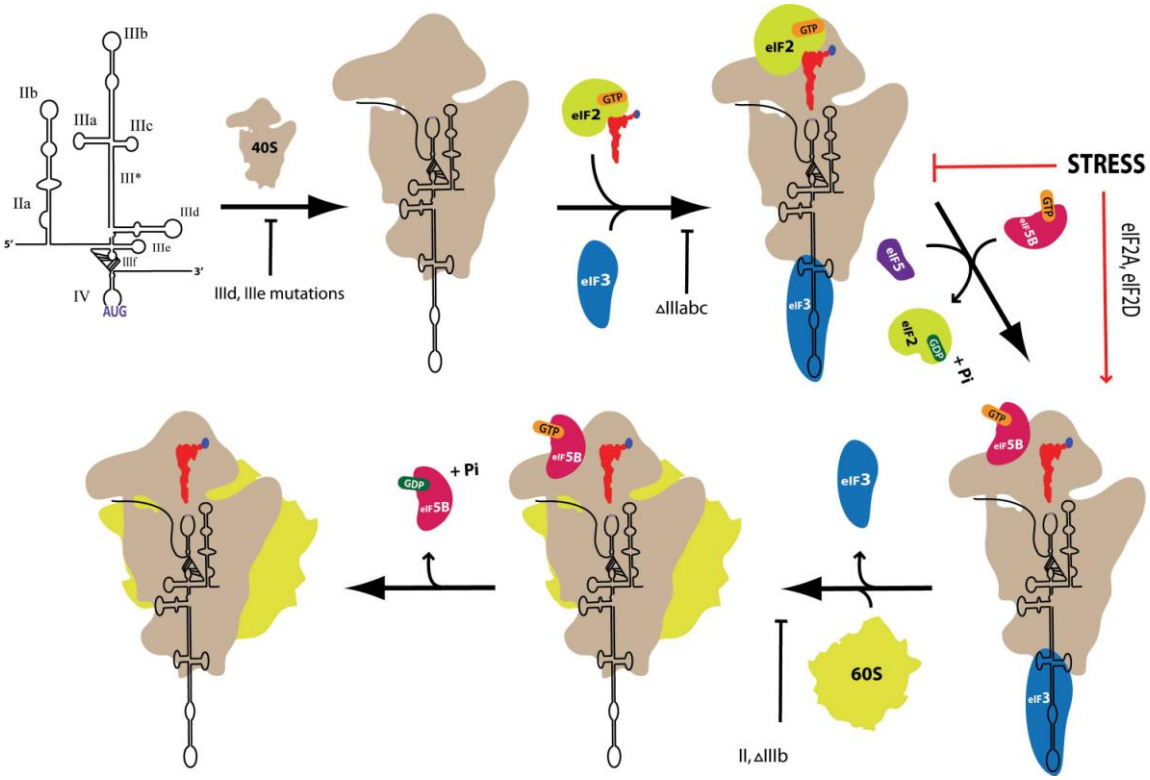

Figure 1: A stepwise demonstration of the HCV-IRES translation initiation pathway. The specified arrangement of the HCV-IRES domains and ribosomal subunits along with the particular placement of a subset of canonical eukaryotic translation initiation factors (eIFs) required to assemble the pre-initiation 48S and 80S complexes are depicted. The red arrows indicate eIF2-GTP-independent translation under stress (upon eIF2α phosphorylation), when protein synthesis continues with eIF5B or other reported factors (eIF2A, eIF2D). Mutations in different HCV-IRES domains that may inhibit particular stages of translation complex formation are shown under the arrows.

subunit at the start AUG codon without the need for any known initiation factor (Pestova et al., 1998). This positioning is followed by the binding of eIF3 and eIF2-GTP-Met $tRNA_i$, the ternary complex (eIF2-TC) that stabilizes the pre-initiation translation assembly (Pestova et al., 1998, Ji et al., 2004). The release of eIFs via GTP hydrolysis further allows the binding of a 60S subunit to assemble an 80S ribosome and initiate protein synthesis (Locker et al., 2007) (Figure 1). The establishment of 48S and 80S pre-initiation complexes requires many other intermediate processes that, if not performed properly, can result in the blockage of efficient translation.

HCV-IRES RNA has a conserved primary sequence and secondary structure that spans ~341 nucleotides composing domains I, II, III and IV (Wang et al., 1993,

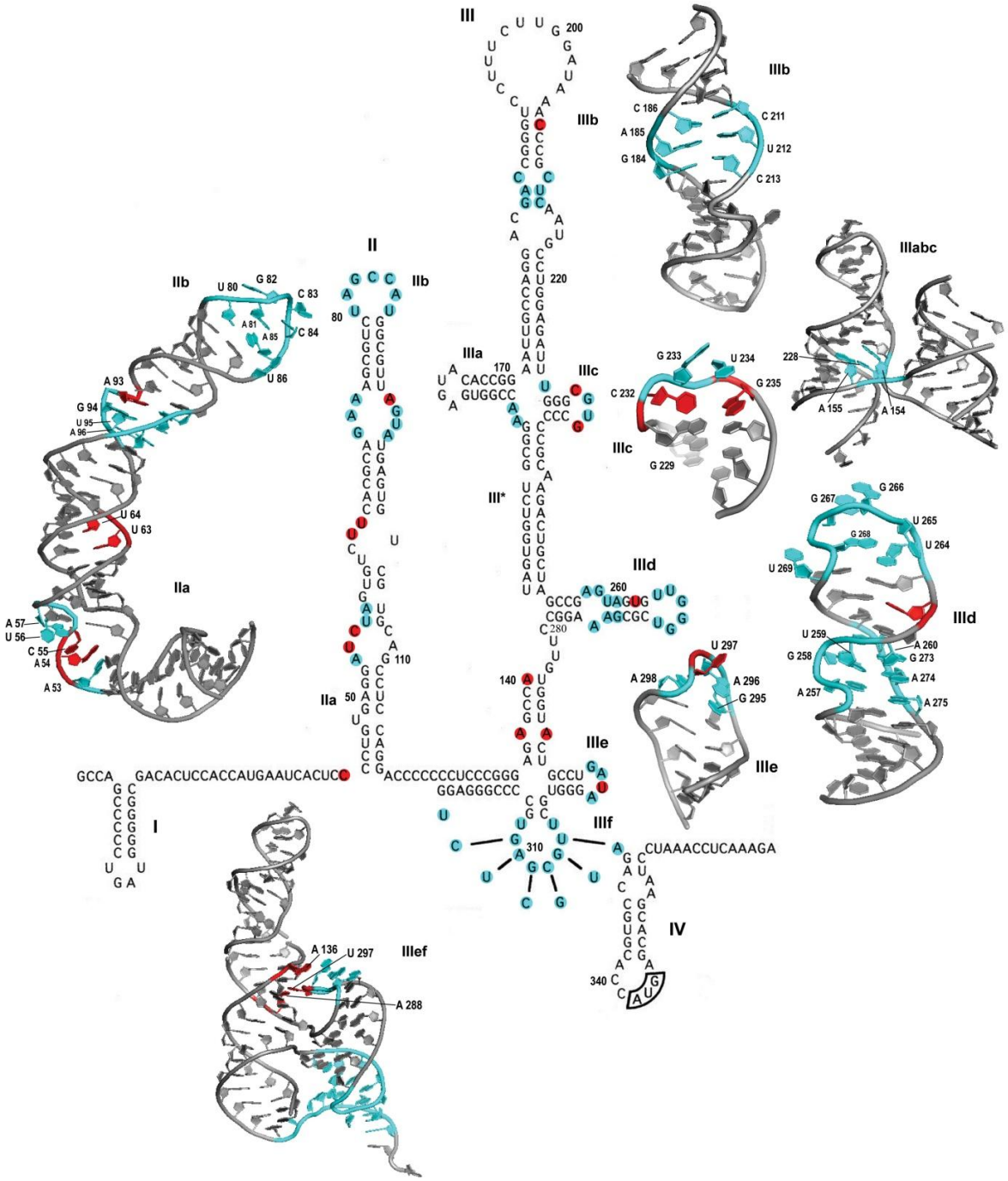

Figure 2: The proposed secondary structure of the hepatitis C virus internal ribosome entry site (HCV-IRES) (Honda et al., 1996a, Zhao and Wimmer, 2001a), with three-dimensional structures of domains II-IV resolved by NMR and X-ray crystallography (Lukavsky et al., 2003, Collier et al., 2002, Kieft et al., 2002, Lukavsky et al., 2000, Rijnbrand et al., 2004, Berry et al., 2011). The domains of the HCV-IRES are labeled. Nucleotides colored in cyan indicate the highly conserved regions of the IRES, whereas nucleotides in red denote mutations in the domains involved in a proposed long-range inter or intra-domain interaction or an unusual translational activity (see results and discussion). The colors in the secondary structure correspond to the 3D structure. The PDB entries for the HCV-IRES domains are domains II (1P5P), IIIb (1KP7), IIIc (1IDV), IIIabc (1KH6), IIId (1F84), IIIe (1F85) and IIIef (3T4B).

Tsukiyamakohara et al., 1992, Brown et al., 1992). The HCV-IRES adopts a characteristic 3D conformation or specific tertiary structure under physiological concentrations of metal ions (Kieft et al., 1999). The conservation of the HCV-IRES structure determines the efficacy of viral protein synthesis. Various structural (Lukavsky et al., 2003, Spahn et al., 2001, Kieft et al., 2002, Lyons et al., 2001), *in vitro* mutagenesis (Odreman-Macchioli et al., 2001, Kalliampakou et al., 2002), biochemical and enzymatic studies (Kolupaeva et al., 2000b, Sizova et al., 1998) have demonstrated the importance of the conservation and specificity of domains II-IV, which are required to interact with translational machinery and perform the HCV-IRES mediated translation. Whereas the deletion of domain I results in an increased level of translational response (Rijnbrand et al., 1995, Honda et al., 1996b), domains II and III have been found to interact with the 40S subunit, which is the first step directed towards protein synthesis. A pancreatic RNase protection assay has been utilized to locate sequences that occur in close proximity to the 40S subunit but are not necessarily in direct contact (Lytle et al., 2001, Lytle et al., 2002). Similarly, other studies employed methodologies such as gel digestion, cross-linking and mass spectrometry analysis in order to directly observe the contact points of HCV-IRES with small ribosomal subunit (Lu et al., 2004, Otto et al., 2002, Fukushi et al., 1999, Yu et al., 2005, Bhat et al., 2015). In addition to making contact with 40S, the basal region of domain III also helps in the placement of the HCV-IRES start codon, which is present in domain IV, in the ribosomal decoding groove (Berry et al., 2010). Domain III also regulates the binding of eIF3 through its own apical half (Kieft et al., 2001, Sizova et al., 1998), while domain II modulates GTP hydrolysis of the TC, mediated by eIF5 (Locker et al., 2007). The HCV-IRES can also initiate translation through an alternative eIF2-independent pathway under stress conditions and increased eIF2-α phosphorylation (Terenin et al., 2008, Pestova et al., 2008).

The attempt here is to describe the current understanding of the HCV-IRES, encompassing the importance of the HCV-IRES structural elements in the mediation of translation initiation. The influence of mutations in the context of possible alterations of the HCV-IRES configuration and its interaction with translational machinery that may disrupt its function is investigated. I have also highlighted combinations of mutations collectively expressing restored translational activity, the individual occurrence of which otherwise displayed reduced HCV-IRES efficiency.

3.1 DOMAIN II STRUCTURE CONSERVATION IS VITAL FOR HCV-IRES FUNCTION

3.1.1 L-shaped domain IIa may serve as a target for new antivirals

Based on phylogenetic and mutation analyses, domain II was found to be highly conserved among the IRES RNA of HCV, pestivirus and GB virus B isolates (Honda et al., 1999a, Zhao and Wimmer, 2001a). The NMR structure of HCV-IRES domain II exhibited an L-shaped conformation that retains an identical configuration in both its free and 40S-bound forms (Lukavsky et al., 2003). The structure of domain IIa (nucleotides 49-69 and 100-115) was also investigated using X-ray crystallography (Dibrov et al., 2007, Zhao et al., 2008). Domain IIa is an independently folded domain of the HCV-IRES that adopts a 90° helical bent structure including nucleotides A^{53}, A^{54}, C^{55}, U^{56} and A^{57}. Continuous base stacking and hydrogen bonding between the nucleotides determine the characteristic structural features of this region comprising an extended conformation and a curve in the RNA backbone with a looped out U^{56} (Figure 2). The divalent metal ions present in the internal bulge of domain IIa interact with different bases, contributing to the stability of the helical bend (Lukavsky et al., 2003, Dibrov et al., 2007, Zhao et al., 2008). The sequence and structural integrity of this asymmetrical loop are important for HCV-IRES function, and the deletion of the loop has deleterious effects on translation efficiency (Odreman-Macchioli et al., 2001). The regions adjacent to the helical bend forming the base and upper stem of domain IIa are also susceptible to substitutions that display lower translation activity (Honda et al., 1999a, Odreman-Macchioli et al., 2001, Tang et al., 1999). Surprisingly, some mutants with multiple mutations in these regions do not display appreciably low translation activity despite the lower translation response observed upon the individual occurrence of these mutations.

A subdomain IIa helical bend has also been identified as a vital target site for a new class of benzimidazole molecule inhibitors that exhibit binding affinity with this region (Seth et al., 2005, Boerneke et al., 2014). Binding of the inhibitory compound changes the structural conformation by widening the interhelical angle in the domain IIa bend in HCV-infected cells. The interference in the structure of the bend helps undock domain IIb from the 40S subunit, eventually restricting the formation of the

80S pre-initiation translational complex to inhibit protein synthesis (Parsons et al., 2009). NMR was employed to study the conformational changes of domain IIa induced by the ligand inhibitor molecule that straightens the helical bend, consequently disturbing 80S assembly (Paulsen et al., 2010). Furthermore, X-ray crystallography data give much deeper insight into the binding interactions of domain IIa in complex with benzimidazole translation initiation inhibitors (Dibrov et al., 2012).

3.1.2 Dynamic domain II induced structural rearrangements in 40S necessitates formation of translation initiation complexes

Subdomain IIb contains an apical loop with a looped out uracil (U^{86}) at its 3' side and a loop E-motif constituting an S-turn. Both the apical loop and the loop E-motif are highly conserved among HCV isolates (Lukavsky et al., 2003). These structural features are somewhat similar to those of domain IIId. However, the location of the looped-out uracil (U^{269}) in the IIId apical loop and the S-turn from the loop E-motif in the same domain are both on the 3' side in contrast to domain IIb, where they reside on opposite sides. These features constitute binding surfaces that are characteristic of both domains IIb and IIId for the interaction with 40S (Lukavsky et al., 2000) (Figure 2). Conservation of the domain II loop E-motif and the apical loop has been shown to be of some significance, with deletions and substitutions abrogating HCV-IRES function in translation initiation (Odreman-Macchioli et al., 2001, Kalliampakou et al., 2002). The substitutions of nucleotides G^{82} and U^{86} at the apical loop somehow do not affect the IRES activity, contrary to substitutions of the other nucleotides in this region. However, the deletion of nucleotides in the apical loop generates low translation feedback, suggesting the importance of structural conservation (Kalliampakou et al., 2002). Cross-linking data have shown that the domain II apical loop interacts with ribosomal proteins S14 and S16 of the 40S subunit, while G^{87} in the IIb stem was not cross-linked to any of the ribosomal proteins or ribosomal RNA (rRNA) (Babaylova et al., 2009, Laletina et al., 2006). The bend conformation allows domain II to reach the E-site upon binding to the 40S subunit. The interaction of domain II with 40S also occurs through ribosomal protein S5 in the head region, which allows a conformational change and movement of the head relative to the 40S

body (Spahn et al., 2001, Fukushi et al., 2001). The similar structural changes were also induced by CrPV-IRES (cricket paralysis virus IRES) to 40S ribosome although RNA-binding sites for both the IRESs are different (Spahn et al., 2004). The apical loop of HCV-IRES domain II lies in close proximity to the loop regions of helices 23 (h23) and 24 of the 40S subunit (Boehringer et al., 2005). The similar interactions of domain II apical loop with the β-hairpin structure of ribosomal protein S5 called uS7 according to the new system for naming ribosomal proteins, (Ban et al., 2014), and S14 (uS11) in E-site of the 40S has been viewed via cryo-EM (Quade et al., 2015, Yamamoto et al., 2015). The introduction of mutations in β-hairpin in the loop of uS7, in yeast, has shown to lower the initiation fidelity *in vivo* (Visweswaraiah et al., 2015), suggesting that uS7 may also play a similar role in start codon recognition in HCV-IRES translation. Additional less-specific contact of ribosomal protein S25 (eS25), which is crucial in translation initiation of HCV, CrPV (Landry et al., 2009) and other IRESs (Hertz et al., 2013), with domain II was also revealed (Quade et al., 2015). By using single-molecule Förster resonance energy transfer (FRET) between rpS25 (eS25) of 40S and domain II of the HCV-IRES it has been demonstrated that 40S recruitment to the HCV-IRES is a single-step process that results in a conformationally dynamic complex formation that continues to undergo structural rearrangements, followed by 80S assembly (Fuchs et al., 2015). A detailed account of structural rearrangements of HCV-IRES/40S upon interaction and assembly crucial for translation initiation has been given in this review (Hellen, 2009). Recent cryo-EM reconstructions of 80S/HCV-IRES assembly has revealed that IRES-induced tilted head position of the 40S is reversed after the attachment of P-site tRNA. It also allows domain II to shift its position from 40S E-site to 60S E-site at h28 of 28S rRNA. The atomic model also exhibits the structure of ribosome bound to HCV-IRES in 80S/HCV-IRES complex at a resolution of 3.9 Å (angstrom) together with the interpretation of molecular interactions between HCV-IRES and ribosome (Yamamoto et al., 2015).

Through chemical probing, it was shown that mutations in domain IIb affect the domain IV structure in the HCV-IRES/40S complex, disrupting its arrangement at the AUG start codon and leading to a decreased translational response. It has been proposed that a long-range structural interaction between two domains that lie in close proximity within the HCV-IRES/40S complex can have important implications

(Filbin and Kieft, 2011). Mutations in the apical loop of domain IIb result in decreased HCV-IRES functionality, but translation is not aborted, and the formation of all translation complexes occurs with equal efficiency to that with the wild type (WT) IRES. However, the interaction of domain IIb apical loop mutants with the β-hairpin of S5 (uS7) in the E-site is altered. The disruptive contact with the β-hairpin of ribosomal protein S5 caused by local structural perturbation of the mutants is suggested to induce a global change in the HCV-IRES/40S conformation that may stall the first event of ribosomal translocation, as observed through toe-printing, after 80S assembly (Filbin et al., 2012). Computational modelling together with RNA-protein interaction assays and UV cross-linking has also shown the interaction of domain II and IV with β-hairpin of ribosomal protein S5. The array of domain II and IV mutants established the contact points with S5 and demonstrated that interaction between S5-domain II is crucial in formation of 48S and 80S assembly (Bhat et al., 2015, Joseph et al., 2014). While the HCV-IRES is still able to bind the 40S ribosomal subunit and recruit eIF2-TC and eIF3 to form a 48S pre-initiation complex when domain II is deleted entirely, it is inefficient at 60S subunit joining and thus 80S ribosome assembly (Otto and Puglisi, 2004, Ji et al., 2004). This inefficient 80S formation is mainly due to the fact that the processes facilitated by domain II that are downstream of 48S assembly are also affected upon domain II deletion. The conformational change induced by domain II in the head region may assist eIF5, a GTPase-activating protein, and eIF2-TC interaction on the 40S subunit. This contact would initiate GTP hydrolysis and subsequent release/dissociation of eIF2-GDP, after the establishment of AUG codon recognition by initiating Met-tRNA$_i$. A reduction in 80S assembly compared to the WT was also encountered when conserved structural motifs of domain II were deleted individually (Locker et al., 2007). The release of inorganic phosphate (Pi) in cap-dependent translation is dependent on AUG initiation codon recognition in eIF5-promoted eIF2-GTP hydrolysis, with eIF1 being a negative regulator of this Pi release (Algire et al., 2005). Once eIF2-GDP is released from the complex, the catalysis of second GTP hydrolysis for the displacement of eIF3 is driven by eIF5B, thereby promoting the formation of the 80S pre-initiation complex by allowing subunit joining (Lee et al., 2002) (Figure 1). The subnanometer modelling of the structure 80S/HCV-IRES-Met-tRNA$_i$-eIF5B-GMPPNP has also revealed the conformational dynamics of domain II which showed release of its apical part from

binding site of the 40S subunit and interacting with the elbow of tRNA upon 60S ribosome joining (Yamamoto et al., 2014).

3.2 IRES DOMAIN III CONSERVED REGIONS ARE IMPORTANT FOR 40S AND eIF3 BINDING

3.2.1 The domain III apical region is crucial for recruitment of eIF3 and for efficient translation

Domain III is a principal domain of the HCV-IRES and is further sub-divided into domains IIIa-IIIf. The structural elements of domain III are involved in making most of the direct contacts with the host translational machinery, 40S and eIF3, as observed through a variety of experimental approaches, including biochemical, mutation and structural analyses (Kieft et al., 2001, Lukavsky et al., 2000, Jubin et al., 2000).

The apical half of domain III comprises domains IIIa, IIIb, IIIc and a four-way junction (jIIIabc). Mutations in domain IIIa were found to affect the stability of the binding affinity with eIF3. Substituting the IIIa loop causes a >6-fold reduction in the eIF3 binding affinity, with loss of HCV-IRES function (Kieft et al., 2001, Ji et al., 2004). It was demonstrated through toe-printing, mutagenesis, and chemical and enzymatic footprinting that for eIF3/HCV-IRES binding, domain IIIb and junction IIIabc are vital determinants of this interaction (Figure 1) (Pestova et al., 1998, Sizova et al., 1998, Buratti et al., 1998, Kieft et al., 2001). Domain IIIb comprises an apical loop and an internal loop variable among HCV isolates that adopts a conserved three-dimensional secondary structure, observed through NMR. The interhelical region, consisting of a C^{186}•C^{211} mismatch and adjacent base pairs G^{184}-C^{213} and A^{185}-U^{212}, is conserved (Figure 2) and mutations in this region leads to low levels of translation. Variations at nucleotides position 182 and 217 result in a range of activities, from as low as 33% to activity equivalent to the WT, which also suggests that sequence preservation in this region might be important for eIF3 recognition. Among the domain IIIb most common variants that were structurally analyzed, C^{183}•A^{214} and A^{183}-U^{214} adopt a similar structural conformation with a signature S-turn and display an efficient HCV-IRES response, contrary to mutations at A^{183}-G^{214} pair

that alter the geometry of the structure within the loop. This finding may suggest that the interaction of eIF3 with this region is more dependent on the geometry and characteristic of its backbone rather than the primary sequence (Collier et al., 2002).

The primary sequence of the domain IIIb apical loop may not be vital for HCV-IRES function as revealed by mutagenesis (Wang et al., 1994b). Deletion of domain IIIb causes a reduced affinity for eIF3, which has been shown to affect the stable association of eIF2 with the 48S complex and the deposition of Met-$tRNA_i$ to the AUG binding site. Domain IIIb deletion mutants may assemble an HCV-IRES/80S complex but do not do so as efficiently as the WT (Ji et al., 2004, Otto and Puglisi, 2004). Similarly, a low translation response upon deletion of the domain IIIb apical loop has also been reported (Buratti et al., 1997). Any nucleotide insertion between nucleotides at positions 206 and 207, adjacent to the 3' end of the IIIb apical loop has been shown to increase HCV-IRES activity in both monocistronic and bicistronic luciferase expression systems compared to the WT. The inserted nucleotide at position 207 is speculated to increase the eIF3 binding affinity for the HCV-IRES, which might be responsible for the increase in cap-independent translation (Zhang et al., 1999)**.**

The helical junctions in HCV-IRES interact with 40S and eIF3 by providing specific recognition sites, as observed using various biochemical assays. Junction IIIabc (jIIIabc) plays an important role in the interaction with eIF3 and 40S (Kieft et al., 2001). Structural analysis using X-ray crystallography determined the arrangement of junction nucleotides and the adjacent helices that provide the specific binding site for eIF3. Junction IIIabc adopts a parallel orientation in which helix IIIc is stacked on the rest of stem III in an almost perfectly coaxial manner. Helixes IIIa and IIIb have disrupted stacking due to the insertion of junction residues (A^{154}, A^{155} and U^{228}) into the minor groove (Figure 2). These features of the junction suggest that the distortion of the helical structure is important for providing a recognition site for eIF3 (Kieft et al., 2002). The introduction of mutations at the junction has been shown to be devastating for HCV-IRES efficiency, as it apparently destroys the fine intricacy of the hydrogen bonding between the junction fold and the adjacent nucleotides. The mutant U^{228C} achieves assembly of the 48S complex with a >15-fold reduced eIF3 binding affinity and a significant reduction in 80S formation, leading to a decrease in IRES activity (Kieft et al., 2002, Kieft et al., 1999, Ji et al., 2004). Junction IIIabc

interacts with the 40S subunit body near expansion segment 6 in HCV-IRES/40S and HCV-IRES/80S complexes. A shift from a parallel to antiparallel orientation of coaxially stacked helices between domain III and IIIb in jIIIabc was suggested for HCV-IRES/80S (Boehringer et al., 2005). The presence of two different parallel and antiparallel conformations of jIIIabc has also been shown to exist in solution using time-resolved fluorescence measurements (Melcher et al., 2003).

Domain IIIc is a conserved region with a stem-loop structure composed of 10 nucleotides. The stem constitutes a three G-C base-pair, along with a tetra-loop (CGUG). The interaction of domain IIIc with eIF3 was established through chemical and enzymatic footprinting. The deletion of stem loop IIIc affects the eIF3 binding stability to this region (Sizova et al., 1998). Using NMR, the first and fourth nucleotides of the tetraloop, C^{232} and G^{235}, were revealed to be a Watson-Crick base pair. The center nucleotides of the loop, G^{233} and U^{234}, are the only residues in the structure that adopt a C2' -endo conformation. The substitutions of G^{233} and U^{234} in the apical loop are well-tolerated. An examination of the mutations at base pair C^{232}-G^{235} showed that conservation of the structure through usual Watson-Crick base-pairing maintains the translation response similar to that of the WT, while other mutations have deleterious effects (Rijnbrand et al., 2004). However, Tang and colleagues (Tang et al., 1999) showed that the compensatory base pair mutations $C^{232}G/G^{235}C$ and $C^{232}U/G^{235}A$ restore only 25% and 40% of the HCV-IRES activity, respectively, which suggests that the primary sequence is equally important to the maintenance of the secondary structure. The stacking of the bases and hydrogen bonding at C^{232}-G^{235} provide stability for this domain (Rijnbrand et al., 2004, Tang et al., 1999).

3.2.2 Monitoring the effects of domain III basal region mutations on 40S/IRES-RNA interaction and ribosome assembly

Data obtained by chemical probing of domains IIId and IIIe demonstrated the importance of the stem loop residues involved in the 40S subunit interaction. Structural analysis using NMR further revealed the configurations of domains IIId and IIIe. A helical stem, an internal loop and an apical hexanucleotide loop constitute

domain IIId. The internal loop is conserved among HCV isolates, and it folds into an E-loop motif. The inversion of backbone direction at A^{257} and G^{258} leads to an S-turn formation, characteristic of an E-loop motif, that further allows the parallel hydrogen bonding of A^{257}•A^{275} (Lukavsky et al., 2000). Transition and transversion at A^{276} causes a reduction in the HCV-IRES response and substituting ^{275}CUC^{277} for AAG generates only 25% activity relative to the WT. Moreover, substituting A^{257}G and A^{275}G of the loop E-motif in combination distorts translation almost completely, with no 48S and 80S formation (Otto and Puglisi, 2004).

The terminal loop of domain IIId (5'-UUGGGU-3') is one of the most important conserved regions of the HCV-IRES. The ^{266}GGG^{268} is strongly conserved and is essential for HCV-IRES translation in all HCV genotypes and HCV-like IRESs (Hellen and de Breyne, 2007)**.** The nucleotide U^{269} is looped out into the solution, favoring the positioning of G^{268} towards the major groove and an inversion in the backbone, which causes the formation of another S-turn between G^{267} and C^{270} (Figure 2). The mutation of ^{266}GGG^{268} to AAA to preserve the structure still results in a 50% decrease in translation. Any disruption of the IIId terminal loop causes a low translation response (Lukavsky et al., 2000, Klinck et al., 2000). The functional capacity of HCV-IRES translation has been shown to be seriously challenged upon mutations in this domain (Jubin et al., 2000, Ji et al., 2004, Kieft et al., 1999). A molecular dynamics simulation followed by circular dichroism (CD) spectroscopy on a G^{266}A/G^{268}U mutant showed a similar structural conformation in domain IIId as that of the WT and at different magnesium concentrations. The activity of this mutant in translation assays, however, was abrogated, which highlighted the importance of conservation of the primary sequence of the domain IIId hexaloop (Barria et al., 2009).

The HCV-IRES contacts the 40S ribosomal subunit at multiple interaction sites that are specific and important. The basal portion of domain III, particularly domain IIId, and domain IV have been observed to be the interaction sites for the 40S subunit. The majority of nucleotides in domain IIId have been shown to be protected from RNase V1, RNase T1 and iodine cleavage upon 40S binding (Kieft et al., 2001, Kolupaeva et al., 2000b). Domain IIId is proposed to contact the 40S subunit near helix 26, which is extended to expansion segment 7 (ES7). The cross-linking experiment also suggests that there are contact points at ribosomal proteins S14 and

S16 (cross-linked to A^{275} or G^{263}). This interaction of IIId-40S is the most extensive and provides stability to the structure (Babaylova et al., 2009, Boehringer et al., 2005). The ^{266}GGG^{268} in domain IIId has been shown to contact 18S rRNA through a ^{1116}CCC^{1118} sequence in the apical loop of ES7 with complementary base-pairing, as analyzed through dimethyl sulphate (DMS) modification (Malygin et al., 2013a) and also demonstrated functionally (Matsuda and Mauro, 2014). A cryo-EM structure of the HCV-IRES bound to 40S ribosome at 3.9 Å has also displayed the specific contact sites of the HCV-IRES domain IIId loop forming a kissing complex within the apical loop of ES7, reinforced by the interaction with domain IIIe (Quade et al., 2015, Yamamoto et al., 2015). Recently, structure probing data such as selective 2'-OH acylation analysed by primer extension (SHAPE) and footprint analysis together with molecular modelling was employed to visualize and reveal the contact sites of domain IIId loop and the 18S rRNA. The relative reactivity of WT (40S/HCV-IRES) and the IIId loop mutants with 40S observed through structural probing alongside 3D model led to the conclusion that domain IIId loop interacts directly with the ribosomal h26 of ES7 and is crucial in coordinated structural re-arrangements of HCV-IRES/18S rRNA upon formation of a binary complex that facilitates HCV mRNA translation (Angulo et al., 2016). Remarkably, a motif conserved among caliciviruses containing three guanosine residues were also observed to interact with h26 stem-loop (CCC) of the 18S rRNA important for translation reinitiation (Zinoviev et al., 2015, Luttermann and Meyers, 2009).

Domain IIIe consists of a tetraloop (5'-GA[U/C]A-3') that is conserved among HCV isolates and HCV-like IRES RNAs. The NMR and crystal structures of this domain display base-pairing of the bottom nucleotides, G^{295} and A^{298}, of the tetraloop. The conformation of the IIIe hairpin alone observed in NMR is distinct from the X-ray crystallographic IIIe tetraloop structure, shown with a pseudoknot and stem III junction, most likely due to its tertiary interaction (Figure 2). The overall configuration of this tetraloop is different (in the positioning and stacking of the bases) from the standard GNRA tetraloop (Lukavsky et al., 2000, Berry et al., 2011). The central nucleotide, U^{297}, flips out of the tetraloop, and it base-pairs with A^{288} from stem III, which is also bulged outward (Figure 2). This tertiary interaction constitutes a second pseudoknot in the HCV-IRES (Berry et al., 2011). The mutation analysis demonstrates that the conservation of a canonical base-pair at 288-297 is important

for HCV-IRES function in translation initiation (Easton et al., 2009). However, it was also demonstrated that the maintenance of purine-pyrimidine—not necessarily the Watson-Crick base pairing—interaction between A^{288} and U^{297}, generates HCV-IRES activity similar to that of wild type and that altering the orientation of this identity disrupts translation. In the context of an existing destabilized IIIf pseudoknot with 40% activity, the compensatory mutations restore the translation activity to 45% of that of the WT (Berry et al., 2010, Berry et al., 2011). Another bulging nucleotide (A^{136}) that is flipped out from stem III is stacked over A^{296} and A^{298} of the IIIe tetraloop. Substituting this nucleotide causes no hindrance to HCV-IRES function (Berry et al., 2011). Interestingly, another study reported that in HCV genotype 1b, the presence of nucleotide substitution $A^{136}G$ along with another mutation, $A^{140}U$, decreases the translation activity to merely 35% relative to that of the WT (Barria et al., 2009).

Deletions and substitutions in the IIIe tetra loop result in a severe deterioration of the translation response from 60-90% in all types of transitions and transversions, demonstrating the importance of the conservation of the sequence of each of the four nucleotides (Psaridi et al., 1999). Using a site-specific crosslinking method, hairpin IIIe has been identified in proximity to ribosomal proteins S3a, S5 and S16 on the solvent side of the 40S subunit (Laletina et al., 2006).

The domain III basal region is composed of another helical junction that comprises a pseudoknot structure. The pseudoknot in domain IIIf, located upstream of the AUG initiation codon, is highly conserved, and maintenance of its secondary structure is necessary for HCV-IRES function (Wang et al., 1994b, Wang et al., 1995). The pseudoknot domain configuration was revealed using computational modelling and X-ray crystallography (Lavender et al., 2010, Berry et al., 2011). The junction arranges itself into a unique, complex double-pseudoknot fold that further allows the formation of two helices, coaxially stacked and non-parallel with a tilt of approximately 40° between the helical stacks. The significance of the pseudoknot has been determined through mutation and functional analyses that revealed the pseudoknot's contribution to the translational activity and its importance in properly positioning the AUG initiation codon at the 40S binding site (Berry et al., 2011, Berry et al., 2010, Kieft et al., 2001, Wang et al., 1995). The pseudoknot contacts the helices 28, 37 and 40 at the back of the 40S subunit head region (Boehringer et al.,

2005). Unspecific interactions of pseudoknot in domain IIIf with ribosomal protein eS28 at the exit of the ribosomal mRNA binding groove has also been reported (Quade et al., 2015, Yamamoto et al., 2015).

The characterization of the HCV IRES structures was also performed with the development and combination of small-angle x-ray scattering data (SAXS) and molecular modeling that gave rise to the ensemble of conformers for the full-length HCV-IRES RNA (Perard et al., 2013).

3.3 CHANGES IN SEQUENCE COMPOSITION DOWNSTREAM OF AUG CAN ALTER THE HCV-IRES ACTIVITY

The initiator AUG codon of the HCV-IRES is localized in a stem-loop IV (Rijnbrand et al., 1996, Reynolds et al., 1996); a conserved domain among the HCV genotypes and distant GB virus B. The chemical and enzymatic probing data have demonstrated the interaction of domain IV with the 40S ribosome (Kolupaeva et al., 2000b, Kieft et al., 2001). The nucleotide changes in domain IV that resulted in an increased stabilization of this domain exhibit translation with decreased efficiency (Honda et al., 1996a). As it binds to the mRNA binding cleft of the ribosome, domain IV is unwound, due to its increased flexibility, to support the correct positioning of the initiation codon and the subsequent binding of eIF2-TC (Filbin and Kieft, 2011). Evidence for a possible long-range structural and functional interaction between domain IIb and domain IV has been discussed earlier in the text.

In our previous work, we studied the necessity of the AUG codon at position 342 in yeast and mammalian systems by introducing mutations. The substitution of UUG for the AUG codon resulted in a substantial inhibition of the IRES-mediated response in both mammals and yeast (Masek et al., 2007). The sequence conservation of nucleotides located downstream of the AUG initiation codon has been proposed to be very crucial for the maintenance and modulation of HCV-IRES activity (Reynolds et al., 1995, Honda et al., 1996b). Our results suggested that a high AT content within the first 15 codons of the HCV polyprotein is important in maintaining the HCV-IRES activity. The study reflects the evolutionary conservation of basic translation processes by displaying similarities of HCV-IRES translation between two unrelated organisms, such as yeast and humans (Masek et al., 2007).

However, no specific requirement has also been reported for either the nucleotides or amino acid sequences downstream of the IRES for efficient translation in HCV and other flaviviruses (Rijnbrand et al., 2001). The core gene stem loops SL47 and SL87 are highly structured elements, which have been concluded to contribute to HCV-IRES translation in JFH-1 viral strain. The introduction of silent mutations in SL47 and SL87 separately and simultaneously led to a decreased translation of the viral RNA and pointed to the importance of these structural elements for robust viral production both *in vitro and in vivo* (Vassilaki et al., 2008). The stability of the long-range RNA-RNA interaction between the HCV-IRES (nucleotide (nt.) 24 to 38) and the core gene (nt. 428 to 442) was shown to be involved in modulating viral gene expression as well. Mutation analysis revealed that disruption in the stability of this interaction increases the HCV-IRES translation. In contrast, compensatory mutations that restore stability led to slightly reduced translation activity compared to the WT, both *in vivo* and *in vitro*. This reduced translation may be important for viral persistence during chronic infections (Kim et al., 2003, Wang et al., 2000). A region between domain I and domain II (nt. 22-28) is also shown to be the site of attachment for the liver miRNA-122, which may cause interference with the long-range interaction between the HCV-IRES and the core gene, and stimulate the HCV translation, although under different conditions, *in vitro* and *in vivo* (Henke et al., 2008, Goergen and Niepmann, 2012). Annealing of miRNA-122 induces a switch from close to an open conformation of HCV RNA, as observed *in vitro* (Diaz-Toledano et al., 2009).

3.4 STABILIZATION OF THE TRANSLATION COMPLEX WITH CHARACTERISTIC BINDING OF eIF3 SUBUNITS TO THE IRES

Function of the HCV-IRES has been repeatedly reported to depend on the specific interaction with the eukaryotic translation initiation factor 3 (eIF3). Mammalian eIF3 is a 13-subunit complex (subunits eIF3a-m) of ~800 kDa (Damoc et al., 2007). Cryo-electron microscopy (EM) revealed the structure of eIF3 and its arrangement on the HCV-IRES/40S subunit. The interaction occurs through the front face of eIF3, a 5-lobed structure consisting of a head, arms and legs. The left arm is composed of an

extended portion located near the HCV-IRES domain II towards the E-site, while the left leg covers S15/rpS13 below the 40S platform blocking its contact with helix 34 of the 60S in order to prevent premature ribosomal subunit association. The domains IIIdef and IIIabc are located near the center and right leg of eIF3 (Siridechadilok et al., 2005).

A number of eIF3 subunits interact with the HCV-IRES, as observed using limited proteolysis and mass spectrometry (MS)**.** The subunits eIF3a, eIF3c, eIF3e, eIF3f, eIF3h and eIF3l reside mostly in the right and left legs of eIF3 and are responsible for high-affinity binding with the IIIabc domain of the HCV-IRES (Cai et al., 2010). However, other studies suggest that the eIF3b and eIF3g subunits are also involved in a direct interaction, whereas the eIF3i and eIF3l:k dimer is proposed to have no immediate contact with the HCV-IRES (Zhou et al., 2008, Perard et al., 2009, Yu et al., 2005).

The association of eIF3 with the HCV-IRES places the C-terminus of eIF3j, a subunit of eIF3, in the 40S mRNA entry channel. The dissociation of eIF3j is necessary for the proper placement of HCV mRNA in the decoding groove. Directed hydroxyl radical probing and toe-printing showed that the dissociation of eIf3j is promoted by a conformational change induced by domain II in the head region of the 40S subunit. Moreover, the recruitment of eIF2-TC is needed to displace the eIF3j, providing more stability to the HCV mRNA in the 40S mRNA binding channel (Fraser et al., 2009). Using cryo-EM, it has been proposed that in the HCV-like CSFV IRES (classical swine fever virus IRES), the eIF3 is displaced from the 40S subunit observed in the 43S pre-initiation complex. Instead, eIF3 interacts with the apical region of domain III, presenting its ribosome-binding surface, eIF3a and eIF3c, thus forming a 40S-IRES-eIF3 complex (Hashem et al., 2013b). The binding of ribosomal protein eS27 with domains IIIa and IIIc occur through specific interactions as shown in 3.9 Å cryo-EM data (Quade et al., 2015). The same ribosomal protein eS27 makes contact with eIF3 complex at the opposite face, in canonical pre-initiation (Hashem et al., 2013a), compared to the HCV-IRES. The IRES-eS27 contact presumably orients HCV-IRES towards the same region above the ribosome as eIF3 and consequently displaces eIF3 from the 40S (Quade et al., 2015), as observed with the above mentioned CSFV-IRES. Another study using low-resolution EM along with

biochemical analysis revealed that conserved regions in eIF3a and eIF3c directly bind with domain IIIabc of the HCV-IRES (Sun et al., 2013).

3.5 eIF2-INDEPENDENT TRANSLATION MECHANISM (STRESS-INDUCED HCV PROTEIN SYNTHESIS)

The protein kinase R (PKR) undergoes activation during HCV infection which leads to eIF2 inhibition and a consequential decrease in protein synthesis, including that of the antiviral interferon-stimulated genes (ISG), in the host cells. While suppressing host protein synthesis by activating PKR to its own advantage, the HCV-IRES persists in translation (Garaigorta and Chisari, 2009) through an eIF2α-independent mechanism. The HCV-IRES domains III-IV have been found to be critical for PKR activation (Shimoike et al., 2009). Domain II has also been reported to act as a potent PKR activator (Toroney et al., 2010). The ability of the HCV-IRES to undergo an alternate bacteria-like mode of translation upon eIF2 inactivation only requires two initiation factors, eIF3 and eIF5B-GTP, which are necessary for 80S pre-initiation assembly. The initiation factor eIF5B was shown to be required to direct Met-tRNA$_i$ to the HCV-IRES/40S complex before 60S subunit attachment and provides stability without any direct binding (Figure 1) (Terenin et al., 2008). The role of eIF5B, in addition to the displacement of eIF3 and other initiation factors in cap-dependent translation, has also been implied in assisting the subunit joining of 60S to 40S, forming an 80S complex (Pestova et al., 2000, Unbehaun et al., 2004). GTP is required because it provides an active conformation for eIF5B (Roll-Mecak et al., 2000). Another factor, eIF2A, is also proposed to help with the recruitment of Met-tRNA$_i$ to the P-site by binding to domain IIId. By using a filter-binding assay, eIF2A was shown to have a binding affinity for Met-tRNA$_i$. Moreover, the deletion of domain IIId or mutations in the IIId hexaloop cause a low binding capability with eIF2A, thus affecting translation activity under stress conditions. eIF2A knockdown using siRNA has an inhibitory effect on HCV-IRES translation under conditions where the eIF2α subunit is phosphorylated (Kim et al., 2011). eIF2D is also one of the factors that have been reported to facilitate Met-tRNA$_i$ assembly with the 40S under stress conditions (Figure 1) (Dmitriev et al., 2010). The alternate mode of translation occurs only when the eIF2α subunit is phosphorylated under stress conditions to avoid the

severe inhibition of viral protein synthesis. Both the HCV and related CSFV-IRES-mediated translation were shown to be resistant to the reduced levels of eIF2-TC (Robert et al., 2006, Pestova et al., 2008, Terenin et al., 2008, Kim et al., 2011)

The HCV-IRES can also constitute a pre-initiation 80S translation complex without the requirement of any of the initiation factors at a higher Mg^{2+} concentration of 5 mM and continue protein synthesis with the aid of elongation factors and tRNAs (Lancaster et al., 2006).

4 MATERIALS and METHODS

PROTOCOLS

4.1 Bacterial strain

E. coli XL-1 Blue: *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´ *proAB lacIqZ_M15* Tn*10 (*Tetr*)*] (Stratagene)

4.1.1 Culturing and handling the bacterial strains

The bacterial colonies were selected and inoculated in a culture medium for the purpose of cultivating bacterial cells in order to isolate DNA plasmid on a larger scale. The liquid medium and agar plates used were 2xTY with appropriate antibiotic (ampicillin 100 µg/ml or kanamycin 50 µg/ml). All the cultivations were carried out at 37^{o}C.

4.1.2 Cultivation media

2xTY: 1.6% peptone; 1% yeast autolysate; 0.5% NaCl; adjust to pH 7.0 with NaOH; antibiotics in selective media (ampicillin 100 µg/ml or kanamycin 50µg/ml).

4.1.3 Agar plates

2xTY: 2xTY medium, 2% agar; antibiotics (ampicillin 100 µg/ml or kanamycin 50 µg/ml)

4.2 Mammalian Cell line

HeLa (Chang Liver) (Organism: Homo sapiens, human / Tissue: HeLa contaminant). (ATCC number CCL-13™).

4.2.1 Culturing and handling the mammalian cells

The human epithelial cell line CCL13 (also known as Chang cells) were cultured in Dulbecco's modified Eagle's medium (DMEM; Sigma) supplemented with 100 units/mL penicillin, 100 µg/mL streptomycin, and 10% fetal bovine serum (GIBCO) at

37°C in 5% CO_2 and 95% relative humidity. Confluent adhesive cells were detached by trypsin-EDTA solution (Gibco) and split into 6-cm dishes at 1:4-1:8. Cells were stored in liquid nitrogen for longer duration in DMEM medium having 20% fetal bovine serum and 10% DMSO.

4.3 Screening for HCV-IRES heterogeneity in viral population(s) from hepatitis C virus infected patients

4.3.1 Full-length HCV-IRES library and recombinant plasmid

RNA isolated from patients (4, 7 and 9) suffering from the hepatitis C virus was reverse transcribed to cDNA by using SuperScript® III Reverse Transcriptase (Invitrogen) and reverse BamHI primer (Table 2). Full-length HCV-IRES fragments (1-385 nt.) were successfully amplified by PCR using HCV-IRES specific BamHI and SalI primers (Table 2) and cloned into a bicistronic vector pRG. The vector comprises two reporter genes DsRED2 and EGFP which were under the influence of CMV (cytomegalovirus) promoter (Figure 3). DsRED2 being the first cistron is translated cap-dependently while EGFP2, the second cistron, carries out its protein synthesis via cap-independent translation. The library of full-length HCV-IRES fragment(s) in pRG vector signifies the HCV-IRES population present in clinical sample at a given time (Masek et al., 2007).

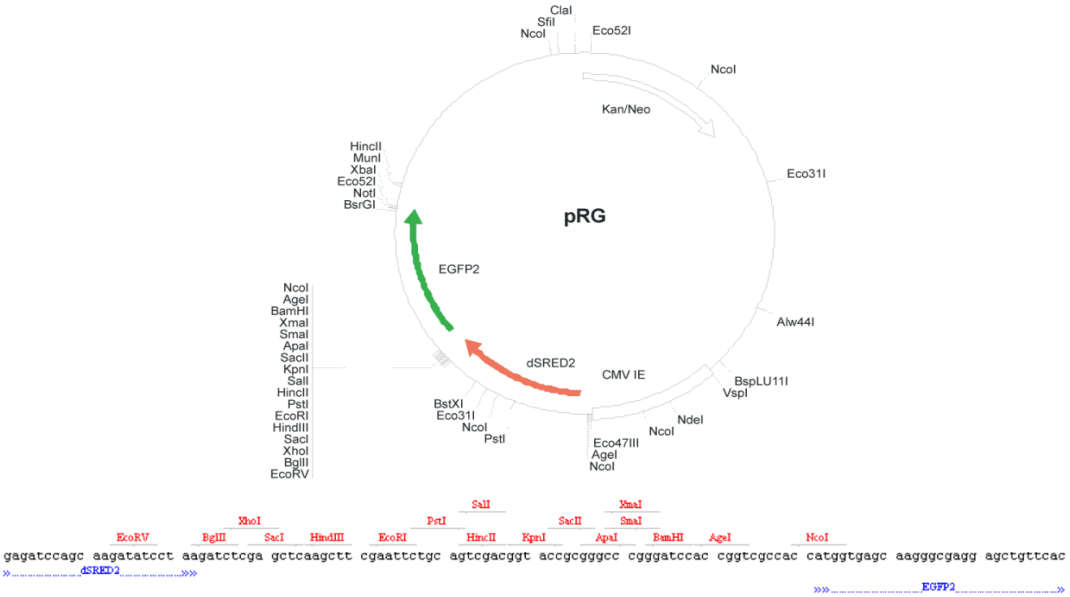

Figure 3: Map of pRG vector with polylinker.

4.3.2 Transformation of *E.coli* by electroporation

We used *E.coli* XL-1 blue competent cells to carry out electroporation. Immediately prior electroporation plasmid sample/ligation mix 1-5 µl was added to 50 µl aliquot of *E.coli* XL-1 blue electrocompetent cells and was kept on ice for 1 min. Electroporator (BIO-RAD GenePulser Xcell^{TM}) was set to 2.5 kV, 200 ohm resistance and 25 µF capacitance. Plasmid sample/XL1-blue mix was added to curvette (0.2 cm electrode gap), electroporated and then 1 ml of 2xTY non-selective growth medium was added immediately after to curvette containing the cell suspension. The contents of curvette were transferred to the eppendorf tube and incubated at 37°C for 30-60 minutes with shaking. The cells were later plated onto desired medium.

Clones from the HCV-IRES library of three patients (4, 7 and 9) were spread on the kanamycin resistant 2xTY agar plates. The plates were incubated overnight at a temperature of 37°C. The plates were checked for the growing colonies the next day.

4.3.3 Plasmid minipreparation from *E. coli*

This protocol was used for quick and easy isolation of plasmids from bacterial cells. First step involves the transfer of maximum bacterial biomass to 400 µl of STET buffer (avoid of agar transfer to eppendorf tube) and deeply resuspend. 5 µl of freshly prepared solution of lysozyme (5 %) was addded and mixed by vortexing for 5 sec. The tubes were placed in a boiling water bath for exactly 40 seconds or incubate for 2 min in dry bath incubator at 95 °C. The lysate was centrifuged at 13,000 g for 10 min at room temperature. The glob of bacterial debris was carefully removed with a sterile toothpick, 400 µl volume of isopropanol was added to the supernatant and mixed by vortexing. The tubes were stored for 30 min at -20°C and centrifuged at 13,000 g for 10 min. The supernatant was removed, pellet washed with 1 ml of 70% ethanol. Pellets were allowed to dry at room temperature and resuspend in 30 µl TE buffer. Samples were analyzed on agrose gel electrophoresis.

STET buffer: 10 % sacharose; 50 mM Tris-HCl pH 8; 50 mM EDTA-NaOH pH 8; 1 % Triton X – 100.

TE buffer: 10 mM Tris-HCl pH 8; 1 mM EDTA-NaOH pH 8.

4.3.4 Plasmid isolation from *E.coli* (Midiprep)

For extraction of DNA plasmid on a larger scale from bacterial cells we used this protocol. It involved the inoculation of 50 ml sterile 2xTY medium containing antibiotic with a single bacterial colony and was grown to saturation (overnight). 50 ml of cells were centrifuged for 10 min at 5000 rpm to pellet. The supernatant was removed. The pellet was resuspend in 2 ml TEG solution and incubated for 5 min at room temperature. 4 ml of cell lysis solution was added and tubes were inverted a few times and were place on ice for 5 min. Tubes were inverted few times after adding 3 ml 3 M potassium acetate solution and placed on ice for 20 min. The cell lysate was centrifuged for 30 min at 5000 rpm to pellet cell debris and precipitate chromosomal DNA. Supernatant was transferred to a fresh tube and mixed it with 0.7 volume of isopropanol and incubate 10 min at 4 °C to precipitate nucleic acids. Centrifugation was carried out for 30 min at 4°C to pellet plasmid DNA. Supernatant was removed and the pellet was washed with 3 ml of 70 % ethanol and dried at room temperature. Pellet was re-suspended in 400 µl TE buffer. 3 µl of RNase A solution was added and incubate for 60 min in 37°C. Samples were extracted by adding 400 µl of phenol (pH 8), 400 µl of phenol/chloroform and 400 µl of chlorofom/isoamyl alcohol to the DNA solution to be purified in a 1.5-ml microcentrifuge tube. Vigorous vortexing was performed for 30 sec and samples were microcentrifuged for 5 min at maximum speed and room temperature. The top (aqueous) phase containing the DNA was carefully removed using a 100 µl pipettor and transferred to a new tube. 1/10 vol of 3 M sodium acetate, pH 5.2 and 2.5 to 3 volume of ice-cold 96% ethanol was also added and mixed by vortexing before placing in -20 °C overnight. Centrifugation for 10 min at maximum speed was attained and supernatant was removed. With addition of 1 ml ice-cold 70 % ethanol tubes were inverted several times and centrifuged for 10 min at maximum speed. The supernatant was removed and pellet was dried at room temperature. Plasmid DNA was resuspended in an appropriate volume of TE buffer (300 µl) and analyzed on agarose gel electrophoresis.

TEG: 25 mM Tris-HCl pH 8; 10 mM EDTA-NaOH pH 8, 50 mM glucose

Cell lysis solution: 1% SDS; 0.2 M NaOH

TE buffer: 10 mM Tris-HCl pH 8; 1 mM EDTA-NaOH pH 8

RNase A solution: Ribonuclease A 110 mg/ml; 0.1 mM Tris-HCl pH 7.5; 0.045 mM NaCl

4.3.5 Agarose gel electrophoresis of DNA

Visualization of all DNA samples was carried out employing DNA agarose gel electrophoresis. Different percentages of agarose gels (varying concentration according to size of DNA fragment) were prepared by resuspension of agarose in 1x TAE buffer and heated until agarose is dissolved completely. The solution was cooled around 50°C and poured in a casting tray with addition of EtBr (final concentration 0.175 µg/ml). A comb was placed and let the gel solidify at room temperature.
50x TAE: 2 M Tris; 1 M acetic acid; 0.5 M EDTA-NaOH pH 8; adjust final volume to 1l with ddH_2O.

4.3.6 Restriction Digestion for insert verification

For confirmation of desired insert in the vector we carried out restriction digestion. All restriction enzymes and corresponding buffers used were from Fermentas. The pRG vector was digested using SalI and BamHI restriction endonucleases. These restriction sites were added to the primers and desired fragment was amplified using optimized PCR conditions. The insert was ligated in pRG vector and further confirmation was carried out by restriction digestion.

Plasmid (500 ng)	2 µl
BamHI (10 U/µl)	0.5 µl
SalI (10 U/µl)	0.5 µl
10x Tango buffer (yellow)	3 µl
RNAse (20 µg/ml)	0.5 µl
ddH_2O	5.5 µl

- Reaction was incubated at 37°C for 2 hours. The presence of insert was further confirmed using gel electrophoresis.

4.3.7 Polymerase chain reaction (PCR)

In all PCR reactions, 10µM primers and 10µM deoxyribonucleotide triphosphates (dNTPs; 10µM dATP, dCTP, dGTP and dTTP in water) were used. Precise quantities of all reagents, temperatures and incubation times are mentioned in each protocol separately. I was using the Biorad C1000™ Thermal cycler or eppendorf *Mastercycler* ep*gradient S*.

4.3.8 Colony PCR

The protocol was used to screen for colonies with desired plasmid that were transformed. Bacterial colonies were selected from the agar plates that were grown overnight with selective antibiotics and transferred to an already filled 300 µl ddH_2O in a tube and mixed. The tubes were vortex and incubated at 96°C for 10 min. The tubes were allowed to cool down at room temperature for 5 min. The tubes were heated again in incubator at 96°C for 10 min and later used it as a template in PCR reaction. We used GC clamp primers (forward) for further usage of this PCR product in DGGE and TGGE analyses.

Name	Sequence 5' – 3'
HCVIRESf-Sal-clamp	CGCCCGCCGCGCCCCGCGCCCGTCCCGCCGCCCCCGCCCGGTCGA CGCCAGCCCCCTGATGGGGGCGACAC
Reverse BamHI	GGATCCGTGTTACGTTTGGTTTTTCTTTGAGGTTTAGG

Table 1: GC- clamp primer sequences (forward and reverse).

For the amplification of HCV-IRES fragment in pRG vector we also used primers without GC-clamp while employing similar PCR reaction volume and conditions.

Name	Sequence 5' – 3'
Forward SalI	AAAGTCGACGCCAGCCCCCTGATGGGGGCGACAC
Reverse BamHI	GGATCCGTGTTACGTTTGGTTTTTCTTTGAGGTTTAGG

Table 2: Primer sequences (forward and reverse) for PCR amplification of HCV-IRES fragment from pRG vector.

Template (plasmid) (1ng)	1 µl
Clamp primer, forward (10 pmol)	1.25 µl
Reverse primer (10 pmol)	1.25 µl
dNTPs (10 mM)	0.5 µl
Taq polymerase (2 U/µl)	1 µl
10x PCR buffer	2.5 µl
ddH_2O	17.5 µl

95°C	3 min	
94°C	30 sec	25 X
56°C	30 sec	
72°C	45 sec	
72°C	10 min	

4.3.9 Denaturing Gradient Gel Electrophoresis (DGGE)

The DGGE was performed following the manufacturer's instructions (INGENYphorU-2). The conditions for the experiments were optimized using 6% separating polyacrylamide gel (30% acrylamide / 0.8% bisacrylamide). The denaturanting gradient for polyacrylamide gel ranged from 55% to 70%. To prepare the 6% polyacylamide denaturant gradient gel, two solutions were prepared containing 55% denaturnats (4.6 ml of 30% acrylamide / 0.8% bisacrylamide, 5.32 g of final 7M urea concentration, 5.06 ml of 40% formamide and fill with ddH_2O to make 23 ml volume) and 70% denaturnats (4.6 ml of 30% acrylamide/0.8% bisacrylamide, 6.77 g of final 7M urea concentration, 6.44 ml of 40% formamide and fill with ddH_2O to make 23 ml volume), respectively. 68 µl of ammonium persulfate (10%) and 31.5 µl of TEMED was added in both the solutions to polymerize the gel. Immediately afterwards, the solutions were poured in the left and right compartments of the gradient maker and stirred. The gel was casted in between the glass plates' sandwich according to the manufacturer's instructions. After 20-30 min of polymerization the stacking gel composed of 2 ml acrylamide (30% acrylamide / 0.8% bisacrylamide), 0.1 ml of 0.5X Tris/Acetate/EDTA (TAE) buffer, 7.9 ml ddH_2O; 31.5 µl ammonium persulfate (10%) and 19 µl TEMED, was poured. The combs were placed in between the glass plates and gel was allowed to polymerize for 2-3 hours. The electrophoresis cassette was placed in a buffer tank filled with 0.5x TAE and the gel was run at 100V for 17 hours at 60°C. The gel was stained for 10 minutes with ethidium bromide (final

concentration 0.175 µg/ml) and further washed with water for another 10 minutes before taking images.

30% acrylamide / 0.8% bisacrylamide: Mix 30.0 g acrylamide and 0.8 g *N,N'*-methylenebisacrylamide with ddH_2O in a total volume of 100 ml. Filter the solution through a 0.45-µm filter and store at 4°C in the dark.

4.3.10 Temperature Gradient Gel Electrophoresis (TGGE)

The TGGE was performed following the instructions of TGGE MAXI System (Version 3.02) from Whatman Biometra. The TGGE was optimized using 6% polyacrylamide gel electrophoresis. The gradient of temperature ranged from 40 to 60^oC. To prepare the 6% polyacrylamide gel with 70% denaturants we used 10 ml acrylamide (acrylamide stock; 30% acrylamide / 0.8% bisacrylamide), 21.02g urea (7M urea), 20 ml of formamide (40% formamide), glycerol 1 ml (2% glycerol) and 1 ml TAE buffer (1X TAE). The rest was filled with ddH_2O to make the volume 50 ml. 110 µl of APS (10% APS) and 120 µl of TEMED was added for polymerization of the gel. The solution was added in between the glass plates' sandwich and let it be polymerized for 2-3 hours. The samples were run for 16 hours at 100V at a temperature gradient between 40 and 60^oC.

4.3.11 Silver Staining

The gels from TGGE were further silver stained to visualize the migration of different cDNA clones of HCV-IRES. The protocols used for silver staining was from Promega. The fix/stop solution was prepared comprising of 100 ml acetic acid (10%) and 900 ml of ddH_2O. The gel was incubated in the fix/stop solution for about 20 minutes. After washing 3x2 minutes each, gel was stained for 30 minutes in a solution of 1g of silver nitrate ($AgNO_3$), 1.5 ml formaldehyde (37%) and 1l ddH_2O. After staining, the gel was put in ddH_2O briefly and transferred to the pre-chilled developing solution containing 30 g Na_2CO_3 and 3.14 g of $N_2S_2O_3.5H_2O$. The gel was incubated till bands were visible. The developing reaction was terminated by adding the remaining fix/stop solution and the gel was fixed while incubating it for another 2-3 minutes. Gel was rinsed with ddH_2O for 2x2 minutes each and dried at room temperature.

4.3.12 Sequencing reaction

The various cDNA clones containing HCV-IRES fragments that were run on DGGE and TGGE were selected. These selected cDNA constructs that have shown differences or similarities while migrating through the gel were sequenced using BigDye® Terminator v3.1 cycle sequencing kit (Applied Biosystems^{TM}). The primer sequence and reaction conditions are as follows:

Name	Sequence 5' – 3'
DsRedSeqPrimer	AGCTGGACATCACCTCCCACAACG

Table 3: Primer used for sequencing reaction.

Template	750 ng (conc.)
Reaction mix (Big Dye)	2 µl
5X sequencing buffer	4 µl
DsRED seq. primer (10 pmol)	0.5 µl
ddH_2O	12.8 µl

96°C	1 min	
96°C	30 sec	24 X
50°C	30 sec	
60°C	4 min	

The PCR product was transferred to a tube (0.5 ml). 2 µl of NaOAc and 50 µl EtOH 96% was added to the PCR product. The mixture was vortexed, incubated for 30 min and later centrifuged at 13,000 rpm for 30 min. EtOH (96%) was removed and sediment was washed with 250 µl EtOH (70%) and again centrifuged at 13,000 rpm for 15 min. EtOH was completely removed and tubes were dried at 37°C (for about 20 min) or 1 min at 90°C in an incubator. The samples were kept at 4°C and were sent to the sequencing lab to be sequenced.

4.4 Protocols: Introduction of varied codon bias by transfer RNA adaptation index (tAI) measure, and mRNA structure in HCV-core gene constructs.

4.4.1 Synthesis of HCV-core gene variants

We constructed 5 synthetic genes for the purpose of measuring the translation efficiency. These constructs translate for the same HCV-core protein containing

different synonymous codons for each construct. We selected synonymous codons whose transfer RNA adaptation index (tAI) measure was reported highest and lowest (Tuller et al., 2010) for that particular codon. These 5 synthetic genes were as follows:

1. **High tAI with loops:** High tAI values for synonymous codons of the core gene keeping the base pair maintenance of core stem-loop secondary structure. The sequences in the loop region of core gene were not changed. If the replacement of any synonymous codon disrupted the base-pairing is either not changed at all or replaced with the synonymous codon having second or third highest tAI value that will restore the base-pairing.
2. **Low tAI with loops:** Low tAI values for synonymous codons of the core gene keeping the base pair maintenance of core stem-loop secondary structure. The sequences in the loop region of core gene were not changed. If the replacement of any synonymous codon disrupted the base-pairing is either not changed at all or replaced with the synonymous codon having second or third lowest tAI value that will restore the base-pairing.
3. **High tAI without loops:** High tAI values for synonymous codons of the core gene disregarding any base pair maintenance of core stem-loop secondary structure. The codons in the loop regions were also replaced with codons having highest tAI value.
4. **Low tAI without loops:** Low tAI values for synonymous codons of the core gene disregarding any base pair maintenance of core stem-loop secondary structure. The codons in the loop regions were also replaced with codons having lowest tAI value.
5. **Core-original:** The original HCV-core gene with no modification.

4.4.2 Gene Assembly

The oligonucleotides for all the genes were synthesized from SIGMA-ALDRICH. The oligonucleotides for specific gene construct(s) were annealed, ligated and amplified using particular set of primers for each gene. The oligonucleotides for each gene and their respective primers are given in the tables below.

Oligonucleotides name	Sequence (5' – 3')
Forw_universal_1	TCCACCATGAATCACTCCCCTGTGAGGAACTACTGTCTTCACGCAGAAAGCGTCTAGCCATGGCG
Forw_universal_2	TTAGTATGAGTGTCGTGCAGCCTCCAGGACCCCCCCTCCCGGGAGAGCCATAGTGGTCTG
Forw_universal_3	CGGAACCGGTGAGTACACCGGAATTGCCAGGACGACCGGGTCCTTTCTTGGATAAACCCGCTCAA
Forw_universal_4	TGCCTGGAGATTTGGGCGTGCCCCCGCAAGACTGCTAGCCGAGTAGTGTTGGGTCGCGAAAG
Rev_universal_4	CGGCTAGCAGTCTTGCGGGGGCACGCCCAAATCTCCAGGCATTGAGCGGGTTTATCCAAGAAAGG
Rev_universal_3	ACCCGGTCGTCCTGGCAATTCCGGTGTACTCACCGGTTCCGCAGACCACTATGGCTCTCCCGGGAG
Rev_universal_2	GGGGGGTCCTGGAGGCTGCACGACACTCATACTAACGCCATGGCTAGACGCTTTCTGCGTGAAGAC
Rev_universal_1	AGTAGTTCCTCACAGGGGAGTGATTCATGGTGGAGTGTCGCCCCCATCAGGGGGCTGGCGTCGACTTT
HCVcore_1F_univ	GCCTTGTGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCACCA
HCVcore_1R_univ	GGGGCACTCGCAAGCACCCTATCAGGCAGTACCACAAGGCCTTTCGCGACCCAACACTACT
HI_wo_loops_2F	TGAGCACGAATCCTAAGCCTCAGCGTAAGACTAAGCGTAACACTAACCGTCGTCCTCAGGACGT
HI_wo_loops_3F	GAAGTTCCCTGGCGGCGGCCAGATCGTGGGCGGCGTGTACCTTCTTCCTCGTCGTGGCCC
HI_wo_loops_4F	TCGTCTTGGCGTGCGTGCTACTCGTAAGACTTCTGAGCGTTCTCAGCCTGGATCCAC
HI_wo_loops_4R	AAGTCTTACGAGTAGCACGCACGCCAAGACGAGGGCCACGACGAGGAAGAAGGTA
HI_wo_loops_3R	CACGCCGCCCACGATCTGGCCGCCGCCAGGGAACTTCACGTCCTGAGGACGACGGTTAGTGTTAC
HI_wo_loops_2R	GCTTAGTCTTACGCTGAGGCTTAGGATTCGTGCTCATGGTGCACGGTCTACGAGACCTCCC
LO_wo_loops_2F	TGAGCACGAATCCGAAACCGCAACGCAAAACAAAACGCAATACAAATCGCCGCCCGCAAGATGT
LO_wo_loops_3F	AAAATTTCCGGGTGGTGGTCAAATAGTAGGTGGTGTATATCTACTACCGCGCCGCGGTCC
LO_wo_loops_4F	GCGCCTAGGTGTACGCGCGACACGCAGTACAAGTGAACGCAGTCAACCGGGATCCAC

LO_wo_loops_4R	TTGTACTGCGTGTCGCGCGTACACCTAGGCGCGGACCGCGGCG CGGTAGTAGATATACACCACCT
LO_wo_loops_3R	ACTATTTGACCACCACCCGGAAATTTTACATCTTGCGGGCGGCGA TTTGTATTGCGTTTT
LO_wo_loops_2R	GTTTTGCGTTGCGGTTTCGGATTCGTGCTCATGGTGCACGGTCTA CGAGACCTCCC
HI_w_loops_2F	TGAGCACGAATCCTAAGCCTCAGCGTAAGACTAAGCGTAACACTA ACCGTCGTCCTCAGGACGT
HI_w_loops_3F	CAAGTTCCCGGGCGGCGGCCAGATCGTTGGCGGGGTGTACTTG TTGCCGCGTAGGGGCCC
HI_w_loops_4F	TCGTCTGGGCGTGCGCGCAACGAGGAAGACTTCTGAGCGGTCTC AGCCTGGATCCAC
HI_w_loops_4R	AAGTCTTCCTCGTTGCGCGCACGCCCAGACGAGGGCCCCTACGC GGCAACAAGTACACCCCGC
HI_w_loops_3R	CAACGATCTGGCCGCCGCCCGGGAACTTGACGTCCTGAGGACG ACGGTTAGTGTTACGCTTA
HI_w_loops_2R	GTCTTACGCTGAGGCTTAGGATTCGTGCTCATGGTGCACGGTCTA CGAGACCTCCC
LO_w_loops_2F	TGAGCACGAATCCGAAACCGCAACGCAAAACAAAACGCAATACAA ATCGCCGCCCGCAAGATGT
LO_w_loops_3F	CAAATTTCCGGGTGGTGGTCAGATTGTTGGTGGTGTATATTTGTT GCCGCGCAGAGGTCC
LO_w_loops_4F	GCGCTTGGGTGTGCGCGCGACAAGAAGTACATCAGAACGATCAC AACCGGGATCCAC
LO_w_loops_4R	ATGTACTTCTTGTCGCGCGCACACCCAAGCGCGGACCTCTGCGC GGCAACAAATATACACCACCA
LO_w_loops_3R	ACAATCTGACCACCACCCGGAAATTTGACATCTTGCGGGCGGCG ATTTGTATTGCGTTTT
LO_w_loops_2R	GTTTTGCGTTGCGGTTTCGGATTCGTGCTCATGGTGCACGGTCTA CGAGACCTCCC

Table 4: The oligonucleotides for the assembly of different (five) HCV-core constructs are given in the table.

Primers' name	Sequence (5' – 3')
HCVIRES f-SalI	AAAGTCGACGCCAGCCCCCTGATGGGGGCGACAC
HIwoloops_R_PCR	GTGGATCCAGGCTGAGAACGCTCAG
LOwoloops_R_PCR	GTGGATCCCGGTTGACTGCGTTCAC
HIwloops_R_PCR	GTGGATCCAGGCTGAGACCGCTCAG
LOwloops_R_PCR	GTGGATCCCGGTTGTGATCGTTCTG

Table 5: Primer sequences for amplification of the respective gene construct(s).

4.4.2.1 Annealing reaction

I annealed the oligonucleotides of each core construct by adding 4 µl from the respective oligonucleotide stock (100 µM). PCR was run under following conditions:

97°C	6 min
50°C	5 min
37°C	10 min

- The tubes were spin after the reaction and let it stand for 30 min at room temperature.

4.4.2.2 Ligation reaction

From annealing reaction we used 4 µl for ligation.

Annealed primers from 4.4.2.1	4 µl
T4 polynucleotide kinase (10 U/µl)	1 µl
T4 DNA Ligase (2 U/µl)	1 µl
10x Ligase buffer	2 µl
ddH_2O	12 µl

The reaction was incubated at room temperature for 1 hour.

4.4.2.3 Gradient PCR

Six annealing temperatures were used for amplifying the core constructs using specific primers.

Template gene from 4.4.2.2	1 µl
Primer; HCVIRES f-SalI (forward) (10 pmol)	1.25 µl
Specific primer for each gene in Table (Reverse) (10 pmol)	1.25 µl
10x PCR buffer	2.5 µl
dNTPs (10 mM)	0.5 µl
Taq polymerase (2 U/µl)	1 µl
ddH_2O	17.5 µl

95°C	3 min	
94°C	30 sec	25 X
45.3°C	30 sec	
50.8°C		
56.1°C		
61.1°C		
63.1°C		
65.1°C		
72°C	45 sec	
72°C	10 min	

Each core construct was amplified at all the annealing temperatures.

PCR Purification

The PCR products were purified to be further ligated using PCR purification kits such as High Pure PCR Product Purification Kit (Roche) and FastBack DNA minispin kit (Renogen Biolab). The protocols were followed according to the manual instructions from the kit.

4.4.2.4 Restriction Digestion

The PCR product and the empty pRG vector (without insert) were digested using similar restriction enzymes SalI and BamHI and same conditions as described earlier. The reaction was run on 0.8% TAE gel to be further eluted and purified.

4.4.2.5 Elution and Purification

The bands were excised from the 0.8% TAE agarose gel and were put into tubes and weighed. The protocol followed for elution and purification was from the FastBack DNA minispin kit (Renogen Biolab).

4.4.2.6 Ligation

The vector and insert were ligated (overnight) using T4 DNA ligase.

pRG vector (without insert) (100 ng)	4 µl
PCR product (eluted and purified)	10 µl
T4 DNA Ligase (2 U/µl)	2 µl
10X Ligase buffer	3 µl
ddH_2O	11 µl

- The samples were ligated overnight at 16°C for 16 hours. The ligation enzymes were heat inactivated at 65°C for 10 min.

Ligation Purification

To increase the efficiency of electroporation we purified and desalt the ligation product using FastBack DNA minispin kit (Renogen Biolab).

4.4.2.7 Electroporation

Ligated product 5 µl was electroporated in 60 µl of bacterial competent cells *E.coli* (XL1 blue). The process is described earlier and the transformed cells are left in shaking incubator for 1 hour at 37°C with 800 rpm.

Plating

The transformed cells were spread onto kanamycin selected agar plates in densities of 10 µl, 100 µl. Rest of the cells were centrifuged at maximum speed for 1 min, vortexed and spread on the agar plate. The plates were left in incubator overnight at 37°C for growth of bacterial cells.

Plasmid isolation

The selected bacterial cells were cultured as described in minipreparation from *E.coli* following manual of Genbond flexispin plasmid isolation kit (Renogen Biolab).

4.4.2.10 Sequencing

The sequencing was carried out using the same procedure as mentioned above.

4.4.3 Transient transfection of CCL13 cells using ExGen 500 *in vitro* transfection reagent (Fermentas)

For transient transfections, cells were plated in 24-well tissue culture plates 24 h before transfection.

Plasmids that were isolated using low endotoxin minipreparation kit (Qiagene) were prepared according to manufacturers' protocol. DNA concentration of 1 µg was diluted in 100 µl of 150 mM NaCl. Vortex and spin down briefly. 3.9 µl of ExGen 500

was added to the solution and incubated for 15 min at room temperature. 100 µl of ExGen 500/DNA mixture was added to each well and incubated at 37°C for 24-48 hours.

Transient transfection of CCL-13 cells using Turbofect transfection reagent (Thermo Scientific): Transfection was carried out following the manufacturer's protocol.

4.4.3.1 Preparation of cell lysate for western blot

24-48 hours after transfection the cell culture media was removed and 100 µl of 2x SBL was added directly to the adherent cells in each well. The cells were collected in an eppendorf tube and further sonicated (Bandelin *Sonopuls* hd 2070) for 5 sec (4 times and power ~70%) to ensure rupturing of the cells and release of proteins. The samples were incubated at 95°C for 10-20 min.

6x SLB: 7 ml 4x Tris-HCl pH 6.8, 5 ml glycerol, 1 ml β-mercaptoethanol, 1 g SDS, 5 mg bromophenol blue; diluted 2x in deionized water

4.4.3.2 Preparation of cells for flow cytometry

The spent medium was pipetted off and discarded. The wells were washed with 1ml PBS. The monolayer was rinsed by gently rocking the flask back and forth with further treatment of 50 µl trypsin and incubation at 37°C for 3-5 minutes. The dislodging of the cells was checked under microscope. 100 µl of culture media DMEM was added and transferred to a 96 well plate and put it on ice. For first three wells non-transfected cells were added as a negative control.

4.4.3.3 SDS-PAGE

For the purpose of separating the proteins we used sodium dodecyl sulfate polyacrylamide gel electrophoresis. To prepare the 15% separating polyacrylamide gel (2 gels) we used 7.5 ml of 30% acrylamide/0.8% bisacrylamide, 3.75 ml of 4x Tris-Cl/SDS, pH 8.8, 3.75 ml of ddH2O, 0.05 ml of 10% w/v APS and 0.01 ml of TEMED. The solution was gently stirred in order to mix properly and poured in between the two clean glass plates sandwich locked to the casting stand. The top of the gel was covered by overlaying 1x Tris-Cl/SDS, pH 8.8. The gel was allowed to

stand for 25-30 min to polymerize. In the meanwhile, stacking gel was prepared by adding 650 µl 30% acrylamide/0.8% bisacrylamide, 1.25 ml 4x Tris-Cl/SDS pH 6.8, 3.05 ml ddH_2O, 25 µl ammonium persulfate and 5 µl TEMED. 1x Tris-Cl/SDS, pH 8.8 layer was poured off and replaced by the stacking gel. The combs were placed and the gel was polymerized for another 25-30 min. After polymerization combs were removed and samples were loaded and run for 1 hour and 20 min at 150 V in 1x SDS-PAGE buffer.

4x Tris-Cl /SDS, pH 6.8; Dissolve 6.05 g Tris base, 0.4 g SDS; adjust to pH 6.8 with 1 N HCl. Add ddH_2O to 100 ml total volume.
4x Tris-Cl/SDS, pH 8.8; Dissolve 91 g Tris base; 2 g SDS; adjust to pH 8.8 with 1 N HCl. Add ddH_2O to 500 ml total volume.
5x SDS-PAGE buffer: 15.1 g Tris; 72 g glycine; 5 g SDS; ddH2O up to 1 l.

4.4.3.4 Western Blot

Western blot is a technique that is being employed for the identification and detection of proteins separated by SDS-PAGE using specific antibodies. After SDS-PAGE, the 15% polyacrylamide gel was incubated in blotting solution for 20-40 minutes. A blotting sandwich consisting of a nitrocellulose membrane placed in between 4 pieces of 3MM Watmann paper (8 × 6 cm) and the gel was assembled. The blot was run on a semi-dry blotting apparatus (Thermo Scientific^{TM} Owl^{TM} HEP-1) by setting current (about 100 mA, may vary) and time (usually 50 minutes). On the membrane mark position of marker, direction and front side (side in contact with the gel) and the membrane was blocked in milk solution for 1 hour. The primary antibody, mouse monoclonal to anti-Hepatitis C Virus Core Antigen antibody (abcam®) and DsRED2 (25) (Santa Cruz), was diluted in milk in appropriate concentration (dilution 1:1000 for both core and DsRED2 Ab) and incubated at 4°C on a rotator overnight. The membrane was washed 3 x 15 minutes next day in TBS/Tween buffer and blocked again in the milk solution for one hour. Incubation of the membrane with secondary antibody, goat anti-mouse IgG-HRP (Santa Cruz), diluted to appropriate concentration (dilution 1:5000) was carried out at 4°C for 2 hours. The membrane was again washed 3 x 15 minutes in TBS/Tween buffer.

Blotting buffer; 200 ml methanol, 2.9 g glycine, 5,8 g Tris, 0.37 g SDS, ddH_2O up to 1 l

TBS/Tween: 50 mM Tris; 150 mM NaCl; 0.5% Tween-20
Milk solution: 5% Sunar baby (Heinz) in 1x TBS/Tween

4.4.3.5 Chemiluminescence detection

We used chemiluminescent method for the detection of horseradish peroxidase signal conjugated to the secondary antibody because it provides better detection sensitivity. Chemiluminescent reagents such as 200 µl of 0.25M luminol and 88 µl of 0.1M p-coumaric acid was mixed with 12 µl of 30% hydrogen peroxide and 2 ml of Tris-HCl pH 8.5 in a total volume of 20 ml. In other reaction tube with a total volume of 20 ml, only 2 ml of Tris-HCl pH 8.5 was added with 12 µl of 30% hydrogen peroxide. The two solutions were mixed and membrane is immersed in it for a while. The membrane was put in a film cassette along with an X-ray film and incubated for 1 hour. The film was put in a developer solution for 6 minutes depending on the signal and washed with water for sometime. The film was put in a fixer for 15 min.

Luminol; 445 mg luminol, 10 ml DMSO (vortex)

*p-Coumaric acid***;** 67.5 mg *p-Coumaric acid*, 4.4 ml DMSO

Tris-HCl pH 8.5

4.4.3.6 Digital Imaging

To further quantify the intensity of protein signals at different time points we employed LAS-4000 (GE Healthcare Life Sciences) that facilitates the chemiluminescence digital imaging. The chemiluminescent solution after mixing (as described above) was pipetted on the membrane and images were taken using modes such as super, standard and ultrahigh resolution. The procedure was followed according to the instructions of the manufacturer.

4.4.3.7 Co-immunoprecipetation from mammalian cells

Transiently transfected cells growing in the 6-cm culture dishes were washed with PBS and 1 ml of IP lysis buffer supplemented with protease inhibitor cocktail (Roche) was added. Cells were scraped into an eppendorf tube, incubated on ice for 30 min and vortexed repeatedly. Samples were spinned for 10 min at 13 000 g in 4°C and supernatants were transferred into new tubes and kept on ice. 100 µl of Buffer B supplemented with protease inhibitor cocktail (Roche) and 100 µM DTT was added to the pellet, vortexed and incubated on ice for 15 min. After a 10 min centrifugation at 13 000 g in 4°C, the Buffer B supernatant was added to the lysate obtained with the IP lysis buffer. 50 µl of protein G agarose slurry (sigma-aldrich) was added to the lysate and incubated for 1 h at 4°C on rotater. Tubes were spinned at 2000 g for 20 sec and the lysate was transferred into a new tube. 70 µl of protein G agarose slurry was added together with 1 µg of the mouse monoclonal anti-Flag antibody (Sigma) and samples were incubated overnight on a rotater at 4°C. Next day the protein G agarose slurry containing bound proteins was washed 3x for 15 min by ice-cold IP lysis buffer, centrifuged at 2000 g for 20 sec and added 60 µl of 2x SLB and incubated at 100 °C for 5 min.

IP lysis buffer: 50 mM Tris-Cl, pH 7.6, 150 mM NaCl, 1,5 mM MgCl2, 1 mM EDTA, 0,5% NP40, 10% glycerol

Buffer B: 0.02 M HEPES; 25% glycerol; 420 mM NaCl; 1.5 mM MgCl2; 0,2 mM EDTA

4.4.3.8 Flow Cytometry

Samples were analyzed by flow cytometry using a BD LSRII device and a Coherent Sapphire 488-20 DPSS laser to excite cells at 488 nm, a 530/30 nm bandpath filter to detect EGFP, and a 585/ 42 nm bandpath filter to detect DsRED2 expression.

5 Results and Discussion

Study of the HCV-IRES variability in HCV infected patients

Hepatitis C virus internal ribosomal entry site is responsible for carrying out protein synthesis of HCV genome cap-independently. We studied diversity in HCV-IRES region in order to understand how mutations in different regions can impact the regulation and mediation of translation initiation.

For this purpose of quasispecies determination at given time, three libraries of random full-length HCV-IRES segments were created from patients (4, 7 and 9) in a bicistronic pRG vector. Employing methodologies such as denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) helped us in visualizing the variability of HCV-IRES in different constructs, on a large-scale, based on their migration pattern in the gel. The migration-derived pattern of samples can assist in elucidating the amount of quasispecies that has persisted in patients over time. We selected HCV-IRES constructs for sequencing and also measured their translation activity using flow cytometry. The presence of mutations in different HCV-IRES domains and the consequent translational activity got us interested in analyzing the impact of mutations on HCV-IRES function on a much larger scale. We developed a comprehensive HCV-IRES variation database (HCVIVdb; www.hcvivdb.org) comprises ~1900 mutations that were collected from majority of the studies dealing with mutations in the HCV-IRES either naturally occurring or *in vitro* mutagenized. The database is a tool that can be helpful in evaluation of mutations that occur in different domains with corresponding influence on the HCV-IRES structure and function.

5.1 Identification of the HCV-IRES mutants through denaturing gradient gel electrophoresis (DGGE) in HCV infected Patients (4, 7 and 9)

5.1.1 Overview of all-around variability in Patient 7

The Patient 7 suffering from HCV genotype 1a was non-responsive to interferon-alpha (IFN-α) and ribavirin treatment. We utilized DGGE to observe the amount of

HCV-IRES variations that were incorporated in the patient overtime. The DGGE was optimized using 55-70% denaturants (7M urea and 40% formamide) in 6% polyacrylamide gel, described in Materials and Methods section. Since DGGE is sequence-dependent and size-independent, by visualizing the DGGE we were able to estimate the similarity and variability that was integrated among the patients'

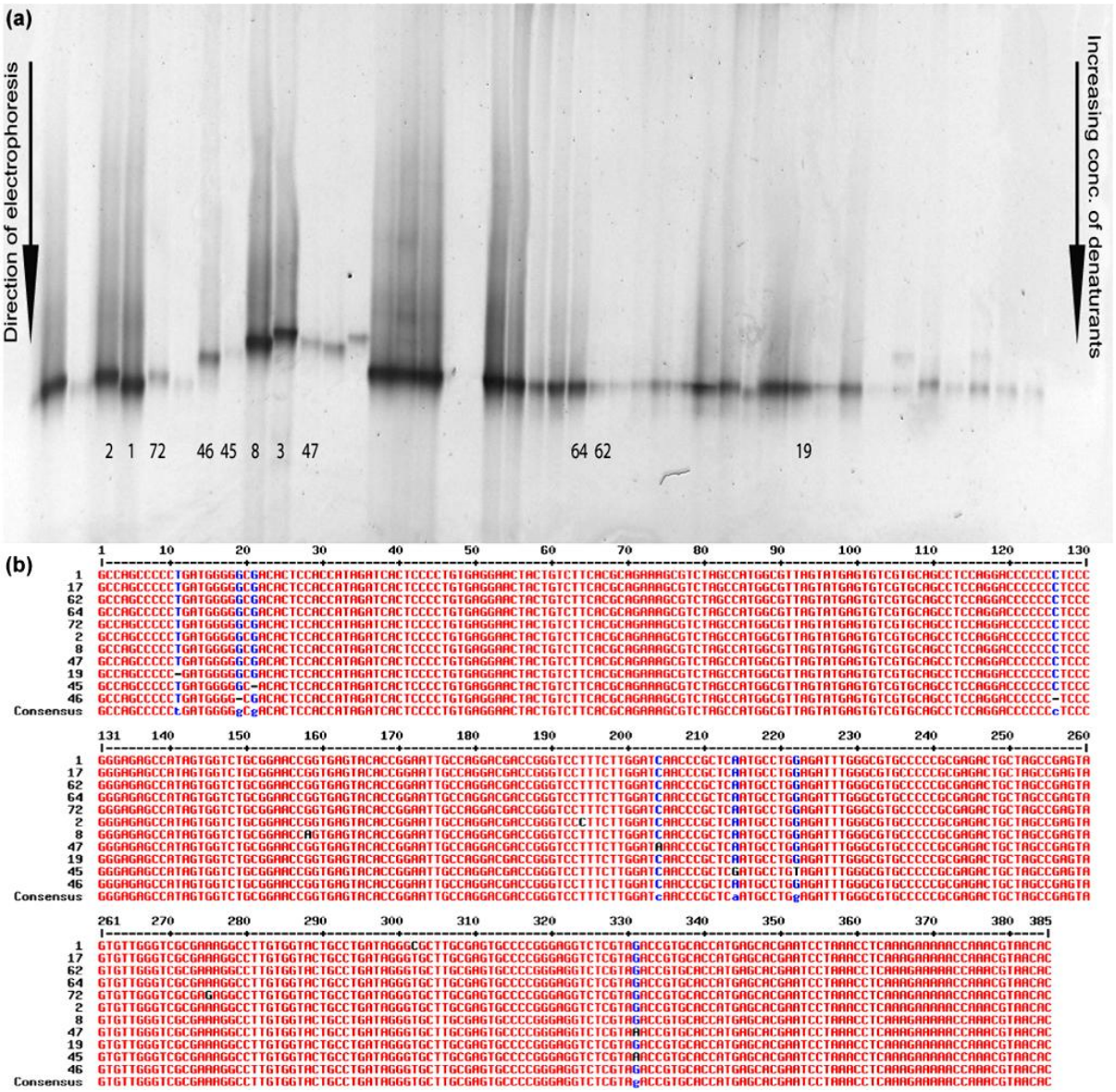

Figure 4: (a) The differences in migration of the HCV-IRES from Patient 7 samples are shown in DGGE. Different cDNA clones, identified by a number under the sample, were selected for sequencing. **(b)** From this gel, 11 constructs were sequenced and aligned to check for mutation(s) in HCV-IRES from the reference sequence of the Patient 7 and further co-related the amount/nature of variation(s) of samples with their movement through the gel.

samples on a large-scale. For Patient 7, differences in migration pattern that hint towards the diversity among patients' samples was observed (Figure 4a). We ran 65 cDNA clones, having positive HCV-IRES fragment, on the DGGE 6% polyacrylamide gel. Out of those 65 we selected 22 samples for sequencing that showed similar or different movement through the gel. The sequencing was carried out to confirm the co-relation that may exist between the amount/nature of variation(s) in the sequences of these cDNA samples and their movement through the gel (Figure 4b).

The samples P7-19, P7-62 and P7-64 are among the general reference sequence or the WT of the HCV-IRES fragment from the Patient 7. The migration pattern of these samples as shown in DGGE (Figure 4a) was similar as their sequences suggested. Other cDNA clones (P7-1, 2, 3, 8, 45, 46, 47 and 72) were selected for the reason that they displayed a varied progression through the gel and were sequenced to further confirm the changes or mutation that may be incorporated in these particular sequences, relative to the reference sequence. With the HCV-IRES fragments (P7-1, 2, 8, 46 and 72) having a single mutation, the respective variation on DGGE was observed within these samples (Figure 4b). The similarity of progression through the gel for these single mutants is not the same except in sample P7-2 and P7-72. The sequence of samples P7-2 and P7-72 shows point mutations $U^{194}C$ and $G^{275}A$, respectively located in domain IIIb apical loop and IIId of the HCV-IRES (Figure 2). Both of these mutations are present in the loop region and do not form a base-pair. Other point mutants (P7-1, 8 and 46) that form a base-pair showed variation among their migration patterns. The samples P7-1 and P7-8 exhibited a transition $U^{302}C$ and $G^{158}A$, respectively. Transition $U^{302}C$, sample P7-1, changes the G^{291}-U^{302} wobble base-pairing to G-C and transition $G^{158}A$, sample P7-8, displaces the Watson-Crick base-pairing G^{158}-C^{167} to A-C. The differences in the progression of these point mutants suggest that the location of these mutations may have important implications in context to HCV-IRES conformation and in turn affect the rate at which these fragments move through the gel.

5.1.2 Overview of all-around variability in Patient 9

The Patient 9 is non-treated with any IFN-α and ribavirin treatment. We ran a total of 91 HCV-IRES cDNA clones from this patient on DGGE. Number of samples that

were sequenced out of these 91 cDNA clones was 24. The selection of samples was based on the distinct migration pattern and sequencing was carried out to co-relate the difference in movement of samples to any changes that is incorporated in the actual or reference sequence. The first three samples (P9-A2, P9-A3 and P9-A4) were among the majority of the Patient 9 that showed the reference sequencing with

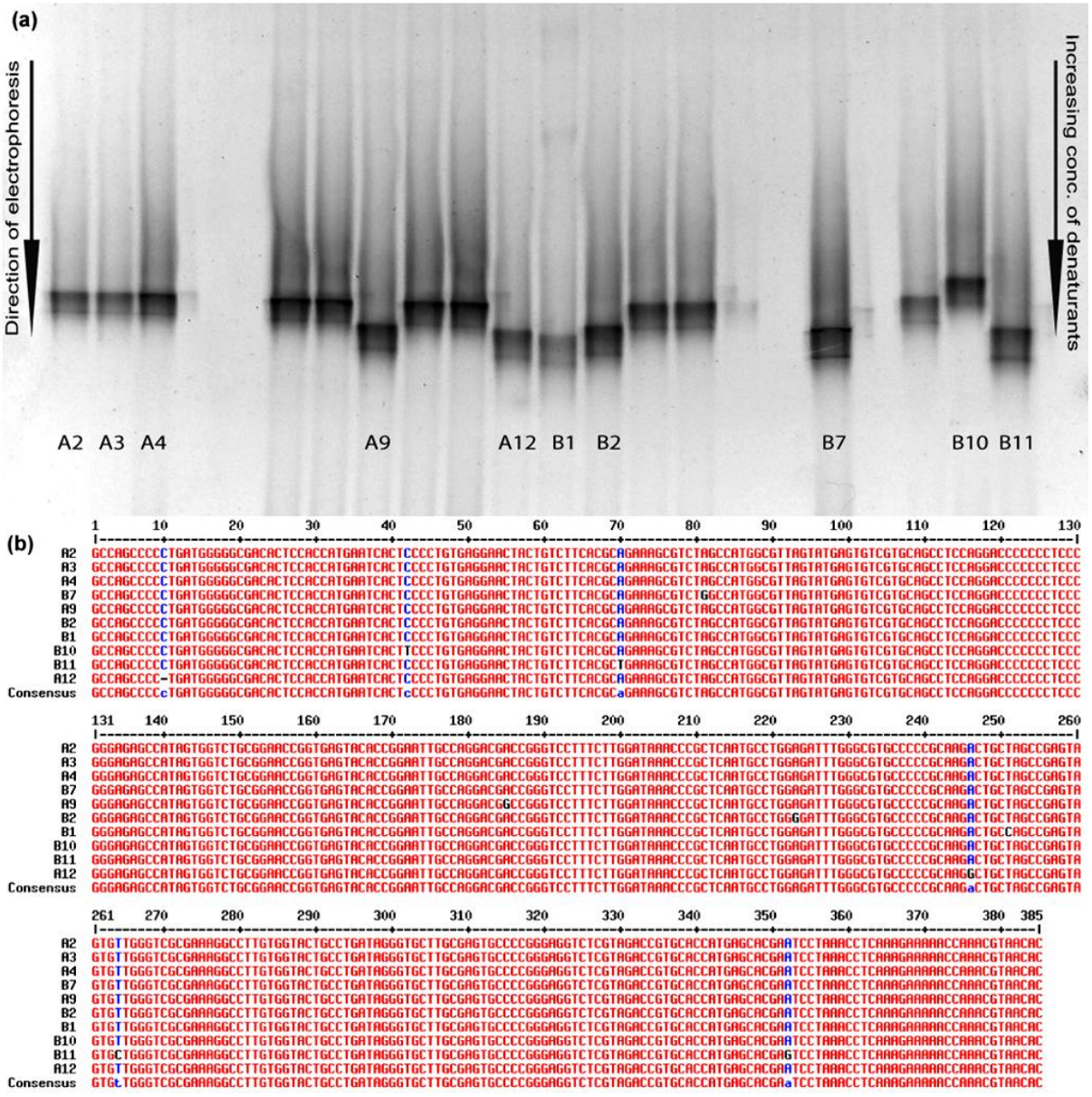

Figure 5: (a) The differences in migration of the HCV-IRES from Patient 9 samples are shown in DGGE. Different cDNA clones, identified by a number under the sample, were selected for sequencing. **(b)** From this gel, 10 constructs were sequenced and aligned to check for mutation(s) in HCV-IRES from the reference sequence of the Patient 9 and further co-related the amount/nature of variation(s) of samples with their movement through the gel.

the migration pattern of these samples also found to be similar in the DGGE (Figure 5a). Other single mutants (P9-A9, A12, B1, B2 and B7) displayed similar distances travelled by these fragments through the gel. They all, upon sequencing, showed transition mutations and were present in the HCV-IRES region where they form base-pair except P9-B7 ($A^{81}G$) which is in the domain IIb loop (Figure 5b). With a triple mutant (P9-B11) having all the transition mutations, one at a base-pair forming region and two at the loop regions showed similar movement of the fragment displayed by the single mutants. Unlike in the Patient 7, where different pattern of separation among single transition mutants were observed, in Patient 9, the migration pattern of point mutants was found to be the same. This can possibly be due the differences in the mutations that were localized in different HCV-IRES regions. It may suggest that variations at distinct locations in the HCV-IRES can lead to different conformation which ultimately results in difference in movement of various fragments through the gel.

5.1.3 Overview of all-around variability in Patient 4

The Patient 4, diagnosed with HCV infection was also non-treated with any IFN-α and ribavirin. From this patient we ran 92 full-length HCV-IRES cDNA clone segments on DGGE to analyze for the overall diversity that has developed over time. Much variation among samples was observed while running the DGGE using same parameters as described in Patient 7. We selected 38 samples to be sequenced. For the most part, movement of fragments through the gel synchronized with the amount of variation that was found through sequencing. The WT or reference sequence samples (e.g., P4-E10 and P4-E11) showed similar sequences and migration pattern in DGGE with majority of the samples analyzed from the Patient 4 (Figure 6). The other fragments mentioned in the gel (Figure 6a) are all single mutants (P4-E9, E12, F2, F3, F4 and F5). All of the samples being single mutants, samples (P4-E9, E12, F3 and F4) exhibited a different shift as compared to samples (P4-F2 and P4-F5). In case of samples P4-E12 and P4-F3, the mutation $U^{234}C$ and $A^{202}G$ occurs in the loop region of IIIc and IIIb, respectively. Whereas, mutation $U^{356}C$, in sample P4-F4, is

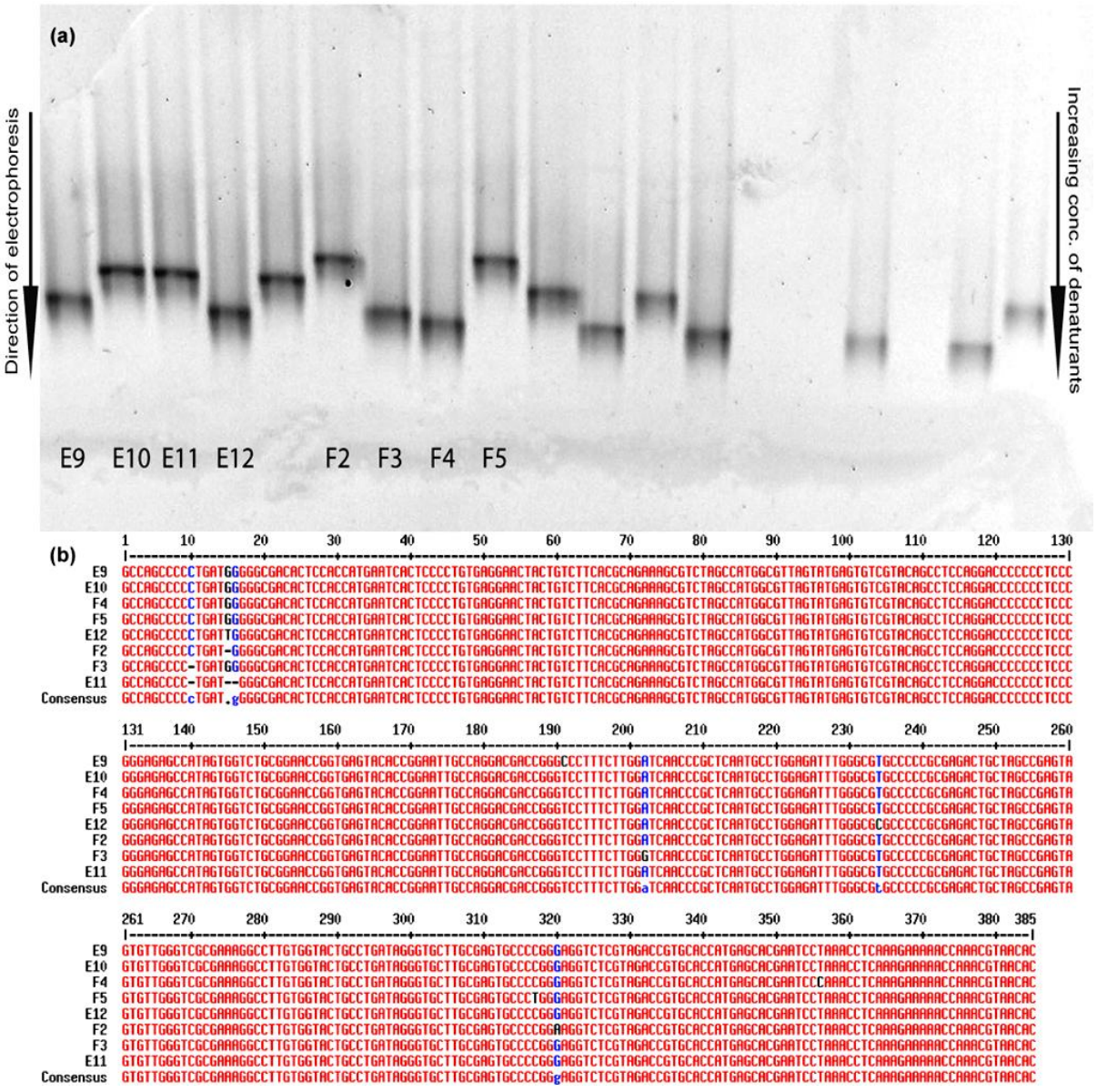

Figure 6: (a) The differences in migration of the HCV-IRES from Patient 4 samples are shown in DGGE. Different cDNA clones, identified by a number under the sample, were selected for sequencing. **(b)** From this gel, 10 constructs were sequenced and aligned to check for mutation(s) in HCV-IRES from the reference sequence of the Patient 4 and further co-related the amount/nature of variation(s) of samples with their movement through the gel.

located in open reading frame which is also single stranded. Sample P4-E9 on the other hand has a mutation $U^{191}C$, adjacent to domain IIIb apical loop, that changes the Watson-Crick base-pair from U^{191}-A^{206} to C-A. These mutations occurring in the loop region may have conformed in a similar way that resulted in the parallel

progression of these samples. Likewise, point substitutions in sample P4-F2 ($G^{320}A$) and P4-F5 ($C^{317}U$) both occur in domain IIIf at the region that makes the pseudoknot (Figure 2). Both of these fragments have shown similar progression through the gel. Although, mutation $G^{320}A$ (P4-F5) changes the Watson-Crick base-pair C^{128}-G^{320} to CA and G^{131}-C^{317} to G-U with parallel progression found in the DGGE, suggests that not only sequence changes but the adjacent secondary structure stacking interactions play a role in molecule's conformation that may affect its rate of movement through the gel.

5.2 Measuring translation activities of the HCV-IRES in selected patients' samples

In order to estimate the translation activities of the HCV-IRES samples that we had found in our patients and verified by sequencing, we employed flow cytometry. Transient expression of the pRG vector with selective HCV-IRES constructs were carried out in CCL-13 cell line using ExGen 500 *in vitro* transfection reagent (Fermentas), described in Materials and Methods section.

5.2.1 Translation efficiency of HCV-IRES in Patient 7

For Patient 7 we evaluated 18 samples for their translation response using flow cytometry (Figure 7). Most of the constructs that contained mutants had shown activities that are near to the WT (Table 6). Mutations were located in different regions of the HCV-IRES that depict the diversity that had incorporated among samples over time in an individual. The analysis can also be widened to study diversity of the HCV-IRES across various groups and populations.

We had found samples with single, double and triple mutants. The expression pattern for EGFP observed by flow cytometry (Figure 7) in most of the constructs was closer to the pRG-IRES control vector translation efficiency. There also had been mutants that displayed relatively lower translation response. For example, sample P7-24 with C insertion at position 127 along with a transition $U^{147}C$ showed very little translational activity. Similarly, samples P7-31 and P7-45 with triple mutations each also have activities that are comparably lower to the WT (Table 6). The difference in

Sr. No.	Sample #	Mutations	Activity (%)
1	P7-10	none	N/A
2	P7-11	none	N/A
3	P7-17	none	N/A
4	P7-19	none	N/A
5	P7-53	none	N/A
6	P7-54	none	N/A
7	P7-57	none	N/A
8	P7-62	none	100
9	P7-64	none	100
Individual mutations			
Domain II			
10	P7-38	G ^{82} A	89
Domain III			
11	P7-8	G ^{158} A	92
12	P7-2	U ^{194} C	93
13	P7-21	C ^{204} A	92
14	P7-49	A ^{275} G	83
15	P7-72	A ^{275} G	85.8
16	P7-1	U ^{302} C	104
17	P7-46	C deletion ^{126}	97.6
Domain IV			
-	-	-	-
Double mutants			
19	P7-24	C insertion ^{127} U ^{147} C	6
20	P7-47	C ^{204} A G ^{331} A	108.2
21	P7-20	G ^{241} A U ^{353} C	84
Triple mutants			
22	P7-31	C ^{83} U A ^{96} G C ^{254} U	54
23	P7-45	A ^{214} G G ^{222} U G ^{331} A	37.4

Table 6: The HCV-IRES nucleotide changes found in Patient 7. The HCV-IRES mutants are categorized based on the number of mutations (individual and multiple) in domains II, III and IV of the HCV-IRES along with the translation activities of these mutant samples measured using flow cytometry.

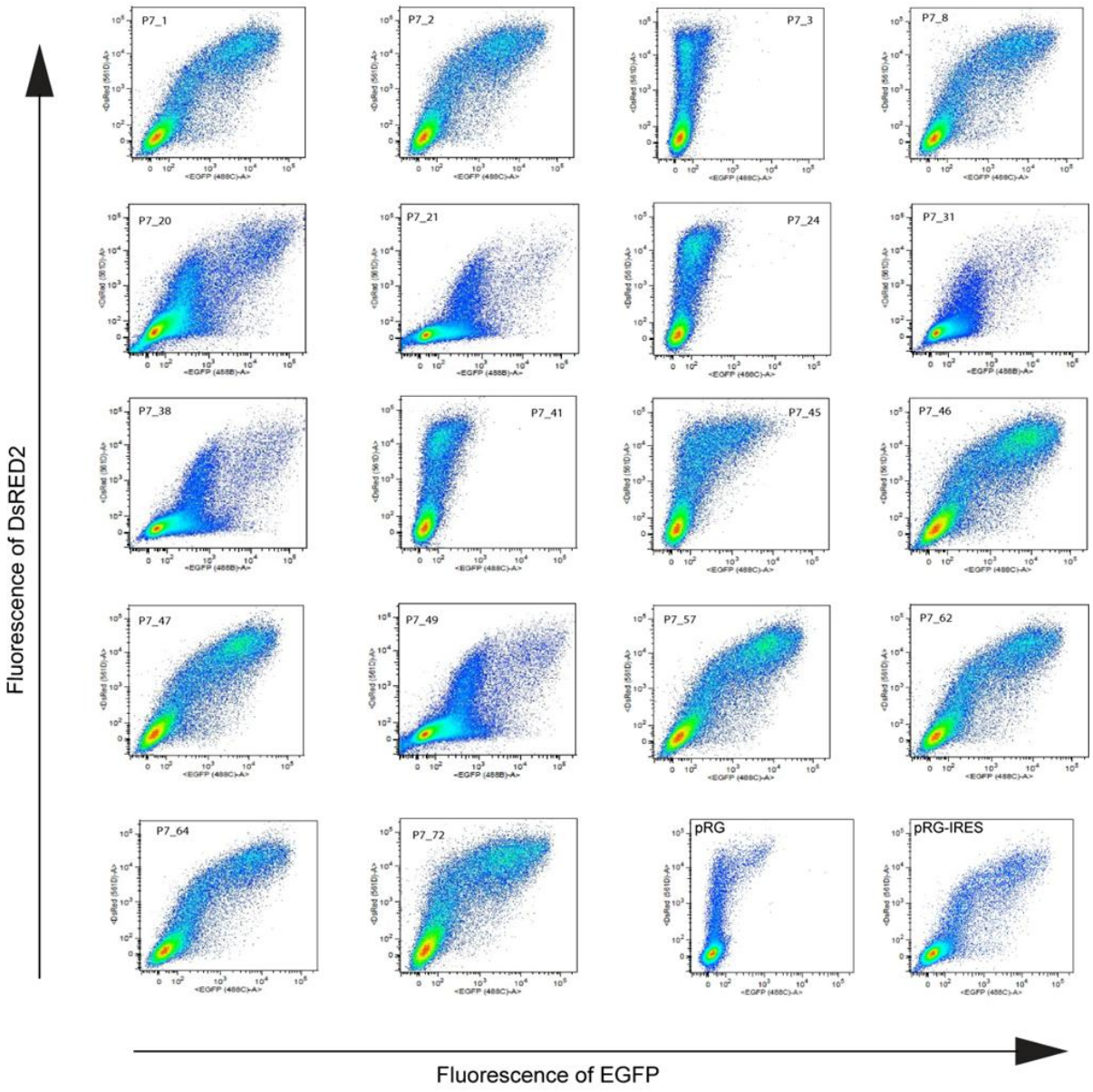

Figure 7: The HCV-IRES translation efficiency from Patient 7 samples were evaluated using flow cytometry. With the aid of bicistronic pRG vector, the HCV-IRES activity was measured by analyzing the transient expression of EGFP reporter gene in mammalian cell culture, employing flow cytometry. The expression pattern exhibits range of activities for P7 constructs with different mutations, suggesting diversity of function that may arise within an individual over time. The name of the individual construct is indicated alongside with patient (P7) and the activity of pRG vector and pRG-IRES control vector translation efficiency is also indicated.

translational efficiency can possibly be subjected to the number of mutations present in a construct. The samples P7-24, P7-31 and P7-45 showing lower activity had two

to three mutations each while all the constructs with point substitutions showed a response that is near or equal to the pRG-IRES control vector translation efficiency. There also had been samples, P7-20 and P7-47, in spite of having double substitutions display an efficient productivity (Figure 7 and Table 6). It may suggest that the location of mutations' occurrence is more significant in determining the HCV-IRES activity. We also found a sample P7-3 with large deletion and multiple substitutions that resulted in minimum to no translational activity.

5.2.2 Translation efficiency of HCV-IRES in Patient 9

We analyzed 16 constructs on flow cytometry for determining the efficiency of HCV-IRES in Patient 9. The samples were evaluated based on the expression pattern of EGFP reporter gene in bicistronic pRG vector, exhibited on fluorescence-activated cell sorting (FACS) (Figure 8). Similar to the Patient 7, mutations were dispersed throughout all domains of the HCV-IRES. In addition, there also had not been any point substitution that shows drastic effect on the HCV-IRES response. However, sample P9-B10 with single substitution $C^{42}U$ at the base of domain II has displayed relatively decreased HCV-IRES activity compared to the pRG-IRES control vector translation efficiency. The lowering of translation activity was exhibited in constructs (P9-B11, P9-F2 and P9-G7) that had multiple mutations (Table 7). There has been no specific pattern that is to be observed that may define the occurrence of these double or triple mutations in various domains with the respective outcome of HCV-IRES response. Double mutant P9-C1 has a mutation $A^{244}G$ adjacent to the mutation $A^{243}G$ found in another sample P9-F2, but the activity shown by the P9-C1 construct is relatively higher than the P9-F2 sample. Other point substitutions at this region $A^{246}G$ and $U^{251}C$ (P9-A12 and P9-B1) have shown not to influence the HCV-IRES translation (Table 7). So, the difference in activity in both these constructs, P9-C1 and P9-F2, is probably due to the impact caused by the second mutation. Lower translational response in the case of sample P9-F2 is because mutation $A^{257}G$ is present in domain IIId which interacts with the 40S ribosome directly and is important for the HCV-IRES activity. At the same time second mutation in sample P9-C1, $U^{353}C$, in domain IV alters the Watson-Crick base-pair A^{332}-U^{353} to A-C (Figure 2)

have not shown to impact HCV-IRES translation in any significant way relative to the pRG-IRES control vector translation efficiency.

Sr. No.	Sample #	Mutations	Activity (%)
1	P9-A2	none	N/A
2	P9-A3	none	N/A
3	P9-A4	none	111
4	P9-D11	none	N/A
5	P9-D12	none	N/A
6	P9-F3	none	N/A
7	P9-F4	none	N/A
8	P9-F5	none	N/A
9	P9-F6	none	N/A
Individual mutations			
Domain II			
10	P9-B10	C ^{42} U	66
11	P9-B7	A ^{81} G	N/A
12	P9-E11	U ^{101} C	85
Domain III			
13	P9-A9	A ^{185} G	89
14	P9-G8	C ^{186} U	93
15	P9-B2	A ^{223} G	89
16	P9-A12	A ^{246} G	93.8
17	P9-B1	U ^{251} C	95
18	P9-E10	U ^{287} C	94
Domain IV			
19	P9-E9	C ^{334} U	95
Double mutants			
20	P9-F2	A ^{243} G A ^{257} G	65
21	P9-G7	A ^{260} G A ^{345} G	42.6
22	P9-C1	A ^{244} G U ^{353} G	92
Triple mutants			
23	P9-B11	A ^{70} U U ^{264} C A ^{352} G	64

Table 7: The display of HCV-IRES mutations found in patient 9. The HCV-IRES mutants are categorized based on the number of mutations (individual and multiple) in domains II, III and IV of the HCV-IRES along with the translation activities of these mutants measured using flow cytometry.

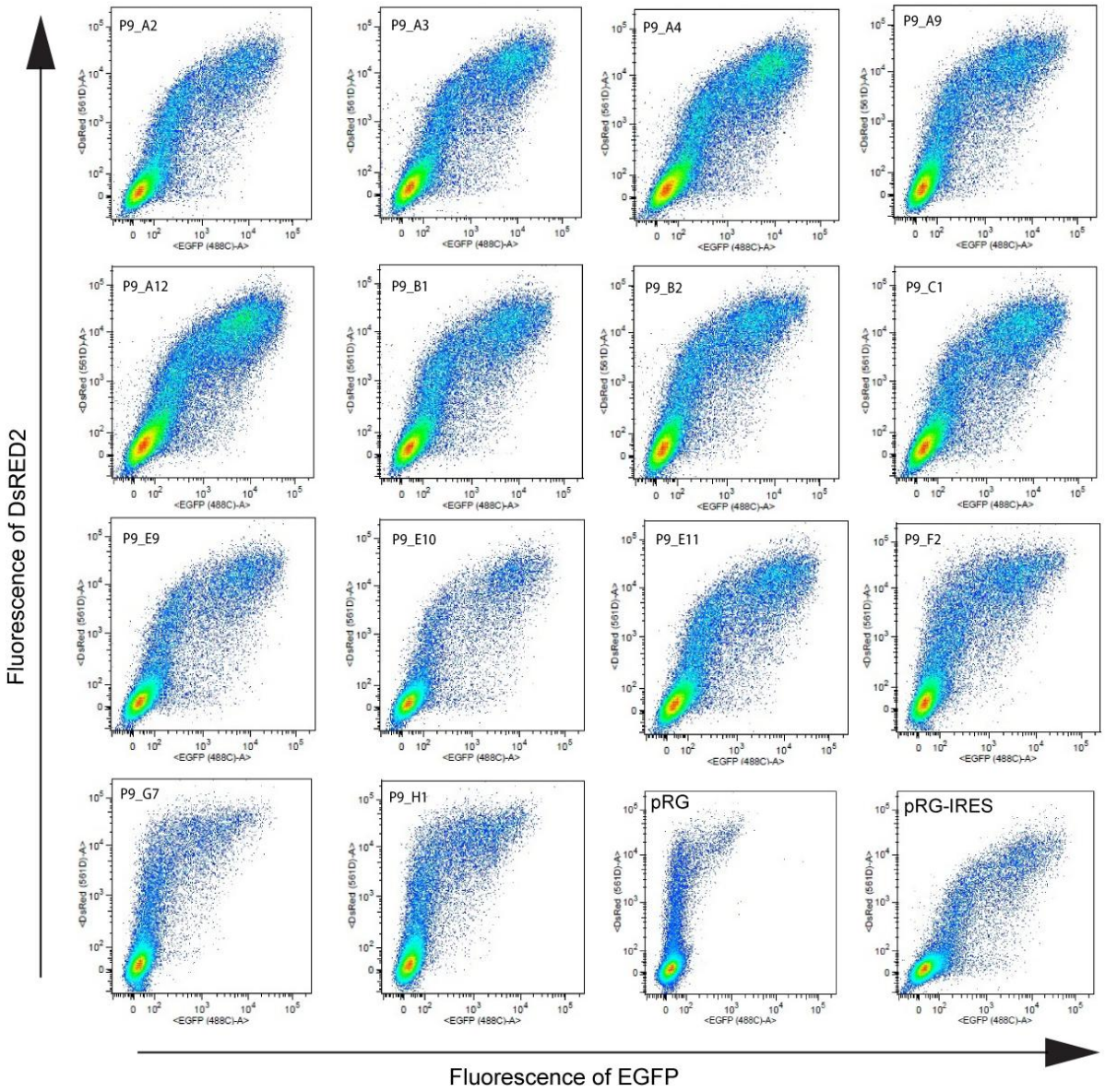

Figure 8: The HCV-IRES translation efficiency from Patient 9 samples were evaluated using flow cytometry. With the aid of bicistronic pRG vector, the HCV-IRES activity was measured by analyzing the transient expression of EGFP reporter gene in mammalian cell culture, employing flow cytometry. The expression pattern exhibits range of activities for P9 constructs with different mutations, suggesting diversity of function that may arise within an individual over time. The name of the individual construct is indicated alongside with patient (P9) and the activity of pRG vector and pRG-IRES control vector translation efficiency is also indicated.

5.2.3 Translation efficiency of HCV-IRES in Patient 4

The samples or the constructs that we sequenced earlier for patient 4 were analyzed on FACS for the estimation of translation activity of the HCV-IRES (Figure 9). We examined 26 samples from Patient 4 to get the overview of various mutations'

Sr. No.	Sample #	Mutations	Activity (%)
1	P4-A2	none	100 – P4 commonIRES
2	P4-A9	none	N/A
3	P4-A10	none	N/A
4	P4-A11	none	N/A
5	P4-B2	none	N/A
6	P4-C1	none	N/A
7	P4-C4	none	N/A
8	P4-G11	none	N/A
9	P4-D12	none	N/A
10	P4-E2	none	N/A
11	P4-E10	none	N/A
12	P4-E11	none	N/A
13	P4-G5	none	N/A
14	P4-G8	none	N/A
Individual mutations			
Domain II			
15	P4-A12	U ^{113} C	90
16	P4-G6	C del. ^{115}	56
Domain III			
17	P4-E9	U ^{191} C	92
18	P4-F3	A ^{202} G	95
19	P4-E1	U ^{284} C	71
20	P4-F5	C ^{317} U	86
21	P4-F2	G ^{320} A	69
Domain IV			
22	P4-A8	C ^{341} U	100
23	P4-F4	U ^{356} C	73
Double mutants			
24	P4-G12	U ^{220} C G ^{303} A	69
25	P4-H1	U ^{220} C G ^{303} A	62
26	P4-C2	G ^{268} A C ^{317} U	37.6
27	P4-C8	A ^{140} G U ^{251} C	86
28	P4-B8	C del. ^{126} U ^{356} C	78
Triple mutants			
29	P4-B3	A ^{66} G A ^{182} C C ^{240} U	92
30	P4-C5	A ^{72} G U ^{234} C U ^{248} A	100

Table 8: The HCV-IRES nucleotide variations observed in patient 4. The HCV-IRES mutants are categorized based on the number of mutations (individual and multiple) in domains II, III and IV of the HCV-IRES along with the translation activities of these mutants measured using flow cytometry.

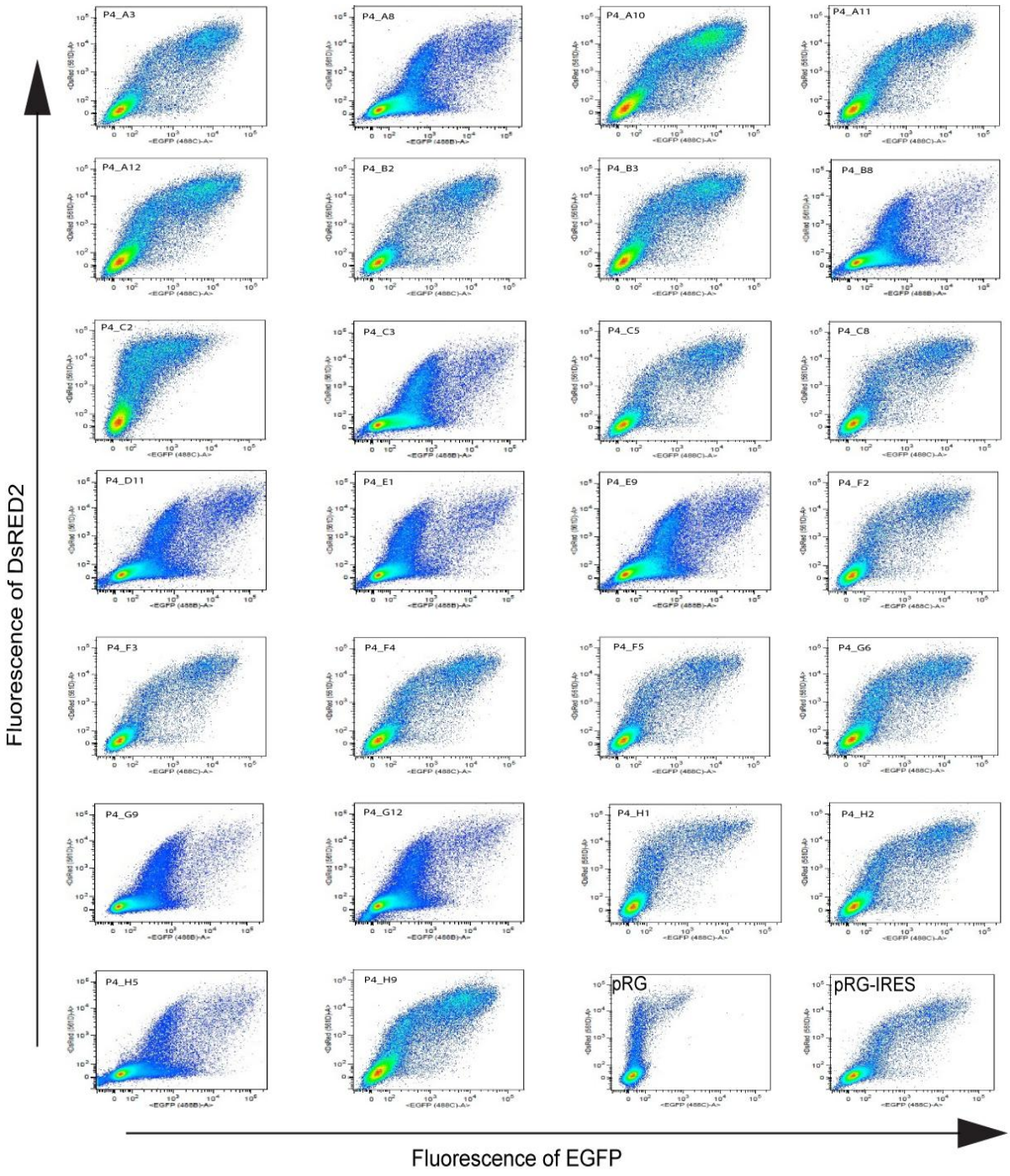

Figure 9: The HCV-IRES translation efficiency from Patient 7 samples were evaluated using flow cytometry. With the aid of bicistronic pRG vector, the HCV-IRES activity was measured by analyzing the transient expression of EGFP reporter gene in mammalian cell culture, employing flow cytometry. The expression pattern exhibits range of activities for P4 constructs with different mutations, suggesting diversity of function that may arise within an individual over time. The name of the individual construct is indicated alongside with patient (P4) and the activity of pRG vector and pRG-IRES control vector translation efficiency is also indicated.

locations on the HCV-IRES domains and their associated impact on translation response (Table 8). Mutations have been found in all of the domains II, III and IV and domain III being the largest incorporates most of the mutations. Single mutations in all the domains have been shown not to impact the activity to any greater extent, except in samples P4-E1, P4-F2 and P4-F4 where the response is somewhat intermediate, 71%, 69% and 73% respectively, relative to the pRG-IRES control vector translation efficiency (Table 8). The samples P4-E1 ($U^{284}C$) and P4-F2 ($G^{320}A$) show a transition which modifies the canonical base-pair A^{140}-U^{284} to A-C and C^{129}-G^{320} to C-A (Figure 2) which can possibly impact the conformation of this region, ultimately leading to lower translational response. A deletion in region close to domain II asymmetrical loop, C^{115}, in P4-G6 construct has also displayed decrease of the HCV-IRES efficiency. Nearly all the double mutants that had been observed in this patient exhibited activities that are relatively reduced as compared to the pRG-IRES control vector translation efficiency (Table 8). The association between the increase in number of mutations and decrease in the HCV-IRES efficiency has also been observed in other patients. But the triple mutants in this patient has shown activities that are near to the pRG-IRES control vector translation efficiency repeatedly emphasizing the importance of location in the HCV-IRES domains and variability in such regions are crucial in specifying the outcome of translational response.

5.3 The HCV-IRES variation database (HCVIVdb) (www.hcvivdb.org)

5.3.1 Database: Overview

The databases are central in providing information that had been collected and assembled from various scientific experiments, published reports and computational analysis. The raw and curated (annotated sequences, structures and expression data) datasets are characterized into different database types depending on the nature of information stored that ranges from complex sequences, three-dimensional (3D) structures, localization, pathways, clinical effects of mutations, literature, nomenclature etc. to how schemes are adopted for efficient data storage, management, representation and easy wide-range access to the contents. The

dataset obtained is generally organized for the optimal usage and assist analyses of biologically relevant datasets such as bio-molecular pathways and interactions, gene/protein sequence and structure, genomes, expression data and understanding the evolution of species. The databases also assist in prediction of genetic diseases, development of targeted sequence and structure drug designs and medications.

With the advent of technologies such as next-generation sequencing (NGS) that provides cheap and accurate genome information, large-sequence dataset volumes have risen exponentially (Metzker, 2010) that demands for stricter data sorting, search and retrieval (Howe et al., 2008) along with the establishment of controlled vocabulary for standardization of biomedical terms within and between databases. The amount of knowledge-bases has grown tremendously in recent years and the focus of databases is diverse (Sharma et al., 2015, Zhulin, 2015, Zou et al., 2015) from the availability of comprehensive knowledge to specialized datasets. The International Nucleotide Sequence Database (INSD) comprising of resources such as GenBank (Benson et al., 2015)**,** European Molecular Biology Laboratory (EMBL) (Brooksbank et al., 2014), and DNA Data Bank of Japan (DDBJ) (Kosuge et al., 2014) are examples of comprehensive and also primary knowledge-base depending on the level of data curation. The examples of databases which are specialized and addresses the specific organism(s) or data are the Human Gene Mutation Database (HGMD) (Auton et al., 2015), Interferome (Rusinova et al., 2013) and DisGeNET (Pinero et al., 2015). The resulting analysis of sequences from primary database can be collated to form a secondary database such as PROSITE (Sigrist et al., 2013), Pfam (Finn et al., 2016), PRINTS (Attwood et al., 2003) etc. Likewise, composite databases comprise combining different sources of primary databases for efficient query, search and retrieval of information from all the primary sources. The biological data from different repositories are combined and collated allowing the users to visualize, compare, analyze and retrieve large datasets. The collaboration and integration of data and annotation is a challenging issue that have its own limitations due to the lack of standard format, quantity of biological data and data errors (Stein, 2003). The examples of integrated resources are SeqMap 2.0 (Hawkins et al., 2011), Dr.Vis (Yang et al., 2015) and RefSeq (Pruitt et al., 2014).

Variety of biological repositories containing information from research areas including signal transduction pathway (Cerami et al., 2011), proteomics (Vizcaino et al., 2013, Andreeva et al., 2014), genomics (Harris et al., 2010), phenotype (Groth et al., 2010), taxonomy etc. are accessible. The development of such resources comprising of data collection and curation provides the users with a computational tool to view, compare, analyze, manipulate and download the related large-scale information from the databases.

5.3.2 HCVIVdb: Data Collation (Construction, Content and Utility)

The evaluation of HCV-IRES diversity in different patients while employing various techniques such as DGGE, TGGE and flow cytometry intrigued us to search more about these mutations and their possible implications on the structure and majorly on function of the HCV-IRES. The influence of nucleotide changes on translational activity was found to be different depending upon the presence of these mutations at specific locations in different HCV-IRES domains. This particular translation response is due to several factors that may include the interference of HCV-IRES interaction with 40S ribosome, translation factors or an effect on HCV-IRES conformation that restricts the translation mediation. To further gain an insight into the workings and behavior of the HCV-IRES and how variations may affect the biology and regulation of viral protein synthesis lead us to the development of an extensive HCV-IRES variation database (HCVIVdb: www.hcvivdb.org). The database consists of collection of mutations from majority of studies dealing with the HCV-IRES. The idea was to understand and estimate the functional capability of distinct HCV-IRES regions on a larger scale. The database comprises the HCV-IRES variations alongside the different translation systems that were being employed by those studies, to test the capability of translation and the impact caused by individual or multiple mutations that either occurs naturally in patients or the result of *in vitro* mutagenesis. We have also organized the plasmid constructs and different reporter genes, listed in these studies, used for the measurement of the HCV-IRES activity. Majority of the studies that dealt with the HCV-IRES belonged to genotype 1. However, other HCV genotypes which were reported in the literature are also documented in the database. Additional information regarding the brief summary of the particular study and (or) the mutation

that may be important in defining IRES role in translation along with the conditions that were employed while testing specific HCV-IRES constructs are also mentioned.

The HCVIVdb (www.hcvivdb.org) is an effective tool that comprises ~1900 published HCV-IRES variations. These variations have been, as mentioned above, categorized based on experimental parameters, frequency of variations, translation efficiency, different genotypes and clinical data. The HCVIVdb gives an easy to access opportunity to browse through different mutation types (point or multiple substitutions, deletions and insertions). Ranging from specific position(s) of nucleotides to the whole of the HCV-IRES domains II, III and IV can be searched displaying the reported variations with or without the translation efficiencies and other features. A single entry searches of a specific nucleotide change while giving all its reported information, would also list identical mutations and the mutations that are present at same position reported by similar or other studies. Hence, gives an overall comparative analysis of the impact caused by particular or neighboring mutations at specific region of the HCV-IRES. The other features of the HCVIVdb include browsing through all the mutations in their respective individual domains I, II, III and IV. The HCV-IRES variations reported in literature can also be searched according to their corresponding genotypes. The different ranges of translational activities have also been grouped that may describe the susceptibility of particular HCV-IRES regions to variations and the impact caused by these mutations in mediation of translation initiation.

5.3.3 Comparative analyses of mutations in patients' samples and the literature

In the HCV infected patients (4, 7 and 9) we had found mutations that were dispersed throughout HCV-IRES domains II, III and IV. We analyzed the quasispecies and the amount of diversity in HCV-IRES that had existed in those patients over some period of time, described in the above section. The measuring of translation activity in patients gave us an understanding of the significance, upon mutations, of different HCV-IRES regions. In an attempt to further analyze the translation efficiency and the effect mutations can have on HCV-IRES domains we developed a comprehensive database (HCVIVdb). With around 1,800 published variations we can compare the published data with the mutations that we had found in our patients. The single and

multiple mutations were categorized based on the position or order of occurrence in different HCV-IRES domains.

5.3.3.1 Identification of point substitutions in domain II and III of patients' data

We had located some individual mutations in domain II of our patients. The translational response shown by the HCV-IRES having these point mutations in domain II do not display much lowering of activity relative to the pRG-IRES control vector translation efficiency. A couple of constructs P4-G6 and P9-B10, however, found in basal region of domain II showed a decrease in the HCV-IRES efficiency (Tables 7, 8). The samples P7-38 and P9-B7 contain point substitutions at the apical loop of domain IIb, $A^{81}G$ and $G^{82}A$, respectively (Tables 6, 7). The activity measured for these mutants are almost near to the reference type. The mutations found at location G^{82} in literature also showed response, 91% and 88%, that is almost similar to the respective wild type (Barria et al., 2009, Kalliampakou et al., 2002). However, deletion of this nucleotide results in reduction of the HCV-IRES activity to a mere 42%, relative to the wild type (Kalliampakou et al., 2002). Likewise, upon deletion of other nucleotides in domain IIb apical loop display a devastating response to the HCV-IRES efficiency, whereas the individual substitutions causes mixed translation responses depending upon position of the mutation (Kalliampakou et al., 2002, Filbin and Kieft, 2011). Introduction of multiple substitutions at the apical loop of domain IIb also generate lower HCV-IRES responses (Kalliampakou et al., 2002, Odreman-Macchioli et al., 2001, Filbin and Kieft, 2011). The apical loop of domain IIb is crucial for the viral genome translation as it makes contact with the 40S ribosome resulting in a conformation change in the head region (Figure 1) (Yamamoto et al., 2015, Spahn et al., 2001, Bhat et al., 2015).

Several point substitutions have also been located in the domain III. The nucleotide changes found at positions A^{185} and C^{186} in constructs P9-A9 and P9-G8 does not impact the HCV-IRES activity significantly (Table 7). These two samples located at the internal loop of domain IIIb have a translational response of 89% and 93%, respectively, relative to the pRG-IRES control vector translation efficiency. A range of activities has been reported at positions nt. 185 and nt. 186 with a combination of mutations that displayed rather decrease in the HCV-IRES response.

Contradictory to our findings in mutant P9-A9 the nucleotide substitution reported has an activity of only 26%. Other combination of mutations at this location disrupting Watson-Crick base-pair A^{185}-U^{212} abrogates translation response (Collier et al., 2002). The nucleotide at position C^{186} forms a mismatch pair with C^{211} (Figure 2) and nucleotide changes that may result in any other base-pair formation at this site has not been able to restore the translation efficiency which suggest the conservation of this region plays a significant role in promoting the translation initiation (Collier et al., 2002). However, the substitution that we had encountered in our patient P9-G8 has not been reported, $C^{186}U$, resulting in a U^{186}-C^{211} mismatch that has exhibited an activity similar to the WT. The reverse of this base-pairing C^{186}-U^{211} restores translation only to 55% (Collier et al., 2002).

Another mutation at basal region of domain IIIb, $A^{223}G$, in mutant P9-B2 causes the lowering of activity only slightly, i.e. 89% (Table 7). However, an identical mutation, $A^{223}G$, reported in a patient whose activity measured using rabbit reticulocyte lysate (RRL) was demonstrated to be as decreased as 55%. A multiple mutant having similar substitution $A^{223}G$ along with $C^{340}U$, nevertheless, restores the response to 86% compared to the WT (Barria et al., 2009). Point mutation at nucletoide position 340 has reported to retain translation efficiency similar to the WT (Barria et al., 2009, Kieft et al., 1999, Motazakker et al., 2007) suggesting that mutation at position 340 while it occurred in a multiple mutant plays a role in restoring the activity.

We also had found mutations in our patients at IIIb apical loop region (Figure 2). The mutations found in IIIb apical loop region have no profound impact on translation. The measured activity of these constructs (P7-2, P7-21 and P4-E9, P4-F3) was all near to the WT (Tables 6, 8). We compared the activity of our patients' mutations in IIIb apical loop to the mutations reported in the literature. The activities in both the literature and in our patients were found to have only a slight influence on the HCV-IRES translation. However, substitution of whole loop, nucleotides at positions 194-199 and 192-205, results in 25% and 50% activity, respectively (Buratti et al., 1997, Kieft et al., 2001).

5.3.3.2 Probable inter-domain long-range interaction of HCV-IRES in patients' samples

We came across an intriguing phenomenon that displays the restoration of translational activity in few of our patients' samples. These samples comprise multiple mutations, the occurrence of which individually restricts the HCV-IRES response to a minimum level. The restoration of activity upon the presence of these mutations that otherwise upon individual appearance will abrogates the HCV-IRES response, is quite interesting. Likewise, we also came across point substitutions that may not affect functioning of the HCV-IRES but combined effect of these mutations shown to have deleterious outcome in terms of the translational response. We have found four examples in our patients' samples that give us an indication of existence of a probable inter-domain long-range interaction among domains of the HCV-IRES.

1. In mutant 1 ($A^{70}U$ / $U^{264}C$ / $A^{352}G$) from Patient 9 sample (P9-B11), we found a triple mutant with an activity of 64% relative to the pRG-IRES control vector translation efficiency (Table 9, colored; red). The substitutions that this mutant incorporates are at domains II, IIId and IV that are important for the HCV-IRES interaction with the 40S ribosome subunit (Figure 10). Hence, the changes in nucleotide sequences that we encountered in literature have shown to generate a response that is far lower compared to the wild type. For example, region that make up domain II loop E-motif has shown to be susceptible to the HCV-IRES variations and changes whether individual or collective can cause a little to no activity (Otto and Puglisi, 2004, Tang et al., 1999, Kalliampakou et al., 2002, Odreman-Macchioli et al., 2001, Locker et al., 2007). It has been reported that deletion of this region can result in complete inhibition with no formation of the 80S ribosome complex (Locker et al., 2007, Odreman-Macchioli et al., 2001). Although, no identical mutation, $A^{70}U$, was found in literature, a transition $U^{97}A$ disrupts Watson-Crick base-pair A^{70}-U^{97} to A-A results in a moderate activity of 75% (Kalliampakou et al., 2002). Other point substitutions that were located in this region in our patients and the literature showed fairly low activity (Table 9).

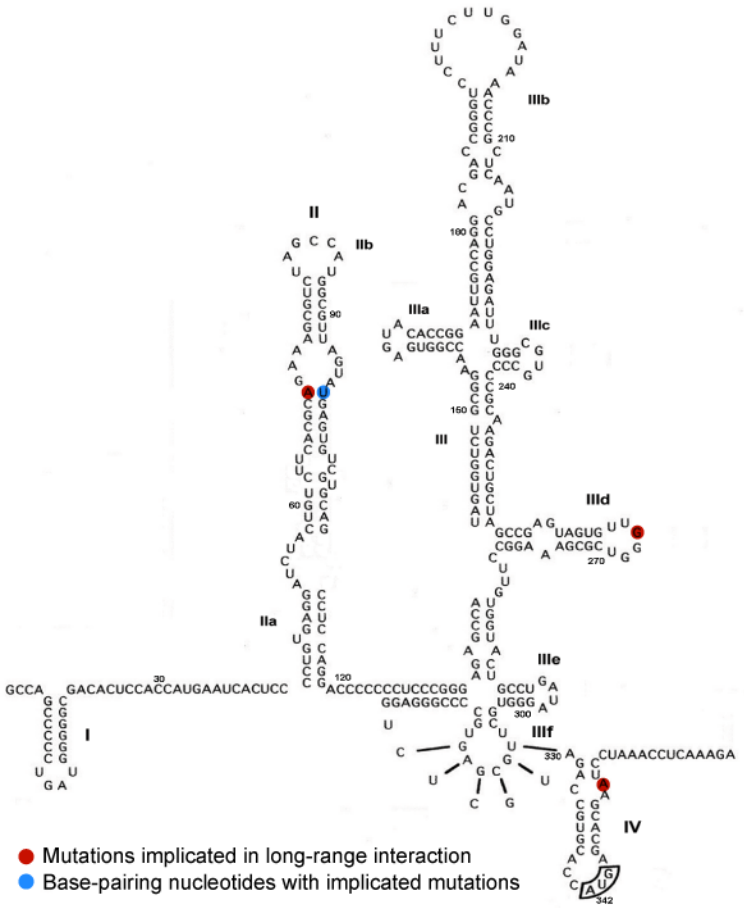

Figure 10: The HCV-IRES secondary structure displaying the mutations implicated in a long-range inter-domain interaction (red) from mutant 1, sample (P9-B11). The respective nucleotides participating in base-pair formation is also highlighted (blue).

The added mutation in mutant 1, $U^{264}C$, occurs in a region that makes the most direct contact with the 40S ribosome and any changes in this region, domain IIId, renders the HCV-IRES functionless. Point substitutions along with the combination of mutations that may include the complementary mutations influence the HCV-IRES response to a great extent (Kieft et al., 1999, Jubin et al., 2000, Klinck et al., 2000) (Table 9). The impact of primary sequence preservation was shown in one study where particular nucleotide changes in the IIId apical loop may retain the structure similarity to the WT, observed through circular dichroism (CD), still manages to display a lower translational response (Barria et al., 2009).

The domain IV has been found to unwind before placing itself in the 40S binding groove (Filbin and Kieft, 2011, Quade et al., 2015). It has been shown through mutation analysis that nucleotide changes that lead to the stability of this domain are proportional to the decrease in translation response and vice versa (Honda et al., 1996a) (Table 9). The point substitutions that have been found in the domain IV stem-loop region in our patients and also in other studies does not seem to impact the translational activity to any significant level (Laporte et al., 2000, Barria et al., 2009, Kieft et al., 1999). However, the additional mutations can be damaging to translational response depending upon the location of occurrence of the

supplementary mutation(s). Here, we found a transition of $A^{352}G$ (Figure 10) and no identical mutation at this location was identified in the published data. The variation of A to G may stabilize the region to some extent that can result in a deficient activity. It is quite intriguing that presence of these triple substitutions causes a restoration in the HCV-IRES translation where upon their individual occurrence is detrimental, as explained above.

	Probable Long-range	Translation activity	Pat. Sample / References
Mutant 1	A^{70} U / U^{264} C / A^{352} G	64	P9-B11
A70	U ^{97} A	75	(Kalliampakou et al., 2002)
	G ^{100} A	68	(Barria et al., 2009)
	U ^{101} C	85	P9-E11
	UUCG $^{70-97}$ dII- mid	14/ 25	(Otto and Puglisi, 2004)
U264	U ^{264} A	13	(Kieft et al., 1999)
	U ^{264} A	~ 43	(Jubin et al., 2000)
	U ^{265} A	47	(Kieft et al., 1999)
	U ^{265} A	~ 43	(Jubin et al., 2000)
A352	A ^{348} G	68	(Laporte et al., 2000)
	G ^{350} A	106	(Barria et al., 2009)
	^{351} insert GGCAGCACG (9 nt.)	34	(Honda et al., 1996a)
	^{351} insert AGCACG (6 nt.)	104	(Honda et al., 1996a)
	^{354} insert GGC	74	(Honda et al., 1996a)

Table 9: Translation response of the HCV-IRES probable long-range interactive mutant 1, sample P9-B11 (colored: red), is shown and compared with reported mutations occurring at the identical positions. The mutations at adjacent locations have also been listed to establish a comprehensive overview with reference to a probable inter-domain long-range interaction that culminates in restoration of viral mRNA translation.

2. The mutant 2 ($A^{72}G$ / $U^{234}C$ / $U^{248}A$) which we have found in Patient 4, sample P4-C5, comprises mutations that are at positions important for the HCV-IRES activity (Figure 11). The first mutation in that triple mutant, $A^{72}G$, is present in loop E-motif of the domain II, a conserved region among all HCV genotypes. An identical substitution was found at similar position, $A^{72}G$, which exhibited a very low translational response (28.6%), relative to the WT (Tang et al., 1999) (Table 10). Other point or multiple substitutions that were located in this region also showed a reduced HCV-IRES activity. The nucleotide changes in this region are deleterious for the HCV-IRES response, as described above in mutant 1.

The nucleotide U^{234} along with G^{233} is involved in the formation of a bi-loop of domain IIIc, observed through NMR spectroscopy (Rijnbrand et al., 2004). The nucleotide G^{233} also makes contact with the eS27 of the 40S subunit (Quade et al., 2015, Malygin et al., 2013b). No exact mutation, $U^{234}C$, at this location was found in any study. Introduction of substitutions to the entire IIIc loop results in considerable decrease in the HCV-IRES activity (Kieft et al., 1999) whereas, complementary mutations can cause a range of responses from 50% decline to as active as the WT (Rijnbrand et al., 2004) (Table 10). The adjacent stem region of domain IIIc is also sensitive to the nucleotide variations and mutations can hinder regulation of the translation initiation.

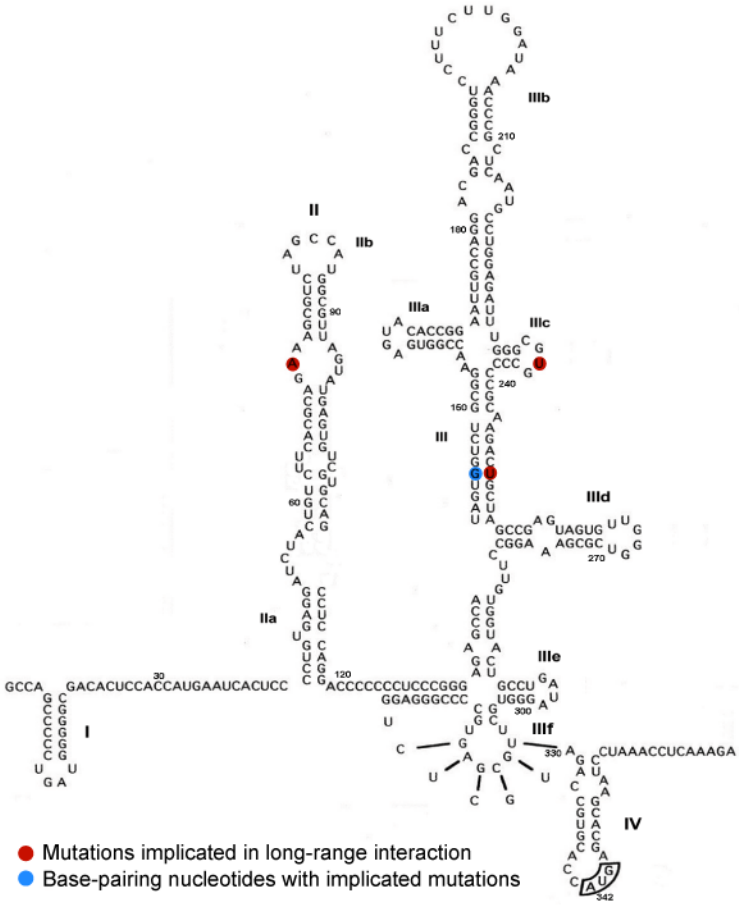

Figure 11: The HCV-IRES secondary structure displaying the mutations implicated in a long-range inter-domain interaction (red) from mutant 2, sample (P4-C5). The respective nucleotides participating in base-pair formation is also highlighted (blue).

The nucleotide change $U^{248}C$ occurs in the stem III* that connects junction IIIabc to the domain IIId. Mutations in this region that have been published do not display much impact on the HCV-IRES mediated translation. A marginal to mediocre decrease in activity relative to the WT is observed (Laporte et al., 2000, Motazakker et al., 2007).

Since, other mutations that we described, $A^{72}G$ and $U^{234}C$, cause a decrease in translation when occurs individually, the mutant 2 shows a recovery of functional response upon collective occurrence of these mutations.

	Probable Long-range	Translation activity (%)	Pat. Sample / Publication
Mutant 2	A ^{72} G / U ^{234} C / U ^{248} A	100	P4-C5
A72	A ^{72} G	28.6	(Tang et al., 1999)
	A ^{74} G	74.8	(Tang et al., 1999)
	A ^{93} U	52	(Kalliampakou et al., 2002)
	A ^{96} U	65	(Kalliampakou et al., 2002)
	A ^{73} G / C ^{340} U / G ^{350} A	50	(Barria et al., 2009)
	GAA del. $^{71-73}$	30	(Odreman-Macchioli et al., 2001)
	AGUA del. $^{93-96}$	Complete inhibition	(Odreman-Macchioli et al., 2001)
	UAGUA $^{92-96}$ UUUC	no 80S assem.	(Locker et al., 2007)
	GAA del. $^{71-73}$ / AGUA del. $^{93-96}$	21	(Odreman-Macchioli et al., 2001)
	$^{71-73}$ AGUA / $^{93-96}$ UUC	23	(Otto and Puglisi, 2004)
U234	G ^{233} G / U ^{234} A	WT	(Rijnbrand et al., 2004)
	G ^{233} A / U ^{234} A	~ 50	(Rijnbrand et al., 2004)
	G ^{233} A / U ^{234} G	~ 50	(Rijnbrand et al., 2004)
	G ^{235} A	3.9 -12.7	(Tang et al., 1999)
	CGUG $^{232-235}$ GCAG	20	(Kieft et al., 2001)
U248	A ^{243} G	92	(Motazakker et al., 2007)
	G ^{245} A	95.8	(Tang et al., 1999)
	A ^{246} G	93.8	Patient 9 - A12
	A ^{142} G	62	(Laporte et al., 2000)

Table 10: Translation response of the HCV-IRES probable long-range interactive mutant 2, sample P4-C5 (colored: red), is shown and compared with reported mutations occurring at the identical positions. The mutations at adjacent locations have also been listed to establish a comprehensive overview with reference to a probable inter-domain long-range interaction that culminates in restoration of viral mRNA translation.

3. The proposed phenomenon of the inter domain long-range interaction was also observed in another multiple mutant sample from the Patient 4. This mutant 3 ($A^{66}G$ / $A^{182}C$ / $C^{240}U$) from sample P4-B3 comprise mutations in regions that make up the domain II, domain IIIb and stem III*, respectively (Figure 12). No similar mutation at location A^{66} that may suggest the affect caused by this nucleotide change on the HCV-IRES activity was found in the literature. A mutation, $U^{101}C$, that alters the base-pairing with A^{66} from a Watson-Crick A^{66}-U^{101} (Figure 12) to A-C, found in

our patient, restrict the translation efficiency only to 85%, relative to pRG-IRES control vector translation efficiency. The nucleotide, U^{101}, is also being protected from iodine cleavage upon the 40S attachment (Kieft et al., 2001). However, a published variation, $G^{100}A$, resulting in the formation of C-A base-pairing causes a decrease in response to 68% (Barria et al., 2009). It may be deduced that primary sequence variations in this region can be tolerated to some level as long as it does not interfere with the secondary structure conformation of the HCV-IRES. Although, there has not been many mutations that were discovered for the domain IIa stem region, the collective mutations observed in this region has exhibited vulnerability to the HCV-IRES function (Honda et al., 1999b) (Table 11) that may hint towards the importance of its sequence and structural preservation.

The second mutation in mutant 3, $A^{182}C$, occurs in the IIIb internal loop region that makes contact with the eIF3 as shown through different mutation, structural and biochemical analysis (Collier et al., 2002, Sizova et al., 1998, Pestova et al., 1998, Siridechadilok et al., 2005). The nucleotide, A^{182}, forms a purine-purine interaction with G^{217} and changes in any of the nucleotides have shown a range of activities with mostly decrease in the translation response. The identical substitution has demonstrated an efficiency of 68%, compared to the WT (Collier et al., 2002, Wang et al., 1994b). The modification interference data while employing diethylpyrocarbonate (DEPC) or

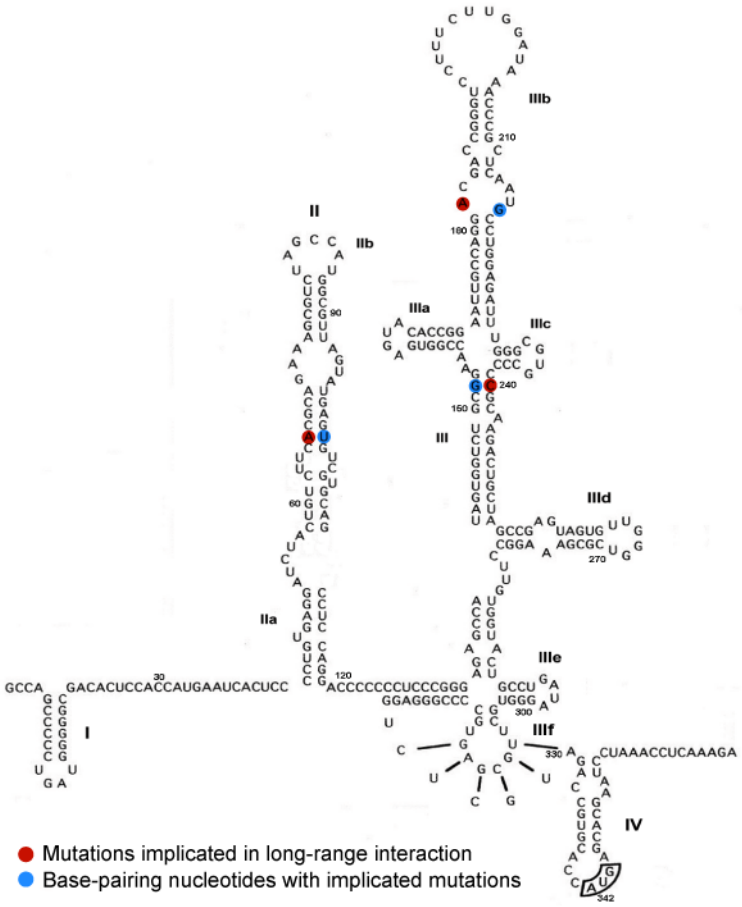

Figure 12: The HCV-IRES secondary structure displaying the mutations implicated in a long-range inter-domain interaction (red) from mutant 3, sample (P4-B3). The respective nucleotides participating in base-pair formation is also highlighted (blue).

aqueous hydrazine treatment has also identified A^{182} along with other nucleotides at positions 214, 215 and 216, in the IIIb internal loop that were protected upon the eIF3 binding suggesting that conservation of the region is important for interaction with the translation machinery. The nucleotide G^{217}, on the other hand, had exhibited a weaker interference pattern (Kieft et al., 2001). The similar nucleotides at locations 214, 215 and 216 were also shown to be protective from RNase I cleavage upon the eIF3 attachment (Sizova et al., 1998). These nucleotides are also involved in the formation of an S-turn motif displayed through the NMR (Collier et al., 2002).

	Probable Long-range	**Translation activity (%)**	**Pat. Sample / Publication**
Mutant 3	A ^{66} G / A ^{182} C / C ^{240} U	98	P4-B3
A66	U ^{101} C	85	P9-E11
	G ^{100} A	68	(Barria et al., 2009)
	U ^{97} A	75	(Kalliampakou et al., 2002)
	CACGC $^{65-69}$ GUGAG	8.3	(Honda et al., 1999b)
	GAGUG $^{98-102}$ CGCAC	5.7	(Honda et al., 1999b)
	$^{65-69}$ GUGAG / $^{98-102}$ CGCAC	32	(Honda et al., 1999b)
A182	A ^{182} C / G ^{217} G	68	(Collier et al., 2002)
	A ^{182} G / G ^{217} G	56	(Collier et al., 2002)
	A ^{182} U / G ^{217} G	76	(Collier et al., 2002)
C240	A ^{243} G	92	(Motazakker et al., 2007)
	G ^{245} A	95.8	(Tang et al., 1999)
	A ^{246} G	93.8	P9-A12

Table 11: Translation response of the HCV-IRES probable long-range interactive mutant 3, sample P4-B3 (colored: red), is shown and compared with reported mutations occurring at the identical positions. The mutations at adjacent locations have also been listed to establish a comprehensive overview with reference to a probable inter-domain long-range interaction that culminates in restoration of viral mRNA translation.

The introduction of mutation $C^{240}U$ may result in decreasing the stability of the structure due to a stretch of adjacent GC base-pairs observed at this region, close to junction IIIabc of the HCV-IRES (Figure 12). The junction nucleotides A^{154}, A^{155} and U^{228} are important for the HCV-IRES activity as shown through the structure and mutation analysis (Kieft et al., 2002, Kieft et al., 1999). Since, junction IIIabc interacts with both the 40S and eIF3, any changes can be deleterious with respect to its interaction with translation machinery and in turn the HCV-IRES translation response

(Ji et al., 2004). The nucleotide variation $C^{240}U$ alters the Watson-Crick base-pair, G^{152}-C^{240}, to G-U which may have some effect on the strong structure observed in this region. The nucleotide G^{152} along with G^{150} is also protected from iodine cleavage upon the 40S attachment (Kieft et al., 2001). Likewise, the protection assay employing RNase V1 shows the protection of nucleotide G^{241} from cleavage upon the 40S subunit binding (Kolupaeva et al., 2000a). The point substitutions that we have found with activities in this region, reported in the literature, have shown not to impact the translation to any greater extent (Tang et al., 1999, Motazakker et al., 2007) (Table 11). After analyzing all the three substitutions in mutant 3 and the possible impact they can have on the HCV-IRES mediated translation, it is still intriguing to know the mechanism HCV-IRES may undertake to compensate for its activity and function. The mutations that we have observed in this case, mutant 3 sample P4-B3, are present in crucial locations of the domain II and III and as described above can have important implications on altering the structural conformation and hence the HCV-IRES function.

4. The other phenomenon that we discovered, parallel to compensation of the translation response upon probable long-range inter-domain interaction was a sharp decline in the HCV-IRES activity in multiple mutants that comprises point substitutions having insignificant effect on translation. The mutant 4, sample P7-24, (Insertion C^{127} / $U^{147}C$) is a double mutant that was found in the Patient 7. This mutant hardly showed any active translation and the reduction in response was found to be as low as 6%, relative to the pRG-IRES control vector translation efficiency (Table 12). The occurrence of C insertion at position 127 in the linker region that connects the domain II with domain III and part of it also forms a double stranded region that makes up a stem region of the domain IIIf pseudoknot (Figure 13). The individual mutations, insertions and deletions observed in this region, in addition to the C insertion at 127 had displayed least impact on the HCV-IRES translation. The reported insertion C^{127} had showed an activity of 96% (Barria et al., 2009). The deletion of adjacent nucleotide C^{126} that have been found in our patients' sample and also reported in a study showed not much difference in activity, compared to the HCV-IRES WT (Laporte et al., 2000). Multiple mutations that were reported to had

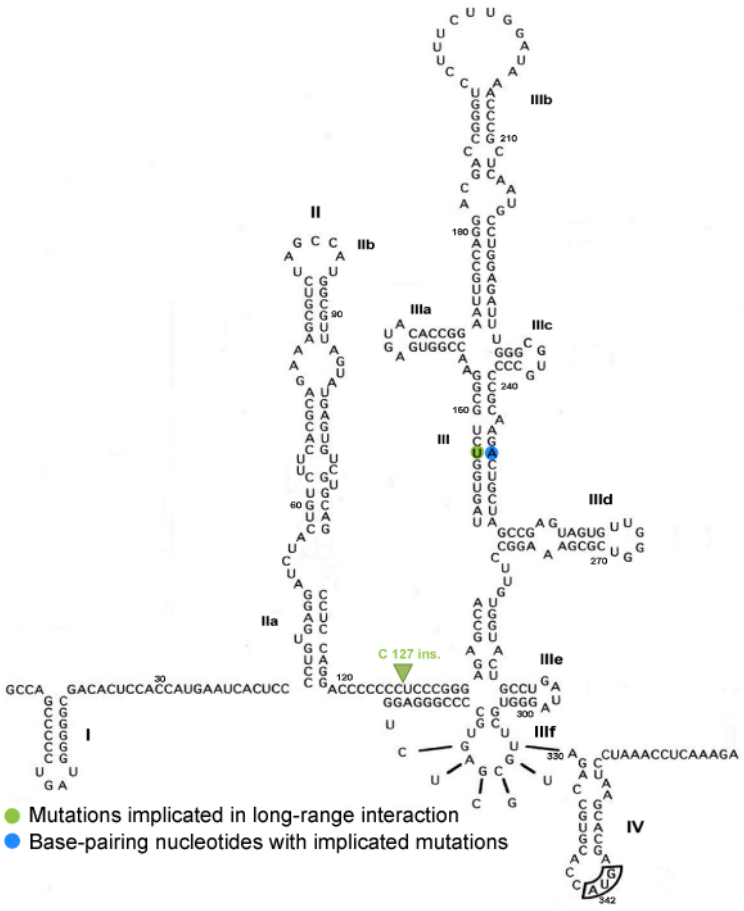

Figure 13: The HCV-IRES secondary structure displaying the mutations implicated in a long-range inter-domain interaction (green) from mutant 4, sample (P7-24). The respective nucleotides participating in base-pair formation is also highlighted (blue).

incorporated insertion C^{127} along with additional mutations, displayed a response seemed to be dependent more so on the activity of the supplementary mutation(s) based on their location rather the C insertion at 127 (Laporte et al., 2000, Barria et al., 2009). This may suggest that the affect of insertion at C^{127} on the HCV-IRES translation is negligible. Other substitutions in the poly(C) linker region also do not hinder function of the HCV-IRES translation (Zhang et al., 1999, Motazakker et al., 2007, Berry et al., 2010) (Table 12). The pseudoknot stem, adjacent to the domain II-III linker, has also shown to be tolerant of single base-pair changes at its 5' terminus. However, switching the purine-pyrimidine identity of two base-pairs at 5' terminus decreases the translation efficiency to moderate levels. The activity is restored to some degree upon introduction of the complementary base-pairing (Berry et al., 2010, Berry et al., 2011). Deletion of two base-pairs in comparison to one has shown to abrogate the translation response (Berry et al., 2010). Collective substitutions were introduced in the upper and lower strands of the first stem of pseudoknot that decreased the activity to much lower level. The lessen activity was completely restored upon complementary base-pairing (Wang et al., 1994b) (Table 12). This may advocate that collective nucleotides' conservation in this region is important for preservation of the HCV-IRES global conformation and in turn its possible direct interaction and placement of AUG initiation codon to the 40S

binding groove. Toe-printing analysis or primer extension inhibition assay has shown stops at positions G^{318}, G^{320} and U^{324} (Otto and Puglisi, 2004, Pestova et al., 1998, Kolupaeva et al., 2000a).

	Probable Long-range	Translation activity (%)	Pat. Sample / Publication
Mutant 4	C insertion ^{127} / U ^{147} C	6	P7-24
127	C Insertion ^{127}	96	(Barria et al., 2009)
	C deletion ^{126}	97.6	Patient 7 - 46
	C deletion ^{126}	80	(Laporte et al., 2000)
	A ^{119} C	84	(Barria et al., 2009)
	A ^{119} C	88	(Motazakker et al., 2007)
	A ^{119} C	~ 90	(Zhang et al., 1999)
	C^{122} + C^{123} del.	73 ± 4	(Berry et al., 2010)
	CC ins. at C^{123}	82 ± 5	(Berry et al., 2010)
	CUCC $^{126 - 129}$ AGAA	7.5	(Wang et al., 1994b)
	GGAG $^{319 - 322}$ UUCU	8.7	(Wang et al., 1994b)
U147	A ^{246} G	93.8	P9-A12
	G ^{245} A	95.8	(Tang et al., 1999)
	T ^{251} C	95	P9-B1
	A ^{142} G	62	(Laporte et al., 2000)
	C ins. ^{127} / A ^{142} G	58	(Laporte et al., 2000)

Table 12: Translation response of the HCV-IRES probable long-range interactive mutant 4, sample P7-24 (colored: green), is shown and compared with reported mutations occurring at the identical positions. The mutations at adjacent locations have also been listed to establish a comprehensive overview with reference to a probable inter-domain long-range interaction that culminates in sharp reduction of viral mRNA translation.

The nucleotide U^{147} occurs in the stem III making a predicted Watson-Crick base-pairing with A^{246}. Transition of this nucleotide, $U^{147}C$ changes the base-pairing to C-A. No exact mutation from the published data at this location was found. We did find a mutation in one of our Patient 9s' sample (P9-A12) at position A^{246} to G which changes the predicted base-pairing from U-A to U-G. This results in an activity that is equivalent to the wild type and causes no changes in productivity of the HCV-IRES. It can possibly be due to the stability that is offered by G-U base-pairing rather than C-A, which helps to keep the integrity of the structure and in turn have no effect on the function. Through toe-printing assay the eIF3/HCV-IRES complexes were displayed to arrest primer extension at nucleotide A^{243} and A^{244} (Pestova et al., 1998). Since,

the enzymatic probing and primer extension inhibition assays had shown the interaction of particular nucleotides in stem III* of the HCV-IRES either with eIF3 or the 40S subunit, it can be assumed that structural preservation of this region is important. However, observing the data from the HCVIVdb suggests that relevance of the primary sequences in context to the nucleotide variations in this region do not influence the HCV-IRES function. The only mediocrity in translation while having a point mutation in the stem III* was observed upon substitution of $A^{142}G$ which culminated in 58% relative to the WT (Laporte et al., 2000). The nucleotide U^{251} forms a predicted base-pair with A^{142}, is protected from iodine cleavage upon the 40S binding (Kieft et al., 1999) (Table 12). The substitution $A^{142}G$ modifies the base-pairing with U^{251} and may impact the conformation of the nucleotide and its probable interaction with the 40S subunit. The sharp decline in HCV-IRES translation that we have observed in the mutant 4 sample (P7-24) is due to the collective effect that may require a probable long-range interaction, since the influence of these mutations occurring individually, as described, do not have such a devastating effect on translation.

5.3.3.3 Probable inter-domain long-range interaction of HCV-IRES in literature

We also came across similar phenomenon of the inter-domain long-range interaction in few of the multiple mutations in our HCVIVdb database. Likewise, those multiple mutants that we had noticed in our patients' samples, the multiple mutants in HCVIVdb also compensate for the HCV-IRES translation while comprising of mutations that have a detrimental impact on the activity upon individual occurrence.

5. The nucleotide transition $C^{43}U$ in mutant 5 ($C^{43}U$ / $U^{262}C$) (Barria et al., 2009) at the basal portion of domain II in the HCV-IRES (Figure 2) can probably be the mutation that resulted in the compensation of translation efficiency in mutant 5. With current data availability the nucleotide, U^{43}, is not shown to be involved in direct interaction with the cellular translational machinery. Mutation analysis has been more focused on the collective substitution of nucleotides which may not conclusively evaluate the impact caused by any point substitution at this region

	Probable Long-range	Translation activity (%)	Pat. Sample / Publication
Mutant 5	C ^{43} U / U ^{262} C	85	(Barria et al., 2009)
C42	C ^{42} U	66	Patient 9 - B10
	CC $^{44-45}$ GG	66	(Zhao and Wimmer, 2001a)
	CU $^{45-46}$ GA	50	(Honda et al., 1999b)
U262	A ^{252} G / G ^{271} A	15 (RRL) / 2 (Huh 7)	(Barria et al., 2009)
	G ^{271} A	59 (RRL) / 26 (Huh 7)	(Barria et al., 2009)
	C^{67}U / A^{252}G / G^{271}A	9	(Barria et al., 2009)
	U ^{259} C	~75-80	(Forton et al., 2004)
	A ^{260} C	40	(Klinck et al., 2000)
	A ^{260} C / G insertion	1	(Klinck et al., 2000)
	A ^{260} U	47	(Klinck et al., 2000)

Table 13: Translation response of the HCV-IRES probable long-range interactive mutant 5 (colored; red) is shown and compared with our patients' samples and other reported mutations occurring at the identical positions. The mutations at adjacent locations have also been listed to establish a comprehensive overview with reference to a probable inter-domain long-range interaction that culminates in restoration of viral mRNA translation.

(Table 13). However, in one of our patients' samples (P9-B10; Table 2) we had found an adjacent transition mutation C^{42}U that showed a fairly moderate activity of 66% compared to the pRG-IRES control vector translation efficiency. The mutations that have been reported in the literature adjacent to this nucleotide occur in predicted base-paired basal region of domain II. These collective mutations have displayed a lower response which restored activity to some level upon covariant substitutions (Honda et al., 1999b, Zhao and Wimmer, 2001a) (Table 13). Since the nucleotide C^{43} occurs in a single-stranded region, introduction of a transition can probably induce a change in its conformation that impact the adjacent regions of the domain II asymmetrical bent, as may well be the case in C^{42}U, the structural integrity of which is vital for the HCV-IRES to make precise contacts with the 40S subunit.

In mutant 5 (C^{43}U / U^{262}C) the substitution U^{262}C occurs in a region that is vulnerable to any nucleotide variations and changes in this domain IIId has largely shown a deleterious effect on the HCV-IRES translation. There have been several studies to observe the functional tolerance of the domain IIId upon introduction of mutations in its stem-loop region and almost all of them had exhibited a reduction in

affinity interaction with the 40S in addition to reduced translation efficiency (Klinck et al., 2000, Otto and Puglisi, 2004, Kieft et al., 1999, Jubin et al., 2000, Lukavsky et al., 2000, Ji et al., 2004) . There is no reported point mutation that could be located at U^{262}. However, the adjacent nucleotide variations, U^{259} and A^{260}, displayed a low activity (Klinck et al., 2000, Forton et al., 2004) (Table 13). The domain IIId has been shown to make the most direct contact with the 40S subunit body, observed through electron microscopy. The HCV-IRES/40S contact takes place near the helix 28 of 18S ribosome which has an expansion segment 7 (Angulo et al., 2016, Boehringer et al., 2005, Joseph et al., 2014). The ribosomal proteins S3a, S14 and S16 have also been shown to cross-linked with the HCV-IRES derivatives with a photoactivatable group at nucleotides G^{263} and A^{275} (Babaylova et al., 2009). The U^{262}-G^{271} base-pair alters to a Watson-Crick C-G upon the $U^{262}C$ substitution. Similarly, point substitution, $G^{271}A$, introduces an A-U pair, the activity of which has been displayed to be 59% and 26% relative to the WT in RRL and Huh-7 cell free translational extracts, respectively (Barria et al., 2009) (Table 13). Enzymatic probing also protected nucleotide U^{262} weakly from RNase V1 cleavage on the 40S ribosomal binding (Kolupaeva et al., 2000a).

6. The substitution $A^{93}U$ in mutant 6 ($A^{93}U$ / $A^{142}G$ / $G^{253}A$) occurs in the domain II loop E-motif (Figure 2). Nucleotide A^{93} is involved in non-canonical base-pairing with A^{73} determined through the NMR (Lukavsky et al., 2003). The nucleotide variation in this position, $A^{93}U$, reported in literature has shown a restricted activity of only 52%, compared to the WT (Kalliampakou et al., 2002) (Table 14). Other mutations both individual and collective in the loop E-motif of domain II, also discussed above, has shown decrease in translational response with no formation of an 80S assembly (Otto and Puglisi, 2004, Tang et al., 1999, Kalliampakou et al., 2002, Odreman-Macchioli et al., 2001). Enzymatic probing has shown protection of most of the nucleotides in the domain II loop E-motif from iodine cleavage when treated with the 40S subunit (Kieft et al., 2001). The mutation analysis on the loop E along with other conserved regions of the domain II, apical loop and asymmetrical bent, has shown that all of these regions are vital in the eIF5 mediated eIF2-GTP hydrolysis and eIF2 release in the HCV and CSFV (Locker et al., 2007). Different functional and biochemical analysis suggests that the role played by the loop E-motif

	Probable Long-range	Translation activity (%)	Pat. Sample / Publication
Mutant 6	$A^{93}U$ / $A^{142}G$ / $G^{253}A$	62	(Laporte et al., 2000)
A93	A ^{93} U	52	(Kalliampakou et al., 2002)
	A ^{96} U	65	(Kalliampakou et al., 2002)
A142	A ^{142} G	62	(Laporte et al., 2000)
	C ins. ^{127} / A ^{142} G	58	(Laporte et al., 2000)
G253	U ^{251} C	91	P9-B1
	A ^{252} G	47 (RRL) / 18 (Huh 7)	(Barria et al., 2009)
	G ^{245} A	95.8	(Tang et al., 1999)
	A ^{243} G	92	(Motazakker et al., 2007)

Table 14: Translation response of the HCV-IRES probable long-range interactive mutant 6 (colored; red) is shown and compared with our patients' samples and other reported mutations occurring at the identical positions. The mutations at adjacent locations have also been listed to establish a comprehensive overview with reference to a probable inter-domain long-range interaction that culminates in restoration of viral mRNA translation.

is both structural and require sequence preservation for the HCV-IRES mediated translation.

The additional substitution in mutant 6, $A^{142}G$, changes the predicted canonical base-pairing with U^{251} from A-U to G-U. The individual mutation that has been reported at this position $A^{142}G$ resulted in a decrease of activity to 58%, relative to the WT (Laporte et al., 2000) (Table 14). The activities that we have found in stem III* of the HCV-IRES are discussed while explaining the mutant 4 sample (P7-24) where a sharp decline of the HCV-IRES translation response is noticed. The overall efficiency upon introduction of point mutations in this region has been rather unaltered compared to the WT from various studies. The same response was identified in collective mutants the activity of which essentially depends on the occurrence of supplementary mutations, observed in the literature (Laporte et al., 2000, Barria et al., 2009, Forton et al., 2004, Motazakker et al., 2007). Considering the nucleotides U^{251} and A^{252} are protected from iodine cleavage upon the 40S binding (Kieft et al., 2001), the alteration of base-pairing nucleotide, $A^{142}G$, can cause a conformational change that may impair the interaction of this region with the 40S subunit. On the other hand, we found a nucleotide variation, $U^{251}C$, in Patient 9 sample (P9-B1) (Table 14) resulting in a predicted A-C base-pair that do not

influence the HCV-IRES translation. Unlike, an adjacent mutation, $A^{252}G$ causes a lowering of translation response to 47% and 18% respectively (Barria et al., 2009) which estimates that a transition resulting in change of predicted canonical base-pairing U^{141}-A^{252} to U-G is not tolerated. However, in both the cases, $A^{142}G$ and $A^{252}G$, end up in a wobble G-U, which seem to stabilize the region and lower the IRES translational response.

The bordering nucleotide to A^{252} is G^{253} that makes the termini of domain IIId, adjoining the stem III*. The impact of nucleotide variation G^{253} on translation was not found to be reported in patients and in the literature. But as we have discussed in above mutant 1 sample (P9-B11) and mutant 5, the individual and multiple mutations occurring in the domain IIId stem-loop region influence the HCV-IRES translation to significant levels. Variety of studies relating to functional and biochemical relevance asserts the importance of primary sequence and secondary structure preservation of the domain IIId region. Through NMR structure, G^{253} forms a predicted Watson Crick base-pair with C^{271} (Lukavsky et al., 2000). The 62% activity found in triple mutant 6 is suggestive of some form of tertiary interaction between nucleotides, spatially distant, that may also structurally adapts to binding with translation machinery and resultantly compensate the HCV-IRES function.

7. The nucleotide deletion C^{55} in mutant 7 (Deletion C^{55}/ $U^{63}A$/ $U^{64}C$/ $A^{93}G$) occurs in a region that makes the asymmetrical bent of domain II (Figure 2). The structural data with the help of X-ray crystallography and NMR has shown domain IIa making a right angle bulge (nucleotide A^{53}-A^{57}) with the upper stem of domain IIa. The stability of bent structure and conformation of the domain IIa is metal-ion dependent. The nucleotides A^{53}, A^{54} and C^{55} make continuous part of the bulge and stack on base pair G52-C111 (Lukavsky et al., 2003, Dibrov et al., 2007, Zhao et al., 2008, Quade et al., 2015). There is no identical point mutation that can be identified in our patients' samples and the HCVIVdb at nucleotide position 55. However, functional analysis with deletion of this helical bent region has shown complete inhibition of the HCV-IRES activity with no formation of the 80S assembly (Odreman-Macchioli et al., 2001, Locker et al., 2007) (Table 15). Similarly, there is a less pronounced bent in between the domain II asymmetrical bent and loop E-motif as observed through the NMR and X-ray crystallography (Lukavsky et al., 2000, Dibrov

et al., 2007, Boerneke et al., 2014). Functional data provided us with one point substitution, $U^{63}C$, with a display of marginal translation response of around 75%, relative to the WT (Tang et al., 1999). The stretch of nucleotides at this region with non-canonical base-pairing may facilitate domain II's interaction with the 40S. The U^{63}-C^{104} base-pairing alteration to C-C, in above example, still able to retain an average translation response is due to the maintenance of pyrimidine-pyrimidine base-pairing which may conserve the structural conformation of this region. Likewise, substitution $C^{104}U$ reported in the literature (Barria et al., 2009) also maintains a pyrimidine-pyrimidine base-pairing, U-U, results in an activity that is near to the WT. However, mutation $U^{63}A$ observed in mutant 7 changes base-pairing interaction to purine-pyrimidine and since we have no particular example no precise conclusion can be drawn out of this substitution in terms of its impact on translation response. A triple substitution reported in one study, introducing purines instead of pyrimidines at positions 62-64 had determined only a quarter of activity, relative to the WT (Honda et al., 1999b) (Table 15).

	Probable Long-range	**Translation activity**	**Pat. Sample /**
Mutant 7	C del.^{55} /U^{63}A / U^{64}C / A^{93}G	80	(Laporte et al., 2000)
C55	A ^{53} G	90	(van Leeuwen et al.,
	AAC del. $^{53\text{-}55}$	0	(Odreman-Macchioli et
	Del AACUA $^{53\text{-}57}$ (ΔdIIa)	no 80S assembly	(Locker et al., 2007)
U63,U64	U ^{63} C	74.8	(Tang et al., 1999)
	C ^{104} U	95	(Barria et al., 2009)
	C^{62}G / U^{63}G / U^{64}A	25	(Honda et al., 1999b)
	U del. ^{103}	99	(Odreman-Macchioli et
	UCUU del. $^{61\text{-}64}$ / U del. ^{103}	26	(Odreman-Macchioli et
A93	A ^{93} U	52	(Kalliampakou et al.,
	A ^{96} U	65	(Kalliampakou et al.,
	A ^{72} G	28.6	(Tang et al., 1999)

Table 15: Translation response of the HCV-IRES probable long-range interactive mutant 7 (colored; red) is shown and compared with our patients' samples and other reported mutations occurring at the identical positions. The mutations at adjacent locations have also been listed to establish a comprehensive overview with reference to a probable inter-domain long-range interaction that culminates in restoration of viral mRNA translation.

Nucleotide variation, $U^{64}C$ in mutant 7, maintains that pyrimidine-pyrimidine base-pairing with nucleotide U^{103}, although modifying it from U-U to U-C. The deletion of U^{103} has shown to have no impact on activity whatsoever (Odreman-Macchioli et al., 2001). However, deletion of this whole loop can be deleterious for the HCV-IRES activity (Honda et al., 1999b). Other collective mutations at this region and the adjacent stem connecting to the loop E-motif of domain II, has shown responses that were on the lower side (Honda et al., 1999b, Odreman-Macchioli et al., 2001, Barria et al., 2009, Kalliampakou et al., 2002) suggesting the preservation of structural and functional integrity is indispensable. Furthermore, all four nucleotides of these two base-pairs U^{63}-C^{104} and U^{64}-U^{103} are protected from iodine cleavage upon the 40S binding (Kieft et al., 2001).

The change in nucleotide, $A^{93}G$ from the mutant 7, occurs in the loop E-motif of domain II. A nucleotide variation at identical location has also been found in mutant 6 and discussed with respect to its impact on the structural and functional conservation of the HCV-IRES. Similarly, other mutations that we found in the loop E-motif in mutant 1 sample (P9-B11) and mutant 2 sample (P4-C5) also gives an insight into the working and importance of this region in maintenance of structural significance and mediation of the HCV-IRES translation initiation.

All of these collective mutations in mutant 7 occur in the domain IIa region. The region has significant value in determining efficiency of the HCV-IRES which is evaluated through introducing nucleotide changes on individual and multiple levels and measuring the overall impact. Considering the occurrence of mutations in mutant 7 at location that is vulnerable to variation(s), the compensation of HCV-IRES translation is quite intriguing.

8. The substitution, $C^{121}U$, found in mutant 8 ($C^{121}U$/ $C^{340}A$/ $G^{350}A$) (Barria et al., 2009) occurs in the poly(C) linker region that connects the domain II with domain III (Yamamoto et al., 2015). Point substitutions that are found or reported in the literature in poly(C) linker have not shown to affect the translation to any higher levels (Table 16). In mutant 4 sample (P7-24), we have viewed a detailed structural, functional and biochemical analysis of this region that along with the pseudoknot is significant for the placement of AUG initiation codon to the 40S mRNA binding cleft.

Base pair deletions and insertions with additional mutations of the poly(C) linker have shown to retain the translation activity (Berry et al., 2010, Berry et al., 2011, Barria et al., 2009, Laporte et al., 2000, Zhang et al., 1999). However, a reported transition mutation $C^{121}U$ displayed a decrease in activity to 60%, relative to the WT (Motazakker et al., 2007) (Table 16).

	Probable Long-range	**Translation activity**	**Pat. Sample / Publication**
Mutant 8	$C^{121}U$ / $C^{340}A$ / $G^{350}A$	107	(Barria et al., 2009)
C121	C ^{121} U	60	(Motazakker et al., 2007)
	A ^{119} C	84	(Barria et al., 2009)
	A ^{119} C	88	(Motazakker et al., 2007)
	A ^{119} C	~90	(Zhang et al., 1999)
C340	C ^{340} A	104	(Barria et al., 2009)
	C ^{340} U	100	(Motazakker et al., 2007)
	U ^{340} C	Retains full act.	(Kieft et al., 1999)
G350	G ^{350} A	106	(Barria et al., 2009)
	C ^{340} U / G ^{350} A	106 ± 6	(Barria et al., 2009)
	C Ins.^{127} / $C^{340}U$ / $G^{350}A$	102	(Barria et al., 2009)
	$C^{121}U$ / $U^{326}C$ / $G^{350}A$	45	(Barria et al., 2009)
	U^{326}-A^{310} (U-A) del.	80 ± 11	(Barria et al., 2009)

Table 16: Translation response of the HCV-IRES probable long-range interactive mutant 8 (colored; red) is shown and compared with our patients' samples and other reported mutations occurring at the identical positions. The mutations at adjacent locations have also been listed to establish a comprehensive overview with reference to a probable inter-domain long-range interaction that culminates in restoration of viral mRNA translation.

Other substitutions in mutant 8, $C^{340}A$ and $G^{350}A$, occurs in the domain IV (Figure 2) and functional response reported in multiple studies suggests that nucleotide variations are tolerant in context to the HCV-IRES function at both the locations (Motazakker et al., 2007, Barria et al., 2009, Kieft et al., 1999) (Table 16). Some nucleotide elements in domain IV are shown to be crucial for rpS5-IRES interaction (Bhat et al., 2015). The analysis carried out employing SHAPE displayed a decrease in modification of nucleotides in predicted domain IV loop containing the start codon, including the nucleotide at position 340, upon the 40S binding. However, strand downstream of the AUG start codon upon the 40S attachment showed

increased SHAPE flexibility and modification that encompass the nucleotide at position 350 (Filbin and Kieft, 2011). It has previously been indicated through mutation analysis that stabilizing the domain IV structure reduces the translation initiation (Honda et al., 1996a). Although, the point substitutions observed for mutations in the domain IV have not shown to decrease the activity and retained a near wild type efficiency. The mutation reported for $C^{121}U$ had an average response which seemed to be improved upon the addition of mutations in domain IV.

5.3.4 Identification and characterization of novel mutations from the patients' samples. Mapping the impact of mutations on HCV-IRES/40S/eIF3 interaction and translation response

The measured translation activities of the HCV-IRES from our patients' samples interested us to collect more mutations that have been reported in the literature. The underlying objective was to illustrate the impact of these mutations on HCV-IRES translation through comparative analysis between our patients' samples and the available published data. In addition, structural and biochemical reports to gain further insight into the working and behavior of the HCV-IRES mediated translation upon mutation(s) at specific regions was also analyzed. We found ~19 new mutations in our patients (Table 17) that we were not able to locate in the HCVIVdb. The novel mutations that we found are mostly point substitutions but there are also some multiple mutations. We have characterized the novel mutations according to the individual and multiple occurrences and order of location of the mutations.

The substitution $C^{42}U$ found in sample P9-B10 was not reported in the HCVIVdb with respect to its impact on the HCV-IRES translation response (Table 17). It was shown to reduce the efficiency to some level, 66%, relative to the pRG-IRES control vector translation efficiency. The decrease in HCV-IRES response can only account for the reason that the introduction of this transition impacts the structural integrity of the adjacent regions important for making precise contacts with the 40S, since this nucleotide does not make a direct contact with either the 40S or eIF3. Functional data is available only on the effect of collective mutations at the adjacent base-paired regions displaying the reduced activity (Zhao and Wimmer,

2001b, Honda et al., 1999b), also described in mutant 5 (Table 13). The mutation $U^{101}C$, sample P9-E11, is located in domain IIa stem that connects to the loop E-motif. The probable impact caused by this mutation on translation and the significance of the region where this nucleotide occurs has been explained in mutant 3 sample (P4-B3). It has also been shown through mutation analysis that substituting the individual strand (nt. 65-69 and nt. 98-102) abrogates the HCV-IRES activity which is restored only to some minor level upon switching the both strands (Honda et al., 1999b). Samples from the Patient 4 (P4-A12, P4-G6), $U^{113}C$ and C^{115} deletion occurs in close proximity at the basal portion of domain II, forming canonical base-pairing with nucleotides A^{50} and G^{47}, respectively, observed through NMR (Lukavsky et al., 2003). The substitution of $U^{113}C$ has shown not to affect the HCV-IRES efficiency. However, a mutation $A^{50}U$ reported in literature ends up with a response of only around 49%, relative to the HCV-IRES WT (Tang et al., 1999). The decrease in activity might be due to the switching of purine-pyrimidine identity to pyrimidine-pyrimidine, whereas in sample P4-A12 this orientation is maintained showing a near pRG-IRES control vector translation activity. The sample P4-G6 also exhibits lower response (Table 17) which may suggest that for structural integrity of this region conservation of primary sequence is significant.

The nucleotide variation $C^{158}A$, in stem region of domain IIIa (Figure 2) found in sample P7-8 has shown a near pRG-IRES control vector translation efficiency activity. The domain IIIa upon substitution of its loop results in reduced binding affinity with the eIF3 and consequently low translational response (Kieft et al., 2001, Ji et al., 2004, Quade et al., 2015). There has not been many point substitutions that were reported in this region. A single substitution at loop region, $A^{164}G$, and stem region, $C^{168}G$, results in 60% and 80% translation activity, relative to the WT, respectively (Shi et al., 2011, van Leeuwen et al., 2004). The switching of base-pairing at position C^{157}-G^{170} to G^{157}-C^{170} results in no activity of the HCV-IRES whereas reversing the base-pair at adjacent nucleotide positions 171-156 restores translation efficiency. It has been proposed that it might be due to the contacts made by the former base-pair with eIF3 and the 40S (Kieft et al., 2002).

Samples #	Novel mutations	Translation activity (%)
P9-B10	**C 42 U**	66
P9-E11	**U 101 C**	85
P4-A12	**U 113 C**	87
P4-G6	**C del. 115**	59
P7-8	**G 158 A**	92
P4-E9	**U 191 C**	92
P9-A12	**A 246 G**	93.80
P9-B1	**U 251 C**	91
P4-E1	**U 284 C**	71
P7-1	**U 302 C**	104
P4-F5	**C 317 U**	86
P4-F2	**G 320 A**	88
P9-E9	**C 334 U**	95
P4-A8	**C 341 U**	100
P4-G12	**U 220 C + G 303 A**	69
P9-C1	**A 244 G + U 353 G**	92
P7-21	**G 241 A + U 353 C**	84
P9-G7	A 260 G + **A 345 G**	42.60
P7-54	C 83 U + A 96 G + **C 254 U**	54

Table 17: Novel HCV-IRES mutations, in bold, that we identified in our patients' samples and their measured translational activities are given. The mutations are characterized according to the order of occurrence in HCV-IRES domains.

The activity displayed by sample P4-E9, $U^{191}C$, is also close to pRG-IRES control vector translation response (Table 17). The mutation $U^{191}C$ occurs in the apical stem loop IIIb which makes direct contact with the eIF3 as shown through mutation, biochemical and structural data (Kieft et al., 2001, Buratti et al., 1997, Odreman-Macchioli et al., 2001, Siridechadilok et al., 2005, Ji et al., 2004, Quade et al., 2015). The overall activity of this apical loop of domain IIIb upon introduction of mutations does not seem to impact the translation (Barria et al., 2009, Motazakker et al., 2007, Forton et al., 2004). However, altering the whole loop region has a devastating impact on the HCV-IRES mediated translation initiation (Buratti et al., 1997, Kieft et al., 2001) which suggests that maintenance of structure and primary sequence to a certain degree is significant for the translational productivity. Similar to other activities reported for this region, the sample P4-E9 ($U^{191}C$) has an equivalent of a fully-functional HCV-IRES.

The description for samples P9-A12 and P9-B1 (Table 17) and the possible impact these mutants have on the surrounding regions and on translation have been discussed in long-range inter-domain mutant 4 sample (P7-24) and mutant 6.

The region connecting domain IIId with IIIe is significant since bulged out nucleotides of this region and domain IIIe, A^{136} and U^{297}, has been shown through structural and mutation data to make tertiary interactions with respective nucleotides (Berry et al., 2011, Easton et al., 2009, Angulo et al., 2016). There have been nucleotides in this stem region that were protected from iodine cleavage upon the 40S binding (Kieft et al., 2001). Employing modification interference data has showed increase flexibility of nucleotide A^{140} with the 40S attachment (Filbin and Kieft, 2011). The substitution in sample P4-E1, $U^{284}C$ (Table 17), changes the predicted base-pairing with A^{140} that probably causes a marginal decrease in translation efficiency.

The variation, $U^{302}C$, found in sample (P7-1) occurrs in domain IIIe displays an activity equivalent to the pRG-IRES control vector translation efficiency (Table 17). However, domain IIIe is susceptible to the nucleotide changes observed in reported mutants exhibiting severe loss of HCV-IRES function (Psaridi et al., 1999, Lukavsky et al., 2000, Easton et al., 2009, Kieft et al., 2001, Berry et al., 2011). The mutation $U^{302}C$ modifies G^{291}-U^{302} into a canonical G-C base pair (Figure 2) that may sustain the structural integrity of this domain and in turn have no effect on translation response. Another reported mutation in the literature adjacent to this nucleotide position, $G^{301}C$, stressed on the functional importance of this nucleotide conservation which upon change resulted only in 10% translation response relative to the WT (Laporte et al., 2000). This probably is due to the changes in canonical base-pairing in this region that may disrupt the stability and integrity of structure since domain IIIe is highly conserved in different genotypes and subtypes.

The translation efficiency of fully-working HCV-IRES was depicted in samples P4-F5 ($C^{317}U$) and P4-F2 ($G^{320}A$) (Table 17). These nucleotides are located in a region that makes part of the pseudoknot stem and through SHAPE probing data has displayed less modification upon the 40S binding (Filbin and Kieft, 2011). Toe-prints have also been located in positions G^{318} and G^{320} (Pestova et al., 1998). Mutation analysis has displayed that most of the mutations that occur in this region do not

impact the HCV-IRES responses to any significant level. However, altering two consecutive purine-pyrimidine base-pairs decreased the efficiency which was restored to some level upon complementary mutations. Similarly, deletion of two base-pairs also resulted in reduction of activity (Berry et al., 2011). Substituting one pseudoknot strand (nt. 126-129) or the other (nt. 319-322) also causes nearly no activity. Although, introduction of complementary substitutions restored the activity to the wild type level (Wang et al., 1994a).

Novel point substitutions from samples P9-E9 ($C^{334}U$) and P4-A8 ($C^{341}U$) that we identified in domain IV maintained a response that is almost equal to the pRG-IRES control vector translation efficiency (Table 12). Most of the single mutations that have been reported in domain IV have shown little or no changes to activity.

The double mutations, $U^{220}C$ and $G^{303}A$, found in sample P4-G12 displayed average translation efficiency 69% compared to the pRG-IRES control vector translation response. Mutation, $U^{220}C$, is adjacent to the internal loop of domain IIIb. The nucleotide U^{220} is protected from RNase V1 cleavage upon the eIF3 interaction and was also identified by modification interference to eIF3 (Kieft et al., 2001, Sizova et al., 1998). There were not many reported nucleotide changes in this stem region of domain IIIb. One nucleotide variation, $A^{223}G$, resulted in a moderate response (Barria et al., 2009). Multiple mutants that consist of a mutation in this region have also been reported which have shown lower response (Barria et al., 2009, Laporte et al., 2000). Considering the functional activity of additional mutations in these multiple mutants it can be suggested that the mutations in region of the domain IIIb may be the cause of lower efficiency and conservation of the sequence and structure of this region is significant.

The additional substitution, $G^{303}A$, occurs in a pseudoknot region that connects with the domain IIIe. This nucleotide forms a predicted Watson-Crick base-pair G^{303}-C^{313} which upon deletion has a reduce translation performance. Similarly, the deletion of the two base-pairs of this region and introduction of two base-pair insertions abrogates the HCV-IRES activity (Berry et al., 2010). No functional data in terms of point substitutions have been located in this region. The changes in base-pairing from G-C to A-C in the case of $G^{303}A$ can be disruptive to the structure. The

nucleotide G^{291} in domain IIIe, through X ray crystallography, has shown to stack on G^{303} providing stability to this region and altering the primary sequence can be devastating (Berry et al., 2011).

The multiple mutants from samples P9-C1 and P7-21 are in the same region and also display an activity illustration of a functional HCV-IRES (Table 17). The explanation of the regions of these mutations' occurrence have been described in detail in above section of long-range inter-domain interaction in mutant samples P9-B11, P4-C52, P4-B3, mutant 6 and mutant 8. Likewise, samples P9-G7 and P7-54 with reduced activities were identified with novel mutations. The regions where these mutations are present have been characterized and illustrated in the above section.

5.3.5 Characterization of HCV-IRES regions based on collective individual and multiple mutations and their average impact on translation response

As specified earlier, the HCVIVdb comprises HCV-IRES variations that were collected from published studies over the course of time. The database constitutes the mutations that have been reported in patients suffering from the HCV and have either gone through standard treatment of interferon and ribavirin or remained non-treated. In addition, the mutations that were introduced artificially into the HCV-IRES, essentially for the purpose of studying the behavior and functional outcome of the viral protein synthesis is also a unit of the HCVIVdb. More than 1,800 mutations were collated by going through majority of the studies that dealt with the HCV-IRES variability and were further assembled in various categories. In an attempt to further understand and simplify the functional capacity of various HCV-IRES regions, we characterized reported individual mutations, the impact of which on HCV-IRES translation efficiency was measured. Similarly, we also assembled multiple mutations located in different domains that were reported and subsequently measured for their effect on translational response. These mutations were further organized according to the region of occurrence summarizing all other identical mutations and mutations present in the same region. It can facilitate in comprehensive and overall comparative evaluation of function of a particular entry which then also display entries of similar position from the same or multiple sources.

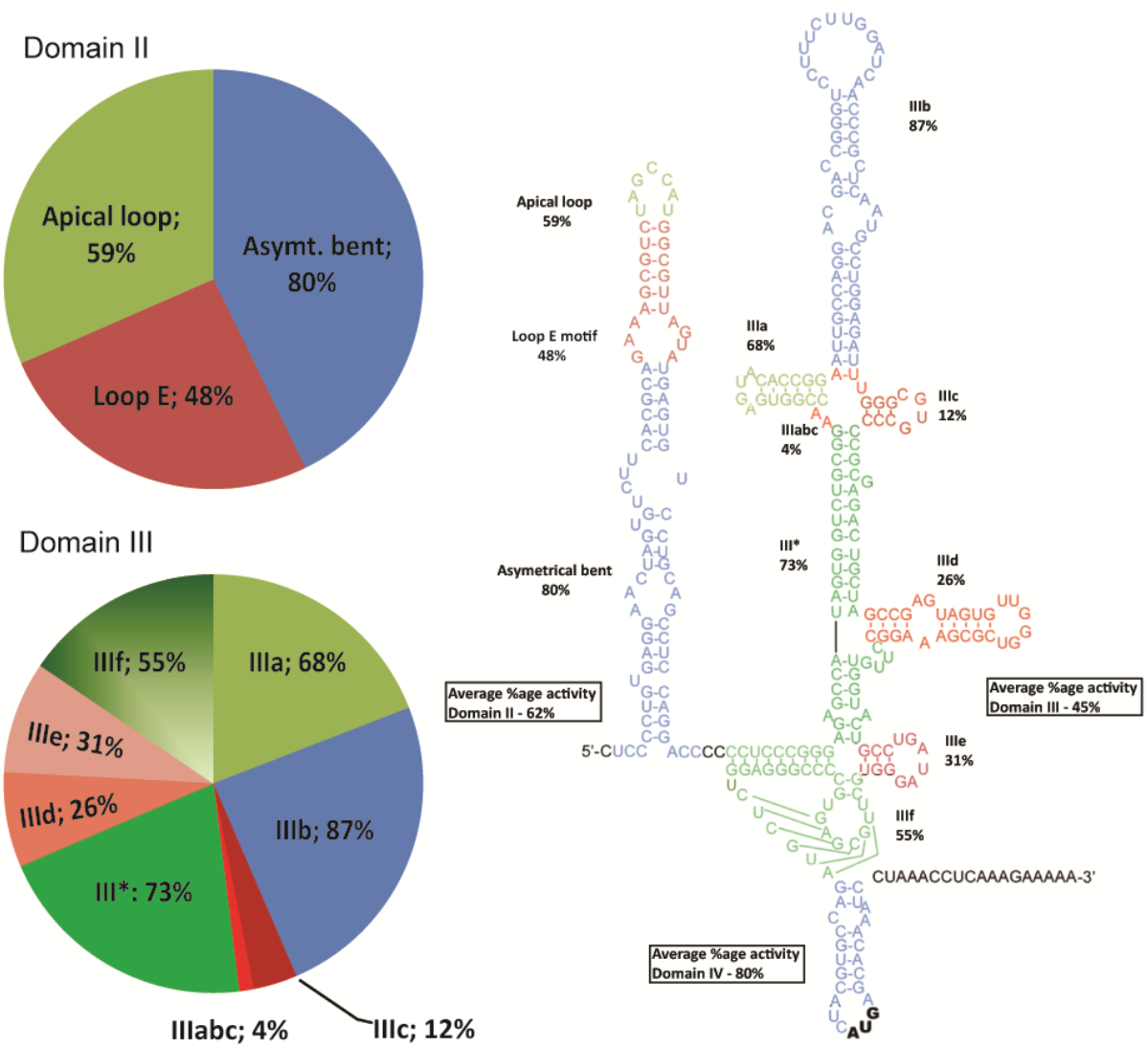

Figure 14: The point substitutions from all the HCV-IRES domains were collected and cumulative average translation activity was calculated for specific and overall domain(s) and sub-domain(s). The HCV-IRES structural elements are identified that are least and most conserved to individual nucleotide variations. The color of the HCV-IRES domains on secondary structure corresponds to color of the pie-chart.

The characterized single mutants with reported translation activity present in a particular domain were estimated for their collective translation efficiency. It was carried out by adding up the respective influence of all the individual mutations on viral protein synthesis in a specific HCV-IRES region and then calculating their collective/overall impact, on average, on translation activity. For example, domain II was categorized into regions comprising asymmetrical bent, loop E-motif and the apical loop. The collective average impact on translation efficiency of all the individual mutations found in asymmetrical bent was shown to be 80%, relative to the HCV-IRES WT (Figure 14). Likewise, for loop E-motif and apical loop this activity was

found to be 48% and 59%, respectively since these loops are involved in interaction with the 40S subunit and stops the assembly of 80S complex upon mutations, also described in detail in the above section. In the same way we analyzed the collective average effect of combined mutations on translation response in specific domains and sub-domains of the HCV-IRES. We found the same pattern for domain II where loop E-motif and apical loop showed the least of collective average productivity with regard to translation. Similar results were found for collective average impact on translation for individual and combined mutations on sub-domains of the major

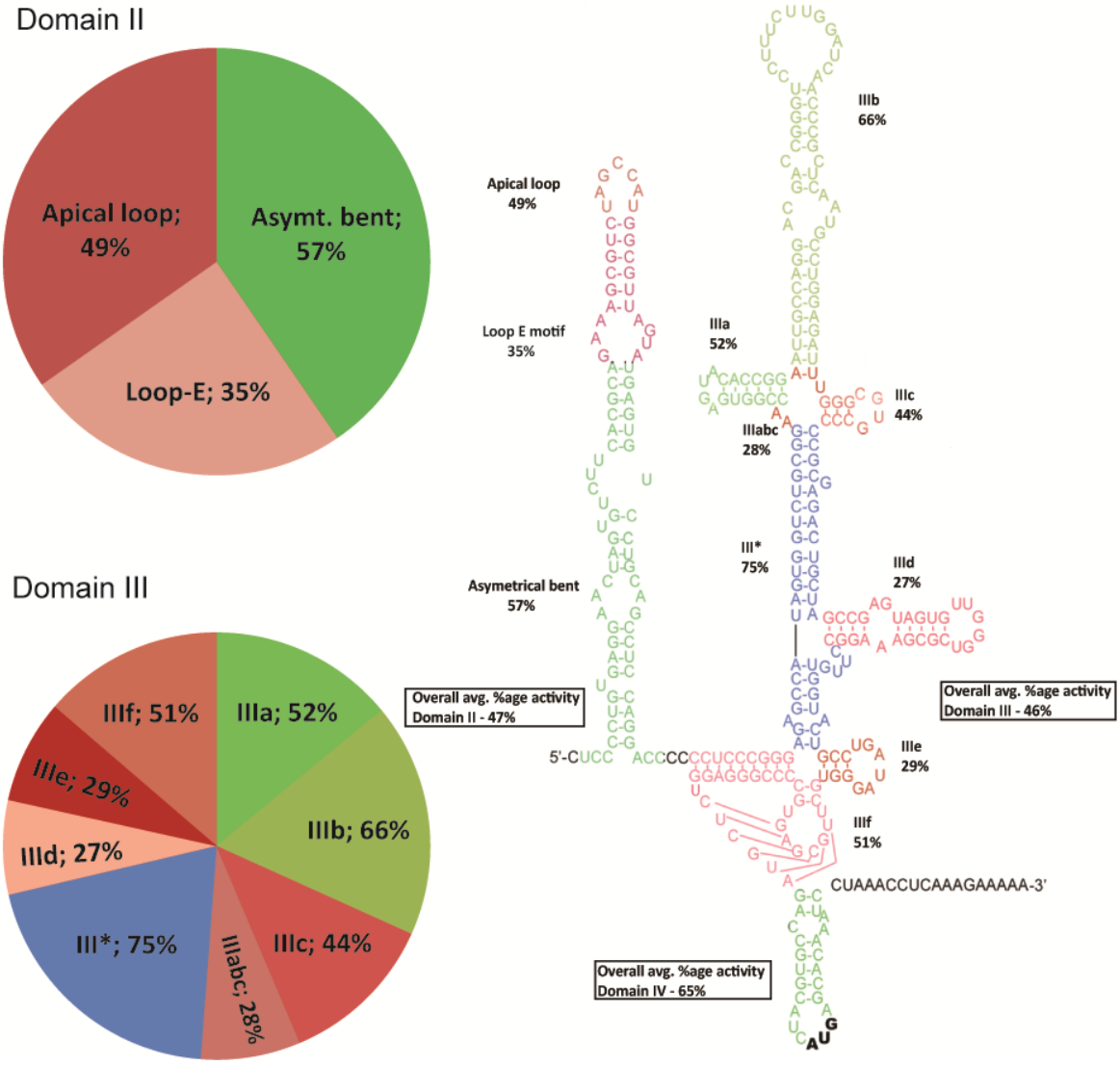

Figure 15: The multiple substitutions from all the HCV-IRES domains were collected and cumulative average translation activity was calculated for specific and overall domain(s) and sub-domain(s). The HCV-IRES structural elements are identified that are least and most conserved to multiple nucleotide variations. The color of the HCV-IRES domains on secondary structure corresponds to color of the pie-chart.

domain III. Such as, the analysis in the both the groups with collective average impact on translation showed least activities in domain IIId, IIIe and junction IIIabc of HCV-IRES (Figure 15). There are some differences with regards to collective average activities of translation in junction IIIabc and domain IIIc of both the groups. This is due to the fact that in combined mutations the complementary mutations in these two domains, junction IIIabc and IIIc, restores the translation efficiency to some level. The region that was least effected from the introduction of mutations was domain IIIb and the stem III region in both the groups. The analyses of both the groups are in consensus, illustrating that different domains and sub-domains of HCV-IRES display essentially an equivalent effect on translation initiation whilst subjected to collective individual and combined mutations.

5.3.6 Link of mutations' dispersal in HCV-IRES domains from treated and non-treated patients and its relevance to treatment outcome

Over the past few years' advances have been made in the development of HCV model system that has allowed us to study the virus-host interactions and its influence on the clinical output. The treatment of HCV thus, has made significant progress with the introduction of various classes of direct-acting antivirals (DAA)-based therapy which improved sustained viral response (SVR) rate in treated patients (Gane et al., 2013, Gentile et al., 2014, Gallay and Lin, 2013). However, the resistance to DAAs has been observed depending on the genotype/subtype and the fitness of the resistant viral populations (Pawlotsky, 2011, Roche et al., 2015, Ferenci et al., 2015, Pawlotsky, 2014).

We also characterized the published patients' data according to the sustained and non-sustained responders to the standard combination therapy of interferon (IFN) and ribavirin. The data was based on the treatment of patients and samples were sequenced before or after the therapy response. The objective was to look for some specific sequences or regions that might be involved in inducing a sustained or non-sustained response in patients by analyzing the frequency and location of variations in both the groups. We found eight studies that dealt with the sustained and/or non-sustained response to IFN and riabvirin therapy with variation of HCV-

IRES sequence reported (El Awady et al., 2009, Thelu et al., 2004, Thelu et al., 2007, Yamamoto et al., 1997, Yasmeen et al., 2006, Ogata et al., 2008, Soler et al., 2002, Lu et al., 1999, Saiz et al., 1999). The numbers of patients from all these studies which were found to be 110, with 61 patients showing sustained response and the other 49 was non-responsive. We mapped all the published variations from these studies on HCV-IRES secondary structure in order to understand the possible correlation that may exist between the occurrence of mutations in different HCV-IRES domains and the patient's response to therapy. As it is evident (Figure 16), there is no obvious pattern and (or) association of mutations' occurrences in HCV-IRES domains and sub-domains to therapeutic responsiveness. The presence of mutations has found to be random in both the populations of patients and no direct link could be observed / developed for the location of variations in the HCV-IRES and their possible effectiveness in inducing a sustained or non-sustained response.

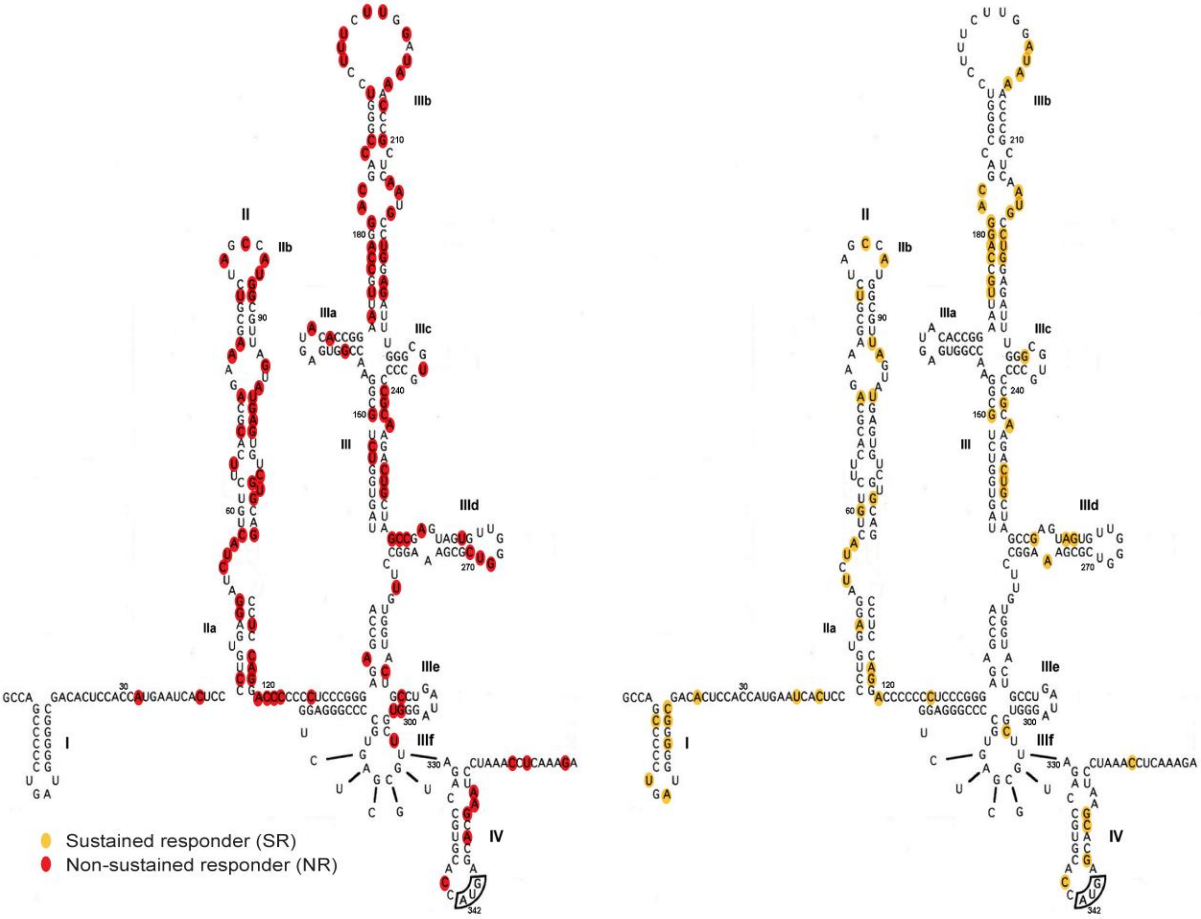

Figure 16: Mapping of the nucleotide variations on HCV-IRES secondary structure from sustained (SR) and non-sustained responders (NR) of the HCV patients.

5.4 Taq polymerase errors and HCV quasispecies

In studies where the functional evaluation in the context to measuring the affect of mutation(s) on a specified gene product is required, the reduction of PCR-generated errors is particularly crucial. There have been studies carried out in order to determine the fidelity of different DNA polymerases. One such study analyzed the error-rates of high and low fidelity DNA polymerases by direct sequencing of cloned PCR products. As compared to the Taq, polymerases such as *Pfu*, *Pwo* and Phusion displayed very low error-rates. The mutations found among these high fidelity enzymes were shown to be comparable (McInerney et al., 2014).

The quasispecies diversity in HCV has been questioned with reference to the artifactual quasispecies that result from induction of the sequence errors by polymerase during amplification or as a genuine biological phenomenon that exists and impacts the HCV evolution and RNA heterogeneity in order to allow the virus to circumvent the immune system and the therapeutic pressures.

The difference of quasispespcies heterogeneity was reported with Taq polymerase and *Pfu* polymerase generated samples of vesicular stomatitis virus (VSV). The Taq-generated substitution errors compared to *Pfu* polymerase were found to be significant with Taq-based analysis can lead to false estimation of nucleotide diversity in quasispecies (Bracho et al., 1998). The HCV hypervariable region of the envelope gene (HVR1-E2) along with interferon sensitivity determining region (ISDR) of NS5a gene was amplified with Taq or *Pwo* DNA polymerase with identical samples. The HCV quasispecies evolution was shown to be overestimated when Taq-treated samples were assessed in comparison to the proof-reading polymerase *Pwo* where use of the Taq polymerase resulted in the rise of minor quasispecies variants (Mullan et al., 2001)**.** However, the use of Taq polymerase was shown not to majorly influence the diversity of HCV quasispecies quantified by heteroduplex mobility assays in cohort-based studies although working with proof-reading polymerase provided more definite analysis of the HCV quasispecies (Polyak et al., 2005)**.** In another study, to estimate the relevance of nucleotides' misincorporation by different DNA polymerases whilst analyzing the genetic viral heterogeneity was also carried out. The amplification of a homogenous cDNA

template containing a 5' UTR of HCV by PCR with Taq polymerase displayed seven times increased mutation rates compared to proof-reading polymerase. Hence, the importance of DNA polymerase with proof-reading activity was emphasized for precise analysis of the natural genetic variability that may present in a population (Malet et al., 2003).

In our study the samples that we analyzed were obtain with Taq DNA polymerase. Most of the studies suggest the use of proof-reading polymerases instead of Taq in order to have an accurate assessment of nucleotide diversity while measuring the quasispecies heterogeneity and genetic diversity. Our approach was more focused on the evaluation of the functional aspect of the acquired substitutions in the HCV-IRES domains. The diversity that was found in different regions of the HCV-IRES was measured concerning the affect that variation(s) can have on protein synthesis of the virus genome. Even if some of the mutations that were functionally analyzed were Taq-induced errors, the impact derived from the estimation of those mutants gives us an insight into the behavior of viral translation irrespective of the nucleotide changes that may or may not be real.

Genetic diversity in HCV is acquired through the presence of an error-prone RNA dependent RNA polymerase (RdRp) that give rise to a pool of genetically distinct but closely related variants termed quasispecies. The absence of proof-reading mechanism results in a mutation rate that is estimated to be 2.5×10^{-5} per nucleotide per genome replication contributing to the viral diversification (Ribeiro et al., 2012). The frequency of mutations in naturally occurring mutants was also analyzed from HCVIVdb. The raw counts from the naturally occurring mutations were ~1400 and displayed a mutation bias towards transitions as compared to the transversions. The transitions accounted for 56% relative to transversions which were 44%, respectively suggesting HCV-IRES mutates more frequently through transitions. The raw counts were further adjusted and normalized to the overall nucleotide frequencies in the HCV-IRES which fetched similar results showing bias for A to G and U to C mutations and vice versa. The rate of transitional changes have been observed occurring more frequently in general in most of the organisms over transversional changes and various methods have been proposed in order to make reliable estimates about the transitional bias (Wakeley, 1996, Kristina Strandberg and

Salter, 2004). The biochemical data correlates with the deep sequencing analysis of NS3 region of HCV-RNA collected from 20 treatment naïve HCV patients displaying a pronounced preference for transitions over transversions, further arguing that nature of the variations may therefore contribute to the genetic barrier in the development of resistance against direct-acting antivirals (DAAs) (Powdrill et al., 2011). Similar evidence demonstrating the mutation bias in favor of transitions rather than transversions was also reported while observing the evolutionary dynamics and genetic variations in HCV replicon system (Kato et al., 2005). Interestingly, we observed similar frequency of naturally occurring mutation patterns in our study, as reported in (Kato et al., 2005), with most observed transition variations were A-to-G and U-to-C respectively, followed by G-to-A and C-to-U mutations. The least common substitution was found to be G-to-U (Figure 17).

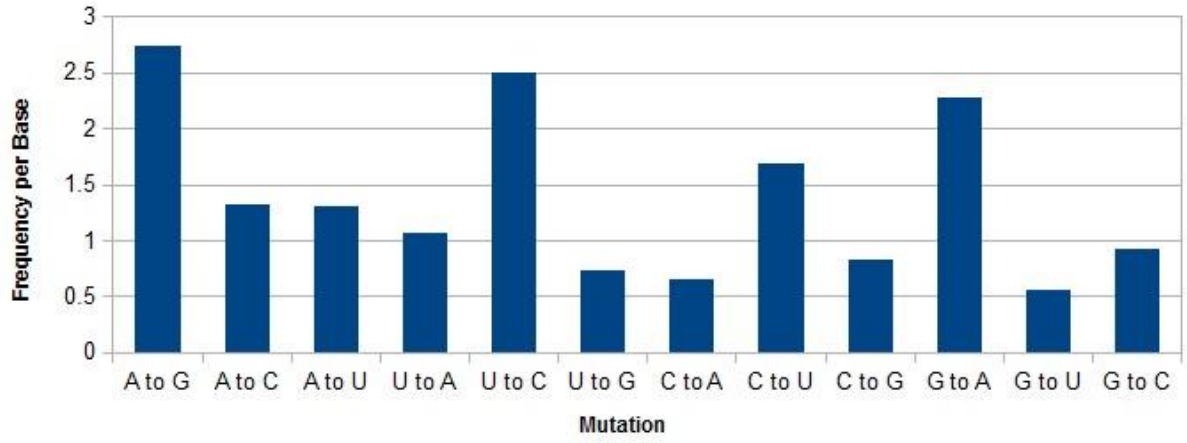

Figure 17: Frequency of naturally occurring HCV-IRES mutations observed in the HCVIVdb. A clear mutation bias for transitions over transversions is evident.

5.5 Validity of HCVIVdb

In an attempt to further understand the functioning of HCV-IRES on nucleotide level, the complete genome sequences of 2006 HCV entries were aligned using the multiple sequence alignment tool MUSCLE (Edgar, 2004, Edgar and Batzoglou, 2006). This was carried out by a colleague Evan Floden that worked on the HCVIVdb in our laboratory. Sequence data was obtained from the NIAID Virus Pathogen

Database and Analysis Resource (ViPR) online through the web site at http://www.viprbrc.org (Pickett et al., 2012b).

The multiple sequence alignment (MSA) was analyzed with the nucleotide counts at each position within the HCV-IRES. Sorting of positions were based on the percentage of nucleotides which are different than the consensus nucleotide (Table 18 and Supplementary file 1). Similarly, the nucleotide changes at each particular position of HCV-IRES was carried out for HCVIVdb and characterized from the most hypervariable of the nucleotides to the most conserved (Table 19 and Supplementary file 1). Further, the nucleotide counts from MSA were compared to the mutations observed within the HCVIVdb at each position to examine for the variability rank of nucleotides between the two databases (Table 20). The dataset from the two databases when plotted for the variance from the consensus sequence showed a positive correlation (Figure 18 and Supplementary file 1).

The nucleotides count from both the databases showed the hypervariability of some particular nucleotides. The first two nucleotides (204 and 243) (Table 20), occur in apical loop of domain IIIb and stem III (Figure 2). The nucleotide 204 has shown to be protected from RNase I cleavage upon the eIF3 attachment (Sizova et al., 1998). With such hypervariability that is observed in nucleotide at position 204, adenine or cytosine occurred most frequent followed by uracil. Surprisingly, in both the databases the presence of a guanine nucleotide at position 204 was negligible (Tables 18, 19). It may suggest the evolutionary preservation or adaptation of nucleotide at position 204 which upon the presence of nucleotides (A, C and U) does not impact the viral protein synthesis. The presence of G may effect the structural stability of the loop IIIb, since it interacts with the eIF3, which can lead to low translational response. Although, we do not have any functional data to support this hypothesis, the near total absence of G hints towards the adaptation to other nucleotides (A, C and U) at nucleotide position 204 over the course of evolution.

In like manner, the nucleotide at position 243 only seem to accommodate A and G bases as observed in HCVIVdb and other database. With all the variations and raw counts at position 243 the presence of C or U occurred only twice and once,

Raw Counts							%				
Sr. #	**Nucleotide position**	**A**	**C**	**G**	**U**	**Sum**	**Nucleotide position**	**A**	**C**	**G**	**U**
1	**204**	905	948	2	123	1978	**204**	45.8	47.9	0.1	6.2
2	**243**	1085	0	919	1	2005	**243**	54.1	0.0	45.8	0.0
3	**183**	465	1359	3	175	2002	**183**	23.2	67.9	0.1	8.7
4	**119**	1367	517	5	64	1953	**119**	70.0	26.5	0.3	3.3
5	**217**	51	0	1949	0	2000	**217**	2.6	0.0	97.5	0.0
6	**340**	89	1544	2	367	2002	**340**	4.4	77.1	0.1	18.3
7	**107**	611	2	1330	0	1943	**107**	31.4	0.1	68.5	0.0
8	**182**	1948	0	50	4	2002	**182**	97.3	0.0	2.5	0.2
9	**214**	1512	9	4	475	2000	**214**	75.6	0.5	0.2	23.8
10	**185**	1996	0	1	5	2002	**185**	99.7	0.0	0.0	0.2
11	**203**	37	30	23	1909	1999	**203**	1.9	1.5	1.2	95.5
12	**212**	1	2	0	1997	2000	**212**	0.1	0.1	0.0	99.9
13	**248**	1	108	5	1888	2002	**248**	0.0	5.4	0.2	94.3
14	**266**	0	1	2004	0	2005	**266**	0.0	0.0	100.0	0.0
15	**297**	1	0	0	2005	2006	**297**	0.0	0.0	0.0	100.0
16	**104**	0	1818	3	86	1907	**104**	0.0	95.3	0.2	4.5
17	**179**	1415	0	587	0	2002	**179**	70.7	0.0	29.3	0.0
18	**224**	118	464	1419	4	2005	**224**	5.9	23.1	70.8	0.2
19	**268**	1	0	2003	1	2005	**268**	0.0	0.0	99.9	0.0
20	**325**	0	2005	0	1	2006	**325**	0.0	100.0	0.0	0.0
21	**288**	2005	1	0	0	2006	**288**	100.	0.0	0.0	0.0
22	**269**	1	6	1	1997	2005	**269**	0.0	0.3	0.0	99.6
23	**350**	729	4	1269	2	2004	**350**	36.4	0.2	63.3	0.1
24	**78**	2	309	11	1314	1636	**78**	0.1	18.9	0.7	80.3
25	**97**	0	69	1	1823	1893	**97**	0.0	3.6	0.1	96.3

Table 18: The nucleotide counts from the multiple sequence alignment of 2006 HCV entries. Raw counts of nucleotide variability at each position with estimated percentages alongside shows the degree of hypervariability and conservation that may exist within each nucleotide of HCV-IRES. The table here shows the first 25 entries of the nucleotide counts. For other entries see the Supplementary file 1.

Position	A→G	A→C	A→U	U→A	U→C	U→G	C→A	C→U	C→G	G→A	G→U	G→C	Sum
204	0	21	0	2	0	0	26	15	1	0	0	0	**65**
243	12	1	0	0	0	0	0	0	0	26	0	1	**40**
183	0	0	0	0	0	0	7	21	4	0	0	0	**32**
119	2	23	2	0	1	0	0	0	0	0	0	0	**28**
217	0	0	0	0	0	0	0	1	0	14	5	6	**26**
340	0	1	0	0	4	0	4	15	0	0	0	0	**24**
107	0	0	0	0	0	0	0	0	0	19	0	0	**19**
182	11	4	4	0	0	0	0	0	0	0	0	0	**19**
214	5	3	11	0	0	0	0	0	0	0	0	0	**19**
185	5	4	7	0	0	0	0	0	0	0	0	0	**16**
203	0	0	1	5	1	5	0	0	0	4	0	0	**16**
212	0	0	0	4	8	4	0	0	0	0	0	0	**16**
248	0	0	0	0	14	2	0	0	0	0	0	0	**16**
266	0	0	0	0	0	0	0	0	0	11	1	4	**16**
297	1	0	0	4	8	2	0	0	0	1	0	0	**16**
104	0	0	0	1	1	1	0	11	1	0	0	0	**15**
179	13	1	0	0	0	0	0	0	0	1	0	0	**15**
224	0	0	0	0	1	0	0	0	0	11	0	3	**15**
268	0	0	0	0	0	0	0	0	0	1	10	4	**15**
325	0	0	0	0	0	0	0	7	8	0	0	0	**15**
288	7	3	4	0	0	0	0	0	0	0	0	0	**14**
269	0	0	0	6	4	2	0	0	0	0	0	1	**13**
350	0	0	0	0	0	0	0	0	0	13	0	0	**13**
78	0	0	0	0	10	1	0	1	0	0	0	0	**12**
97	0	0	0	1	9	0	1	1	0	0	0	0	**12**

Table 19: The nucleotide counts of the HCV-IRES from HCVIVdb. The collected nucleotide changes from the literature are being characterized from high mutation occurrence to the low mutation count at each position that displays the hypervariability and conservation of nucleotides in the HCV-IRES. The table here shows the first 25 entries of the nucleotide counts. For other entries see the Supplementary file 1.

respectively which is negligible. The existence of only purines at this location might have evolutionary advantage in preservation of the wild type efficiency. Considering strictness of the structure with the presence of up to four GC pairs adjacent to nucleotide at position 243 accounts for the stability of this region and probably maintain its conformation which may disrupt upon the introduction of a pyriminde. Since nucleotide at position 243 is in close vicinity to the adjacent base-pair downstream, U^{149}-A^{244}, it can make a wobble with G^{243}-U^{149} and further stabilizes the

region. The region is also closer to junction IIIabc that has shown to interact with both the 40S and eIF3 (Kieft et al., 1999, Kieft et al., 2002, Quade et al., 2015, Yamamoto et al., 2015).

The nucleotide at position 183 showed a variability of 32% from the consensus sequence in MSA of 2006 sequences (Table 18). There is only a 0.1% presence of the nucleotide G which shows that the adaptation to guanine base can have consequences to the structural conformation of the region and function. The reported functional data of the nucleotide at position 183 also showed to have lowest translational response upon substitution to G, compared to A or U, which is 63% relative to the WT (Collier et al., 2002).

Position	Max. Nt. Count%	Seq. Variability Rank	HCVIVdb Variability Rank
204	48	1	1
243	54	2	2
350	63	3	23
183	68	4=	3
107	68	4=	7
119	70	6	4
179	71	7=	17
224	71	7=	18
220	71	7=	29
262	75	10=	88
270	75	10=	120
214	76	12=	9
187	76	12=	108
210	76	12=	111
223	76	12=	30
340	77	16	6
78	80	17	24
352	89	18	75
351	90	19	221
362	91	20=	-

Table 20: Comparison of the HCV-IRES variability rank at each nucleotide position between multiple sequence alignment and the mutation counts in HCVIVdb.

Likewise, in HCVIVdb the variations found in this region has been quite a few. The nucleotide C^{183} forms a base-pair with A^{214} in the internal loop of domain IIIb. The nucleotide A^{214} has also found to be quite variable among the ranks of other HCV-IRES nucleotides, confirmed through HCVIVdb and MSA (Table 20). The highest deviation from consensus sequence, A to U 23.8%, has displayed not much impact on translational activity (Collier et al., 2002). Similarly, the least adapted nucleotide according to MSA and HCVIVdb, guanine, at position 214 also has the most influence in reducing the efficiency of HCV-IRES, 73% relative to the HCV-IRES WT activity, compared with other nucleotide changes (Collier et al., 2002). The nucleotide at position 214 has also been showed to be protected from RNase I cleavage and modification interference upon the eIF3 attachment (Sizova et al., 1998, Kieft et al., 2001). Another nucleotide at position 217 that occurs in this domain IIIb internal loop has shown to be conserved and majority of substitutions are observed to be transitions from G to A. The activity reported on substituting G^{214} to A is equivalent to the WT whereas, variation to other nucleotides shows lower responses (Collier et al., 2002, Wang et al., 1994a). This may suggest the conservation of purines, majority of which is guanine, at position 214.

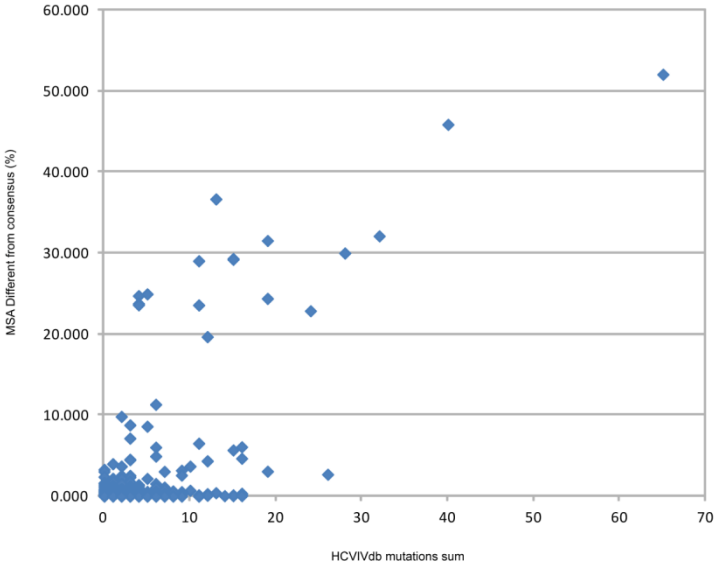

Figure 18: The correlation of the variability count of HCV-IRES nucleotides datasets between HCVIVdb and multiple sequence alignment (MSA).

The nucleotide at position 119 upon counting of the nucleotides, from MSA of 2006 sequences, displayed a 30% of difference from the consensus sequence making it among the most variable of the nucleotides in HCV-IRES. The highest variation observed was a transversion A to C in both the databases (Tables 18 and 19). There is only 0.3% of G detected out of the sum 1953 nucleotide counts at

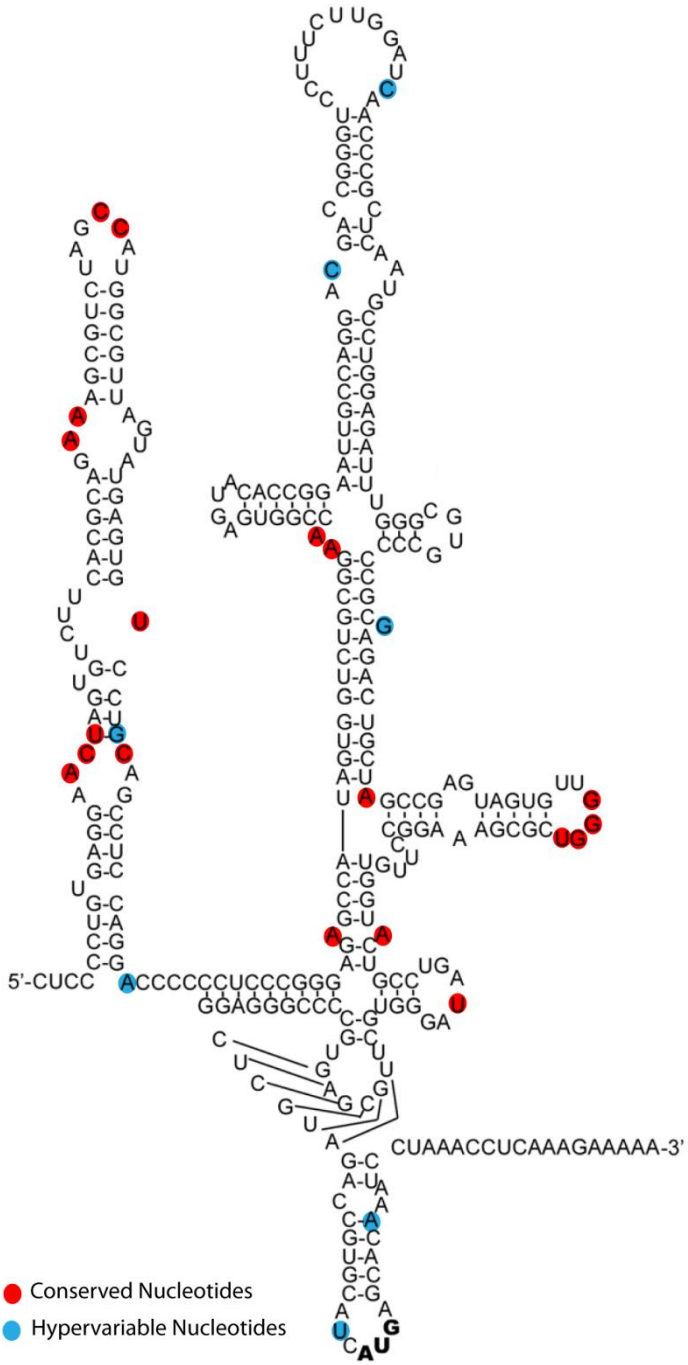

Figure 19: Mapping some of the hypervariable and conserved nucleotides on the HCV-IRES secondary structure (domain II-IV). Hypervariable nucleotides are colored blue whereas conserved nucleotides are colored red.

position 119 which may suggest that base G is evolutionary not preferred at this location. The translation activity of HCV-IRES has hardly displayed any alteration, relative to HCV-IRES WT, upon substitution of A^{119} to C (Barria et al., 2009, Motazakker et al., 2007, Zhang et al., 1999).

The preservation of nucleotide at position 107 has been abandoned by adenine variation that showed to have occurred 26.5% in MSA and also in HCVIVdb. The occurrence of other bases at this position has been negligible. The presence of A instead of G has shown to retain the translational activity similar to the WT (Motazakker et al., 2007). It forms a wobble base-pair with U^{61}, observed through NMR (Lukavsky et al., 2003), and alteration of G^{107} to A may help to maintain the bonding interaction to Watson-Crick base-pair. The nucleotide U^{61} seem to be highly conserved since no reported mutation at this position has been located in HCVIVdb and MSA shows only 0.65% of a difference from the consensus sequence (Tables 18 and 19).

Hypervariable nucleotides	**Translational activity %**	**References**
A 204 C	~100	(Motazakker et al., 2007)
C 204 U	109 ± 9	(Barria et al., 2009)
C 204 U	~ 80	(Forton et al., 2004)
U 204 A	Retains full act.	(Kieft et al., 1999)
A 243 G	92	(Motazakker et al., 2007)
G 203 A + G 243 A	100	(Motazakker et al., 2007)
C 204 A + G 243 A	83	(Forton et al., 2004)
Conserved nucleotides	**Translational activity %**	**References**
G 266 A	5	(Laporte et al., 2000)
G 266 C	3	(Kieft et al., 1999)
G 267 C	>95 reduction in activity	(Jubin et al., 2000)
G 267 C	2	(Kieft et al., 1999)
G 268 C	>95 reduction in activity	(Jubin et al., 2000)
G 268 C	3	(Kieft et al., 1999)

Table 21: The translational efficiency of selected HCV-IRESs containing variations of some hypervariable and/or conserved nucleotides as revealed from the HCVIVdb data analysis.

The most conserved HCV-IRES nucleotides that have been observed while analyzing the MSA of 2006 sequences, showing more than 90% preservation of sequence, are the nucleotides that binds to the translational machinery or are important in maintaining the structural conformation of HCV-IRES domains for proper secondary and/or tertiary interaction(s) in order to sustain the translation functionality (Supplementary file 1). Some of these include nucleotides at positions G^{266}, G^{268}, U^{269}, U^{297}, A^{288}, A^{252}, A^{136} which have been reported to display lower translation efficiency upon mutations and are also significant in making interactions with the translational machinery (Figure 19). Similarly, there have been nucleotides that are

least preserved and nucleotide changes can be tolerated (Table 21). The least conserved nucleotides that have been detected from analyses of both the databases, MSA 2006 and HCVIVdb, these include at positions 204, 243, 350, 183 and others. Some of these nucleotides have been described in detail in the above section.

5.5.1 Comparison with other databases (viral, IRES and RNA databases)

The HCVIVdb is a unique resource which provides specialized information to its users relating to HCV-IRES variability facilitated with analytical tools for multiple analyses and further allows to visualize, compare and download the data. To establish the novelty of our HCVIVdb we compared it with the construction, contents and utility of knowledge-bases dealing with HCV, IRES, RNA or other viruses. Following are some of the main and important resources that comprise the information associated largely to virus and the RNA.

1. **euHCVdb: The European hepatitis C virus database**
   (https://euhcvdb.ibcp.fr/euHCVdb/)

The euHCVdb is an extension of French HCV database (Combet et al., 2004) and is mainly focused on the computer-annotated HCV sequences based on reference genomes combined with the three-dimensional (3D) molecular models of proteins. The euHCVdb aim attentions at the HCV structural biology, protein sequences and functional analyses with the aid of integrated tools provided on the database website. The reference set of 26 well-characterized complete HCV genomes representing 18 sub-types were utilized in order to develop a fully automated annotation procedure (Simmonds et al., 2005). The development of automated annotation establishes the consistency and standardization of nomenclature for all entries and includes annotation such as genomic regions and proteins present in the entry. It further offers cross-references to external databases and to EMBL database, internal cross-references to reference genome, PubMed links, genotypes / subtypes, source of the sequence and 3D structure of the protein models. The cross-references of multiple

databases allows to collect vast amount of additional data and standardize these datasets for further analyses.

The euHCVdb is divided into static and dynamic part. The static part allows the users to view the genome and protein data which can further be visualized in the pre-computed multiple alignments of reference genomes by EditAlignment. The known 3D structures of protein models with links to the Protein Data Bank (PDB) are also available. The dynamic part consists of a query system where user-defined set of sequences can be build submitted through query interface which is divided into sub-sections. There are >30 criteria that can be defined and utilized to build a query. The results are displayed in a table with access to all the data of the entries with hyperlinks to external sources. The data can be exported for further analysis using tools that also gives possibility for transferring the sequence data to NPS@ web server. Other tools such as sequence similarity search and sequence alignment tools can also be used (Combet et al., 2007, Combet et al., 2009). This database is no longer updated.

2. The Los Alamos hepatitis C sequence and immunology database (http://hcv.lanl.gov/content/index)

The Los Alamos sequence database for HCV, officially launched in 2003, is designed and modelled on Los Alamos HIV database (Kuiken et al., 2003). The HCV sequences are collected mainly from the data deposited in the GenBank and annotated manually. Over 50,000 HCV sequences has been published. The data is presented through a web-accessible search interface and variety of sequence analysis tools. The annotation fields include the sequence information such as genotype, subtype, sampling country, -city, -date and –tissue. The annotation field also contains patient information such as health status, age, gender ALT level, treatment and result, co-infection with HIV and hepatitis B, infection date, country, city, HLA type and epidemiological relationship with other patients. The search interface allows the users to search on various fields and also gives the option to automatically exclude sequences that are possibly contaminants, from non-human hosts, synthetic or epidemiologically related. The searched data can be sorted and selected in multiple ways. One way is the graphical representation that displays the

size and position of the sequences in the genome. The search by genomic region allows to assemble or span the same / specific regions. The retrieved sequences can further be downloaded as an alignment. Ready-made nucleotide and amino acid sequence alignments are also provided/downloaded that can be utilized as reference set to align new sequences, subtyping, analyses. The other features include geography tool which can be used to plot the prevalence of specific genotypes in different countries. In addition, large number of tools has been provided to facilitate variety of data analyses (Kuiken et al., 2005, Kuiken et al., 2008).

The HCV immunology database comprises a collection of annotated HCV immunogenic epitopes which can be searched and analyzed. Two sections of the immunological database co-exist, consisting of T-cell responses and antibodies. T-cell database has over 750 epitopes for both cytotoxic T-cell and helper T-cell epitopes. The maximum of 30 amino acids are required to be mapped to a region, though no optimal boundaries are defined, in order to make the database inclusive and comprehensive. The antibody database lists the monoclonal and polyclonal antibodies that specifically binds to the HCV and contains all the relevant publications. The immunology database provides the users with tables, maps and reference HCV-specific epitopes. There are tools that can be utilized to search for epitopes location in reference to H77 polyprotein, sequence, genotype, human HLA, references and notes (Yusim et al., 2005, Hraber et al., 2007)**.** The database is no longer updated.

3. HVDB; Hepatitis virus database (Japanese)

(http://s2as02.genes.nig.ac.jp/)

The HVDB database was developed as a specialized tool for the users to work on/with hepatitis viruses for functions such as phylogenetic relationship among viral variants and the genomic location of the sequences. Initially it was restricted to HCV but later it also combined and coordinated data regarding HBV and HEV. The sequence data is obtained from the international nucleotide sequence database collaboration (INSDC), comprises databases DDBJ/ EMBL/ GenBank, contains approximately 44, 000 HCV, 11,000 HBV and 1600 HEV submitted sequences. Since INSDC is not specialized in hepatitis viruses, a database that is well-equipped,

updated and characterized with the hepatitis related-information retrieval and analyses is imperative. The HVDB is oriented towards phylogenetic relationships and genomic locations of the viruses, providing the essential tools and services for data analyses (Shin et al., 2008). The entries are acquired from the latest release of DDBJ (Okubo et al., 2006) and all the entry sequences are aligned to the reference genome, sequence strain HCV H-77, AF009606 (Kolykhalov et al., 1997)**.** The 'reference map' comprising of genomic loci and subregions from the reference sequence is utilized to compile information by aligning each entry against the reference sequence. It is termed as 'map information'. Divisions are created containing information for all the nucleotide and amino acids sequences of each locus together with the annotation information obtained from the DDBJ, multiple alignment of the sequences, genetic distance matrix and a phylogenetic tree (Shin et al., 2008).

The HVDB has been provided with the tools to view the data, retrieve and further download it via the interface. The data viewers such as map viewer display 'map information' of viruses' data. It shows the loci, base number of the reference genome and the location of each entry in the form of horizontal line corresponding to its specific position relative to the reference loci and genome base number. This could be extended to view specific region or locus on the genome by obtaining similar entries and further downloaded. Tree viewer allows to view the 'tree' or phylogenetic tree of a division. This viewer is used to observe the phylogenetic relationships of different HCV genotypes. Similarly, the align viewer shows the multiple alignment of a division with consensus sequence in the first line followed by all the sequences with bases differing from the original are identified. Users can also do private divisions with other services such as virus genotyping/subtyping and searches by BLAST and FASTA are also provided (Shin et al., 2008)**.**

4. ViPR; The virus pathogen database and analysis resource (http://www.viprbrc.org/brc/home.spg?decorator=vipr)

The ViPR database is an established, well-equipped, organized and large integrated resources of data and web-based analysis tools providing information about human pathogens. The pathogens include viruses belonging to Arenaviridae, Bunyaviridae,

Caliciviridae, Coronaviridae, Flaviviridae, Filoviridae, Hepeviridae, Herpesviridae, Paramyxoviridae, Picornaviridae, Poxviridae, Reoviridae, Rhabdoviridae and Togaviridae families and provides an array of information such as sequence data, curated viral genome records, structural data, immune epitope location, clinical metadata and data that is acqis varietycomparative genomic analysis. There is variety of analytical and visualization tools provided for sequence variation (SNP), phylogenetic tree, multiple sequence alignment, metadata-driven comparative analysis and BLAST comparison. The resulting data can be stored, shared, combined and integrated in personal work benches for future use. The ViPR database is supported by national institute of allergy and infectious diseases (NIAID) and bioinformatics resource centers (BRC) programs. To date the ViPR database consists of more than 460,000 virus strains from 4,350 species belonging to 116 genera and 70 virus families (Pickett et al., 2012b, Pickett et al., 2012a)**.** We gathered HCV sequences from vipr and compared those sequences with the HCV-IRES sequences that we collected from the literature for validity of our data set, as described above.

5. HCV pro: Hepatitis C virus protein interaction database (http://cbrc.kaust.edu.sa/hcvpro/)

The underlying objective of the database is to provide integrated, user-accessible and standardized information directory dealing with hepatitis C protein-protein interactions (Kwofie et al., 2011). The collation of data comprises manually verified hepatitis C viral-viral and viral-human protein interactions from across the literature and associated databases including BIND (Alfarano et al., 2005), VirusMint (Chatr-aryamontri et al., 2009) and VirHostNet (Navratil et al., 2009). It also contains data on hepatocellular carcinoma (HCC) related genes, functional genomics and annotation, gene ontology, pathways and linked to other cross-referenced relevant databases. For user-query of the database interface searches can be performed using the identifier search of proteins, experimental evidence for the interaction detection method linked with the publication source (PMIDs) and chromosomal location of the genes. The HCV pro biological repository also allows retrieval of the analyzed HCV protein-protein interaction data and crucial information relevant to the functional organization of experimental methods. The data integrated is suggestive of

standardizing the protein-protein interaction format besides enhancing the search for therapeutic potential and diagnostic biomarkers (Kwofie et al., 2011).

6. VirusMentha: resource for virus protein interactions (http://virusmentha.uniroma2.it/)

The VirusMentha is an extension of the VirusMint database (Chatr-aryamontri et al., 2009, Licata et al., 2012) which provides a comprehensive, large-scale, well-organized, integrated and self-updating resource and information for the viral-viral and viral-host interactions (Calderone et al., 2015). The development of VirusMentha is based on the manually curated data of the knowledge-bases that deals with the protein-protein interactions and also makes The International Molecular Exchange (IMEx) consortium (Orchard et al., 2012, Orchard, 2012) http://www.imexconsortium.org/home. The automated integration of the curated data of protein-protein interaction from the five databases occurs through implementation of the resource termed mentha (Calderone et al., 2013). The five databases from which the VirusMentha exploits curation potential are MINT, IntAct (Kerrien et al., 2012), MatrixDB (Launay et al., 2015, Chautard et al., 2011), BioGRID (Chatr-Aryamontri et al., 2015) and DIP (Salwinski et al., 2004). Currently, VirusMentha comprise interactions from 24 viral families and it also contains viral interactions data with hosts other than humans displaying that overall focus is not exclusive to specific virus. The web interface is designed to allow the users to directly conduct their searches from the homepage by entering genes, Uniport IDs and keywords. The searches can be limited to specific viral families, host or can include the entire database with the possibility of downloading the viral interaction evidence. The result page presents the relevant or proteins matched with the query with its Uniport IDs and genes. The proteins can be collected in a ‘protein bag’ which shows the ‘list’ of binary interactions. Users can also align proteins from different isolates and ‘preview’ the direct interactions of the selected proteins. The ‘browse’ gives the option of launching the interactive interactome browser. The interface assists the users with the assembly, organization and modification of the interactive networks (Calderone et al., 2015).

We have further summarized the information from other resources in the Supplementary file 2.

5.6 Codon bias and proper mRNA folding both contribute to the translation efficiency of HCV-core RNA

Corelation of codon bias as a determinant of translation efficiency of the HCV-core gene

The HCV-core protein is a vital structural element for the production and assembly of virus particles that also interacts with large number of viral and cellular proteins. HCV-core protein is a basic protein that adopts a dimeric α-helical configuration and is divided into domains D1 and D2. The N-terminal region (120 aa) makes two-third of the HCV-core protein. It is hydrophilic and comprises highly basic residues, constituting the D1 domain. Whereas, the D2 domain that encompass the C-terminus is hydrophobic (Boulant et al., 2005, McLauchlan, 2000, Gawlik and Gallay, 2014). The domain D1 has shown to be involved in RNA-binding interactions (Ivanyi-Nagy et al., 2006, Shimoike et al., 1999) for nucleocapsid formation and oligomerization of the core protein (Stewart et al., 2016, Boulant et al., 2005, Nakai et al., 2006). In addition, the substitution of basic residues in N-terminal of HCV-core domain (grouped in cluster 2, aa 39 to 62) achieved by site-directed mutagenesis leads to disruption of infectious HCV viral particles production (Alsaleh et al., 2010). The HCV-core comprises highly conserved and structured RNA elements with multiple stem-loops (designated as SL47, SL87, SL248 and SL443) (Figure 20) (Vassilaki et al., 2008). The integrity of the SL47 and SL87 is crucial for virus viability and nucleotide changes in these stem-loops have proved to be detrimental for viral mRNA translation (Vassilaki et al., 2008).

Since HCV-core RNA has a very strict secondary structure we initially selected 54 codons (codon 5-58) of the N-terminal region containing 30 hydrophilic, 13 neutral and 11 hydrophobic residues, respectively. The first 4 codons including the AUG initiation codon was not substituted or changed in order to keep the domain IV secondary structure. The truncated HCV-core protein structure (aa 2-45) was determined using Robetta server (http://robetta.bakerlab.org/) which predicted an intrinsically disordered or unstructured protein structure that reflects on domain D1 facilitation for extensive binding and interaction with cellular and viral RNA/proteins (Figure 21) (Uversky, 2011).

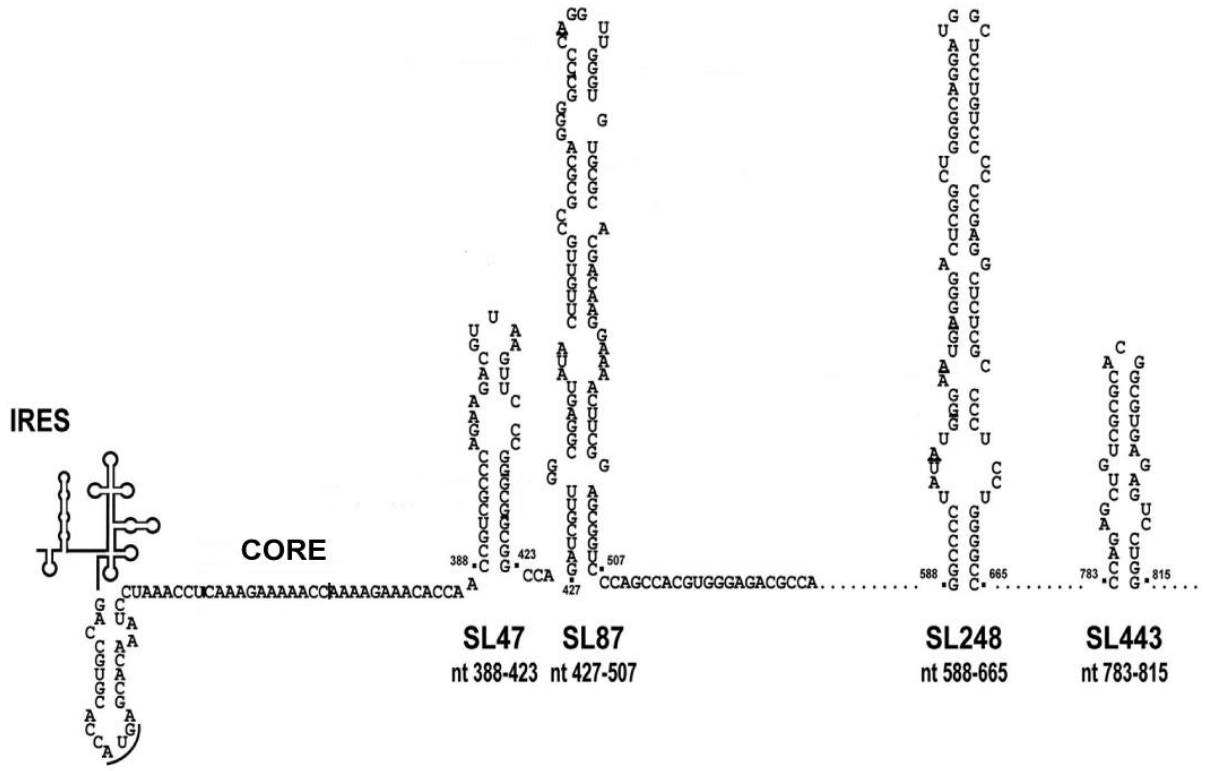

Figure 20: The secondary structure of HCV-core encoding region with multiple stem-loops. The nucleotide numbers refer to JFH-1 isolate (Vassilaki et al., 2008).

We calculated the highest and lowest transfer RNA adaptation index (tAI), for humans (Tuller et al., 2010), of synonymous codons to generate various HCV-core gene constructs in order to detect any correlations that may exist between variable codon bias and the gene expression pattern of the relative core constructs. The tAI is a measure of codon bias in terms of the co-adaptation of tRNA pool or in other words assigns each codon a score based on the availability of corresponding tRNA (Tuller et al., 2010, dos Reis et al., 2004, Waldman et al., 2010, Sabi and Tuller, 2014). The assembly of two of the constructs (HIw and Low) was managed in a way that would maintain the secondary structure of the HCV-core SL-47 and SL-87 domains. Whereas, in the other two HCV-core constructs (HIwo and Lowo) the secondary structure is not taken into account while assigning tAI, high and low, to each codon (Figure 22). The variant HCV-core constructs consists of synonymous

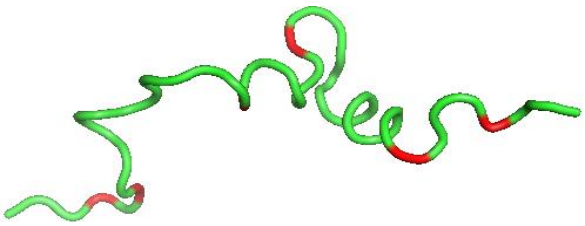

Figure 21: The truncated HCV-core (aa 2-45) display unstructured protein. The structure was analyzed and obtained from Robetta server.

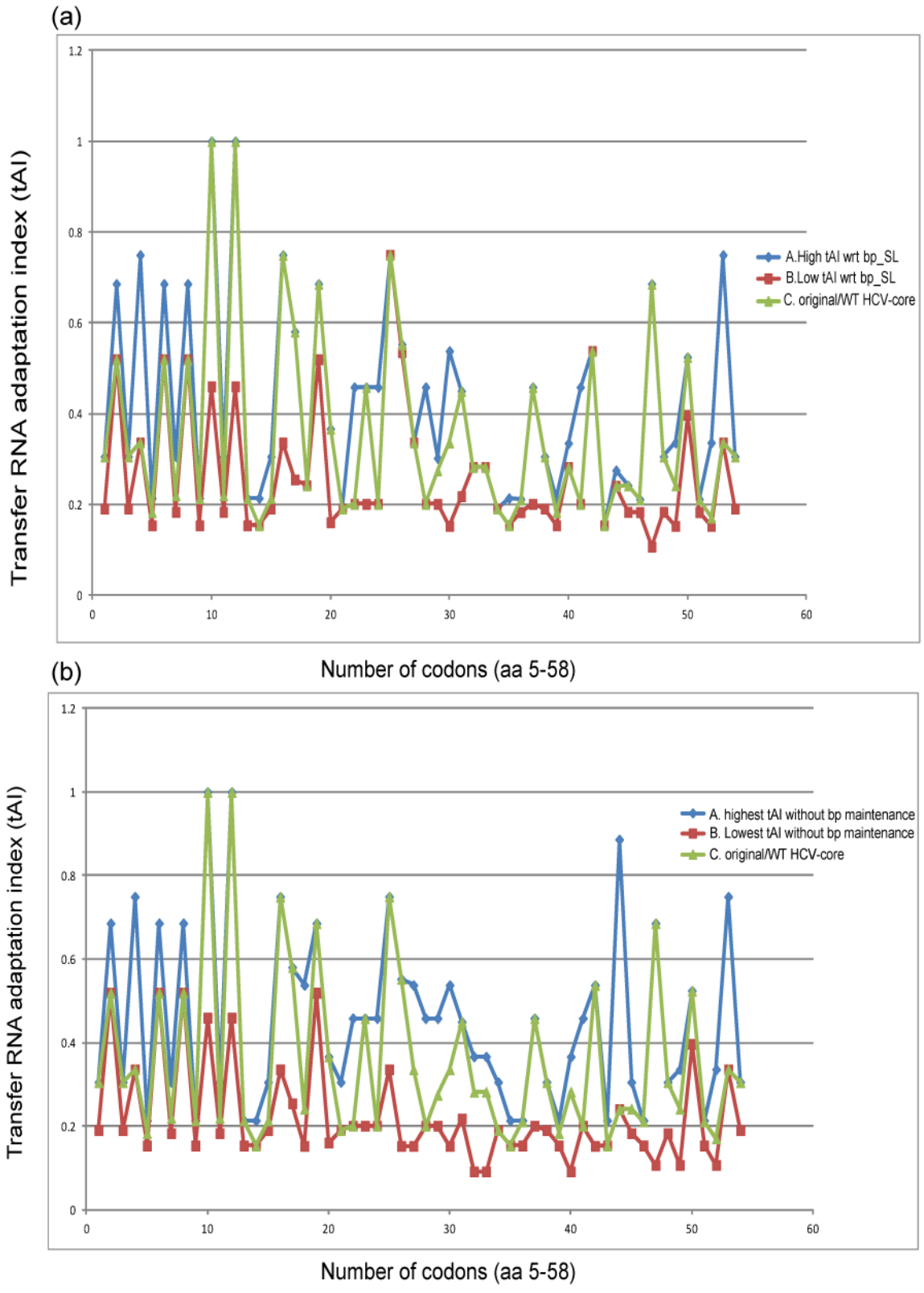

Figure 22: The tAI is used as a measure of codon bias **(a)** The tAI, highest and lowest, for relative HCV-core constructs (from aa 5-58) in which the base-pairing of secondary core structure is maintained (HIw, Low) **(b)** and base-pairing of secondary core structure is not kept (HIwo, Lowo) is plotted, alongside the wild type/original HCV-core.

mutations which were designed to produce the same HCV-core protein (see Materials and Methods, Supplementary file 3). The synthetic oligonucleotides for all the five HCV-core gene constructs (core-original/WT, high without loops (HIwo), high with loops (HIw), low without loops (Lowo), low with loops (Low); see Materials and Methods for details) (Figure 22) were annealed, ligated and the resultant fragment(s) were amplified by the PCR (Figure 23).

The HCV-core segments (nt. 174 / 58 codons) with an upstream full-length HCV-IRES element for all the five constructs were cloned in the bicistronic pRG vector which was also used in earlier study of the HCV-IRES variability (the assigned digits alongside each construct correspond to the number of times it took to successfully clone the HCV-core fragment). We transiently expressed the HCV-core fragments in CCL-13 liver cells as a core-EGFP fusion protein. The cells were cultivated 48 hours after transient transfection and lysate from cell culture was prepared for the western blot. The nitrocellulose membrane was probed with mouse monoclonal anti-HCV-Core Antigen (abcam®) and DsRED2 (25) (Santa Cruz), primary antibodies (dilution 1:1000 for both core and DsRED2 Ab). The signals for gene expression of all the HCV-core constructs were detected by western blot along with the cap-dependent translation of DsRED2 (Figure 24a). The quantification of the expression was achieved using Carestream MI SE software (Kodak) aimed at visualizing the differences of expression that may have risen between the HCV-core constructs relative to their tAI values and the presence or absence of the secondary structure.

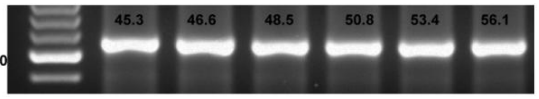

Figure 23: Gradient PCR showing amplified fragments of the core-Lowo construct with all the annealing temperatures (°C) mentioned.

We run 4 biological replicates for each of the HCV-core construct in order to reach accurate and reliable measurement of the protein levels. Successful expression of the core proteins were detected 48h post-transfection of CCL-13 cells by Western blotting (Figure 24a). The expression of core-EGFP fusion proteins were quantified and normalized against expression of the first cistron in pRG vector (Figure 3, Material and methods: page 36) encoding for DsRED2. All values were compard to the WT core proteins (Core B2) (Figure 24b). Experiments yielded variable levels of translation in different core constructs that may correspond to variation in tAI and stability of the SL47 and SL87 stem-loops.

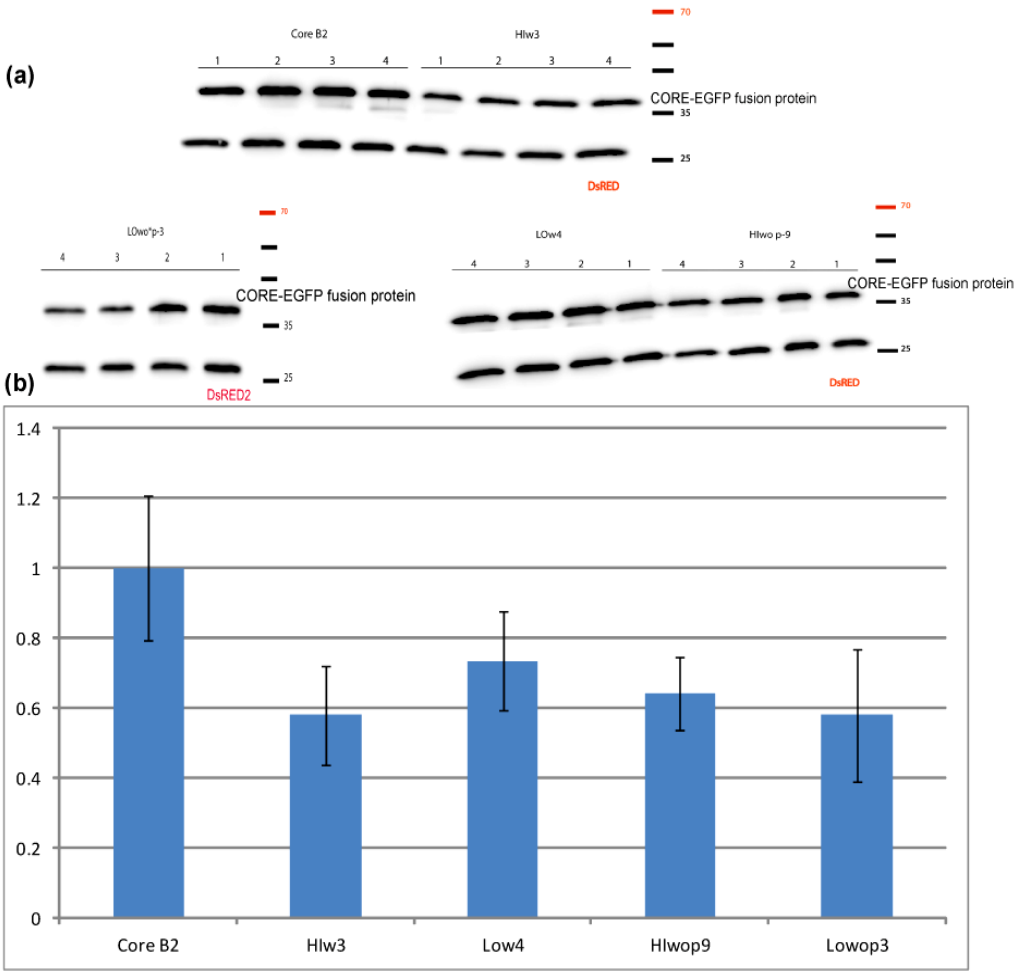

Figure 24: (a) The Western blot analysis of the HCV-core-EGFP fused constructs' (Core B2, Hlw3, Hlwop9, Low4, Lowop3, Lowo4) and DsRED2 protein expression in CCL-13 cells (biological replicates) **(b)** Quantification of the HCV-core protein levels showed inconsistent translation efficiency among varied HCV-core constructs.

The highest core protein amount was displayed by the original WT HCV-core (Core B2) which is suggestive that a balance between codon bias (not necessarily highest or lowest codon bias) and mRNA structure are important in achieving maximum translation efficiency.

Our experiments have shown that both codon bias and the mRNA folding are crucial determinants of translation efficiency on local gene expression level. The observed translation rates of truncated HCV-core protein predicts that efficient translation is linked to the interplay between both the codon bias and mRNA folding during HCV-IRES mediated translation initiation. The balance, however, of codon bias (not essentially high and/or low) and mRNA structure could be important in maximizing

the translation efficiency as shown by the wild type HCV-core protein levels. Although, some mutations in certain positions of SL47 and SL87 were shown to be detrimental for virus viability (Vassilaki et al., 2008), and those mutations were avoided in preparation of two of the HCV-core constructs (HIw, Low). The presence of other synonymous mutations nonetheless, can have impact in shaping the translation and could be the case that increased protein levels seen in the WT HCV-core is due to no nucleotide changes in its primary sequence, contrary to other HCV-core constructs.

In one of the study, no correlation of codon bias in shaping the translation of 154 variants of GFP protein expressed in *E.coli* was established but the presence of mRNA folding was linked to influence the protein levels of individual genes. The codon bias affects the global translation efficiency (Kudla et al., 2009). In difference, the codon bias was shown to be significantly correlated as a determinant of translation efficiency on protein-to-mRNA ratio whereas, folding energy impacts the global translation (Tuller et al., 2010). Intensive progress has been accomplished in the past years to measure the gene expression levels of proteins resulting in various new measures and hypotheses attributed to codon-anticodon adaptation leading to much clear establishment of coding sequence contribution to the translation efficiency.

6 CONCLUSIONS

The present thesis contributes in understanding the role of HCV-IRES in modulation of translation initiation. The HCV-IRES variability is crucial and has a direct effect on translation efficiency. The impact of sequence variation(s) in defining the outcome of mRNA protein synthesis was achieved by comparative analyses of the mutation, structure, biochemical data, from our patients' samples and HCVIVdb, to obtain an overview of the HCV-IRES characteristics and specificity of responses on the level of domain, subdomain and nucleotides. Intriguing results were found that shed insight into the behavior and working of the HCV-IRES and can be utilized to enhance the ongoing research in HCV-IRES field relating to its translation and virus replication.

The experimental observations lead to the following conclusions;

1. The diversity of HCV-IRES found in patients infected with HCV was visualized by employing methods such as DGGE and TGGE, after cloning the HCV-IRES fragments in a bicistronic pRG vector. The migration pattern of the particular HCV-IRES cDNA constructs from all the patients' samples demonstrated the amount of variation that has been introduced in patients over a period of time.

2. The HCV-IRES cDNA amplicon showing variable migration pattern from DGGE and TGGE were selected and sequenced. With the aid of bicistronic pRG vector, the HCV-IRES activity was measured by analyzing the transient expression of EGFP reporter gene in mammalian cell culture, employing flow cytometry. The expression pattern exhibits range of activities for different constructs.

3. To analyze the measured HCV-IRES translation efficiency on a much broader level and determine the accuracy of measured activity, we developed a large-scale HCV-IRES variation database (HCVIVdb; www.hcvivdb.org) that comprises ~1900 mutations from majority of the HCV-IRES mutation-linked published studies. The data were assembled and organized for functional and comparative analyses that allowed wider-range observations of the HCV-IRES mutants and the probable impact they can cause to translation initiation.

4. We identified HCV-IRES least and most vulnerable regions to mutations by calculating the cumulative average translation efficiency of individual mutations and multiple mutations, respectively from the available HCVIVdb dataset.

5. A unique characteristic of the HCV-IRES was found that can be attributed to a long-range inter- and/or intra-domain interaction. The multiple mutantions shown to restore the HCV-IRES translation efficiency, however, the individual occurrences of these mutations proved to be defective for viral protein production. Likewise, a sharp decline of HCV-IRES activity in multiple mutantions was also found having individual mutations with not much impact on viral mRNA translation.

6. We established the functional roles of (~20) novel HCV-IRES mutations that we found in our patients' samples using flow cytometry. The observed HCV-IRES functional response was also described in the context of nucleotide's sequence-structure preservation and interaction with the translation machinery by comparing it with available mutation (HCVIVdb), structure and biochemical data.

7. The naturally occurring HCV-IRES mutations from patients infected with HCV were collected from the HCVIVdb and mapped on the HCV-IRES structure to look for any correlation between (a) frequency and location of mutations (b) and corresponding therapy outcome in sustained and non-sustained viral responders. No direct link could be found.

8. The HCVIVdb dataset was validated by multiple sequence alignment of HCV data from other resources and correlated for the variability count. The datasets showed positive correlation that lead to other interesting finds. The hypervariability and conservation of particular nucleotides were found and correlated with their functional impact on HCV-IRES translation initiation. In addition, the extensive data can be analyzed to further evaluate the overall evolutionary advantage / preservation of some bases over the other at each HCV-IRES nucleotide position.

9. The HCVIVdb contains specialized information that deals with HCV-IRES mutations, assembled and organized for faster and easily accessible functional and comparative analyses. The HCVIVdb is a unique central resource that offers information for researchers in IRES and hepatitis C oriented fields. The contribution

and one-of-a-kind HCVIVdb dataset was also approved by comparing it with the current and existing viral, IRES and RNA knowledge-bases.

10. To find if the codon bias and mRNA structure impacts the translation efficiency of the HCV polypeptide. We prepared truncated HCV-core gene constructs that vary in their codon usage and mRNA structure, using transfer RNA adaptation index (tAI) as a measure of the codon bias. Our results indicate interplay between codon bias and mRNA structure as determinants of the gene expression efficiency in HCV-core RNA, during HCV-IRES mediated translation initiation.

The outcome of HCV-IRES variability research was in the form of three manuscripts and an HCV-IRES variation database (HCVIVdb: www.hcvivdb.org).

Supplementary file 2 A list of online databases related to the hepatitis C virus. Sections are categorized based on the content of the database, tools available and the link to the relevant website and publication.

Name	Database Content	Analytical tools	Links and references
euHCVdb: The European HCV database	HCV: protein sequence, structure and function analyses. In collaboration with Japanese (Hepatitis Virus db) and US databases (Los Alamos). Selection of computer-annotated HCV sequences based on reference genome includes genome mapping of sequences, sub-typing and 3D molecular models of proteins.	Tools: divided into static and dynamic sections Static: Access to genomic regions and proteins with known 3D structures are linked to PDB. Dynamic: consist of a query section, dynamic set of sequences and 3D model with variety of user-defined criteria. The available data can be exported and further analyzed by bioinformatic programs in Web servers @NPS and/or PIG. Links to external sources such as EMBL	https://euhcvdb.ibcp.fr/euHCVdb/jsp/index.jsp (Combet et al., 2007)
FLAVIdb	FLAVIdb is an integrated, comprehensive knowledge-base that catalogues antigenic data of flavivirus collated from multiple sources (GenBank, GenPept, Uniport, IEDB and PDB). The information on 12,858 flavivirus antigenic sequence entries with 184 verified T-cell epitopes and 201 verified B-cell epitopes has been collected, annotated and assembled from the integrated resource databases. The specialized analysis tools allows the users to analyze data (protein and nucleotide sequences, macromolecular structure and	The web interface of the database comprises basic search tools together with the advanced or specialized tools developed for FLAVIdb (BB score, block entropy analysis and viral species classifier). The tools, both generic and specialized, are combined to implement two work flows, summary and query analyzer work flow, that assist in automated complex queries, data mining and report generation. The data mining has been equipped with tools such as multiple sequence alignment and variability analysis and also allow the users to rapidly retrieve the information stored in FLAVIdb.	http://cvc.dfci.harvard.edu/flavi/index.php (Olsen et al., 2011) (Benson et al., 2013) (Berman et al., 2002) (2015)

	immune epitope) oriented towards immunology and vaccinology of the flavivirus.		
HCV pro	HCV: Virus-virus and virus-cellular protein-protein interaction, functional genomics, hepatocellular carcinoma (HCC)-related gene expression and molecular data. Manually verified interaction curated from the literature and databases. Data on Hepatocellular carcinoma related gene expression on genes encoding cellular proteins.	Tools available: protein selection (HCV and humans) chromosomal numbers and evidence that includes experiment type for concluding the interactions along with PubMed ID. Additionally, HCVpro, BIND and VirusMint IDs.	http://cbrc.kaust.edu.sa/hcvpro/index.php Updated (Kwofie et al., 2011)
HCVIVdb: HCV IRES variation database	Focused on the variations that have been found in the HCV IRES sequence across various genotypes and subtypes. The datasets from studies have been collated for the users to understand the function of HCV IRES associated with diverse mutations described in different domains. Allows direct comparative and functional analysis.	Tools that provide extensive search into the variation entries. HCVIVdb entries are grouped into further categories based on genotypes, naturally occurring or engineered location of occurrence, mutation type, range of translation efficiency and original publication. The data related to any category can be retrieved and downloaded.	http://hcvivdb.org/references.php (Khawaja et al., 2015)
HCV-LANL	HCV sequences with details of isolate and patients. Association of HCV variants with infection. Individual and epidemiological outbreaks.	Retrieve data (sequences, clinical information) Tools for sequence variation analysis.	http://hcv.lanl.gov/content/index (Kuiken et al., 2005) Part of database that addresses immunology is not maintained since 2007.
HepSEQ-Research Database System	Hepatitis B infection: molecular, clinical, epidemiological, nucleotide sequence and mutation features. Information on HBV genotype and	Tools provided: SeqMatcher, Genotyper, Gene Mutation, Mutation Annotator. SeqMatcher: Sequence homology searches. Genotyper: Identify HBV genotypes.	http://www.hepseq.org/Public/Web_Front/main.php (Gnaneshan et al., 2007)

	mutation recognition with known clinical impact. Sequence homology searches. All the data are collected from participating centers consisting of virological, clinical, and epidemiological locations, and nucleotide sequences are manually checked and uploaded to the database.	Gene Mutation: Identify mutations in HBV coding regions. Mutation annotator: Annotate sequences known to be linked to anti-viral resistance.	
IEdb: Immune Epitope Database and Analysis Resource	Characterizing; human, non-human, primates. Antibody, T-cell epitopes involved in infectious disease, allergy, auto-immunity, HCV epitopes, experimental and self-antigens.	Epitopes curated manually. Provides intrinsic structural and phylogenetic features. Epitope interaction with host immune system. Tools to query database and analyze epitope information. Also assists in prediction of B-cell and T-cell epitope.	http://www.iedb.org/home_v3.php (Vita et al., 2015)
PhEVER	Interaction; virus-virus, virus-host and drive novel functions through the exchanges of their genetic material (lateral gene transfer). Evolutionary (Viral and protein evolution) and phylogenetic information. Homologous gene family b/w (a) different virus sequences (b) viral seq and seq from cellular organisms. Clustering of homologous protein families.	Complete genomes and their annotations were taken from: (Ensembl, Genome Reviews and RefSeq Viral). Detect sequence alignments and homologies. 2426 non-redundant viral genomes. 1007 non-redundant prokaryotic genomes. 43 non-redundant eukaryotic genomes. Tool to search and analyze horizontal gene transfer.	http://pbil.univ-lyon1.fr/databases/phever/index.php http://www.ncbi.nlm.nih.gov/pubmed/?term=PMC3013642
RNA virus Database	Genome organization. Identify submitted nucleotides. Translated genome sequences for all species.	Analytical tools for 938 known species of RNA virus. Multiple whole-genome alignments. Guidance to other web sources.	http://bioafrica.mrc.ac.za/rnavirusdb/ http://www.ncbi.nlm.nih.gov/pubmed/?term=18948277 The database currently has 1062 viruses (June 2010) Page layout last updated 17 February 2011
ViPR: The virus pathogen	Sequence data. Curation: Viral genomes and	Tools that provide searchable and comprehensive analysis of gene function	For updated sequence data and analytical tools, please visit the ViPR BRC: (http://www.viprbrc.org)

database and resource center	genes Arenaviridae, Bunyaviradae, Filoviradae, Paramyxoviradae, Poxviridae and Togaviridae family including HCV.	with relation of genotype to phenotype and pathogenesis.	
VIDA (The virus database)	Characterize animal virus open reading frame sequences. Homologous protein families from complete and partial viral genomes. Provides means to study evolution and function in viruses. Further cataloging of homologous protein families into various functional classes. Virus families: Arteriviridae, Coronaviridae, Herpesviridae, Papillomaviridae, Poxviridae.	Complete and partial genomes of viral families taken from GenBank and filtered. Links to PDB and CATCH for protein structures. Alternative ways to search for a query in database.	http://www.biochem.ucl.ac.uk/bsm/virus_database/VIDA3/VIDA.html (Alba et al., 2001) Appears that it is no longer being updated
VIRALZONE	Virion structure, replicative cycle, host-virus interaction. Molecular biology and epidemiology for each virus genus with links to proteome UniportKB. 133 viral ontology pages.	Illustrations, text and PubMed references. Links to many databases. List of many annotated proteins in UniportKB.	http://viralzone.expasy.org/ (Masson et al., 2013)
VirBase: a resource for virus–host ncRNA-associated interactions	The database comprises virus-host non-coding RNA (ncRNA) associated interactions and network complexes. The data is manually curated from the literature and aim attention to the development and organization of viral and cellular ncRNAs-related virus-virus, virus-host, host-virus and host-host interactions in viral infection and interaction networks. VirBase includes viruses such as	The total ncRNA categories have been 4 with objective to collate all the information from the respective integrated knowledge-base resource such as long non-coding RNA database (lncRNAdb), miRBase, sno/scaRNAbase and snoRNA-LBME-db. The interface for the database is convenient to use with availability for the users to retrieve interaction data. Within the main category of 'search' and 'browse', the query can further be inspected through sub-categories that include search by keyword, by validated methods and browse by virus	http://www.rna-society.org/virbase/index.html (Li et al., 2015) lncRNAdb: http://lncrnadb.com/ (Quek et al., 2015) miRBase: http://www.mirbase.org/ (Kozomara and Griffiths-Jones, 2011) sno/scaRNAbase: (Xie et al., 2007) snoRNA-LBME-db: https://www-snorna.biotoul.fr/ (Lestrade and Weber, 2006)

	Epstein-Barr virus, human immunodeficiency virus-1 (HIV-1), hepatitis C virus, herpes viruses, papilloma viruses etc. Currently, the database registered 61 viruses and 24 hosts with >12,000 entries of ncRNA associated virus-host interactions.	or ncRNA. The result page offers further brief and descriptive details regarding PubMed ID or associated functions/pathways etc. The visualization panel allows the users to analyze the ncRNA-related interactions by selecting specific virus and host. Users can also submit the new data and download all the host-virus interaction data.	
Virus Variation Resource	The database is an extension of the NCBI influenza virus variation resource. It comprises a comprehensive collection of a large sequence dataset that with the aid of metadata-driven specialized search interface, variety of available tools and displays can be analyzed and retrieved. The database integrates viruses such as West Nile, Dengue, Ebola, Zika virus etc. The nucleotide and protein annotation is carried out through a standardized set of database loading pipelines which is then stored and can be selected using specific biological and clinical criteria. The retrieved sequences can further be downloaded.	The search interface supports number of search criteria that includes sequence patterns, geographic region, sequence, host, strain names in order to display and retrieve protein or DNA sequences from the specified genomes. The retrieved sequences have the option to be downloaded in different formats. A new multiple sequence alignment viewer has been added to facilitate the visualization of large datasets that display dynamic comparison of sequences. The phylogenetic tree viewer allows the users to inspect the phylogenetic trees of the selected nucleotide or protein sequences.	http://www.ncbi.nlm.nih.gov/genome/viruses/variation/ (Brister et al., 2014) NCBI influenza virus variation resource: (Bao et al., 2008)

7 List of Publications

7.1 Publications

Khawaja, A., Vopalensky, V. and Pospisek, M. (2014), Understanding the potential of hepatitis C virus internal ribosome entry site domains to modulate translation initiation via their structure and function. WIREs RNA. doi: 10.1002/wrna.1268

Floden, EW.*, **Khawaja, A.*,** Vopalensky, V. and Pospisek, M. (2016), HCVIVdb: The hepatitis-C IRES variation database. Manuscript submitted in BMC Microbiology.

**Authors contributed equally.*

Khawaja, A., Vopalensky, V., Roznovsky, L., Mrazek, J. and Pospisek, M. (2016), A study of the HCV IRES variability: An experimental approach coupled with design of a large-scale mutation database. Manuscript.

7.2 Conferences

ORAL Presentations

Khawaja, A., Vopalensky, V. and Pospisek, M. Inquiry into the variability of HCV IRES and its impact on function by developing and evaluation of a large-scale mutation database that also unfolds potential of some new nucleotides. RNA society 2013, Davos, Switzerland (June 2013).

Khawaja, A., Vopalensky, V. and Pospisek, M. Extensive investigation of the HCV IRES variability suggested new functions to some nucleotides. International RNA Club meeting, Prague, Czech Republic (Nov., 2011).

Khawaja, A., Vopalensky, V. and Pospisek, M. Extensive investigation of the HCV IRES variability suggested new functions to some nucleotides and revealed possible importance of domain II for the therapy. Society of General Microbiology (SGM), Harrogate International Centre, United Kingdom (April, 2011).

Khawaja, A., Vopalensky, V. and Pospisek, M. Analysis of HCV IRES Quasispecies. International RNA Club meeting, Prague, Czech Republic (Nov., 2008).

POSTER Presentations

Khawaja, A., Vopalensky, V. and Pospisek, M. Extensive investigation of the HCV IRES variability suggested new functions to some nucleotides. EMBO Conference. Series on Protein Synthesis and Translational Control. EMBL Heidelberg, Germany (September, 2011).

Khawaja, A., Vopalensky, V. and Pospisek, M. Extensive investigation of the HCV IRES variability suggested new functions to some nucleotides and revealed possible importance of domains II and IIIb for the therapy. RNA Society 2011, Kyoto, Japan (June, 2011).

Khawaja, A., Vopalensky, V. and Pospisek, M. Point Mutation Determination of HCV-IRES Quasispecies Based on DGGE and TGGE Methods. International Symposium on Hepatitis C Virus and other related viruses, Nice, France (October, 2009).

Khawaja, A., Vopalensky, V. and Pospisek, M. Analysis of HCV- IRES Quasispecies. Federation of European Biochemical Socities (FEBS), Prague, Czech Republic (July, 2009).

8 References

2015. UniProt: a hub for protein information. *Nucleic Acids Research,* 43**,** D204-12.

ALBA, M. M., LEE, D., PEARL, F. M., SHEPHERD, A. J., MARTIN, N., ORENGO, C. A. & KELLAM, P. 2001. VIDA: a virus database system for the organization of animal virus genome open reading frames. *Nucleic Acids Research,* 29**,** 133-6.

ALFARANO, C., ANDRADE, C. E., ANTHONY, K., BAHROOS, N., BAJEC, M., BANTOFT, K., BETEL, D., BOBECHKO, B., BOUTILIER, K., BURGESS, E., BUZADZIJA, K., CAVERO, R., D'ABREO, C., DONALDSON, I., DORAIRAJOO, D., DUMONTIER, M. J., DUMONTIER, M. R., EARLES, V., FARRALL, R., FELDMAN, H., GARDERMAN, E., GONG, Y., GONZAGA, R., GRYTSAN, V., GRYZ, E., GU, V., HALDORSEN, E., HALUPA, A., HAW, R., HRVOJIC, A., HURRELL, L., ISSERLIN, R., JACK, F., JUMA, F., KHAN, A., KON, T., KONOPINSKY, S., LE, V., LEE, E., LING, S., MAGIDIN, M., MONIAKIS, J., MONTOJO, J., MOORE, S., MUSKAT, B., NG, I., PARAISO, J. P., PARKER, B., PINTILIE, G., PIRONE, R., SALAMA, J. J., SGRO, S., SHAN, T., SHU, Y., SIEW, J., SKINNER, D., SNYDER, K., STASIUK, R., STRUMPF, D., TUEKAM, B., TAO, S., WANG, Z., WHITE, M., WILLIS, R., WOLTING, C., WONG, S., WRONG, A., XIN, C., YAO, R., YATES, B., ZHANG, S., ZHENG, K., PAWSON, T., OUELLETTE, B. F. & HOGUE, C. W. 2005. The Biomolecular Interaction Network Database and related tools 2005 update. *Nucleic Acids Research,* 33**,** D418-24.

ALGIRE, M. A., MAAG, D. & LORSCH, J. R. 2005. P-i release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation. *Molecular Cell,* 20**,** 251-262.

ALSALEH, K., DELAVALLE, P. Y., PILLEZ, A., DUVERLIE, G., DESCAMPS, V., ROUILLE, Y., DUBUISSON, J. & WYCHOWSKI, C. 2010. Identification of basic amino acids at the N-terminal end of the core protein that are crucial for hepatitis C virus infectivity. *Journal of Virology,* 84**,** 12515-28.

ANDREEVA, A., HOWORTH, D., CHOTHIA, C., KULESHA, E. & MURZIN, A. G. 2014. SCOP2 prototype: a new approach to protein structure mining. *Nucleic Acids Research,* 42**,** D310-4.

ANGULO, J., ULRYCK, N., DEFORGES, J., CHAMOND, N., LOPEZ-LASTRA, M., MASQUIDA, B. & SARGUEIL, B. 2016. LOOP IIId of the HCV IRES is essential for the structural rearrangement of the 40S-HCV IRES complex. *Nucleic Acids Research,* 44**,** 1309-25.

ATTWOOD, T. K., BRADLEY, P., FLOWER, D. R., GAULTON, A., MAUDLING, N., MITCHELL, A. L., MOULTON, G., NORDLE, A., PAINE, K., TAYLOR, P., UDDIN, A. & ZYGOURI, C. 2003. PRINTS and its automatic supplement, prePRINTS. *Nucleic Acids Research,* 31**,** 400-2.

AUTON, A., BROOKS, L. D., DURBIN, R. M., GARRISON, E. P., KANG, H. M., KORBEL, J. O., MARCHINI, J. L., MCCARTHY, S., MCVEAN, G. A. & ABECASIS, G. R. 2015. A global reference for human genetic variation. *Nature,* 526**,** 68-74.

BABAYLOVA, E., GRAIFER, D., MALYGIN, A., STAHL, J., SHATSKY, I. & KARPOVA, G. 2009. Positioning of subdomain IIId and apical loop of domain II of the hepatitis C IRES on the human 40S ribosome. *Nucleic Acids Research,* 37**,** 1141-1151.

BAN, N., BECKMANN, R., CATE, J. H., DINMAN, J. D., DRAGON, F., ELLIS, S. R., LAFONTAINE, D. L., LINDAHL, L., LILJAS, A., LIPTON, J. M., MCALEAR, M. A., MOORE, P. B., NOLLER, H. F., ORTEGA, J., PANSE, V. G., RAMAKRISHNAN, V., SPAHN, C. M., STEITZ, T. A., TCHORZEWSKI, M., TOLLERVEY, D., WARREN, A. J., WILLIAMSON, J. R., WILSON, D., YONATH, A. & YUSUPOV, M. 2014. A new system for naming ribosomal proteins. *Curr Opin Struct Biol,* 24**,** 165-9.

BAO, Y., BOLOTOV, P., DERNOVOY, D., KIRYUTIN, B., ZASLAVSKY, L., TATUSOVA, T., OSTELL, J. & LIPMAN, D. 2008. The influenza virus resource at the National Center for Biotechnology Information. *Journal of Virology,* 82**,** 596-601.

BARRIA, M. I., GONZALEZ, A., VERA-OTAROLA, J., LEON, U., VOLLRATH, V., MARSAC, D., MONASTERIO, O., PEREZ-ACLE, T., SOZA, A. & LOPEZ-LASTRA, M. 2009. Analysis of natural variants of the hepatitis C virus internal ribosome entry site reveals that primary sequence plays a key role in cap-independent translation. *Nucleic Acids Research,* 37**,** 957-971.

BENSON, D. A., CAVANAUGH, M., CLARK, K., KARSCH-MIZRACHI, I., LIPMAN, D. J., OSTELL, J. & SAYERS, E. W. 2013. GenBank. *Nucleic Acids Research,* 41**,** D36-42.

BENSON, D. A., CLARK, K., KARSCH-MIZRACHI, I., LIPMAN, D. J., OSTELL, J. & SAYERS, E. W. 2015. GenBank. *Nucleic Acids Research,* 43**,** D30-5.

BERMAN, H. M., BATTISTUZ, T., BHAT, T. N., BLUHM, W. F., BOURNE, P. E., BURKHARDT, K., FENG, Z., GILLILAND, G. L., IYPE, L., JAIN, S., FAGAN, P., MARVIN, J., PADILLA, D., RAVICHANDRAN, V., SCHNEIDER, B., THANKI, N., WEISSIG, H., WESTBROOK, J. D. & ZARDECKI, C. 2002. The Protein Data Bank. *Acta Crystallogr D Biol Crystallogr,* 58**,** 899-907.

BERRY, K. E., WAGHRAY, S. & DOUDNA, J. A. 2010. The HCV IRES pseudoknot positions the initiation codon on the 40S ribosomal subunit. *RNA-a Publication of the RNA Society,* 16**,** 1559-1569.

BERRY, K. E., WAGHRAY, S., MORTIMER, S. A., BAI, Y. & DOUDNA, J. A. 2011. Crystal Structure of the HCV IRES Central Domain Reveals Strategy for Start-Codon Positioning. *Structure,* 19**,** 1456-1466.

BHAT, P., SHWETHA, S., SHARMA, D. K., JOSEPH, A. P., SRINIVASAN, N. & DAS, S. 2015. The beta hairpin structure within ribosomal protein S5 mediates interplay between domains II and IV and regulates HCV IRES function. *Nucleic Acids Research,* 43**,** 2888-901.

BOEHRINGER, D., THERMANN, R., OSTARECK-LEDERER, A., LEWIS, J. D. & STARK, H. 2005. Structure of the hepatitis C virus IRES bound to the human 80S ribosome: Remodeling of the HCVIRES. *Structure,* 13**,** 1695-1706.

BOERNEKE, M. A., DIBROV, S. M., GU, J., WYLES, D. L. & HERMANN, T. 2014. Functional conservation despite structural divergence in ligand-responsive RNA switches. *Proc Natl Acad Sci U S A,* 111**,** 15952-7.

BOULANT, S., VANBELLE, C., EBEL, C., PENIN, F. & LAVERGNE, J. P. 2005. Hepatitis C virus core protein is a dimeric alpha-helical protein exhibiting membrane protein features. *Journal of Virology,* 79**,** 11353-65.

BRACHO, M. A., MOYA, A. & BARRIO, E. 1998. Contribution of Taq polymerase-induced errors to the estimation of RNA virus diversity. *Journal of General Virology,* 79 (Pt 12)**,** 2921-8.

BRISTER, J. R., BAO, Y., ZHDANOV, S. A., OSTAPCHUCK, Y., CHETVERNIN, V., KIRYUTIN, B., ZASLAVSKY, L., KIMELMAN, M. & TATUSOVA, T. A. 2014. Virus Variation Resource--recent updates and future directions. *Nucleic Acids Research,* 42**,** D660-5.

BROOKSBANK, C., BERGMAN, M. T., APWEILER, R., BIRNEY, E. & THORNTON, J. 2014. The European Bioinformatics Institute's data resources 2014. *Nucleic Acids Research,* 42**,** D18-25.

BROWN, E. A., ZHANG, H. C., PING, L. H. & LEMON, S. M. 1992. Secondary Structure of the 5' Nontranslated Regions of Hepatitis-C Virus and Pestivirus Genomic Rnas. *Nucleic Acids Research,* 20**,** 5041-5045.

BURATTI, E., GEROTTO, M., PONTISSO, P., ALBERTI, A., TISMINETZKY, S. G. & BARALLE, F. E. 1997. In vivo translational efficiency of different hepatitis C virus 5'-UTRs. *Febs Letters,* 411**,** 275-280.

BURATTI, E., TISMINETZKY, S., ZOTTI, M. & BARALLE, F. E. 1998. Functional analysis of the interaction between HCV 5 ' UTR and putative subunits of eukaryotic translation initiation factor eIF3. *Nucleic Acids Research,* 26**,** 3179-3187.

CAI, Q., TODOROVIC, A., ANDAYA, A., GAO, J. Y., LEARY, J. A. & CATE, J. H. D. 2010. Distinct Regions of Human eIF3 Are Sufficient for Binding to the HCV IRES and the 40S Ribosomal Subunit. *Journal of Molecular Biology,* 403**,** 185-196.

CALDERONE, A., CASTAGNOLI, L. & CESARENI, G. 2013. mentha: a resource for browsing integrated protein-interaction networks. *Nat Methods,* 10**,** 690-1.

CALDERONE, A., LICATA, L. & CESARENI, G. 2015. VirusMentha: a new resource for virus-host protein interactions. *Nucleic Acids Research,* 43**,** D588-92.

CERAMI, E. G., GROSS, B. E., DEMIR, E., RODCHENKOV, I., BABUR, O., ANWAR, N., SCHULTZ, N., BADER, G. D. & SANDER, C. 2011. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Research,* 39**,** D685-90.

CHATR-ARYAMONTRI, A., BREITKREUTZ, B. J., OUGHTRED, R., BOUCHER, L., HEINICKE, S., CHEN, D., STARK, C., BREITKREUTZ, A., KOLAS, N., O'DONNELL, L., REGULY, T., NIXON, J., RAMAGE, L., WINTER, A., SELLAM, A., CHANG, C., HIRSCHMAN, J., THEESFELD, C., RUST, J., LIVSTONE, M. S., DOLINSKI, K. & TYERS, M. 2015. The BioGRID interaction database: 2015 update. *Nucleic Acids Research,* 43**,** D470-8.

CHATR-ARYAMONTRI, A., CEOL, A., PELUSO, D., NARDOZZA, A., PANNI, S., SACCO, F., TINTI, M., SMOLYAR, A., CASTAGNOLI, L., VIDAL, M., CUSICK, M. E. & CESARENI, G. 2009. VirusMINT: a viral protein interaction database. *Nucleic Acids Research,* 37**,** D669-73.

CHAUTARD, E., FATOUX-ARDORE, M., BALLUT, L., THIERRY-MIEG, N. & RICARD-BLUM, S. 2011. MatrixDB, the extracellular matrix interaction database. *Nucleic Acids Research,* 39**,** D235-40.

CHOO, Q. L., RICHMAN, K. H., HAN, J. H., BERGER, K., LEE, C., DONG, C., GALLEGOS, C., COIT, D., MEDINA-SELBY, R., BARR, P. J. & ET AL. 1991. Genetic organization and diversity of the hepatitis C virus. *Proc Natl Acad Sci U S A,* 88**,** 2451-5.

COLLIER, A. J., GALLEGO, J., KLINCK, R., COLE, P. T., HARRIS, S. J., HARRISON, G. P., ABOUL-ELA, F., VARANI, G. & WALKER, S. 2002. A conserved RNA structure within the HCVIRES eIF3-binding site. *Nature Structural Biology,* 9**,** 375-380.

COMBET, C., BETTLER, E., TERREUX, R., GARNIER, N. & DELEAGE, G. 2009. The euHCVdb suite of in silico tools for investigating the structural impact of mutations in hepatitis C virus proteins. *Infect Disord Drug Targets,* 9**,** 272-8.

COMBET, C., GARNIER, N., CHARAVAY, C., GRANDO, D., CRISAN, D., LOPEZ, J., DEHNE-GARCIA, A., GEOURJON, C., BETTLER, E., HULO, C., LE MERCIER, P., BARTENSCHLAGER, R., DIEPOLDER, H., MORADPOUR, D., PAWLOTSKY, J. M., RICE, C. M., TREPO, C., PENIN, F. & DELEAGE, G. 2007. euHCVdb: the European hepatitis C virus database. *Nucleic Acids Research,* 35**,** D363-6.

COMBET, C., PENIN, F., GEOURJON, C. & DELEAGE, G. 2004. HCVDB: hepatitis C virus sequences database. *Appl Bioinformatics,* 3**,** 237-40.

DAMOC, E., FRASER, C. S., ZHOU, M., VIDELER, H., MAYEUR, G. L., HERSHEY, J. W. B., DOUDNA, J. A., ROBINSON, C. V. & LEARY, J. A. 2007. Structural characterization of the human eukaryotic initiation factor 3 protein complex by mass spectrometry. *Molecular & Cellular Proteomics,* 6**,** 1135-1146.

DIAZ-TOLEDANO, R., ARIZA-MATEOS, A., BIRK, A., MARTINEZ-GARCIA, B. & GOMEZ, J. 2009. In vitro characterization of a miR-122-sensitive double-helical switch element in the 5 ' region of hepatitis C virus RNA. *Nucleic Acids Research,* 37**,** 5498-5510.

DIBROV, S. M., DING, K., BRUNN, N. D., PARKER, M. A., BERGDAHL, B. M., WYLES, D. L. & HERMANN, T. 2012. Structure of a hepatitis C virus RNA domain in complex with a translation inhibitor reveals a binding mode reminiscent of riboswitches. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 5223-5228.

DIBROV, S. M., JOHNSTON-COX, H., WENG, Y. H. & HERMANN, T. 2007. Functional architecture of HCVIRES domain II stabilized by divalent metal ions in the crystal and in solution. *Angewandte Chemie-International Edition,* 46**,** 226-229.

DIENSTAG, J. L. & MCHUTCHISON, J. G. 2006. American Gastroenterological Association technical review on the management of hepatitis C. *Gastroenterology,* 130**,** 231-264.

DMITRIEV, S. E., TERENIN, I. M., ANDREEV, D. E., IVANOV, P. A., DUNAEVSKY, J. E., MERRICK, W. C. & SHATSKY, I. N. 2010. GTP-independent tRNA Delivery to the Ribosomal P-site by a Novel Eukaryotic Translation Factor. *Journal of Biological Chemistry,* 285**,** 26779-26787.

DOS REIS, M., SAVVA, R. & WERNISCH, L. 2004. Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Research,* 32**,** 5036-44.

EASTON, L. E., LOCKER, N. & LUKAVSKY, P. J. 2009. Conserved functional domains and a novel tertiary interaction near the pseudoknot drive translational activity of hepatitis C virus and hepatitis C virus-like internal ribosome entry sites. *Nucleic Acids Research,* 37**,** 5537-5549.

EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research,* 32**,** 1792-1797.

EDGAR, R. C. & BATZOGLOU, S. 2006. Multiple sequence alignment. *Current Opinion in Structural Biology,* 16**,** 368-373.

EL AWADY, M. K., AZZAZY, H. M., FAHMY, A. M., SHAWKY, S. M., BADRELDIN, N. G., YOSSEF, S. S., OMRAN, M. H., ZEKRI, A. R. N. & GOUELI, S. A. 2009. Positional effect of mutations in 5'UTR of hepatitis C virus 4a on patients' response to therapy. *World Journal of Gastroenterology,* 15**,** 1480-1486.

FELD, J. J. & HOOFNAGLE, J. H. 2005. Mechanism of action of interferon and ribavirin in treatment of hepatitis C. *Nature,* 436**,** 967-972.

FERENCI, P., KOZBIAL, K., MANDORFER, M. & HOFER, H. 2015. HCV targeting of patients with cirrhosis. *Journal of Hepatology,* 63**,** 1015-22.

FILBIN, M. E. & KIEFT, J. S. 2011. HCV IRES domain IIb affects the configuration of coding RNA in the 40S subunit's decoding groove. *RNA-a Publication of the RNA Society,* 17**,** 1258-1273.

FILBIN, M. E., VOLLMAR, B. S., SHI, D., GONEN, T. & KIEFT, J. S. 2012. HCV IRES manipulates the ribosome to promote the switch from translation initiation to elongation. *Nature Structural & Molecular Biology*.

FINN, R. D., COGGILL, P., EBERHARDT, R. Y., EDDY, S. R., MISTRY, J., MITCHELL, A. L., POTTER, S. C., PUNTA, M., QURESHI, M., SANGRADOR-VEGAS, A., SALAZAR, G. A., TATE, J. & BATEMAN, A. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research,* 44**,** D279-85.

FORTON, D. M., KARAYIANNIS, P., MAHMUD, N., TAYLOR-ROBINSON, S. D. & THOMAS, H. C. 2004. Identification of unique hepatitis C virus quasispecies in the central nervous system and comparative analysis of internal translational efficiency of brain, liver, and serum variants. *Journal of Virology,* 78**,** 5170-5183.

FRASER, C. S., HERSHEY, J. W. B. & DOUDNA, J. A. 2009. The pathway of hepatitis C virus mRNA recruitment to the human ribosome. *Nature Structural & Molecular Biology,* 16**,** 397-404.

FRIED, M. W., BUTI, M., DORE, G. J., FLISIAK, R., FERENCI, P., JACOBSON, I., MARCELLIN, P., MANNS, M., NIKITIN, I., POORDAD, F., SHERMAN, M., ZEUZEM, S., SCOTT, J., GILLES, L., LENZ, O., PEETERS, M., SEKAR, V., DE SMEDT, G. & BEUMONT-MAUVIEL, M. 2013. Once-Daily Simeprevir (TMC435) With Pegylated Interferon and Ribavirin in Treatment-Naive Genotype 1 Hepatitis C: The Randomized PILLAR Study. *Hepatology,* 58**,** 1918-1929.

FUCHS, G., PETROV, A. N., MARCEAU, C. D., POPOV, L. M., CHEN, J., O'LEARY, S. E., WANG, R., CARETTE, J. E., SARNOW, P. & PUGLISI, J. D. 2015. Kinetic pathway of 40S ribosomal subunit recruitment to hepatitis C virus internal ribosome entry site. *Proc Natl Acad Sci U S A,* 112**,** 319-25.

FUKUSHI, S., OKADA, M., KAGEYAMA, T., HOSHINO, F. B. & KATAYAMA, K. 1999. Specific interaction of a 25-kilodalton cellular protein, a 40S ribosomal subunit protein, with the internal ribosome entry site of hepatitis C virus genome. *Virus Genes,* 19**,** 153-161.

FUKUSHI, S., OKADA, M., STAHL, J., KAGEYAMA, T., HOSHINO, F. B. & KATAYAMA, K. 2001. Ribosomal protein S5 interacts with the internal ribosomal entry site of hepatitis C virus. *Journal of Biological Chemistry,* 276**,** 20824-20826.

GALLAY, P. A. & LIN, K. 2013. Profile of alisporivir and its potential in the treatment of hepatitis C. *Drug Des Devel Ther,* 7**,** 105-15.

GANE, E. J., STEDMAN, C. A., HYLAND, R. H., DING, X., SVAROVSKAIA, E., SYMONDS, W. T., HINDES, R. G. & BERREY, M. M. 2013. Nucleotide polymerase inhibitor sofosbuvir plus ribavirin for hepatitis C. *N Engl J Med,* 368**,** 34-44.

GARAIGORTA, U. & CHISARI, F. V. 2009. Hepatitis C Virus Blocks Interferon Effector Function by Inducing Protein Kinase R Phosphorylation. *Cell Host & Microbe,* 6**,** 513-522.

GAWLIK, K. & GALLAY, P. A. 2014. HCV core protein and virus assembly: what we know without structures. *Immunol Res,* 60**,** 1-10.

GENTILE, I., BORGIA, F., COPPOLA, N., BUONOMO, A. R., CASTALDO, G. & BORGIA, G. 2014. Daclatasvir: the first of a new class of drugs targeted against hepatitis C virus NS5A. *Curr Med Chem,* 21**,** 1391-404.

GNANESHAN, S., IJAZ, S., MORAN, J., RAMSAY, M. & GREEN, J. 2007. HepSEQ: International Public Health Repository for Hepatitis B. *Nucleic Acids Research,* 35**,** D367-70.

GOERGEN, D. & NIEPMANN, M. 2012. Stimulation of Hepatitis C Virus RNA translation by microRNA-122 occurs under different conditions in vivo and in vitro. *Virus Research,* 167**,** 343-352.

GRAVITZ, L. 2011. A smouldering public-health crisis. *Nature,* 474**,** S2-S4.

GROTH, P., KALEV, I., KIROV, I., TRAIKOV, B., LESER, U. & WEISS, B. 2010. Phenoclustering: online mining of cross-species phenotypes. *Bioinformatics,* 26**,** 1924-5.

GUTIERREZ, J. A., CARRION, A. F., AVALOS, D., O'BRIEN, C., MARTIN, P., BHAMIDIMARRI, K. R. & PEYTON, A. 2015. Sofosbuvir and simeprevir for treatment of hepatitis C virus infection in liver transplant recipients. *Liver Transpl,* 21**,** 823-30.

HARRIS, T. W., ANTOSHECHKIN, I., BIERI, T., BLASIAR, D., CHAN, J., CHEN, W. J., DE LA CRUZ, N., DAVIS, P., DUESBURY, M., FANG, R., FERNANDES, J., HAN, M., KISHORE, R., LEE, R., MULLER, H. M., NAKAMURA, C., OZERSKY, P., PETCHERSKI, A., RANGARAJAN, A., ROGERS, A., SCHINDELMAN, G., SCHWARZ, E. M., TULI, M. A., VAN AUKEN, K., WANG, D., WANG, X., WILLIAMS, G., YOOK, K., DURBIN, R., STEIN, L. D., SPIETH, J. & STERNBERG, P. W. 2010. WormBase: a comprehensive resource for nematode research. *Nucleic Acids Research,* 38**,** D463-7.

HASHEM, Y., DES GEORGES, A., DHOTE, V., LANGLOIS, R., LIAO, H. Y., GRASSUCCI, R. A., HELLEN, C. U., PESTOVA, T. V. & FRANK, J. 2013a. Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. *Cell,* 153**,** 1108-19.

HASHEM, Y., DES GEORGES, A., DHOTE, V., LANGLOIS, R., LIAO, H. Y., GRASSUCCI, R. A., PESTOVA, T. V., HELLEN, C. U. T. & FRANK, J. 2013b. Hepatitis-C-virus-like internal ribosome entry sites displace eIF3 to gain access to the 40S subunit. *Nature,* advance online publication.

HAWKINS, T. B., DANTZER, J., PETERS, B., DINAUER, M., MOCKAITIS, K., MOONEY, S. & CORNETTA, K. 2011. Identifying viral integration sites using SeqMap 2.0. *Bioinformatics,* 27**,** 720-2.

HELLEN, C. U. & DE BREYNE, S. 2007. A distinct group of hepacivirus/pestivirus-like internal ribosomal entry sites in members of diverse picornavirus genera: evidence

for modular exchange of functional noncoding RNA elements by recombination. *Journal of Virology,* 81**,** 5850-63.

HELLEN, C. U. T. 2009. IRES-induced conformational changes in the ribosome and the mechanism of translation initiation by internal ribosomal entry. *Biochimica Et Biophysica Acta-Gene Regulatory Mechanisms,* 1789**,** 558-570.

HENKE, J. I., GOERGEN, D., ZHENG, J. F., SONG, Y. T., SCHUTTLER, C. G., FEHR, C., JUNEMANN, C. & NIEPMANN, M. 2008. microRNA-122 stimulates translation of hepatitis C virus RNA. *Embo Journal,* 27**,** 3300-3310.

HERTZ, M. I., LANDRY, D. M., WILLIS, A. E., LUO, G. & THOMPSON, S. R. 2013. Ribosomal protein S25 dependency reveals a common mechanism for diverse internal ribosome entry sites and ribosome shunting. *Molecular and Cellular Biology,* 33**,** 1016-26.

HONDA, M., BEARD, M. R., PING, L. H. & LEMON, S. M. 1999a. 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. *Journal of Virology,* 73**,** 1165-1174.

HONDA, M., BROWN, E. A. & LEMON, S. M. 1996a. Stability of a stem-loop involving the initiator AUG controls the efficiency of internal initiation of translation on hepatitis C virus RNA. *RNA-a Publication of the RNA Society,* 2**,** 955-968.

HONDA, M., PING, L. H., RIJNBRAND, R. C. A., AMPHLETT, E., CLARKE, B., ROWLANDS, D. & LEMON, S. M. 1996b. Structural requirements for initiation of translation by internal ribosome entry within genome-length hepatitis C virus RNA. *Virology,* 222**,** 31-42.

HONDA, M., RIJNBRAND, R., ABELL, G., KIM, D. S. & LEMON, S. M. 1999b. Natural variation in translational activities of the 5 ' nontranslated RNAs of hepatitis C virus genotypes 1a and 1b: Evidence for a long-range RNA-RNA interaction outside of the internal ribosomal entry site. *Journal of Virology,* 73**,** 4941-4951.

HOWE, D., COSTANZO, M., FEY, P., GOJOBORI, T., HANNICK, L., HIDE, W., HILL, D. P., KANIA, R., SCHAEFFER, M., ST PIERRE, S., TWIGGER, S., WHITE, O. & RHEE, S. Y. 2008. Big data: The future of biocuration. *Nature,* 455**,** 47-50.

HRABER, P. T., LEACH, R. W., REILLY, L. P., THURMOND, J., YUSIM, K. & KUIKEN, C. 2007. Los Alamos hepatitis C virus sequence and human immunology databases: an expanding resource for antiviral research. *Antivir Chem Chemother,* 18**,** 113-23.

IVANYI-NAGY, R., KANEVSKY, I., GABUS, C., LAVERGNE, J. P., FICHEUX, D., PENIN, F., FOSSE, P. & DARLIX, J. L. 2006. Analysis of hepatitis C virus RNA dimerization and core-RNA interactions. *Nucleic Acids Research,* 34**,** 2618-33.

JI, H., FRASER, C. S., YU, Y. H., LEARY, J. & DOUDNA, J. A. 2004. Coordinated assembly of human translation initiation complexes by the hepatitis C virus internal ribosome entry site RNA. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 16990-16995.

JOSEPH, A. P., BHAT, P., DAS, S. & SRINIVASAN, N. 2014. Re-analysis of cryoEM data on HCV IRES bound to 40S subunit of human ribosome integrated with recent structural information suggests new contact regions between ribosomal proteins and HCV RNA. *RNA Biology,* 11**,** 891-905.

JUBIN, R., VANTUNO, N. E., KIEFT, J. S., MURRAY, M. G., DOUDNA, J. A., LAU, J. Y. N. & BAROUDY, B. M. 2000. Hepatitis C virus internal ribosome entry site (IRES) stem loop IIId contains a phylogenetically conserved GGG triplet essential for translation and IRES folding. *Journal of Virology,* 74**,** 10430-10437.

KALLIAMPAKOU, K. I., PSARIDI-LINARDAKI, L. & MAVROMARA, P. 2002. Mutational analysis of the apical region of domain II of the HCVIRES. *Febs Letters,* 511**,** 79-84.

KATO, N., NAKAMURA, T., DANSAKO, H., NAMBA, K., ABE, K., NOZAKI, A., NAKA, K., IKEDA, M. & SHIMOTOHNO, K. 2005. Genetic variation and dynamics of hepatitis C virus replicons in long-term cell culture. *Journal of General Virology,* 86**,** 645-56.

KERRIEN, S., ARANDA, B., BREUZA, L., BRIDGE, A., BROACKES-CARTER, F., CHEN, C., DUESBURY, M., DUMOUSSEAU, M., FEUERMANN, M., HINZ, U., JANDRASITS, C., JIMENEZ, R. C., KHADAKE, J., MAHADEVAN, U., MASSON, P., PEDRUZZI, I., PFEIFFENBERGER, E., PORRAS, P., RAGHUNATH, A., ROECHERT, B., ORCHARD, S. & HERMJAKOB, H. 2012. The IntAct molecular interaction database in 2012. *Nucleic Acids Research,* 40**,** D841-6.

KHAWAJA, A., VOPALENSKY, V. & POSPISEK, M. 2015. Understanding the potential of hepatitis C virus internal ribosome entry site domains to modulate translation initiation via their structure and function. *Wiley Interdiscip Rev RNA,* 6**,** 211-24.

KIEFT, J. S., ZHOU, K. H., GRECH, A., JUBIN, R. & DOUDNA, J. A. 2002. Crystal structure of an RNA tertiary domain essential to HCVIRES-mediated translation initiation. *Nature Structural Biology,* 9**,** 370-374.

KIEFT, J. S., ZHOU, K. H., JUBIN, R. & DOUDNA, J. A. 2001. Mechanism of ribosome recruitment by hepatitis CIRES RNA. *RNA-a Publication of the RNA Society,* 7**,** 194-206.

KIEFT, J. S., ZHOU, K. H., JUBIN, R., MURRAY, M. G., LAU, J. Y. N. & DOUDNA, J. A. 1999. The hepatitis C virus internal ribosome entry site adopts an ion-dependent tertiary fold. *Journal of Molecular Biology,* 292**,** 513-529.

KIM, J. H., PARK, S. M., PARK, J. H., KEUM, S. J. & JANG, S. K. 2011. eIF2A mediates translation of hepatitis C viral mRNA under stress conditions. *Embo Journal,* 30**,** 2454-2464.

KIM, Y. K., LEE, S. H., KIM, C. S., SEOL, S. K. & JANG, S. K. 2003. Long-range RNA-RNA interaction between the 5 ' nontranslated region and the core-coding sequences of hepatitis C virus modulates the IRES-dependent translation. *RNA-a Publication of the RNA Society,* 9**,** 599-606.

KLINCK, R., WESTHOF, E., WALKER, S., AFSHAR, M., COLLIER, A. & ABOUL-ELA, F. 2000. A potential RNA drug target in the hepatitis C virus internal ribosomal entry site. *RNA-a Publication of the RNA Society,* 6**,** 1423-1431.

KOLUPAEVA, V. G., PESTOVA, T. V. & HELLEN, C. U. 2000a. An enzymatic footprinting analysis of the interaction of 40S ribosomal subunits with the internal ribosomal entry site of hepatitis C virus. *Journal of Virology,* 74**,** 6242-50.

KOLUPAEVA, V. G., PESTOVA, T. V. & HELLEN, C. U. T. 2000b. An enzymatic footprinting analysis of the interaction of 40S ribosomal subunits with the internal ribosomal entry site of hepatitis C virus. *Journal of Virology,* 74**,** 6242-6250.

KOLYKHALOV, A. A., AGAPOV, E. V., BLIGHT, K. J., MIHALIK, K., FEINSTONE, S. M. & RICE, C. M. 1997. Transmission of hepatitis C by intrahepatic inoculation with transcribed RNA. *Science,* 277**,** 570-4.

KOSUGE, T., MASHIMA, J., KODAMA, Y., FUJISAWA, T., KAMINUMA, E., OGASAWARA, O., OKUBO, K., TAKAGI, T. & NAKAMURA, Y. 2014. DDBJ progress report: a new submission system for leading to a correct annotation. *Nucleic Acids Research,* 42**,** D44-9.

KOZOMARA, A. & GRIFFITHS-JONES, S. 2011. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Research,* 39**,** D152-7.

KRISTINA STRANDBERG, A. K. & SALTER, L. A. 2004. A comparison of methods for estimating the transition:transversion ratio from DNA sequences. *Mol Phylogenet Evol,* 32**,** 495-503.

KUDLA, G., MURRAY, A. W., TOLLERVEY, D. & PLOTKIN, J. B. 2009. Coding-sequence determinants of gene expression in Escherichia coli. *Science,* 324**,** 255-8.

KUIKEN, C., HRABER, P., THURMOND, J. & YUSIM, K. 2008. The hepatitis C sequence database in Los Alamos. *Nucleic Acids Research,* 36**,** D512-6.

KUIKEN, C., KORBER, B. & SHAFER, R. W. 2003. HIV sequence databases. *AIDS Rev,* 5**,** 52-61.

KUIKEN, C., YUSIM, K., BOYKIN, L. & RICHARDSON, R. 2005. The Los Alamos hepatitis C sequence database. *Bioinformatics,* 21**,** 379-84.

KWOFIE, S. K., SCHAEFER, U., SUNDARARAJAN, V. S., BAJIC, V. B. & CHRISTOFFELS, A. 2011. HCVpro: hepatitis C virus protein interaction database. *Infect Genet Evol,* 11**,** 1971-7.

LALETINA, E., GRAIFER, D., MALYGIN, A., IVANOV, A., SHATSKY, I. & KARPOVA, G. 2006. Proteins surrounding hairpin IIIe of the hepatitis C virus internal ribosome entry site on the human 40S ribosomal subunit. *Nucleic Acids Research,* 34**,** 2027-2036.

LANCASTER, A. M., JAN, E. & SARNOW, P. 2006. Initiation factor-independent translation mediated by the hepatitis C virus internal ribosome entry site. *RNA-a Publication of the RNA Society,* 12**,** 894-902.

LANDRY, D. M., HERTZ, M. I. & THOMPSON, S. R. 2009. RPS25 is essential for translation initiation by the Dicistroviridae and hepatitis C viral IRESs. *Genes & Development,* 23**,** 2753-2764.

LAPORTE, J., MALET, I., ANDRIEU, T., THIBAULT, V., TOULME, J. J., WYCHOWSKI, C., PAWLOTSKY, J. M., HURAUX, J. M., AGUT, H. & CAHOUR, A. 2000. Comparative analysis of translation efficiencies of hepatitis C virus 5 ' untranslated regions among intraindividual quasispecies present in chronic infection: Opposite behaviors depending on cell type. *Journal of Virology,* 74**,** 10827-10833.

LAUNAY, G., SALZA, R., MULTEDO, D., THIERRY-MIEG, N. & RICARD-BLUM, S. 2015. MatrixDB, the extracellular matrix interaction database: updated content, a new navigator and expanded functionalities. *Nucleic Acids Research,* 43**,** D321-7.

LAVENDER, C. A., DING, F., DOKHOLYAN, N. V. & WEEKS, K. M. 2010. Robust and Generic RNA Modeling Using Inferred Constraints: A Structure for the Hepatitis C Virus IRES Pseudoknot Domain. *Biochemistry,* 49**,** 4931-4933.

LAWITZ, E., MANGIA, A., WYLES, D., RODRIGUEZ-TORRES, M., HASSANEIN, T., GORDON, S. C., SCHULTZ, M., DAVIS, M. N., KAYALI, Z., REDDY, K. R., JACOBSON, I. M., KOWDLEY, K. V., NYBERG, L., SUBRAMANIAN, G. M., HYLAND, R. H., ARTERBURN, S., JIANG, D. Y., MCNALLY, J., BRAINARD, D., SYMONDS, W. T., MCHUTCHISON, J. G., SHEIKH, A. M., YOUNOSSI, Z. & GANE, E. J. 2013. Sofosbuvir for Previously Untreated Chronic Hepatitis C Infection. *New England Journal of Medicine,* 368**,** 1878-1887.

LEE, J. H., PESTOVA, T. V., SHIN, B. S., CAO, C., CHOI, S. K. & DEVER, T. E. 2002. Initiation factor eIF5B catalyzes second GTP-dependent step in eukaryotic translation initiation. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 16689-16694.

LESTRADE, L. & WEBER, M. J. 2006. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Research,* 34**,** D158-62.

LI, Y., WANG, C., MIAO, Z., BI, X., WU, D., JIN, N., WANG, L., WU, H., QIAN, K., LI, C., ZHANG, T., ZHANG, C., YI, Y., LAI, H., HU, Y., CHENG, L., LEUNG, K. S., LI, X., ZHANG, F., LI, K. & WANG, D. 2015. ViRBase: a resource for virus-host ncRNA-associated interactions. *Nucleic Acids Research,* 43**,** D578-82.

LICATA, L., BRIGANTI, L., PELUSO, D., PERFETTO, L., IANNUCCELLI, M., GALEOTA, E., SACCO, F., PALMA, A., NARDOZZA, A. P., SANTONICO, E., CASTAGNOLI, L. & CESARENI, G. 2012. MINT, the molecular interaction database: 2012 update. *Nucleic Acids Research,* 40**,** D857-61.

LOCKER, N., EASTON, L. E. & LUKAVSKY, P. J. 2007. HCV and CSFVIRES domain II mediate eIF2 release during 80S ribosome assembly. *Embo Journal,* 26**,** 795-805.

LU, H., LI, W. Q., NOBLE, W. S., PAYAN, D. & ANDERSON, D. C. 2004. Riboproteomics of the hepatitis C virus internal ribosomal entry site. *Journal of Proteome Research,* 3**,** 949-957.

LU, M., WIESE, M. & ROGGENDORF, M. 1999. Selection of genetic variants of the 5' noncoding region of hepatitis C virus occurs only in patients responding to interferon alpha therapy. *Journal of Medical Virology,* 59**,** 146-53.

LUKAVSKY, P. J., KIM, I., OTTO, G. A. & PUGLISI, J. D. 2003. Structure of HCVIRES domain II determined by NMR. *Nature Structural Biology,* 10**,** 1033-1038.

LUKAVSKY, P. J., OTTO, G. A., LANCASTER, A. M., SARNOW, P. & PUGLISI, J. D. 2000. Structures of two RNA domains essential for hepatitis C virus internal ribosome entry site function. *Nature Structural Biology,* 7**,** 1105-1110.

LUTTERMANN, C. & MEYERS, G. 2009. The importance of inter- and intramolecular base pairing for translation reinitiation on a eukaryotic bicistronic mRNA. *Genes Dev,* 23**,** 331-44.

LYONS, A. J., LYTLE, J. R., GOMEZ, J. & ROBERTSON, H. D. 2001. Hepatitis C virus internal ribosome entry site RNA contains a tertiary structural element in a functional domain of stem-loop II. *Nucleic Acids Research,* 29**,** 2535-2541.

LYTLE, J. R., WU, L. & ROBERTSON, H. D. 2001. The ribosome binding site of hepatitis C virus mRNA. *Journal of Virology,* 75**,** 7629-7636.

LYTLE, J. R., WU, L. & ROBERTSON, H. D. 2002. Domains on the hepatitis C virus internal ribosome entry site for 40s subunit binding. *RNA-a Publication of the RNA Society,* 8**,** 1045-1055.

MALCOLM, B. A., ARASSAPPAN, A., BENNETT, F., BOGEN, S., CHASE, R., CHEN, K., CHEN, T., INGRAVALLO, P., JAO, E., KONG, S., LAHSER, F., LIU, R., LIU, Y. T., LOVEY, R., MCCORMICK, J., NJOROGE, G. F., SAKSANA, A., SKELTON, A., TONG, X., VENKATRAMAN, S., WRIGHT-MINOGUE, J., XIA, E. & GIRIJAVALLABHAN, V. 2005. SCH 503034, a mechanism-based inhibitor of hepatitis C virus (HCV) NS3 protease suppresses polyprotein maturation and enhances the antiviral activity of interferona-211 (INF). *Hepatology,* 42**,** 535a-536a.

MALET, I., BELNARD, M., AGUT, H. & CAHOUR, A. 2003. From RNA to quasispecies: a DNA polymerase with proofreading activity is highly recommended for accurate assessment of viral diversity. *Journal of Virological Methods,* 109**,** 161-70.

MALYGIN, A. A., KOSSINOVA, O. A., SHATSKY, I. N. & KARPOVA, G. G. 2013a. HCV IRES interacts with the 18S rRNA to activate the 40S ribosome for subsequent steps of translation initiation. *Nucleic Acids Research*.

MALYGIN, A. A., SHATSKY, I. N. & KARPOVA, G. G. 2013b. Proteins of the human 40S ribosomal subunit involved in hepatitis C IRES binding as revealed from fluorescent labeling. *Biochemistry (Mosc),* 78**,** 53-9.

MASEK, T., VOPALENSKY, V., HORVATH, O., VORTELOVA, L., FEKETOVA, Z. & POSPISEK, M. 2007. Hepatitis C virus internal ribosome entry site initiates protein synthesis at the authentic initiation codon in yeast. *Journal of General Virology,* 88**,** 1992-2002.

MASSON, P., HULO, C., DE CASTRO, E., BITTER, H., GRUENBAUM, L., ESSIOUX, L., BOUGUELERET, L., XENARIOS, I. & LE MERCIER, P. 2013. ViralZone: recent updates to the virus knowledge resource. *Nucleic Acids Research,* 41**,** D579-83.

MATSUDA, D. & MAURO, V. P. 2014. Base pairing between hepatitis C virus RNA and 18S rRNA is required for IRES-dependent translation initiation in vivo. *Proc Natl Acad Sci U S A,* 111**,** 15385-9.

MCINERNEY, P., ADAMS, P. & HADI, M. Z. 2014. Error Rate Comparison during Polymerase Chain Reaction by DNA Polymerase. *Mol Biol Int,* 2014**,** 287430.

MCLAUCHLAN, J. 2000. Properties of the hepatitis C virus core protein: a structural protein that modulates cellular processes. *J Viral Hepat,* 7**,** 2-14.

MELCHER, S. E., WILSON, T. J. & LILLEY, D. M. J. 2003. The dynamic nature of the four-way junction of the hepatitis C virus IRES. *RNA-a Publication of the RNA Society,* 9**,** 809-820.

METZKER, M. L. 2010. Sequencing technologies - the next generation. *Nat Rev Genet,* 11**,** 31-46.

MIZOKAMI, M., YOKOSUKA, O., TAKEHARA, T., SAKAMOTO, N., KORENAGA, M., MOCHIZUKI, H., NAKANE, K., ENOMOTO, H., IKEDA, F., YANASE, M., TOYODA, H., GENDA, T., UMEMURA, T., YATSUHASHI, H., IDE, T., TODA, N., NIREI, K., UENO, Y., NISHIGAKI, Y., BETULAR, J., GAO, B., ISHIZAKI, A., OMOTE, M., MO, H., GARRISON, K., PANG, P. S., KNOX, S. J., SYMONDS, W. T., MCHUTCHISON,

J. G., IZUMI, N. & OMATA, M. 2015. Ledipasvir and sofosbuvir fixed-dose combination with and without ribavirin for 12 weeks in treatment-naive and previously treated Japanese patients with genotype 1 hepatitis C: an open-label, randomised, phase 3 trial. *Lancet Infect Dis,* 15**,** 645-53.

MOTAZAKKER, M., PREIKSCHAT, P., ELLIOTT, J., SMITH, C. A., MILLS, P. R., OIEN, K., SPENCE, E., ELLIOTT, R. M. & MCCRUDEN, E. A. B. 2007. Translation efficiencies of the 5 '-untranslated region of genotypes 1a and 3a in hepatitis C infected patients. *Journal of Medical Virology,* 79**,** 259-269.

MULLAN, B., KENNY-WALSH, E., COLLINS, J. K., SHANAHAN, F. & FANNING, L. J. 2001. Inferred hepatitis C virus quasispecies diversity is influenced by choice of DNA polymerase in reverse transcriptase-polymerase chain reactions. *Anal Biochem,* 289**,** 137-46.

NAGPAL, N., GOYAL, S., WAHI, D., JAIN, R., JAMAL, S., SINGH, A., RANA, P. & GROVER, A. 2015. Molecular principles behind Boceprevir resistance due to mutations in hepatitis C NS3/4A protease. *Gene,* 570**,** 115-21.

NAKAI, K., OKAMOTO, T., KIMURA-SOMEYA, T., ISHII, K., LIM, C. K., TANI, H., MATSUO, E., ABE, T., MORI, Y., SUZUKI, T., MIYAMURA, T., NUNBERG, J. H., MORIISHI, K. & MATSUURA, Y. 2006. Oligomerization of hepatitis C virus core protein is crucial for interaction with the cytoplasmic domain of E1 envelope protein. *Journal of Virology,* 80**,** 11265-73.

NAVRATIL, V., DE CHASSEY, B., MEYNIEL, L., DELMOTTE, S., GAUTIER, C., ANDRE, P., LOTTEAU, V. & RABOURDIN-COMBE, C. 2009. VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks. *Nucleic Acids Research,* 37**,** D661-8.

NEUKAM, K., MUNTEANU, D. I., RIVERO-JUAREZ, A., LUTZ, T., FEHR, J., MANDORFER, M., BHAGANI, S., LOPEZ-CORTES, L. F., HABERL, A., STOECKLE, M., MARQUEZ, M., SCHOLTEN, S., DE LOS SANTOS-GIL, I., MAUSS, S., RIVERO, A., COLLADO, A., DELGADO, M., ROCKSTROH, J. K. & PINEDA, J. A. 2015. Boceprevir or Telaprevir Based Triple Therapy against Chronic Hepatitis C in HIV Coinfection: Real-Life Safety and Efficacy. *Plos One,* 10**,** e0125080.

ODREMAN-MACCHIOLI, F., BARALLE, F. E. & BURATTI, E. 2001. Mutational analysis of the different bulge regions of hepatitis C virus domain II and their influence on internal ribosome entry site translational ability. *Journal of Biological Chemistry,* 276**,** 41648-41655.

OGATA, K., KASHIWAGI, T., IWAHASHI, J., HARA, K., HONDA, H., IDE, T., KUMASHIRO, R., KOHARA, M., SATA, M., WATANABE, H. & HAMADA, N. 2008. A mutational shift from domain III to II in the internal ribosome entry site of hepatitis C virus after interferon-ribavirin therapy. *Archives of Virology,* 153**,** 1575-1579.

OKUBO, K., SUGAWARA, H., GOJOBORI, T. & TATENO, Y. 2006. DDBJ in preparation for overview of research activities behind data submissions. *Nucleic Acids Research,* 34**,** D6-9.

OLSEN, L. R., ZHANG, G. L., REINHERZ, E. L. & BRUSIC, V. 2011. FLAVIdB: A data mining system for knowledge discovery in flaviviruses with direct applications in immunology and vaccinology. *Immunome Res,* 7.

ORCHARD, S. 2012. Molecular interaction databases. *Proteomics,* 12**,** 1656-62.

ORCHARD, S., KERRIEN, S., ABBANI, S., ARANDA, B., BHATE, J., BIDWELL, S., BRIDGE, A., BRIGANTI, L., BRINKMAN, F. S., CESARENI, G., CHATR-ARYAMONTRI, A., CHAUTARD, E., CHEN, C., DUMOUSSEAU, M., GOLL, J., HANCOCK, R. E., HANNICK, L. I., JURISICA, I., KHADAKE, J., LYNN, D. J., MAHADEVAN, U., PERFETTO, L., RAGHUNATH, A., RICARD-BLUM, S., ROECHERT, B., SALWINSKI, L., STUMPFLEN, V., TYERS, M., UETZ, P., XENARIOS, I. & HERMJAKOB, H. 2012. Protein interaction data curation: the International Molecular Exchange (IMEx) consortium. *Nat Methods,* 9**,** 345-50.

OTTO, G. A., LUKAVSKY, P. J., LANCASTER, A. M., SARNOW, P. & PUGLISI, J. D. 2002. Ribosomal proteins mediate the hepatitis C virus IRES-HeLa 40S interaction. *RNA-a Publication of the RNA Society,* 8**,** 913-923.

OTTO, G. A. & PUGLISI, J. D. 2004. The pathway of HCVIRES-mediated translation initiation. *Cell,* 119**,** 369-380.

PARSONS, J., CASTALDI, M. P., DUTTA, S., DIBROV, S. M., WYLES, D. L. & HERMANN, T. 2009. Conformational inhibition of the hepatitis C virus internal ribosome entry site RNA. *Nature Chemical Biology,* 5**,** 823-825.

PAULSEN, R. B., SETH, P. P., SWAYZE, E. E., GRIFFEY, R. H., SKALICKY, J. J., CHEATHAM, T. E. & DAVIS, D. R. 2010. Inhibitor-induced structural change in the HCV IRES domain IIa RNA. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 7263-7268.

PAWLOTSKY, J. M. 2011. Treatment failure and resistance with direct-acting antiviral drugs against hepatitis C virus. *Hepatology,* 53**,** 1742-51.

PAWLOTSKY, J. M. 2014. New hepatitis C virus (HCV) drugs and the hope for a cure: concepts in anti-HCV drug development. *Semin Liver Dis,* 34**,** 22-9.

PERARD, J., LEYRAT, C., BAUDIN, F., DROUET, E. & JAMIN, M. 2013. Structure of the full-length HCV IRES in solution. *Nature Communications,* 4.

PERARD, J., RASIA, R., MEDENBACH, J., AYALA, I., BOISBOUVIER, J., DROUET, E. & BAUDIN, F. 2009. Human initiation factor eIF3 subunit b interacts with HCV IRES RNA through its N-terminal RNA recognition motif. *Febs Letters,* 583**,** 70-74.

PERNI, R. B., ALMQUIST, S. J., BYRN, R. A., CHANDORKAR, G., CHATURVEDI, P. R., COURTNEY, L. F., DECKER, C. J., DINEHART, K., GATES, C. A., HARBESON, S. L., HEISER, A., KALKERI, G., KOLACZKOWSKI, E., LIN, K., LUONG, Y. P., RAO, B. G., TAYLOR, W. P., THOMSON, J. A., TUNG, R. D., WEI, Y. Y., KWONG, A. D. & LIN, C. 2006. Preclinical profile of VX-950, a potent, selective, and orally bioavailable inhibitor of hepatitis C virus NS3-4A serine protease. *Antimicrobial Agents and Chemotherapy,* 50**,** 899-909.

PESTOVA, T. V., DE BREYNE, S., PISAREV, A. V., ABAEVA, I. S. & HELLEN, C. U. 2008. eIF2-dependent and eIF2-independent modes of initiation on the CSFVIRES: a common role of domain II. *Embo Journal,* 27**,** 1060-1072.

PESTOVA, T. V., LOMAKIN, I. B., LEE, J. H., CHOI, S. K., DEVER, T. E. & HELLEN, C. U. T. 2000. The joining of ribosomal subunits in eukaryotes requires eIF5B. *Nature,* 403**,** 332-335.

PESTOVA, T. V., SHATSKY, I. N., FLETCHER, S. P., JACKSON, R. J. & HELLEN, C. U. 1998. A prokaryotic-like mode of cytoplasmic eukaryotic ribosome binding to the initiation codon during internal translation initiation of hepatitis C and classical swine fever virus RNAs. *Genes Dev,* 12**,** 67-83.

PICKETT, B. E., GREER, D. S., ZHANG, Y., STEWART, L., ZHOU, L., SUN, G., GU, Z., KUMAR, S., ZAREMBA, S., LARSEN, C. N., JEN, W., KLEM, E. B. & SCHEUERMANN, R. H. 2012a. Virus pathogen database and analysis resource (ViPR): a comprehensive bioinformatics database and analysis resource for the coronavirus research community. *Viruses,* 4**,** 3209-26.

PICKETT, B. E., SADAT, E. L., ZHANG, Y., NORONHA, J. M., SQUIRES, R. B., HUNT, V., LIU, M., KUMAR, S., ZAREMBA, S., GU, Z., ZHOU, L., LARSON, C. N., DIETRICH, J., KLEM, E. B. & SCHEUERMANN, R. H. 2012b. ViPR: an open bioinformatics database and analysis resource for virology research. *Nucleic Acids Research,* 40**,** D593-8.

PINERO, J., QUERALT-ROSINACH, N., BRAVO, A., DEU-PONS, J., BAUER-MEHREN, A., BARON, M., SANZ, F. & FURLONG, L. I. 2015. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes. *Database (Oxford),* 2015**,** bav028.

POLYAK, S. J., SULLIVAN, D. G., AUSTIN, M. A., DAI, J. Y., SHUHART, M. C., LINDSAY, K. L., BONKOVSKY, H. L., DI BISCEGLIE, A. M., LEE, W. M., MORISHIMA, C. &

GRETCH, D. R. 2005. Comparison of amplification enzymes for Hepatitis C Virus quasispecies analysis. *Virology Journal,* 2**,** 41.

POWDRILL, M. H., TCHESNOKOV, E. P., KOZAK, R. A., RUSSELL, R. S., MARTIN, R., SVAROVSKAIA, E. S., MO, H., KOUYOS, R. D. & GOTTE, M. 2011. Contribution of a mutational bias in hepatitis C virus replication to the genetic barrier in the development of drug resistance. *Proc Natl Acad Sci U S A,* 108**,** 20509-13.

PRUITT, K. D., BROWN, G. R., HIATT, S. M., THIBAUD-NISSEN, F., ASTASHYN, A., ERMOLAEVA, O., FARRELL, C. M., HART, J., LANDRUM, M. J., MCGARVEY, K. M., MURPHY, M. R., O'LEARY, N. A., PUJAR, S., RAJPUT, B., RANGWALA, S. H., RIDDICK, L. D., SHKEDA, A., SUN, H., TAMEZ, P., TULLY, R. E., WALLIN, C., WEBB, D., WEBER, J., WU, W., DICUCCIO, M., KITTS, P., MAGLOTT, D. R., MURPHY, T. D. & OSTELL, J. M. 2014. RefSeq: an update on mammalian reference sequences. *Nucleic Acids Research,* 42**,** D756-63.

PSARIDI, L., GEORGOPOULOU, U., VARAKLIOTI, A. & MAVROMARA, P. 1999. Mutational analysis of a conserved tetraloop in the 5 ' untranslated region of hepatitis C virus identifies a novel RNA element essential for the internal ribosome entry site function. *Febs Letters,* 453**,** 49-53.

QUADE, N., BOEHRINGER, D., LEIBUNDGUT, M., VAN DEN HEUVEL, J. & BAN, N. 2015. Cryo-EM structure of Hepatitis C virus IRES bound to the human ribosome at 3.9-A resolution. *Nat Commun,* 6**,** 7646.

QUEK, X. C., THOMSON, D. W., MAAG, J. L., BARTONICEK, N., SIGNAL, B., CLARK, M. B., GLOSS, B. S. & DINGER, M. E. 2015. lncRNAdb v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Research,* 43**,** D168-73.

REYNOLDS, J. E., KAMINSKI, A., CARROLL, A. R., CLARKE, B. E., ROWLANDS, D. J. & JACKSON, R. J. 1996. Internal initiation of translation of hepatitis C virus RNA: The ribosome entry site is at the authentic initiation codon. *RNA-a Publication of the RNA Society,* 2**,** 867-878.

REYNOLDS, J. E., KAMINSKI, A., KETTINEN, H. J., GRACE, K., CLARKE, B. E., CARROLL, A. R., ROWLANDS, D. J. & JACKSON, R. J. 1995. Unique Features of Internal Initiation of Hepatitis-C Virus-RNA Translation. *Embo Journal,* 14**,** 6010-6020.

RIBEIRO, R. M., LI, H., WANG, S., STODDARD, M. B., LEARN, G. H., KORBER, B. T., BHATTACHARYA, T., GUEDJ, J., PARRISH, E. H., HAHN, B. H., SHAW, G. M. & PERELSON, A. S. 2012. Quantifying the diversification of hepatitis C virus (HCV) during primary infection: estimates of the in vivo mutation rate. *Plos Pathogens,* 8**,** e1002881.

RIJNBRAND, R., BREDENBEEK, P., VANDERSTRAATEN, T., WHETTER, L., INCHAUSPE, G., LEMON, S. & SPAAN, W. 1995. Almost the Entire 5' Non-Translated Region of Hepatitis-C Virus Is Required for Cap-Independent Translation. *Febs Letters,* 365**,** 115-119.

RIJNBRAND, R., BREDENBEEK, P. J., HAASNOOT, P. C., KEIFT, J. S., SPAAN, W. J. M. & LEMON, S. M. 2001. The influence of downstream protein-coding sequence on internal ribosome entry on hepatitis C virus and other flavivirus RNAs. *RNA-a Publication of the RNA Society,* 7**,** 585-597.

RIJNBRAND, R., THIVIYANATHAN, V., KALUARACHCHI, K., LEMON, S. M. & GORENSTEIN, D. G. 2004. Mutational and structural analysis of stem-loop IIIc of the hepatitis C virus and GB virus B internal ribosome entry sites. *Journal of Molecular Biology,* 343**,** 805-817.

RIJNBRAND, R. C. A., ABBINK, T. E. M., HAASNOOT, P. C. J., SPAAN, W. J. M. & BREDENBEEK, P. J. 1996. The influence of AUG codons in the hepatitis C virus 5' nontranslated region on translation and mapping of the translation initiation window. *Virology,* 226**,** 47-56.

ROBERT, F., KAPP, L. D., KHAN, S. N., ACKER, M. G., KOLITZ, S., KAZEMI, S., KAUFMAN, R. J., MERRICK, W. C., KOROMILAS, A. E., LORSCH, J. R. &

PELLETIER, J. 2006. Initiation of protein synthesis by hepatitis C virus is refractory to reduced eIF2 center dot GTP center dot Met-tRNA(i)(Met) ternary complex availability. *Molecular Biology of the Cell,* 17**,** 4632-4644.

ROCHE, B., COILLY, A., ROQUE-AFONSO, A. M. & SAMUEL, D. 2015. Interferon-Free Hepatitis C Treatment before and after Liver Transplantation: The Role of HCV Drug Resistance. *Viruses,* 7**,** 5155-68.

ROLL-MECAK, A., CAO, C., DEVER, T. E. & BURLEY, S. K. 2000. X-ray structures of the universal translation initiation factor IF2/eIF5B: Conformational changes on GDP and GTP binding. *Cell,* 103**,** 781-792.

RUSINOVA, I., FORSTER, S., YU, S., KANNAN, A., MASSE, M., CUMMING, H., CHAPMAN, R. & HERTZOG, P. J. 2013. Interferome v2.0: an updated database of annotated interferon-regulated genes. *Nucleic Acids Research,* 41**,** D1040-6.

SABI, R. & TULLER, T. 2014. Modelling the efficiency of codon-tRNA interactions based on codon usage bias. *DNA Res,* 21**,** 511-26.

SAIZ, J. C., LOPEZ DE QUINTO, S., IBARROLA, N., LOPEZ-LABRADOR, F. X., SANCHEZ-TAPIAS, J. M., RODES, J. & MARTINEZ-SALAS, E. 1999. Internal initiation of translation efficiency in different hepatitis C genotypes isolated from interferon treated patients. *Archives of Virology,* 144**,** 215-29.

SALWINSKI, L., MILLER, C. S., SMITH, A. J., PETTIT, F. K., BOWIE, J. U. & EISENBERG, D. 2004. The Database of Interacting Proteins: 2004 update. *Nucleic Acids Research,* 32**,** D449-51.

SETH, P. P., MIYAJI, A., JEFFERSON, E. A., SANNES-LOWERY, K. A., OSGOOD, S. A., PROPP, S. S., RANKEN, R., MASSIRE, C., SAMPATH, R., ECKER, D. J., SWAYZE, E. E. & GRIFFEY, R. H. 2005. SAR by MS: Discovery of a new class of RNA-binding small molecules for the hepatitis C virus: Internal ribosome entry site IIA subdomain. *Journal of Medicinal Chemistry,* 48**,** 7099-7102.

SHARMA, D., PRIYADARSHINI, P. & VRATI, S. 2015. Unraveling the web of viroinformatics: computational tools and databases in virus research. *Journal of Virology,* 89**,** 1489-501.

SHI, G., YAGYU, F., SHIMIZU, Y., SHIMIZU, K., OSHIMA, M., IWAMOTO, A., GAO, B., LIU, W., GAO, G. F. & KITAMURA, Y. 2011. Flow cytometric assay using two fluorescent proteins for the function of the internal ribosome entry site of hepatitis C virus. *Cytometry A,* 79**,** 653-60.

SHIMOIKE, T., MCKENNA, S. A., LINDHOUT, D. A. & PUGLISI, J. D. 2009. Translational insensitivity to potent activation of PKR by HCV IRES RNA. *Antiviral Research,* 83**,** 228-237.

SHIMOIKE, T., MIMORI, S., TANI, H., MATSUURA, Y. & MIYAMURA, T. 1999. Interaction of hepatitis C virus core protein with viral sense RNA and suppression of its translation. *Journal of Virology,* 73**,** 9718-25.

SHIN, I. T., TANAKA, Y., TATENO, Y. & MIZOKAMI, M. 2008. Development and public release of a comprehensive hepatitis virus database. *Hepatology Research,* 38**,** 234-43.

SIGRIST, C. J., DE CASTRO, E., CERUTTI, L., CUCHE, B. A., HULO, N., BRIDGE, A., BOUGUELERET, L. & XENARIOS, I. 2013. New and continuing developments at PROSITE. *Nucleic Acids Research,* 41**,** D344-7.

SIMMONDS, P., BUKH, J., COMBET, C., DELEAGE, G., ENOMOTO, N., FEINSTONE, S., HALFON, P., INCHAUSPE, G., KUIKEN, C., MAERTENS, G., MIZOKAMI, M., MURPHY, D. G., OKAMOTO, H., PAWLOTSKY, J. M., PENIN, F., SABLON, E., SHIN, I. T., STUYVER, L. J., THIEL, H. J., VIAZOV, S., WEINER, A. J. & WIDELL, A. 2005. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology,* 42**,** 962-73.

SIRIDECHADILOK, B., FRASER, C. S., HALL, R. J., DOUDNA, J. A. & NOGALES, E. 2005. Structural roles for human translation factor eIF3 in initiation of protein synthesis. *Science,* 310**,** 1513-1515.

SIZOVA, D. V., KOLUPAEVA, V. G., PESTOVA, T. V., SHATSKY, I. N. & HELLEN, C. U. 1998. Specific interaction of eukaryotic translation initiation factor 3 with the 5' nontranslated regions of hepatitis C virus and classical swine fever virus RNAs. *Journal of Virology,* 72**,** 4775-82.

SOFIA, M. J., BAO, D., CHANG, W., DU, J. F., NAGARATHNAM, D., RACHAKONDA, S., REDDY, P. G., ROSS, B. S., WANG, P. Y., ZHANG, H. R., BANSAL, S., ESPIRITU, C., KEILMAN, M., LAM, A. M., STEUER, H. M. M., NIU, C. R., OTTO, M. J. & FURMAN, P. A. 2010. Discovery of a beta-D-2 '-Deoxy-2 '-alpha-fluoro-2 '-beta-C-methyluridine Nucleotide Prodrug (PSI-7977) for the Treatment of Hepatitis C Virus. *Journal of Medicinal Chemistry,* 53**,** 7202-7218.

SOLER, M., PELLERIN, M., MALNOU, C. E., DHUMEAUX, D., KEAN, K. M. & PAWLOTSKY, J. M. 2002. Quasispecies heterogeneity and constraints on the evolution of the 5 ' noncoding region of hepatitis C virus (HCV): Relationship with HCV resistance to interferon-alpha therapy. *Virology,* 298**,** 160-173.

SPAHN, C. M., JAN, E., MULDER, A., GRASSUCCI, R. A., SARNOW, P. & FRANK, J. 2004. Cryo-EM visualization of a viral internal ribosome entry site bound to human ribosomes: the IRES functions as an RNA-based translation factor. *Cell,* 118**,** 465-75.

SPAHN, C. M. T., KIEFT, J. S., GRASSUCCI, R. A., PENCZEK, P. A., ZHOU, K. H., DOUDNA, J. A. & FRANK, J. 2001. Hepatitis C virus IRES RNA-induced changes in the conformation of the 40S ribosomal subunit. *Science,* 291**,** 1959-1962.

STEIN, L. D. 2003. Integrating biological databases. *Nat Rev Genet,* 4**,** 337-45.

STEWART, H., BINGHAM, R. J., WHITE, S. J., DYKEMAN, E. C., ZOTHNER, C., TUPLIN, A. K., STOCKLEY, P. G., TWAROCK, R. & HARRIS, M. 2016. Identification of novel RNA secondary structures within the hepatitis C virus genome reveals a cooperative involvement in genome packaging. *Sci Rep,* 6**,** 22952.

SUN, C., QUEROL-AUDI, J., MORTIMER, S. A., ARIAS-PALOMO, E., DOUDNA, J. A., NOGALES, E. & CATE, J. H. 2013. Two RNA-binding motifs in eIF3 direct HCV IRES-dependent translation. *Nucleic Acids Research,* 41**,** 7512-21.

TAKAMIZAWA, A., MORI, C., FUKE, I., MANABE, S., MURAKAMI, S., FUJITA, J., ONISHI, E., ANDOH, T., YOSHIDA, I. & OKAYAMA, H. 1991. Structure and organization of the hepatitis C virus genome isolated from human carriers. *Journal of Virology,* 65**,** 1105-13.

TANG, S. X., COLLIER, A. J. & ELLIOTT, R. M. 1999. Alterations to both the primary and predicted secondary structure of stem-loop IIIc of the hepatitis C virus ib 5 ' untranslated region (5 ' UTR) lead to mutants severely defective in translation which cannot be complemented in trans by the wild-type 5 ' UTR sequence. *Journal of Virology,* 73**,** 2359-2364.

TERENIN, I. M., DMITRIEV, S. E., ANDREEV, D. E. & SHATSKY, I. N. 2008. Eukaryotic translation initiation machinery can operate in a bacterial-like mode without eIF2. *Nature Structural & Molecular Biology,* 15**,** 836-841.

THELU, M. A., DROUET, E., HILLERET, M. N. & ZARSKI, J. P. 2004. Lack of clinical significance of variability in the internal ribosome entry site of hepatitis C virus. *Journal of Medical Virology,* 72**,** 396-405.

THELU, M. A., LEROY, V., RAMZAN, M., DUFEU-DUCHESNE, T., MARCHE, P. & ZARSKI, J. P. 2007. IRES complexity before IFN-alpha treatment and evolution of the viral load at the early stage of treatment in peripheral blood mononuclear cells from chronic hepatitis C patients. *Journal of Medical Virology,* 79**,** 242-253.

TORONEY, R., NALLAGATLA, S. R., BOYER, J. A., CAMERON, C. E. & BEVILACQUA, P. C. 2010. Regulation of PKR by HCV IRES RNA: Importance of Domain II and NS5A. *Journal of Molecular Biology,* 400**,** 393-412.

TSUKIYAMAKOHARA, K., IIZUKA, N., KOHARA, M. & NOMOTO, A. 1992. Internal Ribosome Entry Site within Hepatitis-C Virus-RNA. *Journal of Virology,* 66**,** 1476-1483.

TULLER, T., WALDMAN, Y. Y., KUPIEC, M. & RUPPIN, E. 2010. Translation efficiency is determined by both codon bias and folding energy. *Proc Natl Acad Sci U S A,* 107**,** 3645-50.

UNBEHAUN, A., BORUKHOV, S. I., HELLEN, C. U. T. & PESTOVA, T. V. 2004. Release of initiation factors from 48S complexes during ribosomal subunit joining and the link between establishment of codon-anticodon base-pairing and hydrolysis of eIF2-bound GTP. *Genes & Development,* 18**,** 3078-3093.

UVERSKY, V. N. 2011. Intrinsically disordered proteins from A to Z. *Int J Biochem Cell Biol,* 43**,** 1090-103.

VAN LEEUWEN, H. C., REUSKEN, C. B. E. M., ROETEN, M., DALEBOUT, T. J., RIEZU-BOJ, J. I., RUIZ, J. & SPAAN, W. J. M. 2004. Evolution of naturally occurring 5 ' non-translated region variants of hepatitis C virus genotype 1b in selectable replicons. *Journal of General Virology,* 85**,** 1859-1866.

VASSILAKI, N., FRIEBE, P., MEULEMAN, P., KALLIS, S., KAUL, A., PARANHOS-BACCALA, G., LEROUX-ROELS, G., MAVROMARA, P. & BARTENSCHLAGER, R. 2008. Role of the Hepatitis C Virus Core+1 Open Reading Frame and Core cis-Acting RNA Elements in Viral RNA Translation and Replication. *Journal of Virology,* 82**,** 11503-11515.

VISWESWARAIAH, J., PITTMAN, Y., DEVER, T. E. & HINNEBUSCH, A. G. 2015. The beta-hairpin of 40S exit channel protein Rps5/uS7 promotes efficient and accurate translation initiation in vivo. *Elife,* 4**,** e07939.

VITA, R., OVERTON, J. A., GREENBAUM, J. A., PONOMARENKO, J., CLARK, J. D., CANTRELL, J. R., WHEELER, D. K., GABBARD, J. L., HIX, D., SETTE, A. & PETERS, B. 2015. The immune epitope database (IEDB) 3.0. *Nucleic Acids Research,* 43**,** D405-12.

VIZCAINO, J. A., COTE, R. G., CSORDAS, A., DIANES, J. A., FABREGAT, A., FOSTER, J. M., GRISS, J., ALPI, E., BIRIM, M., CONTELL, J., O'KELLY, G., SCHOENEGGER, A., OVELLEIRO, D., PEREZ-RIVEROL, Y., REISINGER, F., RIOS, D., WANG, R. & HERMJAKOB, H. 2013. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Research,* 41**,** D1063-9.

WAKELEY, J. 1996. The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. *Trends Ecol Evol,* 11**,** 158-62.

WALDMAN, Y. Y., TULLER, T., SHLOMI, T., SHARAN, R. & RUPPIN, E. 2010. Translation efficiency in humans: tissue specificity, global optimization and differences between developmental stages. *Nucleic Acids Research,* 38**,** 2964-74.

WANG, C., SARNOW, P. & SIDDIQUI, A. 1993. Translation of human hepatitis C virus RNA in cultured cells is mediated by an internal ribosome-binding mechanism. *Journal of Virology,* 67**,** 3338-44.

WANG, C., SARNOW, P. & SIDDIQUI, A. 1994a. A conserved helical element is essential for internal initiation of translation of hepatitis C virus RNA. *Journal of Virology,* 68**,** 7301-7.

WANG, C. Y., LE, S. Y., ALI, N. & SIDDIQUI, A. 1995. An RNA Pseudoknot Is an Essential Structural Element of the Internal Ribosome Entry Site Located within the Hepatitis-C Virus 5'-Noncoding Region. *RNA-a Publication of the RNA Society,* 1**,** 526-537.

WANG, C. Y., SARNOW, P. & SIDDIQUI, A. 1994b. A Conserved Helical Element Is Essential for Internal Initiation of Translation of Hepatitis-C Virus-Rna. *Journal of Virology,* 68**,** 7301-7307.

WANG, T. H., RIJNBRAND, R. C. A. & LEMON, S. M. 2000. Core protein-coding sequence, but not core protein, modulates the efficiency of cap-independent translation directed by the internal ribosome entry site of hepatitis C virus. *Journal of Virology,* 74**,** 11347-11358.

XIE, J., ZHANG, M., ZHOU, T., HUA, X., TANG, L. & WU, W. 2007. Sno/scaRNAbase: a curated database for small nucleolar RNAs and cajal body-specific RNAs. *Nucleic Acids Research,* 35**,** D183-7.

YAMAMOTO, C., ENOMOTO, N., KUROSAKI, M., YU, S. H., TAZAWA, J., IZUMI, N., MARUMO, F. & SATO, C. 1997. Nucleotide sequence variations in the internal ribosome entry site of hepatitis C virus-1b: No association with efficacy of interferon therapy or serum HCV-RNA levels. *Hepatology,* 26**,** 1616-1620.

YAMAMOTO, H., COLLIER, M., LOERKE, J., ISMER, J., SCHMIDT, A., HILAL, T., SPRINK, T., YAMAMOTO, K., MIELKE, T., BURGER, J., SHAIKH, T. R., DABROWSKI, M., HILDEBRAND, P. W., SCHEERER, P. & SPAHN, C. M. 2015. Molecular architecture of the ribosome-bound Hepatitis C Virus internal ribosomal entry site RNA. *Embo Journal,* 34**,** 3042-58.

YAMAMOTO, H., UNBEHAUN, A., LOERKE, J., BEHRMANN, E., COLLIER, M., BURGER, J., MIELKE, T. & SPAHN, C. M. 2014. Structure of the mammalian 80S initiation complex with initiation factor 5B on HCV-IRES RNA. *Nature Structural & Molecular Biology,* 21**,** 721-7.

YANG, X., LI, M., LIU, Q., ZHANG, Y., QIAN, J., WAN, X., WANG, A., ZHANG, H., ZHU, C., LU, X., MAO, Y., SANG, X., ZHAO, H., ZHAO, Y. & ZHANG, X. 2015. Dr.VIS v2.0: an updated database of human disease-related viral integration sites in the era of high-throughput deep sequencing. *Nucleic Acids Research,* 43**,** D887-92.

YASMEEN, A., HAMID, S., GRANATH, F. N., LINDSTROM, H., ELLIOTT, R. M., SIDDIQUI, A. A. & PERSSON, M. A. A. 2006. Correlation between translation efficiency and outcome of combination therapy in chronic hepatitis C genotype 3. *Journal of Viral Hepatitis,* 13**,** 87-95.

YU, Y., JI, H., DOUDNA, J. A. & LEARY, J. A. 2005. Mass spectrometric analysis of the human 40S ribosomal subunit: native and HCV IRES-bound complexes. *Protein Sci,* 14**,** 1438-46.

YUSIM, K., RICHARDSON, R., TAO, N., DALWANI, A., AGRAWAL, A., SZINGER, J., FUNKHOUSER, R., KORBER, B. & KUIKEN, C. 2005. Los alamos hepatitis C immunology database. *Appl Bioinformatics,* 4**,** 217-25.

ZHANG, J., YAMADA, O., ITO, T., AKIYAMA, M., HASHIMOTO, Y., YOSHIDA, H., MAKINO, R., MASAGO, A., UEMURA, H. & ARAKI, H. 1999. A single nucleotide insertion in the 5 '-untranslated region of hepatitis C virus leads to enhanced cap-independent translation. *Virology,* 261**,** 263-270.

ZHAO, Q., HAN, Q., KISSINGER, C. R., HERMANN, T. & THOMPSON, P. A. 2008. Structure of hepatitis C virus IRES subdomain IIa. *Acta Crystallographica Section D-Biological Crystallography,* 64**,** 436-443.

ZHAO, W. D. & WIMMER, E. 2001a. Genetic analysis of a poliovirus/hepatitis C virus chimera: New structure for domain II of the internal ribosomal entry site of hepatitis C virus. *Journal of Virology,* 75**,** 3719-3730.

ZHAO, W. D. & WIMMER, E. 2001b. Genetic analysis of a poliovirus/hepatitis C virus chimera: new structure for domain II of the internal ribosomal entry site of hepatitis C virus. *J Virol,* 75**,** 3719-30.

ZHOU, M., SANDERCOCK, A. M., FRASER, C. S., RIDLOVA, G., STEPHENS, E., SCHENAUER, M. R., YOKOI-FONG, T., BARSKY, D., LEARY, J. A., HERSHEY, J. W., DOUDNA, J. A. & ROBINSON, C. V. 2008. Mass spectrometry reveals modularity and a complete subunit interaction map of the eukaryotic translation factor eIF3. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 18139-18144.

ZHULIN, I. B. 2015. Databases for Microbiologists. *J Bacteriol,* 197**,** 2458-67.

ZINOVIEV, A., HELLEN, C. U. & PESTOVA, T. V. 2015. Multiple mechanisms of reinitiation on bicistronic calicivirus mRNAs. *Molecular Cell,* 57**,** 1059-73.

ZOU, D., MA, L., YU, J. & ZHANG, Z. 2015. Biological databases for human research. *Genomics Proteomics Bioinformatics,* 13**,** 55-63.

9 Seleceted Publications

9.1 Understanding the potential of hepatitis C virus internal ribosome entry site domains to modulate translation initiation via their structure and function.

Anas Khawaja, Vaclav Vopalensky and Martin Pospisek*

Department of Genetics and Molecular Biology, Charles University in Prague, Vinicna 5, Praha 2, 128 43, Czech Republic

WIREs RNA, 2015, Volume 6, Issue 2, pages 211–224

WIREs RNA. doi: 10.1002/wrna.1268

The article is a critical and advanced review that aims to provide an extensive analyses and summary of the available data on hepatitis C virus internal ribosome entry site. The article represents primarily the sequence-structure preservation and its relationship with function, the IRES-induced conformational changes on the ribosome and interaction with the eukaryotic initiation factors. The variability of HCV-IRES in its different structural elements and domains in the context of IRES mediated translation initiation was also highlighted. The article carefully analyzes the large number of HCV-IRES mutation data found in literature, which are divided by various classifications across many studies. Within that HCV-IRES mutation data, we found and reported here a distinct and unusual characteristic of the HCV-IRES which is linked to a probable long-range inter- and/or intra-domain interaction.

The introduction includes the summary of current knowledge about the HCV-IRES translation initiation pathway and the specified arrangements of its domains, ribosomal assembly and subsequent eIFs binding. The chapters that follow encompass individual domains II, III and IV with a broader overview discussing mainly the available structure, biochemical and mutation data along with functional elements downstream to AUG initiation codon and interaction of eIF3 with the HCV-IRES. These chapters also analyze the presence of compensatory mutations that indicates the unique phenomenon of a probable long-range interaction of domains,

contributing to total restoration of HCV-IRES translation efficiency. The chapter detailing the stress-induced HCV protein production by eIF-2 independent translation mechanism is also included. We aimed to display the crucial interlink between sequence-structure conservation of the HCV-IRES and its function in initiation of viral mRNA protein synthesis. Furthermore, the impact of mutations on modifying interactions with ribosome subunits and eIFs was carried out by applying well-analyzed and sorted data of mutations spread across various domains of HCV-IRES elements, was thoroughly discussed.

9.2 HCVIVdb: The hepatitis-C IRES variation database

Evan W. Floden[1,2,3*], Anas Khawaja[1*], Václav Vopálenský[1], Martin Pospíšek[1].

[1]Dept. Genetics & Microbiology, Faculty of Science, Charles University in Prague, Viničná 5, 128 44 Prague 2, Czech Republic

[2]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona, Spain

[3]Universitat Pompeu Fabra (UPF), Barcelona, Spain

The hepatitis C virus shows remarkably high level of nucleotide diversity due to the properties of its viral RdRp that resulted in the emergence of six main genotypes and number of subtypes. We analysed the level of variability in one of the most conserved region of the HCV genome across its genotypes: the internal ribosome entry site. The HCV-IRES regulates the cap-independent translation initiation of the viral genome and sequence variability can have important implications on the overall structural organization of its domains and viral protein synthesis. We found high impact of mutations localized in the HCV-IRES domains from our patients' samples on the viral mRNA translation. In an attempt to carry out comparative and functional analyses of our data we collected HCV-IRES mutations (~1900), naturally occurring or *in vitro* mutagenized, from significant number of the HCV-IRES mutation-linked published per reviewed articles, and established an HCV-IRES variation database (HCVIVdb: www.hcvivdb.org). The HCV-IRES entries acquired manually from the published studies were collated into diverse categories such as HCV genotype, nucleotide variations, translation systems used to analyse translation efficiency, plasmid and reporter genes used, clinical data etc. in an effort for easier and wide-scale analyses. The search engine in HCVIVdb provides immediate access to the information involving mutations' type, location, genotype and measured translation efficiency with corresponding publication references and direct link to the PubMed. The HCVIVdb is equipped with analytical tools that permits the variation queries to be further grouped into sub-categories of individual domains, subdomains, translation activity and genotype. The dataset is organized for functional and comparative analyses. This is achieved by presenting the similar / identical entries alongside submitted ones that allows an extensive evaluation in the

context of function, location and/or experimental parameters with further estimation by respective domains, subdomains, by genotype and range of efficiency. We also found differences in the translation efficiency of similar mutations submitted by different studies. The validation of data stored in HCVIVdb was carried out by comparison with the HCV genome entries from other resources. The frequency of occurrence of every nucleotide at each position of the HCV-IRES (1-356) from other resources correlated positively with variability of natural entries from the HCVIVdb. The agreement of the datasets offers deeper insights into the hypervariability and conservation of specific nucleotides and the evolutionary advantage of particular bases over the others at each HCV-IRES nucleotide position. The HCVIVdb provides a unique central resource for specialized HCV-IRES mutation dataset collated into categories that permits the users to perform multiple analyses, evaluate, manipulate and download. It can also be helpful in the development of site-specific drugs and a guide for future experiments.

9.3 A study of the HCV-IRES variability: An experimental approach coupled with design of a large-scale mutation database

Anas Khawaja[1], Vaclav Vopalensky[1], Evan Floden[1], Ludek Roznovsky[2], Jakub Mrazek[3], [4] and Martin Pospisek[1,*]

[1]Department of Genetics and Molecular Biology, Charles University in Prague, Vinicna 5, Praha 2, 128 43, Czech Republic

The mRNA translation of hepatitis C virus begins by direct recruitment of 40S ribosome to the AUG initiation codon of an IRES element, surpassing the requirement of 5' cap recognition and scanning procedure. The IRES-RNA of HCV, provides an alternate cap-independent mechanism for protein synthesis that rely on the presence of only few of the eIF members compared to cap-dependent translation initiation. We began the study on HCV-IRES, the most conserved region of the viral genome, to comprehend the variability that accumulates and persists over a period of time in patients suffering from the HCV. The genetic heterogeneity of HCV-IRES in HCV infected patients were estimated and visualized by denaturing gradient and temperature gel electrophoresis. Thus, establishing a quick format to examine and evaluate the characterization and extent of genetic variation in a given time among virus population(s). The selected HCV-IRES variants were sequenced and the impact of mutations on translation efficiency of viral mRNA was measured by flow cytometry. In an attempt to evaluate observed phenotype(s) of the HCV-IRES variants from patients samples on a wider and much larger-scale, prompted us to collect mutations from significant number of the HCV-IRES mutations'-linked studies. The acquired mutations' dataset comprises the HCV-IRES variations and additional diverse information, which was collated into a knowledge-base termed HCV-IRES variation database (HCVIVdb). The establishment of HCVIVdb into multiple categories (see manuscript 2) is a facilitation to attain a large overview of all the mutations in HCV-IRES and their contribution to translation efficiency in the context of sequence preservation and structural integrity. Using HCVIVdb data we identified IRES-RNA structural elements that are most and least vulnerable to variations upon calculating the cumulative average translation activity of all the individual mutations and multiple mutations, respectively. The comparative analyses also revealed a

unique HCV-IRES translation response which can be attributed to a probable long-range inter- and intra-domain interaction. We discovered, in some cases, multiple mutations located in various HCV-IRES domains managed to restore the translation efficiency, whereas the individual occurrence of corresponding mutations were found to devastate the viral mRNA translation. Likewise, a reverse phenomenon comprises sharp decline of HCV-IRES activity in multiple mutants was also found. The individual substitutions from these multiple mutants do not have any impact on HCV-IRES mediated translation initiation. The phenomenon of probable long-range inter- or intra-domain interaction was found in both the patients' samples and the collected data from the published articles. Such unusual HCV-IRES characterization is discussed thoroughly with reference to mutations' localization, impact on structure configuration and contacts with ribosome assembly and initiation factors. However, we could not establish any direct link between factors such as (a) mutations' type, frequency and localization and (b) probable long-range interaction that results in restoration of HCV-IRES translation efficiency. We also identified novel mutations (~20) from our patients' data and estimated their impact on HCV-IRES functional response. Moreover, we *in silico* analyzed possible effect of these novel mutations on HCV-IRES structure conformation and interaction with translation machinery, from the available mutation, structure and biochemical data stored in HCVIVdb. Our findings provide a unique perspective into the HCV-IRES behavior and significance of sequence and structure preservation of its domains and subdomains on a broader level. This study helped identifying new features of the HCV-IRES function and may lead to uncover other biological roles of IRES-elements in regulation of HCV-IRES mediated translation initiation.