Université Lille 1 - Sciences et Technologies

Ecole Doctorale Biologie Santé de Lille

THESE

Pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITE

Discipline

Aspects moléculaires et cellulaires de la biologie

Présentée et soutenue publiquement par

Luíza MAMIGONIAN BESSA

Investigation of the hepatitis C virus RNA polymerase NS5B in solution by nuclear magnetic resonance and its interaction with intrinsically disordered domain 2 of the NS5A protein

Thèse dirigée par Docteur Xavier HANOULLE

et co-encadrée par Docteur Robert SCHNEIDER

Soutenue le 21 novembre 2017

JURY

Docteur Jérôme BOISBOUVIER	Rapporteur
Docteur Stéphane BRESSANELLI	Rapporteur
Docteur Carine TISNE-VICROBECK	Examinateur
Docteur Caroline SMET-NOCCA	Examinateur
Docteur Robert SCHNEIDER	Invité
Docteur Xavier HANOULLE	Directeur de Thèse

Abstract

NS5B is the hepatitis C virus (HCV) RNA-dependent RNA polymerase. This protein has been extensively studied by X-ray crystallography and shows an organization in three subdomains called fingers, palm and thumb. Whereas static crystallographic data are abundant, structural studies of this protein in solution are limited. Nuclear magnetic resonance (NMR) spectroscopy was used to study the 65 kDa NS5B in solution as well as its interaction with binding partners. It was characterized using selective isotopic labeling of isoleucine side-chain methyl groups, which gives rise to a simplified NMR spectrum with an improved signal-to-noise ratio. This characterization confirmed the presence of particular dynamics in the subdomains, especially in the thumb, as well as long-range effects that are transmitted through to other subdomains. Furthermore, this system was used to investigate the binding of the domain 2 of NS5A (NS5B-D2), a disordered domain of another HCV protein that has been shown to directly interact with NS5B *in vitro*. Using paramagnetic relaxation enhancement we showed that NS5A-D2 binds to NS5B via, at least, two binding sites on the thumb subdomain. As one of these sites was the binding site of allosteric inhibitor filibuvir, we characterized the binding of this small molecule to NS5B by NMR spectroscopy and found long-range effects of its binding throughout the polymerase. Finally, we studied the binding of a small RNA template strand to NS5B and found that both NS5A-D2 and filibuvir reduce but do not abolish the interaction between the polymerase and RNA. In sum, NMR was used to study dynamic properties of NS5B and its interactions with binding partners.

Keywords: Nuclear magnetic resonance spectroscopy, Hepatitis C virus, NS5B

Résumé

NS5B est l'ARN polymérase du virus de l'hépatite C (VHC). Sa structure a beaucoup été étudiée par radiocristallographie. Elle contient trois sous-domaines appelés doigts, paume et pouce. Cependant, les études structurales de cette protéine en solution sont très limitées. La résonance magnétique nucléaire (RMN) a été utilisée pour étudier NS5B en solution ainsi que son interaction avec différents partenaires. L'emploi d'un échantillon de NS5B (65kDa) perdeuterée et sélectivement enrichie au niveau des méthyles $\delta 1$ des résidus d'isoleucines a permis d'obtenir un spectre simplifié et de bonne qualité. Cette étude a confirmé la présence d'une dynamique particulière dans le pouce et a permis de mettre en évidence des effets à longues distances qui se transmettent aux autres sous-domaines. Cette approche a alors été utilisée pour étudier l'interaction entre NS5B et le domaine 2 de la protéine NS5A (NS5A-D2) du VHC. Celui-ci est un domaine intrinsèquement désordonné qui interagit directement avec NS5B *in vitro*. Nous avons identifié que NS5A-D2 se lie via deux sites d'interaction sur le sous-domaine du pouce. Puisqu'un de ces sites est le site de liaison de l'inhibiteur allostérique filibuvir, nous avons étudié la liaison de cette molécule à la polymérase. Sa liaison cause des effets à longues distances tout au long de NS5B. Enfin, nous avons caractérisé la liaison d'un ARN simple brin à NS5B et nous avons identifié que NS5A-D2 et filibuvir réduisent mais ne suppriment pas l'interaction de NS5B avec l'ARN. L'analyse de NS5B par RMN en solution a permis d'étudier des interactions et d'accéder à des paramètres dynamiques très complémentaires des études cristallographiques.

Mots-clés : Spectroscopie de la résonance magnétique nucléaire, Virus de l'hépatite C, NS5B

Acknowledgements

These last three years have shaped me and helped me grow in a way I could not have expected. I would like to thank everyone who contributed in some way to this work and my growth.

First, I would like to thank my thesis supervisors Dr. Xavier Hanoulle and Dr. Robert Schneider for their full support throughout the duration of this work. They were always willing to help when I had questions. They also taught me important lessons, in particular subjects and on how to develop my research in general, that I will certainly take with me in the future. I cannot thank them enough, their contributions shaped this work and were instrumental to my growth during this period.

Besides my supervisors, I would like to thank the members of my thesis committee, Dr. Stéphane Bressanelli, Dr. Jérôme Boisbouvier, Dr. Carine Tisné and Dr. Caroline Smet-Nocca for the honor of having them review this thesis and provide their insightful comments and questions.

I appreciate the contributions from Dr. Stéphane Bressanelli and Dr. Alain Baulard, who participated in my *Comité de Suivi de Thèse*, their questions and suggestions were always pertinent and helped me frame my work within a larger picture.

I would like to thank the members of CIRI (Centre International de Recherche en Infectologie), Dr. Patrice André, Dr. Olivier Diaz and Dr. Vincent Lotteau, with whom we collaborated on the glucokinase project.

Furthermore, I am grateful to Dr. Guy Lippens for his NMR spectroscopy courses that led me to this laboratory and to work with such a great team. I could not forget Dr. Hélène Launay-Souvi, who first introduced me to NS5B, taught me a lot about NMR and biochemistry during my intership in this team, and whose work is the basis for all that is described in this thesis. I would also like to thank Dr. Marie Dujardin who also taught me so much during my thesis. Dr. François-Xavier Cantrelle, you have been a lifesaver, with every time I was stuck on something with the spectrometers and had to call you to help solve it.

All of the members of the NMR and molecular interactions team, who have not yet been named, have also helped shape this work through shared ideas, suggestions,

discussions and encouragement: Isabelle Landrieu, Emmanuelle Boll, Idir Malki, Elian Dupré, Alessia Lasorsa, Hamida Merzougui, Clément Despres, Clément Danis, João Neves, Haoling Qi, Amina Kamah and Isabelle Huvent.

Finally, I would like to express my gratitude to my parents and to my partner, who have supported and encouraged me throughout these years of research and writing this thesis. I could not have done it without them.

Luíza Mamigonian Bessa

Contents

List of Figures

List of Tables

List of Abbreviations

ADP	Adenosine diphosphate
AMP-PNP	Adenylyl-imidodiphosphate
ATP	Adenosine triphosphate
BEST	Band-selective excitation short-transient
CD	Circular dichroism
CD81	Cluster of differentiation 81 protein
Cdk	Cyclin-dependent kinase
CEST	Chemical exchange saturation transfer
CHC	Chronic hepatitis C
Con1	Consensus genome 1 isolate
CS	Chemical shift
CsA	Cyclosporin A
CSI	Chemical shift index
CSP	Chemical shift perturbation
Cyp	Cyclophilin
CypA	Cyclophilin A
DAA	Direct-acting antiviral
DEST	Dark-state exchange saturation transfer
DNA	Deoxyribonucleic acid
dsRNA	Double-stranded RNA
DTT	1,4-dithiothreitol
E1	Envelope glycoprotein E1
E2	Envelope glycoprotein E2
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
EXSY	Exchange spectroscopy
FDA	U.S. Food and drug administration
FPLC	Fast protein liquid chromatography
GCK	Glucokinase
GST	Glutathione S-transferase
HBV	Hepatitis B virus
HCV	Hepatitis C virus or Hepacivirus C
HIV	Human immunodeficiency virus
HK	Hexokinase
HK2	Hexokinase II
HMQC	Heteronuclear multiple quantum correlation (spectrum)
HSQC	Heteronuclear single quantum correlation (spectrum)
HWHM	Half width at half maximum
IDP	Intrinsically disordered protein
IDPR	Intrinsically disordered protein region
IPTG	Isopropyl-1-thio-β-D-galactopyranoside
IRES	Internal ribosome entry site
JFH-1	Japanese fulminant hepatitis 1 isolate
K_D	Dissociation constant
LB	Luria-Bertani
LCS	Low complexity sequence
LVP	Lipoviroparticle
Methyl-HMQC	$^{1}H,^{13}C$-TROSY HMQC spectrum

miR	MicroRNA
miR-122	MicroRNA 122
mRNA	Messenger RNA
MW	Molecular weight
NI	Nucleos(t)ide inhibitor
NMR	Nuclear magnetic resonance
NNI	Non-nucleos(t)ide inhibitor
NOESY	Nuclear Overhauser effect spectroscopy
NS2	Non-structural protein 2
NS3	Non-structural protein 3
NS3/4A	Non-structural protein complex 3/4A
NS4B	Non-structural protein 4B
NS5A	Non-structural protein 5A
NS5A-D1	Domain 1 of Non-structural protein 5A
NS5A-D2	Domain 2 of Non-structural protein 5A
NS5A-D3	Domain 3 of Non-structural protein 5A
NS5B	Non-structural protein 5B
NTP	Nucleotide triphosphate
OD600	Optical density measured at 600 nm
ORF	Open reading frame
PDB	Protein data bank
PEG	Polyethylene glycol
PEG-IFN	Pegylated interferon
PHHI	Persistent hyperinsulinemic hypoglycemia of infancy
PI4KA	Phosphatidylinositol-4 kinase IIIα
PI4P	Phosphatidylinositol 4-phosphate
PPIase	Peptidyl-prolyl cis/trans isomerase
ppm	Parts per million
PRE	Paramagnetic relaxation enhancement
PTM	Post-translational modification
RAS	Resistance-associated substitution
RDC	Residual dipolar coupling
RdRp	RNA-dependent RNA polymerase
RNA	Ribonucleic acid
RNA16	16-mer RNA strand 5′-CUAAGAUGCUCGCUGC
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
S/N	Signal-to-noise ratio
SSP	Secondary structure propensity
ssRNA	Single-stranded RNA
SVR	Sustained viral response
TCA	Thiophene-2-carboxylic acid
TMS	Tetramethylsilane
TMSP	3-(Trimethylsilyl)propionic-2,2,3,3-d_4 acid sodium salt
TROSY	Transverse relaxation optimized spectroscopy
UTR	Untranslated region
VLDL	Very low-density lipoprotein
WT	Wild-type
XRD	X-ray Diffraction

Introduction

1.1 Hepatitis C

1.1.1 Discovery

During the 1970s, a large fraction of transfusion patients were found to develop hepatitis that was not caused by the viruses responsible for hepatitis that were known at the time, the hepatitis A and B viruses. Therefore, the illness was called non-A, non-B hepatitis [1]. A chimpanzee model of acute and chronic non-A, non-B hepatitis was developed in the late 1970s [2]. The agent responsible for this disease was found to be hard to isolate. With a complementary DNA (cDNA) library derived from infected chimpanzee plasma, antibodies from human patients and multiple assays to assure that the genome was of viral origin, the hepatitis C virus (HCV) was identified in 1989 [3, 4] as being the causative agent of non-A non-B chronic hepatitis. In 1997, it was confirmed that this virus alone is the causative agent of this disease. This was done by infecting a chimpanzee with cDNA coding for a consensus sequence from an HCV-infected patient [5].

1.1.2 Transmission

HCV is transmitted through exposure to infected blood. This can be through the receipt of contaminated blood or blood products, medical procedures with unsafe materials or injection drug use. Sexual transmission of HCV is possible but unlike the human immunodeficiency virus (HIV) sexual transmission is rare and is strongly dependent on prior HIV infection [6].

HCV can be transmitted by vertical transmission, that is from mother to child during the period immediately before or after birth. Vertical transmission is rare but its incidence rate increases with HIV co-infection and if the mother has a high viral load, especially in the third trimester. Breastfeeding does not seem to be related to a larger risk of HCV transmission [7].

Certain populations are considered at risk for HCV infection. In middle- to high-income countries the population with the highest risk for HCV infection are people who inject drugs. Recipients of infected blood products and patients who underwent invasive procedures are also at high risk of infection. This is particularly important for those who have received blood products prior to the introduction of the screening of blood products in the early 1990s.

Transmission rates have been declining since the discovery of the viral agent responsible for hepatitis C. This is attributed to improved medical practices and widespread sterilization, screening of blood products for HIV and HCV, and expansion of needle exchanging programs [6].

1.1.3 Epidemiology

Despite this virus being estimated to infect human populations for at least several centuries [8, 9] most of the infections happened during the later half of the twentieth century. This is due to the increase in medical procedures that required perforating and cutting, and injection drug use that saw a large rise after the 1940s. However, in most patients the infection takes several decades to develop and cause symptoms.

Since its discovery, the burden of the disease has increased. In 1990, 333000 people were estimated to have died from this disease. By 2010 this increased to 499000 and 704000 in 2013 [10, 11]. This is in spite of a decrease in transmission after the discovery of the viral agent and improvement in the medical procedures to reduce transmission. The increased burden is due to the delayed development of symptoms that can take decades to develop after initial infection.

Worldwide, 115 million people are estimated to have a history of HCV infection, testing positive for anti-HCV antibodies. Of these people, 80 million are estimated to be chronically infected. The difference in the values arises from people who have either been treated for an HCV infection or have spontaneously cleared the virus but still present the antibodies [12]. Even though this value is significantly lower than earlier estimates of 184 million with a history of infection and 170 million chronically infected [13], it still amounts to 1.6% of the world population being chronically infected.

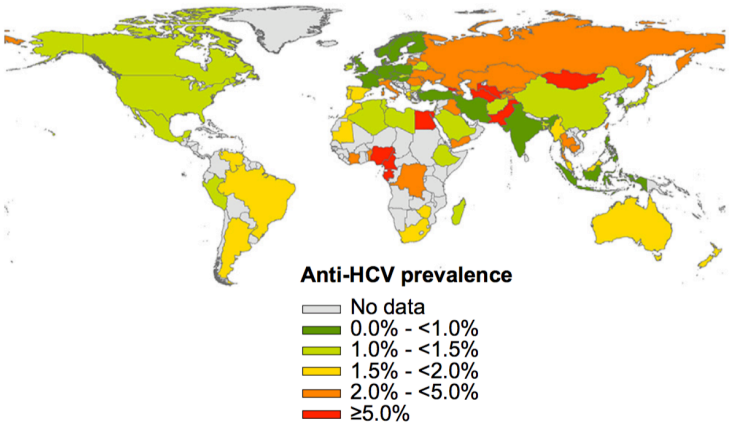

FIGURE 1.1: The prevalence of anti-HCV antibodies by country. This value differs from the number of chronically infected because patients who have cleared the virus are still positive for anti-HCV antibodies. From Gower *et al.* [12], available under a Creative Commons Attribution-Noncommercial-NoDerivatives license.

The prevalence of infection varies throughout the world (Figure 1.1). This has to do with how the virus spread in each region. For example, Egypt has the highest prevalence worldwide, with 14.7% [12] of the population estimated to have a history of HCV infection. This high prevalence is because of exposure during anti-schistosomiasis therapy without proper sterilization practices that was applied community-wide from the 1950s until the 1980s [14] and resulted in a large reservoir of HCV in the country [15].

1.1.4 Disease

After initial infection there is an acute phase of the disease 0 to 24 weeks after contamination (Figure 1.2). In most patients (70-80%) this phase is asymptomatic and the infection is rarely diagnosed. In patients with symptomatic acute hepatitis C it has been shown that treatment during this phase can prevent chronic hepatitis [16–18].

HCV-infected individuals that did not clear the virus after the acute phase (about 70% of infected individuals) develop chronic hepatitis C (CHC). The chronic presentation is characterized by inflammatory changes in the liver. Chronic inflammation causes activation of hepatic stellate cells. These cells are fundamental for the healing of wounds in the liver. They are usually quiescent but in the presence of an immune response or inflammation they become active myofibroblasts. Myofibroblasts cause an increased deposition of fibrous extracellular matrix proteins which become fibrotic

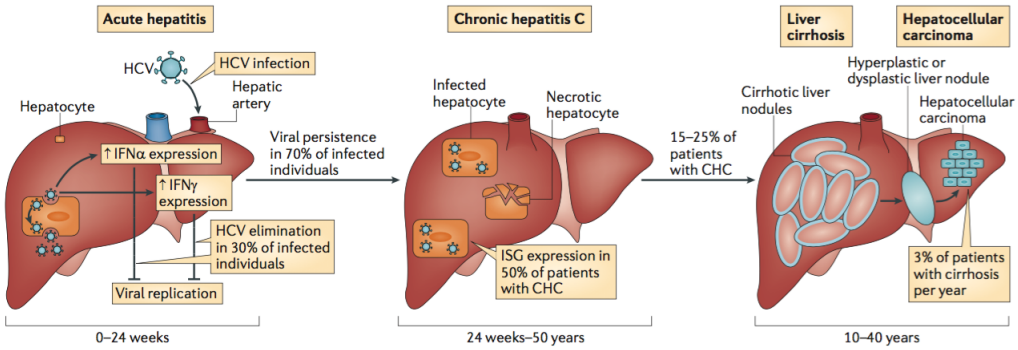

FIGURE 1.2: Evolution of hepatitis C virus infection. After the entry of the virus into the liver through the hepatic artery and portal vein, there is an often asymptomatic acute phase of the infection that can last 0 to 24 weeks. The virus is eliminated during this phase in about 30% of patients. Those that do not clear the virus go on to develop chronic hepatitis C (CHC). Most of these patients do not develop liver fibrosis or liver disease. Nevertheless, cirrhosis develops over 10–40 years in 15%–25% of patients with CHC. Cirrhotic patients have a 1–5% risk of developing hepatocellular carcinoma per year. Decompensated cirrhosis and hepatocellular carcinoma are the main causes of death in end-stage CHC. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Immunology, Heim[19], copyright 2013.

scarring [20]. Initially the fibrotic modifications are confined to the portal triad of the hepatic lobule but it progresses to the centrilobular area [21]. Fibrosis development is dependent on many factors, including age, duration of infection, genotype, male sex and co-infection with HIV or HBV [22]. Over several decades, this CHC develops into cirrhosis in 15–25% of patients [19]. Cirrhosis can become decompensated; that is, presenting either jaundice, ascites, variceal hemorrhage, or hepatic encephalopathy; and increase the risk of death among these patients. Patients with cirrhosis have a 1-5% per year rate of developing hepatocellular carcinoma, a malignant hepatoma [22]. Decompensated cirrhosis and hepatocellular carcinoma are the main causes of death due to HCV [11].

Pathogenesis is not well understood and is believed to be due to an ineffective immune response. This immune reaction is responsible for the liver damage but not capable of clearing the virus. Another possible culprit is oxidative stress. Oxidative stress can be elicited by the expression of the viral proteins, the alterations of the intrahepatic lipid metabolism and inflammatory immune response. Also, oxidative stress can stimulate the conversion of hepatic stellate cells to myofibroblasts [23, 24].

The goal of current treatment is to clear the virus. It is defined as sustained virological response (SVR), the absence of viral RNA, either 12 or 24 weeks after the end of the

treatment. Once SVR is achieved, fibrosis stops developing. However, it does not reverse cirrhosis and hepatocellular carcinoma that have already developed. Reinfection is possible after clearance (spontaneously or due to treatment) and behavioral changes may be needed if the patient has high risk of re-infection [22].

1.2 Hepatitis C virus (HCV)

1.2.1 Classification and genetic distribution

The hepatitis C virus, recently renamed Hepacivirus C [25], is a positive-sense single stranded RNA virus of the family *Flaviviridae* and genus *Hepacivirus*. The *Flaviviridae* family contains four genera: *Flavivirus*, *Pestivirus*, *Pegivirus* and *Hepacivirus* [26]. Among these genera, over 80 species of viruses have been identified. Many of these infect humans. In the genus *Flavivirus*, these include the West Nile virus, dengue virus, yellow fever virus and Zika virus.

GB virus B was the first *Hepacivirus* to be identified after HCV, despite it originially being classified as a *Pegivirus*. It was found in tamarins, a type of new-world monkey, after infection by plasma from individuals with unexplained hepatitis [27, 28].

All currently identified hepaciviruses have mammalian hosts [25]. Other hepaciviruses include primate hepaciviruses [29] and nonprimate hepaciviruses, that infect horses [30, 31], bats [32], rodents [33, 34], cows [35] and dogs [36]. The horse nonprimate hepacivirus is the most closely related to the HCV [37]. The zoonotic origin of HCV is still a subject of debate [38, 39].

The hepatitis C virus shows extensive genetic variability. HCV is classified in 7 genotypes and subtypes within them (Figure 1.3) [40]. The genotypes present variations of 30-35% in the nucleotide sequence while the variation within a subtype is less than 15%.

Genotype 1 is the most common globally, accounting for 46% of infections. However, much like the prevalence of HCV in different countries, the genotypes vary greatly between regions [12]. Genotype 1 is the most prevalent in Europe, the Americas and

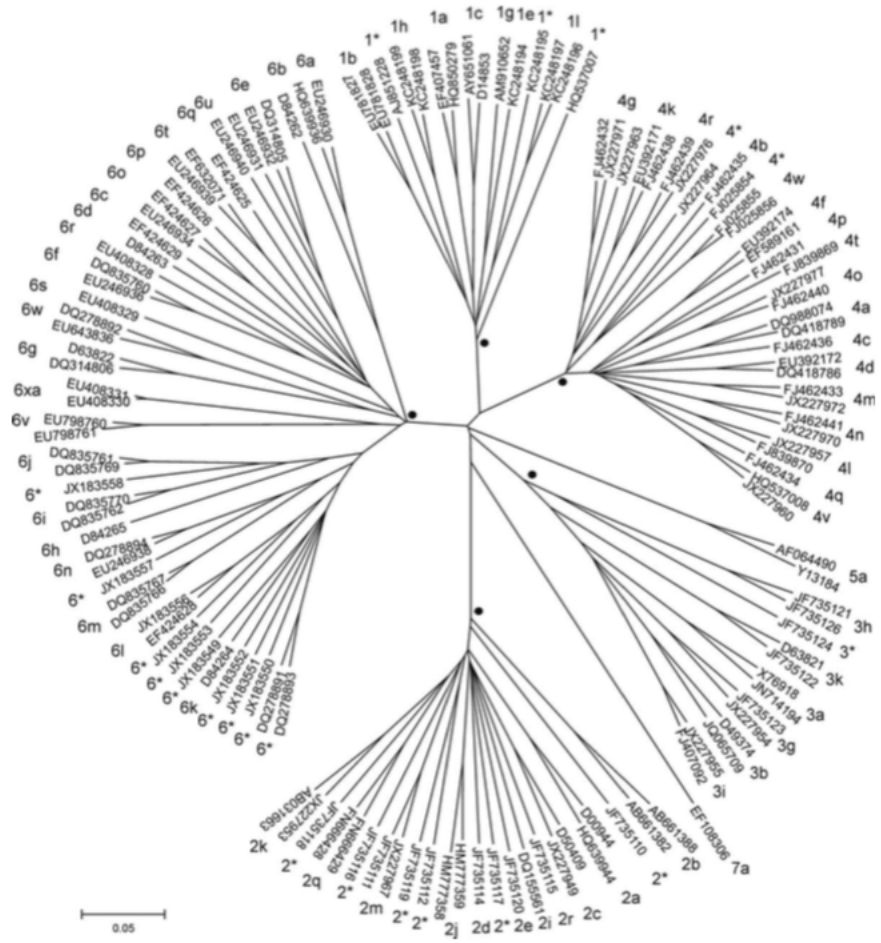

FIGURE 1.3: Phylogenetic tree of complete coding regions. For clarity, the authors included only sequences that were complete (and not just of some regions of the genome) and only one sequence from each subgenotype. From Smith *et al.* [40], available under a Creative Commons Attribution-Unported license.

many regions of Asia (Figure 1.4). Its global prevalence is believed to be due to dissemination during the twentieth century prior to discovery of the virus through contaminated blood and blood products. Genotype 3 is the second most prevalent worldwide and accounts for 40% of infections in Asia. Its spread is associated with injection drug use and population migration from regions where the genotype is dominant. The other genotypes are believed to be more endemic but can become epidemic under certain conditions. For example, Genotype 4 is the most prevalent in North Africa and the Middle East but this is greatly from the contribution of Egypt where the exposure during anti-schistosomiasis therapy caused the spreading of mainly this genotype [41]. Genotype 7, the most recently discovered genotype, has a handful of identified cases, mostly identified in Canada in patients native from central Africa [42].

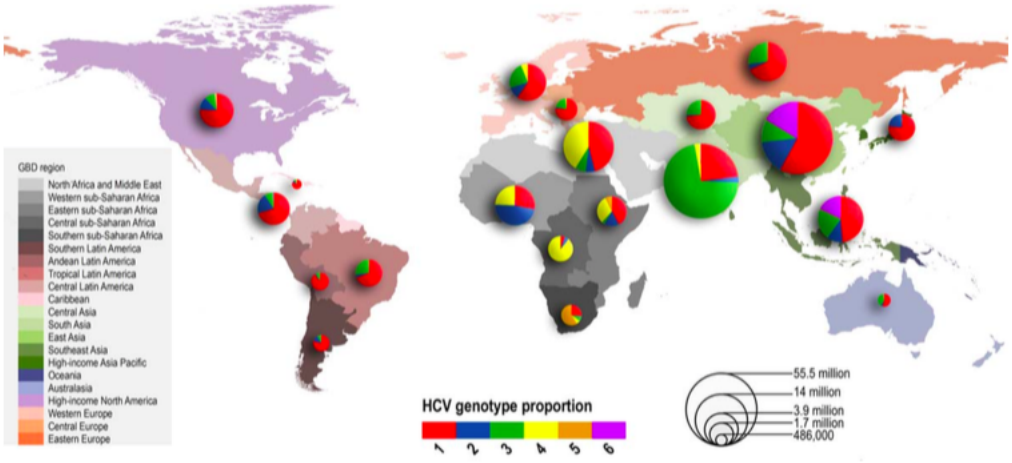

FIGURE 1.4: Relative prevalence of each HCV genotype by region. Size of pie charts represent the estimated number of seropositive cases in each region. From Messina *et al.* [41], available under a Creative Commons Attribution-Unported license.

Identification of the genotype of the infection is important for the patient and physician to identify the best therapeutic procedure as different genotypes have different suggested treatments and prognoses. Prior to 2011, when only interferon-based therapy was available, treatment was significantly less effective on genotype 1-infected patients [43] compared to other genotypes. Current therapies present sustained virological response in over 90% of patients with HCV genotype 1 but now genotype 3 has become the hardest to treat HCV genotype.

1.2.2 Viral particle

The hepatitis C virus is a small enveloped virus whose ribonucleic genome is encapsulated in a capsid composed of the viral protein Core that is further surrounded by a membrane composed of lipids, apolipoproteins (Apo) and viral glycoproteins E1 and E2 [44]. Unlike related *Flaviviridae* viruses, the HCV particle is very heterogeneous. Particles derived from cell culture (as opposed to originary from patients) have sizes varying from 40 to 100 nm in diameter [45]. Ultracentrifugation of patient serum gives rise to particles that have a large range of densities from less than 1.0063 g/mL to 1.055 g/mL [46]. Cell culture-derived HCV particles have a higher average density and are less infectious than chimpanzee-derived HCV particles (which have densities similar to human HCV particles) suggesting that the low-density factors are responsible for infectivity [47].

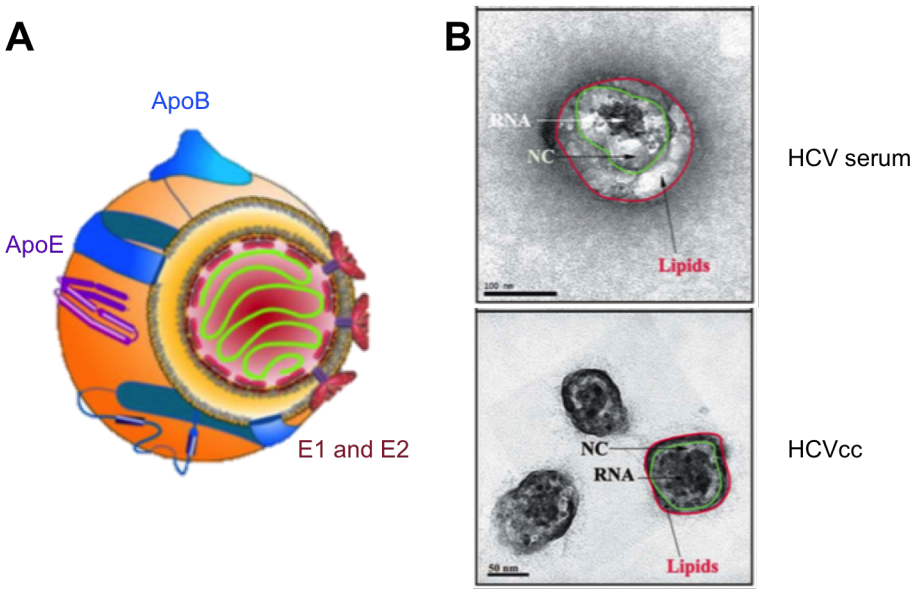

FIGURE 1.5: Structure of HCV virions. **A** Organization from Bartenschlager *et al.* in which HCV is associated with a lipoprotein to form a virolipoprotein particle. RNA (in green) is encircled by the capsid and further encircled by lipids and ER-derived membranes containing host apolipoproteins B and E, and the viral glycoproteins E1 and E2. Reprinted from [48], Copyright 2011, with permission from Elsevier. **B** Electron micrographs from Piver *et al.* showing HCV particles derived from patients (top, HCV serum) and cell culture (bottom, HCVcc) displaying the structure organization. Scale bars represent 100 and 50 nm, respectively. Adapted by permission from BMJ Publishing Group Limited [49].

In 2005, André and colleagues [50] proposed an organization of the infectious lipid-associated viral particle, called a lipoviroparticle (LVP) (Figure 1.5). The viral RNA is encapsulated in a capsid composed of the structural viral protein Core. This capsid is surrounded by the hydrophilic heads of an endoplasmic reticulum-derived phospholipid monolayer forming a hydrophobic particle. The outer monolayer contains host proteins ApoE and ApoB. Between these layers is a neutral lipid core that contains low-density phospholipids and triglycerides. The viral glycoproteins are attached to the surface of these LVPs [51]. Recent electron micrographs of HCV particles are in agreement with this structure organization and show a smaller lipid fraction in cell culture-derived HCV particle compared to patient derived [49].

1.2.3 Genome

The hepatitis C virus contains a 9.6 kbase single-stranded positive-sense RNA genome (Figure 1.6). That is, it can directly act as an mRNA for the cellular machinery. The genome contains a single open reading frame (ORF) that is translated to give a ~3000

amino acid precursor polyprotein. The ORF is flanked by very conserved 5′- and 3′-untranslated regions (UTRs) which aid in RNA replication and translation.

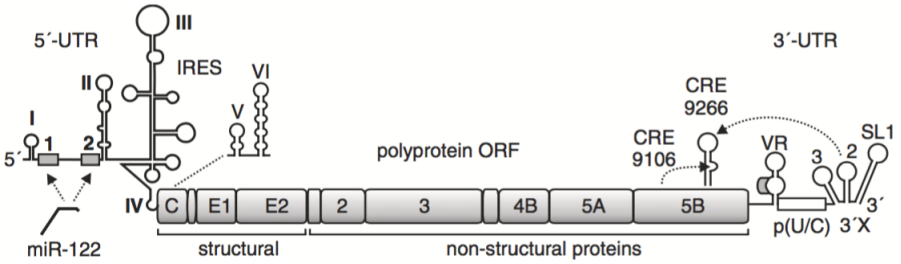

FIGURE 1.6: Organization of the HCV genome. Original caption: *The HCV genome with the 5′- and 3′-UTRs. The two miR-122 target consensus sequences in the 5′-UTR and one in the variable region (VR) of the 3′-UTR are boxed. Cis-acting replicative elements (CREs) and their interactions as well as the stem-loops in the core-coding region are indicated.* From Niepmann [52], Copyright Springer-Verlag Berlin Heidelberg 2013, used with permission of Springer.

5′-Untranslated region

The first 340 nucleotides of the viral genome (JFH-1 isolate) are untranslated and constitute the 5′-untranslated region (5′-UTR). The 5′-UTR contains structures important for replication, notably the first two stem-loops [53]. In this part celullar microRNA-122 (miR-122), a non-coding RNA molecule highly expressed in the liver, binds and stabilizes the viral genome [54]. It also contains a structured region called the internal ribosome entry site (IRES) responsible for recruiting host ribosomes and initiation factors [55] for the translation of viral mRNA by forming a complex with the human 40S ribosomal subunit [56]. It is due to this IRES that the translation of the viral genome by host proteins is possible without the presence of a cap structure [57].

Open reading frame

With the aid of the IRES to align the ribosome, translation starts at a start codon at position 341. The viral genome has a single ORF that codes for a large polyprotein of 3033 amino acids that is cleaved by viral and host proteases to give rise to 10 mature viral proteins.

3′-Untranslated region

The final 236 nucleotides of the viral genome (JFH-1 isolate) are a 3′-untranslated region. It is composed of a variable region, a poly-U/UC sequence and the final 98 nucleotides which form a stem loop called X region. The X region is very conserved. Like the 5′-UTR, this region is highly structured. The 3′-UTR is the site for initiation of viral RNA replication of the positive-sense strand to form the intermediary negative-sense RNA strand. The poly-U/UC and X region are necessary for replication [58, 59]. Additionally, the 3′-UTR stimulates viral translation possibly by retaining host ribosomes at the termination of translation to facilitate efficient subsequent initiation [60].

1.2.4 Proteins

The large polyprotein is cleaved co- and post-translationally to give rise (from amino- to carboxylterminal) to three structural proteins: Core, E1 and E2; and seven non-structural (NS) proteins: p7, NS2, NS3, NS4A, NS4B, NS5A, NS5B (Figure 1.7). These proteins are found within cells associated to the endoplasmic reticulum (ER) membrane. Since they all derive from the same polyprotein these proteins are produced in equimolar amounts [61]. Non-structural proteins NS3 through NS5B are necessary for RNA replication as can be shown from subgenomic replicons that only code for these viral proteins [62, 63].

Core

Core is the first protein to be encoded by the viral ORF. This protein forms the viral capsid. It is cleaved from the following protein, E1, by signal peptidase then further processed by intramembrane cleaving protease SPP to give rise to the 21 kDa mature core protein. Core contains an N-terminal hydrophilic domain that can bind RNA and oligomerize to assist in nucleocapsid formation. Core has nucleic acid chaperone activity to further assist the formation of the nucleocapsid. Its other domain is hydrophobic and mediates the anchoring of Core to lipid droplets by two amphipathic α-helices [66]. Core also assists in virion production by recruiting NS5A to the lipid droplets [67].

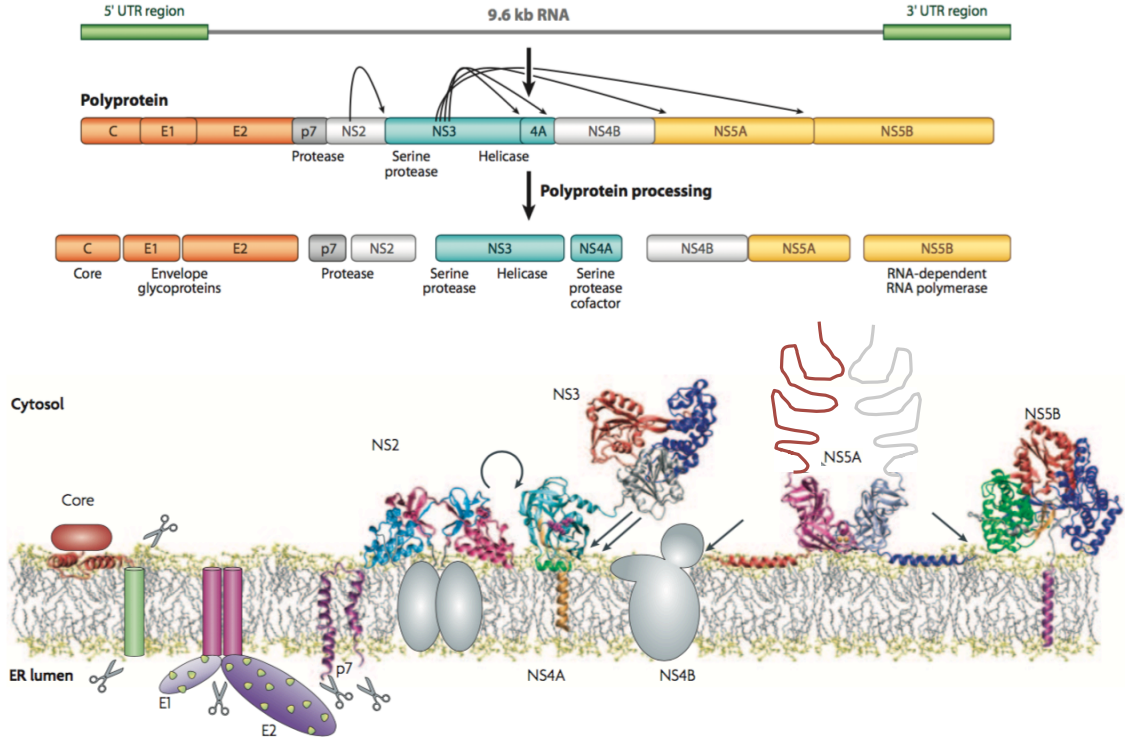

FIGURE 1.7: HCV polyprotein organization. **Top** The viral open reading frame is translated to give rise to a single polyprotein that is cleaved by viral and host proteases. Reprinted from Fusco and Chung [64]. **Bottom** Ten mature ER membrane-associated proteins are produced in this cleavage. Adapted by permission from Macmillan Publishers Ltd: Nature Reviews Microbiology. Moradpour *et al.* [65], copyright 2007.

F-protein

Frameshift (F) or Alternative reading frame protein is a 17 kDa protein produced by a +1 shift of the Core reading frame, with translation starting at the same AUG codon as for Core. It can be detected in patients [68]. Its role is not well understood but recent studies suggest it modulates interferon sensitivity [69].

E1 and E2

E1 and E2 are envelope glycoproteins that are found embedded in the ER-derived phospholipid bilayer that encapsulates infectious virions [70]. Together, these proteins form a heterodimer that is essential to viral entry, fusion and secretion [71]. Each protein possesses a transmembrane helix that anchors it to the membrane and mediates formation of the heterodimer. They also contain structured ectodomains. Parts of these ectodomains have been crystallized independently [72–74]. E2 contains a region that

binds to CD81 (cluster of differentiation 81) to aid in viral entry as well as hypervariable regions that assist in immune evasion. The heterodimers further oligomerize to form a trimer of dimers [75] with three monomers of E1 at the center and three E2 around them.

p7

p7 is a 63 amino acid, 7 kDa integral membrane protein. It is a viroporin, that is, a small viral protein can manipulate membrane permeability for ions [76]. It aids in viral release from host cells. This protein is not necessary for RNA replication but needed for the production of infectious viral particles. Oligomerization in hexameric structures forms a hydrophilic pore that gives rise to this protein's ion channel activity. This activity equilibrates intracellular vesicle pH to promote viral production. Its role appears to be most important during the later stages of viral assembly [77].

NS2

NS2 is an autoprotease that catalyzes the cleavage at the NS2/NS3 junction. NS3 needs to be cleaved to be functional but NS2 itself is dispensable for RNA replication and can be omitted from subgenomic replicon systems [63]. It plays a role in viral assembly since mutations within it can affect HCV infectivity. It has been suggested that it recruits structural and non-structural proteins for viral assembly [78].

NS3/4A protease

NS3 has two domains with distinct activities. Its N-terminal 181 residues, with the aid of NS4A as a cofactor, constitute a serine protease which is responsible for the cleavage of the remaining viral proteins. The role of this protein is not only to cleave the viral polyprotein but also to cleave some host proteins. In particular, through the cleavage of proteins responsible for immune sensing of the virus this protease plays a role in the persistence of HCV [79, 80]. This protein is the target of several direct-acting antivirals that have reached commercialization.

NS3 helicase

The C-terminal two-thirds of NS3 have helicase activity and catalyze ATP hydrolysis to translocate along either RNA or DNA. This region of the protein folds to give rise to three domains responsible for moving along RNA strands to dislodge bound proteins and complementary strands [81]. These three domains all interact with ssRNA during binding, mainly through hydrogen bonds with the backbone of the nucleic acid [82]. The role of this enzyme in HCV replication is still not well understood. Its function can be to help unwind duplex secondary RNA structures for translation and transcription. Additionally, it can also unwind dsRNA after polymerase synthesis and assist in packaging [83].

NS4B

NS4B is an integral membrane protein with amphipathic helices at the amino- and carboxy- terminals [84] and four predicted transmembrane passages. As the only protein without an elucidated crystallographic structure, its function is less well understood. This protein is responsible for the formation of the membranous web, a reorganization of host membranes to form vesicles that contain the HCV replication complexes [85].

NS5A

Nonstructural protein 5A (NS5A) is the most intriguing of the HCV viral proteins as it has no known enzymatic activity [86] but is necessary for replication of the viral genome [87, 88] and formation of viral particles. It is a very promiscuous protein with over 100 possible interaction partners described *in vitro* [86]. Among its partners, it had been shown to interact with the HCV RNA-dependent RNA polymerase, NS5B. However, experiments are conflicting on whether it activates [89–91] or inhibits [92–94] the polymerase. NS5A is associated to the cellular endoplasmic reticulum by an N-terminal amphipathic α-helix membrane anchor [95, 96]. Its cytoplasmic part is organized in three domains separated by two low complexity sequences (LCSs) (Figure

1.8). Domains 1 and 2 are mainly involved in RNA replication, while domain 3 has an essential role in viral assembly.

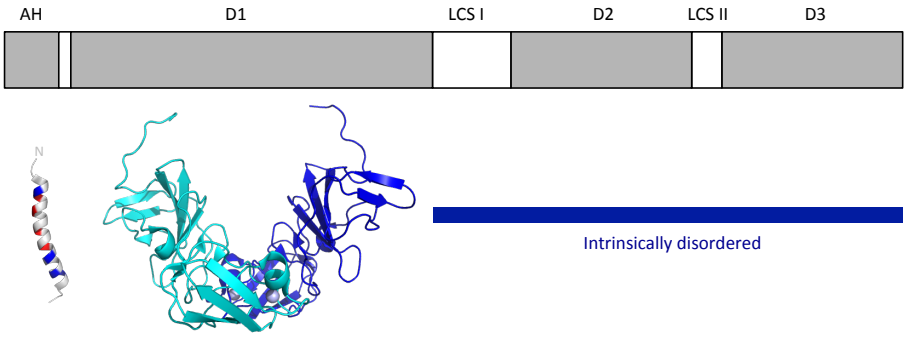

FIGURE 1.8: Domain organization of NS5A. The structures of the amphipathic α-helix (pdb 1R7C) [96] and domain 1 dimers (pdb 1ZH1) [97] are shown under their position on the organization. The blue bar highlights the regions without a fixed three-dimensional structure.

Domain 1 (NS5A-D1) is structured and its crystallographic structures form dimers [97]. Each of its monomers binds a zinc atom in a tetracysteine motif [98]. The dimeric structures have been obtained with multiple organizations [97, 99, 100]. The first of these structures by Tellinghuisen *et al.* shows a basic cleft between the monomers. NS5A has been shown to bind RNA [101]. Taking the structure and RNA binding together, it has been hypothesized that NS5A and, in particular, its domain 1 acts as a track or tether for RNA to coordinate its different fates during HCV replication [102]. Furthermore, this domain and the amphipathic N-terminal α-helix are responsible for DMV formation [103].

Domain 2 (NS5A-D2) is intrinsically disordered as has been shown by nuclear magnetic resonance and circular dichroism [104–106]. It does, however, have regions of transient structure as shown by deviations of carbon chemical shifts from random coil values [107]. NS5A-D2 from the JFH-1 isolate has regions of helical propensity at residues 252-262 and 295-303 and an extended region at residues 278-287 [108]. Interestingly, both NS5A-D2 from the JFH-1 and Con1 isolates present these regions of helical tendency but the higher positive secondary structure propensity (SSP) of NS5A-D2 Con1 suggests that the protein from this isolate samples a helical conformation more often. This is in agreement with SAXS data that shows that NS5A-D2 Con1 presents a smaller radius of gyration [109]. Another structural motif in this domain has been

recently identified in our group. It contains the conserved residues P314 and W316 (numbering from the Con1 isolate, these correspond to P310 and W312 in the JFH-1 isolate) and forms a small turn called PW-turn. This motif is necessary for RNA replication and for the interaction between NS5A and Cyclophilin A (CypA) [110]. CypA is a host protein with peptidyl-prolyl cis/trans isomerase (PPIase) activity, that is, an enzyme that interconverts the conformation of proline residues between cis and trans. This enzyme is necessary for HCV RNA replication [111, 112], a function that seems to be mediated via NS5A-D2. Indeed, this motif is in the overlapped region of the CypA and NS5B binding sites on NS5A [108].

Domain 3 (NS5A-D3) is also intrinsically disordered as shown by NMR spectroscopy [113]. Contrary to domain 2, D3 is dispensable for RNA replication and can support the introduction of inserts [114, 115] in replicon systems. This domain plays an important role in viral packaging and assembly [116]. The role may in part be due to its ability to recruit Apolipoprotein E [117, 118]. For its role in the assembly of infectious viral particles NS5A is recruited by Core [67]. This leads to its association with lipid droplets [119] necessary for the late steps of viral life cycle. Its role is believed to be to help recruit RNA for encapsidation [120]. Supporting this, it was recently shown that D3 contains two distinct regulatory regions that act on the HCV assembly. Firstly, a serine cluster in the C-terminal region recruits the replication complex to Core. Secondly, a highly conserved basic cluster at the N-terminal part delivers the RNA genome to Core [121].

NS5A acts on most steps of the HCV life cycle. It has been suggested this protein is regulated by phosphorylation and different phosphoforms are responsible for early and late steps in the viral life cycle [122]. NS5A is phosphorylated by multiple host kinases such as casein kinase I and II [123, 124], mitogen-activated protein kinase, glycogen synthase kinase 3 [125] and polo-like kinase 1 [126]. This protein exists in two forms depending on its phosphorylation state: p56, the basally phosphorylated state important for RNA replication and the p58 hyperphosphorylated form involved in the production of infectious particles [127]. Evidence for the importance of the basally phosphorylated state for RNA replication comes from the improved RNA replication in the presence of kinase inhibitors [128]. The characterization of the phosphorylation

sites in both forms is still under investigation.

Several phosphorylation sites have been identified. In D1 there is a phosphorylation site that was observed in NS5A from the JFH-1 isolate in S146 that appears to suppress hyperphosphorylation in other sites [122]. However, this serine is not strictly conserved. In the LCS I, there is a cluster of conserved phosphorylation sites. Among these, S222 was identified as phosphorylated in a replicon system and appears to function as a negative regulator of RNA replication [129]. Another site in the LCS I plays in an important role in RNA replication, phosphorylation of S235 promotes RNA replication [124]. This site appears to be phosphorylated by casein kinase Iα [130]. Furthermore, phosphorylation at S235 promotes the phosphorylation at a neighboring site, S238 [131]. Other sites have been identified in LCS II (T348 [122] and T356 [132]).

Recently, NS5A from the Con1 isolate produced in a wheat germ cell-free system and its phosphorylation sites were characterized [109]. Heterogeneous populations arose from the combination and level of phosphorylation. The authors identified sites in LCS I (S222, S230 and S249), D2 (S274 and S306), and D3 (T360, S364, T366, S370, S401, S408, S412, S429, S432, S434, T435 and S437) with varying levels of phosphorylation. Several of these sites in D3 were observed by phosphorylation of D3 with casein kinase II (S401, S408, S412, S429 and S434) [133].

Briefly, NS5A plays important roles in various parts of the viral life cycle possibly modulating the function of other proteins, both viral and from the host.

NS5B

The protein responsible for replication of the viral genome is the RNA-dependent RNA polymerase (RdRp) NS5B. NS5B is capable of replication of RNA through both *de novo* initiation (without the need for a primer) and extension of a primer bound to the template [135, 136]. Its crystallographic structure was determined independently by three groups in 1999 [137–139]. All three crystal structures were of NS5B of the BK isolate (genotype 1b). Since then, over 160 crystallographic structures of the polymerase have been deposited in the Protein Data Bank (PDB) [140], of several genotypes, either on its own, bound to small molecule inhibitors [141–143] and in complex with nucleic acids [144, 145].

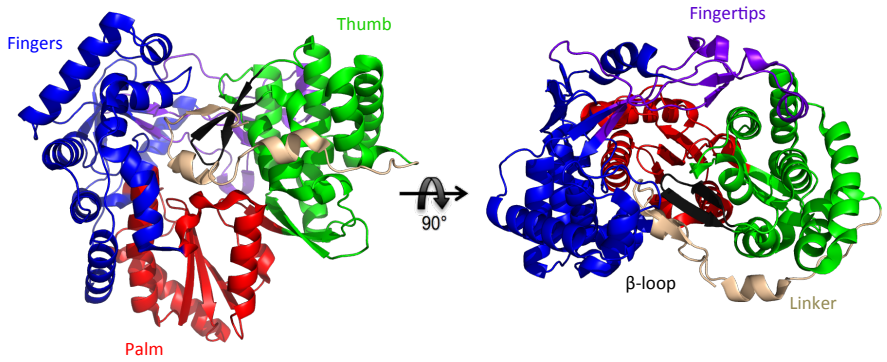

FIGURE 1.9: Subdomain organization of NS5B. Crystallographic structure of NS5B JFH-1 (pdb 3I5K) [134] colored by domain: Fingers (residues 1-188 and 227-287, blue and purple), palm (residues 189-226 and 288-370, red), thumb (residues 371-531, green and black), fingertips (residues 11-45 and 139-161, purple), β-loop (residues 442-454, black), linker (residues 532-566, beige).

NS5B is a protein of about 66 kDa (NS5B from Con1 and JFH-1 isolates weigh 65757.8 and 65668.7 Da, respectively) and 591 amino acids. At its C-terminal end, 21 hydrophobic residues form an α-helix that targets the ER membrane. These residues anchor the polymerase to the the endoplasmic reticulum [146]. Its cytoplasmic part of 570 amino acids has a structure akin to a right hand. It is composed of three subdomains called thumb, fingers and palm (Figure 1.9), like other polymerases, and a linker of 39 amino acids that attaches the globular protein to the transmembrane helix [137]. Its crystal structures present some additional structural features: loops that extend from the finger subdomain (called fingertip loops Λ1 and Λ2) that bind to the thumb subdomain stabilizing a closed conformation and a β-loop that extends into the active site.

Palm The palm is the catalytic subdomain, the most conserved in all known polymerases [145]. It contains the GDD motif responsible for RNA polymerase activity [147]. NS5B binds two Mg^{2+} ions in this domain with conserved aspartates D220, D318 and D319 [137, 148]. These ions can be exchanged for Mn^{2+} and result in 20-fold increased activity [135, 149]. These metal ions are responsible for the coordination of the α and β phosphates of the incoming nucleotides [148]. Allosteric inhibitor binding sites have been mapped to this region and their binding blocks the polymerase in a conformation that is too inaccessible for RNA replication [150].

Fingers The fingers are made up of residues 1-188 and 227-287. This region contains conserved arginine residues R48 and R158 that coordinate the α and β phosphates of the template strand opposite the incoming nucleotides [148]. Additionally, like other *Flaviviridae* polymerases, NS5B has extensions from the fingers to the thumb. These loops called Λ1 and Λ2 or fingertips have contacts with the thumb domain and are responsible for an enclosed active site and leave a narrow tunnel for the access of incoming nucleotides. The Λ1 loop binds to a hydrophobic pocket in the thumb domain at the position of the very conserved Pro 495 [151]. However it is known that these loops can open because the polymerase is able to initiate RNA synthesis from a circular template [152]. Disrupting this contact by introducing hydrophilic residues at the binding region of the Λ1 loop reduces NS5B activity 50-fold [153] suggesting this contact is essential to stabilize an active form of the polymerase.

Thumb The thumb subdomain is the one which varies the most among polymerases [145]. In the case of NS5B, it contains sites for allosteric inhibition. It also contains a motif called β-loop (also β-hairpin or β-flap, in black in Figure 1.9) formed by a two-stranded anti-parallel β-sheet [137]. It is necessary for HCV replication [154] but crystal structures of NS5B in complex with double-stranded RNA with a primer longer than 2 nucleotides required its deletion [145, 148]. This motif extends from the thumb to the RNA binding groove formed between the subdomains reducing the space in the active site for RNA-binding. It has been proposed that this β-hairpin aids in the alignment of the template strand with nucleotides of the complementary strand in *de novo* initiation but is dislocated for primer elongation [148].

C-terminal The C-terminal 21 amino acids form a transmembrane helix that anchors NS5B to the cellular ER membrane. These residues are necessary for RdRp activity *in vivo* but are dispensible for the polymerase activity *in vitro* [155, 156] and are deleted from constructs used in crystallographic structures. The C-terminal part prior to the hydrophobic anchor is found in crystallographic structures wrapped around the thumb domain and with its tip folded into the RNA-binding groove [157]. This part is called the linker (in beige in Figure 1.9). The comparison of crystallographic structures

deleted of 21 and 55 C-terminal amino acids showed that the removal of the linker results in the crystallization of a more open conformation [158]. Additionally, truncations in the linker up to the deletion of 60 amino acids result in increased polymerase activity [157] compared to the deletion of only the transmembrane helix. Molecular dynamics studies show that the presence of the C-terminal linker reduces the flexibility and conformational sampling by NS5B [159]. However, *in vivo* due to the association of the C-terminal to cellular membranes it would be more difficult for the linker to bind to the remainder of the polymerase. Indeed, a recent study using full length NS5B showed improved enzymatic activity after association of the polymerase to a biomimetic lipid bilayer [91].

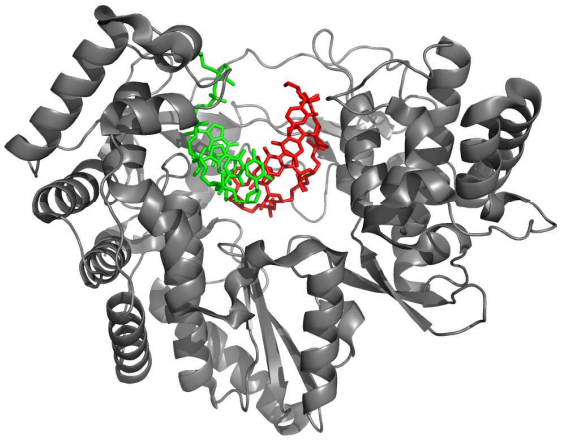

FIGURE 1.10: Crystallographic structure of NS5B bound to double-stranded RNA obtained by Mosley *et al.*[145] with the deletion of the β-loop (PDB 4E78).

The presence of the linker, β-loop and fingertips loops result in an enclosed active site. Until 2012, all known structures of this protein (always crystallized without the transmembrane helix) were considered 'closed' due to the lack of space to fit double-stranded RNA within its active site due to steric hinderance. In 2012, Mosley *et al.*[145] presented an open structure in complex with a short double-stranded RNA fragment (Figure 1.10) where there is a reorganization of the relative positions of the subdomains to allow for the accommodation of the RNA fragment. This was done by deletion of the β-loop. At least two interchangeable conformations are thought to be required

for the function of this polymerase. A closed conformation would be required for *de novo* initiation and an open conformation for elongation of the newly replicated RNA strand. The conformational rearrangement is a rate-limiting event that is believed to occur between *de novo* initiation and processive elongation of the new RNA strand [160, 161].

Recently, Appleby *et al.*[148] published several structures using a different mutant containing 5 point mutations (S15G E86Q E87Q C223H V321I) which stabilize a more open form of NS5B. In some of their structures they used a construct that also had a deletion of the β-loop. This enabled the authors to crystallize the polymerase in different steps of RNA replication and to propose a mechanism for the *de novo* initiation and elongation of RNA during replication (Figure 1.11). In this proposed mechanism initially the linker leaves the active site, allowing for the entry of the template strand and the first nucleotide (*de novo* initiation). The following formation of the first dinucleotide is a slow step. This is followed by another slow step, the rearrangement of the protein, especially the β-loop, to allow for elongation.

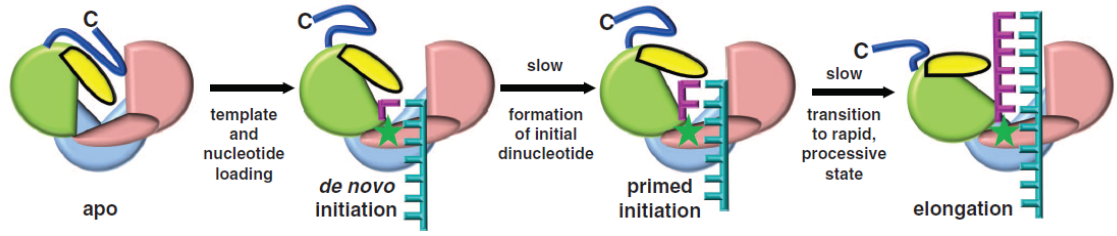

FIGURE 1.11: Mechanism of NS5B RNA replication. In its apo form, a part the RNA binding site is blocked by the C-terminus (blue line) and the β-loop (yellow). For *de novo* initiation, the RNA template and the first nucleotides enter the active site, the β-loop positions the first nucleotide aligned with the beginning of the template strand. The formation of the first dinucleotide is a first slow step. The β-loop leaves the active site and there is a large rearrangement of the relative positions of the subdomains during the transition to the processive elongation state. This rearrangement is a second slow step. From Appleby *et al.*[148]. Reprinted with permission from AAAS.

NS5B is highly error-prone which is the origin of the large genetic variance seen in the hepatitis C virus [65, 162]. These errors give rise to genetic variation even within a single individual giving rise to a population of genetically distinct variants called a "quasispecies" [163]. Depending on the mismatch, the rate of error can reach 10^{-3} per site. However, NS5B can excise nucleotides (and chain-terminating analogues) by

an NTP-mediated mechanism [164] which is nucleotide dependent. The rapid rate of mutation has important implications on the development of resistance to antivirals.

1.2.5 Life cycle

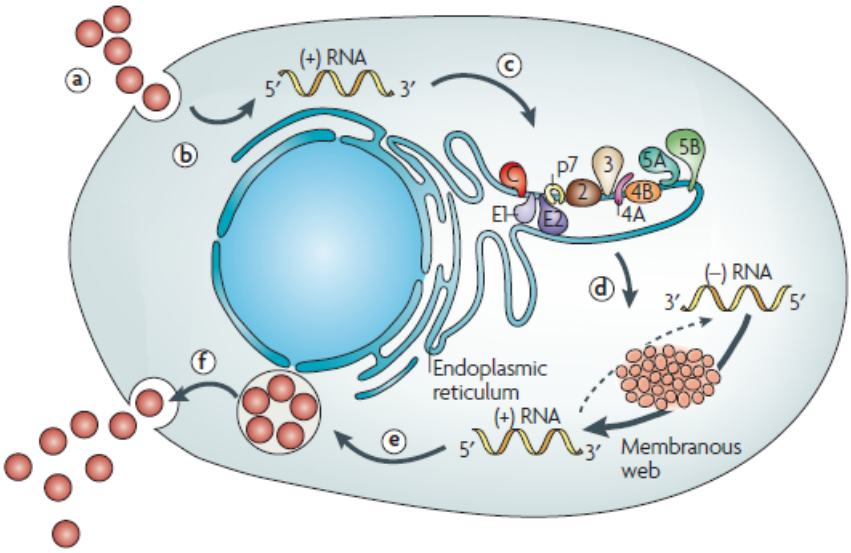

FIGURE 1.12: Life cycle of the hepatitis C virus. a. Entry. b. Release of genomic RNA from nucleocapsid. c. Translation of genome into structural and non-structural proteins. d. Replication of RNA by the replication complex (first to (-)RNA then to (+)RNA). e. Assembly of virions. f. Release of virions. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Microbiology. Moradpour *et al.* [65], copyright 2007.

The life cycle of the hepatitis C virus can be split in several parts: entry and uncoating, translation of the viral genome, replication of viral RNA, assembly of virions and release of infectious particles (Figure 1.12).

Entry and uncoating The first step is the entry of the virus into a host cell (Figure 1.13). The initial step of viral attachment involves the envelope glycoproteins E1 and E2 as well as apolipoproteins on the surface of lipoviroparticles and various cell surface molecules. Firstly, glycosaminoglycans and the low-density lipoprotein receptor seem to be responsible for low affinity binding. Following this initial attachment a multi-receptor complex which involves CD81, scavenger receptor class B member 1, claudin 1, occludin and other host proteins as well as the viral glycoproteins is responsible for the clathrin-mediated endocytosis. The viral lipid membrane then fuses with host endosomal membranes aided by E1 to release the nucleocapsid in the cytoplasm.

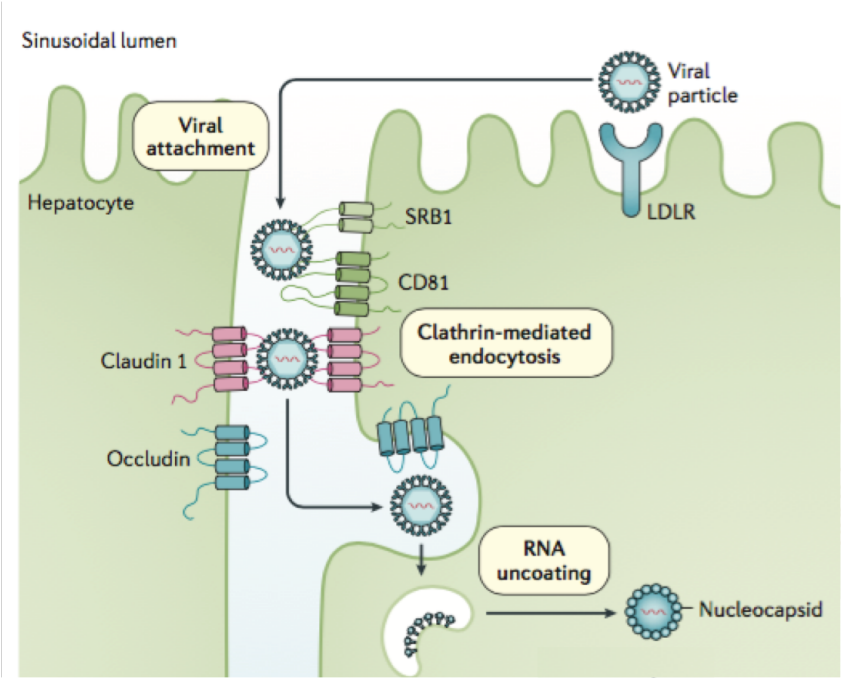

FIGURE 1.13: Entry of the HCV particle into a hepatocyte. Adapted by permission from Macmillan Publishers Ltd: Nature Reviews Disease Primers. Manns *et al.* [165], copyright 2017.

Translation of the viral genome The positive-strand RNA is released into the cytosol with the uncoating of the viral RNA. The 5′-UTR and the first nucleotides of the core coding region form an IRES that together with a 40S ribosomal subunit consitute a binary complex and with eukaryotic initiation factors [55] and a 60S subunit form a 80S complex [166]. Translation of the single open reading frame yields a large polyprotein that is cleaved co- and post-translationally. Core, the envelope glycoproteins and p7 are cleaved by the host endoplasmic reticulum signal peptidase. The NS2/NS3 autoprotease cleaves between NS2 and NS3 and the NS3/4A serine protease is responsible for the cleavage at the remaining sites.

Replication of the viral RNA The non-structural proteins are responsible for RNA replication [62]. They induce a change in the structures of intracellular ER membranes into structures called the membranous web that is made up of double membrane vesicles. These structures emerge as protrusions in the ER membranes but present a different lipid composition compared to the ER membranes [167]. This structure serves as scaffolding for the replication complex. The viral non-structural proteins as well as host proteins are needed to form an effective replication complex. This is centered around

NS5B, the RNA-dependent RNA polymerase. NS5B, from a positive-sense strand, synthesizes a negative-sense strand that acts as a template for new positive strands.

Virus assembly and release Assembly of new virions is believed to happen on the surface of lipid droplets within the cell where Core protein localizes [168]. NS5A can also be found in proximity of the lipid droplets. Given its ability to bind RNA and its presence in the replication complex it has been suggested that NS5A acts as a carrier of RNA between steps of the life cycle. Positive-sense RNA strands serve as new templates and are progressively assembled into infectious virions which are released from the cell by the very low-density lipoprotein (VLDL) secretory pathway[44, 65, 169].

1.2.6 Models

In vitro models

In 1999, Lohmann *et al.* [62] developed a replicon system capable of producing viral RNA in a human hepatoma cell line, Huh7. A replicon is a nucleic acid polymer (DNA or RNA) capable of replication. Lohmann and colleagues achieved this using a subgenomic replicon composed of the 5'-IRES, the neomycin phosphotransferase gene (which grants resistance to the antibiotic G418), the EMCV-IRES and HCV sequences from NS3 to the HCV 3' end. This was done with a consensus genome derived from RNA isolated from a patient chronically infected with genotype 1b (Con1). The replicon was transfected into human Huh7 cells and under G418 selection, colonies capable of replicating the replicon RNA were isolated.

Since then, replicon systems have been developed for various isolates of several genotypes [170–172]. Additionally, replicon systems have been improved to enhance replication by selecting Huh-7 cells that were more permissive for HCV RNA replication (for example Huh-7.5) and introduction of adaptive mutations within the ORF. Such systems have been widely used to study RNA replication in HCV-infected cells [173].

To study viral entry, another viral model was developed. Retroviral pseudoparticles (HCVpp) can be used to study HCV entry into hepatoma cells. To produce these particles, human embryo kidney 293T cells are transfected with three plasmids coding,

firstly, for HCV glycoproteins E1 and E2; secondly, for retroviral Gag and Pol proteins, necessary for RNA packaging and encapsulation; and finally, for a reporter gene (either luciferase or GFP). This results in membrane coated particles displaying E1 and E2 tropism and can be used to infect Huh-7 cells and study the entry step of HCV infection independently [174, 175].

One of the major limitations of these model systems is the lack of production of viral particles. However, full-length HCV genomes were not able to sustain infection that could be subsequently transmitted. Nonetheless, in 2005, cell-culture derived HCV (HCVcc) particles were developed based on a sequence isolated from a Japanese patient with a fulminant form of hepatitis [176]. The full sequence could be transfected into hepatoma Huh-7.5 cells without the need for adaptive mutations and was replicated efficiently [171]. This isolate, called JFH-1 (for Japanese Fulminant Hepatitis), was shown to be capable of producing infectious particles that could be used to infect chimpanzees [177–179] producing the first cellular model of the full viral life cycle.

JFH-1, as a genotype 2a isolate, is not representative of the majority of infections worldwide, which are mainly from genotype 1. To obtain HCVcc of other genotypes, chimeric sequences of JFH-1 and other genotypes have been developed [175] that introduce culture-adaptive mutations into other HCV isolate sequences.

Even though these HCVcc particles are able to reproduce the full HCV life cycle and have been instrumental to the development of antiviral drugs that target the virus specifically, they have limitations that arise from the use of cancer-derived hepatoma cells. These cells present differences compared to non-malignant primary human hepatocytes. For example, hepatoma cell lines are deficient in lipoprotein assembly [51] which has been proposed to result in differing virion structure [48, 49] compared to primary hepatocytes. In 2010, Banaudha *et al.* [180] reported a culture system of primary hepatocytes capable of supporting the replication of infectious clones of HCV genotypes 1a, 1b and 2a without the need for culture-adaptive mutations. This provides a model that represents the host-virus interactions more accurately. For example, recently Harak and coworkers [181] showed that HCV requires only low levels of a host lipid kinase called phosphatidylinositol-4 kinase IIIα (PI4KA) that is activated by NS5A. Huh-7-derived cells overexpress this protein compared to primary hepatocytes.

Culture-adaptive mutations in HCVcc models reduce the activation of these proteins which allow the virus to replicate despite the high expression of the host protein.

Animal models

Chimpanzees (*Pan troglodytes*) are the only species that can develop full hepatitis C after infection. They were used in the initial discovery of the etiological agent responsible for the disease [3]. Many characteristics of human and chimpanzee infection are very similar including viral titers and evolution of pathology to chronicity, despite a milder disease in chimpanzees without the presence of fibrosis [182]. However, the elevated cost and rising ethical concerns as well as the development of other models have widely stopped the use of chimpanzees as models for HCV [183, 184].

The tupaia (*Tupaia belangeri*) is a small tree-shrew that was considered as a small animal model for HCV infection due to the capacity of infecting its primary hepatocytes with HCV from humans [185]. Experimental inoculations of tupaias with HCV resulted in lower viremia, however, studies show that tupaia pathogenesis resembles that of humans with chronicity, liver steatosis and tumorigenesis within 3 years [186].

Ideally, the goal is to be able to research HCV in a small animal model that is largely available. Mice are resistant to HCV infection due to the lack of viral entry factors specific to humans. By adenoviral delivery of genes coding for these entry factors for the transient expression of human CD81 and OCLN proteins it was possible to obtain a mouse model for viral entry [187]. By introducing four human entry factors and using mice deficient in several immune pathways, in 2013, Dorner and colleagues reported the development of genetically humanized mice in which the entire HCV life cycle is reproduced [188].

Another strategy for using mice as models for HCV infection is to transplant human hepatocytes and have them populate the mouse liver. To achieve this, the mice need to be immune-deficient to avoid rejection of the graft [189, 190]. These models are quite robust and can be infected with both HCVcc and patient-derived virus [190, 191]. These liver chimeric mice can be used to test novel antiviral strategies. However, due to their immunodeficiency they cannot be used as models for immune response to HCV. Mice transplanted with both $CD34^{+}$ human hematopoietic stem cells and hepatocyte

progenitors have allowed the study of the T-cell response [192] and can be used for the study of pathogenesis and drug response [193].

1.3 Hepatitis C replication complex

Characteristic of positive-sense RNA viruses, HCV infection causes alterations in the membranes within the host hepatocytes [194]. Due to their appearance these invaginations and protrusions in the case of HCV are called "membranous web" (Figure 1.14). It was later found that the membranous web was made up of mainly of double membrane vesicles (DMVs) of about 200 nm [195]. This sort of organization in vesicles was shown to be induced by the viral non-structural proteins, NS4B and NS5A, in particular [167, 196, 197].

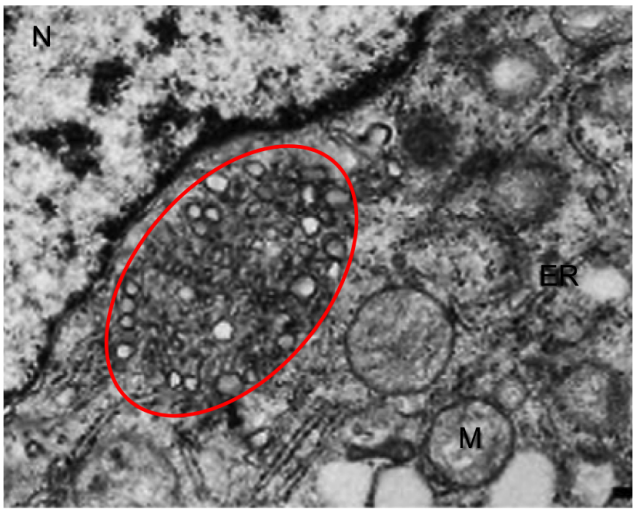

FIGURE 1.14: "Membranous web" structures found in HCV-infected cells. Letters ER, M and N represent the endoplastic reticulum, mitochondria and nucleus, respectively. From Gosert *et al.* [198] and Moradpour *et al.* [156]. Reprinted from Moradpour *et al.* [156], copyright 2003, with permission from Elsevier.

DMVs were found to contain double-stranded RNA (dsRNA) and enzymatically active RNA polymerase [167] as well as all other non-structural proteins [199]. This is in agreement with earlier findings that HCV RNA is found mostly in a detergent-protected environment [200]. These detergent-resistant vesicles are proposed to form the hepatitis C virus replication complex.

Replication complexes are supported onto the modified membranes (Figure 1.15). The main protein of the complex is the RNA-dependent RNA polymerase NS5B that is

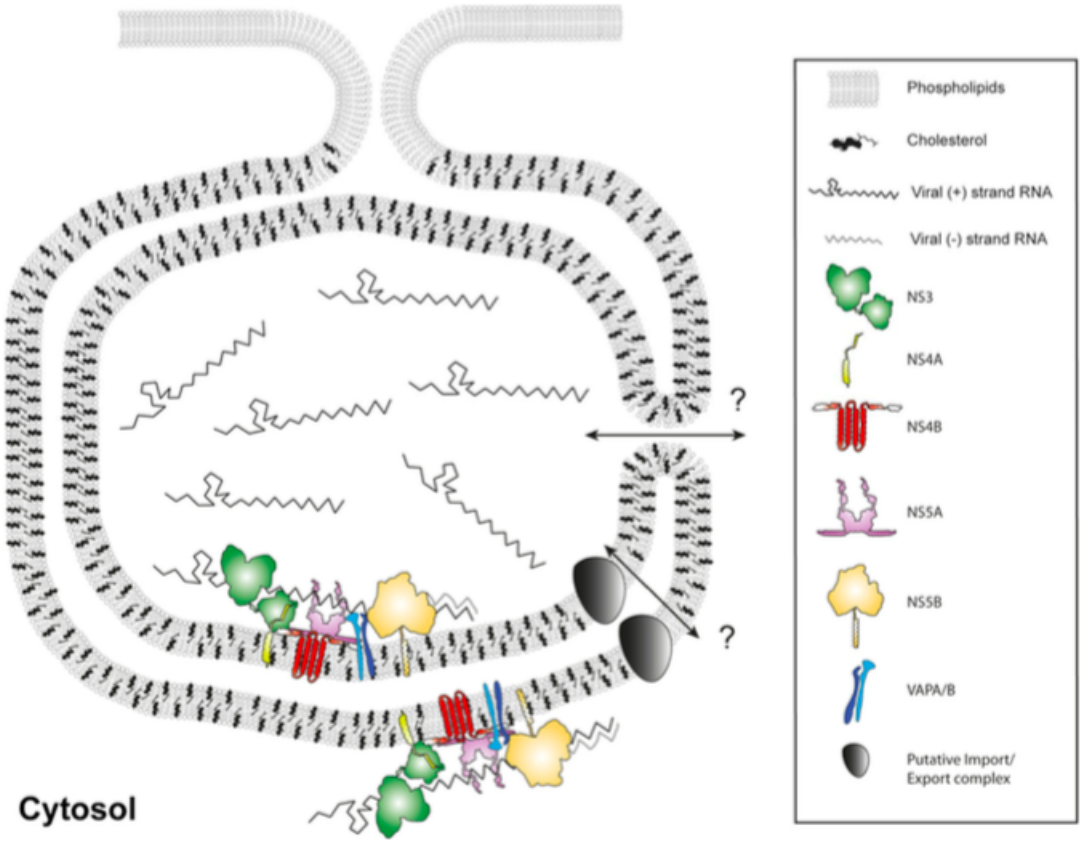

FIGURE 1.15: Schematic model of HCV replication complex in a double membrane vesicle (DMV). The majority of viral RNA is protected from nuclease activity inside the vesicles. Reprinted from Paul *et al.* [167], copyright 2013 American Society for Microbiology.

anchored to the membrane through an carboxyterminal transmembrane helix. However, for the replication complex to be active it needs other viral non-structural proteins. NS3 and NS5A have been shown *in vitro* to bind NS5B when it is tethered to a lipid bilayer. NS5A also increases replicase activity after a lag period under these conditions [91]. The lag period was attributed to a conformational rearrangement of the NS5B-NS5A-RNA system. Immuno-electron microscopy showed that NS5A and dsRNA co-localized within the vesicles [195]. Despite their precise role in replication not being fully understood it is likely related to the helicase activity for NS3 and RNA binding and conformational rearrangement for NS5A.

Host factors also play an important role in the formation of the replication complex. One such protein is the lipid kinase phosphatidylinositol-4 kinase IIIα (PI4KA) [201]. It catalyzes the conversion of phosphatidylinositol to phosphatidylinositol 4-phosphate (PI4P). Mediated by NS5A, its product accumulates in HCV infected cells [202, 203].

PI4P can recruit lipid transport proteins, for example oxysterol-binding protein whose inhibition reduces the transport of cholesterol to DMVs [204]. DMV membranes are enriched in cholesterol 9-fold compared to ER membranes, as cholesterol appears to have a structural function in these vesicles [167].

Vesicle-associated membrane protein-associated proteins (VAPs), in particular VAP-A and VAP-B, are host proteins involved in HCV replication through interaction with NS5A and NS5B [205, 206]. These proteins co-localize with NS5A in the DMVs [167]. The polymerase's C-terminal linker [207] and domain 3 of NS5A [208] interact with overlapping binding sites on VAP-B.

There is however a paradox in the localization of the replicase complex within DMVs. Only 8% of vesicles presented visible openings that would allow the entry of nucleotides and release of nascent RNA [197]. Some hypotheses have been proposed to explain this apparent incoherence [197, 199]. Possibly only a fraction of DMVs are replication-competent or the transport of NTPs and other necessary factors is mediated by unidentified protein channels. Another hypothesis is that the replicase complex is located on the exterior surface of the DMVs but this is inconsistent with the resistance to nuclease activity of the viral RNA [167].

Within infected cells, Quinkert *et al.* [61] found on average 40 copies of negative-strand RNA. Compared to the negative-strand RNA, they found a 5-fold excess of positive-strand RNA and a 20 000 to 40 000-fold excess of viral proteins. This is in agreement with Caroll *et al.* [209] who showed that less than 1% of NS5B is catalytically active. Quinkert *et al.* [61] also proposed that since each replication complex requires a negative-strand RNA, 40 is the upper limit to the number of active complexes per cell.

1.4 Treatment

In the last decade, the treatment of hepatitis C has greatly evolved. Until 2011, treatment was based on interferon and broad spectrum antivirals. Improvement in experimental models, notably the possibility to reproduce the entire cell cycle in culture, has led to the development of drugs that target viral proteins specifically.

1.4.1 Treatment before 2011

Since the identification of non-A, non-B hepatitis there has been a need to treat patients with this form of hepatitis. The rates of "cure" are measured through a value called the sustained virologic response (SVR) which means undectectable serum HCV RNA level 24 weeks after the termination of treatment [210]. In 1986, prior to the identification of the virus responsible for the disease, treatment was suggested with recombinant human alpha interferon [211]. Interferons are a group of human proteins released by host cells in the presence of pathogens to modulate the immune response. Many viruses have evolved resistance mechanisms against interferons. HCV NS5A protein, for example, inhibits intracellular interferon alpha [212, 213]. Treatment with interferon alpha for 48 weeks allowed 15-20% patients to achieve SVR [210].

Through to 1998, interferon continued to be the main treatment option for patients. In this year, ribavirin, a guanosine analogue broad spectrum antiviral was added to improve the percentage of patients that achieved SVR to 31-38% [214]. In 2001, to reduce the clearance of interferon in the body it was fused with polyethylene glycol (PEG) to form pegylated interferon (PEG-IFN) [43]. However, treatment lasted 24-48 weeks depending on genotype with significant side-effects. Additionally, 17% of patients are not eligible for combined interferon and ribavirin treatment [215, 216]. Moreover, the response to interferon-based treatment was very dependent on genotype. Genotype 1, the main genotype in Western countries, responds to interferon-based therapy in less than 50% of patients [217].

The need for improvement of SVR rates, especially for genotype 1 patients, and the reduction of side-effects drove the development of drugs that target the hepatitis C virus specifically and revolutionized its treatment.

The goal of drug development for HCV was to find a treatment that had the following characteristics: (a) short treatment duration; (b) without side-effects; (c) be pangenotypic, that is, be active on all HCV genotypes; (d) be taken orally, ideally with a single pill; (e) be interferon-free, so patients that are not eligible for interferon treatment can also use it; (f) low costs.

1.4.2 Direct-acting antivirals

Direct-acting antivirals (DAAs) specifically target viral proteins. In the case of the hepatitis C virus, these drugs are mainly of three classes depending on the protein they target: (1) Protease inhibitors, the first to be approved, target the NS3/4A protease. (2) NS5A inhibitors target the RNA-binding domain of NS5A. (3) Polymerase inhibitors target NS5B to inhibit RNA replication.

Protease inhibitors

Protease inhibitors block the NS3/NS4A protease complex. This inhibits the cleavage of the polyprotein into the mature non-structural proteins. Boceprevir and telaprevir were the first DAAs to be approved for therapeutic use. Since then, simeprevir, paritaprevir and grazoprevir have been approved by the U.S. food and drug administration (FDA). All of these target the active site of NS3. However, these drugs present narrow genotype specificity and low barrier to resistance [218]. Allosteric inhibitors of the HCV protease have also been described and bind to the interface between the NS3 protease and helicase domains [219, 220].

NS5A inhibitors

Due to the lack of enzymatic activity, NS5A inhibitors appeared as a viable antiviral strategy only after development of cell-based screening which identified the iminothiazolidinone BMS-858 as an inhibitor of RNA replication. After extensive chemical refinement, Daclatasvir (previously known as BMS-790052) was identified as a potent HCV replication inhibitor [221, 222]. Since NS5A is involved with both RNA replication and viral assembly [116] its inhibitors have great potential [223]. Indeed they are shown to be very potent antivirals with low and sub-picomolar activities in cells [221, 224].

Daclatasvir is the basis for the structures of most other NS5A inhibitors (Figure 1.16). The symmetric nature of their structures suggest they bind to a symmetric protein structure [225]. Indeed, mapping of the resistance mutations to daclatasvir suggests that it targets the interface between domain 1 of NS5A and the ER membrane

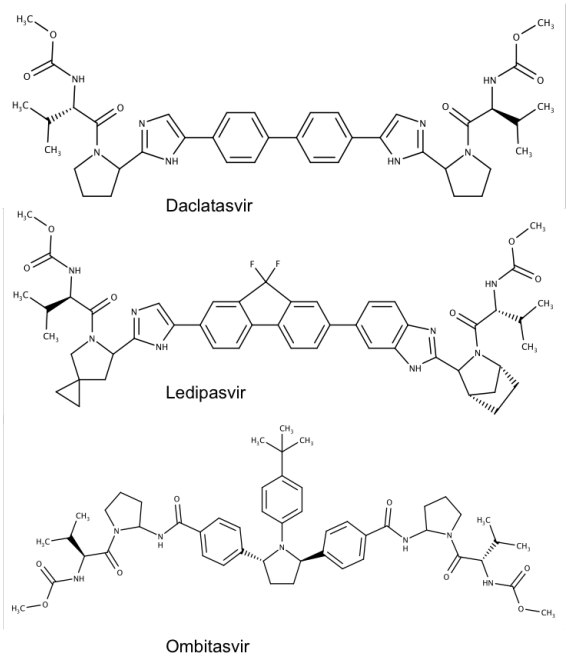

FIGURE 1.16: Examples of NS5A inhibitors that have reached clinical use. Their structures are derived from the symmetric daclatasvir.

[226, 227]. Attempts to co-crystallize NS5A with an inhibitor have not been successful but molecular docking studies suggest that these small molecules favor a dimeric conformation unable to bind RNA [228, 229].

NS5A inhibitors have the highest potencies among HCV antivirals with effective doses in the pM range [218]. Several NS5A inhibitors (names ending in -asvir) have been approved for therapeutic use. As of April 2017, daclatasvir, velpatasvir, ledipasvir, ombitasvir and elbasvir are to be used in combination with other DAAs without the use of interferon [230]. However, resistance-associated variants have high replication fitness. Presence of high baseline levels of resistance substitutions to NS5A inhibitors seems to be an important factor whether NS5A inhibitor-based therapy is effective and screening for these mutations before initiating treatment is currently a subject of debate [231].

Recently, a new class of NS5A inhibitors has been discovered whose resistance mutations are localized in domains 2 and 3 of NS5A without mutations in NS5B [225].

Nucleos(t)ide inhibitors of the HCV polymerase

Nucleos(t)ide inhibitors are nucleoside/nucleotide analogues. They are incorporated in the nascent RNA strand by the viral polymerase and block further incorporation of nucleotides. These inhibitors possess a 3′-hydroxyl group. Their activity arises from steric hindrance at the 2′ position [232]. These analogues are functionalized to be cleaved by hepatic enzymes avoiding systemic exposure [233]. They become active by phosphorylation by host enzymes.

Sofosbuvir is the only compound of this class to have reached approval [230]. It was first approved for use in combination with interferon and ribavirin or just ribavirin in 2013, but has since been approved for use with DAAs of other classes without the need for interferon or ribavirin. This class of inhibitors has the highest barrier to resistance among DAAs. A serine to threonine substitution at position 282 was identified as a major resistance mutation for sofosbuvir [234] and other nucleos(t)ide inhibitors [235]. Crystal structures of NS5B in its elongation assembly with nucleotides or the active metabolite of sofosbuvir showed that S282 is part of the hydrogen bonding network involved in recognizing the incoming nucleotide [148]. This mutation reduces the replication fitness of HCV in replicon models [234].

Allosteric inhibitors of the HCV polymerase

Non-nucleotide inhibitors (NNIs) bind to different allosteric sites on the protein and inactivate it. Structures with NNIs have shown five binding pockets, in the thumb (I and II) and palm (I, II and β) domains (Figure 1.17) [239, 240].

The thumb-I binding pocket is located at the contact between the fingertips Λ1 loop and the thumb domain. Binding of an inhibitor to the hydrophobic pocket at the very conserved [151] Pro 495 disrupts the contact between the two domains of the polymerase. In the crystal structures of NS5B co-crystallized with inhibitors of this type, the fingertips loop becomes too flexible to identify its electron density [143, 241, 242]. Disrupting this contact is believed to uncorrelate the movement of the thumb and fingers. It has been suggested that this locks the enzyme in an open conformation [241]. However, molecular dynamics suggests that the polymerase is not locked but actually

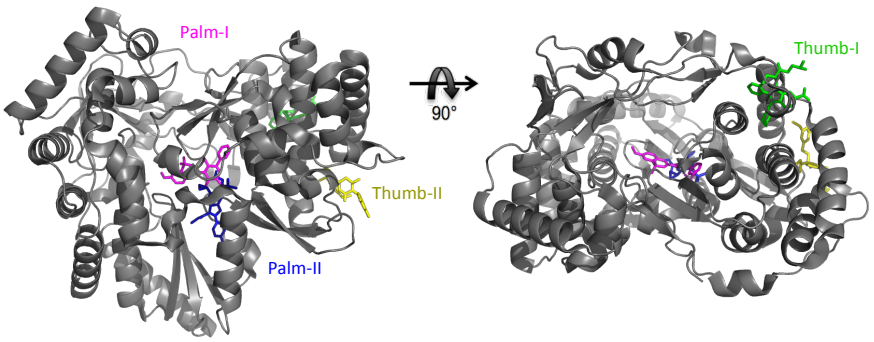

FIGURE 1.17: Allosteric inhibitor sites on NS5B. Cartoon representation of NS5B (PDB 3I5K [134]) featuring the non-nucleotide inhibitor sites. MK-3281 (green, from PDB 2XWY [143]), Filibuvir (yellow, from PDB 3FRZ [236]), a pyridazone inhibitor (magenta, from PDB 3BSC [237]) and Nesbuvir (blue, from PDB 4TLR [238]) were positioned through alignment and are shown in stick representation to highlight the positions of the allosteric inhibitor sites except Palm-β for which no representative crystallographic structure is available. Adapted from Caillet-Saguy et al.[158]

rapidly samples multiple conformations, both open and closed [243]. Of the inhibitors of this class, deleobuvir (previously BI 207127) was the most advanced in clinical trials, having reached phase III but development was discontinued due to insufficient efficacy [244, 245].

The thumb-II binding site is located at the base of the thumb. Classes of molecules that bind to this pocket include thiophene-2-carboxylic acids like lomibuvir (previously VX-222) and dihydropyranones like filibuvir (previously PF-00868554) [246]. Resistance mutations to these inhibitors appear mainly located on the hydrophobic residues of the binding pocket: L419, M423, M426 and I482 (genotype 1). In monotherapy these inhibitors have a low barrier to resistance [247]. These inhbitors cause only small differences in conformation compared to the apo enzyme in crystallographic structures [158]. However, inhibitors of this type are believed to limit the dynamics of the polymerase, in particular blocking the conversion between closed and open conformations needed to switch between *de novo* initiation and elongation [248, 249].

The palm-I pocket is at the contact between the thumb and the palm subdomains, near the enzymatic active site. A system of hydrogen bonds between the inhibitor and the enzyme block NS5B in a closed conformation [243] and prevent the access of substrates to the active site [150]. Molecules that bind to this pocket are benzothiadiazine derivatives. One such example, dasabuvir (ABT-333) has been approved for clinical

use. Resistance to these inhibitors occurs with the mutation S556G within the binding pocket [250]. Analysis of crystallographic structures of NS5B bound to inhibitors of this class showed either a structure indistinguishable from the apo form or a small displacement of the central fingertips and rotation of the thumb [158].

The palm-II pocket partially overlaps with the palm-I pocket. Benzofurans are the only class of molecules found to bind to this site and nesbuvir (HCV-796) is the model inhibitor of this class. Nesbuvir binds to the polymerase through an induced-fit mechanism [142]. Like for palm-I inhibitors, nesbuvir causes a rotation of the thumb but in the opposite direction. In this case, the rotation is due to the displacement of the polymerase to open a binding subpocket [158].

The most recent class of allosteric inhibitors of NS5B bind to the Palm-β site. The resistance mutations to this type of inhibitor appear near the active site as well as in the thumb β-loop. The most advanced inhibitor of this type is tegobuvir (GS-9190). This molecule does not directly bind to the polymerase but undergoes metabolic activation to bind covalently to the polymerase [251].

1.4.3 Host-targeted antivirals

Another strategy for HCV inhibition is targeting the host biomolecules it requires for replication. This approach, alternative to the direct-acting antivirals, has been developed to avoid resistance and achieve pan-genotypic coverage [252]. These include cyclophilin A (CypA), phosphatidylinositol-4 kinase IIIα (PI4KA) and microRNA 122 (miR-122).

MicroRNAs (miR) are non-coding strands of RNA produced by many eukaryotic cells involved in gene expression. In general their role is to negatively regulate expression by binding to mRNA targets [253]. However, in the case of the HCV genome, a liver-specific miR, miR-122, actually positively regulates the replication of HCV RNA [254] by binding to two sites on the 5'-UTR (Figure 1.6). This nucleic acid polymer together with a human argonaute protein, a human protein that also typically downregulates expression, stabilizes the uncapped viral RNA from degradation [255]. A nucleic acid, miravirsen (previously SPC3649) is the main miR-122 antagonist. Its structure is a

modified complementary strand of miR-122 [256]. Clinical trials for this molecule have reached phase 2 [256, 257] and have shown high barrier to resistance [258].

The lipid kinase phosphatidylinositol-4 kinase IIIα is one of the host proteins necessary for RNA replication. It is activated by NS5A-D1 [259] and its product is believed to be an essential component of the replication complex membrane [260]. Conversely, PI4KA regulates the phosphorylation state of NS5A [261]. PI4KA inhibitors were first thought to inhibit HCV replication by targeting NS5A instead of a host protein [262]. One such inhibitor, AZD-2836 reached phase I-II clinical trials but was discontinued [263]. Development of new inhibitors is ongoing since the discovery of a novel mechanism of inhibition [264].

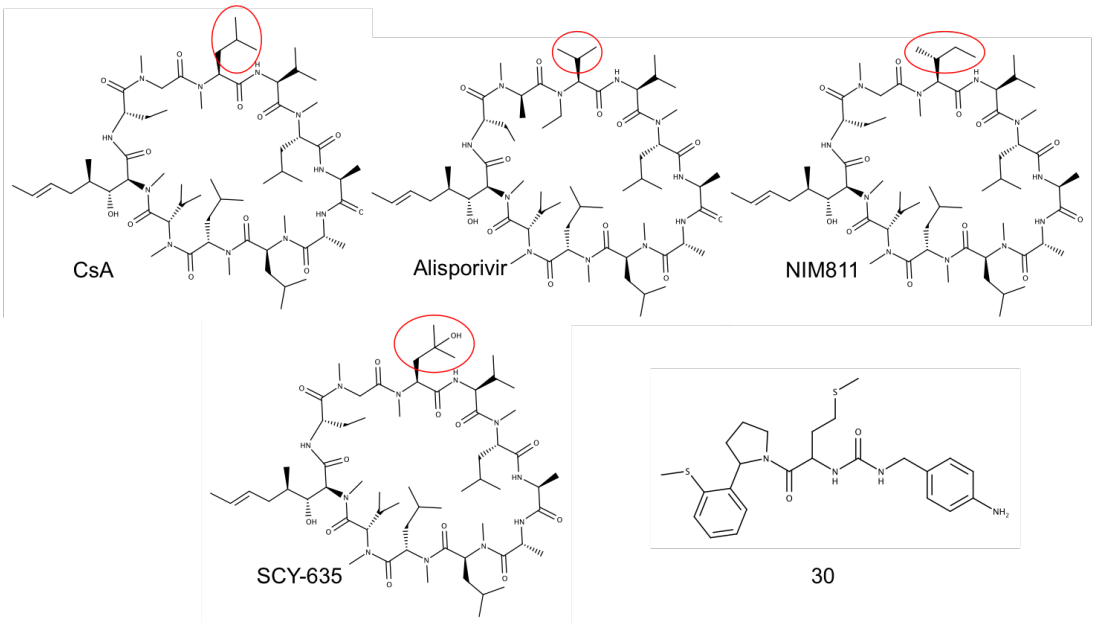

FIGURE 1.18: Structure of Cyclophilin A inhibitors. Cyclosporine A (CsA) and its non-immunosuppressive analogues alisporivir, NIM811 and SCY-635 with their difference in structure highlighted as well as an example of a novel class of cyclophilin inhibitors (30) described by Ahmed-Belkacem *et al.* [265].

Cyclosporine A (CsA), a well-known immunosuppressant, was identified as inhibiting HCV RNA replication in a replicon system [266]. This was shown to be due to its ability to inhibit cyclophilins. Cyclophilins (Cyp) are peptidyl-prolyl cis/trans isomerases (PPIases), thus catalyze the conversion between the trans and cis configurations of the amino acid proline. There was some debate about which cyclophilin isoform is biologically relevant in HCV replication but now evidence points to cyclophilin A [111, 112]. Cyclosporine A is immunosuppressive because, when it is bound to the

Cyp active site, Cyp can bind to calcineurin, a human phosphatase involved in immune signaling [267]. Since HCV replication inhibition is due to the inhibitor binding to the Cyp active site [112], development of Cyp inhibitors focused on removing the immunossupressive interaction with calcineurin.

Alisporivir (DEB025), SCY-635 [268] and NIM811 [269] are the main examples of this class of anti-HCV drugs (Figure 1.18). Clinical trials with alisporivir were put on hold at phase III due to pancreatic toxicity in combination with interferon [270]. Phase I and II trials have since been restarted with alisporivir in the absence of interferon [271].

Resistance to CsA and its derivatives has been mapped mainly to the domain 2 of NS5A and NS5B [272, 273]. Congruently, direct interaction between CypA and NS5A *in vitro* [111] and D2 in particular [105] has been described. Additionally, NS5A-D2 can act as a substrate for the cyclophilin PPIase activity. However, our group has previously shown that CypA does not directly interact with the polymerase [108]. Therefore, the effect that CypA has on RNA replication is expected to be mediated via NS5A-D2.

CypA is involved in the life cycle of various viruses and thus its inhibitors are being investigated to be applied to other diseases [274–276] and novel inhibitors are being researched. Ahmed-Belkacem *et al.* [265] recently described using fragment-base drug discovery to design a non-peptidic cyclophilin inhibitor without immunosuppressive properties.

1.4.4 Current treatment options

Since 2011, treatment has rapidly evolved and changed. The first direct-acting antivirals to be approved, telaprevir and boceprevir, both protease inhibitors, were used in combination with interferon and ribavirin. Direct-acting antivirals are often used in combination to avoid the development of resistance mutations. This combination therapy had improved SVR but had a larger side-effect profile.

Sofosbuvir, a nucleotide inhibitor, was approved in 2013 and revolutionized the treatment of HCV with SVR in over 90% of patients in combination therapy. Soon after, additional direct acting antivirals were approved for use and combination therapies without interferon or ribavirin became standard of care. The effectiveness of current

treatment protocols is such that the first-generation DAAs, telaprevir and boceprevir, have been discontinued [277, 278].

Current standard of care is dependent on genotype and whether the patient has compensated or decompensated cirrhosis (Table 1.1). Interferon-based therapies are no longer recommended by the American FDA.

	sofosbuvir ledipasvir	sofosbuvir velpatasvir	sofosbuvir daclatasvir	sofosbuvir simeprevir	ombitasvir paritaprevir ritonavir dasabuvir	ombitasvir paritaprevir ritonavir ribavirin	elbavir grazoprevir
Genotype 1	■	■	■		■		■
Genotype 2		■	■				
Genotype 3		■	■				
Genotype 4	■	■	■	■		■	■
Genotype 5	■	■	■				
Genotype 6	■	■	■				

TABLE 1.1: Summary of recommended treatment for patients with or without cirrhosis. Boxes filled in black represent drug combinations that are recomended for patients with the genotype on the left. Patients with compensated cirrhosis should not use treatments with protease inhibitors (gray background). Adapted from the 2016 EASL guidelines [279].

For genotype 1, previously the hardest genotype to treat, many combinations have been approved. All of these combine two or more types of DAAs. In particular, there are many treatment options for patients without cirrhosis or with compensated cirrhosis. These include sofosbuvir combined with either ledipasvir, simeprevir, velpatasvir or daclatasvir. In late 2014, an ombitasvir, paritaprevir, ritonavir and dasabuvir combination was approved and has since been included as a recommended treatment for these patients. Most recently, elbasvir and grazoprevir have also been approved [280]. All of these treatments last for 12 weeks and have very low side-effect profiles. This is a major improvement over interferon-era 24–48 week treatments.

All genotypes have treatment options since sofosbuvir in combination with either velpatasvir or daclatasvir is recommended for all genotypes. Indeed, effective treatment is available for most patients, even those who did not have such options previously. In patients with decompensated cirrhosis, defined as the presence of at least one of the following symptoms: jaundice, ascites, variceal hemorrhage, or hepatic encephalopathy, options are available despite the fact that the use of NS3 protease inhibitors is contraindicated because of lower protease inhibitor clearance [279]. Recently, sofosbuvir with either ribavirin or ledipasvir was approved for use in pediatric patients, a group in which the only treatment options were interferon-based therapy [281].

Treatment of patients with HCV/HIV co-infection should choose the HCV antiviral treatment depending on the antiretroviral therapy to avoid drug-drug interactions [6]. Similarly, DAAs can have interactions with lipid lowering drugs, anti-depressants, anti-psychotics, cardiovascular drugs and immunosuppressants. Thus patients taking these medications concomitantly might require adjustments in their DAA dosage, increased monitoring or even contraindicate the use of some DAAs [279].

Treatment failure, despite becoming less frequent, still occurs in 5-10% [280] of patients. It is usually tied to the underlying presence or development of resistance-associated substitutions (RAS) [279]. The resistance profile and resistance barrier is dependent on the class of the DAA [218]. For example, the resistance barrier to NIs is high. For sofosbuvir the main RAS is S282T. Since this mutation reduces the replication fitness of HCV this RAS tends to disappear after termination of an unsuccessful treatment and repeat treatment with sofosbuvir is recommended even in the case of failure. The profiles of NS5A inhibitor RASs are dependent on the particular drug but these mutations result in a virus with high replication fitness. Protease inhibitors and dasabuvir (NS5B palm-I) can develop RASs near their binding sites and, like for NS5A inhibitors, repeat treatment in treatment-experienced patients should select a combination that does not have the same resistance profile and should ideally target different viral proteins [279].

HCV treatment options have become much more varied, efficient and with a smaller side-effect profile in recent years. However, the elevated price tag for the new DAA-based protocols precludes the treatment of all HCV-infected patients at this time. The continued study of HCV and its cellular and biochemical mechanisms can aid in the development of newer, more cost-effective treatment options as well as improve our understanding of mechanisms common to other viruses [282].

1.5 Hepatitis C virus infection affects cellular metabolism

To aid in its adaptation within the host, the hepatitis C virus can modify the metabolism of the host. It has been described that the virus modulates both the glucose and lipid metabolisms [283, 284].

The mechanisms by which HCV perturbs the lipid metabolism can include increased lipogenesis, reduced β-oxidation and fatty acid export. Lipogenesis is increased by activation of sterol regulatory element-binding protein 1c [285] which activates genes responsible for fatty acid and cholesterol synthesis [286]. Oxidative stress caused by HCV induces increased peroxidation of lipids [287] and this seems to act as a regulatory feedback loop that inhibits HCV replication [288]. Additionally, β-oxidation, a catabolic process that breaks down lipids is decreased in infected cells [289]. Furthermore, to allow the virus to co-assemble with VLDL for viral secretion, the VLDL secretion is decreased [290]. Together, these processes are believed to contribute to the development of steatosis [291].

Steatosis is a condition commonly present in HCV-infected individuals. This pathology is characterized by the accumulation of lipid droplets in the liver [292]. The virus uses lipid droplets as platforms for virion assembly. The lipid droplets supply lipids for the VLDL production and release. The virus makes use of this assembly pathway to produce its LVP.

Cholesterol homeostasis is also affected in the HCV infection. There is increased biosynthesis of this sterol in infected cells but chronic HCV patients had lower blood levels of cholesterol compared to uninfected controls [293]. Interestingly, not only virion assembly [294] is affected by cholesterol biosynthesis inhibition but also RNA replication [295]. This may be explained by the role of cholesterol in the DMVs that surround the replication complexes that have much higher concentrations of this lipid than ER membranes.

Beyond the perturbations in the lipid homeostasis, HCV also modulates other metabolic pathways. In particular, the proteomic profiling of HCV-infected Huh-7.5 cells suggested that HCV induces perturbations in several central carbon metabolism pathways, including the glycolytic pathway [283]. The goal of this perturbation may be to support cell viability in spite of high viral replication.

In 2014, Ramière *et al.* [296] showed that Huh-7.5 infected with HCV had increased consumption of glucose, glutamate and pyruvate compared to uninfected cells. They related this increased consumption with an activation of hexokinase 2 (HK2) by NS5A. Hexokinases catalyze the first rate-limiting step of glycolysis, the phosphorylation of

glucose to glucose-6-phosophate. Hepatoma cells, as used in their study, mainly express the HK2 isoform of hexokinase. However, in hepatocytes the main hexokinase isoform present is glucokinase (GCK). Based on this we were interested in investigating the mechanism of activation of GCK by NS5A-D2 (see Chapter 8).

1.6 Intrinsically disordered proteins

Since 1894, when Emil Fischer proposed the "lock and key" analogy to explain the specificity observed in enzymes that catalyze the reduction of sugars [297], protein structure was believed to be necessary for function. His analogy suggested that the substrate of an enzyme acts as a key. This key should have the right shape to fit into the lock, the active site. If the fit was not perfect, no reaction would occur. Later, because certain experimental observations could not be explained by a rigid enzyme, the "lock and key" analogy was questioned in 1958 in the proposal by Koshland [298] called the "induced fit" model. In this model, the binding of the substrate causes a conformational change in the enzyme and a certain flexibility of the enzyme is necessary for it to achieve its catalytically active conformation. However, structure was still believed to be necessary for function.

The necessity of structure for function was called into question only much later during the 1990s. In 1996, Kriwacki and coworkers showed that the N-terminal region of $p21^{Waf1/Cip1/Sdi1}$, that acts as a cyclin-dependent kinase inhibitor, lacks stable secondary and tertiary structure while free in solution. When bound to one of its biological targets, however, the protein undergoes a disorder-order transition. Later, in 1998, Romero *et al.* [299] predicted that over 15000 proteins in the SwissProtein database displayed disordered regions at least 40 amino acids long. In 1999, Wright and Dyson [300] reviewed the growing body of evidence and showed that proteins without stable three-dimensional structure are ubiquitous in various roles and processes throughout the cell. These proteins may undergo folding only in the presence of a binding partner. Additionally, they suggested that disorder may provide functional advantages to a protein.

During the last two decades, our understanding of intrinsically disordered proteins

(IDPs), proteins that do not possess stable tertiary structure, and the biological processes in which they are implicated has greatly improved.

1.6.1 Characterization of IDPs

IDPs do not have a global stable three-dimensional structure over time due to their amino-acid composition. They tend to have a higher percentage of charged residues and prolines and a lower percentage of hydrophobic and aromatic residues, which tend to cause folding to form the cores of globular proteins [301].

Rather than existing as stable conformations, IDPs can sample a large ensemble of exchanging conformations. Thus, one of the main characteristics of IDPs is their flexibility. Unlike folded proteins, these molecules do not have single conformations that are well-defined energetic minima (Figure 1.19). Instead, these proteins are capable of sampling a multitude of conformations in solution that are minima separated by low energy barriers [302].

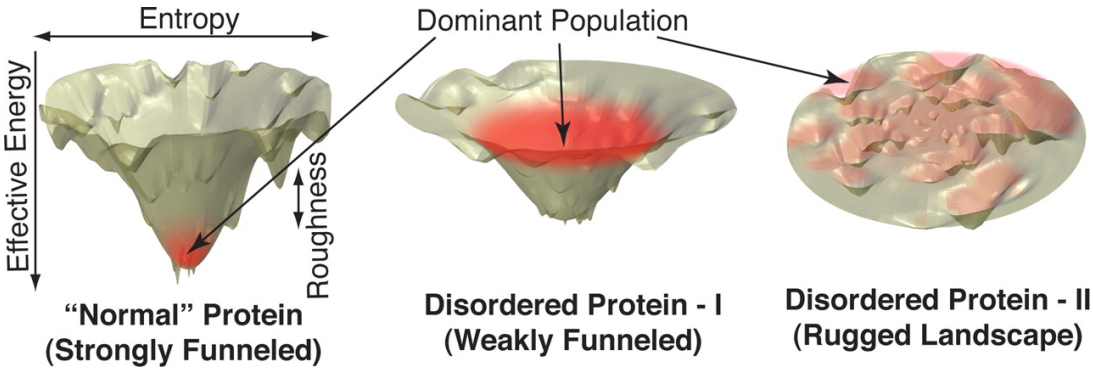

FIGURE 1.19: Energy landscapes of folded proteins and intrinsically disordered proteins. **Left** The energy landscape of a "normal" folded protein is well funneled so that the protein conformation is a well-defined minimum. **Center** Intrinsically disordered proteins can be funneled as well but are weakly funneled and thus present more flexibility being able to sample more conformations. **Right** IDPs can also have a "rugged landscape" where they can change conformations and sample many conformations with small energy barriers between them. From Papoian [302]. Copyright 2008 National Academy of Sciences.

A region or protein without any structural propensities is called "random coil". The orientation of each amino acid within a protein chain is described by its torsion angles ϕ and ψ. In a random coil, for each amino acid there will be a distribution of torsion angles [303] and all allowed conformations are being sampled. However, often certain conformations are populated more often than others. Many IDPs have propensities to form secondary or tertiary structural motifs and are not completely random coil,

which can be called structural propensity. For example, region of an IDP with helical tendencies will be in a helical conformation more often than would be expected if it was random coil. These motifs can be transient or longer lasting [304].

The characterization of the structure of an IDP consists in characterizing either an average structure or an ensemble of possible conformations the IDP can populate. This consists in identifying whether there are secondary structural motifs (transient or not), the residues involved in such structures or even long-range transient structures.

The main challenge in characterizing these mostly unstructured proteins is that X-ray crystallography, a tool commonly used in structural characterization of proteins, cannot be used for the study of the dynamic free state of IDPs, because the intrinsic flexibility of these proteins precludes their crystallization. In some cases where the affinity between a structured protein and its intrinsically disordered partner is high it is possible to crystallize a complex representative of the bound state of this IDP [305]. However, for most studies other experimental techniques are used for the characterization of an IDP. Nuclear magnetic resonance spectroscopy, one of the main experimental techniques used for the study of IDPs, will be discussed in Section 1.7.2.

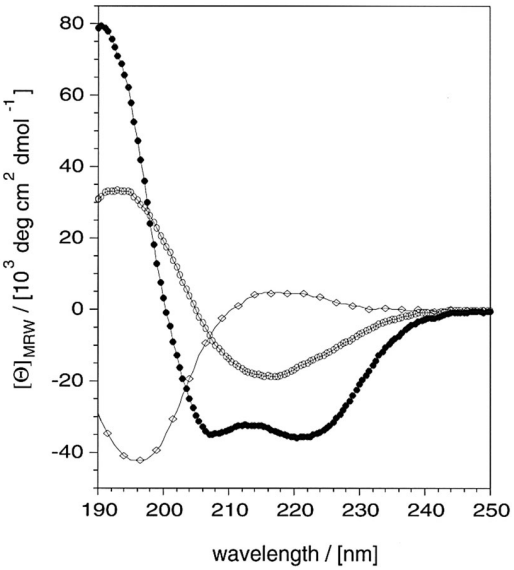

FIGURE 1.20: Far-UV CD spectra of typical of secondary structure and random coil. Poly-L-lysine forms different structures depending on the conditions. The far-UV CD spectrum of typical of an α-helical structure is represented with filled circles (0.1 mg/mL poly-L-lysine freshly dissolved at pH 11.1). A CD spectrum typical of a β-sheet is represented with open circles (0.1 mg/mL poly-L-lysine at pH 11.1 after heating at 52°C for 15 minutes) and the CD spectrum characteristic of random coil conformation (diamonds, 0.1 mg/mL poly-L-lysine in pure water). From Fändrich and Dobson [306]. Copyright 2002 European Molecular Biology Organization.

One powerful technique for the assessment of conformational properties of a protein is circular dichroism (CD) in the far ultraviolet region. This technique involves the differential absorption of left- or right-handed circularly polarized light. This can be used to determine folding properties and secondary structure content of a protein. The structural dependence occurs because when the amides of the backbone are aligned in arrays their optical transitions are shifted [307]. Different structural elements give rise to characteristic CD spectra (Figure 1.20). The CD spectrum of a protein does not provide information of the secondary structure of specific residues but can be used to determine whether a protein is globally intrinsically disordered [308]. Additionally, it can also be used to identify the presence of structural propensity within an otherwise disordered protein as has been done for NS5A-D2 [106]. The CD spectrum of this protein largely resembles a random coil conformation. Additionally, the presence of a negative shoulder at 220 nm that increases with the addition of 2,2,2-trifluoroethanol (which stabilizes helical folding) and temperature increase (that strengthens hydrophobic interactions) suggests that there is not only random coil structure but also structure or structural propensity.

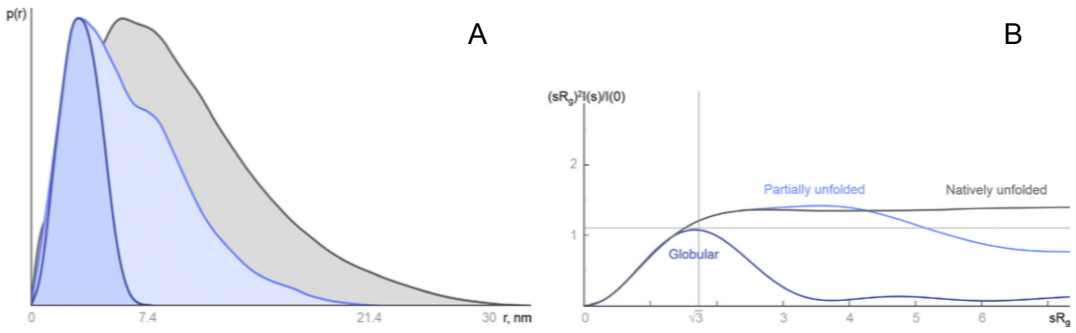

FIGURE 1.21: SAXS characteristics of IDPs. Data simulated from three 60 kDa proteins. A globular protein (dark blue), a protein 50% unfolded (light blue) and fully disordered (gray) are shown. **Left** Distance distribution p(r) showing the larger radii of disordered proteins. **Right** The Kratky plot displays a bell-shape for proteins with globular regions. From Kikhney and Svergun [309], available under a Creative Commons Attribution-Noncommercial-NoDerivatives license.

Another technique that can be used to characterize the global structure of IDPs is small-angle X-ray scattering (SAXS) that can be used to investigate the overall shape

of biomolecules. The scattering pattern of X-rays traversing a protein solution is measured at different angles. The scattered intensity (I) is given as a function of the momentum transfer $s = 4\pi \sin \frac{\theta}{\lambda}$, where 2θ is the scattering angle and λ is the wavelength of the incident light. The scattered pattern is isotropic as the measurement is done in solution. The scattered intensity can be related to the electron density of the particle (ρ). Globular monodisperse systems show electron density that reflects the shape and size of the protein. As natively unfolded proteins are more extended than globular ones of the same number of amino acids, these proteins have larger radii of gyration (Figure 1.21A). Proteins that have both folded and disordered regions present a distribution that is as a combination of both globular and natively unfolded characteristics. This is particularly visible using a dimensionless Kratky plot, given as $(sR_g)^2 I$ as a function of sR_g (Figure 1.21B) [309].

Other analytical techniques that can be used to characterize IDPs include fluorescence spectroscopy, infra-red spectroscopy, dynamic light scattering and size-exclusion chromatography, i.e., techniques that can be used to infer information on secondary structure and on the size/shape of the protein [310].

1.6.2 Biological roles of IDPs

There are many advantages to the increased flexibility of IDPs. These proteins can be shaped by their environment which allows them to recognize a variety of biological targets [300]. Furthermore, their extended structure provides them with a larger surface area which contributes to their binding of multiple partners and allows for easier access by enzymes responsible for post-translational modifications [311]. Binding to partners can occur with the IDP in an (at least locally) ordered configuration [312]. Additionally, due to their lack of structure, IDPs are more sensitive to degradation by proteases. At first glance this may seem as a disadvantage but it provides an additional mode for the cell to regulate their activity.

These advantages provide insight into the biological roles in which we find IDPs or intrinsically disordered protein regions (IDPRs). Based on their functions they can be classified into six categories [313, 314]: entropic chains, display sites, chaperones, effectors, assemblers and scavengers. Firstly, their flexibility allows regions of disorder to

act as entropic chains, in which their ability to populate a large conformational space is central to their activity. An example of this are linkers that connect structured domains of a protein. Other roles of IDPs and IDPRs are related to their recognition of binding partners. Display sites are proteins in which their disordered nature allows for easy access by proteins that introduce post-translational modifications (PTMs). These PTMs can then modulate the interactions between the disordered protein and its partner(s). Chaperones assist in the correct folding of proteins or RNA. Disordered regions make up a large part of both RNA and protein chaperones [315]. Effectors act by modifying the activity of their partner, either activating or, more often, inhibiting it. Assemblers bind to multiple targets to aid in the formation of multi-protein complexes by bringing together different partners. Finally, scavengers are a class of IDPs and IDPRs that store and neutralize small ligands, for example calcium-binding phosphoproteins can solubilize calcium phosphate [316].

Disorder is present in all kingdoms of life and is prevalent even among viruses [317]. For viruses, IDPs provide the advantage of being capable of interacting with a large number of partners, both from the host and the virus itself thus maximizing the functions with a minimal genome. Additionally, disordered regions provide flexibility to mutate to adapt to the host. This was shown to be the case for the C-terminus of the viral polymerase of the Nodamura virus [318]. Indeed, several viral genomes contain regions coding for IDPs. For example, the human immunodeficiency virus 1, hepatitis δ virus, influenza virus, human papillomavirus as well as many others have IDPs [319]. Also, in the hepatitis C virus the C-terminal of NS5A protein is intrinsically disordered, which contributes to the large number of interaction partners this protein is capable of binding, including NS5B and glucokinase.

1.6.3 Binding mechanisms of IDPs

IDPs are often folded when bound to their biological targets [312]. There are two mechanisms in which folding to bind to a partner occurs (Figure 1.22). Firstly, the active conformation already exists in solution and is selected by the binding partner in a mechanism called "conformational selection". Secondly, IDPs can undergo induced fit, in which binding occurs with a somewhat unfolded state that becomes structured upon

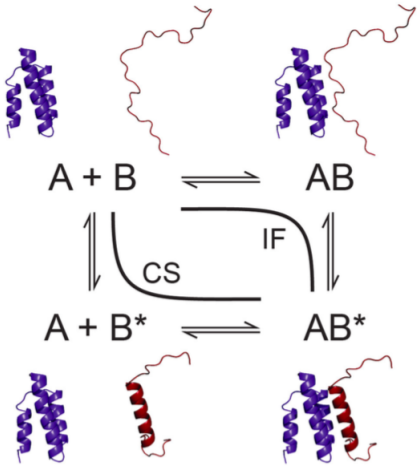

FIGURE 1.22: Binding mechanisms of IDPs. If the IDP is folded in its bound state there are two main binding mechanisms. In an induced fit mechanism (IF) the disordered partner (B) only folds into its final bound state (B*) upon binding. Additionally, binding can occur with a conformation that resembles the bound state that exists in solution in the case of conformational selection (CS). From Mollica *et al.*[320]

binding. Nonetheless, these two mechanisms are only two extremes of possible binding mechanisms. So far, for IDPs neither extreme case has been proven to exist exclusively. Often binding mechanisms are a combination of both cases, for example by binding a pre-existing structure but undergoing further rearrangement upon binding [320, 321].

Despite the fact that most interaction of IDPs occur with folding, some proteins remain disordered even upon binding. These form what are called "fuzzy complexes" [322]. This is the case of the disordered signaling protein Sic1. This protein is recognized by Cdc4 prior to degradation. This only occurs when six sites on the disordered protein are phosphorylated but its binding partner Cdc4 only has one recognition site. In 2008 it was shown that Sic1 is disordered even when bound to Cdc4 and that the phosphate groups take turns binding to the site on Cdc4 [323]. It is through increased electrostatic interactions in the presence of phosphorylation that the presence of six sites on one protein is detected by single site on its partner [324].

1.7 Protein nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy, like other spectroscopic techniques, uses the interaction of electromagnetic waves with matter. In NMR spectroscopy the wavelengths used are radiofrequency or MHz range frequency. Additionally it makes

use of an effect called the Zeeman effect in which the energy levels of of nuclear spin states split in the presence of an external magnetic field (B_0). This splitting is proportional to the applied magnetic field. For NMR experiments, the magnetic field strengths applied are homogeneous and very high (in this work we use spectrometers that have a field strength of 14.1 and 21.1 tesla). Despite the high magnetic fields, the splitting is small compared to the thermal energy at the temperature of the experiments and results in a very small excess (about 1 in 100 000) of the lower energy nuclear spin magnetization state resulting in the low sensitivity that is characteristic of this technique.

The resulting magnetization vector in a magnetic field is aligned with the external magnetic field (Z direction). By exciting the nuclei with a pulse of radiofrequency of the right wavelength the magnetization vector can go to the plane perpendicular to the external magnetic field (XY plane). There it precesses at a specific frequency, called the Larmor frequency (ω_0) which induces a current in the receiver coil of the spectrometer. This current gives the signal known as free induction decay (FID) that decays over time due to relaxation mechanisms. By Fourier transform this time-domain signal is transformed to give the frequency-domain signal known as the NMR spectrum.

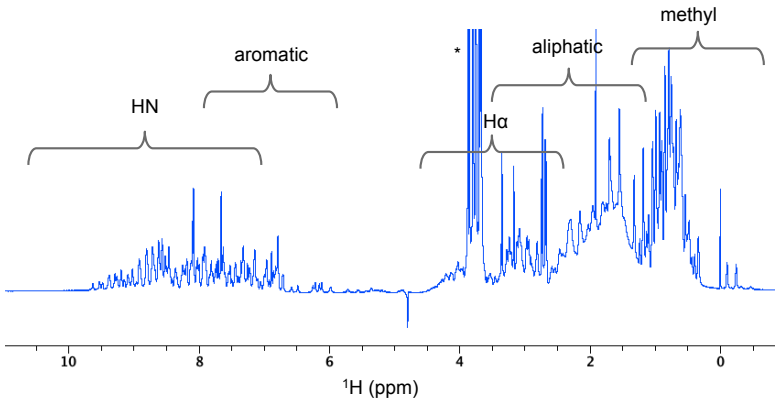

FIGURE 1.23: ^{1}H NMR spectrum of human ubiquitin. This spectrum of ^{13}C,^{15}N-labeled ubiquitin (8 kDa) shows the approximate regions of different protons from the protein. Water was suppressed with a WATERGATE pulse sequence and the residual water signal is visible as the negative peak at 4.8 ppm. Even with such a small protein the 1D spectrum is complex. The peaks labeled with an asterisk are due to the buffer.

The most simple type of NMR experiment is a 1D experiment where there is a single 90° pulse then the FID is acquired. The Fourier-transformed 1D ^{1}H spectrum of ubiquitin obtained with water suppression [325] (to remove the large signal due to the proton of water) is shown Figure 1.23. Simultaneous excitation of all ^{1}H gives rise to a

spectrum with a frequency dimension, where frequency is specified in parts per million relative to a defined standard. The position of the peaks correspond to their chemical shifts.

Some of the most interesting experiments that can be done by NMR spectroscopy are done by manipulating the nuclear spins between the excitation pulse and detection of the signal. By using timed pulses of varying lengths and intensities as well as specific delay times we can manipulate the state of the ensemble of nuclear spins [326]. The nuclei commonly used in protein NMR, ^{1}H, ^{15}N, ^{13}C are spin $\frac{1}{2}$ nuclei that have spherical charge distributions.

In a quantum mechanical description of NMR spectroscopy, the energy operator, the Hamiltonian, of a spin i has the following terms that affect an NMR experiment [326]:

$$\hat{H}_i = \hat{H}_i^{static} + \hat{H}_i^{CS} + \hat{H}_{ij}^{J} + \hat{H}_{ik}^{DD} + \hat{H}_i^{RF} + \hat{H}_i^{other} \tag{1.1}$$

These are terms for the contribution of the static field ($static$), also called Zeeman Hamiltonian, of the additional magnetic field due to the electrons surrounding the nucleus (CS), of scalar coupling (also called J-coupling) with nucleus i (J) summed for all j, of dipole-dipole interaction with nucleus i (DD) summed for all k, of radiofrequency fields like the one caused by the pulse (RF) and any other fields present ($other$).

Chemical shift (CS) arises from the movement of electrons within the sample, induced by the external magnetic field. Since this movement is an electrical current, it creates a smaller local magnetic field which in turn changes the value of the magnetic field effectively detected by a particular nucleus. It can be expressed through its Hamiltonian

$$\hat{H}_i^{static} + \hat{H}_i^{CS} = -\gamma_i B_0 (1 + \delta_i) I_z \tag{1.2}$$

where γ, called the gyromagnetic ratio, is a characteristic of the nucleus that is proportional to the strength of the NMR signal; I_z is the angular momentum operator and δ is the value of the chemical shift. CS contains information on the chemical environment of a nucleus which can be extended to information about the structure of

a molecule. Chemical shift is given in parts per million (ppm) which is a unit that does not depend on the magnitude of the external magnetic field and is relative to the frequency of a reference compound. For ^{1}H and ^{13}C, tetramethylsilane (TMS) or its water-soluble derivative Trimethylsilylpropanoic acid (TMSP) can be used as internal reference compounds at 0 ppm.

Scalar coupling, the isotropic component of J-coupling (given by $\hat{H}_{ij}^{J}$ in Equation 1.1), is due to an indirect interaction between two nuclei that is mediated by electrons in the bonds that connect them. Scalar coupling causes the splitting of signals in NMR. For J-coupling to heteronuclei this can be removed by applying decoupling during the acquisition of the FID. Scalar coupling can also be used to transfer magnetization between nuclei that are connected through bonds, a strategy often used to obtain multidimensional spectra.

The dipole-dipole contribution arises from the effect of the magnetic field of one nucleus on another. Since the effect of the magnetic field is dependent on the relative positions and orientations of the nuclei, this effect is through space and can provide insight into the proximity of two nuclei. However, for most applications in solution NMR, the dipolar contribution to the energy levels of a nuclear spin is averaged because all orientations can be sampled in an isotropic solution. It does, however, occur as a perturbation that causes relaxation.

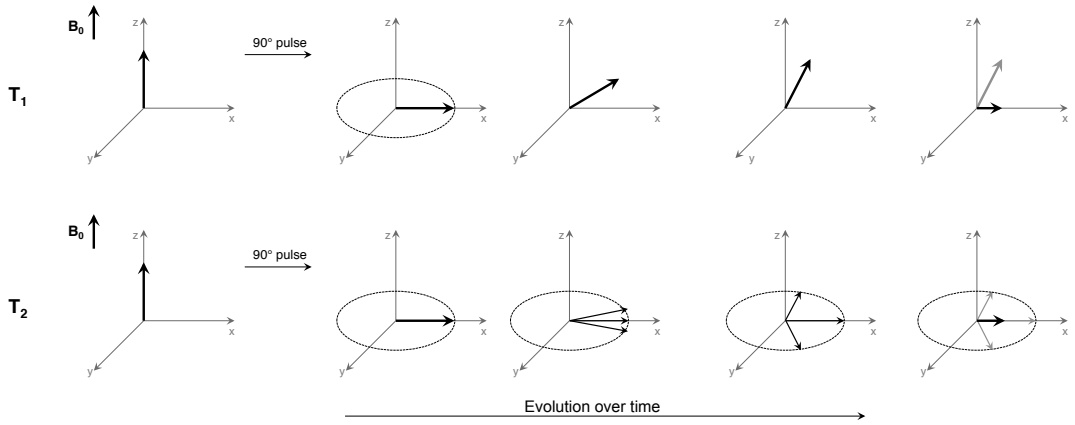

FIGURE 1.24: Spin relaxation mechanisms. The magnetization vector is shown as a black arrow in the rotating-frame (to simplify the precession in the XY plane at the Larmor frequency). **Top** Spin-lattice relaxation (T_1) causes the return of the magnetization from the XY plane to the Z-axis, reducing the projection of the magnetization vector on the XY plane over time. **Bottom** Spin-spin relaxation causes a change in the frequency of precession of the nuclear spins which reduce net magnetization.

Relaxation causes the FID signal that is acquired in an NMR experiment to decay over time. This is due to different mechanisms that result in the loss of coherence on the XY plane (Figure 1.24). The first type of relaxation is longitudinal relaxation. This relaxation is the return of magnetization from the XY plane to the Z-axis. It has a time constant called T_1 and relaxation rate (the inverse of the time constant) called R_1. The second type of relaxation is spin-spin relaxation, also known as transverse relaxation, and is the decay of the magnetization on the XY plane. It is caused by loss of coherence, that is, the nuclear spins to precess in the XY plane at progressively different frequencies and consequently net magnetization is lost. Both types of relaxation originate from random fluctuations of the local magnetic field. Spin-spin relaxation has a time constant called T_2 and a relaxation rate called R_2. Relaxation causes exponential decay of the FID signal. The Fourier transform of a signal that decays slowly is a narrow peak. Inversely, fast decay results in broad peaks. Therefore, quickly relaxing nuclei give rise to broad resonances in the NMR spectrum and the width of the peak can be used to estimate the transverse relaxation rate.

Two sources of nuclear spin relaxation are the anisotropic part of chemical shift interaction and dipole-dipole coupling. These mechanisms turn the the molecular dynamics into fluctuating magnetic fields that result in NMR relaxation. Thus, these are both directly related to dynamic properties of the protein, such as the rotational correlation time (τ_r) and internal motions. T_1 is shortest when the rotational correlation time is $1/\omega_0$, which is of the order of 1 ns for a proton at 900 MHz. Since T_2 is equal or shorter than T_1, T_2 is the main contributor of the two to the broadening of peaks.

Nuclear magnetic resonance spectroscopy allows for the study of proteins in solution with atomic resolution. This can provide information that cannot be obtained by X-ray crystallography where the proteins are locked in a conformation that can be crystallized which might not be representative of the conformation(s) in solution. In general, a spectrum obtained is a weighted average of conformations that exist in solution. Other advantages of protein NMR spectroscopy are that it can be used to study interactions between protein and other biomolecules to obtain both thermodynamic parameters and the regions of the protein(s) that interact and investigate the dynamics of proteins in solution with atomic resolution.

However, as can be seen in Figure 1.23 the abundance of peaks makes the analysis of a 1D ^{1}H spectrum unpractical for even small proteins. Additional dimensions, particularly those of heteronuclei are used to disperse the peaks and gain further information. One of the main types of multidimensional spectra used in protein NMR is the ^{1}H,^{15}N-heteronuclear single quantum coherence (HSQC) spectrum.

1.7.1 ^{1}H,^{15}N-Heteronuclear single quantum coherence spectrum

With specific combinations of pulses and delays we can manipulate the spin polarization to transfer it from one nucleus to the other, through bonds or through space. By exciting ^{1}H nuclei and transferring their spin polarization to ^{15}N through one bond we can do 2D NMR and observe the amide bonds of the protein backbone, as well as other N-H bonds in the protein side chains, through the chemical shift of the ^{1}H and the ^{15}N bound to it.

Most protein NMR makes use of labeling the nitrogens. ^{14}N, the most abundant nitrogen isotope is NMR active but has spin 1 and is therefore quadrupolar, that is, does not have a spherical nuclear charge distribution which causes very broad signals. In its place, we make use of ^{15}N, that has a natural abundance of 0.37% and spin $\frac{1}{2}$. Due to its low natural abundance, isotopic enrichment is necessary for its use. This is achieved using an expression system and introducing ^{15}N as the main source of nitrogen. For example, by growing *Escherichia coli* cells that can overexpress the protein of interest with $[^{15}N]NH_4Cl$ as the sole source of nitrogen, the protein will be uniformly labeled ^{15}N.

Once the protein is uniformly ^{15}N-labeled the spectra acquired are typically a form of heteronuclear 2D spectrum called ^{1}H,^{15}N-Heteronuclear Single Quantum Coherence (HSQC) spectrum. Like other heteronuclear pulse sequences, the HSQC (Figure 1.25) transfers magnetization from one nuclei (here ^{1}H) to the other (^{15}N) and then has an evolution time (t_1) of the magnetization on the second nucleus, then transfers magnetization back to the original nucleus to detect, while decoupling on the second nucleus. This sequence is repeated several times with variable evolution times. Fourier transform is applied twice to give rise to spectra that have two frequency dimensions. In the spectra there are cross-peaks that correspond to the frequencies of ^{1}H (on what is called

direct dimension of the 2D spectrum because the detection is done on this nucleus) that are one bond from ^{15}N atoms, whose frequencies are also visible in the indirect dimension.

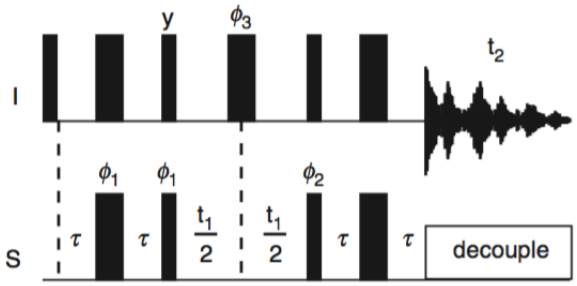

FIGURE 1.25: Basic HSQC pulse sequence. Initially the I nucleus is excited by a 90° pulse. Then its magnetization is transfered to the S nucleus and allowed to evolve for a period of t_1 followed by the return of the magnetization to the I nucleus to be detected while a decoupling pulse sequence on the S nucleus removes J-coupling from the spectrum. This sequence is repeated for variable values of t_1 giving rise to a 2D spectrum. For a ^{1}H,^{15}N-HSQC the hydrogen is the I nucleus and the nitrogen is the S nucleus due to the improved sensitivity of detecting on hydrogen. From Cavanagh *et al.* [327].

In the ^{1}H,^{15}N-HSQC, the region of interest for protein NMR is about from 6.5 to 11 ppm in the ^{1}H dimension where the amide protons can be located. Each amino acid has an amide bond on the protein backbone and except for proline, that is a secondary amine, each amino acid has an amide proton. Therefore, it is expected that this region of the spectrum has one cross-peak for each amino acid except proline. Therefore, this ^{1}H,^{15}N-HSQC spectrum can be imagined as a fingerprint of the protein (Figure 1.7.2).

The ^{1}H-^{15}N cross-peaks are sensitive to their chemical environment and the structure of the protein. In a well-structured protein the environment observed by the amide protons is varied. On the other hand, in an intrinsically disordered protein the amide protons mostly see the solvent. Therefore, for a well-structured protein the proton dimension of the spectrum is well dispersed compared to the spectrum of a denatured or intrinsically disordered protein.

To make full use of a ^{1}H,^{15}N-HSQC spectrum we must assign its resonances to residues in the amino acid sequence. This is typically done by what is called triple resonance backbone assignment. To do this, the protein used is isotopically labeled ^{15}N and ^{13}C because in these NMR experiments the magnetization is transferred from the amide hydrogen to a carbon in either in the same amino acid (i) and/or the amino acid before it in the sequence (i-1) through the nitrogen atom. There are many types

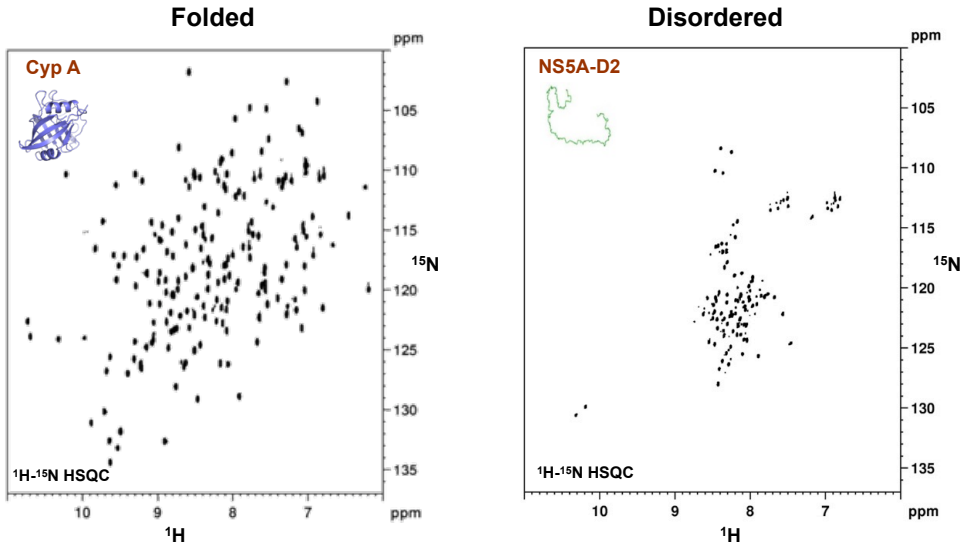

FIGURE 1.26: Comparison of the ^{15}N,^{1}H-HSQC spectra of a folded protein and an IDP. **Left** The ^{1}H,^{15}N-HSQC spectrum of human cyclophilin A (CypA) which is a 18 kDa folded protein. **Right** The ^{1}H,^{15}N-HSQC spectrum of the domain 2 of NS5A (NS5A-D2) a 11.6 kDa IDP shows a much less disperse spectrum characteristic of a disordered protein.

of complementary triple resonance experiments. HNCA and HN(CO)CA have peaks due to to both i and i-1 C_α and only the i-1, respectively. Similarly, HNCACB and HN(CO)CACB have peaks due to C_α and C_β due to both i and i-1 and only the i-1 residues, respectively. HN(CA)CO and HNCO contain signals due to carbonyl carbons (C'). These pairs of spectra allow the discrimination of the carbon chemical shift of the i and i-1 residues.

Beyond assigning the resonances manually, it is also possible using a software to find sequential assignments from a table of chemical shifts. One such program is MARS [328]. It searches for sequential connectivity between spin systems and also compares the chemical shift values to those expected of the amino acid sequence and secondary structure. Furthermore, it does so without penalty for missing chemical shifts.

The assignment of the ^{1}H,^{15}N-HSQC spectrum can be used as a starting point to investigate structure, interactions and dynamics with per residue resolution.

1.7.2 NMR spectroscopy of intrinsically disordered proteins

The lack of a fixed three-dimensional structure of intrinsically disordered proteins makes NMR an important technique for their investigation.

Since the chemical shift is averaged for the conformations that the protein samples (as long as it samples these conformations fast enough) and the spins are often exposed to the solvent. Thus, the main characteristic of the NMR spectra of intrinsically disordered proteins is that the peaks are not very dispersed, especially in the ^{1}H dimension (Figure 1.26). In IDPs the values of the chemical shift of a resonance will be close to the "random coil" value [329], that is, the value of the chemical shift if rotational conformations of the carbon bonds interconvert freely to all the accessible conformations. Therefore, CSs of these nuclei depend mainly on the nature of the residue observed and contributions from the closest neighbors.

Deviations from random coil values occur when there is propensity to form structure. Several methods exist to determine the presence of structure or structural propensity in a residue-specific way.

Using the chemical shift information from $^{1}H_{\alpha}$ [330] and then further extended to use values from the backbone ^{13}C [331], Wishart and colleagues developed a protocol that determines an index called the chemical shift index (CSI) that identifies secondary structure. $^{13}C_{\alpha}$ and $^{13}C'$ present a downfield shift in CS values when in a helical structure and an upfield shift when in a β-strand. Inversely, $^{13}C_{\beta}$ present an upfield shift in CS values when in a helical structure and a downfield shift when in a β-strand. Based on this the residue is given a score of -1, 0 or 1 which when grouped together represent β-strands, coil and α-helices, respectively.

CSI was developed for use with folded proteins and as such it predicts fully formed structure. For IDPs, however, it is interesting to get a measure of structural tendencies. For this, the secondary structure propensities (SSP) score [332] makes use of the deviation of the chemical shift values of multiple nuclei from the random coil values and also compares them to fully α- and β- structures. This gives a weighted score that indicates a propensity to form these secondary structures. Despite being able to include $^{13}C_{\alpha}$, $^{13}C_{\beta}$, $^{13}C'$, $^{1}H_{\alpha}$, $^{1}H^{N}$ and ^{15}N in the calculation, the original authors suggest the use of $^{13}C_{\alpha}$, $^{13}C_{\beta}$ and $^{1}H_{\alpha}$ as these will be the most sensitive to the presence of α-helices and extended structure propensities. An SSP score, calculated for each residue in the protein, gives a value between -1 and 1 with negative values suggesting extended structural propensity, positive values for α-helical propensity and 0 for random coil. Values

of -1 or 1 represent fully formed structure while, for example, a value of 0.3 for a certain residue indicates that roughly 30% of conformers are sampling a helical structure at that residue at a given time.

Originally, the random coil CS values came from experimental measurement of small peptides like GGXAGG [329] where X is the amino acid whose random coil CS is being measured. Instead, Tamiola *et al.* [333] proposed the use of a "neighbor-corrected" value for the expected chemical shift in the absence of structure. In this way, a reference random coil CS value of each amino acid is modulated by the amino acid immediately before and the amino acid after giving more accurate expected chemical shift values for random coil residues which can be used to calculate secondary chemical shifts (the difference between experimental CS and random coil CS) as well as in combination with the previous chemical shift analysis techniques to better assess the presence of secondary structure.

Another way to determine secondary structure in IDPs by NMR spectroscopy is measuring the three-bond J-coupling of between $^{1}H^{N}$ and $^{1}H_{\alpha}$ that can be related to the ϕ dihedral torsion angle through the Karplus equation [334] (Equation 1.3).

$$^{3}J(\phi) = Ccos2\phi + Bcos\phi + A \tag{1.3}$$

where A, B and C are empirically derived parameters [335]. These angles can then be used to determine the structure of the protein backbone. However, the use of this in disordered proteins is limited due to their lack of dispersion [326].

Furthermore, NMR experiments can be used to determine distance restraints in IDPs. Nuclear Overhauser effect spectroscopy (see Section 1.7.6) can give short range distance restraints that may indicate the presence of stable local structure within an IDP and can be used to calculate an ensemble of conformers that conform to these distance restraints to identify the structure as was used to determine the structure of a small structural motif in the domain 2 of NS5A [110]. Somewhat similarly, paramagnetic relaxation enhancement (see Section 1.7.7) provides long-range distance restraints that can be used to identify long-range contacts within the IDP [336].

IDPs are highly dynamic and therefore have short correlation times, giving rise to

slow relaxation and narrow peaks even for very large IDPs. However, this is not the case for large folded proteins.

1.7.3 High molecular weight protein NMR spectroscopy

The study large proteins or complexes is challenging using protein NMR spectroscopy. For globular proteins using "classical" protein NMR with ^{15}N-labeled samples and the ^{1}H,^{15}N-HSQC there is an upper limit of about 30 kDa. There are two main factors that contribute to the difficulty in obtaining spectra with increasing size. Firstly, the amount of signals increases with the number of amino acids in the protein, which can cause signal overlap. This can be remediated by using different strategies. Higher field strength can be used to improve resolution in ppm. Additional dimensions may be needed if two amide N and H have the same chemical shifts. For example, in an HNCO spectrum, magnetization is transferred to the carbonyl of the previous amino-acid and the spectrum gains a third dimension that represents the chemical shift of the carbonyl carbon and it is less likely to have two cross-peaks with the same chemical shift for all three nuclei. Another strategy, which is protein-dependent, is segmental labeling in which one or more domains of the protein are labeled with NMR-active nuclei while the remainder is NMR-inactive. They are then fused together to obtain the entire protein which will give a spectrum that contains only a fraction of the resonances.

Another major difficulty in obtaining and assigning spectra with globular proteins larger than 30 kDa is that as the protein increases in size it tumbles more slowly. This increases its correlation time and shortens its transverse relaxation time (T_2) (Figure 1.27). One of the main contributions to the fast relaxation are dipolar interactions between protons. The effect this has on the spectrum are broader peaks which are less intense and overlap even more. For proteins larger than 30 kDa, peaks can be so broad that they can no longer be detected.

Some strategies have been developed to reduce the broadening of peaks with large proteins.

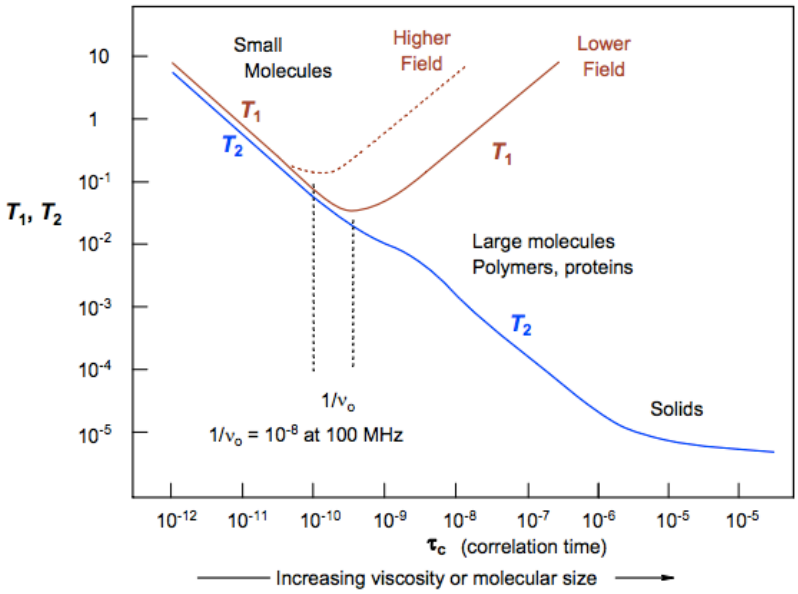

FIGURE 1.27: Effect of rotational correlation time on relaxation times. With long correlation times, the relaxation effects due to orientation-dependent relaxation processes (dipolar interaction and chemical shift anisotropy, for example) result in faster relaxation and shorter T_2. T_1 is most affected by relaxation when the timescale of these processes is of the same order as the (field-dependent) Larmor frequency. For large proteins, slow tumbling (represented by a large τ_c) gives rise to large R_2. From Reich [337] (available at https://www.chem.wisc.edu/areas/reich/nmr/index.htm).

Deuteration

Deuteration is the substitution of ^{1}H by ^{2}H. The resonance frequency of ^{2}H is far from that of ^{1}H. The early applications of deuteration in protein NMR took advantage of this to simplify 1D spectra. In their 1968 article Crespi *et al.* [338] extracted C-phycocyanin from blue-green algae that had been grown in D_2O supplemented with protonated leucine. This resulted in simplified ^{1}H NMR spectra of the leucine side-chain protons. Beyond just the simplification of the spectra, the deuterated background also improved the linewidth of the peaks. The application of deuteration to high molecular weight proteins takes advantage of the dependence of the dipolar coupling on the gyromagnetic ratio (γ). The relaxation due to dipolar coupling (R^{DD}) on nucleus I is related to the gyromagnetic ratios of nucleus I and its neighbor S, that are separated by a distance r by Equation 1.4.

$$R^{DD} \propto \frac{\gamma_I^2 \gamma_S^2}{r^6} \tag{1.4}$$

Since the ratio between the gyromagnetic ratios of ^{1}H and ^{2}H is about 6.5 we can see a theoretical reduction of the relaxation by 42 times by deuteration [339]. There is also a dependence on the distance between the nuclei for the effectiveness of relaxation by dipole-dipole interactions, therefore protons that are isolated from other protons will relax more slowly. Improved relaxation properties result in spectra with smaller linewidths, which is essential for experiments with high molecular weight proteins where the slow tumbling rates result in broadening.

Perdeuteration, that is complete deuteration of all hydrogens, can be done in proteins produced recombinantly by producing them in a medium where the main sources of hydrogens are deuterated [340], notably D_2O and glucose. It can be done with a minimal medium but there are also deuterated rich media available commercially. It is typically done in procaryotic expression systems but protocols exist for perdeuteration in yeast [341] and by cell-free protein synthesis [342] for proteins that cannot be expressed in bacteria.

After the protein was produced in deuterated medium some hydrogens need to be exchanged back to ^{1}H to use as the active nucleus in NMR. This exchange is done by incubating the protein in a protonated aqueous buffer, typically the buffer used for the NMR experiments. Exchangeable hydrogens like those in amide groups are exchanged with the solvent, the degree of exchange depends on the structure of the protein and incubation time. The protonation of the exchangeable hydrogens allows for the acquisition of a ^{1}H,^{15}N-HSQC spectrum with improved relaxation properties.

Transverse relaxation optimized spectroscopy

Another strategy that improves the signal-to-noise ratio in large proteins or complexes is the use of transverse relaxation optimized spectroscopy (TROSY) pulse sequences.

In the HSQC pulse sequence in Figure 1.25 there is decoupling to obtain a single peak. This avoids the splitting caused by J-coupling to nucleus S. By removing the decoupling during t_1 and t_2, splitting produces four signals, called a multiplet, instead of one, which reduces the intensity of the signal and may cause more overlap.

The main contributors to transverse relaxation in large proteins are the chemical shift anisotropy and dipole-dipole interaction. These mechanisms contribute to the

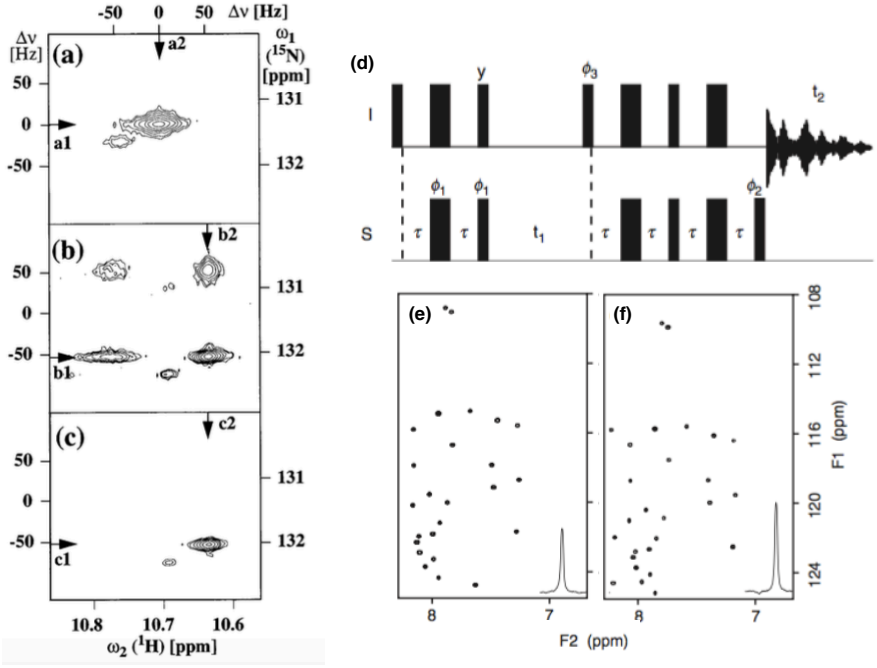

FIGURE 1.28: TROSY-HSQC spectra have smaller linewidths than equivalent HSQC spectra. **(a)** Example of a ^{15}N-^{1}H spin system obtained with an HSQC pulse sequence with decoupling. **(b)** The same spectrum without decoupling displays four peaks forming a multiplet with a 100 Hz J-coupling constant. Each peak has different relaxation contributions from the dipole-dipole interactions and chemical shift anisotropy. The bottom-right peak has a smaller linewidth due to the destructive interference between these effects resulting in slower relaxation. **(c)** Phase-cycling or application of pulsed field gradients is used to select only the coherence transfer pathway with the improved relaxation properties. From Pervushin *et al.* [343]. **(d)** Example of TROSY-HSQC pulse sequence. **(e)** HSQC spectrum of ubiquitin acquired with decoupling during both t_1 and t_2. **(f)** TROSY-HSQC spectrum showing the smaller linewidth when using this type of pulse sequence. From Cavanagh *et al.* [327].

relaxation of the different components of the multiplet with different signs. In one of the peaks of the multiplet, these contributions interfere destructively with each other, which results in reduced relaxation for this signal. A TROSY experiment consists in not decoupling but instead using phase-cycling to remove the three peaks that have less desirable relaxation properties (Figure 1.28). Phase-cycling is the changing of phases in the pulses in the pulse sequence so that undesired signals have opposing signals during successive scans and cancel themselves out when adding all the scans. The result in the spectrum is that only the peak with improved relaxation properties is present from the multiplet. It has a smaller linewidth compared to the decoupled signal in the HSQC, resulting in improved resolution and signal to noise ratio.

Since chemical shift anisotropy is field-dependent and dipole-dipole coupling is field independent the choice of the field strength is important to use this effect. It has

been proposed that about 1 GHz proton resonance frequency (or 23.5 T) results in the optimum reduction of the relaxation [343].

The TROSY strategy can be added to many types of pulse sequences. TROSY-based pulse sequences exist not only for the HSQC but also for triple-resonance pulse sequences [344], NOESY pulse sequences [345], etc.

Selective labeling

The studies of the largest proteins and complexes by NMR were done using selective labeling. Selective labeling is a strategy in which only certain sites on the protein are labeled with NMR active nuclei. This simplifies the spectrum by reducing the amount of resonances that are visible. This is often done in a way where one or a few amino acids are the only residues to be labeled. For example, the introduction of a single ^{15}N-labeled amino acid in an otherwise ^{14}N culture medium results in the labeling of only these residues in the HSQC spectrum, as long as there is no isotope scrambling during the expression of the protein. The evident advantage of this strategy is the simplification of the spectrum which can aid the assignment of the HSQC by identifying the nature of the amino acid that gives rise to certain resonances [346].

While the selective labeling of backbone amides has great advantages, the true power of selective labeling is harnessed by using selective protonation and ^{13}C-labeling on an otherwise perdeuterated and ^{12}C-labeled protein. The chemical group usually labeled in high molecular weight proteins are the methyl groups. In this strategy methyl groups are selectively labeled ^{13}C and protonated within a protein that is otherwise at natural abundance for carbon and perdeuterated. Methyls are often found in hydrophobic cores and interaction surfaces and are less sensitive to pH variations than amide protons. They contain three hydrogens that are equivalent, because of the rapid rotation of the methyl group. Thus, these three hydrogens give a single resonance in the NMR spectrum resulting in an improved signal-to-noise ratio. Additionally, the rapid rotation of the methyl group results in improved relaxation properties that result in smaller linewidths [347]. This increases the size of systems that can be studied by solution protein NMR spectroscopy beyond 100 kDa [348].

Protocols exist for the labeling of side-chain methyl groups of leucine, valine, isoleucine, alanine, methionine and threonine. For alanine [349], methionine [350] and threonine [351] the protocols consist in adding the desired amino acid, that is deuterated and ^{12}C except for the $^{13}C^{1}H_3$-labeled methyl group, just prior to the induction of the cells grown in a deuterated medium. Additional amino acids that share a biosynthetic pathway may be added to avoid isotopic scrambling, i.e. the labeling of unwanted amino acids.

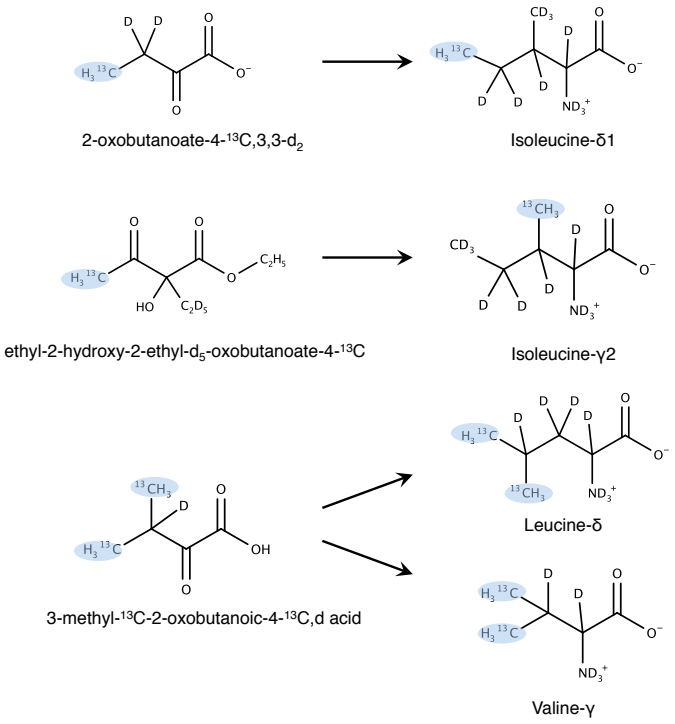

FIGURE 1.29: Precursors used for selective protonation of methyl groups of isoleucine, leucine and valine. The precursors on the left are converted by *E. coli* metabolism to methyl groups selectively ^{13}C-labeled and protonated in the amino acids on the right. The labeled methyl group in the precursors and the amino acids are highlighted in blue. Adapted from Kerfah *et al.* [352].

To label the side chain methyl groups of leucine, valine and isoleucine precursors can be used. 3-methyl-2-oxobutanoic acid or α-isoketovaleric acid is transformed into leucine and valine [353] and 2-oxobutanoic acid or α-ketobutyric acid [354] and 2-hydroxy-2-ethyl-3-oxobutanoate [355] is transformed into isoleucine by the *E. coli*

metabolism. In a deuterated culture, their ^{13}C,^{1}H-methyl labeled derivatives are converted to leucine and valine that are selectively protonated and ^{13}C on their methyl groups and similarly labeled isoleucine $\delta 1$ or $\gamma 2$ [356] methyl groups (Figure 1.29). Protocols also exist for stereospecific labeling of the leucine and valine methyl groups.

Another advantage of the labeling of methyl groups was the discovery that methyl groups are intrinsically TROSY-optimized when using a ^{1}H,^{13}C-heteronuclear multiple quantum coherence (HMQC) pulse sequence [357, 358]. There is a five-fold improvement of the sensitivity when a HMQC pulse sequence is compared to a HSQC-type experiment [359]. This is because for isolated methyl groups in slow tumbling molecules, in the HMQC there is complete cancellation of ^{1}H-^{1}H and ^{1}H-^{13}C dipolar interaction for 50 % of the signal. Unlike the TROSY-HSQC, this mechanism does not involve the cancellation between chemical shift anisotropy (field-dependent) and dipolar interactions (field-independent) and is, therefore, independent of the external magnetic field strength. However, in the presence of external ^{1}H spins dipolar interactions with these spins cause the slowly relaxing coherences to interconvert with the fastly relaxing coherences, losing the benefits of the TROSY effect. Consequently, the ^{1}H-^{13}C-TROSY HMQC spectrum (methyl-HMQC spectrum) should be acquired on a sample with a high level of deuteration, avoiding the dipolar interactions with external ^{1}H [360].

Several large proteins and protein complexes have been investigated using selective labeling of the methyl groups [361]. One of the largest systems for which there is extensive assignment of methyl groups is the 670 kDa 20S proteasome from the archaeabacterium *Thermoplasma acidophilum* CP [362]. In their work, Sprangers and Kay investigated the interaction between the 20S proteasome and the 11S activator complex from *Trypanosoma brucei* to form a 1.1 MDa complex. They took advantage of thermal stability of the proteins and complexes as well as a highly symmetrical nature of the protein complex. Indeed, many of the largest systems to be studied using methyl labeling take advantage of symmetry to simplify the spectra or at least complexes formed of smaller strucutres that are investigated separately (in what is called a divide-and-conquer strategy) [363–365].

1.7.4 The study of protein dynamics by NMR spectroscopy

Proteins are inherently flexible meaning they can sample many conformations in solution. The motions that give rise to these different conformations can be of different timescales, for example methyls rotate at a picosecond timescale [366] while domain motion is often in the microsecond to millisecond timescale. Different NMR techniques can probe conformational exchange with rates that can vary from the picosecond to days ranges (Figure 1.30).

The NMR signal is affected by chemical exchange and its rate between states in solution. We can imagine a theoretical protein that has two conformations in exchange in solution. This chemical exchange causes a change in resonance frequency between two states (A and B) of a spin. In such a two-state system, A is in exchange with B as in Equation 1.5.

$$A \underset{k_b}{\overset{k_a}{\rightleftharpoons}} B \tag{1.5}$$

where k_a and k_b are the forward and backward kinetic rate constants, whose sum is called k_{ex}.

As long as states A and B have different NMR resonance frequencies then the comparison of the exchange rate (k_{ex}) and the difference in the resonance frequencies of the states ($\Delta\omega$) will determine what will be observed in the spectrum (Figure 1.30). At low exchange frequencies (when $k_{ex} << \Delta\omega$, called slow exchange) we will observe two peaks, one for each state with relative peak volumes that give the relative population in each state. At high exchange frequencies (when $k_{ex} >> \Delta\omega$, called fast exchange) these peaks are merged into a single peak. At intermediate frequencies (when $k_{ex} \approx \Delta\omega$, intermediate exchange) these peaks broaden, even beyond detection, when the rate of the chemical exchange is of the same order as the difference in resonance frequencies of the two states. The peaks merge at a condition called coalescence that is shown in Equation 1.6 [368] where p_A and p_B are the populations of the A and B states.

$$2(p_A p_B)^{1/2} k_{ex} \approx \Delta\omega \tag{1.6}$$

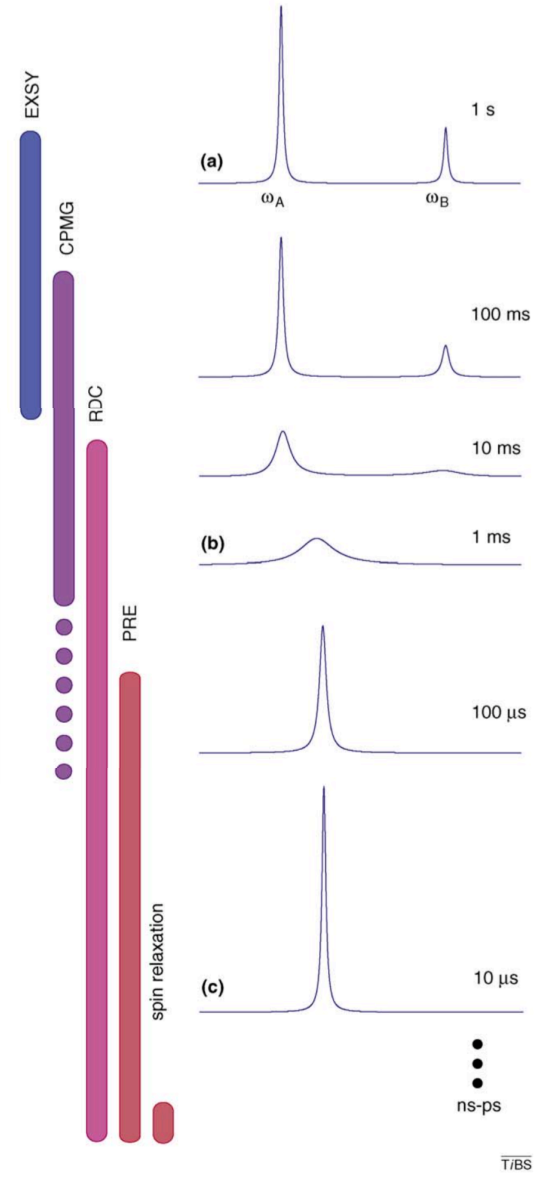

FIGURE 1.30: Two-state chemical exchange and NMR dynamics experiments. The figure shows simulated NMR spectra for the exchange between two states that resonate at the chemical shifts ω_A and ω_B with $\Delta\omega = \omega_A - \omega_B = 100$ Hz. **(a)** Slow exchange occurs when the exchange rate is much slower than $\Delta\omega$ resulting in separate peaks for each state. **(b)** When $k_{ex} \approx \Delta\omega$ the peaks coalesce and broadening is observed, sometimes even causing broadening beyond detection. **(c)** Fast exchange occurs when the exchange rate is much faster than $\Delta\omega$, in this case the resonances merge to give a peak at an average frequency. The experiments that probe different timescales are shown on the left. Reprinted from Trends in Biochemical Sciences, 34, Mittermaier and Kay, Observing biological dynamics at atomic resolution using NMR, 601–611, Copyright (2009), with permission from Elsevier [367].

Depending on the range of the dynamics observed in the protein different NMR strategies can be applied.

When in the two states are in slow exchange, where two peaks are visible, exchange spectroscopy (EXSY) can be used study this exchange. In this experiment, a delay

(called mixing time) is introduced in which the protein can exchange between states. This provides cross-peaks that connect the resonances of both states. By varying the mixing time and measuring the intensity of the peaks the rate constants can be determined. This technique can be applied to systems with rate constants from 0.5 s^{-1} to 50 s^{-1} [367]. When the population of the minor state is too small to be observed chemical exchange saturation transfer (CEST) and dark-state exchange saturation transfer (DEST) can be used to saturate the invisible state and observe the transfer of saturation by chemical exchange on the visible state [369, 370].

In intermediate to fast exchange (with timescales of the order of 100 s^{-1} to 2000 s^{-1}) the Carr-Purcell-Meiboom-Gill (CPMG) measurements are the most used. The CPMG experiment is based on the spin echo that refocuses the dephasing of the magnetization due to field inhomogeneity. In the absence of exchange, after a spin echo, magnetization is refocused. However, the random exchange between different precession frequencies during the spin echo due to chemical exchange prevents full refocusing of the signal. In the CPMG experiment, the refocusing pulses are applied at different frequencies (identified as ν_{CPMG}), as this frequency is increased dephasing is decreased usually giving rise to a lower value of $R_{2,eff}$ (which is composed of the transverse relaxation rate and the contribution from the chemical exchange). The plot of ν_{CPMG} versus $R_{2,eff}$ can be fit to the appropriate model of exchange and provides information on the rate of exchange between the states, their population and difference in chemical shift [371].

An experiment closely related to the CPMG experiment called $R_{1\rho}$ relaxation experiment measures spin relaxation in the rotating frame. With this, it is possible to use higher refocusing frequencies and thus it can probe protein dynamics on a timescale that is one order of magnitude faster than with CPMG [372].

NMR-active nuclei cause a weak magnetic field on other neighboring nuclei, this is called magnetic dipolar coupling. This effect is highly dependent on the orientation of these nuclei with respect to the external magnetic field. In isotropic solutions, no dipolar coupling is observed because molecular tumbling makes all orientations equally likely. However, if the protein is in a partially ordered solution called an alignment

medium the protein may align itself with the particles in the medium and present residual dipolar coupling (RDC). This is often used for structure determination [373, 374] but can also be used for the investigation of dynamics [375]. When intramolecular dynamics cause the protein to change conformations, the observed RDC is a population-weighted average. It is usually necessary to measure the protein in several alignment conditions to obtain a large dataset to obtain information on dynamic systems by RDC. RDCs can probe dynamics that are in timescales that are too fast to detect with CPMG experiments and too slow to characterized by analysis of the spin relaxation [376].

Very fast exchange, from the nanosecond to picosecond exchange, is on the timescale of spin relaxation. Therefore, measurement of the relaxation can probe into dynamics on this timescale such as internal dynamics and rotational diffusion of the molecule. Dynamics in this range is typically investigated by measuring the T_1 and T_2 relaxation rates and heteronuclear NOEs [377].

1.7.5 Interactions by NMR spectroscopy

Besides the study of protein dynamics, one of the main interests of using NMR spectroscopy for the study of proteins is to study interactions. The main technique that is used is the analysis of the chemical shift perturbation. In a titration experiment, a spectrum of the protein of interest labeled with an NMR-active nucleus, ^{15}N for example, is acquired in the absence and presence of increasing concentrations of its unlabeled ligand. This experiment allows for the simultaneous determination of dissociation constants (K_D) and a binding site. This strategy can be used with various ligands including both biomolecules and synthetic small molecules. The binding of a ligand L to a protein P can be described by the following reversible reaction where k_{on} and k_{off} are the forward and backward kinetic rate constants.

$$P + L \underset{k_{\text{off}}}{\overset{k_{\text{on}}}{\rightleftharpoons}} PL. \tag{1.7}$$

The concentrations of the free protein $[P]$, free ligand $[L]$ and complex $[PL]$ are related by the dissociation constant (K_D) by Equation 1.8.

$$K_D = \frac{[P][L]}{[PL]} \tag{1.8}$$

The dissociation constant is related to the rate constants by Equation 1.9.

$$K_D = \frac{k_{\text{off}}}{k_{\text{on}}} \tag{1.9}$$

The rate of exchange between the free and bound states will determine the effect observed in the titration spectra. If the exchange rate (in s^{-1}) is much faster than the chemical shift perturbation in Hz we will observe a gradual shifting of the resonance to the from the frequency of the free protein to the frequency of the complex as in this case the signal is a weighted average of signals due to free and bound states. In fast exchange the position of the resonance as a function of the ligand concentration can be used to determine the dissociation constant.

$$\Delta_{obs} = \Delta_{max} \frac{(K_D + [L]_0 + [P]_0) - \sqrt{(K_D + [L]_0 + [P]_0)^2 - (4[P]_0[L]_0)}}{2[P]_0} \tag{1.10}$$

where $[P]_0$ and $[L]_0$ represent the concentrations in solution of the labeled protein and the ligand, respectively. Δ_{obs} and Δ_{max} are the shifts at the observed concentration of ligand and at saturation, respectively. For two-dimensional spectra the shifts used in this calculation are usually combined chemical shifts perturbations (CSP). In this work we combined them as an Euclidean distance (see Equation 3.1) with a scaling factor to compensate for the larger dispersion in ppm of the ^{13}C and ^{15}N dimensions compared to the ^{1}H dimension.

On the other extreme, if the exchange rate is much slower than the chemical shift difference the resonance there will be only two resonance frequencies, those of free and bound states (Figure 1.31), with intensities corresponding to the populations of each state. If the exchange rate is similar to the chemical shift difference, the resonances will shift and broaden at the same time [378].

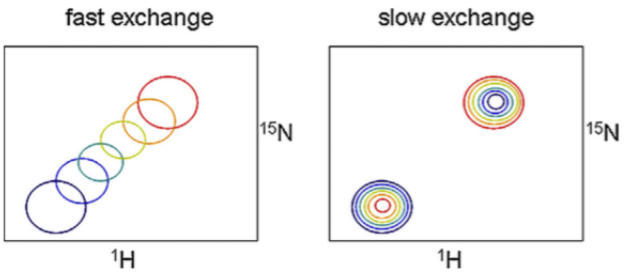

FIGURE 1.31: Effect of exchange rate on titration spectra. In fast exchange peaks that move gradually from the position of the free resonance to the position of the bound resonance. In slow exchange peaks become broadened when in equilibrium and reappear at the position of the bound resonance at saturation. Reprinted from Williamson [378], copyright 2013, with permission from Elsevier.

Broadening can also occur when the complex tumbles more slowly than the free protein even in the fast exchange regime. This is often observed with the binding of an IDP to a globular protein of high molecular weight which can significantly slow the tumbling rate of the residues near the binding site. Complex behavior of the resonances during a titration like non-linear shifts in fast exchange or broadening but not reappearing at a bound frequency at saturation, in case this cannot be explained by slow tumbling, suggest a binding mechanism that cannot be described by a single step, for example a change in conformation with the binding.

The addition of unlabeled ligand to a protein can also be used to determine the binding site of a ligand. In general, atoms that give rise to the signals that shift the most in the spectrum are often near the binding site of the ligand as long as the resonances due to residues at the binding site are present in the spectrum. However, chemical shift is sensitive to any change in the electronic environment of the observed nucleus. This means that a change in conformation upon binding can cause chemical shift perturbations away from the binding site. While the magnitude of the chemical shift perturbations may be indicative of the direct binding (with the largest shifts being due to direct binding) this experiment does not distinguish between direct binding or long-range allosteric effects.

Additionally, when the ligand in this experiment is also a protein, it can be done with each of the partners labeled, one at a time to gain complementary information on the binding sites on each of the partners and should independently give the same dissociation constant.

1.7.6 Nuclear Overhauser effect spectroscopy

The nuclear Overhauser effect (NOE) is the transfer of spin magnetization via a dipole-dipole interaction. Since this type of interaction occurs through space, the multidimensional experiments that makes use of this effect, called nuclear Overhauser effect spectroscopy (NOESY) provide information on atoms that are near in space, including those that may not be close through bonds.

Like for the HSQC experiment, this experiment is repeated multiple times with a variable evolution time, however it also includes a fixed delay called a mixing time in which the transfer of spin magnetization through dipolar interaction occurs. This results in a spectrum with a diagonal that resembles the 1D spectrum and additionally has off-diagonal cross-peaks at the frequencies of residues that are close in space.

The intensity of the NOE cross-peaks at first approximation is proportional to r^{-6}, where r is the distance between the two nuclei, and thus can be used as distance restraints. Most NOE cross-peaks that are visible in protonated proteins are due to nuclei less than 5-6 Å from each other. However, this is dependent on the density of protons in the protein. In the case of a perdeuterated protein with selective incorporation of protons into methyl sites, NOE cross-peaks have been observed for interatomic distances up to 12 Å [379] in a 20 kDa protein.

A 2D ^{1}H-^{1}H-NOESY experiment can be combined with a heteronuclear magnetization transfer, like the HSQC or HMQC to give three-dimensional spectra. In these 3D NOESY-HSQC and 3D NOESY-HMQC experiments the 2D NOESY is resolved along an additional heteronuclear dimension, reducing spectral overlap [380].

This is used to help in the assignment of a ^{1}H,^{13}C-HMQC spectrum of NS5B that has been perdeuterated and selectively labeled $^{13}C^{1}H_3$ on the side-chain methyl groups of isoleucine residues. Since the through-bond connectivities are lost when using this labeling strategy, the through-space magnetization transfer can be used to identify the atoms that are close in space by comparison to the crystallographic structures.

1.7.7 Paramagnetic relaxation enhancement

Much like the NOE, paramagnetic relaxation enhancement (PRE) provides information about distances through space. But PRE makes use of the the large magnetic moment of the electron. This causes dipolar coupling relaxation and consequently larger linewidths on the resonances of atoms in proximity of an unpaired electron. For paramagnetic centers with isotropic spin distribution this effect has a distance dependence of r^{-6}. However, compared to NOE this effect has a much larger effective distance due to the larger magnetic moment of the electron.

In practice, PRE is often done by labeling the protein with a spin label that contains a stabilized nitroxide radical. For this paramagnetic center, broadening is observed up to about 25 Å from the unpaired electron [381]. The most commonly used spin labels are MTSL, TEMPO, PROXYL and their derivatives (Figure 1.32). Depending on their functional groups these can be used to label cysteine side-chains through a disulfide bond (S-S) or a carbon-sulfur bond, which is not sensitive to the presence of reducing agents. Site-directed mutagenesis can be used to assure there is only a single cysteine and therefore a single labeled site on the protein [382].

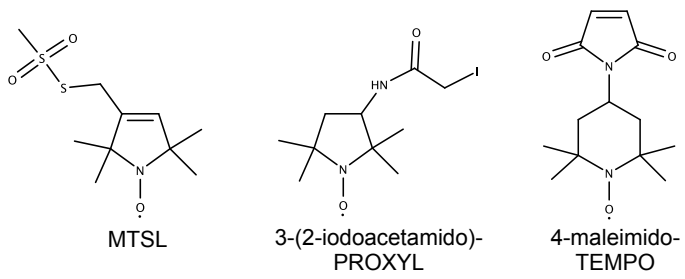

FIGURE 1.32: Examples of nitroxide-based spin labels used in the labeling of proteins. These spin labels can be incorporated into proteins to be used in NMR spectroscopy studies. MTSL (1-oxy(-2,2,5,5-tetramethyl-3-pyrroline-3-methyl) methanethiosulfonate), 3-(2-iodoacetamido)-PROXYL (3-(2-iodoacetamido)-2,2,5,5-tetramethyl-1-pyrolinyloxyl) and 4-maleimido-TEMPO (4-maleimido-2,2,6,6-tetramethyl-1-piperidinyloxyl) can be used to label cysteine residues [382].

The PRE NMR experiment is done by repeating the acquisition of the HSQC or HMQC spectrum with a paramagnetic and diamagnetic sample and comparing the intensities of the resonances. The diamagnetic sample can be obtained by reducing the nitroxide radical with ascorbic acid. The ratio of the intensities of a resonance between

the paramagnetic and diamagnetic samples can be related to distances in solution [381]. PRE can be used to gain intramolecular distance restraints if a single protein is labeled with NMR-active nuclei (^{15}N, for example) and with a single spin-label. PRE can also be used to obtain intermolecular distances between proteins by labeling one protein with the spin-label and another with the NMR-active nuclei. This can be used to determine a binding surface or in some cases to gain distance restraints for determination of the structure of the complex [383]. This has the advantage over CSP for the determination of a binding surface that it is not affected by allosteric effects.

Additionally, PREs can cause broadening even when the conformation where the nucleus is close to the unpaired electron is weakly populated, as long as the exchange rate between conformations is fast. Therefore, even if the paramagnetic atom is close to a nucleus only in a transient conformation (with populations as low as 1%), this will cause broadening that can be detected [381, 384]. This can be used to identify minor states in fast exchange based on a mismatch between PREs calculated from the structure of a major state and the experimental PREs.

The use of paramagnetic labels in protein NMR is not a recent development [385] but has seen a resurgence due to its application in the investigation of long-range order in IDPs [386] and application in studies with selectively labeled proteins where NMR-active nuclei may be too distant for structural information using NOEs, particularly in the case of supramolecular complexes [387].

Objectives

The hepatitis C virus genome codes for 10 viral proteins, one of which is the RNA-dependent RNA polymerase responsible for its RNA replication, NS5B. With a limited number of viral proteins, the virus makes use of proteins that interact with a large number of partners, both host and viral. One such promiscuous protein is NS5A. Its three domains are necessary in different steps of the viral lifecycle: in RNA replication, virion formation and regulation of the host metabolism. However its precise role in these activities is not well elucidated.

The main aim of this work is to gain insight into the structure and dynamics of the HCV RNA-dependent RNA polymerase NS5B in solution. Additionally, we wanted to understand the role of the intrinsically disordered domain 2 of NS5A in RNA replication via its regulation of NS5B.

To achieve these main objectives, initially we characterized the polymerase in solution using NMR spectroscopy. In particular, we were interested in characterizing the conformations of the polymerase that are expected to exist in solution and dynamics present in different timescales. Then, we investigated the binding site of NS5A-D2 on NS5B as well as the effect of the domain on the stability of NS5B. Additionally, we compared the effect of NS5A-D2 on NS5B to the effect of a small molecule inhibitor that has a partially overlapping binding site. Finally, the effect of NS5A-D2 and the small molecule inhibitor on RNA binding by the polymerase were examined and compared to the structural and dynamic data obtained by NMR.

An independent main objective in this work was to investigate the role of domain 2 of NS5A in regulation of the host metabolism. A study by Ramière and coworkers [296] identified the activation of hexokinase II in the context of overexpression of NS5A in hepatoma cell lines. Hexokinases catalyze the phosphorylation of hexoses, the first step of the metabolic glycolysis pathway. Since the hexokinase mainly present in hepatocytes is hexokinase IV, or glucokinase, we investigated the interaction between NS5A-D2 and glucokinase to elucidate the activation of the enzyme by the viral protein.

Experimental Section

3.1 Expression and purification of recombinant NS5B JFH-1

Expression of the recombinant NS5B from HCV JFH-1 isolate (GenBank AB047639) fused to a C-terminal His_6 tag was performed in *Escherichia coli* BL21(DE3). The pET21-NS5B Δ21 wild-type (corresponding to residues 1-570 of NS5B JFH-1) and pET21-NS5B Δ60 wild-type (corresponding to residues 1-531 of NS5B JFH-1) plasmids that contain extra MA- and -HHHHHH extensions at the N and C termini were kindly provided by Dr. S. Bressanelli (CNRS, Gif-sur-Yvette, France) [134]. All other plasmids coding for NS5B were obtained by site-directed mutation. The pET-derived plasmids were introduced into the *E. coli* BL21(DE3) cells by electroporation and transformed cells were selected by resistance to ampicilin on agar plates. A pre-culture of 25 mL Luria-Bertani (LB) culture medium supplemented with ampicillin (100 mg/L) was inoculated with an isolated colony and grown to saturation at 37°C prior to inoculate the culture mediums used for protein expression.

Non-labeled NS5B

To obtain non-labeled $NS5B_{\Delta 21}$ and $NS5B_{\Delta 60}$ 1 liter of LB medium supplemented with ampicillin (100 mg/L) was inoculated with cells from the pre-culture that had been centrifuged at 4000 $\times$ g to remove the liquid culture medium to obtain an initial optical density at 600 nm (OD600) of 0.05. Then, cells were grown at 37°C in ampicillin-supplemented Luria-Bertani medium and when the OD600 of the culture reached about 0.7 the temperature was lowered to 22°C. Protein production was then induced with 0.4 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG) for 23h. Cells were harvested by centrifugation at 5000 $\times$ g for 20 minutes and resuspended in 40 mL of lysis buffer (20 mM Tris-Cl pH 7.5, 500 mM NaCl, 1 tablet protease inhibitors (Complete EDTA-free, Roche Applied Science) 10% v/v glycerol). DNAse I (22.9 mg/L) and RNAse A (12.5 mg/L) were added to the suspension. Cells were lysed with a high-pressure homogenizer (Emulsiflex-B30) at 30 psi then centrifuged at 35000 $\times$ g at 4°C for 45 minutes to obtain the protein-containing supernatant. Subsequently, the supernatant was

filtered at 0.45 μm. NS5B was purified by Ni^{2+} affinity chromatography (1 mL His-Trap column GE Healthcare) using an ÄKTA pure fast protein liquid chromatography (FPLC) system with UV detection at 215, 260 and 280 nm. The supernatant was loaded onto the column previously equilibrated with buffer A (50 mM potassium phosphate buffer pH 7.6, 500 mM NaCl). Then column was washed with buffer A containing 20 mM imidazole to remove non-specific binding. The protein was eluted with buffer A and a gradient of imidazole (20–375 mM over 10 column volumes). NS5B-containing fractions were identified by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) then pooled and dialyzed twice at 15°C against 3L of 20 mM Tris-Cl pH 7.0, 300 mM NaCl, 2 mM $MgCl_2$, 0.5 mM EDTA. 4 mM THP was added after dialysis. NS5B was then concentrated with a VivaSpin 4 (cutoff 30 kDa, Sartorius Stedim Biotech) to about 110 μM or until significant precipitation was present. 350 μL aliquots were flash-frozen in liquid nitrogen and stored at -80°C until use.

U-[^{15}N,^{13}C,^{2}H],-labeled NS5B

To obtain uniformly [^{15}N,^{13}C,^{2}H]-labeled $NS5B_{\Delta 21}$, cells were grown in deuterated M9-based semi-rich medium (M9 medium in 99.9% D_2O with D-glucose-$^{13}C_6d_7$ (3 g/L), [^{15}N]NH_4Cl (1 g/L) and ^{15}N, ^{13}C, ^{2}H-rich medium (0.5 g/L) (Isogro ^{15}N,^{13}C, D, Isotec) supplemented with 100 mg/L of ampicillin. Otherwise, culture and purification were as for non-labeled NS5B. After purification NS5B-containing fractions were dialyzed twice at 15°C against 3L of 300 mM sodium phosphate buffer (pH 6.8), 50 mM NaCl, 2 mM $MgCl_2$, 0.5 mM EDTA, 5 mM DTT. 4 mM THP was added after dialysis. NS5B was then concentrated with a VivaSpin 4 (cutoff 30 kDa, Sartorius Stedim Biotech) to 110 μM and 5 % D_2O was added. 300 μL aliquots were flash-frozen in liquid nitrogen and stored at -80°C until use.

U-[^{15}N,^{2}H]-labeled NS5B

To obtain uniformly [^{15}N,^{2}H]-labeled $NS5B_{\Delta 21}$, cells were grown in deuterated M9-based semi-rich medium (M9 medium in 99.9% D_2O with D-glucose-d_7 (3 g/L), $^{15}NH_4Cl$ (1 g/L) and ^{15}N, ^{2}H-rich medium (0.5 g/L) (Isogro ^{15}N, D, Isotec) supplemented with 100 mg/L of ampicillin. Otherwise, culture and purification were as for non-labeled

NS5B. After purification NS5B-containing fractions were dialyzed twice at 15°C against 3L of 300 mM sodium phosphate buffer (pH 6.8), 50 mM NaCl, 2 mM $MgCl_2$, 1 mM EDTA. 4 mM THP was added after dialysis. NS5B was then concentrated with a VivaSpin 4 (cutoff 30 kDa, Sartorius Stedim Biotech) to 110 μM and 5 % D_2O was added. 350 μL aliquots were flash-frozen in liquid nitrogen and stored at -80°C until use.

U-[^{2}H], Ile δ1-[^{13}C^{1}H$_3$]-labeled NS5B

Transformed *Escherichia coli* BL21(DE3) cells were grown in deuterated M9-based semi-rich medium (M9 medium in 99.9% D_2O with D-glucose-d_7 (3 g/L), NH_4Cl (1 g/L) and ^{15}N,^{2}H-rich medium (Isogro ^{15}N, D, Isotec) (0.5 g/L)) supplemented with ampicillin (100 mg/L). Cells were grown at 37°C. Isoleucine ^{13}C^{1}H$_3$-labeling was achieved by introduction of 2-ketobutryric acid-4-^{13}C,3,3-d_2 sodium salt (60 mg/L)(Sigma Aldrich) when the culture reached OD600 $\approx$ 0.8. 2-ketobutyric acid-4-^{13}C,3,3-d_2 was prepared by the overnight incubation of 2-ketobutyric acid-4-^{13}C in 50 mL D_2O at 45°C with pH adjusted to $\approx$ 10. One hour after this addition the temperature was lowered to 22°C and protein production was induced with 0.4 mM IPTG for 22 h. Cells were harvested by centrifugation and resuspended in 50 mL of lysis buffer per liter of culture (20 mM Tris buffer pH 7.0, 500 mM NaCl, 10% Glycerol v/v, 1 cOmplete EDTA-free protease inhibitor tablet). Purification by nickel-affinity chromatography and storage were as for unlabeled NS5B.

Trp ε1-[^{15}N]-labeled NS5B

Transformed cells were grown in M9 minimum medium. When OD600 $\approx$ 0.8, 50 mg of ^{15}N-indole dissolved in 500 μL of DMSO was added to the culture. Fifteen minutes after addition of the precursor the temperature was lowered to 22°C. Protein production was then induced with 0.4 mM IPTG for 23h. Dialysis was done against 300 mM sodium phosphate buffer (pH 6.8), 50 mM NaCl, 2 mM $MgCl_2$, 1 mM EDTA. Otherwise, culture, purification and storage were as for non-labeled NS5B.

3.2 Expression and purification of NS5A-D2

In most of this work NS5A-D2 from the JFH-1 isolate was used. However, in Chapter 8 both NS5A-D2 from the JFH-1 (genotype 2a) and Con1 (genotype 1b) isolates were used. The sequence of NS5A-D2 JFH-1 used in this study corresponds to residues 248–341 from the HCV JFH-1 isolate (GenBank AB047639 [388], genotype 2a) and the sequence of NS5A-D2 Con1 corresponds to the residues 245–341 from the HCV Con1 isolate (GenBank AJ238799, genotype 1b). Both have extra M- and –LQHHHHHH extensions at the N- and C-termini and were introduced into a pT7-7 expression vector. Other plasmids of NS5A-D2 with single site mutations were obtained by site-directed mutagenesis.

Non-labeled NS5A-D2 (JFH-1 and Con1)

The plasmids were introduced into *E. coli* BL21(DE3) cells by electroporation. Transformed cells were selected by growth on ampicillin-containing (100 mg/L) agar plates. A 25 mL pre-culture in ampicillin-supplemented LB was grown to saturation from an isolated colony. This was used to inoculate 1L of LB with 100 mg/L of ampicillin to an initial OD600 of 0.05. Cells were grown at 37°C. At an OD600 of about 0.7 the protein production was induced with 0.4 mM IPTG and cells were harvested 3.5 hours post-induction by centrifugation at 5000 × g for 20 minutes. DNAse I (22.9 mg/L) and RNAse A (12.5 mg/L) were added to the suspension. Cells were lysed with a high-pressure homogenizer (Emulsiflex-B30) at 30 psi then centrifuged at 35000 × g for 45 min at 4°C to obtain the protein-containing supernatant. NS5A-D2 was purified in two steps. As an intrinsically disordered protein, NS5A-D2 is thermostable therefore a heating step at 75°C for 15 minutes was used to denature most bacterial proteins. Denatured proteins were separated by centrifugation at 35000 × g for 20 minutes at 4°C. Then Ni^{2+} affinity chromatography (1 mL HisTrap column GE Healthcare) using an ÄKTA pure fast protein liquid chromatography (FPLC) system with UV detection at 215, 260 and 280 nm was used. The supernatant was loaded onto the column previously equilibrated with buffer A (50 mM sodium phosphate pH 7.8 and 300 mM NaCl).

Then column was washed with buffer A containing 20 mM imidazole to remove non-specific binding. The protein was eluted with buffer A and a gradient of imidazole (20–250 mM over 5 column volumes) NS5A-D2-containing fractions were identified by SDS-PAGE then pooled and dialyzed twice against 3L of 50 mM sodium phosphate buffer pH 6.4 (pH 6.8 if producing NS5A-D2 Con1), 50 mM NaCl and 0.2 mM EDTA. 1 mM of THP was added after dialysis. The protein was concentrated up to 300 μM using a Vivaspin 4 concentrator (cutoff 5 kDa) (Sartorius Stedim Biotech). 350 μL aliquots were flash-frozen and stored at -80°C until use.

^{15}N-labeled NS5A-D2

To obtain ^{15}N-labeled NS5A-D2, cells were grown in M9-based semi-rich medium supplemented with [^{15}N]NH_4Cl (1 g/L) and Isogro ^{15}N powder growth medium (0.5 g/L, Sigma-Aldrich). Otherwise, culture, purification and storage were as for non-labeled NS5A-D2.

Nitroxide-labeling of NS5A-D2

Wild-type NS5A-D2 of the JFH-1 strain contains two natural cysteines at positions 298 and 338. Three single cysteine constructs of NS5A-D2 were obtained by site-directed mutagenesis: NS5A-D2$_{C254}$, NS5A-D2$_{C298}$, NS5A-D2$_{C338}$. NS5A-D2$_{C298}$ and NS5A-D2$_{C338}$ were obtained by substituting cysteines 338 and 298 for serine, respectively. NS5A-D2$_{C254}$ was obtained by substituting both natural cysteines for serines and the substitution of the methionine 254 for a new cysteine. These constructs were expressed and purified as described for wild-type non-labeled NS5A-D2. The proteins' buffer was exchanged by Zeba Spin desalting column (cutoff 7 kDa, 2 mL, Thermo Scientific) to ammonium carbonate buffer (100 mM NH_4HCO_3, 50 mM NaCl, 1 mM THP) prior to the labeling. The cysteine residues of these proteins were labeled with a nitroxide group by the addition of a 25-fold molar excess of 3-(2-Iodoacetamido)-PROXYL (Sigma Aldrich) and incubated at room temperature for 3 h, protected from light. The unreacted reagent was quenched with 5 mM 1,4-dithiothreitol (DTT). Zeba Spin desalting columns were used to exchange the buffer to the NMR buffer of the partner protein prior to the use in NMR experiments.

3.3 Expression and purification of glucokinase

Human glucokinase (GCK) (Uniprot P35357, residues 16-465) was cloned into a pETNKI-His-3C vector. Expression of the recombinant human GCK fused to a N-terminal hexahistidine tag was performed in *Escherichia coli* BL21(DE3) in a culture medium supplemented with kanamycin (25 mg/L).

Non-labeled GCK

Transformed cells were grown at 37°C in Luria-Bertani medium and when the culture reached OD600 $\approx$ 0.6 the temperature was lowered to 22°C. Protein production was then induced with 0.4 mM isopropyl 1-thio-β-D-galactopyranoside for 23h. Cells were harvested by centrifugation and resuspended in 40 mL of lysis buffer (50 mM sodium phosphate buffer pH 7.8, 300 mM NaCl, protease inhibitors (Complete EDTA-free, Roche Applied Science) 10% v/v glycerol). Cells were lysed with a high-pressure homogenizer (Emulsiflex-B30) at 40 psi then centrifuged at 35000 x g for 45 minutes at 4°C. GCK was purified by Ni^{2+} affinity chromatography (1 mL HisTrap column GE Healthcare). The supernatant was loaded onto the column previously equilibrated with buffer A (50 mM sodium phosphate buffer pH 7.8, 300 mM NaCl). The column was washed with buffer A containing 20 mM imidazole and the protein was eluted with buffer A and a gradient of imidazole (20–250 mM over 8 column volumes). GCK-containing fractions were pooled and dialyzed twice against 3L of 50 mM potassium phosphate buffer pH 7.4, 150 mM NaCl 5 mM DTT 0.2 mM EDTA. 5 mM THP and 10% v/v glycerol was added after dialysis. GCK was then concentrated with a VivaSpin 4 centrifugal concentrator (cutoff 30 kDa, Sartorius Stedim Biotech) to about 400 μM. 400 μL aliquots were flash-frozen in liquid nitrogen and stored at -80°C until use.

U-[^{15}N, ^{2}H], Ile δ1-[^{13}C^{1}H$_3$]-labeled GCK

Transformed cells were grown in semi-rich deuterated M9-based medium (M9 medium in 99.8% D_2O with 1 g/L of $^{15}NH_4Cl$, 3 g/L of D-$^{2}H_7$glucose and 5 mL/L of ^{2}H,^{15}N-rich 10X concentrate culture medium (BioExpress, Cambridge Isotope Laboratories))

and when the optical density at 600 nm reached 0.7, 60 mg/L of 2-ketobutyric acid-4-^{13}C,3,3-d$_2$ sodium salt hydrate was added to the culture. One hour after addition of the precursor the temperature was lowered to 22°C. Protein production was then induced with 0.4 mM isopropyl 1-thio-β-D-galactopyranoside for 23h. Otherwise, culture, purification and storage were as for non-labeled GCK. Before use for NMR experiments protein sample was unfrozen and buffer exchanged to 50 mM potassium phosphate buffer pH 6.8, 50 mM KCl, 0.2 mM EDTA, 5 mM THP, 5% v/v glycerol-d$_5$ in 99.9% D$_2$O) by Zeba Spin Desalting column (cutoff 7 kDa, 2 mL, Thermo Scientific).

Trp ε1-[^{15}N]-labeled GCK

Transformed cells were grown in M9 minimum medium. When OD600 $\approx$, 0.8 50 mg of ^{15}N-indole dissolved in 500 μL of DMSO was added to the culture. Fifteen minutes after addition of the precursor the temperature was lowered to 22°C. Protein production was then induced with 0.4 mM isopropyl 1-thio-β-D-galactopyranoside for 23h. Otherwise, culture, purification and storage were as for non-labeled GCK.

Cleavage of hexahistidine purification tag

The purification tag fused to GCK was cleaved using a recombinant human rhinovirus 3C protease fused to glutathione S-transferase (GST). The buffer containing GCK was exchanged to cleavage buffer (50 mM TrisCl pH 7.0, 150 mM NaCl, 1 mM EDTA and 1 mM DTT) by a 2 mL Zeba Spin desalting column (Thermo Fisher) and 2 units of protease were added per 100 μg of GCK. This mixture was left incubating with agitation at 5°C for 4 hours. The protease was removed using Glutathione Sepharose beads (GE Healthcare) previously equilibrated with the cleavage buffer that were left incubating overnight at 4°C with agitation. The cleaved protein was recovered by centrifugation at 8000 $\times$ g for 15 minutes to separate the beads. The buffer was then exchanged to NMR buffer by a 2mL Zeba Spin desalting column.

3.4 Site-directed mutagenesis

Site-directed mutagenesis for both NS5B$_{\Delta 21}$ and NS5A-D2 single-site mutations was done either by GeneCust (Luxembourg) or in-house using a QuikChange II Site-Directed Mutagenesis kit (Agilent Technologies, Santa Clara, CA, USA).

3.5 Nuclear magnetic resonance spectroscopy

3.5.1 Spectrometers and data treatment

NMR experiments were performed either on a Bruker 900 MHz Avance III HD spectrometer equipped with a CPTCI 5 mm cryoprobe or a Bruker 600 MHz Avance I spectrometer equipped with a CPQCI 5 mm cryoprobe. 5 mm Shigemi D_2O matched NMR tubes with 300–400 μL samples were used for all NMR experiments. 3-(Trimethylsilyl)propionic-2,2,3,3-d$_4$ acid sodium salt (TMSP) was used as an internal standard. Spectra were acquired and processed with TopSpin 3.5 software (Bruker). They were further visualized and analyzed using the NMRFAM-Sparky software package [389, 390]. Combined chemical shift perturbations (CSP) were calculated using Equation 3.1,

$$\Delta\delta = \sqrt{\Delta\delta(^1H)^2 + (\alpha \times \Delta\delta(X))^2} \tag{3.1}$$

where X represents either ^{13}C or ^{15}N depending the on the heteroatom being observed, α is a scaling factor that is representative of the dispersion of the cross-peaks in the indirect dimension. In this work, for ^{15}N-^{1}H correlations the value used was 0.2 and for ^{13}C-^{1}H methyl correlations the value used was 0.16.

3.5.2 NMR experiments on [^{15}N,^{2}H]-labeled NS5B

NMR experiments containing NS5B were performed at 305K on a Bruker 900 MHz Avance III HD spectrometer equipped with a CPTCI 5 mm cryoprobe.

^{1}H,^{15}N-HSQC spectra were acquired with a band-selective excitation short-transient (BEST) TROSY HSQC pulse sequence with 160 scans, 2048 and 512 points and sweep widths of 14 and 31.3 ppm in the direct and indirect dimensions, respectively. The

buffer used was 300 mM sodium phosphate pH 6.8, 50 mM NaCl, 1 mM $MgCl_2$, 0.5 mM EDTA, 4 mM THP and 5 % D_2O.

These experiments were done with samples containing either 89 μM U-[^{15}N,^{2}H]-labeled NS5B$_{\Delta 21}$, both 69 μM U-[^{15}N,^{2}H]-labeled NS5B$_{\Delta 21}$ and 69 μM nitroxide-labeled NS5A-D2$_{C338}$ in the absence and presence of 4 mM ascorbic acid or 89 μM U-[^{15}N,^{2}H]-labeled NS5B$_{\Delta 21}$ saturated with filibuvir (by incubating overnight with 1 mg of solid filibuvir).

Tentative Assignment

The assignment of the HSQC spectrum of NS5B was attempted using a triple resonance backbone assignment strategy with BEST pulse sequences [391]. This was done with uniformly triple-labeled samples (^{15}N, ^{13}C and perdeuterated) in 300 mM sodium phosphate pH 6.8, 50 mM NaCl, 2 mM $MgCl_2$, 1 mM EDTA and 4 mM THP in with 5% D_2O.

A ^{1}H,^{15}N BEST-TROSY-HSQC spectrum was acquired with a ^{15}N, ^{13}C, ^{2}H-labeled NS5B sample at 110 μM with 2048 and 64 complex points in the direct and indirect dimensions respectively and with 80 scans. The same sample was used for BEST-TROSY-HNCA and BEST-TROSY-HNCACB experiments with 2048, 96 and 70 complex points in the direct, indirect (^{13}C) and indirect (^{15}N) dimensions respectively and with 112 scans for the BEST-TROSY-HNCA and 2048, 116 and 40 complex points in the direct, indirect (^{13}C) and indirect (^{15}N) dimensions respectively and with 368 scans for the BEST-TROSY-HNCACB. To obtain further C_β i-1 resonances, a BEST-TROSY-HN(COCA)CB was used with 2048, 42 and 120 complex points in the direct, indirect (^{15}N) and indirect (^{13}C) dimensions and 352 scans. The BEST-TROSY-HNCACB and BEST-TROSY-HN(COCA)CB were acquired with non-uniform sampling 60%.

Furthermore a new sample at 101 μM was used to acquire BEST-TROSY-HNCO and BEST-TROSY-HN(CA)CO spectra with 2048, 48 and 48 points in the direct, indirect (^{15}N) and indirect (^{13}C) dimensions. The BEST-TROSY-HNCO and the BEST-TROSY-HN(CA)CO were acquired with 272 and 688 scans respectively, both with non-uniform sampling 60%.

3.5.3 NMR experiments on selectively isoleucine-labeled NS5B

Spectra of uniformly deuterated and selective isoleucine δ1-[$^{13}C^{1}H_3$]-labeled NS5B$_{\Delta 21}$ samples were acquired with a ^{1}H,^{13}C-TROSY HMQC spectrum centered on the resonances of the δ1 methyl group of the isoleucine side chain (called methyl-HMQC for clarity). Control spectra containing only NS5B$_{\Delta 21}$ were acquired with samples of 30 to 100 μM of polymerase with 2048 and 128 complex points and 14 and 10 ppm in the direct and indirect dimensions respectively. The number of scans varied from 16 to 96.

These experiments were done in one of two buffers. The phosphate buffer used was 300 mM sodium phosphate pH 6.8, 50 mM NaCl, 2 mM $MgCl_2$, 1 mM EDTA and 4 mM THP in D_2O. The RE buffer used was 20 mM TrisCl buffer pH 7.0, 100 mM NaCl, 100 mM arginine, 100 mM glutamate, 2 mM $MgCl_2$, 1 mM EDTA 4 mM THP in 99.9% D_2O. The buffer was changed from the storage buffer to one of these buffers just prior to the NMR experiments by Zeba spin desalting column (2 mL, cutoff 7 kDa).

Assignment of the methyl-HMQC

Isoleucine-to-valine single substitution mutants and NOESY spectra were used to assign the methyl-HMQC spectrum. The resulting assignment was deposited in the Biological Magnetic Resonance Data Bank under accession number under 27073.

Isoleucine-to-valine substitution The methyl-HMQC spectra of several isoleucine-to-valine substituted mutants of NS5B$_{\Delta 21}$ were used to assign resonances on the spectrum of the wild-type protein. The substitutions used were I23V, I119V, I160V, I233V, I363V, I405V, I412V I413V, I419V, I432V, I462V, I463V, I512V and I560V. Methyl-HMQC spectra of these mutants were acquired with samples of about 30 μM with 32 scans, 2048 and 128 complex points, and 14 and 10 ppm in the direct and indirect dimensions, respectively.

NOESY spectra A 3D ^{1}H,^{13}C NOESY-HMQC and a 3D ^{1}H,^{13}C HMQC-NOESY-HMQC were acquired with a sample of 100 μM U-[^{2}H], Ile δ1-[$^{13}C^{1}H_3$]-labeled NS5B$_{\Delta 21}$ in phosphate buffer. The 3D ^{1}H,^{13}C NOESY-HMQC was acquired with 0.9 s mixing time, 16 scans, 2048, 166 and 48 points, and 14, 9.5 and 10 ppm in the direct, indirect (^{1}H) and

indirect (^{13}C) dimensions, respectively The 3D ^{1}H,^{13}C HMQC-NOESY-HMQC was acquired with a 1.0 s mixing time, 16 scans, 2048, 66 and 66 points, and 14, 10 and 10 ppm in the direct, indirect (^{13}C) and indirect (^{13}C) dimensions, respectively.

NS5B open constructs

Methyl-HMQC spectra were acquired of various constructs of NS5B that have been described as being more open than wild-type NS5B$_{\Delta 21}$.

NS5B$_{\Delta 60}$ A methyl-HMQC spectrum of a sample of 20 μM U-[^{2}H], Ile δ1-[^{13}C^{1}H$_3$]-labeled NS5B$_{\Delta 60}$ in phosphate buffer was acquired with 2048 and 128 complex points and 14 and 10 ppm in the direct and indirect dimensions and 160 scans.

NS5B$_{\Delta 21}$ L30S A methyl-HMQC spectrum of a sample of 20 μM U-[^{2}H], Ile δ1-[^{13}C^{1}H$_3$]-labeled NS5B$_{\Delta 21}$ L30S in RE buffer was acquired with 2048 and 140 complex points and 14 and 10 ppm in the direct and indirect dimensions and 228 scans.

NS5B$_{\Delta 21}$ 5mut A methyl-HMQC spectrum of a sample of 37 μM U-[^{2}H], Ile δ1-[^{13}C^{1}H$_3$]-labeled NS5B$_{\Delta 21}$ S15G E86Q E87Q C223H V321I (called 5mut) in RE buffer was acquired with 2048 and 140 complex points and 14 and 10 ppm in the direct and indirect dimensions and 160 scans. A ^{13}C-decoupled 2D ^{1}H-^{1}H NOESY spectrum of a 31 μM sample of U-[^{2}H], Ile δ1-[^{13}C^{1}H$_3$]-labeled NS5B$_{\Delta 21}$ 5mut was acquired with 2048 and 200 complex points and 14 and 11 ppm in the direct and indirect dimensions, respectively with 32 scans.

Experiments with NS5A-D2

Titration with NS5A-D2 A titration of U-[^{2}H], Ile δ1-[^{13}C^{1}H$_3$]-labeled NS5B$_{\Delta 21}$ was done with a sample containing 40 μM selective isoleucine-labeled NS5B$_{\Delta 21}$ to which 0, 20, 40, 80 μM of unlabeled NS5A-D2 JFH-1 was added. Independent NMR samples contained 40 μM selective isoleucine-labeled NS5B$_{\Delta 21}$ and either 200 or 400 μM of NS5A-D2 JFH-1 were also used to avoid significant dilution. The titration done in phosphate buffer was followed by methyl-HMQC spectra with 2048 and 256 complex points, 14 and 20 ppm in the direct and indirect dimensions, respectively with 32 scans.

Paramagnetic relaxation enhancement Intermolecular paramagnetic relaxation enhancement (PRE) experiments were done with samples containing 71 μM of U-[^{2}H], Ile δ1-[^{13}C^{1}H$_3$]-labeled NS5B$_{\Delta 21}$ and an equimolar amount of nitroxide-labeled NS5A-D2 in phosphate buffer. Methyl-HMQC spectra were acquired with 2048 and 280 points, and 14 and 20 ppm in the direct and indirect dimensions, respectively and 112 scans. 4 mM of ascorbic acid was added to an independent and identical sample to reduce the paramagnetic label and obtain a diamagnetic control spectrum acquired in the same conditions. This was done with the three single cysteine constructs of NS5A-D2: NS5A-D2$_{C254}$ (M254C, C298S, C338S), NS5A-D2$_{C298}$ (C338S) and NS5A-D2$_{C338}$ (C298S).

Effect of PW-turn A spectrum of 71 μM U-[^{2}H], Ile δ1-[^{13}C^{1}H$_3$]-labeled NS5B$_{\Delta 21}$ with an equimolar amount of unlabeled NS5A-D2 A311G JFH-1 in phosphate buffer was used to determine the effect of the proline-tryptophan (PW)-turn motif to NS5B$_{\Delta 21}$ binding to NS5A-D2. The methyl-HMQC spectrum was acquired with 2048 and 280 points, and 14 and 20 ppm in the direct and indirect dimensions, respectively and 112 scans. This was compared to a spectrum of U-[^{2}H], Ile δ1-[^{13}C^{1}H$_3$]-labeled NS5B$_{\Delta 21}$ with NS5A-D2 C298S (NS5A-D2$_{C338}$) because both were acquired with NS5B$_{\Delta 21}$ from the same culture batch and under the same conditions.

Experiments with Filibuvir

Filibuvir was added to samples using about 1 mg of solid powder filibuvir (Sigma-Aldrich) directly added to 350-400 μL samples of NS5B$_{\Delta}$21 (or mutant) and incubated for 6-12 hours at 15°C with agitation to obtain NS5B$_{\Delta 21}$ saturated with the inhibitor. Excess solid filibuvir was removed by centrifugation for 10 s at 10 000 x g at room temperature.

NS5B$_{\Delta 21}$ wild-type A methyl-HMQC spectrum was acquired of 81 μM filibuvir-saturated U-[^{2}H], Ile δ1-[^{13}C^{1}H$_3$]-labeled NS5B$_{\Delta 21}$. This spectrum was acquired in phosphate buffer with 2048 and 140 complex points, 14 and 10 ppm and 96 scans.

NS5B$_{\Delta 21}$ L30S A methyl-HMQC spectrum was acquired of 20 μM filibuvir-saturated U-[^{2}H], Ile δ1-[^{13}C^{1}H$_{3}$]-labeled NS5B$_{\Delta 21}$ L30S. This spectrum was acquired in RE buffer with 2048 and 140 complex points, 14 and 10 ppm and 226 scans.

NS5B$_{\Delta 21}$ 5mut A methyl-HMQC spectrum was acquired of 37 μM filibuvir-saturated U-[^{2}H], Ile δ1-[^{13}C^{1}H$_{3}$]-labeled NS5B$_{\Delta 21}$ 5mut. This spectrum was acquired in RE buffer with 2048 and 140 complex points, 14 and 10 ppm and 160 scans.

NS5B$_{\Delta 21}$ with NS5A-D2 A methyl-HMQC spectrum was acquired of 48 μM filibuvir-saturated U-[^{2}H], Ile δ1-[^{13}C^{1}H$_{3}$]-labeled NS5B$_{\Delta 21}$ with an equimolar amount of NS5A-D2 added. This spectrum was acquired in phosphate buffer with 2048 and 240 complex points, 14 and 20 ppm and 112 scans. An independent sample in the same conditions without NS5A-D2 was acquired as control.

Titration with ssRNA

A titration of U-[^{2}H], Ile δ1-[^{13}C^{1}H$_{3}$]-labeled NS5B$_{\Delta 21}$ with a single strand of RNA (5'-CUAAGAUGCUCGCUGC-3') (RNA16) (GeneCust, Luxembourg) was done with independent samples containing 30 μM of NS5B$_{\Delta 21}$ and either 0, 10, 30 or 60 μM of unlabeled RNA16 in RE buffer. This titration was followed by methyl-HMQC spectra acquired with 240 scans, 2048 and 128, and 14 ppm and 10 ppm. After acquisition, 0.5 mg of filibuvir powder was added to the sample containing 30 μM RNA16 and incubated overnight at 15°C, then another methyl-HMQC spectrum was acquired of this sample with the same parameters.

3.5.4 NMR experiment on selectively tryptophan-labeled NS5B

A ^{1}H,^{15}N-BEST-TROSY-HSQC spectrum of a 58 μM Trp ε1-[^{15}N]-labeled sample of NS5B$_{\Delta 21}$ was acquired in 300 mM sodium phosphate pH 6.8, 50 mM NaCl, 2 mM $MgCl_2$, 1 mM EDTA, 4 mM THP and 5% D_2O. This spectrum was acquired with 512 scans, 2048 and 64 points and 14 and 12 ppm in the direct and indirect dimensions, respectively with the center of the ^{15}N dimension at 126 ppm.

3.5.5 NMR experiments on selectively isoleucine-labeled GCK

NMR experiments containing GCK were performed at 305K on a Bruker 900 MHz Avance III HD spectrometer equipped with a CPTCI 5 mm cryoprobe. Experiments on U-[^{15}N,^{2}H], Ile δ1-[^{13}C^{1}H$_3$]-labeled GCK were done in 50 mM potassium phosphate buffer pH 6.8, 50 mM KCl, 0.2 mM EDTA, 5 mM THP and 5% glycerol-d$_5$ in 99.9% D$_2$O.

Control methyl-HMQC spectra were acquired with 313 μM U-[^{15}N,^{2}H], Ile δ1-[^{13}C^{1}H$_3$]-labeled GCK, unliganded, in the presence of 200 mM D-glucose, and with both 200 mM D-glucose and 10 mM adenylyl-imidodiphosphate (AMP-PNP). These spectra were acquired with 2048 and 200 points, 14 and 11 ppm in direct and indirect dimensions, respectively with 32 scans.

A spectrum of 82 μM U-[^{15}N,^{2}H], Ile δ1-[^{13}C^{1}H$_3$]-labeled GCK that had its purification tag previously removed was acquired in the same conditions.

Experiments with NS5A-D2

To 100 μM of U-[^{15}N,^{2}H], Ile δ1-[^{13}C^{1}H$_3$]-labeled GCK 100 μM and 300 μM unlabeled NS5A-D2 (lyophilized) was added. Methyl-HMQC spectra were acquired in the absence of NS5A-D2 and at the two ratios with 2048 and 200 points and 14 and 11 ppm in the direct and indirect dimensions, respectively with 32 scans. After, 100 mM D-glucose was added and another spectrum under the same conditions was acquired.

3.5.6 NMR experiments on selectively tryptophan-labeled GCK

^{15}N,^{1}H-HSQC spectra of 315 μM Trp ε1-[^{15}N]-labeled GCK unliganded, in the presence of 200 mM D-glucose and with both 200 mM D-glucose and 10 mM AMP-PNP were acquired with 2048 and 128 points, and 16 and 12 ppm, in the direct and indirect dimensions, 72 scans and the center of the ^{15}N dimension set to 126 ppm. 50 mM potassium phosphate buffer pH 6.8, 50 mM KCl, 0.2 mM EDTA, 5 mM THP, 5% glycerol-d$_5$ and 5% D$_2$O was used as a buffer.

3.5.7 NMR experiments on ^{15}N-labeled NS5A-D2

NMR experiments containing ^{15}N-labeled NS5A-D2 were performed at 298K on a Bruker 600 MHz spectrometer equipped with a cryoprobe. Assignment of the NS5A-D2 ^{15}N,^{1}H-HSQC has been previously reported by our group and had been deposited in the Biological Magnetic Resonance Data Bank (BMRB) under the accession numbers 16165 (JFH-1) and 19055 (Con1).

Experiments with NS5B

^{15}N,^{1}H-HSQC spectra of 20 μM ^{15}N-labeled NS5A-D2 JFH-1 were acquired in the absence of NS5B, with 100 μM unlabeled NS5B$_{\Delta 21}$ and with 100 μM unlabeled filibuvir-saturated NS5B$_{\Delta 21}$. The samples were in 300 mM sodium phosphate pH 6.8, 50 mM NaCl, 4 mM THP, 2 mM $MgCl_2$ The spectra were acquired with 2048 and 256 complex points, 14 and 25 ppm in the direct and indirect dimensions, respectively and 192 scans.

Experiments with ssRNA

^{15}N,^{1}H-HSQC spectra were acquired of two independent samples of 50 μM ^{15}N-labeled NS5A-D2 JFH-1 in the absence and presence of an equimolar amount of RNA16 in 50 mM Tris-d_{11} pH 6.4, 30 mM NaCl, 0.5 mM EDTA, 1 mM THP 2 mM $MgCl_2$ in 5 % D_2O. The spectra were acquired with the same parameters as the experiments with NS5B.

Experiments with GCK

^{15}N,^{1}H-HSQC spectra were acquired of independent samples of 87 μM ^{15}N-labeled NS5A-D2 JFH-1 in the absence and presence 87 and 174 μM unlabeled GCK in 50 mM sodium phosphate pH 6.4, 150 mM NaCl, 0.2 mM EDTA, 1 mM THP 2 mM $MgCl_2$ in 5 % D_2O. The spectra were acquired with 2048 and 256 complex points, 14 and 25 ppm in the direct and indirect dimensions, respectively and 48 scans. To the sample containing 87 μM ^{15}N-labeled NS5A-D2 JFH-1 and 174 μM unlabeled GCK 100 mM of D-glucose was added and another spectrum was acquired in the same conditions.

A similar titration was done using 83 μM ^{15}N-labeled NS5A-D2 from the Con1 isolate in the absence and presence of 83 μM and 166 μM GCK in 50 mM sodium phosphate pH 6.8, 150 mM NaCl, 0.2 mM EDTA, 1 mM THP 2 mM $MgCl_2$ in 5 % D_2O. The spectra were acquired in the same conditions and 100 mM D-glucose was added to the sample with 166 μM of GCK.

3.5.8 Chemical shift predictions

Chemical shift predictions for NS5B were obtained from crystallographic data using ShiftX2 [392] to determine isoleucine δ1 carbon and hydrogen chemical shifts for the deuterated protein at pH 7 and 305 K. Predictions were obtained from both closed (PDB: 2XXD, 2XYM [393], 3I5K [134], 4AEP, 4AEX [161]) and open (PDB ID 4E76 [145], 4OBC [394], 4WT9 [148]) apo structures of the JFH-1 strain. When multiple NS5B chains were present in a deposited structure only the first chain was used in the prediction. Normalized deviations for each isoleucine cross peak were calculated for ^{13}C and ^{1}H dimensions and combined using Equation 3.2,

$$\sigma = \sigma_H + \alpha \times \sigma_C \tag{3.2}$$

where σ_C and σ_H are the standard deviations taken across the CS predictions, and α is the carbon scaling factor used in the calculation of chemical shift perturbations with a value of 0.16.

3.5.9 CPMG experiments

Data acquisition and treatment of Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiments were kindly done by Dr. Robert Schneider. The methyl-TROSY ^{13}C-^{1}H multiple-quantum CPMG relaxation dispersion experiments relaxation dispe [359] were recorded on samples of about 80 μM $NS5B_{\Delta 21}$ in phosphate buffer (300 mM sodium phosphate pH 6.8, 50 mM NaCl, 2 mM $MgCl_2$, 1 mM EDTA) with or without filibuvir on both 600 and 900 MHz spectrometers. 10 or 12 points were acquired for each dispersion curve with CPMG frequencies (ν_{CPMG}) between 100 and 1000 Hz, with 1 or 2 repeat experiments for error analysis. Spectra were acquired with acquisition

times of 82 and 32 ms; 2048 and 128 real points in ^{1}H and ^{13}C dimensions, respectively at 900 MHz. At 600 MHz, spectra were acquired with acquisition times of 76 and 42 ms; 2048 and 112 real points in ^{1}H and ^{13}C dimensions, respectively. Spectra were processed with NMRPipe [395] and analyzed with Sparky [390]. Peak heights were used to calculate relaxation dispersion profiles, $R_{2,\text{eff}}(\nu_{\text{CPMG}})$ using Equation 3.3.

$$R_{2,\text{eff}}(\nu_{\text{CPMG}}) = -\frac{1}{T} ln \left(\frac{I(\nu_{\text{CPMG}})}{I_0} \right) \tag{3.3}$$

where $I(\nu_{\text{CPMG}})$ and I_0 are the peak heights obtained with CPMG frequency ν_{CPMG} and in a reference spectrum recorded without CPMG pulse train. Errors in $R_{2,\text{eff}}$ were calculated based on repeat measurements as described by Ishima and Torchia [396] and a minimum error of 0.5 s^{-1} was assumed. The relaxation dispersion curves were fit to both a two-site exchange model [359] using the software cpmg_fitd9 [397] kindly provided by Dimitry Korzhnev and to a model assuming no exchange (i.e. a flat horizontal line). Non-flat dispersion curves were defined as those with at least a 2 s^{-1} difference between $R_{2,\text{eff}}$ measured at the minimum and maximum ν_{CPMG} frequencies and where the two-site exchange model improved the fit, with a 98% confidence level according to F test statistics. Individual residues were fit either allowing both ^{13}C and ^{1}H chemical shift differences between states ($\Delta\omega$) to vary or only ^{13}C with ^{1}H $\Delta\omega$ fixed to zero. The ^{1}H $\Delta\omega$ fit parameter did not statistically significantly improve the fits. However, for some residues its inclusion decreased errors of fit substantially. For these residues, both ^{13}C and ^{1}H $\Delta\omega$ variables are given; otherwise only ^{13}C $\Delta\omega$ is reported. Global fits, grouping all non-flat relaxation dispersion curves, were performed allowing both ^{13}C and ^{1}H $\Delta\omega$ to vary for each residue. χ^2, the chi-squared goodness of fit measure was calculated as in Equation 3.4.

$$\chi^2 = \sum\nolimits_i \frac{\left[{R_{2,\text{eff}}}^{exp}(i) - {R_{2,\text{eff}}}^{calc}(i) \right]^2}{\sigma_i^2} \tag{3.4}$$

where ${R_{2,\text{eff}}}^{exp}$ and ${R_{2,\text{eff}}}^{exp}$ are experimental and back-calculated $R_{2,\text{eff}}$ values from the dispersion curves, σ is the experimental error and i runs over all data points fit together. The reduced χ^2 is the χ^2 value divided by the number of degrees of freedom of the fit (number of data points minus the number of fit parameters).

3.6 Fluorescence spectroscopy

The binding of $NS5B_{\Delta 21}$ to an RNA template strand was assayed by fluorescence spectroscopy at 294 K on a PTI fluorescence spectrometer (PTI Monmouth Junction, NJ, USA). The 16-mer RNA single strand (5'-CUAAGAUGCUCGCUGC-3') fused to a 6-carboxyfluorescein (6-FAM) label on the 5' end was obtained from GeneCust (Luxembourg). The excitation of the fluorescein label was done at 492 nm and emission spectrum was scanned from 500 to 530 nm (1 nm step-wise, 2 seconds per scan). Fluorescence intensities were integrated from 515 to 525 nm. The excitation and the emission slit widths were set to 1 nm. This assay was carried out in RE buffer (20 mM TrisCl buffer pH 7.5, 100 mM NaCl, 80 mM arginine, 80 mM glutamate, 5 mM $MgCl_2$,1 mM DTT) in 1 cm length cells. The fluorescence of 50 nM 6-FAM-labeled RNA was detected in the presence of 0-3.5 μM of NS5B. The assay was repeated in the presence of 0, 1.33, 4 and 20 μM of NS5A-D2.

3.7 Size-exclusion chromatography

Size-exclusion chromatography was performed on an ÄKTA Purifier FPLC system (GE Healthcare, Little Chalfont, UK) with a UV detector. Chromatographic separations were done with a Superdex 75 3.2/300 column (GE Healthcare, Little Chalfont, UK) at room temperature. The mobile phase buffer (20 mM Tris pH 7.0, 100 mM arginine, 100 mM glutamate, 100 mM NaCl, 5 mM $MgCl_2$, 2mM THP) was passed through the column at a flow rate of 20 μL/min. 40 μL samples containing 10 μM of 6-FAM-labeled RNA template (as used for fluorescence spectroscopy) were injected and detected at 260, 280 and 495 nm. Samples with RNA and 0, 3, 7, 10, 15 μM of $NS5B_{\Delta 21}$ or RNA and 0, 5, 15, 50 μM of NS5A-D2 were injected. To study the effect of NS5A-D2 on RNA binding to NS5B RNA, 7 μM $NS5B_{\Delta 21}$ and 0, 5, 15, 50 μM of NS5A-D2 were also injected. The proteins were exchanged to the same buffer as the mobile phase prior to the sample preparation by Zeba Spin desalting columns (Thermo Fisher). The effect of Filibuvir on RNA-binding was observed with an $NS5B_{\Delta 21}$ sample incubated with solid filibuvir prior to injection at the same concentrations as the RNA and apo $NS5B_{\Delta 21}$ samples described above. Samples containing only 10 μM of either $NS5B_{\Delta 21}$,

Filibuvir-loaded NS5B$_{\Delta 21}$ or NS5A-D2 were used to determine the elution time for these proteins.

3.8 Differential scanning fluorimetry

In preparation for differential scanning fluorimetry (DSF) the proteins had their buffers exchanged to RE buffer (20 mM Tris, 100 mM NaCl, 100 mM arginine, 100 mM glutamate, 5 mM $MgCl_2$ and 2 mM THP) using Zeba spin desalting columns. The assay performed in the presence of Filibuvir was done with NS5B loaded and incubated overnight with the inhibitor. DSF was carried out on a Stratagene Mx3005P QPCR system (Agilent Technologies, Santa Clara, CA, USA) with a 40 μL reaction volume containing RE buffer and 1:1000 diluted Sypro orange (ThermoFisher) fluorophore. For the assay, NS5B concentration was kept constant at 30 μM. Increasing concentrations of NS5A-D2 (0 – 180 μM) were added to NS5B to determine the effect of the interaction. A control containing 180 μM of NS5A-D2 without NS5B was done to confirm that NS5A-D2 did not present a residual melting temperature. Thermal denaturation was carried out while collecting fluorescence intensity data (excitation at 492 nm and emission at 516 nm) during an increase in temperature from 298 K to 368 K with a rate of 1 K/min. The inflection point of the melting curves, which represents the melting temperature (T_m), was determined by fitting the obtained intensities, up to the plateau, to a Boltzmann model (Equation 3.5),

$$F(T) = F_{base} + \frac{F_{max} - F_{base}}{1 + exp\left(\frac{T_m - T}{a}\right)} \tag{3.5}$$

where F_{max} and F_{base} are the maximum and minimum recorded fluorescence intensities, respectively, and a is the slope of the curve.

Characterization of NS5B by nuclear magnetic resonance spectroscopy

4.1 Introduction

The HCV RNA polymerase, NS5B, is a 66 kDa protein with a C-terminal 21 amino-acid transmembrane helix. Expression of NS5B truncated of its transmembrane helix results in a protein with improved solubility [398]. Cleavage of these 21 residues and preceding 39 amino acids in the linker gives rise to a protein that is catalytically active *in vitro* [155] and can be used for structural and functional studies.

HCV has a large genetic variability with over 90000 genomes deposited in the European hepatitis C virus database [388]. We chose to work with NS5B from the JFH-1 isolate (genotype 2a) because, it was the first isolate for which the entire life cycle of the virus could be reproduced in cell culture [177–179]. Consequently, there is an abundance of structural and functional information available [134, 393, 399] for this isolate.

In this work we used two different constructs of NS5B that varied by amino acid length. One had the C-terminal 21 amino acids (the transmembrane helix) deleted and in the other the C-terminal 60 amino acids (both the transmembrane helix and the linker) were deleted. These were called NS5B$_{\Delta 21}$ and NS5B$_{\Delta 60}$, respectively for clarity. However, due to the better expression, higher stability, higher signal-to-noise ratio in the NMR spectra obtained for the NS5B$_{\Delta 21}$ construct, most of this work was done with this construct.

4.2 Expression and purification of NS5B

Recombinant NS5B from the JFH-1 isolate was expressed in the BL21(DE3) strain of *Escherichia coli* using a pET-derived expression vector [134]. Cell growth and protein expression was done in either Luria-Bertani (LB) rich culture medium for unlabeled samples or supplemented M9 minimal medium for various labeling strategies. D_2O and D-glucose-d_7 in the culture medium were used to perdeuterate the recombinant protein. Similarly, ^{15}N-labeled ammonium chloride and D-glucose-$^{13}C_6$-d_7 were

used to achieve near complete labeling of the nitrogens and carbons respectively when necessary. To selectively label the δ1 methyl groups on the isoleucine side-chains 2-ketobutyric acid-4-^{13}C-3,3-^{2}H was added prior to induction. 2-ketobutyric acid-4-^{13}C-3,3-^{2}H is produced by incubating 2-ketobutyric acid-4-^{13}C for 12 hours at 45°C at pH 10 in D_2O to take advantage of the high acidity of the β-hydrogens to exchange them for deuterium. Subsequent purification of the protein is done by nickel-affinity chromatography and the protein is concentrated to 100 μM or until significant precipitation is observed as it is not stable at higher concentrations. Purification gives rise to the 65 kDa protein shown in Figure 4.1.

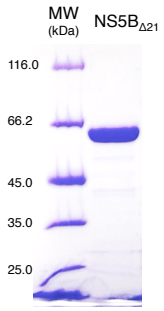

FIGURE 4.1: SDS-PAGE of NS5B$_{\Delta21}$ labeled U-^{2}H Ile δ1-^{13}C^{1}H$_3$ with a 10% acrylamide gel stained with Coomassie blue.

The yields obtained after purification were dependent on labeling strategy and NS5B construct. For wild-type NS5B$_{\Delta21}$ the yield was about 15 mg/liter of culture for unlabeled protein and 8–10 mg/liter of culture for deuterated or selectively labeled on the tryptophan side-chains. For open constructs of NS5B (NS5B$_{\Delta21}$ L30S, NS5B$_{\Delta21}$ 5mut and NS5B$_{\Delta60}$ the yield was about 4–6 mg/liter of culture for unlabeled protein and 3–4 mg/liter of uniformly deuterated and selectively isoleucine-δ1 methyl group ^{13}C^{1}H$_3$-labeled NS5B.

4.3 ^{1}H,^{15}N-Heteronuclear single quantum coherence spectrum of NS5B$_{\Delta21}$

A nuclear magnetic resonance (NMR) experiment frequently used for proteins is a ^{1}H,^{15}N-heteronuclear single quantum coherence (HSQC) experiment. The protein is

enriched in ^{15}N to obtain signal. This experiment results in a two-dimensional spectrum with one axis for ^{1}H and the other for ^{15}N. The resonances in this spectrum represent each unique proton attached to a nitrogen. By focusing on the region between 7 and 10 ppm we observe mainly amide protons, of which the majority are from the protein backbone N-H bonds. Therefore, in the absence of slow exchange (which can give more than one peak per resonance), one peak in this spectrum represents one residue except proline. Some side chain N-H groups like that of tryptophan, asparagine and glutamine can also give signals in the same region of the spectrum. Prolines are secondary amines, consequently they do not have amide protons when they form peptide bonds and thus are invisible in this experiment.

The polymerase of the hepatitis C virus, NS5B, is a large globular protein. The construct we are working with has a molecular weight of 65 kDa in its form cleaved of the 21 amino acids from the transmembrane helix. Protein NMR presents challenges beyond 25–30 kDa [347]. $NS5B_{\Delta 21}$ also presents other characteristics that make NMR experiments with it challenging. Beyond about 90 μM, the polymerase precipitates which limits the concentration that can be used in the study. Additionally, the polymerase needs high salinity to be stable at this concentration. This results in lower sensitivity compared to a lower salinity buffer, especially when using a cryoprobe [400]. In most experiments a buffer composed of 300 mM sodium phosphate pH 6.8, 50 mM NaCl, 2 mM $MgCl_2$, 1 mM EDTA and 4 mM THP (called phosphate buffer) was used. $MgCl_2$ was added as divalent magnesium ions are needed for polymerase activity.

Several strategies are used to improve the quality of the spectrum obtained of $NS5B_{\Delta 21}$. Firstly, most experiments were carried out using a 900 MHz spectrometer equipped with a cryoprobe. High-field spectrometers provide better resolution as the signals have a larger frequency difference between them than they would have with a spectrometer of lower field. Secondly, TROSY-type experiments [343] are used to reduce peak linewidth. The spectrometer frequency also improves the TROSY effect. Thirdly, perdeuteration reduces transverse relaxation due to dipolar coupling to protons.

$NS5B_{\Delta 21}$ was expressed in *E. coli* in a culture medium that contained deuterated D-glucose and D_2O and $[^{15}N]NH_4Cl$ to obtain perdeuterated and ^{15}N-labeled polymerase. In this way, most of the protons responsible for relaxation are exchanged to

deuterium. The amide protons are labile and are exchanged to ^{1}H by doing the purification of protein and NMR experiment in protonated water.

The ^{1}H,^{15}N-TROSY-HSQC spectrum of NS5B$_{\Delta 21}$ is shown in Figure 4.2. This TROSY-HSQC spectrum is well dispersed, as expected of a well-folded protein. The peaks have a large distribution of intensities (Figure 4.3) which can suggest either that there is dynamics in a timescale close to intermediate exchange or that there are multiple slowly exchanging conformations in solution that have variable relative populations.

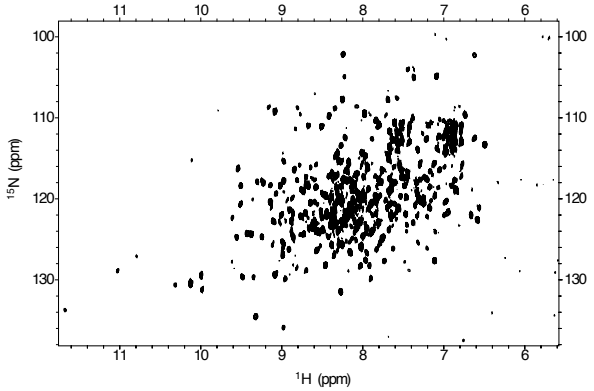

FIGURE 4.2: ^{15}N,^{1}H-HSQC NMR spectrum of NS5B$_{\Delta 21}$. This shows a well dispersed spectrum characteristic of a well-folded protein. However, only 318 of the expected 544 peaks (this protein contains 35 prolines) are present.

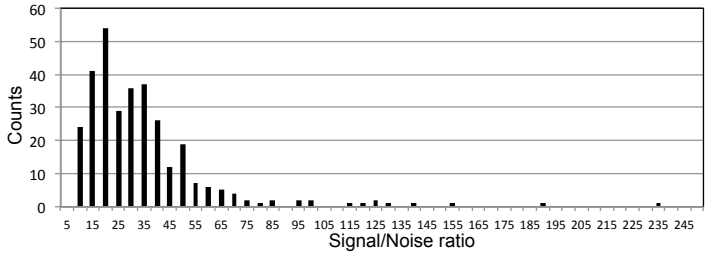

FIGURE 4.3: Histogram of signal-to-noise ratio (S/N) of peaks in ^{15}N,^{1}H-TROSY-HSQC spectrum of NS5B$_{\Delta 21}$. The dispersion in this histogram suggests dynamic processes. The highest S/N is due to the hexahistidine tag.

4.3.1 Assignment

Assignment of the ^{1}H,^{15}N-TROSY-HSQC was attempted by a triple-resonance assignment strategy. HNCA, HNCACB, HN(COCA)CB, HNCO and HN(CA)CO TROSY 3D spectra [344, 391, 401, 402] were acquired on ^{15}N,^{13}C,^{2}H-labeled samples. To improve global signal-to-noise (S/N) in the 3D spectra we used Band-selective Excitation Short

Transient (BEST)-type pulse sequences [402]. These pulse sequences make use of selective pulses, that is, pulses that only excite a subset of the proton spins in the protein. The unexcited protons act as a reservoir of protons to which the excited protons can transfer their spin magnetization, thus reducing the T_1 relaxation time. This allows more scans in the same period of time and therefore obtain an increased S/N per unit time in the center of the spectrum. The S/N ratios are smaller towards the edge of our spectrum as we have a large dispersion of frequencies and the selective pulses irradiate more efficiently at the frequencies at center of the spectra. However, due to the low S/N ratio observed for many peaks in the TROSY-HSQC and 3D spectra we often lack the peaks required to follow the connectivity between spin systems (C_α i-1, for example). Only 318 peaks in the $^{1}H,^{15}N$-TROSY-HSQC spectrum present peaks in one of the 3D spectra (which is necessary for the triple-resonance assignment strategy). There are 544 non-proline residues in our $NS5B_{\Delta 21}$ construct, thus a large fraction of these resonances are not present in the spectrum.

Due to the use of BEST, the information on the carbon resonances (C_α, C_β, C') towards the edge of the spectra was much more limited than for peaks at the center of the spectrum. However, in comparison, we attempted to acquire 3D spectra without BEST but the global reduced S/N per unit time meant these peaks were not visible with these pulse sequences either.

Despite the long acquisition times and many triple resonance spectra used for the assignment, we were not able to assign the $^{1}H,^{15}N$-HSQC spectrum with this strategy.

4.4 Selective isoleucine-δ1 labeling

Due to the challenges of assigning the $^{1}H,^{15}N$-HSQC spectrum, we used another NMR strategy developped for large proteins. Selective labeling of methyl side chains is a strategy widely used in the NMR study of high molecular weight proteins [361]. This can be done for several methyl-containing amino acids, independently or combined: alanine, isoleucine, leucine, methionine, threonine and valine. Methyl groups are often chosen to be labeled because they appear in hydrophobic cores and at biomolecular interfaces and bacterial metabolic pathways can be used to label these side-chain

groups. Furthermore, three ^{1}H contribute to the intensity of the same methyl signal. Their fast rotation give signals with narrow linewidths even in the context of slowly-tumbling large proteins. Additionally, methyls are intrinsically optimized for TROSY experiments by using the heteronuclear multiple quantum coherence (HMQC) spectrum instead of the HSQC [360]. The main challenge with selective labeling is that the sequential connectivity used to assign HSQC spectra is lost and different assignment strategies must be applied.

For NS5B from the JFH-1 isolate cleaved of its transmembrane helix ($NS5B_{\Delta 21}$) we chose to label isoleucine side chains as opposed to other methyl groups because isoleucine 405 has been shown to be a key residue in *de novo* RNA synthesis. This residue is a valine in most HCV isolates. By mutating the residue at position 405, Scrima *et al.* showed that this position has a role in the formation of the first dinucleotide and the transition between initiation and elongation competent conformations [161]. Additionally, the isoleucines of $NS5B_{\Delta 21}$ are well distributed throughout the three subdomains and I560 in the linker. This protein contains 24 natural isoleucines (Table 4.1).

To obtain the samples of NS5B selectively on the isoleucine side-chain methyl groups, a selectively labeled precursor, 2-ketobutyric acid-4-^{13}C-3,3-^{2}H, was added to the deuterated culture medium. The bacterial metabolism transforms this precursor into selectively labeled isoleucine that is incorporated into the expressed protein without scrambling [353]. The resulting perdeuterated protein should give a ^{1}H-^{13}C-TROSY HMQC spectrum (called methyl-HMQC spectrum for clarity) [357] with one resonance per isoleucine $\delta 1$ methyl group at about 9 – 16 ppm in the ^{13}C dimension and 0 – 1.5 ppm in the ^{1}H dimension.

The methyl-HMQC spectrum obtained on uniformly ^{2}H-labeled and selectively isoleucine $\delta 1$ $^{13}C^{1}H_3$-labeled $NS5B_{\Delta 21}$ wild-type shows 20 resonances (Figure 4.4). The lack of 4 peaks (to complete the expected 24) suggests that these residues give resonances that are broadened beyond detection by millisecond to microsecond dynamics, invisible to NMR spectroscopy.

Isoleucine	Location
11 23	Fingertips loop Λ1
71 119 134 138 160 178 233 239 253 293	Fingers
308 323 363	Palm
405 412 413 419 432 462 463 512	Thumb
560	Linker

TABLE 4.1: Location of the 24 isoleucines on NS5B.

4.4.1 Assignment

To assign the resonances in the methyl-HMQC spectrum of wild-type NS5B$_{\Delta 21}$ we used a combination of single-site substitutions with valine, NOE correlations and chemical shift predictions.

A total of 14 isoleucine to valine mutations were done by site-directed mutagenesis to assign the isoleucine resonances in the methyl-HMQC spectrum (Figures 4.5 and 4.6). In the spectrum of the mutant, the resonance corresponding to the mutated isoleucine will disappear. Comparing the spectrum of the mutant with that of wild-type allows for the identification of the resonance that corresponds to each isoleucine in the methyl-HMQC spectrum of wild-type NS5B$_{\Delta 21}$.

The I-to-V mutants allowed us to ascertain that I405, I412, I413 and I560 are the isoleucines whose resonances are broadened beyond detection as no resonances disappeared from these spectra. Using these mutations, isoleucines 23, 119, 160, 233, 363,

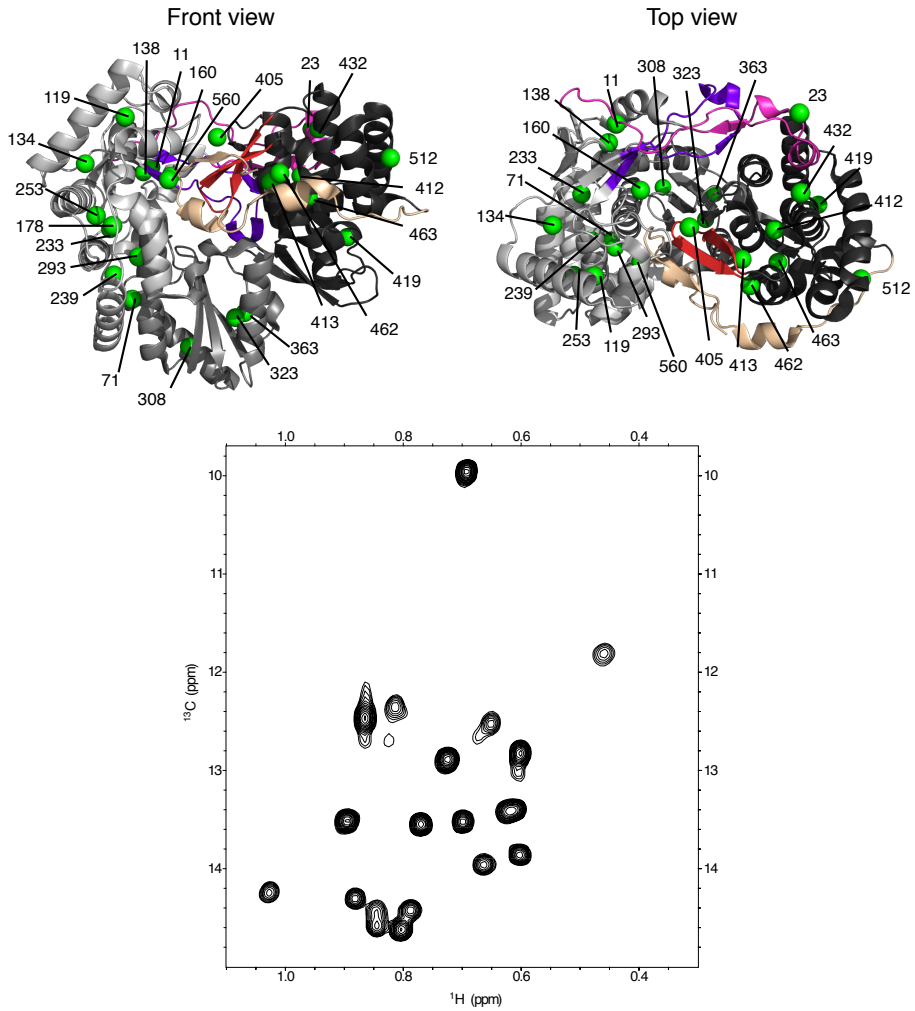

FIGURE 4.4: Selective isoleucine-δ1 labeling of NS5B$_{\Delta 21}$. **Top** Crystallographic structure of NS5B (PDB 3I5K [134]) with the isoleucine δ1 carbons shown as green spheres. The fingers, palm and thumb are shown in light, medium and dark gray, respectively. The structural features called fingertips Λ1 and Λ2 loops, β-loop and linker are highlighted in magenta, purple, red and tan, respectively. **Bottom** Methyl-HMQC spectrum of 97 μM U-^{2}H, Ile δ1-[^{13}C^{1}H$_3$]-NS5B$_{\Delta 21}$ wild-type in phosphate buffer (300 mM sodium phosphate pH 6.8, 50 mM NaCl, 2mM $MgCl_2$, 0.5 mM EDTA, 4 mM THP in D_2O). Obtained with 16 scans.

419, 432, 462, 463 and 512 were assigned to resonances in the spectrum.

Nuclear Overhauser effect spectroscopy (NOESY) spectra were used to assign peaks that are close in space. Two types of 3D HMQC-NOESY spectra were used. The ^{1}H,^{13}C-NOESY-HMQC shows the NOE cross-peaks in a ^{1}H-^{1}H plane while the ^{1}H,^{13}C-HMQC-NOESY-HMQC represents the NOE cross-peaks in a ^{13}C-^{13}C plane. These approaches are complementary because of the overlap of peaks in the carbon and hydrogen dimensions if a single nucleus was used.

NOE cross-peaks between isoleucine resonance pairs 11/138, 178/253, and 323/363

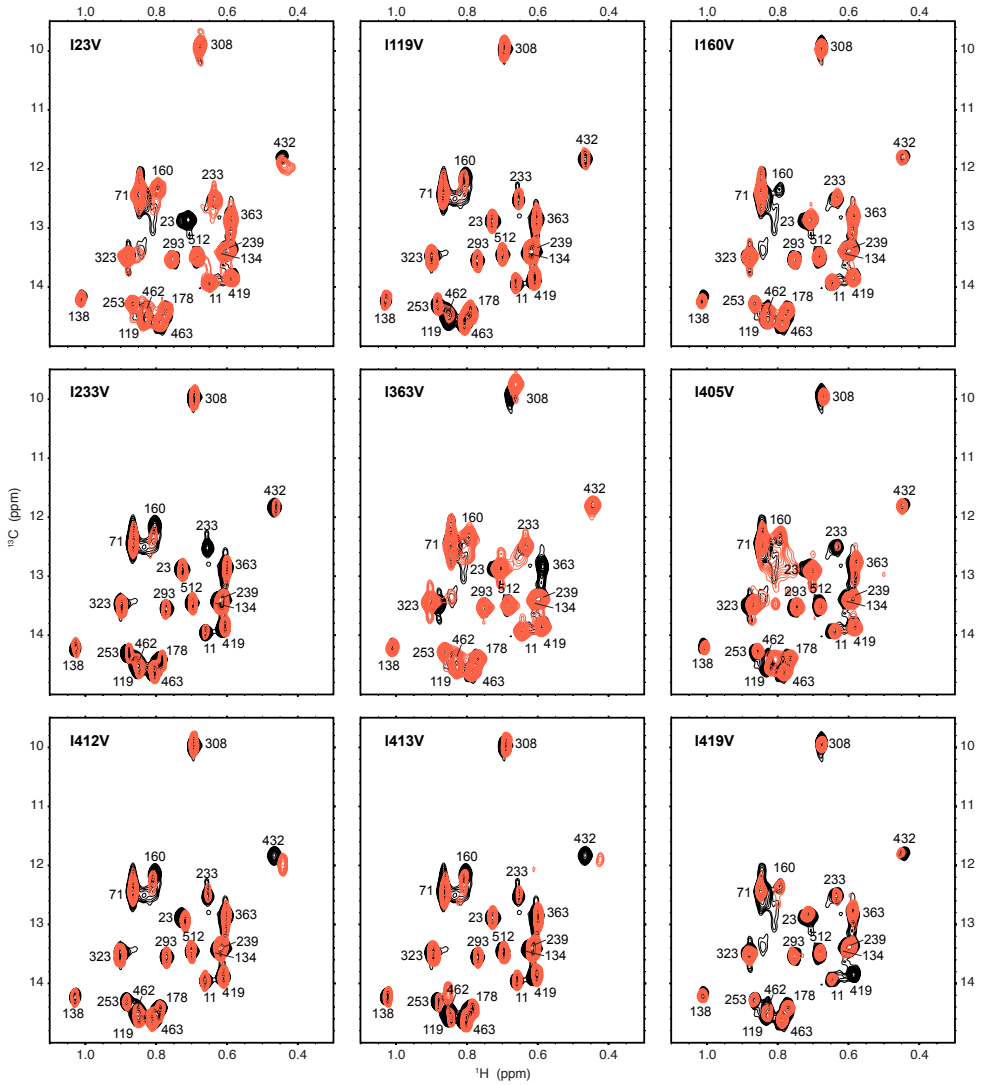

FIGURE 4.5: Isoleucine-to-valine mutants used for assignment (part 1). The disappearance of a resonance in the isoleucine-to-valine mutant methyl-HMQC spectrum (red) compared to wild-type (black) permits the assignment of the corresponding mutated isoleucine residue. I405V, I412V and I413V shown above do not give rise to observable disappearing signals meaning they are broadened beyond detection in the wild-type spectrum.

and between isoleucines 71, 239 and 293 allowed the identification of these residues (Figures 4.7 and 4.8 and Table 4.2). This identification is ambiguous for all of these except 323/363 because neither of the resonances from the other pairs has been assigned by mutation. We used the chemical shift prediction software ShiftX2 [392] to resolve the ambiguity (Tables 4.5 and 4.6, in the next section). This was done by comparing the relative upfield and downfield position of the resonances and the predictions. The

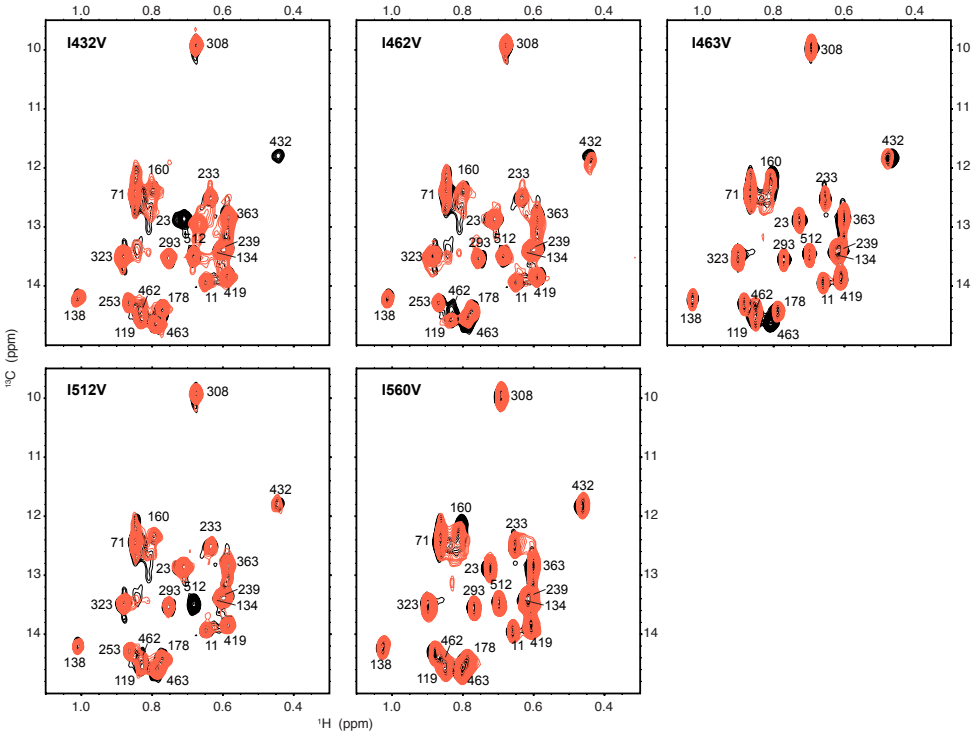

FIGURE 4.6: Isoleucine-to-valine mutants used for assignment (part 2). The disappearance of a resonance in the isoleucine-to-valine mutant methyl-HMQC spectrum (red) compared to wild-type (black) permits the assignment of the corresponding mutated residue. I560V shown above does not give rise to an observable disappearing signal meaning this isoleucine is broadened beyond detection in the wild-type spectrum.

remaining isoleucines, 134 and 308, were also identified by their very different chemical shifts as calculated by chemical shift prediction. The combination of isoleucine-to-valine point mutations, NOE cross-peaks and chemical shift predictions allowed the assignment of the 20 resonances in the methyl-HMQC spectrum of NS5B$_{\Delta 21}$ wild-type (Figure 4.9). This assignment has been deposited in the BMRB under accession number 27073.

The isoleucines that present resonances broadened beyond detection are located around the RNA-binding groove. Ile 405, 412 and 413 are at the top of the thumb subdomain (using the orientation shown in Figure 4.4, left) and Ile 560 within the part of the linker that folds back into the RNA-binding groove in the crystallographic structures of apo NS5B. Their broadening suggests chemical exchange in an intermediate timescale. Thus, it is possible that they are sensitive to exchange between open and closed conformations of NS5B$_{\Delta 21}$, which is in agreement with their orientation towards the central groove.

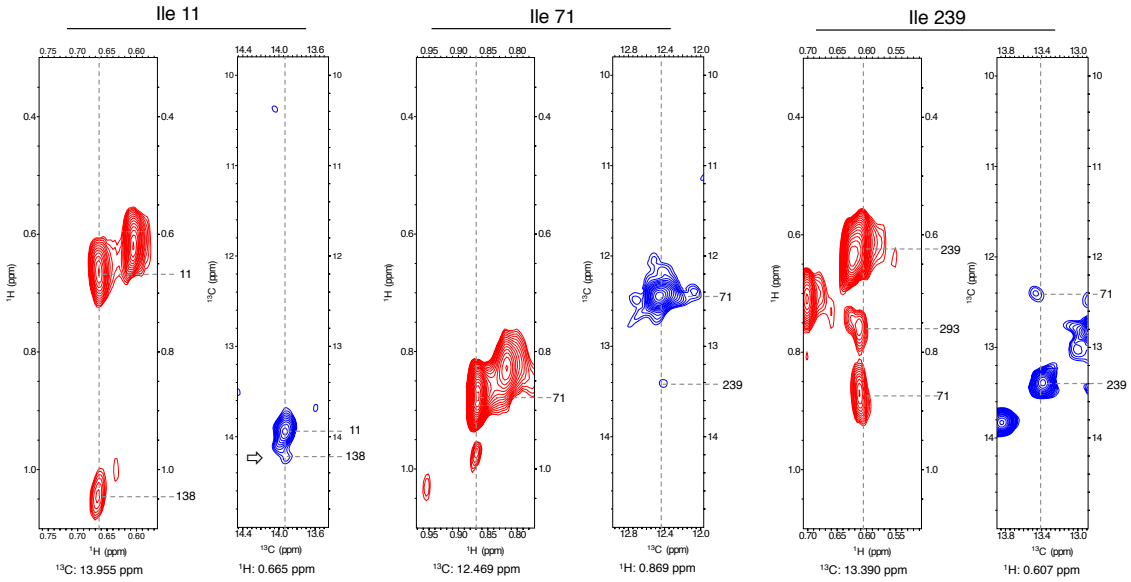

FIGURE 4.7: Strips extracted from the 3D NOESY spectra used to identify resonances whose residues are close in space (Part 1). Both 3D ^{1}H,^{13}C-NOESY-HMQC (red) and ^{1}H,^{13}C-HMQC-NOESY-HMQC (blue) strips are shown for each resonance. NOESY strips shown were used to assign resonances to I11 and I138 (left), I71 and I239 (center and right).

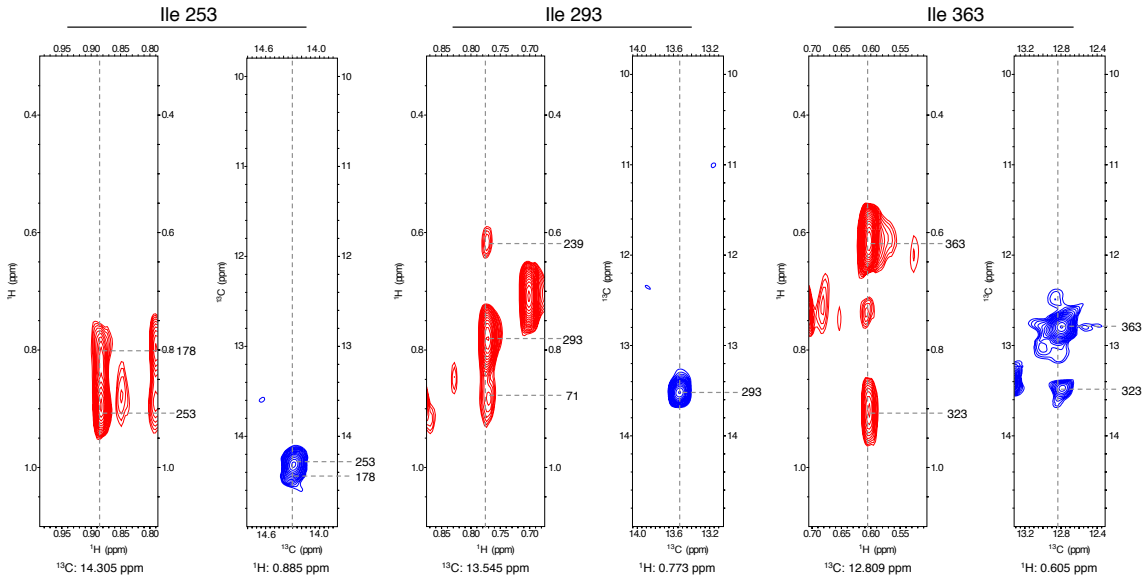

FIGURE 4.8: Strips extracted from the 3D NOESY spectra used to identify resonances whose residues are close in space (Part 2). 3D ^{1}H,^{13}C-NOESY-HMQC (red) and ^{1}H,^{13}C-HMQC-NOESY-HMQC (blue) strips are shown for each resonance. NOESY strips shown were used to assign resonances to I178 and I253 (left), I293 (center), and I323 and I363 (right).

NOE	Distance (Å)
178/253	3.579
11/138	4.153
323/363	4.518
413/462	4.739
239/71	5.442
239/293	5.519
413/463	7.197
412/463	7.303
462/463	7.468
71/293	7.601

TABLE 4.2: Shortest distances between isoleucine $\delta 1$ carbons in the crystallographic structure of $NS5B_{\Delta 21}$ JFH-1 (PDB 2XXD). These are in agreement with the NOE cross-peaks identified in the NOESY spectra. NOE cross-peaks written in gray are not observed because one of the residues is broadened beyond detection in the methyl-HMQC spectrum.

Mutation	CSP	Height increase	Height decrease
I23V	432	160	
I119V	419	138	
I160V	138, 462		
I233V	160	119, 512	
I363V	308, 323		462
I405V	462, 363		134, 178, 233, 253, 308, 512
I412V	432, 23, 160	432	
I413V	160, 432, 462, 419	462	
I419V	363		119
I432V	23		
I462V	432, 463		119
I463V	160, 462, 432		
I512V			
I560V	160, 419		

TABLE 4.3: Residues for which their NMR resonances are affected by an I-to-V mutation. ^{1}H and ^{13}C combined chemical shift perturbations (CSP) higher than 0.015 ppm are shown in black and 0.010-0.015 ppm in gray. Variations in normalized data height are also shown (height increase: > 150% of wild-type, height decrease: < 75% of wild-type.

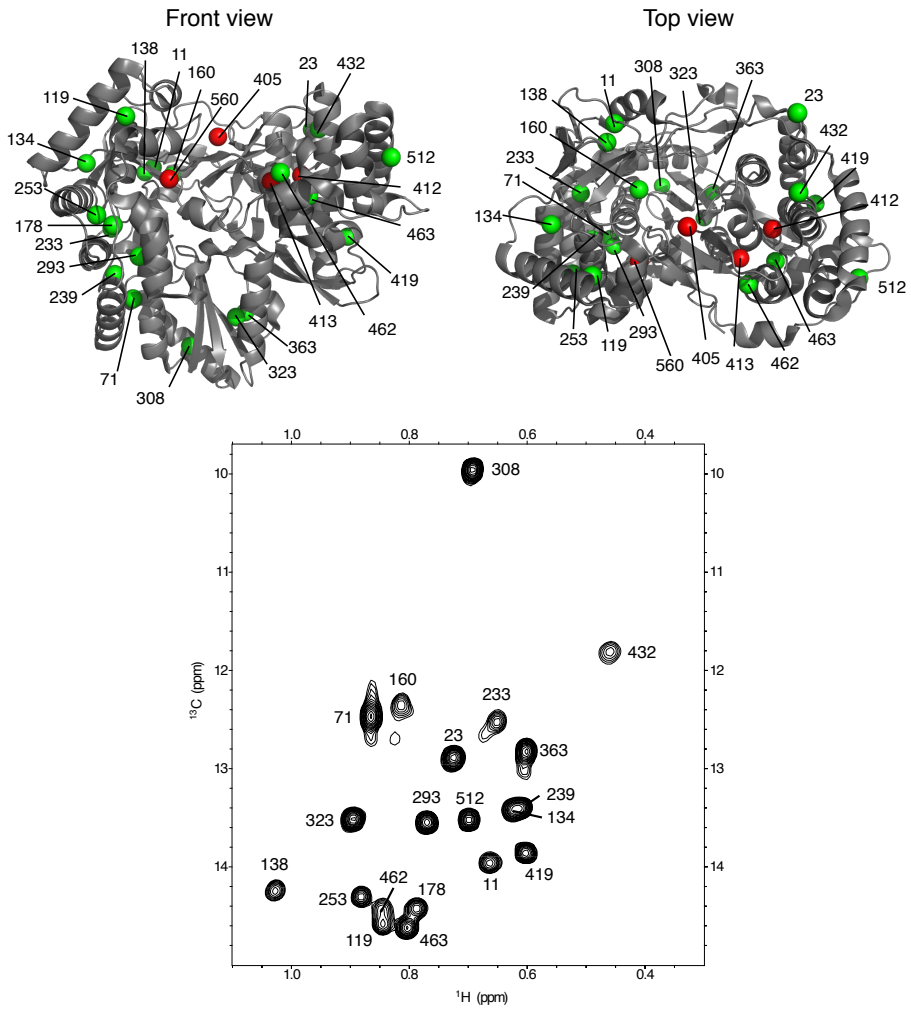

FIGURE 4.9: Assignment of the methyl-HMQC spectrum of NS5B$_{\Delta 21}$. **Top** Crystallographic structure of NS5B (PDB 3I5K) with the isoleucine $\delta 1$ carbons shown as spheres, green for those visible in the spectrum and red for the isoleucines broadened beyond detection. **Bottom** Methyl-HMQC spectrum of NS5B$_{\Delta 21}$ showing the assignment of the 20 resonances.

Interestingly, the I-to-V mutants give us insight into the long-range allosteric networks within the protein (Table 4.3). For example, by mutating I363 we cause not only a combined chemical shift perturbation (CSP) on the resonance of I323 that is 4.5 Å from the mutation but also on the resonance of I308 which is farther away (12.3 Å) but also within the palm subdomain. Usually, the perturbations of isoleucine resonances stay within the same subdomain. However, it also highlights the association of the fingertips $\Lambda 1$ loop to the thumb through the effect I23V has on the I432 resonance and, conversely, I432V has on the I23 resonance. Furthermore, the mutation of I413, an isoleucine for which no peak was detected, causes NMR signal perturbations in the resonances corresponding to isoleucines 462, 432, 419 and 160 that are 4.7, 18.1, 20.3

and 22.2 Å away. In sum, this points to the subdomains of $NS5B_{\Delta 21}$ being connected via a network that we can probe using the methyl-HMQC spectra.

4.5 Chemical shift predictions and measures of conformation

The JFH-1 isolate of HCV was chosen as the object of this study despite not being representative of most infections because of the abundance of structural and functional data that exists on the polymerase of this isolate. As of May 2017, 185 structures of NS5B have been deposited in the protein data bank (PDB) [140] and of these, 24 are of the JFH-1 isolate. In particular, structures that show NS5B in complex with double-stranded RNA, believed to represent the elongation complex, have only been obtained with mutants of this isolate [145, 148].

We looked at all structures of NS5B JFH-1 that were not bound to a ligand in the PDB. These were classified as either closed or open using the structural metrics defined by Davis and colleagues [243] (Figure 4.10). In their work they use two structural metrics to define the degree of closure. Firstly, the template channel width is calculated by measuring the distance between the centers of mass of residues 139 and 405. Secondly, the interdomain angle is defined as the angle between the centers of mass of the fingers, palm and thumb subdomains. Values of 20 Å and 70° are the cutoff values between open and closed structures. To be classified as an open structure both criteria must be above this cutoff, conversely closed structures have values below the cutoff for both criteria. Structures that do not satisfy either are classified as intermediate.

The PDB contained 8 apo structures of NS5B JFH-1, 5 of which were classified as closed and 3 were classified as open (Table 4.4). These structures were input into the chemical shift prediction software ShiftX2 [392] to obtain predicted values for the chemical shifts of the isoleucine $\delta 1$ side-chain methyl groups. Only chain A was chosen when multiple chains are present in the asymmetric unit of the crystallographic structure to avoid oversampling of some orientations. Since we are assuming large scale dynamics to be present in solution it was important to obtain chemical shift predictions from crystallographic structures with open and closed conformations.

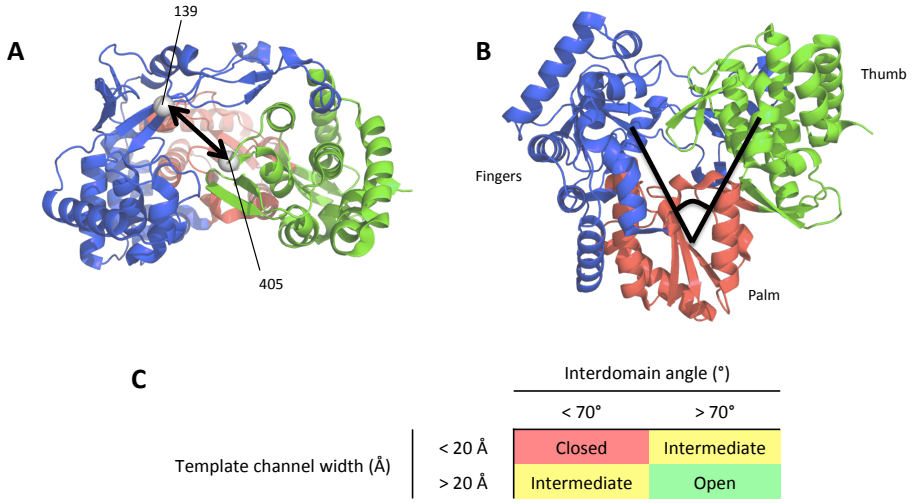

FIGURE 4.10: Structural metrics used to define degree of opening of an NS5B structure. Adapted from Davis *et al.* [243]. **A** The template channel width is defined as the distance between the center of mass of residues 139 and 405. **B** The interdomain angle is the angle formed between the centers of mass of the fingers, palm and thumb subdomains. **C** The structures are classified as open, closed or intermediate depending whether these measures are above or below the cutoff values of 20 Å and 70° for template channel width and interdomain angle, respectively.

PDB structure	Interdomain angles	Template channel width	Classification
2XXD	68.807	16.595	
2XYM	66.844	16.158	
3I5K	68.823	15.872	Closed
4AEP	68.290	16.568	
4AEX	69.193	16.074	
4E76	75.915	26.027	
4OBC	74.311	25.944	Open
4WT9	75.444	26.152	

TABLE 4.4: Crystallographic structures of unliganded NS5B JFH-1 used for chemical shift predictions. Interdomain angles and template channel width are used as structural metrics to define whether these structures are open or closed.

Despite the absolute values of the chemical shift predictions (Tables 4.5 and 4.6) not corresponding directly to the experimental values observed, certain tendencies from the predictions can be used. For example, I308 and I432 have resonances that are shifted upfield compared to the other isoleucine ^{13}C both in the predictions and experimentally. Therefore, we looked at the normalized deviation (standard deviations combined both in proton and carbon dimensions, see Equation 3.2) of the predicted values and found that four of the six with the highest normalized deviations are the residues that are broadened beyond detection (Figure 4.11). In the case of I560, no electron density

Isoleucine	Predicted chemical shift (ppm)		Experimental chemical shift (ppm)
	Closed	Open	
11	0.655 ± 0.055	0.634 ± 0.096	0.664
23	0.746 ± 0.082	0.827 ± 0.021	0.726
71	0.931 ± 0.018	0.936 ± 0.008	0.865
119	0.717 ± 0.040	0.794 ± 0.026	0.846
134	0.734 ± 0.013	0.711 ± 0.007	0.625
138	0.875 ± 0.031	0.817 ± 0.090	1.028
160	0.849 ± 0.104	0.775 ± 0.088	0.815
178	0.690 ± 0.025	0.718 ± 0.011	0.788
233	0.766 ± 0.006	0.771 ± 0.015	0.652
239	0.739 ± 0.017	0.755 ± 0.012	0.610
253	0.803 ± 0.011	0.859 ± 0.019	0.883
293	0.774 ± 0.058	0.766 ± 0.039	0.771
308	0.740 ± 0.015	0.718 ± 0.022	0.693
323	0.841 ± 0.051	0.902 ± 0.026	0.896
363	0.709 ± 0.043	0.732 ± 0.026	0.603
405	0.706 ± 0.032	0.546 ± 0.050	n.d.
412	0.899 ± 0.022	0.267 ± 0.013	n.d.
413	0.440 ± 0.128	0.735 ± 0.058	n.d.
419	0.560 ± 0.052	0.490 ± 0.046	0.603
432	0.658 ± 0.031	0.981 ± 0.013	0.460
462	0.833 ± 0.056	0.885 ± 0.021	0.848
463	0.881 ± 0.027	0.904 ± 0.025	0.804
512	0.771 ± 0.010	0.772 ± 0.027	0.700
560	0.834 ± 0.071	n.a.	n.d.

TABLE 4.5: Prediction of the ^{1}H chemical shifts of NS5B$_{\Delta 21}$. Average chemical shift predictions from ShiftX2 obtained for the δ1 hydrogens of the isoleucine side-chains of NS5B$_{\Delta 21}$ separated by closed and open structures and compared to the experimental chemical shifts. n.a. not available, the open crystallographic structures did not have electronic density for I560. n.d. not determined, resonances in the spectra were not identified for isoleucines 405, 412, 413 and 560.

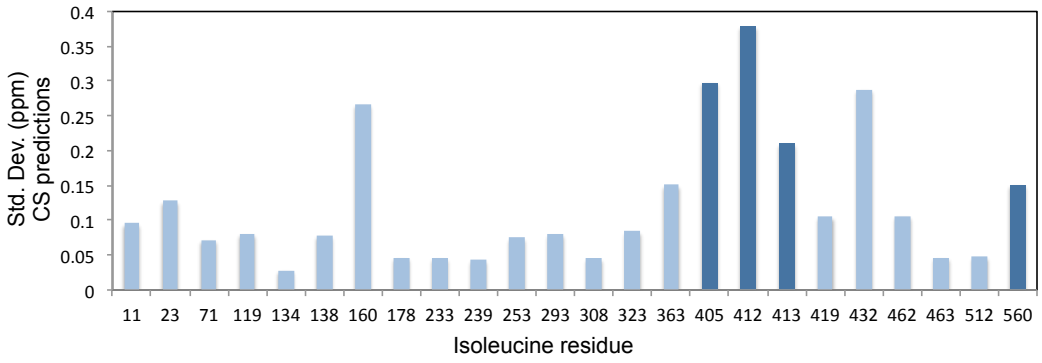

FIGURE 4.11: Standard deviations of chemical shift predictions for δ1-methyl groups from each of the 24 isoleucines. These values are for ^{1}H and ^{13}C combined. Isoleucines in dark blue are those that are broadened beyond detection. There is a correlation between the variation in the predictions and the broadness in the methyl-HMQC spectrum.

Isoleucine	Predicted chemical shift (ppm)		Experimental chemical shift (ppm)
	Closed	Open	
11	14.335 ± 0.202	14.372 ± 0.161	13.961
23	13.714 ± 0.425	13.641 ± 0.082	12.889
71	13.944 ± 0.414	14.223 ± 0.195	12.457
119	14.225 ± 0.181	14.128 ± 0.202	14.590
134	14.040 ± 0.059	13.966 ± 0.091	13.422
138	14.325 ± 0.030	14.288 ± 0.190	14.246
160	13.954 ± 0.457	12.589 ± 1.273	12.350
178	14.229 ± 0.066	14.069 ± 0.182	14.424
233	14.377 ± 0.123	14.063 ± 0.251	12.524
239	14.349 ± 0.117	14.148 ± 0.184	13.398
253	13.941 ± 0.345	13.849 ± 0.066	14.304
293	14.026 ± 0.107	14.360 ± 0.110	13.547
308	12.159 ± 0.174	11.990 ± 0.075	9.958
323	14.495 ± 0.109	14.164 ± 0.138	13.514
363	13.408 ± 0.805	14.046 ± 0.292	12.826
405	12.283 ± 0.951	14.346 ± 0.168	n.d.
412	14.116 ± 0.209	14.642 ± 0.072	n.d.
413	14.171 ± 0.198	14.073 ± 0.195	n.d.
419	13.792 ± 0.300	13.816 ± 0.336	13.861
432	11.577 ± 0.206	12.966 ± 0.149	11.808
462	14.076 ± 0.100	13.851 ± 0.588	14.427
463	14.016 ± 0.116	13.878 ± 0.090	14.621
512	13.633 ± 0.211	13.820 ± 0.124	13.522
560	13.134 ± 0.497	n.a.	n.d.

TABLE 4.6: Prediction of the ^{13}C chemical shifts of $NS5B_{\Delta 21}$. Average chemical shift predictions from ShiftX2 obtained for the $\delta 1$ carbons of the isoleucine side-chains of $NS5B_{\Delta 21}$ separated by closed and open structures and compared to the experimental chemical shifts. n.a. not available, the open crystallographic structures did not have electronic density for I560. n.d. not determined, resonances in the spectra were not identified for isoleucines 405, 412, 413 and 560.

was identified in the crystallographic structures of NS5B in open conformations. Additionally, I160 and I432 that also have large deviations in the predictions have resonances with larger linewidths in the ^{13}C dimension. This suggests that these residues are more affected by changes in conformation which supports possible open/close dynamics.

We measured the ^{1}H, ^{13}C combined chemical shifts perturbations between the experimental positions of chemical shifts and the positions calculated by ShiftX2 for all structures, for closed structures and open structures exclusively (Figure 4.12). The chemical shifts predicted for the closed structures are closer to the experimentally observed position of the resonances for most resonances and in particular for I432, I419 and I363. Contrastingly, the experimental ^{1}H, ^{13}C chemical shifts of I160 are closer to

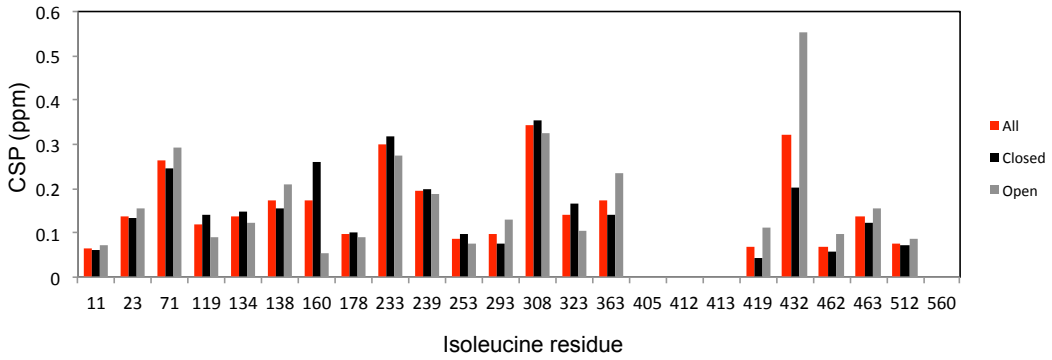

FIGURE 4.12: Comparison between open and closed chemical shift predictions. ^{1}H, ^{13}C combined chemical shifts perturbations (CSP) between the experimental chemical shifts of the resonances of $NS5B_{\Delta 21}$ and the average predicted chemical shifts obtained using ShiftX2. These values are shown for all structures (red), only closed structures (black) and only open structures (gray).

the predictions of the open structures. Nonetheless, this data suggest that the polymerase mainly samples a closed conformation but should also sample open conformations.

One of the major contributions to the $C^{\delta 1}$ chemical shifts of isoleucines is the isoleucine side-chain χ_2 dihedral angle [403]. These correspond to dihedral angles between the isoleucine carbons α, β, $\gamma 1$ and $\delta 1$. The side-chains χ_2 can lead to four conformations, of which trans, gauche- and gauche+ can be populated in solution. These correspond to dihedral angles that are on average 170°, -60° and 60°, respectively. Trans and gauche+ conformations give rise to similar $C^{\delta 1}$ chemical shifts, however gauche- gives rise to smaller values of $C^{\delta 1}$ chemical shift. Due to this relationship between the angle and the chemical shift we measured the χ_2 dihedral angles in the structures used for chemical shift predictions (Figure 4.13).

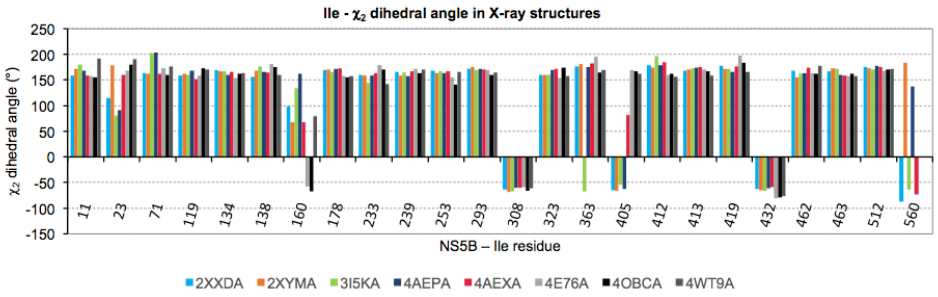

FIGURE 4.13: χ_2 dihedral angles of the isoleucines in the NS5B JFH-1 apo structures. Values around 170° represent trans conformations and values of 60° and -60 represent gauche+ and gauche- conformations respectively. Structures in shades of gray (4E76A, 4OBCA and 4WT9A) are of NS5B in open conformations.

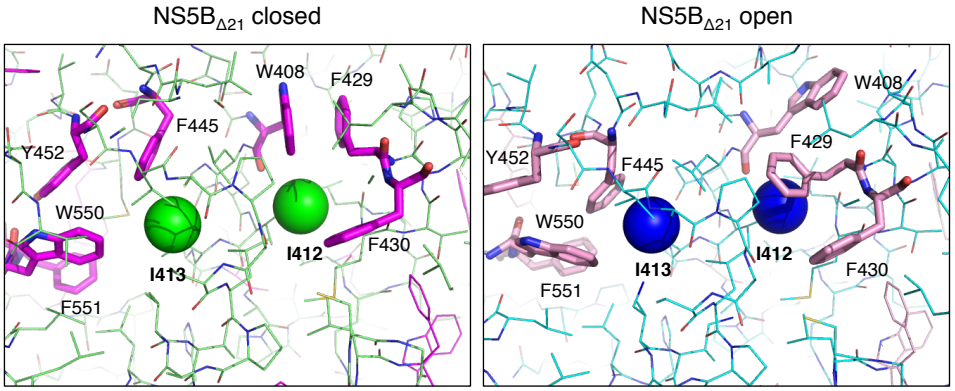

FIGURE 4.14: Reorientation of the aromatic amino acid side-chains around isoleucines 412 and 413. Closed (3I5K, left) and open (4OBC, right) structures are shown in line representation. Ile 412 and 413 $\delta 1$ carbons are shown as spheres. The aromatic amino acids are shown in magenta/pink and those that are reoriented in the open structure are shown as sticks. The reorientation of the aromatic side-chains can affect the proton chemical shift observed by the resonances corresponding to these residues. This can explain the large normalized chemical shift deviations observed for these residues.

The distribution of dihedral angles correlates well to characteristics we observe in the spectrum. In the structures, most isoleucine side-chains are in always in trans, this corresponds well to the observation that most resonances have $C^{\delta 1}$ chemical shifts above 12 ppm. I308 and I432 are the only residues with all structures in gauche- conformation and they give the only resonances with $C^{\delta 1}$ chemical shifts below 12 ppm. However these resonances are not at 9.3 ppm as would be expected if they populated the gauche- conformation exclusively, demonstrating that in solution they should populate other conformations as well. Furthermore, I405 presents a distribution of χ_2 dihedral angles, both in trans and gauche- that is consistent with its broadening beyond detection. None of the open crystallographic structures has electron density for I560 but even the variation of its dihedral angles between trans and gauche- among the closed crystallographic structures suggest that it may be sampling a very large conformational space in solution. I412 and I413 do not show this variation in dihedral angles but the large normalized deviation of chemical shift predictions was mainly due to contributions from the ^{1}H chemical shifts. Indeed, in the crystallographic structures, these neighboring residues are near aromatic amino acids (W408, F429, F430, F445, Y452, W550, F551) whose relative orientation and/or distance appears to change between open and closed conformations of NS5B (Figure 4.14).

4.6 Study of the dynamics of NS5B$_{\Delta 21}$

The first crystallographic structures of NS5B [137–139] were too closed to allow the fit of dsRNA which was needed for the polymerase activity. Consequently, it was hypothesized that the polymerase must undergo a change in conformation to form a competent elongation complex. More recently, crystal structures of mutants have been obtained that are in a conformation that is capable of containing dsRNA [145, 148], in a conformation called "open" as opposed to the "closed" conformation of the wild-type apo enzyme.

In this work, initially, we will consider the global dynamics of the polymerase as an exchange between two states, closed and open, related by an exchange constant k$_{ex}$ (Equation 4.1). This is only an approximation, however, as it is unlikely that the exchange between open and closed states is a single step and also unlikely that there are only one open and one closed states of the polymerase. Indeed, we will see that this two-state model does not correctly describe all of our observations.

$$NS5B_{closed} \rightleftharpoons^{k_{ex}} NS5B_{open} \tag{4.1}$$

The exchange regime of each resonance depends on whether k$_{ex}$ is larger (fast exchange), approximately the same (intermediate exchange) or smaller (slow exchange) than the difference in chemical shift between the states in angular frequency units ($\Delta\omega$). In the methyl-HMQC spectrum we observe mostly one resonance for each isoleucine $\delta 1$ methyl group that is present. If indeed we have two states that are exchanging, for each residue that gives a single resonance we have three possibilities: there is no difference in chemical shift between the states, the second state has a very small population or the exchange is fast between the states. For the residues that are broadened beyond detection these are in the intermediate exchange regime where k$_{ex} \approx \Delta\omega$. If there is only a single chemical exchange occurring in this protein, the presence of two regimes within the same protein is due to a difference in $\Delta\omega$.

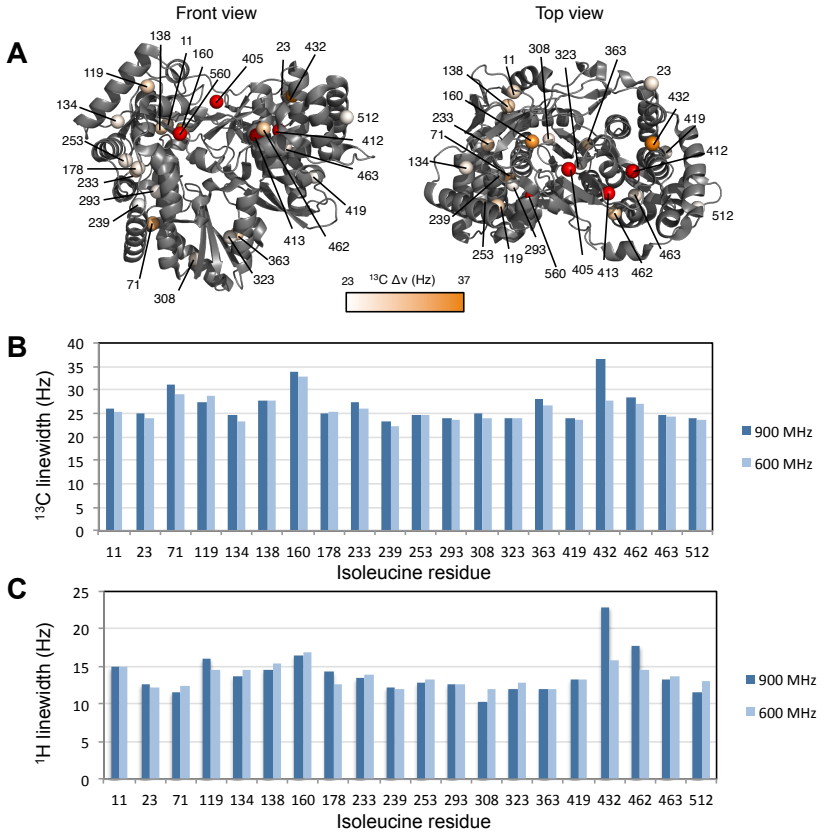

FIGURE 4.15: Linewidths of the methyl-HMQC spectrum. **A** Crystallographic structure of wild-type NS5B (PDB 3I5K) showing the ^{13}C linewidths of the δ1 carbons of the isoleucines (spheres) obtained with a 900 MHz spectrometer. Orange spheres represent the larger linewidths. Red spheres represent the isoleucines with resonances broadened beyond detection. **B** ^{13}C linewidths of the δ1 carbons of the isoleucines obtained with 900 and 600 MHz spectrometers. **C** ^{1}H linewidths of the δ1 hydrogens of the isoleucines obtained with 900 and 600 MHz spectrometers.

4.6.1 Line-broadening as a measure of dynamics

We extracted the linewidths of the methyl resonances in the direct and indirect dimensions at two different field strengths (Figure 4.15). In the ^{1}H dimension only I432 has a larger linewidth than most other visible residues at 900 MHz. Interestingly, at 600 MHz I432 is not broader than other residues, this suggests that its $\Delta\omega$ is already rather close to k_{ex}, becoming closer to k_{ex} at 900 MHz. Analysis of the ^{13}C linewidths show that I71, I160 and I432 have a larger linewidth than other visible residues. For residues I160 and I432 this could be explained by a possible open/close exchange as both are near the residues that are broadened beyond detection. I160 is on the interior side of the RNA binding groove and I432 is at the top of the thumb subdomain, regions expected to be perturbed by this reorientation. This is harder to explain for I71. Despite it being on

the interface between the fingers and the palm, I239 and I293 are on the same interface and are not broadened.

4.6.2 Slow exchange in NS5B$_{\Delta 21}$

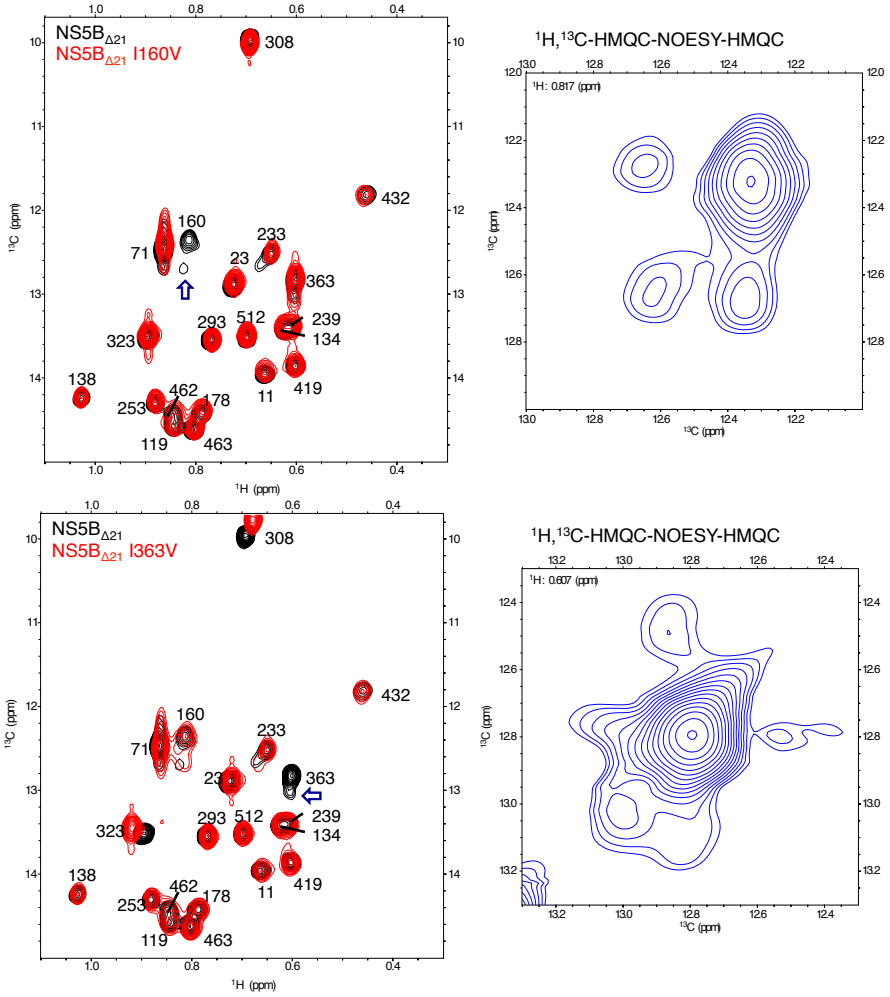

FIGURE 4.16: Identification of minor forms of I160 and I363. **Top** Comparison of the methyl-HMQC of NS5B$_{\Delta 21}$ wild-type (black) and I160V (red) showing that a minor peak (blue arrow) disappears with the mutation. In the 3D ^{1}H,^{13}C-HMQC-NOESY-HMQC spectrum of NS5B$_{\Delta 21}$ there are cross-peaks between the major and minor peaks corresponding to I160. **Bottom** A minor peak is also observed in the methyl-HMQC of NS5B$_{\Delta 21}$ wild-type (black) and I363V (red) that also displays cross-peaks in the 3D ^{1}H,^{13}C-HMQC-NOESY-HMQC spectrum.

The 3D NOESY spectra acquired for the assignment of the isoleucine side-chain resonances also provide insights into the slow exchange regime. Resonances due to isoleucines 160 and 363 have minor peaks adjacent to them that also disappear from the spectrum when these isoleucines are substituted by valines suggesting these are minor conformations of these isoleucines (Figure 4.16, left). By observing their frequency on

the HMQC-NOESY-HMQC we can detect cross-peaks between these peaks and the major peaks of I160 and I363 (Figure 4.16, right). The presence of these cross-peaks is probably because these residues have two distinct conformations that are in slow exchange between each other. The mixing time used for the NOESY transfer can also provide the necessary time for chemical exchange to occur in a slow exchange regime like in an EXSY experiment.

Interestingly, these residues are about 29 Å from each other in the crystallographic structures and are in very different structural parts of the protein. I160 is in the fingers subdomain facing the RNA-binding groove and I363 is in a hinge region between the palm and thumb. Since we do not seem to observe this exchange with other residues, it is likely that the exchange in this timescale observed on these residues is isolated to specific regions of the polymerase instead of a global exchange.

The relative volume of these minor peaks compared to the major peaks can provide information on the population of the minor state but is hard to estimate as these are not well-isolated peaks. The minor resonances of I160 and I363 correspond to 32.8 % and 21.0 % of the total volume due to these isoleucine methyl groups.

We can estimate a range for this slow exchange. I160 and its minor peak are separated by 0.35 ppm in the ^{13}C dimension, which corresponds to a frequency of 498 rad/s. To be in slow exchange, k_{ex} needs to be much slower than this frequency (2 or 3 times slower at least). So the slow exchange on I160 should be much slower than 498 s^{-1} and 285 s^{-1} on I363.

4.6.3 Intermediate exchange in NS5B$_{\Delta 21}$

The lack of signal for the resonances for isoleucines 405, 412, 413 and 560 suggest that they exchange on a ms-μs timescale. To investigate exchange in this timescale, we used methyl-TROSY ^{1}H-^{13}C multiple quantum Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion experiments (Figure 4.17). We repeated the experiment using 900 and 600 MHz spectrometers. The data was fit to a two-state chemical exchange model like in Equation 4.1. The experiment provides an exchange rate k_{ex}, the relative population of the minor state p_B and the difference in chemical shift between the states [367, 404].

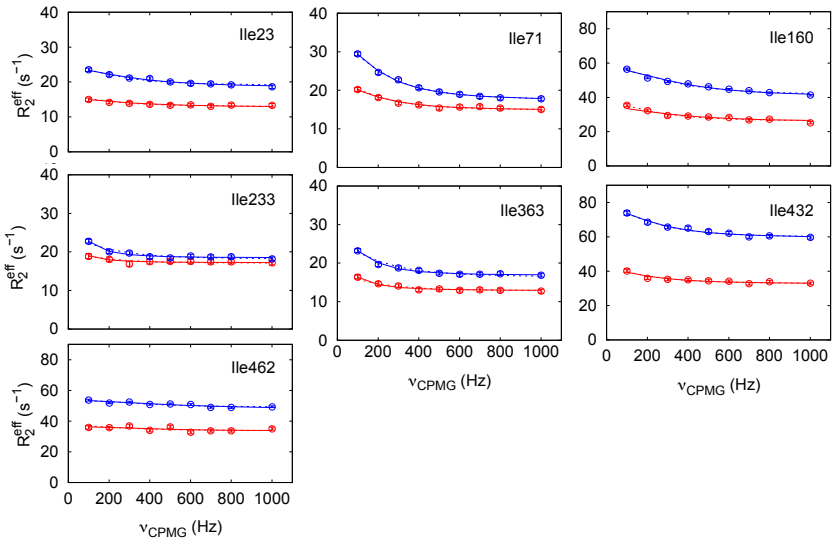

FIGURE 4.17: Relaxation dispersion plots for resonances with significant exchange contributions to transverse relaxation rate $R_{2,eff}$. Red curves were from experiments at acquired with a 600 MHz spectrometer and blue curves were acquired with a 900 MHz spectrometer. Solid lines represent the best fits obtained by fitting the residues individually. Dashed lines represent the best fit obtained by fitting the residues with significant dispersion grouped together.

Firstly in this experiment, we tested whether there were significant exchange contributions to the transverse relaxation, i.e. non-flat dispersion curves, among the resonances. Of the 20 resonances, 7 displayed non-flat dispersion curves according to F test statistics. The exchange rates of the individual resonances (Table 4.7) can be split into three different ranges, around 600, 1450 and above 2000 s^{-1}. Still, there is no apparent correlation between these values and the position on the structure. Moreover, the populations of the minor form are ill-defined and not consistent among these groups.

A global fit of all resonances (Table 4.8) with non-flat dispersion curves gave a global k_{ex} of (1309 $\pm$ 135) s^{-1} and a not well defined p_B of (6.1 $\pm$ 6.6) %. However, as there is a large variation among the individual-fit k_{ex} values and a sizeable increase in the reduced χ^2 measure of goodness of fit between the global and individual-residue fits this indicates that different dynamic processes are acting on different residues.

This data supports that we do not have a single global exchange process occurring in solution. Instead, different parts of the protein may have separate movements with different rates or we may observe an exchange that must be fitted to a model with more states. Caillet-Saguy *et al.* [158] used crystallographic structures of the BK isolate

Isoleucine	k_{ex} (s^{-1})	Δk_{ex} (s^{-1})	p_B (%)	Δp_B (%)	$\Delta\omega(^{13}C)$ (ppm)	$\Delta\Delta\omega(^{13}C)$ (ppm)	$\Delta\omega(^{1}H)$ (ppm)	$\Delta\Delta\omega(^{1}H)$ (ppm)	χ^2	Reduced χ^2
23	2187	481	21.5	217.7	0.20	0.73	-0.032	7.295	2.659	0.222
71	1398	114	25.9	11.2	0.25	0.04	-0.064	0.241	3.961	0.330
160	2024	509	3.8	3.4	0.83	0.39	–	–	26.068	2.005
233	620	1418	26.4	307.7	0.13	0.24	0.041	2.025	4.031	0.336
363	570	117	2.3	0.4	0.59	0.09	–	–	3.232	0.249
432	1502	288	20.7	46.2	0.32	0.32	0.093	0.255	9.818	0.818
462	2675	2068	0.7	0.8	1.38	0.95	–	–	19.707	1.516
								Total:	69.476	0.799

TABLE 4.7: Parameters obtained from fitting a two-state exchange model to the methyl-TROSY multiple quantum CPMG relaxation data of individual resonances of apo $NS5B_{\Delta 21}$. These values and their uncertainties were obtained using the cmpg_fitd9 software [397]. k_{ex}, exchange rate; p_B, population of the minor exchanging state; $\Delta\omega$, chemical shift change between states, χ^2, chi-squared goodness of fit measure. When ^{1}H $\Delta\omega$ are shown both ^{1}H and ^{13}C $\Delta\omega$ were allowed to vary, otherwise ^{1}H $\Delta\omega$ was set to zero. The choice between the two was based on smaller parameter uncertainties as the difference in goodness of fit was not significant.

Isoleucine	k_{ex} (s^{-1})	Δk_{ex} (s^{-1})	p_B (%)	Δp_B (%)	$\Delta\omega(^{13}C)$ (ppm)	$\Delta\Delta\omega(^{13}C)$ (ppm)	$\Delta\omega(^{1}H)$ (ppm)	$\Delta\Delta\omega(^{1}H)$ (ppm)	χ^2	Reduced χ^2
23	1309	135	6.1	6.6	0.32	0.17	-0.070	0.050	88.008	0.917
71					0.49	0.26	0.000	316.4		
160					0.69	0.39	0.090	0.030		
233					0.28	0.14	0.000	115.3		
363					0.35	0.19	0.000	54.8		
432					0.55	0.28	0.030	0.110		
462					0.43	0.25	0.120	0.020		

TABLE 4.8: Parameters obtained from fitting a two-state exchange model to the methyl-TROSY multiple quantum CPMG relaxation data of apo $NS5B_{\Delta 21}$ globally. The non-flat dispersion curves were simultaneously fit to a single exchange rate. Abreviations are as in Table 4.7. These values and their uncertainties were obtained using the cmpg_fitd9 software [397]. For the global fit, both ^{1}H and ^{13}C $\Delta\omega$ were allowed to vary.

(genotype 1b) to determine independent rigid parts and flexible regions of the polymerase. The authors suggest division of the polymerase in fingers, palm and thumb but also other independent rigid groups: top of fingers, outer fingertips, central fingertips, beta-hairpin hinge and beta-flap. They also identified flexible regions that were not part of these rigid groups, like the linker. Despite the difference in genotype, if we use their division of independently moving parts we observe that most of our resonances with non-flat dispersion are in different independent rigid groups (Table 4.9). Those that are in the same independent rigid groups, I71 and I233 (fingers), and I432 and I462 (thumb) do not display k_{ex} of the same order when fit independently. Thus, this does not suggest that they have coordinated movement. This further supports that more resonances would be needed to use the CPMG experiment to determine independently exchanging groups.

Isoleucine	Independent rigid group
23	outer fingertips
71	fingers
160	central fingertips
233	fingers
363	beta-hairpin hinge
432	thumb
462	thumb

TABLE 4.9: Independent rigid groups of the residues that display non-flat relaxation dispersion. Caillet-Saguy *et al.* proposed the further division of NS5B beyond the subdomains based on independent rigid groups observed by comparison of crystallographic structures. Our 7 resonances that display non-flat relaxation dispersion are in 5 different independent rigid groups.

Furthermore, how slow exchange observed for resonances corresponding to Ile 363 and 160 fits into this model is not clear. Since both of these resonances also present intermediate exchange, they should exchange between at least three states, related by two different exchange rates.

Altogether, $NS5B_{\Delta 21}$ is a dynamic protein with dynamics in various timescales that cannot be fit to a single two-state exchange model. Extending the selective labeling to other methyl groups (the polymerase contains 55 leucine and 38 valine residues) or to other amino acids could provide a more complete set of resonances that could aid in identifying the different dynamic processes at play.

4.7 Selective tryptophan-ε1 labeling

Another selective labeling strategy that was used for $NS5B_{\Delta 21}$ was the labeling of the nitrogens in the tryptophan indole side chains. By introducing ^{15}N-labeled indole in the culture medium prior to induction of protein expression, the protein is selectively ^{15}N-labeled on the tryptophan side-chains. This method described by Tarragò *et al.* [405] takes advantage of the favorable relaxation properties of tryptophan side-chains and can be used without deuteration. This should result in a more cost-effective labeling strategy when compared to selective isoleucine labeling with perdeuteration.

$NS5B_{\Delta 21}$ has 9 natural tryptophans of which 5 are in the thumb subdomain. However, only resonances corresponding to five tryptophan side-chains are present in the spectrum (Figure 4.18). The remaining four are broadened beyond detection. This may be due to the same dynamics that broadened four of the isoleucine resonances. None of

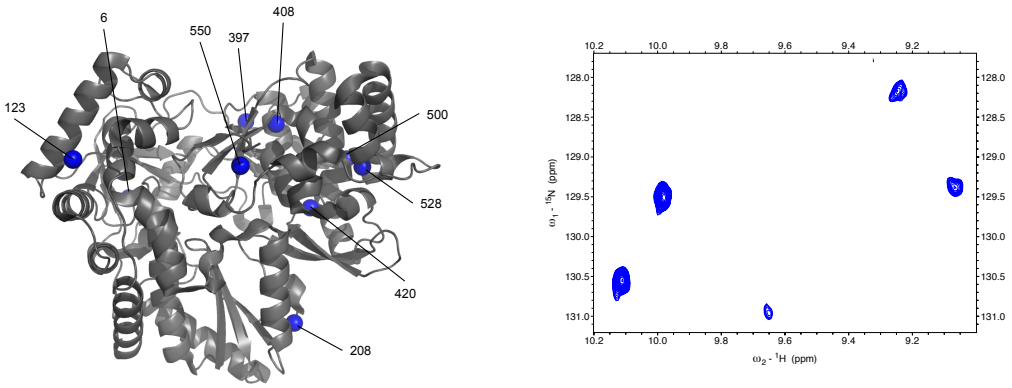

FIGURE 4.18: Tryptophan labeling of NS5B JFH-1. **Left** Cartoon representation of the crystallographic structure of NS5B JFH-1 (pdb 3I5K) with the position of tryptophan ε1 nitrogens marked as blue spheres. **Right** ^{15}N,^{1}-TROSY-HSQC of Trp-ε1-{^{15}N}-labeled NS5B$_{\Delta 21}$ JFH-1.

the tryptophan side-chain nitrogens are within 10 Å of each other so it is not possible to use NOEs cross-peaks to get a preliminary assignment. Since we wanted to gain new probes in a region of the protein in which we did not have many probes and there is a high chance that these residues are broadened beyond detection we did not pursue mutation-based assignment of the detected resonances.

4.8 NS5B open structure constructs

Several different constructs of NS5B were used in this work to try to identify a spectral signature of the potential opening of NS5B. These constructs are described in the literature as being more open or having modified open-close dynamics.

4.8.1 NS5B$_{\Delta 21}$ L30S

In 2002, Labonté and coworkers [406] described a point mutation in NS5B that introduced a polar residue at the apex of the Λ1-loop where it binds to a pocket in the thumb subdomain. In the wild-type protein, this contact is mediated by a strong hydrophobic interaction by leucine 30. This loop is one of the structural characteristics believed to be responsible for the closed conformation of NS5B. The L30S mutant was shown to be almost inactive in a polymerase assay. This mutant was further investigated by Rigat *et al.* [153] that showed a 100-fold decrease in polymerase activity compared to wild-type. Additionally, it was demonstrated that this mutant was more sensitive than wild-type

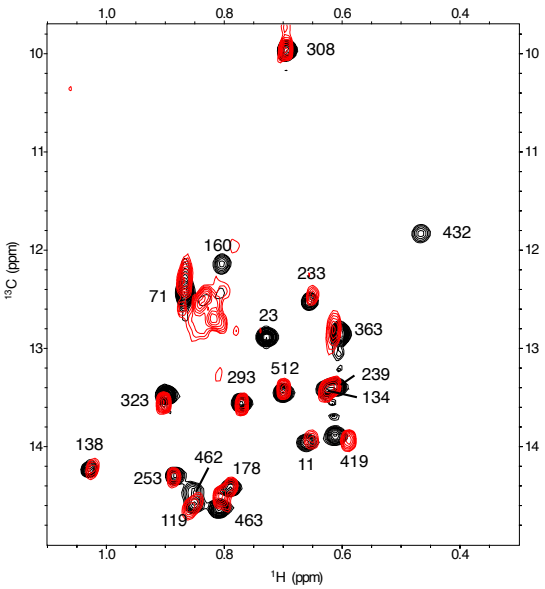

FIGURE 4.19: Methyl-HMQC spectrum of NS5B$_{\Delta 21}$ L30S (red) overlaid on the wild-type spectrum (black). I23, I160 and I432 could not be identified in the spectrum.

to trypsin proteolysis between lysines 50 and 51 at the base of the fingertips and that the same digestion pattern occurred in the presence of inhibitors of the thumb-I class.

These thumb-I inhibitors, first described by Di Marco *et al.* [241], were shown to cause the displacement of the fingertips Λ1-loop. In crystals of the enzyme bound to this type of inhibitor, no electron density was identified for 14 residues of this loop, suggesting that it is moving freely. Additionally, the crystal structure displays an intermediate opening with interdomain angles larger than 70° and template channel lengths smaller than 20 Å. Kinetic studies of the binding of a thumb-I inhibitor to wild-type and L30S suggest that binding to L30S is diffusion controlled, unlike for the case of wild-type, further supporting the hypothesis that the Λ1-loop is free from the thumb-domain in this mutant [407].

Using molecular dynamics simulations, Davis *et al.* [243] proposed that thumb-I inhibitors do not allow the enzyme to stably sample active conformations, instead it rapidly samples both open and closed conformations. Additionally, the movement of the thumb domain, which was previously correlated to the movement of the fingers through the Λ1-loop becomes uncorrelated.

We were interested in observing the effect of the L30S mutation on our methyl-HMQC spectrum as an example of NS5B that has been suggested to sample open conformations more than wild-type. The overlay of the methyl-HMQC spectra of NS5B$_{\Delta 21}$

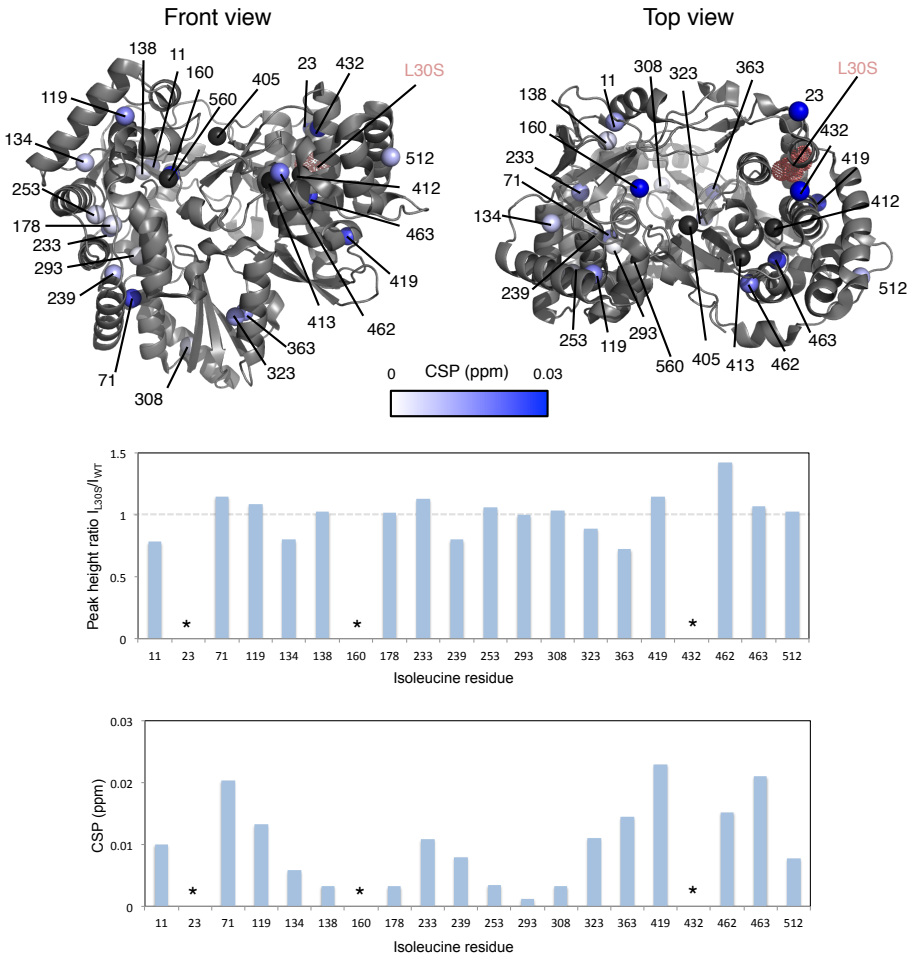

FIGURE 4.20: Effect of L30S on NS5B$_{\Delta 21}$. I23, I432 and I160 could not be identified in the spectrum. **Top** Crystallographic structure of wild-type NS5B (PDB 3I5K) showing the chemical shift perturbations (CSP, blue) due to the introduction of the L30S mutation (in pink). Isoleucines that correspond to resonances not visible in the wild-type spectrum are colored in gray. **Middle** Changes in peak height caused by the mutation of the fingertips loop residue. **Bottom** ^{1}H,^{13}C combined chemical shift perturbations caused by the mutation. Asterisks represent the resonances that could not be identified in the NS5B$_{\Delta 21}$ L30S methyl-HMQC spectrum.

L30S and wild-type is shown in Figure 4.19.

The resonances assigned to I23, I160 and I432 have disappeared showing that this mutation changes the dynamics of the protein, either causing some residues to become disordered or broaden beyond detection. In the spectrum of mutant NS5B$_{\Delta 21}$ we observe additional signals in the random-coil region of the spectrum (0.86 ppm, 12.90 ppm [329]). Based on the crystal structures of NS5B bound to thumb-I inhibitors and the kinetic studies of NS5B L30S binding to such inhibitors we expect that I23, in the Λ1-loop, becomes disordered or is in a different exchange regime. That is, if the fingertips Λ1-loop was previously bound to the thumb and now exchanges between bound

and free states we could be observing intermediate exchange for this residue. I432 is in the binding pocket within the thumb that is in contact with the fingertips Λ1-loop. A chemical exchange between states where the loop is bound and is free could also explain the disappearance of this resonance. I160 presented a slow exchange between two forms, it is possible that the mutation shifted its population towards the minor form in the disordered region of spectrum.

Additionally, residues throughout the protein are affected by this mutation and undergo chemical shift perturbations. Isoleucines 419, 462 and 463 in the thumb, 323 and 363 in the palm, and 71 and 119 in the fingers present the largest combined chemical shift perturbations. The core of the fingers subdomain (with the exception of I71) appears to be the least perturbed region. This suggests a global average conformation that is different than wild-type, possibly with a reorientation of the thumb subdomain toward an open conformation (Figure 4.20).

4.8.2 NS5B$_{\Delta 60}$

Another structural characteristic proposed to be responsible for the closed conformation of NS5B observed in crystal structures is the presence of the 39 amino acid linker, between the C-terminal anchor sequence and the catalytic domain, which appears folded back into the RNA-binding groove in the structures. Deletion of the C-terminal 55 amino acids results in a more active polymerase than the 21-amino acid deleted construct [157] suggesting that the linker precludes the adoption of a functionally active conformation. Molecular dynamics suggest that the presence of the linker limits the conformational sampling of the enzyme [159].

Compared to NS5B$_{\Delta 21}$, the construct deleted of the final 60 amino acids, NS5B$_{\Delta 60}$ is more difficult to express and purify, even when not labeled, 4 mg/L of culture were achieved with this construct, compared to 15 mg/L of culture for NS5B$_{\Delta 21}$. Additionally, only lower concentrations were achieved before precipitation of the protein. This is in agreement with differential scanning fluorimetry experiments by Boyce *et al.* [408] that showed that progressive deletion of the linker results in lower thermal denaturation temperatures for NS5B.

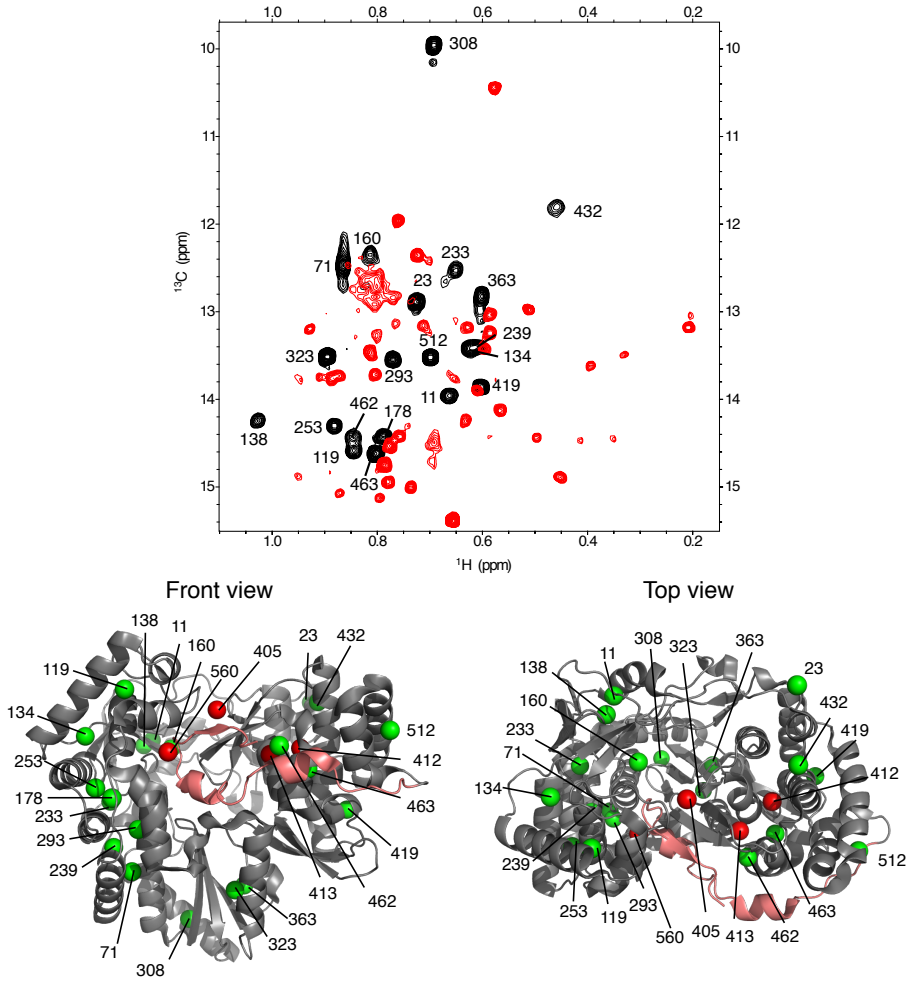

FIGURE 4.21: Methyl-HMQC spectrum of NS5B$_{\Delta 60}$. **Top** Spectrum of NS5B$_{\Delta 60}$ (red) overlaid on the NS5B$_{\Delta 21}$ spectrum (black). The abundance of resonances can be due to slow exchange between two conformations. **Bottom** Crystallographic structure of NS5B with the residues that have been deleted in NS5B$_{\Delta 60}$ (from the linker) highlighted in pink. Isoleucine $\delta 1$ carbons are shown as spheres, green for the resonances visible in the NS5B$_{\Delta 21}$ spectrum and red for the others.

The methyl-HMQC of a selective isoleucine-labeled sample of this construct (Figure 4.21) displays roughly 40 resonances while the construct contains 23 isoleucines. The larger number of resonances compared to total isoleucines suggests that there may be two conformations in slow exchange in solution. However, we do not have an assignment of the resonances in this methyl-HMQC spectrum, thus we cannot fully take advantage of this spectrum.

4.8.3 NS5B$_{\Delta 21}$ 5mut

Lam and colleagues [394, 409] identified three point mutations that provided resistance to a guanosine analogue (a nucleotide inhibitor). Together these mutations, S15G, C223H and V321I resulted in an enzyme that presents 50% higher activity than wild-type [394]. A crystallographic structure with these mutations and two surface stabilization amino acid substitutions, E86Q and E87Q, was obtained and presented extension of the Λ1-loop and rearrangement of the thumb subdomain as well as partial disordering of C-terminal residues. This structure (PDB 4OBC) revealed the same rearrangements that had been observed in a crystal structure of NS5B bound to primer-template duplex RNA in which the β-hairpin loop had been deleted [145].

The crystal structure of NS5B JFH-1 with five amino acid substitutions S15G, E86Q, E87Q C223H and V321I (which here we call 5mut for simplicity) exhibits a structure that is classified as open by the conformational parameters described previously. Therefore, we were interested in observing how these mutations would impact the spectrum. The influence on the thumb subdomain was of particular interest because none of the point mutations are in this region and it experiences the largest rearrangement in the crystal structure.

We acquired the methyl-HMQC spectra of NS5B$_{\Delta 21}$ 5mut (Figure 4.22). Due to the V321I substitution, this mutant possesses an extra isoleucine and indeed 21 resonances can be identified in the spectrum. Most resonances have not shifted with the inclusion of the five point mutations and could be assigned directly from the wild-type spectrum. The remaining resonances were tentatively assigned assuming that the electronic effect of the new isoleucine would affect I323 and I363 similarly and cause in both a downfield shift (in proton and carbon dimensions) as expected by comparing the chemical shift predictions of closed conformations and PDB 4OBC. The remaining resonance was assigned to I321. This was confirmed by a ^{1}H-^{1}H 2D NOESY spectrum (Figure 4.23) that shows the correlation between the resonances assigned to I321, I323 and I363.

The changes in peak height and chemical shift compared to wild-type are shown in Figure 4.24. As expected, the largest changes occur in resonances corresponding

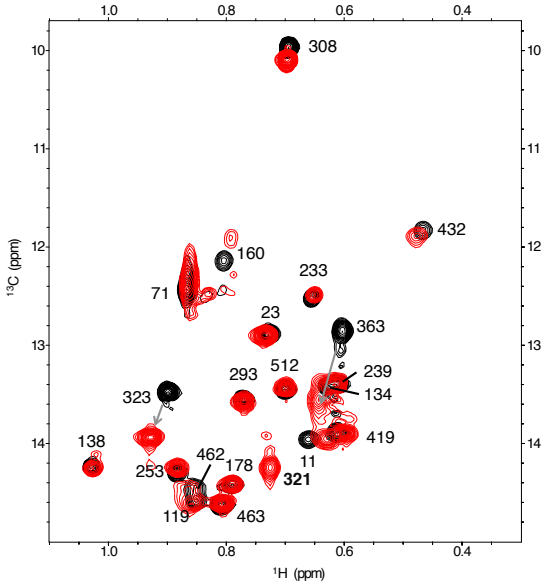

FIGURE 4.22: Methyl-HMQC spectrum of NS5B$_{\Delta 21}$ 5mut (red) overlaid on the wild-type spectrum (black). The new isoleucine, I321 is shown in bold. Arrows show the chemical shift perturbations undergone by I323 and I363.

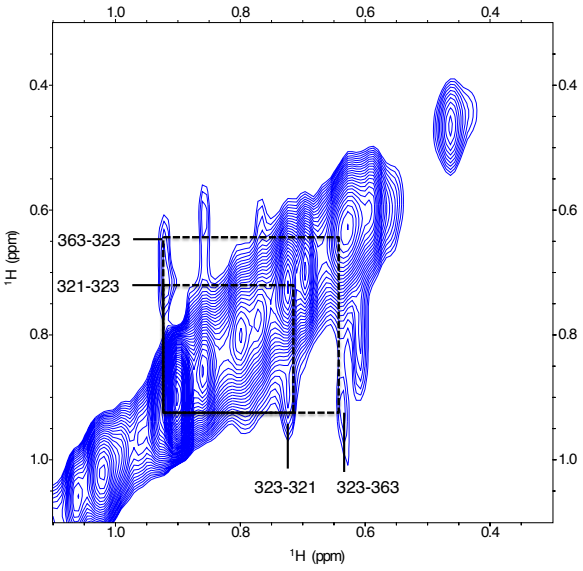

FIGURE 4.23: ^{1}H-^{1}H NOESY spectrum of NS5B$_{\Delta 21}$ 5mut used to confirm the expected assignment of isoleucines 321, 323 and 363. The cross-peaks between the resonances corresponding to I321 and I363 are not visible because they are expected to be too close to the main diagonal of the spectrum.

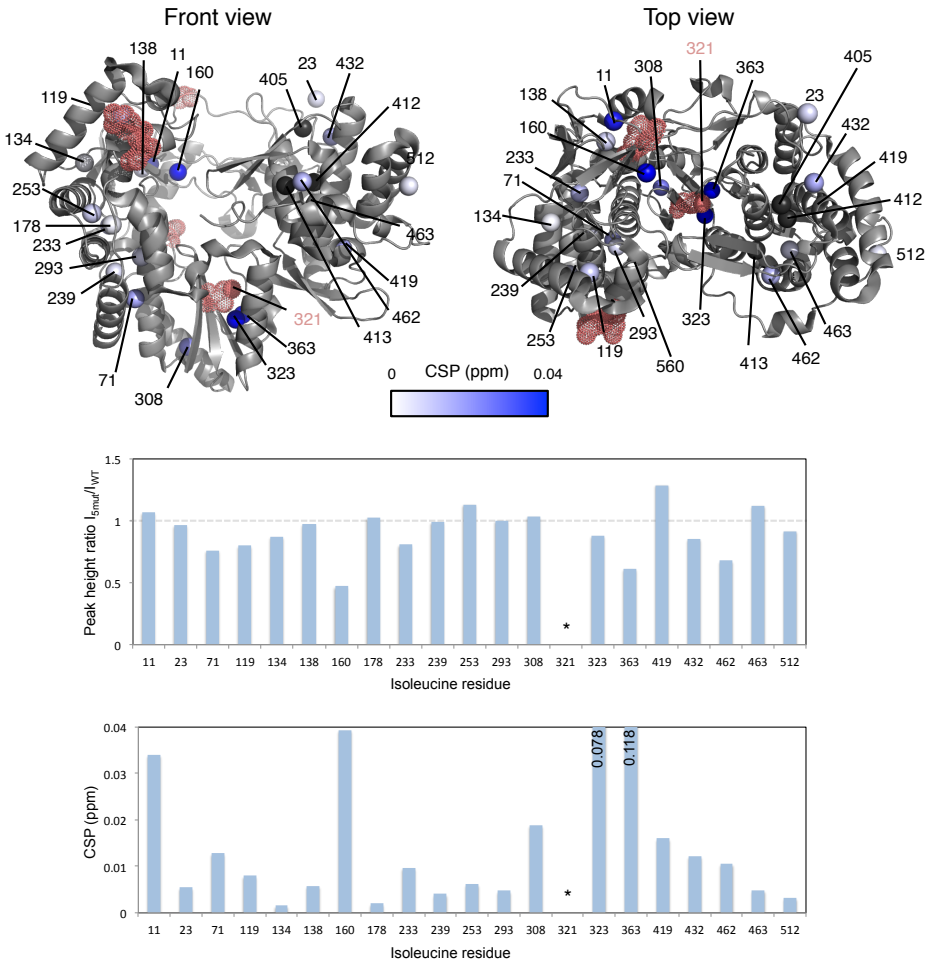

FIGURE 4.24: Effect of 5mut (S15G E86Q E87Q C223H V321I) on $NS5B_{\Delta 21}$. I321 was introduced by the mutations and is therefore highlighted. **Top** Crystal structure of $NS5B_{\Delta 21}$ 5mut (PDB 4OBC) showing the combined chemical shift perturbations (CSP, blue) due to the introduction of the mutations (in pink). **Middle** Changes in peak height caused by the introduction of the mutations. **Bottom** CSP caused by the mutations. I323 and I363 have CSP values larger than the scale, therefore their CSP values are labeled. I321 that was introduced by the mutations is marked with an asterisk.

to residues which are relatively close to the mutations. Isoleucines 11 and 160 are about 15 Å from both S15G and C223H mutations and isoleucines 323 and 363 are 6 Å from the V321I mutation. However, despite the lack of mutations in the thumb region, isoleucines 419 and 432 display $^{1}H,^{13}C$ combined chemical shift perturbations larger than 0.01 ppm. I419 also is also more intense in the mutant than wild-type. Interestingly, resonances corresponding to residues in the fingers subdomain that are not as close to mutations as 11 and 160, like 178 and 233, display smaller CSPs than for I419 and I432, suggesting that indeed it is the thumb subdomain that undergoes a rearrangement as proposed by X-ray crystallography.

4.8.4 Comparison between NS5B open structures

From the chemical shift predictions, the resonances corresponding to I160 and I432 are expected to be the most the sensitive to the opening. However, the resonance of I160 is not visible in the methyl-HMQC spectrum of NS5B$_{\Delta 21}$ L30S and the perturbation may be due to a mutation in close proximity in the NS5B$_{\Delta 21}$ 5mut methyl-HMQC spectrum. The resonance of I432 is broadened in the methyl-HMQC of NS5B$_{\Delta 21}$ L30S, no resonances are close to its expected position in the methyl-HMQC spectrum of NS5B$_{\Delta 60}$ and it is shifted in the NS5B$_{\Delta 21}$ 5mut spectrum. Thus, if there is a concerted opening movement, this residue seems to act as a probe as it is perturbed by these mutations. Nevertheless, it cannot be used to determine a degree of opening because sometimes broadening, sometimes chemical shift perturbations are observed thus it is not consistent enough to use as a probe.

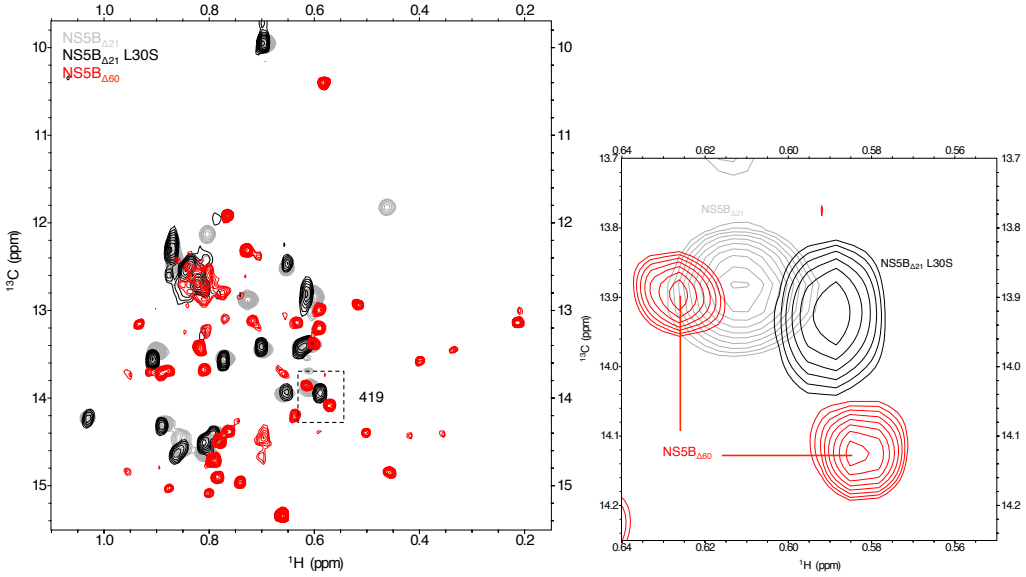

FIGURE 4.25: Comparison of the resonance corresponding to I419 in NS5B$_{\Delta 21}$, NS5B$_{\Delta 21}$ L30S and NS5B$_{\Delta 60}$. **Left** Overlay of methyl-HMQC spectra of NS5B$_{\Delta 21}$ (gray), NS5B$_{\Delta 21}$ L30S (black) and NS5B$_{\Delta 60}$ (red). The position of the resonances estimated to correspond to I419 are shown with a dashed box. **Right** Zoom of the region containing the I419 resonances. In the NS5B$_{\Delta 60}$ spectrum, the two closest resonances to the expected I419 are shown.

Another resonance, the one that corresponds to I419, seems to be an interesting candidate to be a probe of this open/close reorientation. If we compare the methyl-HMQC spectrum of NS5B$_{\Delta 60}$ to that of NS5B$_{\Delta 21}$ and NS5B$_{\Delta 21}$ L30S around the resonance of Ile 419 (Figure 4.25) we observe that one of the peaks of NS5B$_{\Delta 60}$ overlaps with a peak of NS5B$_{\Delta 21}$ and the resonance assigned to I419 in the methyl-HMQC of NS5B$_{\Delta 21}$ L30S is

about halfway between the peak that aligns well with the NS5B$_{\Delta 21}$ wild-type position of I419 and another peak present in the NS5B$_{\Delta 60}$ spectrum. We can hypothesize that NS5B$_{\Delta 21}$ wild-type is in one conformation, while NS5B$_{\Delta 60}$ exchanges slowly between this and another conformation. Crystal structures suggest these could be closed and open conformations. NS5B$_{\Delta 21}$ L30S in this interpretation would be in fast exchange between these conformations as shown by its resonance that is an average of the peaks apparent in the NS5B$_{\Delta 60}$ spectrum. However this is only speculation and does not seem to occur with other peaks in the spectrum.

NS5B construct	CSP (ppm)	Classification
Δ21 wild-type	0	Closed
Δ21 L30S	0.023	Intermediate
Δ60	0.009 / 0.063	Intermediate
Δ21 5mut	0.016	Open

TABLE 4.10: Comparison of combined chemical shift perturbations (CSP) of I419 for NS5B open structures. CSP were calculated compared to NS5B$_{\Delta 21}$ wild-type. Both resonances shown in Figure 4.25 are shown for NS5B$_{\Delta 60}$. The classification of structures as open or closed was as shown in Figure 4.10.

We had observed an apparent correlation between the shifting of the resonance corresponding to I419 in the NS5B$_{\Delta 21}$ L30S and NS5B$_{\Delta 60}$ and the opening of the structure. NS5B$_{\Delta 21}$ 5mut also displays a shifting of this resonance in the same direction as for the other mutants but the degree of shifting does not appear to correlate directly to the degree of opening of the structure as the crystal structure of NS5B$_{\Delta 21}$ 5mut (PDB 4OBC [394]) displays structural metrics that correspond to a fully open structure. Compared to this, structures with the linker cleaved (NS5B$_{\Delta 55}$, 1GX5 and 1CSJ, [137, 410]) display structural metrics that can be classified as intermediate. The only NS5B L30S crystallographic structure available [411] also shows intermediate values but was obtained with a palm-II inhibitor so it might not be representative of the apo enzyme. Thus, the degree of CSP of the resonance corresponding to I419 does not appear to directly correlate to the degree of opening observed in the crystallographic structures (Table 4.10).

Altogether, we do not observe a consistently useful probe into the possible opening of the polymerase. This may be because the change between the conformations in solution is not a single concerted movement, in agreement with the CPMG relaxation dispersion data.

4.9 Conclusions

$NS5B_{\Delta 21}$ is a large protein by NMR standards which makes its study without perdeuteration and TROSY-type pulse sequences unpractical. Despite both of these advances, the assignment of the $^{1}H,^{15}N$–TROSY-HSQC is difficult due to heterogeneous linewidths among the peaks and many that are broadened beyond detection, at least in the triple resonance spectra needed for the assignment even at 900 MHz.

To study this protein despite this challenge we used perdeuteration and selective isoleucine labeling. Combined with the intrinsic TROSY effect of the methyl groups we are able to obtain a methyl-HMQC that has 20 resonances for the 24 isoleucines and these peaks were assigned.

Compared to other proteins and complexes beyond 100 kDa to which this labeling strategy has been applied in solution NMR spectroscopy [361], $NS5B_{\Delta 21}$ is rather small. However, $NS5B_{\Delta 21}$ presents different challenges. Beyond 100 μM (and even less for the mutants) we observe precipitation of the polymerase. Also, unlike the 670 kDa 20S proteasome studied by Sprangers and Kay [362] it is not stable at high temperatures, which can be used to reduce the rotational correlation time (τ_C). Additionally, as it is a monomer so we cannot take advantage of symmetry to improve signal and reduce signal overlap. All of this, combined with the intrinsic exchange dynamics of NS5B in solution create an interesting but challenging system to study by NMR spectroscopy.

We characterized exchange in the millisecond to microsecond timescale using CPMG relaxation dispersion experiments and identified slow exchange on two residues. Based on these experiments, $NS5B_{\Delta 21}$ does not seem to undergo single two-state exchange. Instead, independent movement of different parts and more than two exchanging conformations likely occur.

Using different mutants and constructs, we looked for spectral probes of opening. Based on χ_2 dihedral angles determined using crystallographic structures and chemical shift predictions the resonances that should be the most sensitive to this change in conformation are those that correspond to I160 and I432. $NS5B_{\Delta 60}$ presented a double population of resonances with about the same intensity, suggesting two conformations populated about equally, exchanging slowly. $NS5B_{\Delta 21}$ L30S and 5mut display changes

in the thumb subdomain but only one set of resonances suggesting that these mutants may be quickly exchanging between conformations.

The resonances of I160, I419 and I432 appear to be affected in the spectra of open constructs. I432 is affected by broadening, usually beyond detection, making it not a very useful probe if we wish to use it to measure the degree of opening. Similarly, I419 is shifted in open structures, however, the degree of shifting does not seem to be directly correlated to the the degree of opening. Furthermore, comparing crystallographic structures to determine the degree of opening in solution may be misleading as the structure present in solution can be a combination of exchanging conformations and the crystallographic structure could be a local energetic minimum.

In conclusion, the isoleucine methyl-HMQC spectrum of NS5B$_{\Delta 21}$ was assigned. Based on this spectrum of wild-type NS5B$_{\Delta 21}$, spectra of mutants, a NOESY/EXSY spectrum and CPMG experiments, we observe that NS5B is a dynamic system in solution that cannot be described simply by an exchange between two forms. In the following chapters we will be using this spectrum to study the interactions of NS5B$_{\Delta 21}$ with partners.

Interaction between NS5B and NS5A-D2

5.1 Introduction

NS5A is an intriguing protein because it has been shown to interact with over 100 partners both from the host and virus [86] but for which no enzymatic activity has been described. It has an organization in three domains, of which the domain 1 (D1) is structured and has been crystallized as homodimers [97, 99, 100]. The other domains are mainly intrinsically disordered [104, 106, 412].

Domain 2 of NS5A (NS5A-D2) has been shown to be necessary for RNA replication [87, 88]. The structure of this domain is intrinsically disordered with regions with secondary structure tendencies [105, 107], a stable structural motif called the PW-turn [110] and weak tertiary structure proposed based on paramagnetic relaxation enhancement studies [413].

NS5A-D2 has been shown to interact directly with NS5B [89, 94, 108]. Surface plasmon resonance data suggests that this interaction is at a 1:1 ratio between the polymerase and NS5A-D2 [108]. However, the importance of this interaction for the activity of the polymerase is not well understood. There are studies that show that NS5A can both stimulate [89–91] and deactivate the polymerase [89, 92, 93] but these studies have been done under different conditions, which can partially explain their disparate results.

5.2 Expression and purification of NS5A-D2 JFH-1

Since in this work we are using $NS5B_{\Delta 21}$ from the JFH-1 strain, we used NS5A-D2 from the same strain. *E. coli* containing a pT7-7 plasmid coding for NS5A-D2 with a C-terminal hexahistidine tag were grown in ampicillin-supplemented mediums, either LB medium (for unlabeled protein) or M9 minimal culture medium with $[^{15}N]NH_4$ as the source of nitrogen (for ^{15}N-labeled protein).

Purification was done in two steps. After lysis of the bacterial cells, the lysate was heated at 75°C for 15 minutes. This takes advantage of the disordered nature of NS5A-D2 which is thermoresistant while the majority of the other proteins in the lysate denature, precipitate and can, subsequently, be separated by centrifugation. After this, a nickel affinity-chromatography step is used to further purify the protein. The resulting 11.6 kDa protein migrates in SDS-PAGE gels as if it is over 18 kDa because of its intrinsic disorder (Figure 5.1) [414]. The yields obtained for NS5A-D2 were about 20 mg/liter of culture for unlabeled samples and 10 mg/ liter of culture for ^{15}N-labeled samples.

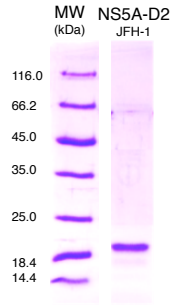

FIGURE 5.1: SDS-PAGE of NS5A-D2 with a 12% acrylamide gel stained with Coomassie blue stain.

5.3 Interactions between NS5B and NS5A-D2

Nuclear magnetic resonance can provide information on interactions with residue-level resolution by using one of the partners labeled with an NMR-active nucleus and adding the other partner or partners unlabeled. The cross-peaks of the labeled protein can undergo two types of changes with the addition of a binding partner. The interaction between partners can cause a shift or a broadening of the cross-peaks or a combination of both, depending on the exchange rate and difference in chemical shift between free and bound states. In general, when the exchange rate is fast compared to the chemical shift difference between states, the resonance appears to shift from the frequency of the free to the frequency of the bound state. Conversely, when the exchange rate is slow, the resonance at the free frequency decreases in intensity and a new resonance appears at the bound frequency. Broadening can occur when the exchange rate is of the order of the difference in chemical shift between the states. Broadening can be

observed even with a fast exchange rate when the bound state has a different rotational correlation time. Otherwise, deviations from this behavior indicate that a model with two states (free/bound) may not be sufficient to describe this interaction. Using NMR to study interactions is particularly interesting when one of the partners is intrinsically disordered as NS5A-D2 because of the difficulty of co-crystallizing these proteins.

5.4 Mapping the interaction on NS5A-D2

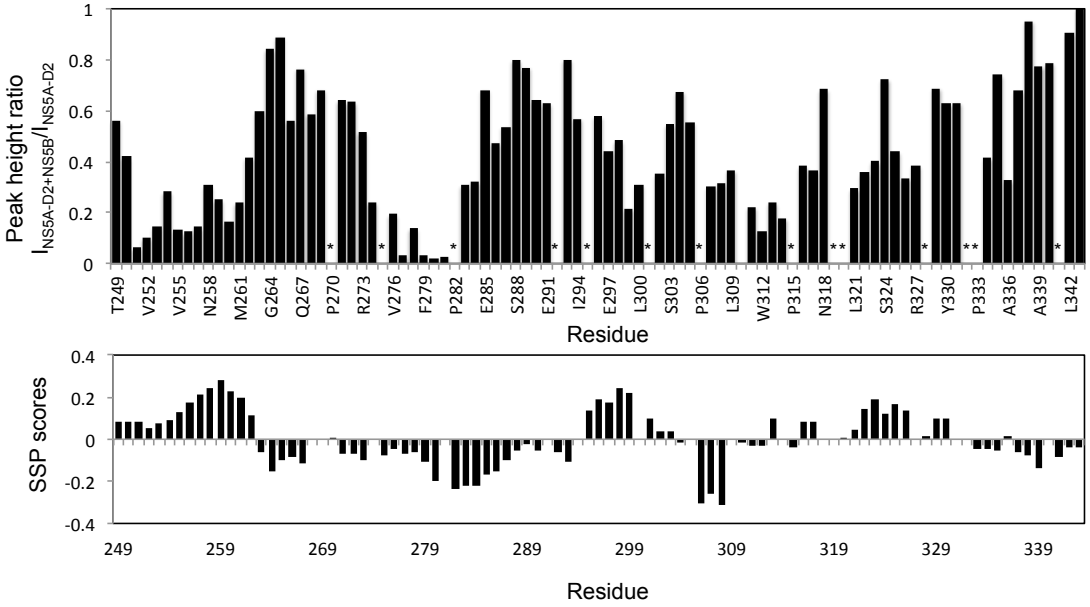

FIGURE 5.2: NMR analysis of the effect of $NS5B_{\Delta 21}$ on NS5A-D2. **Top** Relative peak intensity is given by normalizing the peak height of the interaction spectrum containing 20 μM [^{15}N]-NS5A-D2 and 80 μM unlabeled $NS5B_{\Delta 21}$ with the peak height in the absence of $NS5B_{\Delta 21}$. This profile is in good agreement with the profile obtained by Rosnoblet *et al.* [108]. **Bottom** Secondary structure propensities (SSP) [332] based on C_α and C_β chemical shifts of NS5A-D2 JFH-1 [109]. Positive and negative values represent tendencies to form α-helical and extended structures, respectively.

The interaction between NS5A-D2 and $NS5B_{\Delta 21}$ was studied by our group [108] using ^{15}N-labeled NS5A-D2 from the JFH-1 strain that had been previously assigned by triple resonance backbone assignment. In this work, unlabeled NS5B was added to ^{15}N-labeled NS5A-D2 at 1:1 and 5:1 ratios and mainly broadening of residues was detected. This is likely due to the large size of NS5B that causes the reduction of the rotational correlation time of the residues in contact with it. This allowed the identification of three regions of NS5A-D2 whose resonances broaden with the addition of

NS5B. These correspond to regions on NS5A-D2 that have structural tendencies. The first one (residues 251-261) is in a region with α-helical tendency [106, 413] due to the positive deviation from carbon random coil values [329]. The second interaction region (residues 274-281) shows a negative deviation from these random coil values, representing an extended structure. Finally, the large third region contains a small structural motif that was later identified as a proline-tryptophan turn [110] at residues 310-313.

The region of the small structural motif is highly conserved [105]. It has been shown to be implicated in several other of NS5A-D2's interactions, including with human MOBKL1B and cyclophilin A [94, 415] in the JFH-1 (genotype 2a) and Con1 (genotype 1b), respectively. This suggests that this region and its structure may be essential for the activity of NS5A-D2 on its partners independent of its genotype.

We repeated this experiment to compare it to the previous results. The regions of NS5A-D2 that broaden in the presence of $NS5B_{\Delta 21}$ (Figure 5.2) are in good agreement with Rosnoblet *et al.* [108].

5.5 Mapping the interaction on NS5B

After mapping the interaction on NS5A-D2, our interest shifted to mapping the interaction on NS5B as this can improve our understanding of the functional importance of the interplay between these proteins. $NS5B_{\Delta 21}$ is a 65 kDa protein and, as we saw in the previous chapter, this creates challenges for its study by NMR spectroscopy. Unfortunately, the lack of the assignment of the ^{1}H-^{15}N TROSY-HSQC creates a limitation for residue-level resolution mapping of the interaction. However, taken globally, the effect of the addition of unlabeled NS5A-D2 can be compared to ligands with known interaction zones.

In 2012, our group [108] compared the effect of NS5A-D2 on the ^{1}H,^{15}N TROSY-HSQC of NS5B to the binding of thiophene-2-carboxylic acid (TCA), a thumb-II inhibitor, and recognized that the NS5A-D2 binding site(s) partially overlap with the binding site of TCA. Based on this, it was proposed that NS5A-D2 bound to the thumb-II site on NS5B. In this work, we compared the interaction of NS5B with NS5A-D2 to another thumb-II inhibitor, filibuvir, in Chapter 6.

5.5.1 Selective isoleucine labeling

After the characterization of NS5B selectively labeled on the $\delta 1$ methyl groups of isoleucines, we could take advantage of the methyl-HMQC spectrum to gain information on the protein's interactions with NS5A-D2. To monitor by NMR spectroscopy the effect of NS5A-D2 on $NS5B_{\Delta 21}$, 40 μM isoleucine-labeled $NS5B_{\Delta 21}$ was titrated with 0 – 400 μM of unlabeled NS5A-D2 (Figure 5.3, top). The concentrations were chosen based on the estimation that the affinity constant between NS5B and NS5A-D2 in solution was of the same order as the affinity constant between cyclophilin A and NS5A-D2 of 64 μM [105]. This is because NS5B and CypA share a binding site on NS5A-D2 and have been shown to have affinity constants of the same order. The affinity constant between the polymerase and disordered protein has been determined by surface plasmon resonance as 21$\pm$14 μM [108]. This may be different from the the value determined in solution by NMR because of the immobilization of $NS5B_{\Delta 21}$ [416].

The titration is shown in Figure 5.3. The resonances that display CSPs and/or broadening can provide information on the location of the binding site of NS5A-D2 on NS5B (Figure 5.4). The resonance that has the largest chemical shift perturbation, that corresponds to I419, is at the thumb-II site. This is the same site that was suggested by our group by comparing the effect of the thumb-II site inhibitor TCA and NS5A-D2 on $NS5B_{\Delta 21}$ using ^{1}H,^{15}N-HSQC spectra of NS5B [108]. CSPs are also observed in the fingertips $\Lambda 1$ loop (I23) and palm (I323 and I363). Changes in intensity occur throughout the polymerase, broadening occurs in resonances corresponding to isoleucines 419 and 432. The perturbation of resonances corresponding to residues in all three subdomains point to two possibilities: (1) There are multiple binding sites on $NS5B_{\Delta 21}$, which is consistent with three binding sites observed on NS5A-D2 that cover about half its length. (2) That the interaction can cause both direct and allosteric effects. These possibilities are not mutually exclusive as we may observe multiple sites and also long-range allosteric effects.

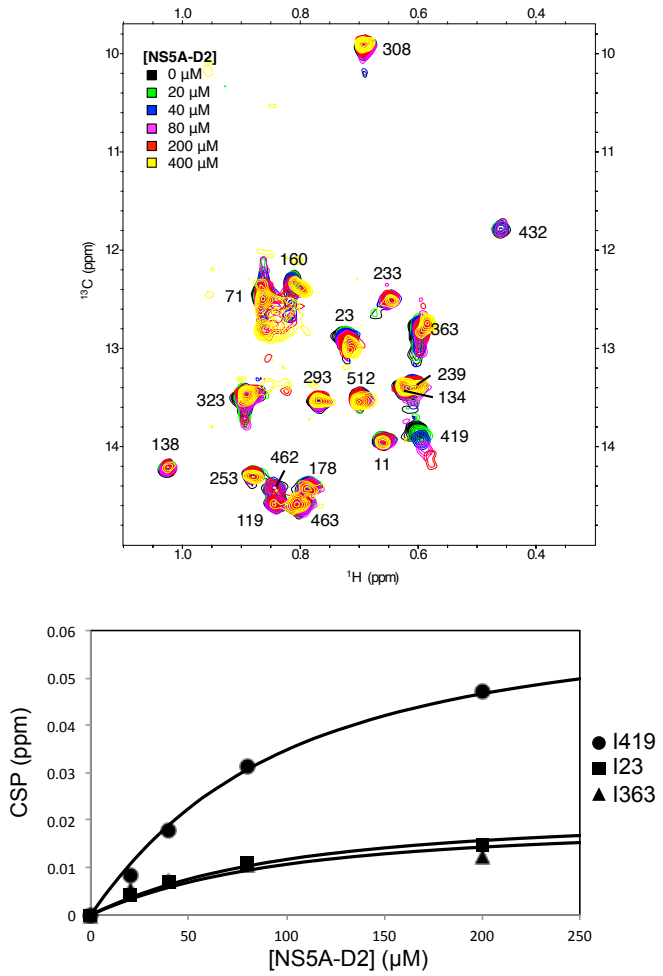

FIGURE 5.3: Titration of isoleucine δ1-labeled NS5B$_{\Delta 21}$ with NS5A-D2 by NMR. Increasing concentrations of NS5A-D2 (0-200 μM) were added to 40 μM of isoleucine labeled NS5B$_{\Delta 21}$. **Top** Overlay of the methyl-HMQC spectra of NS5B$_{\Delta 21}$ in the presence of NS5A-D2. **Bottom** ^{1}H and ^{13}C combined chemical shift perturbations of I419, I23 and I363 were fit together to measure an affinity constant for this interaction of 70 μM.

CSPs from residues shifting in fast exchange can be used to determine the affinity constant (K_D) by using Equation 1.10 [417].

$$\Delta_{obs} = \Delta_{max} \frac{(K_D + [L]_0 + [P]_0) - \sqrt{(K_D + [L]_0 + [P]_0)^2 - (4[P]_0[L]_0)}}{2[P]_0} \tag{5.1}$$

where $[P]_0$ and $[L]_0$ represent the concentrations in solution of the labeled protein and the ligand, respectively. Δ_{obs} and Δ_{max} are the CSPs at the observed concentration of ligand and at saturation, respectively. In this titration the residues with the largest CSP were I419, I23 and I363. By fitting the CSP for these residues simultaneously, the dissociation constant (K_D) of this interaction was estimated to be 70 μM (Figure 5.3, bottom).

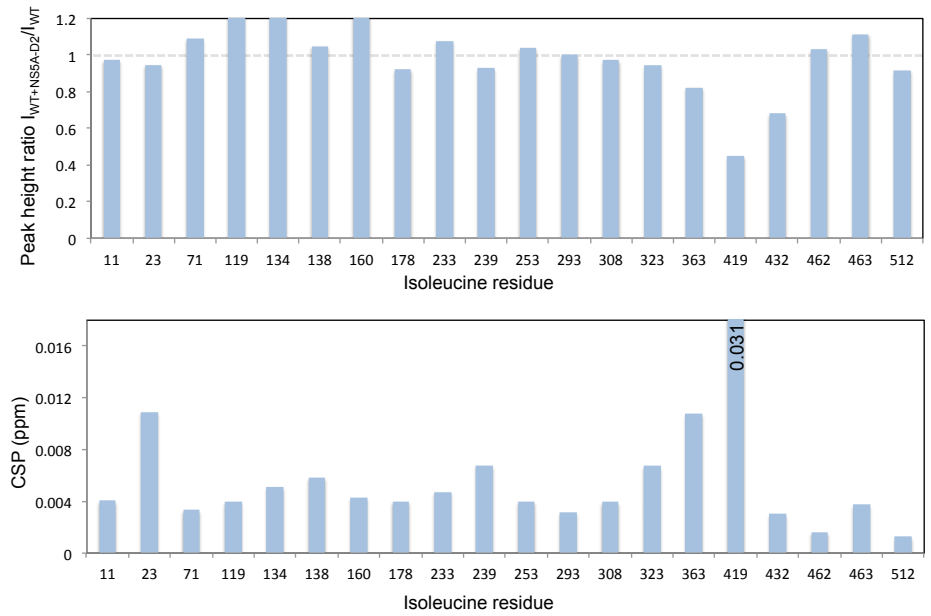

FIGURE 5.4: Effect of NS5A-D2 on NS5B$_{\Delta 21}$ spectra. Data extracted from the comparison of spectra shown in Figure 5.3 containing 40 μM of U-^{2}H, Ile δ1-^{13}C^{1}H$_3$ NS5B$_{\Delta 21}$ and either with 200 μM unlabeled NS5A-D2 or without. **Top** Peak height ratio between the cross-peaks in the spectra with and without NS5A-D2. **Bottom** Chemical shift perturbations due to the addition of NS5A-D2.

Using the experimental K_D, the expected percentage of NS5B$_{\Delta 21}$ bound to NS5A-D2 at each point in the titration can be calculated using Equation 5.2 to estimate the concentration of the complex:

$$[PL] = \frac{([P]_0 + [L]_0 + K_D) - \sqrt{([P]_0 + [L]_0 + K_D)^2 - 4[P]_0[L]_0}}{2} \tag{5.2}$$

where [PL] is the concentration of the bound complex (NS5B:NS5A-D2) and [P]$_0$ and [L]$_0$ are the initial concentrations of the protein (NS5B$_{\Delta 21}$) and ligand (NS5A-D2), respectively. The percentages of NS5B$_{\Delta 21}$ that is expected to be bound in solution during the titration are shown in Table 5.1.

[NS5A-D2] (μM)	% bound NS5B$_{\Delta 21}$
0	0
20	22.2
40	36.4
80	53.3
200	74.1
400	85.1

TABLE 5.1: Proportion of NS5B$_{\Delta 21}$ that is bound to NS5A-D2 at the concentrations used in the titration in Figure 5.3 based on a dissociation constant of 70 μM.

In the spectrum we observe broadening of the resonances of $NS5B_{\Delta 21}$ with the addition of NS5A-D2. However, this is not consistent with a slow-exchange process because we do not detect the apparition of resonances corresponding to the bound state, despite the high ratio $NS5B_{\Delta 21}$ expected to be bound at the end of the titration experiment. Instead, this suggests that the binding mechanism includes more than one step and/or the complex retains some dynamics.

To discriminate between a direct effect due to nearby binding and long-range allosteric effects and to determine whether there are other binding sites on $NS5B_{\Delta 21}$, we continued the investigation of this interaction.

5.5.2 Is this interaction dependent on the PW-turn motif?

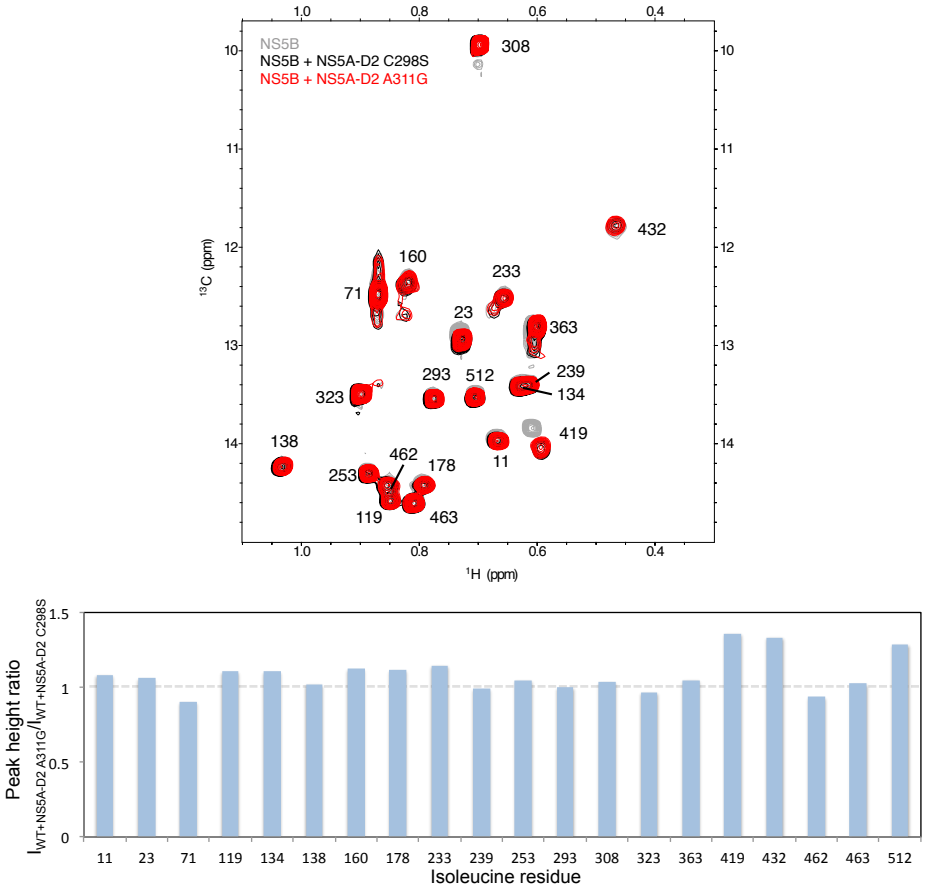

FIGURE 5.5: Effect of the A311G mutation on the ability of NS5A-D2 to bind to NS5B. **Top** Methyl-HMQC spectrum of 71μM of $NS5B_{\Delta 21}$ in the absence of NS5A-D2 (gray) or with 71μM of NS5A-D2, either C298S (black, as a substitute for wild-type) or A311G (red) showing that the mutant that lacks the PW-turn (A311G) still interacts with $NS5B_{\Delta 21}$. **Bottom** Peak height ratios of the cross-peaks in these spectra. There do not seem to be systematic height differences. The peak heights were normalized using I293 because the spectra were not acquired with the same number of scans.

Our group showed the presence of a small structural motif called a PW-turn [110] that contains residues ^{310}PAWA^{314} of NS5A-D2 from the JFH-1 isolate. This structural motif is required for the proper interaction between this viral protein and host cyclophilin A. The resonances of the PW-turn and surrounding residues broaden with the addition of NS5B$_{\Delta 21}$ so we investigated whether this structural motif is required for the interaction between the polymerase and the disordered domain. We used NS5A-D2 A311G, a mutant in which the structural motif is absent. The substitution of the alanine at position 311 for a glycine destabilizes the structural motif in the JFH-1 strain much like the I315G mutation described for the con1 strain [110]. This mutation has been shown to significantly disrupt the ability of a subgenomic replicon to replicate in Huh-7 cells [88]. Additionally, our group had also shown that this mutation enhances the cyclophilin A-catalyzed cis/trans isomerization on the proline adjacent to it [110].

For this we compared the methyl-HMQC spectrum of NS5B in the presence of either NS5A-D2 A311G or NS5A-D2 C298S (as a substitute for wild-type). NS5A-D2 C298S can be used as a substitute for wild-type as it causes the same perturbations as the wild-type NS5A-D2 (see Section 5.5.3).

We expected to see a reduction of the interaction between NS5A-D2 and NS5B when using the NS5A-D2 A311G mutant because the region containing and surrounding the PW-turn was shown to be an NS5B binding site on NS5A-D2. However, there was no significant difference between the interaction spectra (Figure 5.5) suggesting that this motif is not necessary for the interaction between NS5A-D2 and NS5B.

5.5.3 Paramagnetic relaxation enhancement experiments

In the titration experiment described above we cannot discriminate whether the effect (CSP or broadening) on a resonance is direct or due to allostery. This, combined with the limited number of probes (in particular in the RNA-binding groove), made it necessary to use a different strategy to complement the mapping of the interaction. Furthermore, Ngure *et al.* [94] recently suggested that NS5A-D2 inhibits RNA-binding to NS5B competitively, as they suggested the interaction between the polymerase and the disordered domain occurred at the RNA-binding site. Also, we observed that the resonance of I419 was affected through long-range allosteric effects with the binding

of RNA (see Chapter 7). Thus, to discern between direct and allosteric effects and investigate this hypothesis, despite the lack of probes within the RNA-binding groove, we used an NMR technique that gives long-range distance information, paramagnetic relaxation enhancement (PRE).

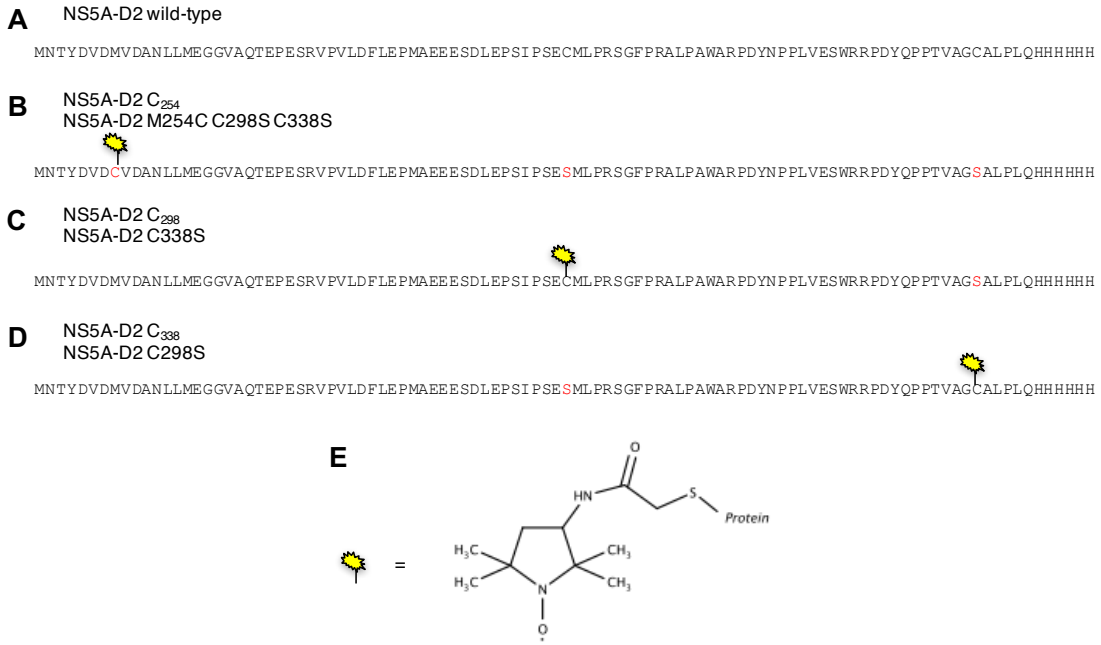

FIGURE 5.6: NS5A-D2 single-cysteine constructs used for PRE. The position of the spin label is represented as a yellow burst symbol. **A** Amino acid sequence of the NS5A-D2 JFH-1 wild-type construct used in this work. It contains two natural cysteines at positions 298 and 338. **B** To obtain a single cysteine near the N-terminal, both natural cysteines were exchanged for serines and methionine 254 was mutated into a cysteine to obtain the NS5A-D2 C_{254} construct. This cysteine was subsequently labeled with 3-(2-iodoacetamido)-PROXYL. **C** The C338 natural cysteine was mutated into a serine to obtain the NS5A-D2 C_{298} construct. **D** The C298 natural cysteine was mutated into a serine to obtain the NS5A-D2 C_{338} construct. **E** After the labeling reaction the 3-(2-iodoacetamido)-PROXYL forms this label covalently bound to the protein by a C-S bond.

Paramagnetic atoms cause increased relaxation that has a r^{-6} distance dependence. This causes broadening of NMR resonances in the presence of a paramagnetic atom. Residues up to 20-25 Å[381] from the paramagnetic label usually display broadened resonances compared to a diamagnetic sample. A smaller ratio between the intensities of the paramagnetic and diamagnetic sample represents a residue that is physically closer to the spin label. Thus we can label NS5A-D2 with a paramagnetic label to gain information on residues in $NS5B_{\Delta 21}$ that are spatially close to the disordered protein. The resonances of residues of $NS5B_{\Delta 21}$ within 25 Å of the spin label on NS5A-D2 are expected to display broadening.

In this work we used 3-(2-iodoacetamido)-PROXYL, that contains a stabilized nitroxide radical, as a spin label. This label was attached to NS5A-D2 through a cysteine residue. To have only a single position labeled at a time, single-cysteine constructs of NS5A-D2 were used. Three such nitroxide radical-labeled single-cysteine constructs of NS5A-D2 were used to determine if there is a preferred orientation of the disordered protein bound to the polymerase. Spin labels were introduced at N-terminal (NS5A-D2$_{C254}$), center (NS5A-D2$_{C298}$) and C-terminal (NS5A-D2$_{C338}$) positions of the disordered protein (Figure 5.6). To obtain a diamagnetic sample we prepared an identical sample and added 4 mM ascorbic acid that reduces the nitroxide radical [418].

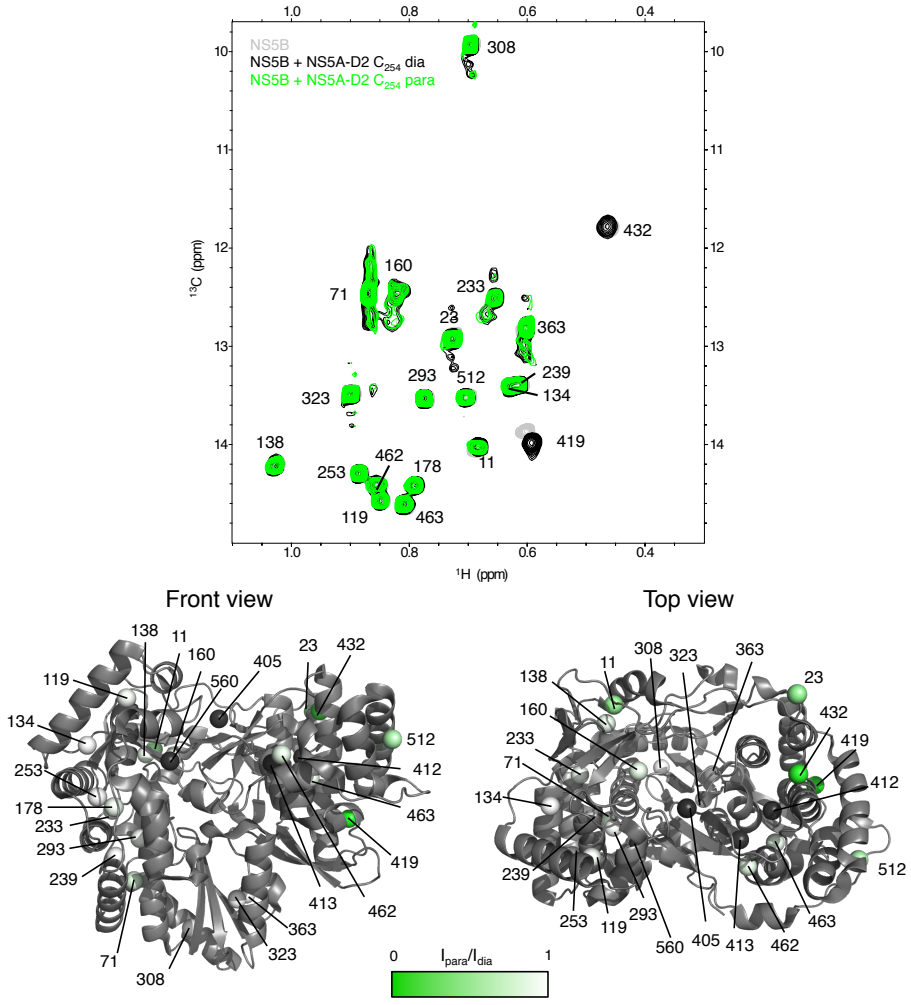

FIGURE 5.7: Methyl-HMQC spectrum of NS5B$_{\Delta 21}$ with spin-labeled NS5A-D2$_{C254}$ **Top** Overlay of methyl-HMQC spectra of NS5B alone (gray) and in the presence of NS5A-D2$_{C254}$ with the spin label (diamagnetic in black and paramagnetic in green). **Bottom** Crystallographic structure of NS5B (PDB 3I5K) showing the broadening caused by the paramagnetic probe (darker green spheres are the isoleucine $\delta 1$ carbons most affected). The unpaired electron causes the broadening of the cross-peaks due to residues physically closest to it.

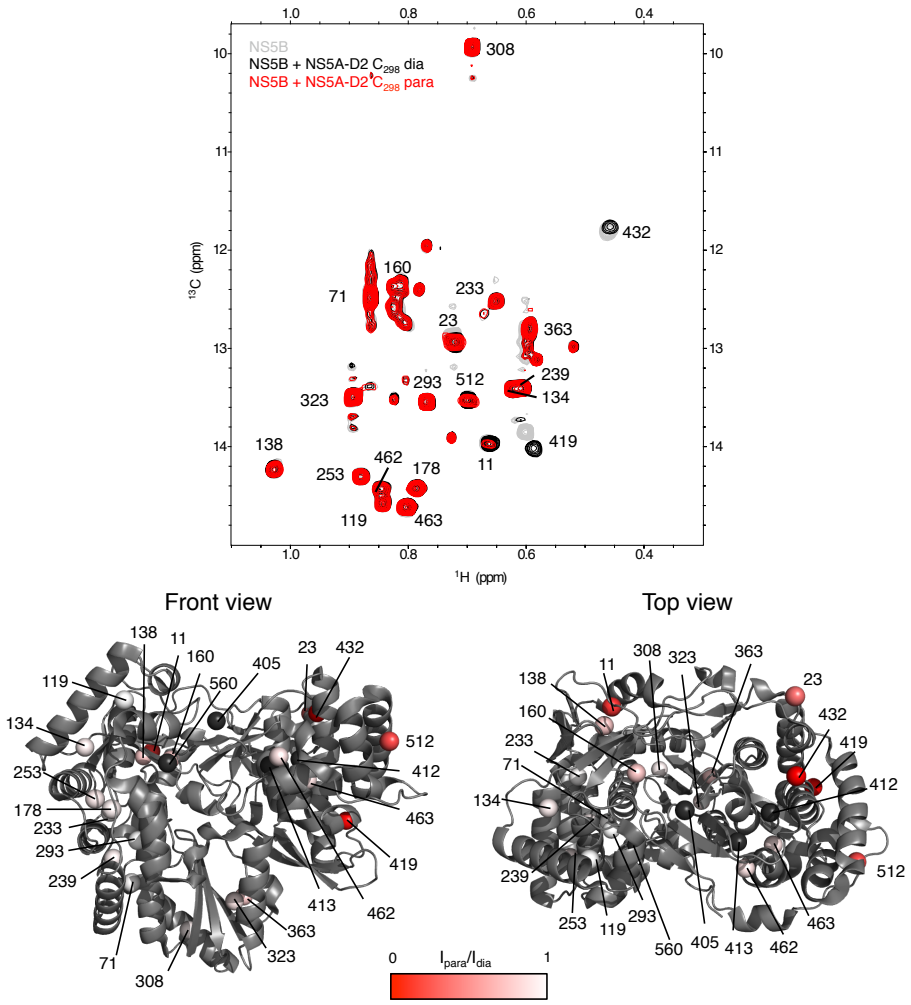

FIGURE 5.8: Methyl-HMQC spectrum of NS5B$_{\Delta 21}$ with spin-labeled NS5A-D2$_{C298}$.**Top** Overlay of methyl-HMQC spectra of NS5B alone (gray) and in the presence of NS5A-D2$_{C298}$ with the spin label (diamagnetic in black and paramagnetic in red). **Bottom** Crystallographic structure of NS5B (PDB 3I5K) showing the broadening caused by the paramagnetic probe (darker red spheres are the isoleucine δ1 carbons most affected). Cross-peaks I419 and I11 are both broadened despite being 40Å from each other in crystal structures.

The mutations and addition of the spin label do not appear to perturb NS5A-D2 binding to NS5B$_{\Delta 21}$. The methyl-HMQC spectra of NS5B$_{\Delta 21}$ with diamagnetic samples of these NS5A-D2 mutants display same the chemical shift perturbations compared to apo NS5B$_{\Delta 21}$, in particular of the resonances corresponding to I419, I23 and I363 that shift with this interaction.

The methyl-HMQC spectra of NS5B$_{\Delta 21}$ in presence of spin-labeled (paramagnetic) NS5A-D2 show that the isoleucines on NS5B$_{\Delta 21}$ closest to NS5A-D2 are those in the thumb subdomain and in the fingertips Λ1-loop (Figures 5.7, 5.8, 5.9 and 5.10). Independent of the position of the spin label on NS5A-D2, the resonances on NS5B$_{\Delta 21}$

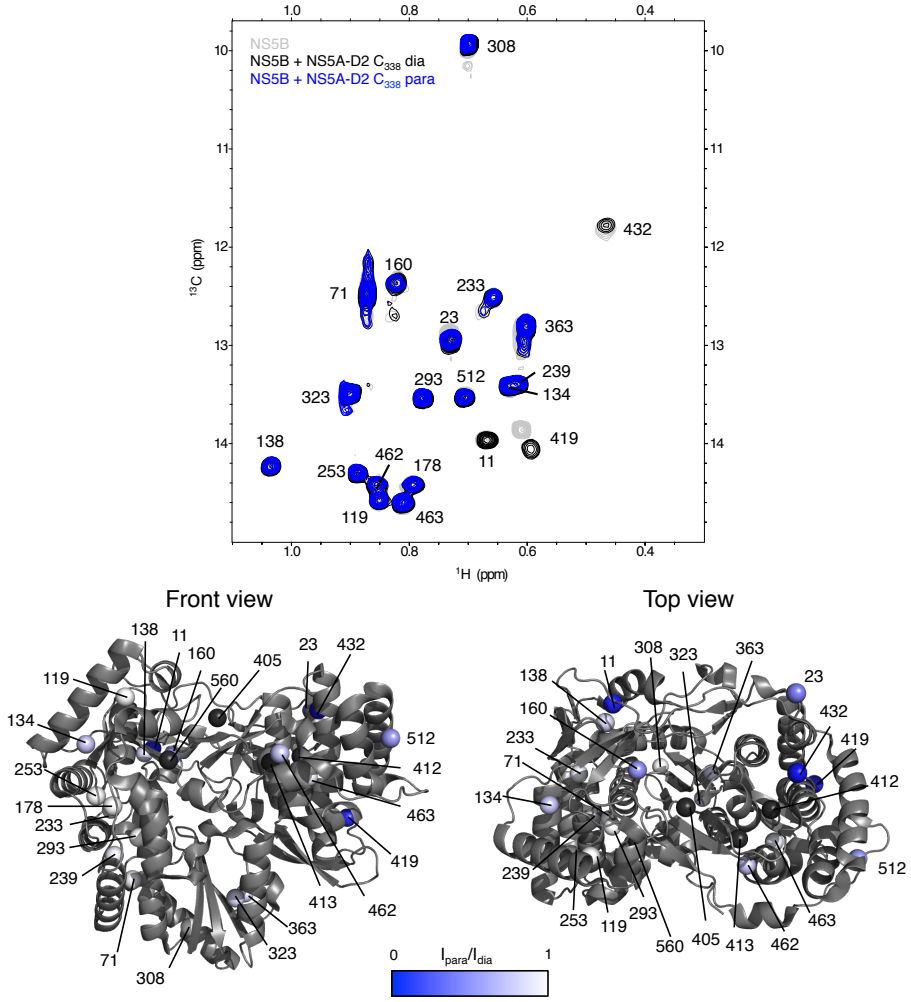

FIGURE 5.9: Methyl-HMQC spectrum of NS5B$_{\Delta 21}$ with spin-labeled NS5A-D2$_{C338}$. **Top** Overlay of methyl-HMQC spectra of NS5B alone (gray) and in the presence of NS5A-D2$_{C338}$ with the spin label (diamagnetic in black and paramagnetic in blue). **Bottom** Crystallographic structure of NS5B (PDB 315K) showing the broadening caused by the paramagnetic probe (darker blue spheres are the isoleucine δ1 carbons most affected).

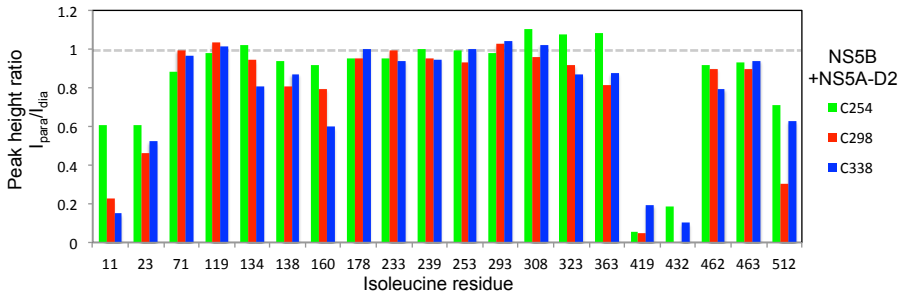

FIGURE 5.10: Intensity ratios of NS5B$_{\Delta 21}$ cross-peaks between paramagnetic and diamagnetic labeled NS5A-D2 for each of the single-cysteine mutants. A lower ratio suggests that the paramagnetic label is located close to the residue that corresponds to this resonance.

that are the most broadened are those of I419 and I432 and to a lesser extent I11, I23 and I512. Little broadening is observed for other resonances from the fingers or any of the resonances from the palm. The isoleucine 419 resonance is the most broadened when the paramagnetic label is on the N-terminal site. In contrast, the isoleucine 432 resonance is the most broadened when the paramagnetic label is on the center and C-terminal regions. These isoleucines are located near the thumb-II and thumb-I inhibitor sites, respectively. One of the main features supporting that the thumb-II site is not the only binding site on NS5B is that the resonance of isoleucine 11 is still affected by the paramagnetic labels, especially those at the center and C-terminal. This isoleucine is located in the fingers subdomain, over 40 Å from the thumb-II site,

Based on the residues with the lowest intensity ratio for each NS5A-D2 mutant, we propose that there is small preference for the orientation of NS5A-D2 when binding to NS5B$_{\Delta 21}$ (Figure 5.11). When the spin-label is close to the N-terminal region, in NS5A-D2$_{C254}$, the most broadened resonance is the one corresponding to I419. C254 is in a region of NS5A-D2 that broadens when NS5B$_{\Delta 21}$ is added, thus it is expected to be within a binding site. The PRE experiment shows that this position of the spin-label causes less broadening on the resonance of I11 than the other two positions of the spin label, as well as less broadening than NS5A-D2$_{C298}$ on resonances of I432 and I512. Based on this, we suggest that the N-terminal region has a preference for binding to the thumb-II site.

Spin-labeled NS5A-D2$_{C298}$ causes the largest broadening on resonance I432. The central position of the spin-label is located in a region that is not a binding site on NS5A-D2 but it is between two sites that display broadening when NS5B$_{\Delta 21}$ is added to NS5A-D2. The spin-label at this position also causes large broadening on resonances I11 and I419 which are 40 Å apart. The midway point between these residues is close to the thumb-I site. Also, resonance I512 displays the largest broadening when the spin label is on this position. All of this supports that the center of NS5A-D2 is preferentially close to the thumb-I site.

Finally, spin-labeled NS5A-D2$_{C338}$ has a similar broadening pattern as when the spin label is on C298, with less broadening on the resonances of I419, I432 and I512. Additionally, it shows broadening (that is less intense than what is observed on the

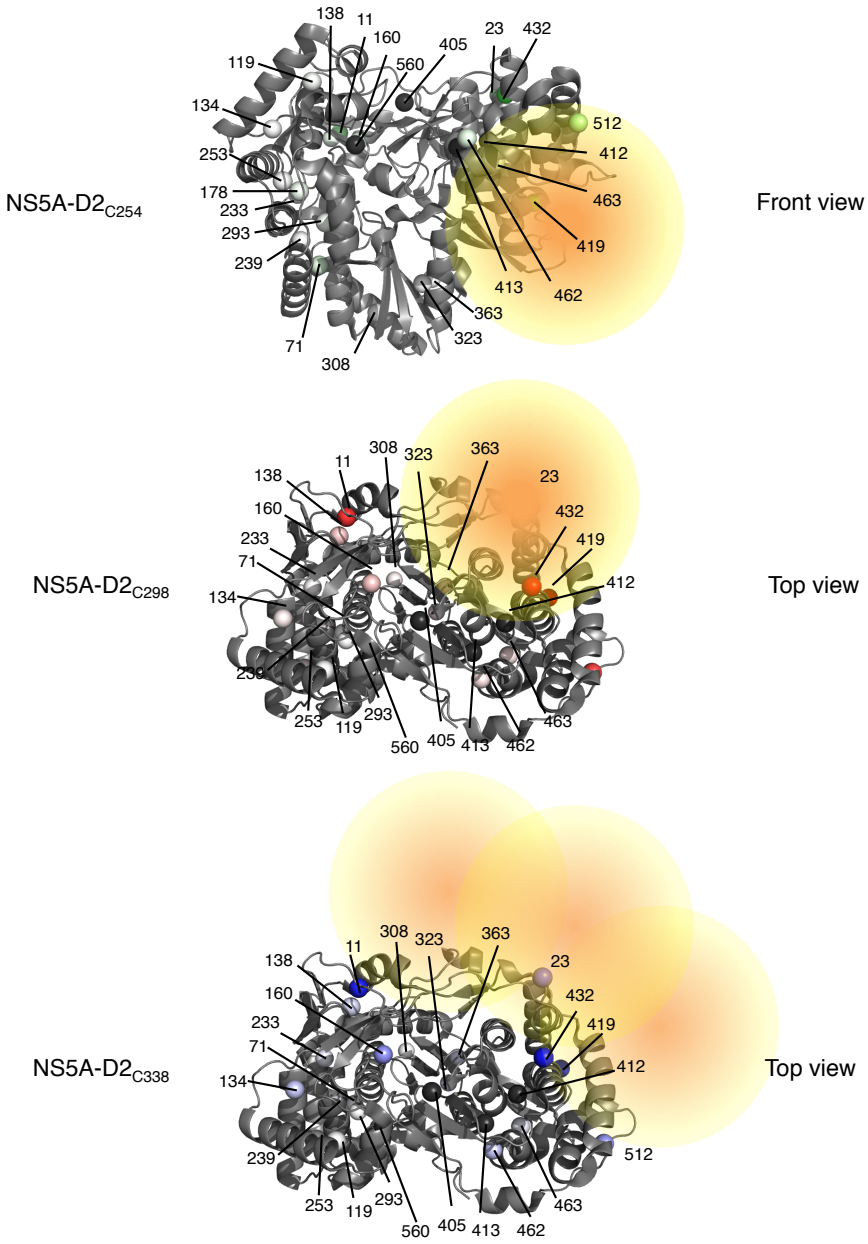

FIGURE 5.11: Proposed position of the spin-label with respect to the NS5B$_{\Delta 21}$ structure. The spin label position is represented by a orange-yellow circles that has about 25 Å radius in the scale of the structures. The top or front view were chosen depending to best represent the position of the spin-label. **Top** Front view of NS5B$_{\Delta 21}$ from Figure 5.7 and the estimated position of the spin label on NS5A-D2$_{C254}$, close to the thumb-II site. **Center** Top view of NS5B$_{\Delta 21}$ from Figure 5.8 and the estimated position of the spin label on NS5A-D2$_{C298}$, close to the thumb-I site. **Bottom** Top view of NS5B$_{\Delta 21}$ from Figure 5.9. Multiple positions of the spin label on NS5A-D2$_{C338}$ are shown to represent the flexibility proposed for this spin label.

resonances from the thumb) on resonances of I134, I138 and I160 in the fingers subdomain. This more "diffuse" broadening, that is spread out over a larger area but less intense, compared to the other spin-label positions leads us to propose that this spin

label is mobile in solution. The resonance of C338 does not display large broadening when NS5B$_{\Delta 21}$ is mixed with NS5A-D2, thus we do not consider it to be within one of the binding sites on NS5A-D2. The PRE data suggests that NS5A-D2 is tethered to NS5B$_{\Delta 21}$ but the C-terminal part containing this residue remains largely mobile in solution.

The interaction between NS5A-D2 and NS5B$_{\Delta 21}$ cannot be described by a single position of NS5A-D2 with respect to NS5B$_{\Delta 21}$ in a conformation like the crystal structures. In this next sections, we will show examples using PRE restraints that show that this interaction cannot be fit with a simple docking method. The intrinsic disorder of NS5A-D2 can cause the paramagnetic label to be in multiple positions with respect to NS5B even if this protein is bound. As a guideline for the discussion, we estimated the distance between the paramagnetic label and the nucleus as a function of the relative intensity.

Estimation of the distance between nitroxide group and methyl groups

To better understand how the paramagnetic label on NS5A-D2 affects NS5B we used some approximations to estimate the relative intensity of the resonances and their dependence on the distance between the methyl groups and the paramagnetic label. This analysis was first described by Battiste and Wagner [381] based on the Solomon-Bloembergen equations and applied to HSQC pulse sequences.

The peak intensities of the paramagnetic and diamagnetic samples are roughly proportional to

$$I_{para} \approx e^{-R_2^* t} \quad \text{and} \quad I_{dia} \approx e^{-R_2 t} \tag{5.3}$$

where t is the total polarization transfer time in the pulse sequence. This protocol originally used the INEPT transfer time in the ^{1}H,^{15}N-HSQC sequence but has been successfully applied to methyl-HMQC spectra by using its total polarization transfer time [363]. In our PRE experiments, this transfer time is 6.9 ms.

The transverse relaxation rate of a resonance in the paramagnetic sample has the contribution of both the R_2 and the paramagnetic relaxation rate enhancement (Γ_2)

$$R_2^* = R_2 + \Gamma_2 \tag{5.4}$$

Γ_2 can be determined from the experimental data by solving Equation 5.5.

$$\frac{I_{para}}{I_{dia}} = \frac{R_2 e^{-\Gamma_2 t}}{R_2 + \Gamma_2} \tag{5.5}$$

Lapinaite *et al.* [364] extended this for ^{1}H,^{13}C-HMQC experiments by taking into account the transverse relaxation rates for both methyl protons (R_2^H) and the HC double quantum coherence (R_2^{HC}) in the diamagnetic state (Equation 5.6). In this analysis, the paramagnetic contribution to the methyl group relaxation was assumed to be the same for the protons and double quantum coherence.

$$\frac{I_{para}}{I_{dia}} = \frac{R_2^H R_2^{HC} e^{-\Gamma_2 t}}{(R_2^H + \Gamma_2)(R_2^{HC} + \Gamma_2)} \tag{5.6}$$

The relationship between the PRE rate and the distance between nuclear and electronic spins is given by Equation 5.7.

$$r = \sqrt[6]{\frac{K}{\Gamma_2}\left(4\tau_c + \frac{3\tau_c}{1+\omega_h^2\tau_c^2}\right)} \tag{5.7}$$

where r is the distance between the nucleus and the electron, ω_h is the Larmor frequency of the proton, K is a constant dependent on the nuclear gyromagnetic ratio, the electronic g factor and the Bohn magneton. This constant is $1.23 \times 10^{-32} cm^6/s^2$ for the relaxation of a proton enhanced by a nitroxide spin label. τ_c is the correlation time between the nuclear spin and the unpaired electron and depends on both the electron spin relaxation time τ_s and the protein rotational correlation time τ_r.

$$\frac{1}{\tau_c} = \frac{1}{\tau_s} + \frac{1}{\tau_r} \tag{5.8}$$

In our case, the estimation of the τ_c is not straightforward because we are observing the effect of a spin label that is attached to an IDP and detecting its effects on the nuclei

of a globular protein. We used two extreme cases: One where the nitroxide is on a completely flexible part of the protein with an estimated τ_c of 5 ns [386]. The other case we considered was that the spin label is on a part of the protein that directly bound to the polymerase. We combined the estimated τ_r of NS5B$_{\Delta 21}$ (about 40 ns) with the τ_s of the spin label (> 100 ns) [419]. Since τ_s is much longer than τ_r, τ_c is approximately the same as τ_r, so about 40 ns.

Comparison of the curves of relative peak intensity (I_{para}/I_{dia}) as a function of the distance of the paramagnetic label show that the distance affected by the spin label is dependent on the correlation time (Figure 5.12).

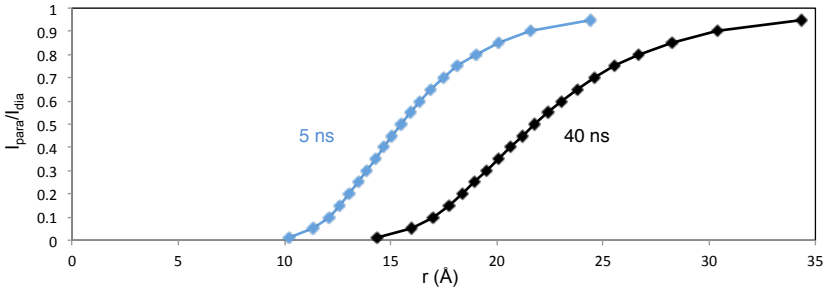

FIGURE 5.12: Simulated relative peak heights relative to the distance between the nucleus and the electron in the nitroxide label. Using Equations 5.6 and 5.7, the relative peak heights can be related to the distance to the spin label. The parameters used were a HMQC transfer time of 6.9 ms, a proton Larmor frequency of 900 MHz, a proton transverse relaxation rate of 38 s^{-1} and a double quantum coherence transverse relaxation rate of 104 s^{-1}.

To calculate these curves there are several approximations that may not strictly be valid. Firstly, values of τ_c calculated using the protein rotational correlation time may not be valid for methyl groups due to their rapid rotation. Secondly, single values of R_2^H and R_2^{HC} estimated from the linewidths in the spectrum were used, but to gain more exact values, R_2^H and R_2^{HC} would have to be determined for each resonance in a diamagnetic sample. Furthermore, the mobility of disordered NS5A-D2 is taken into account by using two extreme examples, where τ_c is that of a disordered protein and where it is the τ_c of NS5B$_{\Delta 21}$ but the true value of τ_c may not be represented by these cases.

These curves can still help us compare the proposed the binding sites of NS5A-D2 on NS5B$_{\Delta 21}$ with the experimental data. In particular, this can be done by comparing

the relative intensity of the resonance and the position of the residue in the crystallographic structures.

Deviations from calculated distances

The curves shown in Figure 5.12 can be used to obtain a range for the expected distance between the nucleus and the unpaired electron based on the experimentally measured relative intensities. These distances can help position the paramagnetic labels with respect to the crystallographic structure of NS5B$_{\Delta 21}$. However, deviations from distances that are expected based on the position of the spin label suggest either reorientation of the polymerase compared to its position in the crystallographic structure or movement of the paramagnetic label. PRE experiments can detect dynamics because transiently occupied conformations where the distances between the paramagnetic center and observed nucleus are small will be enhanced due to the dependence of the PRE on r^{-6}.

The paramagnetic label on NS5A-D2$_{C298}$ causes broadening of resonances of both I11 and I419. These relative intensities are 22% and 5% and suggest that these residues are 13.1 – 18.4 Å and 11.4 – 16.0 Å from the paramagnetic label, respectively. However, crystallographic structures show that these residues are at least 40 Å from each other. The midway point between them, approximately within the thumb-I inhibitor binding site, is about 20 – 30 Å away from each residue. This suggests that there is flexibility in the position of the spin label, because the distance values are smaller than expected. The spin label may be closer to I11 in one conformation and closer to I419 in another, causing this deviation. This can be the case because C298 is in a region of NS5A-D2 that does not display broadening with the addition of NS5B$_{\Delta 21}$ [108], so it is likely to remain mobile and just tethered, but not directly bound, to the polymerase.

The paramagnetic label on NS5A-D2$_{C254}$ causes less broadening on the resonance of I11 than when the spin label is on one of the other positions on NS5A-D2. Previously, we used this to orient the N-terminal region of NS5A-D2 preferentially toward the thumb-II site. C254 is within a binding site on NS5A-D2 that broadens in the presence of NS5B$_{\Delta 21}$ supporting that it should directly interact with the polymerase. However, despite I11 being over 40 Å away from the thumb-II site in the crystal structures, a distance where we do not expect to detect PREs, the resonance due to I11 displays only

66% of its intensity in the presence of the paramagnetic atom. Thus showing that either the N-terminal part also displays mobility and that multiple binding modes between the two proteins potentially exist or that I11 is closer to the thumb-II site than predicted by the crystal structures.

There are also changes in the conformation of the polymerase with respect to the crystallographic structures of apo NS5B that are exposed by this experiment. This becomes apparent when observing the effect of the paramagnetic probes on the resonance that corresponds to Ile 23. The presence of the spin label on C298 results in a relative intensity of 45 % which would correspond to a distance of 15.1 – 21.2 Å between this spin label and isoleucine 23. Since this residue is in the region of the fingertips Λ1 loop, in the crystal structures, that is in contact with the thumb-I site this distance seems incoherent with the binding of NS5A-D2 to this site. Also, the resonance of I11 displays greater or equal PREs with all positions of the paramagnetic label despite appearing farther than I23 from the expected binding sites. Yet, in crystallographic structures of NS5B bound to thumb-I inhibitors [143, 241] the electron density of this loop cannot be seen, consistent with it becoming flexible with the binding of the inhibitor to the site. Thus, the smaller PREs observed for I23, despite it appearing to be in the center of the region affected by PREs, lead us to propose that NS5A-D2 binding to the NS5B thumb-I site can cause the reduction of the contact between the thumb-I site and the fingertip loop, which would distance Ile 23 from the paramagnetic label.

Other resonances that support a perturbation of the fingertips Λ1 loop are those corresponding to I11 and I138. NOE contacts between these resonances (see Chapter 4) showed that these residues are close in unliganded NS5B$_{\Delta 21}$. The PREs observed on the resonance corresponding to I11 are much stronger than those observed on the resonance of neighboring I138. This could be explained by displacement of the Λ1 loop upon binding of NS5A-D2.

While we do seem to observe a tendency for a preferential orientation of NS5A-D2 binding, it is possible that more than one binding mode exists. Different binding modes may be in exchange, thus resulting in a fuzzy complex between NS5B$_{\Delta 21}$ and

NS5A-D2 [322]. This would not be the first time that NS5B would be described forming a fuzzy complex with a binding partner, as this has been described for the interaction of NS5B with the human vesicle-associated membrane protein-associated protein C (VAPC) [420]. Furthermore, the fact that we observe broadening of resonances like those that correspond to I419 and I432 without the appearance of resonances corresponding to the bound state of these residues in the titration experiment suggests that the binding may not be described as a single step and/or the complex remains flexible when bound.

Taken together, these issues result in a large difficulty to model the complex using these distance restraints as both proteins display flexibility not represented by typical docking programs.

For IDPs, an ensemble approach can be used with PRE restraints. In this approach, the conformational space is extensively sampled to construct an ensemble of structures based on the amino acid sequence. Then, using experimental restraints like PRE, chemical shifts and/or RDCs; structures within the ensemble that fit the data are selected [386, 421]. In another approach, the structure of a protein-protein complex containing a flexible peptide has been calculated using PREs by first determining the structure of the bound peptide then fitting that to the other protein with intermolecular PRE restraints [363]. Our case, however, would need a combination of both approaches. An ensemble of conformations of NS5A-D2 that correspond to the intermolecular PRE restraints would need to be selected but the flexibility that $NS5B_{\Delta 21}$ displays in solution would also need to be included in the model.

These paramagnetic resonance enhancement NMR experiments supported that the thumb-II site is a direct interaction site, as opposed to being affected by allostery. This site appears to interact preferentially with the N-terminal binding site on NS5A-D2. These experiments also identified thumb-I as an additional binding site that could not be discerned with chemical shift perturbations alone. We cannot determine which of the three binding sites on NS5A-D2 bind(s) to the thumb-I site on $NS5B_{\Delta 21}$. These experiments highlighted the flexible nature of this interaction. Thus, it is possible that $NS5B_{\Delta 21}$ and NS5A-D2 may form a fuzzy complex with multiple binding modes.

[NS5A-D2] (μM)	T_m (°C)	% bound NS5B
0	41.4	0
30	38.5	24.5
60	36.7	40.6
180	36.1	69.5

TABLE 5.2: Melting temperatures of $NS5B_{\Delta 21}$ with the addition of NS5A-D2. These values are the inflection points of the curves in Figure 5.14 determined by fitting a Boltzmann model (Equation 3.5). The addition of NS5A-D2 reduces the thermal stability of NS5B and causes a lowering of the melting temperature.

5.6 Differential scanning fluorimetry

When mixing NS5A-D2 with $NS5B_{\Delta 21}$ we observe small quantities of white precipitate that form in the NMR tube. Thus, we were interested in the effect of NS5A-D2 on the thermal stability of $NS5B_{\Delta 21}$. To investigate this, we used differential scanning fluorimetry (DSF), also known as thermofluor. DSF can be used to measure the temperature at which a protein unfolds by measuring the increase of fluorescence of a dye like SYPRO orange. The dye has affinity for hydrophobic residues that are exposed during unfolding and denaturation [422]. The melting temperature of the investigated protein using this technique is defined as the inflection of the fluorescence curve under increasing temperature and can be determined by fitting a Boltzmann sigmoidal curve between the minimum and maximum fluorescence. The fluorescence reaches a maximum when hydrophobic residues become exposed and then decreases due to aggregation.

DSF is often used to determine the variation in melting temperature due to binding of a small-molecule ligand to a structured protein [423]. An increase in melting temperature suggests an increase in protein stability and, conversely, a decrease indicates lower protein stability.

In this study, we are detecting the interaction between two proteins which could potentially both contribute to the fluorescence increase. Therefore, we tested whether NS5A-D2 alone presented a thermal denaturation transition that could be detected by the probe. We used NS5A-D2 at 180 μM in the same buffer used in the following study with $NS5B_{\Delta 21}$. No melting transition that could directly interfere with the interaction study was detected (Figure 5.13).

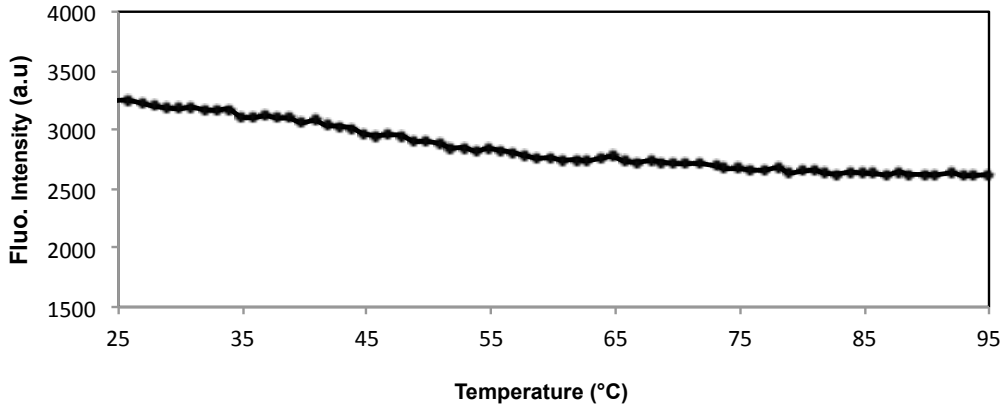

FIGURE 5.13: NS5A-D2 does not present a melting transition between 25-95°C as no sharp increase in fluorescence of SYPRO Orange is detected during the temperature ramp. We can assume that an effect NS5A-D2 has on NS5B melting temperature is thus not due to a melting transition of its own.

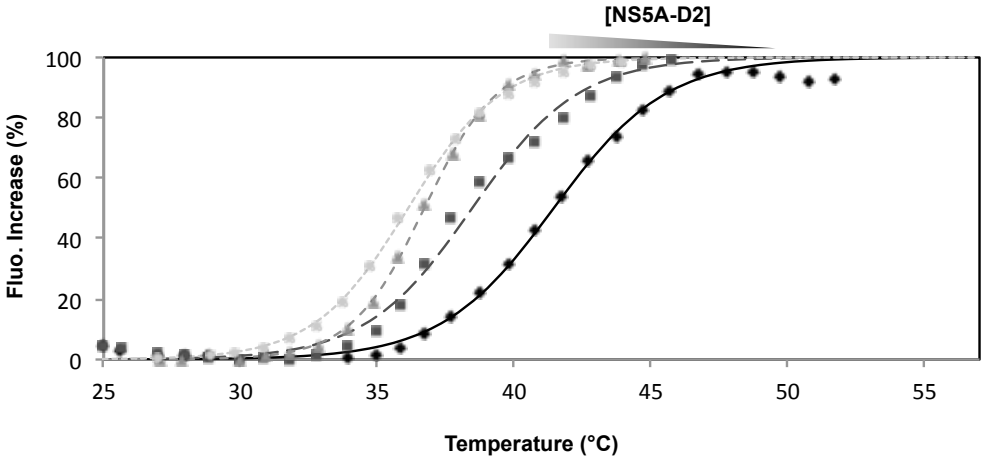

FIGURE 5.14: Effect of NS5A-D2 on $NS5B_{\Delta 21}$ thermal stability. NS5A-D2 was added to 30 μM $NS5B_{\Delta 21}$ at the concentrations of 0 μM (black, $\Diamond$), 30 μM (dark gray, $\square$), 60 μM (medium gray, $\triangle$), 180 μM (light gray, $\bigcirc$). The addition of NS5A-D2 causes the reduction of the melting temperature of $NS5B_{\Delta 21}$, represented by the inflection point of the curves.

The effect of NS5A-D2 on the stability of NS5B was investigated with the addition of 0 – 180 μM of NS5A-D2 to 30 μM $NS5B_{\Delta 21}$ This addition causes a dose-dependent shift to lower values of melting temperature (Figure 5.14 and Table 5.2). A decrease in apparent melting temperature observed by DSF means that the hydrophobic residues of the protein are exposed at a lower temperature meaning that the protein is destabilized by the binding of the ligand [423]. Thus, NS5A-D2 can decrease the stability of $NS5B_{\Delta 21}$.

The reduction of the melting temperature of $NS5B_{\Delta 21}$ observed in the presence of NS5A-D2 is in line with the decrease in thermal stability observed in the presence of

thumb-I inhibitors and with C-terminal deletions of NS5B that were described by Boyce *et al.* [408]. Thus, a lowering of the melting temperature could be expected if NS5A-D2 displaces the fingertips Λ1-loop as a thumb-I inhibitor or results in a less stable open conformation like the C-terminal deletions. In contrast, it was shown that thumb-II inhibitors increase the thermal stability of $NS5B_{\Delta 21}$ (next chapter and [408]).

Unfortunately, this method cannot be used to estimate an affinity constant for the interaction because the affinity constant is a function of temperature.

5.7 Conclusion

In this chapter the interaction between NS5A-D2 and NS5B was investigated by NMR spectroscopy and differential scanning fluorimetry. Interactions between NS5A-D2 and NS5B had been already studied by NMR using ^{15}N-labeled NS5A-D2 and had shown three interaction sites on the intrinsically disordered protein.

Using selectively isoleucine-labeled $NS5B_{\Delta 21}$, direct interaction by NMR shows broadening and chemical shift perturbations throughout the thumb subdomain but most intensely at isoleucine 419 close to the thumb-II inhibitor site. Based on chemical shift perturbations of the most perturbed resonances, we estimated an affinity constant of 70 μM. This value is higher than had been determined by surface plasmon resonance [108] because in the second case NS5B is immobilized, resulting in an different apparent affinity.

To discriminate between a direct interaction and an allosteric effect we used paramagnetic relaxation enhancement. By labeling three positions on NS5A-D2 we observe broadening of resonances in the methyl-HMQC spectrum of $NS5B_{\Delta 21}$ within 20-25 Å of each of the paramagnetic labels. The N-terminal paramagnetic label causes the largest broadening on I419, which corresponds to the thumb-II site. The center paramagnetic label causes the largest broadening on I432 which corresponds to the thumb-I site. Finally, the C-terminal paramagnetic label causes the same broadening as the center site with lesser effect on the intensity, suggesting it is free in solution. The proposed preferential orientation is shown in a tentative model in Figure 5.15. Furthermore, our

data suggest that this complex is dynamic, possibly forming what is called a "fuzzy complex".

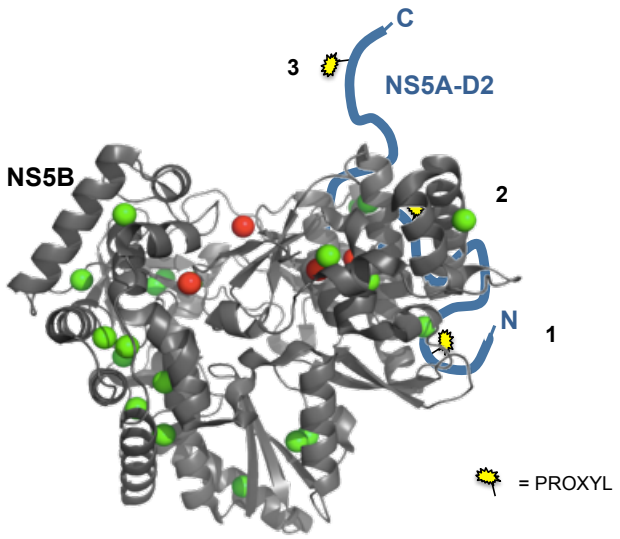

FIGURE 5.15: Tentative model of the binding of NS5A-D2 based on PRE data. $NS5B_{\Delta 21}$ is shown as a cartoon representation colored as in Figure 4.9. NS5A-D2 is shown as a blue line with its N- and C-terminals labeled, the position with respect to NS5B is to show the preferential orientation observed in PRE data, however the conformation is arbitrary to represent its intrinsic disorder. Spin label positions are shown in yellow. **1** The broadening profile of the paramagnetic label closest to the NS5A-D2 N-terminal (C_{254}) suggests that this part of the protein binds to the thumb-II site on NS5B. **2** The broadening profile of the paramagnetic label at the center of NS5A-D2 (C_{298}) suggests that this part of the protein binding to the thumb-I site on NS5B. **3** Since the broadening profile of the paramagnetic label closest to the C-terminal (C_{338}) resembles that of C_{254} with less broadening would suggest that this part of the protein is flexible and not bound to the polymerase.

Other authors have already attempted to map this interaction. Qin *et al.* [424] used alanine mutations and pull-down experiments to identify 4 regions whose mutation abolishes the interaction. These sites are on the back of the thumb at positions 139–145, 149–155, 365–371 and 382–388. These mutated regions correspond to the fingertips loop Λ2, and hinge region between the palm and thumb. While these sites do not correspond to the binding sites we observed, these results do not preclude our own. NS5A-D2 bound to both thumb sites could be wrapped around the thumb subdomain in a pattern that resembles the regions they described. Indeed, we observe a chemical shift perturbation of the resonance of I363, an isoleucine residue within the mutated zones they proposed.

Ngure *et al.* [94] proposed a binding site by mapping basic residues on the surface of NS5B exposed to protecting agents in the absence and presence of NS5A-D2. They

proposed that NS5A-D2 binds to the NS5B RNA-binding groove because most of the residues that became protected with NS5A-D2 binding were located there. Our results by PRE, however, are not in agreement with this result. If NS5A-D2 bound to the RNA-binding groove we would expect a general broadening of the residues upto 25 Å from the RNA-binding groove, which includes residues in the fingers subdomain that are not broadened. Additionally, the authors mutated arginine residues that they had identified and tested their NS5A-D2 binding. Interestingly, the mutation that most decreased NS5A-D2 binding was R380E, which is at the hinge between the thumb and palm subdomains, about 10 Å below the thumb-II site. They did not have any probes near the thumb-I site. The binding site they observe may be due to a conformational change brought on by NS5A-D2 that decreases the exposure of the RNA-binding groove to the protecting agents.

In sum, we propose that NS5A-D2 binds to two allosteric binding sites on the thumb subdomain, possibly with multiple binding modes. These sites have been described as having different effects on NS5B conformational dynamics and stability [243, 408]. This may contribute to the seemingly contradictory results that show that NS5A-D2 reduces polymerase activity [94] but is necessary for RNA replication [87].

Effect of a Thumb-II inhibitor on NS5B

6.1 Introduction

In the previous chapter we showed that NS5A-D2 affects the thumb-II inhibitor site. Based on this, the goal of this chapter is two-fold. Firstly, we want to investigate in solution the effect of a small molecule binding on NS5B to its thumb-II site. Secondly, we want to compare the binding of the small molecule to NS5B with the binding of NS5A-D2.

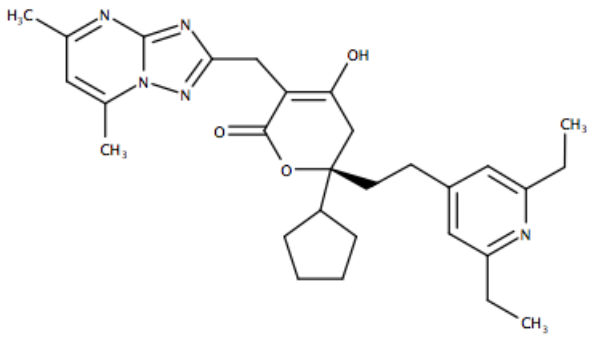

FIGURE 6.1: Chemical structure of Filibuvir.

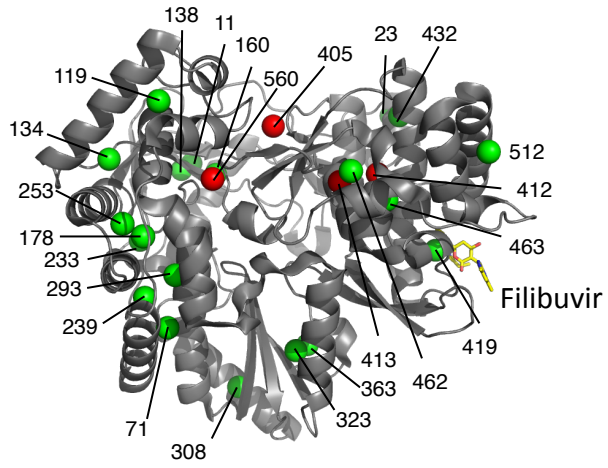

FIGURE 6.2: Filibuvir binds to the thumb-II allosteric inhibitor binding site. Crystallographic structure of NS5B (PDB 3I5K) in cartoon representation with isoleucine $\delta 1$ carbons shown as spheres and colored as in Figure 4.9. Filibuvir, shown in yellow sticks, was positioned by alignment with PDB 3FRZ [236]. Filibuvir is located in the pocket that contains I419.

Filibuvir (Figure 6.1), previously known as PF-00868554, the small molecule inhibitor chosen for this study, is a dihydropyranone which has reached phase 2 trials [425] but development was halted for strategic reasons [426]. Its binding site on NS5B was mapped to the NNI-2 site, which contains I419. Figure 6.2 shows the location of the binding site of this inhibitor on NS5B. This inhibitor does not a cause significant change in conformation of the polymerase in the crystal structure [236].

There are two main hypotheses for the mode of action of thumb-II inhibitors like filibuvir. The first hypothesis is that has been suggested that the inhibitors act by blocking the polymerase in an inactive conformation that is incapable of *de novo* initiation [158], possibly in a closed conformation. Another hypothesis proposes that thumb-II inhibitors prevent a change in conformation between initiation- and elongation-compentent forms of the polymerase [243].

However, few experiments have been done in solution to investigate these hypotheses. Deredge and colleagues [248] have recently shown that the binding of filibuvir and other thumb-II inhibitors significantly slows hydrogen/deuterium exchange, mainly in the fingers and thumb subdomain, compared to apo NS5B. The authors propose that this effect is due to a loss of conformational dynamics that correlates with inhibitor potency.

Firstly, we will map the effect of filibuvir in solution using NMR which gives information on both direct and allosteric effects. In particular, we are interested in the chemical shift perturbations that occur in resonances that are distant from the binding site as these can provide insight into an allosteric communication pathway. We will also measure chemical exchange in a millisecond to microsecond timescale to investigate whether the inhibitor affects the dynamics of the polymerase in this timescale. Finally, we will compare the effects of filibuvir on $NS5B_{\Delta 21}$ to those observed in the presence of NS5A-D2.

6.2 Mapping the interaction on NS5B

To investigate the effect of the binding of Filibuvir to NS5B, the inhibitor was added to labeled $NS5B_{\Delta 21}$ and the changes in the spectra were observed. We avoided the

use of an organic co-solvent for filibuvir, as this can significantly affect NS5B$_{\Delta 21}$. The widespread effect of dimethyl sulfoxide as a co-solvent had been observed by our group previously [108]. Thus, due to its low solubility in water and high affinity for NS5B (K_D of 27 nM) [246] an NS5B-filibuvir complex was obtained by the addition of an excess of solid filibuvir and incubation overnight at 15°C and the excess was removed by centrifugation or filtration at 0.2 μM. This strategy had been effectively applied by our group to investigate the high affinity complex between CypA and CsA [427].

6.2.1 Mapping the interaction with methyl-HMQC spectra

We also used the assigned isoleucine methyl-HMQC spectrum to investigate the binding of filibuvir to NS5B$_{\Delta 21}$. A titration experiment was not acquired with filibuvir, instead, filibuvir was added in excess as a powder, resulting in a spectrum of NS5B$_{\Delta 21}$ saturated with the inhibitor. The comparison of the spectrum of apo NS5B$_{\Delta 21}$ and filibuvir-bound NS5B$_{\Delta 21}$ is shown in Figure 6.3. I419, I432 and I512 all display chemical shift perturbations larger than 0.05 ppm. Of these, only I419 is close to the binding site (about 3.5 Å). I432 and I512 are about 16 and 18 Å from the expected position of filibuvir. The others show that the long-range effects traverse the thumb subdomain. Additionally, I23, I71 and I160 in the fingers; I308, I323 and I363 in the palm; and I463 in the thumb all have CSPs larger than 0.01 ppm. This shows that the effect of filibuvir is not limited to the thumb subdomain as it perturbs resonances in all three subdomains.

These effects throughout the polymerase are consistent with filibuvir acting as an allosteric inhibitor in which the binding at the NNI-2 site causes a change in conformation or dynamics of the polymerase in solution. Furthermore, it is notable that I419 and I432 are on the same α-helix, called Q in the nomenclature described by Bressanelli *et al.* [137]. Since I432 is in the NNI-1 site, where the fingertips Λ1 loop makes contact with the thumb subdomain it is possible that these structural elements, Q-helix and Λ1 loop, mediate the transmission of the perturbations from the thumb-II site through to the fingers subdomain.

The NNI-1 site is in contact with the apex of the Λ1 loop. Thus the perturbation of the this NNI site might be responsible for the transmission of the perturbation to the

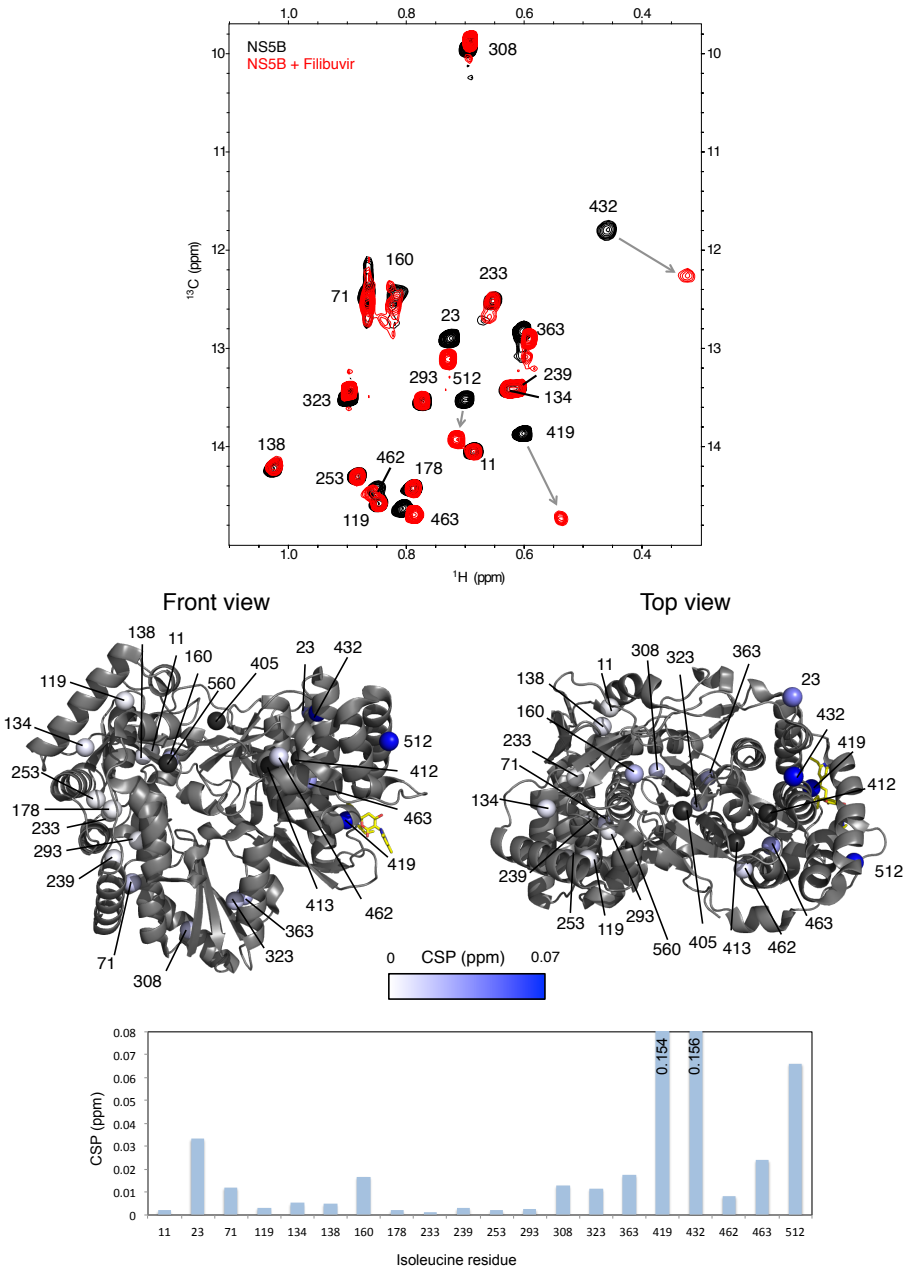

FIGURE 6.3: Methyl-HMQC spectrum of NS5B$_{\Delta 21}$ bound to filibuvir. **Top** Overlay of methyl-HMQC spectra of NS5B$_{\Delta 21}$ in the absence (black) and presence (red) of filibuvir. Filibuvir was added in excess so the spectrum corresponds to NS5B$_{\Delta 21}$ saturated with the inhibitor. **Center** Crystallographic structure of wild-type NS5B (PDB 3I5K) showing the chemical shift perturbations (CSP, blue) caused by the binding of filibuvir. Filibuvir is shown with sticks in yellow. The largest CSPs are in the thumb subdomain but the effect of filibuvir is transmitted throughout the three subdomains. **Bottom** Histogram of the CSP caused by the mutations. I419 and I432 have CSP values larger than the scale, therefore their CSP values are labeled.

fingers subdomain. Evidence for this is present in the perturbations of the resonances of I23 and I160 that may be affected by the motion of the fingertips loop.

Another structural element that may mediate the transmission of perturbations is the T α-helix. The closest part of this helix is less than 4 Å from the expected position

of filibuvir, the other side of this helix contains I512, a residue that displays chemical shift perturbation with this interaction.

We do not observe the apparition of new resonances. The resonances of isoleucines 405, 412, 413 and 560 remain broadened beyond detection, thus filibuvir does not stop the intermediate exchange that is responsible for the broadening of these resonances. I560 is in the C-terminal linker and I405, I412, I413 are near the β-loop. It has been shown that thumb-II inhibitors require the β-loop and C-terminal tail for their polymerase inhibitory activity [408]. However, without the detection of these four resonances we lack probes that can investigate this interplay. Deredge *et al.* [248] identified that filibuvir, unlike thumb-II inhibitors lomibuvir and GS-9669, did not appear to cause a rigidification in the C-terminal linker or the β-loop.

The minor peaks of I160 and I363 are still apparent in the spectrum in the presence of filibuvir. Therefore, the dynamics on a slow timescale that give rise to these minor peaks is still present with the inhibitor.

Based on the perturbations of filibuvir visible in the methyl-HMQC spectrum, we can observe that not only does the binding of the inhibitor cause perturbations around the thumb-II binding site and but also causes perturbations in isoleucines that are distant from the binding site. These perturbations may be mediated mainly by helices Q and T but also appear to depend on the fingertips loop Λ1.

6.2.2 Effect of filibuvir on NS5B open mutants

To continue the investigation of the effect filibuvir has on NS5B, we were interested in investigating the hypothesis that filibuvir stabilized a closed conformation of the polymerase. For this, we used two of the open mutants: NS5B$_{\Delta 21}$ L30S and NS5B$_{\Delta 21}$ 5mut (Figure 6.4).

If the mechanism that filibuvir uses to stabilize a closed conformation opposes the mechanisms that open these mutants we would expect to recover the changes caused by these mutations, that is, observe a spectrum without the changes caused by the apparent opening of these mutants. In particular, the three residues that disappear in the NS5B$_{\Delta 21}$ L30S methyl-HMQC should reappear if this were the case. However, in both mutants the effect of filibuvir is just summed to that of the mutants and we do

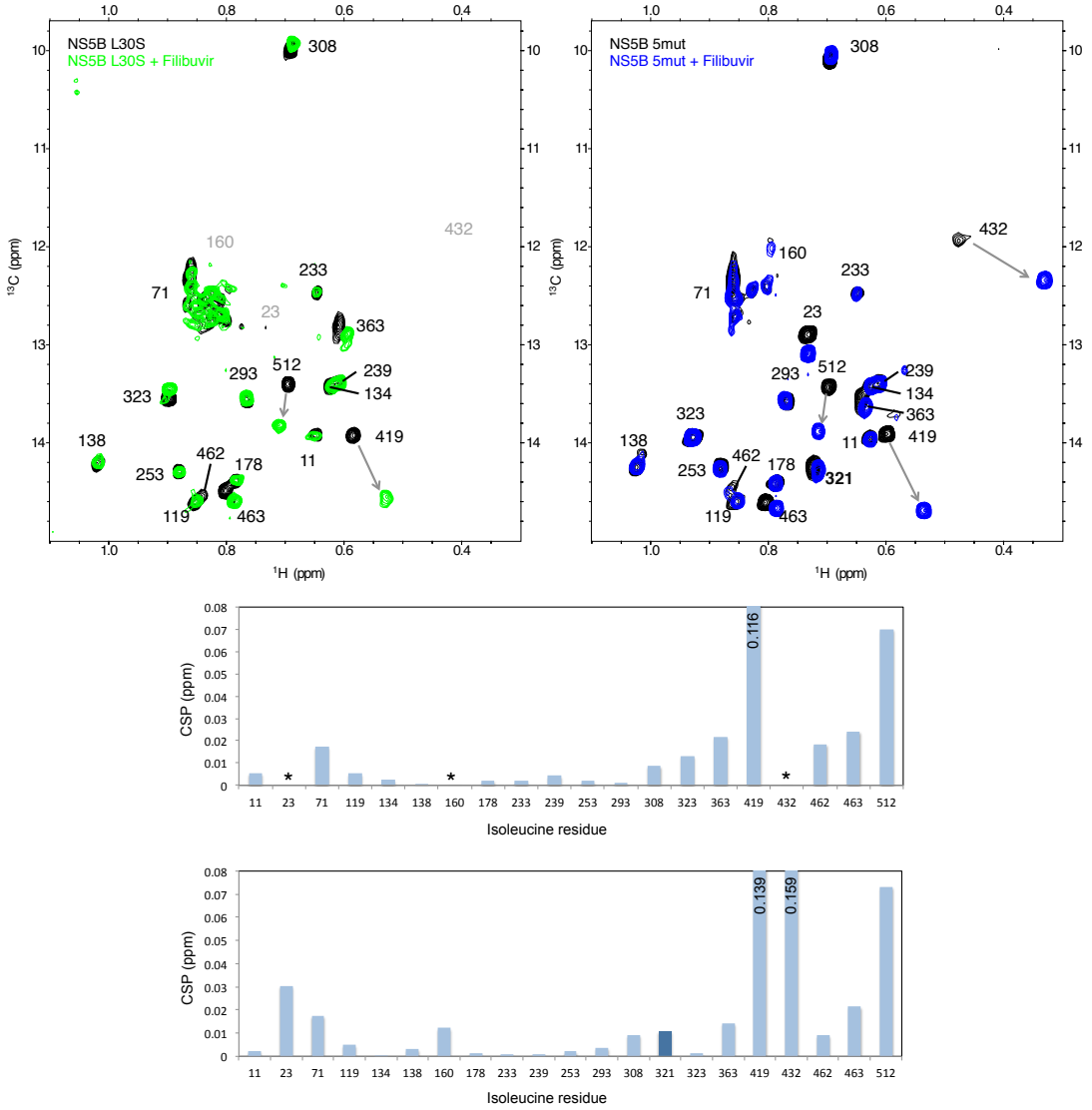

FIGURE 6.4: Effect of filibuvir on open mutants of NS5B$_{\Delta 21}$. Since the effect of filibuvir (seen in the wild-type) is just added to the effect of the mutations, these mechanisms seem independent. **Top left** Overlay of the methyl-HMQC of NS5B$_{\Delta 21}$ L30S in the absence (black) and presence (green) of filibuvir. Isoleucines broadened beyond detection that are visible in wild-type NS5B$_{\Delta 21}$ are labeled in gray. **Top right** Overlay of the methyl-HMQC of NS5B$_{\Delta 21}$ 5mut in the absence (black) and presence (blue) of filibuvir. All these spectra were acquired in RE buffer. **Center** Chemical shift perturbations caused by filibuvir on NS5B$_{\Delta 21}$ L30S. **Bottom** Chemical shift perturbations caused by filibuvir on NS5B$_{\Delta 21}$ 5mut.

not observe an opposing effect. This suggests that the stabilization caused by filibuvir is independent from the mechanisms triggered by these mutations or that an open structure is not incompatible with filibuvir binding.

Isoleucine	k_{ex} (s^{-1})	Δk_{ex} (s^{-1})	p_B (%)	Δp_B (%)	$\Delta\omega(^{13}C)$ (ppm)	$\Delta\Delta\omega(^{13}C)$ (ppm)	$\Delta\omega(^{1}H)$ (ppm)	$\Delta\Delta\omega(^{1}H)$ (ppm)	χ^2	Reduced χ^2
23	742	262	14.8	75.7	-0.23	0.51	–	–	10.015	0.910
71	1570	400	20.6	165.1	0.28	0.82	0.019	5.5	19.318	1.932
160	2266	920	1.3	0.1	3.58	0.41	–	–	6.114	0.556
								Total:	35.447	1.108

TABLE 6.1: Parameters obtained from fitting a two-state exchange model to the methyl-TROSY multiple quantum CPMG relaxation data of individual resonances of filibuvir-bound $NS5B_{\Delta 21}$. These values and their uncertainties were obtained using the cmpg_fitd9 software [397]. k_{ex}, exchange rate; p_B, population of the minor exchanging state; $\Delta\omega$, chemical shift change between states, χ^2, chi-squared goodness of fit measure. When ^{1}H $\Delta\omega$ are shown both ^{1}H and ^{13}C $\Delta\omega$ were allowed to vary, otherwise ^{1}H $\Delta\omega$ was set to zero. The choice between the two was based on smaller parameter uncertainties as the difference in goodness of fit was not significant.

6.3 Intermediate exchange in the presence of filibuvir

While the results with the open mutants suggest that an open structure is not incompatible with filibuvir binding, we were also interested in the effect filibuvir has on the dynamics of the protein. The four resonances that are broadened beyond detection with apo $NS5B_{\Delta 21}$ remain broadened thus exchange in an intermediate timescale appears to occur even in the presence of filibuvir.

Based on molecular dynamics simulations, Davis and Thorpe [428] proposed that the binding of thumb-II inhibitors reduced the overall flexibility of NS5B. Similarly, Deredge and coworkers observed decreased hydrogen/deuterium exchange [248] in NS5B in the presence of thumb-II inhibitors, especially in the thumb and palm subdomains. Thus, we were interested in using NMR spectroscopy to investigate the pertubation filibuvir causes on the dynamics of $NS5B_{\Delta 21}$

To investigate the effect filibuvir has on NS5B dynamics in the millisecond to microsecond timescale we used methyl-TROSY multiple quantum CPMG relaxation dispersion experiments. In Chapter 4 we investigated this dynamics of apo $NS5B_{\Delta 21}$ and fit the data to two-state exchange models. Here, we repeated this experiment in the presence of filibuvir.

By comparing the results (Figure 6.5) obtained for $NS5B_{\Delta 21}$ in the absence (Tables 4.7 and 4.8) and presence (Tables 6.1 and 6.2) of filibuvir we can observe that filibuvir perturbs the exchange occurring in this timescale. In apo $NS5B_{\Delta 21}$ 7 resonances displayed significant non-flat dispersion. The addition of filibuvir reduces this number to 3 resonances. The resonances that display non-flat dispersion with filibuvir are those

Isoleucine	k_{ex} (s^{-1})	Δk_{ex} (s^{-1})	p_B (%)	Δp_B (%)	$\Delta\omega(^{13}C)$ (ppm)	$\Delta\Delta\omega(^{13}C)$ (ppm)	$\Delta\omega(^{1}H)$ (ppm)	$\Delta\Delta\omega(^{1}H)$ (ppm)	χ^2	Reduced χ^2
23	1074	145	3.6	1.1	0.48	0.08	0.000	197.3	54.739	1.610
71					0.60	0.11	0.000	21.4		
160					3.07	0.63	-0.200	0.060		

TABLE 6.2: Parameters obtained from fitting a two-state exchange model to the methyl-TROSY multiple quantum CPMG relaxation data of filibuvir-bound NS5B$_{\Delta 21}$ globally. The non-flat dispersion curves were simultaneously fit to a single exchange rate. These values and their uncertainties were obtained using the cmpg_fitd9 software [397]. For the global fit, both ^{1}H and ^{13}C $\Delta\omega$ were allowed to vary.

that correspond to isoleucines 23, 71 and 160, all in the fingers submain. The residues that no longer display non-flat relaxation dispersion (Ile 233, 363, 432 and 462) suggest that there is a reduction of intermediate exchange dynamics in the thumb subdomain as Ile 432 and 462 are in this region and I363 is at the edge between palm and thumb.

When the relaxation dispersion profiles are fit globally to a single two-state model we observe a modest reduction of the exchange rate from 1309±135 to 1074±145 $s^{-1}$with the addition of filibuvir. However, like for the experiments with apo NS5B$_{\Delta 21}$, the reduced χ^2 goodness of fit parameter increases when using a global fit. Instead, an independent fit to a two-state exchange model shows that Ile 71 and 160 display slightly higher exchange rates compared to apo NS5B$_{\Delta 21}$ that are not statistically significant. On the other hand, the resonance of I23 displays a significantly slower exchange. As this residue is in the fingertips Λ1 loop, it is in contact with the thumb subdomain and may be detecting the reduced dynamics of this region.

The difference in chemical shift between both forms mostly are not well defined for the resonances of filibuvir-bound NS5B$_{\Delta 21}$. This does not help determine the identity of the minor form. Both the individual fit and global fit show a reduction of the population of the minor form p_B, but this population is not well defined in the individual-residue fit.

Altogether, this experiment supports the hypothesis that filibuvir reduces the dynamics of the polymerase. We observe, in particular, a reduction of the dynamics in the thumb and resonances of isoleucines surrounding it. The lower population of the minor form may also support a mechanism of inhibition where an active conformation is sampled less often. This experiment would also benefit from more methyl probes, as has been suggested for apo NS5B$_{\Delta 21}$, because this could help in obtaining more well

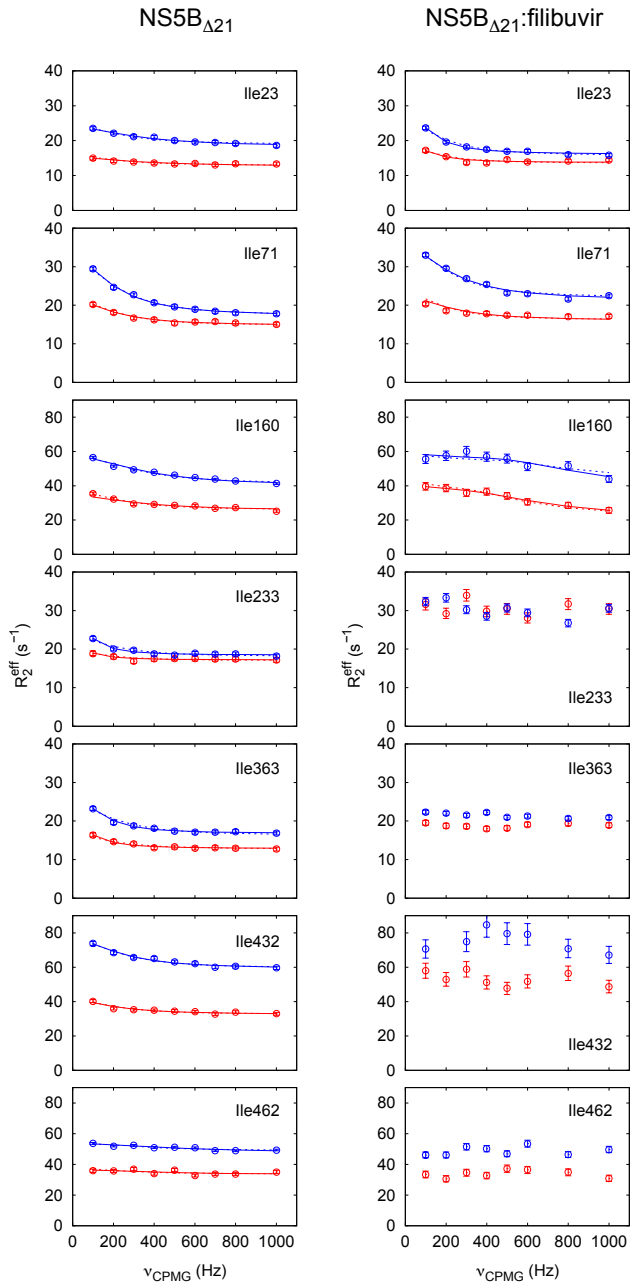

FIGURE 6.5: Methyl-TROSY multiple quantum CPMG relaxation dispersion curves of all NS5B$_{\Delta 21}$ isoleucine residues that display non-flat dispersion in apo NS5B$_{\Delta 21}$. Apo NS5B$_{\Delta 21}$ relaxation dispersion curves are shown on the left and filibuvir-bound NS5B$_{\Delta 21}$ relaxation dispersion curves are shown on the right. Full lines show the best fit to individual residues when significant exchange is present. Dashed lines represent best fit to all residues exhibiting dispersion of each state (apo or filibuvir-bound), respectively. Red and blue represent data acquired on 600 and 900 MHz spectrometers respectively.

defined parameters.

6.4 Comparison between filibuvir and NS5A-D2 binding

In Chapter 5 we identified that one of the binding sites of NS5A-D2 on NS5B$_{\Delta 21}$ is the NNI-2 site. Thus, we were interested in comparing the spectra of the polymerase bound to either filibuvir or the disordered protein.

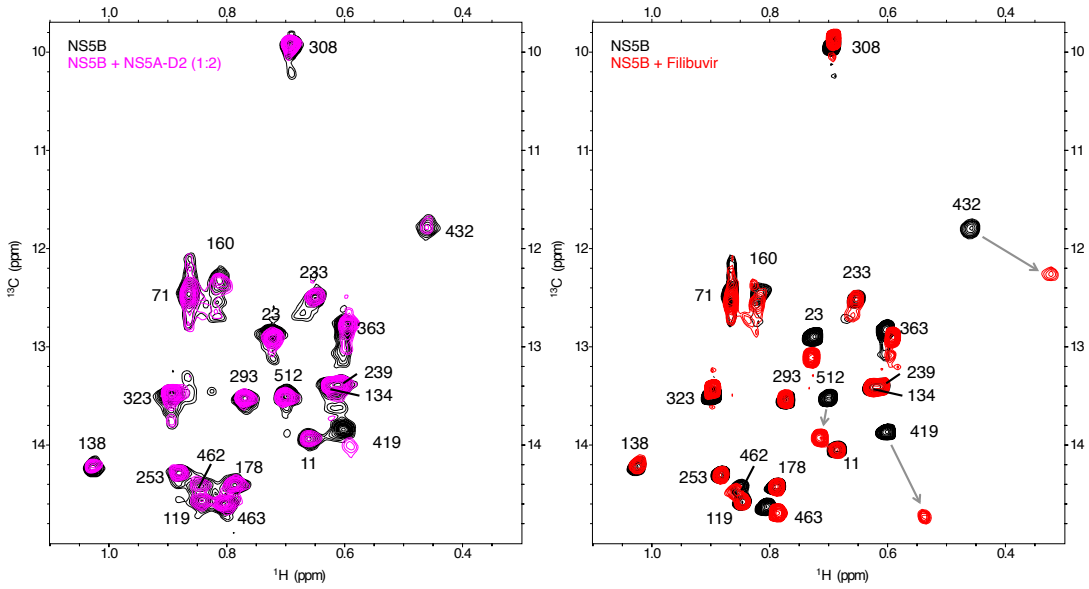

FIGURE 6.6: Comparison of methyl-HMQC spectra of isoleucine δ1-labeled NS5B$_{\Delta 21}$ in the presence of NS5A-D2 or filibuvir. **Left** Overlay of methyl-HMQC spectra of 40 μM NS5BΔ21 in the absence (black) and presence (magenta) of 80 μM NS5A-D2. **Right** Overlay of methyl-HMQC spectra of NS5B$_{\Delta 21}$ in the absence (black) and presence (red) of filibuvir (saturated). Some of the same resonances (I23,I419, I432) display either CSPs or broadening with the addition of the ligands.

Comparing the methyl-HMQC spectra of isoleucine-δ1-labeled NS5B$_{\Delta 21}$ bound to either NS5A-D2 or filibuvir, we observe that similar resonances are affected by both of the partners (Figure 6.6). We cannot directly compare the CSP values because NS5A-D2 is not saturated at the concentrations used in Figure 5.4 and the concentrations of NS5A-D2 needed to saturate NS5B$_{\Delta 21}$ cause it to precipitate. The resonances corresponding to I23, I419, I432 are affected by both ligands. However, the resonances corresponding to I512 is only affected by filibuvir. I419 is at the thumb-II site and on helix Q. I432 is also on helix Q and I23 is on the fingertips Λ1 loop facing helix Q. Thus all three of these resonances seem to be part of the same network that is perturbed by the binding of both ligands. However, I512 is on helix T and is only perturbed by filibuvir. Therefore, it appears that both partners have at least one allosteric network in

common and another one seems only to be affected by filibuvir.

6.4.1 Comparison of $^{1}H,^{15}N$-TROSY HSQC spectra

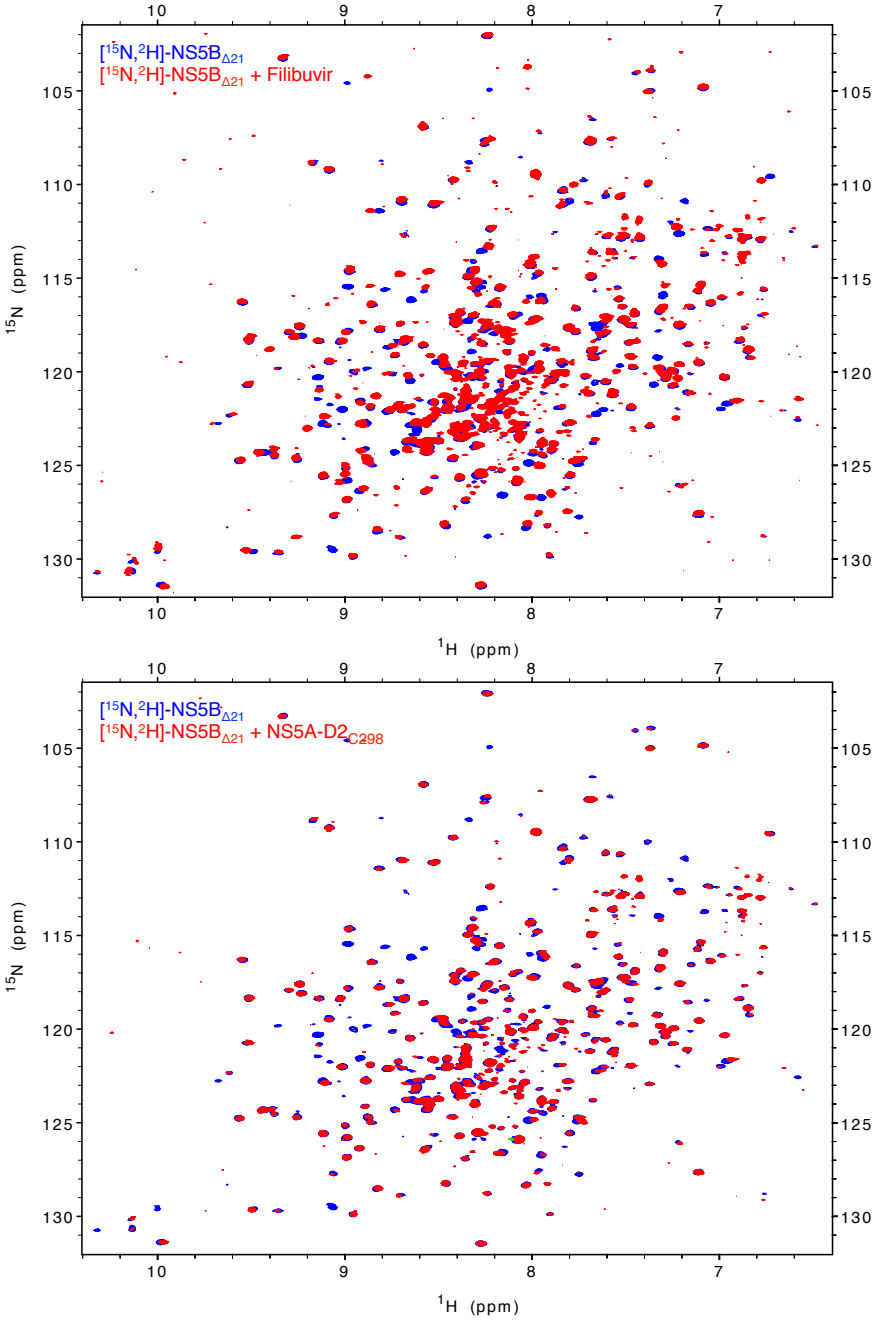

FIGURE 6.7: $^{15}N,^{1}H$-BEST-TROSY-HSQC spectra of $[^{2}H,^{15}N]$-NS5B with filibuvir and NS5A-D2 C338S. **Top** Overlay of apo $NS5B_{\Delta 21}$ (blue) and $NS5B_{\Delta 21}$ bound to filibuvir (red). Many chemical shift perturbations are visible. **Bottom** Overlay of apo $NS5B_{\Delta 21}$ (blue) and $NS5B_{\Delta 21}$ bound to an equimolar amount of spin-labeled NS5A-D2 C338S reduced (red). The main effects are the broadening of cross-peaks.

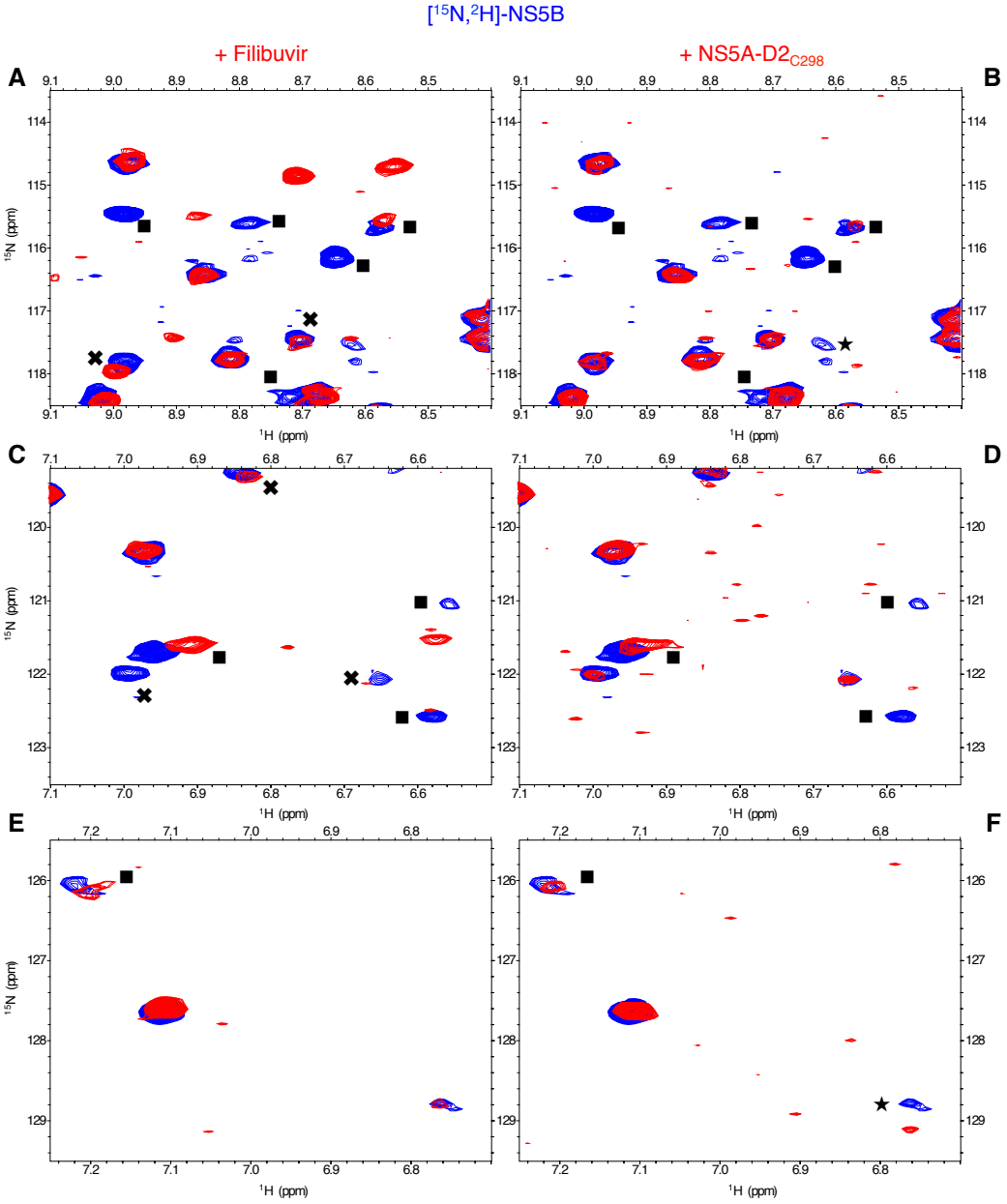

FIGURE 6.8: Enlargements of the ^{15}N,^{1}H-BEST-TROSY-HSQC spectra of [^{2}H,^{15}N]-NS5B with filibuvir and NS5A-D2. **A**, **C** and **E** show the overlay of NS5B$_{\Delta 21}$ in the absence (blue) and presence (red) of filibuvir (saturated). **B**, **D** and **F** show the overlay of NS5B$_{\Delta 21}$ in the absence (blue) and presence (red) of an equimolar amount of NS5A-D2$_{C298}$. Squares, crosses and a star are used to label cross-peaks that are perturbed with both filibuvir and NS5A-D2, only filibuvir and only NS5A-D2, respectively. Adapted from Rosnoblet *et al.* [108].

To extend our insight into the allosteric networks perturbed by NS5A-D2 and filibuvir we used the ^{1}H,^{15}N-TROSY HSQC spectrum of NS5B$_{\Delta 21}$. Ideally, the study of NS5B$_{\Delta 21}$ and its partners would be done with HSQC-type experiments in which we observe the amide protons in the protein and obtain information on the interaction at each backbone amide group, i.e. have a residue-level resolution. However, without the

assignment, the interpretation of the ^{1}H,^{15}N-HSQC of [^{2}H,^{15}N]-NS5B$_{\Delta 21}$ can only be by comparing spectra and the binding sites known by other techniques. We acquired BEST-TROSY ^{1}H,^{15}N-HSQC spectra to improve the linewidth and signal-to-noise ratio compared to HSQC spectra. These spectra were acquired on NS5B$_{\Delta 21}$ apo, bound to filibuvir, bound to nitroxide-labeled NS5A-D2$_{C298}$ and reduced nitroxide-labeled NS5A-D2$_{C298}$. This permits a comparison of the effect of filibuvir with that of NS5A-D2.

Since we do not have an assignment for the resonances in the ^{1}H,^{15}N-BEST-TROSY-HSQC of NS5B$_{\Delta 21}$ the 386 resonances visible in this spectrum (this count is higher than in Chapter 4 because there we included only resonances that had peaks in the 3D spectra) were arbitraily labeled. To assign these labels in the interaction spectra, if a peak from the apo spectrum is not present, we looked for the closest resonance that is not in a position of another apo peak. For some peaks in the apo spectrum a resonance could not be identified in the interaction spectrum.

In the spectrum of NS5B$_{\Delta 21}$ bound to filibuvir, 29 of the resonances from the apo spectrum could not be attributed to peaks in the interaction spectrum. This means they are either broadened beyond detection or too shifted to be identified in the spectrum of NS5B$_{\Delta 21}$ bound to filibuvir (Figure 6.7, top). Since there are only 4 resonances in the NS5B$_{\Delta 21}$-filibuvir spectrum that did not correspond to resonances in the apo spectrum, it means that despite NS5B being saturated with filibuvir, some resonances still broaden beyond detection. Among the peaks that appear to display CSPs, 57 resonances have ^{15}N,^{1}H combined chemical shift perturbations larger than 0.03 ppm.

In comparison, in the ^{1}H,^{15}N-TROSY HSQC spectrum of NS5B$_{\Delta 21}$ with NS5A-D2$_{C298}$, 26 could not be attributed to peaks in the interaction spectrum and only 9 are shifted beyond 0.03 ppm compared to apo NS5B$_{\Delta 21}$. NS5A-D2$_{C298}$ causes mainly partial broadening of the peaks in the spectrum (Figure 6.7, bottom). This difference is probably mainly due to the fact that NS5B was saturated with filibuvir in the sample containing the inhibitor and only 38% of NS5B$_{\Delta 21}$ was expected to be bound to NS5A-D2 at the concentrations used in the sample containing the disordered protein.

Many of the same resonances are affected (either not having a corresponding resonance in the interaction spectrum, displaying a CSP of more than 0.01 ppm or a relative

intensity of less than 30% of the apo NS5B$_{\Delta 21}$ spectrum) by both ligands (Figure 6.8) (101 resonances or 26.2% of the total resonances the spectrum). This suggests that they not only bind to the same site but also cause the same allosteric effects. This is somewhat surprising because filibuvir is a small-molecule inhibitor with a specific binding site on the thumb subdomain. NS5A-D2, on the other hand, is a 95-residue protein with about half its length involved in its interaction with NS5B$_{\Delta 21}$. This highlights the allosteric effect of filibuvir that can cause long-range perturbations on the NS5B$_{\Delta 21}$. Furthermore, we observe more resonances that are only affected by filibuvir (91 resonances or 23.5% of all resonances) compared to resonances that are only affected by NS5A-D2 (25 resonances or 6.5% of all resonances). This is in agreement with what was observed using selective methyl labeling, that showed that there are allosteric pathways (i.e. helix T) only perturbed by filibuvir but not by NS5A-D2.

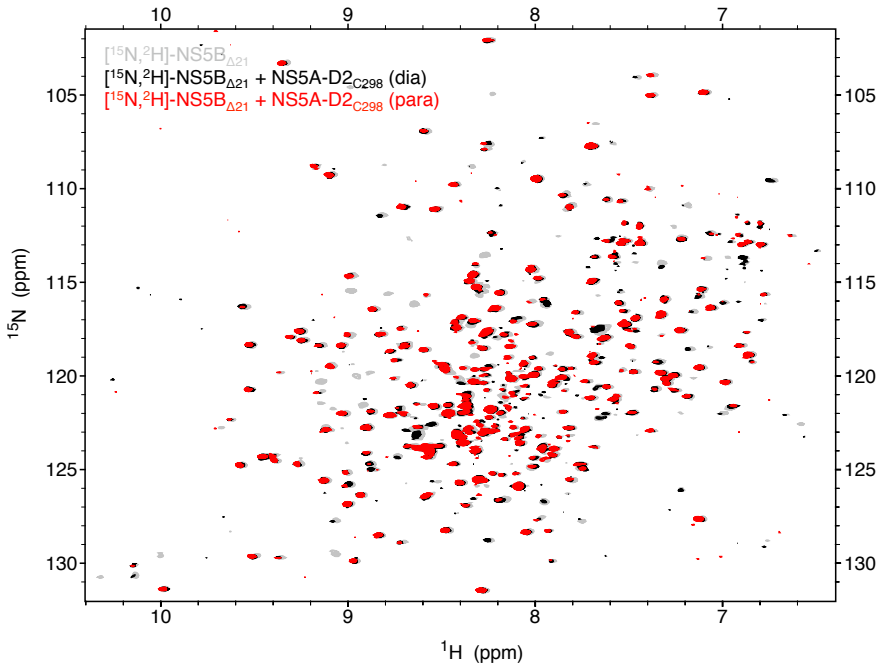

FIGURE 6.9: Overlay of ^{1}H,^{15}N-BEST-TROSY-HSQC spectra of NS5B$_{\Delta 21}$ apo (gray) and in the presence of spin-labeled NS5A-D2$_{C298}$ both diamagnetic (black) and paramagnetic (red).

Furthermore, we used paramagnetic relaxation enhancement experiments (Figure 6.9) to aid in the discrimination between direct effects and very long-range perturbations caused by NS5A-D2. The resonances with largest CSPs due to NS5A-D2 (above 0.03 ppm) are broadened by the spin label. However, there are peaks with CSPs above

0.01 ppm that are not broadened by the paramagnetic label. This may be due to allosteric effects but also binding to $NS5B_{\Delta 21}$ of a part of NS5A-D2 that is distant from the spin label. Thus we cannot ascertain with this PRE experiment alone the presence of allosteric effects.

Altogether, many of the resonances that are affected by NS5A-D2 binding are also affected by filibuvir, this could be used to extrapolate that they bind to the same site and induce similar allosteric effects. However, due to the large amount of peaks that shift in the presence of filibuvir we cannot discriminate between direct binding of the small molecule and indirect perturbations with this experiment alone. A way to discriminate between direct effects and allosteric effects could be to use a NOESY spectrum to determine which resonances on from $NS5B_{\Delta 21}$ are in close proximity to filibuvir.

6.5 Effect of filibuvir on NS5B binding to NS5A-D2

In the previous chapter, we proposed that one of the binding sites of NS5A-D2 on $NS5B_{\Delta 21}$ is at the thumb-II site. Filibuvir binds to this site and apparently cause a widespread effect on $NS5B_{\Delta 21}$ and therefore we are interested in investigating whether its binding would induce a change in conformation or dynamics that would interfere with the interaction between NS5A-D2 and the polymerase. Filibuvir has significantly better affinity for NS5B compared to NS5A-D2 (27 nM compared to 70 μM). Therefore, by saturating $NS5B_{\Delta 21}$ with filibuvir prior to the addition of NS5A-D2, the disordered protein should not be able to efficiently compete with the small-molecule inhibitor for its binding site.

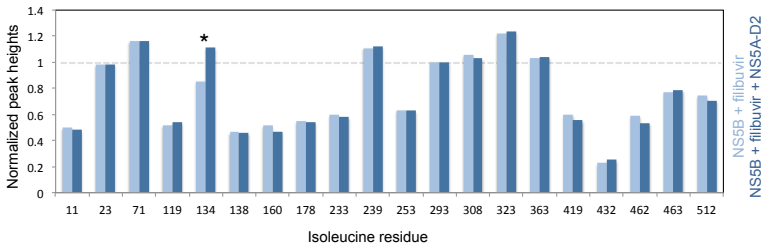

FIGURE 6.10: Effect of NS5A-D2 on $NS5B_{\Delta 21}$ bound to filibuvir. The normalized (with respect to I293) data heights for $NS5B_{\Delta 21}$ bound to filibuvir in the absence (light blue) and presence (dark blue) of NS5A-D2. The only difference is observed on I134 (labeled with an asterisk).

The addition of NS5A-D2 to NS5B$_{\Delta 21}$ previously loaded with filibuvir gives rise to a spectrum with little difference compared to NS5B$_{\Delta 21}$ and filibuvir exclusively (Figure 6.10). The only difference is on the peak height of I134 and since this isoleucine is far from the binding sites of either NS5A-D2 or filibuvir and is overlapped with I239 it is possible that this effect is due to the overlap. Based on this, it is not possible to be certain that NS5A-D2 still interacts with NS5B$_{\Delta 21}$ in the presence of filibuvir, especially since the lack of probes due to the selective labeling might be obscuring the effect.

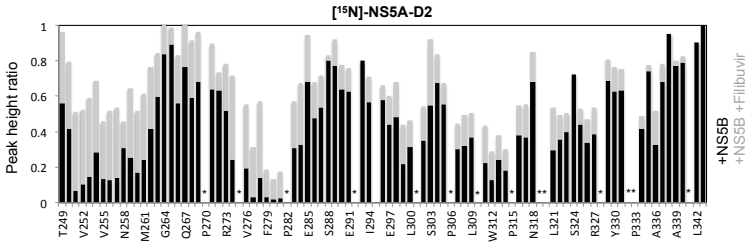

FIGURE 6.11: Effect of filibuvir on NS5A-D2 binding to NS5B$_{\Delta 21}$. Normalized peak heights obtained from the ^{1}H,^{15}N-HSQC of ^{15}N-labeled NS5A-D2. A value of 1 corresponds to the peak intensity of ^{15}N-NS5A-D2 alone. Bars represent the normalized peak height of NS5A-D2 in the presence of NS5B$_{\Delta 21}$ in the absence (black) and presence (gray) of filibuvir. Filibuvir reduces but does not abolish the interaction between NS5A-D2 and NS5B$_{\Delta 21}$. The positions of proline residues in the primary sequence are represented by asterisks.

To ascertain whether filibuvir reduces the interaction between NS5A-D2 and NS5B the converse experiment was also done (Figure 6.11). By adding filibuvir-bound NS5B$_{\Delta 21}$ to ^{15}N-labeled NS5A-D2 we could observe in the ^{1}H,^{15}N-HSQC that the NS5A-D2 still interacts with the polymerase, but a in lesser extent. Filibuvir, however, interferes with the binding between the proteins. Interestingly, the broadening of the resonances in the three binding sites is still visible and for the second and third sites only slightly reduced by the presence of filibuvir. Thus filibuvir does not completely abolish the interaction between the proteins. The intensity of the peaks of the N-terminal binding site however is increased from 20% to 60% of the intensity of NS5A-D2 alone. This can be interpreted as filibuvir directly blocking the binding of NS5A-D2 to the thumb-II site. The decreased interaction mainly in NS5A-D2's N-terminal site further supports the hypothesis presented in Chapter 5 that the N-terminal site of NS5A-D2 preferentially binds to the thumb-II site on NS5B$_{\Delta 21}$.

6.6 Differential Scanning Fluorimetry

DSF was used as in the previous chapter to show the effect of an interaction on the $NS5B_{\Delta 21}$ thermal stability. Here we observe the effect previously described by Boyce *et al.* [408] in which the thumb-II inhibitors cause an increase of the thermal stability of NS5B (Figure 6.12).

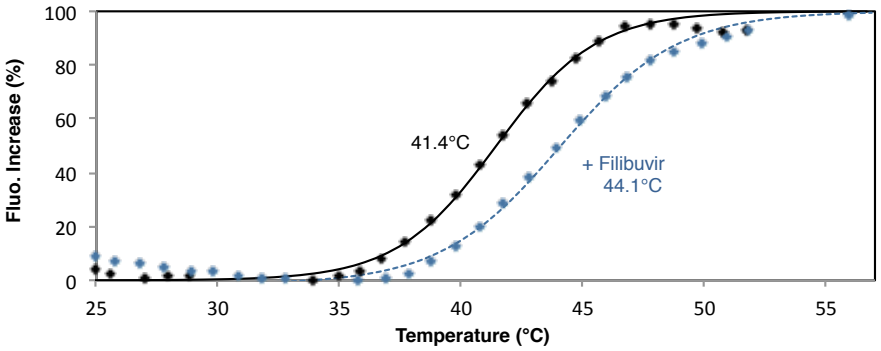

FIGURE 6.12: Effect of Filibuvir on $NS5B_{\Delta 21}$ thermal stability. Filibuvir was incubated with $NS5B_{\Delta 21}$ to saturate the polymerase prior to the measurement (blue curve) and compared to apo $NS5B_{\Delta 21}$ (black curve). The addition of the inhibitor increases of the melting temperature of $NS5B_{\Delta 21}$, represented by the inflection point of the curves.

We measured a melting temperature of 41.4°C for $NS5B_{\Delta 21}$ alone, lower than the temperature described by Boyce and colleagues. This result can be due to the different buffers used for these experiments, in particular in their work they used 1.5% DMSO to act as a co-solvent for the inhibitors which we avoided by saturating with filibuvir powder. Additionally, the authors observed a 3.4°C increase in melting temperature in the presence of filibuvir while we observed a smaller increase of 2.7°C which is within a 1°C error (1°C was the temperature step used between consecutive fluorescence measurements).

This increase in thermal stability due to a thumb-II inhibitor is typically expected in the case of a small molecule that binds a to a protein [423]. This is in contrast with the decrease observed in the presence of a thumb-I inhibitor Boyce *et al.* reported and in the presence of NS5A-D2 that we observed. To observe a decrease in thermal denaturation temperature upon binding, ligands must trigger a partially unfolded configuration of the polymerase that exposes more of its hydrophobic cores to the fluorescent probe.

6.7 Conclusions

In this chapter we investigated the effect of filibuvir binding on NS5B$_{\Delta 21}$. Recently, Li and Johnson [249] showed that filibuvir and other thumb-II site allosteric inhibitors do not inhibit initiation, the binding of the template RNA and the polymerization of the first 2-5 nucleotides, nor the elongation step, the processive synthesis of complementary RNA. Thus, the authors proposed that the inhibition observed for these molecules occurs by reducing the conformational transition that is necessary to go from initiation to elongation. Deredge *et al.* [248] correlate this effect with the protection from hydrogen/deuterium exchange that the inhibitors cause on large regions of the polymerase, distant from the binding site.

Our results obtained using the methyl-HMQC spectrum of NS5B$_{\Delta 21}$ show that the binding of filibuvir to the polymerase affects not only the binding site but also residues that are 12 Å or more away from it. These effects seem to propagate through the protein through the Q and T helices and the fingertips Λ1 loop. Additionally, using two open mutants of NS5B$_{\Delta}$ we observe that filibuvir does not perturb the mechanism that results in their opening, thus suggesting that filibuvir binding is not incompatible with an open structure.

The intermediate exchange that results in the broadening of four resonances in apo NS5B$_{\Delta 21}$ still exists with filibuvir binding. Thus we lack probes to investigate the effect of filibuvir on the C-terminal linker and β-loop that have been shown to be required for NNI-2 activity [408]. Interestingly, Deredge *et al.* [248] showed that filibuvir, unlike other thumb-II inhibitors, did not protect these regions from deuterium exchange. It would be interesting to test whether GS-9669 or lomibuvir, the other inhibitors tested in their study, would allow us to observe the resonances that are broadened beyond detection.

CPMG experiments were used to investigate the millisecond to microsecond dynamics present in the polymerase in the presence of the thumb-II inhibitor. Compared to apo NS5B$_{\Delta 21}$, less resonances present non-flat relaxation dispersion. In particular, the resonances in the thumb subdomain (I432 and I462) display flat relaxation dispersion and the resonance corresponding to I23 in the fingertips Λ1 loop has a markedly

slower exchange rate. Thus, it appears that filibuvir affects the dynamics of the thumb.

Filibuvir can also be used to gain insights into NS5A-D2 binding on NS5B as they share a binding site. The changes in peak height in the $^{1}H,^{15}N$-HSQC of NS5A-D2 in the presence of $NS5B_{\Delta 21}$ with and without filibuvir we can confirm that filibuvir does not completely abolish the interaction between the proteins. It does however, bind to the thumb-II site, blocking the access to this site for NS5A-D2. This allowed us to identify the N-terminal binding site on NS5A-D2 as the one that preferably binds to the thumb-II site on NS5BΔ21. This further supports that two or more binding sites of NS5A-D2 exist on $NS5B_{\Delta 21}$ as not all binding sites on NS5A-D2 display reduced binding despite the thumb-II being blocked. Additionally, filibuvir blocking the binding site of NS5A-D2 might be contribute to its inhibitory activity *in vivo*.

In conclusion, filibuvir acts as a allosteric inhibitor with an effect on the dynamics of the polymerase that results in a stabilization of the polymerase observed by DSF. In particular, the thumb subdomain is most affected by chemical shift perturbations and a reduction in its dynamics. This effect seems to be mediated by helices Q, T and the fingertips Λ1 loop.

Interplay between NS5B, NS5A-D2 and RNA

7.1 Introduction

As a RNA-dependent RNA polymerase, the activity of NS5B is dependent on its capacity to bind RNA. This polymerase is able to synthesize new RNA strands in two ways: *de novo* initiation and primer extension, depending on whether there is a primer to initiate its reaction or not. A recently proposed mechanism [148] highlights the importance of the relative positions of the subdomains in the different steps of *de novo* RNA synthesis (see Figure 1.11). Firstly, the binding of the template to NS5B and subsequent binding of the first complementary nucleotides to the template, has been proposed to occur with the polymerase in a more closed conformation where the β-loop helps align the first nucleotides to the template. Then, a change in conformation to a more open state is necessary to transition to elongation [160, 429]. An X-ray crystallography structure of $NS5B_{\Delta 21}$ from the HC-J4 isolate [144] bound to a small oligonucleotide shows two distinct crystallized conformations of $NS5B_{\Delta 21}$, one that is classified as closed and one that is intermediate by the definitions presented in Chapter 4 (see Figure 4.10). Contrastingly, crystallographic structures of NS5B bound to template-primer RNA [145, 148] are open structures (but the constructs of NS5B that were used had been modified to be more open beforehand). Thus closed and open conformations, in addition to the transition between the conformations all seem to be important for the activity of the polymerase.

In this chapter we will study the effect of the binding of a single-stranded 16-mer RNA to NS5B *in vitro*. This strand of RNA, 5'-CUAAGAUGCUCGCUGC-3' (RNA16), was used by Cramer *et al.* [430] to study the binding of $NS5B_{\Delta 21}$ from the HC-J4 isolate (genotype 1b) to RNA. The binding of a single-strand RNA corresponds to the first step of the RNA synthesis, the binding of the template strand, that is expected to occur in a more closed conformation than the subsequent elongation steps. Furthermore, we will investigate how NS5A-D2 and filibuvir affect this binding.

7.2 Characterization of the interaction between NS5B and RNA

To characterize the interaction between RNA and NS5B we montitored the interaction between $NS5B_{\Delta 21}$ and RNA16 by NMR and fluorescence spectroscopies and size-exclusion chromatography (SEC).

7.2.1 By NMR spectroscopy

To study this interaction by NMR spectroscopy we had to use a different buffer solution than the phophate buffer used in most experiments before. $NS5B_{\Delta 21}$ is only soluble at concentrations necessary for NMR under saline buffer compositions which are not always compatible with NMR experiments due to their high conductivity. The buffer composition used in most of the previous experiments, the phosphate buffer (300 mM sodium phosphate pH 6.8, 50 mM NaCl, 2 mM $MgCl_2$, 1 mM EDTA, 4 mM THP) is not compatible with the study of the interaction with a nucleic acid. Instead, RE (arginine/glutamate) buffer (100 mM arginine, 100 mM glutamate pH 7.0, 100 mM NaCl, 2 mM $MgCl_2$, 1 mM EDTA, 4 mM THP) was used for the experiments with RNA16. This buffer provides improved solubility without significantly impacting the NMR spectrum [431] and is compatible with nucleic acid binding. We compared the methyl-HMQC spectra obtained for $NS5B_{\Delta 21}$ in phosphate buffer and RE buffer (Figure 7.1) and we observe a shift of the resonance corresponding to I160 and also to a lesser extent I512. Interestingly for I160, this suggests that when unliganded this side chain, that is oriented towards the RNA-binding groove, is somewhat solvent-exposed. Furthermore, based on a crystallographic structure obtained for NS5B from the HC-J4 isolate co-crystallized to a ssRNA 4-mer poly-U oligonucleotide, this residue directly interacts with the RNA strand [144]. Also, crystallographic structures of the NS5B ternary elongation complex show that this residue together with Y162 interact with template strand opposite the incoming nucleotide [148].

In their work, Cramer *et al.* [430] determined a dissociation constant of 99 nM between NS5B and RNA16, thus we expected to observe a strong affinity interaction by

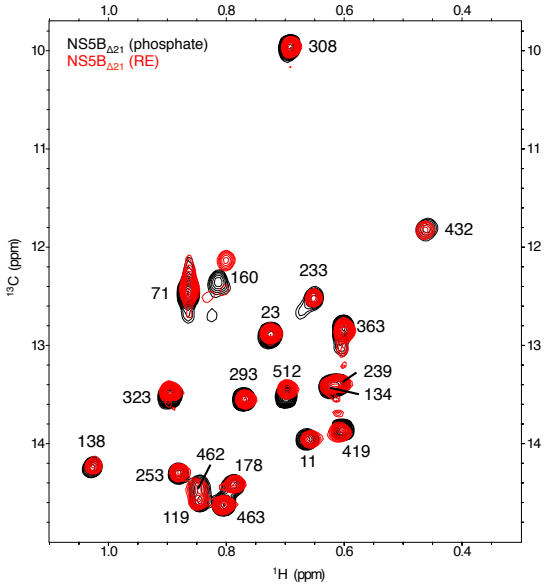

FIGURE 7.1: Effect of arginine/glutamate buffer on methyl-HMQC spectrum of U-^{2}H, Ile δ1-[^{13}C^{1}H$_3$]-labeled NS5B$_{\Delta21}$. Overlay of methyl-HMQC spectra of U-^{2}H, Ile-[^{13}C^{1}H$_3$]-labeled NS5B$_{\Delta21}$ obtained in phosphate buffer (300 mM sodium phosphate pH 6.8, 50 mM NaCl, 2 mM $MgCl_2$, 1 mM EDTA, 4 mM THP; black) and RE buffer (20mM TrisCl pH 7.0, 100 mM arginine, 100 mM glutamate, 100 mM NaCl, 2 mM $MgCl_2$, 1 mM EDTA, 4 mM THP; red). The resonance due to isoleucine 160 is the most affected by the change in buffer.

NMR spectroscopy. Using the RE buffer, perdeuterated and isoleucine δ1-[^{13}C^{1}H$_3$]-labeled NS5B$_{\Delta21}$ was titrated with RNA16 and this was followed using the methyl-HMQC spectrum. Figure 7.2 shows the effect of 0 - 30 μM of RNA16 on 50 μM of NS5B$_{\Delta21}$. Unfortunately, the addition of RNA in the large concentrations necessary for NMR causes the progressive precipitation of NS5B$_{\Delta21}$ out of solution. This may be similar to the precipitation observed by Jin *et al.* [429] after mixing NS5B to its substrates. To compensate for this effect we normalized the peak heights with respect to the signal of I293. The following point in our titration, 60 μM of RNA16, had such low signal-to-noise ratio that proper quantification of the protein signals was not possible.

In the titration, we observe the progressive reduction of the intensities of some NS5B resonances with the addition of RNA16. This reduction is observed on the resonances that correspond to isoleucines 419, 432 and 512 in the thumb and 23, 134, 178 and 233 in the fingers.

We expected to observe perturbation of the resonance corresponding to I160 since this residue appears to interact directly with the template strand in crystallographic structures [144, 148]. However, unexpectedly, this resonance does not shift or broaden.

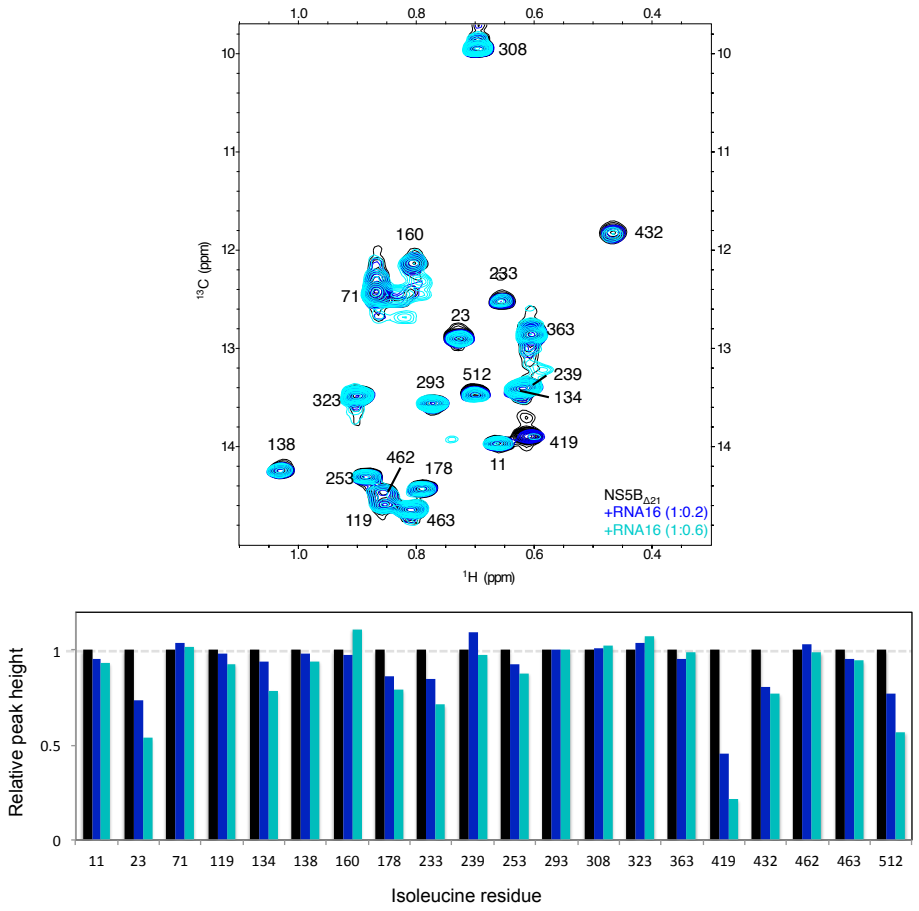

FIGURE 7.2: Interaction between NS5B$_{\Delta 21}$ and RNA16 by NMR spectroscopy. **Top** Methyl-HMQC spectrum of 50 μM U-^{2}H, Ile δ1-[^{13}C^{1}H$_3$] NS5B$_{\Delta 21}$ in the absence (black) and presence of 10 μM (blue) and 30 μM of RNA16. **Bottom** Relative peak height due to the addition of RNA16 colored as the spectrum. Normalization was done assuming that I293 was not perturbed by the addition.

A possible, albeit seemingly unlikely, explanation for this behavior is that the chemical shift of the resonance in the free and bound states are the same.

Owing to the lack of intensity from the residues on the inner surface of the RNA-binding groove (I405, I412, I413) we do not have other direct probes of the RNA-binding interaction. Nevertheless, the intensity of certain resonances is reduced with the addition of RNA16. Based on a crystallographic structure of NS5B bound to single-stranded RNA (PDB 1NB7) [144], these residues are not directly interacting with RNA and are all over 12 Å away from the RNA oligonucleotide. The isoleucine Cδ1 of isoleucine 419, whose resonance is the most affected, is over 28 Å from the expected position of the RNA strand in the crystallographic structure (Figure 7.3). These effects in the spectra highlight long-range allosteric effects. Since resonances from residues

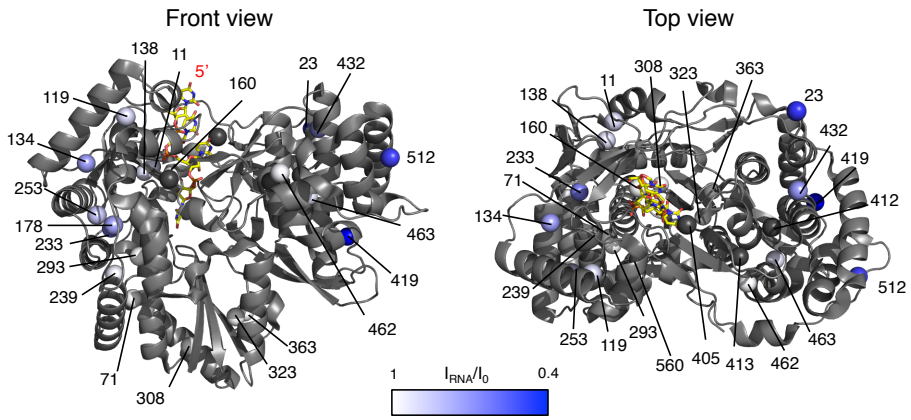

FIGURE 7.3: Effect of RNA16 on peak intensities. Crystallographic structure of $NS5B_{\Delta 21}$ (PDB 3I5K) shown as cartoon with spheres to represent isoleucine $\delta 1$ carbons. Resonances that are broadened in the presence of RNA16 are shown in shades of blue while those unaffected are shown in white. Relative peak intensities as shown in Figure 7.2 for 50 μM $NS5B_{\Delta 21}$ and 30 μM of RNA16. Isoleucines 405, 412, 413 and 560 that do not present intensity in the unliganded spectrum are shown as gray spheres but are not labeled in the front view for clarity. The resonance due to I419 broadens to 21.5% of the expected intensity, which is below the cutoff value in the figure (40%). A 4-mer poly-U oligonucleotide was placed by alignment to PDB 1NB7 as the estimated positioning of RNA16 in the polymerase and is represented as sticks. Its 5′ end is labeled in red in the front view.

in the thumb subdomain and isoleucine 23 in the fingertips $\Lambda 1$ loop (that is in contact with the thumb subdomain) are the most affected. This suggests that the thumb either undergoes a change in conformation to form an initial RNA-polymerase complex or there is a change in the dynamics of the thumb upon binding.

This further highlights the importance of the thumb subdomain as the main subdomain perturbed by the dynamics of the polymerase. The binding of allosteric inhibitors and NS5A-D2 to this region may therefore affect the movement of the thumb and consequently the binding of RNA.

We observe the broadening of resonances during the titration of $NS5B_{\Delta 21}$. Broadening occurs when there is intermediate exchange between the free and bound states. We do not observe new resonances appearing that could correspond to resonances in the bound state. With the addition of 30 μM of RNA16 only a fraction of 50 μM $NS5B_{\Delta 21}$ (about 60% assuming this interaction is strong) that we are observing will be bound but we still would expect to observe new resonances appearing.

In the mechanism proposed by Appleby *et al.*, the first step in RNA synthesis would involve the displacement of the linker. One of the residues whose corresponding resonance is broadened beyond detection is I560, which suggests that this residue is in

intermediate exchange. This isoleucine is in the linker. Additionally, none of the open structures of NS5B JFH-1 have electron density for this residue. Based on this, we were expecting to observe a new disordered resonance due to this isoleucine to show that the linker became disordered in solution. However, we do not observe the appearance of a resonance in the disordered region of the spectrum which suggests that the linker does not become completely disordered with the binding of ssRNA. The linker would remain in intermediate exchange. Indeed, Harrus *et al.* [160] proposed that the C-terminal linker region aids in *de novo* initiation and could be displaced but remain in the RNA binding groove during this first step.

The concentrations needed to investigate this interaction by NMR spectroscopy resulted in precipitation of the complex which precludes the determination of the dissociation constant of this interaction using this technique, thus we used other techniques to complement this investigation.

7.2.2 By fluorescence spectroscopy

To investigate the binding of NS5B$_{\Delta 21}$ to RNA we used fluorescence spectroscopy. For this, the 5'-end of RNA16 was labeled with 6-fluorescein amidite to give 6FAM-RNA16. We observed the quenching of the fluorescence with the addition of NS5B$_{\Delta 21}$ as shown in Figure 7.4. This was done in a RE buffer (20 mM TrisCl pH 7.5, 100 mM NaCl, 80 mM arginine, 80 mM glutamate, 5 mM $MgCl_2$, 1 mM DTT) similar to the buffer used for the NMR experiments with RNA16 and compatible with the observation of fluorescence emission.

Titration data of the interaction of 6FAM-RNA16 with NS5B$_{\Delta 21}$ was fit to an adapted version of Equation 1.10 for fluorescence data shown below (Equation 7.1), where ΔF_{obs} and ΔF_{max} are the absolute variations of fluorescence integration at a given concentration of NS5B$_{\Delta 21}$ and at saturation, respectively, gave a K_D of 0.89 μM. This is 9-fold larger than the K_D of 99 nM observed by Cramer *et al.* [430]. They used the same RNA fragment and tested the activity of NS5B$_{\Delta 21}$ from a genotype 1b isolate whereas we used NS5B$_{\Delta 21}$ from the JFH-1 isolate (genotype 2a). Furthermore, they used a smaller concentration of sodium chloride. In our case, the higher salinity can

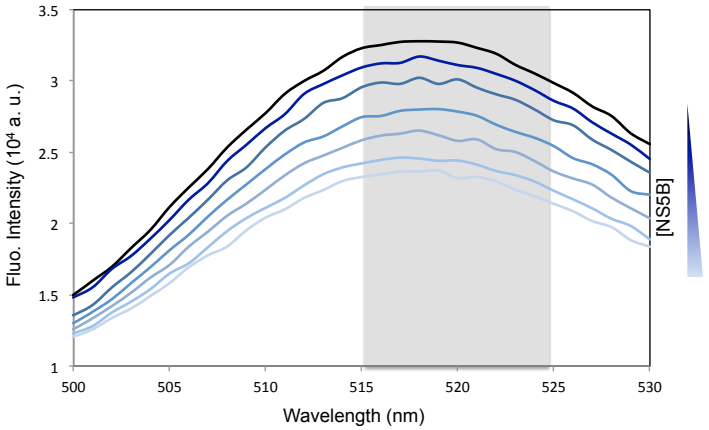

FIGURE 7.4: Quenching of 6FAM-RNA16 fluorescence by NS5B$_{\Delta 21}$. Fluorescence emission spectrum of 50 nM of 6-FAM RNA16 (excitation at 492 nm) in the absence (black) and presence of 70 - 3000 nM of NS5B$_{\Delta 21}$ (shades of blue). NS5B$_{\Delta 21}$ causes a quenching of the fluorescence that was quantified by integrating the fluorescence intensity between 515-525 nm (highlighted in gray).

interfere with the electrostatic nature of the NS5B$_{\Delta 21}$-RNA interaction.

$$\Delta F_{obs} = \Delta F_{max} \frac{(K_D + [L]_0 + [P]_0) - \sqrt{(K_D + [L]_0 + [P]_0)^2 - (4[P]_0[L]_0)}}{2[P]_0} \quad (7.1)$$

7.2.3 By size-exclusion chromatography

To complement the investigation by fluorescence we analyzed the interaction between NS5B and RNA16 by SEC. SEC separates particles by their size (hydrodynamic radius). This is done by using a column packed with porous beads that act as the stationary phase. Smaller particles enter the pores and are retained longer in the column while large particles cannot enter as many pores and pass through the column less hindered. Therefore, the larger the particle the faster it will elute in SEC [432].

Here we used the same fluorescent RNA, 6FAM-RNA16, as used in the fluorescence assays. We used an analytical size-exclusion column with a fractionation range of 3 to 70 kDa for globular proteins. This is well adapted to the biomolecules we are studying since the fluorescein-labeled RNA molecule has a molecular weight of 5.8 kDa, NS5A-D2 11.6 kDa (but migrates as if it was larger because of its disordered nature) and NS5B$_{\Delta 21}$ 65 kDa. We can detect the elution of RNA by the characteristic absorption of the fluorescein label at 495 nm. To detect the elution of the proteins alone we

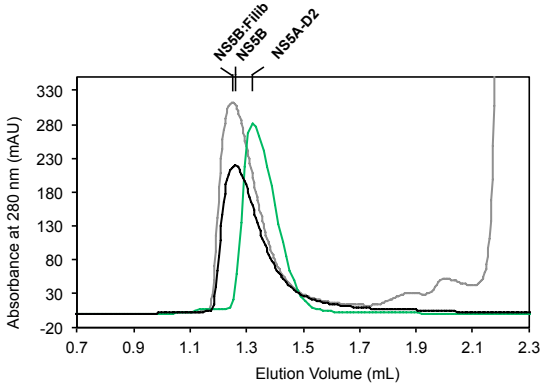

FIGURE 7.5: Elution profiles of NS5B$_{\Delta 21}$ and NS5A-D2. Elution profiles of 10 μM NS5B$_{\Delta 21}$ (black), 10 μM NS5B$_{\Delta 21}$ saturated with filibuvir (gray) and 10 μM NS5A-D2 (green). These were measured in the absence of RNA16, and the absorbance was detected at 280 nm. The strong absorbance at 2.3 mL in the presence of filibuvir is due to excess filibuvir in solution.

[6FAM-RNA16] (μM)	[NS5B$_{\Delta 21}$] (μM)	% bound 6FAM-RNA16
10	0	0
10	3	26.7
10	7	57.8
10	10	74.3
10	15	87.5

TABLE 7.1: Expected ratio of bound 6FAM-RNA16 in the size-exclusion chromatography experiments. The concentrations of NS5B for the SEC experiments were chosen based on the dissociation constant determined by fluorescence spectroscopy (0.89 μM) to give a wide range of ratios of bound 6FAM-RNA.

also observed the absorbance at 280 and 260 nm. The elution volumes for the proteins alone were determined in this way (Figure 7.5). The range of concentrations used to study the interaction between NS5B$_{\Delta 21}$ and 6FAM-RNA was chosen based on the dissociation constant measured by fluorescence spectroscopy so that various ratios of the NS5B:RNA complex formation could be observed (Table 7.1).

If 6FAM-RNA16 binds to NS5B$_{\Delta 21}$ we expect the absorbance peak measured at 495 nm (only the fluorescein-labeled oligonucleotide absorbs at this wavelength) to shift to an earlier elution volume with the addition of NS5B$_{\Delta 21}$, as the complex is significantly larger than 6FAM-RNA16 alone. Indeed, we observe a decrease in the elution peak of 6FAM-RNA16 at 1.54 mL and the progressive formation of a new elution peak at 1.10 mL with the addition of 3 – 15 μM of NS5B$_{\Delta 21}$, corresponding to the the NS5B-RNA complex. This elution is faster than NS5B$_{\Delta 21}$ alone (at 1.25 mL). This is consistent with the dissociation constant of 0.89 μM determined by fluorescence for this interaction. However, a dissociation constant could not be confidently determined by this method.

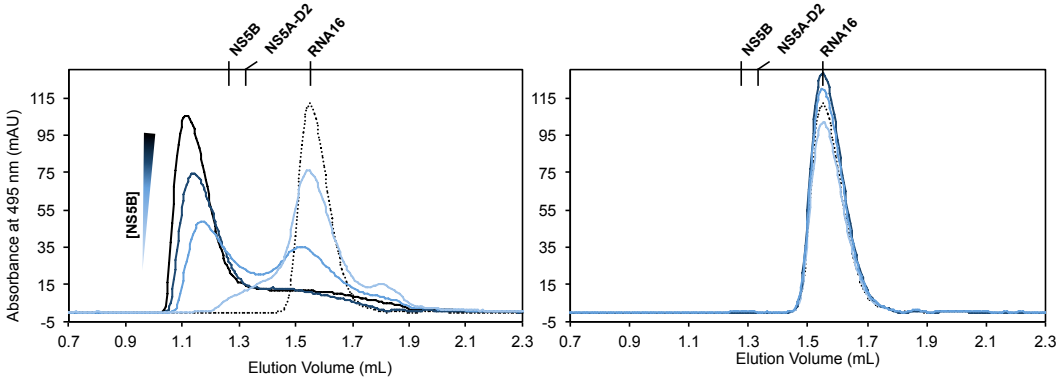

FIGURE 7.6: Interaction between 6FAM-RNA16 and NS5B$_{\Delta 21}$ or NS5A-D2. By size-exclusion chromatography the elution of 10 μM of fluorescent RNA oligonucleotide 6FAM-RNA16 was observed in the absence and presence of NS5B$_{\Delta 21}$ or NS5A-D2. **Left** The effect of the interaction with NS5B$_{\Delta 21}$. 6FAM-RNA16 elution is shown in the absence (black dotted line), and with the addition of 3 μM (light blue), 7μM (medium blue), 10 μM (dark blue) and 15 μM (full black line) NS5B$_{\Delta 21}$. **Right** NS5A-D2 does not interact with 6FAM-RNA16. 10 μM 6FAM-RNA16 (alone, black dotted line) was mixed with 0 (black dotted line), 5 (light blue), 15 (medium blue) and 50 (dark blue) μM of NS5A-D2. Elution times of NS5B$_{\Delta 21}$, NS5A-D2 and 6FAM-RNA16 are labeled at the top of the chromatograms and have been measured at 280 nm.

This is due to two effects: (i) the integration of the peak should be taken into account for this determination and this could not be confidently estimated due to overlap of free- and bound-RNA peaks and (ii) the dissociation of the complex that can occur during the chromatography.

In sum, NMR spectroscopy showed that RNA-binding causes long-range effects and additional intermediate exchange that causes broadening of the resonances. Fluorescence spectroscopy and SEC support a low micromolar interaction between NS5B$_{\Delta 21}$ and RNA16. However, the lack of the resonances corresponding to the four isoleucines that surround the RNA-binding groove limit the usefulness of this analysis. It has still provided the information that the thumb subdomain is affected during this step as seen by the decrease in intensity of resonances due to residues in this subdomain. For this reason, we were interested in the effects of NS5A-D2 and filibuvir on RNA-binding because they bind to the thumb and have been shown to induce long-range effects on NS5B$_{\Delta 21}$.

7.3 Does NS5A-D2 bind RNA?

We were interested in investigating the effect of NS5A-D2 on the ability of $NS5B_{\Delta 21}$ to bind RNA. Full-length NS5A is known to bind RNA. In particular its domain 1 dimers have RNA-binding properties [101]. Furthermore, there have been reports that the disordered domains of NS5A also contribute to RNA-binding [433, 434]. To understand the contribution of NS5A-D2 in the RNA-binding activity of NS5B we first tested whether our RNA template bound to NS5A-D2. This is important to decouple the effect NS5A-D2 has on its own on RNA from the effect mediated through NS5B.

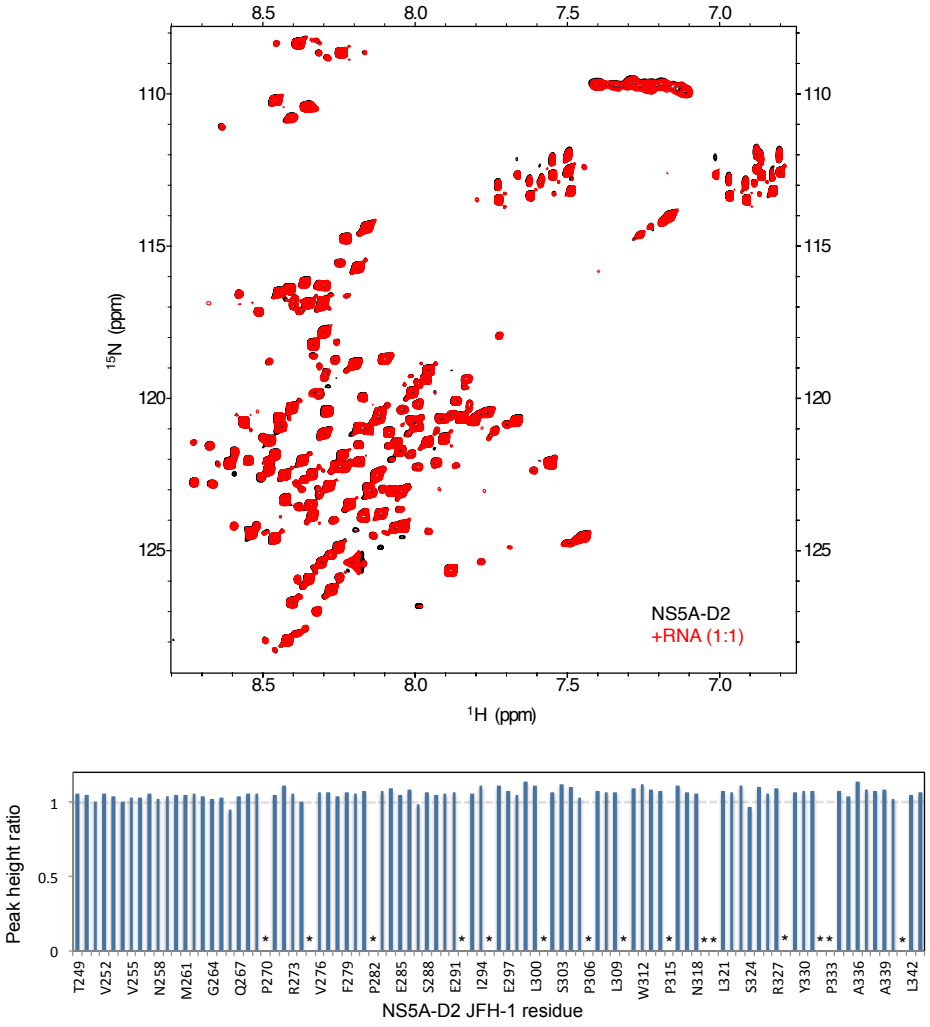

FIGURE 7.7: NS5A-D2 does not bind RNA16. The addition of RNA16 to ^{15}N-labeled NS5A-D2 JFH-1 does not cause changes in chemical shifts or peak heights. **Top** ^{15}N,^{1}H-HSQC of 50 μM of ^{15}N-labeled NS5A-D2 JFH-1 in the absence (black) and presence of an equimolar amount of unlabeled RNA16 (red). **Bottom** Peak height ratios between NS5A-D2 in the presence and absence of RNA16. A value of 1 indicates there is no broadening due to the interaction. Prolines (that do not have amide protons) are shown as asterisks in the graph.

NS5A-D2 binding to RNA was tested by the equimolar addition of RNA16 to ^{15}N-labeled NS5A-D2 (Figure 7.7). No chemical shift perturbations were detected on the resonances corresponding to NS5A-D2 residues. There is only a shift of the resonance of the hexahistidine tag, likely due to a slight change in pH with the addition of the RNA. Additionally, the ratios of peak heights are also shown in Figure 7.7 and do not present changes. Taken together this means that NS5A-D2 does not bind RNA16.

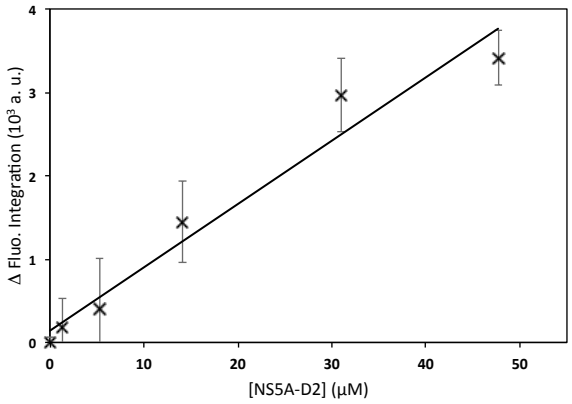

FIGURE 7.8: Effect of NS5A-D2 on 6FAM-RNA16. The titration of 50 nM of 6FAM-RNA16 with NS5A-D2 shows that the fluorescence increases linearly with the protein concentration. This suggests that there is not an interaction between these molecules as a we do not observe a titration curve.

Fluorescence spectroscopy and size-exclusion chromatography further support the lack of an interaction between NS5A-D2 and RNA16. In a fluorescence titration, we observe a linear increase in fluorescence with the addition of NS5A-D2 to 6FAM-RNA16 (Figure 7.8). This can be due to a change in the properties of the solution, for example its viscosity, that can alter the fluorescence of the probe. If there were an interaction we would expect a titration curve as seen with the addition of NS5B$_{\Delta 21}$. SEC does not show elution of a 6FAM-RNA16 and NS5A-D2 complex (Figure 7.6, right). 5 – 50 μM of NS5A-D2 was added to 10 μM of 6FAM-RNA16 to observe whether there was an interaction. At 495 nm we did not detect a change in the elution profile of 6FAM-RNA16.

We did not expect to observe an interaction between NS5A-D2 and RNA as theoretical isoelectric point of our construct of the domain is 4.66 and at pH 7.0 it would be negatively charged and would repel negatively charged RNA. The difference with respect to reports of NS5A-D2 binding RNA could be due to the choice in RNA strand and proteins from different isolates. Firstly, it has been suggested that NS5A has a

preference for poly-U RNA [433]. Secondly, the experiments that showed RNA binding were done with NS5A-D2 from the Con1 isolate and were mapped to the positively charged residues at positions 304-311 (which correspond to residues 300-307 in NS5A-D2 from the JFH-1 isolate). NS5A-D2 from the JFH-1 isolate only has two arginine residues in this region, compared to four arginines and a lysine in the same region in the Con1 isolate.

Therefore, it is possible that this interaction is not present in NS5A-D2 JFH-1 but exists in the Con1 protein. The absence of direct interaction between NS5A-D2 JFH-1 and our RNA oligomer thus allows us to investigate the possible indirect effect that can be mediated through $NS5B_{\Delta 21}$.

7.4 Effect of NS5A-D2 on NS5B RNA-binding

The study of the ternary complex composed of RNA16, $NS5B_{\Delta 21}$ and NS5A-D2 using the methyl-HMQC is not practical. Both NS5A-D2 and RNA16 cause the broadening of many of the same resonances in $NS5B_{\Delta 21}$ and it would be difficult to identify if only a small effect was present. Thus, we used other techniques to understand the effect NS5A-D2 has on the capacity of $NS5B_{\Delta 21}$ to bind RNA.

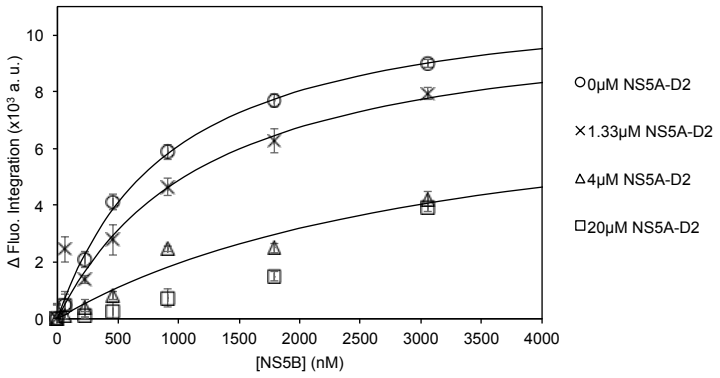

FIGURE 7.9: Effect of NS5A-D2 on $NS5B_{\Delta 21}$ binding to 6FAM-RNA16. Absolute variation of the fluorescence intensity of 50 nM 6FAM-RNA16 during titration with $NS5B_{\Delta 21}$ in the presence of 0 - 20 μM of NS5A-D2. Dissociation constants were measured (see Table 7.2) for concentrations of NS5A-D2 up to 4 μM (fitted curves shown as full lines). With the addition of 20 μM of NS5A-D2 a dissociation constant could no longer be fit to the data. These experiments were done in triplicate.

We repeated this titration in the presence of increasing concentrations of NS5A-D2 (Figure 7.9). We observe an increase in the dissociation constant for the interaction

[NS5A-D2] (μM)	K_D (μM)	% NS5B bound to NS5A-D2
0	0.89	0
1.33	1.20	1.8
4	3.32	5.2
20	n.d.	21.7

TABLE 7.2: Experimental dissociation constant values for the $NS5B_{\Delta 21}$ and 6FAM-RNA16 interaction in the presence of NS5A-D2 determined by fluorescence spectroscopy by fitting Equation 7.1 to the experimental absolute variations in fluorescence integration. With the addition of 20 μM of NS5A-D2 we could not measure a dissociation constant with the experimental data (n.d.). The last column shows the percentage of NS5B at the final titration point ([NS5B] = 3 μ) that would be expected to be bound to NS5A-D2 in the absence of RNA based on an NS5B:NS5A-D2 dissociation constant of 70 μM. The presence of NS5A-D2 decreases the affinity between 6FAM-RNA16 and $NS5B_{\Delta 21}$.

[$NS5B_{\Delta 21}$] (μM)	[NS5A-D2] (μM)	% bound NS5B
7	0	0
7	5	6.1
7	15	16.5
7	50	40.3

TABLE 7.3: Proportion of NS5B:NS5A-D2 complex at the concentrations used in SEC experiments. NS5A-D2 concentrations were chosen to bind $NS5B_{\Delta 21}$ to give different proportions of protein-protein complex to determine how this complex affects the formation of the NS5B:RNA complex. These values are calculated based on the NS5B:NS5A-D2 dissociation constant determined by NMR spectroscopy (70 μM) in the absence of RNA.

between $NS5B_{\Delta 21}$ and RNA16 (Table 7.2), meaning that NS5A-D2 decreases the affinity between these biomolecules since there is no interaction between NS5A-D2 and RNA. We observe more than a 3-fold increase in dissociation constant with the addition of 4 μM of NS5A-D2. With the addition of 20 μM of NS5A-D2, the titration was no longer displayed a quadratic K_D curve and thus could not be used to determine the constant.

To further support these findings on the effect NS5A-D2 has on the binding of $NS5B_{\Delta 21}$ to RNA16 we used SEC. We started with 10 μM 6FAM-RNA16 with 7 μM $NS5B_{\Delta 21}$, which corresponds to about 58 % of RNA bound to NS5B without NS5A-D2. Then, increasing amounts of NS5A-D2 we added to determine if it perturbs the formation of the NS5B:RNA complex. The concentrations chosen of NS5A-D2 in the absence of RNA would be expected to bind 6 – 40 % of the polymerase in solution (Table 7.3) based on the K_D determined by NMR spectroscopy in Chapter 5.

Prior to the addition of NS5A-D2, the elution of the NS5B:RNA complex occurs at 1.15 mL. 5 – 50 μM of NS5A-D2 were mixed with samples containing 10 μM 6FAM-RNA16 and 7 μM $NS5B_{\Delta 21}$ (Figure 7.10). We observe a decrease in the formation of the

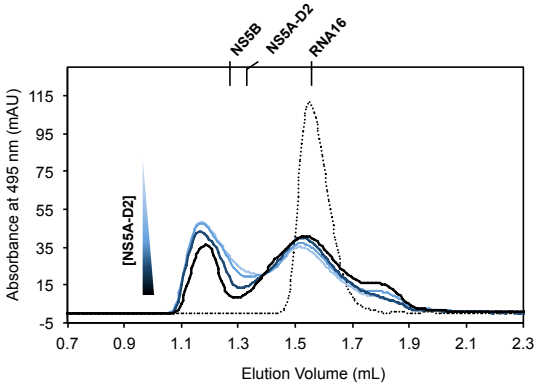

FIGURE 7.10: NS5A-D2 inhibits the interaction between NS5B$_{\Delta 21}$ and RNA16. 10 μM 6FAM-RNA (alone, black dotted line) with 7 μM NS5B$_{\Delta 21}$ was mixed with 0 (light blue), 5 (medium blue), 15 (dark blue) and 50 (black) μM of NS5A-D2. Elution times of NS5B$_{\Delta 21}$, NS5A-D2 and 6FAM-RNA16 alone are labeled at the top of the chromatograms and have been measured at 280 nm.

NS5B:RNA complex visible by the shift of the complex peak to 1.17 mL and decrease of its intensity. Conversely, we observe an increase of the RNA elution peak by its shift toward the elution of free RNA and increase in intensity of the free-RNA peak. However, these peaks become broader suggesting an equilibrium between bound and free forms. This is in agreement with the decrease in the affinity of NS5B$_{\Delta 21}$ for RNA in the presence of NS5A-D2 that had been observed by fluorescence spectroscopy. We cannot, however, determine with certainty whether both NS5A-D2 and RNA can bind to NS5B simultaneously but the later elution of the RNA:NS5B peak in the presence of NS5A-D2 compared to RNA:NS5B alone suggests that a ternary complex is not stable, as the larger ternary complex would be expected to migrate faster in the column.

Since in the previous chapter we showed that NS5A-D2 interacts with NNI binding sites on the thumb subdomain and we do not observe an interaction between NS5A-D2 and RNA, this suggests that NS5A-D2 modulates NS5B RNA-binding via an allosteric mechanism. Moreover, fluorescence spectroscopy and SEC show that NS5A-D2 binding reduces NS5B$_{\Delta 21}$ binding to RNA16.

7.5 Effect of filibuvir on RNA-binding to NS5B$_{\Delta 21}$

We used our methyl-HMQC spectrum of NS5B$_{\Delta 21}$ to test whether filibuvir affects the capacity of NS5B$_{\Delta 21}$ to bind RNA. It has been suggested that filibuvir does not affect or

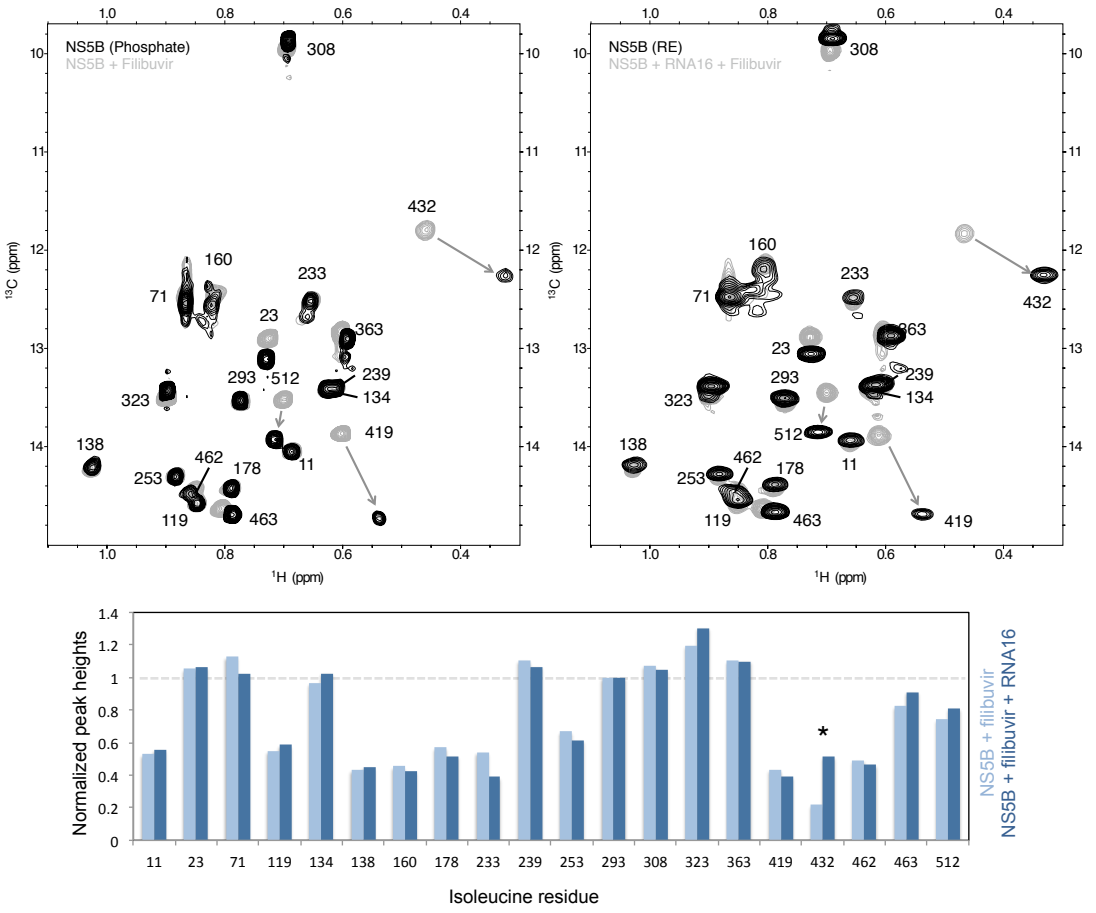

FIGURE 7.11: Effect of filibuvir on RNA-binding of NS5B$_{\Delta 21}$. **Top left** Methyl-HMQC spectra of 81 μM U-^{2}H, Ile-[^{13}C^{1}H$_3$]-labeled NS5B$_{\Delta 21}$ unliganded (gray) and bound to filibuvir (saturated, black) in phosphate buffer. **Top right** Methyl-HMQC spectra of 50 μM U-^{2}H, Ile-[^{13}C^{1}H$_3$]-labeled NS5B$_{\Delta 21}$ unliganded (gray) or bound to filibuvir in the presence of 30 μM of unlabeled RNA16 in RE buffer. **Bottom** Peak heights of NS5B$_{\Delta 21}$ (normalized with I293) bound to filibuvir in the absence (light blue) and presence (dark blue) of 30μM of RNA16. There is an increase of the peak height of I432.

only affects modestly *de novo* initiation [246, 249, 435] and therefore we should expect to observe binding of RNA16 to NS5B$_{\Delta 21}$ when it is saturated with filibuvir.

Our spectrum of NS5B$_{\Delta 21}$ with filibuvir in the presence of RNA16 is very similar to the spectrum of NS5B$_{\Delta 21}$ with only filibuvir (Figure 7.11, top). Analysis of the peak heights shows that the peak height of I432 is increased with RNA16 suggesting an effect of binding (Figure 7.11, bottom) despite being somewhat inconclusive.

To determine whether NS5B$_{\Delta 21}$ binds to RNA16 in the presence of filibuvir, we used SEC. The effect of filibuvir on the binding of NS5B$_{\Delta 21}$ to 6FAM-RNA16 was investigated by the addition of 10 μM of both unliganded and filibuvir-bound NS5B$_{\Delta 21}$ to 10 μM 6FAM-RNA. Both chromatograms were compared to investigate the effect of

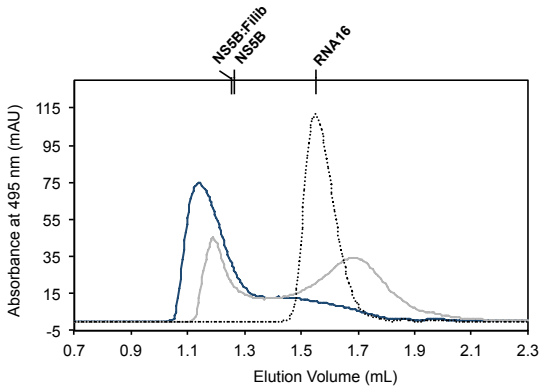

FIGURE 7.12: Filibuvir inhibits the interaction between $NS5B_{\Delta 21}$ and RNA16. 10 μM of 6FAM-RNA16 (dotted black line) in the presence of 10 μM $NS5B_{\Delta 21}$ unliganded (dark blue) or filibuvir-bound (gray).

the presence of the small molecule inhibitor (Figure 7.12). With unliganded $NS5B_{\Delta}21$ most of RNA appears complexed with the polymerase. However, in the presence of filibuvir, NS5B is less capable of binding RNA but it does not completely eliminate this interaction as evidenced by the presence of the complex peak at 1.2 mL.

Overall, both size-exclusion chromatography using fluorescent-labeled RNA showed that filibuvir perturbs the formation of the NS5B:RNA complex but does not completely abolish it.

7.6 Conclusions

Firstly, we investigated the binding of $NS5B_{\Delta 21}$ to a 16-nucleotide RNA oligomer by NMR spectroscopy. The NS5B-RNA interaction was observed by NMR using the methyl-HMQC spectrum characterized in Chapter 4 but due to the lack of probes in the RNA-binding groove we could observe only indirect effects. These indirect effects show us that the thumb subdomain is perturbed during RNA-binding, either due to a change in conformation and/or dynamics. Since many of the same resonances that broaden in the presence of NS5A-D2 broaden also in the presence of RNA16 we chose not to characterize the ternary complex by NMR.

Unlike others in the literature [94, 433, 434] we did not observe an interaction between NS5A-D2 and RNA. This was confirmed by experiments done by NMR spectroscopy, fluorescence spectroscopy and SEC. The basic binding site that has been proposed to bind RNA on NS5A-D2 [94] contains fewer basic residues in the protein from

the JFH-1 isolate compared to the Con1 isolate used in the previous studies. The absence of interaction between these partners allows us to decouple the effect NS5A-D2 has on NS5B-RNA binding and consider it to be only due to its interaction with $NS5B_{\Delta 21}$.

We studied the effect of NS5A-D2 on the RNA:NS5B complex by fluorescence and size-exclusion chromatography using a fluorescein-labeled derivative of our 16-mer RNA. Both experiments support that NS5A-D2 decreases the affinity of NS5B for RNA16. Ngure and coworkers [94] recently suggested that NS5A-D2 inhibits the RNA-binding activity of NS5B by a competitive mechanism in which both the disordered protein and RNA compete to bind to the RNA-binding groove. However, our results show an affinity between NS5B and RNA that is more than an order of magnitude stronger compared to the affinity between NS5B and NS5A-D2. As such, our results do not support that NS5A-D2 reduces RNA-binding by a competitive mechanism. Furthermore, in the previous chapter we did not observe the binding of NS5A-D2 to the RNA-binding groove. Instead, the binding of NS5A-D2 to the NNI sites on the thumb subdomain support an allosteric mechanism for the decrease in affinity observed in the presence of NS5A-D2.

We also investigated the effect of filibuvir on the binding of RNA. It has been suggested that filibuvir has only modest effects on *de novo* initiation [246, 249, 435] and acted on a later step of RNA replication, like the transition between initiation and elongation. We report that the presence of filibuvir reduces the capacity $NS5B_{\Delta 21}$ has for binding our single stranded RNA but does not completely abolish it. This indicates that filibuvir may have an inhibitory role in the early steps of RNA replication by reducing RNA template-binding.

In sum, allostery plays an important role in the interaction between NS5B and RNA templates. In particular, the thumb subdomain seems to undergo long-range perturbations upon RNA-binding, even with the binding of a single strand. Both NS5A-D2 and filibuvir affect $NS5B_{\Delta 21}$ allosterically via the thumb subdomain and, indeed, we observe perturbations on RNA binding in the presence of these molecules.

Regulation of human glucokinase by NS5A-D2

8.1 Introduction

Hepatitis C virus causes metabolic changes in its host which are believed to aid in its replication [291], to avoid immune response and generally facilitating persistance [436]. Furthermore, Diamond *et al.* [283] showed that the virus modifies biosynthetic pathways within the infected cell. At first this is to support the synthesis of biomolecules like nucleotides and lipids. Later, as the infection progresses, the perturbed biosynthetic pathways shift to use amino acids and lipids for energetic purposes. The mechanisms by which the virus reprograms the cellular metabolism are not well understood. Ramière *et al.* [296] studied the effect of HCV infection in Huh7.5 hepatoma cells in the glycolysis pathway. They identified a direct interaction between the viral NS5A protein and cellular hexokinase II (HK2) that activates the host enzyme. HK2 is one of four hexokinase isoforms in mammalian cells. These enzymes catalyze the phosphorylation of hexoses with a molecule of ATP (Figure 8.1). HK2 is the main isoform present in tissues such as heart, skeletal muscle and adipose tissues [437]. It is also associated with hepatoma cells such as the Huh7.5 hepatocellular carcinoma-derived cell line that was used in the the study that identified its activation in the presence of HCV infection.

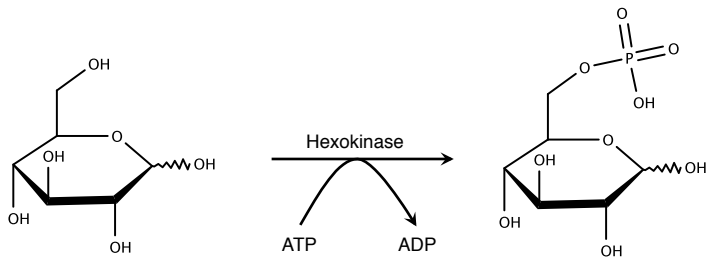

FIGURE 8.1: Hexokinases catalyze the phosphorylation of hexoses. The conversion of glucose to glucose-6-phosphate is one of the main reactions of these enzymes. To carry out the phosphorylation they transfer a phosphate group from an ATP molecule, converting it to ADP.

In collaboration with the research group of Patrice André, Vincent Lotteau and Olivier Diaz, we were interested in investigating the underlying molecular mechanism of the activation of hexokinases by NS5A. Previous work in our group identified a binding site for HK2 on NS5A-D2 and confirmed by NMR that this domain is capable

of activating its enzymatic activity [438]. However, HK2 is not the main hexokinase isoform present in mammalian hepatocytes. In these cells, the main isoform of hexokinase is hexokinase IV, also known as glucokinase [439].

Glucokinase (GCK) is related to other mammalian hexokinase isoforms I-III, all of which also catalyze glucose phosphorylation which is the first step in the glycolysis pathway. Compared to other hexokinase isoforms, GCK has some differences: monomeric structure, low affinity for glucose, its tissue expression, positive cooperativity for glucose (Hill coefficient of 1.7) and it is not inhibited by its product. Due to these characteristics, this protein has a special role as a sort of 'glucose sensor' in humans. This is due to the high glucose concentration it requires ($K_{0.5} \approx 8$ mM, which is within the physiological range), and that its activity has sigmoidal dependence on glucose concentration [440]. Since this enzyme exhibits non-Michaelis-Menten kinetics (due to cooperativity), the Michaelis-Menten constant (K_m) is not used and instead $K_{0.5}$ is used to represent the substrate concentration at half the maximum reaction rate. Similarly, $k_{cat}/K_{0.5}$ represents the catalytic efficiency of such an enzyme instead of k_{cat}/K_m that is used for enzymes with Michaelis-Menten kinetics.

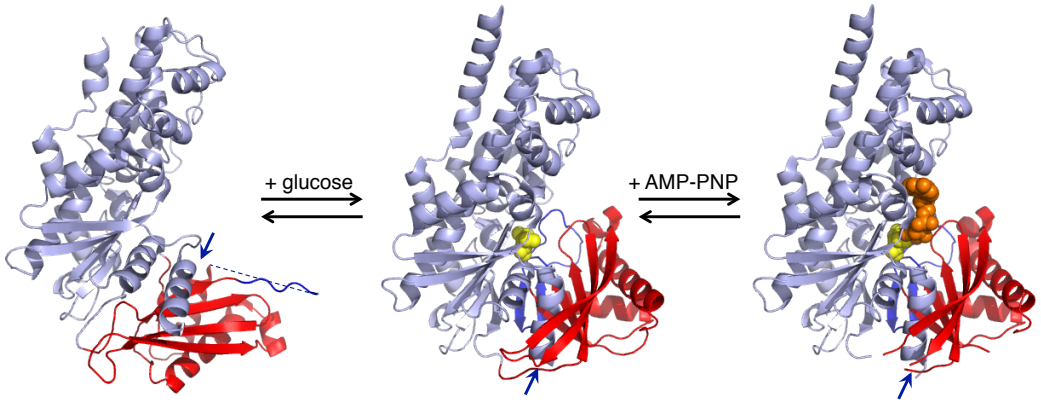

FIGURE 8.2: Allosteric effect of glucose on human glucokinase. The unliganded structure of glucokinase (left, PDB 1V4T [441]) is an 'open' structure with a disordered β-loop 151-180 region (blue dashed line). With the binding of glucose (center, PDB 1V4S [441]) the structure 'closes' to the catalytically-competent form and the β-loop (shown in dark blue) becomes structured. The binding of AMP-PNP (ATP analogue, on the right, PDB 3FGU [442]) does not appreciably alter this active structure. The large domain (residues 5-66 and 204-461) is represented in light blue and the small domain (residues 67-203), that contains the β-loop, in red and blue. Glucose and AMP-PNP are represented in yellow and orange, respectively. The position of the α13-helix is labeled with dark blue arrows in each structure. Adapted from Larion *et al.* [443].

Structurally, GCK has two subdomains called large and small subdomains (Figure 8.2). The active site is at the interface between these subdomains. The binding

of glucose stabilizes the enzyme in a 'closed' conformation. Also, the substrate binding causes a change in position of the C-terminal α-helix called the α13-helix and a reduction of the conformational heterogeneity displayed by the protein [444]. The closed form is the active form of this protein. Larion *et al.*[445] showed by NMR that the small domain exhibits conformational exchange on the millisecond timescale. An order-disorder transition of this domain leads to the cooperativity observed for this enzyme. Most resonances from this subdomain are broadened in the methyl-HMQC spectrum of GCK selectively labeled ^{13}C and ^{1}H on the δ1 methyl groups of isoleucine side chains. The exception are the resonances due to two isoleucines in the disordered 151-180 loop. The binding of glucose reduces the mobility of the small domain (confirmed by the presence of the previously broadened resonances in the spectra) and structures the 151-180 loop. The structuring of the 151-180 loop is evidenced by its presence in the crystallographic structures obtained with bound glucose [441]. Furthermore, the chemical shift of its two previously disordered resonances shift to a region in the methyl-HMQC spectrum that indicates that they are no longer disordered [443]. At the end of the catalytic cycle this loop is believed to become disordered to allow for the release of the product and this step is a limitation for the turnover rate k_{cat} [440].

Small-molecule GCK activators act by stabilizing a closed active form of GCK. This results in a higher affinity for glucose. This is the case of RO-28-0450 and several related activators that interact with Arg 63 and the α13-helix and which, when co-crystallized with GCK, result in a closed crystallographic structure with a structured 151-180 loop [446].

Persistent hyperinsulinemic hypoglycemia of infancy (PHHI) is a form of congenital hyperinsulinism associated with hyperactive GCK variants. These variants contain substitutions in the α13-helix that cause reduced glucose cooperativity. Based on this, Larion and Miller [447] redesigned the α13-helix. Their variant showed increased catalytic activity. This increased activity is due to this variant being in closed conformation even in the absence of glucose [443]. It has an increased affinity for glucose, effectively increasing $k_{cat}/K_{0.5}$.

Another gain-of-function GCK mutant is the R155H-H156M-E157L-K161V β-loop quadruple mutant described by Whittington *et al* [440]. This mutant possesses slightly

lower affinity for glucose than wild-type GCK but increased k_{cat}. It is believed that this is due to increased mobility of the already disordered 151-180 loop which would allow the faster release of the product. Effectively, these two activity-increasing mechanisms are independent as is shown by combining both sets of mutations to obtain a mutant GCK that has 100-fold increased activity [440].

We are interested in investigating the interaction between the viral NS5A-D2 and GCK by NMR spectroscopy. As the small-molecule activators of GCK act by stabilizing the active form, which is readily identified by NMR, we expect to observe this effect also with NS5A-D2 as it acts as an activator with HK2. Compared to HK2, which is a very large protein (103 kDa), the size of GCK (52 kDa) is better adapted to the characterization by NMR spectroscopy. This investigation can provide insight into the molecular mechanism of how the hepatitis C virus modulates glycolysis for its own benefit.

8.2 Expression and purification of recombinant GCK

The coding sequence for residues 16-465 of human GCK (Uniprot P35557) was cloned into a pETNKI-His-3C-LIC-kan vector to give the pETNKI-GCK plasmid. Residues 1-15, that were not included in our construct, are variable and at least three isoforms (Uniprot P35557-1, -2 and -3) have variations in this region without known functional difference [448]. The vector allows for the expression of GCK fused to an N-terminal hexahistidine tag as well as a cleavage site for the human rhinovirus 3C protease. The canonical cleavage site LEVLFQ/GP is cleaved between the glutamine and glycine [449] by this protease Therefore, by using this vector, this cleavage can be used to remove the hexahistidine tag.

GCK was produced recombinantly in *E. coli* BL21(DE3) cells and purified using nickel-affinity chromatography. Several labeling schemes were used in this study. Unlabeled GCK was produced to be used in enzymatic activity assays and in titration experiments with ^{15}N-labeled NS5A-D2. GCK that was uniformly ^{15}N-labeled and perdeuterated and selectively ^{13}C and ^{1}H-labeled on the δ1 methyl group of isoleucine

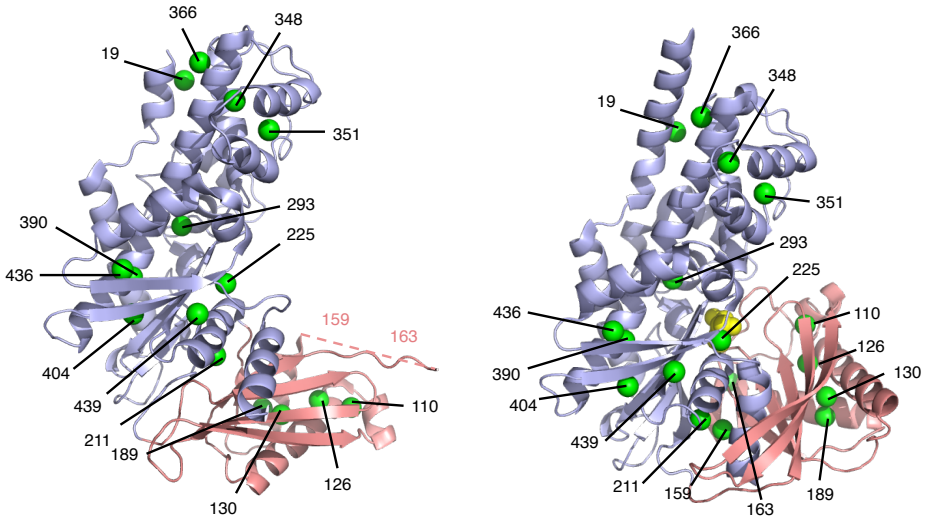

FIGURE 8.3: Distribution of isoleucines residues within GCK. GCK contains 17 isoleucines that are present in both subdomains and also within the 151-180 loop. **Left** Position of the isoleucines in unliganded GCK (PDB 1V4T). I159 and I163 are in the 151-180 loop that does not have electron density in this crystallographic structure. **Right** Position of the isoleucines in glucose-bound GCK (PDB 1V4S). All isoleucines are represented since the 151-180 loop becomes structured. Isoleucine δ1 carbons are shown as green spheres. The large subdomain is represented in light blue and the small subdomain in pink. Glucose is represented by yellow spheres.

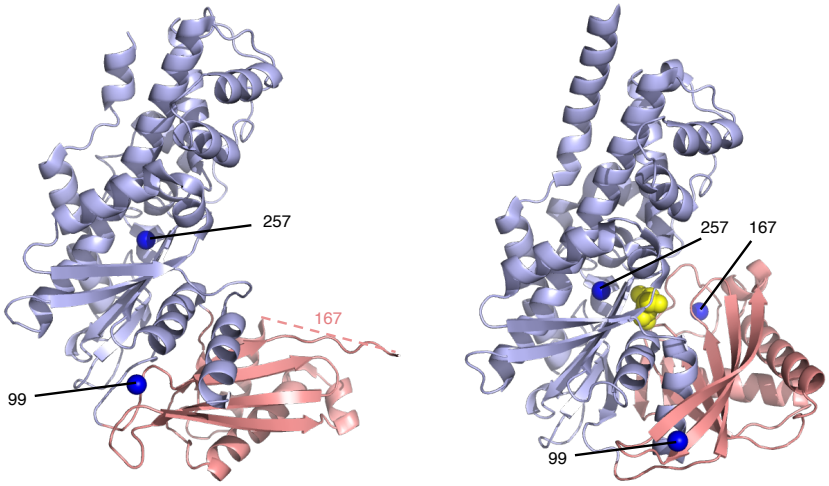

FIGURE 8.4: Distribution of tryptophan residues within GCK. GCK contains 3 tryptophans, these are in regions perturbed by glucose-binding. **Left** Position of the isoleucines in unliganded GCK (PDB 1V4T). W167 is in the 151-180 loop that does not have electron density in this crystallographic structure. **Right** Position of the tryptophans in glucose-bound GCK (PDB 1V4S). Tryptophan ε1 nitrogen atoms are shown as blue spheres. The large subdomain is represented in light blue and the small subdomain in pink. Glucose is represented by yellow spheres.

side chains and GCK selectively ^{15}N-labeled on the tryptophan side chain ε1 nitrogen were used to for most of the NMR experiments. GCK contains 17 isoleucines that are well distributed throughout the protein (Figure 8.3) thus the selective labeling of

these residues can provide information on various parts of the protein. Furthermore, it contains 3 tryptophans (Figure 8.4): Trp 99 in the small subdomain, Trp 167 in the disordered 151-180 loop and Trp 257 in the large subdomain. All of these residues are in regions of the protein that is affected by glucose binding.

The yield obtained for 1 L of culture is about 70 mg of unlabeled GCK; 20 mg of uniformly ^{15}N and deuterated, and selectively isoleucine δ1 ^{13}C and protonated GCK (Figure 8.5); and 14 mg of selectively tryptophan ε1-^{15}N-labeled.

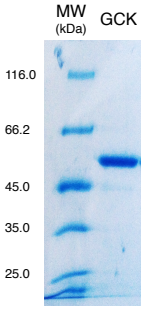

FIGURE 8.5: SDS-PAGE of U-[^{15}N,^{2}H], Ile δ1-[^{13}C^{1}H$_3$]-labeled GCK on a 10% acrylamide gel, stained with Coomassie blue dye.

Unless stated explicitly, all studies were done with GCK that had not been cleaved of its N-terminal purification tag.

8.3 Enzymatic activity of recombinant GCK

Our collaborators Patrice André, Vincent Lotteau and Oliver Diaz (Centre International de Recherche en Infectologie, Lyon) tested the activity of the recombinant GCK using a colorimetric enzymatic assay. This assay consists in coupling the hexokinase activity of GCK with the activity of glucose-6-phosphate dehydrogenase. This second enzyme oxidizes glucose-6-phosphate into 6-phosphogluconate by reducing NAD^+ to NADH (Figure 8.6). NADH can be detected by UV spectrophotometry at 340 nm to follow the reaction.

Firstly, they showed that the recombinant GCK was indeed active and can be used for further studies. Additionally they tested whether NS5A-D2 activates the enzyme. They showed that 1:3 ratio of GCK to NS5A-D2 causes an increase in reaction velocity that is over 1.5 times that of the enzyme alone (Figure 8.7).

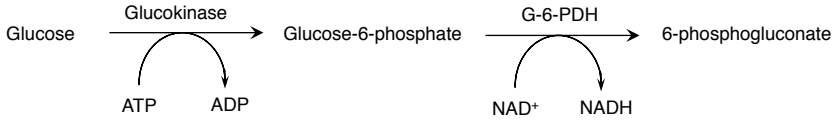

FIGURE 8.6: Coupled enzymatic reactions used to assay the activity of glucokinase. In the first step GCK converts glucose to glucose-6-phosphate consuming a molecule of ATP, followed by the conversion of glucose-6-phosphate to 6-phosphogluconate by Glucose-6-phosphate dehydrogenase. The second enzyme reduces a nicotinamide adenine dinucleotide from NAD^+ (oxidized form) to NADH (reduced form). The product NADH can be detected by UV spectrophotometry at 340 nm.

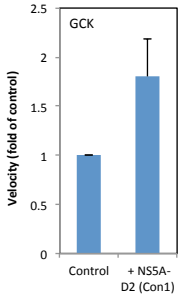

FIGURE 8.7: NS5A-D2 increases the enzymatic activity of GCK. Velocity of conversion of glucose to glucose-6-phosphate in the presence of a 3-fold excess of NS5A-D2 compared to GCK given as a proportion of the velocity of conversion of the control.

8.4 NMR characterization of GCK

As a first step in the investigation of the molecular mechanism of the interaction between NS5A-D2 and GCK we were interested in characterizing our recombinant GCK by NMR spectroscopy. Human GCK has previously been characterized using selective $^{1}H,^{13}C$-labeling on isoleucine $\delta 1$ side-chain methyl groups on an otherwise deuterated protein, as well as selective ^{15}N-labeling on tryptophan $\varepsilon 1$-nitrogen within the indole side-chains by Brüschweiler and colleagues [440, 443–445]. In these previous characterizations, the resonances in the spectra in the presence and absence of glucose have been assigned by point mutations. Furthermore, we used a similar buffer (50 mM potassium phosphate buffer pH 6.8, 50 mM KCl, 0.2 mM EDTA, 5 mM THP and 5% glycerol-d_5 in 99.9% D_2O or 50 mM potassium phosphate buffer pH 6.8, 50 mM KCl, 0.2 mM EDTA, 5 mM THP and 5% glycerol-d_5 in H_2O for the experiments with isoleucine-labeled and tryptophan-labeled GCK, respectively) to what was used in their study but with a higher concentration of buffer and salt to aid in the stability of the protein. Glycerol is needed for the stability of the protein but due to the high concentration needed and

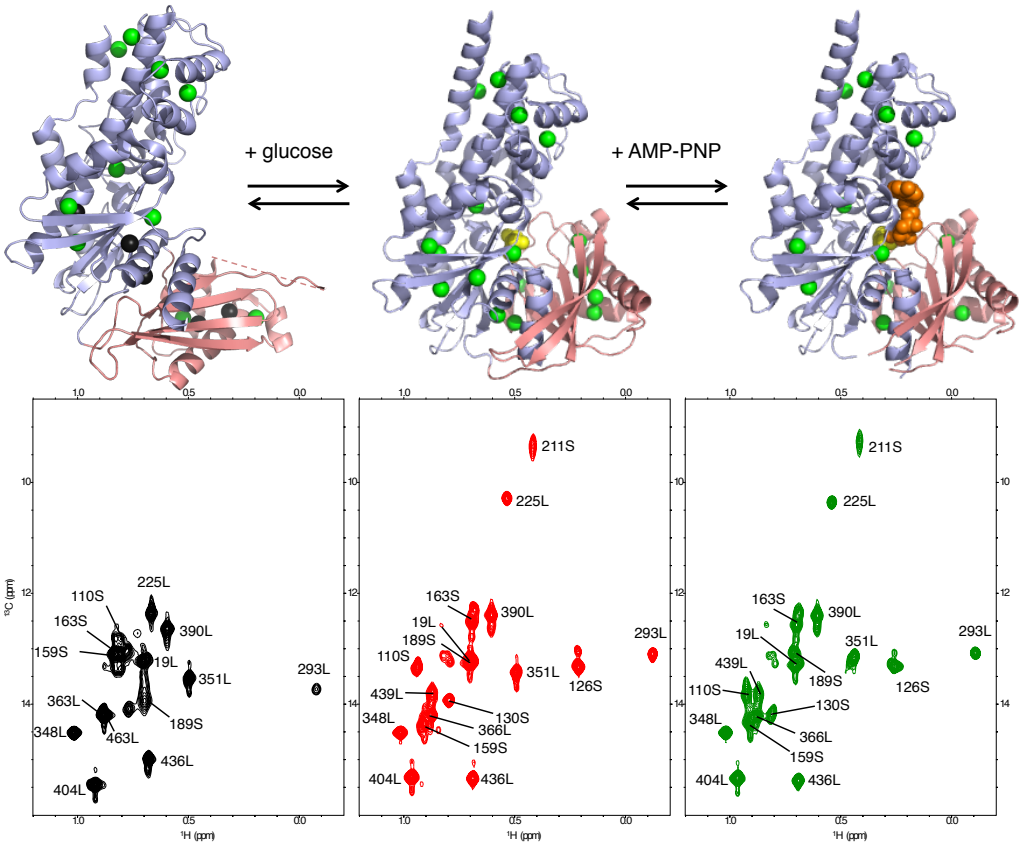

FIGURE 8.8: Methyl-HMQC spectra of U-^{2}H Ile δ1-{^{13}C^{1}H$_3$}-labeled GCK. Ile δ1 carbons are shown with spheres, green for those visible in the spectra and gray if too broad in the spectrum. Glucose and AMP-PNP are represented in yellow and orange respectively. Residues in the small domain are marked with an S and those in the large domain are identified with an L. In the spectrum of unliganded GCK (313 μM GCK, left) the resonances due to most residues in the small domain are weaker than the lowest contour level represented (3.5 $\times 10^6$ a.u). With the addition of 200 mM glucose (center) all isoleucines are present and their large shift suggests large structural changes. Finally, with the subsequent addition of 10 mM AMP-PNP (right) the spectral change is less pronounced.

because it contains protons that can result in large signals in the spectrum, deuterated glycerol was used to avoid the presence of these signals.

We tested whether our recombinant GCK provided spectra in good agreement with the one that was used in the prior studies to see if we could take advantage of the assignments. The methyl-HMQC spectra obtained for selectively ^{1}H,^{13}C isoleucine δ1 methyl groups are shown in Figure 8.8.

For unliganded GCK, the peak heights are very heterogeneous (Figure 8.9). There are two major differences, with respect to what was described by Larion *et al.* [443, 445], in the spectrum obtained for apo GCK. Firstly, the resonances due to I366 and I436 are

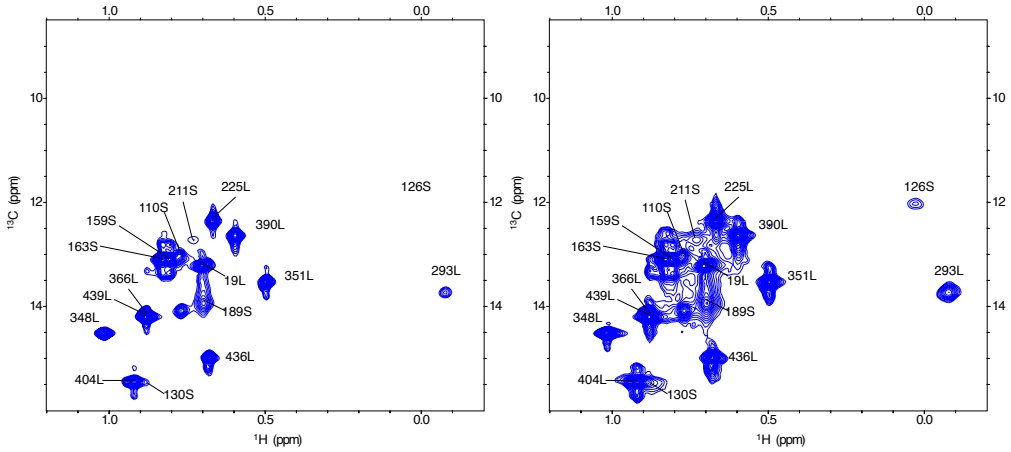

FIGURE 8.9: Resonances due to the residues in small subdomain are less intense. The spectrum of 313 μM U-[^{15}N,^{2}H]; Ileδ1-[^{13}C^{1}H$_3$]-labeled GCK is shown at two different lowest contour levels: 3.5 $\times 10^6$ a.u (right) and 1.5 $\times 10^6$ a.u (left). Some resonances from the small subdomain only become visible by lowering the contour levels. Resonances due to residues in the small and large subdomains are labeled with an S and an L, respectively.

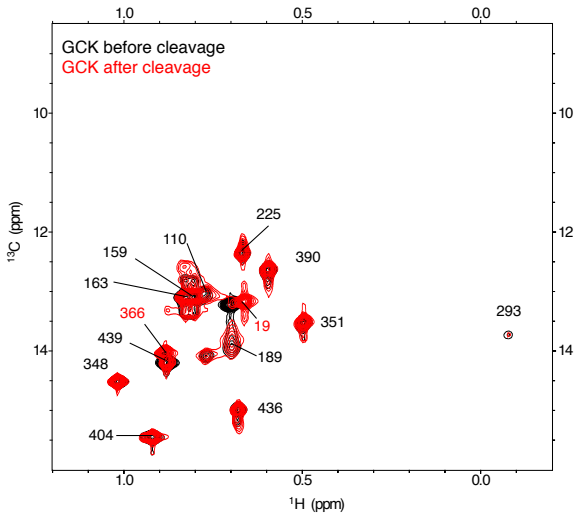

FIGURE 8.10: Cleavage of the hexahistidine tag from GCK. The methyl-HMQC of N-terminal hexahistidine tag-fused GCK is shown in black and in blue after cleavage to remove the tag. The resonances due to I19 and I366 shift because of their proximity to the tag.

overlapped in our spectrum and cannot be well discriminated as the spectra in the literature. Secondly, I19 appears to be shifted in our spectrum. We hypothesized that this is due to the varying constructs of GCK used in each study. In their work, Larion *et al.* used GCK fused with a hexahistidine tag at the C-terminus. Here, we used a construct where the purification tag is fused to the N-terminus of the protein. However, we have a cleavage site that allows us to remove the tag. The cleavage was done with human rhinovirus 3C protease fused to a GST tag to cleave at a LEVLFQ/GP site between the tag and the recombinant protein. This caused a chemical shift displacement in the I19

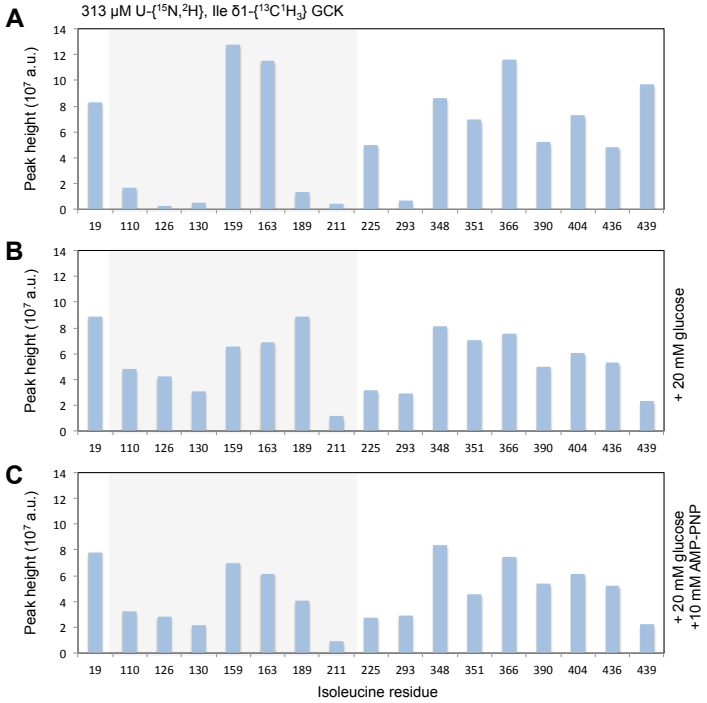

FIGURE 8.11: Change in peak heights upon the binding of glucose to GCK. **Top** Peak heights of resonances in the $^{13}C,^{1}H$-HMQC spectrum of unliganded U-{$^{15}N,^{2}H$} and Ile δ1-{$^{13}C^{1}H_3$}-labeled GCK. Resonances in the small domain are among the broadest of the spectrum except for the resonances due to I159 and I163 which are in the disordered loop. **Middle** Peak heights after the addition of 20 mM glucose to the previous sample. Peak heights become more homogeneous after the addition of glucose which has been attributed to stabilization of the small domain to give a closed conformation.**Bottom** Peak heights after the addition of the ATP analogue AMP-PNP to the glucose bound GCK sample. Only small perturbations are observed compared to the glucose-bound enzyme. A light gray box in the graphs highlights the residues in the small domain.

resonance (Figure 8.10) confirming its assignment. We also observe the separation of the overlapped resonances due to I439 and I366. Isoleucine I366 is oriented towards the N-terminal and therefore its resonance is shifted after the cleavage of the tag.

As was observed in the previous characterization of GCK, we observe large changes in the spectrum with the addition of glucose, congruent with a conformational change in the protein. Subsequent addition of adenylyl-imidodiphosphate (AMP-PNP), a non-hydrolysable analogue of ATP, does not alter the spectrum as significantly (Figures 8.11 and 8.12).

The reproducibility of the NMR experiment was also tested with the selective tryptophan labeling (Figure 8.13) and is in good agreement with the literature. Despite the high protein concentration and a long experiment time in the Trp-^{15}N-HSQC the signal-to-noise ratio of the spectrum obtained is low.

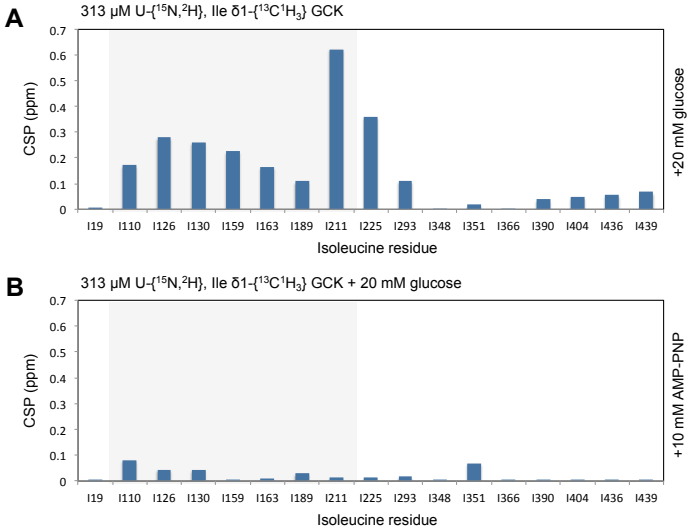

FIGURE 8.12: Combined chemical shift perturbations due to glucose and AMP-PNP. **Top** Chemical shift perturbations in the $^{13}C,^{1}H$-HMQC spectrum due to the addition of glucose. **Bottom** Chemical shift perturbations in the $^{13}C,^{1}H$-HMQC spectrum due to the addition of AMP-PNP to glucose-bound GCK. Glucose causes a large rearrangement in spectrum that has been proposed to be due to the stabilization of a closed conformation of GCK. The subsequent addition of an ATP analogue does not greatly change the spectrum suggesting that GCK does not undergo a large change in conformation with the binding of the ATP analogue.

8.5 Interaction between GCK and NS5A-D2

NS5A-D2 has been shown to be an activator of hexokinase II [296]. Our collaborators also showed that NS5A-D2 from the Con1 isolate increases GCK activity. Larion *et al.* [443] showed that the binding of a small-molecule activator to GCK causes a change in the methyl-HMQC spectrum of GCK and presents a spectrum that resembles that of glucose-bound GCK even in the absence of glucose. They proposed that the small molecule activator shifts the equilibrium in solution toward a closed active conformation. Based on this, our first hypothesis was that NS5A-D2 acted as an activator by also shifting the equilibrium in solution toward an active closed conformation.

To test this hypothesis, U-[$^{15}N,^{2}H$]; Ile δ1-[$^{13}C^{1}H_3$]-labeled GCK was titrated with increasing amounts of lyophilized unlabeled NS5A-D2 JFH-1. A small broadening of peaks, mainly in the small subdomain was observed in the spectrum in the absence (Figure 8.14) of glucose. Chemical shift perturbations after the addition of NS5A-D2 JFH-1 at a 1:3 ratio are smaller than 0.004 ppm for all resonances that were not overlapped. I130 appears to shift about 0.014 ppm but this is difficult to confirm as it is broad and appears close to I404.

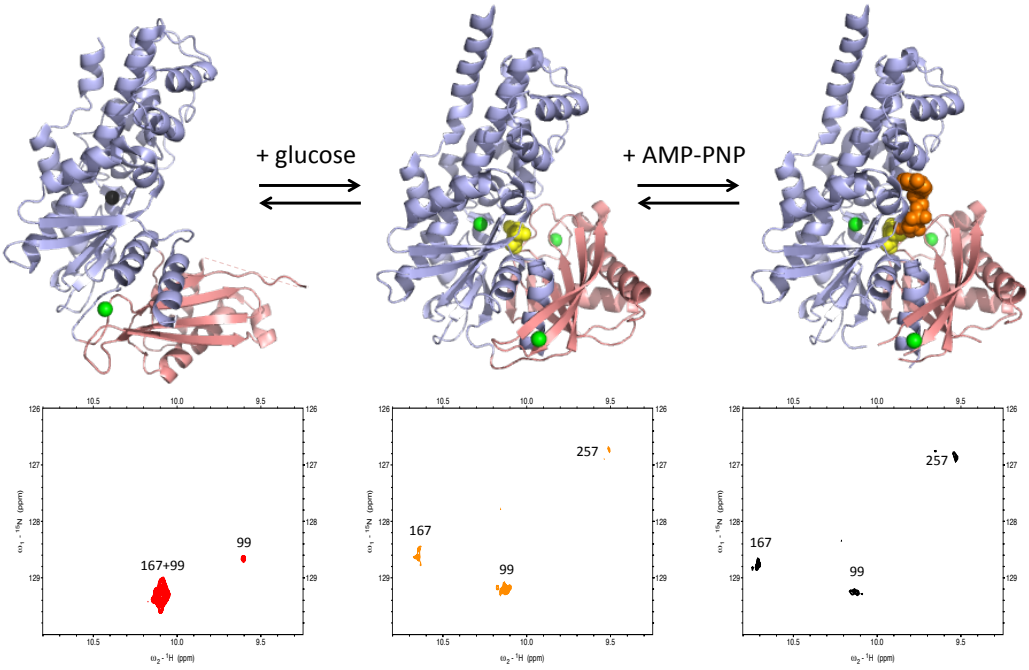

FIGURE 8.13: $^{1}H,^{15}N$-HSQC spectra of Trp ε1-{^{15}N} glucokinase. Trp ε1-N are shown with spheres, green for those visible in the spectra and gray if not present in the spectrum. Glucose and AMP-PNP are represented in yellow and orange, respectively. Unliganded GCK (315 μM GCK, left) has an intense resonance whose intensity is mostly due to Trp 167, present in the disordered β-loop (and therefore not visible in the crystal structure). With the addition of 200 mM glucose (center) all three tryptophans are visible with a large shift for Trp 167. Finally, with the subsequent addition of 10 mM AMP-PNP (right) there is no significant spectral change.

If NS5A-D2 acted as an activator by stabilizing the closed conformation we would expect to have observed a change in the NMR spectrum like the change observed with the addition of glucose. Therefore, NS5A-D2 must activate GCK by another mechanism that does not cause large conformational changes. Furthermore, isoleucine labeling does not introduce enough probes to identify a binding site on GCK.

Additionally, we tested whether NS5A-D2 JFH-1 affects glucose-bound GCK (Figure 8.15). Combined chemical shift perturbations are below 0.003 ppm and after adjusting for differences in GCK concentration, broadening is not observed.

Another hypothesis for the activation mechanism is that NS5A-D2 acts by a mechanism similar to the gain-of-function mutant described by Whittington [440]. These authors observe a 9-fold increase in $k_{cat}/K_{0.5}$ using this mutant but its spectrum does

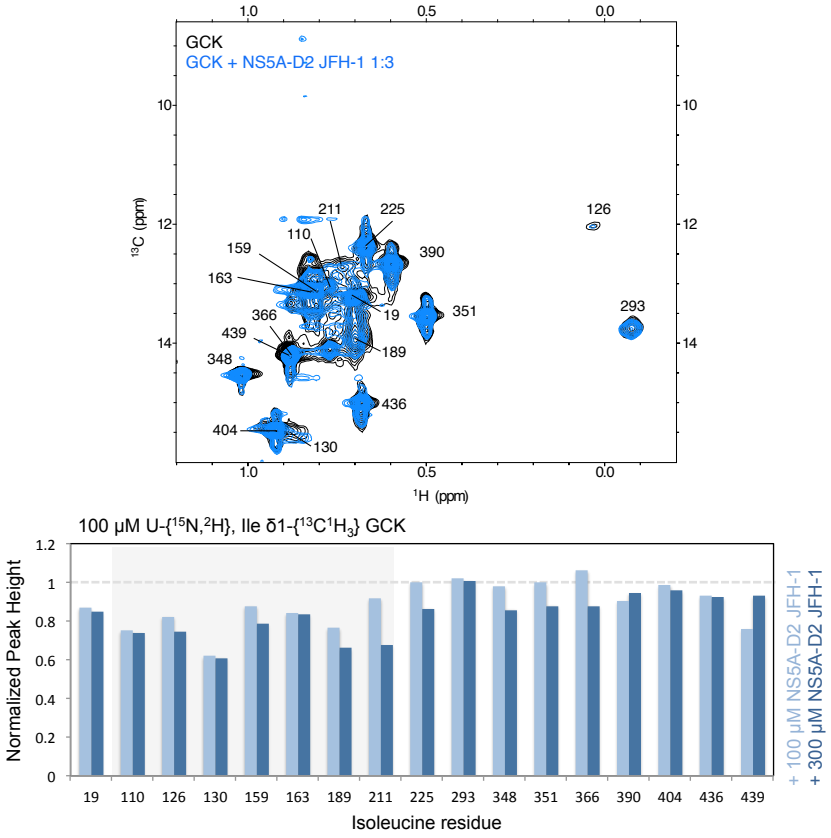

FIGURE 8.14: Titration of labeled GCK with unlabeled NS5A-D2 JFH-1. **Top** Overlay of ^{13}C,^{1}H-HMQC spectra of 100 μM perdeuterated, uniformly ^{15}N and selective labeled Ile δ1-[^{13}C^{1}H$_3$] GCK without and with 300 μM of NS5A-D2 JFH-1. **Bottom** Peak heights of spectra of 100 μM Ile δ1-[^{13}C^{1}H$_3$]-labeled GCK with 100 and 300 μM of NS5A-D2 JFH-1 relative to the peak heights of the spectrum in the absence of NS5A-D2. NS5A-D2 appears to affect mainly the residues in the small domain (highlighted with a light gray box).

not show differences compared to unliganded GCK. Its increased activity is due to increased mobility of the already disordered 151-180 loop which would allow the faster release of the product. If NS5A-D2 acts through this mechanism large changes in the NMR spectrum would not be observed because the effect is on the already disordered loop containing residues 159 and 163. Further enzymatic tests in the presence of NS5A-D2 are needed to support this hypothesis.

As the broadening observed on GCK is small following the addition of NS5A-D2 from the JFH-1 isolate we did the converse titration to confirm that there is an observable interaction between GCK and NS5A-D2. We added unlabeled GCK to ^{15}N-labeled NS5A-D2 from both JFH1 and Con1 strains (Figures 8.16 and 8.17). Moreover, this allows us to map the interaction between the two proteins on NS5A-D2.

We do not observe large chemical shift variations. The resonances that shift the

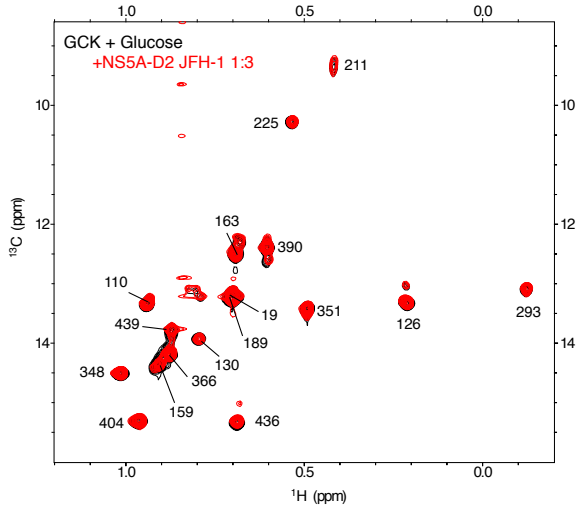

FIGURE 8.15: Effect of NS5A-D2 JFH-1 on glucose-bound GCK. Overlay of 2D ^{1}H-^{13}C HMQC NMR spectra of glucose-bound U-{^{15}N,^{2}H}; Ile δ1-{^{13}C^{1}H$_3$}-labeled GCK in the absence (black, 313 μM GCK and 20 mM glucose) and presence (red, 100 μM GCK, 10 mM glucose) of 300 μM NS5A-D2 JFH-1. Contours were adjusted to account for the difference in GCK concentration.

most corresponds to W316 (NS5A-D2 Con1) and W312 (NS5A-D2 JFH-1). We do, however, observe the effect of the interaction through broadening due to the large size of GCK. The regions implicated in this interaction are mainly regions that had been recognized by our group as having structure or structural tendencies, including both regions with helical tendencies and the PW-turn (where W316/312 is located), a small structural motif [110] that is also implicated in the interaction with cyclophilin A and NS5B. Interestingly, the effect of NS5A-D2 on GCK is more pronounced when using the protein from the Con1 isolate.

If the interaction between NS5A-D2 and GCK is dependent on the presence of structural tendencies, it is possible that this may be the reason we observe a difference between the viral protein from both isolates. NS5A-D2 from the Con1 isolate has been shown by our group to present higher SSP scores in its regions of secondary structure propensity and present more compact structures in SAXS analyses [109], which could suggest a mechanism where the already-formed secondary structure facilitate the interaction.

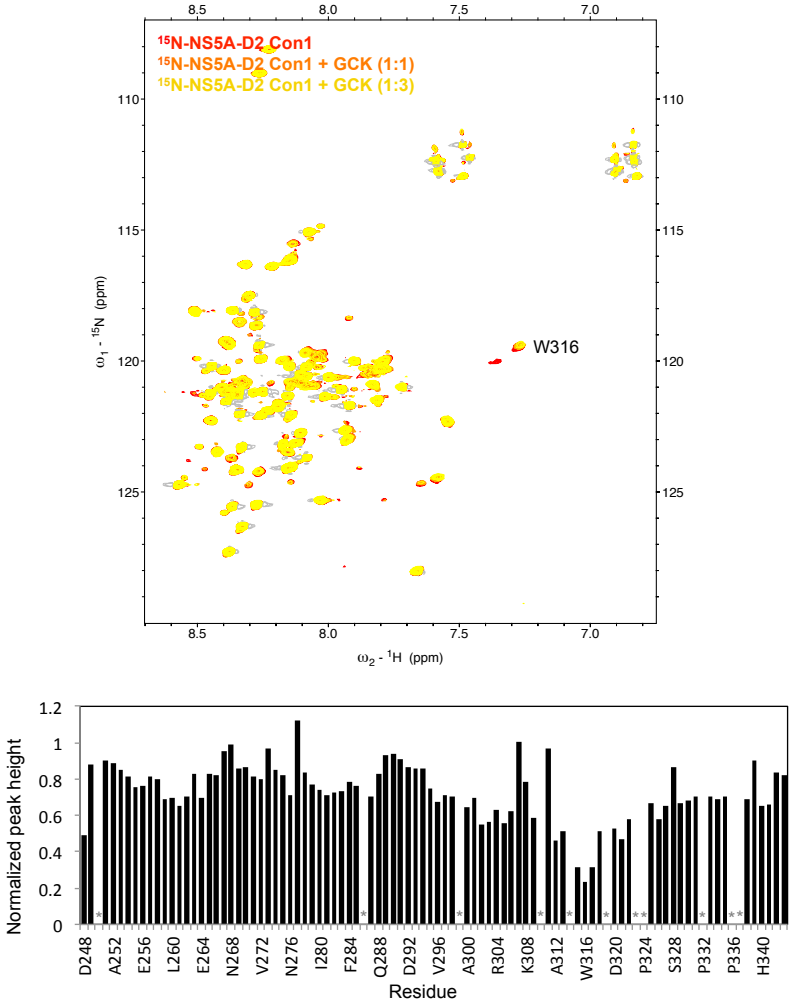

FIGURE 8.16: Titration of labeled NS5A-D2 Con1 with unlabeled GCK. **Top** ^{15}N,^{1}H-HSQC spectra of 87 μM ^{15}N-labeled NS5A-D2 Con1 with 0 (red), 83 (orange) and 166 (yellow) μM of unlabeled glucokinase. **Bottom** Peak intensity ratios between the spectrum of ^{15}N-NS5A-D2 Con1 in the presence of 2 molar equivalents and absence of GCK. We can observe broadening, notably around the PW-turn (^{314}PIWA^{317}) [110]. Prolines are shown as gray asterisks.

8.6 Conclusions

Previous studies had shown in HCV-infected hepatoma cells that the virus enhances glycolysis during infection. Furthermore, NS5A activates hexokinase II in these cells [296]. However, GCK is the main hexokinase isoform in non-tumorous hepatocytes. In this work we were interested in investigating the activation by the viral NS5A-D2 of the human enzyme glucokinase which is the main enzyme involved in the first step of glycolysis, the phosphorylation of glucose to glucose-6-phosphate, in hepatocytes.

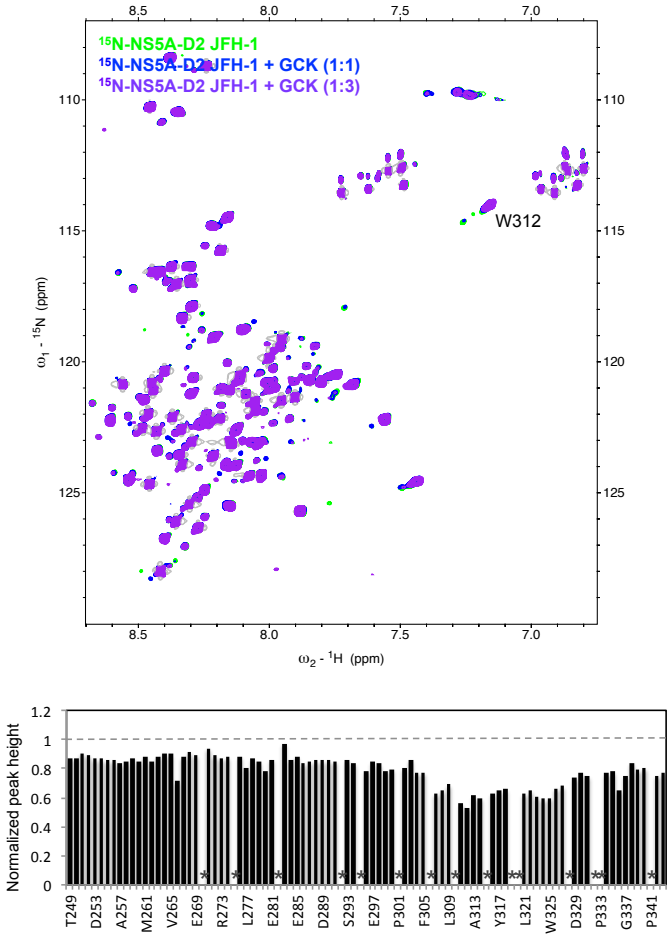

FIGURE 8.17: Titration of labeled NS5A-D2 JFH-1 with unlabeled GCK. **Top** ^{15}N,^{1}H-HSQC spectra of 87 μM ^{15}N-labeled NS5A-D2 JFH-1 with 0 (green), 87 (blue) and 174 (purple) μM of unlabeled glucokinase. **Bottom** Peak intensity ratios between the spectrum of ^{15}N-NS5A-D2 JFH-1 in the presence of 2 molar equivalents and absence of GCK. The broadening of resonances in the spectra is less pronounced in the JFH-1 isolate. Prolines are shown as gray asterisks.

Firstly, we produced recombinant glucokinase, both unlabeled and selectively labeled on isoleucine and tryptophan side-chains. Our collaborators tested the enzymatic activity of the recombinant GCK both in the absence and presence of NS5A-D2. They showed that the recombinant GCK is active and that NS5A-D2 from the Con1 isolate acts as an activator.

To investigate the mechanism in which NS5A-D2 acts as an activator of GCK by NMR spectroscopy, we first characterized our recombinant enzyme. The spectra obtained were in good agreement with spectra of GCK that had been previously described and therefore we could use the published assignments.

We had expected that the viral protein stabilizes an active conformation as in the

mechanism of action of small molecule GCK activators. Instead, the addition of NS5A-D2 from the JFH-1 isolate caused only small changes in the spectrum of the enzyme, mainly in the small domain.

To confirm this interaction and it interaction on the viral protein, the effect of the addition of GCK on the ^{1}H,^{15}N-HSQC spectrum of ^{15}N-labeled NS5A-D2 was observed. This was done for NS5A-D2 from both JFH-1 and Con1 isolates. Broadening occurred in regions with structure, mapping the interaction to these regions. Larger broadening is observed with NS5A-D2 from the Con-1 isolate. Since this protein has a higher tendency for the formation of secondary structure, this secondary structure may be needed for binding to GCK.

Our results suggest NS5A-D2 activates GCK by a different mechanism than the one described for small molecule activators of GCK that stabilize the active closed conformation of GCK. One hypothesis is that it acts by a mechanism akin to that described by Whittington *et al.* where the disordered 159–180 loop becomes more flexible, reducing the k_{cat} by releasing the product faster. With this mechanism, no large changes are expected in the NMR spectrum. Enzymatic assays that measure enzymatic parameters like $K_{0.5}$ and k_{cat} could be used to investigate this hypothesis. Furthermore, intra-methyl ^{1}H–^{1}H cross-correlated spin relaxation rates can be used as probes to investigate the amplitude of motion of methyl-containing side-chains [450], thus could also contribute to the testing of this hypothesis.

Further prospects in this investigation include testing the effect of NS5A-D2 from the Con1 isolate on labeled GCK since protein from this isolate showed a greater response to GCK in the converse titration. Tests to determine enzymatic parameters could provide insights into the activation mechanism, as well as the measuring of ^{1}H–^{1}H spin relaxation rates. Additionally, PRE experiments using the NS5A-D2 single cysteine mutants used in Chapter 5 could aid in mapping the interaction on GCK.

Small molecule glucokinase activators have been studied for application in the treatment of type 2 diabetes [451–453]. However, in phase II trials they have been shown to lose their efficacy after months of treatment, give rise to a high incidence of hypoglycemia and disturb lipid homeostatis [454]. If NS5A-D2 activates GCK through an independent mechanism, the investigation of its GCK activation mechanism might

provide insight into the development of novel GCK activators for use with type 2 diabetes.

General conclusions and perspectives

NS5B, the HCV RNA-dependent RNA polymerase

In this work we characterized the HCV polymerase NS5B and its interactions with biological partners using NMR spectroscopy in solution.

Firstly, NS5B$_{\Delta 21}$ was selectively labeled on isoleucine $\delta 1$ side-chain methyl groups to investigate the polymerase by NMR spectroscopy in solution using methyl-HMQC spectra. The methyl-HMQC spectrum of NS5B$_{\Delta 21}$ contains 20 resonances corresponding to a subset of the 24 isoleucines present in the protein. These resonances were assigned by a combination of NOESY experiments, isoleucine-to-valine site mutations and chemical shift predictions. The missing resonances highlight the presence of intermediate exchange in this protein.

Based on the comparison of chemical shift predictions from both open and closed crystallographic structures, NS5B$_{\Delta 21}$ appears to adopt mainly a closed conformation in solution but also samples open conformations. Overall, the polymerase is highly dynamic in solution and displays dynamics in multiple timescales.

Using methyl-HMQC spectra, we characterized constructs of NS5B that have been described as being more open or having modified open-close dynamics to attempt to identify a spectral signature of the opening of the polymerase. We identified that the resonances corresponding to I160, I419 and I432 display spectral changes in the open constructs but they do not display consistent chemical shift perturbations to use directly as a probe for the opening of the protein.

We were interested in the effect of NS5A-D2 on NS5B$_{\Delta 21}$. NS5A-D2 is an intrinsically disordered domain of HCV NS5A that had been shown to interact with the polymerase but the role of this interaction is not well understood. By NMR titration, we determined that the dissociation constant of this interaction is 70 μM. We also observed that the largest chemical shift perturbation was near the NNI-2 site on the resonance of I419 but spectral changes appear in all three subdomains of the polymerase.

As the spectral changes observed could be due to direct interactions or allosteric effects, intermolecular paramagnetic relaxation enhancement experiments were used

to determine the regions of the polymerase that are closest to NS5A-D2. This allowed us to identify two binding sites of NS5A-D2 on NS5B$_{\Delta 21}$ in the thumb subdomain, the thumb-I and thumb-II allosteric inhibitor binding sites and a preferential orientation for this interaction. However, multiple binding modes are possible for this interaction and these two proteins may form a "fuzzy complex". Due to the resonances of NS5B$_{\Delta 21}$ displaying spectral changes in the presence of NS5A-D2 and the results of the PRE experiments, NS5A-D2 appears to perturb residues in helix Q on the polymerase and fingertips Λ1 loop by binding to the thumb allosteric sites.

Since the thumb-II site is one of the binding sites of NS5A-D2 on the polymerase, we used the methyl-HMQC spectrum of NS5B$_{\Delta 21}$ to investigate the binding of a small molecule inhibitor that binds to this site, filibuvir. Our experimental data is compatible with the small molecule inhibiting NS5B activity by binding to the thumb-II site. This triggers perturbations via at least two allosteric pathways, including the helix Q and fingertips Λ1 loop that are also perturbed by NS5A-D2 but also through helix T that contains I512. Filibuvir perturbs the dynamics of the polymerase as CPMG experiments in the presence of filibuvir showed that the inhibitor reduces the μs–ms dynamics present in apo NS5B$_{\Delta}$ especially in the thumb subdomain. Furthermore, as filibuvir and NS5A-D2 appear to compete for a binding site on NS5B, it is possible that, at least in part, the inhibitory activity of filibuvir be via the perturbation of the interaction between the viral proteins.

Finally, we characterized the interaction between NS5B$_{\Delta 21}$ and a 16-mer strand of RNA using fluorescence and NMR spectroscopy and SEC. By fluorescence spectroscopy, a dissociation constant for this interaction of 0.89 μM was measured. We observed long-range effects of this interaction by NMR spectroscopy, observable on the resonances of I419 and I512 in the thumb subdomain and I23 in the fingertips Λ1.

Since the interaction between NS5A-D2 and RNA had been previously described for NS5A-D2 from the Con1 isolate [433], we tested whether NS5A-D2 JFH-1, which we used in this study, interacted with our RNA molecule. Experimental data obtained by SEC, fluorescence and NMR spectroscopy do not show an interaction between this protein and RNA.

Despite not interacting directly with RNA, NS5A-D2 decreases the affinity of NS5B$_{\Delta 21}$

for RNA. Since NS5A-D2 binds to sites known to bind allosteric inhibitors of $NS5B_{\Delta 21}$, this supports that the decreased affinity observed is via allosteric modulation of the polymerase. Likewise, filibuvir also reduces the capacity of $NS5B_{\Delta 21}$ of binding single-stranded RNA but does not completely eliminate it. Thus, filibuvir may have an inhibitory role by reducing RNA template-binding by the polymerase.

Altogether, the characterization of NS5B showed that this polymerase is highly dynamic in solution and that its conformation and/or dynamics can be perturbed via allosteric pathways that involve helices in the thumb subdomain and the fingertips loops. The thumb subdomain appears to be central to this modulation as resonances in this region are perturbed by NS5A-D2, filibuvir and RNA binding. Thus, this work provides new insights into the molecular properties of NS5B and its interactions with selected partners. Furthermore, it can act as a basis for the investigation of NS5B with other biological partners.

Perspectives

The characterization of $NS5B_{\Delta 21}$ selectively labeled isoleucine $\delta 1$ methyl groups in solution by NMR spectroscopy provides a potential starting point for multiple studies with the polymerase.

Firstly, we showed that NS5A-D2 binding both thumb-I and thumb-II inhibitor sites. We characterized the binding of a thumb-II inhibitor. Similarly, the characterization of the binding of a thumb-I inhibitor can provide further insight into the allosteric pathways in the polymerase. The spectrum of the polymerase with an NNI-1 can be compared to the spectrum in the presence of NS5A-D2 and $NS5B_{\Delta 21}$ L30S, as the effect of this mutation has been described as similar to the effect of these inhibitors, to see if similar resonances are perturbed. Furthermore, a spectrum of ^{15}N-labeled NS5A-D2 with apo $NS5B_{\Delta 21}$ and NNI-1-bound $NS5B_{\Delta 21}$ could help identify the region of NS5A-D2 that binds to the polymerase thumb-I site.

Secondly, a characterization of the μs–ms dynamics of the polymerase by CPMG relaxation dispersion experiments in the presence of NS5A-D2 may provide insight into the perturbations caused by NS5A-D2 to the dynamics of NS5B in this timescale.

Thirdly, the effect of NS5A-D2 on open constructs of NS5B both by NMR and by DSF can provide insight into the mechanism of allosteric modulation caused by the disordered protein. If NS5A-D2 causes an opening of the polymerase, we may observe that long-range effects of the binding are less present when using an open construct. Similarly, we may not observe a decrease in the stability of the polymerase in the presence of NS5A-D2 with these constructs.

The main limitation of the methyl-HMQC spectrum are the limited number of probes present on the polymerase and the lack of probes in structural elements like the β-loop and the C-terminal. The labeling of other side-chain methyl groups is conceivable. There are 55 leucines and 39 valines in NS5B$_{\Delta 21}$. The assignment of the isoleucine resonances can aid in the assignment using NOESY spectra. This could significantly reduce the number of potential single-site mutations necessary for the assignment of spectra with these side-chain methyl groups labeled. More probes could aid in identifying independently moving parts in solution to compare to the study using crystallographic structures by Caillet-Saguy *et al.* [158].

Furthermore, extending this study to NS5B from other isolates can contribute to the understanding of the differences between HCV isolates. NS5B from isolates that are more representative of patient infections or harder to treat genotypes are of particular interest.

Regulation of glucokinase by NS5A-D2

We also investigated the interaction between HCV NS5A-D2 and an enzyme from the glycolysis pathway, GCK. NS5A-D2 activates GCK activity by a mechanism that is distinct from the mechanism described for small molecule activators. Small molecule inhibitors lock the enzyme in a closed active conformation, which causes a significant change in the methyl-HMQC spectrum compared to unliganded GCK [443]. Instead, the spectrum of GCK with NS5A-D2 JFH-1 is similar to the spectrum of unliganded GCK with broadening in the small domain. Thus, on GCK, the binding site of NS5A-D2 was mapped to the small domain.

Conversely, the binding site of GCK on NS5A-D2 is mainly surrounding the PW-turn structural motif that has been implicated in other interactions of the viral protein. This is the case for both NS5A-D2 from JFH-1 and Con1 isolates, however, NS5A-D2 Con1 displays a stronger interaction with GCK. Since NS5A-D2 Con1 displays larger secondary structural propensity than NS5A-D2 JFH-1 [109] it is possible that protein-protein interaction occurs through a conformational selection mechanism.

As the activation appears to occur with a mechanism that has not yet been described for activators, the investigation of this interaction and its mechanism can potentially lead to the development of novel activators of GCK.

Perspectives

In the study of the interaction between GCK and NS5A-D2, we observed using ^{15}N-labeled NS5A-D2 that the viral protein from the Con1 isolate interacts more strongly than the one from JFH-1. Methyl-HMQC spectra of GCK in the presence of NS5A-D2 Con1 could confirm this. Furthermore, PRE experiments with a single-cysteine mutant of NS5A-D2 could help investigate whether the small domain is indeed the region of GCK that interacts with NS5A-D2.

Another perspective for this study is the investigation of the mechanism of activation. Potentially it may act in a similar way as the mutant described by Whittington *et al.* [440] that displays increased flexibility of the disordered loop. Enzymatic assays could be used to investigate this hypothesis. Furthermore, intra-methyl ^{1}H–^{1}H dipolar cross-correlated spin relaxation rates can be used to determine order parameters and thus the amplitude of motion of methyl side-chains which would also aid in the investigation of this mechanism.

Resumé substantiel en français

Le virus de l'hépatite C (VHC) est un virus à ARN simple brin responsable d'hépatites aiguës et chroniques, lesquelles peuvent évoluer vers une cirrhose du foie et un carcinome hépatocellulaire. Les antiviraux développés récemment sont très efficaces, ils ciblent les protéines virales et en particulier NS5B, l'ARN polymérase. La structure de NS5B a beaucoup été étudiée par radiocristallographie (RX). Elle contient trois sous-domaines appelés doigts, paume et pouce par analogie avec une main droite. Ces structures révèlent une conformation de NS5B incompatible avec la liaison d'un ARN double-brin, ce qui est nécessaire pour son activité de polymérase. Par conséquence, la protéine doit être capable de changer entre des conformations ouvertes et fermées. Tandis que des données statiques de cristallographie sont abondantes, les études structurales de cette protéine en solution sont très limitées.

Nous avons utilisé la résonance magnétique nucléaire (RMN) pour caractériser NS5B en solution. Cette approche permet d'examiner la dynamique locale des sous-domaines et d'étudier des interactions avec des molécules qui lient NS5B avec une faible affinité ou avec des protéines intrinsèquement désordonnées qui ne peuvent pas être étudiées par cristallographie.

La caractérisation de NS5B (65kDa) par RMN a été faite avec la protéine perdeuterée et sélectivement enrichie (^{13}C, ^{1}H) au niveau des méthyles δ1 des résidus d'isoleucines. Cette stratégie permet d'obtenir un spectre simplifié, avec un nombre réduit de signaux, ainsi qu'un bon rapport signal sur bruit. Cette étude a confirmé la présence d'une dynamique particulière dans le sous-domaine pouce et a permis de mettre en évidence des effets à longues distances qui se transmettent aux autres sous-domaines.

Nous avons alors utilisé cette approche expérimentale pour étudier l'interaction entre NS5B et le domaine 2 de la protéine NS5A (NS5A-D2). Ce dernier est un domaine intrinsèquement désordonné d'une autre protéine du VHC qui est nécessaire pour la réplication du virus et dont l'interaction directe avec NS5B a été montrée in vitro. En utilisant le spectre RMN de NS5B, son interaction avec NS5A-D2 a été caractérisée et cartographiée sur la structure de NS5B. Pour distinguer entre les effets directs et les effets induits par l'interaction, des expériences RMN de type paramagnetic relaxation enhancement ont été utilisées. Ainsi nous avons déterminé que NS5A-D2 se lie via,

au moins, deux sites de liaison (pouce-I et –II) sur le sous-domaine du pouce. Ceux-ci correspondent à des sites de liaison d'inhibiteurs allostériques qui provoquent des changements de la conformation et la dynamique de la polymérase.

Le spectre RMN de NS5B a été aussi été utilisé pour caractériser la liaison du Filibuvir, un inhibiteur allostérique qui se lie au site pouce-II. La structure RX de la polymérase liée à cette molécule n'est pas significativement différente de celle de NS5B apo. Cependant, dans notre étude RMN en solution a permis de montrer que l'inhibiteur provoque des perturbations dans tous les sous-domaines de la protéine. De plus, par compétition avec le Filibuvir, nous avons identifié la zone de NS5A-D2 qui se lie à ce site pouce-II. Enfin, par RMN nous avons observé que la liaison d'un ARN simple brin dans le site actif de NS5B provoquait des effets à longues distances sur le pouce. Nous avons pu constater que NS5A-D2 et Filibuvir réduisent l'affinité de NS5B pour la matrice d'ARN mais ne l'éliminent pas.

L'analyse de NS5B par RMN en solution a permis d'étudier des interactions et d'accéder à des paramètres dynamiques très complémentaires des études cristallographiques. D'autres interactions entre NS5B et différents partenaires (molécules, protéines…) pourront ensuite être caractérisées grâce à cette étude.

Bibliography

1. Feinstone, S. M., Kapikian, A. Z., Purcell, R. H., Alter, H. J. & Holland, P. V. Transfusion-Associated Hepatitis Not Due to Viral Hepatitis Type A or B. *New England Journal of Medicine* **292,** 767–770. ISSN: 0028-4793 (1975).
2. Alter, H., Holland, P., Purcell, R. & Popper, H. TRANSMISSIBLE AGENT IN NON-A, NON-B HEPATITIS. *The Lancet. Originally published as Volume 1, Issue 8062* **311,** 459–463. ISSN: 0140-6736 (1978).
3. Choo, Q. L., Kuo, G., Weiner, A. J., Overby, L. R., Bradley, D. W. & Houghton, M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* **244,** 359–362. ISSN: 0036-8075, 1095-9203 (1989).
4. Houghton, M. Discovery of the hepatitis C virus. *Liver International* **29,** 82–88. ISSN: 1478-3231 (2009).
5. Kolykhalov, A. A., Agapov, E. V., Blight, K. J., Mihalik, K., Feinstone, S. M. & Rice, C. M. Transmission of Hepatitis C by Intrahepatic Inoculation with Transcribed RNA. *Science* **277,** 570–574. ISSN: 0036-8075, 1095-9203 (1997).
6. World Health Organization. *Guidelines for the screening, care and treatment of persons with chronic hepatitis C infection. Updated Version, April 2016* tech. rep. ISBN 978 92 4 154961 5 (World Health Organization, 2016), 138. `http://www.who.int/hepatitis/publications/hepatitis-c-guidelines-2016/en/` (2017).
7. Yeung, C.-Y., Lee, H.-C., Chan, W.-T., Jiang, C.-B., Chang, S.-W. & Chuang, C.-K. Vertical transmission of hepatitis C virus: Current knowledge and perspectives. *World Journal of Hepatology* **6,** 643–651. ISSN: 1948-5182 (2014).
8. Pybus, O. G., Charleston, M. A., Gupta, S., Rambaut, A., Holmes, E. C. & Harvey, P. H. The Epidemic Behavior of the Hepatitis C Virus. *Science* **292,** 2323–2325. ISSN: 0036-8075, 1095-9203 (2001).
9. Iles, J. C., Raghwani, J., Harrison, G. L. A., Pepin, J., Djoko, C. F., Tamoufe, U., LeBreton, M., Schneider, B. S., Fair, J. N., Tshala, F. M., Kayembe, P. K., Muyembe, J. J., Edidi-Basepeo, S., Wolfe, N. D., Simmonds, P., Klenerman, P. & Pybus, O. G. Phylogeography and epidemic history of hepatitis C virus genotype 4 in Africa. *Virology* **464–465,** 233–243. ISSN: 0042-6822 (2014).
10. Lozano, R. *et al.* Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet* **380,** 2095–2128. ISSN: 0140-6736, 1474-547X (2012).
11. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet* **385,** 117–171. ISSN: 0140-6736, 1474-547X (2015).
12. Gower, E., Estes, C., Blach, S., Razavi-Shearer, K. & Razavi, H. Global epidemiology and genotype distribution of the hepatitis C virus infection. *Journal of Hepatology* **61,** S45–S57. ISSN: 0168-8278, 1600-0641 (2014).
13. Mohd Hanafiah, K., Groeger, J., Flaxman, A. D. & Wiersma, S. T. Global epidemiology of hepatitis C virus infection: New estimates of age-specific antibody to HCV seroprevalence. *Hepatology* **57,** 1333–1342. ISSN: 1527-3350 (2013).
14. Strickland, G. T. Liver disease in Egypt: Hepatitis C superseded schistosomiasis as a result of iatrogenic and biological factors. *Hepatology* **43,** 915–922. ISSN: 1527-3350 (2006).

15. Frank, C., Mohamed, M. K., Strickland, G. T., Lavanchy, D., Arthur, R. R., Magder, L. S., Khoby, T. E., Abdel-Wahab, Y., Ohn, E. S. A., Anwar, W. & Sallam, I. The role of parenteral antischistosomal therapy in the spread of hepatitis C virus in Egypt. *The Lancet* **355,** 887–891. ISSN: 0140-6736, 1474-547X (2000).

16. Jaeckel, E., Cornberg, M., Wedemeyer, H., Santantonio, T., Mayer, J., Zankel, M., Pastore, G., Dietrich, M., Trautwein, C. & Manns, M. P. Treatment of Acute Hepatitis C with Interferon Alfa-2b. *New England Journal of Medicine* **345,** 1452–1457. ISSN: 0028-4793 (2001).

17. Wiegand, J., Buggisch, P., Boecher, W., Zeuzem, S., Gelbmann, C. M., Berg, T., Kauffmann, W., Kallinowski, B., Cornberg, M., Jaeckel, E., Wedemeyer, H. & Manns, M. P. Early monotherapy with pegylated interferon alpha-2b for acute hepatitis C infection: The HEP-NET acute-HCV-II study. *Hepatology* **43,** 250–256. ISSN: 1527-3350 (2006).

18. Deterding, K., Grüner, N., Buggisch, P., Wiegand, J., Galle, P. R., Spengler, U., Hinrichsen, H., Berg, T., Potthoff, A., Malek, N., Großhennig, A., Koch, A., Diepolder, H., Lüth, S., Feyerabend, S., Jung, M. C., Rogalska-Taranta, M., Schlaphoff, V., Cornberg, M., Manns, M. P. & Wedemeyer, H. Delayed versus immediate treatment for patients with acute hepatitis C: a randomised controlled non-inferiority trial. *The Lancet Infectious Diseases* **13,** 497–506. ISSN: 1473-3099, 1474-4457 (2013).

19. Heim, M. H. 25 years of interferon-based treatment of chronic hepatitis C: an epoch coming to an end. *Nature Reviews Immunology* **13,** 535–542. ISSN: 1474-1733 (2013).

20. Hernandez-Gea, V. & Friedman, S. L. Pathogenesis of Liver Fibrosis. *Annual Review of Pathology: Mechanisms of Disease* **6,** 425–456 (2011).

21. Yamane, D., McGivern, D. R., Masaki, T. & Lemon, S. M. Liver injury and disease pathogenesis in chronic hepatitis C. *Current Topics in Microbiology and Immunology* **369,** 263–288. ISSN: 0070-217X (2013).

22. Westbrook, R. H. & Dusheiko, G. Natural history of hepatitis C. *Journal of Hepatology* **61,** S58–S68. ISSN: 0168-8278 (2014).

23. Brenner, D. A. Molecular Pathogenesis of Liver Fibrosis. *Transactions of the American Clinical and Climatological Association* **120,** 361–368. ISSN: 0065-7778 (2009).

24. Bataller, R., Paik, Y.-h., Lindquist, J. N., Lemasters, J. J. & Brenner, D. A. Hepatitis C virus core and nonstructural proteins induce fibrogenic effects in hepatic stellate cells. *Gastroenterology* **126,** 529–540. ISSN: 0016-5085, 1528-0012 (2004).

25. Smith, D. B., Becher, P., Bukh, J., Gould, E. A., Meyers, G., Monath, T., Muerhoff, A. S., Pletnev, A., Rico-Hesse, R., Stapleton, J. T. & Simmonds, P. Proposed update to the taxonomy of the genera Hepacivirus and Pegivirus within the Flaviviridae family. *Journal of General Virology* **97,** 2894–2907 (2016).

26. International Committee on Taxonomy of Viruses. *ICTV Taxonomy* 2017. `https://talk.ictvonline.org/taxonomy/` (2017).

27. Simons, J. N., Pilot-Matias, T. J., Leary, T. P., Dawson, G. J., Desai, S. M., Schlauder, G. G., Muerhoff, A. S., Erker, J. C., Buijk, S. L. & Chalmers, M. L. Identification of two flavivirus-like genomes in the GB hepatitis agent. *Proceedings of the National Academy of Sciences of the United States of America* **92,** 3401–3405. ISSN: 0027-8424 (1995).

28. Simons, J. N., Leary, T. P., Dawson, G. J., Pilot-Matias, T. J., Muerhoff, A. S., Schlauder, G. G., Desai, S. M. & Mushahwar, I. K. Isolation of novel virus-like sequences associated with human hepatitis. *Nature Medicine* **1,** 564–569. ISSN: 1078-8956 (1995).

29. Lauck, M., Sibley, S. D., Lara, J., Purdy, M. A., Khudyakov, Y., Hyeroba, D., Tumukunde, A., Weny, G., Switzer, W. M., Chapman, C. A., Hughes, A. L., Friedrich, T. C., O'Connor, D. H. & Goldberg, T. L. A Novel Hepacivirus with an Unusually Long and Intrinsically Disordered NS5A Protein in a Wild Old World Primate. *Journal of Virology* **87,** 8971–8981. ISSN: 0022-538X, 1098-5514 (2013).

30. Lyons, S., Kapoor, A., Sharp, C., Schneider, B. S., Wolfe, N. D., Culshaw, G., Corcoran, B., McGorum, B. C. & Simmonds, P. Nonprimate Hepaciviruses in Domestic Horses, United Kingdom. *Emerging Infectious Diseases* **18,** 1976–1982 (2012).

31. Tanaka, T., Kasai, H., Yamashita, A., Okuyama-Dobashi, K., Yasumoto, J., Maekawa, S., Enomoto, N., Okamoto, T., Matsuura, Y., Morimatsu, M., Manabe, N., Ochiai, K., Yamashita, K. & Moriishi, K. Hallmarks of Hepatitis C Virus in Equine Hepacivirus. *Journal of Virology* **88,** 13352–13366. ISSN: 0022-538X, 1098-5514 (2014).

32. Quan, P.-L. *et al.* Bats are a major natural reservoir for hepaciviruses and pegiviruses. *Proceedings of the National Academy of Sciences* **110,** 8194–8199. ISSN: 0027-8424, 1091-6490 (2013).

33. Kapoor, A., Simmonds, P., Scheel, T. K. H., Hjelle, B., Cullen, J. M., Burbelo, P. D., Chauhan, L. V., Duraisamy, R., Leon, M. S., Jain, K., Vandegrift, K. J., Calisher, C. H., Rice, C. M. & Lipkin, W. I. Identification of Rodent Homologs of Hepatitis C Virus and Pegiviruses. *mBio* **4,** e00216–13. ISSN: , 2150-7511 (2013).

34. Drexler, J. F. *et al.* Evidence for Novel Hepaciviruses in Rodents. *PLOS Pathogens* **9,** e1003438. ISSN: 1553-7374 (2013).

35. Baechlein, C., Fischer, N., Grundhoff, A., Alawi, M., Indenbirken, D., Postel, A., Baron, A. L., Offinger, J., Becker, K., Beineke, A., Rehage, J. & Becher, P. Identification of a Novel Hepacivirus in Domestic Cattle from Germany. *Journal of Virology* **89,** 7007–7015. ISSN: 0022-538X (2015).

36. El-Attar, L. M. R., Mitchell, J. A., Brooks Brownlie, H., Priestnall, S. L. & Brownlie, J. Detection of non-primate hepaciviruses in UK dogs. *Virology* **484,** 93–102. ISSN: 0042-6822 (2015).

37. Thézé, J., Lowes, S., Parker, J. & Pybus, O. G. Evolutionary and Phylogenetic Analysis of the Hepaciviruses and Pegiviruses. *Genome Biology and Evolution* **7,** 2996–3008. ISSN: , 1759-6653 (2015).

38. Simmonds, P. The origin of hepatitis C virus. *Current Topics in Microbiology and Immunology* **369,** 1–15. ISSN: 0070-217X (2013).

39. Pybus, O. G. & Thézé, J. Hepacivirus cross-species transmission and the origins of the hepatitis C virus. *Current Opinion in Virology. Emerging viruses• Viral immunology* **16,** 1–7. ISSN: 1879-6257 (2016).

40. Smith, D. B., Bukh, J., Kuiken, C., Muerhoff, A. S., Rice, C. M., Stapleton, J. T. & Simmonds, P. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: Updated criteria and genotype assignment web resource. *Hepatology* **59,** 318–327. ISSN: 1527-3350 (2014).

41. Messina, J. P., Humphreys, I., Flaxman, A., Brown, A., Cooke, G. S., Pybus, O. G. & Barnes, E. Global distribution and prevalence of hepatitis C virus genotypes. *Hepatology* **61,** 77–87. ISSN: 1527-3350 (2015).

42. Murphy, D. G., Sablon, E., Chamberland, J., Fournier, E., Dandavino, R. & Tremblay, C. L. Hepatitis C Virus Genotype 7, a New Genotype Originating from Central Africa. *Journal of Clinical Microbiology* **53,** 967–972. ISSN: 0095-1137, 1098-660X (2015).

43. Fried, M. W., Shiffman, M. L., Reddy, K. R., Smith, C., Marinos, G., Gonçales, F. L. J., Häussinger, D., Diago, M., Carosi, G., Dhumeaux, D., Craxi, A., Lin, A., Hoffman, J. & Yu, J. Peginterferon Alfa-2a plus Ribavirin for Chronic Hepatitis C Virus Infection. *New England Journal of Medicine* **347,** 975–982. ISSN: 0028-4793 (2002).

44. Dubuisson, J. & Cosset, F.-L. Virology and cell biology of the hepatitis C virus life cycle – An update. *Journal of Hepatology* **61,** S3–S13. ISSN: 0168-8278 (2014).

45. Catanese, M. T., Uryu, K., Kopp, M., Edwards, T. J., Andrus, L., Rice, W. J., Silvestry, M., Kuhn, R. J. & Rice, C. M. Ultrastructural analysis of hepatitis C virus particles. *Proceedings of the National Academy of Sciences of the United States of America* **110,** 9505–9510. ISSN: 0027-8424 (2013).

46. André, P., Komurian-Pradel, F., Deforges, S., Perret, M., Berland, J. L., Sodoyer, M., Pol, S., Bréchot, C., Paranhos-Baccalà, G. & Lotteau, V. Characterization of Low- and Very-Low-Density Hepatitis C Virus RNA-Containing Particles. *Journal of Virology* **76,** 6919–6928. ISSN: 0022-538X, 1098-5514 (2002).

47. Lindenbach, B. D., Meuleman, P., Ploss, A., Vanwolleghem, T., Syder, A. J., McKeating, J. A., Lanford, R. E., Feinstone, S. M., Major, M. E., Leroux-Roels, G. & Rice, C. M. Cell culture-grown hepatitis C virus is infectious in vivo and can be recultured in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **103,** 3805–3809. ISSN: 0027-8424 (2006).

48. Bartenschlager, R., Penin, F., Lohmann, V. & André, P. Assembly of infectious hepatitis C virus particles. *Trends in Microbiology* **19,** 95–103. ISSN: 0966-842X, 1878-4380 (2011).

49. Piver, E., Boyer, A., Gaillard, J., Bull, A., Beaumont, E., Roingeard, P. & Meunier, J.-C. Ultrastructural organisation of HCV from the bloodstream of infected patients revealed by electron microscopy after specific immunocapture. *Gut,* gutjnl–2016–311726. ISSN: 0017-5749, 1468-3288 (2016).

50. André, P., Perlemuter, G., Budkowska, A., Bréchot, C. & Lotteau, V. Hepatitis C Virus Particles and Lipoprotein Metabolism. *Seminars in Liver Disease* **25,** 93–104. ISSN: 0272-8087, 1098-8971 (2005).

51. Icard, V., Diaz, O., Scholtes, C., Perrin-Cocon, L., Ramière, C., Bartenschlager, R., Penin, F., Lotteau, V. & André, P. Secretion of Hepatitis C Virus Envelope Glycoproteins Depends on Assembly of Apolipoprotein B Positive Lipoproteins. *PLOS ONE* **4,** e4233. ISSN: 1932-6203 (2009).

52. Niepmann, M. *Hepatitis C Virus: From Molecular Virology to Antiviral Therapy* (ed Bartenschlager, R.) *Current Topics in Microbiology and Immunology* 369. DOI: 10.1007/978-3-642-27340-7_6, 143–166 (Springer Berlin Heidelberg, 2013). ISBN: 978-3-642-27339-1 978-3-642-27340-7.

53. Friebe, P., Lohmann, V., Krieger, N. & Bartenschlager, R. Sequences in the 5 Nontranslated Region of Hepatitis C Virus Required for RNA Replication. *Journal of Virology* **75,** 12047–12057. ISSN: 0022-538X, 1098-5514 (2001).

54. Sedano, C. D. & Sarnow, P. Hepatitis C virus subverts liver-specific miR-122 to protect the viral genome from exoribonuclease Xrn2. *Cell Host & Microbe* **16,** 257–264. ISSN: 1934-6069 (2014).

55. Jaafar, Z. A., Oguro, A., Nakamura, Y. & Kieft, J. S. Translation initiation by the hepatitis C virus IRES requires eIF1A and ribosomal complex remodeling. *eLife* **5,** e21198. ISSN: 2050-084X (2016).

56. Quade, N., Boehringer, D., Leibundgut, M., Heuvel, J. v. d. & Ban, N. Cryo-EM structure of Hepatitis C virus IRES bound to the human ribosome at 3.9-Å resolution. *Nature Communications* **6,** 7646. ISSN: 2041-1723 (2015).

57. Piñeiro, D. & Martinez-Salas, E. RNA Structural Elements of Hepatitis C Virus Controlling Viral RNA Translation and the Implications for Viral Pathogenesis. *Viruses* **4,** 2233–2250 (2012).

58. Friebe, P. & Bartenschlager, R. Genetic Analysis of Sequences in the 3 Nontranslated Region of Hepatitis C Virus That Are Important for RNA Replication. *Journal of Virology* **76,** 5326–5338. ISSN: 0022-538X, 1098-5514 (2002).

59. Shi, S. T. & Lai, M. M. C. *Hepatitis C Viruses: Genomes and Molecular Biology* (ed Tan, S.-L.) 49–87 (Horizon Bioscience, Norfolk (UK), 2006). ISBN: 978-1-904933-20-5. `http://www.ncbi.nlm.nih.gov/books/NBK1624/` (2017).

60. Bai, Y., Zhou, K. & Doudna, J. A. Hepatitis C virus 3UTR regulates viral translation through direct interactions with the host translation machinery. *Nucleic Acids Research* **41,** 7861–7874. ISSN: 0305-1048 (2013).

61. Quinkert, D., Bartenschlager, R. & Lohmann, V. Quantitative Analysis of the Hepatitis C Virus Replication Complex. *Journal of Virology* **79,** 13594–13605. ISSN: 0022-538X, 1098-5514 (2005).

62. Lohmann, V., Körner, F., Koch, J.-O., Herian, U., Theilmann, L. & Bartenschlager, R. Replication of Subgenomic Hepatitis C Virus RNAs in a Hepatoma Cell Line. *Science* **285,** 110–113. ISSN: 0036-8075, 1095-9203 (1999).

63. Blight, K. J., Kolykhalov, A. A. & Rice, C. M. Efficient Initiation of HCV RNA Replication in Cell Culture. *Science* **290,** 1972–1974. ISSN: 0036-8075, 1095-9203 (2000).

64. Fusco, D. N. & Chung, R. T. Novel Therapies for Hepatitis C: Insights from the Structure of the Virus. *Annual Review of Medicine* **63,** 373–387 (2012).

65. Moradpour, D., Penin, F. & Rice, C. M. Replication of hepatitis C virus. *Nature Reviews Microbiology* **5,** 453–463. ISSN: 1740-1526 (2007).

66. Gawlik, K. & Gallay, P. A. HCV core protein and virus assembly: what we know without structures. *Immunologic Research* **60,** 1–10. ISSN: 0257-277X, 1559-0755 (2014).

67. Masaki, T., Suzuki, R., Murakami, K., Aizaki, H., Ishii, K., Murayama, A., Date, T., Matsuura, Y., Miyamura, T., Wakita, T. & Suzuki, T. Interaction of Hepatitis C Virus Nonstructural Protein 5A with Core Protein Is Critical for the Production of Infectious Virus Particles. *Journal of Virology* **82,** 7964–7976. ISSN: 0022-538X (2008).

68. Xu, Z., Choi, J., Yen, T., Lu, W., Strohecker, A., Govindarajan, S., Chien, D., Selby, M. J. & Ou, J.-h. Synthesis of a novel hepatitis C virus protein by ribosomal frameshift. *The EMBO Journal* **20,** 3840–3848. ISSN: 0261-4189 (2001).

69. Park, S. B., Seronello, S., Mayer, W. & Ojcius, D. M. Hepatitis C Virus Frameshift/Alternate Reading Frame Protein Suppresses Interferon Responses Mediated by Pattern Recognition Receptor Retinoic-Acid-Inducible Gene-I. *PLOS ONE* **11,** e0158419. ISSN: 1932-6203 (2016).

70. Freedman, H., Logan, M. R., Law, J. L. M. & Houghton, M. Structure and Function of the Hepatitis C Virus Envelope Glycoproteins E1 and E2: Antiviral and Vaccine Targets. *ACS Infectious Diseases* **2,** 749–762 (2016).

71. Lindenbach, B. D. & Rice, C. M. The ins and outs of hepatitis C virus entry and assembly. *Nature Reviews Microbiology* **11,** 688–700. ISSN: 1740-1526 (2013).

72. Omari, K. E., Iourin, O., Kadlec, J., Sutton, G., Harlos, K., Grimes, J. M. & Stuart, D. I. Unexpected structure for the N-terminal domain of hepatitis C virus envelope glycoprotein E1. *Nature Communications* **5,** 4874. ISSN: 2041-1723 (2014).

73. Kong, L., Giang, E., Nieusma, T., Kadam, R. U., Cogburn, K. E., Hua, Y., Dai, X., Stanfield, R. L., Burton, D. R., Ward, A. B., Wilson, I. A. & Law, M. Hepatitis C Virus E2 Envelope Glycoprotein Core Structure. *Science* **342,** 1090–1094. ISSN: 0036-8075, 1095-9203 (2013).

74. Khan, A. G., Whidby, J., Miller, M. T., Scarborough, H., Zatorski, A. V., Cygan, A., Price, A. A., Yost, S. A., Bohannon, C. D., Jacob, J., Grakoui, A. & Marcotrigiano, J. Structure of the core ectodomain of the hepatitis C virus envelope glycoprotein 2. *Nature* **509,** 381–384. ISSN: 0028-0836 (2014).

75. Falson, P., Bartosch, B., Alsaleh, K., Tews, B. A., Loquet, A., Ciczora, Y., Riva, L., Montigny, C., Montpellier, C., Duverlie, G., Pécheur, E.-I., Maire, M. l., Cosset, F.-L., Dubuisson, J. & Penin, F. Hepatitis C Virus Envelope Glycoprotein E1 Forms Trimers at the Surface of the Virion. *Journal of Virology* **89,** 10333–10346. ISSN: 0022-538X, 1098-5514 (2015).

76. Steinmann, E. & Pietschmann, T. Hepatitis C Virus P7—A Viroporin Crucial for Virus Assembly and an Emerging Target for Antiviral Therapy. *Viruses* **2,** 2078–2095 (2010).

77. Atoom, A. M., Taylor, N. G. A. & Russell, R. S. The elusive function of the hepatitis C virus p7 protein. *Virology* **462–463,** 377–387. ISSN: 0042-6822 (2014).

78. Popescu, C.-I., Callens, N., Trinel, D., Roingeard, P., Moradpour, D., Descamps, V., Duverlie, G., Penin, F., Héliot, L., Rouillé, Y. & Dubuisson, J. NS2 Protein of Hepatitis C Virus Interacts with Structural and Non-Structural Proteins towards Virus Assembly. *PLoS Pathogens* **7,** e1001278. ISSN: 1553-7366 (2011).

79. Li, X.-D., Sun, L., Seth, R. B., Pineda, G. & Chen, Z. J. Hepatitis C virus protease NS3/4A cleaves mitochondrial antiviral signaling protein off the mitochondria to evade innate immunity. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 17717–17722. ISSN: 0027-8424, 1091-6490 (2005).

80. Li, K., Foy, E., Ferreon, J. C., Nakamura, M., Ferreon, A. C. M., Ikeda, M., Ray, S. C., Gale, M. & Lemon, S. M. Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 2992–2997. ISSN: 0027-8424, 1091-6490 (2005).

81. Frick, D. N. *Hepatitis C Viruses: Genomes and Molecular Biology* (ed Tan, S.-L.) (Horizon Bioscience, Norfolk (UK), 2006). ISBN: 978-1-904933-20-5. `http://www.ncbi.nlm.nih.gov/books/NBK1614/` (2017).

82. Appleby, T. C., Anderson, R., Fedorova, O., Pyle, A. M., Wang, R., Liu, X., Brendza, K. M. & Somoza, J. R. Visualizing ATP-Dependent RNA Translocation by the NS3 Helicase from HCV. *Journal of Molecular Biology* **405,** 1139–1153. ISSN: 0022-2836 (2011).

83. Belon, C. A. & Frick, D. N. Helicase inhibitors as specifically targeted antiviral therapy for hepatitis C. *Future virology* **4,** 277–293. ISSN: 1746-0794 (2009).

84. Gouttenoire, J., Montserret, R., Paul, D., Castillo, R., Meister, S., Bartenschlager, R., Penin, F. & Moradpour, D. Aminoterminal Amphipathic α-Helix AH1 of Hepatitis C Virus Nonstructural Protein 4B Possesses a Dual Role in RNA Replication and Virus Production. *PLoS Pathogens* **10,** e1004501. ISSN: 1553-7366 (2014).

85. Sklan, E. H. & Glenn, J. S. *Hepatitis C Viruses: Genomes and Molecular Biology* (ed Tan, S.-L.) 245–266 (Horizon Bioscience, Norfolk (UK), 2006). ISBN: 978-1-904933-20-5. `http://www.ncbi.nlm.nih.gov/books/NBK1633/` (2016).

86. Ross-Thriepland, D. & Harris, M. Hepatitis C virus NS5A: enigmatic but still promiscuous 10 years on! *Journal of General Virology* **96,** 727–738 (2015).

87. Tellinghuisen, T. L., Foss, K. L., Treadaway, J. C. & Rice, C. M. Identification of Residues Required for RNA Replication in Domains II and III of the Hepatitis C Virus NS5A Protein. *Journal of Virology* **82,** 1073–1083. ISSN: 0022-538X, 1098-5514 (2008).

88. Ross-Thriepland, D., Amako, Y. & Harris, M. The C terminus of NS5A domain II is a key determinant of hepatitis C virus genome replication, but is not required for virion assembly and release. *The Journal of General Virology* **94,** 1009–1018. ISSN: 0022-1317 (2013).

89. Shirota, Y., Luo, H., Qin, W., Kaneko, S., Yamashita, T., Kobayashi, K. & Murakami, S. Hepatitis C virus (HCV) NS5A binds RNA-dependent RNA polymerase (RdRP) NS5B and modulates RNA-dependent RNA polymerase activity. *The Journal of Biological Chemistry* **277,** 11149–11155. ISSN: 0021-9258 (2002).

90. Quezada, E. M. & Kane, C. M. The Hepatitis C Virus NS5A Stimulates NS5B During In Vitro RNA Synthesis in a Template Specific Manner. *The Open Biochemistry Journal* **3,** 39–48. ISSN: 1874-091X (2009).

91. Cho, N.-J., Pham, E. A., Hagey, R. J., Lévêque, V. J., Ma, H., Klumpp, K. & Glenn, J. S. Reconstitution and Functional Analysis of a Full-Length Hepatitis C Virus NS5B Polymerase on a Supported Lipid Bilayer. *ACS Central Science* **2,** 456–466. ISSN: 2374-7943 (2016).

92. Chinnaswamy, S., Murali, A., Li, P., Fujisaki, K. & Kao, C. C. Regulation of De Novo-Initiated RNA Synthesis in Hepatitis C Virus RNA-Dependent RNA Polymerase by Intermolecular Interactions. *Journal of Virology* **84,** 5923–5935. ISSN: 0022-538X, 1098-5514 (2010).

93. Ranjith-Kumar, C. T., Wen, Y., Baxter, N., Bhardwaj, K. & Cheng Kao, C. A Cell-Based Assay for RNA Synthesis by the HCV Polymerase Reveals New Insights on Mechanism of Polymerase Inhibitors and Modulation by NS5A. *PLoS ONE* **6,** e22575. ISSN: 1932-6203 (2011).

94. Ngure, M., Issur, M., Shkriabai, N., Liu, H.-W., Cosa, G., Kvaratskhelia, M. & Götte, M. Interactions of the Disordered Domain II of Hepatitis C Virus NS5A with Cyclophilin A, NS5B, and viral RNA Show Extensive Overlap. *ACS Infectious Diseases* **2,** 839–851 (2016).

95. Brass, V., Bieck, E., Montserret, R., Wölk, B., Hellings, J. A., Blum, H. E., Penin, F. & Moradpour, D. An Amino-terminal Amphipathic α-Helix Mediates Membrane Association of the Hepatitis C Virus Nonstructural Protein 5A. *Journal of Biological Chemistry* **277,** 8130–8139. ISSN: 0021-9258, 1083-351X (2002).

96. Penin, F., Brass, V., Appel, N., Ramboarina, S., Montserret, R., Ficheux, D., Blum, H. E., Bartenschlager, R. & Moradpour, D. Structure and Function of the Membrane Anchor Domain of Hepatitis C Virus Nonstructural Protein 5A. *Journal of Biological Chemistry* **279,** 40835–40843. ISSN: 0021-9258, 1083-351X (2004).

97. Tellinghuisen, T. L., Marcotrigiano, J. & Rice, C. M. Structure of the Zinc-Binding Domain of an Essential Replicase Component of Hepatitis C Virus Reveals a Novel Fold. *Nature* **435,** 374–379. ISSN: 0028-0836 (2005).

98. Tellinghuisen, T. L., Marcotrigiano, J., Gorbalenya, A. E. & Rice, C. M. The NS5A Protein of Hepatitis C Virus Is a Zinc Metalloprotein. *Journal of Biological Chemistry* **279,** 48576–48587. ISSN: 0021-9258, 1083-351X (2004).

99. Love, R. A., Brodsky, O., Hickey, M. J., Wells, P. A. & Cronin, C. N. Crystal Structure of a Novel Dimeric Form of NS5A Domain I Protein from Hepatitis C Virus. *Journal of Virology* **83,** 4395–4403. ISSN: 0022-538X, 1098-5514 (2009).

100. Lambert, S. M., Langley, D. R., Garnett, J. A., Angell, R., Hedgethorne, K., Meanwell, N. A. & Matthews, S. J. The crystal structure of NS5A domain 1 from genotype 1a reveals new clues to the mechanism of action for dimeric HCV inhibitors. *Protein Science* **23,** 723–734. ISSN: 1469-896X (2014).

101. Huang, L., Hwang, J., Sharma, S. D., Hargittai, M. R. S., Chen, Y., Arnold, J. J., Raney, K. D. & Cameron, C. E. Hepatitis C Virus Nonstructural Protein 5A (NS5A) Is an RNA-binding Protein. *Journal of Biological Chemistry* **280,** 36417–36428. ISSN: 0021-9258, 1083-351X (2005).

102. Brass, V., Gosert, R. & Moradpour, D. *Hepatitis C* (ed Tang, H.) *Methods in Molecular Biology™* 510. DOI: 10.1007/978-1-59745-394-3_15, 195–209 (Humana Press, 2009). ISBN: 978-1-58829-970-3. http://dx.doi.org/10.1007/978-1-59745-394-3_15 (2017).

103. Romero-Brey, I., Berger, C., Kallis, S., Kolovou, A., Paul, D., Lohmann, V. & Bartenschlager, R. NS5A Domain 1 and Polyprotein Cleavage Kinetics Are Critical for Induction of Double-Membrane Vesicles Associated with Hepatitis C Virus Replication. *mBio* **6,** e00759–15. ISSN: , 2150-7511 (2015).

104. Liang, Y., Ye, H., Kang, C. B. & Yoon, H. S. Domain 2 of Nonstructural Protein 5A (NS5A) of Hepatitis C Virus Is Natively Unfolded. *Biochemistry* **46,** 11550–11558. ISSN: 0006-2960 (2007).

105. Hanoulle, X., Badillo, A., Wieruszeski, J.-M., Verdegem, D., Landrieu, I., Bartenschlager, R., Penin, F. & Lippens, G. Hepatitis C Virus NS5A Protein Is a Substrate for the Peptidyl-prolyl cis/trans Isomerase Activity of Cyclophilins A and B. *Journal of Biological Chemistry* **284,** 13589–13601. ISSN: 0021-9258, 1083-351X (2009).

106. Hanoulle, X., Badillo, A., Verdegem, D., Penin, F. & Lippens, G. The Domain 2 of the HCV NS5A Protein Is Intrinsically Unstructured. *Protein & Peptide Letters* **17,** 1012–1018. ISSN: 09298665 (2010).

107. Feuerstein, S., Solyom, Z., Aladağ, A., Hoffmann, S., Willbold, D. & Brutscher, B. 1H, 13C, and 15N resonance assignment of a 179 residue fragment of hepatitis C virus non-structural protein 5A. *Biomolecular NMR Assignments* **5,** 241. ISSN: 1874-2718, 1874-270X (2011).

108. Rosnoblet, C., Fritzinger, B., Legrand, D., Launay, H., Wieruszeski, J.-M., Lippens, G. & Hanoulle, X. Hepatitis C Virus NS5B and Host Cyclophilin A Share a Common Binding Site on NS5A. *Journal of Biological Chemistry* **287,** 44249–44260. ISSN: 0021-9258, 1083-351X (2012).

109. Badillo, A., Receveur-Brechot, V., Sarrazin, S., Cantrelle, F.-X., Delolme, F., Fogeron, M.-L., Molle, J., Montserret, R., Bockmann, A., Bartenschlager, R., Lohmann, V., Lippens, G., Ricard-Blum, S., Hanoulle, X. & Penin, F. Overall Structural Model of NS5A Protein from Hepatitis C Virus and Modulation by Mutations Conferring Resistance of Virus Replication to Cyclosporin A. *Biochemistry* **56,** 3029–3048. ISSN: 0006-2960 (2017).

110. Dujardin, M., Madan, V., Montserret, R., Ahuja, P., Huvent, I., Launay, H., Leroy, A., Bartenschlager, R., Penin, F., Lippens, G. & Hanoulle, X. A Proline-Tryptophan Turn in the Intrinsically Disordered Domain 2 of NS5A Protein is Essential for Hepatitis C Virus RNA Replication. *Journal of Biological Chemistry* **290,** 19104–19120. ISSN: 0021-9258, 1083-351X (2015).

111. Yang, F., Robotham, J. M., Nelson, H. B., Irsigler, A., Kenworthy, R. & Tang, H. Cyclophilin A Is an Essential Cofactor for Hepatitis C Virus Infection and the Principal Mediator of Cyclosporine Resistance In Vitro. *Journal of Virology* **82,** 5269–5278. ISSN: 0022-538X (2008).

112. Chatterji, U., Bobardt, M., Selvarajah, S., Yang, F., Tang, H., Sakamoto, N., Vuagniaux, G., Parkinson, T. & Gallay, P. The Isomerase Active Site of Cyclophilin A Is Critical for Hepatitis C Virus Replication. *Journal of Biological Chemistry* **284,** 16998–17005. ISSN: 0021-9258, 1083-351X (2009).

113. Verdegem, D., Badillo, A., Wieruszeski, J.-M., Landrieu, I., Leroy, A., Bartenschlager, R., Penin, F., Lippens, G. & Hanoulle, X. Domain 3 of NS5A Protein from the Hepatitis C Virus Has Intrinsic α-Helical Propensity and Is a Substrate of Cyclophilin A. *The Journal of Biological Chemistry* **286,** 20441–20454. ISSN: 0021-9258 (2011).

114. Moradpour, D., Evans, M. J., Gosert, R., Yuan, Z., Blum, H. E., Goff, S. P., Lindenbach, B. D. & Rice, C. M. Insertion of green fluorescent protein into nonstructural protein 5A allows direct visualization of functional hepatitis C virus replication complexes. *Journal of Virology* **78,** 7400–7409. ISSN: 0022-538X (2004).

115. Appel, N., Pietschmann, T. & Bartenschlager, R. Mutational Analysis of Hepatitis C Virus Nonstructural Protein 5A: Potential Role of Differential Phosphorylation in RNA Replication and Identification of a Genetically Flexible Domain. *Journal of Virology* **79,** 3187–3194. ISSN: 0022-538X (2005).

116. Appel, N., Zayas, M., Miller, S., Krijnse-Locker, J., Schaller, T., Friebe, P., Kallis, S., Engel, U. & Bartenschlager, R. Essential Role of Domain III of Nonstructural Protein 5A for Hepatitis C Virus Infectious Particle Assembly. *PLOS Pathogens* **4,** e1000035. ISSN: 1553-7374 (2008).

117. Cun, W., Jiang, J. & Luo, G. The C-Terminal α-Helix Domain of Apolipoprotein E Is Required for Interaction with Nonstructural Protein 5A and Assembly of Hepatitis C Virus. *Journal of Virology* **84,** 11532–11541. ISSN: 0022-538X, 1098-5514 (2010).

118. Benga, W. J. A., Krieger, S. E., Dimitrova, M., Zeisel, M. B., Parnot, M., Lupberger, J., Hildt, E., Luo, G., McLauchlan, J., Baumert, T. F. & Schuster, C. Apolipoprotein E interacts with hepatitis C virus nonstructural protein 5A and determines assembly of infectious particles. *Hepatology* **51,** 43–53. ISSN: 1527-3350 (2010).

119. Shi, S. T., Polyak, S. J., Tu, H., Taylor, D. R., Gretch, D. R. & Lai, M. M. C. Hepatitis C Virus NS5A Colocalizes with the Core Protein on Lipid Droplets and Interacts with Apolipoproteins. *Virology* **292,** 198–210. ISSN: 0042-6822 (2002).

120. Miyanari, Y., Atsuzawa, K., Usuda, N., Watashi, K., Hishiki, T., Zayas, M., Bartenschlager, R., Wakita, T., Hijikata, M. & Shimotohno, K. The lipid droplet is an important organelle for hepatitis C virus production. *Nature Cell Biology* **9,** 1089–1097. ISSN: 1465-7392 (2007).

121. Zayas, M., Long, G., Madan, V. & Bartenschlager, R. Coordination of Hepatitis C Virus Assembly by Distinct Regulatory Regions in Nonstructural Protein 5A. *PLoS Pathog* **12,** e1005376 (2016).

122. Ross-Thriepland, D. & Harris, M. Insights into the Complexity and Functionality of Hepatitis C Virus NS5A Phosphorylation. *Journal of Virology* **88,** 1421–1432. ISSN: 0022-538X (2014).

123. Kim, J., Lee, D. & Choe, J. Hepatitis C Virus NS5A Protein Is Phosphorylated by Casein Kinase II. *Biochemical and Biophysical Research Communications* **257,** 777–781. ISSN: 0006-291X (1999).

124. Chong, W. M., Hsu, S.-C., Kao, W.-T., Lo, C.-W., Lee, K.-Y., Shao, J.-S., Chen, Y.-H., Chang, J., Chen, S. S.-L. & Yu, M.-J. Phosphoproteomics Identified an NS5A Phosphorylation Site Involved in Hepatitis C Virus Replication. *Journal of Biological Chemistry* **291,** 3918–3931. ISSN: 0021-9258, 1083-351X (2016).

125. Reed, K. E., Xu, J. & Rice, C. M. Phosphorylation of the hepatitis C virus NS5A protein in vitro and in vivo: properties of the NS5A-associated kinase. *Journal of Virology* **71,** 7187–7197. ISSN: 0022-538X (1997).

126. Chen, Y.-C., Su, W.-C., Huang, J.-Y., Chao, T.-C., Jeng, K.-S., Machida, K. & Lai, M. M. C. Polo-Like Kinase 1 Is Involved in Hepatitis C Virus Replication by Hyperphosphorylating NS5A. *Journal of Virology* **84,** 7983–7993. ISSN: 0022-538X, 1098-5514 (2010).

127. Yamauchi, S., Takeuchi, K., Chihara, K., Sun, X., Honjoh, C., Yoshiki, H., Hotta, H. & Sada, K. Hepatitis C Virus Particle Assembly Involves Phosphorylation of NS5A by the c-Abl Tyrosine Kinase. *Journal of Biological Chemistry,* jbc.M115.666859. ISSN: 0021-9258, 1083-351X (2015).

128. Neddermann, P., Quintavalle, M., Di Pietro, C., Clementi, A., Cerretani, M., Altamura, S., Bartholomew, L. & De Francesco, R. Reduction of Hepatitis C Virus NS5A Hyperphosphorylation by Selective Inhibition of Cellular Kinases Activates Viral RNA Replication in Cell Culture. *Journal of Virology* **78,** 13306–13314. ISSN: 0022-538X (2004).

129. LeMay, K. L., Treadaway, J., Angulo, I. & Tellinghuisen, T. L. A Hepatitis C Virus NS5A Phosphorylation Site That Regulates RNA Replication. *Journal of Virology* **87,** 1255–1260. ISSN: 0022-538X, 1098-5514 (2013).

130. Lee, K.-Y., Chen, Y.-H., Hsu, S.-C. & Yu, M.-J. Phosphorylation of Serine 235 of the Hepatitis C Virus Non-Structural Protein NS5A by Multiple Kinases. *PLOS ONE* **11,** e0166763. ISSN: 1932-6203 (2016).

131. Hsu, S.-C., Lo, C.-W., Pan, T.-C., Lee, K.-Y. & Yu, M.-J. Serine 235 Is the Primary NS5A Hyper-Phosphorylation Site Responsible for HCV Replication. *Journal of Virology,* JVI.00194–17. ISSN: 0022-538X, 1098-5514 (2017).

132. Cordek, D. G., Croom-Perez, T. J., Hwang, J., Hargittai, M. R. S., Subba-Reddy, C. V., Han, Q., Lodeiro, M. F., Ning, G., McCrory, T. S., Arnold, J. J., Koc, H., Lindenbach, B. D., Showalter, S. A. & Cameron, C. E. Expanding the proteome of an RNA virus by phosphorylation of an intrinsically disordered viral protein. *Journal of Biological Chemistry,* jbc.M114.589911. ISSN: 0021-9258, 1083-351X (2014).

133. Secci, E., Luchinat, E. & Banci, L. The Casein Kinase 2-Dependent Phosphorylation of NS5A Domain3 from HepatitisC Virus Followed by Time-Resolved NMR Spectroscopy. *ChemBioChem* **17,** 328–333. ISSN: 1439-7633 (2016).

134. Simister, P., Schmitt, M., Geitmann, M., Wicht, O., Danielson, U. H., Klein, R., Bressanelli, S. & Lohmann, V. Structural and Functional Analysis of Hepatitis C Virus Strain JFH1 Polymerase. *Journal of Virology* **83,** 11926–11939. ISSN: 0022-538X, 1098-5514 (2009).

135. Luo, G., Hamatake, R. K., Mathis, D. M., Racela, J., Rigat, K. L., Lemm, J. & Colonno, R. J. De Novo Initiation of RNA Synthesis by the RNA-Dependent RNA Polymerase (NS5B) of Hepatitis C Virus. *Journal of Virology* **74,** 851–863. ISSN: 0022-538X, 1098-5514 (2000).

136. Zhong, W., Ferrari, E., Lesburg, C. A., Maag, D., Ghosh, S. K., Cameron, C. E., Lau, J. Y. & Hong, Z. Template/primer requirements and single nucleotide incorporation by hepatitis C virus nonstructural protein 5B polymerase. *Journal of Virology* **74,** 9134–9143. ISSN: 0022-538X (2000).

137. Bressanelli, S., Tomei, L., Roussel, A., Incitti, I., Vitale, R. L., Mathieu, M., De Francesco, R. & Rey, F. A. Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Proceedings of the National Academy of Sciences of the United States of America* **96,** 13034–13039. ISSN: 0027-8424 (1999).

138. Lesburg, C. A., Cable, M. B., Ferrari, E., Hong, Z., Mannarino, A. F. & Weber, P. C. Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. *Nature Structural & Molecular Biology* **6,** 937–943. ISSN: 1072-8368 (1999).

139. Ago, H., Adachi, T., Yoshida, A., Yamamoto, M., Habuka, N., Yatsunami, K. & Miyano, M. Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Structure (London, England: 1993)* **7,** 1417–1426. ISSN: 0969-2126 (1999).

140. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. The Protein Data Bank. *Nucleic Acids Research* **28,** 235–242. ISSN: 0305-1048 (2000).

141. Love, R. A., Parge, H. E., Yu, X., Hickey, M. J., Diehl, W., Gao, J., Wriggers, H., Ekker, A., Wang, L., Thomson, J. A., Dragovich, P. S. & Fuhrman, S. A. Crystallographic Identification of a Noncompetitive Inhibitor Binding Site on the Hepatitis C Virus NS5B RNA Polymerase Enzyme. *Journal of Virology* **77,** 7575–7581. ISSN: 0022-538X, 1098-5514 (2003).

142. Hang, J. Q., Yang, Y., Harris, S. F., Leveque, V., Whittington, H. J., Rajyaguru, S., Ao-Ieong, G., McCown, M. F., Wong, A., Giannetti, A. M., Pogam, S. L., Talamás, F., Cammack, N., Nájera, I. & Klumpp, K. Slow Binding Inhibition and Mechanism of Resistance of Non-nucleoside Polymerase Inhibitors of Hepatitis C Virus. *Journal of Biological Chemistry* **284,** 15517–15529. ISSN: 0021-9258, 1083-351X (2009).

143. Narjes, F., Crescenzi, B., Ferrara, M., Habermann, J., Colarusso, S., del Rosario Rico Ferreira, M., Stansfield, I., Mackay, A. C., Conte, I., Ercolani, C., Zaramella, S., Palumbi, M.-C., Meuleman, P., Leroux-Roels, G., Giuliano, C., Fiore, F., Di Marco, S., Baiocco, P., Koch, U., Migliaccio, G., Altamura, S., Laufer, R., De Francesco, R. & Rowley, M. Discovery of (7R)-14-Cyclohexyl-7-{[2-(dimethylamino)ethyl](methyl) amino}-7,8-dihydro-6H-indolo[1,2-e][1,5]benzoxazocine-11-carboxylic Acid (MK-3281), a Potent and Orally Bioavailable Finger-Loop Inhibitor of the Hepatitis C Virus NS5B Polymerase. *Journal of Medicinal Chemistry* **54,** 289–301. ISSN: 0022-2623 (2011).

144. O'Farrell, D., Trowbridge, R., Rowlands, D. & Jäger, J. Substrate Complexes of Hepatitis C Virus RNA Polymerase (HC-J4): Structural Evidence for Nucleotide Import and De-novo Initiation. *Journal of Molecular Biology* **326,** 1025–1035. ISSN: 0022-2836 (2003).

145. Mosley, R. T., Edwards, T. E., Murakami, E., Lam, A. M., Grice, R. L., Du, J., Sofia, M. J., Furman, P. A. & Otto, M. J. Structure of Hepatitis C Virus Polymerase in Complex with Primer-Template RNA. *Journal of Virology* **86,** 6503–6511. ISSN: 0022-538X, 1098-5514 (2012).

146. Schmidt-Mende, J., Bieck, E., Hügle, T., Penin, F., Rice, C. M., Blum, H. E. & Moradpour, D. Determinants for Membrane Association of the Hepatitis C Virus RNA-dependent RNA Polymerase. *Journal of Biological Chemistry* **276,** 44052–44063. ISSN: 0021-9258, 1083-351X (2001).

147. Miller, R. H. & Purcell, R. H. Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. *Proceedings of the National Academy of Sciences* **87,** 2057–2061. ISSN: 0027-8424, 1091-6490 (1990).

148. Appleby, T. C., Perry, J. K., Murakami, E., Barauskas, O., Feng, J., Cho, A., Fox, D., Wetmore, D. R., McGrath, M. E., Ray, A. S., Sofia, M. J., Swaminathan, S. & Edwards, T. E. Structural basis for RNA replication by the hepatitis C virus polymerase. *Science* **347,** 771–775. ISSN: 0036-8075, 1095-9203 (2015).

149. Ranjith-Kumar, C. T., Kim, Y.-C., Gutshall, L., Silverman, C., Khandekar, S., Sarisky, R. T. & Kao, C. C. Mechanism of De Novo Initiation by the Hepatitis C Virus RNA-Dependent RNA Polymerase: Role of Divalent Metals. *Journal of Virology* **76,** 12513–12525. ISSN: 0022-538X (2002).

150. Beaulieu, P. L. Recent advances in the development of NS5B polymerase inhibitors for the treatment of hepatitis C virus infection. *Expert Opinion on Therapeutic Patents* **19,** 145–164. ISSN: 1744-7674 (2009).

151. Di Maio, V. C., Cento, V., Mirabelli, C., Artese, A., Costa, G., Alcaro, S., Perno, C. F. & Ceccherini-Silberstein, F. Hepatitis C Virus Genetic Variability and the Presence of NS5B Resistance-Associated Mutations as Natural Polymorphisms in Selected Genotypes Could Affect the Response to NS5B Inhibitors. *Antimicrobial Agents and Chemotherapy* **58,** 2781–2797. ISSN: 0066-4804 (2014).

152. Ranjith-Kumar, C. T. & Kao, C. C. Recombinant viral RdRps can initiate RNA synthesis from circular templates. *RNA* **12,** 303–312. ISSN: 1355-8382 (2006).

153. Rigat, K., Wang, Y., Hudyma, T. W., Ding, M., Zheng, X., Gentles, R. G., Beno, B. R., Gao, M. & Roberts, S. B. Ligand-induced changes in hepatitis C virus NS5B polymerase structure. *Antiviral Research* **88,** 197–206. ISSN: 0166-3542 (2010).

154. Cheney, I. W., Naim, S., Lai, V. C. H., Dempsey, S., Bellows, D., Walker, M. P., Shim, J. H., Horscroft, N., Hong, Z. & Zhong, W. Mutations in NS5B Polymerase of Hepatitis C Virus: Impacts on in Vitro Enzymatic Activity and Viral RNA Replication in the Subgenomic Replicon Cell Culture. *Virology* **297,** 298–306. ISSN: 0042-6822 (2002).

155. Tomei, L., Vitale, R. L., Incitti, I., Serafini, S., Altamura, S., Vitelli, A. & De Francesco, R. Biochemical characterization of a hepatitis C virus RNA-dependent RNA polymerase mutant lacking the C-terminal hydrophobic sequence. *Journal of General Virology* **81,** 759–767 (2000).

156. Moradpour, D., Gosert, R., Egger, D., Penin, F., Blum, H. E. & Bienz, K. Membrane association of hepatitis C virus nonstructural proteins and identification of the membrane alteration that harbors the viral replication complex. *Antiviral Research. Proceedings of the Tenth International Symposium on Viral Hepetitis* **60,** 103–109. ISSN: 0166-3542 (2003).

157. Adachi, T., Ago, H., Habuka, N., Okuda, K., Komatsu, M., Ikeda, S. & Yatsunami, K. The essential role of C-terminal residues in regulating the activity of hepatitis C virus RNA-dependent RNA polymerase. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1601,** 38–48. ISSN: 1570-9639 (2002).

158. Caillet-Saguy, C., Simister, P. C. & Bressanelli, S. An Objective Assessment of Conformational Variability in Complexes of Hepatitis C Virus Polymerase with Non-Nucleoside Inhibitors. *Journal of Molecular Biology* **414,** 370–384. ISSN: 0022-2836 (2011).

159. Davis, B. C. & Thorpe, I. F. Molecular Simulations Illuminate the Role of Regulatory Components of the RNA Polymerase from the Hepatitis C Virus in Influencing Protein Structure and Dynamics. *Biochemistry* **52,** 4541–4552. ISSN: 0006-2960 (2013).

160. Harrus, D., Ahmed-El-Sayed, N., Simister, P. C., Miller, S., Triconnet, M., Hagedorn, C. H., Mahias, K., Rey, F. A., Astier-Gin, T. & Bressanelli, S. Further Insights into the Roles of GTP and the C Terminus of the Hepatitis C Virus Polymerase in the Initiation of RNA Synthesis. *Journal of Biological Chemistry* **285,** 32906–32918. ISSN: 0021-9258, 1083-351X (2010).

161. Scrima, N., Caillet-Saguy, C., Ventura, M., Harrus, D., Astier-Gin, T. & Bressanelli, S. Two Crucial Early Steps in RNA Synthesis by the Hepatitis C Virus Polymerase Involve a Dual Role of Residue 405. *Journal of Virology* **86,** 7107–7117. ISSN: 0022-538X, 1098-5514 (2012).

162. Powdrill, M. H., Tchesnokov, E. P., Kozak, R. A., Russell, R. S., Martin, R., Svarovskaia, E. S., Mo, H., Kouyos, R. D. & Götte, M. Contribution of a mutational bias in hepatitis C virus replication to the genetic barrier in the development of drug resistance. *Proceedings of the National Academy of Sciences* **108,** 20509–20513. ISSN: 0027-8424, 1091-6490 (2011).

163. Pawlotsky, J.-M. Hepatitis C Virus Population Dynamics During Infection, 261–284 (2006).

164. Jin, Z., Leveque, V., Ma, H., Johnson, K. A. & Klumpp, K. NTP-mediated nucleotide excision activity of hepatitis C virus RNA-dependent RNA polymerase. *Proceedings of the National Academy of Sciences* **110,** 1575–1576. ISSN: 0027-8424, 1091-6490 (2013).

165. Manns, M. P., Buti, M., Gane, E., Pawlotsky, J.-M., Razavi, H., Terrault, N. & Younossi, Z. Hepatitis C virus infection. *Nature Reviews Disease Primers* **3,** 17006. ISSN: 2056-676X (2017).

166. Otto, G. A. & Puglisi, J. D. The Pathway of HCV IRES-Mediated Translation Initiation. *Cell* **119,** 369–380. ISSN: 0092-8674, 1097-4172 (2004).

167. Paul, D., Hoppe, S., Saher, G., Krijnse-Locker, J. & Bartenschlager, R. Morphological and Biochemical Characterization of the Membranous Hepatitis C Virus Replication Compartment. *Journal of Virology* **87,** 10612–10627. ISSN: 0022-538X, 1098-5514 (2013).

168. Herker, E. & Ott, M. Unique ties between hepatitis C virus replication and intracellular lipids. *Trends in Endocrinology & Metabolism* **22,** 241–248. ISSN: 1043-2760 (2011).

169. Chevaliez, S. & Pawlotsky, J.-M. *Hepatitis C Viruses: Genomes and Molecular Biology* (ed Tan, S.-L.) (Horizon Bioscience, Norfolk (UK), 2006). ISBN: 1-904933-20-3 978-1-904933-20-5. `http://www.ncbi.nlm.nih.gov/books/NBK1630/` (2015).

170. Blight, K. J., McKeating, J. A., Marcotrigiano, J. & Rice, C. M. Efficient Replication of Hepatitis C Virus Genotype 1a RNAs in Cell Culture. *Journal of Virology* **77,** 3181–3190. ISSN: 0022-538X, 1098-5514 (2003).

171. Kato, T., Date, T., Miyamoto, M., Furusaka, A., Tokushige, K., Mizokami, M. & Wakita, T. Efficient replication of the genotype 2a hepatitis C virus subgenomic replicon. *Gastroenterology* **125,** 1808–1817. ISSN: 0016-5085, 1528-0012 (2003).

172. Saeed, M., Scheel, T. K. H., Gottwein, J. M., Marukian, S., Dustin, L. B., Bukh, J. & Rice, C. M. Efficient Replication of Genotype 3a and 4a Hepatitis C Virus Replicons in Human Hepatoma Cells. *Antimicrobial Agents and Chemotherapy* **56,** 5365–5373. ISSN: 0066-4804, 1098-6596 (2012).

173. Woerz, I., Lohmann, V. & Bartenschlager, R. Hepatitis C virus replicons: dinosaurs still in business? *Journal of Viral Hepatitis* **16,** 1–9. ISSN: 1352-0504 (2009).

174. Bartosch, B., Dubuisson, J. & Cosset, F.-L. Infectious Hepatitis C Virus Pseudoparticles Containing Functional E1–E2 Envelope Protein Complexes. *The Journal of Experimental Medicine* **197,** 633–642. ISSN: 0022-1007 (2003).

175. Steinmann, E. & Pietschmann, T. *Hepatitis C Virus: From Molecular Virology to Antiviral Therapy* (ed Bartenschlager, R.) *Current Topics in Microbiology and Immunology* 369. DOI: 10.1007/978-3-642-27340-7_2, 17–48 (Springer Berlin Heidelberg, 2013). ISBN: 978-3-642-27339-1 978-3-642-27340-7.

176. Kato, T., Furusaka, A., Miyamoto, M., Date, T., Yasui, K., Hiramoto, J., Nagayama, K., Tanaka, T. & Wakita, T. Sequence analysis of hepatitis C virus isolated from a fulminant hepatitis patient*. *Journal of Medical Virology* **64,** 334–339. ISSN: 1096-9071 (2001).

177. Lindenbach, B. D., Evans, M. J., Syder, A. J., Wölk, B., Tellinghuisen, T. L., Liu, C. C., Maruyama, T., Hynes, R. O., Burton, D. R., McKeating, J. A. & Rice, C. M. Complete Replication of Hepatitis C Virus in Cell Culture. *Science* **309,** 623–626. ISSN: 0036-8075, 1095-9203 (2005).

178. Wakita, T., Pietschmann, T., Kato, T., Date, T., Miyamoto, M., Zhao, Z., Murthy, K., Habermann, A., Kräusslich, H.-G., Mizokami, M., Bartenschlager, R. & Liang, T. J. Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. *Nature Medicine* **11,** 791–796. ISSN: 1078-8956 (2005).

179. Zhong, J., Gastaminza, P., Cheng, G., Kapadia, S., Kato, T., Burton, D. R., Wieland, S. F., Uprichard, S. L., Wakita, T. & Chisari, F. V. Robust hepatitis C virus infection in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **102,** 9294–9299. ISSN: 0027-8424, 1091-6490 (2005).

180. Banaudha, K., Orenstein, J. M., Korolnek, T., St. Laurent, G. C., Wakita, T. & Kumar, A. Primary hepatocyte culture supports hepatitis C virus replication: A model for infection-associated hepatocarcinogenesis. *Hepatology* **51,** 1922–1932. ISSN: 1527-3350 (2010).

181. Harak, C., Meyrath, M., Romero-Brey, I., Schenk, C., Gondeau, C., Schult, P., Esser-Nobis, K., Saeed, M., Neddermann, P., Schnitzler, P., Gotthardt, D., Perez-del-Pulgar, S., Neumann-Haefelin, C., Thimme, R., Meuleman, P., Vondran, F. W. R., Francesco, R. D., Rice, C. M., Bartenschlager, R. & Lohmann, V. Tuning a cellular lipid kinase activity adapts hepatitis C virus to replication in cell culture. *Nature Microbiology* **2,** 16247. ISSN: 2058-5276 (2016).

182. Couto, L. B. & Kolykhalov, A. A. *Hepatitis C Viruses: Genomes and Molecular Biology* (ed Tan, S.-L.) 353–374 (Horizon Bioscience, Norfolk (UK), 2006). ISBN: 978-1-904933-20-5. `http://www.ncbi.nlm.nih.gov/books/NBK1627/` (2017).

183. Wadman, M. Animal rights: Chimpanzee research on trial. *Nature News* **474,** 268–271. ISSN: 0028-0836 (2011).

184. Kaiser, J. NIH to end all support for chimpanzee research. *Science | AAAS*. `http://www.sciencemag.org/news/2015/11/nih-end-all-support-chimpanzee-research` (2017) (2015).

185. Zhao, X., Tang, Z.-Y., Klumpp, B., Wolff-Vorbeck, G., Barth, H., Levy, S., Weizsäcker, F. v., Blum, H. E. & Baumert, T. F. Primary hepatocytes of Tupaia belangeri as a potential model for hepatitis C virus infection. *The Journal of Clinical Investigation* **109,** 221–232. ISSN: 0021-9738 (2002).

186. Amako, Y., Tsukiyama-Kohara, K., Katsume, A., Hirata, Y., Sekiguchi, S., Tobita, Y., Hayashi, Y., Hishima, T., Funata, N., Yonekawa, H. & Kohara, M. Pathogenesis of Hepatitis C Virus Infection in Tupaia belangeri. *Journal of Virology* **84,** 303–311. ISSN: 0022-538X, 1098-5514 (2010).

187. Dorner, M., Horwitz, J. A., Robbins, J. B., Barry, W. T., Feng, Q., Mu, K., Jones, C. T., Schoggins, J. W., Catanese, M. T., Burton, D. R., Law, M., Rice, C. M. & Ploss, A. A genetically humanized mouse model for hepatitis C virus infection. *Nature* **474,** 208–211. ISSN: 0028-0836 (2011).

188. Dorner, M., Horwitz, J. A., Donovan, B. M., Labitt, R. N., Budell, W. C., Friling, T., Vogt, A., Catanese, M. T., Satoh, T., Kawai, T., Akira, S., Law, M., Rice, C. M. & Ploss, A. Completion of the entire hepatitis C virus life cycle in genetically humanized mice. *Nature* **501,** 237–241. ISSN: 0028-0836 (2013).

189. Mercer, D. F., Schiller, D. E., Elliott, J. F., Douglas, D. N., Hao, C., Rinfret, A., Addison, W. R., Fischer, K. P., Churchill, T. A., Lakey, J. R. T., Tyrrell, D. L. J. & Kneteman, N. M. Hepatitis C virus replication in mice with chimeric human livers. *Nature Medicine* **7,** 927–933. ISSN: 1078-8956 (2001).

190. Bissig, K.-D., Wieland, S. F., Tran, P., Isogawa, M., Le, T. T., Chisari, F. V. & Verma, I. M. Human liver chimeric mice provide a model for hepatitis B and C virus infection and treatment. *The Journal of Clinical Investigation* **120,** 924–930. ISSN: 0021-9738 (2010).

191. Bukh, J., Meuleman, P., Tellier, R., Engle, R., Feinstone, S., Eder, G., Satterfield, W., Govindarajan, S., Krawczynski, K., Miller, R., Leroux-Roels, G. & Purcell, R. Challenge pools of hepatitis C virus genotypes 1-6 prototype strains: Replication fitness and pathogenicity in chimpanzees and human liver-chimeric mouse models. *Journal of Infectious Diseases* **201,** 1381–1389 (2010).

192. Washburn, M. L., Bility, M. T., Zhang, L., Kovalev, G. I., Buntzman, A., Frelinger, J. A., Barry, W., Ploss, A., Rice, C. M. & Su, L. A Humanized Mouse Model to Study Hepatitis C Virus Infection, Immune Response, and Liver Disease. *Gastroenterology* **140,** 1334–1344. ISSN: 0016-5085, 1528-0012 (2011).

193. Keng, C. T., Sze, C. W., Zheng, D., Zheng, Z., Yong, K. S. M., Tan, S. Q., Ong, J. J. Y., Tan, S. Y., Loh, E., Upadya, M. H., Kuick, C. H., Hotta, H., Lim, S. G., Tan, T. C., Chang, K. T. E., Hong, W., Chen, J., Tan, Y.-J. & Chen, Q. Characterisation of liver pathogenesis, human immune responses and drug testing in a humanised mouse model of HCV infection. *Gut* **65,** 1744–1753. ISSN: 0017-5749, 1468-3288 (2016).

194. Romero-Brey, I. & Bartenschlager, R. Membranous Replication Factories Induced by Plus-Strand RNA Viruses. *Viruses* **6,** 2826–2857 (2014).

195. Ferraris, P., Blanchard, E. & Roingeard, P. Ultrastructural and biochemical analyses of hepatitis C virus-associated host cell membranes. *Journal of General Virology* **91,** 2230–2237 (2010).

196. Egger, D., Wölk, B., Gosert, R., Bianchi, L., Blum, H. E., Moradpour, D. & Bienz, K. Expression of Hepatitis C Virus Proteins Induces Distinct Membrane Alterations Including a Candidate Viral Replication Complex. *Journal of Virology* **76,** 5974–5984. ISSN: 0022-538X, 1098-5514 (2002).

197. Romero-Brey, I., Merz, A., Chiramel, A., Lee, J.-Y., Chlanda, P., Haselman, U., Santarella-Mellwig, R., Habermann, A., Hoppe, S., Kallis, S., Walther, P., Antony, C., Krijnse-Locker, J. & Bartenschlager, R. Three-Dimensional Architecture and Biogenesis of Membrane Structures Associated with Hepatitis C Virus Replication. *PLOS Pathogens* **8,** e1003056. ISSN: 1553-7374 (2012).

198. Gosert, R., Egger, D., Lohmann, V., Bartenschlager, R., Blum, H. E., Bienz, K. & Moradpour, D. Identification of the hepatitis C virus RNA replication complex in Huh-7 cells harboring subgenomic replicons. *Journal of Virology* **77,** 5487–5492. ISSN: 0022-538X (2003).

199. Wang, H. & Tai, A. W. Mechanisms of Cellular Membrane Reorganization to Support Hepatitis C Virus Replication. *Viruses* **8,** 142 (2016).

200. Shi, S. T., Lee, K.-J., Aizaki, H., Hwang, S. B. & Lai, M. M. C. Hepatitis C Virus RNA Replication Occurs on a Detergent-Resistant Membrane That Cofractionates with Caveolin-2. *Journal of Virology* **77,** 4160–4168. ISSN: 0022-538X, 1098-5514 (2003).

201. Berger, K. L., Cooper, J. D., Heaton, N. S., Yoon, R., Oakland, T. E., Jordan, T. X., Mateu, G., Grakoui, A. & Randall, G. Roles for endocytic trafficking and phosphatidylinositol 4-kinase III alpha in hepatitis C virus replication. *Proceedings of the National Academy of Sciences* **106,** 7577–7582. ISSN: 0027-8424, 1091-6490 (2009).

202. Reiss, S., Rebhan, I., Backes, P., Romero-Brey, I., Erfle, H., Matula, P., Kaderali, L., Poenisch, M., Blankenburg, H., Hiet, M.-S., Longerich, T., Diehl, S., Ramirez, F., Balla, T., Rohr, K., Kaul, A., Bühler, S., Pepperkok, R., Lengauer, T., Albrecht, M., Eils, R., Schirmacher, P., Lohmann, V. & Bartenschlager, R. Recruitment and activation of a lipid kinase by hepatitis C virus NS5A is essential for integrity of the membranous replication compartment. *Cell host & microbe* **9,** 32–45. ISSN: 1931-3128 (2011).

203. Berger, K. L., Kelly, S. M., Jordan, T. X., Tartell, M. A. & Randall, G. Hepatitis C Virus Stimulates the Phosphatidylinositol 4-Kinase III Alpha-Dependent Phosphatidylinositol 4-Phosphate Production That Is Essential for Its Replication. *Journal of Virology* **85,** 8870–8883. ISSN: 0022-538X, 1098-5514 (2011).

204. Wang, H., Perry, J. W., Lauring, A. S., Neddermann, P., De Francesco, R. & Tai, A. W. Oxysterol-Binding Protein is a Phosphatidylinositol 4-kinase Effector Required for HCV Replication Membrane Integrity and Cholesterol Trafficking. *Gastroenterology* **146,** 1373–1385.e11. ISSN: 0016-5085 (2014).

205. Gao, L., Aizaki, H., He, J.-W. & Lai, M. M. C. Interactions between Viral Nonstructural Proteins and Host Protein hVAP-33 Mediate the Formation of Hepatitis C Virus RNA Replication Complex on Lipid Raft. *Journal of Virology* **78,** 3480–3488. ISSN: 0022-538X, 1098-5514 (2004).

206. Hamamoto, I., Nishimura, Y., Okamoto, T., Aizaki, H., Liu, M., Mori, Y., Abe, T., Suzuki, T., Lai, M. M. C., Miyamura, T., Moriishi, K. & Matsuura, Y. Human VAP-B Is Involved in Hepatitis C Virus Replication through Interaction with NS5A and NS5B. *Journal of Virology* **79,** 13473–13482. ISSN: 0022-538X, 1098-5514 (2005).

207. Gupta, G. & Song, J. C-Terminal Auto-Regulatory Motif of Hepatitis C Virus NS5B Interacts with Human VAPB-MSP to Form a Dynamic Replication Complex. *PLoS ONE* **11,** e0147278. ISSN: 1932-6203 (2016).

208. Gupta, G., Qin, H. & Song, J. Intrinsically Unstructured Domain 3 of Hepatitis C Virus NS5A Forms a "Fuzzy Complex" with VAPB-MSP Domain Which Carries ALS-Causing Mutations. *PLoS ONE* **7,** e39261. ISSN: 1932-6203 (2012).

209. Carroll, S. S., Sardana, V., Yang, Z., Jacobs, A. R., Mizenko, C., Hall, D., Hill, L., Zugay-Murphy, J. & Kuo, L. C. Only a Small Fraction of Purified Hepatitis C RNA-Dependent RNA Polymerase Is Catalytically Competent: Implications for Viral Replication and in Vitro Assays. *Biochemistry* **39,** 8243–8249. ISSN: 0006-2960 (2000).

210. McHutchison, J. G., Gordon, S. C., Schiff, E. R., Shiffman, M. L., Lee, W. M., Rustgi, V. K., Goodman, Z. D., Ling, M.-H., Cort, S. & Albrecht, J. K. Interferon Alfa-2b Alone or in Combination with Ribavirin as Initial Treatment for Chronic Hepatitis C. *New England Journal of Medicine* **339,** 1485–1492. ISSN: 0028-4793 (1998).

211. Hoofnagle, J. H., Mullen, K. D., Jones, D. B., Rustgi, V., Di Bisceglie, A., Peters, M., Waggoner, J. G., Park, Y. & Jones, E. A. Treatment of Chronic Non-A, Non-B Hepatitis with Recombinant Human Alpha Interferon. *New England Journal of Medicine* **315,** 1575–1578. ISSN: 0028-4793 (1986).

212. Zhang, T., Lin, R.-T., Li, Y., Douglas, S. D., Maxcey, C., Ho, C., Lai, J.-P., Wang, Y.-J., Wan, Q. & Ho, W.-Z. Hepatitis C virus inhibits intracellular interferon alpha expression in human hepatic cell lines. *Hepatology (Baltimore, Md.)* **42,** 819–827. ISSN: 0270-9139 (2005).

213. Lan, K.-H., Lan, K.-L., Lee, W.-P., Sheu, M.-L., Chen, M.-Y., Lee, Y.-L., Yen, S.-H., Chang, F.-Y. & Lee, S.-D. HCV NS5A inhibits interferon-α signaling through suppression of STAT1 phosphorylation in hepatocyte-derived cell lines. *Journal of Hepatology* **46,** 759–767. ISSN: 0168-8278 (2007).

214. Poynard, T., Marcellin, P., Lee, S. S., Niederau, C., Minuk, G. S., Ideo, G., Bain, V., Heathcote, J., Zeuzem, S., Trepo, C. & Albrecht, J. Randomised trial of interferon α2b plus ribavirin for 48 weeks or for 24 weeks versus interferon α2b plus placebo for 48 weeks for treatment of chronic infection with hepatitis C virus. *The Lancet* **352,** 1426–1432. ISSN: 0140-6736, 1474-547X (1998).

215. Talal, A. H., LaFleur, J., Hoop, R., Pandya, P., Martin, P., Jacobson, I., Han, J. & Korner, E. J. Absolute and relative contraindications to pegylated-interferon or ribavirin in the US general patient population with chronic hepatitis C: results from a US database of over 45 000 HCV-infected, evaluated patients. *Alimentary Pharmacology & Therapeutics* **37,** 473–481. ISSN: 1365-2036 (2013).

216. Rowan, P. J., Tabasi, S., Abdul-Latif, M., Kunik, M. E. & El-Serag, H. B. Psychosocial factors are the most common contraindications for antiviral therapy at initial evaluation in veterans with chronic hepatitis C. *Journal of Clinical Gastroenterology* **38,** 530–534. ISSN: 0192-0790 (2004).

217. Pang, P. S., Planet, P. J. & Glenn, J. S. The Evolution of the Major Hepatitis C Genotypes Correlates with Clinical Response to Interferon Therapy. *PLOS ONE* **4,** e6579. ISSN: 1932-6203 (2009).

218. Götte, M. & Feld, J. J. Direct-acting antiviral agents for hepatitis C: structural and mechanistic insights. *Nature Reviews Gastroenterology & Hepatology* **13,** 338–351. ISSN: 1759-5045 (2016).

219. Saalau-Bethell, S. M., Woodhead, A. J., Chessari, G., Carr, M. G., Coyle, J., Graham, B., Hiscock, S. D., Murray, C. W., Pathuri, P., Rich, S. J., Richardson, C. J., Williams, P. A. & Jhoti, H. Discovery of an allosteric mechanism for the regulation of HCV NS3 protein function. *Nature chemical biology* **8,** 920–925. ISSN: 1552-4450 (2012).

220. McGivern, D. R., Masaki, T., Lovell, W., Hamlett, C., Saalau-Bethell, S. & Graham, B. Protease Inhibitors Block Multiple Functions of the NS3/4A Protease-Helicase during the Hepatitis C Virus Life Cycle. *Journal of Virology* **89,** 5362–5370. ISSN: 0022-538X, 1098-5514 (2015).

221. Gao, M., Nettles, R. E., Belema, M., Snyder, L. B., Nguyen, V. N., Fridell, R. A., Serrano-Wu, M. H., Langley, D. R., Sun, J.-H., O'Boyle, D. R., Lemm, J. A., Wang, C., Knipe, J. O., Chien, C., Colonno, R. J., Grasela, D. M., Meanwell, N. A. & Hamann, L. G. Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect. *Nature* **465,** 96–100. ISSN: 1476-4687 (2010).

222. Lemm, J. A., Leet, J. E., O'Boyle, D. R., Romine, J. L., Huang, X. S., Schroeder, D. R., Alberts, J., Cantone, J. L., Sun, J.-H., Nower, P. T., Martin, S. W., Serrano-Wu, M. H., Meanwell, N. A., Snyder, L. B. & Gao, M. Discovery of Potent Hepatitis C Virus NS5A Inhibitors with Dimeric Structures. *Antimicrobial Agents and Chemotherapy* **55,** 3795–3802. ISSN: 0066-4804 (2011).

223. Scheel, T. K. H. & Rice, C. M. Understanding the hepatitis C virus life cycle paves the way for highly effective therapies. *Nature medicine* **19,** 837–849. ISSN: 1078-8956 (2013).

224. Scheel, T. K. H., Gottwein, J. M., Mikkelsen, L. S., Jensen, T. B. & Bukh, J. Recombinant HCV Variants With NS5A From Genotypes 1–7 Have Different Sensitivities to an NS5A Inhibitor but Not Interferon-α. *Gastroenterology* **140,** 1032–1042.e6. ISSN: 0016-5085, 1528-0012 (2011).

225. Lougiakis, N., Frakolaki, E., Karmou, P., Pouli, N., Marakos, P., Madan, V., Bartenschlager, R. & Vassilaki, N. Novel nucleoside analogues targeting HCV replication through an NS5A-dependent inhibition mechanism. *Chemical Biology & Drug Design,* 1–16. ISSN: 1747-0285 (2017).

226. Fridell, R. A., Qiu, D., Valera, L., Wang, C., Rose, R. E. & Gao, M. Distinct Functions of NS5A in Hepatitis C Virus RNA Replication Uncovered by Studies with the NS5A Inhibitor BMS-790052. *Journal of Virology* **85,** 7312–7320. ISSN: 0022-538X, 1098-5514 (2011).

227. Nakamoto, S., Kanda, T., Wu, S., Shirasawa, H. & Yokosuka, O. Hepatitis C virus NS5A inhibitors and drug resistance mutations. *World Journal of Gastroenterology : WJG* **20,** 2902–2912. ISSN: 1007-9327 (2014).

228. Ascher, D. B., Wielens, J., Nero, T. L., Doughty, L., Morton, C. J. & Parker, M. W. Potent hepatitis C inhibitors bind directly to NS5A and reduce its affinity for RNA. *Scientific Reports* **4,** 4765. ISSN: 2045-2322 (2014).

229. Ahmed, M., Pal, A., Houghton, M. & Barakat, K. A Comprehensive Computational Analysis for the Binding Modes of Hepatitis C Virus NS5A Inhibitors: The Question of Symmetry. *ACS Infectious Diseases* **2,** 872–881 (2016).

230. FDA. *Hepatitis B and C - Hepatitis B and C Treatments* 2017. `https://www.fda.gov/ForPatients/Illness/HepatitisBC/ucm408658.htm` (2017).

231. AALSD. *Recommendations for Testing, Managing, and Treating Hepatitis C* 2017. `http://hcvguidelines.org/sites/default/files/HCV-Guidance_April_2017_a.pdf` (2017).

232. Clercq, E. D. The design of drugs for HIV and HCV. *Nature Reviews Drug Discovery* **6,** 1001–1018. ISSN: 1474-1776 (2007).

233. Delang, L., Neyts, J., Vliegen, I., Abrignani, S., Neddermann, P. & Francesco, R. D. *Hepatitis C Virus: From Molecular Virology to Antiviral Therapy* (ed Bartenschlager, R.) *Current Topics in Microbiology and Immunology* 369. DOI: 10.1007/978-3-642-27340-7_12, 289–320 (Springer Berlin Heidelberg, 2013). ISBN: 978-3-642-27339-1 978-3-642-27340-7.

234. Lam, A. M., Espiritu, C., Bansal, S., Micolochick Steuer, H. M., Niu, C., Zennou, V., Keilman, M., Zhu, Y., Lan, S., Otto, M. J. & Furman, P. A. Genotype and Subtype Profiling of PSI-7977 as a Nucleotide Inhibitor of Hepatitis C Virus. *Antimicrobial Agents and Chemotherapy* **56,** 3359–3368. ISSN: 0066-4804 (2012).

235. Migliaccio, G., Tomassini, J. E., Carroll, S. S., Tomei, L., Altamura, S., Bhat, B., Bartholomew, L., Bosserman, M. R., Ceccacci, A., Colwell, L. F., Cortese, R., Francesco, R. D., Eldrup, A. B., Getty, K. L., Hou, X. S., LaFemina, R. L., Ludmerer, S. W., MacCoss, M., McMasters, D. R., Stahlhut, M. W., Olsen, D. B., Hazuda, D. J. & Flores, O. A. Characterization of Resistance to Non-obligate Chain-terminating Ribonucleoside Analogs That Inhibit Hepatitis C Virus Replication in Vitro. *Journal of Biological Chemistry* **278,** 49164–49170. ISSN: 0021-9258, 1083-351X (2003).

236. Li, H., Tatlock, J., Linton, A., Gonzalez, J., Jewell, T., Patel, L., Ludlum, S., Drowns, M., Rahavendran, S. V., Skor, H., Hunter, R., Shi, S. T., Herlihy, K. J., Parge, H., Hickey, M., Yu, X., Chau, F., Nonomiya, J. & Lewis, C. Discovery of (R)-6-cyclopentyl-6-(2-(2,6-diethylpyridin-4-yl)ethyl)-3-((5,7-dimethyl-[1,2,4]triazolo[1,5-a]pyrimidin-2-yl)methyl)-4-hydroxy-5,6-dihydropyran-2-one (PF-00868554) as a potent and orally available hepatitis C virus polymerase inhibitor. *Journal of Medicinal Chemistry* **52,** 1255–1258. ISSN: 1520-4804 (2009).

237. Zhou, Y., Webber, S. E., Murphy, D. E., Li, L.-S., Dragovich, P. S., Tran, C. V., Sun, Z., Ruebsam, F., Shah, A. M., Tsan, M., Showalter, R. E., Patel, R., Li, B., Zhao, Q., Han, Q., Hermann, T., Kissinger, C. R., LeBrun, L., Sergeeva, M. V. & Kirkovsky, L. Novel HCV NS5B polymerase inhibitors derived from 4-(1,1-dioxo-1,4-dihydro-1λ6-benzo[1,2,4]thiadiazin-3-yl)-5-hydroxy-2H-pyridazin-3-ones. Part 1: Exploration of 7-substitution of benzothiadiazine. *Bioorganic & Medicinal Chemistry Letters* **18,** 1413–1418. ISSN: 0960-894X (2008).

238. Barnes-Seeman, D., Boiselle, C., Capacci-Daniel, C., Chopra, R., Hoffmaster, K., Jones, C. T., Kato, M., Lin, K., Ma, S., Pan, G., Shu, L., Wang, J., Whiteman, L., Xu, M., Zheng, R. & Fu, J. Design and synthesis of lactam–thiophene carboxylic acids as potent hepatitis C virus polymerase inhibitors. *Bioorganic & Medicinal Chemistry Letters* **24,** 3979–3985. ISSN: 0960-894X (2014).

239. Powdrill, M. H., Bernatchez, J. A. & Götte, M. Inhibitors of the Hepatitis C Virus RNA-Dependent RNA Polymerase NS5B. *Viruses* **2,** 2169–2195. ISSN: 1999-4915 (2010).

240. Eltahla, A. A., Luciani, F., White, P. A., Lloyd, A. R. & Bull, R. A. Inhibitors of the Hepatitis C Virus Polymerase; Mode of Action and Resistance. *Viruses* **7,** 5206–5224. ISSN: 1999-4915 (2015).

241. Di Marco, S., Volpari, C., Tomei, L., Altamura, S., Harper, S., Narjes, F., Koch, U., Rowley, M., Francesco, R. D., Migliaccio, G. & Carfí, A. Interdomain Communication in Hepatitis C Virus Polymerase Abolished by Small Molecule Inhibitors Bound to a Novel Allosteric Site. *Journal of Biological Chemistry* **280,** 29765–29770. ISSN: 0021-9258, 1083-351X (2005).

242. Martin Hernando, J. I., Ontoria, J. M., Malancona, S., Attenni, B., Fiore, F., Bonelli, F., Koch, U., Di Marco, S., Colarusso, S., Ponzi, S., Gennari, N., Vignetti, S. E., del Rosario Rico Ferreira, M., Habermann, J., Rowley, M. & Narjes, F. Optimization of Thienopyrrole-Based Finger-Loop Inhibitors of the HepatitisC Virus NS5B Polymerase. *ChemMedChem* **4,** 1695–1713. ISSN: 1860-7187 (2009).

243. Davis, B. C., Brown, J. A. & Thorpe, I. F. Allosteric Inhibitors Have Distinct Effects, but Also Common Modes of Action, in the HCV Polymerase. *Biophysical Journal* **108,** 1785–1795. ISSN: 0006-3495 (2015).

244. Zeuzem, S., Soriano, V., Asselah, T., Bronowicki, J.-P., Lohse, A. W., Müllhaupt, B., Schuchmann, M., Bourlière, M., Buti, M., Roberts, S. K., Gane, E. J., Stern, J. O., Vinisko, R., Kukolj, G., Gallivan, J.-P., Böcher, W.-O. & Mensa, F. J. Faldaprevir and Deleobuvir for HCV Genotype 1 Infection. *New England Journal of Medicine* **369,** 630–639. ISSN: 0028-4793 (2013).

245. Ingelheim, B. *Clinical Trial Results* `https://trials.boehringer-ingelheim.com/trial_results/clinical_trials_overview/clinical_trial_result.c=i.i=33.html` (2017).

246. Yi, G., Deval, J., Fan, B., Cai, H., Soulard, C., Ranjith-Kumar, C. T., Smith, D. B., Blatt, L., Beigelman, L. & Kao, C. C. Biochemical Study of the Comparative Inhibition of Hepatitis C Virus RNA Polymerase by VX-222 and Filibuvir. *Antimicrobial Agents and Chemotherapy* **56,** 830–837. ISSN: 0066-4804 (2012).

247. Di Bisceglie, A. M., Sulkowski, M., Gane, E., Jacobson, I. M., Nelson, D., DeSouza, C., Alves, K., George, S., Kieffer, T., Zhang, E. Z., Kauffman, R., Asmal, M. & Koziel, M. J. VX-222, a non-nucleoside NS5B polymerase inhibitor, in telaprevir-based regimens for genotype 1 hepatitis C virus infection. *European Journal of Gastroenterology & Hepatology* **26,** 761–773. ISSN: 1473-5687 (2014).

248. Deredge, D., Li, J., Johnson, K. A. & Wintrode, P. L. Hydrogen/Deuterium Exchange Kinetics Demonstrate Long Range Allosteric Effects of Thumb Site 2 Inhibitors of Hepatitis C Viral RNA-dependent RNA Polymerase. *Journal of Biological Chemistry* **291,** 10078–10088. ISSN: 0021-9258, 1083-351X (2016).

249. Li, J. & Johnson, K. A. Thumb Site 2 Inhibitors of Hepatitis C Viral RNA-dependent RNA Polymerase Allosterically Block the Transition from Initiation to Elongation. *Journal of Biological Chemistry* **291,** 10067–10077. ISSN: 0021-9258, 1083-351X (2016).

250. Feld, J. J., Kowdley, K. V., Coakley, E., Sigal, S., Nelson, D. R., Crawford, D., Weiland, O., Aguilar, H., Xiong, J., Pilot-Matias, T., DaSilva-Tillmann, B., Larsen, L., Podsadecki, T. & Bernstein, B. Treatment of HCV with ABT-450/r–Ombitasvir and Dasabuvir with Ribavirin. *New England Journal of Medicine* **370,** 1594–1603. ISSN: 0028-4793 (2014).

251. Hebner, C. M., Han, B., Brendza, K. M., Nash, M., Sulfab, M., Tian, Y., Hung, M., Fung, W., Vivian, R. W., Trenkle, J., Taylor, J., Bjornson, K., Bondy, S., Liu, X., Link, J., Neyts, J., Sakowicz, R., Zhong, W., Tang, H. & Schmitz, U. The HCV Non-Nucleoside Inhibitor Tegobuvir Utilizes a Novel Mechanism of Action to Inhibit NS5B Polymerase Function. *PLOS ONE* **7,** e39163. ISSN: 1932-6203 (2012).

252. Li, Y., Yamane, D., Masaki, T. & Lemon, S. M. The yin and yang of hepatitis C: synthesis and decay of hepatitis C virus RNA. *Nature Reviews Microbiology* **13,** 544–558. ISSN: 1740-1526 (2015).

253. MacFarlane, L.-A. & Murphy, P. R. MicroRNA: Biogenesis, Function and Role in Cancer. *Current Genomics* **11,** 537–561. ISSN: 1389-2029 (2010).

254. Jopling, C. L. Regulation of hepatitis C virus by microRNA-122. *Biochemical Society Transactions* **36,** 1220–1223. ISSN: 0300-5127, 1470-8752 (2008).

255. Shimakami, T., Yamane, D., Jangra, R. K., Kempf, B. J., Spaniel, C., Barton, D. J. & Lemon, S. M. Stabilization of hepatitis C virus RNA by an Ago2–miR-122 complex. *Proceedings of the National Academy of Sciences* **109,** 941–946. ISSN: 0027-8424, 1091-6490 (2012).

256. Janssen, H. L., Reesink, H. W., Lawitz, E. J., Zeuzem, S., Rodriguez-Torres, M., Patel, K., van der Meer, A. J., Patick, A. K., Chen, A., Zhou, Y., Persson, R., King, B. D., Kauppinen, S., Levin, A. A. & Hodges, M. R. Treatment of HCV Infection by Targeting MicroRNA. *New England Journal of Medicine* **368,** 1685–1694. ISSN: 0028-4793 (2013).

257. Van der Ree, M. H., van der Meer, A. J., van Nuenen, A. C., de Bruijne, J., Ottosen, S., Janssen, H. L., Kootstra, N. A. & Reesink, H. W. Miravirsen dosing in chronic hepatitis C patients results in decreased microRNA-122 levels without affecting other microRNAs in plasma. *Alimentary Pharmacology & Therapeutics* **43,** 102–113. ISSN: 1365-2036 (2016).

258. Ottosen, S., Parsley, T. B., Yang, L., Zeh, K., van Doorn, L.-J., van der Veer, E., Raney, A. K., Hodges, M. R. & Patick, A. K. In Vitro Antiviral Activity and Preclinical and Clinical Resistance Profile of Miravirsen, a Novel Anti-Hepatitis C Virus Therapeutic Targeting the Human Factor miR-122. *Antimicrobial Agents and Chemotherapy* **59,** 599–608. ISSN: 0066-4804 (2015).

259. Lim, Y.-S. & Hwang, S. B. Hepatitis C Virus NS5A Protein Interacts with Phosphatidylinositol 4-Kinase Type IIIα and Regulates Viral Propagation. *Journal of Biological Chemistry* **286,** 11290–11298. ISSN: 0021-9258, 1083-351X (2011).

260. Lindenbach, B. D. Understanding How Hepatitis C Virus Builds Its Unctuous Home. *Cell host & microbe* **9,** 1–2. ISSN: 1931-3128 (2011).

261. Reiss, S., Harak, C., Romero-Brey, I., Radujkovic, D., Klein, R., Ruggieri, A., Rebhan, I., Bartenschlager, R. & Lohmann, V. The Lipid Kinase Phosphatidylinositol-4 Kinase III Alpha Regulates the Phosphorylation Status of Hepatitis C Virus NS5A. *PLOS Pathogens* **9,** e1003359. ISSN: 1553-7374 (2013).

262. Bianco, A., Reghellin, V., Donnici, L., Fenu, S., Alvarez, R., Baruffa, C., Peri, F., Pagani, M., Abrignani, S., Neddermann, P. & Francesco, R. D. Metabolism of Phosphatidylinositol 4-Kinase IIIα-Dependent PI4P Is Subverted by HCV and Is Targeted by a 4-Anilino Quinazoline with Antiviral Activity. *PLOS Pathogens* **8,** e1002576. ISSN: 1553-7374 (2012).

263. Desai, M. C., Meanwell, N. A. & Thurston, D. E. *Successful Strategies for the Discovery of Antiviral Drugs* Google-Books-ID: vN6NIoNWolUC. ISBN: 978-1-84973-657-2 (Royal Society of Chemistry, 2013).

264. Leivers, A. L., Tallant, M., Shotwell, J. B., Dickerson, S., Leivers, M. R., McDonald, O. B., Gobel, J., Creech, K. L., Strum, S. L., Mathis, A., Rogers, S., Moore, C. B. & Botyanszki, J. Discovery of Selective Small Molecule Type III Phosphatidylinositol 4-Kinase Alpha (PI4KIIIα) Inhibitors as Anti Hepatitis C (HCV) Agents. *Journal of Medicinal Chemistry* **57,** 2091–2106. ISSN: 0022-2623 (2014).

265. Ahmed-Belkacem, A., Colliandre, L., Ahnou, N., Nevers, Q., Gelin, M., Bessin, Y., Brillet, R., Cala, O., Douguet, D., Bourguet, W., Krimm, I., Pawlotsky, J.-M. & Guichou, J.-F. Fragment-based discovery of a new family of non-peptidic small-molecule cyclophilin inhibitors with potent antiviral activities. *Nature Communications* **7,** 12777. ISSN: 2041-1723 (2016).

266. Watashi, K., Hijikata, M., Hosaka, M., Yamaji, M. & Shimotohno, K. Cyclosporin A suppresses replication of hepatitis C virus genome in cultured hepatocytes. *Hepatology* **38,** 1282–1288. ISSN: 1527-3350 (2003).

267. Jin, L. & Harrison, S. C. Crystal structure of human calcineurin complexed with cyclosporin A and human cyclophilin. *Proceedings of the National Academy of Sciences* **99,** 13522–13526. ISSN: 0027-8424, 1091-6490 (2002).

268. Hopkins, S., Bobardt, M., Chatterji, U., Garcia-Rivera, J. A., Lim, P. & Gallay, P. A. The Cyclophilin Inhibitor SCY-635 Disrupts Hepatitis C Virus NS5A-Cyclophilin A Complexes. *Antimicrobial Agents and Chemotherapy* **56,** 3888–3897. ISSN: 0066-4804, 1098-6596 (2012).

269. Ma, S., Boerner, J. E., TiongYip, C., Weidmann, B., Ryder, N. S., Cooreman, M. P. & Lin, K. NIM811, a Cyclophilin Inhibitor, Exhibits Potent In Vitro Activity against Hepatitis C Virus Alone or in Combination with Alpha Interferon. *Antimicrobial Agents and Chemotherapy* **50,** 2976–2982. ISSN: 0066-4804 (2006).

270. Keller, D. M. Pancreatitis Events Halt Development of Alisporivir for HCV. *Medscape.* `http://www.medscape.com/viewarticle/762399` (2017) (2012).

271. Owens, C. M., Brasher, B. B., Polemeropoulos, A., Rhodin, M. H. J., McAllister, N., Peng, X., Wang, C., Ying, L., Cao, H., Lawitz, E., Poordad, F., Rondon, J., Box, T. D., Zeuzem, S., Buggisch, P., Lin, K., Qiu, Y.-L., Jiang, L., Colvin, R. & Or, Y. S. Preclinical Profile and Clinical Efficacy of a Novel Hepatitis C Virus NS5A Inhibitor, EDP-239. *Antimicrobial Agents and Chemotherapy* **60,** 6207–6215. ISSN: 0066-4804, 1098-6596 (2016).

272. Coelmont, L., Hanoulle, X., Chatterji, U., Berger, C., Snoeck, J., Bobardt, M., Lim, P., Vliegen, I., Paeshuyse, J., Vuagniaux, G., Vandamme, A.-M., Bartenschlager, R., Gallay, P., Lippens, G. & Neyts, J. DEB025 (Alisporivir) Inhibits Hepatitis C Virus Replication by Preventing a Cyclophilin A Induced Cis-Trans Isomerisation in Domain II of NS5A. *PLoS ONE* **5,** e13687. ISSN: 1932-6203 (2010).

273. Robida, J. M., Nelson, H. B., Liu, Z. & Tang, H. Characterization of Hepatitis C Virus Subgenomic Replicon Resistance to Cyclosporine In Vitro. *Journal of Virology* **81,** 5829–5840. ISSN: 0022-538X (2007).

274. Qing, M., Yang, F., Zhang, B., Zou, G., Robida, J. M., Yuan, Z., Tang, H. & Shi, P.-Y. Cyclosporine Inhibits Flavivirus Replication through Blocking the Interaction between Host Cyclophilins and Viral NS5 Protein. *Antimicrobial Agents and Chemotherapy* **53,** 3226–3235. ISSN: 0066-4804 (2009).

275. Bekerman, E. & Einav, S. Combating emerging viral threats. *Science* **348,** 282–283. ISSN: 0036-8075, 1095-9203 (2015).

276. Chiramel, A. I., Banadyga, L., Dougherty, J. D., Falzarano, D., Martellaro, C., Brees, D., Taylor, R. T., Ebihara, H. & Best, S. M. Alisporivir Has Limited Antiviral Effects Against Ebola Virus Strains Makona and Mayinga. *The Journal of Infectious Diseases* **214,** S355–S359. ISSN: 1537-6613 (2016).

277. HIVandHepatitis.com. *Vertex to Discontinue Sale of Telaprevir (Incivek) for Hepatitis C* 2014. `http://www.hivandhepatitis.com/hcv-treatment/approved-hcv-drugs/4808-vertex-to-discontinue-sale-of-telaprevir-incivek-for-hepatitis-c` (2017).

278. Bader, P. J. *Merck Voluntarily Discontinuing VICTRELIS® (boceprevir) 200 mg Capsules* 2015. `https://www.fda.gov/downloads/drugs/drugsafety/drugshortages/ucm430818.pdf` (2017).

279. EASL. EASL Recommendations on Treatment of Hepatitis C 2016. *Journal of Hepatology* **66,** 153–194. ISSN: 0168-8278, 1600-0641 (2017).

280. Yao, Y., Yue, M., Wang, J., Chen, H., Liu, M., Zang, F., Li, J., Zhang, Y., Huang, P. & Yu, R. Grazoprevir and Elbasvir in Patients with Genotype 1 Hepatitis C Virus Infection: A Comprehensive Efficacy and Safety Analysis. *Canadian Journal of Gastroenterology & Hepatology* **2017,** 8186275. ISSN: 2291-2789 (2017).

281. Morin, S., Klein, R. & Struble, K. *FDA approves two Hepatitis C drugs for pediatric patients* 2017. `https://content.govdelivery.com/accounts/USFDA/bulletins/192d767` (2017).

282. Moradpour, D., Grakoui, A. & Manns, M. P. Future landscape of hepatitis C research – Basic, translational and clinical perspectives. *Journal of Hepatology* **65,** S143–S155. ISSN: 0168-8278, 1600-0641 (2016).

283. Diamond, D. L., Syder, A. J., Jacobs, J. M., Sorensen, C. M., Walters, K.-A., Proll, S. C., McDermott, J. E., Gritsenko, M. A., Zhang, Q., Zhao, R., Metz, T. O., Ii, D. G. C., Waters, K. M., Smith, R. D., Rice, C. M. & Katze, M. G. Temporal Proteome and Lipidome Profiles Reveal Hepatitis C Virus-Associated Reprogramming of Hepatocellular Metabolism and Bioenergetics. *PLOS Pathogens* **6,** e1000719. ISSN: 1553-7374 (2010).

284. Popescu, C.-I., Riva, L., Vlaicu, O., Farhat, R., Rouillé, Y. & Dubuisson, J. Hepatitis C Virus Life Cycle and Lipid Metabolism. *Biology* **3,** 892–921. ISSN: 2079-7737 (2014).

285. Lerat, H., Kammoun, H. L., Hainault, I., Mérour, E., Higgs, M. R., Callens, C., Lemon, S. M., Foufelle, F. & Pawlotsky, J.-M. Hepatitis C Virus Proteins Induce Lipogenesis and Defective Triglyceride Secretion in Transgenic Mice. *The Journal of Biological Chemistry* **284,** 33466–33474. ISSN: 0021-9258 (2009).

286. Horton, J. D., Goldstein, J. L. & Brown, M. S. SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *The Journal of Clinical Investigation* **109,** 1125–1131. ISSN: 0021-9738 (2002).

287. Ivanov, A. V., Bartosch, B., Smirnova, O. A., Isaguliants, M. G. & Kochetkov, S. N. HCV and Oxidative Stress in the Liver. *Viruses* **5,** 439–469. ISSN: 1999-4915 (2013).

288. Paul, D. & Bartenschlager, R. A sensor at the lipid-protein interface: Lipid peroxidation controls hepatitis C virus replication. *Hepatology* **61,** 1083–1085. ISSN: 1527-3350 (2015).

289. Amako, Y., Munakata, T., Kohara, M., Siddiqui, A., Peers, C. & Harris, M. Hepatitis C Virus Attenuates Mitochondrial Lipid β-Oxidation by Downregulating Mitochondrial Trifunctional-Protein Expression. *Journal of Virology* **89,** 4092–4101. ISSN: 0022-538X, 1098-5514 (2015).

290. Domitrovich, A. M., Felmlee, D. J. & Siddiqui, A. Hepatitis C Virus Nonstructural Proteins Inhibit Apolipoprotein B100 Secretion. *Journal of Biological Chemistry* **280,** 39802–39808. ISSN: 0021-9258, 1083-351X (2005).

291. Syed, G. H., Amako, Y. & Siddiqui, A. Hepatitis C Virus Hijacks Host Lipid Metabolism. *Trends in endocrinology and metabolism: TEM* **21,** 33. ISSN: 1043-2760 (2010).

292. Filipe, A. & McLauchlan, J. Hepatitis C virus and lipid droplets: finding a niche. *Trends in Molecular Medicine* **21,** 34–42. ISSN: 1471-4914 (2015).

293. Siagris, D., Christofidou, M., Theocharis, G. J., Pagoni, N., Papadimitriou, C., Lekkou, A., Thomopoulos, K., Starakis, I., Tsamandas, A. C. & Labropoulou-Karatza, C. Serum lipid pattern in chronic hepatitis C: histological and virological correlations. *Journal of Viral Hepatitis* **13,** 56–61. ISSN: 1365-2893 (2006).

294. Liefhebber, J. M. P., Hague, C. V., Zhang, Q., Wakelam, M. J. O. & McLauchlan, J. Modulation of Triglyceride and Cholesterol Ester Synthesis Impairs Assembly of Infectious Hepatitis C Virus. *Journal of Biological Chemistry* **289,** 21276–21288. ISSN: 0021-9258, 1083-351X (2014).

295. Saito, K., Shirasago, Y., Suzuki, T., Aizaki, H., Hanada, K., Wakita, T., Nishijima, M. & Fukasawa, M. Targeting Cellular Squalene Synthase, an Enzyme Essential for Cholesterol Biosynthesis, Is a Potential Antiviral Strategy against Hepatitis C Virus. *Journal of Virology* **89,** 2220–2232. ISSN: 0022-538X, 1098-5514 (2015).

296. Ramière, C., Rodriguez, J., Enache, L. S., Lotteau, V., André, P. & Diaz, O. Activity of Hexokinase Is Increased by Its Interaction with Hepatitis C Virus Protein NS5A. *Journal of Virology* **88,** 3246–3254. ISSN: 0022-538X, 1098-5514 (2014).

297. Lichtenthaler, F. W. 100 Years "Schlüssel-Schloss-Prinzip": What Made Emil Fischer Use this Analogy? *Angewandte Chemie International Edition in English* **33,** 2364–2374. ISSN: 1521-3773 (1995).

298. Koshland, D. E. The Key–Lock Theory and the Induced Fit Theory. *Angewandte Chemie International Edition in English* **33,** 2375–2378. ISSN: 1521-3773 (1995).

299. Romero, P., Obradovic, Z., Kissinger, C. R., Villafranca, J. E., Garner, E., Guilliot, S. & Dunker, A. K. Thousands of proteins likely to have long disordered regions. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, 437–448. ISSN: 2335-6936 (1998).

300. Wright, P. E. & Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of Molecular Biology* **293,** 321–331. ISSN: 0022-2836 (1999).

301. Lise, S. & Jones, D. T. Sequence patterns associated with disordered regions in proteins. *Proteins: Structure, Function, and Bioinformatics* **58,** 144–150. ISSN: 1097-0134 (2005).

302. Papoian, G. A. Proteins with weakly funneled energy landscapes challenge the classical structure–function paradigm. *Proceedings of the National Academy of Sciences* **105,** 14237–14238. ISSN: 0027-8424, 1091-6490 (2008).

303. Smith, L. J., Fiebig, K. M., Schwalbe, H. & Dobson, C. M. The concept of a random coil: Residual structure in peptides and denatured proteins. *Folding and Design* **1,** R95–R106. ISSN: 1359-0278 (1996).

304. Konrat, R. NMR contributions to structural dynamics studies of intrinsically disordered proteins. *Journal of Magnetic Resonance. A special "JMR Perspectives" issue: Foresights in Biomolecular Solution-State NMR Spectroscopy – From Spin Gymnastics to Structure and Dynamics* **241,** 74–85. ISSN: 1090-7807 (2014).

305. Kingston, R. L., Hamel, D. J., Gay, L. S., Dahlquist, F. W. & Matthews, B. W. Structural basis for the attachment of a paramyxoviral polymerase to its template. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 8301–8306. ISSN: 0027-8424 (2004).

306. Fändrich, M. & Dobson, C. M. The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *The EMBO Journal* **21,** 5682–5690. ISSN: 0261-4189, 1460-2075 (2002).

307. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nature protocols* **1,** 2876–2890. ISSN: 1754-2189 (2006).

308. Woody, R. W. *Instrumental Analysis of Intrinsically Disordered Proteins* (eds Uversky, V. N. & Longhi, S.) DOI: 10.1002/9780470602614.ch10, 303–321 (John Wiley & Sons, Inc., 2010). ISBN: 978-0-470-60261-4. http://onlinelibrary.wiley.com/doi/10.1002/9780470602614.ch10/summary.

309. Kikhney, A. G. & Svergun, D. I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Letters* **589,** 2570–2577. ISSN: 1873-3468 (2015).

310. Uversky, V. & Longhi, S. *Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure and Conformation* 1 edition. ISBN: 978-0-470-34341-8 (Wiley, Hoboken, N.J, 2010).

311. Bah, A. & Forman-Kay, J. D. Modulation of Intrinsically Disordered Protein Function by Post-Translational Modifications. *Journal of Biological Chemistry*, jbc.R115.695056. ISSN: 0021-9258, 1083-351X (2016).

312. Tompa, P. Intrinsically unstructured proteins. *Trends in Biochemical Sciences* **27,** 527–533. ISSN: 0968-0004 (2002).

313. Tompa, P. The interplay between structure and function in intrinsically unstructured proteins. *FEBS Letters* **579,** 3346–3354. ISSN: 1873-3468 (2005).

314. Van der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D. T., Kim, P. M., Kriwacki, R. W., Oldfield, C. J., Pappu, R. V., Tompa, P., Uversky, V. N., Wright, P. E. & Babu, M. M. Classification of Intrinsically Disordered Regions and Proteins. *Chemical Reviews* **114,** 6589–6631. ISSN: 0009-2665 (2014).

315. Tompa, P. & Csermely, P. The role of structural disorder in the function of RNA and protein chaperones. *The FASEB Journal* **18,** 1169–1175. ISSN: 0892-6638, 1530-6860 (2004).

316. Holt, C. Unfolded phosphopolypeptides enable soft and hard tissues to coexist in the same organism with relative ease. *Current Opinion in Structural Biology. New contructs and expressions of proteins / Sequences and topology* **23,** 420–425. ISSN: 0959-440X (2013).

317. Chen, J. W., Romero, P., Uversky, V. N. & Dunker, A. K. Conservation of Intrinsic Disorder in Protein Domains and Families: I. A Database of Conserved Predicted Disordered Regions. *Journal of proteome research* **5,** 879–887. ISSN: 1535-3893 (2006).

318. Gitlin, L., Hagai, T., LaBarbera, A., Solovey, M. & Andino, R. Rapid Evolution of Virus Sequences in Intrinsically Disordered Protein Regions. *PLoS Pathogens* **10.** ISSN: 1553-7366. doi:10.1371/journal.ppat.1004529. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4263755/ (2014).

319. Xue, B., Blocquel, D., Habchi, J., Uversky, A. V., Kurgan, L., Uversky, V. N. & Longhi, S. Structural Disorder in Viral Proteins. *Chemical Reviews* **114,** 6880–6911. ISSN: 0009-2665 (2014).

320. Mollica, L., Bessa, L. M., Hanoulle, X., Jensen, M. R., Blackledge, M. & Schneider, R. Binding Mechanisms of Intrinsically Disordered Proteins: Theory, Simulation, and Experiment. *Frontiers in Molecular Biosciences* **3,** 52. ISSN: 2296-889X (2016).

321. Dogan, J., Gianni, S. & Jemth, P. The binding mechanisms of intrinsically disordered proteins. *Physical Chemistry Chemical Physics* **16,** 6323–6331. ISSN: 1463-9084 (2014).

322. Tompa, P. & Fuxreiter, M. Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends in Biochemical Sciences* **33,** 2–8. ISSN: 0968-0004 (2008).

323. Mittag, T., Orlicky, S., Choy, W.-Y., Tang, X., Lin, H., Sicheri, F., Kay, L. E., Tyers, M. & Forman-Kay, J. D. Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proceedings of the National Academy of Sciences of the United States of America* **105,** 17772–17777. ISSN: 0027-8424 (2008).

324. Mittag, T., Marsh, J., Grishaev, A., Orlicky, S., Lin, H., Sicheri, F., Tyers, M. & Forman-Kay, J. D. Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. *Structure (London, England : 1993)* **18,** 494–506. ISSN: 0969-2126 (2010).

325. Liu, M., Mao, X.-a., Ye, C., Huang, H., Nicholson, J. K. & Lindon, J. C. Improved WATERGATE Pulse Sequences for Solvent Suppression in NMR Spectroscopy. *Journal of Magnetic Resonance* **132,** 125–129. ISSN: 1090-7807 (1998).

326. Mulder, F. A. A., Lundqvist, M. & Scheek, R. M. *Instrumental Analysis of Intrinsically Disordered Proteins* (eds Uversky, V. N. & Longhi, S.) DOI: 10.1002/9780470602614.ch3, 59–87 (John Wiley & Sons, Inc., 2010). ISBN: 978-0-470-60261-4. `http://onlinelibrary.wiley.com/doi/10.1002/9780470602614.ch3/summary`.

327. Cavanagh, J., Fairbrother, W. J., III, A. G. P., Skelton, N. J. & Rance, M. *Protein NMR Spectroscopy: Principles and Practice* ISBN: 978-0-08-047103-7 (Academic Press, 2010).

328. Jung, Y.-S. & Zweckstetter, M. MARS - robust automatic backbone assignment of proteins. *Journal of Biomolecular NMR* **30,** 11–23. ISSN: 0925-2738, 1573-5001 (2004).

329. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. & Sykes, B. D. 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *Journal of Biomolecular NMR* **5,** 67–81. ISSN: 0925-2738, 1573-5001 (1995).

330. Wishart, D. S., Sykes, B. D. & Richards, F. M. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31,** 1647–1651. ISSN: 0006-2960 (1992).

331. Wishart, D. S. & Sykes, B. D. The 13C Chemical-Shift Index: A simple method for the identification of protein secondary structure using 13C chemical-shift data. *Journal of Biomolecular NMR* **4,** 171–180. ISSN: 0925-2738, 1573-5001 (1994).

332. Marsh, J. A., Singh, V. K., Jia, Z. & Forman-Kay, J. D. Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. *Protein Science : A Publication of the Protein Society* **15,** 2795–2804. ISSN: 0961-8368 (2006).

333. Tamiola, K., Acar, B. & Mulder, F. A. A. Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. *Journal of the American Chemical Society* **132,** 18000–18003. ISSN: 0002-7863 (2010).

334. Li, F., Lee, J. H., Grishaev, A., Ying, J. & Bax, A. High Accuracy of Karplus Equations for Relating Three-Bond J Couplings to Protein Backbone Torsion Angles. *Chemphyschem : a European journal of chemical physics and physical chemistry* **16,** 572–578. ISSN: 1439-4235 (2015).

335. Vögeli, B., Ying, J., Grishaev, A. & Bax, A. Limits on Variations in Protein Backbone Dynamics from Precise Measurements of Scalar Couplings. *Journal of the American Chemical Society* **129,** 9377–9385. ISSN: 0002-7863 (2007).

336. Feuerstein, S., Solyom, Z., Aladag, A., Favier, A., Schwarten, M., Hoffmann, S., Willbold, D. & Brutscher, B. Transient Structure and SH3 Interaction Sites in an Intrinsically Disordered Fragment of the Hepatitis C Virus Protein NS5A. *Journal of Molecular Biology* **420,** 310–323. ISSN: 0022-2836 (2012).

337. Reich, H. J. *8.1 Relaxation in NMR Spectroscopy* 2017. `https://www.chem.wisc.edu/areas/reich/nmr/08-tech-01-relax.htm` (2017).

338. Crespi, H. L., Rosenberg, R. M. & Katz, J. J. Proton Magnetic Resonance of Proteins Fully Deuterated except for 1H-Leucine Side Chains. *Science* **161,** 795–796. ISSN: 0036-8075, 1095-9203 (1968).

339. Sattler, M. & Fesik, S. W. Use of deuterium labeling in NMR: overcoming a sizeable problem. *Structure* **4,** 1245–1249. ISSN: 0969-2126 (1996).

340. Venters, R. A., Huang, C.-C., Farmer, B. T., Trolard, R., Spicer, L. D. & Fierke, C. A. High-level 2H/13C/15N labeling of proteins for NMR studies. *Journal of Biomolecular NMR* **5,** 339–344. ISSN: 0925-2738, 1573-5001 (1995).

341. Morgan, W. D., Kragt, A. & Feeney, J. Expression of deuterium-isotope-labelled protein in the yeast Pichia pastoris for NMR studies. *Journal of Biomolecular NMR* **17,** 337–347. ISSN: 0925-2738, 1573-5001 (2000).

342. Etezady-Esfarjani, T., Hiller, S., Villalba, C. & Wüthrich, K. Cell-free protein synthesis of perdeuterated proteins for NMR studies. *Journal of Biomolecular NMR* **39,** 229–238. ISSN: 0925-2738, 1573-5001 (2007).

343. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences* **94,** 12366–12371. ISSN: 0027-8424, 1091-6490 (1997).

344. Salzmann, M., Wider, G., Pervushin, K., Senn, H. & Wüthrich, K. TROSY-type Triple-Resonance Experiments for Sequential NMR Assignments of Large Proteins. *Journal of the American Chemical Society* **121,** 844–848. ISSN: 0002-7863 (1999).

345. Pervushin, K. V., Wider, G., Riek, R. & Wüthrich, K. The 3D NOESY-[1H,15N,1H]-ZQ-TROSY NMR experiment with diagonal peak suppression. *Proceedings of the National Academy of Sciences of the United States of America* **96,** 9607–9612. ISSN: 0027-8424 (1999).

346. Jaipuria, G., Krishnarjuna, B., Mondal, S., Dubey, A. & Atreya, H. S. *Isotope labeling in Biomolecular NMR* DOI: 10.1007/978-94-007-4954-2_6, 95–118 (Springer, Dordrecht, 2012). ISBN: 978-94-007-4953-5 978-94-007-4954-2. `https://link-springer-com/chapter/10.1007/978-94-007-4954-2_6` (2017).

347. Kay, L. E. & Gardner, K. H. Solution NMR spectroscopy beyond 25 kDa. *Current Opinion in Structural Biology* **7,** 722–731. ISSN: 0959-440X (1997).

348. Ruschak, A. M. & Kay, L. E. Methyl groups as probes of supra-molecular structure, dynamics and function. *Journal of Biomolecular NMR* **46,** 75–87. ISSN: 0925-2738, 1573-5001 (2010).

349. Ayala, I., Sounier, R., Usé, N., Gans, P. & Boisbouvier, J. An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. *Journal of Biomolecular NMR* **43,** 111–119. ISSN: 0925-2738, 1573-5001 (2009).

350. Zheng, X., Mueller, G. A., DeRose, E. F. & London, R. E. Solution characterization of [methyl-13C]methionine HIV-1 reverse transcriptase by NMR spectroscopy. *Antiviral Research* **84,** 205–214. ISSN: 0166-3542 (2009).

351. Sinha, K., Jen-Jacobson, L. & Rule, G. S. Specific Labeling of Threonine Methyl Groups for NMR Studies of Protein–Nucleic Acid Complexes. *Biochemistry* **50,** 10189–10191. ISSN: 0006-2960 (2011).

352. Kerfah, R., Plevin, M. J., Sounier, R., Gans, P. & Boisbouvier, J. Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. *Current Opinion in Structural Biology* **32.** WOS:000359032300016, 113–122. ISSN: 0959-440X (2015).

353. Goto, N. K., Gardner, K. H., Mueller, G. A., Willis, R. C. & Kay, L. E. A robust and cost-effective method for the production of Val, Leu, Ile (δ1) methyl-protonated 15N-, 13C-, 2H-labeled proteins. *Journal of Biomolecular NMR* **13,** 369–374. ISSN: 0925-2738, 1573-5001 (1999).

354. Gardner, K. H. & Kay, L. E. Production and Incorporation of 15N, 13C, 2H (1H-δ1 Methyl) Isoleucine into Proteins for Multidimensional NMR Studies. *Journal of the American Chemical Society* **119,** 7599–7600. ISSN: 0002-7863 (1997).

355. Ruschak, A. M., Velyvis, A. & Kay, L. E. A simple strategy for 13C,1H labeling at the Ile-γ2 methyl position in highly deuterated proteins. *Journal of Biomolecular NMR* **48,** 129–135. ISSN: 0925-2738, 1573-5001 (2010).

356. Ayala, I., Hamelin, O., Amero, C., Pessey, O., Plevin, M. J., Gans, P. & Boisbouvier, J. An optimized isotopic labelling strategy of isoleucine-γ2 methyl groups for solution NMR studies of high molecular weight proteins. *Chemical Communications* **48,** 1434–1436. ISSN: 1364-548X (2012).

357. Tugarinov, V., Hwang, P. M., Ollerenshaw, J. E. & Kay, L. E. Cross-Correlated Relaxation Enhanced 1H13C NMR Spectroscopy of Methyl Groups in Very High Molecular Weight Proteins and Protein Complexes. *Journal of the American Chemical Society* **125,** 10420–10428. ISSN: 0002-7863 (2003).

358. Tugarinov, V. & Kay, L. E. An Isotope Labeling Strategy for Methyl TROSY Spectroscopy. *Journal of Biomolecular NMR* **28,** 165–172. ISSN: 0925-2738, 1573-5001 (2004).

359. Korzhnev, D. M., Kloiber, K., Kanelis, V., Tugarinov, V. & Kay, L. E. Probing Slow Dynamics in High Molecular Weight Proteins by Methyl-TROSY NMR Spectroscopy: Application to a 723-Residue Enzyme. *Journal of the American Chemical Society* **126,** 3964–3973. ISSN: 0002-7863 (2004).

360. Ollerenshaw, J. E., Tugarinov, V. & Kay, L. E. Methyl TROSY: explanation and experimental verification. *Magnetic Resonance in Chemistry* **41,** 843–852. ISSN: 1097-458X (2003).

361. Wiesner, S. & Sprangers, R. Methyl groups as NMR probes for biomolecular interactions. *Current Opinion in Structural Biology. Catalysis and regulation Protein-protein interactions* **35,** 60–67. ISSN: 0959-440X (2015).

362. Sprangers, R. & Kay, L. E. Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* **445,** 618–622. ISSN: 0028-0836 (2007).

363. Gelis, I., Bonvin, A. M., Keramisanou, D., Koukaki, M., Gouridis, G., Karamanou, S., Economou, A. & Kalodimos, C. G. Structural basis for signal sequence recognition by the 204-kDa translocase motor SecA determined by NMR. *Cell* **131,** 756–769. ISSN: 0092-8674 (2007).

364. Lapinaite, A., Simon, B., Skjaerven, L., Rakwalska-Bange, M., Gabel, F. & Carlomagno, T. The structure of the box C/D enzyme reveals regulation of RNA methylation. *Nature* **502,** 519–523. ISSN: 0028-0836 (2013).

365. Cvetkovic, M. A., Wurm, J. P., Audin, M. J., Schütz, S. & Sprangers, R. The Rrp4-exosome complex recruits and channels substrate RNA by a unique mechanism. *Nature Chemical Biology* **13,** 522–528. ISSN: 1552-4450 (2017).

366. Xue, Y., Pavlova, M. S., Ryabov, Y. E., Reif, B. & Skrynnikov, N. R. Methyl Rotation Barriers in Proteins from 2H Relaxation Data. Implications for Protein Structure. *Journal of the American Chemical Society* **129,** 6827–6838. ISSN: 0002-7863 (2007).

367. Mittermaier, A. K. & Kay, L. E. Observing biological dynamics at atomic resolution using NMR. *Trends in Biochemical Sciences* **34,** 601–611. ISSN: 0968-0004 (2009).

368. Palmer, A. G., Kroenke, C. D. & Patrick Loria, J. Nuclear Magnetic Resonance Methods for Quantifying Microsecond-to-Millisecond Motions in Biological Macromolecules. *Methods in Enzymology. Nuclear Magnetic Resonance of Biological Macromolecules - Part B* **339,** 204–238. ISSN: 0076-6879 (2001).

369. Anthis, N. J. & Clore, G. M. Visualizing transient dark states by NMR spectroscopy. *Quarterly Reviews of Biophysics* **48.** WOS:000350304300002, 35–116. ISSN: 0033-5835 (2015).

370. Yuwen, T., Huang, R. & Kay, L. E. Probing slow timescale dynamics in proteins using methyl 1H CEST. *Journal of Biomolecular NMR* **68,** 215–224. ISSN: 0925-2738, 1573-5001 (2017).

371. Korzhnev, D. M., Kloiber, K. & Kay, L. E. Multiple-Quantum Relaxation Dispersion NMR Spectroscopy Probing Millisecond Time-Scale Dynamics in Proteins: Theory and Application. *Journal of the American Chemical Society* **126,** 7320–7329. ISSN: 0002-7863 (2004).

372. Kovermann, M., Rogne, P. & Wolf-Watz, M. Protein dynamics and function from solution state NMR spectroscopy. *Quarterly Reviews of Biophysics* **49.** ISSN: 0033-5835, 1469-8994. doi:10.1017/S0033583516000019. https://www.cambridge.org/core/journals/quarterly-reviews-of-biophysics/article/protein-dynamics-and-function-from-solution-state-nmr-spectroscopy/12660D14E19E7A261932103F7F3761B6 (2016).

373. Simon, B. & Sattler, M. De Novo Structure Determination from Residual Dipolar Couplings by NMR Spectroscopy. *Angewandte Chemie International Edition* **41,** 437–440. ISSN: 1521-3773 (2002).

374. Chen, K. & Tjandra, N. *NMR of Proteins and Small Biomolecules* DOI: 10.1007/128_2011_215, 47–67 (Springer, Berlin, Heidelberg, 2011). ISBN: 978-3-642-28916-3 978-3-642-28917-0. https://link-springer-com/chapter/10.1007/128_2011_215 (2017).

375. Tolman, J. R. & Ruan, K. NMR Residual Dipolar Couplings as Probes of Biomolecular Dynamics. *Chemical Reviews* **106,** 1720–1736. ISSN: 0009-2665 (2006).

376. Lange, O. F., Lakomek, N.-A., Farès, C., Schröder, G. F., Walter, K. F. A., Becker, S., Meiler, J., Grubmüller, H., Griesinger, C. & Groot, B. L. d. Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution. *Science* **320,** 1471–1475. ISSN: 0036-8075, 1095-9203 (2008).

377. Charlier, C., Cousin, S. F. & Ferrage, F. Protein dynamics from nuclear magnetic relaxation. *Chemical Society Reviews* **45,** 2410–2422. ISSN: 1460-4744 (2016).

378. Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy* **73,** 1–16. ISSN: 0079-6565 (2013).

379. Sounier, R., Blanchard, L., Wu, Z. & Boisbouvier, J. High-Accuracy Distance Measurement between Remote Methyls in Specifically Protonated Proteins. *Journal of the American Chemical Society* **129,** 472–473. ISSN: 0002-7863 (2007).

380. Muhandiram, R. & Kay, L. E. *eMagRes* DOI: 10.1002/9780470034590.emrstm0563 (John Wiley & Sons, Ltd, 2007). ISBN: 978-0-470-03459-0. http://onlinelibrary.wiley.com/doi/10.1002/9780470034590.emrstm0563/abstract.

381. Battiste, J. L. & Wagner, G. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* **39,** 5355–5365. ISSN: 0006-2960 (2000).

382. Su, X.-C. & Otting, G. Paramagnetic labelling of proteins and oligonucleotides for NMR. *Journal of Biomolecular NMR* **46,** 101–112. ISSN: 0925-2738, 1573-5001 (2010).

383. Huang, S., Umemoto, R., Tamura, Y., Kofuku, Y., Uyeda, T. Q. P., Nishida, N. & Shimada, I. Utilization of paramagnetic relaxation enhancements for structural analysis of actin-binding proteins in complex with actin. *Scientific Reports* **6,** 33690. ISSN: 2045-2322 (2016).

384. Tang, C., Schwieters, C. D. & Clore, G. M. Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR. *Nature* **449,** 1078–1082. ISSN: 0028-0836 (2007).

385. Cohn, M. & Reuben, J. Paramagnetic probes in magnetic resonance studies of phosphoryl transfer enzymes. *Accounts of Chemical Research* **4,** 214–222. ISSN: 0001-4842 (1971).

386. Salmon, L., Nodet, G., Ozenne, V., Yin, G., Jensen, M. R., Zweckstetter, M. & Blackledge, M. NMR Characterization of Long-Range Order in Intrinsically Disordered Proteins. *Journal of the American Chemical Society* **132,** 8407–8418. ISSN: 0002-7863 (2010).

387. Tzeng, S.-R., Pai, M.-T. & Kalodimos, C. G. NMR Studies of Large Protein Systems. *Methods in molecular biology (Clifton, N.J.)* **831,** 133–140. ISSN: 1064-3745 (2012).

388. Combet, C., Garnier, N., Charavay, C., Grando, D., Crisan, D., Lopez, J., Dehne-Garcia, A., Geourjon, C., Bettler, E., Hulo, C., Mercier, P. L., Bartenschlager, R., Diepolder, H., Moradpour, D., Pawlotsky, J.-M., Rice, C. M., Trepo, C., Penin, F. & Deleage, G. euHCVdb: the European hepatitis C virus database. *Nucleic Acids Research* **35,** D363–D366. ISSN: 0305-1048 (2007).

389. Goddard, T. D. & Kneller, D. G. *SPARKY 3* San Francisco, 2007.

390. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31,** 1325–1327. ISSN: 1367-4803 (2015).

391. Lescop, E., Schanda, P. & Brutscher, B. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *Journal of Magnetic Resonance* **187,** 163–169. ISSN: 1090-7807 (2007).

392. Han, B., Liu, Y., Ginzinger, S. W. & Wishart, D. S. SHIFTX2: significantly improved protein chemical shift prediction. *Journal of Biomolecular NMR* **50,** 43–57. ISSN: 0925-2738, 1573-5001 (2011).

393. Schmitt, M., Scrima, N., Radujkovic, D., Caillet-Saguy, C., Simister, P. C., Friebe, P., Wicht, O., Klein, R., Bartenschlager, R., Lohmann, V. & Bressanelli, S. A Comprehensive Structure-Function Comparison of Hepatitis C Virus Strain JFH1 and J6 Polymerases Reveals a Key Residue Stimulating Replication in Cell Culture across Genotypes. *Journal of Virology* **85,** 2565–2581. ISSN: 0022-538X, 1098-5514 (2011).

394. Lam, A. M., Edwards, T. E., Mosley, R. T., Murakami, E., Bansal, S., Lugo, C., Bao, H., Otto, M. J., Sofia, M. J. & Furman, P. A. Molecular and Structural Basis for the Roles of Hepatitis C Virus Polymerase NS5B Amino Acids 15, 223, and 321 in Viral Replication and Drug Resistance. *Antimicrobial Agents and Chemotherapy* **58,** 6861–6869. ISSN: 0066-4804, 1098-6596 (2014).

395. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of biomolecular NMR* **6,** 277–293. ISSN: 0925-2738 (1995).

396. Ishima, R. & Torchia, D. A. Error estimation and global fitting in transverse-relaxation dispersion experiments to determine chemical-exchange parameters. *Journal of Biomolecular NMR* **32,** 41–54. ISSN: 0925-2738, 1573-5001 (2005).

397. Korzhnev, D. M., Neudecker, P., Mittermaier, A., Orekhov, V. Y. & Kay, L. E. Multiple-Site Exchange in Proteins Studied with a Suite of Six NMR Relaxation Dispersion Experiments: An Application to the Folding of a Fyn SH3 Domain Mutant. *Journal of the American Chemical Society* **127,** 15602–15611. ISSN: 0002-7863 (2005).

398. Yamashita, T., Kaneko, S., Shirota, Y., Qin, W., Nomura, T., Kobayashi, K. & Murakami, S. RNA-dependent RNA Polymerase Activity of the Soluble Recombinant Hepatitis C Virus NS5B Protein Truncated at the C-terminal Region. *Journal of Biological Chemistry* **273,** 15479–15486. ISSN: 0021-9258, 1083-351X (1998).

399. Abe, K.-i., Ikeda, M., Ariumi, Y., Dansako, H., Wakita, T. & Kato, N. HCV genotype 1b chimeric replicon with NS5B of JFH-1 exhibited resistance to cyclosporine A. *Archives of Virology* **154,** 1671. ISSN: 0304-8608, 1432-8798 (2009).

400. Kelly, A. E., Ou, H. D., Withers, R. & Dötsch, V. Low-Conductivity Buffers for High-Sensitivity NMR Measurements. *Journal of the American Chemical Society* **124,** 12013–12019. ISSN: 0002-7863 (2002).

401. Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wüthrich, K. TROSY in triple-resonance experiments: New perspectives for sequential NMR assignment of large proteins. *Proceedings of the National Academy of Sciences of the United States of America* **95,** 13585–13590. ISSN: 0027-8424 (1998).

402. Schanda, P., Van Melckebeke, H. & Brutscher, B. Speeding Up Three-Dimensional Protein NMR Experiments to a Few Minutes. *Journal of the American Chemical Society* **128,** 9042–9043. ISSN: 0002-7863 (2006).

403. Hansen, D. F., Neudecker, P. & Kay, L. E. Determination of Isoleucine Side-Chain Conformations in Ground and Excited States of Proteins from Chemical Shifts. *Journal of the American Chemical Society* **132,** 7589–7591. ISSN: 0002-7863 (2010).

404. Korzhnev, D. M., Mittermaier, A. K. & Kay, L. E. Cross-correlated spin relaxation effects in methyl 1H CPMG-based relaxation dispersion experiments: complications and a simple solution. *Journal of biomolecular NMR* **31,** 337–342. ISSN: 0925-2738 (2005).

405. Tarragó, T., Claasen, B., Kichik, N., Rodriguez-Mias, R. A., Gairí, M. & Giralt, E. A Cost-Effective Labeling Strategy for the NMR Study of Large Proteins: Selective 15N-Labeling of the Tryptophan Side Chains of Prolyl Oligopeptidase. *ChemBioChem* **10,** 2736–2739. ISSN: 1439-7633 (2009).

406. Labonté, P., Axelrod, V., Agarwal, A., Aulabaugh, A., Amin, A. & Mak, P. Modulation of Hepatitis C Virus RNA-dependent RNA Polymerase Activity by Structure-based Site-directed Mutagenesis. *Journal of Biological Chemistry* **277,** 38838–38846. ISSN: 0021-9258, 1083-351X (2002).

407. Rigat, K. L., Lu, H., Wang, Y.-K., Argyrou, A., Fanslau, C., Beno, B., Wang, Y., Marcinkeviciene, J., Ding, M., Gentles, R. G., Gao, M., Abell, L. M. & Roberts, S. B. Mechanism of Inhibition for BMS-791325, a Novel Non-nucleoside Inhibitor of Hepatitis C Virus NS5B Polymerase. *Journal of Biological Chemistry,* jbc.M114.613653. ISSN: 0021-9258, 1083-351X (2014).

408. Boyce, S. E., Tirunagari, N., Niedziela-Majka, A., Perry, J., Wong, M., Kan, E., Lagpacan, L., Barauskas, O., Hung, M., Fenaux, M., Appleby, T., Watkins, W. J., Schmitz, U. & Sakowicz, R. Structural and Regulatory Elements of HCV NS5B Polymerase – β-Loop and C-Terminal Tail – Are Required for Activity of Allosteric Thumb Site II Inhibitors. *PLoS ONE* **9,** e84808 (2014).

409. Lam, A. M., Espiritu, C., Bansal, S., Steuer, H. M. M., Zennou, V., Otto, M. J. & Furman, P. A. Hepatitis C Virus Nucleotide Inhibitors PSI-352938 and PSI-353661 Exhibit a Novel Mechanism of Resistance Requiring Multiple Mutations within Replicon RNA. *Journal of Virology* **85,** 12334–12342. ISSN: 0022-538X, 1098-5514 (2011).

410. Bressanelli, S., Tomei, L., Rey, F. A. & Francesco, R. D. Structural Analysis of the Hepatitis C Virus RNA Polymerase in Complex with Ribonucleotides. *Journal of Virology* **76,** 3482–3492. ISSN: 0022-538X, 1098-5514 (2002).

411. Eastman, K. J. *et al.* The discovery of a pan-genotypic, primer grip inhibitor of HCV NS5B polymerase. *MedChemComm* **8,** 796–806. ISSN: 2040-2511 (2017).

412. Hanoulle, X., Verdegem, D., Badillo, A., Wieruszeski, J.-M., Penin, F. & Lippens, G. Domain 3 of non-structural protein 5A from hepatitis C virus is natively unfolded. *Biochemical and Biophysical Research Communications* **381,** 634–638. ISSN: 0006-291X (2009).

413. Sólyom, Z., Ma, P., Schwarten, M., Bosco, M., Polidori, A., Durand, G., Willbold, D. & Brutscher, B. The Disordered Region of the HCV Protein NS5A: Conformational Dynamics, SH3 Binding, and Phosphorylation. *Biophysical Journal* **109,** 1483–1496. ISSN: 0006-3495 (2015).

414. Iakoucheva, L. M., Kimzey, A. L., Masselon, C. D., Smith, R. D., Dunker, A. K. & Ackerman, E. J. Aberrant mobility phenomena of the DNA repair protein XPA. *Protein Science : A Publication of the Protein Society* **10,** 1353–1362. ISSN: 0961-8368 (2001).

415. Chung, H.-Y., Gu, M., Buehler, E., MacDonald, M. R. & Rice, C. M. Seed sequence-matched controls reveal limitations of small interfering RNA knockdown in functional and structural studies of hepatitis C virus NS5A-MOBKL1B interaction. *Journal of Virology* **88,** 11022–11033. ISSN: 1098-5514 (2014).

416. Karlsson, R., Roos, H., Fägerstam, L. & Persson, B. Kinetic and Concentration Analysis Using BIA Technology. *Methods* **6,** 99–110. ISSN: 1046-2023 (1994).

417. Fielding, L. NMR methods for the determination of protein–ligand dissociation constants. *Progress in Nuclear Magnetic Resonance Spectroscopy* **51,** 219–242. ISSN: 0079-6565 (2007).

418. Paleos, C. M. & Dais, P. Ready reduction of some nitroxide free radicals with ascorbic acid. *Journal of the Chemical Society, Chemical Communications* **24,** 345–346. ISSN: 0022-4936 (1977).

419. Iwahara, J., Tang, C. & Clore, G. M. Practical Aspects of 1H Transverse Paramagnetic Relaxation Enhancement Measurements on Macromolecules. *Journal of magnetic resonance (San Diego, Calif. : 1997)* **184,** 185–195. ISSN: 1090-7807 (2007).

420. Goyal, S., Gupta, G., Qin, H., Upadya, M. H., Tan, Y. J., Chow, V. T. K. & Song, J. VAPC, an Human Endogenous Inhibitor for Hepatitis C Virus (HCV) Infection, Is Intrinsically Unstructured but Forms a "Fuzzy Complex" with HCV NS5B. *PLOS ONE* **7,** e40341. ISSN: 1932-6203 (2012).

421. Rezaei-Ghaleh, N., Blackledge, M. & Zweckstetter, M. Intrinsically Disordered Proteins: From Sequence and Conformational Properties toward Drug Discovery. *ChemBioChem* **13,** 930–950. ISSN: 1439-7633 (2012).

422. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols* **2,** 2212–2221. ISSN: 1754-2189 (2007).

423. Scott, D. E., Spry, C. & Abell, C. *Fragment-based Drug Discovery Lessons and Outlook* (eds Erlanson, D. A. & Jahnke, W.) DOI: 10.1002/9783527683604.ch07, 139–172 (Wiley-VCH Verlag GmbH & Co. KGaA, 2016). ISBN: 978-3-527-68360-4.

424. Qin, W., Yamashita, T., Shirota, Y., Lin, Y., Wei, W. & Murakami, S. Mutational analysis of the structure and functions of hepatitis C virus RNA–dependent RNA polymerase. *Hepatology* **33,** 728–737. ISSN: 1527-3350 (2001).

425. Manns, M. P. & von Hahn, T. Novel therapies for hepatitis C — one pill fits all? *Nature Reviews Drug Discovery* **12,** 595–610. ISSN: 1474-1776 (2013).

426. Gentile, I., Buonomo, A. R., Zappulo, E. & Borgia, G. Discontinued drugs in 2012 – 2013: hepatitis C virus infection. *Expert Opinion on Investigational Drugs* **24,** 239–251. ISSN: 1354-3784 (2015).

427. Landrieu, I., Hanoulle, X., Fritzinger, B., Horvath, D., Wieruszeski, J.-M. & Lippens, G. Ranking High Affinity Ligands of Low Solubility by NMR Spectroscopy. *ACS Medicinal Chemistry Letters* **2,** 485–487. ISSN: 1948-5875 (2011).

428. Davis, B. C. & Thorpe, I. F. Thumb inhibitor binding eliminates functionally important dynamics in the hepatitis C virus RNA polymerase. *Proteins: Structure, Function, and Bioinformatics* **81,** 40–52. ISSN: 1097-0134 (2013).

429. Jin, Z., Leveque, V., Ma, H., Johnson, K. A. & Klumpp, K. Assembly, Purification, and Pre-steady-state Kinetic Analysis of Active RNA-dependent RNA Polymerase Elongation Complex. *The Journal of Biological Chemistry* **287,** 10674–10683. ISSN: 0021-9258 (2012).

430. Cramer, J., Jaeger, J. & Restle, T. Biochemical and Pre-Steady-State Kinetic Characterization of the Hepatitis C Virus RNA Polymerase (NS5BΔ21, HC-J4)†. *Biochemistry* **45,** 3610–3619. ISSN: 0006-2960 (2006).

431. Golovanov, A. P., Hautbergue, G. M., Wilson, S. A. & Lian, L.-Y. A Simple Method for Improving Protein Solubility and Long-Term Stability. *Journal of the American Chemical Society* **126,** 8933–8939. ISSN: 0002-7863 (2004).

432. Fekete, S., Guillarme, D., Sandra, P. & Sandra, K. Chromatographic, Electrophoretic, and Mass Spectrometric Methods for the Analytical Characterization of Protein Biopharmaceuticals. *Analytical Chemistry* **88,** 480–507. ISSN: 0003-2700 (2016).

433. Foster, T. L., Belyaeva, T., Stonehouse, N. J., Pearson, A. R. & Harris, M. All Three Domains of the Hepatitis C Virus Nonstructural NS5A Protein Contribute to RNA Binding. *Journal of Virology* **84,** 9267–9277. ISSN: 0022-538X, 1098-5514 (2010).

434. Foster, T. L., Gallay, P., Stonehouse, N. J. & Harris, M. Cyclophilin A Interacts with Domain II of Hepatitis C Virus NS5A and Stimulates RNA Binding in an Isomerase-Dependent Manner. *Journal of Virology* **85,** 7460–7464. ISSN: 0022-538X, 1098-5514 (2011).

435. Winquist, J., Abdurakhmanov, E., Baraznenok, V., Henderson, I., Vrang, L. & Danielson, U. H. Resolution of the interaction mechanisms and characteristics of non-nucleoside inhibitors of hepatitis C virus polymerase. *Antiviral Research* **97,** 356–368. ISSN: 0166-3542 (2013).

436. Heim, M. H. & Thimme, R. Innate and adaptive immune responses in HCV infections. *Journal of Hepatology* **61,** S14–S25. ISSN: 0168-8278 (2014).

437. Roberts, D. J. & Miyamoto, S. Hexokinase II integrates energy metabolism and cellular protection: Akting on mitochondria and TORCing to autophagy. *Cell Death and Differentiation* **22,** 248–257. ISSN: 1350-9047 (2015).

438. Dujardin, M. *Etude des relations structurales et fonctionnelles entre le domaine 2 de la protéine NS5A du virus de l'hépatite C et la Cyclophiline A humaine* PhD thesis (Université Lille 1 - Sciences et Technologies, Lille, 2015).

439. Reyes, A. & Cárdenas, M. L. All hexokinase isoenzymes coexist in rat hepatocytes. *Biochemical Journal* **221,** 303–309. ISSN: 0264-6021 (1984).

440. Whittington, A. C., Larion, M., Bowler, J. M., Ramsey, K. M., Brüschweiler, R. & Miller, B. G. Dual allosteric activation mechanisms in monomeric human glucokinase. *Proceedings of the National Academy of Sciences* **112,** 11553–11558. ISSN: 0027-8424, 1091-6490 (2015).

441. Kamata, K., Mitsuya, M., Nishimura, T., Eiki, J.-i. & Nagata, Y. Structural Basis for Allosteric Regulation of the Monomeric Allosteric Enzyme Human Glucokinase. *Structure* **12,** 429–438. ISSN: 0969-2126 (2004).

442. Petit, P., Antoine, M., Ferry, G., Boutin, J. A., Lagarde, A., Gluais, L., Vincentelli, R. & Vuillard, L. The active conformation of human glucokinase is not altered by allosteric activators. *Acta Crystallographica Section D: Biological Crystallography* **67,** 929–935. ISSN: 0907-4449 (2011).

443. Larion, M., Salinas, R. K., Bruschweiler-Li, L., Miller, B. G. & Brüschweiler, R. Order–Disorder Transitions Govern Kinetic Cooperativity and Allostery of Monomeric Human Glucokinase. *PLoS Biol* **10,** e1001452 (2012).

444. Larion, M., Salinas, R. K., Bruschweiler-Li, L., Brüschweiler, R. & Miller, B. G. Direct Evidence for Conformational Heterogeneity in Human Pancreatic Glucokinase from High-Resolution NMR. *Biochemistry* **49,** 7969–7971. ISSN: 0006-2960 (2010).

445. Larion, M., Hansen, A. L., Zhang, F., Bruschweiler-Li, L., Tugarinov, V., Miller, B. G. & Brüschweiler, R. Kinetic Cooperativity in Human Pancreatic Glucokinase Originates from Millisecond Dynamics of the Small Domain. *Angewandte Chemie* **127,** 8247–8250. ISSN: 1521-3757 (2015).

446. Paczal, A., Bálint, B., Wéber, C., Szabó, Z. B., Ondi, L., Theret, I., De Ceuninck, F., Bernard, C., Ktorza, A., Perron-Sierra, F. & Kotschy, A. Structure–Activity Relationship of Azaindole-Based Glucokinase Activators. *Journal of Medicinal Chemistry* **59,** 687–706. ISSN: 0022-2623 (2016).

447. Larion, M. & Miller, B. G. 23-Residue C-terminal alpha-helix governs kinetic cooperativity in monomeric human glucokinase. *Biochemistry* **48,** 6157–6165. ISSN: 1520-4995 (2009).

448. Iynedjian, P. B. Molecular Physiology of Mammalian Glucokinase. *Cellular and Molecular Life Sciences* **66,** 27–42. ISSN: 1420-682X (2009).

449. Ullah, R., Shah, M. A., Tufail, S., Ismat, F., Imran, M., Iqbal, M., Mirza, O. & Rhaman, M. Activity of the Human Rhinovirus 3C Protease Studied in Various Buffers, Additives and Detergents Solutions for Recombinant Protein Production. *PLOS ONE* **11,** e0153436. ISSN: 1932-6203 (2016).

450. Tugarinov, V., Sprangers, R. & Kay, L. E. Probing Side-Chain Dynamics in the Proteasome by Relaxation Violated Coherence Transfer NMR Spectroscopy. *Journal of the American Chemical Society* **129,** 1743–1750. ISSN: 0002-7863 (2007).

451. Bonadonna, R. C., Heise, T., Arbet-Engels, C., Kapitza, C., Avogaro, A., Grimsby, J., Zhi, J., Grippo, J. F. & Balena, R. Piragliatin (RO4389620), a Novel Glucokinase Activator, Lowers Plasma Glucose Both in the Postabsorptive State and after a Glucose Challenge in Patients with Type 2 Diabetes Mellitus: A Mechanistic Study. *The Journal of Clinical Endocrinology & Metabolism* **95,** 5028–5036. ISSN: 0021-972X (2010).

452. Meininger, G. E., Scott, R., Alba, M., Shentu, Y., Luo, E., Amin, H., Davies, M. J., Kaufman, K. D. & Goldstein, B. J. Effects of MK-0941, a Novel Glucokinase Activator, on Glycemic Control in Insulin-Treated Patients With Type 2 Diabetes. *Diabetes Care* **34,** 2560–2566. ISSN: 0149-5992 (2011).

453. Matschinsky, F. M., Zelent, B., Doliba, N., Li, C., Vanderkooi, J. M., Naji, A., Sarabu, R. & Grimsby, J. Glucokinase Activators for Diabetes Therapy. *Diabetes Care* **34,** S236–S243. ISSN: 0149-5992 (2011).

454. Nakamura, A. & Terauchi, Y. Present status of clinical deployment of glucokinase activators. *Journal of Diabetes Investigation* **6,** 124–132. ISSN: 2040-1116 (2015).