

**Structural studies of 5' UTR of Hepatitis C viral RNA by  
NMR-based structural biology**

Von der Naturwissenschaftlichen Fakultät  
der Gottfried Wilhelm Leibniz Universität Hannover

zur Erlangung des Grades

Doktorin der Naturwissenschaften  
Dr. rer. nat.

genehmigte Dissertation von

specialist: vrac biofizik (Diplom), RNIMU  
Olga Alexandrovna Vitsyna

2023

Referent: Prof. Dr. rer. nat. Andreas Kischning

Korreferent: Prof. Dr. rer. nat. Wulf Blankenfeldt

Tag der Promotion: 18.09.2023

## Summary

RNA is a single-stranded biopolymer that plays a myriad of roles in physiological and pathological processes and is the carrier of genetic information in many human pathogens. Hepatitis C virus (HCV) is one of the most impactful representatives of RNA viruses. Liver-abundant human microRNA-122 (miR-122) binds to two tandem sites within domain I of the 5' untranslated region (5' UTR) of HCV, ultimately resulting in upregulation of viral propagation. Despite many studies of the interaction between HCV and miR-122, the exact mechanism by which this recognition event leads to increased viral propagation is unknown. In this thesis, I have studied the 5' UTR HCV–miR-122 interaction at different levels of structural complexity (domain I, domains I-II and the full 5' UTR) using an integrative NMR-based structural biology approach.

First, I have performed the near-complete assignment of domain I resonances and determined its secondary structure. Isolated domain I binds two copies of miR-122 with different affinities, and the binding kinetics fall into the slow-to-intermediate exchange-regime on the NMR chemical-shift timescale. Magnesium ions promote structural rearrangement of domain I, which in turn changes its interaction pattern with miR-122.

Next, I have determined the secondary structures of the isolated domain II and a domain I-II construct, both in their apo (without miR-122) and holo (bound to miR-122) states. The data demonstrates that, in the domain I-II construct, domains I and II maintain independent folds; furthermore, the secondary structure of domain II remains intact upon domain I binding two copies of miR-122. However, the binding of miR-122 to the domain I-II construct does lead to a structural rearrangement that changes the relative orientation of the two domains, resulting in more open and extended conformation. Finally, I have investigated the interaction of miR-122 with the full 5' UTR. Since the differences between the low-resolution scattering data of the 5' UTR in the apo and holo states were minimal, no major structural changes in the 5' UTR upon miR-122 binding appear to occur. To study the local structural details of the 5' UTR, I have explored the use of solid-state NMR. While there were clear changes in chemical shifts of the 5' UTR upon miR-122 binding, indicating conformational changes in the 5' UTR, acquisition of solid-state NMR data on segmentally labeled samples and isolated domain I was challenging and could not provide definitive answers at this stage. Overall, using an NMR-based integrative structural biology approach, I could show that miR-122 binding to domain I causes both widespread local rearrangements within domain I and a significant reorientation of domain I relative to domain II, while the effect of miR-122 binding on the overall structure of the full 5' UTR was found to be minimal.

Keywords: RNA, HCV, miR-122, 5' UTR, IRES, NMR



## Zusammenfassung

RNA ist ein einzelsträngiges Biopolymer, das eine Vielzahl von Funktionen in physiologischen und pathologischen Prozessen erfüllt und bei vielen menschlichen Krankheitserregern als Träger der genetischen Information fungiert. Das Hepatitis-C-Virus (HCV) ist ein von den einflussreichsten Vertretern der RNA-Viren. Die in der Leber reichlich vorhandene menschliche microRNA-122 (miR-122) bindet an zwei Tandemstellen in der Domäne I der 5' untranslatierten Region (5' UTR) des HCV und letztlich zur Hochregulierung der Virusausbreitung führt. Trotz zahlreicher Studien zur Interaktion zwischen HCV und miR-122 ist der genaue Mechanismus, der für die erhöhte virale Ausbreitung durch dieses Erkennungsevent verantwortlich ist, noch nicht bekannt.

In dieser Arbeit habe ich die 5' UTR HCV–miR-122 Interaktion auf den verschiedenen Ebenen der strukturellen Komplexität (Domäne I, Domänen I-II und vollständige 5' UTR) mithilfe des integrativen NMR-basierten strukturbioologischen Ansatzes untersucht.

Zunächst habe ich eine nahezu vollständige Zuordnung der Resonanzen der Domäne I vorgenommen und ihre Sekundärstruktur bestimmt. Die isolierte Domäne I bindet zwei Kopien von miR-122 mit unterschiedlicher Affinität, wobei die Bindungskinetik in einem langsamen bis intermediären Austauschregime an der Zeitskala von NMR-chemischen Verschiebungen erfolgt. Magnesiumionen fördern die strukturelle Umordnung der Domäne I, was wiederum ihr Interaktionsmuster mit miR-122 verändert.

Anschließend habe ich mit Hilfe von NMR im Lösungszustand die Sekundärstruktur der isolierten Domäne II und Domäne I-II Konstrukt, im Apo-Zustand (ohne miR-122) als auch in Holo-Zustand (gebunden an miR-122) bestimmt. Die Daten zeigen, dass im Domäne I-II-Konstrukt die Domäne I und die Domäne II unabhängige Faltungen beibehalten; außerdem bleibt die Sekundärstruktur der Domäne II bei der Bindung von zwei Kopien von miR-122 an die Domäne I intakt. Die Bindung von miR-122 an das Domäne I-II-Konstrukt führt jedoch zu struktureller Umgestaltung, die die relative Orientierung von beiden Domänen ändert, was zu einer offeneren und ausgedehnteren Konformation führt.

Zuletzt habe ich die Wechselwirkungen von miR-122 mit der gesamten 5' UTR untersucht. Da die Unterschiede im niedrig auflösenden Streudata der 5' UTR im apo- und holo-Zustand minimal waren, wurden keine größeren strukturellen Veränderungen in der 5' UTR nach der miR-122 Bindung festgestellt. Um lokale Strukturdetails der 5'-UTR zu untersuchen, habe ich den Einsatz von Festkörper-NMR erforscht. Während es bei der Interaktion mit miR-122 deutliche Veränderungen der chemischen Verschiebungen der 5'-UTR gab, was auf Konformationsänderungen in der 5'-UTR hindeutet, war die Datenerfassung für die segmental markierte isolierte Domäne I eine Herausforderung und konnte zu diesem Zeitpunkt keine endgültige Antwort liefern.

Insgesamt konnte ich mit Hilfe des NMR-basierten integrativen strukturbioologischen Ansatzes zeigen, dass die Bindung von miR-122 an Domäne I starke lokale Umstrukturierungen in Domäne I, sowie eine auffällige relative Neuausrichtung von Domäne I relativ zu Domäne II verursacht, während die Wirkung der miR-122-Bindung auf die Gesamtstruktur von 5'-UTR minimal war.

Schlüsselwörter: RNA, HCV, miR-122, 5' UTR, IRES, NMR



## Abbreviations and symbols

A: adenosine

ATP: adenosine triphosphate

bp: base pairs

C: cytidine

CTP: cytidine triphosphate

CSP: chemical shift perturbation

DTT: dithiothreitol

D max : maximum dimension of the particle

D \*: domain \* of 5' untranslated region of hepatitis C

DNA: deoxyribonucleic acid

EDTA: ethylenediaminetetraacetic acid

eIF\*: eukaryotic initiation factor \*

EM: electron microscopy

EMSA: electrophoretic mobility shift assay

G: guanosine

GTP: guanosine-5'-triphosphate

HEPES: (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HH: hammerhead (ribozyme)

HMQC: heteronuclear multiple-quantum coherence

HSQC: heteronuclear single-quantum coherence

IRES: internal ribosomal entry site

Kd: dissociation constant

MALS: multi angle light scattering

MD: molecular dynamics

min: minute (s)

miR: microRNA

Mw: molecular weight

NMR: nuclear magnetic resonance

NTP: nucleotide triphosphate

nt: nucleotide

NOESY: nuclear Overhauser effect spectroscopy

o/n: over night

PCR: polymerase chain reaction  
PEG: polyethylene glycol  
ppm: parts per million  
Rg: radius of gyration  
RNA: ribonucleic acid  
rpm: revolutions per minute  
Rz: ribozyme  
ssNMR: solid-state nuclear NMR  
SANS: small-angle neutron scattering  
SAXS: small-angle X-ray scattering  
SAS: small-angle scattering  
SEC: size exclusion chromatography  
SLD: scattering length density  
TROSY: transverse relaxation optimized spectroscopy  
U: uridine  
UTP: uridine-5'-triphosphate  
UTR: untranslated region



# Table of Contents

<b>Summary</b>	<b>3</b>
<b>Zusammenfassung</b>	<b>5</b>
<b>Abbreviations and symbols</b>	<b>7</b>
<b>Table of Contents</b>	<b>9</b>
<b>1. Introduction</b>	<b>13</b>
<i>1.1. HCV genome organization</i>	<i>13</i>
<i>1.2. Function and organization of HCV 5' and 3' untranslated regions</i>	<i>15</i>
1.2.1. 5' UTR	15
1.2.2. 3' UTR	15
<i>1.3. HCV lifecycle</i>	<i>16</i>
<i>1.4. 5' UTR architecture and translation initiation of HCV</i>	<i>17</i>
1.4.1. Internal ribosomal entry site mediated initiation of HCV translation	17
1.4.2. Structural organization of the 5' UTR	18
<i>1.5. The role of host miR-122 in HCV lifecycle</i>	<i>21</i>
<i>1.6. miR-122 – 5' UTR interactions</i>	<i>21</i>
<i>1.7. Stimulating effect of miR-122 on the HCV propagation and proposed mechanisms of miR-122 action</i>	<i>22</i>
<i>1.8. HCV therapy</i>	<i>25</i>
1.8.1. Direct-acting antiviral treatment	25
1.8.2. miR-122 as a candidate target for HCV treatment	26
<b>2. Methodological introduction</b>	<b>29</b>
<i>2.1. Size-exclusion chromatography and Multi-angle light scattering</i>	<i>29</i>
2.1.1. Size-exclusion chromatography	29
2.1.2. Multi-angle light scattering	30
<i>2.2. Small-angle scattering</i>	<i>32</i>
<i>2.3. Nuclear magnetic resonance (NMR)</i>	<i>39</i>

2.3.1. Quantum spin Hamiltonian theory	43
2.3.2. Internal spin interactions	44
2.3.3. Multi-dimensional NMR	48
2.3.4. Chemical shift perturbation analysis	52
2.4. <i>Solid-state NMR</i>	55
2.4.1. Cross-polarization experiment	56
2.4.2. Double cross-polarization	56
2.4.3. Recoupling	57
2.4.4. Sample preparation for ssNMR	59
2.5. <i>NMR of RNA</i>	59
2.5.1. Information about base pairing	61
<b>3. Aims</b>	<b>63</b>
<b>4. Material and Methods</b>	<b>65</b>
4.1. <i>Cloning</i>	65
4.1.1. Plasmid linearization	65
4.1.2. Insert annealing	65
4.1.3. Ligation of the insert into the plasmid	66
4.1.4. Heat-shock transformation	66
4.2. <i>Mutagenesis</i>	66
4.2.1. 5' Hammerhead ribozyme insertion in cis	67
4.2.2. miRcopy plasmid construction	67
4.3. <i>DNA template production and purification</i>	68
4.3.1. DNA template production in <i>E.coli</i> cells	68
4.3.2. DNA template production by PCR	69
4.4. <i>RNA transcription and purification</i>	70
4.4.1. Domains I-II.	72
4.4.2. miR-122	72
4.5. <i>Site-specific RNaseH cleavage guided by 2'-O-methyl-RNA/DNA chimera</i>	73
4.6. <i>Ligation of RNAs</i>	74
4.7. <i>Size-exclusion Chromatography and Multi-angle Light Scattering</i>	74

4.8. Electrophoretic mobility shift assays (EMSA)	74
4.9. NMR	75
4.9.1. Solution-state NMR	75
4.9.2. Solid-state NMR	77
4.10. SANS	78
4.11. SAXS	79
4.12. RNA Modeling	79
<b>5. Results and Discussion</b>	<b>83</b>
5.1. Domain I	83
5.1.1. Domain I interacts with miR-122 and forms ternary complex	83
5.1.2. Domain I binds two copies of miR-122 and requires magnesium ions	85
5.1.3. Resonance assignment of domain I	88
5.1.4. Elucidation of base pairing and secondary structure	90
5.1.5. Features and dynamics of the domain I–miR-122 interaction	91
5.1.6. Solid-state NMR of domain I	101
5.2. Domains I-II	103
5.2.1. Sample preparation	103
5.2.2. NMR studies of domain II folding in the presence of domain I and upon miR-122 titration	106
5.2.3. Characterization of conformational changes in domains I-II upon miR-122 binding by small angle neutron scattering	112
5.2.4. RNA modeling	114
5.3. 5' UTR	119
5.3.1. HCV 5' UTR binds two copies of miR-122 without major structural changes	119
5.3.2. Stoichiometry of the 5' UTR–miR-122 interaction	121
5.3.3. miR-122 binding to 5' UTR does not introduce significant conformational changes in 5' UTR	121
<b>6. Conclusions and Outlook</b>	<b>129</b>
<b>7. Bibliography</b>	<b>133</b>

<b>8. Extended data</b>	<b>155</b>
<b>9. Acknowledgements</b>	<b>169</b>
<b>10. Curriculum Vitae</b>	<b>171</b>

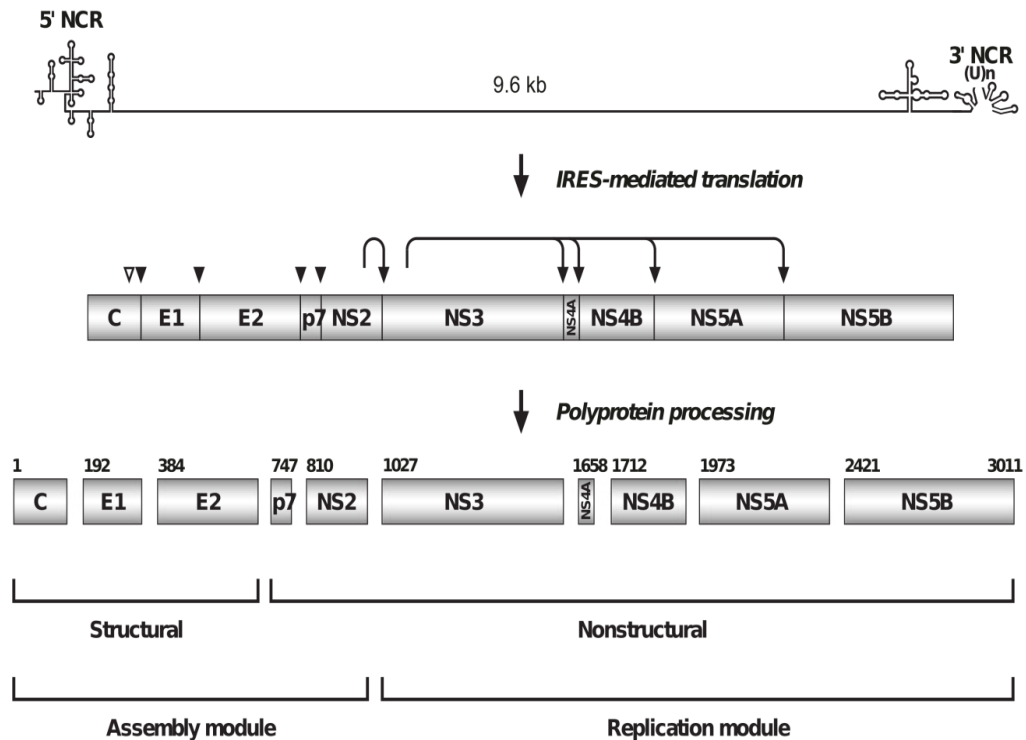
# 1. Introduction

Hepatitis C virus (HCV) is a global health threat with over 70 million chronically infected people worldwide. 10 to 20% of infected individuals develop liver-related sequelae such as decompensated cirrhosis, liver failure and hepatocellular carcinoma (Spearman et al. 2019). In addition, chronic HCV infection in many cases leads to extrahepatic manifestation that harms metabolic, neurological and cardiovascular systems via autoimmune and inflammatory mediated mechanisms (Negro et al. 2015). In the light of public health burden related to viral hepatitis, in 2016 World Health Organization proposed a strategy, which is aimed at the elimination of viral hepatitis as a major health treat by 2030 with the focus on the Hepatitis C and B. Among the others, the main strategic directions are the development of new efficient therapeutics and their costs reduction (WHO 2016).

## 1.1. HCV genome organization

HCV belongs to the family Flaviviridae. Its genome is uncapped single-stranded positive-sense RNA ~9.6 kb in length that consists of an open reading frame (ORF), which is flanked by highly structured 5' and 3' untranslated regions (UTR). ORF encodes for a long polyprotein, which is translated at the rough endoplasmatic reticulum and is subsequently cleaved by viral and cellular proteases into ten viral proteins (**Fig. 1.1.1**). Among them there are three structural proteins (core, envelope glycoproteins E1 and E2), which form the virion, and seven non-structural proteins (p7, NS2, NS3, NS4A, NS4B, NS5A and NS5B (Moradpour and Penin 2013). P7 viroporin is a 63 amino acids membrane protein, which has cation channel activity, which is dispensable in RNA replication but is essential for the release and assembly of the virus both *in vivo* and *in vitro* (Steinmann and Pietschmann 2013). NS2 is a dimeric cysteine protease, which in tandem with NS3 protease forms a complex and accomplishes the self-cleavage at the NS2/NS3 junction and liberates fully functional NS3 (Jirasko et al. 2008). NS3, in turn, is a multifunctional enzyme with N-terminal protease activity and C-terminal helicase and NTPase activities (Raney et al. 2010). NS4A is a cofactor, which non-covalently binds to NS3 protein, and forms NS3-NS4 complex. The last enhances the protease activity and anchors the complex to the endoplasmic reticulum membrane (Morikawa et al. 2011). NS4B is a poorly characterized membrane protein that is an essential component of the membranous web. It has NTPase activity and participates in virus assembly (Jones et al. 2009; Thompson et al. 2009). NS5A is a

multifunctional protein that is involved in viral assemble and replication. It facilitates viral propagation by the modulation the host cell interferon response (He, Staschke, and Tan 2006). Being an essential component of the replication complex, it is a main target of the direct-acting antiviral therapy (Gitto, Gamal, and Andreone 2017). NS5A was shown to modulate the activity of NS5B protein, an RNA-dependent RNA polymerase (RdRp), which catalyzes the polymerization of rNTPs during the replication (Jin et al. 2012). Similar to NS5A, NS5B is an important objective for DAAs (Geddawy et al. 2017).



**Figure 1.1.1. HCV genome organization and polyprotein processing.** Schematic representation of the positive single-stranded 9.6 kb RNA genome of HCV, consisting of open reading frame flanked by 5' and 3' non-coding regions (NCR) is shown on the top. IRES-mediated translation results in polyprotein which is co- and post-translationally processed into 10 structural and non-structural viral proteins. The figure is adapted from (Moradpour and Penin 2013).

According to International Committee on Taxonomy of Viruses (ICTV), as of May 2019, there are 90 confirmed HCV subtypes classified into 8 genotypes (Donald B. Smith, Jens Bukh, Carla Kuiken, A. Scott Muerhoff, Charles M. Rice n.d.). Such extraordinary genetic diversity is determined by high error rate caused by the nature of replication machinery (Powdrill et al. 2011) and should be considered to achieve efficient therapy with optimal direct-acting antiviral (DAA) regimen (Welzel et al. 2017).

## 1.2. Function and organization of HCV 5' and 3' untranslated regions

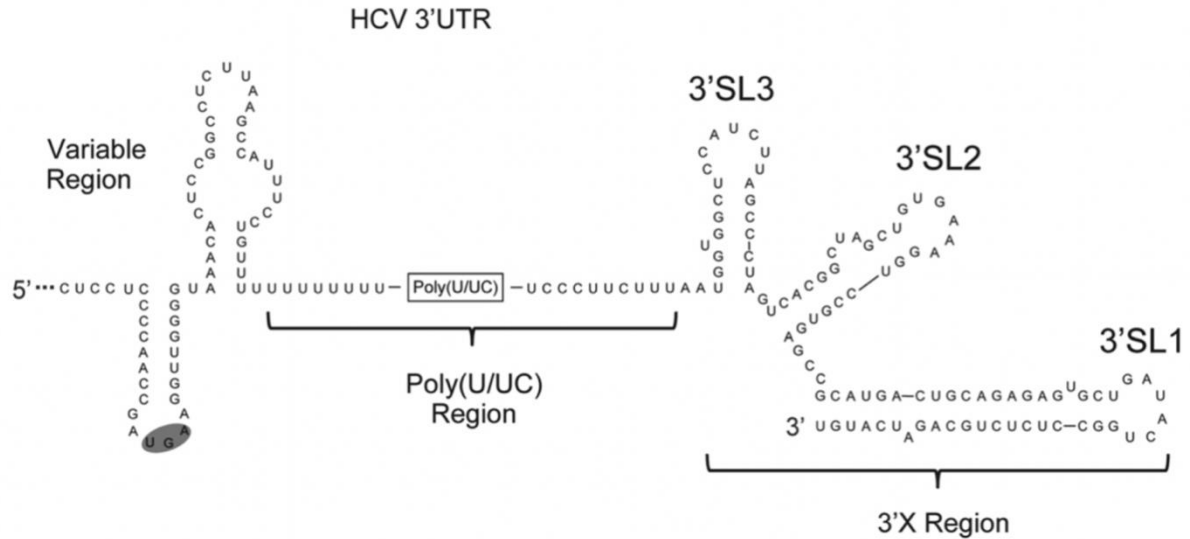
Despite large sequence variability between different genotypes, both 5' and 3' UTR of HCV enclose conserved structural elements that are crucial for translation and replication of the virus. In most of the HCV subtypes 5' UTR spans 341 terminal nucleotides and shares sequential and structural similarities not only within various HCV genotypes, but also between classical swine fever virus (CSFV) (Pestova et al. 1998), flavivirus GB virus B and pestiviruses (Honda et al. 1999).

### 1.2.1. 5' UTR

5' UTR carries internal ribosomal entry site (IRES), the structural element responsible for the cap-independent translation initiation (Honda et al. 1999; Rijnbrand et al. 1995; Tsukiyama-Kohara et al. 1992) and upstream stem-loop I that bears triphosphate (Li et al. 2013). Domain I of 5' UTR was shown to be not important for the translation of the virus (Reynolds et al. 1995), but, in contrast, the first 5' terminal 115 nucleotides, which include domains I-II are essential for the replication of the virus as a cis-acting elements (Friebe et al. 2001).

### 1.2.2. 3' UTR

The 3' UTR has a length of 200-235 nucleotides and is organized into three distinct regions from 5' to 3' direction: a hypervariable region, a poly U/UC stretch, which varies in length, and highly conserved 98-nt X-tail (**Fig. 1.2.1**) (Friebe and Bartenschlager 2002; Kolykhalov, Feinstone, and Rice 1996). HCV replication initiates at the 3' UTR and results in the synthesis of the negative strand RNA, which in turn serves as a template for positive strands synthesis of viral genome with replication initiation at the 3' UTR as well (Lohmann 1999). While the 5' UTR is known to stimulate replication, it was revealed that the 3' UTR binds directly to the 40S ribosomal subunit and eIF-3 transcription factor and stimulates translation of HCV *in cis*, most probably by keeping translation machinery on the RNA strand, facilitating its turnover (Bai, Zhou, and Doudna 2013).

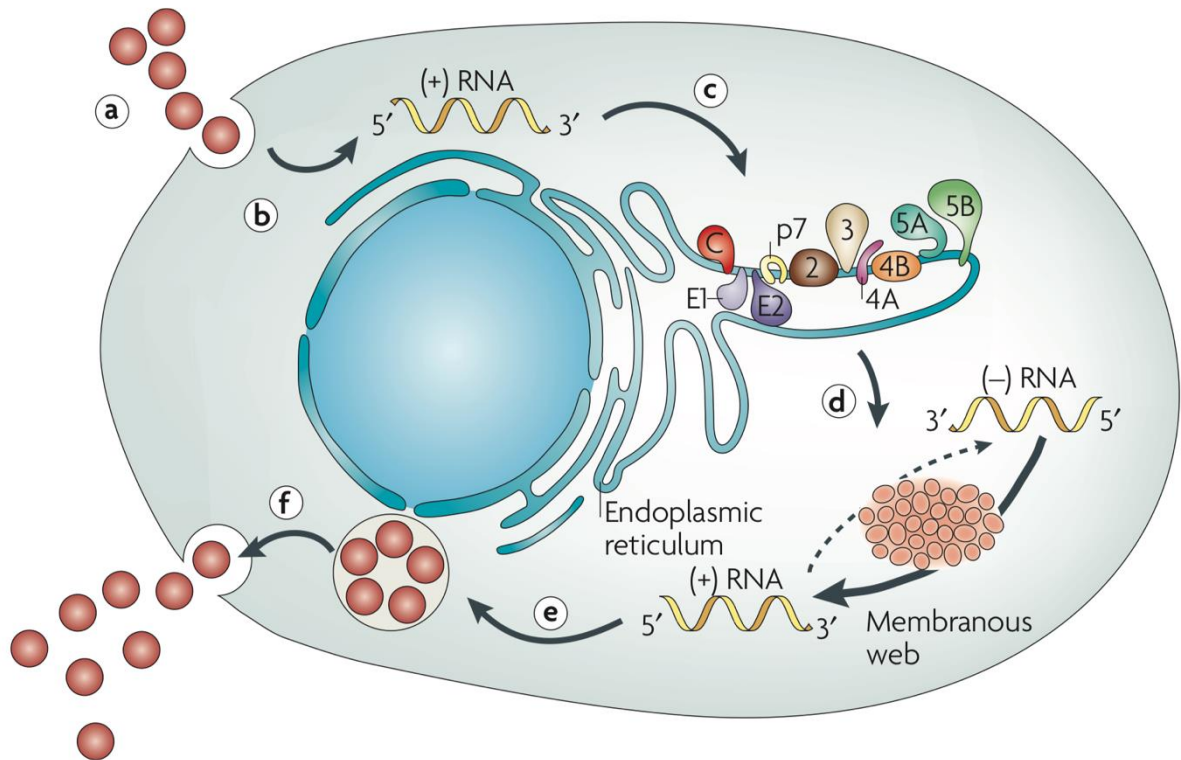


**Figure 1.2.1. Secondary structure of the positive-strand HCV 3' UTR.** The figure is adapted from (Bai et al. 2013)

### 1.3. HCV lifecycle

The HCV lifecycle is illustrated in **Fig. 1.3.1**. After the entry of hepatitis C virus to the cell, the released positive RNA strand interacts with endoplasmic reticulum (ER), where the polyprotein is translated by the ribosomes. The translated and cleaved by host and viral proteases non-structural proteins (NS3, NS4A, NS4B, NS5A, and NS5B) assemble in replicase complex on the ER membranes and form a membranous web (Chatel-Chaix and Bartenschlager 2014), where positive RNA strand is replicated to the negative, which in turn serves as a template for multiple copies of positive viral RNA. These positive strands are packed to the envelope that is built from structural proteins (core, envelope 1, E1 and E2) and then released from the cell or stay there for further translation and replication cycles (Li et al. 2015).





**Figure 1.3.1. HCV lifecycle.** a) The virus enters the cell and uncoats. b) Release of the positive RNA strand genome to the cytoplasm. c) IRES-mediated translation and polyprotein processing. d) Replication of RNA. e) Virus packaging and assembly. f) Virus exit. The figure is adapted from (Moradpour, Penin, and Rice 2007)

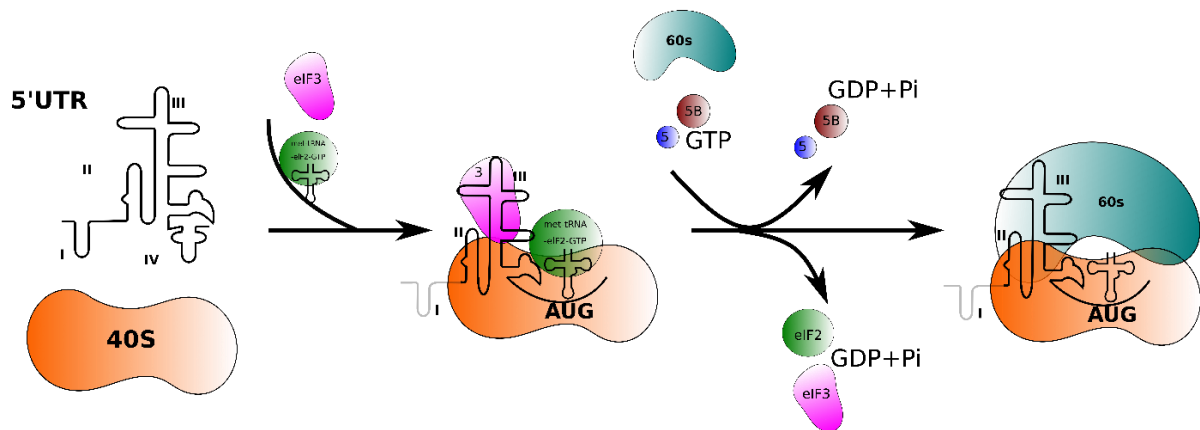
## 1.4. 5' UTR architecture and translation initiation of HCV

In this section detailed overview of the HCV 5' UTR is given, as this is the main subject of the study.

### 1.4.1. Internal ribosomal entry site mediated initiation of HCV translation

The 5' UTR bears the structural element responsible for 5' cap-independent translation initiation, so-called internal ribosomal entry site (IRES) (Brown et al. 1992; Fukushi et al. 1994). In contrast to the ordinary eukaryotic translation initiation, IRES-mediated initiation requires just a limited number of cellular eukaryotic initiation factor (eIF) and is driven by high affinity of IRES structure to the 40S ribosomal subunit, which results in the direct position of the initiator codon at the ribosomal peptidyl (P)-site without ribosomal scanning (Pestova et al. 1998). The initiation of translation commences after the 40S ribosomal subunit binds directly to the IRES with high affinity ( $K_d = 1.9$  nM) (Kieft et al. 2001). Domain II of IRES was proved to mediate conformational changes in 40S subunit (Spahn et al. 2001), while flexible domain III provides tight binding to the platform side of the 40S. This, in turn,

recruits eIF2–tRNA<sup>iMet</sup>-GTP ternary complex and eIF3, which interacts with domains III and IV of IRES. eIF3 was shown to be not essential for the correct placement of 40S subunit, but its role is to enhance the 48S complex formation (Ji et al. 2004). Subsequently recruited eIF5 induces the activation of GTP hydrolysis by eIF2 and releases eIF2/GDP with the help of domain II of IRES. eIF3 release is mediated by eIF5B and GTP hydrolysis, which lets the 60S ribosomal subunit to associate and form translation active 80S ribosome (Locker, Easton, and Lukavsky 2007). A schematic representation of IRES-mediated translation initiation is illustrated in **Fig. 1.4.1**.



**Figure 1.4.1. The scheme of HCV IRES translation initiation.** In the first step, HCV IRES directly binds to 40S subunit, then eIF-3 and eIF2–tRNA<sup>iMet</sup>-GTP ternary complex form 48S complex. Subsequently, 60S subunit assembles and forms functional 80S ribosome. This happens with the help of energy release from the GTP hydrolysis with the participation of eIF5 and eIF-5B and eIF-2, eIF-3 dissociation.

#### 1.4.2. Structural organization of the 5' UTR

The structural and functional organization of IRES elements was studied in great details. It comprises its tertiary fold in the presence of magnesium ions and consists of two mayor domains II and III, and a small stem-loop domain IV, which bears the initiator AUG start codon (nucleotides 342-344) (reviewed in (Lukavsky 2009)). IRES structural elements are conserved not only among different HCV strains but also between classical swine fever virus (CSFV) (Pestova et al. 1998), flavivirus GB virus B and pestiviruses (Honda et al. 1999).

While high natural flexibility of RNA complicates its study by X-ray crystallography and cryo-EM, its low chemical diversity of nucleotides and intrinsic size limitation of solution-state NMR hamper large RNA structure determination by this technique. Several atomic resolution structures of small IRES subdomains have been obtained by NMR and X-ray (reviewed in (Lukavsky 2009)). The first cryo-EM structure of IRES has been resolved at 20

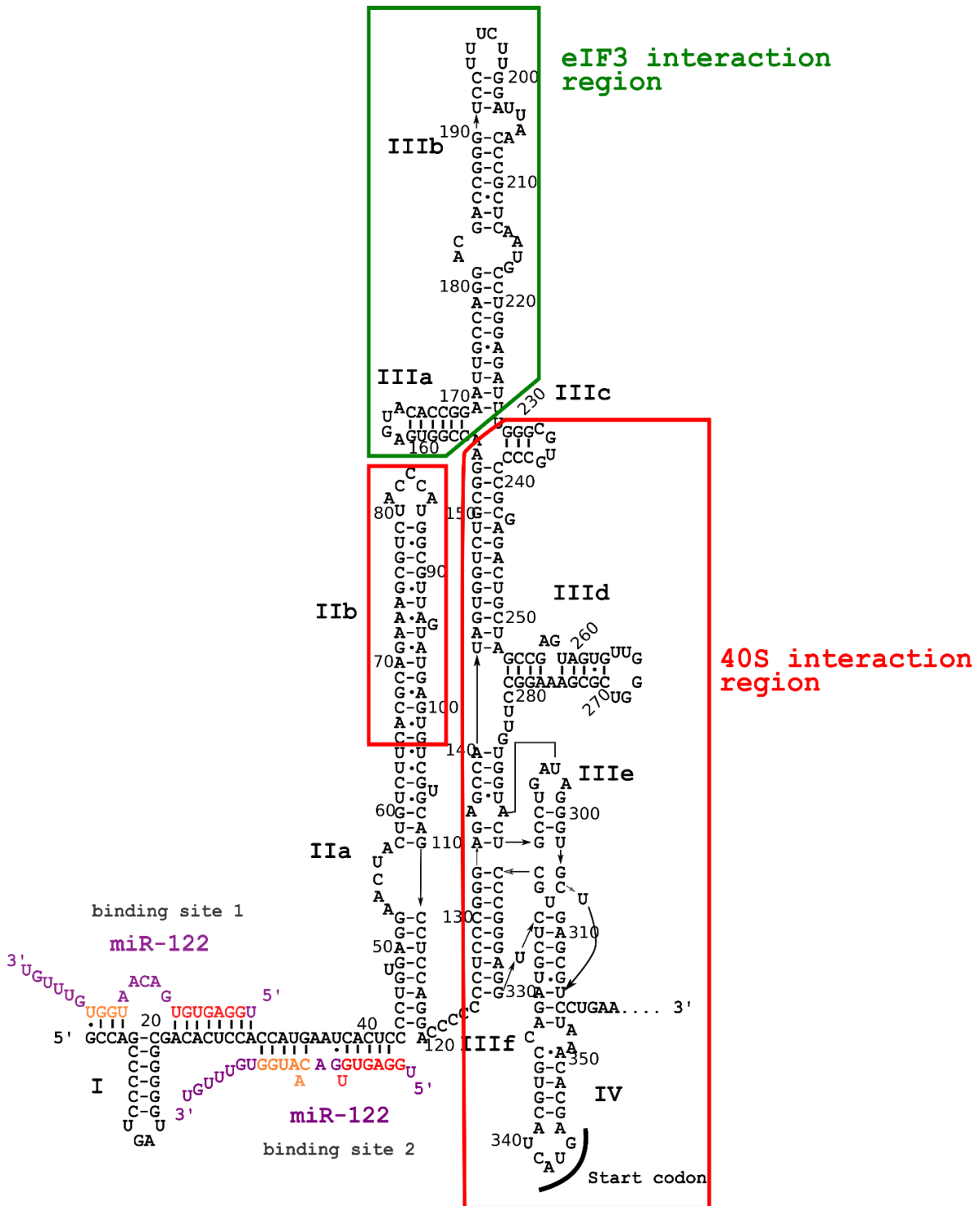
Å resolution in complex with rabbit 40S ribosomal subunit, which has been shown to have a conformational switch closing the mRNA binding cleft (Spahn et al. 2001). The best available cryo-EM structure of IRES in complex with human 40S ribosomal subunit has been resolved to 3.9 Å and it reveals that the expansion segment 7 (ES7) of the 18S rRNA is a central anchor point for the HCV IRES (Quade et al. 2015). IRES adopts its fold independently of the ribosome (Kieft et al. 2001), but this fold is distinct from the bound state. The structure of the free IRES is available only at low-resolution and was obtained with the help of SAXS. It represents an ensemble of conformers, which best fits experimental data (Pérard et al. 2013). Most importantly, structural studies of 5′ UTR cover just IRES and do not include domain I.

**Domain I** of 5′ UTR consists of G-C stem-loop and extended single-stranded region. It bears two microRNA seed regions that bind liver abundant miR-122. This interaction is important for viral propagation and will be reviewed in detail later. There is no three-dimensional structure of domain I available to date.

While several secondary structures of domain II have been predicted (Honda et al. 1999; Honda, Brown, and Lemon 1996; Zhao and Wimmer 2001), finally, its structure in the free state has been determined by solution-state NMR (Lukavsky et al. 2003), which agrees well with the one, based on phylogenetical prediction (Zhao and Wimmer 2001). Domain II folds independently and adopts a distorted L-shape structure, which is very similar to the one bound to the 40S ribosomal subunit complex (Quade et al. 2015; Spahn et al. 2001).

**Domain III** of 5′ UTR is a mayor branched stem-loop structure (abcdf), which bears three- and four-way junctions. The upper part of domain III forms a four-way junction (IIIabc), in the middle part there is a conserved stem-loop IIIc, which comprises a three-way junction. The basal part of domain III represents a four-way junction structure, which consists of pseudoknot IIIf and a stem-loop IIIe. During transcription initiation basal domain IIIdef and IIIc bind to the E-site of 40S ribosomal subunit, while the apical domain IIIab serves as a platform for eIF-3 (reviewed in (Lukavsky 2009)). For most of the small subdomains X-ray and NMR structures are available: 1F85 (IIIe) and 1F84 (IIIc) (Lukavsky et al. 2000), 1IDV (IIIc) (Rijnbrand et al. 2004), 1KH6 (IIIabc junction) (Kieft et al. 2002), and 1KP7 (IIIb) (Collier et al. 2002).

**Domain IV** is formed around the start codon AUG. In solution, domain IV adopts a small stem-loop structure, but most probably it unwinds upon the 40S binding (Honda et al. 1996). Schematic representation of 5′ UTR secondary structure, the regions of miR-122, 40S ribosomal subunit and eIF-3 interactions are illustrated in **Fig. 1.4.2**.



**Figure 1.4.2. Secondary structure of the HCV 5' UTR.** Two copies of miR-122 bound to domain I are shown in violet (seed region, 2-8 nt, is in red, auxiliary region is in orange). Green color box indicates eIF-3 interacting region of domain III, red color box encircles the regions of domain II, III and IV, that are involved in 40S ribosomal subunit interaction. The secondary structure and binding regions are from (Lukavsky 2009; Mortimer and Doudna 2013).

## 1.5. The role of host miR-122 in HCV lifecycle

MicroRNAs (miRNAs) are small (22-24 nt in length) non-coding RNAs. The sequence of miRNAs is highly conserved and tissue specific. In a complex with Argonaute (Ago) protein miRNAs usually target 3' end of mRNA in a sequence-dependent manner, forming RNA-induced silencing complex (RISC), which results in a cleavage of mRNA or repression of translation (Bartel 2004). Such mechanism to control the gene-expression is conserved among nematodes, insects, plants and vertebrates (Lewis et al. 2003).

miR-122 is a liver-specific abundant microRNA (Lagos-Quintana et al. 2002). It constitutes ~ 70% of the total liver miRNAs and each hepatocyte was estimated to have 120.000 copies of miR-122 (Valdmanis et al. 2018). Within HCV genome three putative binding sites were identified: two at the 5' UTR, one with 7 nucleotides and another with only 6 nucleotides complementarity to the miR-122 seed region, and the third binding site at the 3' UTR, which demonstrated independence from miR-122.

miR-122 is not a unique miRNA, which interacts with 5' UTR. Some other miRs were identified to target 5' UTR and coding mRNA regions (Fang and Rajewsky 2011; Lee et al. 2009). The stimulation effect on viral propagation by miR-122 is rare, but is not exceptional. For example, bovine viral diarrhea virus (BVDV) and classical swine fever virus (CSFV), which belong to the Pestivirus genus within the Flaviviridae family, depend on miR-17 and let-7 that interact with 3' UTR (Scheel et al. 2016). Nevertheless, neither for miR-122–HCV nor for the other host miRs-virus interactions there is a good mechanistic understanding of how exactly the virus exploits host miRNAs.

## 1.6. miR-122 – 5' UTR interactions

miR-122 interacts with the HCV 5' UTR in a single-stranded region located between stem-loop I and II. This region is involved in the regulation of both translation and replication of the virus.

Two copies of argonaute protein 2 (Ago2) can simultaneously occupy domain I increasing overall miR-122 binding affinity from the nanomolar to picomolar range. Ago2 was shown to attenuate miR-122 auxiliary binding at the second binding site to accommodate second copy of the Ago2–miR-122 complex at the first binding site. Computational modeling predicted the further interaction of Ago2 through PIWI and Mid domains with IRES SL II that stabilizes IRES-40S complex and promotes efficient translation (Chahal et al. 2019).

Previous studies have shown the importance of miR-122 binding to domain I of HCV RNA. Both sites at the HCV 5' UTR are equally important for virus replication (Jopling 2005, 2008; Thibault et al. 2015) and translation stimulation (Henke et al. 2008). Not only annealing to the seed region (2-8 nt) but also the auxiliary nucleotide binding (13-17 nt) was determined to be important for efficient viral propagation (Shimakami et al. 2012), which is very similar to the typical binding pattern of others microRNAs to 3' mRNA, which leads to suppression of gene expression in those cases. Interestingly, 3' terminal nucleotides of miR-122 are important for its HCV upregulation, whereas for its gene silencing function they are not necessary (Machlin, Sarnow, and Sagan 2011). However, recent finding, that siRNA, which binds to the nucleotides 23-35 of 5' UTR in Ago-2 knock-out cells, can provide same level of the replication stimulation as two copies of miR-122, propose that rather only two regions, which include seed region 1 and auxiliary site 2 are important. Additionally, this study shows that the binding to nucleotides 2-3 of 5' UTR may further enhance HCV replication but is not crucial. The role of 3' overhangs was also criticized as siRNA bound to 5' UTR without overhanging region was even less effective in replication stimulation than the one, which lacks 3' terminal 7 nucleotides (Kunden, Ghezelbash, et al. 2020).

Meanwhile, two groups of microRNAs, which can mimic miR-122 function and upregulate HCV, have been found (Ono 2020). The first group is miR-122-like RNAs, which target two binding sites via seed region, and miR-122-unlike RNAs, which target the bridge between two binding sites masking G28 and C29. This confirms the fact that HCV may be stimulated by other microRNAs (Ono 2020).

## 1.7. Stimulating effect of miR-122 on the HCV propagation and proposed mechanisms of miR-122 action

Since the discovery of the pivotal role of miR-122 for hepatitis C virus propagation, three independent positive effects of its interaction with 5' UTR were identified:

- Stimulation of viral replication
- Amplification of translation
- Viral genome stabilization

All three effects can be co-dependent: during translation non-structural proteins are expressed, subsequently they assemble to form the replication complex. Therefore, efficient replication is dependent on the translation. And vice versa: upon replication stimulation more copies of HCV RNA are produced, which are then available for translation. The stabilization

of viral genome, namely protection of HCV RNA from cellular exonucleases, has a positive impact on both translation and replication of the virus.

Thus, most of the scientific groups, that investigate the role of miR-122 on HCV propagation, including ours, are currently focused on the understanding of the individual contribution and mechanism of those effects. Although there are numerous studies, exact mechanisms of HCV upregulation by miR-122 are still mostly elusive, especially on the structural level.

### ***Stimulation of HCV replication***

Several mechanisms for miR-122-induced stimulation have been proposed. One of them is the displacement of poly(rC)- binding protein 2 (PCBP2), which is required for HCV translation, thus shifting the involvement of HCV RNA from translation to the replication (Masaki et al. 2015). Whereas the stimulation of the HCV replication is enormous (1000 times) in comparison to the translation stimulation (2 times), there is still no clear evidence, whether this effect is independent and is not a consequence of the RNA stabilization effect and translation stimulation provided by miR-122 (Kunden, Khan, et al. 2020; Schult et al. 2018).

### ***Stabilization of the HCV genome***

miR-122 protects 5' end of HCV RNA from Xrn2 exonuclease, which mediates 5'→3' RNA decay. Usually, Xnr2 terminates RNA Polymerase I and Polymerase II transcription. The depletion of Xnr2 increases HCV RNA amount by a factor of four, while following sequestration of miR-122 did not change this ratio (Sedano and Sarnow 2014).

Doudna group has demonstrated that a stable ternary complex of HCV and two copies of miR-122 resists the degradation mediated by abundant Xnr1 exonuclease. However, the same effect occurs upon the masking of 5' end of HCV RNA by the mutant version of miR-122 that is not able to enhance the propagation of HCV in cells but still binds to the 5' UTR, meaning that other than genome stabilization effect, different miR-122 mediated mechanisms of the HCV upregulation, should be present (Mortimer and Doudna 2013). This finding is in agreement with study by the Lemon group, showing, that the depletion of Xnr1 exonuclease does not rescue the replication of HCV with p6m mutation in both miR-122 binding sites. This mutant cannot bind miR-122, but the replication might be rescued by complementary mutated miR-122p6 (Li et al. 2013)

### ***Amplification of the HCV translation***

After the discovery of miR-122 enhancing effect on viral HCV propagation, experiments on HCV replicons missing replication ability have shown independent enhancement of the HCV translation via two equally important miR-122 binding sites located upstream from IRES and acceleration of association of the small ribosomal subunit with the viral RNA at the early initiation stage of transcription (Henke et al. 2008).

### ***miR-122 alters the structure of stem-loop II***

miR-122 binding may stabilize energetically less beneficial conformation of SL II that provides functional IRES ready for translation. By now, this mechanism was shown by two groups (Chahal et al. 2019; Schult et al. 2018) mainly by SHAPE method (Merino et al. 2005), computational studies and in vivo experiments on reporter constructs predicted to adopt classical or alternative domain II folding. However, there are no direct structural evidences of alternative conformations as yet.

Nevertheless, the chimeric construct of 5' UTR, containing first two domains of HCV 5' UTR and the III domain substituted with CSFV analog, which has similar secondary structure to HCV, showed a significant reduction in miR-122 translation stimulation in comparison to the wild type HCV 5' UTR. This implies that maximal miR-122 HCV translation stimulation requires the structures located downstream of domain II (Roberts, Lewis, and Jopling 2011).

miR-122 interacts with 5' terminus of HCV 5' UTR simultaneously but with different affinities. It was shown that binding at site 2 occurs with higher affinity (Chahal et al. 2019; Mortimer and Doudna 2013), whereas for the isolated domain I different groups have obtained ambivalent results. While for the 47 nt long construct of genotype 1b RNA the binding site 2 has higher affinity to miR-122 ( $K_d1$  (site 1) =  $845 \pm 354$  nM vs  $K_d2$  (site 2) =  $90 \pm 52$  nM) (Mortimer and Doudna 2013), 42 nt long construct of genotype 2a has lower  $K_d$  and, therefore, tighter binding for miR-122 at the binding site 1 ( $K_d1 = 20.45 \pm 11.8$  nM) than at the binding site 2 ( $K_d2 = 186.08 \pm 57.76$  nM) (Chahal et al. 2019). Currently it is not clear, whether these contradictory findings are due to the difference in the sequence or the length of the constructs.

Interestingly, in the absence of miR-122, HCV evolves adoptive mutations, so that miR-122 becomes obsolete for efficient viral replication, e.g., an emerging of G28A substitution in genotype 2a HCV in miR-122 decoy cells (Israelow et al. 2014) or C3U mutation in genotype 1b HCV in patients with virologic rebound after anti-miR treatment (Ottosen et al. 2015; van der Ree et al. 2017). Low levels of replication of the wild-type virus are often



detected in non-hepatic cells in chronic HCV patients (Castillo 2005; Wilkinson, Radkowski, and Laskus 2009). Ono et al. demonstrated that HCV might utilize other microRNAs in miR-122 deficient cells. Among these miRNAs two groups are distinguished: miR-122-like miRNAs which target 6 nucleotides at two binding sites and non-miR-122-like miRNAs with at least 7 matching nucleotides that bind at a single site between sites I and II at domain I of 5' UTR. Importantly, those miRNAs must provide stabilization of IRES three-dimensional structure (Ono 2020). Even more miR-122 independent HCV mutants have been discovered via siRNA-directed mutagenesis of 5' UTR (Amador-Cañizares et al. 2018). Despite existence of a variety of miR-122 independent HCV mutants, such mutations are not common in infected patients and mainly occur as a response to the therapy, which sequesters miR-122 or in chronic patients with weakened immune system (Israelow et al. 2014). Such observations can be explained as an evolutionary pressure on HCV to maintain miR-122 dependence. This miRNA dependence provides tropism to the liver, the organ, which is due to its immune tolerance favors chronic infection (Amador-Cañizares et al. 2018; Yu et al. 2017).

## 1.8. HCV therapy

Up to 2011, the conventional hepatitis C therapy included weekly injections of pegylated recombinant interferon alfa in combination with daily oral ribavirin intake for 24 to 48 weeks. This treatment provided sustained viral response (SVR) for up to 80-90% of patients infected with HCV genotype II and III, meaning that there was no virus detected in the blood over 6 months after the completion of the treatment. However, for the patients infected with genotypes I and IV SVR could be achieved only in 50% cases. In addition to the limited success and long duration, in many patients the treatment was associated with different adverse effects (Rong and Perelson 2010).

### 1.8.1. Direct-acting antiviral treatment

The development of highly effective direct-acting antiviral (DAA) therapy has revolutionized the treatment of HCV over the past decade. Initially it was used in combination with interferon-based therapy and could provide better therapy outcome (Hézode et al. 2013). DAAs are targeting the virus at the different stages of its lifecycle. Three major viral objects for DAAs are NS3/4A protease, NS5A protein, and NS5B RNA-dependent RNA polymerase, which are usually targeted simultaneously to increase the efficiency of the therapy and minimize the rate of HCV resistance (Götte and Feld 2016). Inhibition of

NS3/4A protease is carried out by peptidomimetic protease inhibitor (PI), which occupies the shallow cleft between the N-terminal and C-terminal domains of NS3 instead of the substrate, thus, interrupting the release of non-structural HCV proteins, which are necessary for replication of the virus. The most recent PI, the second-generation drug Grazoprevir, was shown to be less prone to resistance development and affect a broader range of genotypes in comparison to the first-generation PIs (Summa et al. 2012). Though structure and its relation to the function of multifunctional RNA-binding NS5A protein remain under investigation (Ross-Thriepland and Harris 2015), its inhibition shows great potential for antiviral treatment. The screening of compound libraries on a cell-based replicon system conceived the first NS5A inhibitor daclatasvir (Belema and Meanwell 2014) and later Ombitasvir (Gentile, Buonomo, and Borgia 2014) and ledipasvir (Link et al. 2014). The proposed antiviral effect of NS5A inhibition is the inhibition of membranous web formation (Berger et al. 2014). There are two big classes of NS5B inhibitors, which act at different stages of RNA synthesis: non-nucleotide inhibitors (NNIs) and nucleotide inhibitors (NIs). NNIs act as steric inhibitors at the initiation phase of the transcription preventing the switch of NS5B to the elongation initiation conformation. NIs are nucleotide analogs, which disrupt the binding and inclusion of rNTPs (e.g., sofosbuvir is uridine analog with 2'-fluoro-C-methyl motif). Unlike NNIs, sofosbuvir was shown to have a pangenotypic activity and high resistance barrier (Keating 2014). Currently, PEG-interferon-free DAAs can provide SVR better than 95% across all genotypes with a few to no adverse effects. The duration of the treatment usually takes 8-12 weeks and is oral needle-free, which makes the therapy more convenient for the patients (Martinello, Bajis, and Dore 2020). According to the cohort study DAA treatment was proved to be associated with the significant decrease in mortality rate among treated patients regardless of disease progression (Kalidindi et al. 2020).

### **1.8.2. miR-122 as a candidate target for HCV treatment**

So far, no microRNA-based therapies have been FDA-approved, though the field is promising and there are several drugs at the different stages of clinical trials (Bajan and Hutvagner 2020).

Due to the strong dependence of HCV RNA propagation on the miR-122 presence, miR-122 has been considered a reasonable target for pharmaceutical intervention strategies. Miravirsen is an anti-miR-122 therapeutic agent, which represents locked nucleic acid (LNA) antisense to miR-122. Miravirsen is a synthetic RNA chemically modified ribose that increases its stability and affinity to the target (miR-122) and is the first miRNA-based drug,

which has entered phase II of clinical trials. It has demonstrated half maximal effective concentration (EC<sub>50</sub>) of 0.67  $\mu$ M in the treatment of genotype 1b HCV infection (Ottosen et al. 2015). It has shown a long-lasting suppression of HCV replication in infected patients (Lindow and Kauppinen 2012). Additionally, miravirsen was shown to have an additive and synergistic antiviral effect in the combination therapy with DAAs for complicated cases (Liu et al. 2016). However, several subjects have demonstrated viral rebound associated with the C3U substitution (Mata et al. 2019).

Another miR-122 antagomiR drug, RG-101, was introduced by Regulus Therapeutics. RG-101 is hepatocyte targeting N-acetylgalactosamine conjugated antagomiR oligonucleotide. Similarly to miravirsen, it has demonstrated sustainable HCV replication inhibition and has entered phase II trials (van der Ree et al. 2017), but in Spain, Greece and Hungary the study is discontinued due to identification of inhibition of the bilirubin transporter in some patients (Jolla, Regulus, and Rgls 2017).



## 2. Methodological introduction

### 2.1. Size-exclusion chromatography and Multi-angle light scattering

#### 2.1.1. Size-exclusion chromatography

Size-exclusion chromatography (SEC), also known as gel-filtration, is a biophysical method for both analytical and preparative-scale molecules separation in solution based on their apparent size. The apparent size of the molecule depends on the molecular weight, its shape (folding) and the hydration shell.

In SEC the stationary phase consists of polymer beads with pores of different sizes typically packed into the column. The total volume of the column (also called bed volume) ( $V_t$ ) is equal to the sum of the volume which is not occupied by the beads, or void volume ( $V_0$ ), volume of the pores inside the beads ( $V_i$ ) and volume of the solid beads material ( $V_b$ ):

$$V_t = V_0 + V_b + V_i \quad (1)$$

As a mobile phase, which consists of the buffer with dissolved molecules of interest, passes through the equilibrated column, molecules passively distribute between the  $V_0$  and  $V_i$ . If the size of the molecule is very large, it cannot enter the pores at all and is consequently eluted in the void volume. The smaller the size of the molecules, the bigger pore volume they occupy, which is reflected by the bigger elution volume ( $V_e$ ) necessary to elute the molecules from the column.

Elution volume ( $V_e$ ) is an individual characteristic of the molecules of particular size. It also depends on the column volume, flow rate of the mobile phase, and pore size distribution. When it is necessary to compare the runs performed on the columns of the same matrix type but with different dimensions, distribution coefficient ( $K_d$ ) is used. This parameter also characterizes the elution of the molecule, but it is independent of the column volume.

$$K_d = (V_e - V_0)/V_i \quad (2)$$

In structural biological studies SEC is commonly used for proteins and nucleic acids purification and has the advantage to preserve their native folding and therefore biological activity. Since the molecules do not interact with the matrix and the elution is isocratic, wide range of possible buffers can be chosen, to keep the biomolecules in a stable and folded state.

SEC might be used not only to separate the molecules but also to characterize the sample oligomerization and complex formation between various biomolecules if it affects their apparent size.

The detection of the eluted biomolecules is typically done by measuring the UV-absorption, which is proportional to the sample concentration. After separation by their size, molecules can be characterized not only by UV-absorption but by wide range of techniques. For example, in this study Multi-angle light scattering and Small-angle X-ray scattering were performed in-line with SEC.

### **2.1.2. Multi-angle light scattering**

While SEC provides information about the apparent size of the molecules, it has a limited capacity for shape-independent molecular weight estimation. Multi-angle light scattering (MALS) technique is used for the determination of the molecular weight of molecules and their complexes in a shape-independent manner, as well as their radius of gyration ( $R_g$ ) and second virial coefficient  $A_2$ , a thermodynamic parameter, which is used to characterize weak self- and cross-interactions of the particles in solution. MALS can be applied both in batch and in line with SEC. When it is used in batch modus, measured parameters are the average values for all different molecules present in the solution. This regime can be informative for the monodisperse sample composition having the advantage of short measurement time and no need of the inline SEC setup. However, for the characterization of mixture of molecules of different size, appropriate SEC separation prior to MALS is necessary. MALS is non-destructive and then the sample can be used for further studies.

The technique measures the light scattered by the particle at different angles  $\Theta$ . Depending on the experimental setup, both multiple static detectors or single movable detector can be used. Typically, a light source is a collimated laser beam. When a laser beam meets the matter, its oscillating electric component induces the polarization of the matter, or in other words, induces the oscillating dipole in the sample, which re-radiates or scatters the light.

The theoretical background of MALS is described by Zimm`s equation (Zimm 2004):

$$\frac{K^*c}{R(\theta, c)} = \frac{1}{M_w P(\theta)} + 2A_2c \quad (3)$$

where  $R(\theta, c)$  is the excess Rayleigh ratio as a function of the scattering angle  $\theta$  and the concentration,  $c$ .  $M_w$  molecular weight of the molecule,  $P(\theta)$  describes the angular

dependence of the scattered light.  $A_2$  is the second virial coefficient.  $K^*$  is the constant equal to:

$$K^* = \frac{4\pi^2(\frac{dn}{dc})^2 n_0^2}{N_a \lambda} \quad (4)$$

where  $N_a$  is Avogadro's number  $6.022 \times 10^{23}$ ,  $\frac{dn}{dc}$  is the refractive index increment. It describes the change in the refractive index,  $\Delta n$ , with the concentration change,  $\Delta c$ , and reflects the polarization difference between the molecules and the solvent in a concentration-dependent manner. This value can be measured with refractive index detectors, but typically it has a little variation for a certain molecule type. Thus, its Wyatt Technology recommended value of 0.185 mL/g for proteins and 0.17 mL/g for nucleic acids was used.

$P(\theta)$  relates to the radius of gyration,  $R_g$  as:

$$\frac{1}{P(\theta)} = 1 + \frac{16\pi^2 n_0^2 R_g^2}{3\lambda_0^2} \sin^2\left(\frac{\theta}{2}\right) \quad (5)$$

where  $n_0$  is the refractive index of the solvent,  $\lambda_0$  is the wavelength of the incident beam in vacuum, and  $\theta$  is the measurement angle.

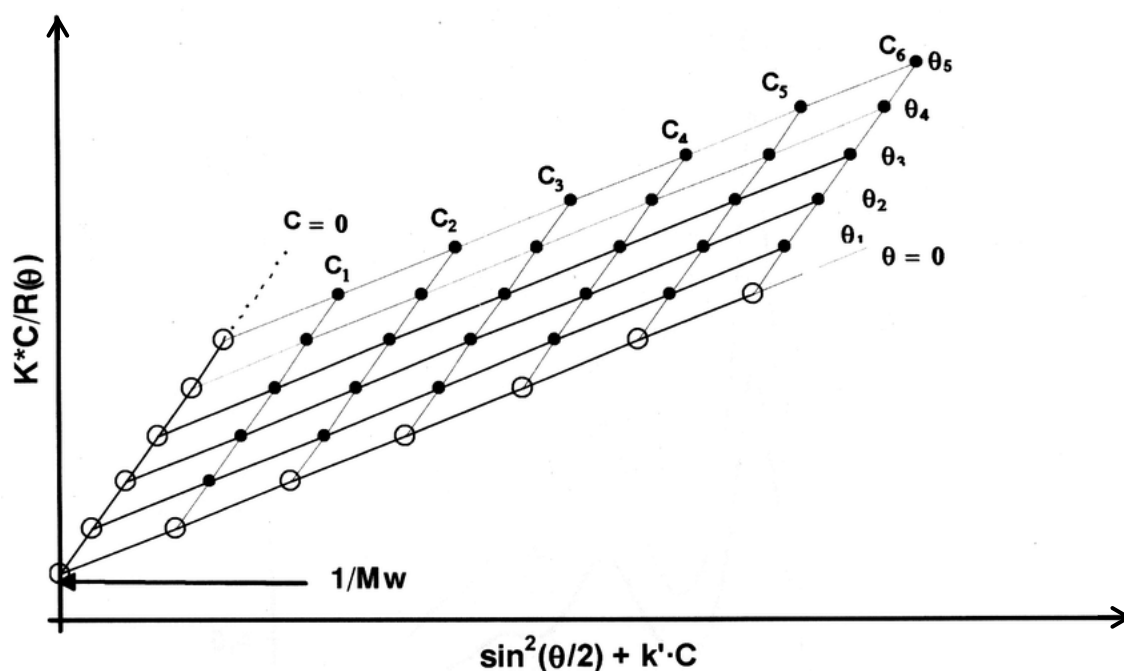
For the molecules, which size is greater than  $\sim 10$  nm (for the incident light at  $\lambda=660$  nm), the phase of the light scattered from the different parts of the molecule varies. This results in the angular variation of the intensity of the scattered light. Thus, for the molecules bigger than 10 nm,  $R_g$  can be determined.

In practice, the excess of Rayleigh ratio,  $R(\Theta)$ , is proportional to the intensity of the scattered light,  $I(\Theta)$ , which in turn is proportional to molecular weight,  $M_w$ , the concentration of particles in solution,  $c$ , and the refractive index increment,  $\frac{dn}{dc}$ .

$$I(\theta) \propto M_w c \left(\frac{dn}{dc}\right)^2 \quad (6)$$

As the concentration of the sample,  $R(\Theta, c)$ ,  $P(\Theta)$  can be directly measured in the experiment,  $M_w$  can be estimated via Zimm plot, where the scattering data for each slice of the elution volume (different concentrations) is plotted against varying angles and then extrapolates to zero angle and zero concentration. Second virial coefficient  $A_2$  can be extracted from the

slope of the projection to  $\theta = 0$ , while the slope of the projection to  $C = 0$  gives  $R_g$  value (Fig. 2.1.1).



**Figure 2.1.1.** Zimm plot with the experimental and extrapolated to zero light scattering data obtained as a series of measurements of the same sample at different concentrations,  $c$ , at different angles,  $\theta$ . The figure is adapted from (Oliva, Llabres, and Farina 2005).

For my thesis the use of MALS was indispensable, as the shape of the RNA molecule is rarely close to spherical and rather elongated, therefore, the elution volume on SEC in many cases can be very misleading for the sample  $M_w$  determination.

## 2.2. Small-angle scattering

Small-angle scattering (SAS) technique is a powerful tool to obtain low-resolution nanoscale atomic models of biological macromolecules in solution. SAS can be used complementary to other structural biology methods (NMR, X-ray crystallography, cryo-EM, etc.) or as an individual approach for those systems, which are not eligible for crystallization or cryo-EM studies due to their high flexibility or if their molecular weight is above the limit of solution-state NMR. Both NMR and SAS allow studying of biomacromolecules in nearly physiological conditions. In SAS two types of radiation could be used: the thermal neutrons ( $\lambda=1-10 \text{ \AA}$ ), then the technique is called small angle neutron scattering (SANS), and the X-rays ( $\lambda= 0.5-2 \text{ \AA}$ ) for small angle x-ray scattering (SAXS). SAXS has an advantage in the measurement time (a couple of minutes compared to 1-2 hours for a typical SANS

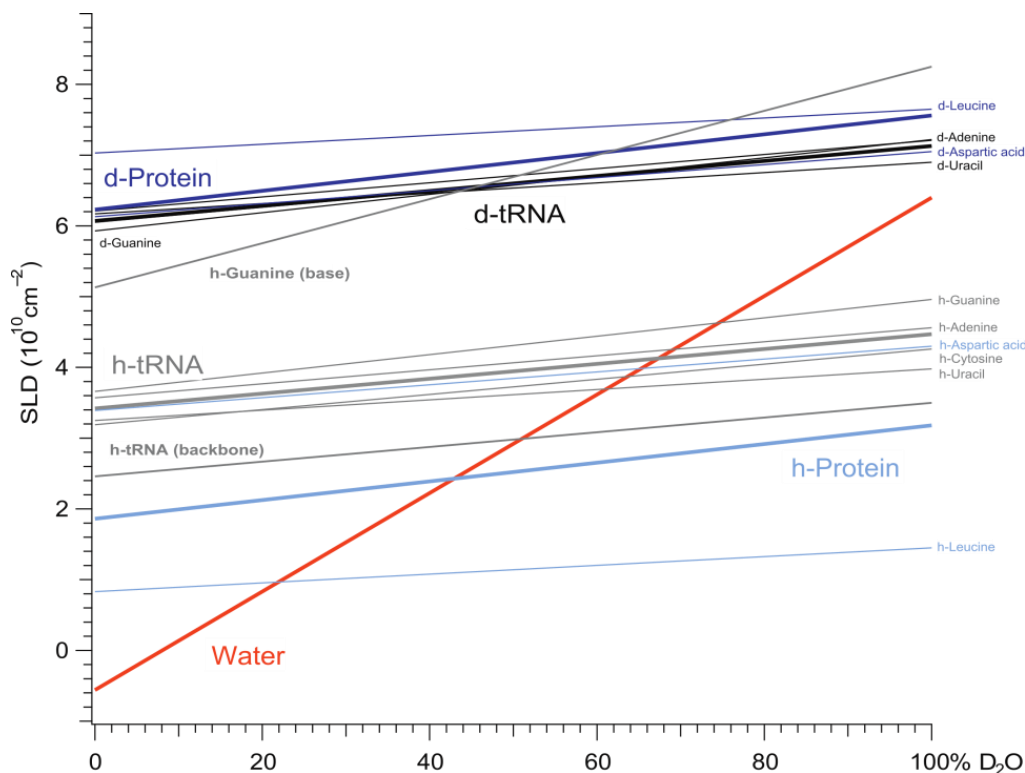


experiment) and low sample amount (10-30  $\mu\text{l}$  compared to 250  $\mu\text{l}$  for SANS); however, the strong limitation of SAXS over SANS is the radiation damage of the sample, which may severely affect the quality of the data (Feigin and Svergun 1987; Gabel 2015).

Apart from the absence of the radiation damage, another great advantage of SANS over SAXS or light scattering techniques is the possibility to selectively observe different components of the sample using a perdeuteration approach. Hydrogen  $^1\text{H}$  and its heavy isotope deuterium  $^2\text{H}$  exhibit distinct neutron scattering properties, (neutron coherent scattering length of proton and deuterium are  $b_c=-3.7423$  fm and  $b_c=6.675$  fm, respectively) and therefore very different scattering length densities (SLDs) for  $\text{H}_2\text{O}$  and  $\text{D}_2\text{O}$  ( $-0.562\times 10^{10}$   $\text{cm}^{-2}$  and  $6.404\times 10^{10}$   $\text{cm}^{-2}$ , respectively) (Svergun et al. 2013). By adjusting specific  $\text{H}_2\text{O}/\text{D}_2\text{O}$  ratios for the system buffer, it is possible to get a solution with any SLD within this range. Accordingly, it is possible to vary the level of perdeuteration of biological macromolecules as well as to exploit their natural neutron scattering properties dissimilarities, e.g., RNA and protein, to contrast-match them, e.g., make them “invisible” in solution, or, contrariwise, to highlight them (**Fig. 2.2.1**). Thus, it has a key advantage to study individual structural properties, such as conformation or relative orientation of complex subunits.

For example, protonated RNA is typically matched out in 68%  $\text{D}_2\text{O}$  and protein is matched out in ca. 42%  $\text{D}_2\text{O}$  buffer, while deuterated molecules at the same buffer conditions will have non-zero scattering contrast (Dunne et al. 2017; Jeffries et al. 2016). Based on the literature data, SLD ( $\rho$ ) can be calculated for any molecule as a sum of the neutron scattering lengths for individual atoms ( $b_i$ ) divided by the solvent-excluded volume of the molecule ( $V$ ) (Jacrot 1976; Voss and Gerstein 2005):

$$\rho = \frac{\sum_i b_i}{V} \quad (7)$$



**Figure 2.2.1. Perdeuteration approach in SANS.** Neutron scattering length densities for deuterated (d-) and hydrogenated (p-) protein (PDB entry 1GTR), t-RNA and selected amino acids and nucleotides as a function of the percentage of D<sub>2</sub>O in the solvent. The figure is adapted from (Gabel 2015).

Both techniques can be performed in line with both SEC and MALS, when it is necessary (Bucciarelli et al. 2018; Jordan et al. 2016). Though, X-rays are scattered by the electrons and neutrons of the nuclei, there is no fundamental difference between SANS and SAXS theory of scattering.

For the diluted, isotropically oriented mixtures the direct neutron or X-ray beam scatters symmetrically over the 2D detector plane. The scattered intensity,  $I$ , can be represented as a function of the absolute value of the momentum transfer,  $Q$  (the distance of the direct beam to the detector in inverse angstroms,  $\text{\AA}^{-1}$ ) (Guinier and Fournet 1955; Svergun et al. 2013)

$$I(Q) \propto N \left| \int_V \Delta\rho(\vec{r}) e^{i\vec{r}\vec{Q}} dV \right|^2 \quad (8)$$

where  $N$  is the number of scattering particles dissolved in the solvent. The integral runs over the whole solvent-excluded volume  $V$  of a particle with  $\vec{r}$  being the vector (of arbitrary

origin) pointing to a volume element  $dV$  of scattering contrast,  $\Delta\rho(\vec{r})$  is a difference between the SLDs of the particle and the solvent (also called the scattering contrast).

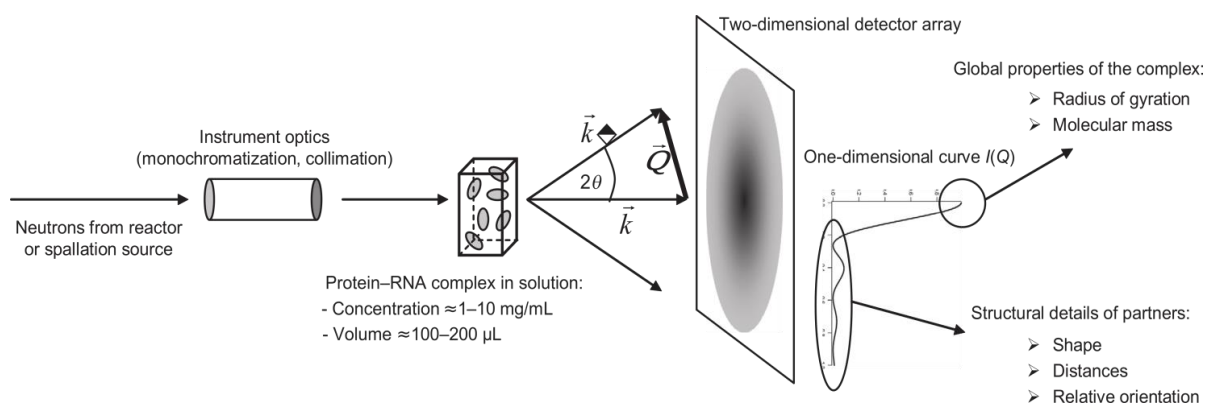
Q is defined as:

$$Q = \frac{4\pi}{\lambda} \sin \frac{\theta}{2} \quad (9)$$

where  $\lambda$  is the radiation wavelength and  $2\theta$  is the scattered angle.

For X-ray scattering SLD is the difference in electron density between the scattering molecule and the solvent. In neutron scattering, SLD depends on the nuclei type and can vary a lot between isotopes of the same element.

Standard SAS setup is also similar in SANS and SAXS (in **Fig. 2.2.2** SANS setup is shown) and consists of a radiation source, instrument optics, which includes a monochromator to select a certain range of wavelengths and a collimation system to align the beam, scattering system (a sample itself), sample holder and a detector of scattered neutrons or X-rays. Since the neutron flux is relatively low (in comparison to the one for X-ray scattering) larger volume of the sample is needed (ca. 100-200  $\mu\text{l}$  with the concentration of several mg/ml), but in contrast to SAXS the same sample can be measured for a long time (hours), because of the absence of the radiation damage (Hammouda 2008; Jeffries et al. 2016).



**Figure 2.2.2. SANS layout scheme.** The incident neutron beam passes instrumental optic, scatters on the sample. The scattering intensities are recorded by two-dimensional detector. The figure is adapted from (Gabel, 2015)

The measurement is performed in reciprocal (Fourier transform) space as a density-density correlation function, therefore the information of the phase is lost, thus making impossible the reconstruction of the complete image of the molecule from one sample.

However, it is possible to extract the information about the distances distribution  $P(r)$  within the molecule by applying Fourier transformation (FT) to the scattering function  $I(q)$ , therefore transforming the data to the real space:

$$P(r) = \frac{1}{2\pi^2} \int_0^\infty I(q)qr \sin(qr) dq \quad (10)$$

The limitation of the FT is the absence of the scattering data for the zero and infinite values of  $q$ . Here, the indirect Fourier transformation for a series of curves is used to fit experimental data with a given maximum dimension of the molecule ( $D_{\max}$ ).  $D_{\max}$  has to be chosen such, that resulting distance distribution curve smoothly approximates the zero value. For this, GNOM software can be used (Svergun 1992).

The interpretation of scattering data is based on the set of standard linear 2D plots, which represent functions of the scattered intensity  $I(Q)$  plotted against functions of the scattering range  $Q$ . The most important model-free parameters which can be extracted from the scattering data are molecular weight (MW) of the molecule and its radius of gyration ( $R_g$ ). Molecular weight value can be extracted from the equation 8 at the approximation of  $Q=0$ ,

$$I(0) \propto N \left| \int_V \Delta\rho(\vec{r}) dV \right|^2 \quad (11)$$

Indeed, MW is inversely proportional to the number of the particles, N, at the given concentration and square proportional to V. Therefore:

$$I(0) = A * \frac{1}{MW} * MW^2 \quad (12)$$

In order to find the proportionality constant A, which depends on several particle-independent parameters, including the radiation wavelength and geometric parameters of the instrument, usually the calibration for H<sub>2</sub>O is performed (Mylonas and Svergun, no date; Jacrot and Zaccai, 1981).

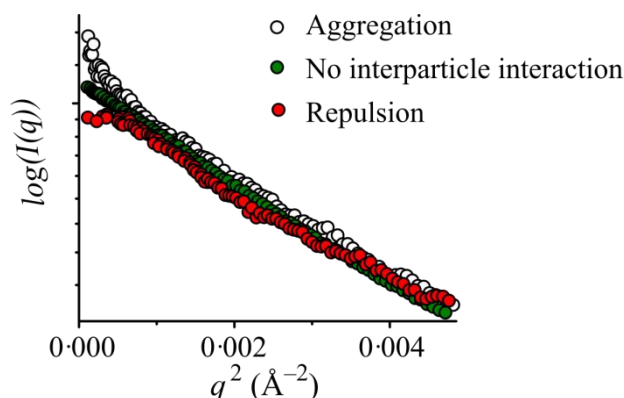
It is important to note, that the scattering intensity is highly sensitive to the oligomeric state of the sample, since it is proportional to the square of the particle volume. Therefore, the probable oligomerization state should be checked prior to the SAS measurements by other techniques (e.g., gel-filtration), since even the small fraction of the sample in oligomeric state drastically affects I(0) value.

Another important model-free parameter that is obtained from the data is radius of gyration (R<sub>g</sub>). R<sub>g</sub> is determined as a root mean square distance of each scattering centre to the centre of mass and represents an effective size of the molecule. It can be extracted from the Guinier plot (natural logarithm of scattering intensity (ln(I)) plotted against the square of the scattering variable (Q<sup>2</sup>)), using the Guinier approximation (Guinier 1939) and applying a linear fit at very low angles:

$$I(Q) \approx I(0) \exp\left(-\frac{q^2 R_g^2}{3}\right) \quad (13)$$

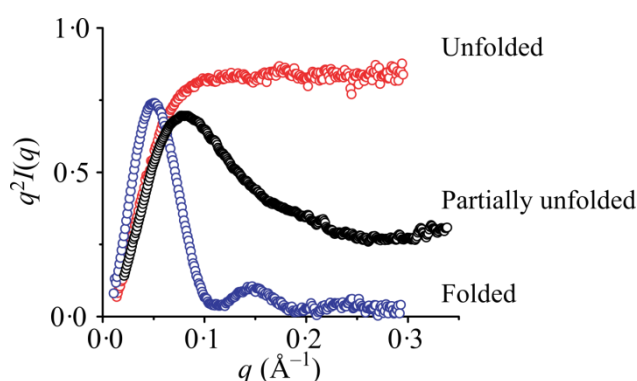
$$\ln(I(Q)) \approx \ln(I(0)) - \frac{q^2 R_g^2}{3} \quad (14)$$

The Guinier plot can be also used to assess the quality of the sample. A Guinier region with an upturn at q→0 is a sign of aggregation, whereas a Guinier plot with a downturn at q→0 can be an indication of intermolecular repulsion (**Fig. 2.2.3**).



**Figure 2.2.3. The Guinier plot.** Deviation of the Guinier region from the linear helps to identify interparticle interaction in the sample. Slope up (white) is typical for the aggregates in the sample. Turn down (red) is a sign of intermolecular repulsion. A good sample has a linear Guinier region (green). The figure is adapted from (Putnam et al. 2007).

The Kratky plot represents scattering data as  $I(q)q^2$  against  $q$  and serves as a good tool to visually estimate the folding of the macromolecule (**Fig. 2.2.4**). When the molecule is folded, Kratky plot has a bell-shaped form with an extremum, which provides an estimation for the overall size of the molecule. Unfolded extended molecules demonstrate the tendency to plateau or increased  $I(q)q^2$  at the larger  $q$ -values and miss the extremum peak (Putnam et al. 2007).



**Figure 2.2.4. Kratky plot.** The shape of the Kratky plot can be used to assess the folding degree of the sample (see detailed explanation in the text). The figure is adapted from (Putnam et al. 2007)

Apart from the model-free structural parameters, that can be extracted directly from the scattering curve, it is possible to obtain *ab initio* models generating a dummy atom envelope using DAMMIN, DAMMIF and MONSA software (Franke and Svergun 2009; Svergun 1999; Svergun and Nierhaus 2000). However, the main issue of the modeling of the 3D shape from the 1D scattering pattern is that several distinct models may exist, that all would have an identical scattering curve. On the opposite, the given structure yields only a single scattering curve. With the use of a CRYSON (for neutron scattering data) or CRY SOL (for X-ray scattering data) software, one can validate a set of the structures obtained by another structural biology technique and select the one, that satisfies experimental scattering data (Svergun et al. 1998; Svergun, Barberato, and Koch 1995).

The software compares the experimental scattering curve to the theoretical one calculated from the PDB file by calculating chi-square parameter ( $\chi^2$ ), which is defined as:

$$\chi^2 = \frac{1}{N-1} \sum_{i=1}^N \left[ \frac{I_{\text{exp}}(q_i) - \alpha I_{\text{mod}}(q_i)}{\sigma(q_i)} \right]^2 \quad (15)$$

where N is the number of data points,  $I_{\text{exp}}(q)$  is the experimental scattering intensity curve,  $I_{\text{mod}}(q)$  is the theoretical scattering curve calculated from the PDB file,  $\alpha$  is a scaling factor for the theoretical curve to minimize the value of  $\chi^2$ ,  $\sigma(q)$  is the standard error for each data point. Accordingly, the best fit is determined as the one with the minimal  $\chi^2$  value.

### 2.3. Nuclear magnetic resonance (NMR)

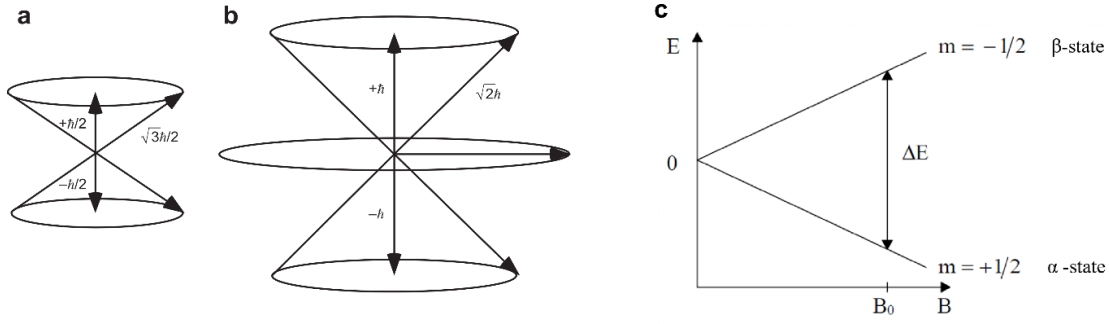
To provide better understanding of the experimental results obtained in this project, a brief introduction to the theory of NMR is given. Most of the theory description has been adapted from James Keeler's book "Understanding of NMR spectroscopy", the book of John Cavanagh "Protein NMR spectroscopy" and the course of the lectures of Jun.-Prof. Alexander Marchanka given at Leibniz Universität Hannover.

The nuclear spin angular momentum,  $P$ , is the complex vector quantum-mechanical property, which defines the existence of NMR spectroscopy. It is characterized by the nuclear spin quantum number,  $I$ . The nuclear spin number can be equal to zero, integer (1, 2, 3...) or half-integer (1/2, 1 1/2, 2 1/2...). For the conventional NMR spectroscopy, the nuclei with the non-zero spin number are used. The most important nuclei used in biomolecular NMR have either a spin quantum number  $I=1/2$  ( $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^{19}\text{F}$  and  $^{31}\text{P}$ ) or integer spin-quantum number  $I=1$  ( $^2\text{H}$ ).

The magnitude of nuclear spin angular momentum is related to the nuclear spin quantum number as

$$|\vec{P}| = \sqrt{\hbar[I(I+1)]} \quad (16)$$

where  $\hbar$  is the reduced Planck's constant.



**Figure 2.3.1. Angular momentum** vectors  $P$  and allowed  $z$ -component  $P_z$  of **a)** spin-1/2 and **b)** spin-1 particles (Cavanagh et al. 2010). **c)** The projections of the angular momentum of the spin-1/2-nuclei onto the  $z$ -axis of the laboratory frame in the presence of a static magnetic field,  $B_0$ , resulting in equally spaced energy levels ( $\alpha$ - and  $\beta$ -states)

According to Heisenberg's uncertainty, which states that the position and momentum of the particle cannot be measured simultaneously exactly, only the module and one of the Cartesian's components can be specified. Due to the instrumental setup, it is convenient to work with the  $z$ -component of the nuclear spin angular momentum,  $P_z$ , which is equal to:

$$P_z = \hbar m \quad (17)$$

where  $m$  is the magnetic quantum number, which can accept values from  $I$  to  $-I$  ( $2I+1$  in total), and thus, determines the orientation of the spin angular momentum in the space. For example, nuclei with  $I=1/2$  can have only two possible spin orientations. Nuclei with a nonzero spin possess nuclear magnetic momentum,  $\mu$ :

$$\vec{\mu} = \gamma \vec{P}, \quad (18)$$

$$\mu_z = \gamma P_z = \gamma \hbar m \quad (19)$$

Here,  $\gamma$  is gyromagnetic (magnetogyric) ratio, a nucleus characteristic constant.

When the external magnetic field is applied along  $z$ -axis, the energy of the spin states of the nucleus,  $E_m$ , can be written as:

$$E_m = -\gamma P_z B_0 = -\gamma \hbar m B_0 \quad (20)$$

Due to the quantization, the projections of the angular momentum of the nuclei onto the  $z$ -axis result in equally spaced energy levels (Zeeman levels). The number of the levels is  $2I+1$ .



Thus, for the nuclei with the spin  $\frac{1}{2}$  there are two energy levels, which are named  $\alpha$  and  $\beta$  and the energy difference for these levels is:

$$\Delta E = \gamma \hbar B_0 \quad (21)$$

The gap between the levels defines resonance frequency, e.g., amount of energy that is necessary to apply to excite a transition between energy levels. For  $\alpha$ -state  $m=+1/2$  and the spin is parallel to the magnetic field, with  $\alpha$ -state being energetically more favorable. In  $\beta$ -state  $m=-1/2$ , the spin is antiparallel and energetically less favorable. These discrete energy levels are populated according to the Boltzmann distribution:

$$\frac{N_\alpha}{N_\beta} = e^{-\Delta E/k_B T} \quad (22)$$

where  $N_\alpha$  and  $N_\beta$  are the numbers of spins on the respective energy levels,  $T$  is the temperature in Kelvin and  $k_B$  is the Boltzmann constant ( $1.380649 \times 10^{-23}$  J/K).

The difference in the spin population of  $\alpha$  and  $\beta$  levels results in small microscopically observable magnetization that is a sum of all magnetic moments. The bulk magnetization precesses around z-axis with a frequency  $\omega$ , which is proportional to the strength of the applied magnetic field and the gyromagnetic ratio of nuclei. This frequency is called Larmor frequency:

$$\omega = \gamma B_0 \quad (23)$$

Each individual spin has a certain chemical and therefore electron environment. The external magnetic field,  $B_0$ , generates currents in electron clouds, which in turn generate a very small ( $10^{-4}$  of  $B_0$ ) induced magnetic field,  $B_{ind}$ . Thus, in reality every nucleus experiences a slightly different local magnetic field,  $B_{loc}$ , which is equal to the sum of the main field and the induced field:

$$B_{loc} = B_0 + B_{ind} \quad (24)$$

Though this difference is very small, NMR is sensitive enough to detect it and exploits this important feature to distinguish between otherwise identical nuclei in a certain chemical

environment. This phenomenon is called a chemical shift and is of critical importance in NMR.

The distribution of electrons in a molecule is anisotropic and chemical shift is described as the second-rank tensor, represented by a 3x3 matrix. The chemical shift tensor for each nuclear site is characterized in orthogonal coordinates, the Principal Axes System (PAS) with respect to the nucleus position and  $\delta_{xx}$ ,  $\delta_{yy}$ ,  $\delta_{zz}$  principal components. If the i-th principal axis of the chemical shift tensor is oriented along the z-axis of the static field, then the net magnetic field at the nucleus is given by:

$$B_{loc} = \delta_{ii} B_0 \quad (25)$$

In gases or liquids, the molecular tumbling is rapid and nuclei experience averaged local fields, chemical shift is then isotropic and is defined as:

$$\delta_{iso} = \frac{\delta_{xx} + \delta_{yy} + \delta_{zz}}{3} \quad (26)$$

However, for molecules with distinct orientations (e.g., in solids) chemical shift anisotropy is defined as:

$$\delta_{\Delta} = \delta_{xx} - \frac{\delta_{yy} + \delta_{zz}}{2} \quad (27)$$

The asymmetry of the chemical shift tensor is:

$$\eta = \frac{\delta_{xx} - \delta_{yy}}{\delta_{zz} - \delta_{iso}} \quad (28)$$

The absolute value of the nucleus frequency depends on the  $B_0$  and its chemical environment. To compare the values of the resonance frequencies from spectrometers with different field strengths, the chemical shift value is expressed via the reference component (TMS), which chemical shift value is taken as 0, and is expressed in parts per million (ppm):

$$\delta = \frac{\omega - \omega_{reference}}{\omega_{reference}} \cdot 10^6 \quad (29)$$

There are different conventions for chemical shift definition. Sometimes (e.g., in ssNMR field), the shielding tensor is used instead, which is denoted with the symbol  $\sigma$  and is the same magnitude but opposite sign as  $\delta$ .

### 2.3.1. Quantum spin Hamiltonian theory

To describe the state of the system, its evolution in time and space, a quantum mechanical explanation is necessary. The extended theoretical background is outside of the scope of this thesis and the reader is referred to the excellent textbook «Spin dynamics: Basics of nuclear magnetic resonance» (Levitt 2009) and articles cited in this chapter.

The operator which represents the total energy of the system is called Hamiltonian,  $\hat{\mathcal{H}}$ . The Schrödinger equation relates the wave function,  $\psi$ , for the state of the system with its energy operator as:

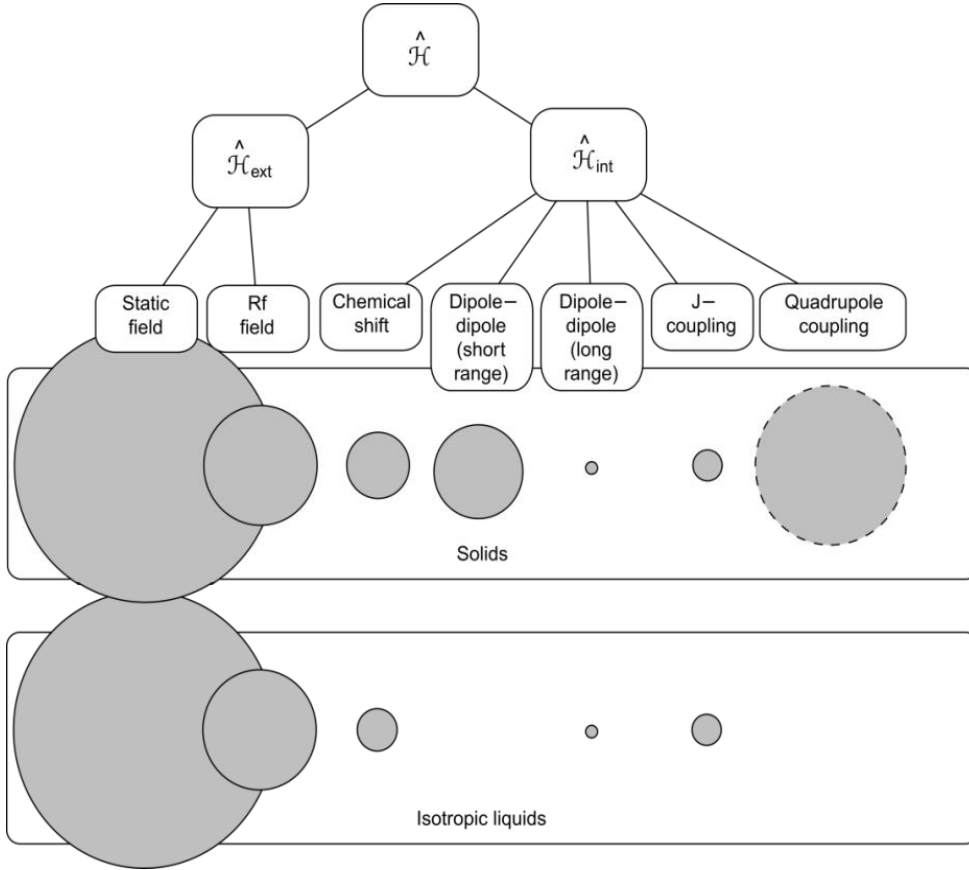
$$\frac{d}{dt}|\psi_{spin}(t)\rangle \cong -i\hat{\mathcal{H}}_{spin}|\psi_{spin}(t)\rangle \quad (30)$$

where  $\psi_{spin}$  is the spin state of the nuclei and  $\hat{\mathcal{H}}_{spin}$  is nuclear spin Hamiltonian. In this equation an important approximation is made: Hamiltonian depends only on the directions of the nuclear spin polarization and the influence of the surrounding electrons' motion is taken as an average.

Hamiltonian can be written as the sum of external (static and gradient magnetic fields, rf pulses) and internal (chemical shift isotropic and anisotropic, dipole-dipole, J-couplings and quadrupole couplings) components:

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_{ext} + \hat{\mathcal{H}}_{int} = \hat{\mathcal{H}}_{static} + \hat{\mathcal{H}}_{grad} + \hat{\mathcal{H}}_{rf} + \hat{\mathcal{H}}_{cs,iso} + \hat{\mathcal{H}}_{CSA} + \hat{\mathcal{H}}_j + \hat{\mathcal{H}}_D + \hat{\mathcal{H}}_Q + \dots \quad (31)$$

Depending on the aggregate state of the sample under the NMR study, some internal interactions are observed while others are not (**Fig. 2.3.2**). In gases and isotropic liquids, where all orientations are equal, all intramolecular interactions are averaged and keep only their isotropic values, while in solids, due to the lack of fast isotropic molecular tumbling, internal spin Hamiltonian has an anisotropic component and depends on the orientation of the sample relative to the magnetic field.



**Figure 2.3.2. Relative magnitudes for the motionally averaged spin Hamiltonians in solids and isotropic liquids.** Dashed line for quadrupolar interactions reminds that for spin 1/2 nuclei quadrupolar coupling vanishes. The figure is adapted from (Levitt 2009)

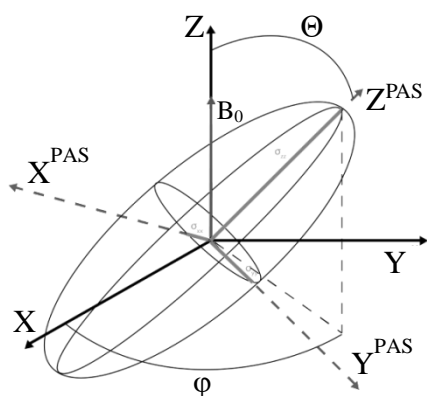
### 2.3.2. Internal spin interactions

**Chemical shift anisotropy (CSA)** originated from the anisotropy of the local magnetic field for a given nucleus for all the directions and the chemical shift is of tensorial nature (**Fig.**

**2.3.3**). Considering a system of  $\frac{1}{2}$ -spin, I, Hamiltonian for the chemical shift,  $\hat{\mathcal{H}}_{CS}$ :

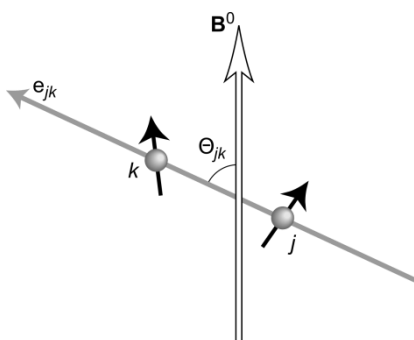
$$\hat{\mathcal{H}}_{CS} = (\omega_0 \sigma_{iso} + \frac{\omega_0 \delta_\sigma}{2} [(3 \cos^2 \theta - 1) - \eta_\sigma \sin^2 \theta \cos 2\phi]) \hat{I}_z \quad (32)$$

where  $\omega_0$  is the spin Larmor frequency,  $\sigma_{iso}$  is the isotropic chemical shift,  $\delta_\sigma$  is chemical shift anisotropy component,  $\theta$  is the angle between orientation of spins in solids and the z-axis,  $\phi$  determines the position of the spin magnetization relative to the xy-plane,  $\eta_\sigma$  is the anisotropy parameter,  $\hat{I}_z$  is z-component of the spin operator (Polenova, Gupta, and Goldbourt 2015).



**Figure 2.3.3. Graphical representation of chemical shift tensor (ellipsoid) in the principal axis system (PAS) and its relation to the laboratory frame coordinates. Angles  $\Theta$  and  $\phi$  define its orientation. Magnetic field  $B_0$  is directed along z-axis. Modified from source (Kaseman 2023)**

**Dipolar coupling** originates from the interaction between magnetic dipole moments of two spins (**Fig. 2.3.4**). This interaction depends on the distance and spins orientation with respect to the external magnetic field.



**Figure 2.3.4. Direct dipole-dipole interaction of two spins k and j. The unit vector  $e_{jk}$  connects the centers of two spins. The angle  $\Theta_{jk}$  defines the relative position to the vector of magnetic field,  $B_0$ . The figure is adapted from (Levitt 2009)**

Hamiltonian for homonuclear dipole-dipole interaction in a secular approximation (high magnetic field along z-axis):

$$\hat{\mathcal{H}}_{jk}^{DD} = -\frac{\mu_0 \gamma_j \gamma_k \hbar}{16\pi r_{jk}^3} (3\cos^2\theta_{jk} - 1)(3\hat{I}_{jz}\hat{I}_{kz} - \hat{I}_j - \hat{I}_k) \quad (33)$$

and for heteronuclear interaction:

$$\hat{\mathcal{H}}_{jk}^{DD} = -\frac{\mu_0 \gamma_j \gamma_k \hbar}{16\pi r_{jk}^3} (3\cos^2\theta_{jk} - 1)2\hat{I}_{jz}\hat{I}_{kz} \quad (34)$$

Where  $r_{jk}$  is the distance between two spins j and k,  $\theta_{jk}$  is the angle between the orientation of an external magnetic field vector  $B_0$  and vector  $e_{jk}$ , which connects centres of two spins.  $\hat{I}_j$

and  $\hat{I}_k$  are angular momentum operators and  $\hat{I}_{jz}, \hat{I}_{kz}$  are their secular components. Thus, secular dipolar coupling depends on the distance between two spins and their orientation with respect to the external magnetic field.

Both CSA and dipole-dipole interactions are not observable in liquid-state NMR but are the origin of the fluctuating magnetic field, which leads to relaxation phenomena and is of utmost importance for an NOE (*vide infra*).

**J coupling** originates from the indirect interaction between two nuclear spins via electrons, which constitute the bond between two nuclei. J coupling is independent of the magnetic field strength applied. The sign of the J-coupling can be both positive and negative according to the spin orientation. J-coupling is exclusively intramolecular and is detectable only for nuclei, which are linked together through a small number of chemical bonds.

The full form of intermolecular J-coupling between two spins  $\hat{I}_j$  and  $\hat{I}_k$  is:

$$\hat{\mathcal{H}}_{jk}^{J,full} = 2\pi\hat{I}_j\mathbf{J}_{jk}\hat{I}_k \quad (35)$$

where  $\mathbf{J}_{jk}$  is the J-coupling tensor 3x3 real matrix.

Isotropic component of the J-coupling,  $J_{jk}$  is called scalar coupling and is preserved in isotropic liquids:

$$\hat{\mathcal{H}}_{jk}^{J,iso} = 2\pi J_{jk}\hat{I}_j\hat{I}_k \quad (36)$$

Scalar coupling is equal to the average of the diagonal elements of J-coupling tensor matrix:

$$J_{jk} = \frac{1}{3}(J_{xx}^{jk} + J_{yy}^{jk} + J_{zz}^{jk}) \quad (37)$$

**Quadrupolar coupling** is a result of electrical interaction of the nucleus with surrounding electrons. Quadrupolar interactions are averaged to zero in isotropic solutions for all nuclei and for spin-1/2 nuclei in solids. For nuclei with the spin  $>1/2$  the Hamiltonian for the first-order nuclear quadrupolar interaction is written as:

$$\hat{\mathcal{H}}_Q^1 = -\frac{eV_{zz}\gamma_k Q}{4I(2I-1)\hbar} [(3\cos^2\theta_{jk} - 1) - \eta_\sigma \sin^2\theta_{jk}\cos 2\varphi] (3\hat{I}_z^2 - \hat{I}^2) \quad (38)$$

where  $V_{zz}$  is the largest component of the electric field gradient tensor,  $Q$  is the nuclear quadrupolar moment (Polenova et al. 2015).

Irradiation at corresponding Larmor frequencies (radio frequencies, RF), e.g., **external spin interactions** can manipulate the bulk magnetization. The strength of the pulse is defined by nutation frequency, which specifies *how fast* the pulse can rotate the magnetization.

The most common pulses used in NMR are  $90^\circ$  and  $180^\circ$  pulses, e.g., they rotate the magnetization by 90 or 180 degrees around corresponding axis. The  $90^\circ$  pulse along x- or y-axis transfers the bulk magnetization ( $M_z$ ) out of z-axis to the xy-plane (transverse plane), where it *precesses* with Larmor frequency around z-axis. The  $180^\circ$  pulse along x- or y-axis inverts the magnetization vector from  $M_z$  to  $-M_z$  making  $N_\beta$  larger than  $N_\alpha$ , e.g., performing population *inversion*.

Once RF pulse is off, the magnetization starts to return to the equilibrium state due to relaxation. There are two types of relaxation:

1) Spin-lattice or also called longitudinal relaxation, which results in the recovery of the magnetization along z-axis. This process is driven by the interaction with the surroundings (lattice) of the spin and is described by a first-order rate expression:

$$\frac{dM_z}{dt} = -\frac{M_z - M_0}{T_1} \quad (39)$$

$$M_z = M_0(1 - e^{-\frac{t}{T_1}}) \quad (40)$$

where  $M_0$  is the magnetization along the z-axis at thermal equilibrium and  $T_1$  is the spin-lattice or longitudinal relaxation time constant.

2) Spin-spin relaxation, also called transverse relaxation, results in a loss of the magnetization in xy-plane by transferring the energy to the neighboring spins and therefore the misplacement of the spin's coherence. It is characterized by the  $T_2$  relaxation constant and is described by the following expression:

$$\frac{dM_{xy}}{dt} = -\frac{M_{xy}}{T_2} \quad (41)$$

$$M_{xy} = M_0 e^{-\frac{t}{T_2}} \quad (42)$$

According to Faraday's law, the precession of the bulk magnetization around the static field induces an electromotive force in a coil, which can be detected by the receiver coil and in NMR experiment is called free induction decay (FID). FID is typically digitized and then transformed by Fourier transformation (FT) into the frequency domain.

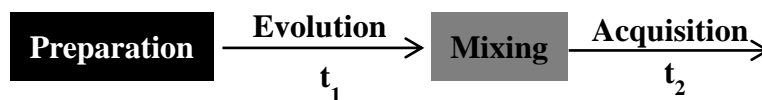
To get a reasonable signal-to-noise ratio (SNR), multiple FIDs at identical conditions must be recorded. The length of  $T_1$  determines the time of the recovery of the bulk magnetization to the equilibrium state and therefore the possibility to start the new experiment. Typically, the delay between two experiments is taken as five  $T_1$ . Thus, the length of  $T_1$  dictates the experiment duration, while the value of  $T_2$  has an impact on the spectral line-width, with short  $T_2$  being a reason for the low spectral resolution and sensitivity.

### **2.3.3. Multi-dimensional NMR**

Application of NMR spectroscopy for the structure determination of biomacromolecules has massively advanced with the development of multi-dimensional experiments.

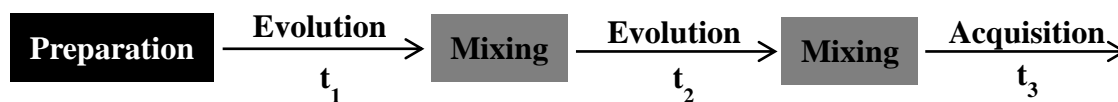
The general scheme of 2D NMR experiment is illustrated in **Fig. 2.3.5**. It includes four basic blocks: preparation, evolution, mixing and acquisition. During the preparation block, the pulse or a pulse sequence brings the system out of equilibrium and transforms the magnetization into some coherence. The coherence evolves throughout the evolution period, which is characterized by the time delay  $t_1$ .  $t_1$  varies over the measurement series with a certain increment and a series of experiments acquired at various  $t_1$  forms an indirect dimension. Importantly, there is no detection during evolution time. Mixing block is a period, during which the information about one spin is related to another and a correlation between two dimensions is established. The transfer of coherence can be established either through the bond via scalar couplings, or through space via Nuclear Overhauser Effect (NOE), exploiting dipolar couplings. A set of free induction decays (FIDs) with variable indirect evolution times  $t_1$  are recorded during direct acquisition time,  $t_2$ , and form the array of FIDs. Finally, time domain data is Fourier transformed (FT) both in  $t_1$  and  $t_2$  dimensions to obtain frequency domain data.





**Figure 2.3.5. The principal scheme of the two-dimensional NMR experiment.** The typical experiment has four blocks. During the preparation period, a pulse or a pulse sequence perturbs equilibrium magnetization that is later evolves during the evolution block. Mixing block serves for the manipulation of the coherence left after the evolution period to transform it into the recordable signal. This signal is recorded during the acquisition time  $t_2$ . The most important 2D experiments such as NOESY, HSQC and TROSY used in this project are discussed in the text.

Often the assignment and structural calculations of biomacromolecules are complicated due to the spectral overlap. To overcome this problem a third dimension can be introduced. Three-dimensional NMR experiment can be shown as two tandem two-dimensional experiments linked via additional evolution and mixing period, which form a third indirect dimension (Fig. 2.3.6).



**Figure 2.3.6. The general scheme of three-dimensional experiment.** Six blocks of 3D NMR experiment include additional mixing and evolution block as compared to 2D NMR experiment

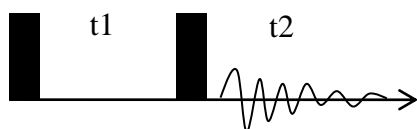
All 3D NMR experiments can be typically separated into two groups: the triple resonance experiments and two combined 2D double resonance experiments.

Triple resonance experiments are usually applied for protein/RNA assignment. They correlate three types of nuclei (typically  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$ ) via scalar coupling and therefore require double (typically  $^{13}\text{C}$  and  $^{15}\text{N}$ ) isotope labeling of the molecule.

Combined 2D double resonance experiments normally consist of 2D  $^1\text{H}$ ,  $^1\text{H}$  NOESY or e.g., 2D  $^{13}\text{C}$ ,  $^{13}\text{C}$ -TOCSY experiment, which is additionally resolved by 2D HSQC in the third dimension.

### *Correlation spectroscopy (COSY)*

COSY experiment is one of the most useful and basic homonuclear 2D experiments. It helps to identify nuclei (e.g., protons) that are coupled to each other by scalar couplings. The scalar-coupled spins generate cross-peaks at the position with corresponding chemical shifts of individual spins. The pulse sequence of this experiment is straightforward and is illustrated in Fig. 2.3.7.



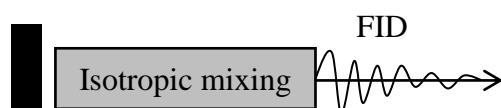
**Figure 2.3.7. Pulse sequence for COSY experiment.** Black rectangles represent  $90^\circ$  pulses. First  $90^\circ$  pulse is followed by the evolution time  $t_1$ , during which the magnetization evolves under the offset of the spin one and scalar coupling between two spins. Second  $90^\circ$  pulse is followed by acquisition during time  $t_2$ .

First  $90^\circ$  pulse generates  $y$ -magnetization. This magnetization is then evolved during  $t_1$  under the offset of the spin one and the scalar coupling between two spins. The second  $90^\circ$  pulse transfers the phase magnetization from one spin to another via coherence transfer and makes the signal detectable.

### ***Total correlation spectroscopy (TOCSY)***

In TOCSY experiment the isotropic mixing sequence is used to transfer in-phase magnetization through the bond between two strongly coupled spins (**Fig. 2.3.8**). A TOCSY experiment has a great advantage over a COSY experiment, since the magnetization can be transferred through several couplings thus greatly simplifying the assignment. If there is no significant relaxation, the cross-peaks between all resonances within one spin system are formed, also called the unbroken chain of couplings.

TOCSY can correlate all protons in the RNA ribose ring in single experiment.



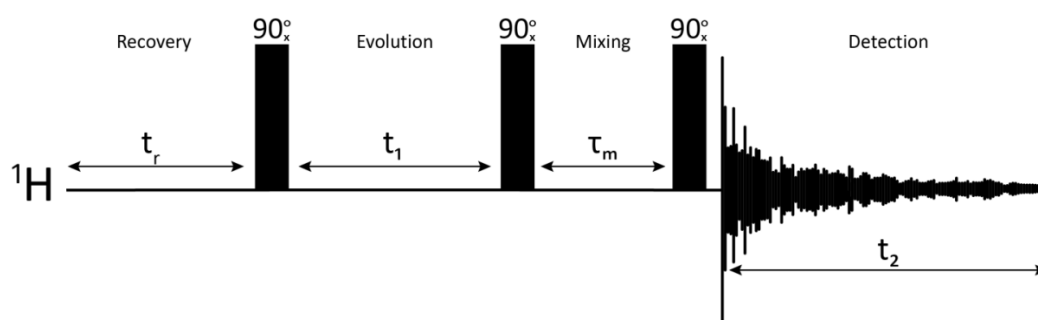
**Figure 2.3.8. The basic TOCSY pulse sequence scheme.** Isotropic mixing is carefully selected train of pulses (e.g., MLEV-17 or DIPSI-2), which serves to suppress chemical shifts for energy matching to allow spin exchange between protons, which are connected via unbroken chain of J-coupling.

### ***Nuclear Overhauser Effect Spectroscopy (NOESY)***

The spins can be connected not only by through-bond (J) couplings, but also by through-space interactions, e.g., by indirect dipolar couplings also known as Nuclear Overhauser Effect (NOE). This phenomenon arises from the cross-relaxation between two adjacent spins, ( $< 5 \text{ \AA}$ ), while one spin is saturated.

The most popular experiment, which exploits NOE and serves to determine structural features of biomolecules is 2D  $^1\text{H}$ - $^1\text{H}$  Nuclear Overhauser Effect Spectroscopy (NOESY). The transfer of the coherence is carried out through indirect dipolar coupling interaction. The experiment identifies nuclear spins, which undergo cross-relaxation as well as to measure their cross-relaxation rates. The basic pulse sequence for the 2D NOESY experiment is shown in **Fig. 2.3.9**. The basic experiment consists of three  $90^\circ$  pulses: the first pulse creates transverse spin magnetization, which evolves during the time delay  $t_1$ , the second pulse

rotates the magnetization component back along z-axis, thus, transforming the transverse magnetization to the longitudinal. During the mixing period magnetization transfer between spins that are close in space happens via dipolar coupling-mediated cross-relaxation mechanism. The third  $90^\circ$  pulse brings the remaining longitudinal magnetization back to the xy-plane for the detection during  $t_2$ . The NOESY mixing time  $t_m$  is typically selected in a range of 150 to 500 ms and depends on many factors, e.g., molecular weight of the molecule and temperature. The cross-peaks for the spins that are within the distance range of  $<5\text{\AA}$  can be observed since an NOE effect scales as  $1/r^6$ . Most importantly, the intensity of these cross-peaks provides an estimation for the inter-spin distances, using the peak intensity between two spins of known distance as a reference.

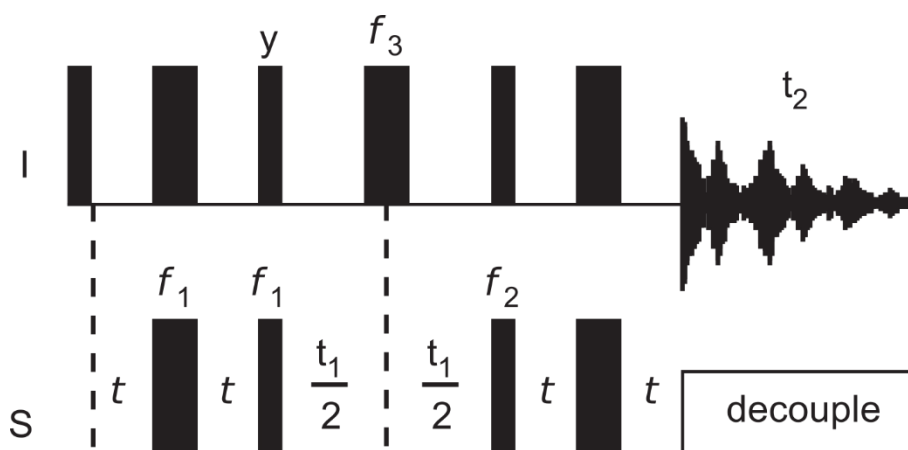


**Figure 2.3.9. Standard pulse sequence for two-dimensional  $^1\text{H}$ - $^1\text{H}$  NOESY NMR experiment.** See description in text. By Janopus - Own work, CC BY-SA 4.0, <https://commons.wikimedia.org/w/index.php?curid=76533364>

### ***Heteronuclear single quantum coherence experiment (HSQC)***

Heteronuclear single quantum coherence (HSQC) spectrum is one of the most used experiments for acquisition of one-bond correlation spectra between heteroatoms. HSQC can be acquired in 2D fashion to get a fingerprint of the molecule and to follow conformational changes or as a part of a 3D experiment to get additional resolution for the assignment or structural calculations. Typically, 2D  $^1\text{H}$ ,  $^{15}\text{N}$  and 2D  $^1\text{H}$ ,  $^{13}\text{C}$  fingerprint spectra are acquired, which utilize one bond  $^1J_{\text{HN}}$  and  $^1J_{\text{HC}}$  couplings, respectively.

The typical HSQC pulse sequence is illustrated in **Fig. 2.3.10**. During the first step, the magnetization is transferred from one spin type (I, typically  $^1\text{H}$ ) to another spin (S, typically  $^{15}\text{N}$  or  $^{13}\text{C}$ ) by Insensitive Nuclei Enhanced by Polarization Transfer (INEPT) pulse sequence to create an antiphase heteronuclear single-quantum (SQ) coherence. The magnetization of the spin S evolves during variable delay  $t_1$  and becomes frequency labeled, and then it is transferred back to the spin I ( $^1\text{H}$ ) by an inverse INEPT pulse sequence for the detection during  $t_2$ . Thus,  $^1\text{H}$  chemical shifts are recorded in direct dimension, while chemical shifts of a heteroatom ( $^{15}\text{N}$  or  $^{13}\text{C}$ ) are recorded in the indirect dimension.



**Figure 2.3.10. Typical pulse sequence for the HSQC experiment.** Thick bars represent  $180^\circ$  pulses and thin bars illustrate  $90^\circ$  pulses. Spin I is decoupled from spin S nuclei during the  $t_2$  acquisition time.  $f_x$  represent phase cycling. The figure is adapted from (Cavanagh et al. 2010)

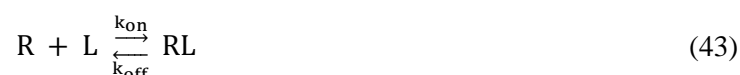
### **2.3.4. Chemical shift perturbation analysis**

Chemical shift perturbation (CSP) analysis can be used to study intermolecular interactions (RNA-RNA, RNA-protein or protein-protein), ligand or ion binding, as well as to investigate structural changes in biomolecules (e.g., RNA) under different environmental conditions (temperature, pH, pressure). CSP originate from changes in a chemical environment around the nuclei. The comparison of peak positions in a titration series can provide a lot of information about the character of interaction, binding affinity, exchange regime and give

insight about the localization of the binding site as well as reveal conformational or structural rearrangements.

A comprehensive theoretical overview of this approach for studies of proteins-ligand interaction is given by Williamson (Williamson 2013). Obviously, it can be also applied for CSP analysis in RNA. Here the most important ideas are summarized to give a brief introduction to the technique.

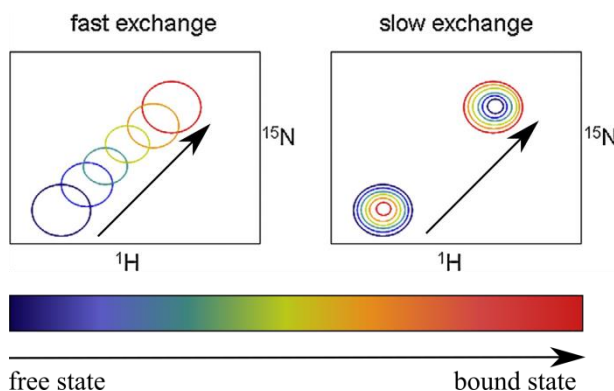
The reversible binding of the ligand (L) to the single site of RNA (R) with formation of RNA-ligand complex RL can be described as:



where  $k_{\text{on}}$  is the rate constant of the forward reaction and  $k_{\text{off}}$  is the rate constant of the back reaction. The dissociation constant  $K_d$  is defined as

$$K_d = \frac{[R][L]}{[RL]} = \frac{k_{\text{off}}}{k_{\text{on}}} \quad (44)$$

If the  $k_{\text{off}}$  rate constant is significantly slower (smaller) (in Hz) than the difference between the chemical shifts of the bound and free RNA, we talk about slow exchange rate modulus (**Fig. 2.3.11, right panel**). For such interaction regime, the intermediate states of the complex cannot be seen by the NMR. Upon the titration of the ligand the disappearance of the free RNA signal and the growth of the signal from the ligand-bound RNA will be observed. The intensity of each peak reflects the concentration of the molecule in one (free) or another (ligand-bound) state. Contrariwise, when the exchange rate is fast (weak interactions with  $K_d$  in  $\mu\text{M}$  range or above), the peak will steady and smoothly move, reflecting an averaged intermediate state of the free and bound RNA (**Fig. 2.3.11, left panel**). Finally, when the exchange rate is similar to the shift difference between bound and free states, the signal gets broad and shifts. In some cases, the  $k_{\text{on}}$ -rate is much slower than the binding to sterically available site, which results in a slow exchange even for the weak binding.



**Figure 2.3.11. Fast and slow exchange modes.** An overlap of 2D NMR spectra of a titration series, which represents two examples of the peak behavior upon the ligand titration in fast (left) and slow (right) exchange modes. The blue color represents the peak for the apo state of the macromolecule, and the red is for the holo state. The figure is adapted from (Williamson 2013)

For the systems in fast exchange, the  $K_d$  value can be extracted following averaged CSPs during the titration if the total RNA ( $R_t$ ) and total ligand ( $L_t$ ) concentrations are known. Both  $R_t$  and  $L_t$  can be represented as a sum of the free and bound equilibrium concentrations, respectively:

$$R_t = [R] + [RL] \quad (45)$$

$$L_t = [L] + [RL] \quad (46)$$

Observed chemical shift ( $\delta_{obs}$ ) in fast exchange regime is the weighted average of the free and bound states and can be expressed as:

$$\delta_{obs} = \delta_{free}f_{free} + \delta_{bound}f_{bound} \quad (47)$$

where  $f$  is the fraction of free and bound state.

Then the observed CSP ( $\Delta\delta_{obs}$ ) can be expressed as:

$$\Delta\delta_{obs} = \Delta\delta_{max} \frac{(R_t + L_t + K_d) - \sqrt{(R_t + L_t + K_d)^2 - 4R_tL_t}}{2R_t} \quad (48)$$

where  $\Delta\delta_{max}$  is a maximal CSP, when the RNA is completely saturated (usually this value cannot be measured experimentally and is obtained as part of the fitting procedure). Then, the dissociation constant  $K_d$  can be calculated by plotting the chemical shift value as a function of ligand concentration (titration curve) following the fitting procedure.

If the target-ligand complex interaction deviates from the fast exchange,  $K_d$  cannot be obtained from the titration curve, since the observed signal is not the weighted average of bound and unbound states. In slow exchange regime, the  $K_d$  can be elucidated by plotting the

intensity of the bound signal against the ligand concentration and fitting using the same equation.

The nuclei with large CSPs can be considered to participate in the binding. One or two standard deviation values are usually taken as a cut-off. However, the binding of the ligand can also introduce allosteric changes to the RNA residues far away from the binding site. Therefore, the number of binding sites and their affinity should be confirmed additionally by other methods.

## 2.4. Solid-state NMR

While solution-state NMR has an intrinsic molecular weight limit and cannot be applied to large biomolecules, solid-state nuclear magnetic resonance (ssNMR) can be in principle used to obtain structural information with atomic-level resolution on biomolecules of any size.

Since there is no fast isotropic molecular tumbling in solids, all anisotropic interactions are preserved. Though, these anisotropic interactions result in severe spectrum broadening, they also contain a lot of important structural and dynamic information. While dipolar coupling provides information about internuclear distances, quadrupolar interaction and CSA shed light on bonding and electronic structure. As a typical solid-state sample is a sample of low-order organization (powder or microcrystals), there is a uniform distribution of all possible molecular orientations, which results in a broad spectrum, which contains all possible resonance frequencies.

Since extremely broad static ssNMR spectra forbid any study, several approaches have been invented to make analysis of ssNMR spectra possible: magic angle spinning (MAS), decoupling and specific labeling.

The most important of these techniques, *MAS*, mimics the rapid molecular isotropic tumbling by rotation of the sample at the angle  $\text{ca } 54.74^\circ$  with respect to the orientation of the magnetic field,  $B_0$ . CSA, dipolar and first-order quadrupolar interactions have orientation dependence for an axially symmetric tensor as  $(3\cos^2\theta_{jk} - 1)$ , which turns to zero at MAS. Faster MAS eliminates stronger anisotropic interactions and today's technologies allow MAS spinning frequencies of 100 kHz and above. Heteronuclear *decoupling* techniques remove strong interactions that cannot be removed by MAS (e.g.,  $^1\text{H}$ - $^{13}\text{C}$ ,  $^1\text{H}$ - $^{15}\text{N}$  couplings during acquisition), while specific *labeling* simplifies spectra even further. Finally, information about spin connectivities is recovered using various *recoupling* techniques.

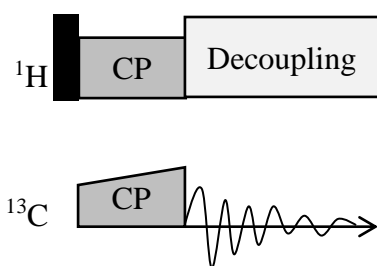
### 2.4.1. Cross-polarization experiment

Unlike in solution-state NMR, where  $^1\text{H}$  spectra have high importance, protons are not convenient to directly acquire in conventional, slow-to-moderate MAS ssNMR spectroscopy. Their high natural abundance and high  $\gamma$ -value lead to a large network of strongly coupled protons, which makes the  $^1\text{H}$  spectra very broad. For this reason, low  $\gamma$  nuclei with spin  $1/2$  ( $^{13}\text{C}$  or  $^{15}\text{N}$ ) are usually observed. Obviously, due to the superior gyromagnetic ratio of protons, the polarization is initially transferred from protons to less sensitive nuclei (e.g.,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ) to enhance the sensitivity of the experiment. Such basic experimental block, which is a solid-state NMR equivalent of solution-state NMR transfer INEPT, is called cross-polarization (CP) and is usually combined with MAS (Pines, Gibby, and Waugh 2003). The pulse sequence of a CP experiment is shown in **Fig. 2.4.1**. The transfer of the magnetization between the nuclei with distinct gyromagnetic ratios,  $\gamma$ , is achieved by irradiating the sample with the radiofrequency pulses that satisfy Hartmann-Hahn condition (Hartmann and Hahn, 1962), keeping the dipolar coupling active under MAS conditions:

$$\gamma_1 B_{1,I} = \gamma_2 B_{1,S} \pm n\omega_r \quad (49)$$

where  $n$  is an integer number,  $\omega_r$  is the MAS spinning rate, and  $B_{1,I}$  and  $B_{1,S}$  are the pulse strengths (nutating frequencies) for nuclei I and S, respectively. To achieve more efficient and/or selective transfer, the pulses with ramped or varying (composite) amplitude have been developed.

Obviously, CP transfer will work only if dipolar couplings are present, e.g., sample should be in the solid form.



**Figure 2.4.1. Basic 1D  $^{13}\text{C}$  cross-polarization pulse sequence.** Black narrow rectangle represents  $90^\circ$  pulse on the proton frequency channel. Next, the magnetization is transferred to heteronuclei (here,  $^{13}\text{C}$ ) by CP pulses, which match the Hartmann-Hahn condition. Finally, the  $^{13}\text{C}$  signal is acquired under high-power proton decoupling to attenuate line-broadening due to e.g.,  $^1\text{H}$ - $^{13}\text{C}$  heteronuclear decoupling.

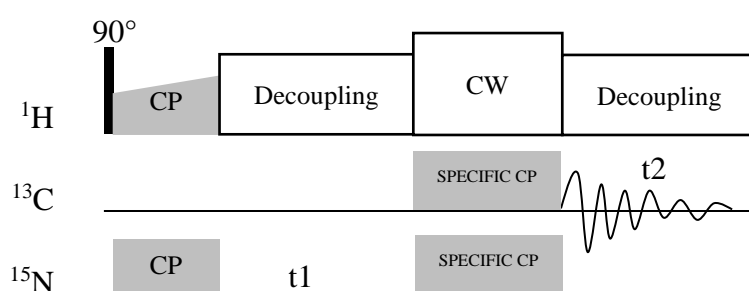
### 2.4.2. Double cross-polarization

For spectral assignment of  $^{13}\text{C}$ ,  $^{15}\text{N}$  labeled molecules double cross-polarization experiment is frequently used (BALDUS et al. 1998). Furthermore, frequency-specific transfer of



magnetization can be achieved via small-to-medium size RF pulses (1-25 kHz) under Hartmann-Hahn conditions and is referred as spectrally induced filtering in combination with CP (SPECIFIC CP).

In 2D SPECIFIC-CP  $^{13}\text{C}$ ,  $^{15}\text{N}$  correlation experiment (**Fig. 2.4.2**) during first cross-polarization step the  $^1\text{H}$  transversal magnetization, generated by  $90^\circ$  pulse is transferred from protons to  $^{15}\text{N}$ . Then,  $^{15}\text{N}$  chemical shifts are evolved during  $t_1$  and recorded. Then, during SPECIFIC-CP, magnetization is transferred to  $^{13}\text{C}$ , where it is detected during  $t_2$  under high-power  $^1\text{H}$ -decoupling.

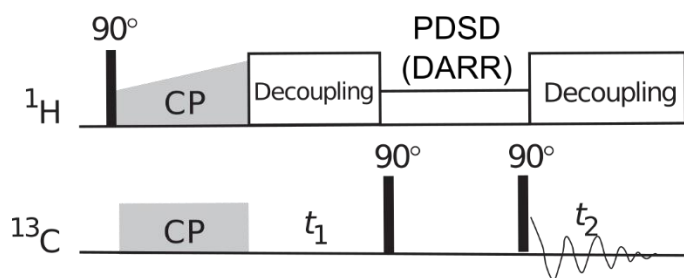


**Figure 2.4.2.** The pulse sequence scheme for 2D heteronuclear NC double CP experiment. CW is continuous-wave decoupling. See text for details.

### 2.4.3. Recoupling

#### *Proton-Driven Spin Diffusion*

The *recoupling* technique is used to recover dipolar interactions under MAS conditions. There are several ways to recouple spins. One of the most commonly used recoupling experiments is the homonuclear 2D proton-driven spin diffusion (PDS) (**Fig. 2.4.3**). In this experiment low- $\gamma$  nuclei magnetization is transferred from one nucleus (e.g.,  $^{13}\text{C}$ ) to all nearby  $^{13}\text{C}$  nuclei. In the first step, CP is used to create transversal  $^{13}\text{C}$  magnetization. After  $t_1$  evolution on carbons,  $^{13}\text{C}$  magnetization is brought back to the z-axis by a  $90^\circ$  pulse, where it then diffuses during the mixing time,  $t_{\text{mix}}$ , to other carbons. During spin diffusion, additional recoupling might be introduced, which could facilitate the spin diffusion (so-called DARR (Dipolar Assisted Rotational Resonance) experiment). The final  $90^\circ$  pulse brings carbon magnetization back to the xy-plane for the detection under proton decoupling. While the PDS experiment is easy to implement, typically it can provide only qualitative distance information.



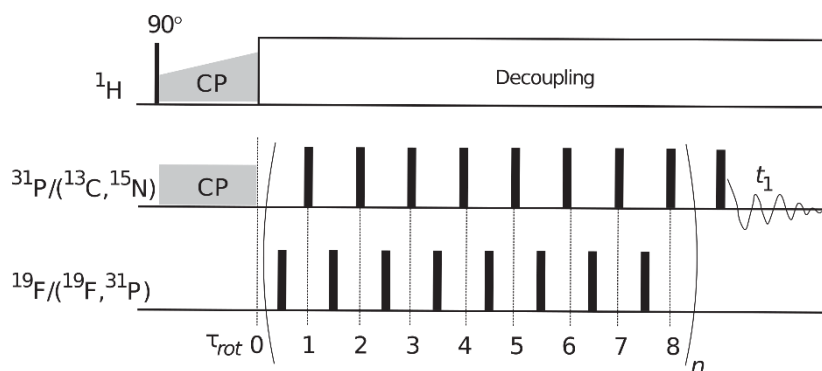
**Figure 2.4.3. The pulse sequence of 2D  $^{13}\text{C}$ ,  $^{13}\text{C}$  PDS (DARR) experiment.** The figure is adapted from (Marchanka and Carlomagno 2014).

Since dipolar couplings are distance-dependent, their accurate measurements can be used to determine internuclear distances. Homonuclear dipolar coupling can be recoupled by rotational resonance technique. Such resonance condition is achieved, when the frequency of sample spinning,  $\omega_r$ , is matched to a small integer,  $n$ , multiplied by the difference in the isotropic chemical shift frequencies of two spins, I and S:

$$n\omega_r = |\omega_I^{iso} - \omega_S^{iso}| \quad (50)$$

Spin diffusion-mediated multidimensional correlation (PDS) also can shed light on the inter-carbon or inter-proton distances.

Heteronuclear dipolar-coupled spins are usually recoupled by the application of RF pulses in rotor-synchronized manner to counteract the effect of MAS. Rotational-echo double-resonance (REDOR) is the simplest pulse sequence to recouple two spins and obtain the internuclear distances (Gullion and Schaefer 1989). In this experiment the intensity from echo-refocused evolution of the magnetization under both chemical shift and heteronuclear decoupling,  $S_0$ , is compared to that from the experiment, where several  $180^\circ$  pulses spaced by a half of a rotor period are applied to recover heteronuclear couplings, yielding intensity  $S$ . Then the ratio  $S/S_0$  gives relaxation-free dipolar dephasing curves, from which in a row internuclear distance can be calculated.



**Figure 2.4.4. Pulse sequence scheme for heteronuclear rotational-echo double resonance**

**(REDOR) based experiment** with alternating rotor-synchronized  $180^\circ$ -pulses on both the observing and dephasing channels. CP is a cross-polarization,  $\tau_{\text{rot}}$  is rotor period (Marchanka and Carlomagno 2014).

#### 2.4.4. Sample preparation for ssNMR

One of the main advantages of ssNMR is no need for a sample to be crystalline as it required for X-ray crystallography. Many objects can be studied in their native states: powders, pastes, membranes, fibrils, gels. Whereas certain biopolymers already have an intrinsic local order (e.g., fibrils or rigid binding sites), most of the biomacromolecules should undergo certain preparation to obtain a certain degree of local order.

There are four main preparation techniques: flash-freezing (Hu, Yau, and Tycko 2010), lyophilization (Huang et al. 1984), crystallization (Martin and Zilm 2003), and sedimentation (Bertini et al. 2011; Wiegand et al. 2020). While freezing and lyophilization are fast, resulting samples show spectra of poor resolution. Nano- or microcrystallization is typically the most appropriate form of biomolecule preparation that yields high-quality spectra in ssNMR. Hanging-drop or sitting-drop crystallization screening is routinely used to find good conditions for the microcrystallization. After optimal conditions are found, the sample is concentrated to 2-20 mg/ml, mixed with an appropriate crystallization solution in 1:1 ratio and then water is partially removed by evaporation.

Lately, sedimentation of the protein sample directly into the MAS rotor using the ultracentrifuge became a widely used approach to study proteins, as the method is robust, fast and yields resolution comparable to the crystalline samples (Lacabanne et al. 2019; Wiegand et al. 2020).

Since RNA molecules are very flexible, crystallization is often more challenging than for proteins. A simple ethanol precipitation procedure of RNA has been reported to provide spectra of comparable resolution to the crystalline sample (Zhao et al. 2019a).

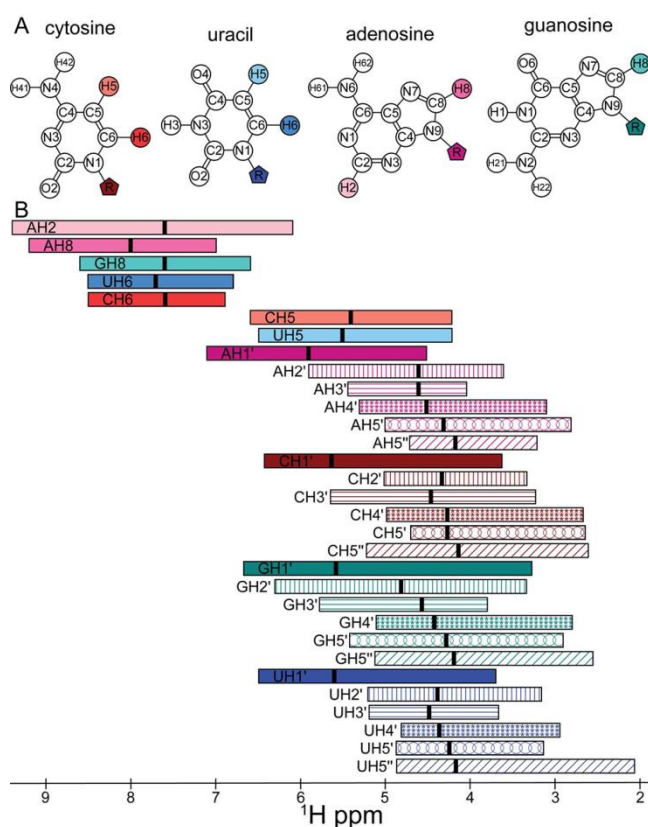
## 2.5. NMR of RNA

NMR remains the leading technique for studies of RNA structure and dynamics at the atomic-resolution level. Though, only 12.2% of the overall RNA structures available to date are obtained by NMR, it is a unique method, that allows studying dynamics of flexible functional RNA under almost physiological conditions (Marušič, Schlagnitweit, and Petzold 2019). Nevertheless, the studies of RNA by NMR are challenging due to poor chemical diversity of RNA nucleotides, with two purines and two pyrimidines, and poor chemical shift dispersion of ribose and non-exchangeable base protons, which results in significant spectra

crowding (**Fig. 2.5.1**) (Oleingski et al. 2021). Another limiting factor for solution-state NMR is the size of RNA of interest. As molecular weight increases, line-width becomes broader due to faster  $T_2$  relaxation. However, this limiting factor is not present in ssNMR due to the theoretical absence of molecular weight limitation.

In last two decades NMR spectroscopy of RNA has moved forward due to the availability of advanced isotopic labeling and development of novel heteronuclear experiments, alleviating the problem of poor proton dispersion.

Isotope-labeled RNA for NMR studies can be obtained by several methods. *In vitro* transcription using T7 RNA polymerase, DNA template and isotope-labeled rNTPs is the most robust and commonly used method to prepare RNAs of virtually any required size. It has its own advantages and disadvantages, which are described in chapter 5.2.1. Due to the relaxation-dependent line broadening, the maximal length of RNA that can be studied by solution-state NMR is around 130 nt (40 kDa); however, due to poor resonances dispersion and therefore severe spectral overlap, the realistic size of RNA that can be efficiently studied by solution-state NMR without any advanced approaches (e.g., perdeuteration, site-specific labeling) is only up to 50 nt (15 kDa). The larger RNA can be studied using segmental labeling strategies (Duss et al. 2010), atom-specifically labeled rNTPs (Alvarado et al. 2014; Longhini et al. 2016), utilizing post-transcriptional position-specific incorporation of labeled nucleotides (Büttner, Javadi-Zarnaghi, and Höbartner 2014) or sophisticated fragmentation-based segmental  $^2\text{H}$ -labelling approach (Keane et al. 2015). Apart from T7 *in vitro* transcription, chemical solid-phase RNA synthesis can deliver site-specific labeled RNA and/or incorporate a labeled nucleotide to any specific position, though the maximal length of RNA that could be obtained by the chemical synthesis is limited by 80 nt (Kremser et al. 2017).



**Figure 2.5.1. Chemical shifts of non-exchangeable RNA protons.** A) Chemical structure of four nucleobases, the ribose ring is depicted as pentagon “R”. B) Chemical shift ranges of non-exchangeable protons are represented as colored rectangles with a black bar, which indicates the average chemical shift value. The figure is adapted from (Kotar et al. 2020)

### 2.5.1. Information about base pairing

To access information about RNA folding or secondary structure of RNA, the base-pairing pattern should be elucidated. Imino protons, which typically undergo rapid solvent exchange, produce NMR resonances in 10-15 ppm  $^1\text{H}$  spectral area upon involvement in a hydrogen bond. A simple 1D proton spectrum can provide information not only about the number of the base pairs in the molecule but also about their chemical nature, since solvent-protected imino protons resonate at distinct frequencies depending on their electronic environment. For example, the guanosine imino proton induces a signal at 12–13.5 ppm, when involved in a G:C Watson-Crick (WC) base pair, and at 10–12 ppm if involved in G-U base pair; uridine has its imino proton peak in a range of 13-15 ppm if involved in A:U WC base pair (Fürtig et al. 2003).

The most familiar NMR experiment for the detection of base-pairs is  $^1\text{H}$ - $^1\text{H}$  NOESY, as it can correlate the protons within 5 Å distance. The sample doesn't need to be isotopically labeled and the experiment is relatively robust and fast. In helical RNA imino-imino NOEs can be observed between strands and sequential nucleotides, which sometimes can lead to ambiguity in secondary structure determination. Furthermore,  $^1\text{H}$ ,  $^1\text{H}$  NOESY spectrum might not show all base pairs due to their elevated dynamics which can broaden the cross-

peak beyond detection. Therefore, the spectrum is often recorded at lower temperatures to stabilize dynamic base pairs. This approach might be challenging as (i) lines get broader due to slower tumbling at lower temperatures and (ii) sometimes it is difficult to correlate temperature-shifted peaks to the peaks obtained in other NMR experiments acquired at higher temperatures. Nonetheless, several large RNAs have been successively assigned via  $^1\text{H}$ - $^1\text{H}$  imino NOE correlations as it provides not only inter- but also intracatenar information and sequential assignment walk (Burke et al. 2012; Imai et al. 2016; Reining et al. 2013).

Another favored experiment that provides information about base-pairs in RNA is HNN-COSY, which directly detects hydrogen bonds utilizing the through-space J-coupling between  $^{15}\text{N}$  imino donor nuclei and the corresponding acceptor  $^{15}\text{N}$  of the base (N1 for A/G with N3 of U/C) (Dingley and Grzesiek 1998). This experiment paved the way for the development of more advanced pulse schemes, e.g., modified quantitative  $^2\text{J}_{\text{HN}}$  HNN-COSY experiment, which correlates nitrogen atoms via non-exchangeable base protons and thus allows the direct observation of unstable base pairs which lack NOE imino signal or detection in 100%  $\text{D}_2\text{O}$  (Hennig and Williamson 2000). Another sensitivity-optimized sequences developed by the Sattler group are pyrimidine BEST-selective long-range (BESTsellr) HNN-COSY for A-U base pairs and (Py) H(CC)NN-COSY for simultaneous detection of A:U and G:C WC base pairs. BESTsellr can be run at higher temperatures and detect both unstable and isolated base pairs.

### 3. Aims

In this study, I aimed to elucidate the structural bases of miR-122 function in the HCV life cycle and understand the conformational changes that occur in the HCV 5' UTR upon miR-122 binding at three different levels of structural complexity (domain I, domains I-II and the entire 5' UTR) using an integrative structural biology approach.

It is known that two copies of miR-122 bind seed and auxiliary regions on domain I and that magnesium ions are indispensable for this interaction (Mortimer and Doudna 2013). First, we sought to confirm the stoichiometry of the domain I-miR-122 complex and obtain structural details of the interaction using solution-state NMR. By performing a series of NMR titration experiments (titrating unlabeled miR-122 onto labeled domain I), I aimed to understand the dynamics of their interplay under near-physiological conditions and also under varying magnesium concentrations.

Next, I set out to explore the influence of domain II on domain I itself and the domain I-miR-122 interaction. As discussed in the Introduction, one of the proposed mechanisms of miR-122 action is a shift from the energetically more favorable conformation of domain II, incorporating the single-stranded region of domain I, so-called microRNA binding region (MBR), within the functional fold, in such a way that IRES of HCV becomes functional and drives translation of the viral genome (Schult et al. 2018). However, this hypothesis has not been confirmed on the structural level. Using solution-state NMR, I wanted to compare the secondary structure and follow the chemical-shift perturbations of isolated domain II and the domain I-II construct upon addition of miR-122, to obtain information on the structural changes in the domain I-II construct induced by the binding to miR-122. To study conformational changes upon binding miR-122 on a global scale, I aimed to carry out titrations on the domain I-II construct using contrast-matching SANS. Next, I constructed structural models based on the experimental data and using molecular dynamics simulations. Finally, I set out to investigate how miR-122 affects long-range interactions between domain I and IRES in the 5' UTR. Small-angle X-ray scattering in line with SEC can provide low-resolution information about the 5' UTR conformation in the apo and holo states. I also assessed the feasibility of structural studies on the 5' UTR using ssNMR, in particular using the segmental labeling approach (Duss et al. 2010) to selectively observe separate domains.





## 4. Material and Methods

### 4.1. Cloning

Each DNA construct for the subsequent *in vitro* T7 transcription of RNA was designed with EcoRI restriction site at 5' end and HindIII restriction site at 3' end for cloning into a plasmid, class III T7 promoter precluding DNA sequence and PstI restriction site for plasmid linearization (**Fig. 4.1.1**).



**Figure 4.1.1. Schematic representation of typical DNA construct used for *in vitro* transcription.**

DNA template for 5' UTR (359mer) of HCV genotype 1b was ordered from GENEWIZ (Sigma-Aldrich) as a synthetic gene in a recombinant pUC57 vector having a kanamycin resistance. The sequence was as in (Mortimer and Doudna 2013) (**Extended data. Fig. 1 and 2**). The coding- and non-coding templates for domain I, binding site 1, extended binding site 1 (34mer), miR-122-HH were ordered at HPLC-grade quality from Sigma as oligonucleotides with cloning overhangs for the cloning into pUC19 vector, T7 promoter at 5' end of DNA sequence and PstI restriction site at 3' end (**Extended data. Table 1**).

#### 4.1.1. Plasmid linearization

For the subsequent re-cloning pUC 19 plasmid was linearized with EcoRI and HindIII restriction enzymes (New England Biolabs, NEB). A typical 50 µl reaction mix contained: 1 µg of plasmid DNA, 1x rCutSmart Buffer (NEB) and 20 units of EcoRI-HF and HindIII-HF restriction enzymes each. The reaction was incubated at 37°C for 15 min. Next, the reaction was stopped by the addition of 6x purple DNA loading dye (NEB) and purified on the 1% TBE-agarose gel electrophoresis with addition of SYBR Safe™ dye (ThermoFisher) to visualize the DNA bands under UV light. The band was cut from the gel and extracted with QIAquick Gel Extraction Kit (QIAGEN).

#### 4.1.2. Insert annealing

DNA oligonucleotides were dissolved in Milli-Q water to 100 µM concentration, non-coding and coding strands were mixed in 30 µl annealing buffer (10 mM Tris, pH 7.5 - 8.0, 50 mM NaCl, 1 mM EDTA) at 2 µM concentration and annealed at 95 °C for 3minutes, then slowly cooled down to room temperature.

#### 4.1.3. Ligation of the insert into the plasmid

Annealed oligonucleotides were then ligated into linearized pUC 19 vector: 20 µl reaction contained 0.5 µM of annealed oligonucleotides, 50 ng of linearized pUC19, 1x ligation buffer (NEB), 2 units of T4 DNA ligase (NEB). The reaction was incubated at room temperature for 2 hours.

#### 4.1.4. Heat-shock transformation

The vector with DNA template for the RNA production was transformed into chemically competent Top10 *E.coli* cells by heat-shock method. 50 µl of glycerol stock cells (stored at -80°C) was thawed on ice. 10 µl of the ligation reaction or 5 ng of synthetic gen plasmid were added to the tube with cells and gently mixed. Tube was incubated on ice for 30 minutes, heat-shocked for 45 seconds at 42°C and then immediately placed back on ice for at least 2 minutes. 800 µl of LB media was added to the mixture and the tube was then incubated at 37 °C with shaking for 2 hours. Afterward, the cells were spun for 1 min at 5000 rpm and supernatant was discarded. The pellet was resuspended in ca. 50 µl of medium and uniformly distributed over the LB-agar plate containing appropriate antibiotic (kanamycin for pUC57 vector or ampicillin for pUC19 vector). The plate was incubated at 37 °C overnight. Few single colonies were picked and resuspended in 6 ml LB medium and grown during 5-7 hours at 37 °C in a shaker at 200 rpm. 5 ml of cells were used for QIAGEN miniPrep to extract the plasmid for sequencing with Mix2Seq Kit (Eurofin genomics) using M13 sequencing primers (**Extended data. Table 2**). The remaining volume of cells was mixed with glycerol at 1:1 ratio and stored at -80 °C for the subsequent big-scale plasmid preparation.

## 4.2. Mutagenesis

The mutagenesis protocol used in this study was adapted from Q5 Site-Directed Mutagenesis Protocol (NEB) and was utilized to perform modifications on 5' UTR and miR-122 constructs. First, primers for insertion and/or substitution were designed using NEBaseChanger tool (**Extended data. Table 3**) and ordered from Sigma. Next, 50 µl PCR reaction with primers, was set up according to the Q5 High-Fidelity DNA Polymerase (M0491) protocol (NEB). The reaction contained 1x Q5 Buffer, 1.25 µM of forward and reverse primers, 0.5 mM NTPs, 1 unit of Q5 polymerase and 10 ng of plasmid DNA. The reaction was set up on ice and performed in T100 Thermal Cycler (BioRad). The PCR conditions are summarized in **Table 4.2.1**.

Cycle step	Cycle s	Temperature	Time
<b>Initial denaturation</b>	1	98 °C	30 s
<b>Denaturation</b>	25	95 °C	10 s
<b>Annealing</b>		T <sub>a</sub>	25 s
<b>Extension</b>		72 °C	20 s
<b>Final extension</b>	1	72 °C	120 s
<b>Hold</b>	1	4 °C	∞

**Table 4.2.1. The PCR conditions for a Site Directed Mutagenesis PCR with Q5 Polymerase.** T<sub>a</sub> indicates annealing temperature.

Afterwards, the PCR reaction was mixed with a cocktail of enzymes: DpnI restriction enzyme (NEB) to degrade the template DNA plasmid, T4 kinase (NEB) to phosphorylate 5′ end of PCR product, and T4 ligase (NEB) to ligate PCR product into plasmid DNA. A one-pot reaction was set in 20 μl volume containing 2 μl of PCR mix, 1x CutSmart Buffer (NEB), 10 mM DTT, 1 mM ATP, 10 units of T4 DNA ligase, 0.25 units of T4 Polynucleotide kinase and 0.5 U of DpnI restriction enzyme. The reaction was performed at room temperature for 20 minutes. The whole reaction mix was then used for the transformation into heat-shock competent Top10 *E. coli* cells as described above.

#### 4.2.1. 5′ Hammerhead ribozyme insertion in cis

Hammerhead ribozyme in *cis* was inserted at 5′-end of 5′-UTR construct to improve 5′ homogeneity. The Hammerhead ribozyme is co-transcriptionally cleaved by an intramolecular transesterification reaction (Epstein and Gall 1987). This construct did not work well for the full-length 5′ UTR due to the incomplete cleavage and the difficulty to separate cleaved and uncleaved RNA. However, this construct has shown good performance for HH-domains I-II DNA amplification (**Extended data. Fig. 3 and 4**). The primers “HH-UTR F” and “HH-UTR R” (**Extended data. Table 3**) were used to insert HH sequence:

5′-GGGGGCTGGCCTGATGAGTCCGTGAGGACGAAACGAGCTAGCTCGTC-3′.

#### 4.2.2. miRcopy plasmid construction

The plasmid construct utilized for the production of miRcopy RNA (pre-miR-122 construct) was generated from pUC19 plasmid contained miR-122 sequence and 3′ extension for cleavage with hammerhead ribozyme in trans (miR-122-HH) (**Extended data. Fig. 4**). miR-122-HH did not work optimally due to unnatural nucleotides at both 5′ and 3′ ends and suffered from the low yield as a result of double purification for the cleavage in *trans*.

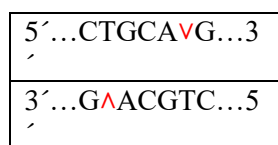
Therefore, new miRcopy construct was generated to yield miR-122 as described (Feyrer et al. 2020).

miRcopy construct was produced by two consecutive site-directed mutagenesis steps. In the first step, an optimal initiation sequence (5'-GGAAAG-3') (Milligan et al. 1987), the last 8 nucleotides of miR-122 sequence (5'-GTCTGTCGCC-3') and one copy of miR-122 sequence (5'-TGGAGTGTGACAATGGTGTGTTTGT-3') were inserted upstream the 5' end of miR-122 sequence using the Q5 SDM protocol described above (**Extended data. Fig. 5**). Primers "5' insertion F" and "5' insertion R" (**Extended data. Table 3**) were used. After the transformation, plasmid sequencing, and amplification, this new plasmid was used as a template for the substitution of the HH-extension and another insertion at the 3' end of miR-122 sequence, which contained one additional copy of miR-122 and the first 7 nucleotides of miR-122 (5'-TGGAGTG-3') at the 3' end. Primers "5' subins F" and "5' subins R" (**Extended data. Table 3**) were used. By this step, eleven more copies of miR-122 were inserted (**Extended data. Fig. 6**).

### 4.3. DNA template production and purification

#### 4.3.1. DNA template production in *E.coli* cells

For the transcription of most of the short RNAs plasmid DNA as a template was used, which was amplified in Top10 *E. coli* cells. Typically, PstI restriction site was used to linearize the plasmid. The PstI restriction site leaves terminal guanosine at the 5' end of the non-coding (template) DNA strand (**Fig. 4.3.1**). Thus, the produced RNA always bears an extra cytidine at 3' end (**Extended data. Table 4**).



**Figure 4.3.1. PstI restriction site.** After cleavage with PstI restriction enzyme (NEB), 5' terminal guanosine remains on template DNA strand.

1 ml of glycerol stock cells was thawed on ice and resuspended in 1L LB medium with appropriate antibiotic (ampicillin or kanamycin). The media was incubated at 37 °C while shaking at 200 rpm overnight. The cells were harvested by centrifugation at 4500 rpm at 4 °C for 30 minutes. Plasmid DNA was extracted and purified using Qiagen Plasmid Mega or Giga Kit (Qiagen). DNA concentration was measured with NanoDrop Spectrophotometer.

The circular plasmid DNA was linearized with PstI restriction enzyme (NEB). Typical linearization reaction of e.g., 5 mg DNA was performed in 5 ml volume and contained: 500

μl 10x Buffer 3.1, 120 μl PstI enzyme, 4400 μl of DNA + Milli-Q H<sub>2</sub>O. The reaction was incubated at 37 °C overnight.

Restriction enzyme was removed from the reaction by phenol/chloroform/isoamyl alcohol (Carl Roth) extraction followed by the chloroform/isoamyl alcohol (Carl Roth) extraction. Afterwards, DNA was precipitated by addition of 2.3 volumes of absolute ethanol and 0.1 volume of 5 M NaCl. The mix was incubated at -20°C for at least 2 hours to precipitate DNA. As a next step, DNA precipitate was spin-down by centrifugation at 8500 rpm for one hour, the supernatant was then discarded and the pellet was washed with 70% ethanol to remove residual salt, and then spin-down for another hour at 8500 rpm. The pellet was dried at room temperature or 37°C and then resuspended in Milli-Q water typically at 1 mg/ml concentration and stored at -20 °C until it was used for RNA transcription. DNA concentration was measured by NanoDrop Spectrophotometer.

#### 4.3.2. DNA template production by PCR

PAGE purification of large RNAs cannot provide single nucleotide resolution. Introduction of two 2'-O-methylated nucleotides at the 5'-end of template DNA strand can significantly improve 3'- homogeneity of the transcribed RNA (Kao, Zheng, and Rüdiger 1999). Thus, DNA templates of some of the RNA constructs were produced by PCR amplification using Phusion High-Fidelity DNA Polymerase (NEB). The template for the PCR reaction was a DNA plasmid with an appropriate sequence insertion. The reaction was set up on ice and performed in T100 Thermal Cycler (BioRad). For the preparative scale PCR the complete reaction volume (typically 10 ml) was split in 100 μl per PCR tube. The typical reaction mix contained 1x Phusion HF buffer, 0.2 mM NTPs, 0.5 μM of forward (F) and reverse (R) primers, 3 % DMSO, 2 pg/ml of plasmid DNA and Phusion HF Polymerase (20 units/ml). The typical PCR cycling conditions are summarized in **Table 4.3.1**. Primers and their annealing temperatures are listed in **Extended data. Table 5**. Next, the whole PCR reaction mix was used for the subsequent transcription reaction.

Cycle step	Cycles	Temperature	Time
<b>Initial denaturation</b>	1	98 °C	30 s
<b>Denaturation</b>	25-35	98 °C	5 s
<b>Annealing</b>		45-72 °C	15 s
<b>Extension</b>		72 °C	15 s
<b>Final extension</b>	1	72 °C	10 min
<b>Hold</b>	1	4 °C	∞

**Table 4.3.1. PCR cycling conditions for a Routine PCR with High Fidelity Phusion Polymerase**

## 4.4. RNA transcription and purification

With some exceptions, when RNA was ordered from Integrated DNA Technologies (IDT) with HPLC grade purification (miR modified), all RNAs in this study were produced in-house by in vitro transcription reaction with T7 polymerase (class III) promoter. All RNA constructs used in the current study are listed in **Extended data. Table 4**. The stocks for unlabelled NTPs used for transcription reaction were prepared from ATP, GTP, CTP and UTP obtained in lyophilized form (Carl Roth) by dissolving in Milli-Q water to the concentration of 100 mM, pH was adjusted to 7.0 with 1 M NaOH. Labeled ( $^{13}\text{C}$ ,  $^{15}\text{N}$ , and  $^2\text{H}$ ) NTPs were purchased from Silantes in solubilized form as lithium salt at 100 mM concentration buffered at pH 7.0. Typically, rNTPs were mixed at the ratio 1:1:1:1, but in some cases (miRcopy), when the nucleotide population in the construct was non-uniform, the NTPs mixture was customized (40 % Ura, 39 % Gua, 17 % Ade, 4 % Cyt) to improve the reaction yield.

The transcription reaction mix contained 1x transcription buffer (TB) (40 mM Tris-HCl, 1 mM spermidine, 5 mM DTT, 0.01% (v/v) Triton X-100, pH 8.0), various amounts of linearized plasmid or PCR mix as DNA template, rNTPs,  $\text{MgCl}_2$ , 0.1 % vol of PEG8000, T7 polymerase. 20% DMSO was added to the reaction mix to improve both reaction yield and 3' homogeneity of the RNA product (Helmling et al. 2015).

Since the yield of the transcription reaction is highly dependent on the salt and buffer content, every reaction was optimized on a 20  $\mu\text{l}$  scale in 96-well microcrystallization plates in two steps. In the 1<sup>st</sup> step, concentrations of  $\text{MgCl}_2$  (20 mM  $\rightarrow$  50 mM) and rNTPs (10 mM  $\rightarrow$  40 mM) were optimized, while the concentrations of plasmid DNA (25  $\rightarrow$  200 ng/ $\mu\text{L}$ ) or PCR template (5%  $\rightarrow$  15% vol.) and in house produced T7 polymerase (25  $\rightarrow$  100 ng/ $\mu\text{L}$ ) were optimized in the 2<sup>nd</sup> step. 20  $\mu\text{l}$  reactions were carried out for 2-3 hours at 37  $^\circ\text{C}$ , stopped by addition of 2x denaturing RNA loading dye (2xTris/Borate/EDTA (TBE) buffer, pH 8.0, 8M UREA, 2.5% (w/v) bromophenol blue, 2.5% (w/v) xylene cyanol) and loaded on analytical size UREA-PAGE (1xTBE pH 8.0, 6M UREA, ROTIPHORESE®NF-Acrylamide/Bis-solution 40 (29:1)) and run in 1xTBE buffer. Small RNAs (length < 50 nt) were typically run at 15W on 20x25cm gel using in house made vertical gel-electrophoresis chamber for 2.5-3 hours, larger RNA could be run at mini 8x10 cm gels in Mini-PROTEAN Tetra Vertical Electrophoresis Cell (BioRad) at 5 W for 30-45 minutes. To achieve optimal resolution, the percentage of the gel was adjusted to the size of the RNA product (**Table**

**4.4.1).** RNA bands were stained by ethidium bromide or SYBR® Safe dye and visualized on Gel Doc XR+ Gel Documentation System (BioRad).

<b>RNA length,nucleotides</b>	<b>Gel percentage</b>
< 30 nt	20%
31-60 nt	15%
61-100 nt	12%
101-150 nt	10%
151-500 nt	8%

**Table 4.4.1 Percentage of acrylamide in the denaturing UREA-PAGE for the optimal separation depending on the RNA size.**

The conditions with the best yield were used to set up a 5-10 ml reaction, which was carried out for 5-6 hours at 37 °C. To remove pyrophosphates, which can slow down the transcription rate, thermostable inorganic pyrophosphatase (TIPP) (2000 U/mL, NEB) was added in 1:1000 ratio. Large scale reaction was stopped by addition of 0.1 volume 0.5 M EDTA, concentrated by evaporation in a speedVac or using Amicon® Ultra Centrifugal Filters (Merck) down to 2-3 ml, mixed with 2x denaturing loading dye and run on preparative scale gel using in-house made vertical gel-electrophoresis chamber for 8-12 hours.

RNA was visualized by UV-shadowing and extracted from the gel by crush and soak procedure (Petrov et al. 2013). Briefly, gel slices, containing RNA were cut out, crushed with a 5 ml sterile plastic syringe into 15 ml falcons, soaked with extraction buffer (40 mM MES, 10 mM EDTA, pH 6.0), frozen at -80 °C for 30 minutes and left on a rotary horizontal shaker overnight at 4 °C. The supernatant containing RNA was separated from gel pieces by centrifugation at 4000 rpm at 4 °C for 30 minutes, collected and filtered with a sterile 0.22 µm syringe filter. Pelleted gel slurry was washed again with extraction buffer, left to shake at 4 °C for 1-2 hours, centrifuged and the second portion of the supernatant was also filtered. RNA in extraction buffer was mixed with 2.3 volumes of absolute ethanol and 0.1 volume of 5 M NaCl. The mix was incubated at -20°C for at least 2 hours or overnight to precipitate RNA. In next step, RNA precipitate was spin-down by centrifugation at 8500 rpm for 1.5 hours, the supernatant was discarded and the pellet was washed with 70% ethanol to remove residual salt, spin-down for another hour at 8500 rpm. The pellet was dried at room temperature or 37°C and then resuspended in Milli-Q water. RNA concentration of the RNA was measured by NanoDrop Spectrophotometer. RNA was stored at -20°C. The purity of the RNA was checked on an analytical gel.

While most of the RNA constructs used in this study, were produced and purified in a straightforward one-pot reaction as described above, some constructs (domains I-II and miR-

122) were obtained from pre-RNAs constructs, which allowed to improve both reaction yield and homogeneity of RNA 5' and 3' ends.

#### **4.4.1. Domains I-II.**

Domains I-II RNA (118mer) was produced from the construct, which contains self-cleaving hammerhead (HH) ribozyme *in-trans* at the 5' end (HH-118mer). A preparative scale transcription reaction was carried out as described above in a 5 ml volume. After reaction has proceed for 5 hours, the reaction was terminated by addition of 0.1 V of 0.5 M EDTA, pH 8. Then, the transcription reaction mix was washed 4 times with Milli-Q water to remove EDTA and MgCl<sub>2</sub> on Amicon® Ultra 15 mL Centrifugal Filters with 10 kDa MWCO. 10x cleavage buffer and Milli-Q water were added up to 4 ml to yield the final concentration of 40 mM Tris pH 7.4, 0.1 M NaCl, 2 mM EDTA. The reaction mix was heated up for 6-8 minutes at 95 °C and then cooled down at the bench for 15-20 min. Then, MgCl<sub>2</sub> was added to yield final concentration of 50 mM. The reaction mix was incubated at 37 °C overnight, concentrated to 2 ml and purified on the preparative scale 10 % denaturing PAGE.

#### **4.4.2. miR-122**

miRcopy (322mer) transcription reaction was set with a customized rNTPs mix (40 % UTP, 39 % GTP, 17 % ATP and 4 % CTP) and 20% DMSO. This customized rNTPs mix has improved RNA yield from a 0.2-0.5 mg/ml-reaction to 4 mg/ml-reaction. DNA traces were removed by weak anion-exchange chromatography using a 5 ml DEAE column with a long salt gradient. This step was critical to attenuate RNA degradation during the subsequent site-specific RNaseH chimera-guided cleavage. Buffers A and B both contained 50 mM sodium phosphate, pH 6.5 and 0.2 mM EDTA; buffer A contained 150 mM NaCl, while buffer B contained 2 M NaCl, and the gradient of 10-30% buffer B in 100 ml was applied. The DNA-less fractions were concentrated to 50 µM for subsequent cleavage with RNase H as described below followed by further PAGE purification. The DNA-containing fractions were also concentrated and added to the next batch of transcription purification to improve overall reaction yield.

#### ***RNA annealing and complex formation***

To ensure the proper folding of RNA, optimized annealing procedure and exchange to the appropriate buffer was carried out before subsequent structural studies.

**Domain I (48mer), binding site 1 (30mer) and 34 mer.** Purified RNA was washed 3 times on Amicon® Ultra 15 mL Centrifugal Filters with 3 kDa MWCO first with Milli-Q water



and then with buffer (100 mM HEPES pH 7.5, 100 mM KCl). RNA was annealed at 100 µg/ml concentration to avoid potential dimerization for 5 minutes at 95 °C, followed by a slow cool-down to room temperature. Afterward, RNA was back concentrated to milligrams/ml concentration and MgCl<sub>2</sub> was added up to 5 mM concentration.

**Domains I-II (118mer)** RNA was annealed following the same procedure described above for domain I, but the final buffer composition was 100 mM HEPES pH 7.5, 10 mM KCl, 5 mM MgCl<sub>2</sub>.

**5' UTR (359mer)** was washed 3 times on Amicon® Ultra 15 mL Centrifugal Filters with 30 kDa MWCO and annealed in H<sub>2</sub>O at 95 °C for 5min at 1200 µg/ml concentration with the following addition of ice-cold 2x buffer. The final buffer composition was 100 mM HEPES, pH 7.5, 100 mM NaCl, 5 mM MgCl<sub>2</sub>.

**miR-122** was washed 3 times on Amicon® Ultra 15 mL Centrifugal Filters with 3 kDa MWCO and buffer exchanged to the same buffer as HCV RNA construct.

The HCV RNA–miR-122 complex formation was carried out by incubation of corresponding HCV RNA with miR-122 in an appropriate ratio for 20-30 minutes at 37 °C.

#### 4.5. Site-specific RNaseH cleavage guided by 2'-O-methyl-RNA/DNA chimera

Site-specific RNaseH cleavage was used for the cleavage of miRcopy RNA for the miR-122 production as described by (Feyrer et al. 2020) and also for the site-specific cleavage of HCV 5' UTR constructs for the segmental labeling approach (Duss et al. 2010). RNA-DNA chimeras for cleavage were designed such that the 4-nucleotides DNA fragments were flanked by two 2'-O-methyl-RNA stretches of 4-15 nucleotides length (**Extended data. Table 6**) and were purchased from Sigma at HPLC-grade quality. The optimal conditions for the cleavage reaction for each construct were determined on a small 20 µl scale reaction by optimization RNA/chimera ratio, RNaseH concentration and reaction temperature. First, RNA at 5 µM concentration was annealed together with chimera at the optimal ratio (5 to 100%) at 90 °C for 2 minutes with a following cool down to room temperature. Then, after annealing, MgCl<sub>2</sub> at 10 mM concentration and RNaseH enzyme (2.5 to 10 ng/µL) were added, and the mixture was incubated at 37/20 °C for 1 hour. The reaction was stopped by the addition of 0.1 volume 0.5 M EDTA, pH 8.0 and 2x denaturing dye and loaded on the PAGE for analysis. For the big-scale reactions concentration of RNaseH was scaled down 5 times to attenuate unspecific cleavage.

## 4.6. Ligation of RNAs

Ligation of RNAs was carried out using complement DNA splint, which brings 3'-OH and 5'-monophosphate ends of RNAs close by (**Extended data. Table 7**). The splints were purchased as oligonucleotides with HPLC-grade purification from Sigma. The ligation reaction was optimized on 20 µl scale reaction by varying RNAs concentration (3.3 µM/10 µM), and reaction time (1/5/12 h) and then performed at a large scale. The best yield of ligated product was obtained in the reaction contained 10 µM RNAs, 15 µM DNA splint and 2.6 units of T4 DNA ligase (NEB) per 1 pmol of RNA, while it was incubated at 37 °C for 5 hours.

In the overall protocol for RNA ligation, RNAs were first annealed with appropriate DNA splint at a ratio 1:1:1.5 (10 µM:10 µM:15 µM) in the annealing buffer containing 25 mM HEPES, pH 7.5, 15 mM NaCl, 0.1 mM EDTA at 80 °C for 2 minutes with subsequent slow cool down to room temperature. Then, T4 DNA ligase buffer and T4 DNA ligase were added, and the reaction was incubated at 37 °C for 5 hours. Finally, RNA was purified on the denaturing PAGE.

## 4.7. Size-exclusion Chromatography and Multi-angle Light Scattering

Purification of the HCV RNA-miR-122 complexes and RNAs purification for SANS was carried out by SEC on Äkta pure system (Cytiva). As a reference, each RNA was analysed in its apo-state. 5' UTR (359 mer) and domains I-II (118mer) were analysed on a Superdex 200 Increase 10/300 GL SEC column (Cytiva), while domains I and binding site 1 constructs were analysed on a Superdex 75 Increase 10/300 GL SEC column (Cytiva).

Molecular weight of RNA and RNA-RNA complexes was measured by MALS using MiniDawn TREOS system (Wyatt Technologies) and Optilab T-rex refractive index detector (Wyatt Technologies) coupled to Äkta Pure system. Prior to each measurement, the system was calibrated with BSA protein in the RNA buffer. Data analysis was performed with the ASTRA 7.0 software package (Wyatt Technologies).

## 4.8. Electrophoretic mobility shift assays (EMSA)

EMSA were performed on domain I RNA (48mer) and binding site 1 of domain I (30mer) titrated with miR-122. RNAs were synthesized by T7 *in vitro* transcription and purified as described above and exchanged to the buffer containing 100 mM HEPES pH 7.5, 100 mM

NaCl and optionally 5 mM MgCl<sub>2</sub>, depending on the experimental setup. Before addition of miR-122, 48mer and 30mer at 15 μM concentration were heated to 95 °C for 3 mins and cooled to room temperature on the bench. After addition of miR-122 in molar ratios of 0, 0.5, 1.0, 2.0 and 3.0, RNA complexes were incubated at 37 °C for 30 mins in 5 μl volume. Equal volume of gel-loading dye (30% glycerol, 0.5 TBE, and optionally 5 mM MgCl<sub>2</sub>) was added prior to the loading on the gel. Non-denaturing gel electrophoresis (10.0 x 8.0 cm) was performed at 4 °C at 120 V for 2.5 hours in 0.5x TBE buffer, optionally containing 5 mM MgCl<sub>2</sub>. RNA was visualized by staining with ethidium bromide.

## 4.9. NMR

### 4.9.1. Solution-state NMR

Annealed RNA sample was transferred into a 5 or 3 mm Wilmad NMR tube. RNA concentrations varied from 100 to 400 μM for different samples. For the experiments recorded in H<sub>2</sub>O-based buffer, 10% D<sub>2</sub>O with 0.05 wt. % 3-(trimethylsilyl) propionic-2,2,3,3-d<sub>4</sub> acid, sodium salt (TMS) (Sigma) was added. The assignment experiments were carried out in 100% D<sub>2</sub>O-based buffer. The experiments were acquired on a 850 MHz Bruker Avance III HD NMR spectrometer (NMR850) equipped with TCI cryoprobe or a 600 MHz Bruker Avance III HD NMR spectrometer (NMR600) equipped with CPP3 probe at 298K or 308 K if not specified otherwise. All data was acquired with TopSpin 3.5 pl 2 software (Bruker Biospin, Rheinstetten, Germany) and analyzed with CcpNMR Analysis 2.4.2 (Vranken et al. 2005).

*Base pair elucidation* was carried out via **2D <sup>1</sup>H-<sup>1</sup>H NOESY experiment** with a mixing time of 300 ms in aqueous buffer. The spectra for sequential walk between imino-protons were recorded at 298 K, while the correlation spectra of guanosine H1 protons to the cytidine H5 protons were recorded at 308 K. (Fürtig et al. 2003). The same 2D <sup>1</sup>H-<sup>1</sup>H NOESY experiment was used to study domains I-II and domain II secondary structures via chemical shift perturbation analysis (Williamson 2013).

The chemical shift distance (CSP) was calculated as:

$$\Delta\delta = |\delta_1 - \delta_2| \quad (51)$$

Uncertainty for the CSP,  $\sigma(\Delta\delta)$  can be calculated as:

$$\sigma(\nu) \approx 0.7 \left( \frac{LW}{S/N} \right) = 0.7 \left( \frac{LW * RMS(noise)}{Peak\ height} \right) \quad (52)$$

where  $\delta$  is a chemical shift value in ppm,  $\sigma$  is chemical shift uncertainty,  $\nu$  is a resonance frequency, RMS is a Root Mean Square, LW is a Line Width.

*Chemical shift perturbation mapping* was performed by **2D  $^1\text{H}$ ,  $^{13}\text{C}$  constant time (CT) HSQC** experiment, which was recorded according to a standard protocol (Santoro and King 1992). This experiment correlates all  $^1\text{H}$ - $^{13}\text{C}$  spin-pairs and additionally helps to distinguish atoms with different number of attached carbons. While the carbons with no or two directly attached aliphatic carbons give rise to negative peaks (C2 of adenosine, C8 from purines, C2', C3', C4' and C5), the carbons with only one attached aliphatic carbon yield positive resonances (C1', C5' and C6).

2D  $^1\text{H}$ - $^{13}\text{C}$  HSQC experiment aimed at C1'-H1' ribose or C8-H8, C6-H6 base atoms regions was applied for  $\text{Mg}^{2+}$ - and miR-122 titration experiments to follow chemical shift perturbations and intensity changes. 2D  $^1\text{H}$ - $^{15}\text{N}$  HSQC experiment with the focus on imino N1-H1 spins was used in the titration experiments on domains I-II construct (Dieckmann and Feigon 1994).

For heteronuclear correlation spectra the averaged chemical shift difference  $d$  is calculated by the following equation:

$$d = \sqrt{\frac{1}{2} [\Delta\delta_{\text{H}}^2 + (\alpha \cdot \Delta\delta_{\text{X}}^2)]} \quad (53)$$

where  $\Delta\delta_{\text{H}}$  is the difference between  $^1\text{H}$  chemical shifts,  $\Delta\delta_{\text{X}}$  is the difference between  $^{13}\text{C}$  or  $^{15}\text{N}$  chemical shifts,  $\alpha$  is the scaling factor:

$$\alpha = \frac{\Delta(^1\text{H})}{\Delta(\text{X})} = 0.5 \quad (54)$$

where  $\Delta(^1\text{H})$  and  $\Delta(\text{X})$  are the shift range of protons and nitrogens or carbons, respectively.

Uncertainty for the averaged chemical shift difference,  $\sigma(d)$  is calculated as:

$$\sigma(d) \approx \sqrt{\frac{(\frac{1}{2}\Delta\delta_{\text{H}}\sigma(\Delta\delta_{\text{H}}))^2 + (\frac{1}{2}\alpha\Delta\delta_{\text{X}}\sigma(\Delta\delta_{\text{X}}))^2}{d^2}} \quad (55)$$

*Assignment of domain I resonances* was crucial for the CSP mapping and this task was performed by Philipp Innig Aguion in his master thesis project under my co-supervision.

Resonances of ribose spin systems (H1', H2', H3', H4', H5'a, H5'b, C1', C2', C3', C4' and C5') were assigned via **3D HCCH-COSY-TOCSY** experiment (Hu et al. 1998). To establish the link between ribose and nucleobase spin systems, a combination of through-bond and through-space experiments was used. Through-bond **3D TROSY-HCN triple-resonance** experiment was applied to establish the link via glycosidic nitrogen atoms (H6-C6-N1-C1'-H1' and H8-C8-N9-C1'-H1' correlations in pyrimidines and purines, respectively). To resolve ambiguous links due to resonance overlaps, through-space **3D CT <sup>1</sup>H,<sup>13</sup>C HSQC-NOESY** experiment with a mixing time of 150 ms was used. This experiment has also provided sequential links between nucleotides. H2' (i-1) – H6/H8 (i) and H5 (i) – H6 (i) contacts provided strong NOE cross-peaks, medium NOE cross-peaks were obtained from H3' (i) – H6/H8 (i), H3' (i-1) – H6/H8 (i) contacts, while H1' (i) – H6/H8 (i), H1'(i-1) – H6/H8 (i), H2' (i) – H6/H8 (i) and H5 (i+1) – H6 (i) contacts resulted in weak NOE peaks (Fürtig et al. 2003). To distinguish intra- and inter-residual NOE cross-peaks, 3D <sup>1</sup>H,<sup>13</sup>C HSQC-<sup>13</sup>C-filtered NOESY and 3D <sup>1</sup>H,<sup>13</sup>C HSQC-<sup>13</sup>C-edited NOESY with a mixing time of 300 ms (Zwahlen et al. 1997) were acquired on <sup>13</sup>C, <sup>15</sup>N nucleotide-type specific labeled samples. In an <sup>13</sup>C-filtered experiment only NOEs between protons of the labeled nucleotides and proximal protons of the *unlabeled* nucleotides are visible, while in an <sup>13</sup>C-edited experiment only NOEs between protons of labeled nucleotides and proximal protons of *labeled* nucleotides are visible.

#### **4.9.2. Solid-state NMR**

Sample preparation for ssNMR was performed with two different methods. For the microcrystallization method the MORPHEUS crystallization screen has been set according to the standard protocol (Gorrec 2009). The conditions with best microcrystals contained only PEG 1000 to 4000 with the buffering agent being any of HEPES, Bis-Tris, or Tris with pH in the range of 6.5-8. 1-3 mg of RNA was concentrated to 5-10 mg/ml in an appropriate buffer and subsequently mixed with an equal amount of precipitation solution (100 mM HEPES, pH 7.5, 100 mM NaCl, 30% PEG 4000/3000/1000). RNA was micro-crystallized by slow precipitation using a SpeedVac concentrator at room temperature for 2-3 hours. For the method of ethanol precipitation (Zhao et al. 2019a), RNA was concentrated to 10-12 mg/ml in a final volume of 300 µL. 30 µl of 2 M NaCl was added aiming for the final concentration of 200 mM NaCl. 900 µl of absolute ethanol was subsequently added to the mixture and it

was incubated at -20 °C overnight. Following, the mixture was centrifuged at 10000xg for 3 min and pellet was collected. In both methods, the precipitate was central-packed into the ssNMR rotor (3.2 mm) by ultracentrifugation.

NMR measurements were performed on a 600 MHz Bruker Avance III spectrometer equipped with 3.2 mm Bruker triple-resonance  $^1\text{H},^{13}\text{C},^{15}\text{N}$  MAS probe (Bruker Biospin, Rheinstetten, Germany). All experiments were acquired at 13.5 kHz  $\pm$  3 Hz MAS frequency. 1D  $^{13}\text{C}$  CP solid-state NMR spectra (Pines et al. 2003) were used to judge the quality of microcrystals. 2D  $^{13}\text{C},^{13}\text{C}$  DARR (Takegoshi, Nakamura, and Terao 2001) was recorded following a standard procedure. The spectra were recorded in a temperature range of 265-285K aiming for the optimal line width.

#### 4.10. SANS

Domain I-II RNA and miR-122 were extracted from Urea-PA as described above and dissolved in water. Residual salts were washed out with  $\text{H}_2\text{O}$  on Amicon® Ultra 15 mL Centrifugal Filters with 10 kDa MWCO (3x dilution/concentration cycles 15 ml  $\rightarrow$  500  $\mu\text{l}$ ) and then buffer exchanged to 100 mM HEPES pH 7.5, 10 mM KCl (3x dilution/concentration cycles 15 ml  $\rightarrow$  500  $\mu\text{l}$ ). RNA was annealed at 100  $\mu\text{g/ml}$  concentration at 95 °C for 6 minutes, cooled down to RT and concentrated to  $\sim$ 10-15 mg/ml. 1 M  $\text{MgCl}_2$  was added to yield the final concentration of 5 mM. To remove residual large polyacrylamide particles, RNA was additionally purified on Superdex Increase 200 SEC column. Collected fractions were concentrated to 10-15 mg/ml. For the samples “domains I-II( $^2\text{H}$ ) + 1x(miR-122)( $^1\text{H}$ )” and “domains I-II( $^2\text{H}$ )+4x(miR-122)( $^1\text{H}$ )” two RNAs were mixed in 1:1 and 1:4 molar ratios, respectively and then incubated for 30 minutes at 37 °C. Each sample was buffer exchanged to 68%  $\text{D}_2\text{O}$  100 mM HEPES pH 7.5, 10 mM KCl, 5 mM  $\text{MgCl}_2$  (SANS buffer) on Amicon® Ultra 15 mL Centrifugal Filters with 10 kDa MWCO (3x dilution/concentration cycles 15ml  $\rightarrow$  500  $\mu\text{l}$ ). Samples were stored on ice. Before the measurements samples were spin-down on the benchtop centrifuge at 13000 rpm to remove possible aggregates and heated up to RT.

SANS experiments were performed at D22 - a small-angle neutron scattering diffractometer at the Institut Laue-Langevin (ILL, Grenoble, France). Sample-detector distance was 1.3 m and 5.6 m, collimation length - 5.6 m and a neutron wavelength of 6 Å. The measurements were carried out on uniformly deuterated domains I-II RNA (118mer) alone and in complex with protonated miR-122 in the SANS buffer. The concentration of 118mer RNA was  $\sim$  7.5

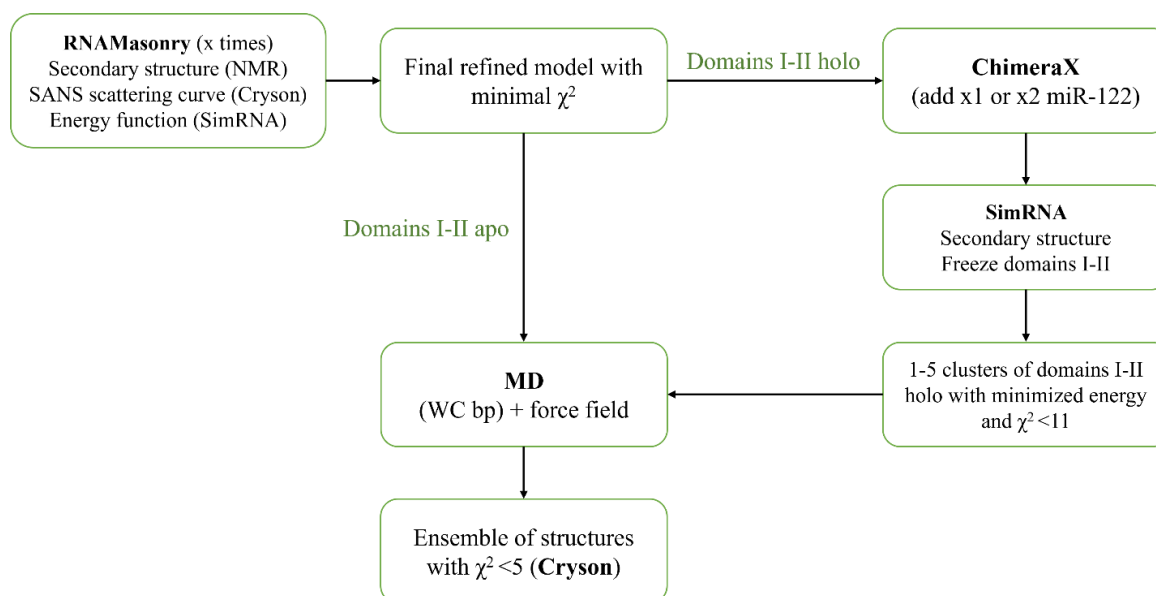
mg/ml. All data was acquired at 298 K. Data reduction and radial integration was performed with a custom ILL software package; mathematical buffer subtraction, data analysis and visualization were carried out with ATSAS software suite (Manalastas-Cantos et al. 2021).

#### 4.11. SAXS

RNA preparation protocol was the same as described above for SANS. All measurements were performed in SAXS buffer containing 100 mM HEPES pH 7.5, 100 mM NaCl, 5 mM MgCl<sub>2</sub>. Scattering curves were recorded for the apo 5' UTR and 5' UTR mixed with miR-122 in molar ratios 1:1 and 1:3. SAXS measurements were performed at beamline P12 at the Petra III storage ring at DESY (Deutsches Elektronen-Synchrotron) (Hamburg, Germany). Both in-batch and in-line SEC-SAXS methods were used. For the in-batch measurements a standard set-up was used to collect the data of samples in batch (an automated robot transferred the sample to capillary and 10 frames with 10 s exposure time were acquired). In-batch measurement series contained 5 dilution points (4.5 mg/ml down to ~ 200 µg/ml). When samples were measured via SEC-SAXS, the same running conditions as for SEC-MALS were used. The injection volume was 50-100 µl with a concentration of 5-10 mg/ml. All measurements were performed at 25 °C. SEC-SAXS data has been processed with CROMIX software (Panjkovich and Svergun 2018). Subsequent pipeline for the data analysis of SEC-SAXS data after CHROMIX processing and of in-batch data was analogous to the SANS data analysis described above.

#### 4.12. RNA Modeling

Since both NMR data and SANS data indicate significant flexibility of domains I-II construct, we have decided to describe our system as several structures and not as a single unique structure (**Fig. 4.12.1**). These structures have to satisfy both SANS curve shape and NMR-derived secondary structure. Since we could find several structures with a  $\chi^2 < 2$  it was not necessary to use an ensemble approach.



**Figure 4.12.1. Modeling approach for domains I-II constructs.** RNAMasonry generates a structure of domains I-II RNA employing the energy function, secondary structure derived from NMR and SANS scattering curve restraints. Final refined models with minimal  $\chi^2$  value are directly subjected to MD simulation (for domains I-II construct) or used to add one or two copies of miR-122 in Chimera software. Afterward, holo constructs are modeled using SimRNA tool. Domains I-II are frozen and secondary structure for miR-122 bound to domain I is provided. The structure clusters with minimized energy and  $\chi^2 < 11$  are subjected to MD simulation under force field and provided WC base pairs restraints. Finally, the pool of the structures is probed by Cryson, which compares experimental and modeled scattering curves.

The models were obtained using RNAMasonry software, which assembles RNA fragments from the RNA Bricks database into geometrically possible models considering provided secondary structure and Small Angle Scattering restraints (Chojnowski et al. 2021). Each program run had 100 folding steps and for each structure I have executed 10-15 runs in total. RNAMasonry program can process only a single RNA chain, so this way I have created a model of domains I-II structure in apo. To create models of domains I-II:(miR-122)<sub>1</sub> and domains I-II:(miR-122)<sub>2</sub>, first I created a model of domains I-II chain with SANS data acquired on holo samples and then added one or two copies of miR-122. First, helical miR-122 was added to RNAMasonry-derived model ChimeraX software (Pettersen et al. 2004) and then processed with SimRNA tool, which can model biologically relevant structures of RNAs and RNA complexes (Boniecki et al. 2016). SimRNA uses a Monte Carlo method for sampling conformational space and utilizes a statistical potential to approximate the energy of relevant conformers. Afterward, all structures with  $\chi^2 < 11$  were used as starting structures for molecular dynamic (MD) simulations in Amber 20 (Case et al. 2020; Götz et al. 2012; Le Grand, Götz, and Walker 2013; Salmon-Ferrer et al. 2013).



Briefly, domains I-II (118mer RNA), domains I-II:(miR-122)<sub>1</sub> and domains I-II:(miR-122)<sub>2</sub> complexes were simulated in explicit TIP3P water (Price and Brooks 2004), applying the AMBER ff14SB (Maier et al. 2015) and OL3 (Cheatham and Case 2013) force-fields. The total charge of the system was neutralized by the addition of 117, 139 and 161 Na<sup>+</sup> ions for domains I-II, domains I-II:(miR-122)<sub>1</sub> and domains I-II:(miR-122)<sub>2</sub>, respectively, prior to adding a truncated octahedral box of water around RNA. Prior to MD, the system was relaxed during two energy minimization stages. In the first stage a steepest-descent algorithm with 1500 steps followed by a conjugated gradient algorithm with 500 steps was applied on ions and solvent, while RNA was kept restrained. In the second stage, the entire system was minimized with 1000 steps of the steepest-descent algorithm followed by 1500 steps of the conjugated gradient algorithm. Afterward, the system was heated up to 300 K during 60 ps, when ions and solvent were allowed to move, while RNA was kept restrained.

In the following 50-200 ns restrained MD the geometry of WC base pairs in domain I and domain II for domains I-II construct (**Fig. 1.4.2**) was preserved, while the rest of the system (ions, solvent, RNA) were allowed to move. For domains I-II:(miR-122)<sub>1</sub> and domains I-II:(miR-122)<sub>2</sub> complexes additionally WC base pairs between domain I and miR-122 (**Fig. 1.4.2**) were restrained. To enhance the sampling, additionally to conventional MD, accelerated Molecular Dynamics (aMD) was used (Hamelberg, Mongan, and McCammon 2004; Pierce et al. 2012).

The final trajectories were analyzed with cpptraj (Roe and Cheatham 2013). Every 50 frames (100 ps) structure was extracted from the MD trajectory and was scored against experimental SANS curve in Cryson (Svergun et al. 1998) and structures with the lowest  $\chi^2$  values were selected.

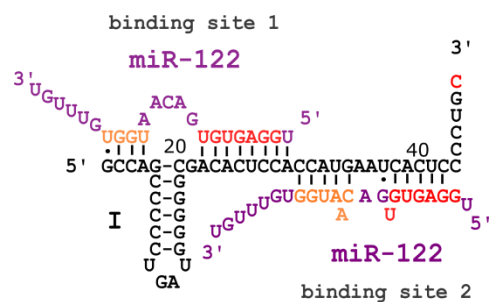


## 5. Results and Discussion

### 5.1. Domain I

#### 5.1.1. Domain I interacts with miR-122 and forms ternary complex

To study in details the 5' UTR–miR-122 interaction, first we have focused on the domain I, which is composed of stem-loop and single-stranded region and has two binding sites for miR-122 (**Fig. 5.1.1**).



**Figure 5.1.1 The sequence and secondary structure of domain I.** Two copies of miR-122 bind to domain I. Non-base-paired nucleotides of miR-122 are shown in violet, seed region (2-8 nt) is shown in red and auxiliary region is shown in orange. The 3'-terminal cytidine of domain I is an artifact from the PstI restriction site and is shown in red. The sequence and binding patterns are from (Mortimer and Doudna 2013).

#### *Sample preparation*

##### Domain I

RNA was synthesized by *in vitro* T7 transcription reaction using a linearized PUC-19 plasmid template with insertion of the first 47 nucleotides of 5' UTR. The final construct contained 48 nucleotides with an artificial 3'-terminal cytidine being a part of PstI restriction linearization site. According to the predicted minimum free energy (MFE) secondary structure (Lorenz et al. 2011), this terminal cytidine affects neither the folding and MFE of domain I itself, nor disturbs the interaction of domain I with miR-122, since it is located far from the miR-122 binding sites (**Table 5.1.1**).

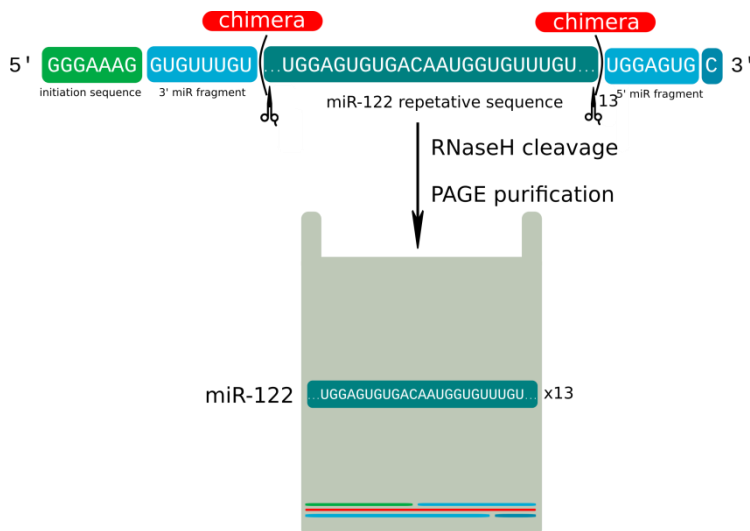
**Table 5.1.1. Secondary structure comparison of domain I constructs in complex with one copy of miR-122.** 48mer construct used in this study possesses an additional 3'-terminal cytidine. 47mer is a native RNA construct. Predictions were made on the RNAfold webserver (Lorenz et al. 2011).

RNA	Sequence and predicted secondary structure in dot-bracket notation	MFE, kcal/mol
<b>48mer+miR-122</b>	GCCAGCCCCCUGAUGGGGG CGACACUCCACCAUGAAUC ACUCCCCUGC&UGGAGUGU GACAAUGGUGUUUGU ...((((((.....)))))).....((((((.....((((..... ...&.)))))).....)))))).....	-33.78
<b>47mer+miR-122</b>	GCCAGCCCCCUGAUGGGGG CGACACUCCACCAUGAAUC ACUCCCCUG&UGGAGUGU GACAAUGGUGUUUGU ...((((((.....)))))).....((((((.....((((..... ..&.)))))).....)))))).....	-33.78

#### miR-122

miR-122 is 23 nucleotides long and has uridines at both 5' and 3' termini. Therefore, it was not possible to synthesize this RNA by the insertion of the DNA sequence flanked with T7 promoter on the 5' and PstI restriction site on the 3' end, as it would introduce both artificial guanosine and cytidine at 5' and 3' ends, respectively, which could affect the binding affinity of miR-122 to the HCV RNA.

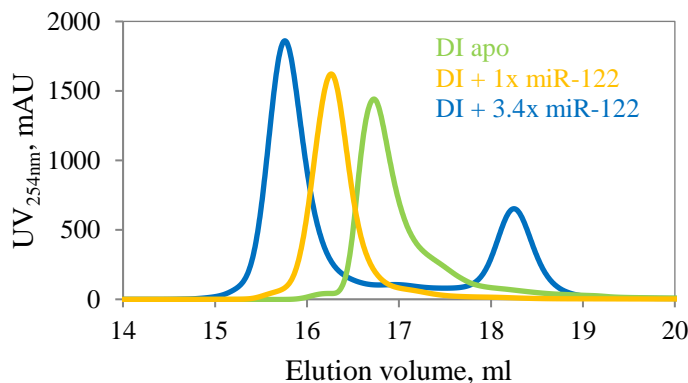
We have applied recently proposed strategy, that delivers RNA with homogenous 3'-OH and 5'-monophosphate ends with any desired sequence (Feyrer et al. 2020). The template for the transcription is the repetitive sequence of the target RNA (miR-122), framed with T7 promoter, optimal initiation sequence and 3'-terminal miR-122 fragment at the 5' end, and 5'-terminal miR-122 fragment bearing the PstI restriction site at the 3' end. After one-pot transcription, the reaction was purified on weak-anion-exchange DEAE-column and a long RNA was specifically cleaved by RNaseH guided by 2'-O-methyl-RNA/DNA chimera and purified on the denaturing PAGE (**Fig. 5.1.2**). This approach has several advantages, as it allows synthesizing of RNA of any desired sequence with a high yield due to the presence of an optimal initiation sequence and the usage of abortive transcription products. Moreover, RNase cleavage results in 5'-monophosphate, which is considered to be important for the miR activity (Rivas et al. 2005; Salzman et al. 2016), while conventional T7-mediated transcription results in 5' triphosphate end.



**Figure 5.1.2. Schematic representation of miR-122 synthesis protocol.** After *in vitro* T7 transcription of long RNA sequence, which consists of an optimal initiation sequence, 5'-miR-122 fragment, 13 copies of miR-122 sequence, 3'-miR-122 fragment and 3'terminal cytidine from the PstI restriction site, the long product is site- specifically cleaved by RNaseH guided by 2'-O-methyl-RNA/DNA chimera. After the cleavage, the reaction is purified by denaturing PAGE.

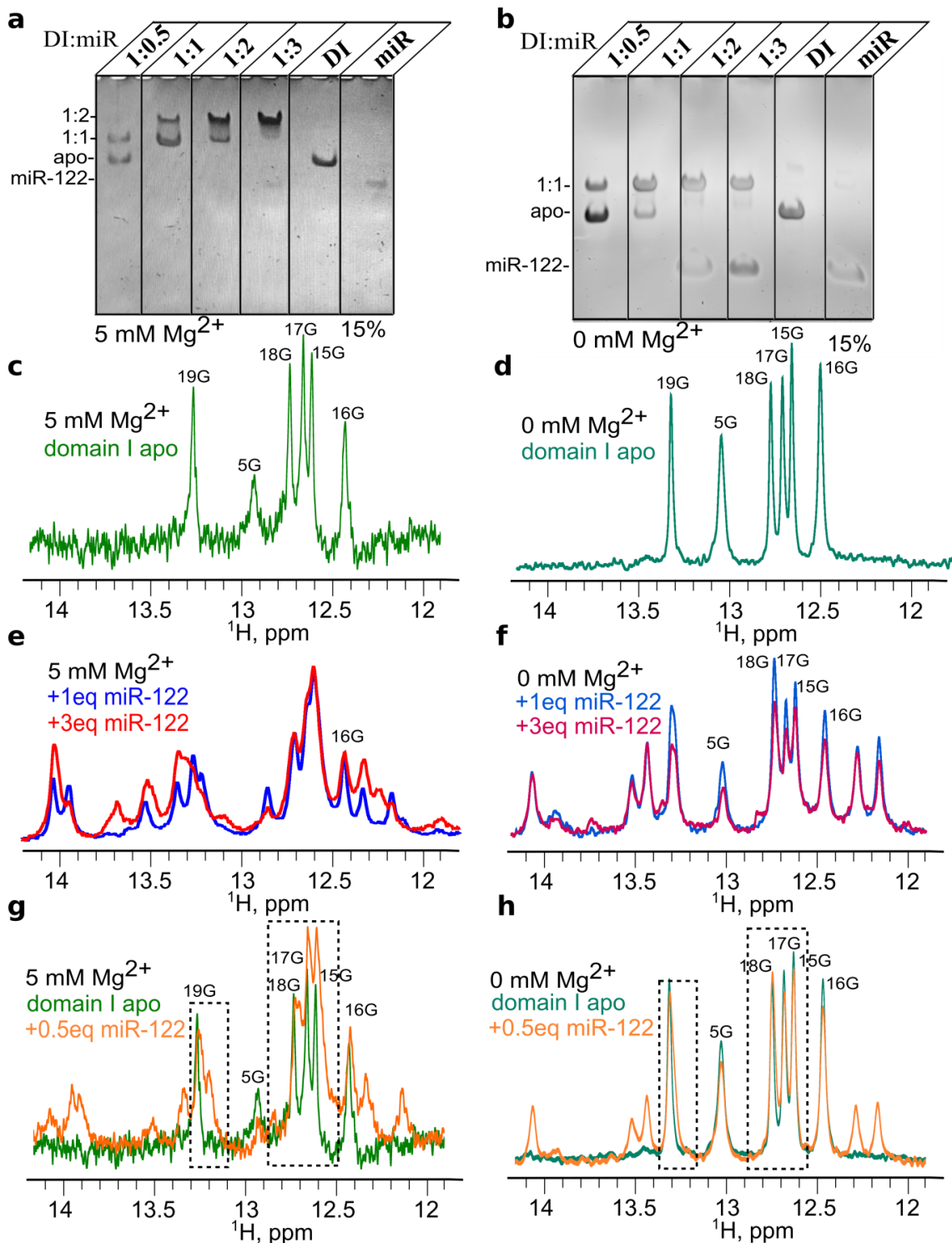
### 5.1.2. Domain I binds two copies of miR-122 and requires magnesium ions

First, the stoichiometry of the domain I–miR-122 interaction was characterized by SEC (**Fig. 5.1.3**). As shown in previous studies domain I binds two copies of miR-122 (Jopling, Schütz, and Sarnow 2008; Machlin et al. 2011; Mortimer and Doudna 2013). This can be concluded from the comparison of elution volumes of domain I in apo, after the addition of 1 equivalent and 3.4 equivalents of miR-122. Domain I absorption peak consistently shifts upon addition of miR-122, indicating formation of the complex with higher molecular size. Furthermore, upon addition of 3.4 equivalents of miR-122 a second peak with an elution volume of 18.5 ml appears, which corresponds to free miR-122.



**Figure 5.1.3. The domain I (DI)–miR-122 interaction.** DI is titrated with miR-122 RNA in increasing ratios (0, 1 and 3.4). The shift of the absorption peak on size-exclusion chromatograms towards smaller elution volumes indicates formation of larger size complexes. Right peak of the “domain I + 3.4x miR-122” sample corresponds to free miR-122.

Electrophoretic mobility shift assay (EMSA) confirms the binding of two copies of miR-122 (**Fig. 5.1.4a**) as well. Reduced mobility of domain I band is visible after the addition of 1 equivalent of miR-122 and further reduction is observed after addition of the second equivalent of miR-122. The band, which corresponds to the ternary complex, is not shifting anymore upon addition of the third equivalent and the excess of miR-122 might be seen as a pale band with higher mobility.



**Figure 5.1.4. The domain I–miR-122 interaction.** Mg<sup>2+</sup> affects miR-122 binding to the domain I. **a)** 15% EMSA of domain I (DI) mixed with miR-122 (miR) in different ratios (from left to right: 1:0.5, 1:1, 1:2, 1:3 1:0, 0:1) in the Mg<sup>2+</sup>-containing buffer (100 mM HEPES pH 7.5, 100 mM KCl, 5 mM MgCl<sub>2</sub>) **b)** 15% EMSA of domain I (DI) mixed with miR-122 (miR) in different ratios (from left to right: 1:0.5, 1:1, 1:2, 1:3 1:0, 0:1) in the buffer without Mg<sup>2+</sup> (100 mM HEPES pH 7.5, 100 mM KCl). **c)** 1D NMR of <sup>1</sup>H imino region of domain I, 6 peaks correspond to 6 G:C WC base pairs in the buffer with Mg<sup>2+</sup>. **d)** 1D NMR of <sup>1</sup>H imino region of domain I, 6 peaks correspond to 6 WC G:C base pairs in the buffer without Mg<sup>2+</sup>. **e)** Overlay of imino proton resonances of domain I+1 equivalent of miR-122 (domain I + 1x miR-122) and domain I+3 equivalents of miR-122 (domain I + 3x miR-122) in buffer with Mg<sup>2+</sup>. **f)** Overlay of imino proton resonances from domain I+1x

miR-122 and domain I+3x miR-122 in the buffer without Mg<sup>2+</sup>. **g-h**) Overlay of imino proton resonances from domain I and domain I+0.5x miR-122 in the Mg<sup>2+</sup>-containing buffer (g) and in the buffer without Mg<sup>2+</sup>(h). Dashed boxes in **g**) and **h**) indicate spectral regions in domain I + 0.5x miR-122 sample with the major differences M between Mg<sup>2+</sup> -containing and Mg<sup>2+</sup>-less buffers. All spectra were acquired at a temperature of 308K (except **e**), at 298K). The spectra **d**), **e**), (f) and **h**) were recorded at a <sup>1</sup>H field-strength of 850 MHz, spectra **a**) and **g**) were acquired at a <sup>1</sup>H field-strength of 700 Mhz.

Number of resonances in the imino region of 1D <sup>1</sup>H NMR spectrum (ca 15-10 ppm) can provide a good estimation for the number of base pairs present in the sample. Indeed, in the imino spectrum of free domain I 6 peaks from imino H1 protons of guanosines are visible, which correspond to 6 WC G:C base pairs of the stem-loop region (**Fig. 5.1.4c**), see section 5.1.4 for the details on imino protons assignment (*vide infra*) After addition of one molar equivalent of miR-122 the number of peaks increases, which is evidence of a base pairs formation between domain I and miR-122. Upon addition of 3 equivalents of miR-122, the number of base pairs keeps growing indicating a new binding event (**Fig. 5.1.4 e**).

Mg<sup>2+</sup> cations are required for the interaction of domain I with two copies of miR-122 (Mortimer and Doudna 2013). To explore the role of Mg<sup>2+</sup> in the HCV–miR-122 interaction, I have measured both EMSA and NMR in the absence of Mg<sup>2+</sup>. In EMSA performed in the buffer without MgCl<sub>2</sub>, the mobility of the RNA band corresponding to the domain I–miR-122 complex did not change beyond 1:1 ratio (**Fig 5.1.4 b**), indicating that in the absence of Mg<sup>2+</sup> only one copy of miR-122 can bind to the domain I.

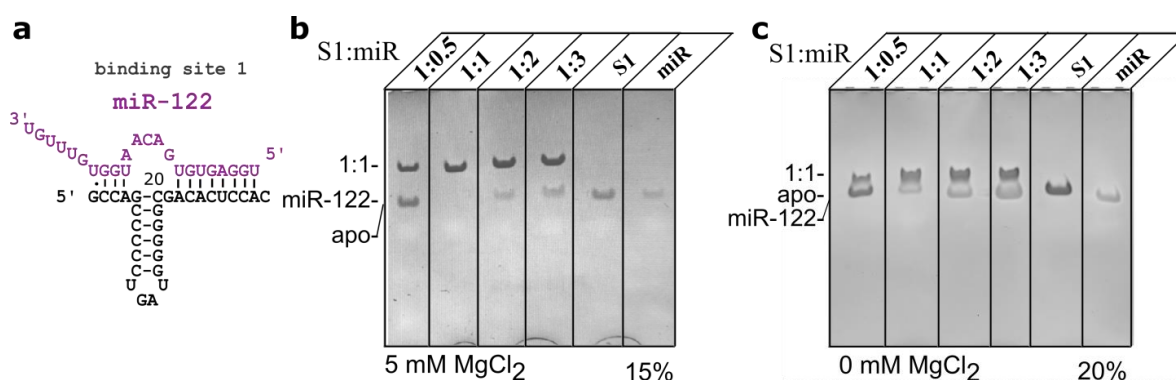
In the absence of Mg<sup>2+</sup> stem-loop of domain I adopts similar secondary structure with 6 WC G:C base pairs as in presence of Mg<sup>2+</sup>, as evident from NMR spectrum (**Fig 5.1.4 d**). 1D NMR imino spectrum of the domain I+ 1x miR-122 complex perfectly matches the spectrum of the domain I + 3x miR-122, indicating that just one copy of miR-122 binds to domain I in buffer without Mg<sup>2+</sup> (**Fig. 5.1.4 f**).

The obstruction of the second miR-122 copy binding upon the absence of Mg<sup>2+</sup> ions can be explained by two possible mechanisms. Without Mg<sup>2+</sup> ions, (i) one of the binding sites might be structurally unavailable for the binding of the second copy of miR-122 or (ii) after the binding of one miR-122 copy in an unconventional manner, both sites become unavailable for the subsequent binding.

To figure out, whether the first copy of miR-122 binds to domain I in the buffer with and without Mg<sup>2+</sup> in a similar way, we have compared spectra of domain I + 0.5x miR-122 sample in Mg<sup>2+</sup>-containing buffer (**Fig 5.1.4 g**) and in the buffer without Mg<sup>2+</sup>(**Fig 5.1.4 h**). We assumed that at this chosen ratio only one binding site will be occupied, e.g., there will

be no ternary complex formation. It was not possible to directly compare spectra acquired with and without  $Mg^{2+}$ , because, (i)  $Mg^{2+}$  may affect domain I secondary structure (see **Fig. 5.1.11**, *vide infra*), and (ii)  $Mg^{2+}$  alters chemical shift values. Thus, we decided to compare relative position of the new imino proton peaks, which appear upon miR-122 binding with and without  $Mg^{2+}$ . We have observed, that in the sample measured in the  $Mg^{2+}$ -containing buffer, more new peaks compared to the sample in the buffer without  $Mg^{2+}$  have appeared with many of them situated in the spectral region typical for guanosine imino protons ((**Fig 5.1.4 g**). This could be an indication that without  $Mg^{2+}$  miR-122 binds in a distinct manner rather than to one of the binding sites of domain I and both binding sites become sterically unavailable.

Our hypothesis is also supported by the EMSA experiment on the binding site 1 (S1, 30mer), which shows that the “weak” first binding site is not actually “magnesium dependent” as it was suggested before (Mortimer and Doudna 2013). Our data demonstrates, that both with and without  $Mg^{2+}$ , the S1–miR-122 complex is formed, which is better resolved on 20% PAGE than on 15% (**Fig. 5.1.5**).



**Figure 5.1.5. The binding site 1 (S1)–miR-122 interaction.** S1 binds one copy of miR-122 both in the absence and in the presence of  $Mg^{2+}$ . **a)** The sequence and secondary structure of the first binding site of domain I (S1, 30mer). **b)** EMSA on 15% PAGE of S1 mixed with miR-122 (miR) in different ratios (from left to right: 1:0.5, 1:1, 1:2, 1:3 1:0, 0:1) in the  $Mg^{2+}$ -containing buffer (100 mM HEPES pH 7.5, 100 mM KCl, 5 mM  $MgCl_2$ ) and **c)** EMSA in the buffer without  $Mg^{2+}$  on the 20% PAGE.

### 5.1.3. Resonance assignment of domain I

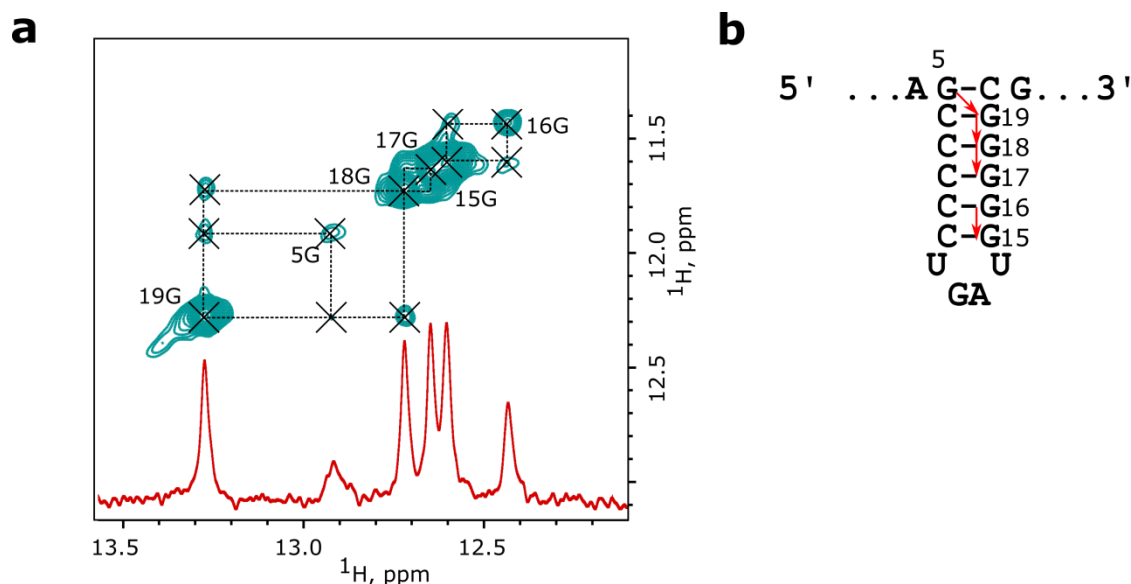
For the further study of domain I and its interaction with miR-122, we have assigned resonances of ribose and base C6/H6 and C8/H8 resonances in pyrimidines and purines, respectively. This part of the work has been started by the author of the thesis and finalized by Philipp Innig Aguion in his master thesis project.



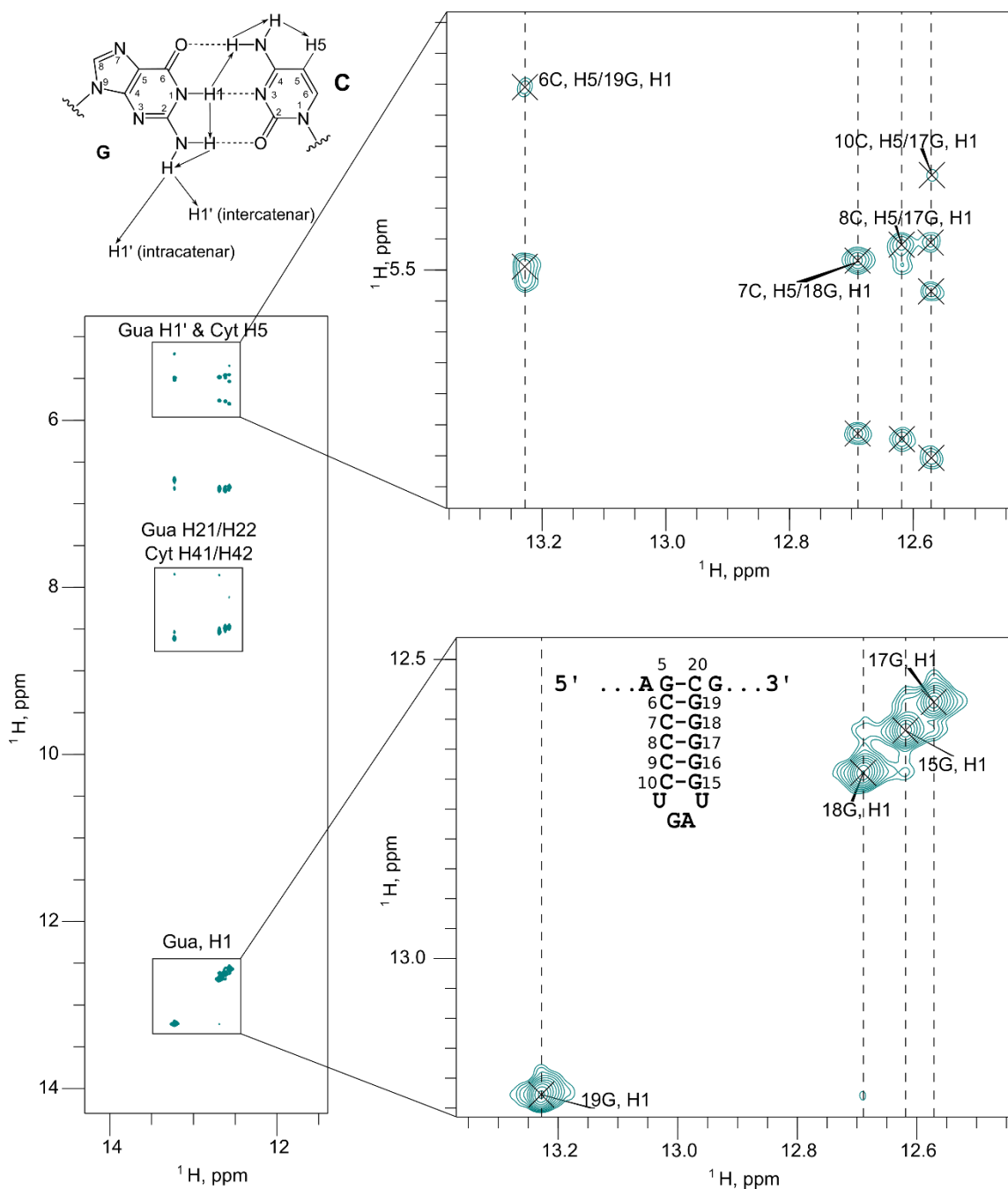


#### 5.1.4. Elucidation of base pairing and secondary structure

Imino protons region of 1D  $^1\text{H}$  NMR spectrum of isolated domain I shows 6 well-separated peaks from the imino protons involved in the WC base pairs (**Fig. 5.1.7**). This observation is in agreement with the established secondary structure of domain I in the context of 5' UTR (Brown et al. 1992). To assign imino protons and establish the sequential walk between them, we have acquired 2D  $^1\text{H}$ - $^1\text{H}$  NOESY spectrum at a temperature of 298K (since at a temperature of 308K, at which most of the assignment experiments were performed, some of the peaks were missing due to enhanced dynamics). The less intense peak belongs to H1 proton of 5G and was a starting point for a sequential walk (**Fig. 5.1.7**). However, due to the absence of the 16G-17G cross-peak, 15G and 16G could not be unambiguously assigned, so that, additionally we have used NOE contacts with the H5 proton of base-paired cytidine (Fürtig et al. 2003) (**Fig. 5.1.8**).



**Figure 5.1.7. The sequential imino-protons walk for the domain I.** a) Imino region of 2D  $^1\text{H}$ - $^1\text{H}$  NOESY (300 ms mixing time) and 1D  $^1\text{H}$  spectra. The lines indicate the imino protons connectivity. b) The secondary structure of stem-loop in domain I. Red arrows indicate NOE-contacts that were observed by NMR. All spectra were acquired at a temperature of 298 K and a  $^1\text{H}$  field-strength of 600 MHz.



**Figure 5.1.8.** 2D  $^1\text{H}$ - $^1\text{H}$  NOESY spectrum of domain I. Left, whole range of intranucleotide G H1–G H21/H22, G H1–G H1' and internucleotide G H1–C H5, H41/H42 correlations. Right, zoom-in on G H1–H1' and G H1–C H5 region (up), zoom-in on the imino-imino region of guanosines (bottom). The cross-peaks between guanosine H1 protons and cytidine H5 protons are shown and serve to assign imino-protons. The dashed lines indicate the imino protons – H5 connectivity. The NOESY mixing time was 300 ms. The spectrum was acquired at a temperature of 308 K and at a  $^1\text{H}$  field-strength of 850 MHz.

**5.1.5. Features and dynamics of the domain I–miR-122 interaction**

In this section I will provide detailed picture of the domain I–miR-122 interaction based on the analysis of the NMR titration series. To follow chemical shift changes upon complex formation, first, we have recorded 2D HSQC spectra of uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$  labeled domain I

and titrated it with unlabeled miR-122. The analysis of the data was challenging due to severe spectra broadening, bleaching of some resonances, appearing of the new peaks, which could not be assigned, so that just a few resonances could be followed after the addition of one equivalent of miR-122 (**Fig. 5.1.9**). Such picture is typical for dynamical systems in slow-to-intermediate exchange regime (exchange between free and miR-122-bound domain I and between different domain I conformers). Moreover, the ternary domain I:(miR-122)<sub>2</sub> complex has twice the size of apo domain I RNA (15.5 kDa vs 31.1 kDa, respectively). The molecules with higher molecular weight tumble slower, thus having shorter T<sub>2</sub> relaxation times, which ultimately leads to the overall line broadening and reduction of the signal intensity (Williamson 2013).

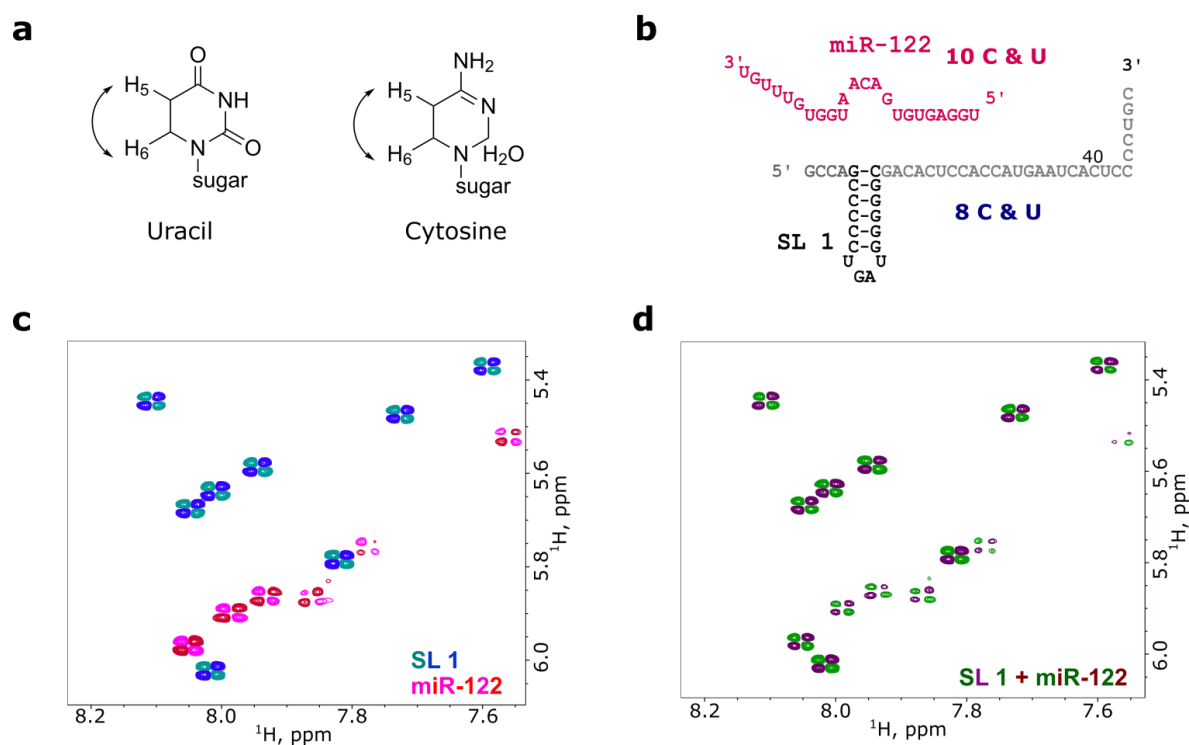
While the system in a slow-to-intermediate exchange regime may show barely any CSPs, interactions can be characterized by the mapping of intensity changes in the NMR titration series, anticipating that the faster drop of intensity indicates tighter binding. Indeed, after the addition of one equivalent of miR-122 the median value (interquartile range) of relative peak intensity for the nucleotides in the binding site 1 (22A – 28C) was 23.7% (15.9%, 24.5%), in the binding site 2 (37U-43C) - 18.8% (18.4%, 21.3%), while for the stem-loop region (5G-20C), the average intensity was 35.3% (32.4%, 44.8%). In holo-state (fully saturated domain I after the addition of 6 equivalents of miR-122) peak intensities decreased even further with most of the peaks becoming weak beyond detection. While the stem-loop region has an average residual intensity of only 12.3% (0.0%, 17.0%), both for binding site 1 and binding site 2 the average intensity was 0.0%. The results are summarized in the **Table 5.1.2**.

The reduction of resonances' intensities in the stem-loop region was the smallest and is caused only by the molecular weight increase upon the complex formation. To prove the absence of interactions between stem-loop and miR-122, we have acquired <sup>1</sup>H-<sup>1</sup>H Double Quantum Filtered (DQF) COSY experiment on the unlabeled stem-loop (SL1) construct mixed with unlabeled miR-122. In this experiment, H5 and H6 atoms of pyrimidines are correlated (**Fig. 5.1.10 a**). First, we have recorded DQF spectra of SL1 and miR-122 separately (**Fig. 5.1.10 c**) and then we have compared these spectra with the spectra acquired on the mixture of both RNAs (**Fig. 5.1.10 d**). The spectrum of the SL1 construct shows 8 resonances from 6 cytidines and 2 uridines as expected, while only 7 out of 10 resonances were visible for miR-122, probably, due to some inherent dynamics or overlap of the resonances. The spectrum of SL1 and miR-122 mixture is almost identical to the sum of spectra of the individual RNAs, confirming that miR-122 indeed does not interact with SL1.



**Table 5.1.2. Average relative intensity of C1'-H1' resonances upon miR-122 binding.** The data for the auxiliary site 1 (2C – 4A), binding site 1 (22A – 28C), auxiliary site 2 (30C – 34G), binding site 2 (37U – 43C) and stem-loop (5G – 20C), acquired on  $^{13}\text{C}$ ,  $^{15}\text{N}$  labeled domain I upon addition of 1, 2, and 6 equivalents of miR-122. Number of the nucleotides used for calculation is indicated in parentheses. The data is shown as a median and interquartile range in parentheses.

	+1 eq miR-122	+2 eq miR-122	+6 eq miR-122
<b>Auxiliary site 1 (3)</b>	42.4% (38.1%, 53.3%)	45.6% (22.8%, 59.3%)	34.2% (17.1%, 38.9%)
<b>Binding site 1 (7)</b>	23.7% (15.9%, 24.5%)	0.0%	0.0%
<b>Auxiliary site 2 (5)</b>	16.1% (15.2%, 17.8%)	0.0%	0.0%
<b>Binding site 2 (7)</b>	18.8% (18.4%, 21.3%)	0.0%	0.0%
<b>Stem-loop (15)</b>	35.3% (32.4%, 44.8%)	15.6% (0.0%, 22.5%)	12.3% (0.0%, 17.0%)

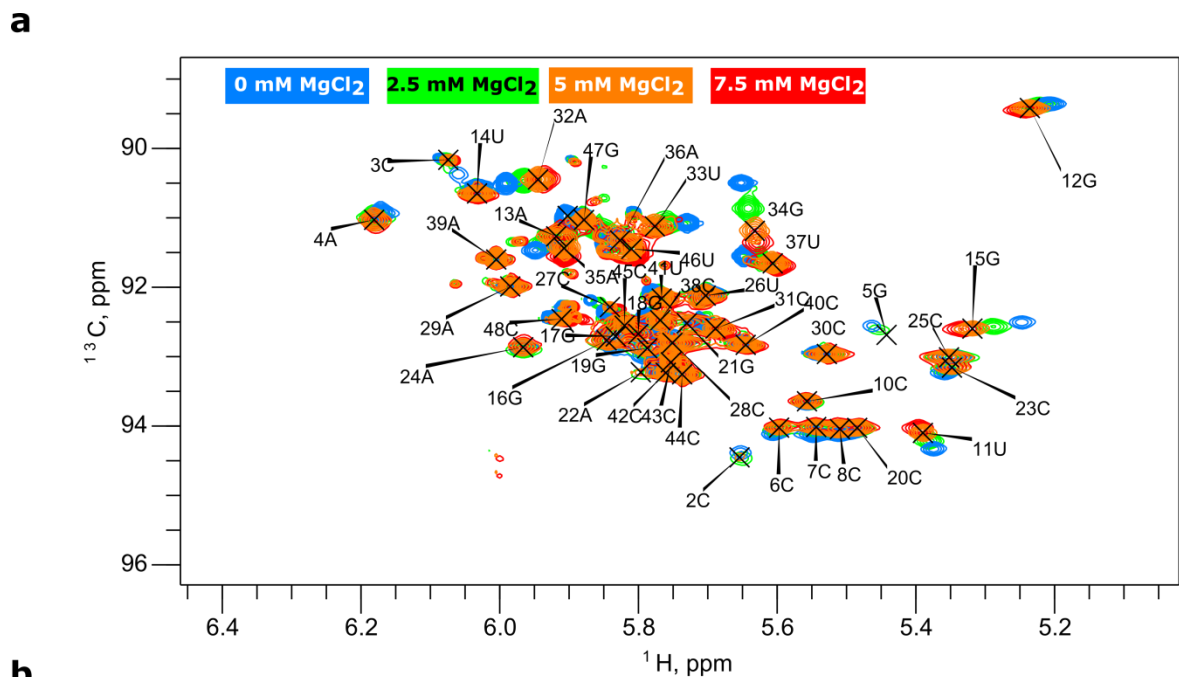


**Figure 5.1.10. miR-122 does not interact with stem-loop 1 (SL1).** **a**) Chemical structure of uracil and cytosine, whose H5 and H6 protons are correlated in the  $^1\text{H}$ - $^1\text{H}$  DQF-COSY experiment. **b**) miR-122 (pink) and SL1 construct (bold black), the rest of domain I is shown in grey and is not part of the RNA construct used in this experiment. **c**) Overlay of  $^1\text{H}$ - $^1\text{H}$  DQF-COSY spectra of the unlabeled SL 1 (green-blue) and miR-122 (pink-red). **d**)  $^1\text{H}$ - $^1\text{H}$  DQF-COSY spectra of the mixture of SL1 and miR-122 (green-purple). All spectra were acquired at a temperature of 308 K and a  $^1\text{H}$  field-strength of 600 MHz.

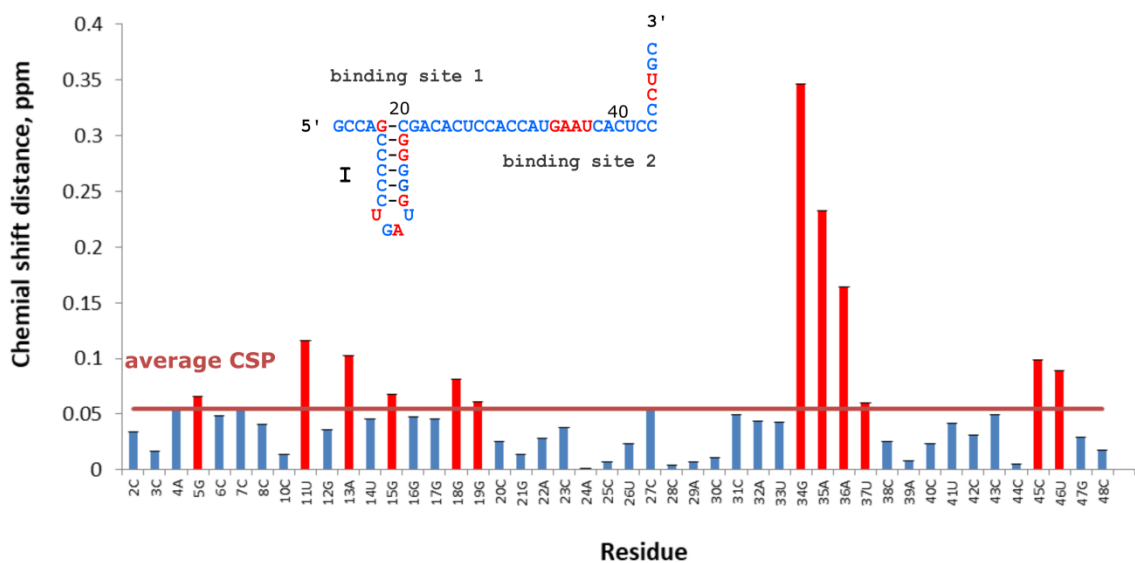
We have made several attempts to stabilize the domain I–miR-122 complex to get reasonable NMR signal that will allow structural studies.

First, we have studied domain I–miR-122 interactions in the buffer without  $Mg^{2+}$ , as we already knew that at this condition only one copy of miR-122 can be bound. The spectrum of the binary domain I:(miR-122)<sub>1</sub> complex with reduced molecular weight would be not interfered by the resonances from the larger tertiary domain I:(miR-122)<sub>2</sub> complex. Furthermore, we have utilized nucleotide-type selective labeling to reduce spectra crowding and allow better CSPs tracking. We did not perform resonance assignment in the buffer without  $MgCl_2$  anew but have recorded several spectra on domain I with gradually increasing  $MgCl_2$  concentration (0, 2.5, 5, 7.5 mM) (**Fig 5.1.11 a**) to reliably transfer resonance assignment.

Due to its high charge density, magnesium ions drive the RNA folding settling around the negatively charged phosphate backbone (Draper 2004). It has been shown that the folding of HCV IRES is driven by charge neutralization and cooperative uptake of  $Mg^{2+}$  (Kieft et al. 1999) and at 5 mM  $MgCl_2$  concentration IRES is competent for translation without initiation factors (Lancaster, Jan, and Sarnow 2006). Although, domain I is not a part of IRES, we have noticed significant CSPs of some resonances upon  $MgCl_2$  titration. While the residues located in stem-loop region (5G, 11U, 13A, 18G, 19G) show only moderate CSPs, substantial CSPs are observed for the binding site 2 nucleotides (34G, 35A, 36A, 37U), indicating some conformational changes in this region (**Fig 5.1.11 b**). These rearrangements upon the  $MgCl_2$  titration might be important for the binding of two copies of miR-122, making both sites accessible for the interaction or promoting the binding in the unconventional manner.



### Mg<sup>2+</sup> (0–5mM) induced CSPs in domain I of 5' UTR HCV

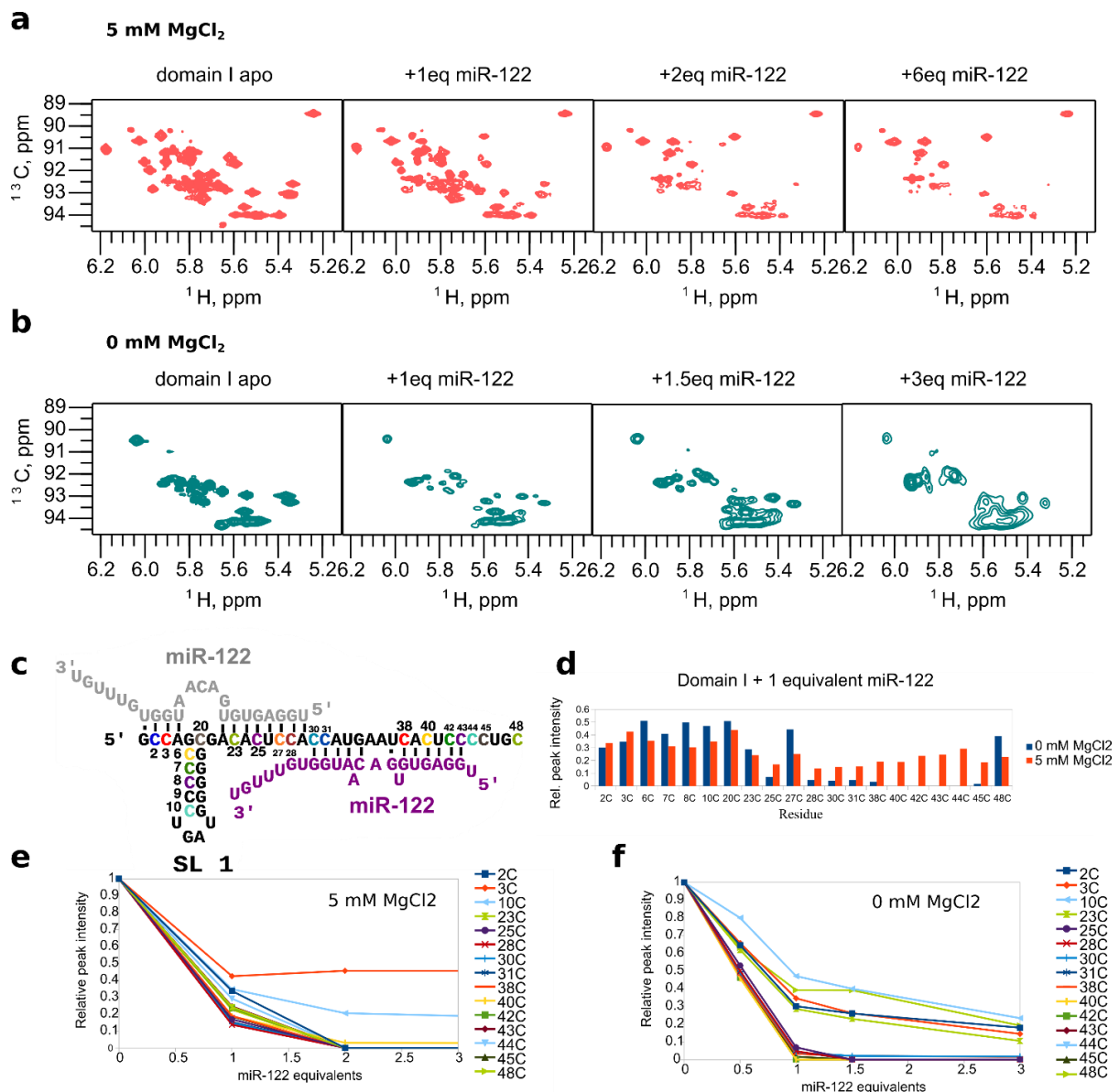


**Figure 5.1.11. Mapping of CSPs in domain I induced by the addition of MgCl<sub>2</sub>.** **a**) C1'-H1' region of 2D <sup>1</sup>H, <sup>13</sup>C HSQC spectra of uniformly <sup>13</sup>C, <sup>15</sup>N labeled domain I at different MgCl<sub>2</sub> concentrations. MgCl<sub>2</sub> was gradually added in 2.5 mM steps (titration points: 0 mM (blue), 2.5 mM (green), 5 mM (orange), 7.5 mM (red)). The resonance assignment is shown for the spectrum acquired at 5 mM MgCl<sub>2</sub>. **b**) CSPs upon the addition of 5 mM MgCl<sub>2</sub>. The calculated average CSP is shown as an orange line, residues with CSP larger than the average value are shown in red both on the bar diagram and on the secondary structure of domain I. The spectra were acquired at a temperature of 308 K and a <sup>1</sup>H field-strength of 850 MHz.



We have acquired a series of 2D HSQC spectra of  $^{13}\text{C}$ ,  $^{15}\text{N}$  cytidine-labeled domain I with gradual addition of 1, 1.5 and 3 molar equivalents of unlabeled miR-122 in the buffer without magnesium ions. Despite formation of only binary domain I:(miR-122)<sub>1</sub> complex, we have faced the same problem with the significant intensities reduction (**Fig 5.1.12 b**), which was even larger for the binding site 2 than in the buffer with  $\text{Mg}^{2+}$  (**Fig 5.1.12 d, e, f**). This finding is a clear indication that the signal broadening is mainly caused by an exchange regimen and not by the molecular weight increase.

The cytidines in the stem-loop region (C6 - C10, C20) keep 49.7% (47.1%, 50.8%) of their intensity upon addition of the first equivalent of miR-122 in the buffer without  $\text{Mg}^{2+}$ , while in the buffer with  $\text{Mg}^{2+}$ , the average residual intensity of cytidines is 34.7% (31.0%, 35.3%) (median value and interquartile range in parentheses). Since intensity reduction of the resonances in stem-loop is mainly caused by MW increase, we can assume that in the presence of magnesium ions the partial binding of the second copy of miR-122 occurs already at the equimolar domain I:miR-122 ratio (**Fig 5.1.4 a**), and, therefore the molecular weight of the complex is higher. Though, as suggested in section 5.1.3, we cannot exclude the possibility of miR-122 to bind differently, when there is no  $\text{Mg}^{2+}$  in the buffer. Surprisingly, we have noticed that the second binding site (38C, 40C, 42C, 43C) has more residual intensity in the buffer with 5 mM  $\text{MgCl}_2$ , than without  $\text{MgCl}_2$  (21.2% (18.9%, 23.8%) vs 0.0% (0.0%, 0.8%)) (**Fig 5.1.12 d**). Nevertheless, this large intensity attenuation of binding site 2 residues can't be explained by tighter binding without  $\text{MgCl}_2$ , because reported  $K_d$  of the binding event is on the contrary slightly higher without  $\text{MgCl}_2$  ( $130 \pm 33$  nM) than with  $\text{MgCl}_2$  ( $90 \pm 52$  nM) (Mortimer and Doudna 2013). Furthermore, EMSA (**Fig 5.1.4 b**) shows that at equimolar ratio domain I is not fully saturated by miR-122 in the buffer without  $\text{Mg}^{2+}$ . Moreover, if we estimate the difference in  $K_d$  based on our EMSA, we conclude that both for full domain I and first binding site S1,  $K_d$  in the buffer without  $\text{MgCl}_2$  is ca. 50-100% higher than in the buffer with  $\text{MgCl}_2$ . Thus, we can speculate, that the binding of miR-122 to domain I in the buffer without  $\text{MgCl}_2$  is more dynamic and is more towards slow-to-moderate-exchange, than with  $\text{MgCl}_2$ . The differing nature of the miR-122 binding is also seen, when we further follow intensity changes of individual nucleotides upon the addition of more equivalents of miR-122 (**Fig. 5.1.12 e, d**).



**Figure 5.1.12. Impact of  $Mg^{2+}$  ions on the domain I–miR-122 interactions.** Line broadening and reduction of resonance intensities upon formation of the domain I–miR-122 complex. **a)** Zoom-in on the C1'-H1' region of  $^1H, ^{13}C$  HSQC NMR spectra of uniformly  $^{13}C, ^{15}N$  labeled domain I in apo-form and after addition of 1, 2 and 6 equivalents of unlabeled miR-122. The buffer contains 5 mM  $MgCl_2$ . **b)** Zoom-in on the C1'-H1' region of  $^1H, ^{13}C$  HSQC spectra of  $^{13}C, ^{15}N$  cytidine-labelled domain I in apo-form and after addition of 1, 1.5 and 3 equivalents of unlabeled miR-122. The buffer contains no  $MgCl_2$ . All spectra were acquired at a temperature of 308 K and a  $^1H$  field-strength of 850 MHz. **c)** The sequence and secondary structure of domain I with two copies of miR-122 bound. All cytidines are numbered and colored, with the color-scheme as in e) and f). **d)** The comparison of C1'H1' relative peak intensity changes of cytidines in domain I upon the addition of 1 equivalent of miR-122 in the buffer containing 5 mM  $MgCl_2$  (red bars) and 0 mM  $MgCl_2$  (blue bars). **e)** Relative intensity changes of C1'-H1' cytidine resonances in uniformly labeled domain I upon titration with miR-122 in the buffer containing 5 mM  $MgCl_2$ . **f)** Relative intensity changes of C1'-H1' cytidine resonances in  $C^{lab}$  domain I upon titration with miR-122 in the buffer without  $MgCl_2$ . The cytidines are color-coded as in c).

Since NMR characterization of the full domain I–miR-122 complex was extremely challenging due to highly dynamic character of their interactions, we have attempted to study only the first binding site (S1, 30mer) (**Fig. 5.1.13 a**). We have recorded 2D HSQC spectrum of  $^{13}\text{C}$ ,  $^{15}\text{N}$  cytidine-labeled S1; C1'–H1' region has 13 resonances as expected and most of the resonances match resonances of the full-length domain I construct (48mer) and could be assigned (**Fig. 5.1.13 b**). However, the analysis of NMR spectra acquired on this 30mer construct bound to unlabeled miR-122 was hampered by the drop of intensity and peak broadening upon the complex formation (**Fig. 5.1.13 c**), indicating highly dynamic character of this interaction even in isolated S1.

To stabilize interaction of S1 with miR-122, we have extended base pairing at 5' end of domain I with two additional G:C base pairs, as only three base pairs in the auxiliary region could be the reason for the high dynamic behavior of the complex. miR-122 also had to be modified to match the base pairing (**Fig. 5.1.13 d**). While according to the prediction, secondary structure of the modified construct should not differ from the one of the native sequence, NMR spectrum revealed at least two conformers (20 C1'–H1' cytidine peaks visible in  $^{13}\text{C}$ ,  $^{15}\text{N}$  cytidine-labeled 34mer instead of expected 15 peaks) (**Fig 5.1.13 e**), with one additional base pair present (**Fig. 5.1.13 f**). Moreover, NMR resonances of this construct also indicate signal broadening upon binding with corresponding miR (**Fig. 5.1.13 g**). We anticipate that there might be a way to stabilize domain I–miR-122 interactions to obtain the structure of the ternary complex by NMR, though it represents a challenge, since it would require the insertion of several non-native nucleotides to miR-122 and both binding sites of domain I and such mayor modification put the biological relevance of the new construct into question.

Overall, we can conclude that the domain I–miR-122 interactions are highly dynamic in nature so that ternary complex may exist in multiple interchanging conformers which can be studied by e.g., relaxation dispersion experiments.



from the additional base pair. **g)** Overlay of 2D  $^1\text{H}$ ,  $^{13}\text{C}$  HSQC spectra of  $^{13}\text{C}$ ,  $^{15}\text{N}$  cytidine-labelled 34mer in apo and after addition of 1.3eq of unlabeled modified miR. Spectra in **e)-g)** were acquired at a temperature of 308 K and a  $^1\text{H}$  field-strength of 600 MHz.

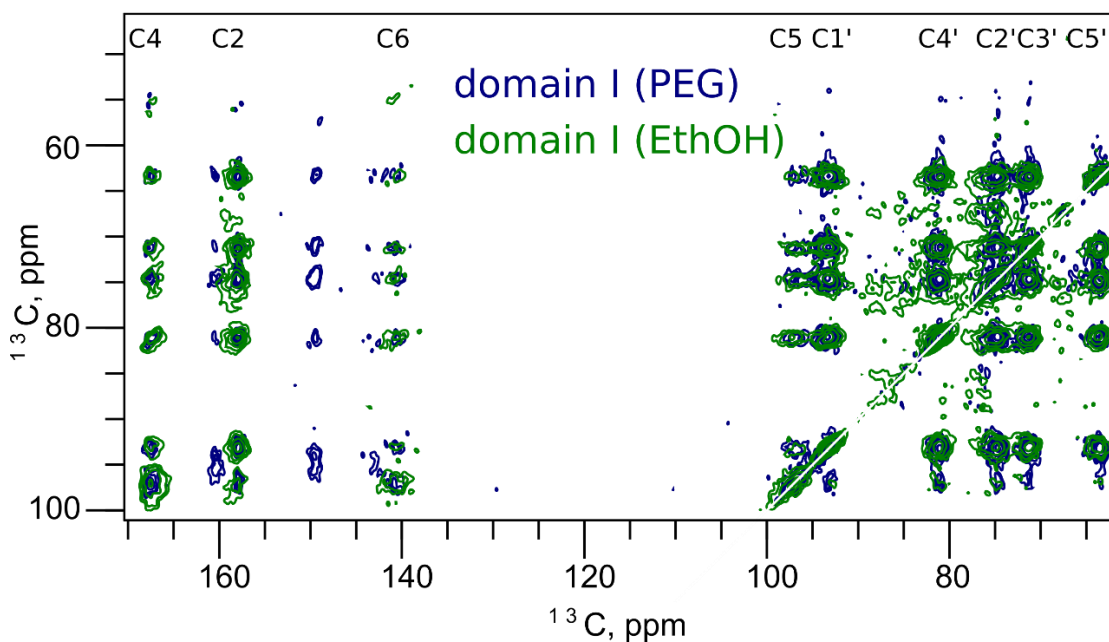
### 5.1.6. Solid-state NMR of domain I

RNA resonance assignment for the domain I was accomplished by solution-state NMR and independent, de novo assignment RNA resonances of domain I by solid-state NMR could serve as a proof-of-principle for this approach. While ssNMR is aimed to obtain spectral fingerprint of 5' UTR, it was interesting to extrapolate the interaction of domain I with miR-122 to the full-length 5' UTR. To compare the structural features of the isolated domain I and domain I as a part of 5' UTR, we have attempted to get its spectral fingerprint in the microcrystalline state by ssNMR.

In a view of the fact, that PEG was a good crystallization agent for ssNMR studies on 5' UTR (section 5.3, *vide infra*), we tested different concentrations of PEG 2000 and 4000 (50% and 30%) both in  $\text{H}_2\text{O}$  and in the RNA buffer (100 mM HEPES, pH 7.5, 100 mM KCl). The crystallization solution was added in 1:1 ratio to the pre-annealed domain I and then  $\text{H}_2\text{O}$  from the reaction mixture was evaporated using a SpeedVac. While the conditions with 30% and 50% PEG aqueous solution resulted in very fast and too compact precipitation, 30% PEG 4000 in 100 mM HEPES pH 7.5, 100 mM KCl crystallization buffer (the same conditions we used for 5' UTR RNA) has yielded good quality microcrystals.

To get a fingerprint of domain I by ssNMR,  $^{13}\text{C}$ ,  $^{15}\text{N}$ -cytidine-labelled sample was prepared, as cytidine are abundant nucleotides in a structured region of domain I (G-C stem-loop). However, despite all our efforts, RNA resonances were too broad with typical  $^{13}\text{C}$  line-widths of 1.15-1.65 ppm, which lead to severe resonance overlap and did not allow resonance assignment, which would require  $^{13}\text{C}$  line width in the range of 0.5 ppm (Aguion and Marchanka 2021; Marchanka et al. 2015).

An attempt to produce an ssNMR sample according to the recently published strategy using ethanol precipitation (Zhao et al. 2019b) also did not result in a spectrum of better resolution. Moreover, a lot of peaks characteristic for non-canonical regions of RNA, that were present in PEG-precipitated sample, were lost in EtOH-precipitated sample (**Fig. 5.1.14**).



**Figure 5.1.14. solid-state NMR on domain I RNA.** Overlay of 2D  $^{13}\text{C}, ^{13}\text{C}$ -PDSD spectra (200 ms mixing time) of  $\text{C}^{\text{lab}}$ -domain I prepared by PEG-microcrystallization (green) and ethanol precipitation (blue) methods acquired on a 600 MHz spectrometer at a temperature of 275 K.

Here, we can speculate that EtOH precipitation approach, while keeping RNA-secondary structure intact, may significantly disrupt tertiary structure in flexible RNA. Advanced sample preparation techniques (e.g., sedimentation by ultracentrifugation) and RNA stabilization by the addition of Ago2 protein to the domain I–miR-122- complex would be the next steps to improve ssNMR spectra and allow structural studies of RNA.

## 5.2. Domains I-II

### 5.2.1. Sample preparation

Domain II folds independently from the rest of the IRES both in free form in solution (Lukavsky et al. 2003) and when bound to 40S ribosomal subunit (Spahn et al. 2001), therefore, divide and conquer approach was justifiable. First, we have optimized sample preparation procedure, which yields the optimal *in vitro* conditions for domains I-II RNA, suitable for solution-state NMR and SANS study, meaning that RNA should have homogeneous 5' and 3' ends, RNA should stay monomeric at the concentrations of at least up to 0.5 mM and, ideally, adopt single secondary fold.

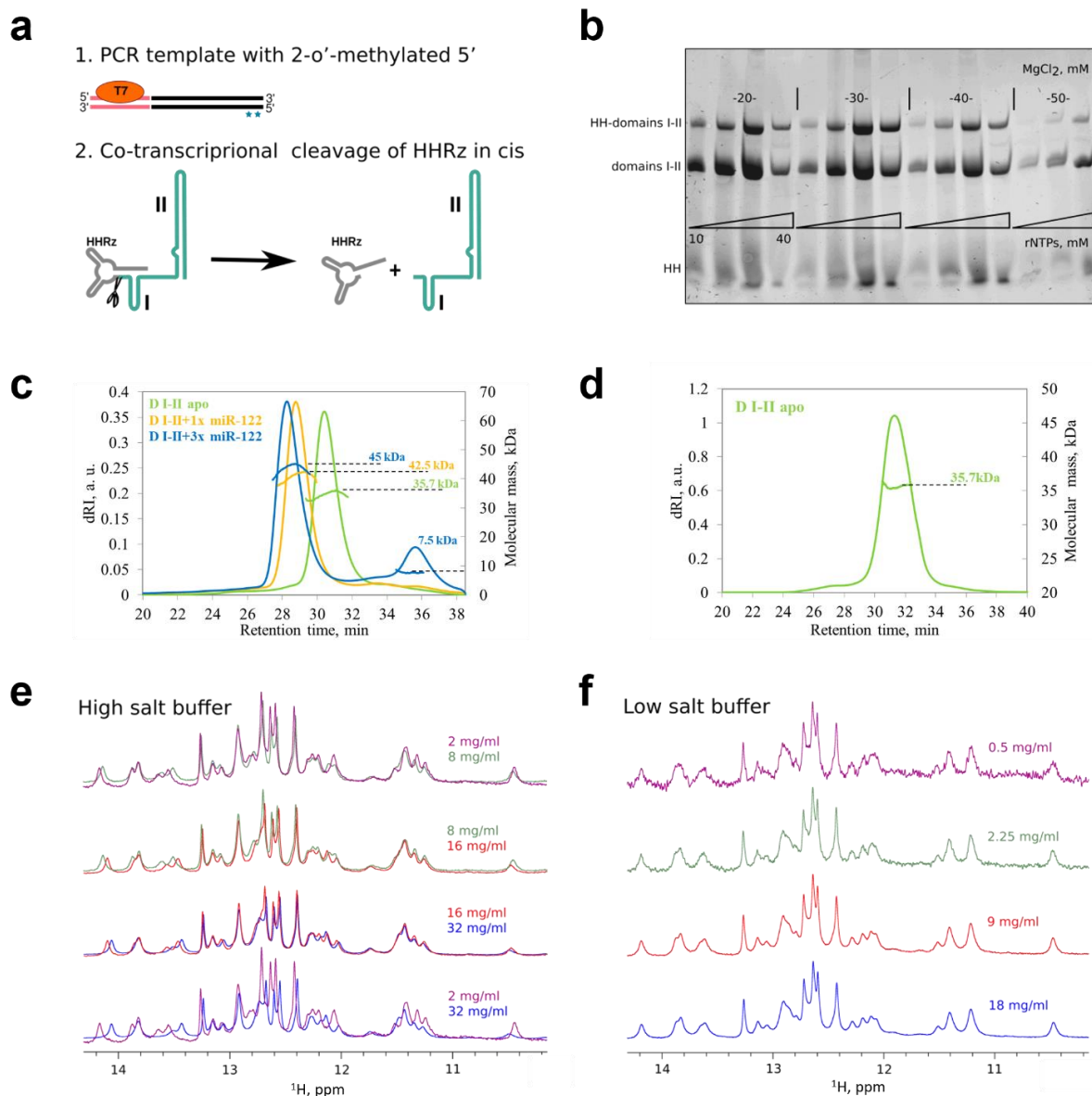
Domains I-II RNA construct was transcribed *in vitro* from the PCR template, which contains T7 promotor site upstream of the 118 nucleotides RNA sequence. *In vitro* transcription with T7 polymerase is commonly used to get milligram quantities of RNA for structural studies but can suffer from both 5' and 3' inhomogeneity during the run-off transcription (Gholamalipour, Karunanayake Mudiyansele, and Martin 2018; Helm et al. 1999; Pleiss, Derrick, and Uhlenbeck 1998). This problem is more challenging for the constructs longer, than 40-50 nucleotides, since PAGE purification cannot provide single-nucleotide resolution for RNA constructs > 50 nucleotides long.

Fortunately, several approaches exist to obtain well-defined homogeneous 5' and 3' ends. To obtain a homogeneous 5' end of RNA, we have positioned 47-nucleotides hammerhead ribozyme (HHRz) *in cis* upstream the 118nt sequence (**Fig. 5.2.1. a**). To minimize the chance that the ribozyme will not adopt the correct fold due to the interaction with the sequence of domains I-II, we have predicted its secondary structure using RNAfold web server (Lorenz et al. 2011). An important advantage of the utilization of 5' ribozyme *in cis* is the possibility to prime the template with five GTPs, a known approach to increase the RNA yield for T7 transcription reaction (Conrad et al. 2020). Another advantage of 5' HHRz is the absence of the downstream sequence specificity. All three potential transcription products can be easily separated on the 10% PAGE due to distinct size differences between 165 nt (uncleaved RNA), 118 nt (desired domains I-II construct) and 47 nt (ribozyme) (**Fig. 5.2.1. b**). Usually, HHRz was completely co-transcriptionally self-cleaved, providing well-defined 5' end. When the cleavage was not complete by the end of the transcription reaction, the RNA was additionally annealed. For this, the crude transcription reaction mix was first buffer exchanged to the cleavage buffer (40 mM Tris-HCl, pH 7.4, 100 mM NaCl, 2 mM EDTA), using Amicon® Ultra 15 mL Centrifugal Filters with MWCO 10 kDa to remove magnesium

ions, which are known to accelerate RNA degradation at higher temperatures (AbouHaidar and Ivanov 1999). Then the buffer-exchanged reaction mix was heated up to 90 °C in the water bath for 6 minutes and cooled down to RT on the bench. To achieve proper ribozyme fold and accomplish the self-cleavage, MgCl<sub>2</sub> was added in up to 50 mM concentration and the sample was incubated at 37 °C for 4-6 hours. Consequently, the sample was purified using 10% urea-PAGE.

To improve 3' end homogeneity, we have used a PCR product with two 2'-O-methylated nucleotides at the 5' end of the template strand as a DNA template for T7 *in vitro* transcription (Kao et al. 1999) (**Fig. 5.2.1 a**). The PCR template was synthesized from PUC-57 plasmid, which contained the full-length HCV 5' UTR sequence with precluding HH-ribozyme at 5' end. The reverse primer was designed to contain two modified 5' terminal nucleotides. This method allows rapid DNA template synthesis of the different length constructs from the same plasmid and does not require any additional purification steps, so that the crude PCR reaction product can be directly used for T7 *in vitro* transcription (Helmling et al. 2015). After denaturing purification followed by an annealing procedure (heat 5-6 minutes in water bath at 95 °C at 100 µg/ml in 100 mM HEPES pH 7.5, 10 mM KCl, cool down on the bench to RT, back concentrate, add MgCl<sub>2</sub> up to 5 mM) we have checked the oligomeric state, molecular weight of the domains I-II RNA, as well as binding stoichiometry of miR-122 and domains I-II. Apo domains I-II construct was eluted as a single peak with molecular weight matching the monomeric state of RNA. Upon addition of one or three molar equivalents of miR-122, domains I-II consecutively bind two copies of it, which is indicated by the increase of the retention time and molecular weight on SEC-MALS (**Fig. 5.2.1. c**).





**Figure 5.2.1. Sample preparation of domains I-II construct.** **a)** The scheme of T7 run-off *in vitro* transcription designed to gain 3'-end homogeneity, using PCR template with two terminal 2'-O-methylated nucleotides (stars) (1), and hammerhead ribozyme (HHRz) *in cis* to obtain well-defined 5' end (2). **b)** Optimization of  $MgCl_2$  and rNTPs concentrations for domains I-II transcription visualized on the 10% Urea-PAGE. Three bands correspond to domains I-II with uncleaved HHRz, domains I-II RNA and cleaved HHRz. The gel is stained with DNA Stain G (SERVA) and visualized using Gel Doc XR+ Gel Documentation System (BioRad). **c)** SEC-MALS of domains I-II apo (green) at 3 mg/ml concentration and domains I-II mixed with miR-122 in ratios 1:1 (orange) and 1:3 (blue) in high salt buffer (100 mM KCl, 100mM HEPES pH 7.5, 5 mM  $MgCl_2$ ). Measured molecular mass corresponds to monomeric state and confirms the consecutive binding of two copies of miR-122. **d)** SEC-MALS of domains I-II apo (green) with concentration 10 mg/ml in low salt buffer (10 mM KCl, 100 mM HEPES pH 7.5, 5 mM  $MgCl_2$ ) with molecular mass that correspond to the monomeric state. **e)** Pairwise comparison of 1D imino protons NMR spectra of apo domains I-II upon serial dilution in the high salt buffer CSPs as a sign of RNA oligomerization or interparticle interaction are observed. **f)** 1D imino protons NMR spectra of apo domains I-II upon serial dilution from 18 to 0.5 mg/ml in the low salt buffer. Absence of CSPs indicates homogeneity and stability of the sample at different concentrations.

## 5.2.2. NMR studies of domain II folding in the presence of domain I and upon miR-122 titration

### *Domain II as a part of domain I-II RNA*

As a next step, we aimed to investigate the possibility of alternative fold of domain II in the presence of domain I as reported previously.

To assign imino resonances of domain II, we have compared our experimental spectrum of  $^{13}\text{C}$ ,  $^{15}\text{N}$  guanosine-specific labeled ( $\text{G}^{\text{lab}}$ ) domains I-II with previously published spectrum for domain II (Lukavsky et al. 2003). Overall, a number of the peaks and resonances pattern in our spectrum were very similar to domain II spectrum obtained in that study, though some of the chemical shifts were perturbed or absent (e.g., 94G). This can be explained by the difference in buffer conditions and that the construct by Lukavsky et al (Lukavsky et al. 2003) has nucleotide modifications in the base region of the stem II, which were introduced to stabilize their construct. We were able to assign tentatively most of the G imino resonances, still, additional experiments to confirm the assignment of our construct and to probe iminos of uridines were necessary.

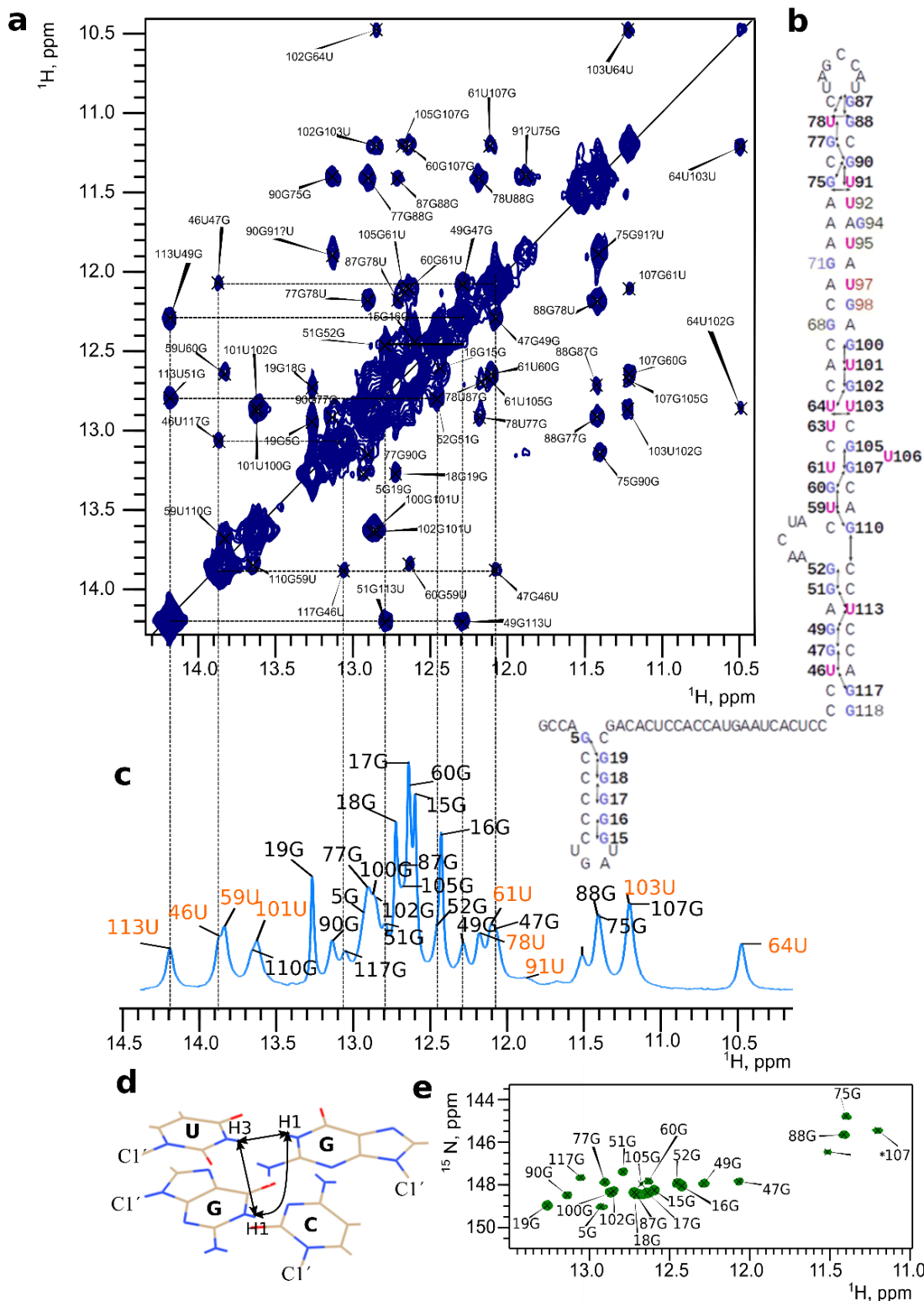
When we proceeded to imino proton assignment by  $^1\text{H}$ ,  $^1\text{H}$  NOESY, we have noticed that the chosen buffer conditions were not optimal for the measurements at high concentrations. The 1D  $^1\text{H}$  imino proton spectra acquired at different concentrations were not the same (**Fig 5.2.1 e**), indicating either oligomerization or interparticle interaction of the RNA. By lowering salt concentration in the buffer from 100 mM KCl to 10 mM KCl, we effectively removed CSP of proton resonances upon concentration variation, thus solving this issue (**Fig 5.2.1 f**). We have repeated the SEC-MALS for domains I-II using the low salt buffer to confirm that the new buffer conditions would not change the monomeric state of the sample (**Fig 5.2.1 d**). The experiments on  $\text{G}^{\text{lab}}$  samples were recorded in high salt buffer, but at relatively low concentrations (2.5 mg/ml), at which no oligomerization and interparticle interactions were observed.

As domains I-II is a relatively large RNA construct (37.9 kDa) with expected  $T_1$  time constant that is rather short, first we have optimized the proton-proton mixing time for the 2D NOESY experiment. Most peaks were visible better at shorter mixing time of 150 ms. Interestingly, the  $T_1$  values for the resonances in domain I were overall longer than for those in domain II, indirectly confirming the independence of domains I and II (**Table 5.2.1**).

Chemical shift range, ppm	Residue	$T_1$ , ms
13.32-13.18	19G	91.6±0.7
14.28-14.07	113U	54.5±0.8
12.29-12.24	49G	46.5±0.6
11.26-11.16	103U, 107G	37.5±0.3
10.57-10.39	64U	38.1±0.4

**Table 5.2.1.  $T_1$  relaxation times in selected residues in the unlabeled domains I-II construct.** RNA concentration was 350  $\mu$ M, measurements were carried out at a temperature 298 K and a  $^1\text{H}$  field-strength of 850 MHz in 10 %  $\text{D}_2\text{O}$  buffer.

We have performed sequential imino resonance assignment both for domains I-II construct and isolated domain II and we have assigned most of the visible imino resonances except the peak at 11.53 ppm (**Fig. 5.2.2**). Most probably, this peak belongs to either 68G or 71G, because its chemical shift value differs from the values typical for canonical WC G:C base pair. We were not able to observe resonances of a few nucleotides from the region in the middle of SL II (68G, 71G, 94G, 95G, 97U, 98G), most probably due to very short  $T_2$ . In total, 80.6% of imino resonances of all base-paired nucleotides were assigned (**Extended data. Table 9**).

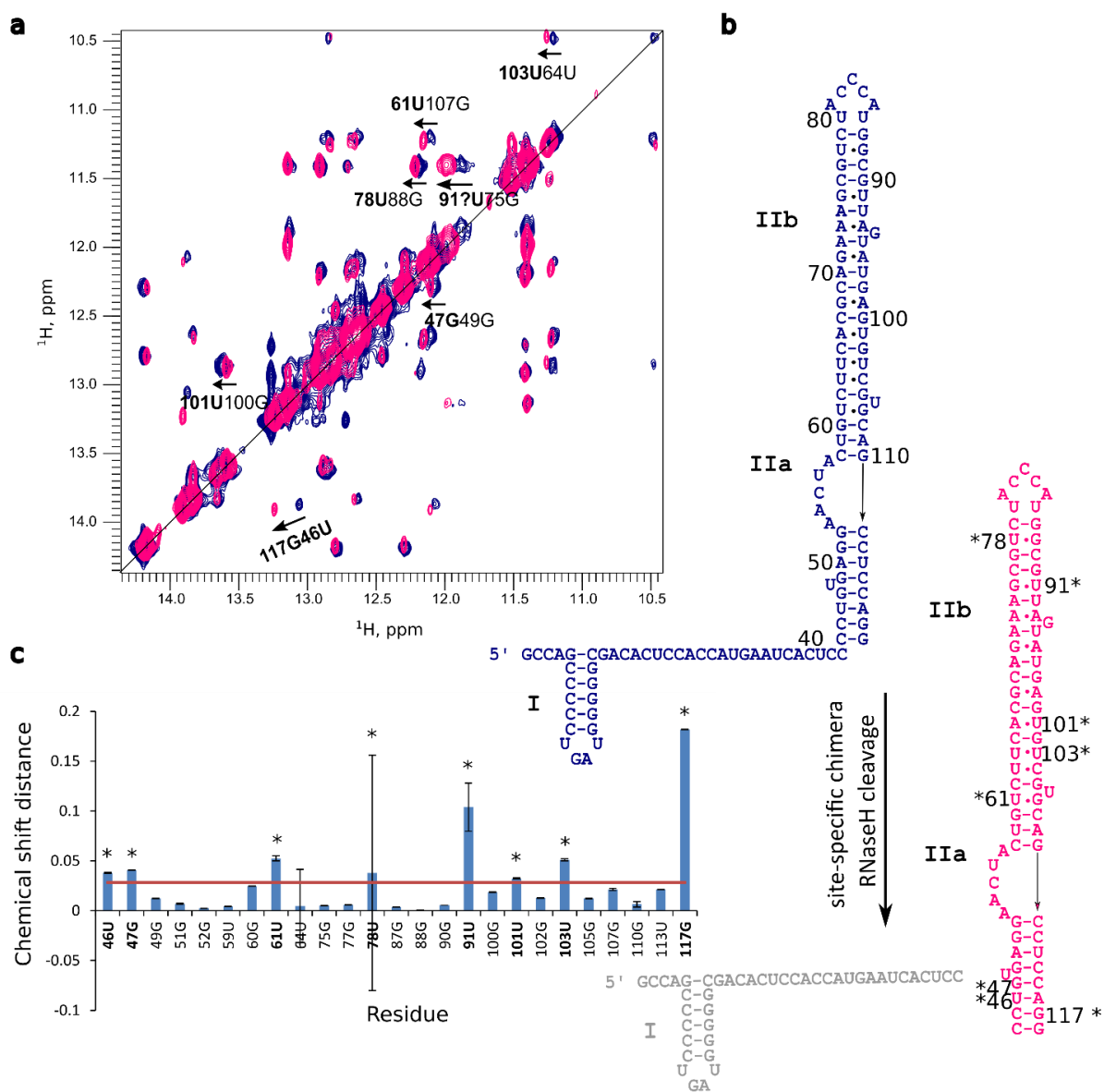


**Figure 5.2.2. NMR-derived secondary structure of domains I-II.** **a)** 2D  $^1\text{H}$ ,  $^1\text{H}$  NOESY spectrum of domains I-II acquired with a mixing time of 150 ms. NOE cross-peaks between imino protons of guanosine (H1) and uridine (H3) were used for the sequential walk. Dashed-lines display the sequential walk from 117G to 52G. **b)** Secondary structure of domains I-II derived from the NOESY assignment; arrows indicate imino contacts. Assigned nucleotides are shown in bold. **c)** 1D  $^1\text{H}$  spectra of domains I-II imino region. Assigned guanosines and uridines are in black and orange, respectively. **d)** An example of NOE contacts in RNA helix (77G, 78U 88G, 89C of domain II, PDB entry 1P5P (Lukavsky et al. 2003)). **e)** 2D  $^1\text{H}$ ,  $^{15}\text{N}$  TROSY spectrum of  $\text{G}^{\text{lab}}$  domains I-II, zoom-in on

guanosine imino region. An asterisk indicates unassigned peak, which belongs either to 68G or 71G. All measurements were carried out at a temperature of 298 K and a  $^1\text{H}$  field-strength of 850 MHz in 10 %  $\text{D}_2\text{O}$  buffer.

As expected, the residues of domain II, which show the most pronounced CSPs in domain I-II construct are located in SL II proximal to domain I in the (117G, 47G, 46U). Additionally, we observe pronounced CSPs for nucleotides 61U, 78U, 91U, 101U and 103U, which could be explained by *tertiary* but not secondary conformational changes in domain II in the presence of domain I (**Fig. 5.2.3 a**).

To provide a quantitative estimate of conformational changes in domain II in the presence of domain I, we have calculated CSPs according to equations (51) and (52) (**Fig. 5.2.3 c**).

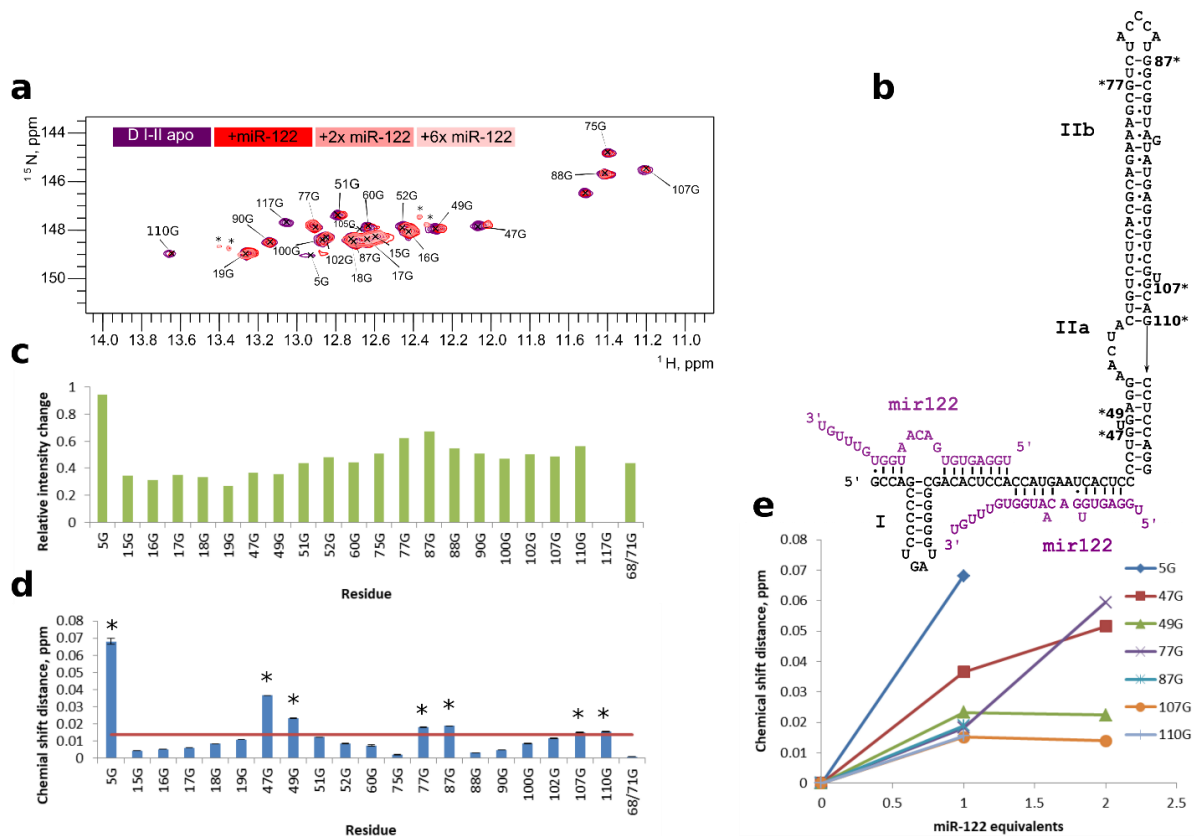


**Figure 5.2.3. Comparison of imino-imino proton NOESY-derived secondary structures of domains I-II and domain II. a) 2D  $^1\text{H}$ - $^1\text{H}$  NOESY spectrum of domains I-II (blue) and domain II**

(magenta) acquired at a temperature of 298K in 10 % D<sub>2</sub>O with a mixing time of 150 ms. The residues that show CSP larger than the average value, are labeled (46U, 47G, 61U, 78U, 91U, 101U, 103U, 117G). **b)** Secondary structure of domains I-II (blue) and domain II (magenta). Domain II construct was produced by site-specific chimera-aided RNaseH cleavage between nucleotides C43 and C44. **c)** CSP mapping of isolated domain II and domain II in domain I-II construct. The residues, which show CSP larger than the average value (brown line), are shown in bold and labeled with asterisks. Uncertainties in CSP values were calculated according to equations (51) and (52).

### ***Does miR-122 binding affects domain II structure?***

To investigate, how miR-122 binding affects the structure of domains I-II, we have recorded a series of <sup>1</sup>H-<sup>15</sup>N TROSY spectra of G<sup>lab</sup> domains I-II RNA titrated with unlabeled miR-122, following their CSPs and relative peak intensity changes (**Fig. 5.2.4 a**). The most pronounced CSPs were observed for basal iminos: 47G and 49G. In addition to CSPs, there was overall reduction of resonances intensity, due to (i) higher molecular weight of the complex and therefore slower tumbling rate, and (ii) most probably some reduction of flexibility. To map CSPs for unambiguously assigned imino groups, we have combined chemical shift changes of both imino protons and nitrogens. Although miR-122 does not interact with domain II, we observe pronounced (on the average >0.03 ppm) CSPs for some of imino resonances (47G, 49G, 77G, 87G, 107G, 110G) of the G<sup>lab</sup> domains I-II construct, which happens up to the addition of the second equivalent of miR-122 and is not any more visible after addition of 6 equivalents due to the signal attenuation (**Fig. 5.2.4 d**).



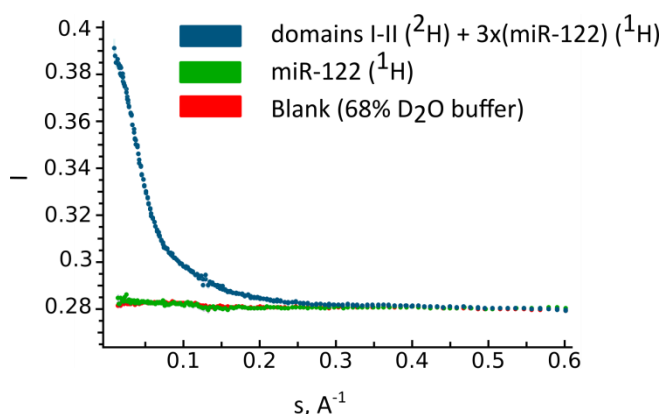
**Figure 5.2.4. miR-122 binding to domains I-II introduces CSPs of the domain II resonances. a)** Overlap of 2D  $^1\text{H}$ ,  $^{15}\text{N}$  TROSY spectra of  $\text{G}^{\text{lab}}$ -domains I-II titrated with unlabeled miR-122 (titration points: 0, 1, 2 and 6 molar equivalents).  $^1\text{H}$ ,  $^{15}\text{N}$  imino region of guanosines recorded at a temperature of 298 K in 10%  $\text{D}_2\text{O}$  buffer. The asterisks indicate the signals from the degradation products, which intensity and linewidth do not change over the time and titration series. **b)** Secondary structure of domains I-II with two copies of miR-122 bound to domain I. The nucleotides with significant CSPs upon addition of miR-122 are indicated with asterisks. **c)** Relative intensity attenuation of imino resonances of domains I-II upon addition of one equivalent of miR-122. **d)** CSPs in domains I-II upon the addition of one equivalent of miR-122. The residues with CSPs larger than the average value (brown line) are labeled with asterisks. **e)** CS mapping of the nucleotide resonances, which show significant CSPs upon miR-122 addition. Uncertainties in CSP values were calculated according to equations (53) and (55).

Summarizing, our NMR experiments on the domains I-II construct cannot provide any evidence of the alternative fold of domain II in the domain I-II construct, due to the preservation of sequential walk pattern in domain II. Furthermore, no major changes in domain II were observed upon binding of two copies of miR-122 to domain I. Still, we could observe some CSPs in domain II upon miR-122 binding to domains I-II, which could be a sign of some tertiary conformational changes.

### 5.2.3. Characterization of conformational changes in domains I-II upon miR-122 binding by small angle neutron scattering

Since solution-state NMR have shown that domain II does not adopt the alternative secondary structure and does not show any signs of secondary structure perturbation upon miR-122 binding, we aimed to obtain low resolution structural data on domains I-II apo and holo using SANS to gain further insides into the role of miR-122. The CSPs observed for domain II resonances indicate that miR-122 binding to domains I-II may introduce conformational changes in domain II or alter relative orientation of domains I and II.

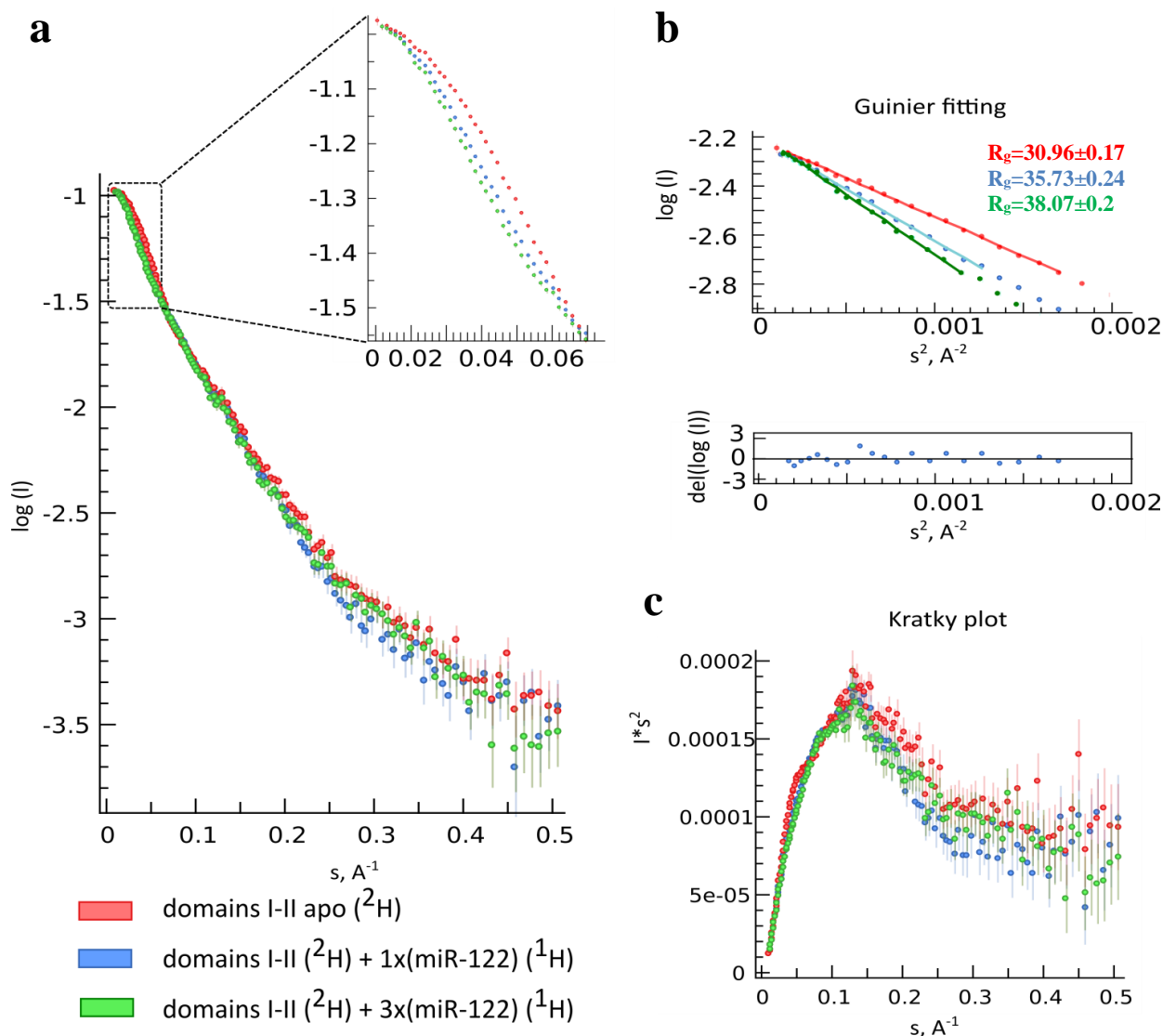
To exclude the impact of the miR-122 on the scattering, we used the advantage of contrast matching in SANS, so that domains I-II RNA was fully deuterated and miR-122 was 100% protonated and therefore matched out, i.e., invisible in 68% D<sub>2</sub>O buffer (**Fig. 5.2.5**). Since small angle scattering techniques are highly sensitive to large particles (e.g., aggregates and residual polyacrylamide), after the PAGE purification we have thoroughly washed samples on Amicon® Ultra Centrifugal Filters and additionally purified them via SEC. As a result, we have not observed any signs of sample aggregation or polyacrylamide contamination.



**Figure. 5.2.5. The contrast matching in SANS of domains I-II.** Overlay of scattering intensity curves of deuterated domains I-II (<sup>2</sup>H) titrated with 3 equivalents of protonated miR-122 (<sup>1</sup>H) (blue), protonated miR-122 (<sup>1</sup>H) (green) and buffer (red). The scattering signal from the protonated RNA is perfectly matched out in 68% D<sub>2</sub>O buffer.

SANS curve changes considerably already after addition of one molar equivalent of miR-122, indicating conformational change, with further changes occurring after the complete saturation of domains I-II with miR-122 (+ 4 equivalents of miR-122) (**Fig. 5.2.6 a**). Both increase of R<sub>g</sub> (**Fig. 5.2.6 b**) and the change of the scattering curve shape upon miR-122 addition indicate more open and extended conformation of domains I-II in complex with miR-122 in comparison to the free, apo state. Since the I(0) intensity does not change within the titration series, the oligomerization of the domains I-II can be excluded.





**Figure 5.2.6. Domains I-II–miR-122 complex monitored by SANS.** Deuterated domains I-II ( $^2\text{H}$ ) is titrated with protonated miR-122 ( $^1\text{H}$ ) with 0 (red), 1 (blue), and 3 (green) equivalents in 68%  $\text{D}_2\text{O}$  buffer. **a)** Overlay of scattering intensity curves. Experimental data is shown as dots with error bars. The region with relevant conformational changes is zoomed-in. **b)** The Guinier plot of the titration series. The fitting curves are shown as solid lines and experimental points are shown as dots. The Guinier region of every measurement is linear, indicating absence of the interparticle interaction in all samples. The  $R_g$  value of domains I-II increases upon addition of miR-122. Below the plot with residuals shows flat and random distribution around zero. **c)** The Kratky plots overlay depicts a typical profile for a partially folded or flexible molecule.

The Kratky plot provides qualitative information about the degree of conformational disorder in biomacromolecules. All studied samples show a shape that is typical for partially unfolded or flexible molecule (**Fig. 5.2.6 c**)

#### 5.2.4. RNA modeling

For structural modeling of apo domains I-II, domains I-II:(miR-122)<sub>1</sub> and holo state (domains I-II:(miR-122)<sub>2</sub>) by RNAMasonry I have tested secondary structure restraints with different number of base pairs (**Table 5.2.3**). It was appropriate to not restrain terminal nucleotides of stem-loop in domain II (44C:118G (w/o last bp) and 44C:118G, 45C:117G (w/o last two bp)) due to the absence of the observable NMR peak for corresponding imino-protons. A good model was determined as a structure with  $\chi^2 < 2$ , without structural clashes, chain brakes and undesired base-pairing. Indeed, the best structures for apo domains I-II were found from runs with the secondary structure without 44C:118G base pair (w/o the last base pair), while for domains I-II:(miR-122)<sub>1</sub> and domains I-II:(miR-122)<sub>2</sub> complexes the best structures were obtained with the secondary structure, where both 44C/118G and 45C/117G were not base-paired.

**Table 5.2.3. Secondary structure input** for the modeling of apo domains I-II RNA, domains I-II:(miR-122)<sub>1</sub> and domains I-II:(miR-122)<sub>2</sub> by RNAMasonry tool.

Name	Secondary structure (5'→3')
All bp	....((((((((.....)))))).....((((((((((((((((((((((((((((((((((((.....))))))))))))))))))))))))))))))
W/o last bp	....((((((((.....)))))).....((((((((((((((((((((((((((((((((((((.....))))))))))))))))))))))))))))))
W/o last two bp	....((((((((.....)))))).....((((((((((((((((((((((((((((((((((((.....))))))))))))))))))))))))))))))

#### Domains I-II apo

For apo domains I-II I have obtained two models, which fit well to the SANS data with  $\chi^2$  values of 1.91 and 1.47, correspondingly (**Fig. 5.2.7 a**).

#### Domains I-II:(miR-122)<sub>1</sub>

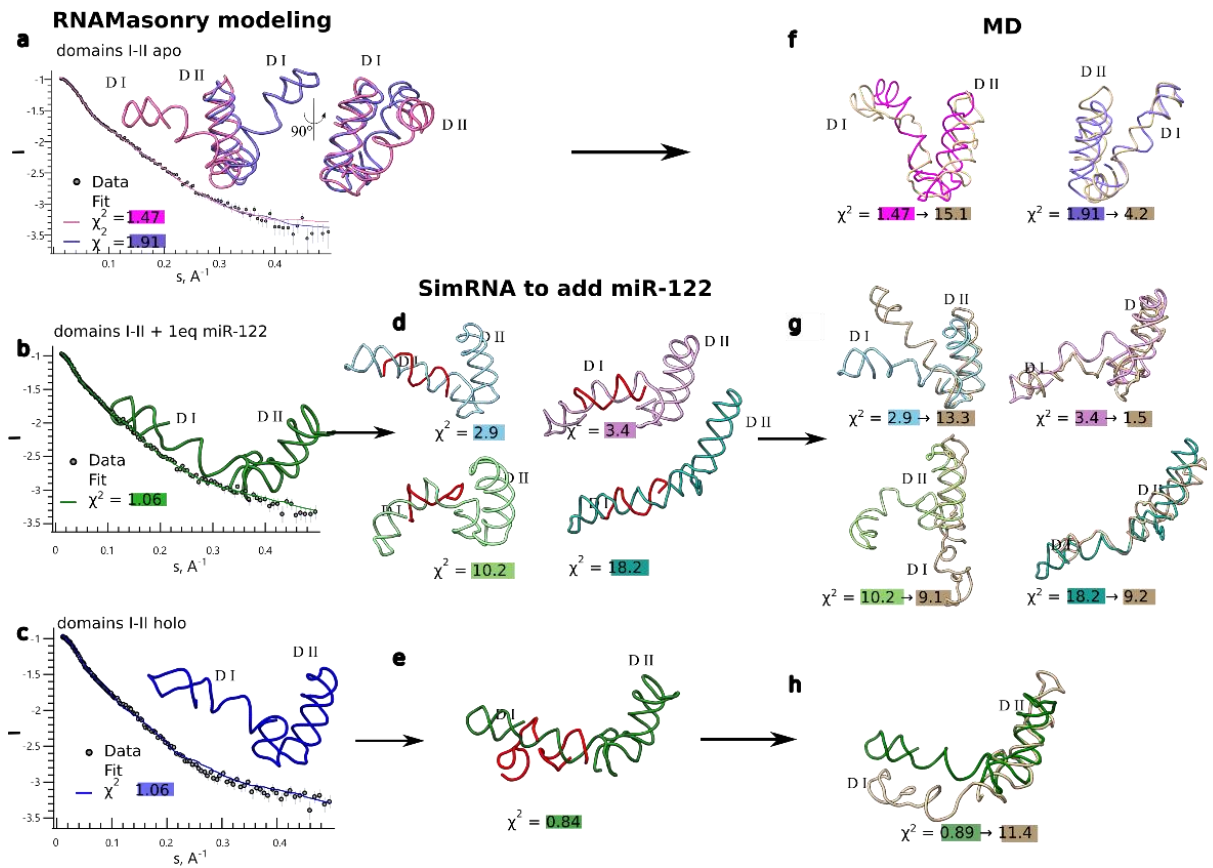
Different secondary structures inputs for domains I-II:(miR-122)<sub>1</sub> provided models with either undesired additional base pairing, non-helical linker between stem-loops I and II and/or way too high  $\chi^2$  values (>10). I have used an option in RNAMasonry tool, that allows to perform the modeling using some starting structure The model of domains I-II:(miR-122)<sub>2</sub> (explained below) was used as a starting structure. The final model has  $\chi^2$  value of 1.23, had correct secondary structure and a helical linker to position miR-122 (**Fig. 5.2.7 b**).

#### Domains I-II:(miR-122)<sub>2</sub>

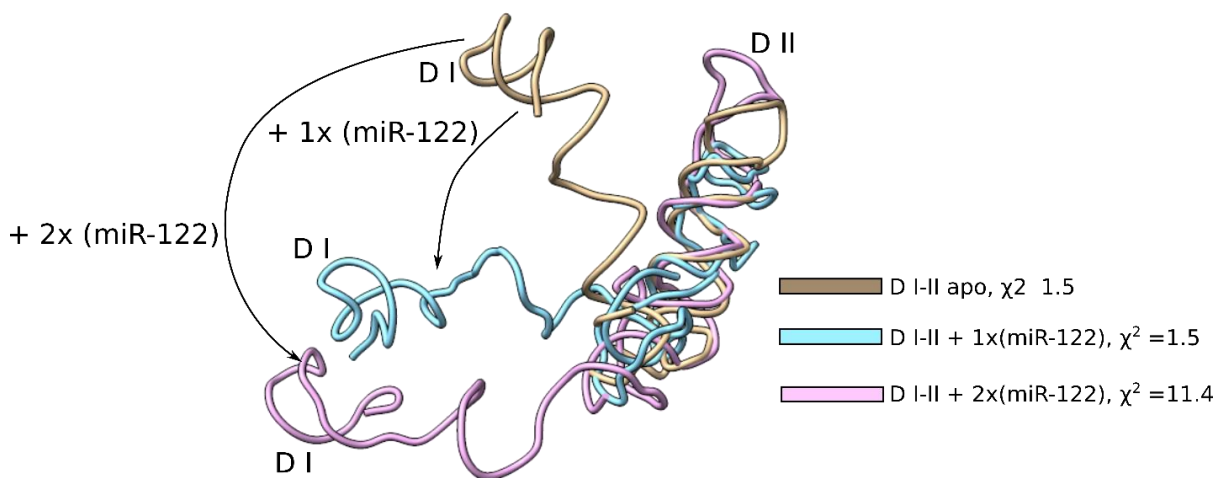
To build the model of domains I-II in holo (based on the SANS scattering curve of domains I-II + 4 x miR-122) I have tested several secondary structure inputs, including all base pairs, without bp 44C:118G (last base pair) and without 44C:118G, 45C:117G (last two base-pairs). The best model (without undesired base pairs,  $\chi^2$  of 1.06 and helical linker between stem-loops I and II, which can accommodate two copies of miR-122) was built with the input without the last two base pairs (**Fig. 5.2.7 c**).

Good models of domains I-II:(miR-122)<sub>1</sub> and domains I-II:(miR-122)<sub>2</sub> were used to add one or two copies of miR-122 in ChimeraX as one or two A-helical chains and then were subjected to SimRNA to position them to the seed region 1 and 2. Several possible inputs were given to SimRNA. The model for domains I-II:(miR-122)<sub>1</sub> with  $\chi^2$  in a range of 2.9 to 18.2 were obtained with the input, where the entire domains I-II RNA was frozen and miR-122 was paired via seed region only (2-8nt) (**Fig. 5.2.7 d**). The best model with  $\chi^2$  of 0.84 for domains I-II:(miR-122)<sub>2</sub> was obtained with the input, where domain II (44-118nt) was frozen and both miR-122 copies were paired via seed region (2-8nt) (**Fig. 5.2.7 e**),

These models were subjected to MD simulations to explore wider conformational space (**Fig. 5.2.7 f-h**). Typically, three different MD runs were performed: regular MD, accelerated MD (aMD) to enhance conformational space sampling and aMD with extra energy variation. The structures with minimal  $\chi^2$  values were chosen.



**Figure 5.2.7. Modeling of domains I-II in apo, domains I-II:(miR-122)<sub>1</sub> and domains I-II:(miR-122)<sub>2</sub> complexes. a-c) RNAMasonry models built with experimental secondary structure and SANS restraints. d-e) SimRNA models with added miR-122 (red). f-h) Additional models after MD simulations with minimal  $\chi^2$  value in beige overlapped with starting models from d, e), miR-122 is omitted for clarity.**



**Figure 5.2.8. Overlay of domains I-II (D I-II) models** in apo, domains I-II:(miR-122)<sub>1</sub> and domains I-II:(miR-122)<sub>2</sub>. Arrows indicate conformational change upon miR-122 binding towards open orientation of domains I and II. miR-122 is omitted for clarity.

Apo and holo models mainly differ in relative orientation of domain I and domain II, resulting in more open and extended conformation of domains I-II in complex with miR-122 in comparison to the apo state. **Fig. 5.2.8** shows an overlay of the models for domains I-II (D I-II) in apo with minimal  $\chi^2$  value after RNAMasonry modeling and for domains I-II:(miR-122)<sub>1</sub> and domains I-II:(miR-122)<sub>2</sub> complexes after MD simulation to account for the presence of miR-122. The conformational change of domain I-II towards open orientation of domains I and II upon miR-122 binding is obvious.



## 5.3. 5' UTR

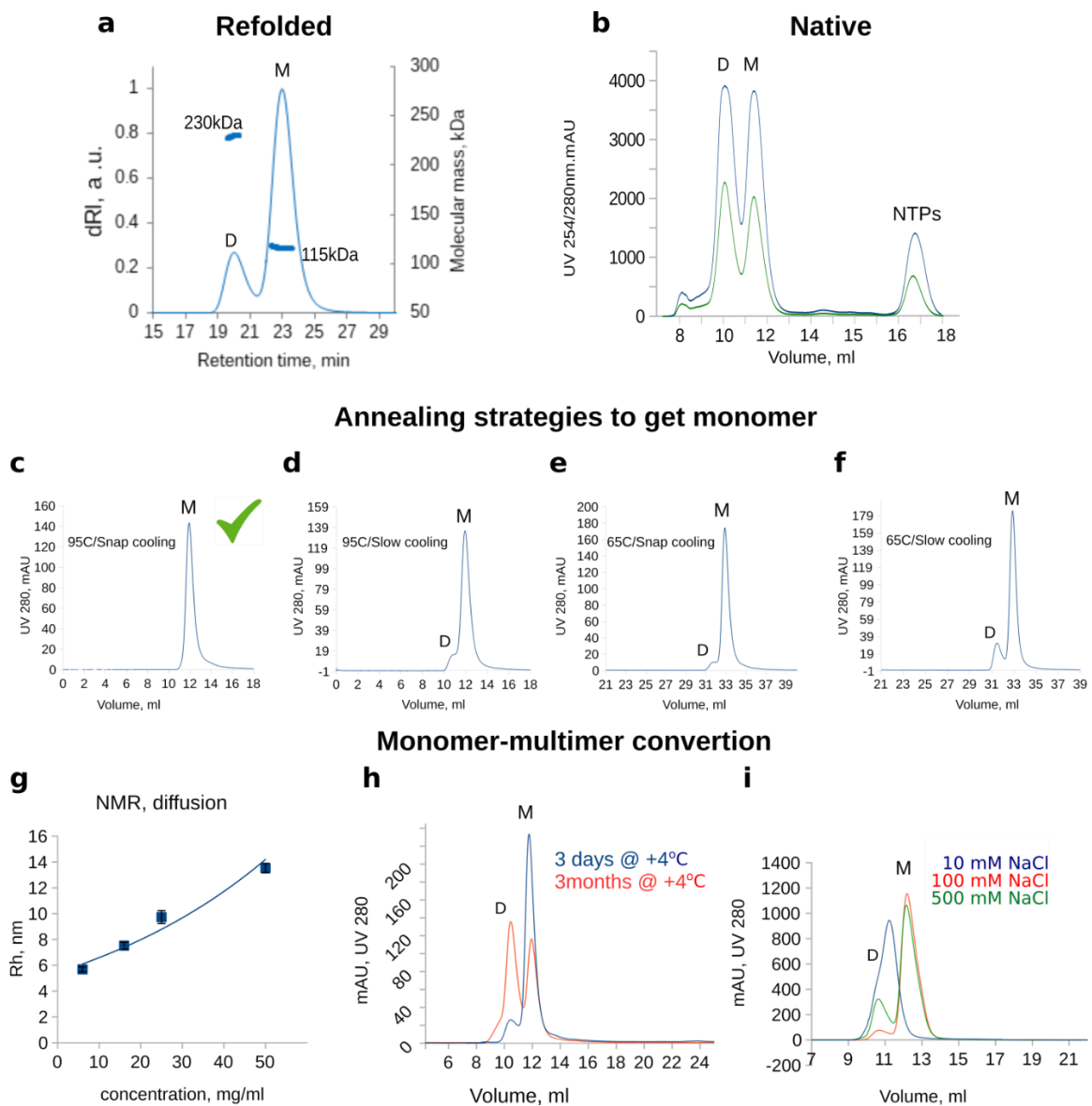
### 5.3.1. HCV 5' UTR binds two copies of miR-122 without mayor structural changes

#### *Sample preparation*

As other RNA constructs, 5' UTR has been synthesized *in vitro* from PCR template with two 2'-O-methylated nucleotides at the 5'-end of the template strand, which serve to improve 3'-end homogeneity of the transcript (Kao et al. 1999). However, no ribozyme was introduced at 5'-end, as neither preparative PAGE nor SEC could provide sufficient separation of cleaved and uncleaved RNA due to the large size of 5' UTR (359 nt).

SEC-MALS has revealed a tendency towards 5' UTR dimerization *in vitro* after our standard denaturing RNA purification protocol (**Fig. 5.3.1 a**). Although, in some cases multimerization of RNA has been shown to have an important function (Jambor, Brunel, and Ephrussi 2011; Moore and Hu 2009), to the best of our knowledge, for HCV 5' UTR it was not reported to occur *in vivo*. The non-physiological oligomerization of RNA is known nuisance in the laboratory that impede structural studies (Bou-Nader and Zhang 2020). We set out to get the monomer, establishing an annealing procedure and switching to native purification approach.

It is believed that most of the *in vitro* multimers are formed via complementary base-pairing of short fragments during annealing procedure after denaturing purification. Firstly, we have tested whether native purification protocol would yield only a monomer of 5' UTR RNA. The explicit protocol is described in Materials and methods chapter. Shortly, the procedure consists of removal of T7 polymerase and TIPP from the transcription mixture by PCI/CI extraction, purification on the desalting column to exchange the buffer and remove the traces of PCI/CI, followed by the final SEC step on S200 column. Secondly, we have tested different annealing strategies (**Fig. 5.3.1 c-f**) including heating to 95 °C or 65 °C followed by snap or slow cool down directly in the working buffer or first in H<sub>2</sub>O and consecutive addition of the refolding buffer. While native purification approach yields a monomer-dimer mixture (**Fig. 5.3.1 b**) similarly to the denaturing purification with refolding (**Fig. 5.3.1 a**), an annealing of 5' UTR in H<sub>2</sub>O, followed by addition of ice-cold 2x buffer has yield monomeric RNA (**Fig. 5.3.1 c**). However, the obtained monomer tends to oligomerize at higher concentrations (**Fig. 5.3.1 g**) and has not demonstrated stability over the time (**Fig. 5.3.1 h**). Furthermore, the salt concentration in the buffer influences the dimer-monomer ratio as well (**Fig. 5.3.1 i**).



**Figure 5.3.1. 5' UTR oligomerization *in vitro*.** Both denaturing and native purification protocols yield a mixture of a dimer (D) and monomer (M). **a)** SEC-MALS of 5' UTR purified by UREA-PAGE and heat refolded in 1x buffer (100 mM HEPES, pH 7.5, 100 mM NaCl, 5 mM MgCl<sub>2</sub>) at 95 °C. **b)** SEC of transcription reaction mixture purified natively. **c)-f)** SEC of 5' UTR with different annealing strategies. **c)** RNA at concentration 1.2 mg/ml in H<sub>2</sub>O was heated to 95 °C for 5 minutes and cooled by addition of ice-cold 2x buffer. **d)** RNA at concentration 1.2 mg/ml in H<sub>2</sub>O was heated to 95 °C for 5 minutes, slowly cooled down and equal volume of ice-cold 2x buffer was added. **e)** RNA at concentration 0.6 mg/ml in 1x buffer was heated to 65 °C for 10 mins, and snap-cooled on ice. **f)** RNA at concentration 0.6 mg/ml in 1x buffer was heated to 65 °C for 10 mins, and slowly cooled. Condition e) yield single peak with an elution volume corresponding to monomeric 5' UTR. **g)** The NMR diffusion experiment at different concentrations of 5' UTR. **h)** Overlay of SEC runs of 5' UTR stored at 4 °C over 3 days and 3 months. **i)** Overlay of 5' UTR SEC runs in buffer with different salt concentration. All SEC runs were performed on Superdex ® 200 Increase 10/300 GL column.

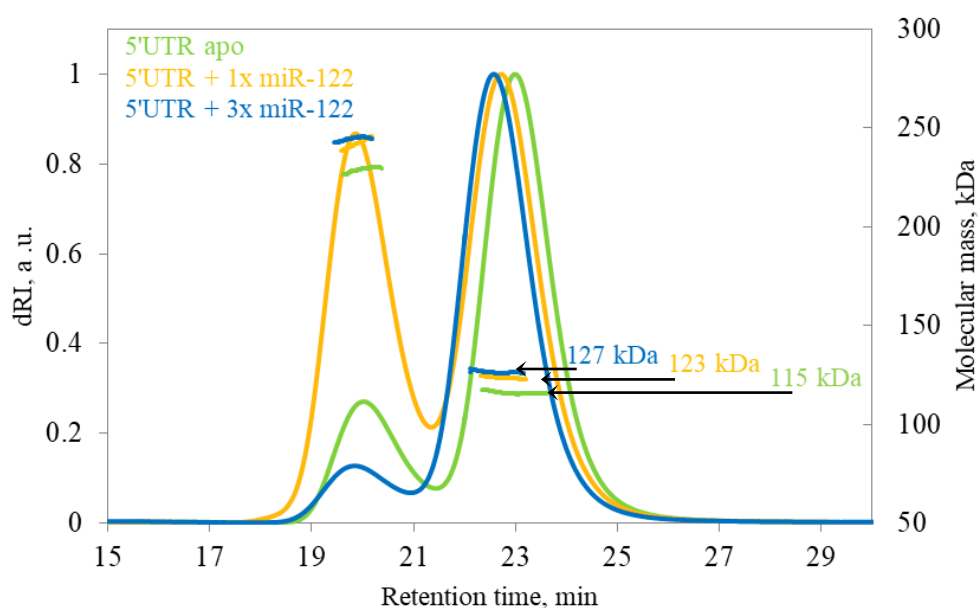


Nevertheless, since the monomer of 5' UTR can be properly separated from the dimer with SEC and remains monomeric for several days, we could perform structural studies both with SEC-SAXS and ssNMR.

### 5.3.2. Stoichiometry of the 5' UTR–miR-122 interaction

In order to confirm the stoichiometry of the 5' UTR–miR-122 interaction, we have performed SEC combined with online detection by multi-angle light scattering and refractometry (SEC-MALS). As expected the changes of both molecular weight and elution time occur upon miR-122 binding (**Fig. 5.3.2**).

5' UTR is present in well separated monomeric and dimeric forms, while both are able to bind two or four copies of miR-122. For the holo sample (5' UTR + 3 equivalents of miR-122) molecular weight difference was equal to 1.5 copy of miR-122, indicating that the full saturation of the second binding site probably was not reached. Nevertheless, the peak of free miR-122 was observed. This incomplete saturation can be explained by previously determined low binding affinity of the miR-122 binding site 1 (Mortimer and Doudna 2013).



**Figure 5.3.2. SEC-MALS of 5' UTR upon titration with miR-122.** Overlap of elution profiles of 5' UTR apo (MW=115.2 kDa), 5' UTR+1 equivalent of miR-122 (MW=122.6 kDa) and 5' UTR+3 equivalents of miR-122 (MW=129.9 kDa). 5' UTR present as a mixture of dimer (left peak) and monomer (right peak). SEC column is Superdex® 200 Increase 10/300 GL

### 5.3.3. miR-122 binding to 5' UTR does not introduce significant conformational changes in 5' UTR

#### *Small angle X-ray scattering (SAXS)*

After confirming that 5' UTR binds two copies of miR-122, we set out to determine, whether miR-122 binding introduces conformational changes in 5' UTR. We have performed small angle X-ray scattering (SAXS) in-line with SEC to exclude the scattering effect of the dimer (Fig. 5.3.3 a). Scattering curves recorded from the apo 5' UTR and 5' UTR mixed with miR-122 in molar ratios 1:1 and 1:3 are not identical (outside of Guinier region), but nearly similar (Fig. 5.3.3 b),  $R_g$  and  $D_{max}$  values vary within 3% for apo, holo and 5' UTR in complex with one copy of miR-122, indicating that no major structural conformational rearrangements happen in 5' UTR upon miR-122 binding that can be detected using low resolution envelope of SAXS (Table 5.3.1).

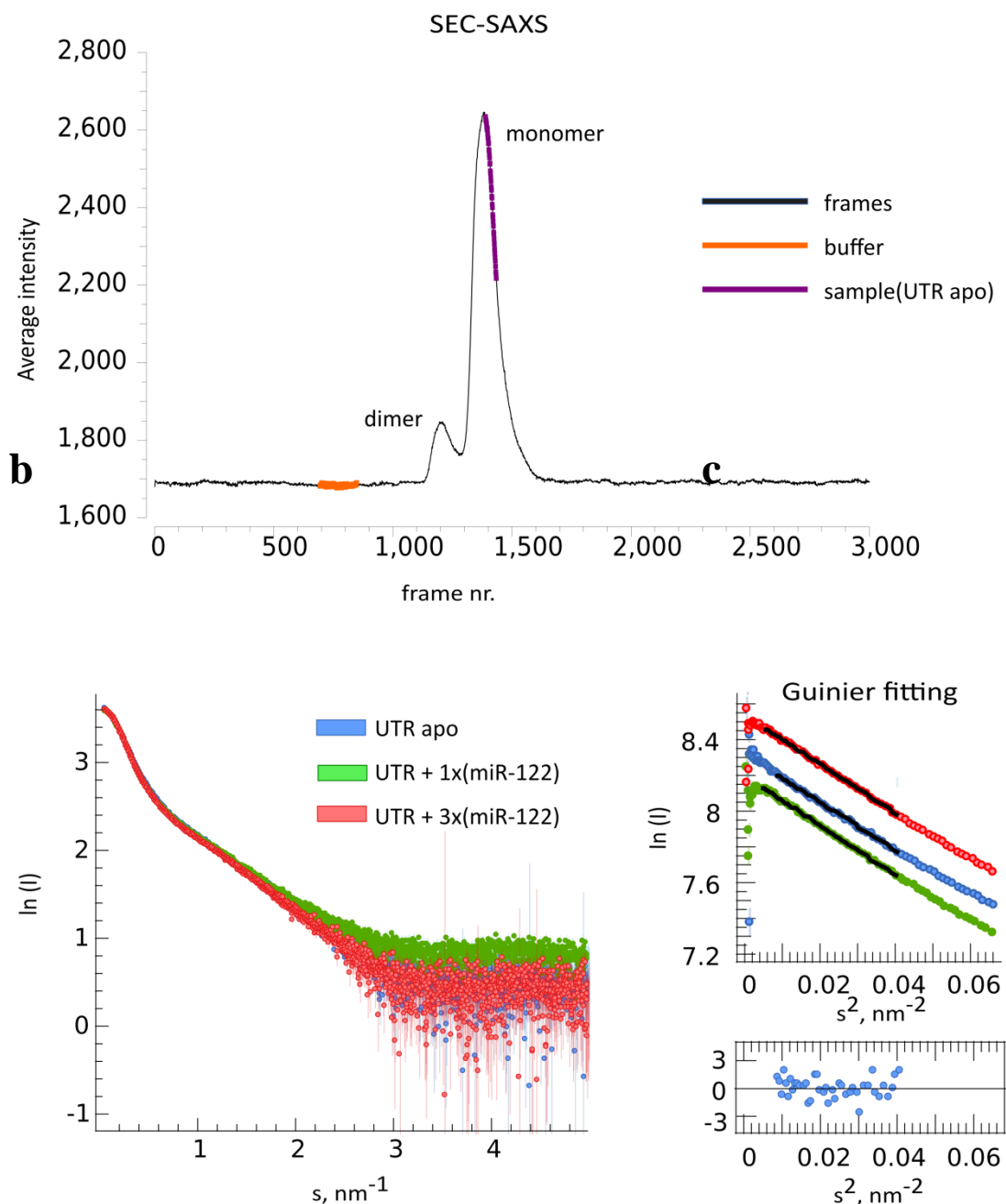


Figure 5.3.3. SEC-SAXS of 5' UTR titrated with miR-122. a) Elution profile of 5' UTR apo

shown as a plot of the average scattering intensity plotted against frame number. The region colored in orange are the frames chosen for the buffer subtraction, purple – monomeric sample region frames. The samples with added one and three copies of miR-122 were processed similarly and are not shown here. **b)** Overlap of scattering intensity plots of 5' UTR titration series without added miR-122 (apo) and with 1 and 3 equivalents of miR-122. **c)** The Guinier fitting for 5' UTR apo, 5' UTR +1x miR-122 and 5' UTR +3x miR-122. The Guinier region is linear. Below is the plot with residuals showing flat and random distribution around zero.

Sample	R <sub>g</sub> , nm	D <sub>max</sub> , nm
UTR apo	6.34±0.1	23.35
UTR +1x miR-122	6.47±0.1	23.30
UTR +3x miR-122	6.40±0.1	23.75

**Table 5.3.1. SEC-SAXS of 5' UTR titrated with miR-122** (titration points: 0, 1 and 3 equivalents). R<sub>g</sub> is calculated by Primus Guinier Wizard and D<sub>max</sub> values are selected manually in Primus distance distribution wizard (Manalastas-Cantos et al. 2021).

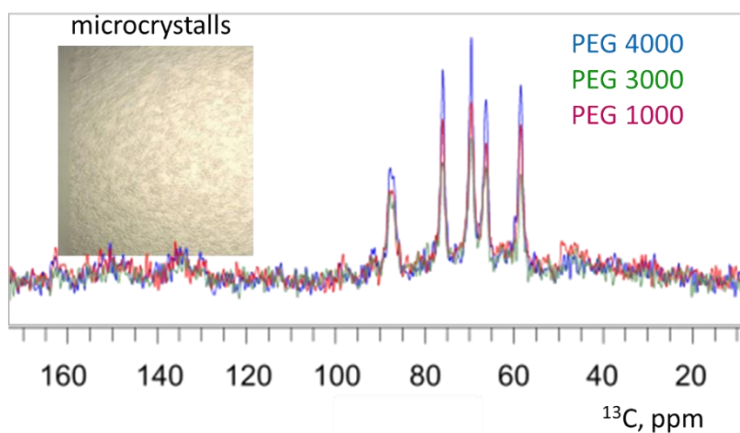
### ***Solid-state Nuclear Magnetic Resonance (NMR)***

To the best of our knowledge, there is no high-resolution structural data on the full-length 5' UTR–miR-122 complex yet. Its large size (115 kDa) and flexibility due to the presence of single-stranded stretches, which connect several hairpin loops, make 5' UTR a challenging object for structural studies either by cryo-EM, x-ray crystallography or solution-state NMR. To investigate structural changes in 5' UTR upon miR-122 binding, we have attempted to use solid-state NMR technique. ssNMR has no intrinsic molecular weight limitations and it is widely applied to study large protein assemblies. Lately, ssNMR was demonstrated to be a promising tool for structural studies of RNA, though the methodology is still under development (Aguion and Marchanka, 2021).

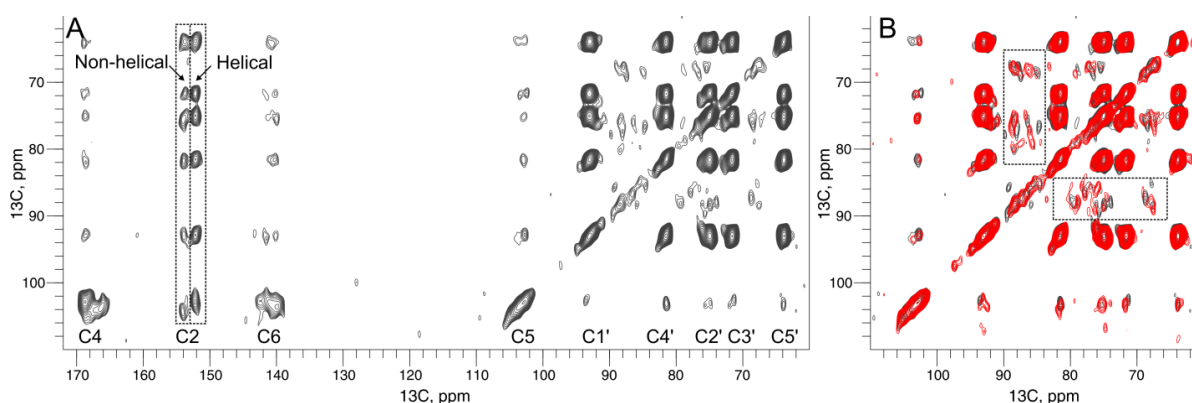
First, we have optimized sample preparation of 5' UTR (359 mer) for ssNMR. One of the most common approaches to prepare an ssNMR sample is the microcrystallization; this technique can be also applied for RNA (Huang et al. 2012; Marchanka et al. 2015). We have set up several crystallization screens and have obtained micro/nanocrystals at many different conditions. To select the condition that was best suited for ssNMR we have acquired 1D <sup>13</sup>C, <sup>1</sup>H cross-polarization spectra on three samples that were crystallized with different PEG in the crystallization buffer aiming for best quality spectra. A slightly better spectrum has been obtained with 30% PEG4000 100 mM HEPES, pH 7.5, 100 mM NaCl buffer, these conditions were then used for the subsequent experiments (**Fig. 5.3.4**).

First, we have acquired 2D <sup>13</sup>C, <sup>13</sup>C PDS spectra of full <sup>13</sup>C, <sup>15</sup>N-uridine-labelled 5' UTR (U<sup>lab</sup>-5' UTR) in both apo- (free) and holo-state (with two copies of miR-122 bound) (**Fig.**

5.3.5). We have chosen uridine- specific  $^{13}\text{C}$ ,  $^{15}\text{N}$  labeling, since uridines are the prevalent nucleotide type in non-helical regions of 5' UTR.



**Figure 5.3.4: Optimization of ssNMR sample preparation.** 1D  $^{13}\text{C}$  CP ssNMR spectra of unlabeled 5' UTR crystallized under different PEGs. Crystallization buffers contained: 100 mM HEPES pH 7.5, 100 mM NaCl, 5 mM  $\text{MgCl}_2$  and PEG 4000/3000/1000.



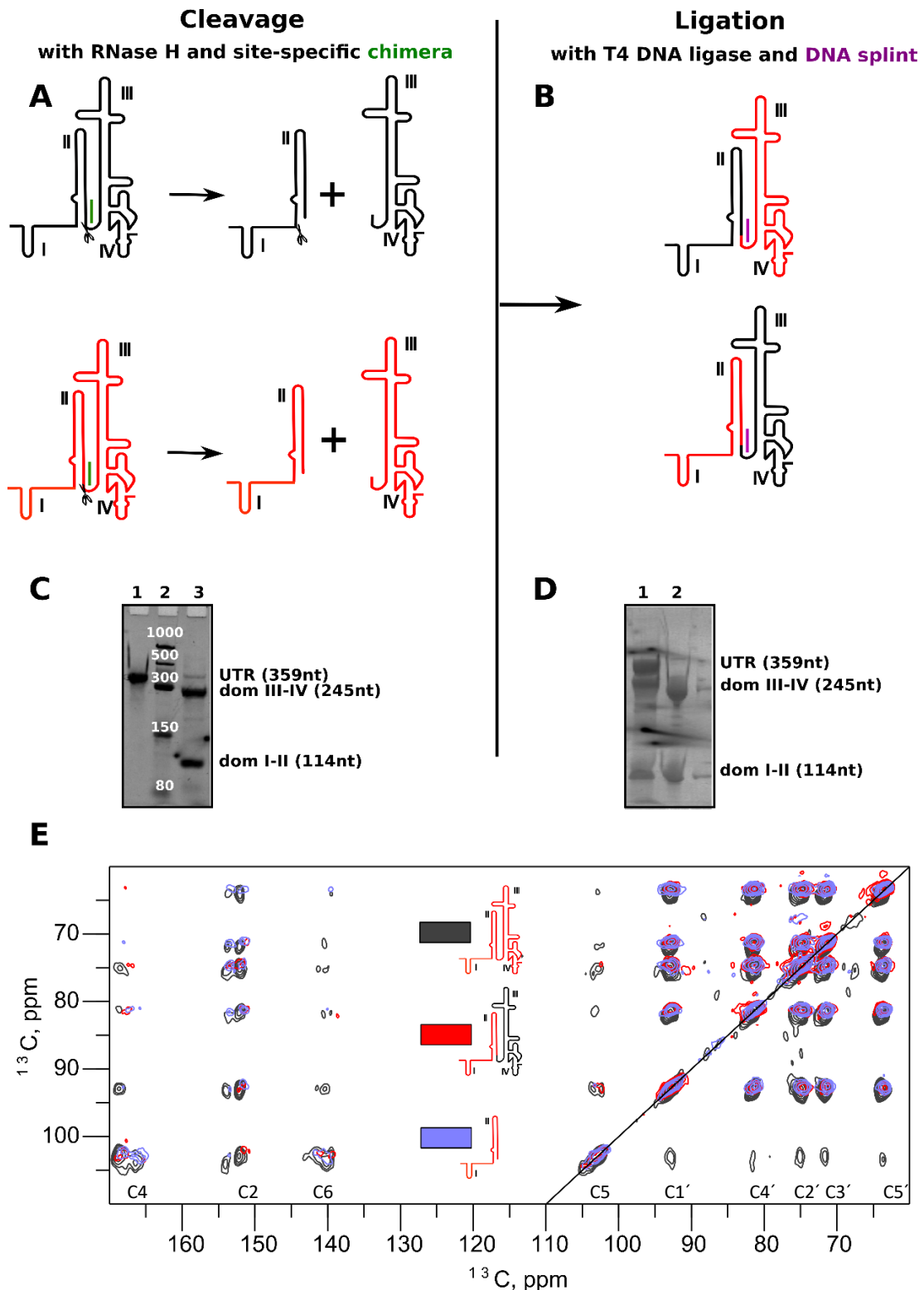
**Figure 5.3.5. 2D  $^{13}\text{C}$ ,  $^{13}\text{C}$  PDSM spectra of  $\text{U}^{\text{lab}}$ -5' UTR RNA with a microcrystalline sample preparation** acquired at a temperature of 270 K on a 600 MHz Bruker Avance III WB spectrometer equipped with a 2.5 mm  $^1\text{H}/^{13}\text{C}$  MAS probehead. A total of 1.5 mg of RNA (300  $\mu\text{g}$  of labeled material) was used. The PDSM mixing time was 100 ms. The line width of separate resonances of non-canonical RNA elements is estimated to be 1.0 ppm. **A)** Full spectrum. The dashed box shows two different regions of C2 chemical shifts that can be tentatively assigned to helical and non-helical areas. **B)** Zoom-in on the ribose region. Reproducible chemical shift changes (dashed boxes) are observed between the apo 5' UTR (gray) and 5' UTR upon the addition of 3 equivalents of unlabeled miR-122 (red), providing strong evidence for conformational changes in the 5' UTR.

The 2D  $^{13}\text{C}$ ,  $^{13}\text{C}$  PDSM spectrum of the  $\text{U}^{\text{lab}}$ -5' UTR shows a promising  $^{13}\text{C}$  line width of 1.0 ppm for separate resonances located in non-canonical secondary structure elements (loops, bulges). Due to the high number of uridines (74) in the 5' UTR RNA, site-specific assignment is not possible with this preparation. Reduction of spectral crowding could in principle be achieved by implementation of segmental labeling of RNA as described below. However, even with this simple preparation, two regions of C2 chemical shifts in bases are

identified that can be tentatively assigned to helical and non-helical areas (**Fig. 5.3.5 A**) (Farès, Amata and Carlomagno, 2007). Moreover, clear changes are observed in chemical shifts upon binding of microRNA, indicating conformational changes in the 5' UTR (**Fig. 5.3.5 B**).

It is not clear at this stage, whether these observed chemical shifts originate from domain I or the rest of 5' UTR. To answer this question and overcome the signal overlap we attempted to label 5' UTR segmentally.

Although, the protocol for segmental labeling for the first two domains of 5' UTR has been optimized (**Fig. 5.3.6 a-d**), we were not able to acquire ssNMR spectra with reasonable resolution neither of U<sup>lab</sup>-domains I-II covalently bound to unlabeled domains III-IV nor U<sup>lab</sup>-domains I-II alone (**Fig. 5.3.6 e**). While for segmentally labeled full 5' UTR suboptimal spectral can be partially explained by low amount of RNA used for ssNMR experiments (140 µg), for other samples different reasons might be responsible.



**Figure 5.3.6. ssNMR on segmentally labeled 5' UTR.** A-D) Segmental labeling scheme. **A)** Site-specific RNase H cleavage of 5' UTR with RNA-DNA chimera (green) between nucleotide position 114 and 115. Unlabeled RNA and  $^{13}\text{C},^{15}\text{N}$ -uridine-labelled RNA are shown in black and red, respectively. **B)** Ligation of labeled and unlabeled segments using DNA splint (green), final product yield was 1.3mg. **C)** 10% denaturing PAGE of the cleavage reaction. Lane 1: 5' UTR (359mer), lane 2: ladder, lane 3: cleavage reaction. **D)** 10% denaturing PAGE of the ligation reaction. Lane 1: ligation reaction of  $^{13}\text{C},^{15}\text{N}$ -uridine-labelled domains I-II (114nt) and unlabelled domains III-IV (245nt), lane 2: ligation reaction mix *before* addition of the ligase. **E)** Overlay of 2D  $^{13}\text{C},^{13}\text{C}$  PDS

spectra (200 ms mixing time) of U<sup>lab</sup>-5' UTR RNA (gray), segmentally labeled U<sup>lab</sup>-domains I-II-unlabelled domains III-IV (red) and isolated U<sup>lab</sup>-domains I-II (blue). All spectra were acquired on a 600 MHz spectrometer at 13 kHz MAS. A total of 1.5 mg, 2.1 mg and 1.2 mg of RNA (300, 140, 80 µg of labeled material) was packed into ssNMR rotor.

At this point we had to conclude, that the structure of this RNA is hardly accessible by ssNMR method in the timeframe of the project. It could be speculated, that partial or complete loss of tertiary structure of RNA during the sample preparation, insufficient hydration of the sample, the absence of the order in microcrystals due to the high RNA flexibility, or the presence of multiple conformers, which previously were reported for free IRES in solution (Pérard et al. 2013), hinder the study of 5' UTR by ssNMR. These challenges can be addressed by advanced ssNMR sample preparation techniques, e.g., direct sedimentation of RNA into ssNMR rotor. Furthermore, acquisition of ssNMR spectra at ultra-fast MAS using <sup>1</sup>H detection can significantly improve both spectral resolution and signal-to-noise (Aguion and Marchanka 2021; Marchanka et al. 2018).

Nevertheless, using SEC-SAXS we have observed the formation of ternary complex 5' UTR:(miR-122)<sub>2</sub>. While major structural changes in 5' UTR upon miR-122 binding were not detected either by scattering or by ssNMR, small structural changes were observable by both techniques. It remains obscure at this stage if these structural changes happen only in domain I or also in domains II-IV. Further research combining segmental labeling of RNA with ssNMR, utilization of <sup>1</sup>H detected ssNMR and SANS on segmentally-deuterated domains should be able to provide the definitive answer to this question.





## 6. Conclusions and Outlook

The thesis describes the investigation of the 5' UTR–miR-122 interaction by an integrative structural biology approach at different levels of structural complexity.

The 5' UTR of HCV consists of four independently folded domains. Domain I is essential for the translation of the virus (Reynolds et al. 1995) and 115 5' terminal nucleotides (domains I-II) are indispensable for the replication of HCV (Friebe et al. 2001). Domains II-IV form an internal ribosomal entry site (IRES), the structural element responsible for the cap-independent translation initiation and positioning of viral RNA on the 40S ribosomal subunit (Tsukiyama-Kohara et al., 1992; Rijnbrand et al., 1995; Honda et al., 1999). Previous SAXS studies have showed that the free structure of IRES in solution is an ensemble of conformers corresponding to a molecule comprised of several domains whose internal structure is rigid, but which show flexibility in their positions and orientations relative to one another (Pérard et al. 2013). This type of molecular system is particularly challenging for structural studies by NMR for a number of reasons, including the limited chemical shift dispersion of RNA nucleotides, the high flexibility of the 5' UTR itself, its large molecular weight (115 kDa) and slow-to-intermediate exchange regime of the interaction with miR-122. An integrative structural biology approach, including solution and solid-state NMR techniques, small-angle scattering, structural modeling and molecular dynamic simulations allowed me to study this system at different levels of structural intricacy.

First, I studied domain I, which contains two tandem binding-sites for liver-abundant human microRNA miR-122. The domain I–miR-122 interaction leads to viral propagation by various mechanisms, such as translation and replication stimulation and protection of the viral genetic material from cellular exonucleases (Henke et al. 2008; Masaki et al. 2015; Schult et al. 2018; Sedano and Sarnow 2014). I have obtained the near-complete  $^1\text{H}$  and  $^{13}\text{C}$  resonance-assignment of the nucleobase and ribose spin-systems of domain I by solution-state NMR for the first time (**Section 5.1.3**). Furthermore, I have determined the secondary structure of domain I by NMR, which agrees both with predicted (Brown et al. 1992) and SHAPE-derived structures (Mortimer and Doudna 2013). Domain I presents a stem-loop with six canonical G:C base pairs, with the rest of the chain adopting an extended and flexible single-stranded structure.

Then, by titration of miR-122 RNA on domain, I have confirmed that miR-122 can bind to two sites on domain I simultaneously, but that the binding affinities of the two interaction-sites are different (**Section 5.1.5**). Studying the domain I–miR-122 interaction by solution-

state NMR was found to be particularly challenging due to the reduction of signal intensity and the severe line broadening of domain I resonances upon miR-122 binding; these effects are likely to be due to a combination of slow conformational exchange of domain I itself and the slow-to-intermediate exchange regime of its interaction with miR-122.

NMR data revealed that  $Mg^{2+}$  ions have an impact on the tertiary fold of domain I in the apo state, which also results in distinct interaction patterns of domain I with miR-122 in the presence and absence of  $Mg^{2+}$  ions. The EMSA and NMR data showed that in the absence of  $Mg^{2+}$  only one copy of miR-122 can be bound by domain I, and furthermore that it is not simply bound to one of the known binding sites of domain I — as suggested in previous studies (Mortimer and Doudna 2013) — but with a different binding mode altogether. In addition, the data showed that  $Mg^{2+}$  modulates the interaction regime between domain I and miR-122, changing it even more towards slow-to-intermediate exchange.

In contrast to *in vitro* conditions, the concentration of magnesium in the cell is not constant and  $Mg^{2+}$  ions exist not only in the free form but also chelated with different affinities to other cell-metabolites (Yamagami, Sieg, and Bevilacqua 2021). Magnesium ions actively participate in RNA folding processes by binding to the negatively charged phosphate backbone, neutralizing the charge and thereby allowing a more compact tertiary fold (Misra, Shiman, and Draper 2003). While the total magnesium concentration in eukaryotic cells is 15-20 mM for most of the mammalian cell-types that have been studied (Romani and Scarpa 1992), the concentration of freely available  $Mg^{2+}$  is much lower. For mammalian liver cells, the concentration of free magnesium was estimated at 0.39 mM (Corkeys et al. 1986) and 0.59 mM (Raju et al. 1989). Thus, it can be speculated that this feature of a  $Mg^{2+}$ -dependent mode of interaction between HCV and miR-122 has some biological relevance; for example, by modulating the balance between the different effects of miR-122 binding. miR-122 has been shown to play numerous roles in the lifecycle of HCV, such as stimulation of both translation and replication of the virus and also genome stabilization by protecting the viral RNA from endogenous nucleases (Henke et al. 2008; Masaki et al. 2015; Mortimer and Doudna 2013). Therefore, further studies in this direction are necessary.

Next, using solution-state NMR, I have determined the secondary structure of isolated domain II and the domain I-II construct (**Section 5.2.2**). The secondary structure of domain II, derived from a 2D  $^1H$ ,  $^1H$  NOESY experiment agrees with the structure published by Lukavsky's group (Lukavsky et al. 2003). Most importantly, domain II maintains the same secondary structure both in isolation and when part of the domain I-II construct. Furthermore, the secondary structure of domain II stays intact upon binding of two copies of

miR-122 to domain I. Nevertheless, some resonances in domain II do show CSPs both between the isolated domain II and the domain I-II construct, and upon miR-122 binding to domain I of the domain I-II construct, indicating conformational change at the tertiary structural level.

Indeed, the contrast-matched SANS data collected on the domain I-II construct titrated with miR-122 suggest that the conformational changes upon miR-122 binding result in a more extended or open conformation with a larger radius of gyration (**Section 5.2.3**).

Using experimentally derived structural restraints (NMR-derived secondary structure and SANS intensity curves), I constructed structural models of domains I-II in the apo- and holo-states; these models were built using RNAMasonry and SimRNA tools and MD-simulations in AMBER. The models reveal a conformational switch between a more closed V-shape architecture in the apo-state to an elongated and more open L-shaped structure in the holo-state (**Section 5.2.4**). Such a conformational switch is likely to be important for the correct positioning of the HCV IRES on the 40S ribosomal subunit and also the opening of 40S subunit, which is necessary for the loading of the initiation codon into the mRNA-binding channel.

Finally, I have studied possible conformational changes in the full-length 5' UTR upon miR-122 binding. I established that the 5' UTR in solution can exist as a mixture of monomer, dimer and a small fraction of higher order oligomers (**Section 5.3.1**). Whether or not the dimerization of 5' UTR is an *in vitro* artifact or has biological relevance and also happens *in vivo*, as has been reported for some RNAs (Jambor et al. 2011; Moore and Hu 2009), is a matter of speculation, but unfortunately it imposes severe restrictions on *in vitro* structural studies. I have used SEC-SAXS, a technique that allows in-line separation of monomer from dimer, and compared the scattering intensity profiles of monomeric 5' UTR in apo- and in holo-states. Since the difference between scattering curves was minimal (**Section 5.3.3**), I conclude that no major structural changes occur in the 5' UTR upon miR-122 binding, although it is possible that there are small local structural rearrangements. To explore this possibility, I have attempted to use ssNMR, which is an emerging technique for structural studies of large RNAs at atomic-level resolution.  $^{13}\text{C}$ ,  $^{13}\text{C}$  ssNMR spectra of uniformly labeled 5' UTR — which is to the best of my knowledge the largest RNA studied by ssNMR — showed chemical shift perturbations upon miR-122 binding in the non-canonical regions. I have optimized a segmental labeling approach in which the first two domains of the 5' UTR are  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled and domains III-IV unlabeled (and vice versa) to gain structural insights into the individual domains in the context of the full-length 5' UTR. Unfortunately, I was not

able to obtain good-quality spectra, which is probably due to a combination of the intrinsic flexibility and structural heterogeneity of this large RNA and the low concentration of the final samples, which was limited by the high cost of isotope-labeled materials. My attempts to record ssNMR on the 5' UTR subdomains microcrystallized under various conditions also yielded spectra with broad  $^{13}\text{C}$  linewidths (1.15–1.65 ppm), which were not narrow enough for structural studies at this stage. Here, novel sample preparation techniques, e.g., sedimentation directly into ssNMR rotor, optimized atom-specific labeling of RNA, and state-of-the-art ssNMR pulse-sequences and hardware, including  $^1\text{H}$ -detection under ultra-fast MAS, will provide significant improvements and allow further studies of this challenging RNA complex.

## 7. Bibliography

- AbouHaidar, Mounir G., and Ivan G. Ivanov. 1999. 'Non-Enzymatic RNA Hydrolysis Promoted by the Combined Catalytic Activity of Buffers and Magnesium Ions'. *Zeitschrift Fur Naturforschung - Section C Journal of Biosciences* 54(7–8):542–48. doi: 10.1515/znc-1999-7-813.
- Aguion, Philipp Innig, and Alexander Marchanka. 2021. 'Strategies for RNA Resonance Assignment by <sup>13</sup>C/<sup>15</sup>N- and <sup>1</sup>H-Detected Solid-State NMR Spectroscopy'. *Frontiers in Molecular Biosciences* 8:900. doi: 10.3389/fmolb.2021.743181.
- Alvarado, Luigi J., Regan M. Leblanc, Andrew P. Longhini, Sarah C. Keane, Niyati Jain, Zehra F. Yildiz, Blanton S. Tolbert, Victoria M. D'Souza, Michael F. Summers, Christoph Kreutz, and T. Kwaku Dayie. 2014. 'Regio-Selective Chemical-Enzymatic Synthesis of Pyrimidine Nucleotides Facilitates RNA Structure and Dynamics Studies'. *ChemBioChem* 15(11):1573–77. doi: 10.1002/CBIC.201402130.
- Amador-Cañizares, Yalena, Mamata Panigrahi, Adam Huys, Rasika D. Kunden, Halim M. Adams, Michael J. Schinold, and Joyce A. Wilson. 2018. 'MiR-122, Small RNA Annealing and Sequence Mutations Alter the Predicted Structure of the Hepatitis C Virus 5' UTR RNA to Stabilize and Promote Viral RNA Accumulation'. *Nucleic Acids Research*. doi: 10.1093/nar/gky662.
- Bai, Yun, Kaihong Zhou, and Jennifer A. Doudna. 2013. 'Hepatitis C Virus 3'UTR Regulates Viral Translation through Direct Interactions with the Host Translation Machinery'. *Nucleic Acids Research* 41(16):7861–74. doi: 10.1093/nar/gkt543.
- Bajan, Sarah, and Gyorgy Hutvagner. 2020. 'RNA-Based Therapeutics: From Antisense Oligonucleotides to MiRNAs'. *Cells* 9(1).
- BALDUS, MARC, ANETA T. PETKOVA, JUDITH HERZFELD, and ROBERT G. GRIFFIN. 1998. 'Cross Polarization in the Tilted Frame: Assignment and Spectral Simplification in Heteronuclear Spin Systems'. *Molecular Physics* 95(6):1197–1207. doi: 10.1080/00268979809483251.
- Bartel, David P. 2004. 'MicroRNAs: Genomics, Biogenesis, Mechanism, and Function'. *Cell*.
- Belema, Makonen, and Nicholas A. Meanwell. 2014. 'Discovery of Daclatasvir, a Pan-Genotypic Hepatitis C Virus NS5A Replication Complex Inhibitor with Potent Clinical Effect.' *Journal of Medicinal Chemistry* 57(12):5057–71. doi: 10.1021/jm500335h.

- Berger, Carola, Inés Romero-Brey, Danijela Radujkovic, Raphael Terreux, Margarita Zayas, David Paul, Christian Harak, Simone Hoppe, Min Gao, Francois Penin, Volker Lohmann, and Ralf Bartenschlager. 2014. 'Daclatasvir-like Inhibitors of NS5A Block Early Biogenesis of Hepatitis C Virus-Induced Membranous Replication Factories, Independent of RNA Replication'. *Gastroenterology* 147(5):1094-1105.e25. doi: 10.1053/J.GASTRO.2014.07.019.
- Bertini, Ivano, Frank Engelke, Claudio Luchinat, Giacomo Parigi, Enrico Ravera, Camilla Rosa, and Paola Turano. 2011. 'NMR Properties of Sedimented Solutes'. *Physical Chemistry Chemical Physics* 14(2):439–47. doi: 10.1039/C1CP22978H.
- Boniecki, Michal J., Grzegorz Lach, Wayne K. Dawson, Konrad Tomala, Pawel Lukasz, Tomasz Soltysinski, Kristian M. Rother, and Janusz M. Bujnicki. 2016. 'SimRNA: A Coarse-Grained Method for RNA Folding Simulations and 3D Structure Prediction'. *Nucleic Acids Research* 44(7):e63–e63. doi: 10.1093/nar/gkv1479.
- Bou-Nader, Charles, and Jinwei Zhang. 2020. 'Structural Insights into RNA Dimerization: Motifs, Interfaces and Functions'. *Molecules* 2020, Vol. 25, Page 2881 25(12):2881. doi: 10.3390/MOLECULES25122881.
- Brown, Edwin A., Hangchun Zhang, Li hua Ping, and Stanley M. Lemon. 1992. 'Secondary Structure of the 5' Nontranslated Regions of Hepatitis C Virus and Pestivirus Genomic RNAs'. *Nucleic Acids Research* 20(19):5041–45. doi: 10.1093/nar/20.19.5041.
- Bucciarelli, Saskia, Søren Roi Midtgaard, Martin Nors Pedersen, Søren Skou, Lise Arleth, and Bente Vestergaard. 2018. 'Size-Exclusion Chromatography Small-Angle X-Ray Scattering of Water Soluble Proteins on a Laboratory Instrument'. *Journal of Applied Crystallography* 51(6):1623–32. doi: 10.1107/S1600576718014462.
- Burke, Jordan E., Dipali G. Sashital, Xiaobing Zuo, Yun Xing Wang, and Samuel E. Butcher. 2012. 'Structure of the Yeast U2/U6 SnRNA Complex'. *RNA* 18(4):673–83. doi: 10.1261/RNA.031138.111.
- Büttner, Lea, Fatemeh Javadi-Zarnaghi, and Claudia Höbartner. 2014. 'Site-Specific Labeling of RNA at Internal Ribose Hydroxyl Groups: Terbium-Assisted Deoxyribozymes at Work'. *Journal of the American Chemical Society* 136(22):8131–37. doi: 10.1021/ja503864v.
- Case, D. A., K. Belfon, I. Y. Ben-Shalom, S. R. Brozell, D. S. Cerutti, T. E. Cheatham, V. W. D. Cruzeiro III, T. A. Darden, R. E. Duke, and G. Giambasu. 2020. 'AMBER2020, University of California, San Fransisco'. *J. Amer. Chem. Soc* 142:3823–35.

- Castillo, I. 2005. 'Hepatitis C Virus Replicates in Peripheral Blood Mononuclear Cells of Patients with Occult Hepatitis C Virus Infection'. *Gut* 54(5):682–85. doi: 10.1136/gut.2004.057281.
- Cavanagh, John, Wayne J. Fairbrother, III Arthur G. Palmer, Mark Rance, and Nicholas J. Skelton. 2010. *Protein NMR Spectroscopy: Principles and Practice*. 2nd ed. Elsevier Science.
- Chahal, Jasmin, Luca F. R. F. R. Gebert, Hin Hark Gan, Edna Camacho, Kristin C. Gunsalus, Ian J. MacRae, and Selena M. Sagan. 2019. 'MIR-122 and Ago Interactions with the HCV Genome Alter the Structure of the Viral 5' Terminus'. *Nucleic Acids Research* 47(10):5307–24. doi: 10.1093/nar/gkz194.
- Chatel-Chaix, Laurent, and Ralf Bartenschlager. 2014. 'Dengue Virus- and Hepatitis C Virus-Induced Replication and Assembly Compartments: The Enemy inside--Caught in the Web.' *Journal of Virology* 88(11):5907–11. doi: 10.1128/JVI.03404-13.
- Cheatham, Thomas E., and David A. Case. 2013. 'Twenty-Five Years of Nucleic Acid Simulations'. *Biopolymers* 99(12).
- Chojnowski, Grzegorz, Rafał Zaborowski, Marcin Magnus, and Janusz M. Bujnicki. 2021. 'RNA Fragment Assembly with Experimental Restraints'. *BioRxiv* 2021.02.08.430198. doi: 10.1101/2021.02.08.430198.
- Collier, Adam J., José Gallego, Roscoe Klinck, Paul T. Cole, Stephen J. Harris, Geoffrey P. Harrison, Fared Aboul-Ela, Gabriele Varani, and Stephen Walker. 2002. 'A Conserved RNA Structure within the HCV IRES EIF3-Binding Site'. *Nature Structural Biology* 9(5):375–80. doi: 10.1038/NSB785.
- Conrad, Thomas, Izabela Plumbom, Maria Alcobendas, Ramon Vidal, and Sascha Sauer. 2020. 'Maximizing Transcription of Nucleic Acids with Efficient T7 Promoters'. *Communications Biology* 3(1):1–8. doi: 10.1038/s42003-020-01167-x.
- Corkeys, Barbara E., Jerzy Duszynski, Terrell L. Richf, Benno Matschinsky, and John R. Williamson. 1986. 'Regulation of Free and Bound Magnesium in Rat Hepatocytes and Isolated Mitochondria.' *Journal of Biological Chemistry* 261(6):2567–74. doi: 10.1016/S0021-9258(17)35825-8.
- Dieckmann, Thorsten, and Juli Feigon. 1994. 'Heteronuclear Techniques in NMR Studies of RNA and DNA'. *Current Opinion in Structural Biology* 4(5):745–49. doi: 10.1016/S0959-440X(94)90174-0.

- Dingley, Andrew J., and Stephan Grzesiek. 1998. 'Direct Observation of Hydrogen Bonds in Nucleic Acid Base Pairs by Internucleotide  $^2 J_{NN}$  Couplings'. *Journal of the American Chemical Society* 120(33):8293–97. doi: 10.1021/ja981513x.
- Donald B. Smith, Jens Bukh, Carla Kuiken, A. Scott Muerhoff, Charles M. Rice, Jack T. Stapleton and Peter Simmonds. n.d. 'International Committee on Taxonomy of Viruses (ICTV)'. Retrieved 9 April 2020 ([https://talk.ictvonline.org/ictv\\_wikis/flaviviridae/w/sg\\_flavi/56/hcv-classification](https://talk.ictvonline.org/ictv_wikis/flaviviridae/w/sg_flavi/56/hcv-classification)).
- Draper, David E. 2004. 'A Guide to Ions and RNA Structure'. *RNA (New York, N.Y.)* 10(3):335–43. doi: 10.1261/RNA.5205404.
- Dunne, O., M. Weidenhaupt, P. Callow, A. Martel, M. Moulin, S. J. Perkins, M. Haertlein, and V. T. Forsyth. 2017. 'Matchout Deuterium Labelling of Proteins for Small-Angle Neutron Scattering Studies Using Prokaryotic and Eukaryotic Expression Systems and High Cell-Density Cultures'. *European Biophysics Journal* 46(5):425–32. doi: 10.1007/s00249-016-1186-2.
- Duss, Olivier, Christophe Maris, Christine von Schroetter, and Frédéric H. T. Allain. 2010. 'A Fast, Efficient and Sequence-Independent Method for Flexible Multiple Segmental Isotope Labeling of RNA Using Ribozyme and RNase H Cleavage'. *Nucleic Acids Research* 38(20). doi: 10.1093/nar/gkq756.
- Epstein, Lloyd M., and Joseph G. Gall. 1987. 'Self-Cleaving Transcripts of Satellite DNA from the Newt'. *Cell* 48(3):535–43. doi: 10.1016/0092-8674(87)90204-2.
- Fang, Zhuo, and Nikolaus Rajewsky. 2011. 'The Impact of MiRNA Target Sites in Coding Sequences and in 3'UTRs'. *PLoS ONE*. doi: 10.1371/journal.pone.0018067.
- Farès, Christophe, Irene Amata, and Teresa Carlomagno. 2007. '13C-Detection in RNA Bases: Revealing Structure-Chemical Shift Relationships.' *Journal of the American Chemical Society* 129(51):15814–23. doi: 10.1021/ja0727417.
- Feigin, L. A., and D. I. Svergun. 1987. 'Structure Analysis by Small-Angle X-Ray and Neutron Scattering'. *Structure Analysis by Small-Angle X-Ray and Neutron Scattering*. doi: 10.1007/978-1-4757-6624-0.
- Feyrer, Hannes, Raluca Munteanu, Lorenzo Baronti, and Katja Petzold. 2020. 'One-Pot Production of RNA in High Yield and Purity Through Cleaving Tandem Transcripts'. *Molecules (Basel, Switzerland)* 25(5). doi: 10.3390/MOLECULES25051142.
- Franke, Daniel, and Dmitri I. Svergun. 2009. 'DAMMIF, a Program for Rapid Ab-Initio Shape Determination in Small-Angle Scattering'. *Journal of Applied Crystallography* 42(Pt 2):342–46. doi: 10.1107/S0021889809000338.



- Friebe, P., V. Lohmann, N. Krieger, and R. Bartenschlager. 2001. 'Sequences in the 5' Nontranslated Region of Hepatitis C Virus Required for RNA Replication'. *Journal of Virology* 75(24):12047–57. doi: 10.1128/jvi.75.24.12047-12057.2001.
- Friebe, Peter, and Ralf Bartenschlager. 2002. 'Genetic Analysis of Sequences in the 3' Nontranslated Region of Hepatitis C Virus That Are Important for RNA Replication'. *Journal of Virology* 76(11):5326–38. doi: 10.1128/JVI.76.11.5326-5338.2002.
- Fukushi, Shuetsu, Kazuhiko Katayama, Chie Kurihara, Narihiro Ishiyama, Fuminori B. Hoshino, Takao Ando, and Akira Oya. 1994. 'Complete 5' Noncoding Region Is Necessary for the Efficient Internal Initiation of Hepatitis C Virus RNA'. *Biochemical and Biophysical Research Communications*. doi: 10.1006/bbrc.1994.1246.
- Fürtig, Boris, Christian Richter, Jens Wöhnert, and Harald Schwalbe. 2003. 'NMR Spectroscopy of RNA'. *ChemBioChem* 4(10):936–62. doi: 10.1002/cbic.200300700.
- Gabel, Frank. 2015. *Small-Angle Neutron Scattering for Structural Biology of Protein-RNA Complexes*. Vol. 558. 1st ed. Elsevier Inc.
- Geddawy, Ayman, Yasmine F. Ibrahim, Nabil M. Elbahie, and Mohammad A. Ibrahim. 2017. 'Direct Acting Anti-Hepatitis C Virus Drugs: Clinical Pharmacology and Future Direction'. *Journal of Translational Internal Medicine* 5(1):8–17. doi: 10.1515/JTIM-2017-0007.
- Gentile, Ivan, Antonio Riccardo Buonomo, and Guglielmo Borgia. 2014. 'Ombitasvir: A Potent Pan-Genotypic Inhibitor of NS5A for the Treatment of Hepatitis C Virus Infection'. *Expert Review of Anti-Infective Therapy* 12(9):1033–43. doi: 10.1586/14787210.2014.940898.
- Gholamalipour, Yasaman, Aruni Karunanayake Mudiyansele, and Craig T. Martin. 2018. '3' End Additions by T7 RNA Polymerase Are RNA Self-Templated, Distributive and Diverse in Character—RNA-Seq Analyses'. *Nucleic Acids Research* 46(18):9253–63. doi: 10.1093/NAR/GKY796.
- Gitto, Stefano, Nesrine Gamal, and Pietro Andreone. 2017. 'NS5A Inhibitors for the Treatment of Hepatitis C Infection'. *Journal of Viral Hepatitis* 24(3):180–86. doi: 10.1111/JVH.12657.
- Gorrec, Fabrice. 2009. 'The MORPHEUS Protein Crystallization Screen'. *Journal of Applied Crystallography* 42(Pt 6):1035. doi: 10.1107/S0021889809042022.
- Götte, Matthias, and Jordan J. Feld. 2016. 'Direct-Acting Antiviral Agents for Hepatitis C: Structural and Mechanistic Insights'. *Nature Reviews Gastroenterology and Hepatology* 13(6):338–51.

- Götz, Andreas W., Mark J. Williamson, Dong Xu, Duncan Poole, Scott Le Grand, and Ross C. Walker. 2012. 'Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born'. *Journal of Chemical Theory and Computation* 8(5). doi: 10.1021/ct200909j.
- Le Grand, Scott, Andreas W. Götz, and Ross C. Walker. 2013. 'SPFP: Speed without Compromise - A Mixed Precision Model for GPU Accelerated Molecular Dynamics Simulations'. *Computer Physics Communications* 184(2). doi: 10.1016/j.cpc.2012.09.022.
- Guinier, A., and G. Fournet. 1955. *Small Angle Scattering of X-Rays*.
- Guinier, André. 1939. 'La Diffraction Des Rayons X Aux Très Petits Angles : Application à l'étude de Phénomènes Ultramicroscopiques'. *Annales de Physique* 11(12):161–237. doi: 10.1051/ANPHYS/193911120161.
- Gullion, Terry, and Jacob Schaefer. 1989. 'Rotational-Echo Double-Resonance NMR'. *Journal of Magnetic Resonance (1969)* 81(1):196–200. doi: 10.1016/0022-2364(89)90280-1.
- Hamelberg, Donald, John Mongan, and J. Andrew McCammon. 2004. 'Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules'. *Journal of Chemical Physics* 120(24). doi: 10.1063/1.1755656.
- Hammouda, Boualem. 2008. 'Insight into the SANS Technique'. *National Institute of Standards and Technology* 1–717.
- Hartmann, S. R., and E. L. Hahn. 1962. 'Nuclear Double Resonance in the Rotating Frame'. *Physical Review* 128(5):2042. doi: 10.1103/PhysRev.128.2042.
- He, Yupeng, Kirk A. Staschke, and Seng-Lai Tan. 2006. 'HCV NS5A: A Multifunctional Regulator of Cellular Pathways and Virus Replication'. *Hepatitis C Viruses: Genomes and Molecular Biology*.
- Helm, M., H. Brule, R. Giege, and C. Florentz. 1999. 'More Mistakes by T7 RNA Polymerase at the 5' Ends of in Vitro-Transcribed RNAs'. *RNA* 5(5):618–21. doi: 10.1017/S1355838299982328.
- Helmling, Christina, Sara Keyhani, Florian Sochor, Boris Fürtig, Martin Hengesbach, and Harald Schwalbe. 2015. 'Rapid NMR Screening of RNA Secondary Structure and Binding'. *Journal of Biomolecular NMR* 63(1):67–76. doi: 10.1007/s10858-015-9967-y.
- Henke, Jura Inga, Dagmar Goergen, Junfeng Zheng, Yutong Song, Christian G. Schüttler, Carmen Fehr, Christiane Jünemann, and Michael Niepmann. 2008. 'MicroRNA-122

- Stimulates Translation of Hepatitis C Virus RNA'. *The EMBO Journal* 27(24):3300–3310. doi: 10.1038/emboj.2008.244.
- Hennig, Mirko, and James R. Williamson. 2000. 'Detection of N-H···N Hydrogen Bonding in RNA via Scalar Couplings in the Absence of Observable Imino Proton Resonances'. *Nucleic Acids Research* 28(7):1585–93. doi: 10.1093/NAR/28.7.1585.
- Hézode, Christophe, Hélène Fontaine, Céline Dorival, Dominique Larrey, Fabien Zoulim, Valérie Canva, Victor de Ledinghen, Thierry Poynard, Didier Samuel, Marc Bourlière, Jean Pierre Zarski, Jean Jacques Raabe, Laurent Alric, Patrick Marcellin, Ghassan Riachi, Pierre Henri Bernard, Véronique Loustaud-Ratti, Sophie Métivier, Albert Tran, Lawrence Serfaty, Armand Abergel, Xavier Causse, Vincent di Martino, Dominique Guyader, Damien Lucidarme, Véronique Grando-Lemaire, Patrick Hillon, Cyrille Feray, Thong Dao, Patrice Cacoub, Isabelle Rosa, Pierre Attali, Ventzislava Petrov-Sanchez, Yoann Barthe, Jean Michel Pawlotsky, Stanislas Pol, Fabrice Carrat, and Jean Pierre Bronowicki. 2013. 'Triple Therapy in Treatment-Experienced Patients with HCV-Cirrhosis in a Multicentre Cohort of the French Early Access Programme (ANRS CO20-CUPIC) - NCT01514890'. *Journal of Hepatology* 59(3):434–41. doi: 10.1016/j.jhep.2013.04.035.
- Honda, Masao, Michael R. Beard, Li-Hua Ping, and Stanley M. Lemon. 1999. 'A Phylogenetically Conserved Stem-Loop Structure at the 5' Border of the Internal Ribosome Entry Site of Hepatitis C Virus Is Required for Cap-Independent Viral Translation'. *Journal of Virology* 73(2):1165–74. doi: 10.1128/JVI.73.2.1165-1174.1999.
- Honda, Masao, Edwin A. Brown, and Stanley M. Lemon. 1996. 'Stability of a Stem-Loop Involving the Initiator AUG Controls the Efficiency of Internal Initiation of Translation on Hepatitis C Virus RNA'. *RNA*.
- Hu, Kan-Nian, Wai-Ming Yau, and Robert Tycko. 2010. 'Detection of a Transient Intermediate in a Rapid Protein Folding Process by Solid-State Nuclear Magnetic Resonance.' *Journal of the American Chemical Society* 132(1):24–25. doi: 10.1021/ja908471n.
- Hu, Weidong, Lazaros T. Kakalis, Licong Jiang, Feng Jiang, Xiaomei Ye, and Ananya Majumdar. 1998. '3D HCCH-COSY-TOCSY Experiment for the Assignment of Ribose and Amino Acid Side Chains in <sup>13</sup>C Labeled RNA and Protein'. *Journal of Biomolecular NMR* 12(4):559–64. doi: 10.1023/A:1008365301124.

- Huang, Tai Huang, William W. Bachovchin, Robert G. Griffin, and Christopher M. Dobson. 1984. 'High-Resolution Nitrogen-15 Nuclear Magnetic Resonance Studies of  $\alpha$ -Lytic Protease in Solid State. Direct Comparison of Enzyme Structure in Solution and in the Solid State'. *Biochemistry* 23(25):5933–37. doi: 10.1021/BI00320A007/ASSET/BI00320A007.FP.PNG\_V03.
- Huang, Wei, Michael F. Bardaro, Gabriele Varani, and Gary P. Drobny. 2012. 'Preparation of RNA Samples with Narrow Line Widths for Solid State NMR Investigations'. *Journal of Magnetic Resonance* 223:51–54. doi: 10.1016/j.jmr.2012.07.018.
- Imai, Shunsuke, Parimal Kumar, Christopher U. T. Hellen, Victoria M. D'Souza, and Gerhard Wagner. 2016. 'An Accurately Preorganized IRES RNA Structure Enables EIF4G Capture for Initiation of Viral Translation'. *Nature Structural & Molecular Biology* 23:9 23(9):859–64. doi: 10.1038/nsmb.3280.
- Israelow, Benjamin, Gavriel Mullokandov, Judith Agudo, Marion Sourisseau, Ali Bashir, Andres Y. Maldonado, Arvin C. Dar, Brian D. Brown, and Matthew J. Evans. 2014. 'Hepatitis C Virus Genetics Affects MiR-122 Requirements and Response to MiR-122 Inhibitors'. *Nature Communications*. doi: 10.1038/ncomms6408.
- Jacrot, B. 1976. 'The Study of Biological Structures by Neutron Scattering from Solution'. *Reports on Progress in Physics* 39(10):911. doi: 10.1088/0034-4885/39/10/001.
- Jacrot, Bernard, and Giuseppe Zaccai. 1981. 'Determination of Molecular Weight by Neutron Scattering'. *Biopolymers* 20(11):2413–26. doi: 10.1002/BIP.1981.360201110.
- Jambor, Helena, Christine Brunel, and Anne Ephrussi. 2011. 'Dimerization of Oskar 3' UTRs Promotes Hitchhiking for RNA Localization in the Drosophila Oocyte'. *RNA* 17(12):2049. doi: 10.1261/RNA.2686411.
- Jeffries, Cy M., Melissa A. Graewert, Clément E. Blanchet, David B. Langley, Andrew E. Whitten, and Dmitri I. Svergun. 2016. 'Preparing Monodisperse Macromolecular Samples for Successful Biological Small-Angle X-Ray and Neutron-Scattering Experiments'. *Nature Protocols* 11(11):2122–53. doi: 10.1038/nprot.2016.113.
- Ji, Hong, Christopher S. Fraser, Yonghao Yu, Julie Leary, and Jennifer A. Doudna. 2004. 'Coordinated Assembly of Human Translation Initiation Complexes by the Hepatitis C Virus Internal Ribosome Entry Site RNA'. *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0407402101.
- Jin, Zhinan, Vincent Leveque, Han Ma, Kenneth A. Johnson, and Klaus Klumpp. 2012. 'Assembly, Purification, and Pre-Steady-State Kinetic Analysis of Active RNA-

- Dependent RNA Polymerase Elongation Complex'. *The Journal of Biological Chemistry* 287(13):10674. doi: 10.1074/JBC.M111.325530.
- Jirasko, Vlastimil, Roland Montserret, Nicole Appel, Anne Janvier, Leah Eustachi, Christiane Brohm, Eike Steinmann, Thomas Pietschmann, Francois Penin, and Ralf Bartenschlager. 2008. 'Structural and Functional Characterization of Nonstructural Protein 2 for Its Role in Hepatitis C Virus Assembly'. *Journal of Biological Chemistry* 283(42):28546–62. doi: 10.1074/JBC.M803981200.
- Jolla, L. A., Prnewswire Regulus, and Nasdaq Rgls. 2017. *Regulus Announces Pipeline Updates and Advancements*. Vol. 4326.
- Jones, Daniel M., Arvind H. Patel, Paul Targett-Adams, and John McLauchlan. 2009. 'The Hepatitis C Virus NS4B Protein Can Trans-Complement Viral RNA Replication and Modulates Production of Infectious Virus'. *Journal of Virology* 83(5):2163–77. doi: 10.1128/JVI.01885-08.
- Jopling, Catherine L. 2005. 'Modulation of Hepatitis C Virus RNA Abundance by a Liver-Specific MicroRNA'. *Science* 309(5740):1577–81. doi: 10.1126/science.1113329.
- Jopling, Catherine L. 2008. 'Regulation of Hepatitis C Virus by MicroRNA-122'. *Biochemical Society Transactions*. doi: 10.1042/BST0361220.
- Jopling, Catherine L., Sylvia Schütz, and Peter Sarnow. 2008. 'Position-Dependent Function for a Tandem MicroRNA MiR-122 Binding Site Located in the Hepatitis C Virus RNA Genome'. *Cell Host & Microbe* 4(1):77–85. doi: 10.1016/j.chom.2008.05.013.Position-dependent.
- Jordan, Ashley, Mark Jacques, Catherine Merrick, Juliette Devos, V. Trevor Forsyth, Lionel Porcar, and Anne Martel. 2016. 'SEC-SANS: Size Exclusion Chromatography Combined in Situ with Small-Angle Neutron Scattering'. *Journal of Applied Crystallography* 49(Pt 6):2015. doi: 10.1107/S1600576716016514.
- Kalidindi, Yamini, Jeah Jung, Roger Feldman, and Thomas Riley. 2020. 'Association of Direct-Acting Antiviral Treatment With Mortality Among Medicare Beneficiaries With Hepatitis C'. *JAMA Network Open* 3(7):e2011055. doi: 10.1001/jamanetworkopen.2020.11055.
- Kao, C., M. Zheng, and S. Rüdisser. 1999. 'A Simple and Efficient Method to Reduce Nontemplated Nucleotide Addition at the 3' Terminus of RNAs Transcribed by T7 RNA Polymerase'. *Rna* 5(9):1268–72. doi: 10.1017/S1355838299991033.
- Kaseman, Derrick. n.d. 'Chemical Shift (Shielding)'. Retrieved 26 May 2023 (<https://chem.libretexts.org/@go/page/1824>).

- Keating, Gillian M. 2014. ‘Sofosbuvir: A Review of Its Use in Patients with Chronic Hepatitis C’. *Drugs* 74(10):1127–46. doi: 10.1007/S40265-014-0247-Z.
- Kieft, Jeffrey S., Kaihong Zhou, Angie Grech, Ronald Jubin, and Jennifer A. Doudna. 2002. ‘Crystal Structure of an RNA Tertiary Domain Essential to HCV IRES-Mediated Translation Initiation’. *Nature Structural Biology* 2002 9:5 9(5):370–74. doi: 10.1038/nsb781.
- Kieft, Jeffrey S., Kaihong Zhou, Ronald Jubin, and Jennifer A. Doudna. 2001. ‘Mechanism of Ribosome Recruitment by Hepatitis C IRES RNA’. *RNA*. doi: 10.1017/S1355838201001790.
- Kieft, Jeffrey S., Kaihong Zhou, Ronald Jubin, Michael G. Murray, Johnson Y. N. Lau, and Jennifer A. Doudna. 1999. ‘The Hepatitis C Virus Internal Ribosome Entry Site Adopts an Ion-Dependent Tertiary Fold’. *Journal of Molecular Biology* 292(3):513–29. doi: 10.1006/JMBI.1999.3095.
- Kolykhalov, A. A., S. M. Feinstone, and C. M. Rice. 1996. ‘Identification of a Highly Conserved Sequence Element at the 3’ Terminus of Hepatitis C Virus Genome RNA.’ *Journal of Virology*. doi: 10.1128/jvi.70.6.3363-3371.1996.
- Kotar, Anita, Hannah N. Foley, Kirk M. Baughman, and Sarah C. Keane. 2020. ‘Advanced Approaches for Elucidating Structures of Large RNAs Using NMR Spectroscopy and Complementary Methods’. *Methods* 183(January 2020):93–107. doi: 10.1016/j.ymeth.2020.01.009.
- Kremser, J., E. Strebiter, R. Plangger, M. A. Juen, F. Nußbaumer, H. Glasner, K. Breuker, and C. Kreutz. 2017. ‘Chemical Synthesis and NMR Spectroscopy of Long Stable Isotope Labelled RNA’. *Chemical Communications* 53(96):12938–41. doi: 10.1039/C7CC06747J.
- Kunden, RD., Sarah Ghezelbash, Juveriya Q. Khan, and Joyce A. Wilson. 2020. ‘Location Specific Annealing of MiR-122 and Other Small RNAs Defines an Hepatitis C Virus 5’ UTR Regulatory Element with Distinct Impacts on Virus Translation and Genome Stability’. *Nucleic Acids Research* 48(16):9235–49. doi: 10.1093/nar/gkaa664.
- Kunden, RD., Juveriya Q. Khan, Sarah Ghezelbash, and Joyce A. Wilson. 2020. ‘The Role of the Liver-Specific MicroRNA, MiRNA-122 in the HCV Replication Cycle’. *International Journal of Molecular Sciences* 21(16):5677. doi: 10.3390/ijms21165677.
- Lacabanne, Denis, Marie Laure Fogeron, Thomas Wiegand, Riccardo Cadalbert, Beat H. Meier, and Anja Böckmann. 2019. ‘Protein Sample Preparation for Solid-State NMR

- Investigations'. *Progress in Nuclear Magnetic Resonance Spectroscopy* 110:20–33. doi: 10.1016/J.PNMRS.2019.01.001.
- Lagos-Quintana, Mariana, Reinhard Rauhut, Abdullah Yalcin, Jutta Meyer, Winfried Lendeckel, and Thomas Tuschl. 2002. 'Identification of Tissue-Specific MicroRNAs from Mouse'. *Current Biology* 12(9):735–39. doi: 10.1016/S0960-9822(02)00809-6.
- Lancaster, Alissa M., Eric Jan, and Peter Sarnow. 2006. 'Initiation Factor-Independent Translation Mediated by the Hepatitis C Virus Internal Ribosome Entry Site'. *RNA* 12(5):894. doi: 10.1261/RNA.2342306.
- Lee, Inhan, Subramanian S. Ajay, In Yook Jong, Sil Kim Hyun, Hyung Hong Su, Hee Kim Nam, Saravana M. Dhanasekaran, Arul M. Chinnaiyan, and Brian D. Athey. 2009. 'New Class of MicroRNA Targets Containing Simultaneous 5'-UTR and 3'-UTR Interaction Sites'. *Genome Research*. doi: 10.1101/gr.089367.108.
- Levitt, Malcolm H. 2009. *Spin Dynamics: Basics of Nuclear Magnetic Resonance, 2nd Edition*. Vol. 34A.
- Lewis, Benjamin P., I. Hung Shih, Matthew W. Jones-Rhoades, David P. Bartel, and Christopher B. Burge. 2003. 'Prediction of Mammalian MicroRNA Targets'. *Cell*. doi: 10.1016/S0092-8674(03)01018-3.
- Li, You, Takahiro Masaki, Daisuke Yamane, David R. McGivern, and Stanley M. Lemon. 2013. 'Competing and Noncompeting Activities of MiR-122 and the 5' Exonuclease Xrn1 in Regulation of Hepatitis C Virus Replication'. *Proceedings of the National Academy of Sciences* 110(5):1881–86. doi: 10.1073/pnas.1213515110.
- Li, You, Daisuke Yamane, Takahiro Masaki, and Stanley M. Lemon. 2015. 'The Yin and Yang of Hepatitis C: Synthesis and Decay of Hepatitis C Virus RNA'. *Nature Reviews Microbiology* 13(9):544–58. doi: 10.1038/nrmicro3506.
- Lindow, Morten, and Sakari Kauppinen. 2012. 'Discovering the First MicroRNA-Targeted Drug'. *Journal of Cell Biology* 199(3):407–12. doi: 10.1083/jcb.201208082.
- Link, John O., James G. Taylor, Lianhong Xu, Michael Mitchell, Hongyan Guo, Hongtao Liu, Darryl Kato, Thorsten Kirschberg, Jianyu Sun, Neil Squires, Jay Parrish, Terry Keller, Zheng-Yu Yang, Chris Yang, Mike Matles, Yujin Wang, Kelly Wang, Guofeng Cheng, Yang Tian, Erik Mogalian, Elsa Mondou, Melanie Cornpropst, Jason Perry, and Manoj C. Desai. 2014. 'Discovery of Ledipasvir (GS-5885): A Potent, Once-Daily Oral NS5A Inhibitor for the Treatment of Hepatitis C Virus Infection'. *Journal of Medicinal Chemistry* 57(5):2033–46. doi: 10.1021/jm401499g.

- Liu, Fanwei, Tetsuro Shimakami, Kazuhisa Murai, Takayoshi Shirasaki, Masaya Funaki, Masao Honda, Seishi Murakami, Minkyung Yi, Hong Tang, and Shuichi Kaneko. 2016. 'Efficient Suppression of Hepatitis C Virus Replication by Combination Treatment with MIR-122 Antagonism and Direct-Acting Antivirals in Cell Culture Systems'. *Scientific Reports* 6(1):1–11. doi: 10.1038/srep30939.
- Locker, Nicolas, Laura E. Easton, and Peter J. Lukavsky. 2007. 'HCV and CSFV IRES Domain II Mediate EIF2 Release during 80S Ribosome Assembly'. *EMBO Journal*. doi: 10.1038/sj.emboj.7601549.
- Lohmann, V. 1999. 'Replication of Subgenomic Hepatitis C Virus RNAs in a Hepatoma Cell Line'. *Science* 285(5424):110–13. doi: 10.1126/science.285.5424.110.
- Longhini, Andrew P., Regan M. Leblanc, Owen Becette, Carolina Salguero, Christoph H. Wunderlich, Bruce A. Johnson, Victoria M. D'Souza, Christoph Kreutz, and T. Kwaku Dayie. 2016. 'Chemo-Enzymatic Synthesis of Site-Specific Isotopically Labeled Nucleotides for Use in NMR Resonance Assignment, Dynamics and Structural Characterizations'. *Nucleic Acids Research* 44(6):e52–e52. doi: 10.1093/NAR/GKV1333.
- Lorenz, Ronny, Stephan H. Bernhart, Christian Höner zu Siederdisen, Hakim Tafer, Christoph Flamm, Peter F. Stadler, and Ivo L. Hofacker. 2011. 'ViennaRNA Package 2.0'. *Algorithms for Molecular Biology* 6(1):1–14. doi: 10.1186/1748-7188-6-26.
- Lukavsky, Peter J. 2009. 'Structure and Function of HCV IRES Domains'. *Virus Research* 139(2):166–71. doi: 10.1016/j.virusres.2008.06.004.
- Lukavsky, Peter J., Insil Kim, Geoff A. Otto, and Joseph D. Puglisi. 2003. 'Structure of HCV IRES Domain II Determined by NMR'. *Nature Structural Biology* 10(12):1033–38. doi: 10.1038/nsb1004.
- Lukavsky, Peter J., Geoff A. Otto, Alissa M. Lancaster, Peter Sarnow, and Joseph D. Puglisi. 2000. 'Structures of Two RNA Domains Essential for Hepatitis C Virus Internal Ribosome Entry Site Function'. *Nature Structural Biology* 7(12):1105–10. doi: 10.1038/81951.
- Machlin, Erica S., Peter Sarnow, and Selena M. Sagan. 2011. 'Masking the 5' Terminal Nucleotides of the Hepatitis C Virus Genome by an Unconventional MicroRNA-Target RNA Complex'. *Proceedings of the National Academy of Sciences* 108(8):3193–98. doi: 10.1073/pnas.1012464108.
- Maier, James A., Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E. Hauser, and Carlos Simmerling. 2015. 'Ffl4SB: Improving the Accuracy of Protein



- Side Chain and Backbone Parameters from Ff99SB'. *Journal of Chemical Theory and Computation* 11(8). doi: 10.1021/acs.jctc.5b00255.
- Manalastas-Cantos, Karen, Petr v. Konarev, Nelly R. Hajizadeh, Alexey G. Kikhney, Maxim v. Petoukhov, Dmitry S. Molodenskiy, Alejandro Panjkovich, Haydyn D. T. Mertens, Andrey Gruzinov, Clemente Borges, Cy M. Jeffries, Dmitri I. Svergun, and Daniel Franke. 2021. 'ATSAS 3.0: Expanded Functionality and New Tools for Small-Angle Scattering Data Analysis'. *Urn:Issn:1600-5767* 54(1):343–55. doi: 10.1107/S1600576720013412.
- Marchanka, Alexander, and Teresa Carlomagno. 2014. 'Solid-State NMR and RNA Structure: A New Partnership?' *EMagRes* 3(2):119–28.
- Marchanka, Alexander, Bernd Simon, Gerhard Althoff-Ospelt, and Teresa Carlomagno. 2015. 'RNA Structure Determination by Solid-State NMR Spectroscopy'. *Nature Communications* 6. doi: 10.1038/NCOMMS8024.
- Marchanka, Alexander, Jan Stanek, Guido Pintacuda, and Teresa Carlomagno. 2018. 'Rapid Access to RNA Resonances by Proton-Detected Solid-State NMR at >100 KHz MAS'. *Chemical Communications* 54(65):8972–75. doi: 10.1039/c8cc04437f.
- Martin, Rachel W., and Kurt W. Zilm. 2003. 'Preparation of Protein Nanocrystals and Their Characterization by Solid State NMR'. *Journal of Magnetic Resonance* 165(1):162–74. doi: 10.1016/S1090-7807(03)00253-2.
- Martinello, Marianne, Sahar Bajis, and Gregory J. Dore. 2020. 'Progress Toward Hepatitis C Virus Elimination: Therapy and Implementation'. *Gastroenterology Clinics of North America* 49(2):253–77.
- Marušič, Maja, Judith Schlagnitweit, and Katja Petzold. 2019. 'RNA Dynamics by NMR Spectroscopy'. *Chembiochem* 20(21):2685. doi: 10.1002/CBIC.201900072.
- Masaki, Takahiro, Kyle C. Arend, You Li, Daisuke Yamane, David R. McGivern, Takanobu Kato, Takaji Wakita, Nathaniel J. Moorman, and Stanley M. Lemon. 2015. 'MiR-122 Stimulates Hepatitis C Virus RNA Synthesis by Altering the Balance of Viral RNAs Engaged in Replication versus Translation'. *Cell Host and Microbe* 17(2):217–28. doi: 10.1016/j.chom.2014.12.014.
- Mata, Miguel, Steven Neben, Karim Majzoub, Jan Carette, Muthukumar Ramanathan, Paul A. Khavari, and Peter Sarnow. 2019. 'Impact of a Patient-Derived Hepatitis C Viral Rna Genome with a Mutated MicroRNA Binding Site'. *PLoS Pathogens* 15(5):e1007467. doi: 10.1371/journal.ppat.1007467.

- Merino, Edward J., Kevin A. Wilkinson, Jennifer L. Coughlan, and Kevin M. Weeks. 2005. 'RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE)'. *Journal of the American Chemical Society* 127(12). doi: 10.1021/ja043822v.
- Milligan, John F., Duncan R. Groebe, Gary W. Witherell, and Olke C. Uhlenbeck. 1987. 'Oligoribonucleotide Synthesis Using T7 RNA Polymerase and Synthetic DNA Templates'. *Nucleic Acids Research* 15(21):8783–98. doi: 10.1093/NAR/15.21.8783.
- Misra, Vinod K., Ross Shiman, and David E. Draper. 2003. 'A Thermodynamic Framework for the Magnesium-Dependent Folding of RNA'. *Biopolymers* 69(1):118–36. doi: 10.1002/BIP.10353.
- Moore, Michael D., and Wei Shau Hu. 2009. 'HIV-1 RNA Dimerization: It Takes Two to Tango'. *AIDS Reviews* 11(2):91.
- Moradpour, Darius, and François Penin. 2013. 'Hepatitis C Virus Proteins: From Structure to Function.' *Current Topics in Microbiology and Immunology* 369:113–42. doi: 10.1007/978-3-642-27340-7\_5.
- Moradpour, Darius, François Penin, and Charles M. Rice. 2007. 'Replication of Hepatitis C Virus'. *Nature Reviews. Microbiology* 5(6):453–63. doi: 10.1038/NRMICRO1645.
- Morikawa, K., C. M. Lange, J. Gouttenoire, E. Meylan, V. Brass, F. Penin, and D. Moradpour. 2011. 'Nonstructural Protein 3-4A: The Swiss Army Knife of Hepatitis C Virus'. *Journal of Viral Hepatitis* 18(5):305–15. doi: 10.1111/J.1365-2893.2011.01451.X.
- Mortimer, Stefanie A., and Jennifer A. Doudna. 2013. 'Unconventional MiR-122 Binding Stabilizes the HCV Genome by Forming a Trimolecular RNA Structure.' *Nucleic Acids Research* 41(7):4230–40. doi: 10.1093/nar/gkt075.
- Mylonas, Efstratios, and Dmitri I. Svergun. 2007. 'Accuracy of Molecular Mass Determination of Proteins in Solution by Small-Angle X-Ray Scattering'. *Journal of Applied Crystallography* 40(s1):s245–49. doi: 10.1107/S002188980700252X.
- Negro, Francesco, Daniel Forton, Antonio Craxì, Mark S. Sulkowski, Jordan J. Feld, and Michael P. Manns. 2015. 'Extrahepatic Morbidity and Mortality of Chronic Hepatitis C'. *Gastroenterology* 149(6):1345–60.
- Olenginski, Lukasz T., Kehinde M. Taiwo, Regan M. Leblanc, and Theodore K. Dayie. 2021. 'Isotope-Labeled RNA Building Blocks for NMR Structure and Dynamics Studies'. *Molecules* 26(18). doi: 10.3390/MOLECULES26185581.

- Oliva, Alexis, Matias Llabres, and Jose Farina. 2005. 'Applications of Multi-Angle Laser Light-Scattering Detection in the Analysis of Peptides and Proteins'. *Current Drug Discovery Technologies* 1(3):229–42. doi: 10.2174/1570163043334938.
- Ono, C. 2020. 'Various MiRNAs Compensate the Role of MiR- 122 on HCV Replication'. 1–9.
- Ottosen, Søren, Todd B. Parsley, Lu Yang, Karin Zeh, Leen Jan van Doorn, Eva van der Veer, Anneke K. Raney, Michael R. Hodges, and Amy K. Patick. 2015. 'In Vitro Antiviral Activity and Preclinical and Clinical Resistance Profile of Miravirsin, a Novel Anti-Hepatitis C Virus Therapeutic Targeting the Human Factor MiR-122'. *Antimicrobial Agents and Chemotherapy* 59(1):599–608. doi: 10.1128/AAC.04220-14.
- Panjkovich, Alejandro, and Dmitri I. Svergun. 2018. 'CHROMIXS: Automatic and Interactive Analysis of Chromatography-Coupled Small-Angle X-Ray Scattering Data'. *Bioinformatics* 34(11):1944–46. doi: 10.1093/BIOINFORMATICS/BTX846.
- Pérard, Julien, Cédric Leyrat, Florence Baudin, Emmanuel Drouet, and Marc Jamin. 2013. 'Structure of the Full-Length HCV IRES in Solution'. *Nature Communications* 4(1):1612. doi: 10.1038/ncomms2611.
- Pestova, Tatyana v., Ivan N. Shatsky, Simon P. Fletcher, Richard J. Jackson, and Christopher U. T. Hellen. 1998. 'A Prokaryotic-like Mode of Cytoplasmic Eukaryotic Ribosome Binding to the Initiation Codon during Internal Translation Initiation of Hepatitis C and Classical Swine Fever Virus RNAs'. *Genes & Development* 12(1):67–83. doi: 10.1101/gad.12.1.67.
- Petrov, Alexey, Tinghe Wu, Elisabetta Viani Puglisi, and Joseph D. Puglisi. 2013. 'RNA Purification by Preparative Polyacrylamide Gel Electrophoresis'. *Methods in Enzymology* 530:315–30. doi: 10.1016/B978-0-12-420037-1.00017-8.
- Pettersen, Eric F., Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. 2004. 'UCSF Chimera--a Visualization System for Exploratory Research and Analysis'. *Journal of Computational Chemistry* 25(13):1605–12. doi: 10.1002/JCC.20084.
- Pierce, Levi C. T., Romelia Salomon-Ferrer, Cesar Augusto F. De Oliveira, J. Andrew McCammon, and Ross C. Walker. 2012. 'Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics'. *Journal of Chemical Theory and Computation* 8(9). doi: 10.1021/ct300284c.

- Pines, A., M. G. Gibby, and J. S. Waugh. 2003. 'Proton-Enhanced Nuclear Induction Spectroscopy. A Method for High Resolution NMR of Dilute Spins in Solids'. *The Journal of Chemical Physics* 56(4):1776. doi: 10.1063/1.1677439.
- Pleiss, Jeffrey A., Maria L. Derrick, and Olke C. Uhlenbeck. 1998. 'T7 RNA Polymerase Produces 5' End Heterogeneity during in Vitro Transcription from Certain Templates'. *T7 RNA Polymerase Produces 5' End Heterogeneity during in Vitro Transcription from Certain Templates*. 1313–17.
- Polenova, Tatyana, Rupal Gupta, and Amir Goldbourt. 2015. 'Magic Angle Spinning NMR Spectroscopy: A Versatile Technique for Structural and Dynamic Analysis of Solid-Phase Systems'. *Anal. Chem* 87:14. doi: 10.1021/ac504288u.
- Powdrill, Megan H., Egor P. Tchesnokov, Robert A. Kozak, Rodney S. Russell, Ross Martin, Evguenia S. Svarovskaia, Hongmei Mo, Roger D. Kouyos, and Matthias Götze. 2011. '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 of the United States of America* 108(51):20509–13. doi: 10.1073/pnas.1105797108.
- Price, Daniel J., and Charles L. Brooks. 2004. 'A Modified TIP3P Water Potential for Simulation with Ewald Summation'. *Journal of Chemical Physics* 121(20). doi: 10.1063/1.1808117.
- Putnam, Christopher D., Michal Hammel, Greg L. Hura, and John A. Tainer. 2007. 'X-Ray Solution Scattering (SAXS) Combined with Crystallography and Computation: Defining Accurate Macromolecular Structures, Conformations and Assemblies in Solution'. *Quarterly Reviews of Biophysics* 40(3):191–285.
- Quade, Nick, Daniel Boehringer, Marc Leibundgut, Joop van den Heuvel, and Nenad Ban. 2015. 'Cryo-EM Structure of Hepatitis C Virus IRES Bound to the Human Ribosome at 3.9-Å Resolution'. *Nature Communications* 6(May):1–9. doi: 10.1038/ncomms8646.
- Raju, B., E. Murphy, L. A. Levy, R. D. Hall, and R. E. London. 1989. 'A Fluorescent Indicator for Measuring Cytosolic Free Magnesium'. <https://doi.org/10.1152/Ajpcell.1989.256.3.C540> 256(3). doi: 10.1152/AJPCCELL.1989.256.3.C540.
- Raney, Kevin D., Suresh D. Sharma, Ibrahim M. Moustafa, and Craig E. Cameron. 2010. 'Hepatitis C Virus Non-Structural Protein 3 (HCV NS3): A Multifunctional Antiviral Target \*'. *Journal of Biological Chemistry* 285(30):22725–31. doi: 10.1074/JBC.R110.125294.

- van der Ree, Meike H., J. Marleen de Vree, Femke Stelma, Sophie Willemse, Marc van der Valk, Svend Rietdijk, Richard Molenkamp, Janke Schinkel, Ad C. van Nuenen, Ulrich Beuers, Salah Hadi, Marten Harbers, Eva van der Veer, Kai Liu, John Grundy, Amy K. Patick, Adam Pavlicek, Jacqueline Blem, Michael Huang, Paul Grint, Steven Neben, Neil W. Gibson, Neeltje A. Kootstra, and Hendrik W. Reesink. 2017. 'Safety, Tolerability, and Antiviral Effect of RG-101 in Patients with Chronic Hepatitis C: A Phase 1B, Double-Blind, Randomised Controlled Trial'. *The Lancet* 389(10070):709–17. doi: 10.1016/S0140-6736(16)31715-9.
- Reining, Anke, Senada Nozinovic, Kai Schlepckow, Florian Buhr, Boris Fürtig, and Harald Schwalbe. 2013. 'Three-State Mechanism Couples Ligand and Temperature Sensing in Riboswitches'. *Nature* 2013 499:7458 499(7458):355–59. doi: 10.1038/nature12378.
- Reynolds, Joanna E., Ann Kaminski, Hilka J. Kettinen, Ken Grace, Berwyn E. Clarke, Anthony R. Carroll, David J. Rowlands, and Richard J. Jackson. 1995. 'Unique Features of Internal Initiation of Hepatitis C Virus RNA Translation'. *EMBO Journal*. doi: 10.1002/j.1460-2075.1995.tb00289.x.
- Rijnbrand, René, Peter Bredenbeek, Tahar van der Straaten, Linda Whetter, Geneviève Inchauspé, Stanley Lemon, and Willy Spaan. 1995. 'Almost the Entire 5' Non-Translated Region of Hepatitis C Virus Is Required for Cap-Independent Translation'. *FEBS Letters*. doi: 10.1016/0014-5793(95)00458-L.
- Rijnbrand, Rene, Varatharasa Thiviyanathan, Kumaralal Kaluarachchi, Stanley M. Lemon, and David G. Gorenstein. 2004. 'Mutational and Structural Analysis of Stem-Loop IIIc of the Hepatitis C Virus and GB Virus B Internal Ribosome Entry Sites'. *Journal of Molecular Biology* 343(4):805–17. doi: 10.1016/J.JMB.2004.08.095.
- Rivas, Fabiola v., Niraj H. Tolia, Ji Joon Song, Juan P. Aragon, Jidong Liu, Gregory J. Hannon, and Leemor Joshua-Tor. 2005. 'Purified Argonaute2 and an siRNA Form Recombinant Human RISC'. *Nature Structural & Molecular Biology* 2005 12:4 12(4):340–49. doi: 10.1038/nsmb918.
- Roberts, Ashley P. E., Andrew P. Lewis, and Catherine L. Jopling. 2011. 'MiR-122 Activates Hepatitis C Virus Translation by a Specialized Mechanism Requiring Particular RNA Components'. *Nucleic Acids Research* 39(17):7716–29. doi: 10.1093/nar/gkr426.
- Roe, Daniel R., and Thomas E. Cheatham. 2013. 'PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data'. *Journal of Chemical Theory and Computation* 9(7). doi: 10.1021/ct400341p.

- Romani, A., and A. Scarpa. 1992. 'Regulation of Cell Magnesium'. *Archives of Biochemistry and Biophysics* 298(1):1–12. doi: 10.1016/0003-9861(92)90086-C.
- Rong, Libin, and Alan S. Perelson. 2010. 'Treatment of Hepatitis c Virus Infection with Interferon and Small Molecule Direct Antivirals: Viral Kinetics and Modeling'. *Critical Reviews in Immunology* 30(2):131–48.
- Ross-Thriepfand, Douglas, and Mark Harris. 2015. 'Hepatitis C Virus NS5A: Enigmatic but Still Promiscuous 10 Years On!' *Journal of General Virology* 96(4):727–38.
- Salmon-Ferrer, Romelia, Andreas W. Goetz, Duncan Poole, Scott Le Grand, and Ross C. Walker. 2013. 'Routine Microsecond Molecular Dynamics Simulations with AMBER - Part II: Particle Mesh Ewald'. *Journal of Chemical Theory and Computation* 9(9).
- Salzman, David W., Kotoka Nakamura, Sunitha Nallur, Michelle T. Dookwah, Chanatip Methetrairut, Frank J. Slack, and Joanne B. Weidhaas. 2016. 'MiR-34 Activity Is Modulated through 5'-End Phosphorylation in Response to DNA Damage'. *Nature Communications* 7. doi: 10.1038/NCOMMS10954.
- Santoro, Jorge, and Garry C. King. 1992. 'A Constant-Time 2D Overbodenhausen Experiment for Inverse Correlation of Isotopically Enriched Species'. *Journal of Magnetic Resonance (1969)* 97(1):202–7. doi: 10.1016/0022-2364(92)90250-B.
- Scheel, Troels K. H., Joseph M. Luna, Matthias Liniger, Eiko Nishiuchi, Kathryn Rozen-Gagnon, Amir Shlomai, Gaël Auray, Markus Gerber, John Fak, Irene Keller, Rémy Bruggmann, Robert B. Darnell, Nicolas Ruggli, and Charles M. Rice. 2016. 'A Broad RNA Virus Survey Reveals Both MiRNA Dependence and Functional Sequestration'. *Cell Host and Microbe*. doi: 10.1016/j.chom.2016.02.007.
- Schult, Philipp, Hanna Roth, Rebecca L. Adams, Caroline Mas, Lionel Imbert, Christian Orlik, Alessia Ruggieri, Anna M. Pyle, and Volker Lohmann. 2018. 'MicroRNA-122 Amplifies Hepatitis C Virus Translation by Shaping the Structure of the Internal Ribosomal Entry Site'. *Nature Communications* 9(1):2613. doi: 10.1038/s41467-018-05053-3.
- Sedano, Cecilia D., and Peter Sarnow. 2014. 'Hepatitis C Virus Subverts Liver-Specific MiR-122 to Protect the Viral Genome from Exoribonuclease Xrn2'. *Cell Host and Microbe* 16(2):257–64. doi: 10.1016/j.chom.2014.07.006.
- Shimakami, T., D. Yamane, C. Welsch, L. Hensley, R. K. Jangra, and S. M. Lemon. 2012. 'Base Pairing between Hepatitis C Virus RNA and MicroRNA 122 3' of Its Seed Sequence Is Essential for Genome Stabilization and Production of Infectious Virus'. *Journal of Virology* 86(13):7372–83. doi: 10.1128/jvi.00513-12.

- Spahn, C. M. T., J. S. Kieft, R. A. Grassucci, P. A. Penczek, K. Zhou, J. A. Doudna, and J. Frank. 2001. 'Hepatitis C Virus IRES RNA-Induced Changes in the Conformation of the 40S Ribosomal Subunit'. *Science* 291(5510):1959–62. doi: 10.1126/science.1058409.
- Spearman, C. Wendy, Geoffrey M. Dusheiko, Margaret Hellard, and Mark Sonderup. 2019. 'Hepatitis C'. *The Lancet* 394(10207):1451–66. doi: 10.1016/S0140-6736(19)32320-7.
- Steinmann, Eike, and Thomas Pietschmann. 2013. 'Hepatitis C Virus: Hepatitis C Virus RNA Translation'. 369:17–49. doi: 10.1007/978-3-642-27340-7.
- Summa, Vincenzo, Steven W. Ludmerer, John A. McCauley, Christine Fandozzi, Christine Burlein, Giuliano Claudio, Paul J. Coleman, Jillian M. DiMuzio, Marco Ferrara, Marcello di Filippo, Adam T. Gates, Donald J. Graham, Steven Harper, Daria J. Hazuda, Carolyn McHale, Edith Monteagudo, Vincenzo Pucci, Michael Rowley, Michael T. Rudd, Aileen Soriano, Mark W. Stahlhut, Joseph P. Vacca, David B. Olsen, Nigel J. Liverton, and Steven S. Carroll. 2012. 'MK-5172, a Selective Inhibitor of Hepatitis C Virus NS3/4a Protease with Broad Activity across Genotypes and Resistant Variants'. *Antimicrobial Agents and Chemotherapy* 56(8):4161. doi: 10.1128/AAC.00324-12.
- Svergun, D., C. Barberato, and M. H. Koch. 1995. 'CRY SOL – a Program to Evaluate X-Ray Solution Scattering of Biological Macromolecules from Atomic Coordinates'. *Urn:Issn:0021-8898* 28(6):768–73. doi: 10.1107/S0021889895007047.
- Svergun, D. I. 1992. 'Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria'. *Journal of Applied Crystallography* 25(pt 4):495–503. doi: 10.1107/S0021889892001663.
- Svergun, D. I. 1999. 'Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing'. *Biophysical Journal* 76(6):2879–86. doi: 10.1016/S0006-3495(99)77443-6.
- Svergun, D. I., S. Richard, M. H. J. Koch, Z. Sayers, S. Kuprin, and G. Zaccai. 1998. 'Protein Hydration in Solution: Experimental Observation by x-Ray and Neutron Scattering'. *Proceedings of the National Academy of Sciences* 95(5):2267–72. doi: 10.1073/PNAS.95.5.2267.
- Svergun, Dmitri I., Michel H. J. Koch, Peter A. Timmins, and Roland P. May. 2013. *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*.

- Svergun, Dmitri I., and Knud H. Nierhaus. 2000. 'A Map of Protein-RRNA Distribution in the 70 S Escherichia Coli Ribosome'. *Journal of Biological Chemistry* 275(19):14432–39. doi: 10.1074/JBC.275.19.14432.
- Takegoshi, K., Shinji Nakamura, and Takehiko Terao. 2001. '13C–1H Dipolar-Assisted Rotational Resonance in Magic-Angle Spinning NMR'. *Chemical Physics Letters* 344(5–6):631–37. doi: 10.1016/S0009-2614(01)00791-6.
- Thibault, Patricia A., Adam Huys, Yalena Amador-Cañizares, Julie E. Gailius, Dayna E. Pinel, and Joyce A. Wilson. 2015. 'Regulation of Hepatitis C Virus Genome Replication by Xrn1 and MicroRNA-122 Binding to Individual Sites in the 5' Untranslated Region'. doi: 10.1128/JVI.03631-14.
- Thompson, Aaron A., Aihua Zou, Jiangli Yan, Rohit Duggal, Weidong Hao, David Molina, Ciarán N. Cronin, and Peter A. Wells. 2009. 'Biochemical Characterization of Recombinant Hepatitis C Virus Nonstructural Protein 4B: Evidence for ATP/GTP Hydrolysis and Adenylate Kinase Activity.' *Biochemistry* 48(5):906–16. doi: 10.1021/bi801747p.
- Tsukiyama-Kohara, K., N. Iizuka, M. Kohara, and A. Nomoto. 1992. 'Internal Ribosome Entry Site within Hepatitis C Virus RNA.' *Journal of Virology* 66(3):1476–83. doi: 10.1128/jvi.66.3.1476-1483.1992.
- Valdmanis, Paul N., Hak Kyun Kim, Kirk Chu, Feijie Zhang, Jianpeng Xu, Elizabeth M. Munding, Jia Shen, and Mark A. Kay. 2018. 'MiR-122 Removal in the Liver Activates Imprinted MicroRNAs and Enables More Effective MicroRNA-Mediated Gene Repression'. *Nature Communications*. doi: 10.1038/s41467-018-07786-7.
- Voss, N. R., and M. Gerstein. 2005. 'Calculation of Standard Atomic Volumes for RNA and Comparison with Proteins: RNA Is Packed More Tightly'. *Journal of Molecular Biology* 346(2):477–92. doi: 10.1016/J.JMB.2004.11.072.
- Vranken, Wim F., Wayne Boucher, Tim J. Stevens, Rasmus H. Fogh, Anne Pajon, Miguel Llinas, Eldon L. Ulrich, John L. Markley, John Ionides, and Ernest D. Laue. 2005. 'The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline'. *Proteins: Structure, Function, and Bioinformatics* 59(4):687–96. doi: 10.1002/PROT.20449.
- Welzel, Tania M., Neeru Bhardwaj, Charlotte Hedskog, Krishna Chodavarapu, Gregory Camus, John McNally, Diana Brainard, Michael D. Miller, Hongmei Mo, Evguenia Svarovskaia, Ira Jacobson, Stefan Zeuzem, and Kosh Agarwal. 2017. 'Global Epidemiology of HCV Subtypes and Resistance-Associated Substitutions Evaluated by



- Sequencing-Based Subtype Analyses'. *Journal of Hepatology* 67(2):224–36. doi: 10.1016/j.jhep.2017.03.014.
- WHO. 2016. 'WHO | Global Health Sector Strategy on Viral Hepatitis 2016-2021'.
- Wiegand, Thomas, Denis Lacabanne, Anahit Torosyan, Julien Boudet, Riccardo Cadalbert, Frédéric H. T. Allain, Beat H. Meier, and Anja Böckmann. 2020. 'Sedimentation Yields Long-Term Stable Protein Samples as Shown by Solid-State NMR'. *Frontiers in Molecular Biosciences* 7:17. doi: 10.3389/FMOLB.2020.00017/BIBTEX.
- Wilkinson, Jeffrey, Marek Radkowski, and Tomasz Laskus. 2009. 'Hepatitis C Virus Neuroinvasion: Identification of Infected Cells'. *Journal of Virology* 83(3):1312–19. doi: 10.1128/JVI.01890-08.
- Williamson, M. P. 2013. 'Using Chemical Shift Perturbation to Characterise Ligand Binding'. *Progress in Nuclear Magnetic Resonance Spectroscopy* 73:1–16. doi: 10.1016/j.pnmrs.2013.02.001.
- Yamagami, Ryota, Jacob P. Sieg, and Philip C. Bevilacqua. 2021. 'Functional Roles of Chelated Magnesium Ions in RNA Folding and Function'. *Biochemistry* 60(31):2374–86. doi: 10.1021/acs.biochem.1c00012.
- Yu, Yingpu, Troels K. H. Scheel, Joseph M. Luna, Hachung Chung, Eiko Nishiuchi, Margaret A. Scull, Natalia Echeverría, Inna Ricardo-Lax, Amit Kapoor, Ian W. Lipkin, Thomas J. Divers, Douglas F. Antczak, Bud C. Tennant, and Charles M. Rice. 2017. 'MiRNA Independent Hepacivirus Variants Suggest a Strong Evolutionary Pressure to Maintain MiR-122 Dependence'. *PLoS Pathogens*. doi: 10.1371/journal.ppat.1006694.
- Zhao, Sha, Yufei Yang, Yujie Zhao, Xinming Li, Yi Xue, and Shenlin Wang. 2019a. 'High-Resolution Solid-State NMR Spectroscopy of Hydrated Non-Crystallized RNA'. *Chemical Communications* 55(93):13991–94. doi: 10.1039/C9CC06552K.
- Zhao, Sha, Yufei Yang, Yujie Zhao, Xinming Li, Yi Xue, and Shenlin Wang. 2019b. 'High-Resolution Solid-State NMR Spectroscopy of Hydrated Non-Crystallized RNA'. *Chemical Communications* 55(93):13991–94. doi: 10.1039/C9CC06552K.
- Zhao, Wei Dong, and Eckard Wimmer. 2001. 'Genetic Analysis of a Poliovirus/Hepatitis C Virus Chimera: New Structure for Domain II of the Internal Ribosomal Entry Site of Hepatitis C Virus'. *Journal of Virology* 75(8):3719–30. doi: 10.1128/jvi.75.8.3719-3730.2001.
- Zimm, Bruno H. 2004. 'The Scattering of Light and the Radial Distribution Function of High Polymer Solutions'. *The Journal of Chemical Physics* 16(12):1093. doi: 10.1063/1.1746738.

Zwahlen, Catherine, Pascale Legault, Sébastien J. F. Vincent, Jack Greenblatt, Robert Konrat, and Lewis E. Kay. 1997. 'Methods for Measurement of Intermolecular NOEs by Multinuclear NMR Spectroscopy: Application to a Bacteriophage  $\lambda$  N-Peptide/BoxB RNA Complex'. *Journal of the American Chemical Society* 119(29):6711–21. doi: 10.1021/JA970224Q.

## 8. Extended data

**Table 1. DNA oligonucleotide sequence** for insertions to clone into the pUC19 plasmid

Name	Primer	Sequence (5'→3')
miR-HH	Coding	AATTCTAATACGACTCACTATAGTGGAGTGTGAC AATGGTGTTTGTCTGTCTGTCGCCCTGCAGA
	Non-Coding	AGCTTCTGCAGGGGCGACAGAC/GACAAACACCAT TGTCACACTCCACTATAGTGAGTCGTATTAG
Domain I	Coding	AATTCTAATACGACTCACTATAGCCAGCCCCCTG ATGGGGGCGACACTCCACCATGAATCACTCCCCT GCTGCAGA
	Non-Coding	AGCTTCTGCAGCAGGGGAGTGATTCATGGTGGA GTGTCGCCCCCATCAGGGGGCTGGCTATAGTGAG TCGTATTAG
Binding site 1	Coding	AATTCTAATACGACTCACTATAGCCAGCCCCCTG ATGGGGGCGACACTCCACTGCAGA
	Non-Coding	AGCTTCTGCAGTGGAGTGTCTGCCCCCATCAGGGG GCTGGCTATAGTGAGTCGTATTAG
34mer	Coding	AATTCTAATACGACTCACTATAGGCCGCCAGCCC CCTGATGGGGGCGACACTCCACTGCAGA
	Non-Coding	AGCTTCTGCAGTGGAGTGTCTGCCCCCATCAGGGG GCTGGCGGCCTATAGTGAGTCGTATTAG

Primer	Sequence (5'→3')
M13 Forward	GTTTCCCAGTCACGAC
M13 Reverse	CAGGAAACAGCTATGAC

**Table 2. The sequence of M13 primers** used for sequencing of pUC plasmid

**Table 3. Primers for site directed mutagenesis.** The nucleotides, that belong to overhanging primer's ends with the insertion sequence are in lowercase, and the nucleotides complementary to plasmid's sequence are in uppercase. T<sub>a</sub> indicates annealing temperature used in PCR reaction.

Primer's name	Sequence (5'→3')	T <sub>a</sub> , °C
HH-UTR F	aggacgaaacgagctagctcgteGCCAGCCCCCTGATGGGG	71
HH-UTR R	cacggactcatcaggccagccccTATAGTGAGTCGTATTAGAATTCGT TCCATCCCAATAC	
5' insertion F	tgtgacaatggtgtttgtTGGAGTGTGACAATGGTG	59
5' insertion R	ctccaacaacacctttccCTATAGTGAGTCGTATTAGAATTC	
3' subins F	gtgtgacaatggtgtttgttgagtgCTGCAGAAGCTTGGCGTA	66
3' subins R	tccaacaacaccattgtcacactccaACAAACACCATTGTCACACTC	

**Table 4. RNA constructs** used in this study. Unnatural nucleotide from PstI restriction site is labeled in red.

Name	Sequence (5' →3')
5' UTR (359mer)	GCCAGCCCCCUGAUGGGGGCGACACUCCACCAUGAAU CACUCCCCUGUGAGGAACUACUGUCUUCACGCAGAAA GCGUCUAGCCAUGGCGUUAGUAUGAGUGUCGUGCAGC CUCCAGGACCCCCCUCCCGGGAGAGCCAUAUGUGGUC UGC GGAACCGGUGAGUACACCGGAAUUGCCAGGACGA CCGGGUCCUUUCUUGGAUUAACCCGCUCAAUGCCUGG AGAUUUGGGCGUGCCCCCGCGAGACUGCUAGCCGAGU AGUGUUGGGUCGCGAAAGGCCUUGUGGUACUGCCUGA UAGGGUGCUUGCGAGUGCCCCGGGAGGUCUCGUAGAC CGUGCAUCAUGAGCACAAAUCCUGAA
Domains I-II (118mer)	GCCAGCCCCCUGAUGGGGGCGACACUCCACCAUGAAU CACUCCCCUGUGAGGAACUACUGUCUUCACGCAGAAA GCGUCUAGCCAUGGCGUUAGUAUGAGUGUCGUGCAGC CUCCAGG
Domain I (48mer)	GCCAGCCCCCUGAUGGGGGCGACACUCCACCAUGAAU CACUCCCCUGC
Binding site 1 (30mer)	GCCAGCCCCCUGAUGGGGGCGACACUCCAC
34mer	GGCCGCCAGCCCCCUGAUGGGGGCGACACUCCAC
Stem-loop 1 (SL 1)	GCCCCUGAUGGGGGC
miR-copy (322mer)	GGGAAAGGUGUUUGUUGGAGUGUGACAAUGGUGUUU GUUGGAGUGUGACAAUGGUGUUUGUUGGAGUGUGAC AAUGGUGUUUGUUGGAGUGUGACAAUGGUGUUUGU GGAGUGUGACAAUGGUGUUUGUUGGAGUGUGACAAU GGUGUUUGUUGGAGUGUGACAAUGGUGUUUGUUGGA GUGUGACAAUGGUGUUUGUUGGAGUGUGACAAUGGU GUUUGUUGGAGUGUGACAAUGGUGUUUGUUGGAGUG UGACAAUGGUGUUUGUUGGAGUGUGACAAUGGUGUU UGUUGGAGUGUGACAAUGGUGUUUGUUGGAGUGC
miR-122 (23mer)	UGGAGUGUGACAAUGGUGUUUGU
miR-122 modified	UGGAGUGUGACAAUGGUGGCCGU

**Table 5. Primers and their annealing temperatures used for DNA template production by PCR.**

RNA construct	PCR template	Primers (5'→3')		T <sub>a</sub> , °C	DNA product (reverse chain 5'→3')
5' UTR (359mer)	UTR	UTR F	TAATACGACTCACT ATAGCCAGCC	64	[mT][mT]CAGGATTTGTGCTCATGATGCACGGTC TACGAGACCTCCCGGGGCACTCGCAAGCACCCCT ATCAGGCAGTACCACAAGGCCTTTCGCGACCCA ACACTACTCGGCTAGCAGTCTCGCGGGGGCACG CCCAAATCTCCAGGCATTGAGCGGGTTAATCCA AGAAAGGACCCGGTCGTCCTGGCAATTCCGGTG TACTCACCGGTTCCGCAGACCACTATGGCTCTCC CGGGAGGGGGGGTCTGGAGGCTGCACGACACT CATACTAACGCCATGGCTAGACGCTTTCTGCGT GAAGACAGTAGTTCCTCACAGGGGAGTGATTCA TGGTGGAGTGTCGCCCCCATCAGGGGGCTGGCT ATAGTGAGTCGTATTA
		UTR Rmet	[mT][mT]CAGGATTT GTGCTCATGATGC		
HH-domains I-II	HH-UTR	HH-118 F	ATTCTAATACGACTC ACTATAGGGGGCT	66	[mC][mC]TGGAGGCTGCACGACACTCATACTAAC GCCATGGCTAGACGCTTTCTGCGTGAAGACAGT AGTTCCTCACAGGGGAGTGATTTCATGGTGGAGT GTCGCCCCCATCAGGGGGCTGGCGACGAGCTAG CTCGTTTCGTCTCACGGACTCATCAGGCCAGCC CCCTATAGTGAGTCGTATTAGAATT
		HH-118 Rmet	[mC][mC]TGGAGGCTG CACGACACTCATACTA AC		

**Table 6. Chimeras used for site-specific RNaseH cleavage.** mN = methylated ribonucleotide; dN = deoxyribonucleotide, “/” indicates the cleavage site.

<b>Chimera's name</b>	<b>Sequence (5'→3')</b>
miRchimera	[mC][mU][mC][mC][mA]/[dA][dC][dA][dA][mA][mC][mA][mC]
1-2 chimera	[mU][mC][mA][mC][mU][mG][mG]/dGdGdAdG[mU][mG][mA][mU][mU][mC]
2-3 shortchimera	[mG][mG][mU][mC][mC][mU][mG]/GAGG[mC][mU][mG][mC][mA][mC][mG]
2-3 longchimera	[mG][mG][mA][mG][mG][mG][mG][mG][mG][mU][mC][mC][mU][mG]/GAGG[mC][mU][mG][mC][mA][mC][mG][mA][mC][mA][mC]

**Table 7. DNA splints** used for segmental RNA labeling.

<b>Splint's name</b>	<b>Sequence (5'→3')</b>
1-2 long	TCCTCACAGGGGAGTGATTC
2-3 short	GGGGGGTCCTGGAGGCTGCAC
1-2 long	GACAGTAGTTCCTCACAGGGGAGTGATTCATGGTGGAGTG
2-3 long	CTCCCGGGAGGGGGGGTCCTGGAGGCTGCACGACACTCATAC

**Table 8. Isotropic chemical shifts of domain I.** The resonances, which could not be assigned, are indicated with a dash and nuclear spins that do not exist for the specified atom type, are indicated by grey background. (P. Innig Aguion, Master thesis, 2020)

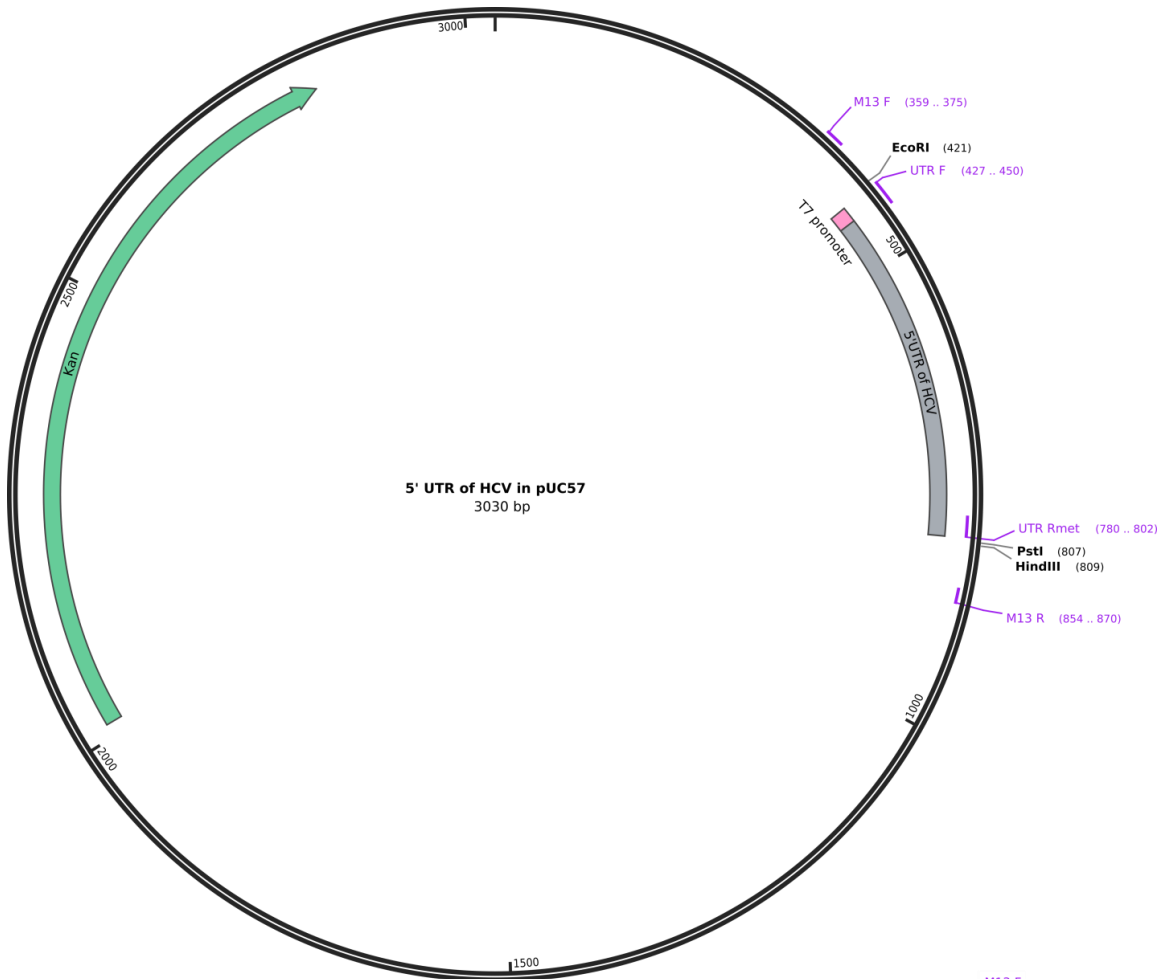
	C1'	C2	C2'	C3'	C4'	C5	C5'	C6	C8	H1'	H2	H2'	H3'	H4'	H5	H6	H8	N1	N9
2C	94.4		-	-	-	-	-	-		5.65		-	-	-	-	-		148.6	
3C	90.2		77.5	72.7	86.1	99.5	-	142.8		6.07		4.57	4.48	4.36	5.65	7.64		-	
4A	91.0	155.6	76.0	76.1	84.6		68.1		142.2	6.18	8.24	4.9	4.78	4.63	4.34.4.39		8.34		-
5G	92.6		-	-	-		-		137.5	5.45		-	-	-	-		7.81		-
6C	94.0		75.5	72.2	-	96.7	-	141.7		5.59		4.41	4.51	-	5.2	7.81		151.8	
7C	94.0		75.6	72.3	81.9	97.6	-	141.5		5.54		4.41	4.52	4.46	5.5	7.87		151.4	
8C	94.0		75.5	72.3	81.9	97.5	-	141.5		5.5		4.37	4.5	4.44	5.47	7.84		151.4	
10C	93.6		75.8	73.0	82.2	97.5	-	141.3		5.56		4.52	4.33	4.38	5.36	7.63		151.6	
11U	94.0		75.4	74.4	83.5	104.4	-	142.1		5.39		4.16	4.41	4.33	5.68	7.72		145.3	
12G	89.4		74.5	78.3	84.7		67.5		141.1	5.24		4.68	4.5	3.84	3.87.4.00		7.7		165.9
13A	91.2	155.3	76.6	77.6	84.9		-		142.7	5.91	8.09	4.9	5.17	4.39	-		8.06		169.7
14U	90.7		75.4	77.1	85.2	105.4	68.5	144.2		6.03		4.56	4.72	4.65	4.26.4.34.5.92	7.92		143.6	
15G	92.6		75.3	74.0	82.8		-		137.3	5.32		4.63	4.45	4.55	-		7.66		169.4
16G	92.8		75.5	72.8	81.8		65.5		136.4	5.84		4.6	4.52	4.52	4.13		7.34		169.8
17G	92.7		-	-	-		-		136.1	5.82		-	-	-	-		7.27		-
18G	92.7		-	-	-		-		136.0	5.79		-	-	-	-		7.22		-
19G	92.9		75.5	-	-		-		135.8	5.79		4.52	-	-	-		7.2		169.6
20C	94.0		75.5	72.1	81.9	97.8	-	140.1		5.49		4.36	4.49	4.42	5.14	7.37		150.5	
21G	92.5		75.7	-	-		-		137.1	5.73		4.47	-	-	-		7.76		169.8
22A	93.2	-	-	-	-		-		141.9	5.8	-	-	-	-	-		7.95		-
23C	93.2		75.8	-	-	97.7	-	140.0		5.35		4.35	-	-	5.39	7.5		151.3	
24A	92.8	-	75.9	-	-		-		139.4	5.95	-	4.39	-	-	-		7.87		170.6
25C	93.0		76.0	73.3	82.7	97.4	-	141.1		5.35		4.09	4.32	4.29	5.32	7.36		152.2	
26U	92.1		75.7	74.6	83.9	104.7	66.1	142.4		5.7		4.27	4.46	4.36	4.07.4.26.5.60	7.69		145.2	
27C	92.3		-	-	-	98.6	-	142.6		5.83		-	-	-	5.83	7.79		-	
28C	92.8		76.0	73.9	-	98.8	-	142.4		5.75		4.39	4.52	-	5.8	7.77		152.1	
29A	91.9	155.1	75.9	74.4	83.7		66.2		141.1	5.98	8.05	4.69	4.65	4.51	4.19.4.39		8.28		169.7
30C	93.0		76.0	73.8	83.2	97.7	65.7	141.6		5.52		4.22	4.39	4.35	4.13.5.48	7.49		151.8	
31C	92.6		76.1	74.6	83.3	98.7	66.0	142.2		5.68		4.24	4.51	4.27	4.09.4.23.5.73	7.68		152.1	
32A	90.4	-	76.4	76.3	84.6		67.0		141.5	5.94	-	4.63	4.75	4.48	4.16.4.31		8.3		169.1



33U	91.2		75.8	-	-	104.9	-	143.0		5.78		4.36	-	-	5.65	7.68		144.2	
34G	91.3		75.2	75.3	84.0		66.8		139.4	5.63		4.71	4.74	4.45	4.19.4.31		7.86		-
35A	91.5	154.8	75.8	74.6	83.7		66.3		141.2	5.91	7.91	4.68	4.75	4.53	4.21.4.43		8.16		169.6
36A	91.3	155.3	75.6	74.5	83.8		66.1		140.7	5.82	8.06	4.52	4.64	4.5	4.18.4.43		8.09		-
37U	91.7		75.7	75.0	84.2	104.4	66.4	142.3		5.6		4.26	4.49	4.37	4.11.4.31.5.45	7.55		144.6	
38C	92.5		-	-	-	98.7	-	142.8		5.76		-	-	-	5.81	7.73		151.7	
39A	91.5	155.2	75.9	74.9	83.9		66.6		141.3	6.01	8.07	4.71	4.71	4.53	4.21.4.37		8.31		169.6
40C	92.9		76.0	74.4	83.4	98.1	66.0	142.2		5.64		4.28	4.49	4.41	4.16.5.62	7.61		152.0	
41U	92.2		76.0	74.8	84.1	104.9	66.3	142.7		5.76		4.37	4.54	4.43	4.15.4.32.5.72	7.79		145.2	
42C	93.2		-	-	-	98.5	-	142.6		5.76		-	-	-	5.9	7.86		152.4	
43C	93.1		76.1	-	-	98.7	-	142.1		5.75		4.32	-	-	5.86	7.82		152.4	
44C	93.3		76.1	73.8	-	98.8	-	142.6		5.74		4.35	4.51	-	5.89	7.77		152.4	
45C	92.6		76.1	74.8	83.5	98.9	66.1	142.5		5.81		4.3	4.55	4.37	4.12.4.33.5.93	7.84		152.2	
46U	91.5		75.7	75.8	84.3	105.1	66.8	143.0		5.8		4.34	4.6	4.37	4.09.4.18.5.78	7.78		144.7	
47G	91.0		75.4	75.8	84.6		67.2		139.7	5.87		4.78	4.76	4.5	4.19.4.32		7.98		168.4
48C	92.5		77.1	71.5	84.9	98.5	66.8	143.5		5.9		4.22	4.3	4.25	4.15.4.34.5.87	7.82		152.8	

	Domains I-II		Domain II	
	H1	H3	H1	H3
<b>46U</b>		13.87		13.91
<b>47G</b>	12.07		12.11	
48U		–		–
49G	12.29		12.30	
51G	12.79		12.80	
52G	12.46		12.46	
56U		–		–
59U		13.84		13.83
60G	12.64		12.66	
<b>61U</b>		12.10		12.16
63U		–		–
64U		10.48		10.48
68G	–		–	
71G	–		–	
75G	11.40		11.40	
77G	12.90		12.91	
<b>78U</b>		12.18		12.22
80U		–		–
82G	–		–	
86U		–		–
87G	12.71		12.70	
88G	11.41		11.41	
90G	13.14		13.14	
<b>91U</b>		11.88		11.98
92U		–		–
94G	–		–	
95U		–		–
97U		–		–
98G	–		–	
100G	12.87		12.89	
<b>101U</b>		13.63		13.59
102G	12.85		12.84	
<b>103U</b>		11.21		11.26
105G	12.68		12.60 9	
107G	11.21		11.23	
110G	13.65		13.66	
113U		14.19		14.17
<b>117G</b>	13.05		13.24	
118G	–		–	

**Table 9. Chemical shift values of imino proton resonances of domains I-II and domain II constructs.** Chemical shift values of guanosine (H1) and uridine (H3) are listed. Unassigned resonances are indicated as “– “. The resonances, that do not exist for the specific residue, are indicated by grey background. The residues, having large CsP between domains I-II and domain II construct are shown in bold.



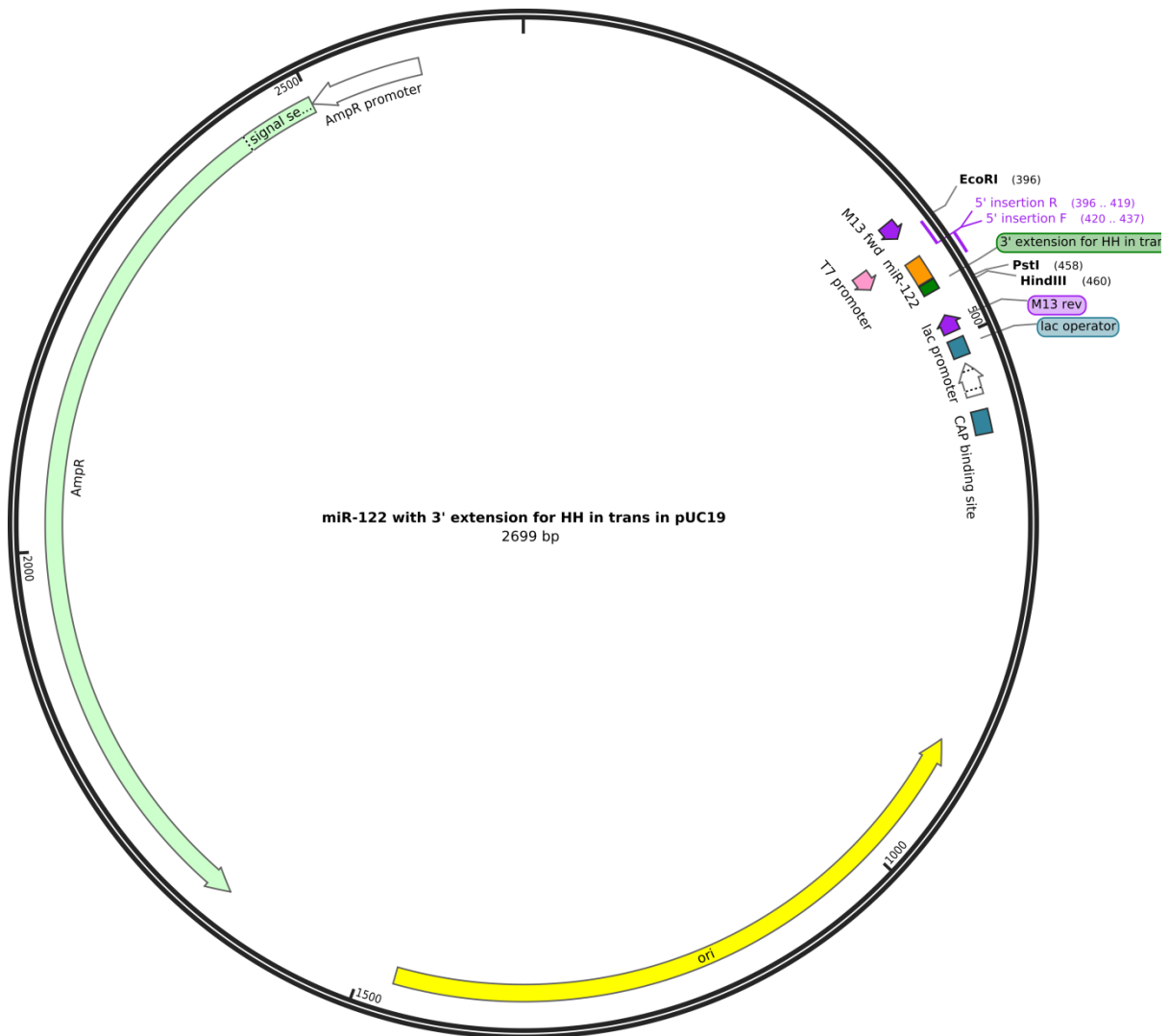
**Figure 1. Plasmid map for pUC57 vector with kanamycin resistance with an insertion of DNA template for HCV 5' UTR.** The primers used for amplification of 5' UTR template are indicated (UTR F, UTR Rmet). Two 5' terminal nucleotides of reverse primer were 2'-O-methylated.



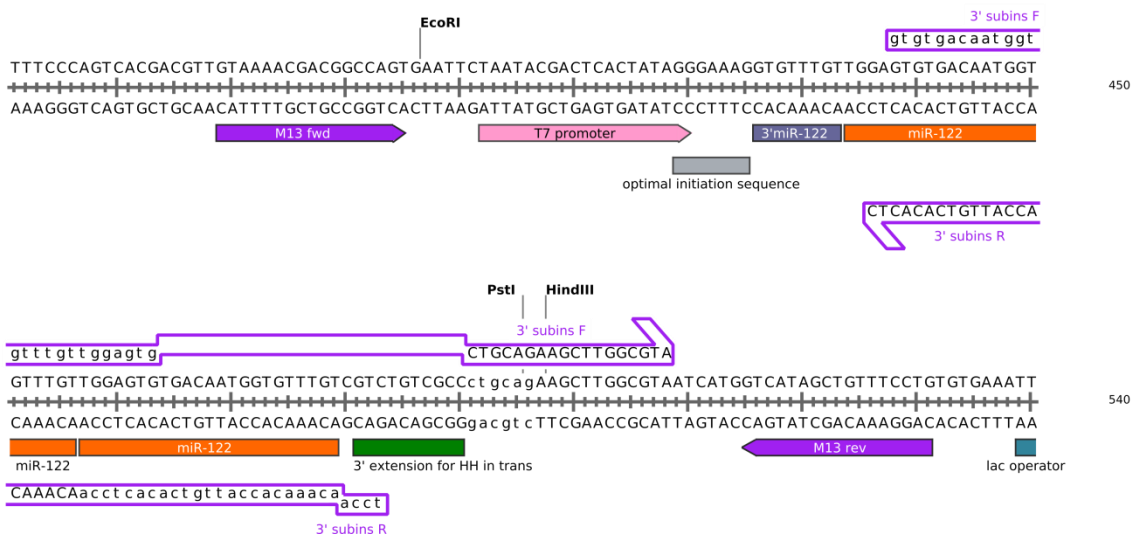
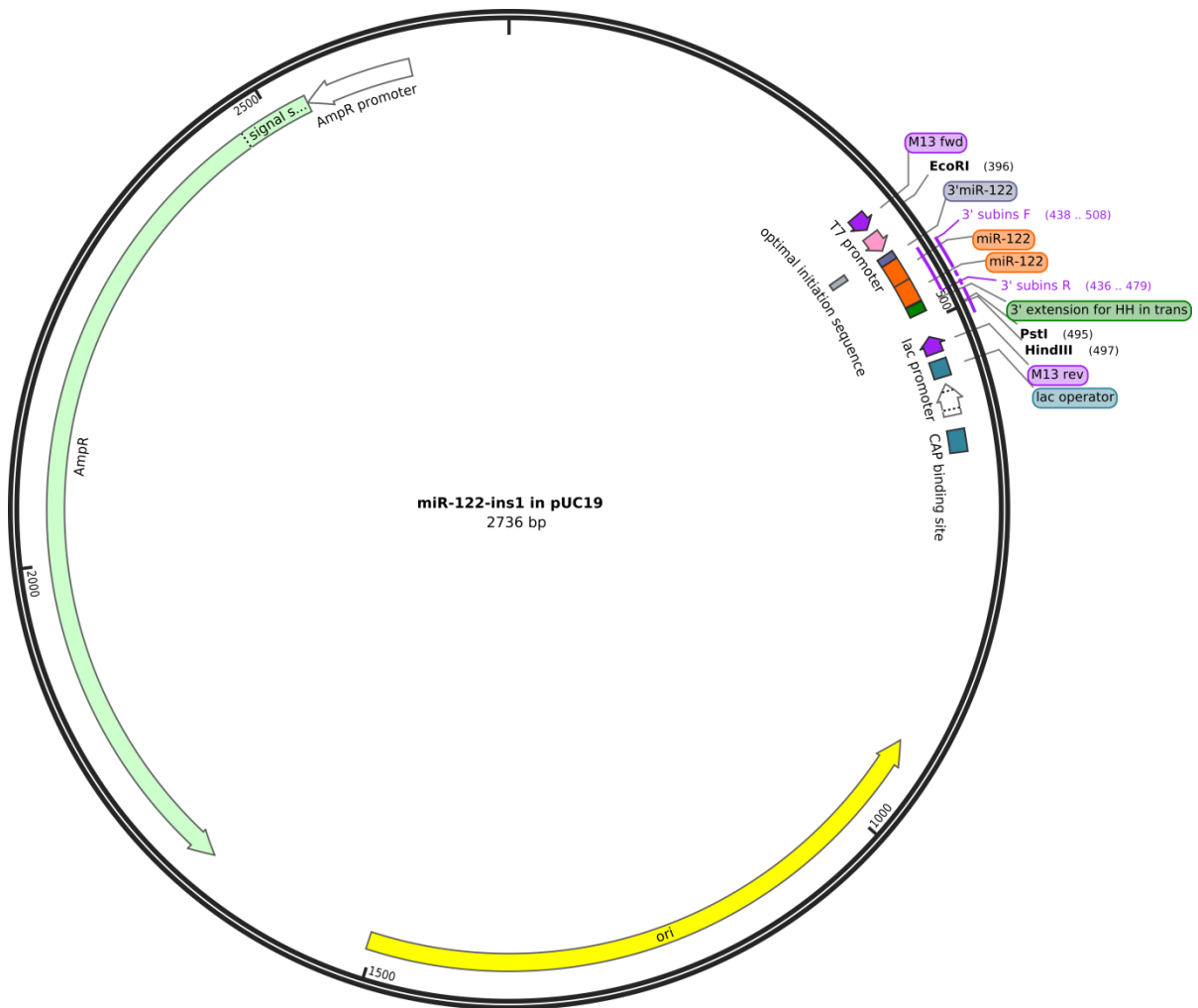
**Figure 2. The sequence of 5' UTR construct cloned into pUC19 vector used for 5' UTR DNA amplification.** Forward and reverse primers are mapped: UTR F and UTR Rmet. First two 5' nucleotides of reverse primer were 2'-O-methylated. M13 F and R are forward and reverse primers, which were used for plasmid sequencing.







**Figure 5. Plasmid map and sequence for pUC19 vector with ampicillin resistance with an insertion of DNA template for miR-122 with 3' extension for HH in trans (miR-HH) construct.** Primers (5' insertion F and R) for SDM to insert an optimal initiation sequence, 3' miR-122 end and one more copy of miR-122 are mapped. Last 5' nucleotides of reverse primer were 2'-O-methylated.



**Figure 6. Plasmid map and sequence for pUC19 vector with ampicillin resistance with an insertion of DNA template for an optimal initiation sequence, 3' miR-122 end, two copies of miR-122 with 3' extension for HH in trans (miRins1-HH) construct. Primers (3' subins F and R) for SDM to substitute 3' extension for HH in trans and insert 5' miR-122 end and more copies of miR-122 are mapped. First two 5' nucleotides of reverse primer were 2'-O-methylated.**



## 9. Acknowledgements

It has been an honour for me and a great privilege to complete my doctoral thesis at Centre of Biomolecular Drug Research of Leibniz University of Hannover.

First of all, I would like to thank my supervisor **Jun.-Prof. Dr. Alexander Marchanka**, for giving me the opportunity to work on this interesting and challenging project, for always being there to support me during difficult stages of my PhD, for my introduction to the beautiful world of RNA, NMR and the real science and for giving me the chance to learn a lot from him.

I express my gratitude to my official supervisor **Prof. Dr. Andreas Kirschning** for being in charge of my project, when my supervisor left the institution, for being always in contact with me and helping with the organization of defense and thesis reviewing.

I also want to express my gratitude to the thesis advisory committee, **Prof. Dr. Wulf Blankenfeldt and Prof. Dr. Thomas Brüser**, who kindly agreed to review my thesis, and for taking an interest in this research.

I would like to thank **Prof. Dr. Teresa Carlomagno**, who also was always there with her valuable advises on the project, who allowed me to collaborate with her group members and learn from them. I deeply appreciate the contribution of **Dr. John Kirkpatrick** to this project, helping with solution-state NMR experiments and sharing his knowledge with me with particular patience and kindness. I also want to thank **Dr. Luca Codutti** for introducing me to MALS and ITC techniques and for providing technical and scientific support.

I would like to express my gratitude to **Philipp Innig Aguion** for making an important impact to this thesis in the framework of his master project, completing assignment of domain I, carrying on the part of miR-122 and Mg<sup>2+</sup> titration experiments on domain I and setting up miR-122 synthesis protocol. It was a big pleasure to work with such a talented student.

I would like to thank **Dr. Frank Gabel** for his assistance with SANS data collection and his contribution to the data analysis. I want to thank **Dr. Peer Lukat** for setting up various crystallization screens and for fruitful discussions.

Furthermore, I want to thank each present and past members of Dr. Alexander Marchanka and Prof. Dr. Teresa Carlomagno groups, with whom I had an honor to work over these years of PhD: my best office-mate **Arun Kumar Sreemantula**; coffee mates, good friends and great scientists **Dr. Megha Karanth, Dr. Natalia Danilenko, Dr. Mumdooh Ahmed, Dr. Andrea Graziadei, Dr. Veena Hegde**; kindest lab-mates **Dr. Georg Krüger, Dr. Neha**

**Dhimole, Dr. Ying Wang, Dr. Simone Höfler, Dr. Marco Aurelio Wust, Dr. Michelangelo Marasco, Dr. Julia Buschmann, Kim Walbrunn and Vittoria Nanna.**

I would like to give special thanks to my friends **Sima Yurova** and **Dr. Alena Klochkova** for all of your support, hours on the phone in both bad and good times, even though there are thousands of kilometers between us.

I want to thank my husband **Vicente** for being the motivation to start this project, for having no doubts I can defend. I want to thank my mother- and father- in-laws **Toñi** and **Vicente** for the huge help during my maternity leave, without it I could not complete this thesis. I want to thank my beautiful son **Sasha**, who is my main inspiration and love.

Last but not least, I want to thank my family members: my deceased father **Alexandr**, mother **Maria** and sister **Irina** for being an important support in all areas from the day one and while working on the project and staying far from home, for constantly believing in me and their true care.

## 10. Curriculum Vitae

### Olga Vitsyna

#### Education

- 09.2016-exp. 09.2023 **Leibniz Universität Hannover, BMWZ (Hannover, Germany)**  
Dr. rer. nat., Chemistry, expected 2023  
*10.2021-03.2023* **Maternity leave**
- 09.2010 – 06.2016 **Russian National Research Medical University, (Moscow, Russia)**  
Diploma in Medical Biophysics (Hons.)

#### Research experience

- 09.2016-present **Leibniz Universität Hannover, BMWZ (Hannover, Germany)**  
Doctoral research  
Advisor: J. Prof. Dr. A. Marchanka,  
Thesis: Structural studies of 5' UTR of hepatitis c viral RNA by NMR-based structural biology
- 02.2016 – 07.2016 **Research Centre of Medical Genetics, Laboratory of Inherited Metabolic Disorders (Moscow, Russia)**  
Diploma project  
Advisor: Leading Researcher G. Baydakova  
Thesis: Improvement of the diagnostic criteria for biophysical methods in biotin dependency diagnosis

#### Conferences

- 03.2019 **Advanced isotope labeling methods for integrated structural biology (Grenoble, France)**  
Selected talk and Poster: Structural studies of the 5'-UTR of HCV-RNS by solid-state NMR-based structural biology
- 02.2018 **18<sup>th</sup> Leibniz Symposium (Hannover, Germany)**  
Poster: Insights into the structural and functional features of miR-122 – HCV RNAs interaction

#### Training and Workshops

- 29.03.2019 – 05.04.2019 **Institut de Biologie Structurale, Grenoble, France**  
Advanced isotope labeling methods for biomolecular NMR

#### Languages

- |         |                    |
|---------|--------------------|
| English | Advanced           |
| Spanish | Upper intermediate |
| German  | Intermediate       |
| Russian | Native             |

