TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Biomolekulare NMR-Spektroskopie,
Department Chemie

Protein and small molecule recognition by native and non-canonical nucleic acids

Alexander Beribisky

TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Biomolekulare NMR-Spektroskopie,
Department Chemie

Protein and small molecule recognition by native and non-canonical nucleic acids

Alexander Beribisky

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Bernd Reif

Prüfer der Dissertation: 1.Univ.-Prof. Dr. Michael Sattler
2. Priv.-Doz. Dr. Dierk Niessing

Die Dissertation wurde am 26.06.2014 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 23.07.2014 angenommen.

Table of Contents

Summary

In this thesis, the molecular recognition of non-canonical structural features in RNA by small molecules and proteins has been investigated focusing on two systems. The first system is an RNA/protein complex which serves as a junction point between two important cellular processes – RNA editing and RNA interference. The second is a RNA/small molecule interaction which modulates the expression of the protein thymidylate synthase.

The protein p100, a component of the RNA-induced silencing complex, is involved in binding and cleaving inosine-edited RNA duplexes, thereby excluding them from downstream events in RNA interference. This function is performed by p100's tandem staphylococcal nuclease domains which were shown to be implicated in edited RNA binding and cleavage. However, many key aspects of this interaction remain unclear. What is the structural basis for the discrimination between edited and unedited RNA sequences? Does p100 possess RNA binding ability alone or also RNAse activity? If so, which structural elements modulate this activity? To address these questions, we set out to study the three-dimensional structure of the p100's inosine-edited RNA substrate of as well as study the structural changes that occur in p100 upon RNA binding.

Biochemical and biophysical characterization of the p100 RNA substrate, shows that the presence of inosine-uracil base-pairs causes a distortion in its helical geometry. This distortion destabilizes the RNA dimer structure and facilitates p100 recognition. Multidimensional nuclear magnetic resonance spectroscopy experiments were used to conduct a backbone assignment of the p100 construct encompassing staphylococcal nuclease domains 3 and 4 up to 90%. Comparison of NMR spectra of individual and dual domain constructs demonstrate an interaction between these two domains, which contributes to the overall dual domain fold. Inosine-edited RNA/p100 binding studies indicate that the interaction is taking place in loop regions of both domains, which rigidifies these local structural elements in the p100 protein. Finally, p100 possesses RNAse activity however, its mechanism seems to differ from that of a classic staphylococcal nuclease

Taken together, these findings show that inosine-uracil tandems can severely destabilize RNA helical structure, even though inosine possesses significant structural similarity to two other canonical two-ringed nucleotides, guanine and adenine. Also of importance is the role p100 loops in RNA substrate recognition.

Thymidylate synthase overexpression has been implicated as one of the factors leading to the development of resistance in anti-cancer therapy. Therefore, repressing thymidylate synthase translation by targeting its own messenger RNA could help to overcome this problem. Here, binding of the small molecule Hoechst 33258 to a CC-mismatch containing RNA construct derived from the thymidylate synthase messenger RNA, was investigated. The interaction predominantly involves the intercalation of Hoechst 33258 at a CC mismatch region of the RNA. Additional work shows that Hoechst 33258 reduces thymidylate synthase protein but not messenger RNA levels in cells. The results presented here may serve as a blueprint for future endeavours in designing thymidylate synthase-targeting pharmacophores.

A number of conclusions can be drawn from the study of the two aforementioned systems. First, a seemingly conservative structural substitution in the RNA sequence, can bring about significant changes in its structure, its ability to interact with its potential binding partners and consequentially, its function. Second, disordered protein regions can play a role in ligand recognition. Finally, non-canonical regions in nucleic acids play a central role in the recognition of their binding partners, making them indispensable in the modulation of various biological processes.

Zusammenfassung

In dieser Arbeit wurde die molekulare Erkennung nicht-kanonischer, struktureller Einheiten von RNA durch kleine Moleküle und Proteine anhand von zwei Systemen untersucht. Das erste System befasst sich mit einem RNA/Protein-Komplex, welcher zwei wichtige zelluläre Prozesse, die RNA-Editierung und RNA-Interferenz, verknüpft. Im zweiten System wurde die Interaktion von einer RNA mit kleinen Molekülen untersucht, welche die Expression des Proteins Thymidylat-Synthase moduliert.

Das Protein p100, eine Komponente des RNA-induced silencing complex, ist an der Bindung und Spaltung Inosin-editierter RNA-Duplexe beteiligt, wobei es sie von weiteren Vorgängen des RNA-Interferenz Signalweges ausschließt. Die Tandem staphylokokkalen Nuklease Domänen von p100 sind für diese Funktion verantwortlich. Es konnte gezeigt werden, dass diese Domänen an der Bindung und Spaltung editierter RNA miteinbezogen sind. Dennoch konnten einige Schwerpunkte dieser Interaktion bisher nicht aufgeklärt werden. Was ist die strukturelle Basis, aufgrund welcher zwischen editierten und uneditierten RNA-Sequenzen unterschieden wird? Besitzt p100 neben der Fähigkeit RNA zu binden auch RNAse-Aktivität? Falls ja, welche strukturellen Elemente modulieren diese Aktivität? Um diese Fragestellungen anzugehen haben wir die 3-dimensionale Struktur des Inosin-edierten RNA-Substrates von p100 untersucht, sowie die durch RNA-Bindung verursachten strukturellen Veränderungen in p100 erforscht.

Biochemische und biophysikalische Charakterisierung des RNA-Substrates von p100 haben gezeigt, dass die Gegenwart eines Inosin-Uracil-Basenpaars eine Verformung in der helikalen Geometrie des Substrates verursacht. Diese Verformung destabilisiert die RNA-Dimer Struktur und erleichtert die Erkennung von p100.

Multidimensionale Kernspinresonanz-Spektroskopie Experimente wurden eingesetzt um bis zu 90% des Proteinrückgrats des p100 Konstrukts zuzuordnen, welches die staphylokokkalen Nuklease Domänen 3 und 4 umfasst. Ein Vergleich der NMR-Spektren der einzelnen Proteindomänen mit den Tandem-Proteindomänen zeigte, dass beide Proteindomänen miteinander interagieren und somit zur globalen Faltung der Tandem-Proteindomänen beitragen. Inosin-editierte RNA/p100 Bindungsstudien zeigten weiterhin, dass die Interaktion beider Domänen jeweils in den Schlaufenbereichen stattfindet. Die Interaktion beider Domänen führt zudem zu einer lokalen Versteifung von diesen strukturellen Einheiten des Proteins p100.

Die Ergebnisse zusammenfassend scheint sich der Mechanismus der p100 RNAse-Aktivität von dem einer klassichen Staphylokokkalenl Nuclease zu unterscheiden. Zudem zeigt sich, dass Inosin-Uracil-Tandems helikale Strukturen von RNA beträchtlich destabilisieren können, obwohl Inosin eine hohe strukturelle Ähnlichkeit mit den beiden anderen kanonischen Nukleotiden, Guanin und Adenin, besitzt. Weiterhin spielen die Schlaufenbereiche des Proteins p100 eine wichtige Rolle in der RNA-Substrat-Erkennung.

Es wird angenommen, dass die Überexpression der Thymidylat-Synthase eine Rolle bei der Resistenzentwicklung bei Krebsbehandlungen spielt. Daher sollte eine Unterdrückung der Thymidylat-Synthase-Translation durch Hemmung seiner Boten-RNA zur Lösung dieses Problems beitragen. In der vorliegenden Arbeit wurde die Bindung des kleinen Moleküls Hoechst 33258 an ein RNA-Konstrukt untersucht, welches von der Thymidylat-Synthase Boten-RNA abgeleitet wurde, und eine CC-Fehlpaarung enthält. Die Interaktion wird überwiegend durch eine Interkalation von Hoechst 33258 an einer CC-Fehlpaarung nahe der Translationsstartstelle bedingt. Anhand zusätzlicher Daten konnte gezeigt werden, dass Hoechst 33258 in der Lage ist den Pegel des Thymidylat-Synthase-Proteins, aber nicht der Boten-RNA in Zellen zu reduzieren. Die hier gezeigten Ergebnisse könnten wegweisend sein für die Entwicklung Thymidylat-Synthase-gezielter Pharmakophore.

Zusammenfassend konnten einige Erkenntnisse aus der Untersuchung beider Systeme gewonnen werden. Erstens können auch scheinbar minimale strukturelle Veränderungen in der RNA-Sequenz zu großen Veränderungen der RNA-Struktur führen. Sie beeinflussen hierdurch ihre Fähigkeit zur Wechselwirkung mit potentiellen Bindungspartners und ihre Funktion. Zweitens können auch unstrukturierte Bereiche in Proteinen wichtig für die Erkennung von Liganden sein. Zum Schluss konnte gezeigt werden, dass nicht-kanonische Regionen in Nukleinsäuren eine zentrale Rolle in der Erkennung ihrer Bindungspartner spielen, was sie unverzichtbar in der Regulierung biologischer Prozesse macht.

Chapter 1 – Introduction to the main research themes of this work

1.1 Introduction to RNA Structure

Ribonucleic Acid (RNA) is a polymer, which consists of repeating monomers called nucleotides (Bloomfield, Crothers et al. 2000). The nucleotide monomers are composed of a D-ribose sugar (Figure 1.01A), a phosphate group and a nitrogenous aromatic base attached to the D-ribose via its C1' anomeric carbon. The nitrogenous bases are most commonly either one or two-ringed and are called pyrimidines and purines, respectively. Among pyrimidines, the most common bases are cytosine (Figure 1.01B) and uracil (Figure 1.01C) while adenine (Figure 1.01D) and guanine (Figure 1.01E) are most common among purines (Bloomfield, Crothers et al. 2000).

(A)   **(B)**   **(C)**   **(D)**   **(E)**
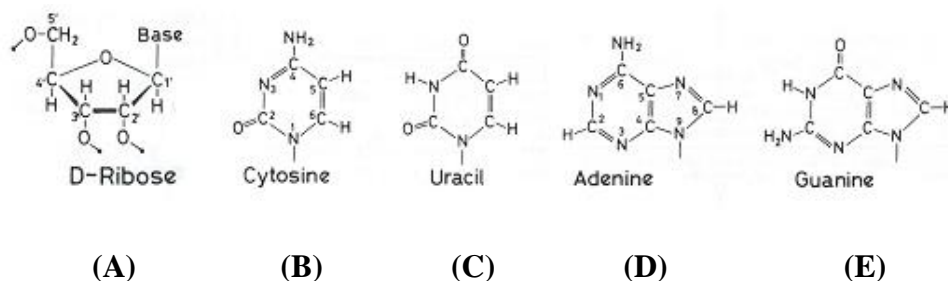
Figure 1.01 – Components of nucleotide structure: **(A)** D-ribose sugar; either one of the nitrogenous bases **(B)** Cytosine; **(C)** Uracil; **(D)** Adenine; **(E)** Guanine and a phosphate group. Adapted from: (Wüthrich 1986)

The nitrogenous bases can hydrogen-bond to one another, creating so-called Watson-Crick base-pairs. Adenine usually base-pairs with uracil forming two hydrogen bonds (Figure 1.02A), while guanine normally base-pairs with cytosine (Figure 1.02B), forming three (Bloomfield, Crothers et al. 2000). Non-standard base pairing such as G-U, A-G and C-C may also occur (Bloomfield, Crothers et al. 2000). Base-pairing allows the RNA to form secondary helical structures that most commonly adopt an A-form helix (Bloomfield, Crothers et al. 2000).

(A)   **(B)**
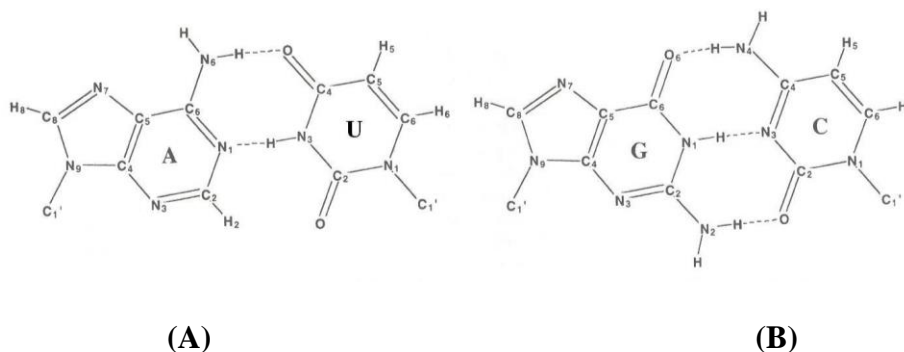
Figure 1.02 – Two common base-pairs in nucleic acids: (A) Adenine-Uracil and (B) Guanine-Cytosine base-pairs are depicted. (Bloomfield, Crothers et al. 2000).

The RNA ribose sugar can adopt two major sugar pucker conformations (Bloomfield, Crothers et al. 2000). In the first, the C2' carbon is coplanar with its C5'counterpart. This conformation is called *C2' endo* (Figure 1.03A). In the other, C3' is coplanar with the C5', and this is called *C3' endo* conformation (Figure 1.03B). The *C3' endo* conformation can be found in helical portions of RNA while *C2' endo* may be present in various non-helical portions of the molecule such as loops, mismatches or bulges (Bloomfield, Crothers et al. 2000).

(A)   **(B)**
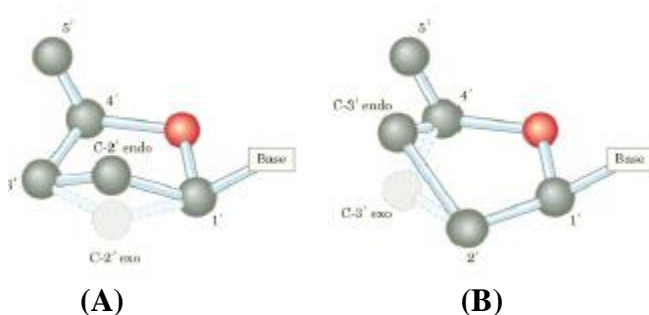
Figure 1.03 – The two types sugar pucker conformations are displayed: (A) *C2' endo* and (B) *C3' endo*. (Lehninger, Nelson et al. 2000).

The nitrogenous base can adopt one of two conformations around the so-called glycosidic angle: One in which the base is situated directly on top of the ribose sugar which is called *syn* (Figure 1.04A) and another where the base points away from the ribose sugar which is called *anti* (Figure 1.04B). *Anti* conformation is prevalent in helical/base-paired regions of the RNA while *syn* is more commonly found in RNA loops (Bloomfield, Crothers et al. 2000).

(A)   **(B)**
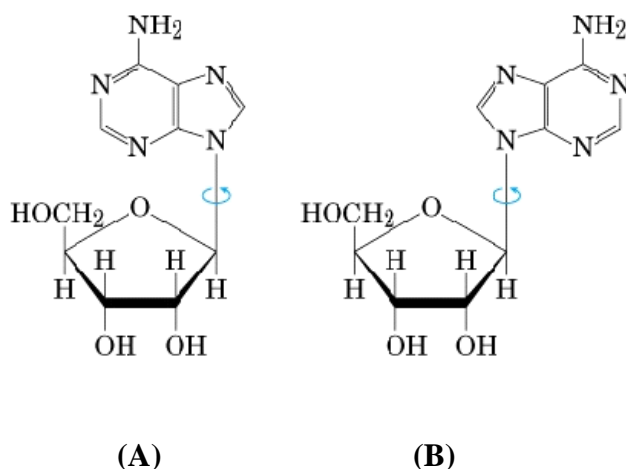
Figure 1.04 – Base Conformations (A) *Syn*; (B) *Anti.* (Lehninger, Nelson et al. 2000).

RNA properties allow it to fold into large variety of secondary and tertiary structures such as bulges, hairpins, loops which play crucial roles in many biological processes such as gene expression, RNA interference and RNA editing (Liu, Reches et al. 1998, Scadden and Smith 2001, Bagasra and Prilliman 2004, Singh and Valcarcel 2005). In addition to playing a structural role in these processes, RNA can also act as a catalyst via its C2' hydroxyl moiety (Bloomfield, Crothers et al. 2000). RNA also serves as a genome for various viruses and plays a crucial role in their life cycles (Weik, Modrof et al. 2002). It is due to this functional diversity, that RNA is hypothesized to be the so-called "first form of life" according to the "RNA world" theory (Bloomfield, Crothers et al. 2000) as it is capable of carrying genetic information as well as reproducing it. Over the course of evolution, however, the more structurally versatile proteins took over the RNA's catalytic functions, while a more stable form of nucleic acid, DNA became the main form of genetic information storage (Bloomfield, Crothers et al. 2000).

As RNA is the centerpiece of so many cellular processes, the study of its structure and function, as well as interactions with its various binding partners is of central importance. The field that deals with such studies is called structural biology. Structural biology is defined as a branch of molecular biology, biochemistry, and biophysics which concerns itself with the molecular structure of biological macromolecules, especially proteins and nucleic acids, how they acquire the structures they have, and how alterations in their structures affect their function. In order to probe biomolecular structure, various methods can be utilized such as mass spectrometry, cryo-electron microscopy and X-Ray crystallography. However, for the purposes of this work, the primary method used for structural investigation, will be Nuclear Magnetic Resonance (NMR) spectroscopy.

1.2 Introduction to Nuclear Magnetic Resonance spectroscopy

Developed in the late 1940's and applied mostly to smaller compounds upon its inception, NMR has become a powerful tool not only in organic chemistry but also in studies of biomolecular structure. In addition to structure determination, NMR can also provide information regarding the dynamics of the molecule. This section will provide an introduction into the basic principles of NMR spectroscopy and its applications in structural biology of proteins and RNA.

1.2.1 Basic principles of NMR

1.2.1.1 Nuclear spin

The basic phenomenon explored in NMR spectroscopy is the nuclear spin. This property is described by the quantum number I (Becker 2000). The nuclear spin varies from atom to atom depending on the number of protons and neutrons in its nucleus by a multiple of 0.5 (Sanders and Hunter 1993). For instance hydrogen, (^{1}H) has I = 0.5 (spin one-half) due to the presence of only one proton in its nucleus. On the other hand, ^{12}C would have I = 0 due to the even number of protons and neutrons (six each). The ^{13}C atom however, has an extra neutron in its nucleus, compared to its more abundant ^{12}C counterpart and hence, like ^{1}H, has I = 0.5. Additional atoms which possess a spin one-half include ^{15}N and ^{31}P. Finally atoms such as ^{2}H and ^{14}N have a nuclear spin of one, as their nuclei contain an odd number of neutrons and protons.

In their natural state, nuclear spins are randomized and their motion along a z-axis can be described by the following formula

$$\mu_z = \gamma I_z = \gamma \hbar m$$

Where μ is the magnetic moment, γ is the nucleus specific gyromagnetic ratio (ratio of the nucleus's magnetic dipole moment to its angular momentum), $\hbar$ is the Planck constant, and m is the so-called magnetic quantum, whose value can vary from –I to +I by integer steps. Hence, nuclei with I = 0.5, have two possible nuclear spin states. In the first state, where m = 0.5, is referred to as "spin-up" or the α state. The other state where m = - 0.5 is referred to as "spin-down" or the β state (Keeler 2002). These states have the same energy and are hence in a thermal equilibrium.

However the spin behaviour changes when nuclei are exposed to an external magnetic field B_0, the effective frequency of which ($\omega_{spectrometer}$) is dependent on the nucleus's gyromagnetic ratio. In its presence, the α spins rotate (or precess) along this field (or the z-axis) assuming the lower energy state, while the β spins orient themselves against B_0 assuming a higher energy state (Keeler

2002). The energy difference between these states, ΔE can be described as the product of $\omega_{spectrometer}$ and γ, with a small population bias towards the lower energy, α state (Figure 1.05).

$$\Delta E = \gamma \, \omega_{spectrometer} = -\gamma \, \hbar B_0$$
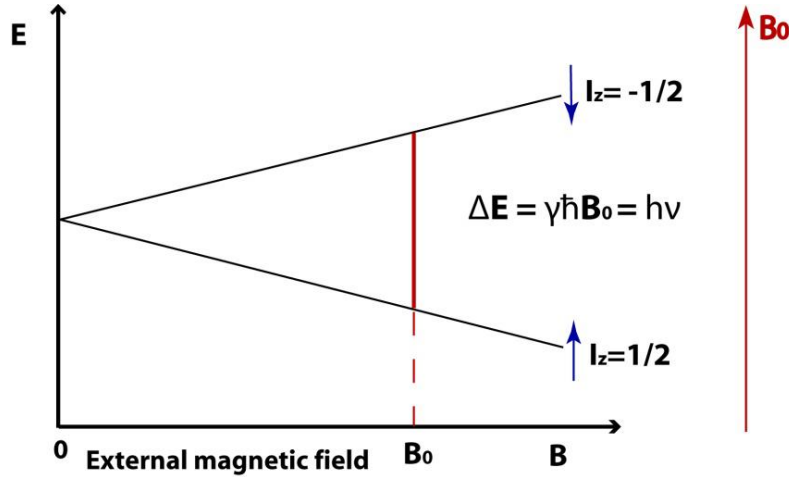
Figure 1.05 – The difference in energy between the two spin states, when an external magnetic field, B_0 is present.

In order to manipulate these spin systems, an electromagnetic radiation with the frequency to match the energy difference between the nuclear spin levels (the so-called Larmor frequency) in a constant magnetic field of the appropriate strength is applied (Keeler 2002). Its value is in the radio-frequency (RF) range. The absorption of resonant spin is detected as an induced electric current in the spectrometer's coils. These coils are aligned with the xy plane (that is perpendicular to B_0 which along in the z-axis) (Keeler 2002) and detect the x-component of the nuclear magnetization M_x, defined by:

$$M_x = M_0 \sin\beta$$

where M_0 is the magnetization in the z-plane, and β the angle with which this magnetization is tilted towards the x-plane. They y-component can be worked out by determining the rotation angle of the magnetization, ε, given by the rate of rotation of the magnetization vector ω_{rf} after a time period t (Keeler 2002):

$$\varepsilon = \omega_{rf} t$$

1.2.1.2 Relaxation

As mentioned, at B_0, the spins exist in an energetic equilibrium. When various pulses are applied, this equilibrium is perturbed. The time it takes for the spins to return their original state is referred to as relaxation.

When the spins are shifted away from their equilibrium along the longitudinal z-axis, they will eventually return to their original equilibrium population. The time it takes for 63% of these spins to do so, by exchanging energy with their surroundings, is called spin-lattice relaxation time, T_1 (Keeler 2002, Levitt 2008), defined by

$$M_z(t) = M_{z,eq} \left(1 - e^{-t/T_1} \right)$$

Where $M_z(t)$ is the magnetization as a function of time t and $M_{z,eq}$ the magnetization at equilibrium.

Another type of relaxation involves the dephasing of spins in the transverse xy plane(M_{xy}). The time it takes for 37% of the spins to lose coherence on this plane, is referred to as spin-spin relaxation time or T_2.

$$M_{xy} = M_{xy}(0)e^{-t/T_2}$$

Both T_1 and T_2 relaxation are related to molecular tumbling defined by the correlation time, τ_c. As τ_c is related to the molecular weight through its hydrodynamic radius, consequentially, so are T_1 and T_2. For instance, small molecules which tumble quickly have both high T_1 and T_2 values (Figure 1.06). Larger molecules such as proteins and polymers which have a slower tumbling rate have a high T_1 but have a low T_2 value.
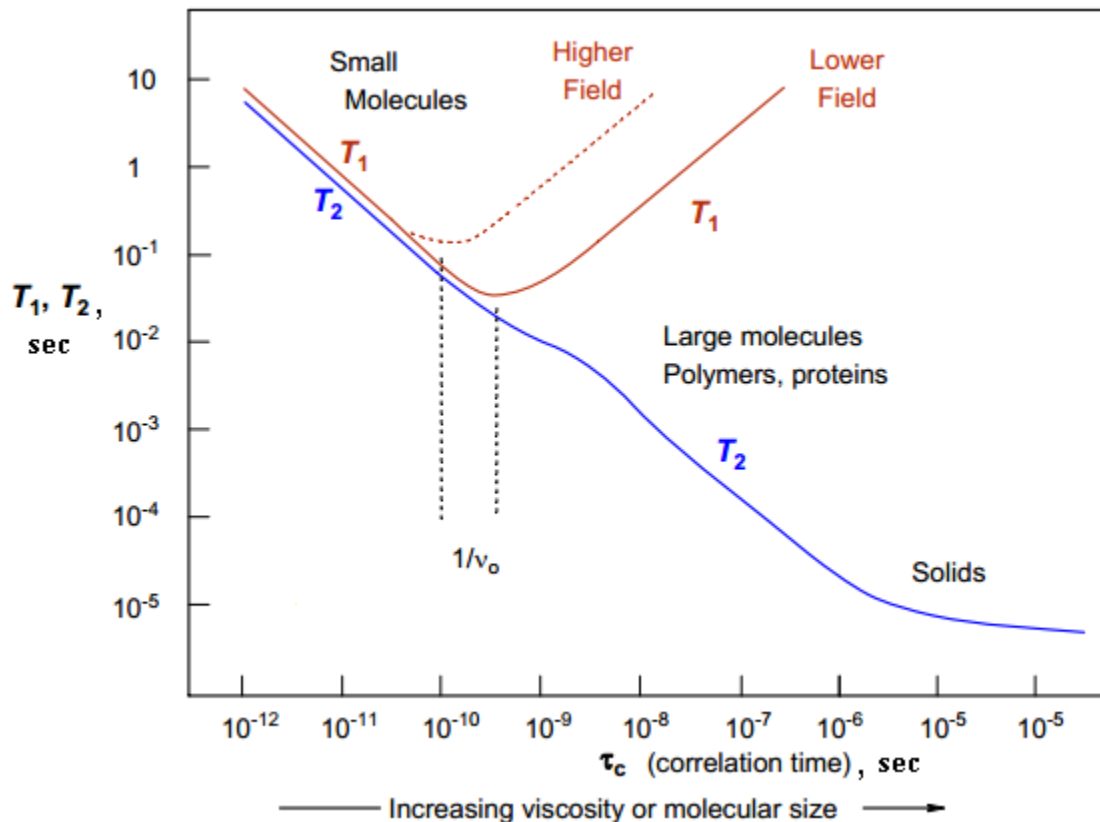
Figure 1.06 – T_1 (in red) and T_2 (in blue) values as a function of τ_c, both in seconds

The magnitude of T_2 is also inversely proportional to half the spectral linewidth, $\nu_{1/2}$ (James 1998, Levitt 2008):

$$\nu_{1/2} = 1/(\pi T_2)$$

As such, smaller molecules with large T_2 values, would give sharper lines in the NMR spectrum compared to proteins whose shorter T_2 would result in broader signals.

1.2.1.3 The chemical shift

The resonance frequency of a given nucleus is also influenced by its surrounding electrons. These electrons rotate, producing a magnetic field, which normally reduces the effective nuclear magnetic field and as a result, its resonant frequency (Martin and Zektzer 1988). This effect is referred to as shielding (Figure 1.07).
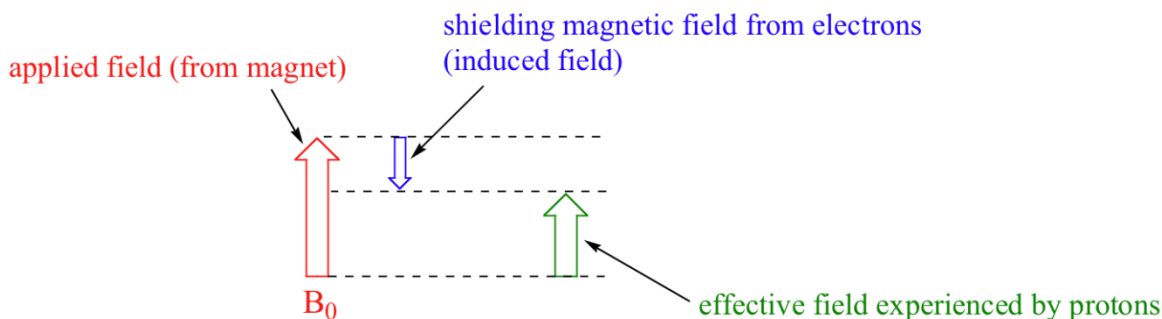
Figure 1.07 – Effective magnetic field experienced by protons as a result of shielding from electrons (Wade 1999).

Each given nucleus and its surrounding electrons are located in their own unique chemical environment. This environment is determined by factors such as the electronegativity of the atoms located close to the atom in question, the presence of electron withdrawing or electron donating groups, anisotropic magnetic fields (Martin and Zektzer 1988). Affected by these factors, the electrons remain close to or move further away from their nucleus. This, in turn, changes the nucleus' resonance frequency. The resulting phenomenon, the chemical shift, δ, defined as:

$$\delta_{\text{,in parts per million (ppm)}} = [(\omega_{atom} - \omega_{reference})/ (\omega_{spectrometer})]\ 10^6$$

(where ω_{atom} is the resonant frequency of a given atom, $\omega_{reference}$ is the resonant frequency of a reference compound, normally tetramethylsilane (Wade 1999), and $\omega_{spectrometer}$ is the operating frequency of the spectrometer), produces a variety of resonant frequencies for different nuclei.

For example, if a certain hydrogen is located in close proximity to an atom which has high electronegativity, such as oxygen or nitrogen, this atom would pull the hydrogen's electron away from its nucleus. A similar effect would occur if an electron withdrawing group such as a nitro group, or a carbon-carbon double bond were present close to the aforementioned hydrogen atom. As a result, the effective resonance frequency of this atom in these three cases would be higher, and as a consequence, so would its chemical shift.

As such, in the proton (^{1}H) chemical shift scale, hydrogens located next to polar functional groups, such carbonyl, carboxylic acids, as well as close to carbon-carbon double bonds, have a larger chemical shift value than hydrogen atoms located in more hydrophobic regions of the molecule such as methyl groups (Figure 1.08A). For instance in the one-dimensional ^{1}H NMR

17

spectrum of 1,2,2-trichloropropane (Figure 1.08B), the signals from the molecule's methylene resonate downfield than their methyl counterpart, due to the closer proximity of the methylene group to the more electronegative chlorine atom.

(A)   (B)
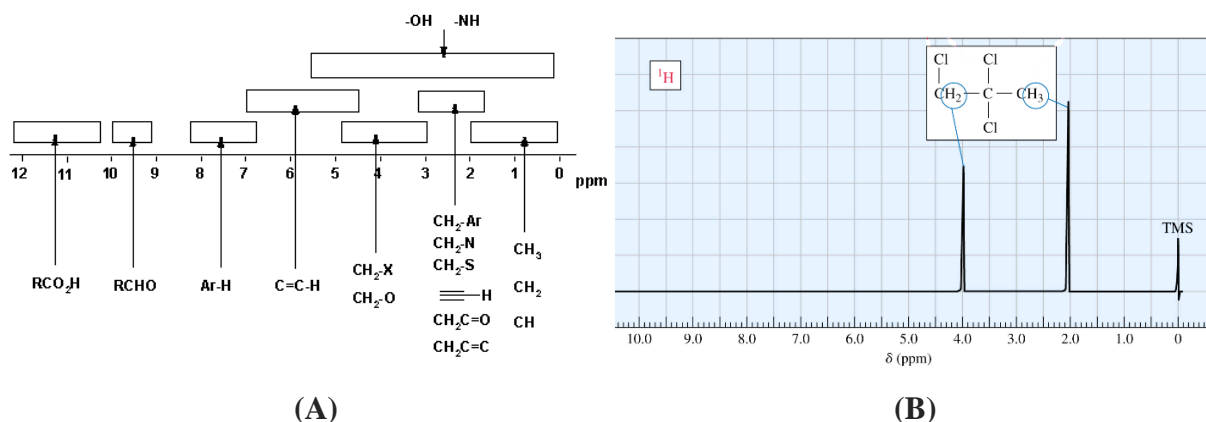
Figure 1.08 – **(A)** Chemical shift scale for ^{1}H in ppm, showing the chemical shift values for protons located in various chemical environments.(B) ^{1}H NMR spectrum of 1,2,2-trichloropropane. Adapted from (Wade 1999)

1.2.1.4 Fourier Transform NMR

The first spectrometers used to sweep the entire frequency range of the spins. As NMR is intrinsically insensitive technique (due to the small difference in the α and β spin populations), the signal quality from these experiments was poor. In addition since one has to spend one second per each Hertz (Hz) in order to obtain a resolution of 1 Hz, this made the measurements very long and yielded poor resolution (Sanders and Hunter 1993).

However, in the 1970's a novel development in NMR data acquisition has emerged – Fourier Transform NMR (FT-NMR).When using FT-NMR, instead of exciting one frequency at a time, a short pulse in the microsecond or millisecond scale is used to excite a whole range of spins in the molecule. The resulting magnetization vector being away from its equilibrium oscillates (Sanders and Hunter 1993), giving rise to a so-called free induction decay (FID), a time dependent signal pattern which is picked up by the spectrometer coils. This FID contains the sum of all the excited spins. This information can be extracted by a Fourier Transform (specifically, via the integration of the time domain) to give rise to the frequency domain and consequentially, the entire NMR spectrum.

The advantages of FT-NMR over the older continuous wave spectrometers are numerous. First, as only one pulse is required to excite the entire NMR spectrum, scan times were drastically

18

reduced. As Fourier Transform allows for addition of FIDs, multiple scans can be performed and combined to improve the signal-to-noise ratio. Finally, manipulation of spins via pulses and delays has opened the door for a larger variety of experiments, both single and multi-dimensional (Sanders and Hunter 1993).

1.2.2 Multidimensional NMR spectroscopy

A one-dimensional NMR (1D NMR) experiment contains just one time domain, t_1 (Figure 1.09A). The information from this time domain is recorded after an RF pulse is applied during the so-called preparation time to tilt the magnetization into the detectable xy plane ("labelling" it with the spins' chemical shift) and is then allowed to go back to its equilibrium state. The time domain is later converted into a frequency domain by a Fourier Transform.

A multi-dimensional NMR spectrum on the other hand contains two or more time domains. The first time domain is recorded just as in the one-dimensional experiment, with RF pulses applied during the preparation time, followed by a period during which the magnetization goes back to its equilibrium state. Instead of immediately detecting this magnetization however, the sample is subjected to another series of pulses (Figure 1.09B) during a so-called mixing time (Keeler 2002). The excited spins go back to their equilibrium states again, during the second time domain, t_2. In case of a two-dimensional NMR (2D NMR) experiment, the FID from this time domain is detected, while if a three-dimensional NMR (3D NMR) experiment is acquired, there is another mixing period and a third domain, t_3, which is then recorded (Figure 1.09C). The experiment is then repeated with increasing values of t_1 (starting from $t_1 = 0$) using a fixed sampling interval (Keeler 2002) and is then detected as a function of t_2 as before (t_3, in case of a three-dimensional experiment). Both time domains are then Fourier transformed to give rise to two (in case of 3D NMR, three) frequency domains: F1 and F2.

As the magnetization during the mixing time is transferred from one group nuclei to another, signals in a multi-dimensional NMR spectrum are correlated based on a certain property, depending which specific kind of pulses were used during the given experiment (Keeler 2002). This correlation can be through bond or through space, and depending on which nuclei were subjected to these pulses, can occur either between the same or different kinds of nuclei. When the

experiment involves same type of nuclei, it is referred to as a homonuclear experiment, while where more than one type of nucleus are involved, the experiment is termed heteronuclear.
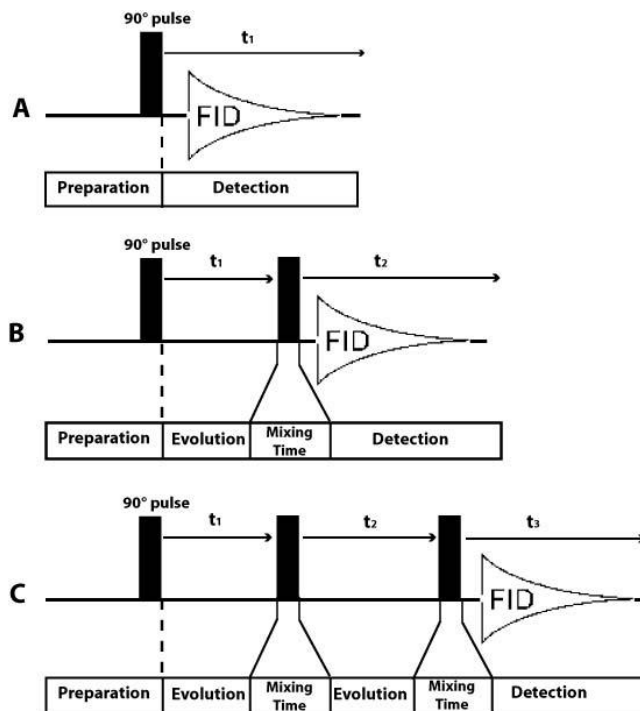
Figure 1.09 – A schematic representation of various NMR experiments. **(A)** A 1D NMR experiments contains just a preparation stage and following the pulse a single time domain is recorded. **(B)** In 2D NMR another series of pulses during the mixing time results in another time domain, which is then detected. **(C)** A 3D NMR experiments has an additional series of pulses which gives rise to a third time domain (Keeler 2002).

Multidimensional NMR experiment offers clear advantages in the area of spectral resolution. A 1D NMR spectrum especially that of a large molecule such as a protein or a nucleic acid, gives rise to a multitude of signals. Coupled with the intrinsically broad peaks produced by larger molecules due to their low T_2 values (1.2.1.2), this results in a severe spectral overlap (Figure 1.10A) making comprehensive spectral assignment of anything but the smallest peptides impossible using 1D NMR techniques alone. In 2D NMR, the presence of signals from two spins (called cross-peaks), from two frequency domains correlated to each other by various properties, vastly improves spectral resolution, thereby simplifying the task of spectral assignment (Figure 1.10B). 3D NMR provides an even superior resolution over its 2D counterpart due to the presence of a third frequency dimension. Different two and three-dimensional homonuclear and heteronuclear experiments provide complementary information not only regarding the chemical

shifts of the molecule's atoms, but also on their proximity to other atoms as well as on the molecule's secondary structure elements (Wüthrich 1986, Cavanagh 2007).

(A)   **(B)**
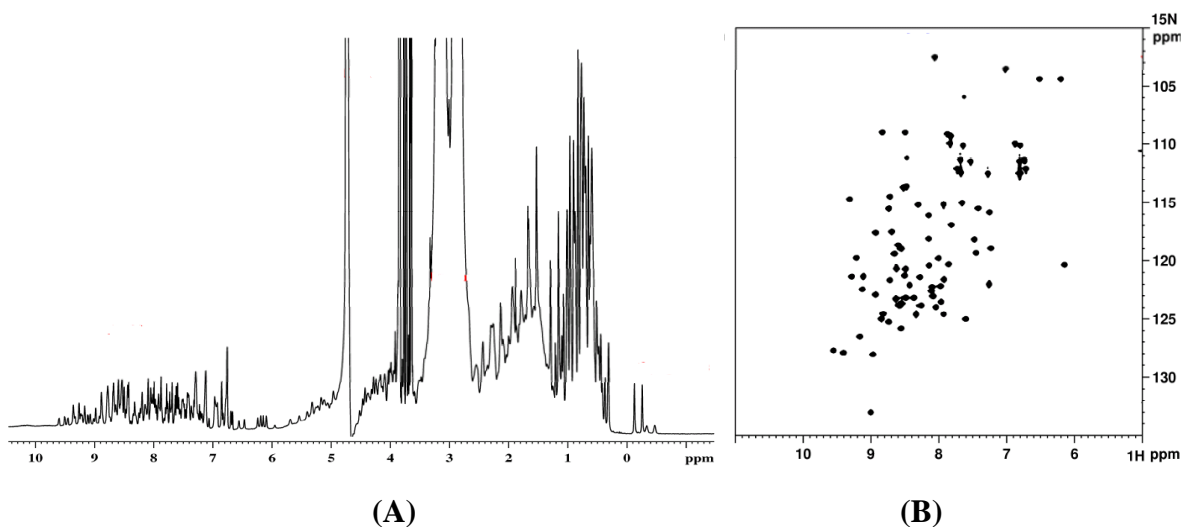
Figure 1.10 – Improvement of spectral resolution using 2D NMR as seen on a sample of ubiquitin. (A) A 1D NMR spectrum of this protein. A very large number of signals from different parts of the protein gives rise to a severe spectral overlap, making assignment impossible (B) A 2D spectrum on the other, give rise to distinct, clearly cross-peaks between various atoms in the molecule. (Cavanagh 2007)

1.2.2.1 Through-bond correlation in multi-dimensional NMR

Multi-dimensional experiments correlating various spins through their bonds, do so using a property called J-coupling. J-coupling is a through-bond interaction between neighboring atoms mediated by their nuclei and surrounding electrons. This interaction can extend from one to three bonds and can be quantified in Hertz. Therefore, upon excitation of a given nucleus, this excitation can be transferred to its neighbor. Consequentially this pair of atoms can be determined to be correlated through bonds in the NMR spectrum. Additional information such as bond distance and angles can also be obtained (Sanders and Hunter 1993).

One such experiment is the two-dimensional Heteronuclear Single Quantum Coherence (HSQC), correlates a hydrogen atom to its directly bound carbon or nitrogen (also called heteroatoms). In the t_1 stage of this experiment the hydrogen spin acquires the chemical shift of its heteroatom, while in the second, t_2 stage it will evolve its own chemical shift. The resulting magnetization is detected and processed, giving rise in the given example to proton-carbon cross-

21

peaks in a specific region of a 19mer-RNA (Figure 1.11A), or proton-nitrogen connectivities in a protein (Figure 1.11B).

(A)   **(B)**
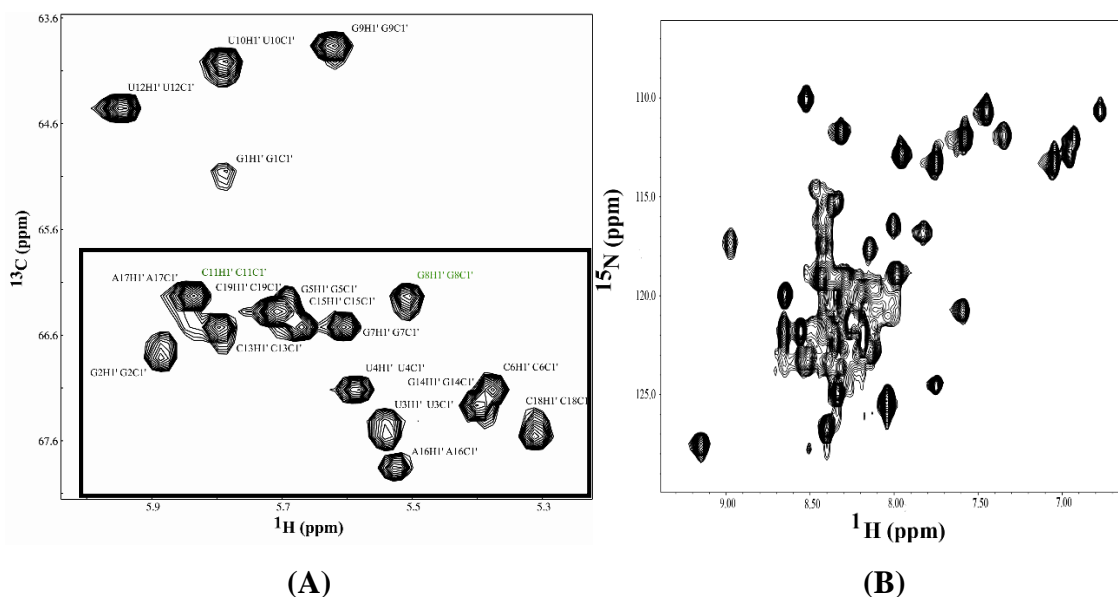
Figure 1.11 – (A) The anomeric hydrogen-carbon cross-peaks in a 2D ^{1}H, ^{13}C HSQC experiment of a 19mer RNA construct (Beribisky, Tavares et al. 2007). (B) A 2D ^{1}H, ^{15}N HSQC of a 15 kDa protein dimer, showing hydrogen-nitrogen cross-peaks for the backbone and side-chain for of the protein. Note: The residues downfield in the carbon dimension are marked with a box.

The 2D ^{1}H, ^{15}N HSQC experiment is of central importance in protein NMR. As this experiment gives rise to proton-nitrogen cross-peaks, one can in principle observe an amide connectivity for the backbone of every non-proline amino acid residue (in addition to the N-H connectivities for the amino acid side chains). The presence of too few or too many cross-peaks in this spectrum might indicate that the protein is not properly folded, or may be degrading, respectively (Cavanagh 2007). Therefore the HSQC serves as a so-called protein "fingerprint spectrum" and is a good starting point for measurements on protein sample (i.e. to test its viability) before proceeding to more complex and time consuming experiments.

Additional through-bond experiments of importance in biomolecular NMR are the homonuclear (although heteronuclear variants also exist) Correlation Spectroscopy (COSY) and Total Correlation Spectroscopy (TOCSY) experiments. The former contains of two 90-degree pulses, after which t_2 modulated by the chemical shift of t_1, is recorded, processed and the resulting spectrum gives rise to cross-peaks between two atoms correlated to each other through three bonds or less via their J-coupling (Keeler 2002). The later experiment is similar to the COSY, however,

here cross-peaks are observed not only between nuclei which are directly coupled, but also between two nuclei coupled to a third nuclei (Figure 1.12), but not to each other, creating so-called a "spin system" (Keeler 2002). This is achieved by using composite spin-lock pulses, causing the magnetization to spread across a larger number of bonds (Inagaki, Shimada et al. 1989).
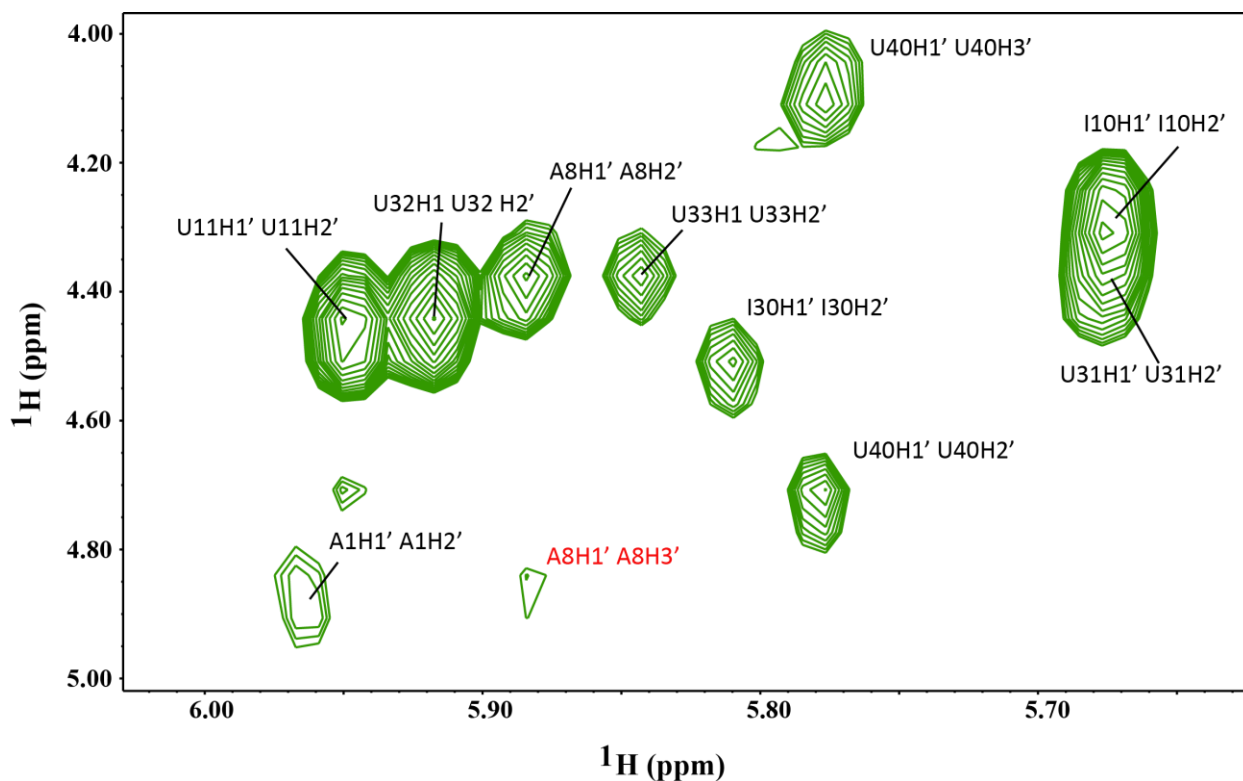
Figure 1.12 – A 2D ^{1}H, ^{1}H TOCSY of a 20mer RNA duplex. A. Cross-peaks between hydrogen atoms which are three bonds apart are marked in black, while the signal which correlates atoms further removed from one another, is marked in red.

Before three-dimensional NMR experiments have become more widespread, COSY and TOCSY experiments were used for assignment of the protein backbone and measuring torsion angles (Cavanagh 2007). In nucleic acids, the homonuclear TOCSY is still used to verify the positions of certain resonances as well as determine the geometry of the ribose/deoxyribose sugar (Varani, Aboul-ela et al. 1996).

Three-dimensional through-bond NMR experiments are essential for complete structural characterization of biological molecules larger than a few kilodaltons. Experiments such as the HNCA, HNCO, and HNCACB, HN(CO)CACB, are used to assign the protein backbone. HCCH-TOCSY, and HCCH-COSY are employed in protein side chain and nucleic acid sugar assignment

(Wüthrich 1986, Fürtig, Richter et al. 2003, Cavanagh 2007). A more detailed description of spectral assignment of biological molecules, in particular nucleic acids, will be provided later (1.2.3).

1.2.2.2 Through-space correlation in multi-dimensional NMR

Through-space information in multi-dimensional NMR is obtained by exploiting the phenomenon of polarized spins transferring their polarization to neighboring spins via cross-relaxation (Williamson 2009), through a dipole-dipole interaction. This transfer is called the Nuclear Overhouser Effect (NOE). The NOE intensity between spins I and S, $\sigma_{IS,}$ is approximated by the following formula

$$\sigma_{IS} \ \alpha \ \upsilon^{-6} \ \tau_c \ ((6/[1 + [\omega_I + \omega_S]^2 \tau_c^2]) - (1/[1 + [\omega_I - \omega_S]^2 \tau_c^2]))$$

Where υ is the internuclear distance between I and S, ω_I and ω_S are the frequencies of the two spins and τ_c is the correlation time. As seen from the formula, the transfer efficiency of the NOE is strongly distance-dependent. It is inversely proportional to the sixth power of the distance between the two nuclei, with a detection limit of 5 Å– 6 Å (Williamson 2009). The sign of the NOE is also related to τ_c (which is in turn, is related to molecular weight – 1.2.1.2). At low correlation times, the NOE is positive (as the first bracket term in the equation above dominates). At high correlation times (corresponding to those in biological molecules), the NOE assumes a negative value (Figure 1.13). For molecules of an intermediate size, for which the NOE would be zero, another technique called rotating frame NOE (ROE) is used, which has a non-zero value independent of molecular weight (Figure 1.13).

Another important aspect of the NOE is spin diffusion. In NOE spreads along the chain of spins, skewing the correlation between the inter-atomic distance and NOE intensity (Williamson 2009). As such, spins which are closer to each other in space, may appear further than they really are and vice versa. Even more so, spin diffusion will give rise to signals for atoms which are beyond the NOE detection limit. Spin diffusion is especially predominant in molecules with high molecular weight and at higher mixing times. Therefore experiments with multiple mixing times should be recorded to account for spin diffusion.
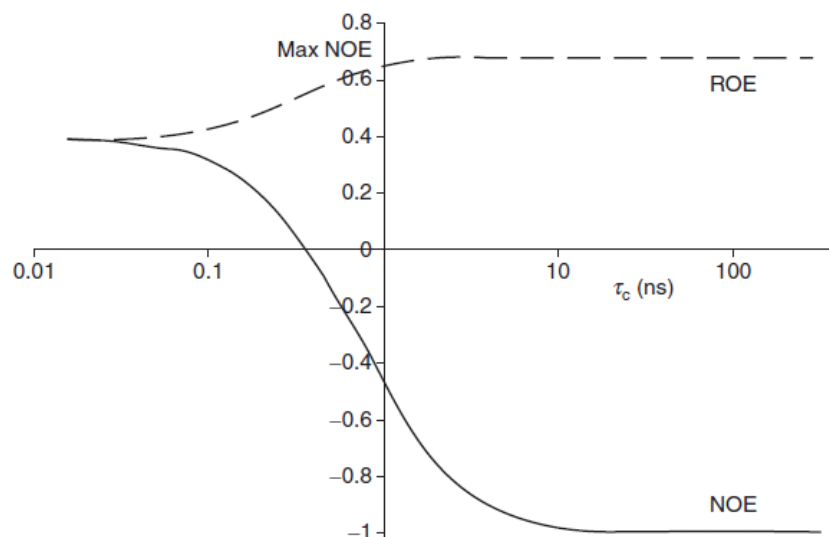
Figure 1.13 – NOE and ROE value as a function of τ_c (and hence molecular weight). At low τ_c, both the NOE and ROE values are positive, while at increasing τ_c, the former becomes negative and the latter remains positive (Williamson 2009).

The 2D homonuclear Nuclear Overhauser effect spectroscopy (NOESY) experiment, consists of three 90-degree pulses. The first pulse tilts the magnetization into the y-dimension, and is followed by t_1. A second 90-degree pulse places all the spins back to the z-axis, initiating the mixing time. It is during this period, that NOE evolves. The mixing time can be varied from one NOESY experiment to another, depending on the NOE intensity but also accounting for spin diffusion. The final 90-degree pulse transfers the magnetization to the transverse axis again, where it is detected during t_2. The resulting two-dimensional spectrum (Figure 1.14) gives rise to cross-peaks between hydrogen atoms located in close proximity to one another in space.

NOE distance restraints form the core of the spectral information required for structure calculation. This experiment provides distance restraints between atoms which are not necessarily close to each other in the amino acid or nucleotide sequence but are closer together in space. This way, information regarding overall three-dimensional fold of the biological molecule is obtained. In 3D and 4D NMR, the NOESY is combined with the HSQC, which in turn provides distance information between hydrogen atoms directly bound to ^{13}C or ^{12}C atoms, providing one with even more restraints for structure determination.
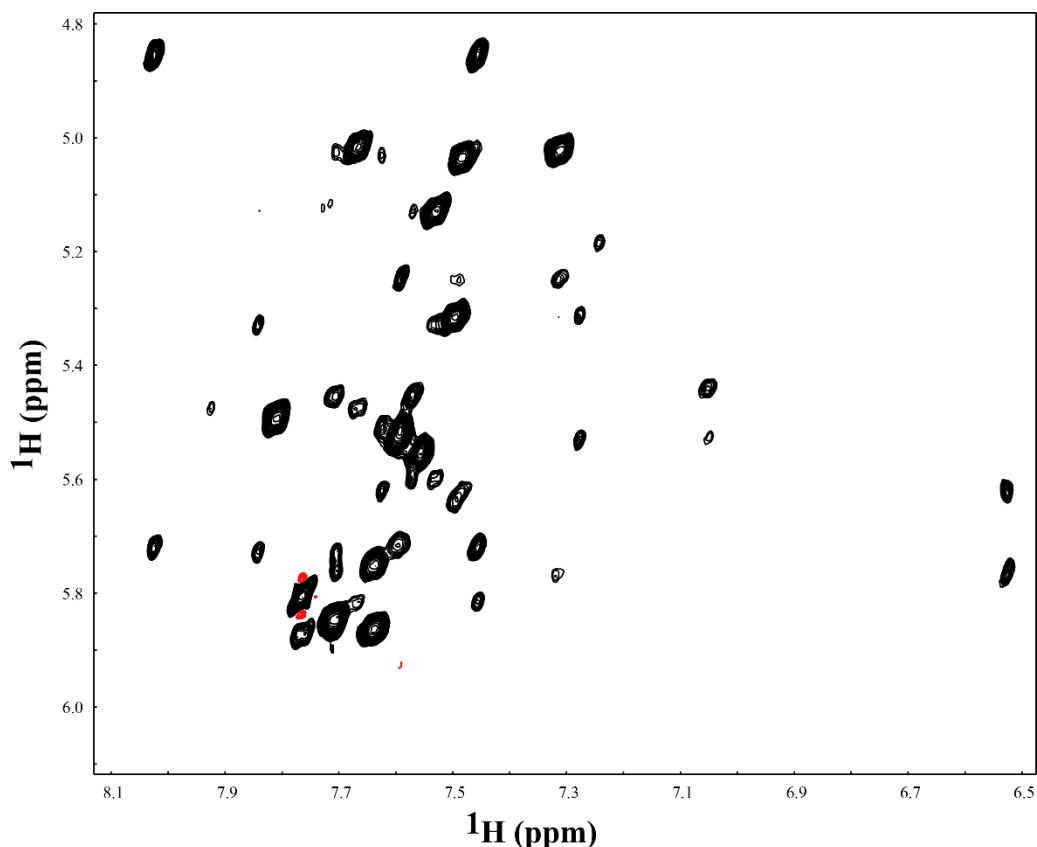
Figure 1.14 – A section of 2D, ^{1}H, ^{1}H NOESY of a 19mer seleoncysteine insertion sequence from *Moorella thermoacetica* RNA (SECIS-MT) construct (Beribisky, Tavares et al. 2007).

1.2.3 NMR in structure determination of biological molecules

Multidimensional NMR spectroscopy has emerged as a powerful tool for determination of the three-dimensional structures of biological molecules. With the advent of multidimensional NMR techniques (Vuister and Bax 1992, Sattler, Schleucher et al. 1999), determination of structures of macromolecules became possible. As of February 2014, macromolecules whose structure was determined by NMR, account for about 10% of all biological structures determined in the Protein Databank. For nucleic acids, this number is much higher, standing at 40%. For his efforts in advancing biomolecular NMR spectroscopy, the Swiss biophysicist Kurt Wüthrich was awarded the 2002 Nobel Prize in Chemistry.

As this work deals mostly with studies on nucleic acids, the process of their structure determination by NMR will be examined in more detail. A brief overview regarding protein NMR will also be provided here.

1.2.3.1 NMR in structural studies of proteins

The protein to be studied is expressed in bacteria (normally *Escherichia coli*) or yeast using a plasmid containing the coding sequence for the protein of interest. As the natural abundance of the NMR active ^{13}C and ^{15}N atoms is very low, enrichment with these isotopes is performed, so that the resulting protein sample can be subjected to heteronuclear experiments. In order to achieve this, the bacteria are grown in a medium which contains ^{13}C-labelled glucose and ^{15}N labelled ammonium chloride as the only sources of ^{13}C and ^{15}N respectively (Fernandez and Wider 2006). For larger proteins (above ~30 kDa), deuteration is also used, by growing the bacteria in a medium containing ^{2}H$_2$O, rather than water. This is done in order to reduce the relaxation rates (due to the deuterium's lower gyromagnetic ratio compared to hydrogen), resulting in sharper spectral lines and consequentially, less spectral overlap (Markus, Dayie et al. 1994, Gardner and Kay 1998, Fernandez and Wider 2006).

Following expression, the protein sample is purified using various chromatographic techniques such as nickel affinity, ion exchange and size exclusion chromatography. In order to obtain a sufficient amount of sample for NMR analysis, the final protein yield should in the milligram range.

The first experiment usually performed on a protein sample is a 2D ^{1}H, ^{15}N HSQC (Figure 1.12B). As mentioned above, the appearance of a cross-peak for the vast majority of N-H connectivities in the protein (assuming that it is properly folded and stable), makes this spectrum a good starting point to test the viability of the sample. If the protein is not properly folded or is degrading, this would result in less N-H signals, or a lack of their dispersion.

The next step is the assignment of the protein backbone. For this purpose three-dimensional through-bond experiments are used (Cavanagh 2007). One such experiment, the HNCA (Figure 1.15A) gives rise to cross-peaks from the amide proton, through its directly bound nitrogen, to two C$_\alpha$ resonances: One to the C$_\alpha$ of its own residue (*i*), and another to the C$_\alpha$ of the previous amino acid (*i-1*) in the protein sequence. In order to distinguish between the two C$_\alpha$ resonances, the HN(CO)CA (Figure 1.15A) is utilized, as this experiment is selective for the C$_\alpha$ of the *i-1* residue. Similar information regarding the C$_\beta$ and carbonyl resonances can be obtained using the pairs of HNCACB/ HN(CO)CACB (Figure 1.15B) and HNCO/HNCACO (Figure 1.15C) experiments. Using these methods, one can assign the backbone resonances as well as correlate them by

connecting these resonances in the assigned residues to each other along the protein sequence. The assignment can be performed manually or with the help of various software, such as MARS (Jung and Zweckstetter 2004).

C_α and C_β chemical shifts also provide insight regarding the protein's secondary structure. The difference in chemical shift between the assigned and the random coil values of the aforementioned resonances may indicate the presence of structural elements such as an α-helix or a β-sheet (For more information, refer to 2.2.2.4.3).
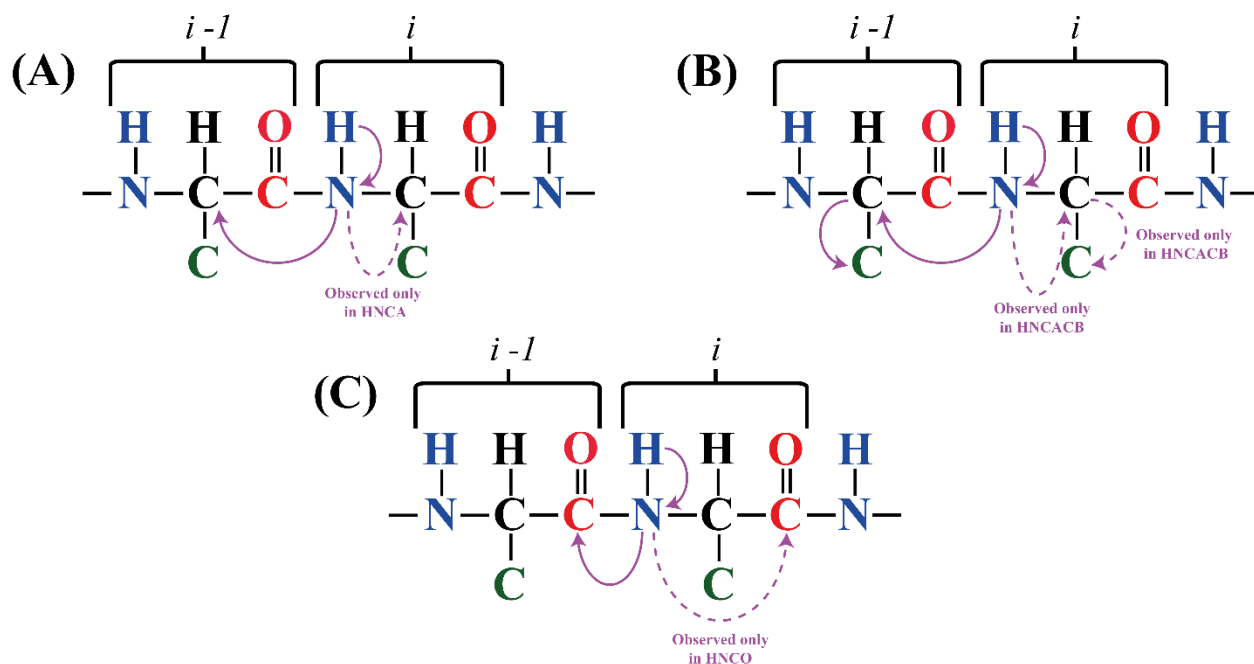
Figure 1.15 – A representation of the cross-peak connectivities showing up for a pair of **(A)** HNCA and HN(CO)CA, **(B)** HNCACB and HN(CO)CACB, **(C)** HNCO and HN(CA)CO. The nomenclature in the figure is as follows: *i* – A given amino acid residue, *i-1* – The residue previous to *i*. Color coding: Blue – amide group, Red – carbonyl group, Black – C_α, Green - C_β. Solid Arrow – Cross-peak connectivities observed for both experiments in each subsection. Dashed Arrow – Cross-peak connectivities unique to the experiment specified.

Once the protein's backbone is assigned, it is now possible to correlate every backbone signal in the 2D ^1H, ^{15}N HSQC to a certain amino acid residue in the protein. This experiment then becomes a valuable tool for titration of the protein with various ligands, as changes that occur in the spectrum upon ligand addition, can be mapped to a certain region in the protein, allowing one to obtain information regarding which regions are affected by binding and map the binding site (Cavanagh 2007)

The protein side chains are assigned. The assignment is performed with the help of three-dimensional heteronuclear experiments such as the HCCH-TOCSY and HCCH-COSY, a three-dimensional variation of the experiments which are described above (Cavanagh 2007). These experiments correlate hydrogen and carbon atoms which are three bonds apart or less or are part of the same spin system. The H(CCO)NH and HBHA(CO)ONH experiments provide connectivities between the side-chain and backbone resonances (Cavanagh 2007).

After completing resonance assignment to the most comprehensive extent possible, inter-atomic distance information is obtained, using 3D ^{1}H, ^{1}H, ^{13}C and/or ^{1}H, ^{1}H, ^{15}N NOESY-HSQC. These experiments produce an NOE cross-peak between any hydrogen atom and another hydrogen bound to a ^{13}C or ^{15}N. As mentioned, in addition to giving insight into the secondary structure of the protein, this experiment provides information regarding atoms close together in space which are not necessarily close to each other in the protein sequence. It is this inter-atomic distance information which forms the core of the restraints used for structure calculation.

Additional restraints obtained from NMR experiments used for protein structure calculations, include secondary chemical shifts, torsion angles, information extracted from Residual Dipolar Couplings (RDCs) measurements (Tjandra and Bax 1997) and Paramagnetic Relaxation Enhancements (Simon, Zanier et al. 2001).

These restraints are inputted into a computer program that generates structures based on those restraints via simulated annealing protocols. The structure calculation programs most commonly used today are ARIA (Rieping, Habeck et al. 2007), XPLOR-NIH (Schwieters, Kuszewski et al. 2003) and CYANA (Guntert 2004). Following the generation of these structures, they undergo multiple steps of refinement. This process involves going back to the spectra, correcting wrong assignments and repeating structure calculation. An ensemble of 10 structures overlaid on one another with minimized Root Mean Square Deviation (RMSD) as a result of those refinements is normally published along with the individual structures that make it up.

1.2.3.2 NMR in structural studies of nucleic acids

1.2.3.2.1 Sample preparation

The RNA sample of interest is either purchased or synthesized using *in vitro* transcription (Milligan et al., 1987, Milligan & Uhlenbeck, 1989). Similar to protein, the NMR-active ^{13}C and ^{15}N nuclei are introduced to enable multi-dimensional NMR analysis and as a result, reduce

spectral overlap. This is achieved by the preparation of ^{13}C and/or ^{15}N labelled RNA through the use of ^{13}C and/or ^{15}N labelled nucleoside triphosphates (NTPs) as the substrate for RNA synthesis (Batey, Inada et al. 1992, Nikonowicz and Pardi 1992). For larger RNA (>30 nt), various selective as well as segmental labelling approaches have been used to further simplify NMR spectra. Those approaches involve either specific label incorporation into individual NTPs through their chemical (Wenter, Reymond et al. 2006), or enzymatic (Schultheisz, Szymczyna et al. 2008, Schultheisz, Szymczyna et al. 2011) synthesis, cleavage, and re-ligation of an unlabelled 5' or 3' RNA fragment to its 3' or 5' labelled counterpart respectively (Tzakos, Easton et al. 2007, Nelissen, van Gammeren et al. 2008, Duss, Maris et al. 2010) and *in vitro* splicing of a labelled nucleotide into an unlabelled RNA chain or vice versa (Kawahara, Haruta et al. 2012). The RNA can be purified using various techniques such as gel ectrophoresis in denaturing conditions (Figure 1.16A), ion exchange (Figure 1.16B) as well as size exclusion chromatography (Milligan and Uhlenbeck 1989, Lukavsky and Puglisi 2004, Batey and Kieft 2007, Garg, Beribisky et al. 2013).

Before conducting NMR experiments, it is important to make sure that the nucleic acid in question adopts the biologically relevant conformation. This is done by running the sample on a polyacrylamide gel in non-denaturing conditions and based on its migration, determine whether it adopts the desired fold (Figure 1.16C).

(A)   **(B)**   **(C)**
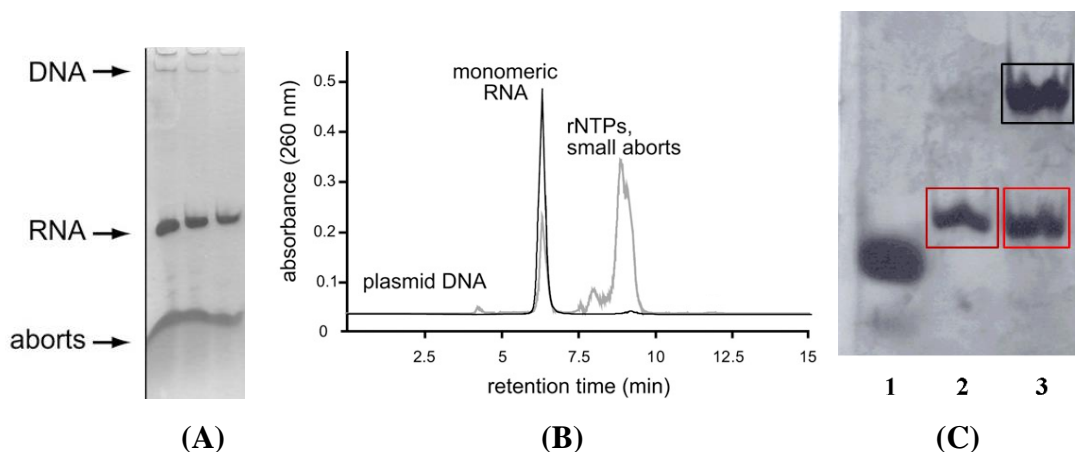
Figure 1.16 – RNA purification methods using **(A)** denaturing polyacrylamide gelectrophoresis and **(B)** anion exchange chromatography. **(C)** A non-denaturing 20% Acrylamide gel Lane 1 – 18 residue marker. Lane 2 – A19mer RNA monomer Lane 3 – Dimer formation of the same RNA is induced by the addition of 1 M NaCl to the sample. A mixture of monomer and dimer is observed. The monomer species in lanes 2 and 3 are boxed in red, while the dimer in lane 3 is boxed in black (Beribisky, Tavares et al. 2007, Easton, Shibata et al. 2010)

1.2.3.2.2 Locating signals from various parts of the nucleic acid on the chemical shift scale

When the sample is prepared, 1D NMR spectra are acquired to ensure the sample's viability for NMR analysis. As there are only four monomeric building blocks for RNA or DNA as opposed to twenty for proteins, this results in a lower chemical shift dispersion and in an overlap problem more severe than with proteins (Fürtig, Richter et al. 2003). Signals from nucleic acids show up in five different regions in the ^{1}H NMR spectrum (Wüthrich 1986). The first region from 0.5 ppm to 2.0 ppm contains signals from the thymine methyl and the methylene group atoms from the DNA's deoxyribose sugar (Figure 1.17A). The second, most crowded region in the RNA/DNA NMR spectrum from 4.0 to 5.0 ppm contains peaks from the rest of the ribose or deoxyribose sugar of the nucleic acid (Figure 1.17A, and 1.18), with the exception of the anomeric H1'. The signals of this atom are located in the region between 5.0 and 6.0 ppm, together with peaks from the H5 of uracil and cytosine (Figure 1.17A and 1.18). The region from 6.5 ppm to 8.0 ppm gives rise to signals from aromatic protons such as H2 of adenine, H8 of purines and H6 of pyrimidines as well as to the exchangeable cytosine amino moieties (Figure 1.17A). Finally, from 11.0 to 14.0 ppm lies the fifth region where the imino protons of uracil and guanine residues show up (Figure 1.17A and 1.18). Normally uracil imino signals show up downfield of their guanine counterparts (Wüthrich 1986, Varani, Aboul-ela et al. 1996). Protons in this part of the spectrum normally only appear if they are solvent-protected and are engaged in a canonical base-pairing interaction. Due to the absence of any other signals from other biological molecules, this region is commonly referred to as the nucleic acid's "fingerprint" region similar to the protein's amide region in the 2D ^{1}H, ^{15}N HSQC and is also commonly used for titration experiments.

For RNA, the amount of overlap is even higher due to the lack of the methylene group in the deoxyribose sugar and thymine methyl group which show up in a unique part of the spectrum. Consequentially, the overlap in the ribose region is even higher. Hence, structure determination of RNA constructs larger than 10-12 residues requires isotopic labelling, as above this molecular weight, spectral crowding precludes unambiguous assignment which in turn does not allow one to generate enough restraints to properly define the structure of the RNA of interest.

(A)

(B)

(C)
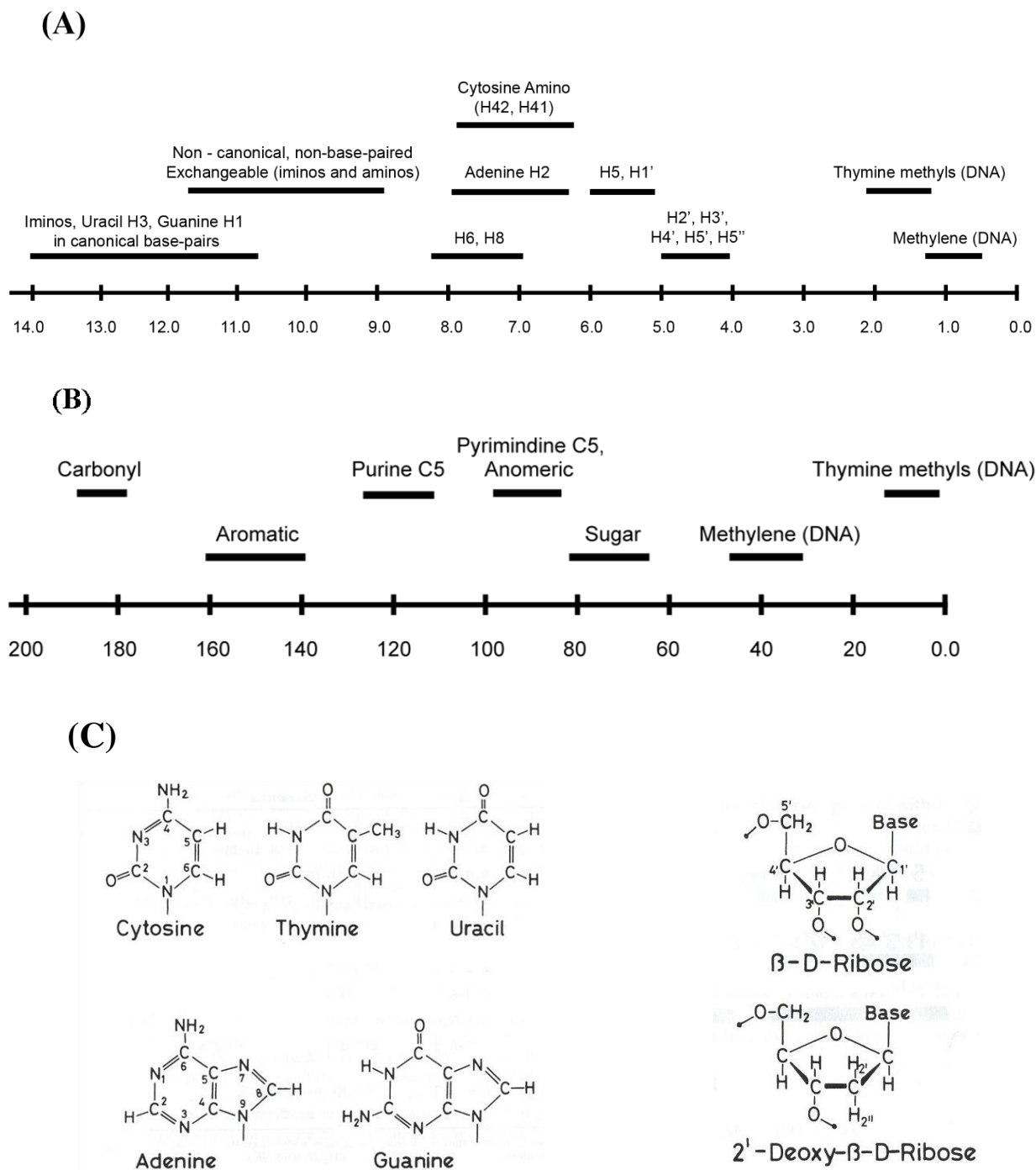
Figure 1.17 – The nucleic **(A)** [1]H and **(B)** [13]C chemical shift scales for DNA and RNA. **(C)** The bases along with the ribose and deoxyribose sugars with the appropriate numbering nomenclature. (Wüthrich 1986)
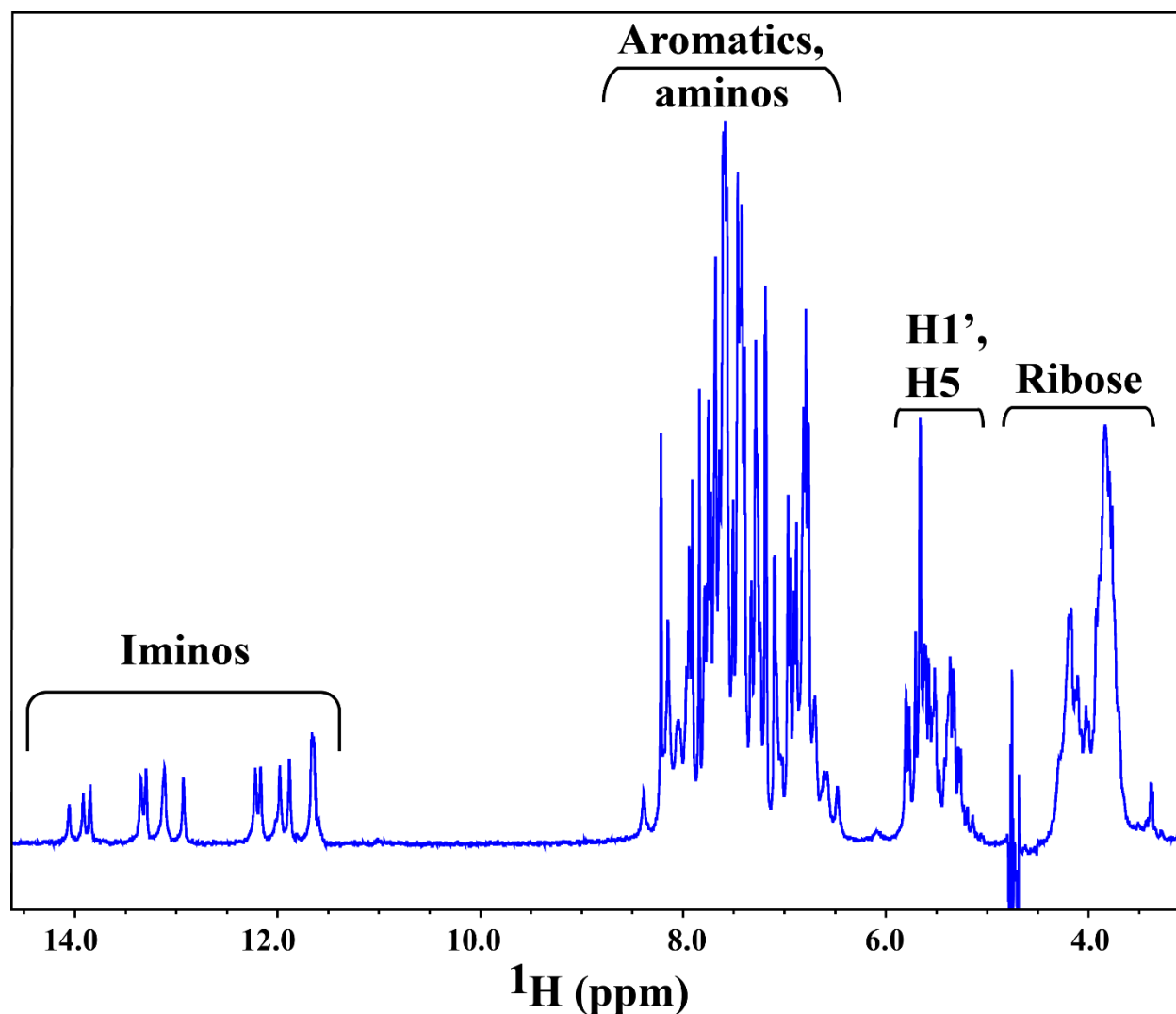
Figure 1.18 – A representative one-dimensional ^{1}H NMR of a 20mer RNA duplex, denoting the spectrum's different regions.

The ^{13}C chemical shift for nucleic acids is more dispersed than its ^{1}H counterpart. The ribose carbons are dispersed along a range of from 60 to 82 ppm (approximately 60-65 ppm for C5', 70-77 ppm for C2' and C3' and 79-82 ppm for C4'), with the exception of the carbon of the DNA's methylene group which shows up at an average value of 30 ppm (Figure 1.17B) and the C1' which shows up together with pyrimidine C5 in a region spanning 5.0 to 6.0 ppm. Signals from the aromatic C8 and C6 carbons appear in a region from 140 to 160 ppm, while guanine, uracil (thymine in DNA) and cytosine carbonyl carbon resonate downfield of 180 ppm (Figure 1.17B).

1.2.3.2.3 Homonuclear two-dimensional NMR experiments used on nucleic acids

<u>2D ^{1}H, ^{1}H water NOESY</u>

One of the first two dimensional NMR experiments performed on RNA samples, is a 2D ^{1}H, ^{1}H NOESY, recorded in 90% H_2O+10% ^2H_2O (water NOESY). This experiment allows one to obtain valuable information regarding exchangeable resonances of the molecule such as imino and amino hydrogen atoms (Wüthrich 1986, Varani and Tinoco 1991, Varani, Aboul-ela et al. 1996). In particular, since every canonical base-pair contains one imino atom, and these base-pairs are stacked about 4.5 Å apart from each other, that is, within the 5.0 Å NOESY detection limit, an NOE cross-peak between those imino resonances will be observed throughout the RNA's helical regions (Figure 1.19) allowing one to conduct a so-called imino-imino walk along the helical regions of RNA. Hence, the NOESY allows one to gain insight regarding the RNA secondary structure and determine which of its regions are helical and which are not.
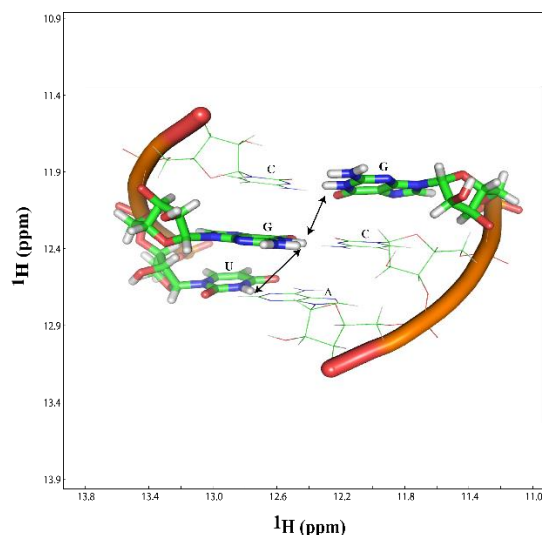
Figure 1.19 – An illustration of the distances between imino atoms between canonical base-pairs in standard, A-form RNA. The proximity of these base-pairs will give rise to NMR signals in the imino-imino region of the water NOESY spectrum.

The imino-amino region of the water NOESY is located upfield of the imino-imino region and also contains quite a bit of useful information. It allows to more clearly distinguish adenine-uracil from guanine-cytosine base-pair. Cross-peaks between the imino hydrogen atoms of guanine to cytosine amino protons can be observed in this region (Figure 1.20). The amino proton involved in the base-pairing normally resonates downfield of its non-base paired counterpart (Varani,

Aboul-ela et al. 1996). Another important feature of the imino-amino region is the very strong cross-peak (Varani, Aboul-ela et al. 1996) from the uracil imino hydrogen (UH3) to an adenine H2 (Figure 1.20). Due to its unambiguous nature, the UH3-AH2 peak very strongly suggests that an adenine-uracil base pair is present in the given assignment strip, further allowing to distinguish uracil-adenine from cytosine-guanine base-pair.
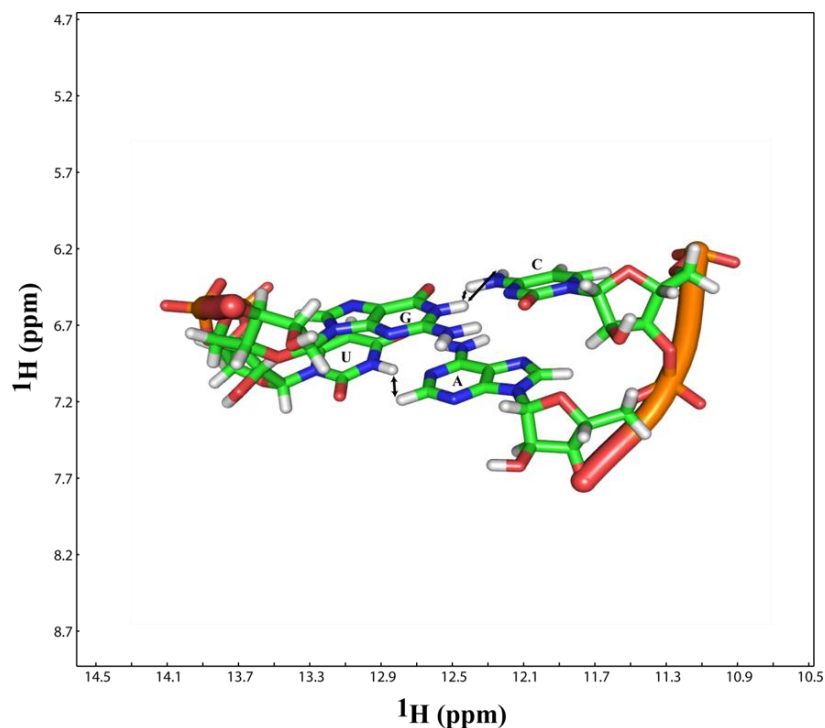
Figure 1.20 – An illustration of the distances between imino atoms of guanine and the amino hydrogens in of its cytosine base-pairing partner, as well as the distance between the uracil imino and the adenine H2 between canonical base-pairs in standard, A-form RNA. The proximity of these base-pairs will give rise to NMR signals in the imino-amino region of the NOESY spectrum.

2D ^{1}H, ^{1}H ^{2}H$_2$O NOESY

A water NOESY provides valuable information regarding exchangeable resonances in the nucleic acid such as imino and amino hydrogen atoms (Wüthrich 1986, Varani and Tinoco 1991, Varani, Aboul-ela et al. 1996). However since all RNA hydrogen atoms show up in spectra collected in water, this creates large amounts of spectral overlap resulting in difficulties in data interpretation especially for non-exchangeable resonances. Hence NOESY spectra with the sample in 100% ^{2}H$_2$O (^{2}H$_2$O NOESY) are also acquired. In this experiment exchangeable resonances are not observed due to their exchanges with the deuterated solvent.

Of great interest in the ^2H_2O NOESY is its aromatic/anomeric region which shows NOE cross-peaks between the aromatic H8 and H6 (for purine and pyrimidine respectively) atoms the H1' resonances. Exclusively for pyrimidine residues (with the exception of thymine in DNA), very strong H6/H5 correlations are also present. The assignment scheme used for this region of the spectrum is the so-called aromatic-anomeric "walk" (Figure 1.21) which is based on the fact that in helical regions of the nucleic acid, each aromatic H6 or H8 atom gives rise to an NOE cross-peak to the H1' of its own ribose sugar and the H1' of the residue 5' to it (Fürtig, Richter et al. 2003). Using this scheme, the aromatic anomeric cross-peaks for the helical parts of the RNA can be connected.

The assignments of aromatic-anomeric walk can be verified by numerous means. Adenine H2 resonances in this region of the spectrum gives NOE signals to the H1' atom of its 3' residue (A16H2 cross-peaks in the furthest to the right in Figure 1.22A), as well to the H1' of the residue 3' of its base-paring partner (Varani, Aboul-ela et al. 1996). Spin diffusion cross-peaks from the water NOESY which correlate the guanine imino to the H5 of its cytosine base-pairing partner, are another way to corroborate the aromatic-anomeric walk. Aromatic-aromatic cross-peaks next to the diagonal of the spectrum, also allow one to cross-check the completed assignments.

In addition to information regarding the location of helical regions in the DNA/RNA of interest, cross-peak intensities in the ^2H_2O NOESY indicate base conformations. A weak intra-residue H8 (or H6)/H1' signal points to an *anti* conformation, while a strong signal suggests a *syn* conformation (marked in red on Figure 1.22A and 1.22B). In addition, very sharp H6/H5 cross-peaks (marked in blue on Figure 1.22A and 1.22B), strongly suggest that the base of the residue in question extrudes from the structure (Johnson and Donaldson 2006, Beribisky, Tavares et al. 2007). Finally an unusually downfield H2' for the 5' residue in an RNA, produces a distinctive cross-peak to its own H8 resonance at about 5.0 ppm (marked in a box in Figure 1.20A). The H8 atom of the 5' itself shows up at around 8.0 ppm. These resonances provide a good starting point for spectral assignments in the ^2H_2O NOESY.

There are two other experiments which can facilitate the assignment of this part of the ^2H_2O NOESY. These experiments are 2D $^1H,^1H$ TOCSY, and the 2D $^1H, ^{13}C$ HSQC, and they will be discussed below.
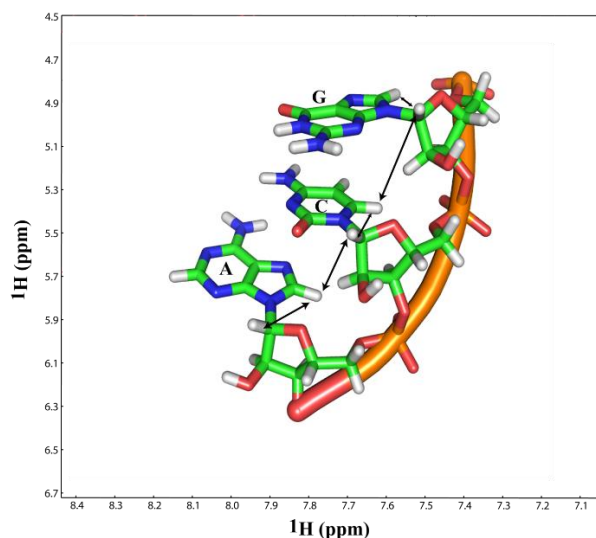
Figure 1.21 – An illustration of the distances between the aromatic H8 (or H6) and the anomeric H1' atoms in standard, A-form RNA. Each aromatic atom is about 4.5 Å from the H1' of its own as well that of its 5' residue. This distance will give rise to NMR signals in the aromatic-anomeric region of the ^2H_2O NOESY spectrum.

(A)   **(B)**
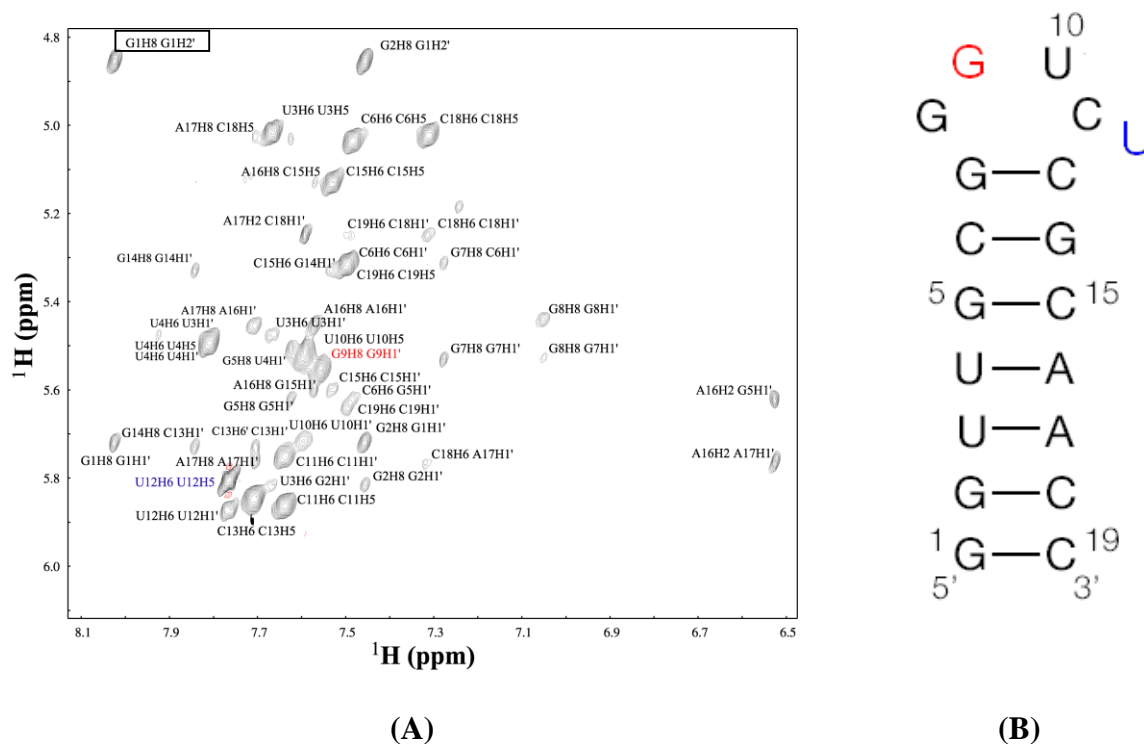
Figure 1.22 – (A) Aromatic–anomeric region of the 2D ^1H, ^1H ^2H_2O NOESY for a 19mer SECISM-MT RNA. Assignments are labeled in the spectrum. The strong G9H8/G9H1' and the sharp U12H6/U12H5 signals are marked in red and blue, respectively. The G1H8/G1H2' cross-peak is marked in a box. **(B)** The secondary structure of the RNA construct whose spectrum is shown in (A). The *syn* G9 and the extruded U12 are labeled in red and blue respectively, all the other residues labelled in black, adopt an *anti* conformation (Beribisky, Tavares et al. 2007).

37

2D ^{1}H, ^{1}H ^{2}H$_2$O TOCSY

The 2D ^{1}H, ^{1}H TOCSY, is commonly used on nucleic acids. In its aromatic-anomeric region, this experiment gives rise to just one type of signals – the cytosine and uracil H6/H5 intra-residue cross-peaks (Figure 1.23A).

The number of H6/H5 correlations in the TOCSY spectrum is a good indicator of whether the nucleic acid in question adopts a single conformation. When it does, the number of H6/H5 cross-peaks should match the number of pyrimidine residues in the RNA sequence. If extra peaks are present, then the RNA adopts more than one conformation and optimization of buffer and/or salt conditions may be required. Assuming that the former is the case, then when overlaid with the 2D ^{1}H, ^{1}H ^{2}H$_2$O NOESY, the location of the H6/H5 cross-peaks in that spectrum becomes immediately known, further simplifying the spectral assignments in the aromatic-anomeric region.

The TOCSY aromatic-anomeric region is also often used to conduct titrations (Tavares, Beribisky et al. 2009, Garg, Beribisky et al. 2013) because of the small number of signals, and also due to the fact that one pyrimidine per Watson-Crick base-pair is always present, allowing one to gauge the effect of ligand binding on every base-pair in the helical region, as well as other pyrimidine-containing parts of the RNA (Figure 1.23B).

(A)   (B)
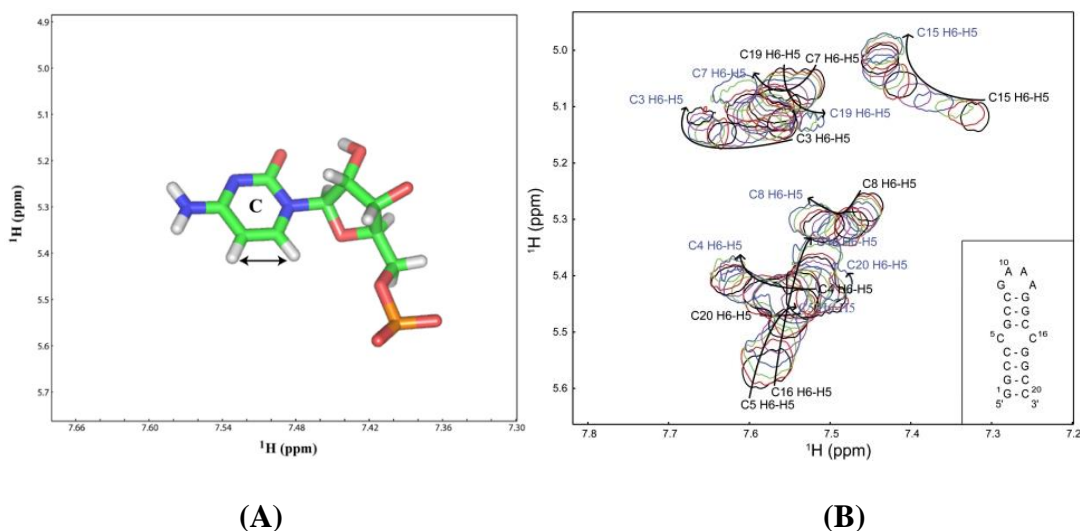
Figure 1.23 – (A) An illustration of the coupling between the H6 and H5 atoms in a pyrimidine. These atoms are three bonds apart. This distance will give rise to NMR signals in the aromatic-anomeric region of the 2D ^{1}H, ^{1}H ^{2}H$_2$O TOCSY spectrum. **(B)** The aromatic-anomeric region of the 2D ^{2}H$_2$O ^{1}H, ^{1}H TOCSY titration of a CC mismatch-containing RNA with an aminoglycoside antibiotic, causing chemical shift perturbation in its H6/H5 cross-peaks (Tavares, Beribisky et al. 2009).

38

As mentioned above (1.2.2.2), a different region of the 2D ^{1}H, ^{1}H ^{2}H$_2$O TOCSY allows one to gain insight regarding the sugar pucker conformation of the nucleic acid of interest. In anomeric-ribose region experiment gives rise to an intra-residue H1'/H2' signal for the residues which adopt a *C2' endo* conformation, due to the larger value of three-bond coupling constant between H1' an H2' (Fürtig, Richter et al. 2003). Residues which adopt the C3' *endo* conformation, on the other hand, do not give rise to such cross-peaks. A short mixing time must be used for this experiment as with longer mixing times, the magnetization is allowed to travel further along the spin system, which may cause the appearance of H1'/H2' cross-peaks which do not represent the true sugar pucker conformation.

1.2.3.2.4 Heteronuclear NMR experiments used on nucleic acids

Once an isotopically labelled sample is prepared, it is analyzed using one-dimensional and possibly two-dimensional NMR experiments to ensure its viability for further analysis, one of the first experiments performed is a 2D ^{1}H, ^{13}C HSQC (Note: this experiment can also be performed on an unlabelled sample, however since the natural abundance of ^{13}C is about 1%, the measurement for this experiment increases significantly). As each hydrogen-carbon correlation is shown in this experiment, it provides a good starting point for their assignment.

Two-dimensional HSQC experiments

The 2D ^{1}H, ^{13}C HSQC, helps in distinguishing between the cytosine and uracil resonances using their carbon chemical shift. Since cytosine C5 has a more upfield chemical counterpart, the H5/C5 correlations for this residue will be found above those of uracil. As the C5 assignment can be directly correlated to H5, this also helps in differentiating between the different types of pyrimidine residues in the aromatic-anomeric region of the ^{2}H$_2$O NOESY (Figure 1.24).

The H1'-C1' region of a 2D ^{1}H, ^{13}C HSQC, provides insight regarding the presence or absence of base-pairing in a given residue. A downfield carbon chemical shift for C1' indicates that the residue to which this C1' belongs may be involved in base-pairing (Marked in a black box in Figure 1.11). This information is particularly important when signals for the exchangeable atoms for this residue are absent, as this alone does not exclude the participation of this residue in base-pairing interactions.
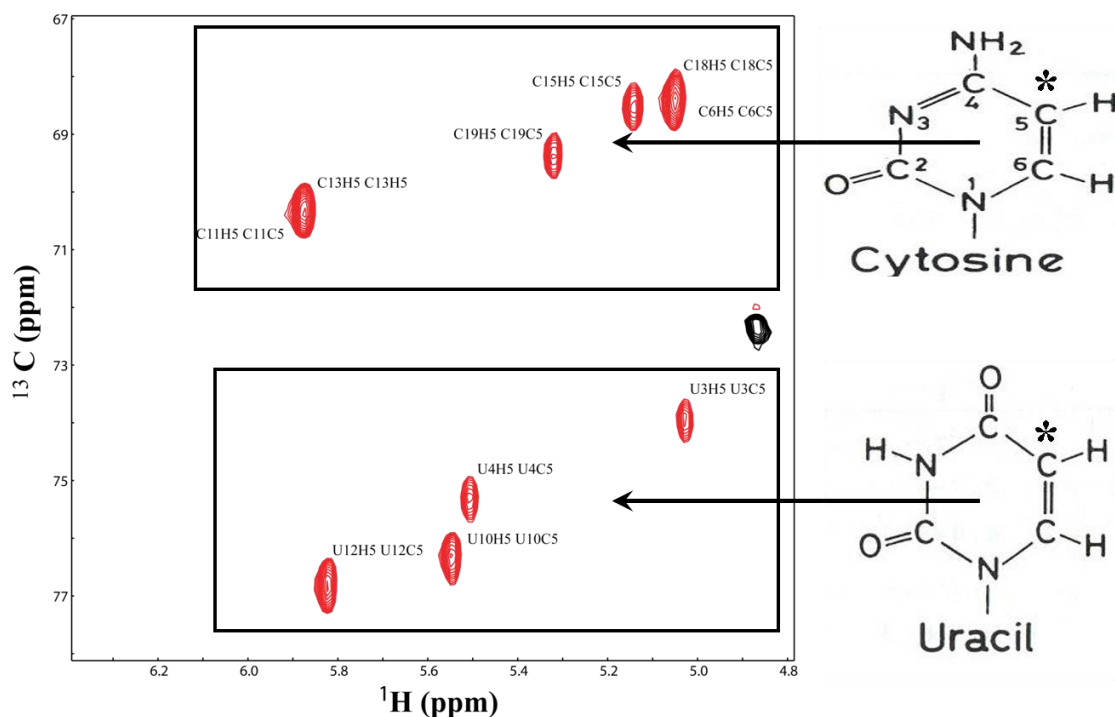
Figure 1.24 – The H5/C5 region in a 2D ^{1}H, ^{13}C HSQC of SECIS-MT. Due to the smaller C5 chemical shift of the cytosine, the H5/C5 cross-peaks for this residue are located upfield in the carbon dimension compared to the H5/C5 correlation for uracil (Beribisky 2008).
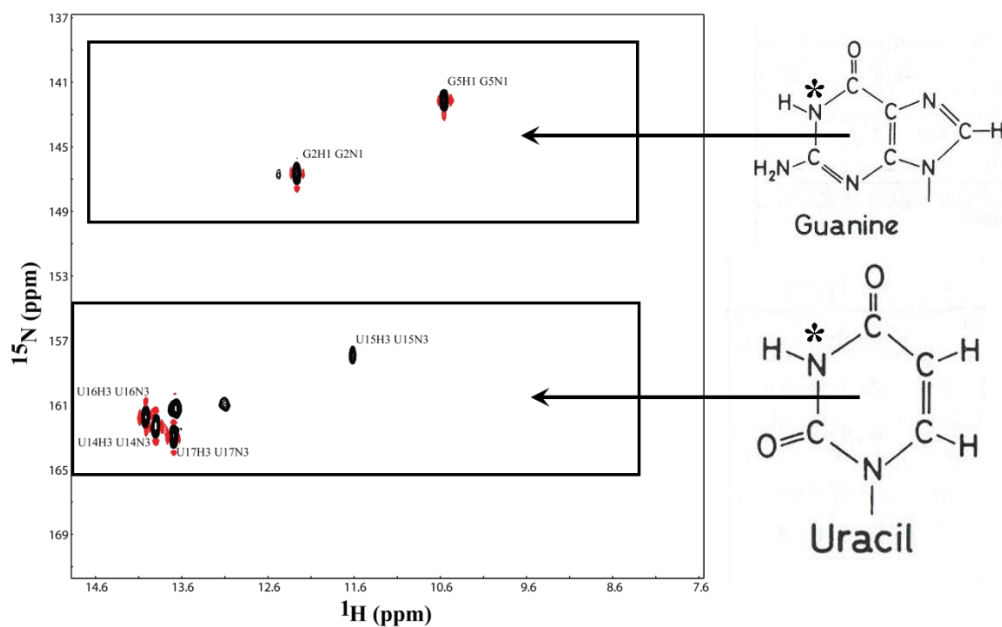
Figure 1.25 – The imino region in a 2D ^{1}H, ^{15}N HSQC. Due to the smaller guanine N1 chemical shift compared to the N3 of uracil, the guanine imino resonances appear upfield of their uracil counterparts in the nitrogen dimension (Beribisky 2008).

To complement the 2D ^{1}H, ^{13}C HSQC a 2D ^{1}H, ^{15}N HSQC is also performed. This experiment provides spectral information regarding the amino resonances of cytosine and any other amino signals which may be observed. In addition, the upfield ^{15}N chemical shift of N1 imino of guanine compared to N3 of uracil, allows one to differentiate between the guanine and uracil imino atoms (Figure 1.25), provided it was not possible to do so in the homonuclear spectra.

The use of three-dimensional experiments in assigning the ribose/deoxyribose sugar

To assign the ribose sugar, one has to resort to three-dimensional experiments. These experiments are 3D ^{1}H, ^{13}C, ^{1}H HCCH-TOCSY and 3D ^{1}H, ^{13}C, ^{1}H HCCH-COSY. As described above (1.2.2.1), the latter correlates resonances which are three bonds apart or less, while the former would give rise to cross-peaks along an entire spin system. As such, in a HCCH-TOCSY, the H1' (through its directly bound C1') would give rise to signals to all the other hydrogen atoms in the ribose spin system – H2', H3', H4', H5' and H5'' (Figure 1.26A). While the H4', H5' and H5'' can all be resolved in the carbon frequency as they all show up in unique regions in the ^{13}C chemical shift scale (1.2.3.2.4), the H2' and H3' are clustered together even in this dimension. In order to distinguish between them the HCCH-COSY is used. As in this experiment the H1' (through its directly bound C1') would only produce one cross-peak – to its neighboring H2'. The H3' for the given ribose sugar can be determined by elimination. For DNA, as the methylene carbon and hydrogen occupy a distinct region in their chemical shift scales, and so an HCCH-COSY is not required.

Three-dimensional NOESY experiments

The same experiments as for proteins, that is 3D ^{1}H, ^{1}H, ^{13}C and ^{1}H, ^{1}H, ^{15}N NOESY-HSQC are used. If a sample is selectively labelled, then one can resort yo 4D NOESY experiments which employ editing and filtering. These experiments would allow or suppress hydrogen resonances which are either ^{13}C or ^{12}C bound. Editing and filtering is discussed in further detail in sections 2.3.4.3 and 2.3.4.4.
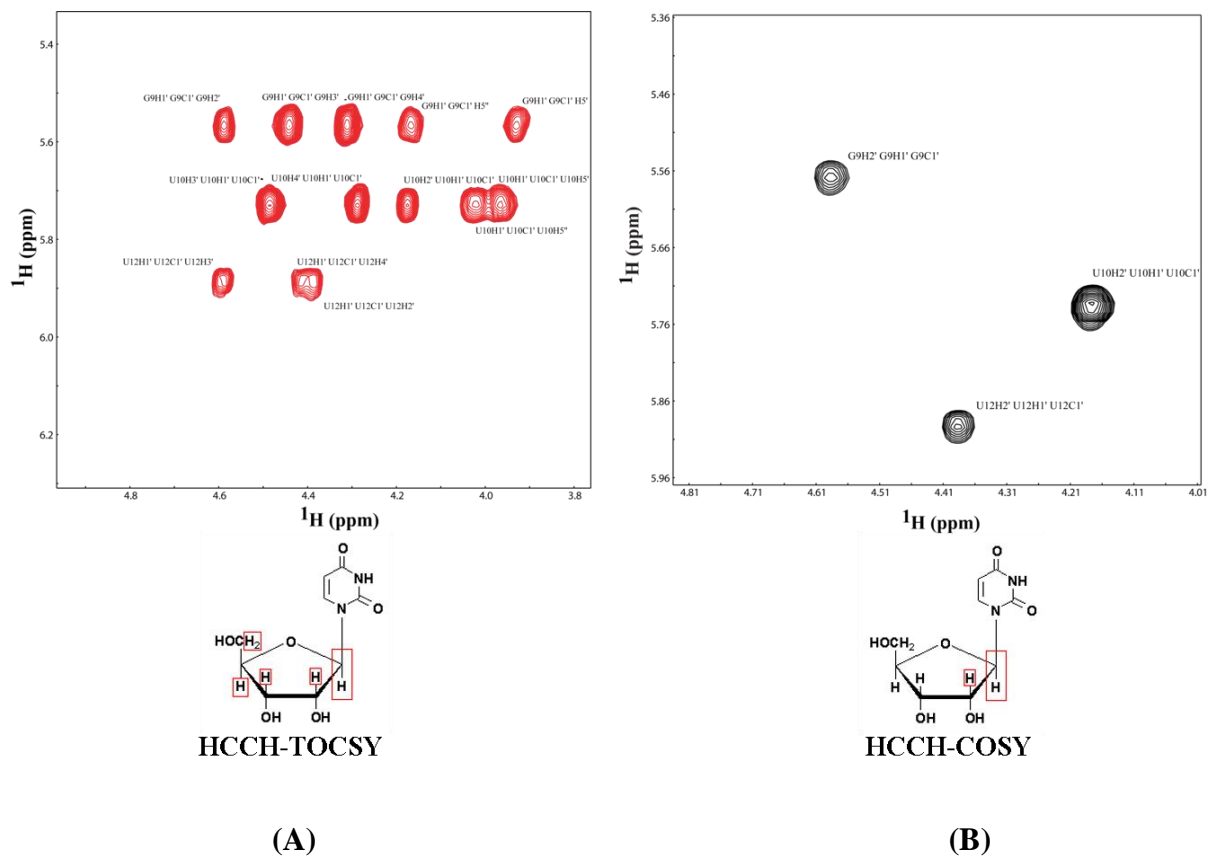
Figure 1.26 – (A) A two-dimensional plane of a 3D ^{1}H, ^{13}C, ^{1}H HCCH-TOCSY showing cross-peaks from the anomeric H1' atom (through its directly bound C1'), to all the other hydrogen atoms in the ribose spin system. **(B)** A two-dimensional plane of a 3D ^{1}H, ^{13}C, ^{1}H HCCH-COSY showing a cross-peak from the anomeric H1' atom (through its directly bound C1'), to its neighboring H2'(Beribisky 2008).

1.2.3.2.5 Generation of distance restraints and structure calculation

In addition to distance restraints from the various NOESY spectra, restraints such as dihedral (which include sugar puckers) and glycosidic angle values are used. In order to flatten the bases and better define base-pairing, hydrogen-bonding and planarity restraints are also introduced. RDCs and more recently PRE data sometimes also complement the traditional restraints outlined here.

For structure calculations programs used for protein structure determination such as ARIA (Rieping, Habeck et al. 2007) and XPLOR-NIH (Schwieters, Kuszewski et al. 2003) are utilized for nucleic acid calculations as well. As with proteins, after multiple steps of refinement, by correcting spectral assignments as well as computational methods such as the AMBER water protocol (Case, Cheatham et al. 2005, Perez, Marchan et al. 2007), a final structure is generated.

An ensemble of 10 structures overlaid on one another with minimized (RMSD) as a result of those refinements is normally published along with the individual structures that make it up (Figure 1.27).

(A)   (B)
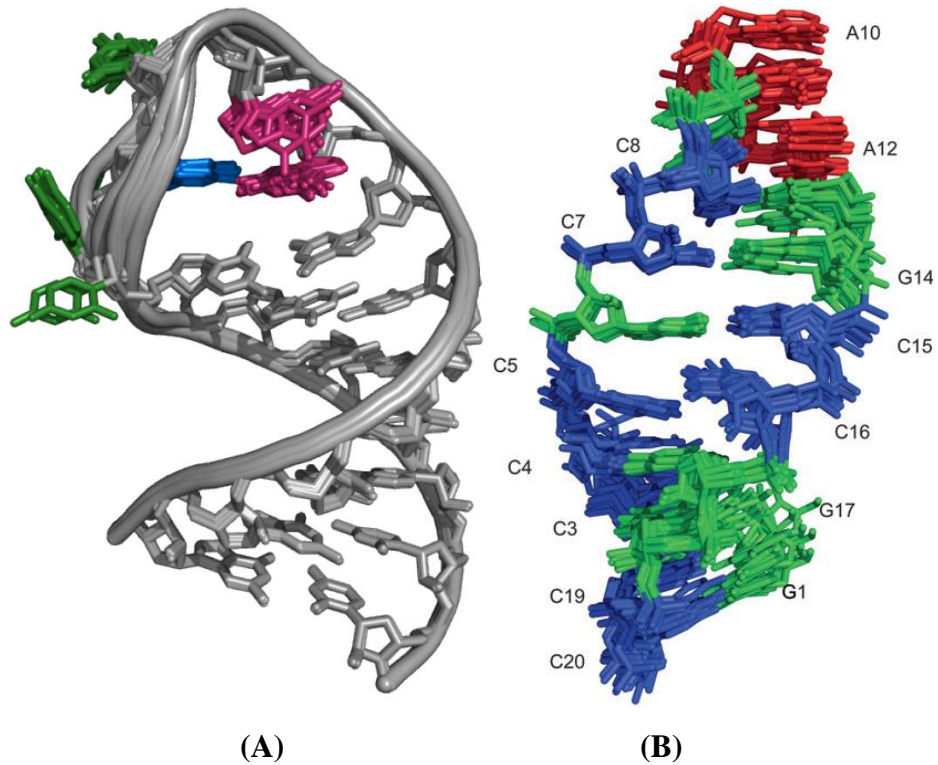
Figure 1.27 – An overlay of the ten lowest energy structures for an (A) 19mer and (B) 20mer RNA constructs whose structure was solved by NMR spectroscopy (Beribisky, Tavares et al. 2007, Tavares, Beribisky et al. 2009).

1.3 Aims and scope

The aims of this work is to explore the structure of two biologically important nucleic acids and the interaction with their respective binding partners. The first nucleic acid is an RNA duplex, containing the non-canonical residue inosine, called the inosine-edited RNA duplex (IRD). It is thought to interact with p100, a multi-functional protein from a multi-protein complex involved in RNA interference (Scadden 2005, Li, Yang et al. 2008). This interaction leads to possible RNA cleavage. Structural work on this protein will also be conducted.

The other is a RNA hairpin construct derived from a larger messenger RNA coding for the protein thymidylate synthase (TS), termed the TS mRNA construct (TSMC). This RNA contains a CC mismatch in its stem (Chu, Voeller et al. 1993). This CC mismatch is a known binding target for various small molecules. The interaction with one such compound may inhibit the translation of the protein which this mRNA codes for.

The interaction between IRD and p100 is an important junction point between two important cellular processes: RNA interference and RNA editing (Scadden and Smith 2001, Scadden 2005). Misregulation of any of these processes is associated with various disorders. The interplay between the TS protein and TS mRNA is an example of autoregulatory feedback inhibition (Cheng, Moore et al. 2003, Maas, Kawahara et al. 2006). TS is also the only *de novo* source of thymine in the cell (Chu, Koeller et al. 1991) and is hence a target of various anti-cancer drugs. Therefore obtaining structural information for these molecules would provide valuable insights into the mechanisms of regulation of such processes.

In order to conduct these studies, various techniques will be employed. NMR spectroscopy will be of central importance in this work, as this method will allow us to obtain structural information on the atomic resolution level. In addition to structural information, NMR relaxation experiments will provide us with insight regarding the dynamics of the biological molecule of study. NMR will also be complemented by various biophysical and biochemical studies to further characterize the structure of the two RNA constructs and their interactions with their binding partners. These methods include UV melts, mobility shift and nuclease assays.

Chapter 2 –Structural study of the inosine edited RNA the RISC component p100

2.1 Introduction

2.1.1 Introduction to RNA interference

RNA is known to be the centrepiece of various cellular processes. One such process is RNA interference (RNAi). The phenomenon was first described in 1998 by Craig Mello and Andrew Fire (Fire, Xu et al. 1998). For their work in this field, these authors were awarded the 2006 Nobel Prize in Medicine.

RNAi has a number of biological functions. One such function is facilitation of immune response. In *Drosophila melanogaster*, for instance, RNAi has been shown to confer innate immunity from the Drosophila X virus (Zambon, Vakharia et al. 2006). Plants were shown to encode RNA which undergo RNAi in response to bacterial infection (Katiyar-Agarwal, Morgan et al. 2006). RNAi is also involved in gene downregulation by translational inhibition or by mRNA degradation in *Caenorhabditis elegans* (Lee, Feinbaum et al. 1993), in plants (Zhang, Pan et al. 2006), and in humans, where RNAs involved in RNAi, can act as both oncogenes and tumor suppressors (Zhang, Pan et al. 2007). Conversely, RNAi can also upregulate numerous genes by enhancing transcription, possibly via histone demethylation (Check 2007).

There are two kinds of RNA which impart silencing in RNAi – microRNAs (miRNAs), and small interfering RNAs (siRNAs). These RNAs are processed by two separate pathways before converging in the RNA-induced silencing complex (RISC). miRNAs are endogenous, non-coding RNAs which are transcribed in the nucleus as primary (pri)-miRNA before being cleaved by a micro-processor complex composed of the RNAse III protein Drosha and a double stranded RNA (dsRNA)-binding protein DGCR8 (Figure 2.01) into a 70 nucleotide stem-loop (Gregory, Chendrimada et al. 2006). The resulting pre-miRNA is exported to the cytoplasm by the exportin-5 (Figure 2.01) protein (Yi, Qin et al. 2003) and has its double stranded portion further processed by the protein Dicer to produce a mature miRNA (Figure 2.01) which is then integrated into the RISC complex (Gregory, Chendrimada et al. 2006).

Endogenous siRNAs are also processed by Dicer (Bernstein, Caudy et al. 2001), to give rise to 20 bp dsRNA with a dinucleotide overhang at the 3' end (Vermeulen, Behlen et al. 2005). Exogenous long dsRNAs, however are detected and processed by an effector protein (RDE-4 in *C. elegans* and R2D2 in *D. melanogaster*) which incorporates the resultant RNA into the RISC complex (Liu, Rand et al. 2003, Parker, Eckert et al. 2006).

RISC is a multi-protein complex whose function is to use the processed miRNAs and siRNAs to give rise to the RNAi effector function (Figure 2.01). It does so by cleaving a target mRNA which is base-paired with the mature miRNA or siRNA by argonaute proteins (Liu, Carmell et al. 2004). As only one siRNA strand is required to hybridize to its target mRNA, and the siRNAs produced by Dicer are double-stranded, the two strands have to be separated by the protein Ago2 (Leuschner, Ameres et al. 2006). The strand that is to be base-paired (guide strand) is normally more stable at its 5' end (Schwarz, Hutvagner et al. 2003). For siRNAs, strand selection is carried out by R2D2 which binds the guide strand (Tomari, Matranga et al. 2004).
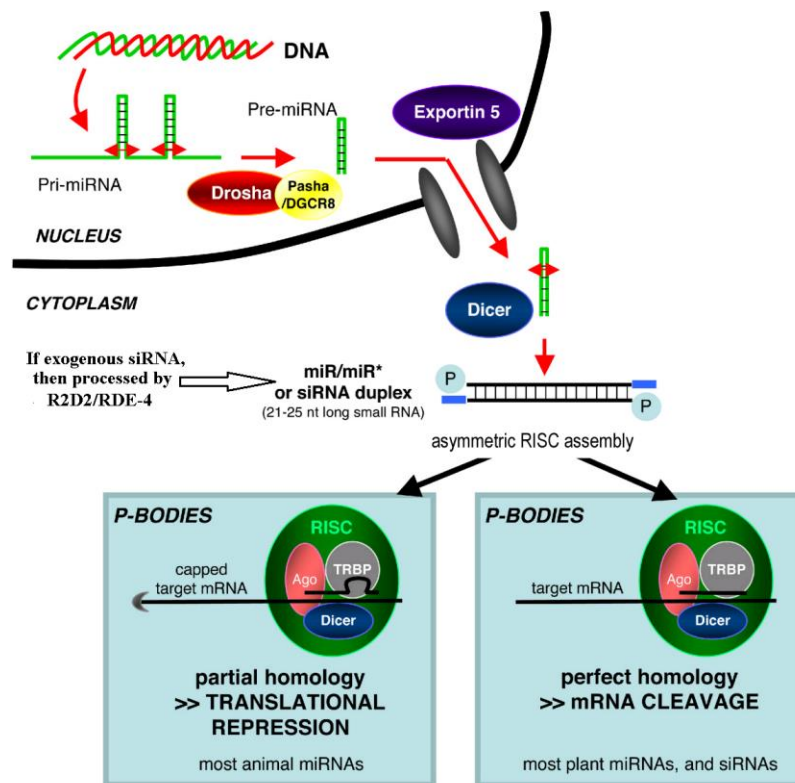
Figure 2.01 – A representation of the RNAi pathway. miRNAs are processed by DGCR8 in the nucleus, before being exported to the cytoplasm, where they are further processed by Dicer. siRNAs are also processed by Dicer, and the mature siRNA or miRNA are then incorporated into the RISC complex which exerts RNAi function by separating one miRNA or siRNA strand from the other and inducing translational repression or mRNA cleavage (Saumet and Lecellier 2006).

As mentioned, RISC is a heterogeneous complex, and is composed of various proteins. One component of RISC is the protein p100 (also known as Tudor-SN, SND1) (Caudy, Ketting et al. 2003). p100 is a highly versatile protein which modulates numerous cellular events. It is known to act as a transcriptional co-activator, interacting with transcription factors EBNA-2 (Tong,

Drapkin et al. 1995), c-Myb (Leverson, Koskinen et al. 1998), STAT6 (Valineva, Yang et al. 2005, Valineva, Yang et al. 2006) and STAT5 (Paukku, Yang et al. 2007). The protein was implicated in spliceosome assembly (Shaw, Zhao et al. 2007, Yang, Valineva et al. 2007). In addition, p100 was shown to be a target for cleavage by caspases, hence making it involved in the pathway for programmed cell death (Sundstrom, Vaculova et al. 2009). More recently, p100 was implicated in miRNA processing, slowing down miR-17-92a cluster maturation under hypoxic conditions (Heinrich, Wagner et al. 2013). Finally another function of p100, which will be examined in this work, is serving as the junction point between RNAi and another important cellular process, RNA editing.

2.1.2 Introduction to RNA editing

RNA editing is a specific post-transcriptional modification of an RNA sequence of some sort. RNA editing can be thought of as another level of post-transcriptional control, where changes to an mRNA sequence can be introduced after it was synthesized, thereby potentially changing the type of amino acid that this mRNA codes for. This in turn, can alter the resultant protein sequence, as well as possibly its structure and function.

There are two general kinds of editing: Editing by insertion of deletion, or editing by nucleotide modification. The former was shown to occur in *Trypanosoma brucei*, where an unedited primary transcript hybridizes with a so-called guide RNA (gRNA), which contains the complementary sequences that is to be edited in and lacks the sequences to be edited out. The resulting dsRNA is then enveloped by a multi-protein editosome complex, which catalyzes the editing process (Alfonzo, Thiemann et al. 1997).

Editing by nucleotide modification involves a modification to the structure of a particular nucleotide. This kind of editing has been widely documented in higher eukaryotes and has two sub-types, both of which involve nucleotide deamination: Cytosine to uracil (C to U editing) and adenine to inosine (A to I editing). C to U editing was first discovered in apoliprotein B (apoB) in humans. Its mRNA sequence contains the codon triplet CAA which codes for glutamine (Powell, Wallis et al. 1987). When not subjected to editing, a full-length apoB (apoB100) is translated as a result, which is localized to the liver. However in the small intestine (Powell, Wallis et al. 1987), where the cytosine residue in the aforementioned triplet undergoes editing by Apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1 (APOBEC1) (Nakamuta, Oka et al. 1995), the

triplet is changed to UAA, which serves as a termination signal in protein synthesis. As a result, a truncated version of apoB (apoB48) is produced.

C to U editing was also implicated in various aspects of immune response. Activation-induced (DNA-cytosine) deaminase performs C to U conversions in B lymphocyte cells making it implicated in the generation of antibody diversity (Muramatsu, Kinoshita et al. 2000). The protein Apobec3G has been shown to act against mutant Human Immunodeficiency Virus (HIV) strains (Sheehy, Gaddis et al. 2002) by producing a C to U hypermutation in the negative strand of the virus's DNA (Mbisa, Barr et al. 2007), resulting in complete abolition of HIV activity (Sheehy, Gaddis et al. 2002). Apobec3G has also shown anti-viral activity against other viruses (Komohara, Yano et al. 2006).

A to I editing is a form of RNA editing in which an adenine's exocylic N6 amino group is deaminated giving rise to the non-canonical residue inosine (Figure 2.02). This reaction is catalysed by a type of enzymes (Bass and Weintraub 1988) called Adenosine deaminases acting on RNA (ADARs). Since inosine is recognized as guanine by the cell's translational machinery (Basilio, Wahba et al. 1962), this process provides a level of post-transcriptional regulation during gene expression, increases transcript diversity and as a consequence, protein structural and functional diversity.

A to I editing is performed only on dsRNAs by the ADAR's C-terminal catalytic region (Lai, Drakas et al. 1995). The enzyme's N-terminal double stranded RNA binding domain is responsible for RNA binding (Ryter and Schultz 1998). For RNA substrates larger than 100 bp, the rate of editing can be very high and random, as more than 50% of adenine residues in such RNAs may be converted to inosine (Cattaneo, Schmid et al. 1988). On the other hand, the amount of editing with shorter, 20-30 bp dsRNAs is much lower, but is also much more specific (Scadden and Smith 1997, Lehmann and Bass 1999, Scadden and Smith 2001). In the latter case, ADARs preferably edit particular sites on the RNA, depending on the identity of the residue 5' of the adenine to be edited. ADAR1 and ADAR2 have the following 5' residue preference: A = U (both adenine and uracil have a similar preference> C > G (Lehmann and Bass 2000).
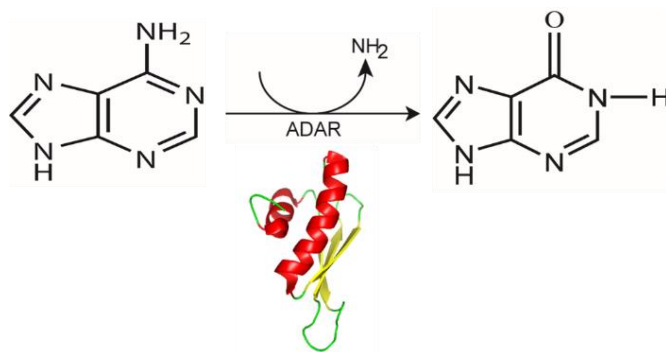
Figure 2.02 – The deamination of adenine's amino to a carbonyl by ADARs to give rise to inosine. ADAR structure adapted from Schade et al. (Schade, Turner et al. 1999)

A to I editing was shown to be implicated in various cellular processes. In the nervous system, ion channel-encoding mRNAs were found to be edited (Burns, Chu et al. 1997, Bhalla, Rosenthal et al. 2004). RNA editing was also shown to regulate alternative splicing (Rueter, Dawson et al. 1999). Various viral RNAs were also shown to be heavily modified by ADARs (Bass, Weintraub et al. 1989, Polson, Bass et al. 1996).

2.1.3 The interplay between RNAi and RNA editing

A to I editing was also found to antagonize RNAi (Scadden and Smith 2001, Yang, Wang et al. 2005, Kawahara, Zinshteyn et al. 2007). Inosine edited RNAs could not be recognized by Dicer (Bass 2000, Scadden and Smith 2001, Kawahara, Zinshteyn et al. 2007). It is thought that Dicer-processed inosine-edited siRNA would not be able to properly base-pair with its cognate mRNA in RISC. Conversely, it was shown that the RNA produced as a result of Dicer processing, are too short to be cleaved by ADARs (Bass 2002).

In one another example, a sequence found in exons 2 and 3 of the rat α-tropomyosin gene has been subjected to editing (Scadden and Smith 1997, Scadden and Smith 2001). As a result, a hyper-edited motif with four IU base-pairs (Figure 2.03B, top construct) is formed with the complementary sequences IIUI and UUIU in the sense and anti-sense strands respectively (Scadden and Smith 1997, Scadden and Smith 2001). A 20 nucleotide dimer fragment comprising this motif was shown to be the binding target, as well as the potential cleavage substrate of the RISC component p100 contained within *Xenopus laevis* extracts (Scadden 2005, Li, Yang et al. 2008). p100 specifically binds and cleaves (Figure 2.03A) this inosine-edited RNA duplex (IRD), while showing no binding or cleavage activity towards a non-edited RNA duplex where the inosine residues have been replaced with guanine (Figure 2.03B, middle construct) (Scadden 2005), and

lack of cleavage activity towards another non-edited RNA construct where inosine was replaced with adenine (Figure 2.03B, bottom construct) (Li, Yang et al. 2008). By binding and cleaving IRD, p100 is thought to exclude this RNA from further downstream events (Figure 2.03A) in the RNAi pathway (Scadden 2005).

(A)   (B)
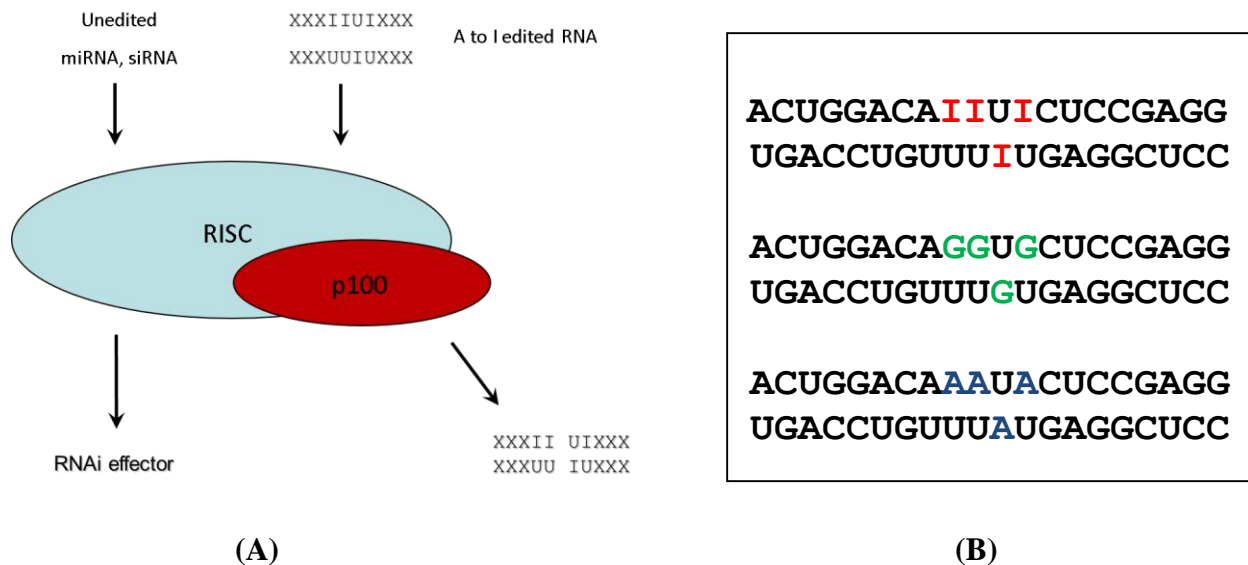
Figure 2.03 – **(A)** Depiction of the role of p100 in RNAi. If an inosine-edited RNA duplex enters RISC, it is bound and cleaved by p100. This cleavage prevents this RNA to exert its effector function in RNAi. Unedited RNA on the other hand, is not bound by p100 and is processed by RISC as usual. **(B)** The three RNA sequences used in binding studies with p100. The top sequence, from the the rat α-tropomyosin gene identified by Scadden et al., the IRD, contains four inosine residues base-paired to four uracil residues. This sequence was used by Scadden et al., and Li et al. The middle sequence, where the four inosine residues are replaced with guanines, the guanine RNA duplex (GRD), used as a control sequence for binding p100 by Scadden et al. The bottom sequence, where the four inosine residues are replaced with adenines, the adenine RNA duplex (ARD), used as a control sequence for binding p100 by Li et al.

2.1.4 Aims and scope of the project

In order to investigate the basis for the p100 discrimination towards IRD, two structural questions have to be addressed. First, characterization of such RNA, both alone and in complex with p100 would prove invaluable. So far, one duplex structure containing tandem I:U base-pairs was characterized by X-Ray crystallography (Pan, Mitra et al. 1998). Importantly, the authors claim that I:U base-pairs possess properties similar to their G:U counterparts and their presence does not strongly destabilize A-form RNA helices (Pan, Mitra et al. 1998). However, other biophysical studies using solution methods show that I:U base pairs have a much more severe effect on the thermodynamic stability of A-form RNA than G:U ones (Serra, Smolter et al. 2004).

NMR analysis of the IRD can provide useful information to characterize the structure and conformational dynamics of the IRD in solution and thus resolve these ambiguities.

Inosine-specific isotope labelling presents itself with numerous unique challenges. *In vitro* transcription with inosine triphosphate in the place of its guanine triphosphate counterpart has been performed before (Scadden and Smith 1997, Scadden and Smith 2001). However, due to the presence of both guanine and inosine residues in the RNA studied here, site-specific incorporation of inosine residues into the RNA product is impossible. Post-transcriptional RNA deamination *in vitro,* is extremely challenging as high levels of editing by ADARs that can generate the IIUI motif can only be achieved with longer RNA substrates (Scadden and Smith 1997, Scadden and Smith 2001). In addition to increasing the size of the RNA and further complicating spectral interpretation, such editing usually is not very specific and may result in the modification of other adenine residues in the edited construct (Nishikura, Yoo et al. 1991, Scadden and Smith 2001). Chemical synthesis of the corresponding phosphoramidites and their subsequent incorporation into RNA, offers an attractive alternative for specific inosine labelling. Moreover, chemical synthesis provides the versatility for the site-specific incorporation of isotopes (even of the same kind), and enables selective isotope labelling of specific atoms within the residues of interest (Wenter, Reymond et al. 2006, Kloiber, Spitzer et al. 2011, Wunderlich, Spitzer et al. 2012). For this work, a 20-mer insoine-edited RNA duplex (Figure 2.03B, top construct) which contains an isotopically labelled four inosine residues, was synthesized by collaborators. NMR spectroscopy on this labelled sample, greatly reduces spectral overlap and provides unambiguous structural information for a central inosine-containing motif in the IRD, showing that the central inosine residue adopts a non-standard orientation.

Another step in understanding the basis for p100 preference towards edited RNA, will be the interaction IRD with p100. p100 is composed of four staphylococcal nuclease (SN) domains, and one mixed SN-Tudor domain (Figure 2.04A). This domain has been shown to interact with symmetrically dimethylated arginines (Figure 2.04C) (Friberg, Corsini et al. 2009) and was also demonstrated to be dispensable for RNA binding (Li, Yang et al. 2008). The first four SN domains however were shown to bind edited, as well as the non-edited adenine RNA duplex (ARD – Figure 2.03B, bottom construct), (Scadden 2005, Li, Yang et al. 2008), in varying affinities depending on the construct used. In particular, the construct spanning SN domains three and four, was shown to interact with IRD and ARD with the affinities of 77 nM and 160 nM respectively (Li, Yang et al.

2008). Both domains were shown to be important in the interaction with the RNA as the use of a shorter, single domain construct, abolishes binding (Li, Yang et al. 2008).

The structure of a p100 construct encompassing SN domains 3 and 4, as well as the fifth SN-Tudor domain shows that the overall SN3-SN4 fold has a crescent-like shape (Li, Yang et al. 2008). A large domain-domain interface between SN3 and SN4 contributes to the overall construct fold (Li, Yang et al. 2008). The surfaces of SN3 and SN4 are basic, and co-crystallization of citrate anions in the vicinity of some of these basic patches, provides clues as to the location of the RNA binding site. When compared to the classical SN protein (Theobald, Mitton-Fry et al. 2003), the two domains overlay quite well, with the exception of differences of the loops in SN3 and the presence of an extra β-hairpin and a α-helix in SN4 (Li, Yang et al. 2008).

(A)

(B)   **(C)**
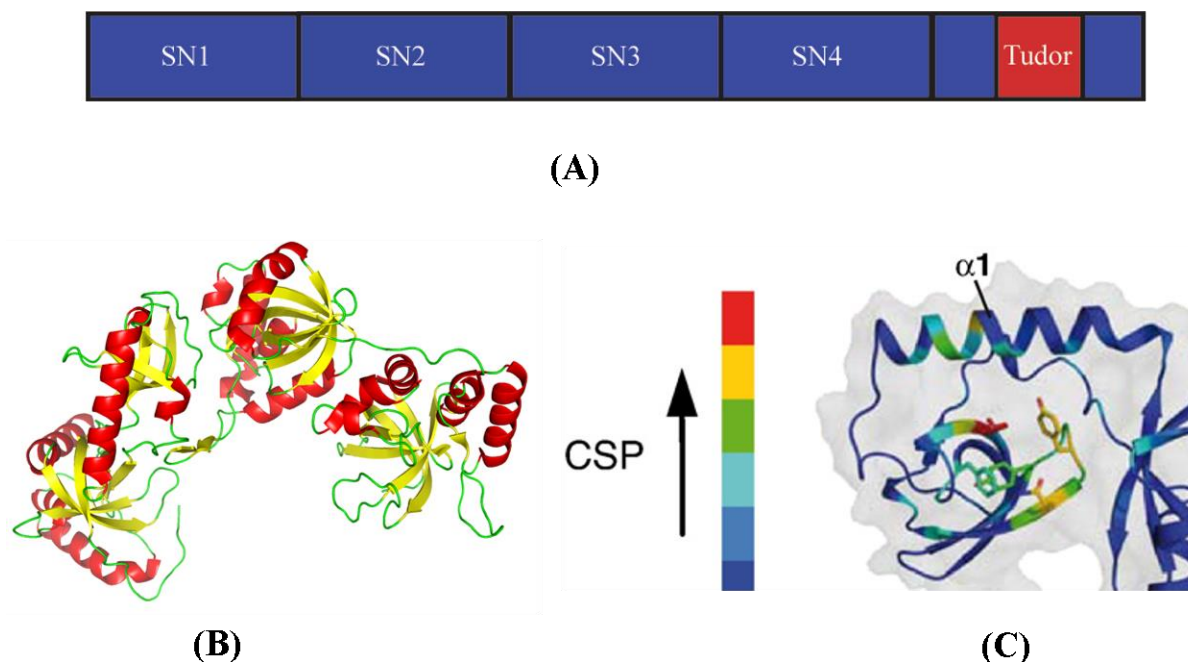
Figure 2.04 – **(A)** Domain representation of p100: Four SN domains, with a Tudor domain inserted in the middle of with a fifth SN domain **(B)** Crystal structure of the domains 3-5 of the human p100 solved by Li et al., **(C)** The interaction of SN5-Tudor domains with symmetrically dimethyated arginines, as studied by Friberg et al. Adapted from (Li, Yang et al. 2008, Friberg, Corsini et al. 2009).

For this work, the interaction between IRD and p100 was studied by NMR spectroscopy. The p100 protein construct used for this work, spans domains 3 and 4 and is from *Drosophila melanogaster* (p100DM34). NMR spectral assignments of p100 both alone and in complex with

RNA, allowed us to identify the RNA binding site in p100, as well as study the dynamics of p100 both alone and in complex with IRD using relaxation experiments. Single domain constructs p100DM3 and p100DM4 were also used to study the contribution of the SN3-SN4 interaction to the overall p100DM34 fold.

2.2 Materials and Methods

2.2.1 Experiments done on RNA

2.2.1.1 Preparation of RNA

2.2.1.1.1 Shorter inosine containing RNA homodimers and heterodimers

All RNAs were purchased from Biospring GmbH and dissolved in sterile water. If the RNA was self-annealing (palinodromic), it was heated to 95 C and then cooled at room temperature for 1.5-2 hours to promote dimer formation. In case of heterodimers, the solutions containing each the sense and anti-sense strands were mixed to produce at a 1:1 sense to anti-sense strand molar ratio. The resulting mixture was heated to 95 C and then cooled at room temperature for 1.5-2 hours to promote dimer formation.

2.2.1.1.2 Unlabelled IRD, ARD, GRD RNA

Unlabelled IRD, ARD and GRD RNAs were purchased from IBA GmbH. The final RNA amount following synthesis and purification was 400 nanomoles. The two RNA strands which comprise the RNA dimer for each construct were delivered as lyophilized powder in two separate vials. To anneal the two strands, each powder was dissolved in sterile water, and the two solutions were mixed to produce at a 1:1 sense to anti-sense strand molar ratio, for each construct. The resulting mixture was heated to 95 C and then cooled at room temperature for 1.5-2 hours to promote dimer formation.

2.2.1.1.3 Labelled IRD RNA

Specifically labelled IRD RNA was produced in the laboratory of Thomas Carell as part of a collaboration. The ^{13}C, ^{15}N labelled inosine ribophosphoramidite was first synthesized, and then incorporated into an otherwise unlabelled RNA using an ABI 394 DNA/RNA Synthesizer (Applied Biosystems). The two RNA strands which comprise the RNA dimer were delivered as lyophilized powder in two separate vials. To anneal the two strands, each powder was dissolved in sterile H_2O, and the two solutions were mixed to produce at a 1:1 sense to anti-sense strand molar ratio. The resulting mixture was heated to 95 C and then cooled at room temperature for 1.5-2 hours to promote dimer formation.

2.2.1.2 Study of RNA dimerization by native-polyacrylamide gelelectrophoresis

RNA dimerization was studied using a polyacrylamide gelectrophoresis (PAGE) at non-denaturing conditions. To prepare a 20% Acrylamide gel, in a 50 mL Falcon tube, the following solutions were mixed: 5 mL of 40% Acrylamide (19:1 Acrylamide to Bis-Acrylamide ratio), 1 mL of 10 X Tris-Boric-Acid-Ethylenediaminetetraacetic acid (TBE) buffer. The total volume was adjusted to 10 mL using sterile water. To polymerize the gel solution, 4 μL of Tetramethylethylenediamine (TEMED) and 70 μL of 10% Ammonium persulfate (APS) were added to the solution. The mixture was poured into a space between two Biorad® plates with 0.5 mm integrated spacers, capped with a 10-well forming comb and left to polymerize for one hour. Meanwhile, the RNA samples were prepared, by mixing 2 μL of approximately 100 mM RNA with 0.5 μL of 50% Glycerol. If required, NaCl or $MgCl_2$ was added to the appropriate final concentration, and if not, sterile water was added up to a final volume of 4 μL. When the gel has polymerized, it was placed in a Biorad® running chamber filled with 1X TBE and the comb was removed. Into each well, 4 μL of sample was added. In one well, a 1:1 mixture of bromophenol blue (BB) and xylene cyanol (XC) in 50% glycerol was added to monitor the migration of the RNA, as well as a size reference (Table 2.01). The gel was ran for approximately 1.5 to 2 hours at 120V at 4 C. After the run, the gel was stained with 0.3% (v/v) Toulidine Blue solution for 15 minutes, and de-stained with water for one hour.

Gel %	Bromophenol Blue	Xylene Cyanol
3.5	100bp	460bp
5.0	65bp	260bp
8.0	45bp	160bp
12.0	20bp	70bp
15.0	15bp	60bp
20.0	12bp	45bp

Table 2.01 – Migration of bromophenol blue and xylene cyanol corresponding to nucleic acid size in base-pairs in in non-denaturing gels. Adapted from: (Sambrook, Fritsch et al. 1989).

Buffers

1 X TBE (90 mM Trisbase, 90 mM Boric acid, 3 mM Ethylenediaminetetraacetic acid (EDTA)).

2.1.1.3 NMR experiments

2.1.1.3.1 NMR experiments on IRD RNA

For all NMR experiments, unlabelled or ^{15}N, ^{13}C-labelled IRD RNA was dissolved in 250 μL or 500 μL of water and 10% (v/v) ^2H_2O was added to all samples. The sample was placed in a standard (in case of 500 μL) or a Shigemi (in case of 250 μL) NMR tube.

Spectra for a 2 mM unlabeled IRD RNA sample were acquired at 278 K or 298 K in 90% H_2O+10% ^2H_2O and 100% ^2H_2O, respectively. The experiments in water were acquired on a Bruker Avance I 600 MHz equipped with a cryogenically cooled probe and on a Bruker Avance III 750 MHz spectrometer. All experiments in ^2H_2O were acquired on a Bruker Avance I 900 MHz spectrometer equipped with a cryogenically cooled probe. Spectra were processed with Topspin and NMRPipe (Delaglio, Grzesiek et al. 1995) and analyzed using NmrView (Johnson and Blevins 1994). Experimental acquisition parameters are described in Table 2.02. For NOESY experiments in ^2H_2O and H_2O (Kumar, Ernst et al. 1980), five spectra each were acquired with 50 ms, 100 ms, 150 ms, 200 ms and 300 ms mixing times respectively. Two TOCSY experiments (Griesinger, Otting et al. 1988) were acquired with 40 ms and 80 ms mixing times respectively.

Spectrum	Solvent	Time domain size (F2/F1)	Acquisition times F2/ F1 (ms)	Sweepwidth F2/F1 (kHz)	No. of scans	Total experiment time (h)
NOESY	^2H_2O	2048/512	113.9/28.5	8.99/8.99	16	5.0
TOCSY	^2H_2O	2048/256	67.0/8.4	15.29/15.29	16	1.25
HSQC	^2H_2O	8192/512	525.9/16.2	7.79/15.85	96	21
NOESY	H_2O	2048/512	75.8/18.9	13.52/13.52	16	5.5
SF-HMQC	H_2O	1642/128	49.9/28.1	16.45/2.28	6144	60

Table 2.02 – Outline of all experiments conducted on an unlabelled IRD RNA, outlining the experiment name, solvent, points in the various time domains, acquisition times, sweepwidth, number of scans and total experiment time.

Spectra of a 0.5 mM labelled IRD were acquired at 298 K in ^2H_2O on a Bruker Avance III 800 MHz spectrometer equipped with a cryogenic probe (filtered/edited NOESY spectra) or on a Bruker Avance III 600 MHz spectrometer equipped with a cryogenic probe (HCCH-TOCSY/COSY spectra). Spectra were processed with Topspin and NMRPipe (Delaglio, Grzesiek et al. 1995) and analyzed using NMRVIEW (Johnson and Blevins 1994). For the assignment of the IRD ribose, HCCH-COSY and HCCH-TOCSY were acquired. Acquisition parameters

described in Table 2.03 were used. For the heteronuclear NOESY spectra, a uniform mixing time of 100 ms was used.

Spectrum	Time domain size (F3/)F2/F1	Acquisition. times (F3/)F2/F1 (ms)	Spectral width (F3/)F2/F1 (kHz)	No. of scans	Total Exp. time (h)
HCCH-COSY	1536/48/176	159.7/39.8/18.3	4.81/6.04/4.81	32	93
HCCH-TOCSY	1536/48/176	159.7/39.8/18.3	4.81/6.04/4.81	32	93
HSQC	2048/256	160.2/6.4	6.39/20.12	16	1
Edited NOESY	1536/256	120.1/20.0	6.40/6.40	256	44
Edited-Edited NOESY	1536/256	120.1/20.0	6.40/6.40	256	44
Edited-Filtered NOESY	1536/256	120.1/20.0	6.40/6.40	256	44

Table 2.03 – Outline of all experiments conducted on a labelled IRD RNA, outlining the experiment name, points in the various time domains, acquisition times, sweepwidth, repetition delays, number of scans and total experiment time.

2.1.1.3.2 NMR experiments on ARD and GRD RNA

For all NMR experiments, unlabelled ARD and GRD RNA were dissolved in 275 µL 90% H_2O+10 % (v/v) ^2H_2O. Each sample was placed in a Shigemi NMR tube. A 2D NOESY of the unlabeled ARD and GRD RNA were acquired at 278 K in 90% H_2O+10% ^2H_2O on a Bruker Avance I 600 MHz spectrometer equipped with a cryogenic probe with the same parameters as for the unlabelled IRD indicated in Table 2.02, with the exception of the of the sweepwidth for F2 and F1 which was set to 14.1 kHz and 14.0 kHz, respectively. Spectra were processed and assigned as above (2.1.1.3.1).

2.1.1.3.3 NMR experiments on 12mer palindromic RNAs

The RNA samples were dissolved in 500 µL 90% H_2O+10 % (v/v) ^2H_2O. Each sample was placed in a standard NMR tube. 2D NOESY experiments were acquired at 278 K on a Bruker Avance I 750 MHz spectrometer with the following parameters: 3072/512 time domain, 87.8/14.5 ms acquisition time, 17.5/17.6 kHz sweepwidth, 250 ms mixing time, 96 scans for a total experiment time of 96 hours. The NOESY spectra were processed and assigned as above (2.1.1.3.1).

2.1.1.4 Ultra Violet (UV) melts

For UV melts, IRD, ARD and GRD RNAs, an OD_{260} of approximately 0.5 were used. Each RNA-containing solution was placed in a 3 mL quartz cuvette, which was placed in an AVIV model 14, UV-vis spectrophotometer with temperature control was used. The RNA OD_{260} was observed as the sample temperature was raised from 5 C to 95 C, at a one degree increment. The resulting data was plotted as melting curve and the melting temperature, T_m, was extracted. The experiment was repeated with increasing NaCl concentration for each RNA.

2.2.2 Experiments on p100

2.2.2.1 Cloning of p100DM34

2.2.2.1.1 Construct and primer design

All constructs except p100HS34, were cloned into the plasmid pETZZ-1a (Appendix A1). This plasmid contains an N-terminal Histidine tag (His-tag) which is as a binding target in nickel affinity chromatography and a ZZ-tag (tandem IgG binding domain of protein A) used for enhancement of protein solubility. p100HS34 was cloned into the plasmid pETM11 (Appendix A2) which contains an N-terminal His-tag. Both plasmids encode for an amino acid sequence located downstream of their tags, which serves as a cleavage target for Tobacco Etch Virus (TEV) protease used in the purification process to cleave those tags off (2.2.2.3.2). All construct, plasmid, primer, restriction enzyme, DNA and protein sequences are listed in Appendix B.

2.2.2.1.2 Amplification of p100DM34 insert by PCR

The DNA fragment corresponding to the coding sequence of p100DM34 was amplified using the Polymerase Chain Reaction (PCR). The following components were used in the reaction mix:

Component	Volume	Final Conc.
Template/primer mix		
Plasmid DNA template	1 μL	1 ng/μL
Forward Primer	0.5 μL	2.0 μM
Reverse Primer	0.5 μL	2.0 μM
Master Mix		
10 X Pfu Buffer	5 μL	1 X
10 mM (each) dNTPs mix	1 μL	0.2 mM
Go Taq Polymerase	0.2 μL	1 unit
Pfu Polymerase	0.5 μL	1.5 units
ddH$_2$O	41.3 μL	N/A
Total Volume	50 μL	N/A
Pfu Polymerase Buffer: 20 mM Tris-HCl, 10 mM KCl, 2 mM MgSO$_4$,10 mM (NH$_4$)$_2$SO$_4$, 0.1% Triton® X-100 and 0.1 mg/ml nuclease-free BSA, pH = 8.8.		

Note: The 10 mM dNTP solution was prepared by mixing 2 μL of 100 mM of each dNTP (dATP, dTTP, dGTP, dCTP) and diluting the resulting mixture with 12 μL ddH$_2$O to a total volume of 20 μL.

The two mixes were prepared separately and combined together, in a PCR tube which was placed in a PCR thermocycler. The following PCR program was used:

Step type	Temperature	Time	No. of cycles
Initial Denaturation	95 ºC	4 minutes	1
Denaturation	95 ºC	30 seconds	
Annealing	52.5 ºC	40 seconds	25
Extension	72 ºC	1 minute	
Final Extension	72 ºC	10 minutes	1

To the PCR product 12.5 μL of 5X Gel loading dye was added and the resulting 62.5 μL solution was loaded on a 1% agarose gel.

2.2.2.1.3 PCR Product purification using agarose gelectrophoresis

The running buffer used for agarose gelectrophoresis was 1X TAE. The buffer was prepared as a 50X stock and diluted 50 fold for running conditions, as required. In a 200 mL Erlenmeyer flask, 0.5 grams of agarose and 1 X TAE were added. The resulting 1% mixture was heated in a microwave for 2-3 minutes, until the agarose dissolved. Once the solution was cooled down to about 40 ºC, 6 μL of ethidium bromide were added, and the mixture was poured into a gel caster with a well comb and was left for 20-30 minutes to allow it harden. Once the gel has solidified, it was placed in the gel running chamber filled with 1X TAE running buffer and the comb was removed. The entire PCR product was loaded on the gel in one well. In another, 5 μL of 100 bp ladder was loaded for size reference. The gel was then run at 80 V for approximately 30-40 minutes. The PCR product was then illuminated under UV light, excised and transferred to a pre-weighed Eppendorf tube. The mass of the gel-piece was then determined, and for every 10 mg of the piece 10 μL of membrane binding solution was added. The gel piece and the solution in the tube were heated at 60 C for 10 minutes, rotating at 1000 RPM. The plasmid DNA was then isolated using a Promega Wizard® SV Gel PCR Clean-up Kit.

2.2.2.1.4 Digestion of p100DM34 PCR product by restriction endonucleases

The purified p100DM34 PCR product was cut by the restriction endonucleases NcoI and Acc65I. For the cleavage reaction, 50 μL of plasmid 5 μL of 10X NEB Buffer 3 (50 mM TrisHCl, 100 mM NaCl, 10 mM $MgCl_2$, 100 μg/mL BSA, pH = 7.9 – All concentrations are at 1 X), and 1 μL (5 units) of each enzyme. The mixture was incubated at 37 C shaking at 1300 RPM for three

hours. The enzymes were then heat-inactivated by increasing the temperature to 65 C and incubating the sample for 20 minutes at 1300 RPM.

2.2.2.1.5 Digested PCR Product isolation using agarose gelectrophoresis

The same procedure was followed as at 2.2.2.1.3.

2.2.2.1.6 Isolation of plasmid DNA

A glycerol stock containing the *E. coli* strain XL1 carrying the plasmid was spread on a Lysogeny Broth (LB) agar plate containing 50 μg/mL kanamycin, and the plate was incubated at 37 C overnight. A clearly distinguished colony was inoculated into a starting culture consisting of 5 mL of rich medium and 50 μg/mL of kanamycin. The resulting mixture was incubated for 12-16 hours at 37 C shaking at 200 RPM to an OD_{600} value of an approximately 2.0. The next morning, the cells in the culture were pelleted by centrifugation at 4000 RPM for 20 min at 4 C using a swinging bucket rotor in the Eppendorf centrifuge. The plasmid DNA was then isolated using a Promega Wizard® Miniprep DNA Purification Kit.

2.2.2.1.7 Restriction digest and plasmid 5' dephosphorylation

The plasmid pETZZ-1a, was cut by the restriction endonucleases NcoI and Acc65I. The same protocol as for the p100DM34 digestion was used (2.2.2.1.4). After heat inactivation, 1 μL of alkaline phosphatase was added and the resulting mixture was incubated 37 C shaking at 1300 RPM for one hour.

2.2.2.1.8 Separation of the large and small plasmid fragments using agarose gelectrophoresis

A 1% agarose gel was prepared as described earlier (2.2.2.1.3). The entire digest mixture was loaded in one lane on the gel. In another, 5 μL of 100 bp ladder was loaded for size reference. The gel was then run at 80 V for approximately one hour. The top plasmid product was excised under UV light, and transferred to a pre-weighed Eppendorf tube. The mass of the gel piece was then determined, and for every 10 mg of the piece 10 μL of membrane binding solution was added. The gel piece and the solution in the tube were heated at 60 C for 10 minutes, rotating at 1000 RPM. The plasmid DNA was then isolated using a Promega Wizard® SV Gel PCR Clean-up Kit.

2.2.2.1.9 Ligation of p100DM34 insert into pETZZ-1a

In a 1.5 mL Eppendorf tube, 2.5 μL of cut plasmid, 1.5 μL of p100DM34 insert, 0.75 μL of 10X NEB Ligase Buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT, pH = 7.5 – all concentrations are 1X), 2.75 μL of sterile water and 1 μL (3 units) of T4 DNA ligase were mixed. The resulting mixture was incubated at 16 C overnight.

2.2.2.1.10 Transformation of the ligated pETZZ-1a/p100DM34 into XL1 E. Coli

The ligated solution was mixed with 80 μL of XL1 *E. coli* cells. Following a 10 minute incubation on ice, the mixture was subjected to a 40 second heat shock at 42 C (shaking at 1300 RPM). The sample was then returned on ice for five minutes. Then, 700 μL of LB medium were added and the sample was incubated at 37 C, shaking at 1300 RPM for one hour. Following incubation, the sample was centrifuged in a benchtop centrifuge at 13000 RPM for 10 minutes. The supernatant was decanted, and the pellet was dissolved in 100 μL of fresh LB. The cells were then plated on an LB Agar plate containing 50 μg/mL kanamycin. The plates were incubated at 37 C overnight.

2.2.2.1.11 p100DM34 Insert screening

The next day, distinct colonies were picked from the transformation plate. These colonies were used to inoculate a 5 mL culture containing 50 μg/mL kanamycin. The resulting mixture was incubated for 12-16 hours at 37 C shaking at 200 RPM to an OD_{600} value of an approximately 2.0. The next morning, the cells in the culture were pelleted by centrifugation at 5000 RPM for 20 min at 4 C using a swinging bucket rotor in the Eppendorf centrifuge. The plasmid DNA was then isolated using a Promega Wizard® Miniprep DNA Purification Kit. The plasmid was digested with the same enzymes used to remove its internal fragment (2.2.2.1.3) and the mixtures were ran on a 1% agarose gel using the same protocol as above. Where an insert corresponding the expected size of p100DM34, was observed (Figure 2.05), the colony from which this DNA was isolated, and sent for sequencing.
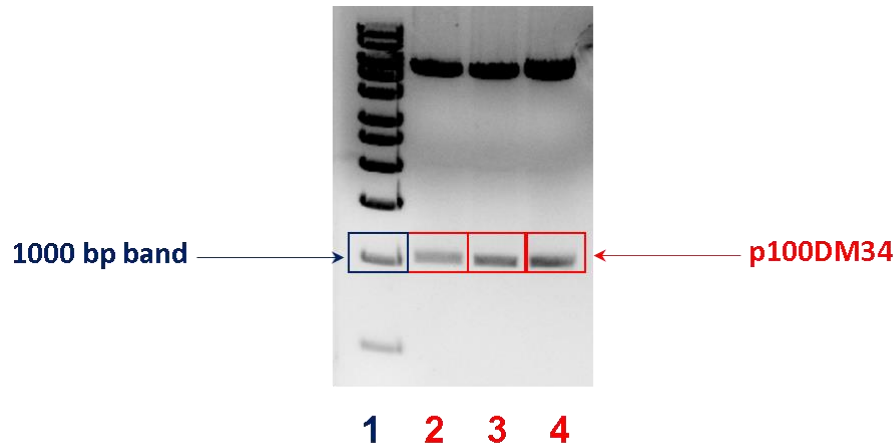
Figure 2.05 – 1 % agarose gel showing a restriction digest screen. From left to right, Lane 1: 1Kb ladder marked in blue; lanes 2-4, show the restriction digest of DNA isolated from three colonies screened for the p100DM34 fragment (marked in red). All three digestions show a large plasmid fragment at the top of the gel and a smaller fragment just above the 1000 bp band of the marker (marked in a blue rectangle) corresponding in size to the expected molecular weight of p100DM34.

2.2.2.1.12 Transformation of pETZZ-1a /p100DM34 into BL21 (DE3) E. Coli

In a 1.5 mL Eppendorf tube, of 1 μL of the plasmid was mixed with 20 μL of BL21 (DE3) *E. coli* cells. Following a 10 minute incubation on ice, the mixture was subjected to a 40 second heat shock at 42 C (shaking at 1300 RPM). The sample was then returned on ice for five minutes. Then, 400 μL of LB medium were added and the sample was incubated at 37 C, shaking at 1300 RPM for one hour. Following incubation, the sample was centrifuged in a benchtop centrifuge at 13000 RPM for 10 minutes. The supernatant was decanted, and the pellet was dissolved in 100 μL of fresh LB. The cells were then plated on an LB Agar plate containing 50 μg/mL kanamycin. The plates were incubated at 37 C overnight.

2.2.2.1.13 Mutagenesis of p100DM34 via Quickchange PCR

The plasmid containing DNA fragment corresponding to the coding sequence of wild type p100DM34 was used as a template for a Quickchange mutagenesis reaction using the primer carrying the necessary mutations in their sequences. Sequences for the mutagenesis primers are outlined in Appendix C. The reaction mix was as follows:

Component	Volume	Final Conc.
Template/primer mix		
Plasmid DNA template	1.0 μL	~ 1 ng/ μL
Forward Mutated Primer	1.0 μL	2.0 μM
Reverse Mutated Primer	1.0 μL	2.0 μM
Master Mix		
10 X Pfu Buffer	5.0 μL	1 X
10 mM (each) dNTPs mix	1.0 μL	0.2mM
Go Taq Polymerase	0.2 μL	1 unit
Pfu Polymerase	0.5 μL	1.5 units
DMSO	5.0 μL	10%
ddH$_2$O	35.3 μL	N/A
Total Volume	50 μL	N/A

Note: The 10 mM dNTP solution was prepared by mixing 2 μL of 100 mM of each dNTP (dATP, dTTP, dGTP, dCTP) and diluting the resulting mixture with 12 μL ddH$_2$O to a total volume of 20 μL.

Step type	Temperature	Time	No. of cycles
Initial Denaturation	95 ℃	2 minutes	1
Denaturation	95 ℃	1 minute	
Annealing	59-52 ℃	1 minute	18
Extension	68 ℃	12 minutes	
Final Extension	68 ℃	12 minutes	1

Following Quickchange PCR, the template plasmid was digested by adding 1 μL of Dpn I to the PCR mixture and then incubating it for 3 hours at 37 C. Dpn I was then heat inactivated by incubation at 80 C for 20 C.

To check for the presence of the new plasmid, the sample was loaded and ran on a 1% agarose gel as described previously (2.2.2.1.3). The DNA product was illuminated under UV light. If the plasmid product was detected, the plasmid was transformed into XL1 *E. Coli* cells as described previously (2.2.2.1.10). The solution was plated on an LB agar plate with 50 μg/mL kanamycin which was incubated for 12-16 hours. The next day, 2-3 colonies were picked from the LB agar plate and used to inoculate a 5 mL culture with 50 μg/mL kanamycin which was incubated for 12-16 hours at 37 C shaking at 200 RPM till an approximate OD$_{600}$ of 2.0. The next day the DNA was isolated using a Promega Wizard® Miniprep DNA Purification Kit and sent for sequencing.

2.2.2.1.14 Cloning of p100HS34, p100DM3 and p100DM4

Cloning of p100HS34, p100DM3 and p100DM4 was carried out using the same protocols as for p100DM34, with two changes. EcoRI was used as the second restriction endonuclease to

digest inserts and vectors the construct were cloned into instead of Acc65I. Also, p100HS34 was cloned into a plasmid pETM-11 (Appendix A2).

2.2.2.2 Expression of p100 protein constructs

2.2.2.2.1 Expression of unlabelled p100DM34 in rich (LB) medium

A clearly distinguished colony was inoculated into a starting culture consisting of 50 mL of rich medium and 50 μg/mL of kanamycin. The resulting mixture was incubated for 12-16 hours at 37 C shaking at 200 RPM to an OD_{600} value of an approximately 2.0. The next morning, the culture was transferred into a growth flask containing the 1 L rich growth medium. The cells were grown at 37 C shaking at 200 RPM to an OD_{600} value of 0.9 and expression was induced with 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). The induced culture was incubated for 16 hours overnight at 20 C. The next day, the cells were harvested by centrifugation at 5,000 RPM for 20 min shaking at 200 RPM at 4 C using the SLC-6000 rotor in the Sorval centrifuge.

Rich growth medium

Note: All amounts and concentration listed here, are for a 1 L solution and 1X final concentration.

1g Bacto Tryptone, 0.5g Yeast Extract, 0.5g NaCl, 50 μg/mL kanamycin

2.2.2.2.2 Expression of ^{15}N labelled p100DM34 in minimal medium

A clearly distinguished colony was inoculated into a starting culture consisting of 50 mL of rich medium and 50 μg/mL of kanamycin. The resulting mixture was incubated for 12-16 hours at 37 C shaking at 200 RPM to an OD_{600} of an approximately 2.0. The next morning, the cells in the culture were pelleted by centrifugation at 4,000 RPM for 20 min at 4 C using a swinging bucket rotor in the Eppendorf centrifuge. The supernatant was decanted and the resulting pellet was re-suspended in approximately 50 mL of minimal growth medium. The re-suspended mixture was transferred into the growth flask containing the 1 L minimal growth medium. The cells were grown at 37 C shaking at 200 RPM to an OD_{600} value of 0.9 and expression was induced with 0.5 mM IPTG. The induced culture was incubated 16 hours overnight at 20 C. The next day, the cells were harvested by centrifugation at 5,000 RPM for 20 min at 4 C using the SLC-6000 rotor in the Sorval centrifuge.

Minimal growth medium

Note: The solutions marked in red and blue are prepared as 10X (M9 medium) and 100X (trace elements) stocks respectively. For preparation of the medium, 100 mL of the M9 medium and 10 mL of trace elements solution were used. All concentrations listed here are final concentrations for 1X solution.

35 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.5 mM NaCl, 9 mM $^{15}NH_4Cl$, 0.17 mM Ethylenediaminetetraacetic acid (EDTA), 3 μM $FeCl_3$ x $6H_2O$, 6 μM $ZnCl_2$, 0.76 μM $CuCl_2$ x $2H_2O$, 0.60 μM $CoCl_2$ x $6H_2O$, 1.62 μM H_3BO_3, 0.068 μM $MnCl_2$ x $6H_2O$, 0.4% (w/v) ^{12}C glucose, 1 mM $MgSO_4$, 0.3 mM $CaCl_2$ (always the last component to be added, after dilution to 1 L), 1 μg/mL Biotin, 1 μg/mL Thiamine, 50 μg/mL kanamycin. Bring up the total volume to 1 L with sterile water. The final pH should be 7.5.

2.2.2.2.3 Expression of ^{13}C, ^{15}N labelled p100DM34 in ^2H_2O medium

A clearly distinguished colony was inoculated into a starting culture consisting of 50 mL of rich medium and 50 μg/mL of kanamycin. The resulting mixture was incubated for 12-16 hours at 37 C shaking at 200 RPM to an OD_{600} value of an approximately 2.0. The next morning, the cells in the culture were pelleted by centrifugation at 5,000 RPM for 20 min at 4 C using a swinging bucket rotor in the Eppendorf centrifuge. The supernatant was decanted and the resulting pellet was re-suspended in approximately 50 mL of ^2H_2O growth medium. The re-suspended culture was grown for 12-16 hours at 37 C shaking at 200 RPM. The next morning, the culture was transferred into a growth flask containing the 1 L ^2H_2O growth medium. The cells were grown to an OD_{600} value of 0.9 and expression was induced with 0.5 mM IPTG. The induced culture was incubated 16 hours overnight at 20 C. The next day, the cells were harvested by centrifugation at 5,000 RPM for 20 min at 4 C using the SLC-6000 rotor in the Sorval centrifuge.

^2H_2O growth medium

Note: The solution marked in blue is prepared as and 800X (trace elements). For the preparation of the medium, 1.25 mL of trace elements solution was used. All concentrations listed here, are final concentrations for 1X solution.

35 mM Na_2HPO_4, 22 mM KH_2PO_4, 8.5 mM NaCl, 9 mM $^{15}NH_4Cl$, 0.17 mM EDTA, 3 μM $FeCl_3$ x $6H_2O$, 6 μM $ZnCl_2$, 0.76 μM $CuCl_2$ x $2H_2O$, 0.60 μM $CoCl_2$ x $6H_2O$, 1.62 μM H_3BO_3, 0.068

μM MnCl$_2$ x 6H$_2$O, 2 mg/L ^{13}C-C6 deuterated glucose, 1 mM MgSO$_4$, 0.3 mM CaCl$_2$ (always the last component to be added, after dilution to 1 L), 1 μg/mL Biotin, 1 μg/mL Thiamine, 50 μg/mL kanamycin. Bring up the total volume to 1 L with sterile ^{2}H$_2$O. The final pH was 7.5.

2.2.2.2.4 Expression of p100HS34, p100DM3 and p100DM4

All three protein constructs were expressed in LB and minimal media, using the same protocol as for p100DM34.

2.2.2.2.5 Expression of p100DM34 mutants

All p100DM34 mutants were expressed in minimal media, using the same protocol as for p100DM34.

2.2.2.3 Purification of p100DM34 p100HS34, p100DM3, p100DM4 and p100DM34 mutants

2.2.2.3.1 Cell Breakage

Following cell harvesting, the pellet was re-suspended in ~20 mL of Nickel Column (NC) Buffer. If previously frozen, the cell pellet was thawed on ice. A small amount of lysozyme powder and approximately 0.01 mg/mL of DNAse I were added and the sample was left incubating for 20-30 minutes at room temperature. Protease inhibitors were then added in 1:1000 volume ratios (i.e 20 μL for a 20 mL pellet). The cells were then lysed by being passed three times through a French Pressure Cell at 1.8 kbar that was pre-washed with the NC buffer at 1.4 kbar. Following cell disruption, soluble and insoluble cellular fractions were separated by centrifuging at 15,000 RPM at 4 C for 45-50 minutes in the Sigma centrifuge.

2.2.2.3.2 Nickel Column Purification and TEV Cleavage

After centrifugation, the supernatant was loaded on the nickel column that has been prewashed twice with 20 mL of NC buffer. The supernatant was then loaded again to maximize protein binding to the column, and the flowthrough is collected and kept. The protein on the column was then washed with 90 mL of NC Buffer containing 2 M NaCl, to remove non-specifically bound nucleic acids. The column was washed twice with 20 mL of NC Buffer and the protein is then eluted with NC buffer containing 300 mM imidazole. A sodium dodecyl sulfate Polyacrylamide gel electrophoresis (SDS-PAGE) was ran to ensure the fusion protein is present in

the elution fraction. There should be a band at ~55 kDa between the third and fourth marker bands from the top. The fusion protein was then concentrated to a volume of 2.5 mL or slightly less in a 30,000 Molecular Weight cutoff (MWCO) concentrator using the Eppendorf centrifuge at 4 C at 4,000 RPM. If the volume of the protein sample is less than 2.5 mL it was toped up to this volume with NC Buffer. The protein was then loaded on to a PD-10 buffer exchange column that was prewashed twice with 10 mL of TEV Cleavage (TC) Buffer. The protein was then eluted with 4.5 mL of TC buffer into a 50 mL Falcon tube. Another 15.5 mL of TC Buffer were added for a total volume of 20 mL. Two vials of TEV protease per one nickel column preparation were added to the protein sample and the mixture was incubated overnight at 4 C. The PD-10 column was washed twice with 10 mL of water for long term storage. The next day the mixture (Figure 2.06, lanes 2-4) was loaded on the nickel column that has been pre-washed twice with 20 mL of TC buffer. The sample was passed over the nickel column three times to make sure that the His-tag/ZZ-tag fusion protein (His-tag alone for p100HS34) as well as TEV protease will bind to the column as much as possible. The flowthrough (where the p100 protein should now be located) was collected in a 50 mL Falcon tube (Figure 2.06, lanes 5-7). The His-tag/ZZ-tag and TEV were then eluted with 20 mL TC Buffer containing 300 mM imidazole (Figure 2.06, lanes 8-10). An SDS-PAGE was ran to confirm that TEV cleavage was successful and that the p100 protein is indeed located in flowthrough, which also had to be free of any visible His-tag/TEV protease bands (those should be present in the elution sample). The gel band for p100 should be about 38 kDa in size, located between the fourth and fifth marker bands (Figure 2.06, lane 1) from the top (Note: For RNAse assays outlined in 2.2.2.7, the purification of unlabeled p100DM34 was completed at this step, all other protein constructs were subjected to additional steps outlined below).
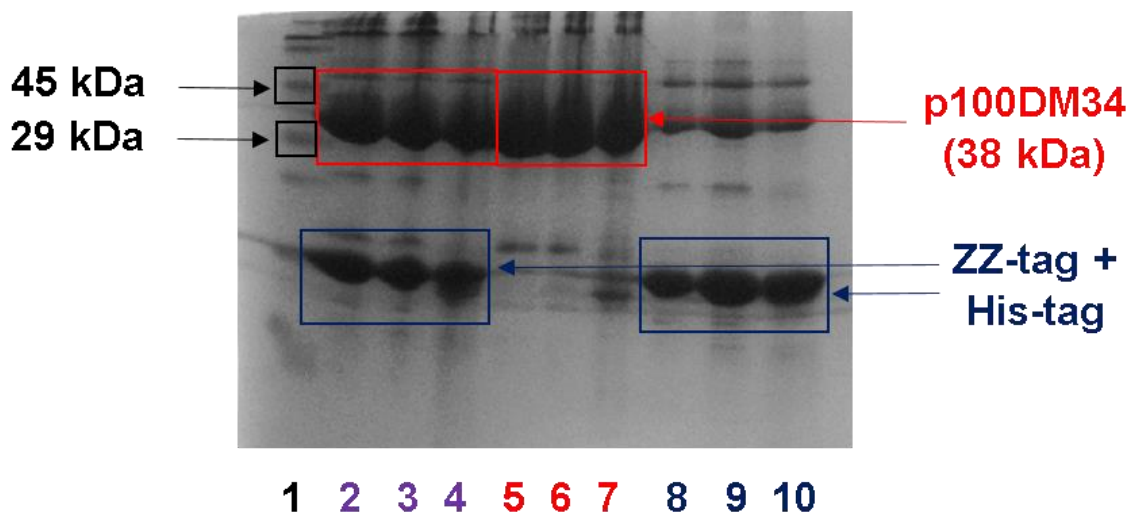
Figure 2.06 – An SDS-PAGE of the p100DM34 construct after cleavage with TEV protease. Lane 1: Protein molecular weight marker, the 29 kDa and 45 kDa, are boxed in black. Lanes 2-4: p100DM34 after TEV protease cleavage before loading on the nickel column. The p100 protein bands are boxed in red, while the cleaved ZZ-tag and His-tag are boxed in blue. Lanes 5-7: p100DM34, after a second nickel column purification in the flowthrough, boxed in red. Lanes 8-10: The cleaved ZZ-tag and His-tag in the elution, boxed in blue.

The protein sample (flowthrough) was then concentrated to 2.5 mL or slightly less using a 10,000 MWCO concentrator (prewashed with 10 mL of TC Buffer for 10 minutes at 4,000 RPM at 4 C) in an Eppendorf centrifuge at 4 C at 4,000 RPM and exchanged on PD-10 column into 4.5 mL of NMR Buffer in a manner similar to that of NC to TC buffer. The protein was then concentrated to 2 mL using a 10,000 MWCO concentrator and placed into a 2 mL Eppendorf tube.

2.2.2.3.3 Gel filtration chromatography

The S75 16/60 gel filtration column was pre-washed with one column volume of water and then with one column volume of the NMR buffer. To save time, this step was normally performed the night before, so that the column will be ready for the run the next morning. The protein sample was spun down in a benchtop centrifuge for five minutes at 13,000 RPM at room temperature, to pellet any precipitate. Meanwhile, a 2 mL syringe was washed first with water, then with NMR buffer, and used to clean a 2 mL loading loop into which the sample would be injected. A fraction collector with 15 mL Falcon tubes is be placed on the fraction collector tray. Following centrifugation, the sample was transferred to the syringe and carefully injected into the system. The run program was then started, fractionation would not begin until after 0.3 column volumes

(40-50 mL) of buffer are passed through. The protein eluted approximately at 70 mL (Figure 2.07A).
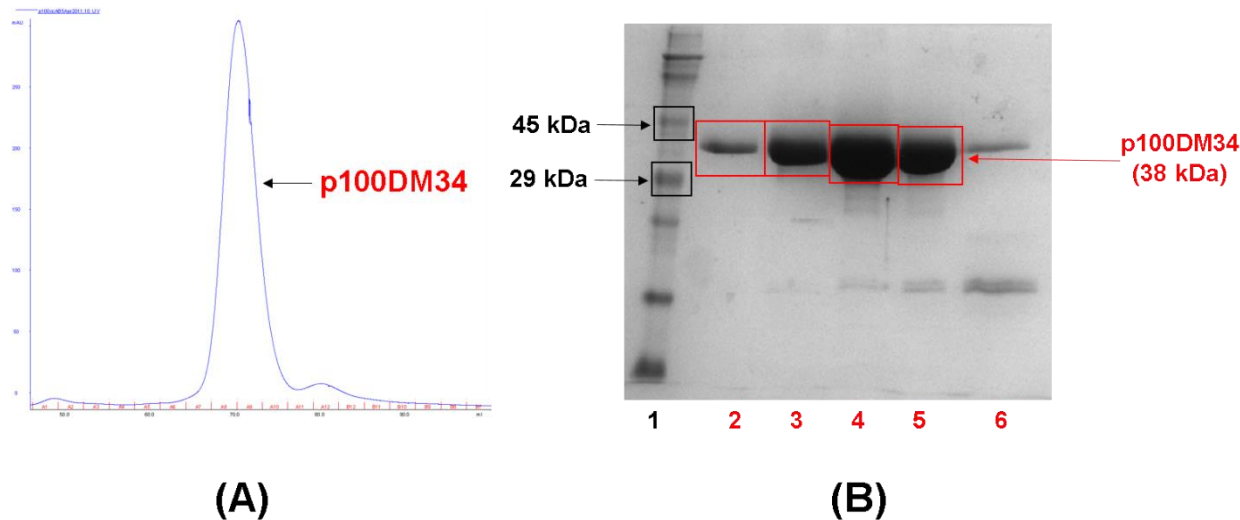

Figure 2.07 – (A) p100DM34 elution profile from the S75 16/60 gel filtration column. The protein elutes at approximately 70 mL. (B) An SDS-PAGE with the fractions containing p100DM34 eluted from the gel filtration column. Lane 1: Protein molecular weight marker, the 29 kDa and 45 kDa, are boxed in black. Lanes 2-5: The p100DM34 protein in the fractions eluted from the gel filtration column, boxed in red.

From each fraction a 20 μL sample was taken for SDS-PAGE which was then run to confirm the presence of the protein (MW ~ 38 kDa, Figure 2.07B). The column was then washed with one column volume of water and one column volume of 20% Ethanol for final storage. The protein fractions are concentrated to a final volume of 0.5 mL using 10,000 MWCO concentrator in an Eppendorf centrifuge at 4 C at 4,000 RPM. The purified protein sample was then transferred to a 1.5 mL Eppendorf tube and 25 μL (5% by volume) $^{2}H_2O$ were added. The final protein concentration was quantified using the NanoDrop.

Buffer compositions

NC Buffer

20 mM TrisHCl, 100 mM NaCl, 5 mM BME (β-mercaptoethanol), 10 mM imidazole, pH = 7.55

20 mM TrisHCl, 2 M NaCl, 5 mM BME (β-mercaptoethanol), 10 mM imidazole, pH = 7.55

20 mM TrisHCl, 100 mM NaCl, 5 mM BME (β-mercaptoethanol), 300 mM imidazole, pH = 7.55

TC Buffer

20 mM Trisbase, 300 mM NaCl, 1 mM TCEP, 10 mM imidazole, pH = 8.00

20 mM Trisbase, 300 mM NaCl, 1 mM TCEP, 300 mM imidazole, pH = 8.00

NMR Buffer 1

50 mM phosphate, 300 mM NaCl, 1 mM DTT, pH = 6.50

NMR Buffer 2.1

20 mM HEPES, 50 mM NaCl, 5 mM BME, pH = 7.00

NMR Buffer 2.2

20 mM HEPES, 50 mM NaCl, 5 mM BME, 10 mM EDTA, pH = 7.00

Note: All buffers were filtered and degassed prior to use

2.2.2.3.4 Purification of the p100HS34, p100DM3, p100DM4 and p100HDM34 mutants

For purification of p100HS34, p100DM3, p100DM4 and p100DM34 mutants, the same protocol was used as for p100DM34.

2.2.2.4 NMR spectroscopy of p100DM34

For all NMR experiments a ^{2}H, ^{15}N, ^{13}C (triple labelled – TL) p100DM34 was dissolved in 250 μL or 500 μL of one of the NMR buffers and 5% (v/v) ^{2}H$_2$O was added to all samples. The sample was placed in a standard (in case of 500 μL) or a Shigemi (in case of 250 μL) NMR tube. All experiments were performed on a Bruker Avance I 900 MHz spectrometer equipped with a cryogenic probe.

2.2.2.4.1 p100DM34 backbone experiments

An initial 2D Transverse relaxation optimized spectroscopy – Heteronuclear single quantum coherence (TROSY-HSQC)(Pervushin, Riek et al. 1997) experiment was acquired to ensure sample viability. Then, the following three-dimensional experiments were acquired for p100DM34 backbone assignment: HNCA (Salzmann, Pervushin et al. 1998, Eletsky, Kienhofer et al. 2001), HNCACB (Salzmann, Pervushin et al. 1999, Eletsky, Kienhofer et al. 2001), and HNCO (Salzmann, Pervushin et al. 1998). The experimental parameters used are outlined in Table 2.03. All data was processed by NMRPipe (Delaglio, Grzesiek et al. 1995) and analysed in Sparky (Goddard, Kneller et al.) and NMRView (Johnson and Blevins 1994). Analysis of p100DM34 torsion angles was performed by TALOS+ (Shen, Delaglio et al. 2009).

Spectrum	Time domain size (F2/F1)	Acquisition times F2/ F1 (ms)	Sweepwidth F2/F1 (kHz)	No. of scans	Total experiment time (h)
HSQC	2048/256	71.2/40.9	14.3/3.1	16	0.6
HNCA	2048/64/128	71.2/10.8/9.7	14.3/2.9/6.6	24	60
HNCACB	1536/64/172	70.0/10.8/6.5	11.9/2.9/13.2	24	60
HNCO	1536/64/64	70.2/10.9/8.3	11.0/2.9/3.85	16	21
Note	All the above experiments are TROSY-based, and were carried out in 90% $H_2O + 10$ ^2H_2O				

Table 2.03 – List of experiments performed on a TL p100DM34 sample for backbone assignment, outlining experiment name, time domain size, acquisition times, sweepwidth, number of scans and experiment time.

After peak assignment, signal-to-residue correlations were obtained using the program MARS (Jung and Zweckstetter 2004) . Resonance assignment were then verified manually, and the program was re-ran with corrected assignments, as required.

2.2.2.4.2 NMR studies on the p100DM34/IRD complex

To create the protein/RNA complex, a 500 µM or 100 µM TL p100DM34 sample was dissolved with IRD RNA of the same concentration in a total volume of 5 mL. The resulting mixture was co-concentrated to a final volume of 250 µL using 10,000 MWCO concentrator at 4000 RPM at 4 C, using a swinging bucket rotor in an Eppendorf centrifuge. To the sample to be measured 10% (v/v) ^2H_2O was added, and the resulting solution was transferred to a Shigemi NMR tube. The experiments used to study both the free and the IRD-bound p100DM34 were TROSY-HSQC and TROSY-HNCO (Salzmann, Pervushin et al. 1998). They were processed and analyzed the same as above (2.2.2.4.1).

2.2.2.4.3 Calculation of p100DM34 secondary chemical shifts

Secondary chemical shifts were calculated in order to predict the secondary structure of p100DM34. For the calculation of the secondary chemical shift, the following formula was applied:

$$\Delta\delta C = (\Delta C\alpha_{observed} - \Delta C\alpha_{randomcoil}) - (\Delta C\beta_{observed} - \Delta C\beta_{randomcoil})$$

$\Delta\delta C$ represents the combined secondary chemical shift, $\delta C\alpha$ and $\delta C\beta$ is the respective chemical shift of the of Cα and Cβ resonances of the protein backbone. The $\delta C\alpha_{observed}$ and $\delta C\beta_{observed}$ are

the values extracted from p100DM34 assignments. As chemical shifts are strongly dependent on the identity of the amino acid residue, the observed chemical shifts were normalized by subtraction of a random coil chemical shift ($\delta C\alpha_{randomcoil}$ and $\delta C\beta_{randomcoil}$) that reflects the amino acid-specific component, based on the Wishart random coil shift database (Wishart and Sykes 1994). For cysteine residues, the values of the reduced state were used.

2.2.2.5 NMR spectroscopy of p100DM34 mutants

All four p100DM34 mutants wer dissolved in 500 μL of NMR buffer 1 and 5% (v/v) ^2H_2O was added to all samples. All samples were placed in standard NMR tubes. For R379E, K447E and R572E mutants, measurements were performed on a Bruker Avance I 600 MHz spectrometer equipped with a room temperature probe, while for S548A, a Bruker Avance I 500 MHz spectrometer equipped with cryogenically cooled probe, was used. The experiment performed on all four samples was a 2D ^1H, ^{15}N HSQC (Bodenhausen and Ruben 1980) with: 1024/256 time domain, 48.7/47.9 ms acquisition time, 10.5/2.7 kHz sweepwidth, 16 scans for a total experiment time of 40 minutes. The experiments were processed as analyzed the same as in 2.2.2.4.1.

2.2.2.6 Study of IRD RNA/p100DM34 interactions by Electrophoretic mobility shift assay

For electrophoretic mobility shift assay (EMSA), non-denaturing PAGE gels of various acrylamide percentages, as well as acrylamide and bis-acrylamide ratios were used. A summary of all gel conditions is provided below (Table 2.04).

Gel %	Acrylamide:Bis-acrylamide ratio	TBE
10%	19:1, 29:1	1 X
15%	19:1, 29:1, 55:1	1 X
20%	19:1, 29:1, 37.5:1 55:1	1 X

Table 2.04 – Outline of the gel percentages and acrylamide/bis-acrylamide ratios used for p100DM34/IRD EMSAs.

Appropriate amounts of protein and RNA were mixed in a 1.5 mL Eppendorf tube. Sodium chloride, magnesium chloride, were added as appropriate for a total volume of 20 μL. As required,

some samples were incubated at 37 C shaking at 200 RPM for various time periods, others left at room temperature or at 4 C. Gel casting and running protocols were the same as above (2.2.1.2).

2.2.2.7 RNAse tests on p100HS34 and p100DM34 mutants

Following a second nickel column, unlabelled p100HS34 was subjected to an RNAse test using the Ambion RNaseAlert® Lab Test Kit. The samples to be tested were incubated at 37 C overnight. The next day they were then illuminated under UV light, to detect the presence and the extent of RNAse activity. As controls, samples with RNAse-free water and RNAse A and the p100HS34 fusion tag which was cleaved by TEV protease (2.2.2.3.2) were used.

2.3 Results

2.3.1 Length and sequence requirements for inosine-containing RNA dimerization

Since the preparation of an isotopically labelled inosine-containing RNA via *in vitro* transcription is impossible due to a lack of sequence-specific incorporation ability of inosine into a nascent RNA chain, preliminary NMR and native PAGE studies were attempted on a 12-nucleotide homodimeric RNA (I12HO) construct (Figure 2.08A), as well as on a non-edited RNA where the two inosine residues are replaced with two adenines (A12HO – Figure 2.08B). It was thought that by using this shorter RNA (I12HO), structure determination can be attempted without the need for isotopic labelling.
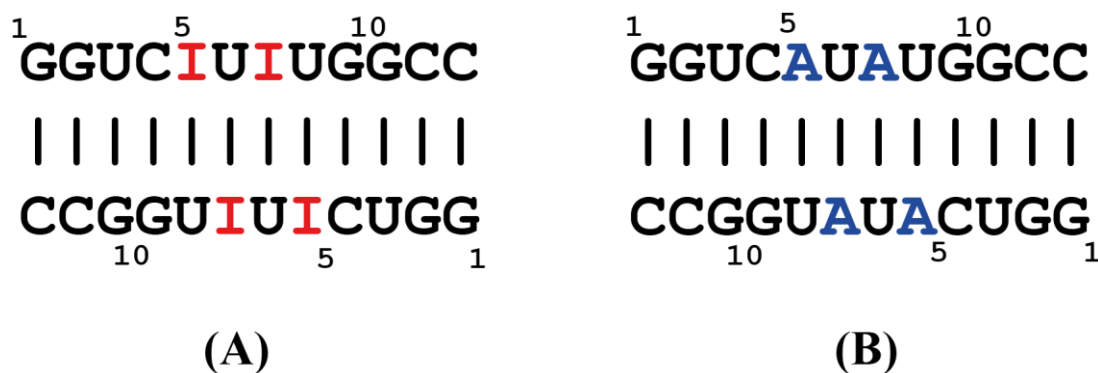

Figure 2.08 – (A) A 12 nucleotide palindromic inosine duplex, containing four IU base-pairs. All inosine residues are marked in red **(B)** A control construct where the inosine residues, are replaced with adenine. The adenine residues replacing their inosine counterparts, are marked in blue.

A2D ^{1}H, ^{1}H water NOESY of I12HO, did not give rise to any exchangeable imino signals for the inosine-containing region of this RNA (Figure 2.09A). The imino-imino walk could only be performed from G2 to G9 that is in the regions flanking the two IU base-pairs (Figure 2.09B). This is an indication that at the very least, there is a deviation from A-form geometry in the I12HO's inosine containing region. The same experiment of A12HO, has given rise to imino resonances for every base-pair in the RNA (Figure 2.09C – Since this RNA is palindromic, only half the imino-imino cross-peaks are observed, while the other half directly overlaps), indicated that this RNA is completely helical in its entirety (Figure 2.09D). Further analysis by native-PAGE, has shown that I12HO migrates significantly faster than its A12HO counterpart (Figure 2.10A). Based on the NMR and the native-PAGE results, it is concluded that I12HO is not a dimer but is rather a monomer in solution. Given the fact that this RNA's imino-imino connectivities for its canonical portion are observed, the exact conformation is probably that of a hairpin (Figure

2.10B). As this conformation is not biologically relevant, due to the fact that p100 interacts with inosine-edited duplexes, this construct was not used for further structural studies.

(A)

(B)

(C)

(D)
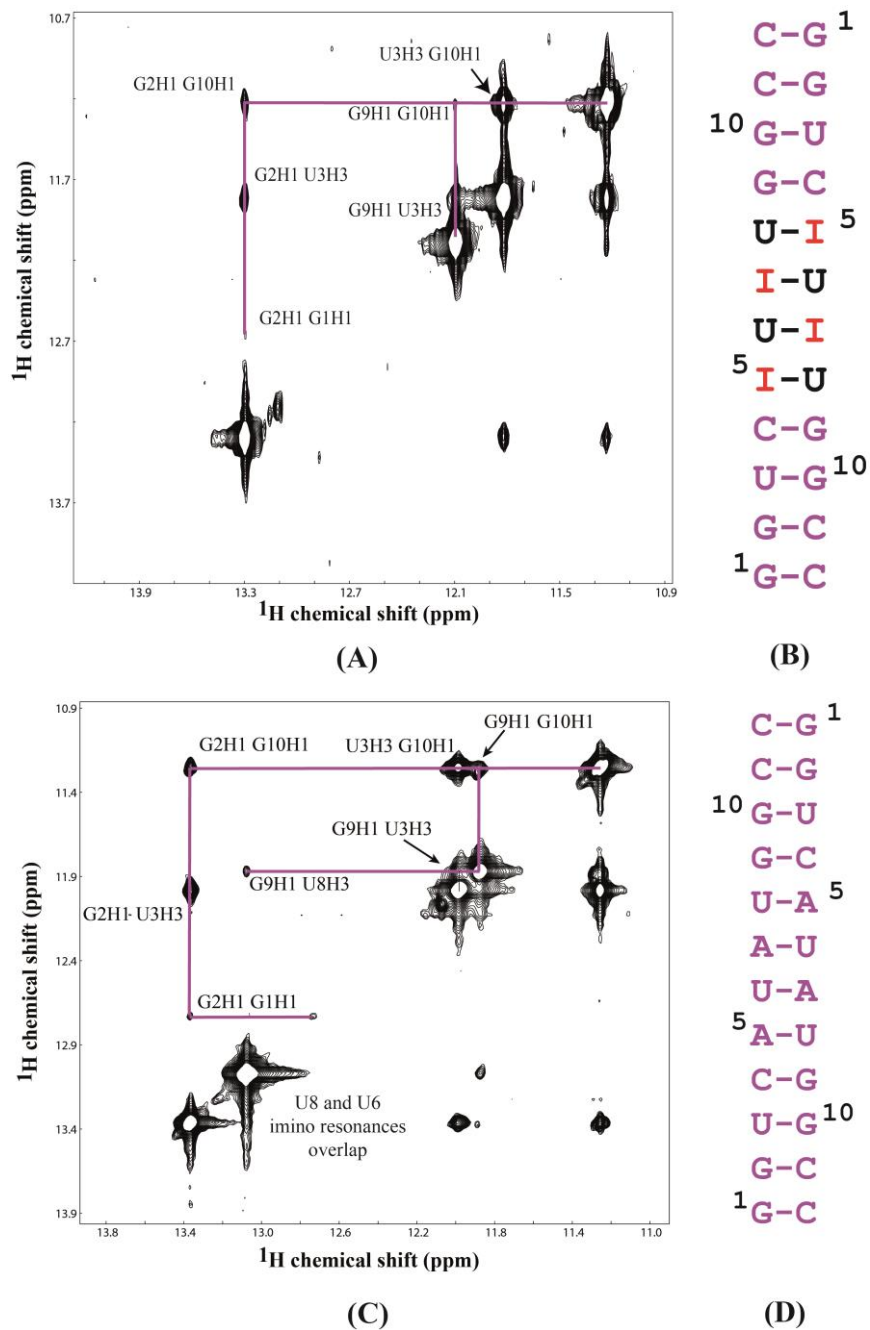
Figure 2.09 – **(A)** The imino-imino region of 2D water NOESY spectrumI12HO in showing imino resonances for the regions flanking the four IU base-pairs, (and the connectivities between them, marked in purple lines) indicating that this region is helical. No imino signals are observed for the IU portion of the RNA. **(B)** Regions of I12HO which adopt helical geometry are marked in purple. **(C)** The imino-imino region of 2D water NOESY of A12HO in showing imino resonances for every base-pair in the sequence (and the connectivities between them marked in purple lines). **(D)** Regions of A12HO which adopt helical geometry are marked in purple.
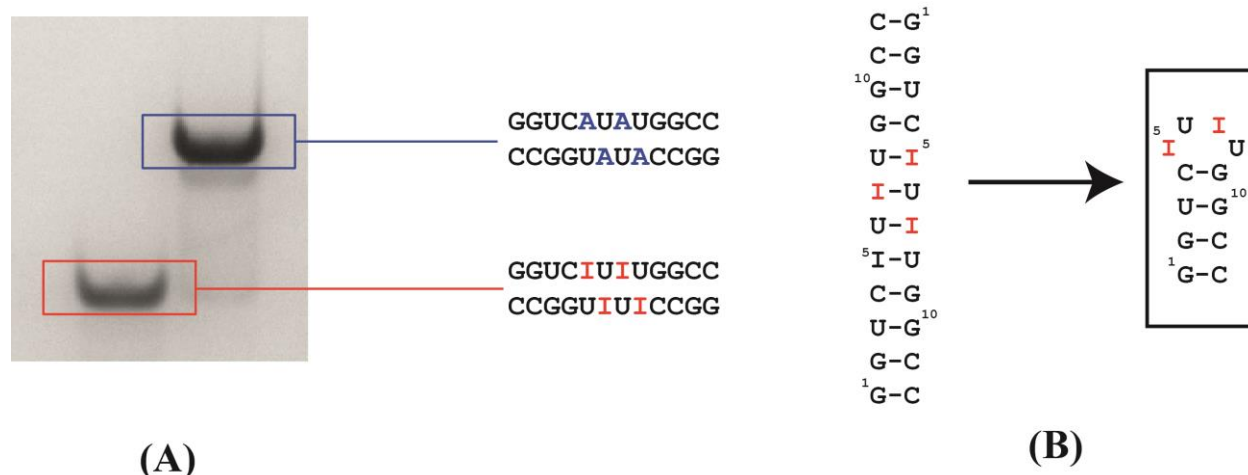
81

Figure 2.10 – **(A)** A 20% Acrylamide native-PAGE, showing the fast migration of I12HO, compared to A12HO, which migrates slower in the gel **(B)** I12HO adopts a monomeric hairpin conformation.

A number of additional palindromic inosine containing RNA constructs were then studied by native-PAGE in order to investigate their ability to dimerize. In one such construct, the flanking GU wobble was replaced with a canonical GC base-pair (I12HOGC – Figure 2.11, sequence is denoted above lanes 2-5). It was thought that increased stability of the flanking regions would promote dimerization more readily. Another longer, 16 nucleotide construct (I16HO – Figure 2.11, sequence is denoted above lanes 6-10) was also used, with the expectation, that increasing the size of the flanking regions would also induce dimerization. This time various concentration of monovalent (NaCl) as well as divalent (MgCl$_2$) salts were also added to the RNAs, to further promote dimerization.

I12HOGC has failed to dimerize under any of the given conditions as it migrates the same way as I12HO (Figure 2.11, lanes 2-5). Dimer formation was observed for I16HO, on the other hand, however only to approximately 30%-40% completion at 0.25 M and 0.75 M NaCl (Figure 2.11, lanes 7 and 8). This dimer-monomer mixture is not amenable for NMR structural studies, where one homogeneous species is a necessary perquisite.

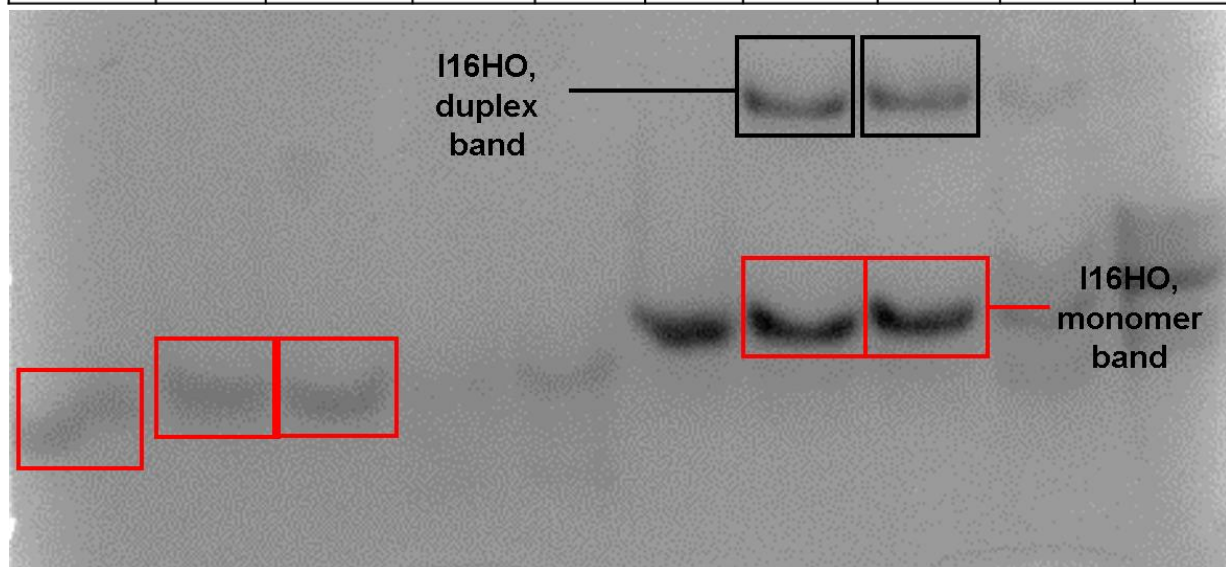| I12HO	GGCCIUIUGGCC CCGGUIUICCGG				GCGGUCIUIUGGCCGC CGCCGGUIUICUGGCG				
1	2	3	4	5	6	7	8	9	10
N/A	N/A	0.75 M	0.1 M	0.5 M	N/A	0.25 M	0.75 M	0.1 M	0.5 M
None	None	NaCl	MgCl$_2$	MgCl$_2$	None	NaCl	NaCl	MgCl$_2$	MgCl$_2$

Figure 2.11 – A 20% Acrylamide native-PAGE of I12HOGC and I16HO RNAs. Dimerization of both RNAs were studied at with no salt, 0.75 M NaCl, 0.1 M MgCl$_2$, and 0.5 M MgCl$_2$ for I12HOGC (lanes 2-5) and with no salt, 0.25 M NaCl, 0.75 M NaCl, 0.1 M MgCl$_2$, and 0.5 M MgCl$_2$ for I16HO (lanes 6-10). A sample of I12HO was also included as a control (lane 1).

The 20meric IRD RNA studied by Scadden et al., dimerizes readily without addition of any salt (Figure 2.12A. lane 3). However since this RNA is heterodimeric, this would require an assignment of all forty residues (while in a case of a homodimeric RNA, only half of the residues would have to be assigned, because of overlapping chemical shifts for its two parts), making structure determination challenging even with isotope labelling. It was therefore attempted to test the dimerization of a GC rich 20mer inosine-containing RNA palindrome (I20HO). However this RNA did not dimerize at any of the experimental conditions listed below (Figure 2.12B). Hence, not only size but also sequence requirements are important for dimerization of inosine-containing RNAs.

It was then decided to use IRD for further structural and functional studies. While complex for NMR studies, this RNA holds the advantage of also being the biologically relevant construct, in its interaction with p100 as identified earlier (Scadden 2005, Li, Yang et al. 2008).
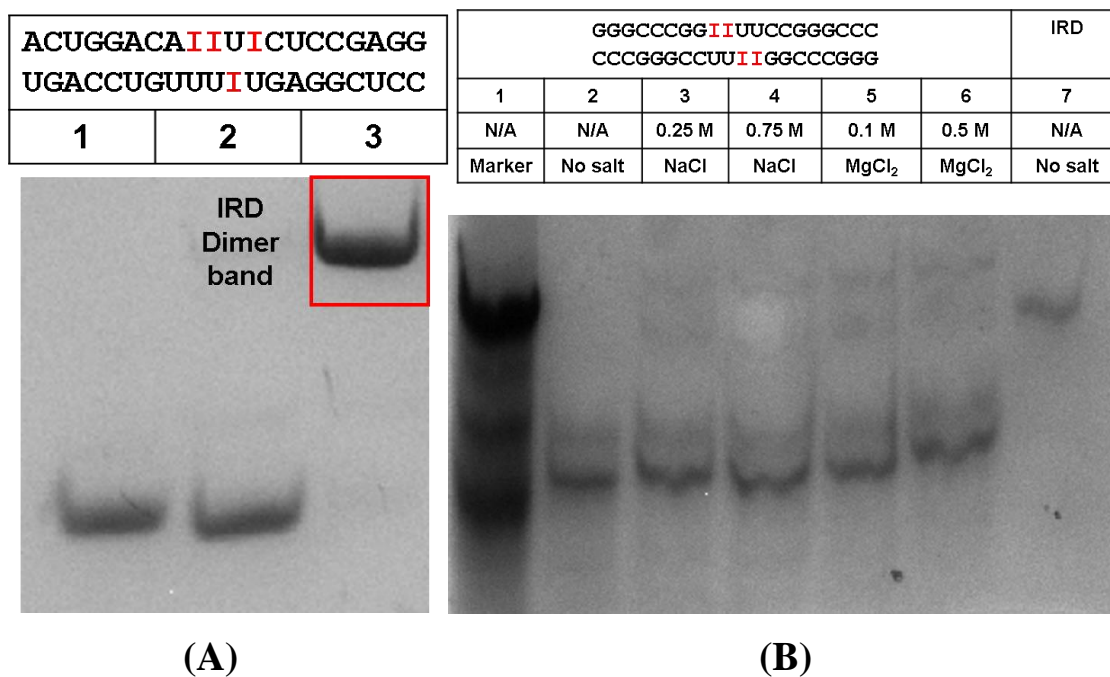
ACUGGACAIIUICUCCGAGG		
UGACCUGUUUIUGAGGCUCC		
1	2	3

GGGCCCGGIIUUCCGGGCCC						IRD
CCCGGGCCUUIIGGCCCGGG						
1	2	3	4	5	6	7
N/A	N/A	0.25 M	0.75 M	0.1 M	0.5 M	N/A
Marker	No salt	NaCl	NaCl	$MgCl_2$	$MgCl_2$	No salt

(A)   (B)

Figure 2.12 – (A) A 20% Acrylamide native-PAGE showing the IRD dimer formation (lane 3), upon the mixture of the top (lane 1) and bottom (lane 2) strands. **(B)** A 20% Acrylamide native-PAGE showing I20HO RNA, at various NaCl (lanes 3-4) and $MgCl_2$ (lanes 5-6), as well as at salt-free (lane 2) conditions. In all lanes, I20HO migrates significantly faster than IRD RNA (lane 7) which is known to dimerize.

2.3.2 UV melts on IRD, GRD and ARD RNA

To study the differences in duplex stabilities of IRD RNA versus its ARD and GRD counterparts, UV melting experiments were performed. The absorbance at 260 nm for each RNA was observed at one degree increments from 5 C to 95 C. As the temperature increases, the dimer melts to yield two monomeric strands, resulting in an increase in absorbance. This increase occurs at different temperatures depending on the structure and stability of the RNA in question.

When subjected to melting, IRD duplex exhibited the least amount thermal stability compared to its counterparts. Its melting temperature, (T_m) was found to be 35 C (Figure 2.13A). ARD RNA was found to be more stable as its T_m was determined to be 42 C (Figure 2.13B). Similar melting temperature of 41.5 C was also observed for GRD (Figure 2.13C). To investigate the effect of salt on the T_m of the three duplexes, the measurements were repeated at 20 mM and 200 mM NaCl. The melting temperatures of all three RNAs have risen with increasing salt concentration (Table 2.05).
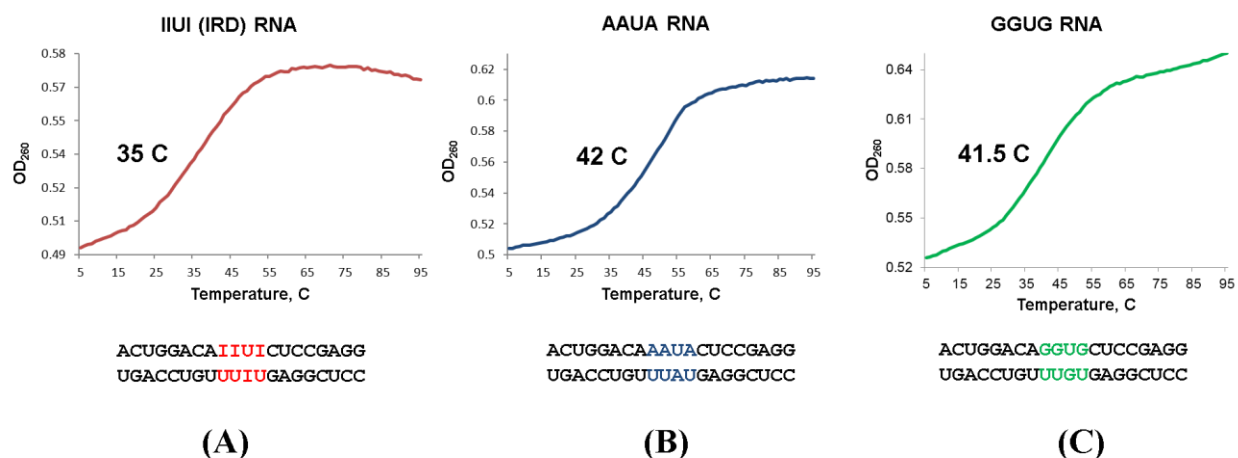
Figure 2.13 – Melting curves for (**A**) IRD, (**B**) ARD and (**C**) GRD RNA. The T_m for each RNA is indicated.

[NaCl], mM	T_m of IRD, C	T_m of ARD, C	T_m of GRD, C
0.00	34.00	42.00	41.50
20.00	50.90	65.50	56.80
200.00	64.20	80.20	71.50

Table 2.05 – Melting temperatures of IRD, ARD, GRD, in salt-free, at 20 mM and 200 mM NaCl.

2.3.3 NMR analysis of unlabelled IRD

Once the ability of IRD to dimerize was confirmed, we proceeded to conduct structural studies on an unlabelled IRD sample by NMR.

2.3.3.1 2D NOESY in water and comparison to ARD and GRD

Similar to the shorter RNAs, the first experiment performed on IRD was the 2D water NOESY, in order to gain insight into this RNA's secondary structure. The imino-imino portion of this spectrum (Figure 2.14A) gave rise to signals for parts of the RNA up to one base-pair below (G39 to G34) and above (G28 to G19) the inosine-containing motif. No imino resonances were observed for U40 and G20 due to terminal base-pair fraying (Varani and Tinoco 1991, Varani, Aboul-ela et al. 1996).

Consequentially, an imino-imino walk could be performed in the regions flanking the inosine containing motif, that is from the C2-G39 to C7-G34 (Figure 2.14A, and 2.14B – marked

in red on the spectrum and the IRD sequence) and C13-G28 to G19-C22 base-pairs (Figure 2.14A, and 2.14B – marked in brown on the spectrum and the IRD sequence). No imino signals were observed for any of the inosine residues or their uracil base-pairing partners. These findings indicate that portions of IRD which flank the inosine-containing motif adopt an A-form helix with standard Watson-Crick base-pairing (Figure 2.14B) while the latter region shows a deviation from this conformation.
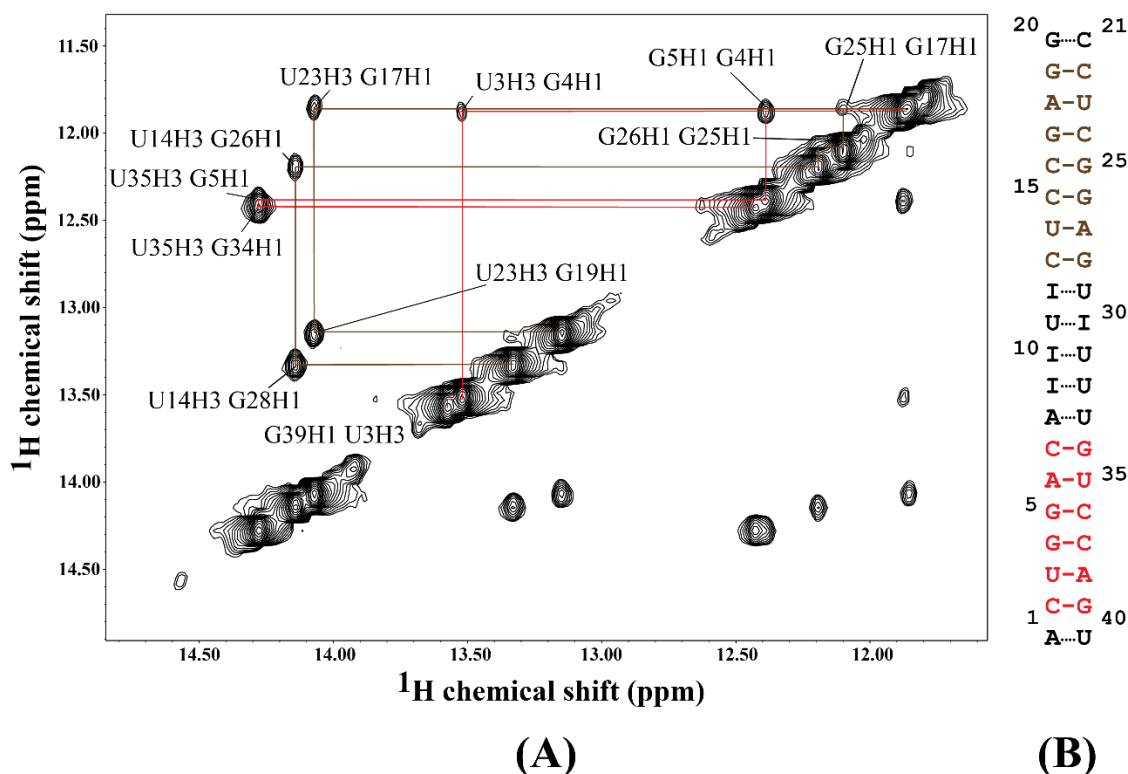

(A)　　　　　　　　　　　　　　　　　**(B)**

Figure 2.14 – **(A)** The IRD 2D water NOESY imino-imino regions showing the cross-peaks for the regions flanking the inosine containing motif. The imino-imino walk for the base-pairs below and above the inosine containing motif, are marked in red and in brown respectively. **(B)** IRD secondary structure with the base-pairs for which the imino-imino walk could be performed are marked with the same color code as in the NOESY spectrum.

In order to further investigate the changes which IU base-pairs impart on the helical conformation of the IRD, the imino-imino portion of the water NOESY for this RNA was compared to that of ARD and GRD (Figure 2.15A). The stem signals of all three RNA constructs largely overlap (boxed signals in Figure 2.15A), indicating that their stem structure is likely to be similar, only showing deviation at the resonances flanking their central motifs.

In contrast to IRD however, all imino signals (except, like in the IRD, the terminal imino ones) were observed for the ARD, including the four central AU base-pairs. The imino-imino walk

could be conducted along the entire length of this RNA (marked in blue in Figure 2.15A and 2.15B, center sequence), indicating that this RNA is helical in its entirety. All GRD imino signals were also observed (except, like in the IRD and ARD the terminal ones), however, no imino-imino connectivities could be made between the U11-G30 and G12-U29 base-pairs (spectrum and base-pairs for which imino resonances were observed are marked in green in Figure 2.15A and 2.15B, right construct). The lack of imino-imino connectivity between these two base-pairs suggests that a kink is formed is formed in this part of the GRD which, disrupts its helical geometry, resulting in the imino resonances in the two base-pairs to be located further than five angstroms.
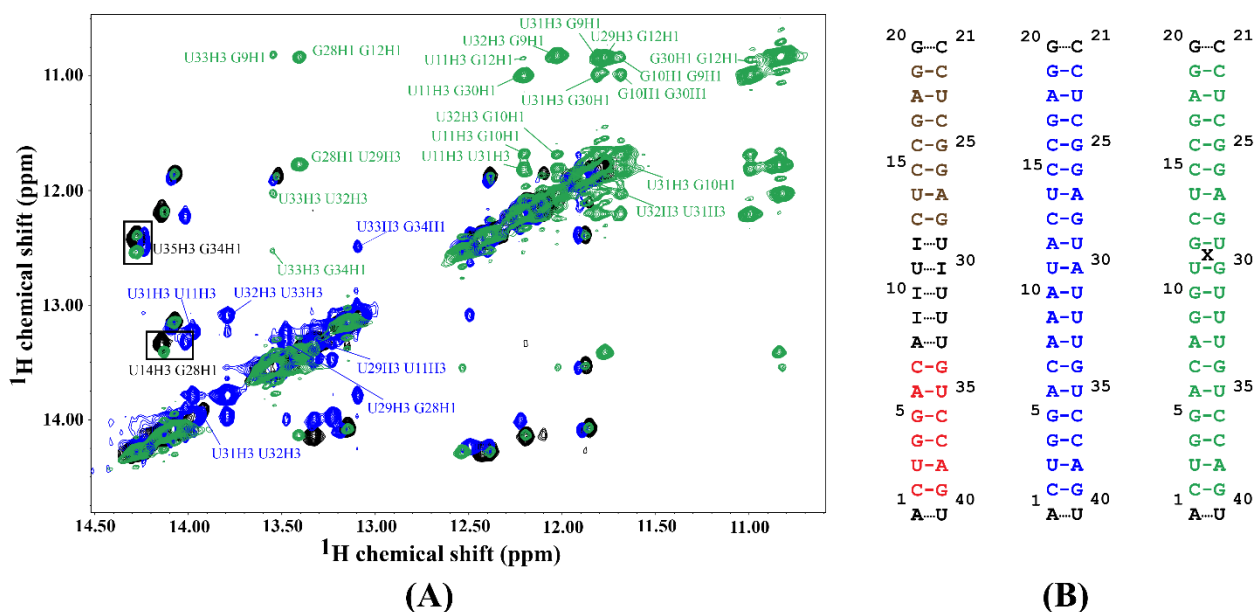

Figure 2.15 – **(A)** An overlay of the imino-imino region of 2D water NOESY of IRD, ARD, and GRD, marked in black, blue, and green respectively. Assignments of the ARD and GRD central regions, as well as the regions flanking them in all three constructs are denoted with the appropriate colour coding. Signals from the residues flanking the central motifs in all three RNA constructs, are marked in a box. **(B)** Secondary structures of IRD (left sequence) with the same colour coding as in Figure 2.14B compared with ARD (centre sequence) and GRD RNAs (right sequence). Regions where the imino resonances were observed are marked in colors other than black. The two base-pairs in the GRD which could not be connected using the imino-imino walk, are marked with an X in between.

2.3.3.2 Assignment of the aromatic/anomeric resonances of IRD

Using the H8 or H6 to H1' correlations in a 2D ^{1}H, ^{1}H NOESY acquired in ^{2}H$_2$O and the with the help of 2D ^{1}H, ^{1}H TOCSY and 2D ^{1}H, ^{13}C HSQC experiments (refer to sections 1.2.3.2.3 and 1.2.3.2.4 for more details), the non-exchangeable ^{1}H signals in IRD were successfully assigned

(Figure 2.16) The aromatic-anomeric "walk" (Fürtig, Richter et al. 2003) was performed where possible, specifically along the helical IRD regions flanking the inosine-containing motif. It was once again confirmed that regions between A1/U40 to A8/U33 and C13/G28 to G20/C21 assume a standard A-form helical conformation. These findings were also supported by assignments of the adenine H2 resonances, signals in the aromatic-aromatic region of the ^2H_2O NOESY, as well as water NOESY spin diffusion cross-peaks.

Some assignments could also be done on parts of the inosine containing motif – the aromatic anomeric "walk" was used to unambiguously determine the H8 (or H6)/H1' intra as well as inter-residue connectivities of I9, I12, on the top strand, and U29, I30 on the bottom strand indicating that at least this part of the inosine-containing motif possesses a geometry resembling A-form RNA.

Another important finding was an absence of sharp H8 and H6 signals in this part of the NOESY spectrum. Such signals would indicate that the residue giving rise to this signal may extrude from the RNA helix (Johnson and Donaldson 2006, Beribisky, Tavares et al. 2007). A lack of those kinds of signals suggests that no residues in IRD stick out from the helix and all of them point into the RNA. In addition, lack of strong intra-residue H8 (or H6) /H1' cross-peaks indicate that all residues in IRD assume the more common *anti* rather than *syn* conformation.

2.3.3.3 Non-standard sugar puckers in the IRD's inosine-containing motif

To examine the RNA for the presence of non-standard sugar pucker conformations, a 2D ^1H, ^1H ^2H_2O TOCSY with a short-mixing time was used. In this spectrum (Figure 2.17A), H1'/H2' correlations were observed for A8, I10, U11, I30, U31, U32, U33 as well as U34. All these residues are located in or near (in case of A8 and U33) the inosine containing motif (Figure 2.17B). These results strongly suggest yet again that the inosine-containing motif shows a marked deviation from standard helical geometry, as the majority of the residues there adopt the less common, *C2' endo* conformation besides I9, I12 and U29. All other residues (with the exception of A1 and U40, which adopt non-standard sugar puckers due to fraying in the RNA ends) adopt the standard *C3' endo* sugar puckers. The sugar pucker conformations for the inosine residues, were also later partially confirmed by the calculation of the can1 value (Fürtig, Richter et al. 2003) derived from the chemical shifts of C1', C4' and C5' atoms (2.3.4.5).
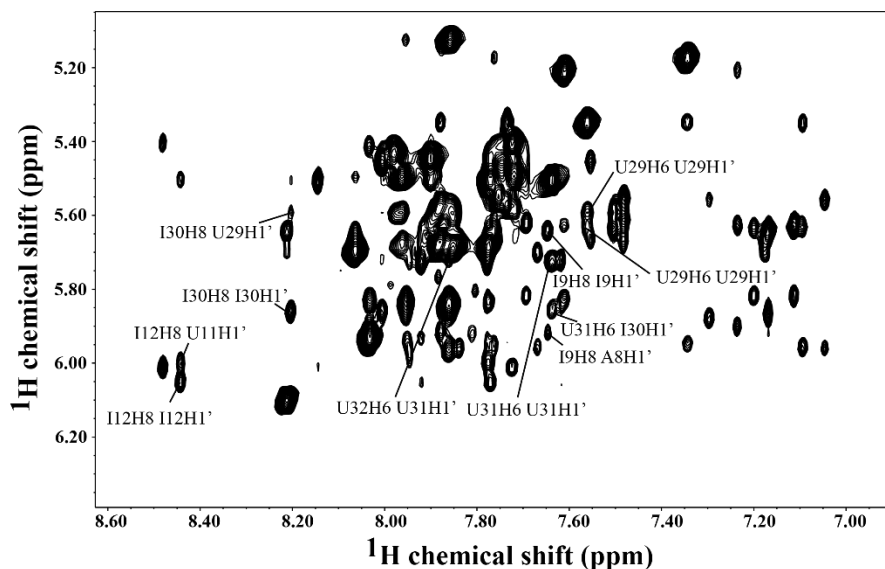
Figure 2.16 – The 2D ^{1}H, ^{1}H ^{2}H$_2$O NOESY of IRD, the aromatic-anomeric region. Important assignments of the IIUI motif are denoted on the spectrum.
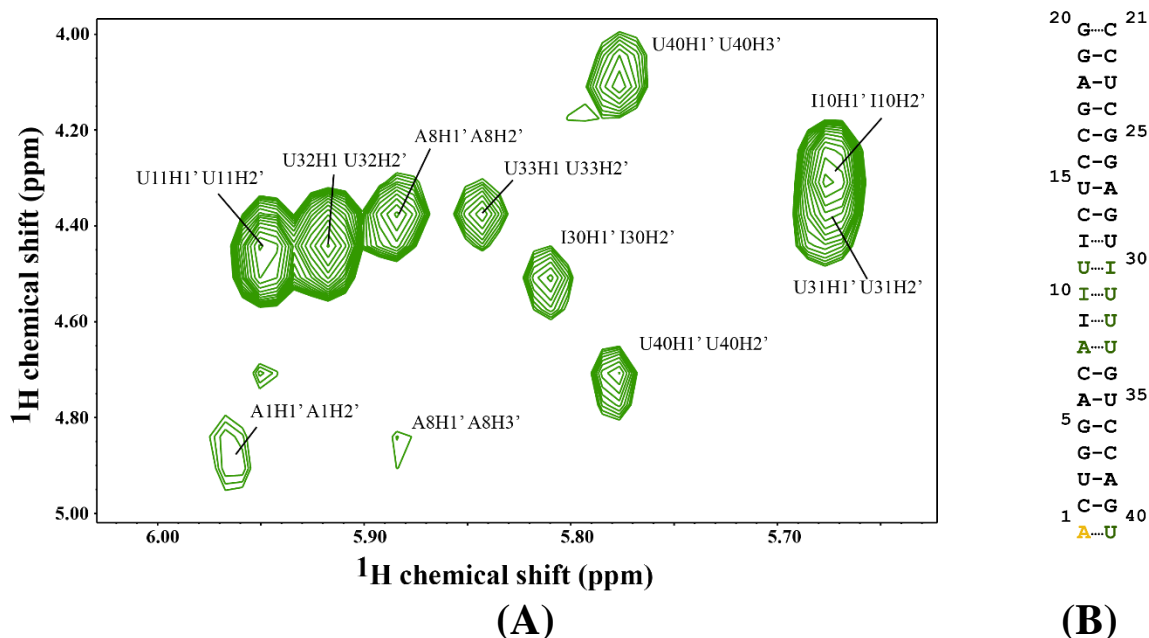
Figure 2.17 – **(A)** A 2D TOCSY of IRD RNA showing the H1'/H2' correlations and their correlations. **(B)** The IRD sequence, with the sugar pucker conformations denoted. The residues colored in black adopt a *C3' endo*, while the residues in green adopt the *C2' endo* conformation. A1 is colored in yellow, as its sugar pucker is a mixture of the two conformations.

89

2.3.4 The use of a labelled sample for further IRD structural analysis

However further resonance assignment of the inosine-containing motif, even in the aromatic-anomeric region, proved ambiguous. Since the A-form geometry may be distorted in this of part of the RNA, standard cross-peaks which would be indicative of this conformation (such as inter-residue H8 (or H6)/H1') may not be present. Due to the large extent of spectral crowding, the presence or the absence of such signal may be impossible to ascertain (Figure 2.20A). Such is the case with potential I10H8/I9H1', U11H6/I10H1' signals. In addition, the assignment of some inosine H2 resonances, in particular I9 and I10 is also complicated by overlap, which makes it difficult to determine whether those protons give rise to an NOE pattern similar to that of adenine H2 (Fürtig, Richter et al. 2003) or whether other and/or additional signals are also observed.

While partial assignment of the aromatic-anomeric region was achieved, assignment of ribose regions of the IRD NOESY is impossible, as the overlap in those parts of the spectra is extremely severe. Of particular importance is the assignment of the ribose resonances of the residues of the inosine-containing motif, as this region of the IRD does not adopt a standard A-form helix and is of special interest. To address this issue, the inosine residues were selectively labelled in the IRD (Figure 2.18A).

2.3.4.1 Choice of IRD labelling scheme

The IRD labelling scheme was designed with the purpose to gain as much insight as possible into the orientation of the inosine residues in the RNA as well as their position relative to their neighbouring residues and their base-pairing partners. The ribose signals tend to overlap in one region. The base signals, on the other hand, are better dispersed and provide the most useful information about the duplex conformation. However, limitations in the inosine phosphoramidite synthesis pathway precluded uniform labelling in the inosine base. In addition, after assignment of the unlabelled spectra the missing assignments for both I10 H2 and H8 suggested that they might be overlapped. It was decided to specifically label the C8 and not the C2 position in order to be able to distinguish between the two and maximise the information gain about the central inosine motif. Hence only the specific positions C8 and N7 of the base as well as all the ribose carbons were ^{13}C and ^{15}N labelled in the inosine residues (Figure 2.18B). The former gives rise to signals which indicate whether the base in question adopts a more common *anti* or the less common *syn* conformation. The latter provides insight into whether the residue is involved in non-standard

hydrogen bonding interactions such as Hoogsteen base-pairing (Gaffney, Kung et al. 1995, Nikolova, Gottardo et al. 2012).

As part of a collaboration project, the labelled IRD sample was prepared in the group of Thomas Carell in two steps. First the labelled inosine phosphoramidite, which is a solid phase RNA synthesis pre-cursor, was prepared. In a second step this phosphoramidite was incorporated into an otherwise IRD RNA using an RNA synthesizer. The RNA product was purified using high-pressure liquid chromatography and its purity was analyzed by mass spectrometry.

(A)   **(B)**
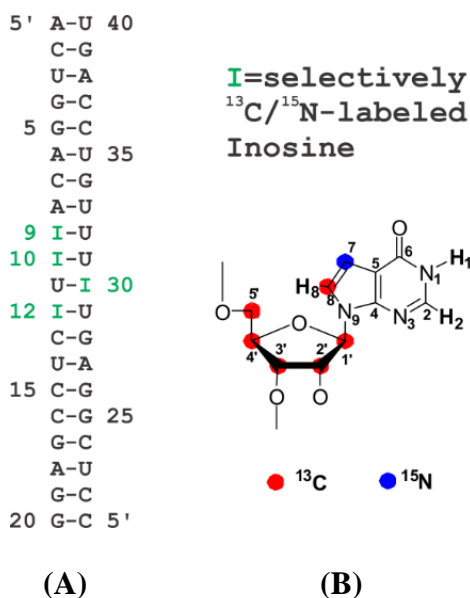
Figure 2.18 – (A) RNA sequence with the inosine residues highlighted in green. **(B)** Site-selective isotope-labeling scheme and atomic numbering for inosine. Positions in red and blue are isotopically ^{13}C and ^{15}N labelled, respectively.

2.3.4.2 Assignment of the inosine ribose sugars in IRD

After the labelled IRD sample was shown to dimerize on a native-PAGE gel, and homonuclear NMR experiments on it gave rise to spectra similar to those of its unlabelled counterpart, we proceeded with heteronuclear experiments. First, assignment of the inosine ribose sugar was performed with the help of the 3D ^{1}H, ^{13}C, ^{1}H HCCH-TOCSY as well as 3D ^{1}H, ^{13}C, ^{1}H HCCH-COSY experiments. With the inosine ribose resonances assigned, various three-dimensional NOESY-based experiments were acquired on the labelled IRD sample (Table 2.03) with the goal of obtaining more distance restraints which would yield more structural information on the inosine-containing motif.

2.3.4.3 IRD labelling simplifies assignments of the inosine-containing motif

The 2D ^{1}H, ^{13}C edited NOESY-HSQC experiment (marked in black on Figures 2.19B, 2.20A, and 2.21A) only gives rise to NOE cross-peaks between two hydrogen atoms if one of them is bound to ^{13}C labelled carbon. Therefore in such experiment, only NOE cross-peaks between the ribose protons or the base H8 proton of inosine (Figure 2.18A) and any other proton close in space are observed (either bound to ^{13}C or ^{12}C). As such, when performed on labelled IRD, this experiment simplifies spectral assignments of residues in the centrally located inosine-containing motif and its adjacent base pairs by reducing spectral overlap significantly (compare figure 2.19 and 2.19B).

The exact position of the H8 proton of I10, obscured due to signal overlap in the unlabelled IRD NOESY spectrum (Figure 2.16) could be unambiguously determined in the NOESY-HSQC (Figure 2.19B). In addition, the lack of an inter-residue I10H8/I9H1' cross-peak was also conclusively demonstrated indicating a clear deviation from A-form geometry for this part of the motif. In the homonuclear NOESY spectra, this claim could not be conclusively made, as it would have been possible that this resonance exists but could not be assigned due to spectral overlap. The NOESY-HSQC spectrum also allows one to conclusively determine the chemical shifts of the inosine H2 resonances. In the homonuclear NOESY, I9H2 and I10H2 were partially overlapped with other signals, while the I12H2 and I30H2 signals were of a very weak intensity. The edited NOESY unambiguously resolves the location of these four peaks (Position of I9H2 and I10H2 are indicated on Figure 2.19B). All four inosine H2 resonances display an NOE pattern similar to that of the adenine H2 counterparts (Fürtig, Richter et al. 2003): One cross-peak to the H1' atom of the residue to the 3'-side of the inosine in question, and another, to the H1' of the residue to the 3'-side of the inosine's base-pairing partner. In addition, I30H2 also displays a signal to the H1' of I10. This NOE would correspond to an inter-atom distance which is not in line with an A-form helix geometry, further confirming that I10 and I30 adopt a non-standard orientation.
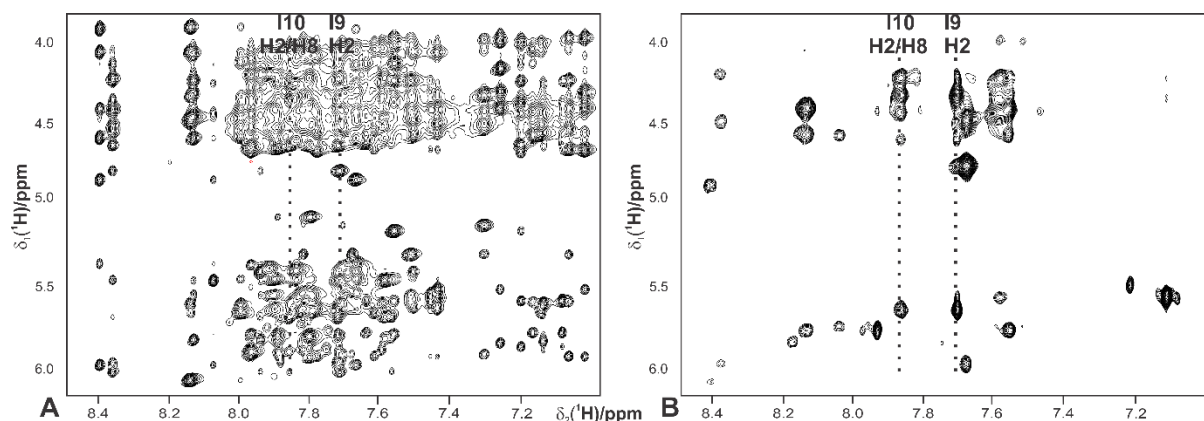
Figure 2.19 – (A) Aromatic-anomeric/sugar proton region of the homonuclear NOESY of the Inosine RNA duplex. The severe overlap especially in the aromatic-sugar proton region renders unambiguous chemical shift assignment impossible and necessitates spectral filtering approaches.

2.3.4.4 Isotope edited and filtered NOESY experiments resolve signal overlap

However, 2D ^{1}H, ^{13}C edited NOESY-HSQC alone still does not allow for a more comprehensive IRD assignment. The aromatic/ribose region of this RNA, still gives rise to a very large degree of spectral overlap and the overlap between the I10 H2 and H8 resonances prohibits unambiguous cross-peak assignments. Hence more intricate experiments were performed on IRD which utilized double editing and editing-filtering. The former is a pseudo 4D NOESY-HSQC which gives rise to an inter-proton cross-peak if both of them are bound to a ^{13}C carbon (marked in blue in Figure 2.20A, 2.20B). This means an NOE cross-peak can be observed between an inosine ribose and H8 base protons in a close spatial proximity. For the edited-filtered experiment, an NOE cross-peak is observed only if one proton is bound to a ^{13}C carbon and the other is bound to a ^{12}C carbon (marked in red in Figure 2.20A and Figure 2.20B). This translates to an NOE cross-peak being observed between an inosine ribose or H8 base proton and H2 of inosine (since it is not ^{13}C-bound) or any non-inosine protons close in space.

As such, the edited-edited NOESY has largely facilitated the assignment of inosine intra-residue as well as I9 and I10 inter-residue cross-peaks, as those two labelled inosine moieties are located next to each other in the IRD. Using this experiment, cross-peaks such as I10H8/I10H2', I10H8/I9H3', I30H8/I30H3' and I12H8/I12H2' (the first two signals are labelled in blue in Figure 2.21A) were successfully assigned. The edited-filtered NOESY on the other hand, has proven valuable in assigning cross-peaks between labelled and unlabelled carbon resonances. These are mostly inter-residue cross-peaks between inosine ribose protons and various uracil aromatic

93

protons located next to those inosine residues or protons of residues flanking the motif, also located next to the labelled inosine moieties. Thus, the edited-filtered NOESY experiment has allowed for unambiguous assignment of C13H6/I12H2', U31H6/I30H2' (Figure 2.20A), and A8H2 I9H2' cross-peaks.

Employing those two complementary experiments has also allowed us to assign completely overlapping signals. In one such example, I10H8/I10H1' and I10H2/I10H1' resonances had exactly the same chemical shift values, and hence appear in the same location in the single-edited NOESY (Figure 2.20B, left strip). This greatly complicates spectral assignments of these resonances. However, since the I10H8 resonance is ^{13}C-bound, and I10H2 is not, the first cross-peak appears in the edited-edited NOESY spectrum (Figure 2.20B, right strip), while the other was observed in its edited-filtered NOESY counterpart (Figure 2.20B, centre strip). Similarly, otherwise completely overlapping, I10H8/I10H2', I10H8/I9H2' (Figure 2.20B, right strip) and I10H2/I10H2', I10H2/I9H2' (Figure 2.20B, centre strip) cross-peaks were successfully assigned using the combination of those two experiments. This shows the strength and necessity for specific labelling in the base moiety of inosine since a uniformly labelled inosine would have failed to resolve this ambiguity.

Using a combination of those experiments, almost all cross-peaks in the aromatic ribose and the anomeric/ribose regions (data not shown) were assigned. Additional cross-peaks in the anomeric-ribose, as well as some outlying ribose-ribose signals were also assigned.
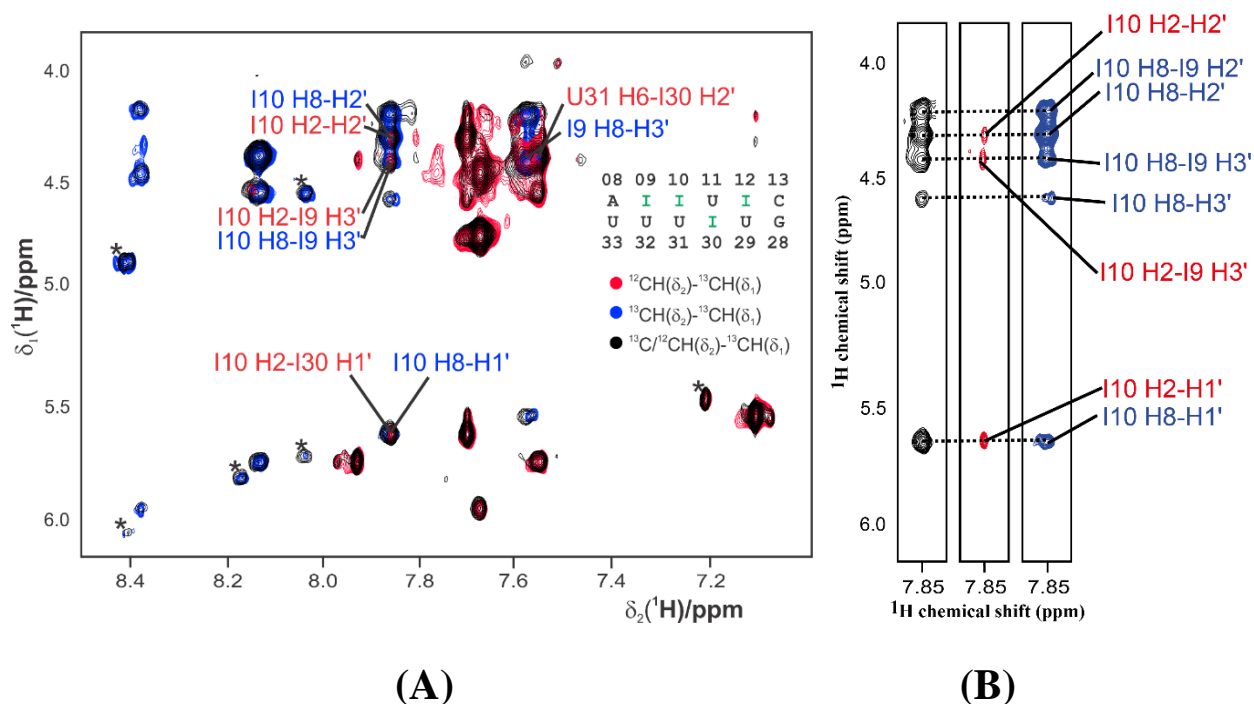
Figure 2.20 – **(A)** An overlay of an edited (marked in black), edited-edited (marked in blue) and edited-filtered (marked in red) NOESY spectra of labelled IRD RNA. Representative resonance assignments, performed with the help of either the edited-edited or edited-filtered experiments are colour-coded appropriately. **(B)** Three strips of a single edited, edited- filtered, edited-edited NOESYs of IRD. The colour coding is the same as in **(A)**. Using the edited- filtered and edited-edited NOESY experiments, it was possible to resolve the ambiguity in spectral assignments of signals involving I10H8 and I10H2.

2.3.4.5 Determination of the sugar pucker conformation from ribose chemical shifts

The ribose chemical shift also provides insight into the sugar pucker conformation of the given residue. A so-called can1 (Fürtig, Richter et al. 2003) value which is defined by the following formula

$$\text{can1} = 0.179 * \delta_{C1'} - 0.225 * \delta_{C4'} - 0.0585 * \delta_{C5'}$$

where the $\delta_{C1'}$, $\delta_{C4'}$, and $\delta_{C5'}$ are the chemical shifts for C1', C4' and C5' respectively. If the can1 value is above -6.25, then residue in question has a *C3' endo* conformation, if it is not, then it is *C2' endo*. Since the ribose moieties of the inosine residues in the IRD sample are labelled, their chemical shifts were unambiguously determined, and their can1 values calculated (Table 2.06). These values indicate that residues I10 and I30, that is the residues in the middle of the inosine-containing motif, adopt a clear *C2' endo* conformation (marked in bold on Table 2.06), while I9

and I12 sugar puckers, most likely adopt a mixture between the two conformations. When comparing these findings to those from the short-mixing time TOCSY above (2.3.3.3), the results agree for I10 and I30 (as the TOCSY also suggests that I10 and I30 adopt a *C2' endo* conformation), but not for I9 and I12. In some cases, when a certain residue has a mixed sugar pucker conformation, it results in an intermediate intensity H1'/H2' cross-peak, which is not the case here. However, since the use of the short-mixing time TOCSY for sugar pucker determination is not a quantitative, the can1 value provides a more accurate description to study the sugar pucker conformation. Hence a revised sugar pucker scheme for the IRD is depicted in Figure 2.21.

	C1'	C2'	C3'	C4'	C5'	Can 1
I9	91.69	76.30	74.45	83.70	65.98	-6.28
I10	90.52	76.82	76.22	84.81	67.20	**-6.81**
I12	93.58	76.20	73.60	85.16	66.63	-6.31
I30	90.68	76.32	75.95	85.70	66.00	**-6.91**

Table 2.06 – The chemical shift and the calculated can1 values for the inosine residues in the labelled IRD sample. The residues clearly adopting a *C2' endo* conformation, are marked in bold.
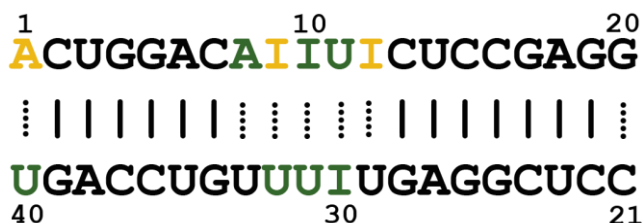

Figure 2.21 – Revised sugar pucker conformation of IRD RNA, taking the determined inosine can1 values into account. The residues colored in black adopt a *C3' endo*, while the residues in green adopt the *C2' endo* conformation. A1 is colored in yellow, as its sugar pucker is a mixture of the two conformations (the colour coding is the same as in Figure 2.18).

2.3.4.6 Analysis of inosine N7 chemical shifts

The specific ^{15}N-labelling of the N7 in inosine residues allowed us to probe for possible Hoogsteen base-pairing. A distinctive upfield ^{15}N chemical shift of N7 is indicative of the existence of such interaction (Gaffney, Kung et al. 1995, Nikolova, Gottardo et al. 2012). The lack of observable imino proton signals for the residues of the inosine containing motif, as well as for the flanking U33 (Figure 2.22A) in the natural ambundance SOFAST-HMQC (Figure 2.22B) indicates the lack of stable base pairing interactions as also demonstrated by the 2D ^{1}H, ^{1}H NOESY on the unlabelled IRD (2.3.3.1). Moreover, the N7 chemical shifts observed in the long-range [1]H-

^{15}N-HSQC experiment (Figure 2.22C) demonstrate that the inosine N7 is not involved in hydrogen bond formation. Additionally, the line-broadening observed for NMR signals of the inosine residues flanking the IIUI motif (I9 and I12) compared to the central ones (I10 and I30). This finding suggests that I9 and I12 are more tightly integrated into the A-form helical stack, than I10 and I30. Nevertheless, I10, I30 and the uracil residues located opposite to them on the other strand (U31 and U11 respectively) continue stacking interactions in the RNA duplex region. This can be concluded from the lack of intense intra-residue H8/H1' and H6/H5 NMR signals in the ^{2}H$_2$O NOESY (2.3.3.1) experiment as well as the significant number of inter-residue NOEs that are observed for these two inosine residues.
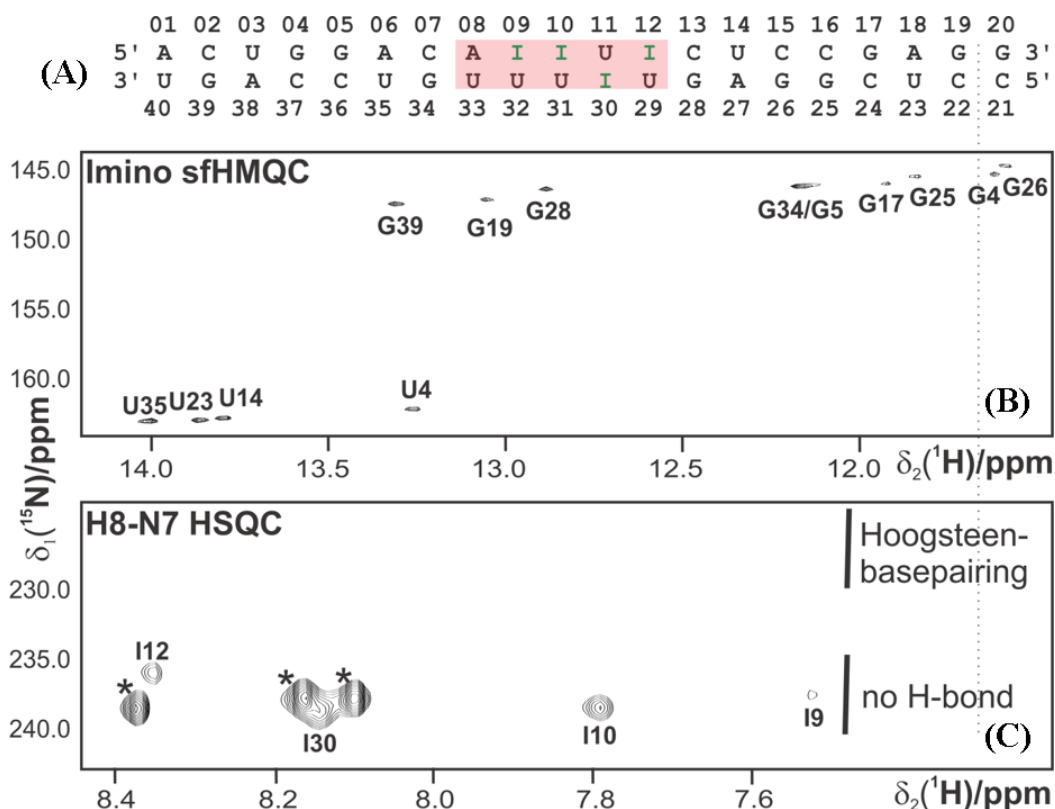
Figure 2.22 – (A) Secondary structure information on IRD. **(B)** Natural abundance Imino-sfHMQC, showing that the imino protons for the inosine-containing motif and the adjacent AU base pair are not observed. **(C)** A long-range sfHSQC spectrum of the labelled IRD, excluding the possibility of Hoogsteen base pair formation due to the absence of upfield N7 chemical shifts. The I9 and I12 signals are more broadened compared to I10 and I30.

2.3.5 The use of p100DM34 construct for structural studies

Due to problems with purification of large amounts of the human p100 construct spanning domains SN3 and SN4, it was decided to use the homologue p100 construct from *Drosophila melanogaster* (p100DM34). The p100DM34 construct was expressed and purified in milligram amounts, as described above (2.2.2.1, 2.2.2.2 and 2.2.2.3) in both hydrogenated as well as deuterated media. Construct solubility was above 1.5 mM in 500 µL and the sample was shown to be stable for at least two weeks at room temperature, before noticeable degradation began to occur.

For a triple labelled p100DM34, a large degree of signal dispersion in its fingerprint two-dimensional HSQC spectrum (Figure 2.23A) has shown that p100DM34 is folded in the given conditions. Once the viability of the construct for further NMR analysis was established, three dimensional experiments described above (2.2.2.4.1), have facilitated p100DM34 backbone assignment. The protein was assigned to 90% completion (the proline residues whose backbone cannot be assigned due to the lack of an amide hydrogen atom were excluded from the total number of residues in this calculation – Figure 2.23B).
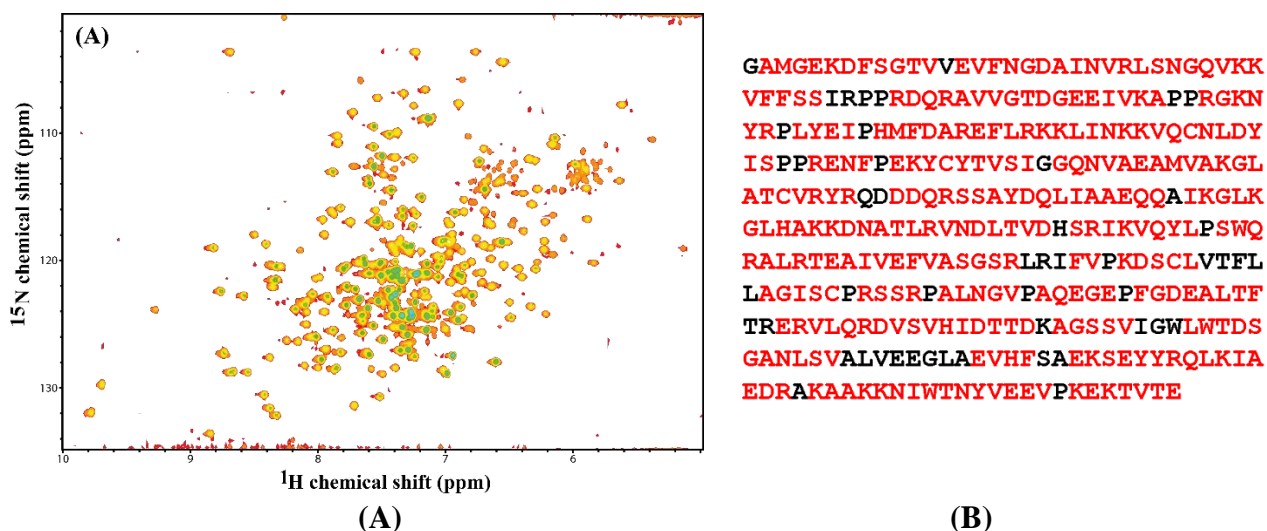

Figure 2.23 – **(A)** 2D ^{1}H, ^{15}N HSQC of a triple labelled p100DM34 sample. Buffer conditions: NMR Buffer 2.1 **(B)** The sequence of p100DM34, with the assigned residues marked in red.

2.3.5.1 p100DM34 homology model in backbone assignment analysis

A homology model of p100DM34 was created with the web based program Swiss Model (Arnold, Bordoli et al. 2006) based on the 1.9 Å resolution crystal structure of the human homolog (PDB-code: 3BDL). The homolog has a sequence identity of 53.7%. The model of the whole construct was generated based on the sequence alignment. The backbone assignments marked on this model (Figure 2.23) span almost all the regions of the construct, except a small portion of the β-barrel in SN4, allowing one to track any changes p100DM34 may undergo upon RNA binding at any location of the protein.
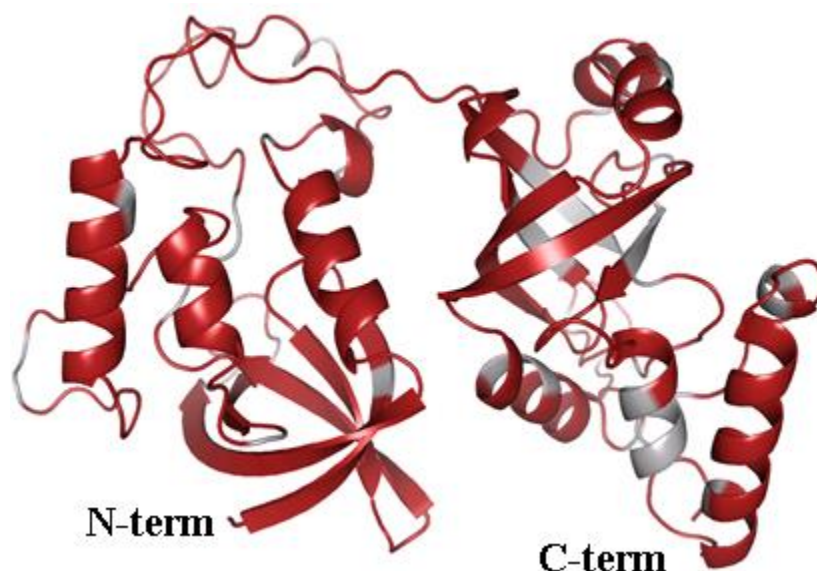
Figure 2.23 – Backbone assignments plotted on a homology model of p100DM34, with the assigned regions marked in red.

2.3.5.2 Analysis of p100DM34 secondary chemical shifts and comparison with p100HS34

From assignments of the C_α and C_β resonances in p100DM34, secondary chemical shifts were calculated using the formula above (2.2.2.4.3), as well as using the program TALOS+. The resultant value for every residue from this formula is strongly indicative of the secondary structure this residue assumes. A positive value points to a helical structure, while a negative one suggests a β-strand.

The p100DM34 model (Figure 2.24), as well as the secondary structure predicted from the secondary chemical shifts (Figure 2.25) show nearly the same secondary structure elements as the

crystal structure of the human homolog, with the exception of 2 short β-strands in the loop between strand β9 and helix α6 (magenta in Figure 2.24) and a longer β-strand between strand β8 and β9 (cyan in Figure 2.24), which are not predicted by TALOS+ in p100DM34 (Figure 2.25). The TALOS+ program also predicts a short two-residue helix at the C-terminus of the model, which will not be counted as a helix here. The loop between strand β3 and helix α1 is extended by ten residues in the drosophila homology model compared to the human crystal structure (highlighted in green in Figure 2.25). In contrast, the loop between β-sheets β5 and β6 is shortened in the drosophila model by six residues (highlighted in red in Figure 2.25). Interestingly, the additional parts in each structure occupy the same space. Residues in the N-terminal border of β11 and the C-terminal border of α7 are not assigned and have no predicted secondary structure from TALOS+. For the model, it was assumed that these secondary structural elements are the same as in the human homolog, as predicted by Swiss Model.
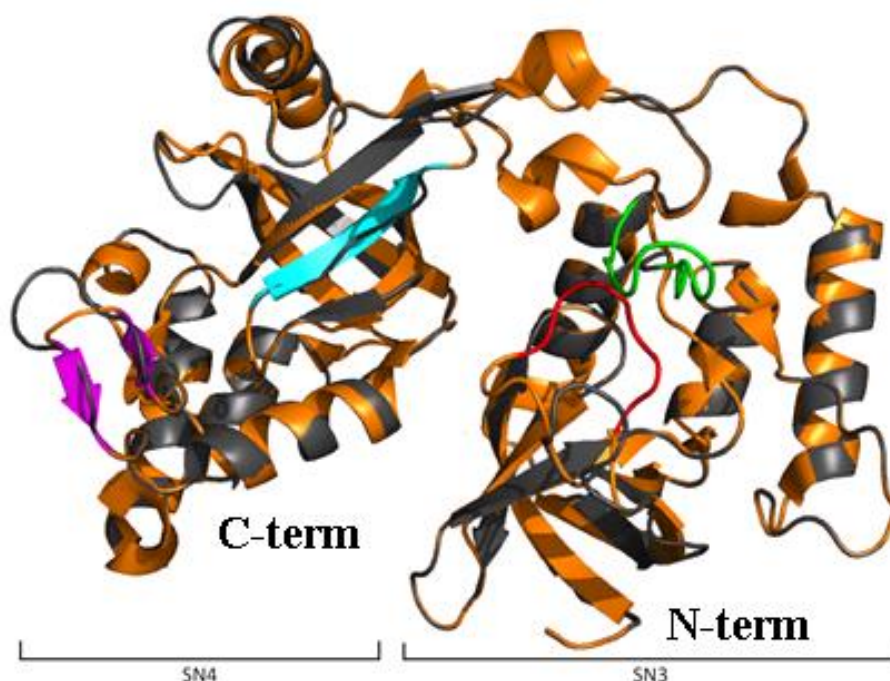
Figure 2.24 – Overlay of the *Drosophila* p100DM34 (grey) model and the human p100 crystal structure (orange). In the *Drosophila* model, elongated loop between β3 and α1 is highlighted in magenta, the shortened loop between β5 and β6 is highlighted in the human structure in green and the in the drosophila model missing β-strands are highlighted in magenta and cyan.

The overall secondary structure of the SN3 domain has the classical elements of an OB-Fold (oligonucleotide/oligosaccharide binding motif fold), which is a five-stranded β-barrel built from two three-stranded β-sheets (Theobald, Mitton-Fry et al. 2003). The β-strand β1 is shared by

both β-sheets. Between β-strand β3 and β4 there is a α-helix, α1, which packs against the β-barrel and encloses it on one site. In addition to this classical OB-fold, there are two further α-helices (α2 and α3) (Figure 2.26A).

SN4 domain has a fold very similar to that of SN3, but the second strand of the first three-stranded β-sheet (loop between β8 and β9) is not predicted as a β-strand by TALOS+ (cyan in Figure 2.25). In addition, the strand that is part of both three stranded β-sheets in SN3 is split in two (β7 and β8) in SN4. These β-strands (β7, β8 and β9) are followed again by an α-helix (α6) and two β-strands (β10 and β11) of the second three stranded β-sheet. N-terminal of this OB-fold is an additional β-strand (β6) and a α-helix, which is split in two parts (α4 and α5). C-terminal of the OB-fold are again two α-helices (α7 and α8). Between these two helices is a short three-residue β-strand (β12) (Figure 2.26B).
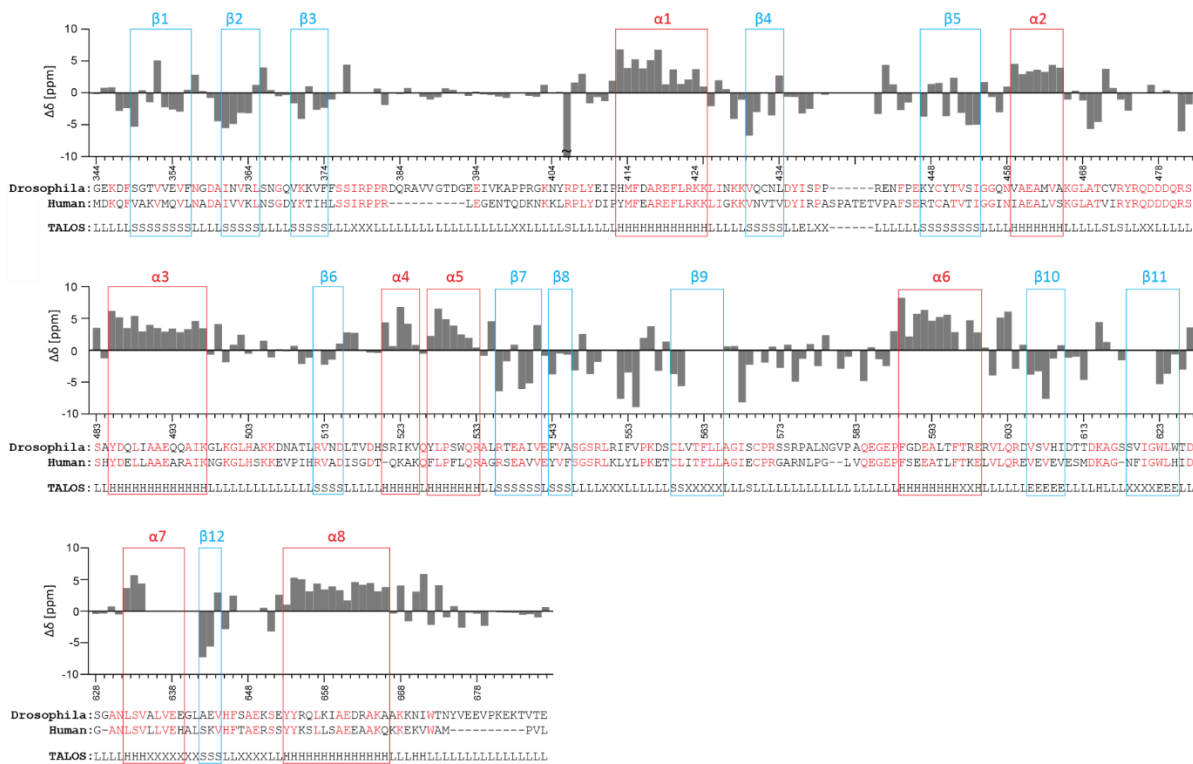
Figure 2.25 – Homology model verification. Secondary chemical shift data of p100DM34 are plotted on the sequence alignment with the human homolog. Below the sequence alignment the secondary structure predictions of the drosophila homology model by the program TALOS+ is indicated (H indicates α-helices S β-sheets, L loops and X unassigned residues). TALOS+ used chemical shift data of C_α, C_β, N, H_N and C'. The secondary structure in the homology model is highlighted by red (α-helices) and blue (β-sheets) boxes. The red letters in the sequence alignment indicate identical residues (53.7% sequence identity), the bars indicate gaps in the sequence alignment.

(A) **(B)**
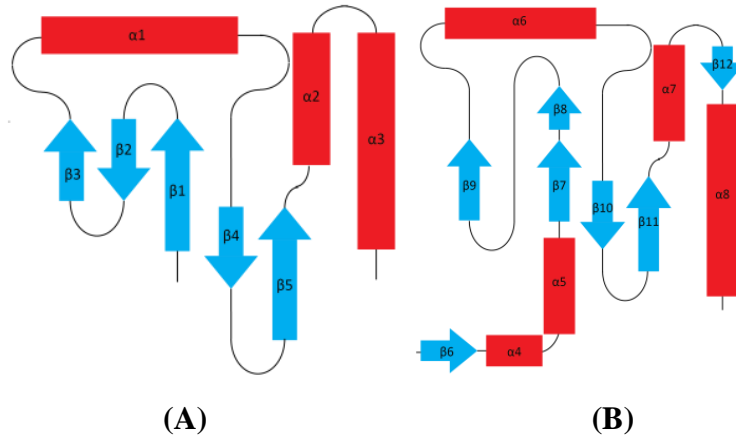
Figure 2.26 – **(A)** Organization of secondary structure elements in the p100DM SN3 domain **(B)** Organization of secondary structure elements in the p100DM SN4 domain.

2.3.6 Contribution of domain-domain interaction to p100DM34 fold

To investigate how the interaction between SN3 and SN4, contribute to the overall fold of p100DM34, ^{15}N-labelled protein constructs spanning just these individual respective domains were cloned, expressed in minimal media and purified. A superposition of the HSQC spectra for SN3 (Figure 2.27A) and SN4 (Figure 2.27B) with that of p100DM34, shows that the signals for the single domain constructs do not overlay well with that of the tandem domain. These findings indicate that an SN3-SN4 interaction contributes to the overall dual construct conformation.
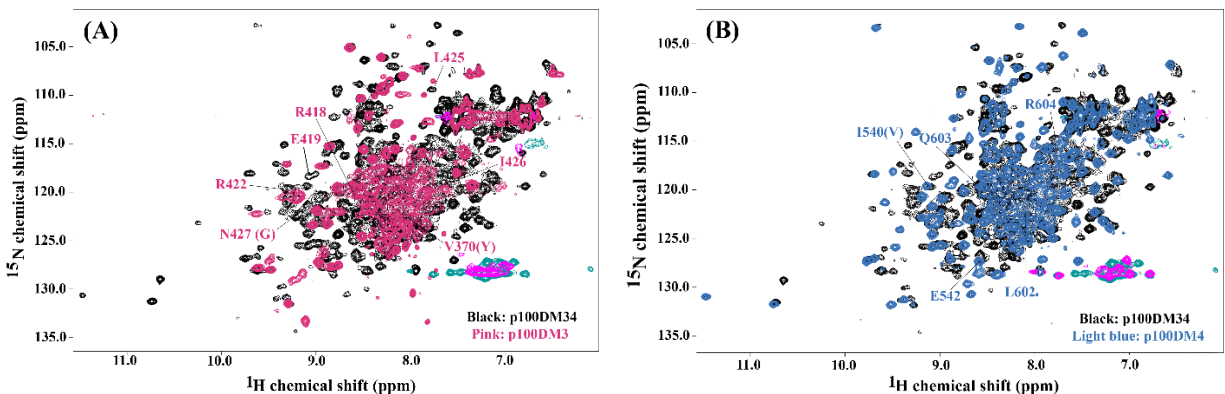
Figure 2.27 – A 2D ^{1}H, ^{15}N HSQC superposition of p100DM34 with **(A)** p100DM3 and **(B)** p100DM4 single domain constructs. A lack of signal overlay suggests that the tandem domain fold is defined in part by a domain-domain interaction. The signals for p100DM34 are marked in black, while those for SN3 and SN4 are marked in pink and light blue, respectively. Residues located at the interface of the two domains are denoted in both spectra with the same colour coding, and differences between the *Drosophila* and the human p100 are marked in brackets when appropriate.

A number of residues at the domain interface, may contribute to the overall tandem domain fold (Figures 2.27A, 2.27B and 2.28). In SN3, R418, R422 most likely engage in ionic interactions with E542 in SN4. R422 makes hydrophobic contacts with Q603 in SN4, which in turn also interacts with the side chain of I426 in SN3. Most of these residues, with the exception of V370, G427 and V540 (Figure 2.28) are conserved in both the human and the *Drosophila* variants of p100, pointing to their importance in shaping the overall conformation of the tandem domain constructs. Differences in the chemical shift in the above mentioned resonances in the tandem and single domain proteins could not be determined, due to their absence in the latter construct.
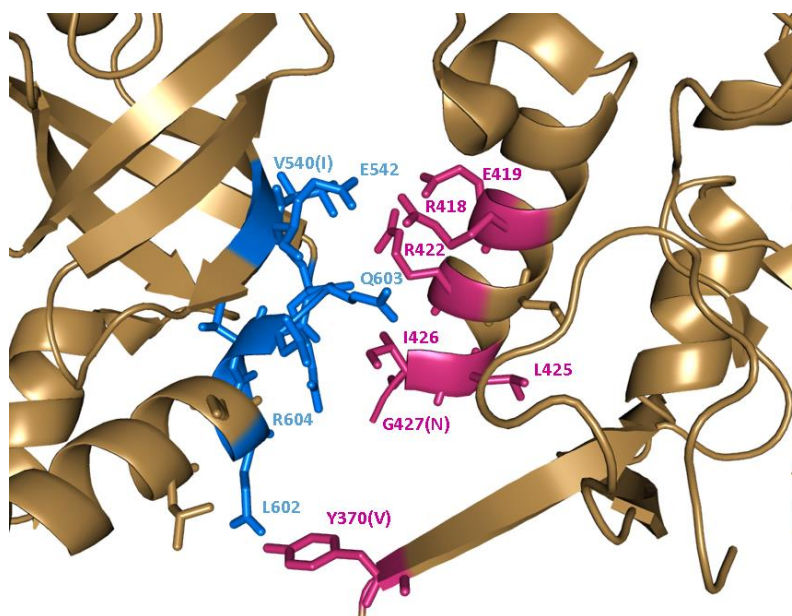
Figure 2.28 – Depiction of the residues at SN3/SN4 the domain interface plotted on the human p100 crystal structure of SN domains 3 and 4. Residues belonging to SN3 are marked in pink, while residues in SN4 are marked in blue. Differences in amino acid sequence between the human and *Drosophila* proteins are indicated in brackets (those contain the *Drosophila* residues).

2.3.7 p100DM34 relaxation experiments

2.3.7.1 Overall p100DM34 relaxation properties

In a work performed by Martin Rübbelke, a Masters' student, ^{15}N-relaxation data was recorded. This data contains information about the dynamics and flexibility of the protein backbone. They are useful for determining which parts of the protein are structured or flexible and for comparing the overall dynamics of entire domains. In addition, they can provide criteria to decide which residues are rigid enough to be used for domain orientation studies using RDCs.

From NMR relaxation experiments (T_1 and $T_{1\rho}$), the rotational correlation time τ_C was determined. The results for the τ_C, R_1 (=$1/T_1$), R_2 (calculated from T_1 and $T_{1\rho}$) are plotted in Figure 2.29. The [1]H-^{15}N heteronuclear Nuclear Overhauser Efffect (hetNOE) values, which provide information regarding the flexibility of the N-H bond in a given residue (Keeler 2002), thereby giving additional insight into the local dynamics of the protein, are also displayed (Figure 2.29).
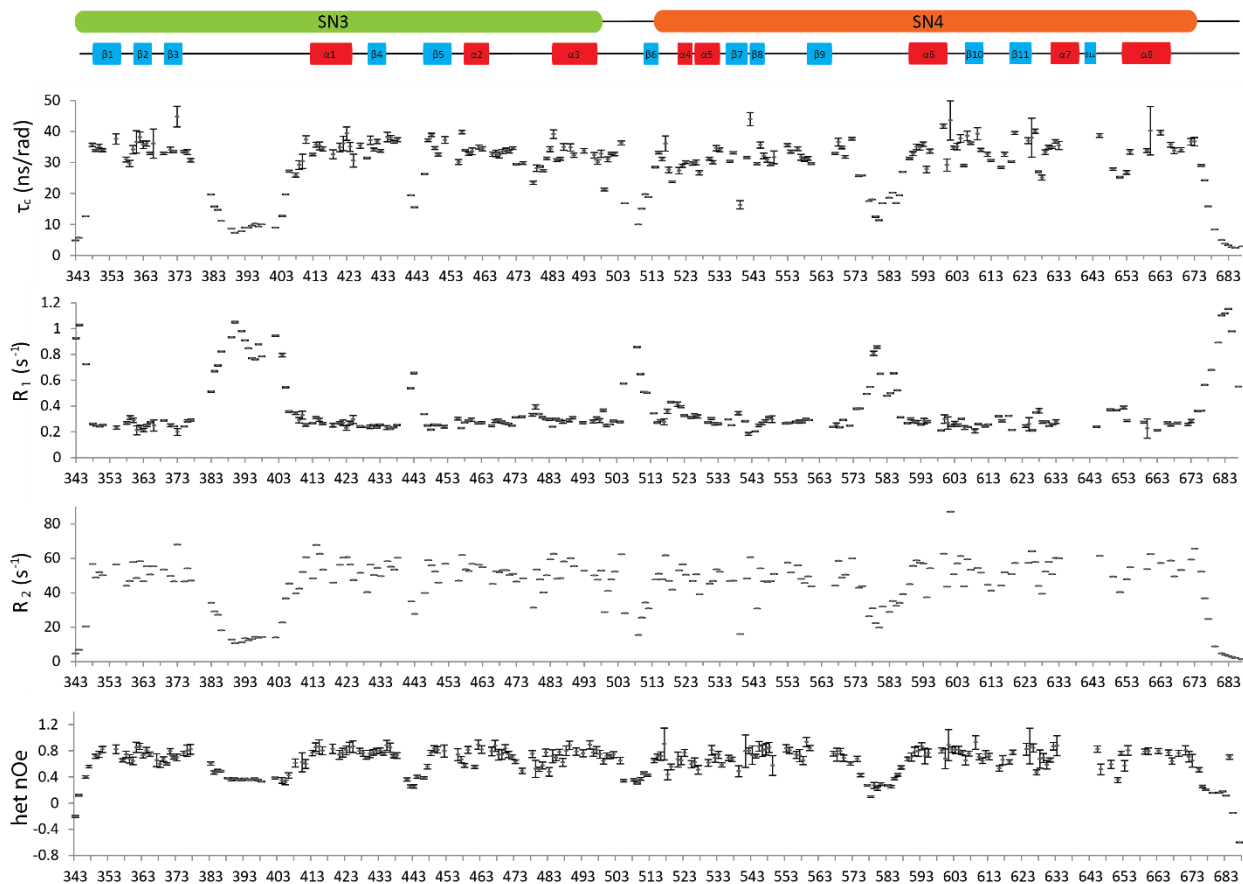
Figure 2.29 – Correlation time (τ_C) in ns/rad, the R_1 and R_2 relaxation constant in s[-1] and the hetNOE data. On top, the 2 domains and the secondary structure elements are indicated (α-helices are indicated by red and β-strands by blue bars).

The four plots show rigid parts, which contain all areas with secondary structure elements (α-helices and β-sheets), and flexible parts. The correlation time of the rigid parts of the protein, which were identified by a hetNOE value larger than 0.6, is 33.8 ns/rad. A protein with a molecular weight comparable to p100 DM34 (38.8 kDa) is expected to tumble much faster with a correlation time around 23 ns/rad (Cavanagh 2007). The higher apparent correlation time might be a concentration effect. The relaxation experiments were recorded at a protein concentration of 950 µM, which might induce dimerization or higher order aggregation.

To address this issue, a static light scattering (SLS) experiment was performed, but showed no signs for dimerization, and only a small fraction of higher order aggregation, which might cause a higher apparent τ_c. However, the SLS could only be performed at a much lower concentration (around 100 µM) compared to the NMR measurement (around 950 µM) due to saturation of the detector at higher protein concentrations. Additionally, dilution effects on the column further decrease the effective concentration in the SLS.

The correlation time of the rigid parts of the individual domains are 33.7 ns/rad for SN3 and 33.9 ns/rad for SN4. Taken together, the high correlation time and the similarity of the correlation times of the single domains indicate that it is likely that the two domains tumble as a single unit and not independently, which is in line with previous findings (2.3.6).

RDC measurements which were performed by Martin Rübbelke, show that SN3 and SN4 have similar magnitudes of the alignment tensor as well as rhombicity values. These findings also suggest that the solution orientation of the two domains is similar to the crystal structure based-homology model. If the orientation of the two domains were different in solution compared to the homology model, then the alignment tensor would be different.

(A)   **(B)**
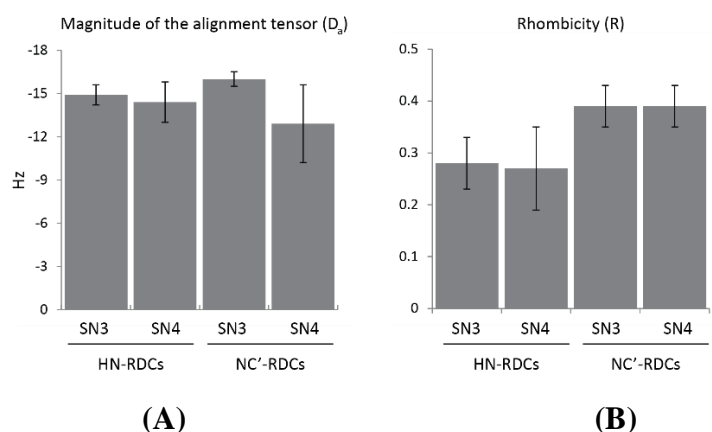
Figure 2.30 – Shown are bar diagrams of **(A)** the magnitude of the alignment tensor and the **(B)** rhombicity for the HN-RDCs and the NC'-RDCs of domains SN3 and SN4.

2.3.7.2 Loops in p100DM34 have a shorter correlation time than elsewhere

There are several flexible parts in the molecule that correspond to loops in the domains and the linker connecting the two domains. The longest and most flexible loop is located in the SN3 domain between β-sheet β3 and α-helix α1. This loop has a length of 38 residues and is elongated by 10 residues compared to the human homolog. In this loop, 26 residues have a smaller correlation

time compared to the average of the rigid parts and can be considered as flexible (green in Figure 2.30). The correlation time of this loop is around 9 ns/rad and the hetNOE is around 0.4, indicating high flexibility. The comparable loop in the SN4 domain between the first set of β-strands (β9) and the first α-helix (α6) of the OB-fold is shorter (15 flexible residues, red in Figure 2.30), but also flexible with a correlation time around 13 ns/rad and a hetNOE value around 0.3. Other flexible parts of the protein are the loop between β4 and β5 in domain SN3 (5 flexible residues, magenta in Figure 2.31), the linker between the two domains (8 flexible residues, orange in Figure 2.31) and the C-terminus of the protein construct (last 13 C-terminal residues, cyan in Figure 2.31). All three flexible loops and the C-terminus of the construct are oriented to the same site of the protein and flank a positively charged cleft, which is the hypothetical dsRNA binding site. These flexible residues might be involved in the binding of dsRNA and this interaction with the dsRNA might rigidify these parts of the protein.
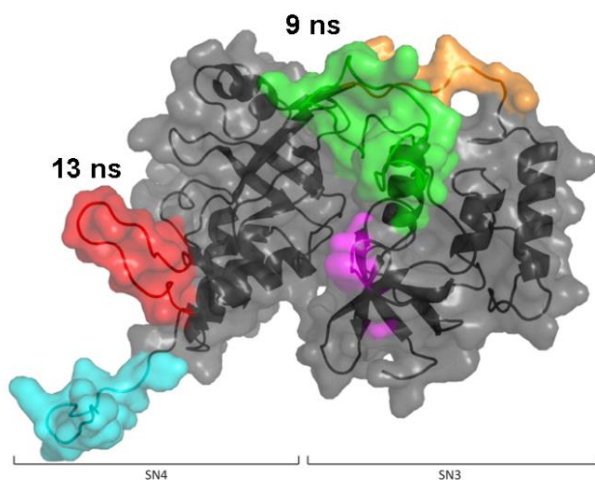
Figure 2.31 – Flexible parts of p100 DM34. Indicated are the flexible parts of p100DM34, which were identified by a smaller correlation time compared to the average (loop between β3 and α1 in green, loop between β4 and β5 in magenta, linker in orange, loop between β9 and α6 in red and C-terminus in cyan). The lowest correlation times for the first and fourth loops are marked.

2.3.8 RNAse tests on p100HS34

To determine whether p100 possesses RNAse activity p100HS34 was subjected to an RNAse activity test using the RNaseAlert® Lab Test Kit. The kit tube contain a lyophilized RNA with a flourophore and a quencher. If the solution to be added to this tube has RNAses, they cleave the RNA, which results in fluorescence which can be detected.

As negative and positive controls, samples containing nuclease-free water and RNAse A respectively, were used. To account for endogenous RNAse contamination in the cellular *E. coli*

extract, the His-tag co-purified with p100HS34, cleaved and separated using a second nickel column step in 2.2.2.3.2 was used as an internal control.

Following the addition of p100HS34 to the lyophilized RNA and the subsequent overnight incubation at 37 C, RNAse activity was clearly observed. Some minor RNAse activity was also present in the sample containing the His-tag (Figure 2.32). Importantly, p100HS34's RNAse activity could be abolished by addition of 20 mM EDTA (a divalent metal chelator) to the sample containing the protein (Figure 2.32). Since the activity of the staphylococcal nuclease proteins is known to be commonly mediated by divalent metals (Hale, Poole et al. 1993), this finding further suggests that the RNAse activity is from p100HS34 in origin.
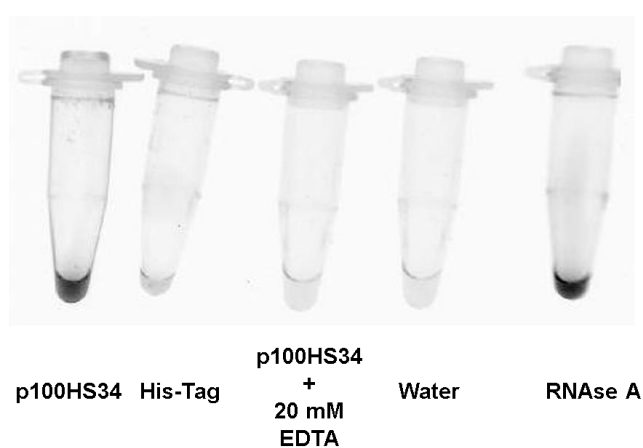
Figure 2.32 – Vials of the RNaseAlert® Lab Test Kit containing from left to right: p100HS34, the cleaved off His-tag, p100HS34 with 20 mM EDTA, nuclease free water, RNAse A. RNAse activity is clearly visible in the sample containing p100HS34 and the RNAse A positive control. Some background RNAse activity is also present in in the His-tag sample

2.3.9 Study of p100-IRD interaction

Now that structural as well as biochemical characterization of p100 and IRD was performed, we set out to study their interaction. In order to determine the structural determinants responsible for the p100/IRD contacts, as well as possibly determine the basis for p100 discrimination towards edited RNA in general, this interaction was studied by NMR spectroscopy. To avoid protein and/complex precipitation upon RNA addition, the solutions containing the two molecules were mixed at a larger volume and then co-concentrated to an NMR sample volume (2.2.2.4.2). No precipitation was encountered, when the RNA/protein complex was prepared in this manner.

Addition of IRD has caused minor chemical shift changes (Table 2.07) in a triple labelled, ^{2}H, ^{15}N, ^{13}C p100 protein sample (Figure 2.33A). These changes are mostly localized to three regions in the protein: The second half of the long loop between β3 and α1 in the SN3 domain (between residues 395 and 410 – Figure 2.32B), the end of the linker and the N-terminus of the SN4 domain (between residues 508 and 524 – Figure 2.33B) and the loop between β9 and α6 (between residues 568 and 585 – Figure 2.32B). In particular, residues V396 and E410 in SN3, C570 and E595 in SN4 experienced the largest chemical shift changes upon IRD addition. These changes occur both in NMR Buffer 2.1 and 2.2 (at the presence or absence of EDTA), so they most likely are not related to p100's potential nuclease ability. Interestingly, the first and third regions most affected by RNA binding, were the ones which were shown to be more flexible in the relaxation studies above (Figure 2.31 and 2.33B).
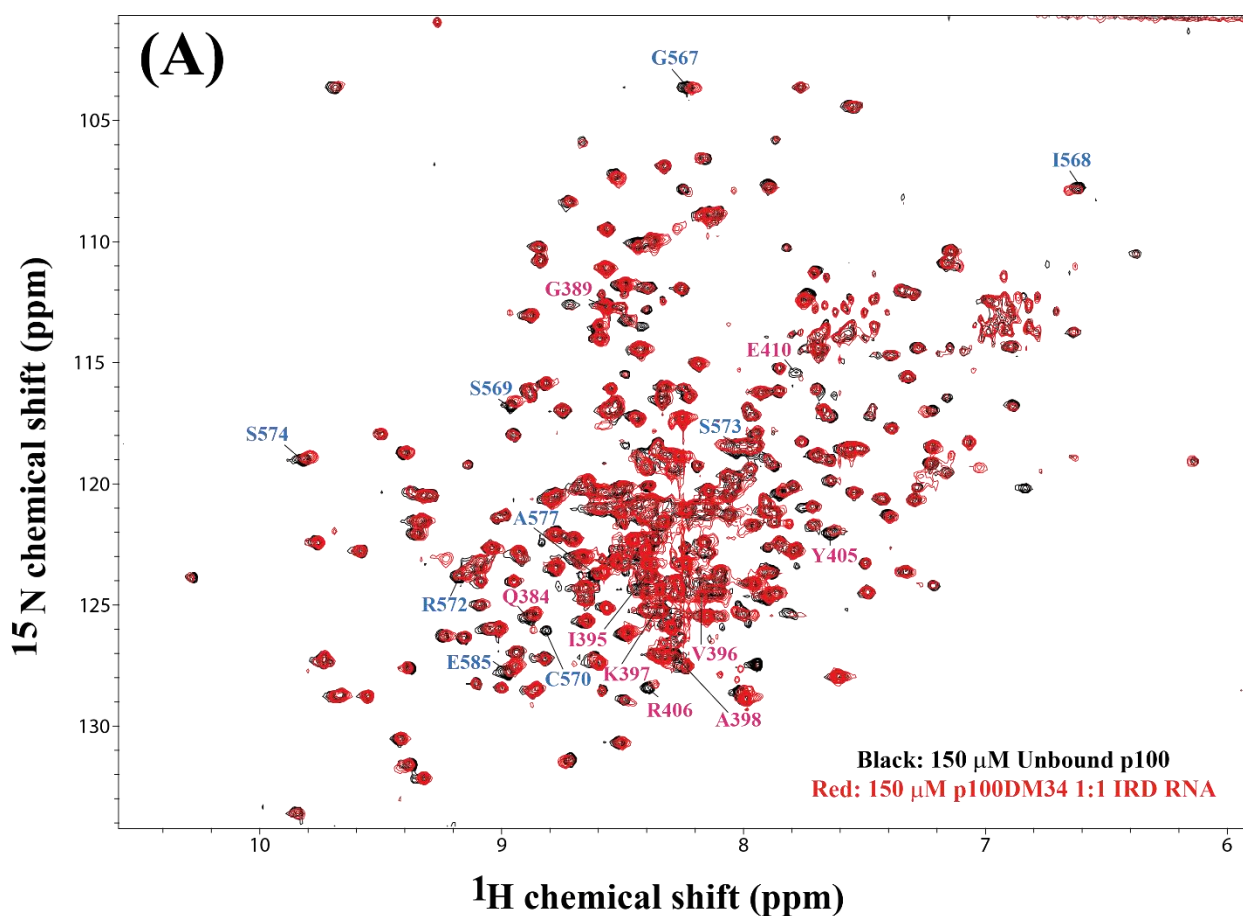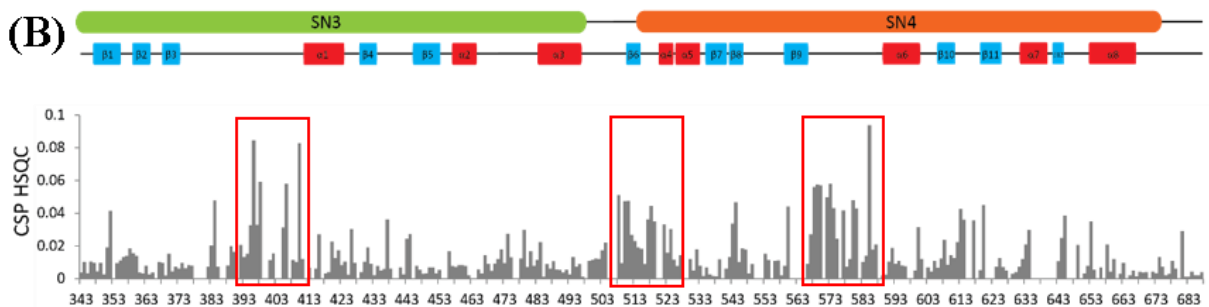
Figure 2.33 – **(A)** An overlay of two 2D ¹H, ¹⁵N HSQC of free and IRD (in black) and IRD RNA-bound p100DM34. Signals experiencing the largest degree of chemical shift changes upon IRD addition are marked in pink and light blue for SN3 and SN4 domains respectively. Buffer conditions: NMR Buffers 2.1. Spectrum and chemical shift changes remain unchanged when NMR buffer 2.2 is used. **(B)** Chemical shift changes in p100DM34 upon addition of IRD RNA. Domain and secondary structure elements are denoted above the respective residues. Loops with significant chemical shift changes are boxed in red.

In work performed by Martin Rübbelke, hetNOEs were measured for the p100/IRD RNA complex. The data for the flexible loops have smaller errors compared to the rigid parts of the protein. The hetNOE values for the loop between β3 and α1 in SN3 (residues 392-394 and 396-397 – Figure 2.34A) and the loop between β9 and α6 in SN4 (residues 577-578, 580-581, 583-584 and 587; Figure 2.34B) increase slightly, but are not significantly higher in the complex compared to the free protein. Some of these residues also experience a marked chemical shift changes upon IRD addition. This might be caused by binding of the flexible loops to the edited dsRNA, which would make the flexible loops more rigid.

(A)

(B)
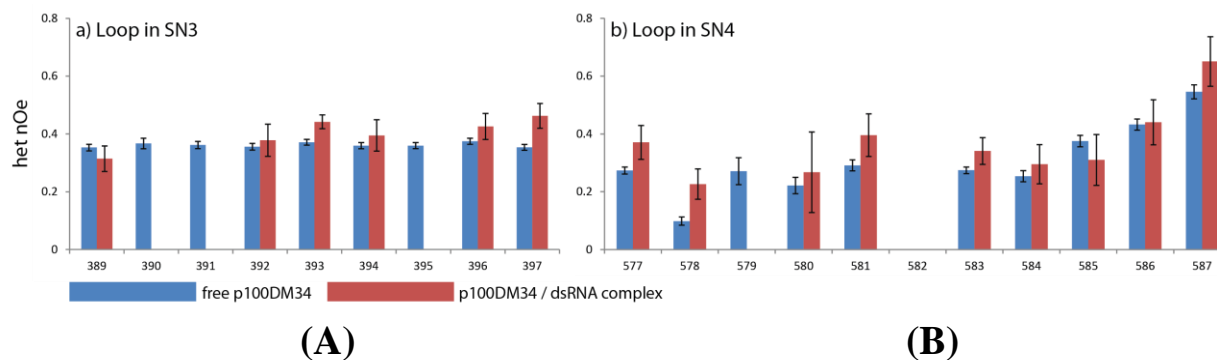
Figure 2.34 – A comparison of hetNOE values for free (blue) and IRD- bound (red) p100 DM34. **(A)** Data from residues of the loop between β3 and α1 in SN3. **(B)** Data from residues of the loop between β9 and α6 in SN4.

SN3		SN4	
Residue	$\Delta\delta_{1H,15N}$	Residue	$\Delta\delta_{1H,15N}$
D383	0.020	L560	0.044
Q384	0.048	A566	0.009
R385	0.007	G567	0.027
V388	0.008	I568	0.056
G389	**0.020**	S569	0.058
T390	0.016	C570	0.057
G392	**0.021**	R572	0.050
E393	**0.013**	S573	0.058
E394	**0.015**	S574	0.043
I395	**0.033**	R575	0.024
V396	**0.085**	**A577**	**0.042**
K397	**0.033**	**L578**	**0.007**
A398	0.059	**N579**	**0.012**
R401	0.011	**G580**	**0.048**
G402	0.015	**V581**	**0.043**
Y405	0.031	**A583**	**0.010**
R406	0.058	**Q584**	**0.014**
L408	0.011	**E585**	**0.094**
Y409	0.010	**G586**	**0.018**
E410	0.083	**E587**	**0.021**

Table 2.07 – List of p100DM34 experiencing significant chemical shift changes upon IRD RNA addition. Residues which exhibit an accompanying increase in their hetNOE values upon complexation, are marked in bold.

The effect of p100DM34 on IRD was also examined, by monitoring the changes in the imino region of this RNA which occur upon protein addition using a one-dimensional experiment (Figure 2.35A). As no imino resonances were observed for the inosine-containing motif, it is not possible to determine whether the inosine residues and their uracil counter-bases are affected by p100DM34 addition using this experiment. However the largest signal shifts were seen for the observable imino resonances close to this motif, namely for G34, G28, U14 and to a lesser extent, U35 (Figure 2.35B). Other imino signals of residues in the IRD further removed from the inosine-containing motif did not experience any changes upon the addition of p100DM34 (Figure 2.35A and 2.35B). These findings strongly suggest that the region of IRD RNA most strongly affected by RNA binding is the vicinity of the inosine-containing motif.
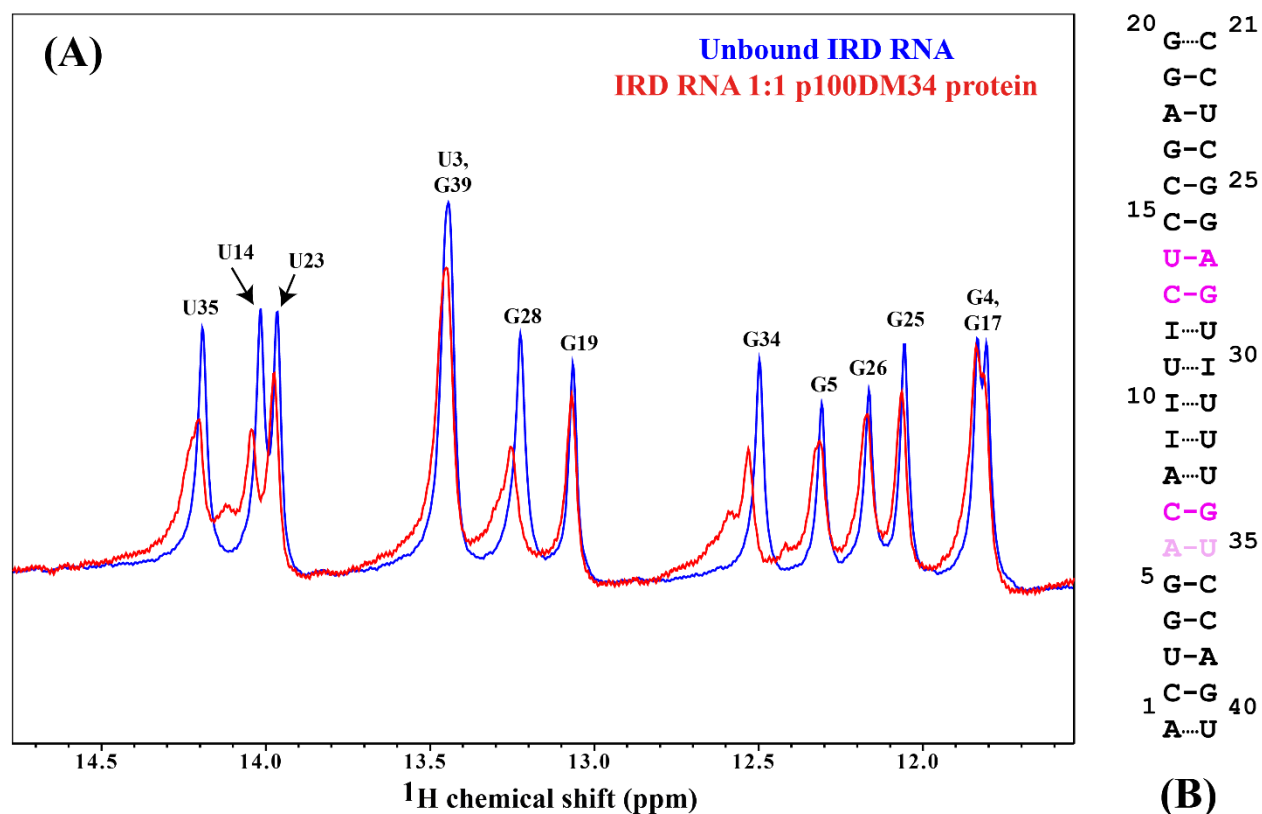
Figure 2.35 – (A) The imino region of the one-dimensional spectrum of IRD RNA in the free (in blue) form and upon addition of p100DM34 (red). The largest changes occur for signals closest to the inosine-containing motif **(B)** Changes in IRD upon p100DM34 plotted on this RNA's secondary structure. Residues (as well as their base-pairing partners) that undergo the most significant chemical shift changes are marked in bright pink, residues (as well as their base-pairing partners) which undergo smaller shifts are marked in pale pink, while other residues are marked in black.

2.3.10 p100DM34 mutants

To investigate whether certain p100DM34 mutants affect the protein fold and function, four such mutants were generated, expressed and purified and their NMR spectra compared to that of the wild type form of p100DM34. The residues chosen for mutagenesis were found to coordinate ions in the crystal structure of the human p100 (Li, Yang et al. 2008). Four such residues are outlined in Table 2.08. In SN3 and SN4 and Arg374, Arg438, Arg562 are basic, and hence may be implicated in the recognition of the phosphodiester backbone of the IRD RNA. Ser538 in SN4 may have a catalytic function in potentially cleaving the inosine-containing motif in the IRD. Three of the four residues are conserved both in the human and the *Drosophila* variants of p100, with Arg438 in the former is replaced with Lys447, an amino acid with similarly basic properties. Hence

they would be good candidates for mutagenesis, as their alterations may result in the abolition of p100 binding and catalytic function.

Domain	Human	*Drosophila*	Mutation
SN3	Arg374	**Arg379**	Glu
SN3	Arg438	**Lys447**	Glu
SN4	Ser538	**Ser548**	Ala
SN4	Arg562	**Arg572**	Glu

Table 2.08 – Four residues coordinating citrate ions in the human p100 crystal structure, their *Drosophila* equivalent, their number in the p100DM34 construct and the resulting mutant.

The residues were mutated as follows: The three basic amino acids (R379, K447 and R572) were mutated to glutamate which has a negatively charged side-chain which should result in a reduction of affinity to the negatively charged backbone of the RNA, while S548 was mutated into an alanine replacing its potentially catalytic polar hydroxyl with a hydrophobic methyl group (Table 2.08).

The 2D ^{1}H, ^{15}N HSQC spectra of all four mutant protein constructs show large similarity to the spectrum of wild type p100DM34 (Figure 2.36). This is an indication that these single mutations do not affect the global protein fold. Mutant binding studies with IRD were not performed because of the already very weak binding affinity of the wild type p100DM34 to this RNA (2.3.9).

When subjected to an RNAse activity test, all mutant have displayed RNAse activity similar to that of the wild type protein (Figure 2.37). One exception was the K447E (Figure 2.37B) mutant which has displayed a somewhat reduced RNAse activity, however this reduction was quite minor. The mutation of S548 to alanine, did not decrease RNAse activity either (Figure 2.37C), indicating that this residue is most likely not participate in the cleavage of edited RNA.
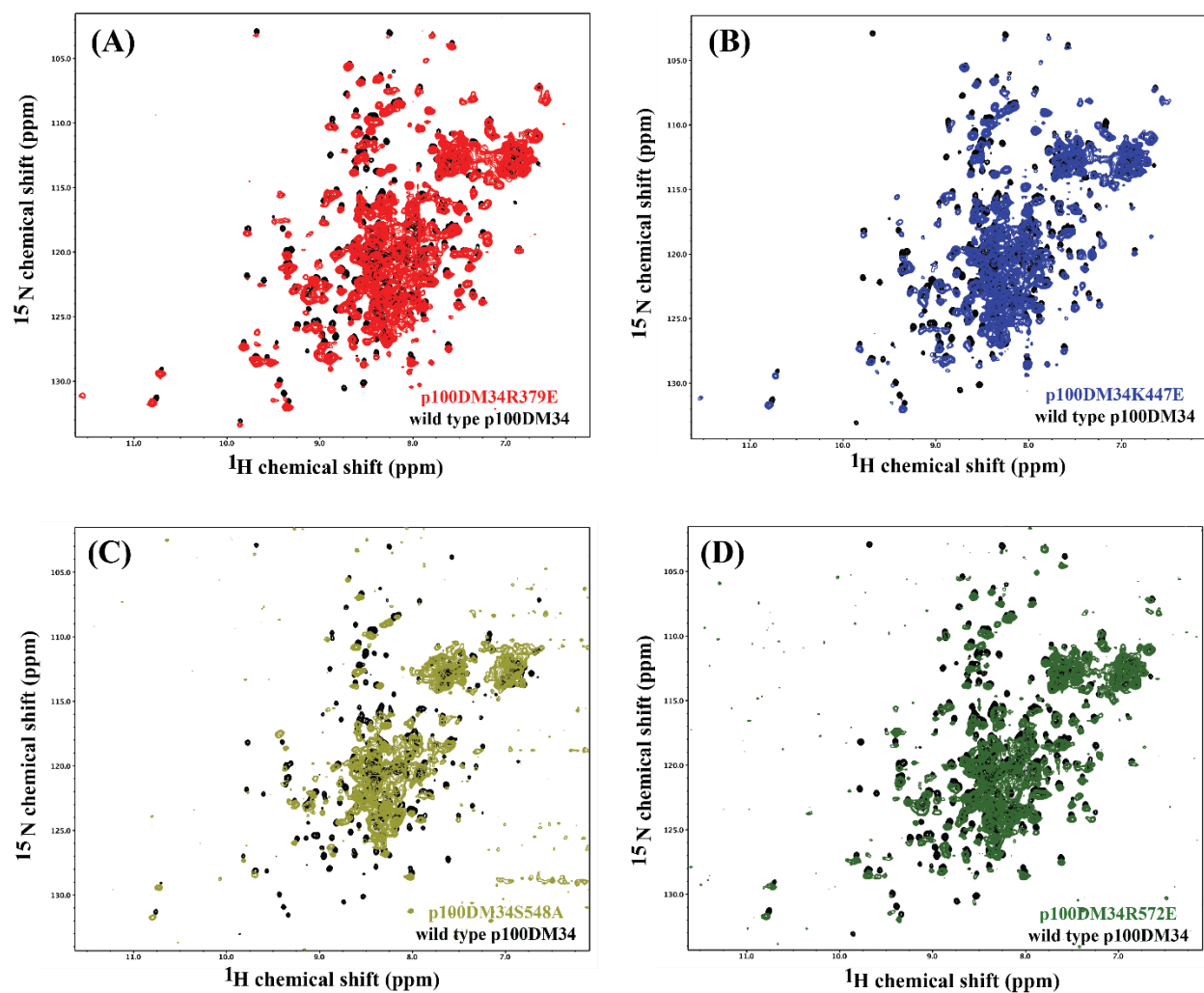
Figure 2.36 – 2D ^{1}H, ^{15}N HSQC overlays of wild type p100DM34 in black with **(A)** R379E in red, **(B)** K447E in blue, **(C)** S548A in beige, **(D)** R572E in green. Buffer conditions: NMR Buffer 1.
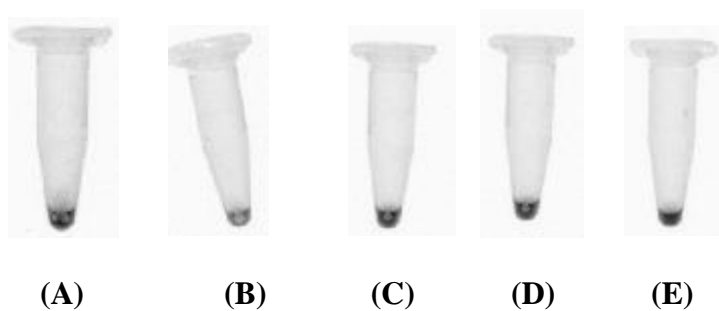
Figure 2.37 – RNAse activity test for **(A)** R379E, **(B)** K447E, **(C)** S548A, **(D)** R572E, **(E)** wild type p100DM34 protein constructs.

2.4 Discussion

2.4.1 Properties of IU-containing RNA

In this work, various inosine containing RNA constructs were studied using native-PAGE as well as NMR spectroscopy. One such construct, a heterodimeric 20mer RNA which was earlier identified to be the binding partner as well as the potential cleavage substrate for the protein p100 (Scadden 2005, Li, Yang et al. 2008), was subjected to a more rigorous structural as well as biophysical analysis using additional NMR and UV melt experiments respectively.

2.4.1.1 Orientation of the IU base-pairs in IRD

In an IU wobble, two base-pairing configurations via its Watson-Crick edge are possible (Janke, Riechert-Krause et al. 2011). In one such configuration, the inosine's N1 imino hydrogen-bonds to the uracil's O2 carbonyl group, while the uracil imino in turn acts as a hydrogen-bond donor for the inosine's O6 (Figure 2.38A). This configuration is commonly found in GU base-pairs. In another configuration, the latter hydrogen bond remains the same, while in the former the uracil O4 carbonyl replaces O2 a hydrogen bond acceptor (Figure 2.38B).
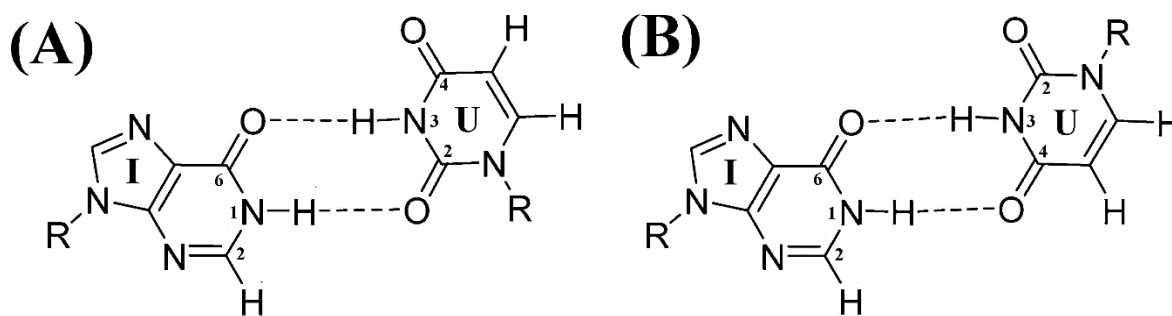
Figure 2.38 – Two different IU base-pairing orientations. **(A)** Inosine imino (N1) hydrogen bonds with O2, while the uracil N3 imino makes contact with the O6 inosine carbonyl. **(B)** Inosine imino hydrogen bonds with O6, while the uracil N3 imino makes contact with the O6 inosine carbonyl as in A (Janke, Riechert-Krause et al. 2011).

However, the second configuration requires one of the residues participating in the interaction to adopt a *syn* conformation (Das and Lyngdoh 2008). Since no IRD residues were found to adopt this conformation (Section 2.3.3.2), it is the first base-pairing configuration which is exclusively present in this RNA.

2.4.1.2 Stability of IU base-pairs

The lack of observable imino resonances in the IIUI motif in the IRD water NOESY (Section 2.3.3.1), as well as broadened long-range HSQC signals for I10 and I30 (Section 2.3.4.6) points to an increased conformational flexibility for the IU base-pairs of this motif.

Previous structural work on an RNA homodimer containing two consecutive interchanging IU base-pairs suggests that these base-pairs are similar to GU base-pairs in their C1'-C1' distances, B-factors and stacking in the crystal structure (Pan, Mitra et al. 1998). However, the angles between the glycosidic bonds and the C1'-C1' vector vart from one crystal to another (Pan, Mitra et al. 1998). In addition, when superimposed onto an AU base-pair, the orientation IU/UI tandem base-pairs varies in different crystals (Pan, Mitra et al. 1998). In the first crystal, I4 and I12 rotate towards the minor groove, U13 (I4's counterpart) points towards the major groove, while the position of U5 (counterpart of I4) remains unchanged (Figures 2.39A and 2.39B). In the second and third crystals however, the uracil residues are rotated towards the major groove, while the inosine residues do not move (Figure 2.39C). These two findings suggest that IU base-pairs possess larger conformational flexibility than its AU counterparts.
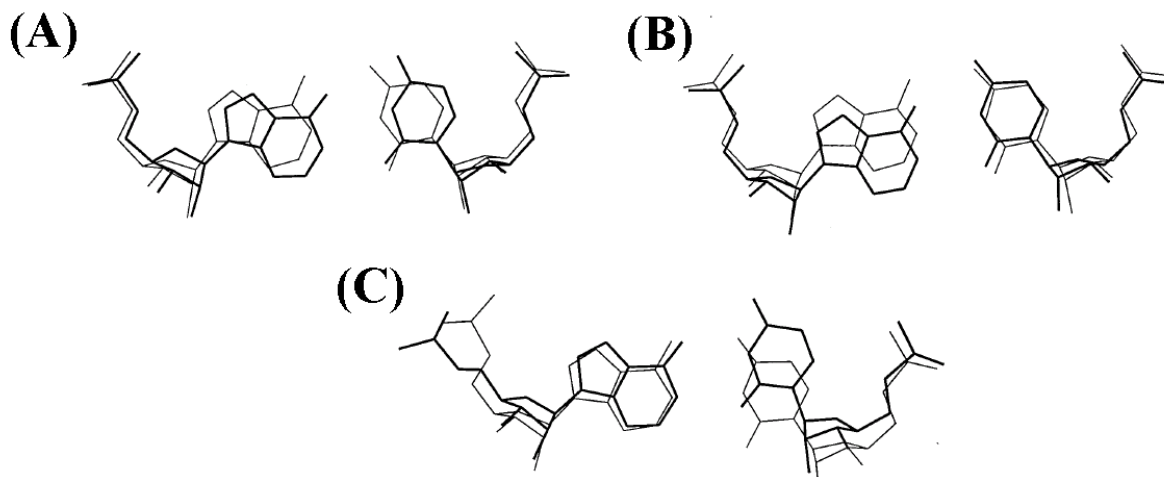
Figure 2.39 – Three different rotation patterns of the wobble bases observed on superimposition of the IU wobbles (thick lines) on the Watson-Crick A·U base pairs (thin lines). **(A)** I4-U13 in duplex 1, **(B)** U5-I12 in duplex 1 and **(C)** IU wobbles in duplexes 2 and 3 (Pan, Mitra et al. 1998).

When comparing IU (Figure 2.40A) to GU (Figure 2.40B) and IC (Figure 2.40C) base-pairs, important differences emerge. While all three base-pairs form two hydrogen bonds, the structural basis of their stability is different.

In a GU base-pair, the wobble mismatch causes the participating residues to rotate in order to accommodate the N1(G)-O2(U) and N3(U)-O6(G) hydrogen bond formation (Figure 2.40B), resulting in a kink in the RNA. However a bridging water molecule between the N2(G) amino and O6(U) carbonyl moieties acts in lieu of a third hydrogen bond (Biswas and Sundaralingam 1997, Biswas, Wahl et al. 1997), stabilizing this base-pair (Figure 2.40B).

IC base-pairs do not possess a bridging water molecule, due to the lack of an N2 amino moiety and the inability of the H2 atom to form hydrogen bonds. However, since both the inosine and cytosine are located in an ideal position to form the N1(I)-N3(C) and N4(C)-O6(I) hydrogen bonds (Figure 2.40C), the base twist present in a GU base-pair is absent here, resulting in a reduced strain on the RNA's helical geometry.
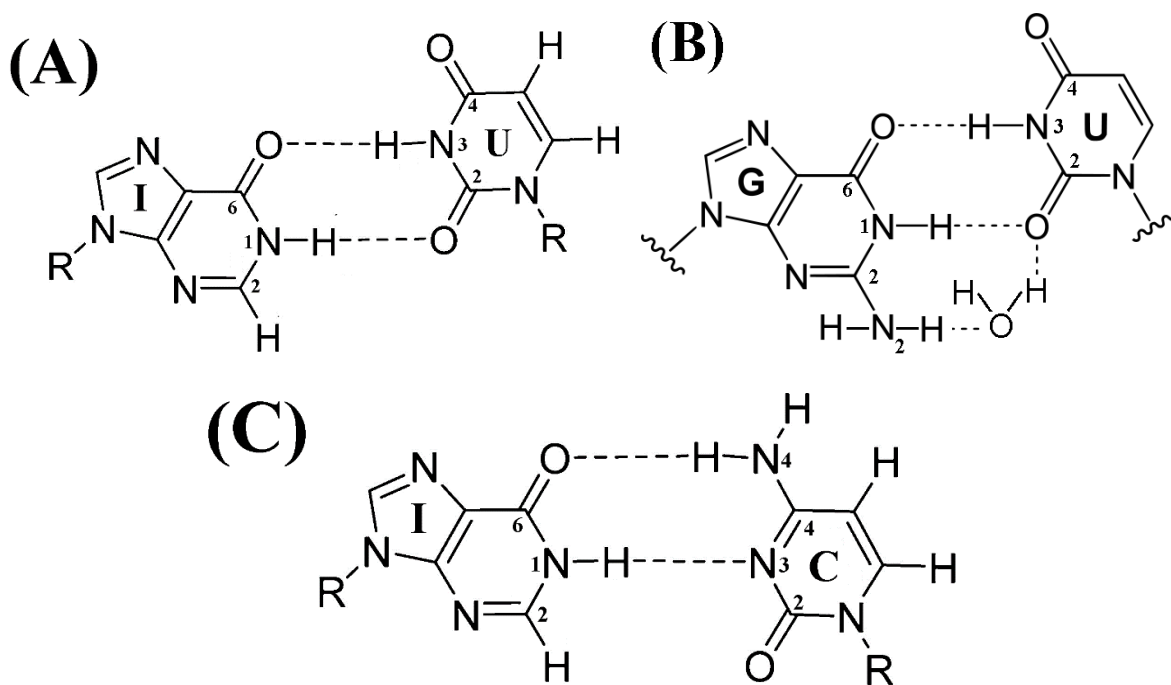
Figure 2.40 – A comparison between three kinds of non-canonical base-pairing. **(A)** An IU base-pair possesses the twist necessary to facilitate hydrogen bonding similar to a GU base-pair **(B)** A GU base-pair, where the bases are rotated to accommodate the hydrogen bond formation, however, the interaction is stabilized by a bridging water molecule. **(C)** In an IC base-pair, the bases are already located in an optimal position for the two-hydrogen bond interaction to take place.

IU base-pairs (Figure 2.40A) lack both of the stabilizing factors present in IC and GU base-pairs. Since the hydrogen bonding network in an IU base-pair is identical to the one in GU wobble (Figure 2.40A and 2.40B), a kink in the RNA (and consequentially a distortion in its helicity) also occurs. However, unlike in a GU wobble (and similar to an IC base-pair), the lack of an N2 amino

precludes the presence of a bridging water molecule. Therefore IU base-pairs are clearly less stable than its IC and GU counterparts, and their presence leads to a greater destabilization of the RNA helical structure, as has been demonstrated by the native gel and UV melt studies in this work (sections 2.3.1 and 2.3.2 respectively). Superior stability of GU base-pair over IU ones, is also corroborated by the water NOESY spectra of GRD and IRD (Section 2.3.3.1), where all the imino resonances of the four GU wobbles are observed, while no IIUI motif imino signals are present in the IRD spectrum, respectively.

Additional work also supports the above-mentioned conclusion. When inosine and uracil nucleosides are mixed at 1:1 ratio at 123 K, only one uracil in twenty base-pairs with inosine (Janke, Riechert-Krause et al. 2011). The rest of the uracil dimerizes with itself. Significant homodimerization is also present for insoine (Janke, Riechert-Krause et al. 2011). In contrast, when inosine and cytosine nucleosides are mixed together, the heterodimeric species forms almost exclusively, and only trace amounts of inosine homodimers are observed (Janke, Riechert-Krause et al. 2011). In another study where the impact of non-canonical base-pairing on DNA stability was investigated IC base-pairs were found to be significantly more stable than their IT counterparts (Watkins and SantaLucia 2005).

2.4.1.3 Impact of IU base-pairs on RNA structure

I12HO (Figure 2.10, lane 1) and I16HO (Figure 2.11) palindromic RNA constructs containing four IU base-pairs failed to dimerize (in case of I16HO, dimerize to a significant extent). In contrast, A12HO (Figure 2.10, lane 2) where the inosine residues are replaced with adenine, as well as palindromic 12mer RNA construct containing four GU base-pairs dimerize to completion (data not shown). While the heterodimeric IRD readily dimerizes (Figure 2.12A), the palindromic I20HO construct did not (Figure 2.12B). As the I20HO flanking sequence exclusively contains GC base-pairs, and the IRD has about the same amount of GC and AU base-pairs, pointing to the importance of the flanking sequence in duplex stability. Although GC base-pairs should in principle impart more stability on duplexes, compared to their AU counterparts (due to the fact that they form three hydrogen bonds as opposed to two), evidently, a certain ratio between them has to be maintained in order for the construct to fully dimerize, as larger amounts GC base-pairs, evidently produces a destabilizing effect on duplex formation. These findings indicate that the presence of IU base-pairs impairs the stability of RNA A-form helices.

Previous work has clearly demonstrated that IU motifs reduce the helical stability of RNA constructs. The introduction of an IU tandem in the place of an AU or GU base-pairs, has shown to reduce the melting temperature of RNA duplexes by up to 25 C and 16 C, respectively (Serra, Smolter et al. 2004). This significant reduction further emphasizes the intrinsic instability of IU base-pairs compared to canonical as well as non-canonical GU as well as IC base-pairs, further corroborating our findings regarding the poor dimerization ability of IU containing constructs less than 20 nucleotides in length.

The effect exerted by IU base-pairs on RNA duplexes may help to address a certain property of RNA editing. While larger RNAs are hyper-edited by ADARs quite readily, the amount of editing in smaller duplexes is drastically reduced (Scadden and Smith 1997, Lehmann and Bass 1999, Scadden and Smith 2001). RNAs smaller than 15 nucleotides cannot undergo RNA editing at all (Bass 2002). It is possible that RNA editing can only take place on an RNA duplex of a minimal size, because A to I conversion on shorter RNA may result in the unwinding of the edited dimer product. Therefore the ADARs have evolved double stranded RNA binding motifs (dsRBMs) which facilitate the recognition of a substrate with a certain minimal length(Macbeth, Lingam et al. 2004) to ensure the integrity of the edited dimer product.

2.4.2 RNA-binding and catalytic properties of p100

In order to cleave the nucleic acid substrate, the staphylococcal nuclease domain requires the presence of certain residues. Asp40, Val41, Glu43 all coordinate a Ca^{2+} ion which in turn polarizes the phosophodiester bond. This bond also forms hydrogen bonds with Arg87 and Arg35 (Cotton, Hazen et al. 1979). The catalysis itself is initiated by the side chain of Glu43 which acts as a general base, to abstract a proton from an attacking water molecule (Cotton, Hazen et al. 1979). The attack results in a trigonal bypyrimidal transition state, and the subsequent cleavage of the phosophodiester bond by the protonation of one of its of oxygen by the guanadium side chain of Arg87 or Tyr113 (Cotton, Hazen et al. 1979). Also of importance are Lys84 and Tyr85 which for hydrogen bonds with the 3' phosophodiester group (Cotton, Hazen et al. 1979).

All of the aforementioned residues, with the exception of Lys84 are not present neither in the human or the *Drosophila* variants of p100 (Figure 2.41). Tyr113 is replaced by Phe which is a similar amino acid residue, however since the former phenol group may be implicated in catalysis, this substitution is significant. The lack of so many catalytically important residues indicates that

either p100 does not cleave the edited RNA which has been previously suggested (Scadden 2005), or the cleavage takes place via another mechanism. Unfortunately further insight into this potential mechanism could not be obtained, as the interaction of p100 was very weak and no clear IRD cleavage products were observed in the NMR spectra as well as in the gel nuclease assays (data not shown).

```
SN        MTFRLLLVDVPETKHPK-KGV-----EKYGPEASAFTKKMVENAKKIEVEFNKGQRTDKY    85
p100DM34  VTFLLAGISCPRSSRPALNGVPAQEGEPFGDEALTFTRERVLQ-RDVSVHIDT---TDKA   616
p100HS34  ITFLLAGIECPRGAR-NLPGL-VQEGEPFSEEATLFTKELVLQ-REVEVEVES----DKA   579


SN        GRG-LAYIYAD-GKMVNEALVRQGLAKVAYVYKPNNTHEQHLRKSEAQAKKEKLNIWS    141
p100DM34  GSSVIGWLWTDSGANLSVALVEEGLAEVHFSAEKSEYYRQ-LKIAEDRAKAAKKNIWT   673
p100HS34  G-NFIGWLHID-GANLSVLLVEHALSKVHFTAERSSYYKS-LLSAEEAAKQKKEKVWA   645
```

Figure 2.41 – Sequence alignment of the staphylococcal nuclease (SN), the *Drosophila* (p100DM34) and human (p100HS34) p100 proteins. Residues essential in the SN catalytic mechanism are denoted in yellow. In the p100DM34 sequence, Lys84 which is also present in its SN counterpart and Phe113 which substitutes for Tyr are marked in yellow and dark yellow respectively, all other corresponding residues are marked in blue.

However, the mutant studies do provide some insight into this matter. All four mutant candidates are residues which are conserved in both the human and the *Drosophila* variant of p100 (with the exception of the second mutant where the arginine in p100HS34 which is replaced with the similar lysine in p100DM34 – Figure 2.43). These residues all coordinate two co-crystallized citrate ions in the p100 human crystal structure (Li, Yang et al. 2008). Citrate ions provide a possible hint as to the location of a nucleic acid binding site, as they mimic its phosophodiester backbone. A slightly RNAse reduced activity in the K447E mutant, may indicate a further loss in binding and consequentially cleavage ability of the IRD substrate. However, as mentioned, this reduction is quite minor. p100DM34 Constructs containing multiple mutants should be expressed, purified and assayed for RNAse activity to determine whether a multitude of mutations can further reduce the RNA binding and cleavage activity of p100.

2.4.3 Structural basis of p100/IRD interaction

Since the imino resonances of the IIUI motif are not observed, it is impossible to determine the direct effect of p100 addition on this portion of the RNA. However, IRD chemical shift changes occurring after p100 addition in the resonances flanking this motif (Figure 2.35) points to IRD recognition by p100 this region of the RNA. The importance of the IIUI motif in p100 recognition

has also been demonstrated in earlier work (Scadden 2005, Li, Yang et al. 2008). It is likely, that the dynamic nature of the IU mismatches (especially that of I10-U31 and U11-I30 – Figure 2.22C), and their flanking base-pairs creates a potential bubble which serves as a binding target for p100 (Figure 2.42). In contrast no such bubble is formed in the more stable ARD and GRD duplexes which precludes p100 recognition. This hypothesis has been raised previously (Li, Yang et al. 2008), and the results presented here lend it further support
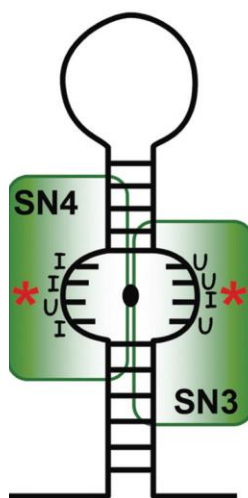
Figure 2.42 – The IIUI motif of the IRD may have an open conformation. The phosphate backbones may displace from a canonical A-form conformation so that they can be bound by the loops in the SN3 and SN4 domains in p100 (Li, Yang et al. 2008).

From the protein side, the bulk of the interaction seems to take place in loops of the SN3 and SN4 domains (Figure 2.33). The residues of these loops are partially conserved in both the human and *Drosophila* p100 (Figure 2.43). Increase in the hetNOE values for some of these residues upon IRD addition to p100DM34, suggests that complexation results in the reduction of their flexibility (Figure 2.34). However the global p100 domain fold remains largely the same in its RNA-bound form, as the HSQC peak pattern do not experience significant changes upon incubation with p100 (Figure 2.33A).

```
p100DM34    EKDFSGTVVEVFNGDAINVRLSNGQVKKVFFSSIRPPRDQRAVVGTDGEEIVKAPPRGKN    404
p100HS34    DKQFVAKVMQVLNADAIVVKLNSGDYKTIHLSSIRPPR-------LEGE---NTQDKNKK    386

p100DM34    YRPLYEIPHMFDAREFLRKKLINKKVQCNLDYI---SPPRENFP---EKYCYTVSIGGQN    458
p100HS34    LRPLYDIPYMFEAREFLRKKLIGKKVNVTVDYIRPASPATETVPAFSERTCATVTIGGIN    446

p100DM34    VAEAMVAKGLATCVRYRQDDDQRSSAYDQLIAAEQQAIKGLKGLHAKKDNATLRVNDLTV    518
p100HS34    IAEALVSKGLATVIRYRQDDDQRSSHYDELLAAEARAIKNGKGLHSKKEVPIHRVADISG    506

p100DM34    DHSRIKVQYLPSWQRALRTEAIVEFVASGSRLRIFVPKDSCLVTFLLAGISCPRSSR--P    576
p100HS34    DTQKAK-QFLPFLQRAGRSEAVVEYVFSGSRLKLYLPKETCLITFLLAGIECPRGARNLP    565

p100DM34    ALNGVPAQEGEPFGDEALTFTRERVLQRDVSVHIDTTDKAGSSVIGWLWTDSGANLSVAL    636
p100HS34    GL----VQEGEPFSEEATLFTKELVLQREVEVEVESMDKAG-NFIGWLHID-GANLSVLL    621

p100DM34    VEEGLAEVHFSAEKSEYYRQLKIAEDRAKAAKKNIWTNYVEEVPKEKTV     685
p100HS34    VEHALSKVHFTAERSSYYKSLLSAEEAAKQKKEKVWAHY-EEQPVEEVM     670
```

Figure 2.43 – Sequence alignment of p100DM34 and p100HS34. Regions in p100DM34 experiencing chemical shift changes upon IRD addition, are marked in light green. In p100HS34, identical residues are also marked in light green, residues similar in structure (i.e. belong to the same amino acid class, for instance both are basic, or acidic) are marked in dark green, while completely different residues are marked in brown. Three out of the four mutants are marked in red in both sequences, while R572 (R562 in the human p100) is marked in red contours and green interior as in the wild type p100DM34 it experiences chemical perturbation upon IRD addition.

The p100DM34/IRD interaction observed here is of a very weak affinity. It was previously demonstrated that truncated versions of the human p100 bind the IRD with an affinity about three and half times lower than the wild type (Li, Yang et al. 2008), however here the affinity is even further reduced. The results presented here are more in line with those published by Scadden where recombinant p100 alone showed very weak binding activity towards IRD (Scadden 2005). Perhaps additional factors are required to enhance p100 binding and cleavage activity. Elucidating those potential factors and dissecting their effect (if any) on p100 binding and cleavage activity can be a subject of future studies in this field.

Chapter 3 –The interaction between CC mismatch containing RNA with Hoechst 33258

This chapter is based on the following manuscript: Garg, D., A. V. Beribisky, G. Ponterini, A. Ligabue, G. Marverti, A. Martello, M. P. Costi, M. Sattler and R. C. Wade (2013). "Translational repression of thymidylate synthase by targeting its mRNA." <u>Nucleic Acids Res</u> **41**(7): 4159-4170.

3.1 Introduction

3.1.1 Introduction to Thymidylate synthase

Thymidylate synthase (TS) catalyses the reductive methylation of 2′-deoxyuridine-5′-monophosphate (dUMP) to 2′-deoxythymidine-5′-monophosphate (dTMP). It plays a key role in the biosynthetic pathway that, in human cells, provides the sole *de novo* source of thymidylate. As thymidylate is an essential precursor required for DNA replication and repair (Garg, Henrich et al. 2010), TS is an attractive anti-cancer target. Drugs inhibiting the TS protein lead to its overexpression and consequent development of drug resistance. The effect has been attributed, at least in part, to the interruption of the autoregulatory feedback loop caused by disruption of the TS protein–TS mRNA interaction (Chu, Koeller et al. 1991). TS protein binds its own mRNA at two locations. One location is the so-called site I which spans residues 75-110 and contains the AUG start codon shown to be necessary for TS binding (Chu, Voeller et al. 1993). The other area located further downstream in the mRNA, and is termed site II (Chu, Voeller et al. 1993). It is thought that as TS levels rise, this protein binds its own mRNA, at those two sites, thereby inhibiting its own translation. Therefore, repression of TS translation by targeting the TS mRNA could provide an effective means of overcoming the development of resistance.

3.1.2 The CC mismatch in TS mRNA as a small molecule binding site

The TS-mRNA sequence spanning nucleotides 75–110, referred to as Site I, encompasses the start codon and has been shown to be relevant for the TS mRNA–protein interaction. It is predicted to have a stable stem-loop structure with a CC mismatch in the stem (Figure 3.01A) (Chu, Voeller et al. 1993). Previous work on dyes such as HOECHST 33258 (HT), as well as aminoglycosides (Figure 3.01C), with Site I-like RNA constructs, among them the TS mRNA construct (TSMC – Figure 3.01B, leftmost construct), suggests that these bind at or in the vicinity of the CC mismatch (Tok, Cho et al. 1999, Cho and Rando 2000, Tavares, Beribisky et al. 2009). When the CC mismatch was replaced with a GC base-pair (TSGC – Figure 3.01B middle construct), binding was abolished for both of the aforementioned molecule classes (Tok, Cho et al. 1999, Cho and Rando 2000, Tavares, Beribisky et al. 2009). Therefore, Site I provides a functionally relevant structural motif that may be targeted to develop novel anti-cancer drugs.
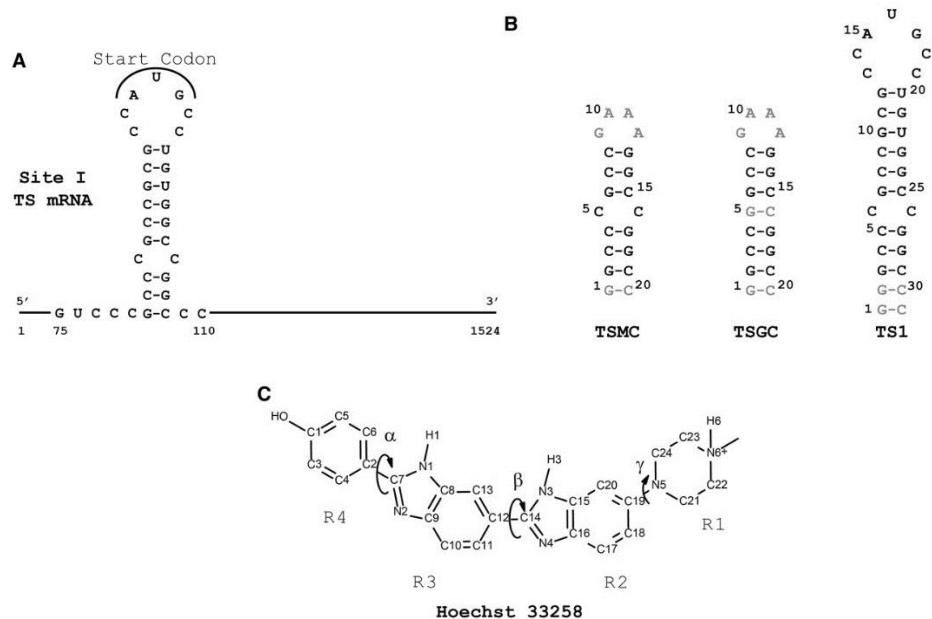
Figure 3.01 (A) Depiction of TS mRNA with the predicted stem-loop secondary structure of Site I (nucleotides 75–110) containing the start codon. **(B)** Predicted secondary structures of the RNA constructs, from left to right, TSMC, TSGC and TS1. The sequence and structural elements of Site I preserved in the RNA constructs are shown in black, whereas the rest of the RNA construct is gray. **(C)** 2D structure of Hoechst 33258 (HT) with the atom, ring and torsion angle nomenclature used in the text. Adapted from Garg et al., 2013 (Garg, Beribisky et al. 2013).

HT is a fluorescent dye widely known to bind at the minor groove of AT-rich B-DNA (Vega, Garcia Saez et al. 1994). However in GC-rich DNA regions, partial intercalation has been observed (Bailly, Colson et al. 1993, Colson, Houssier et al. 1995, Adhikary, Buschmann et al. 2003). Based on electric linear dichroism studies, HT has also been proposed to intercalate at the pyrimidine bulge in the trans-activation response (TAR) RNA of HIV-1, although the authors questioned the structural viability of intercalation due to the steric hindrance of the bulky methylpiperazine and hydroxyl phenyl groups on either side of the bis-benzimidazole fragment of HT (Bailly, Colson et al. 1996). HT is readily taken into cells and is known to have moderate anti-leukemic and anti-helminthic activity, although the mechanism of cytotoxicity is not well understood (Baraldi, Bovero et al. 2004). HT has been observed to bind specifically to a TS RNA Site I construct with a dissociation constant of 60 ± 13 nM; the binding is facilitated by the presence of a CC mismatch and is comparable to the binding of the aminoglycoside, paromomycin (Cho and Rando 2000). Mutation and RNAse footprinting experiments indicated that the specific binding of HT requires non-duplex RNA, as well as the presence of GC base pairs adjacent to the mismatch. However binding independent of the base type at the bubble (Cho and Rando 2000).

126

3.1.3 Aims of the work

To investigate the biological relevance of the interaction of HT with the TS mRNA, we performed cell-based assays and monitored the effect of HT on the levels of TS mRNA and protein. Surprisingly, we observed that HT reduced the TS protein levels by acting at the level of translational regulation, raising the possibility that HT might directly interact with the TS mRNA in the cell. To exploit HT as a lead compound for the design of anti-cancer agents targeting the TS mRNA, a detailed structural characterization of HT-mRNA binding is desirable. Since the CC-mediated HT binding site on TS mRNA (Cho and Rando 2000) is structurally distinct from the HT binding site observed for the TAR RNA (Bailly, Colson et al. 1996), a direct deduction of the binding mode from that for TAR RNA is not feasible. We therefore studied the molecular details of HT-TS mRNA interactions using nuclear magnetic resonance (NMR), UV-Vis and fluorescence spectroscopy techniques, complemented with computational docking and molecular simulations. For this purpose, we analyzed three RNA constructs: TSMC, TSGC, TS1 (Figure 3.01B, left to right). TS1 has the native predicted stem loop structure of Site I, stabilized by two additional GC base pairs; its interaction with HT was reported by Cho *et al.* (Cho and Rando 2000). TSMC is a shorter construct that has the same three base pairs as Site I flanking the CC mismatch in both directions; its interaction with paromomycin has previously been studied by NMR (Tavares, Beribisky et al. 2009). The CC mismatch of TSMC has been replaced by a GC base pair in TSGC. Our data show that HT binds the Site I-like RNA constructs in an ensemble of modes with intercalation at the site of the CC bubble being the dominant binding mode.

3.2 Materials and Methods

3.2.1 NMR spectroscopy

TSMC, TSGC RNA were purchased from IBA GmbH. RNA samples were measured at 100 μM concentration, in 90% H_2O + 10% ^2H_2O or 100% ^2H_2O, as required. The pH value was adjusted to 6.4 with dilute HCl/NaOH. HT was dissolved at 10 mM concentration in the same solvent as the RNA, and the pH was adjusted to 6.4. All experiments were performed on a 900-MHz Bruker spectrometer with a cryogenically cooled probe. Measurements were done at 293 K which was found to be the optimal temperature for both water and ^2H_2O measurements.

Spectrum	Solvent	Time domain size (F2/F1)	Acquisition times F2/ F1 (ms)	Sweepwidth F2/F1 (kHz)	No. of scans	Total experiment time (h)
NOESY	H_2O	4096/1024	151.6/37.9	13.51/13.50	64	12
NOESY	^2H_2O	2048/300	113.9/27.8	8.99/5.41	96	11.5
TOCSY	^2H_2O	2048/256	113.9/21.5	8.99/5.95	64	5.5

Table 3.01 – Experimental acquisition parameters for the experiments on TSMC and TSGC, including time domain sizes, acquisition times, sweepwidth, number of scans and overall experiment time.

RNA chemical shift changes upon HT addition were monitored using homonuclear 2D TOCSY (Piotto, Saudek et al. 1992, Sklenar, Piotto et al. 1993) ($\tau_m = 80$ ms) and 2D NOESY (Liu, Mao et al. 1998)($\tau_m = 300$ and 80 ms) experiments at HT/TSMC ratios of 0:1, 0.5:1, 1:1 and 1.5:1. Additional experimental parameters are outlined in Table 3.01. For the former experiment the sample was dissolved in 100% ^2H_2O, while for the latter, measurements in both 90% H_2O + 10% ^2H_2O and 100% ^2H_2O were conducted. Significant precipitation at the highest ratio precluded the use of further titration points. The chemical shift assignments for the free TSMC reported by Tavares et al. (Tavares, Beribisky et al. 2009) were used.

3.2.2 Molecular Modelling

3.2.2.1 Preparation and parameterization of the starting structures

Model 6 from the solution structure ensemble of TSMC reported by Tavares *et al.* (PDB ID 2RPT) (Tavares, Beribisky et al. 2009) and HT coordinates from solvation dynamics studies kindly supplied by Dr S.A. Corcelli (Furse, Lindquist et al. 2008) were used for modeling the complex of TSMC with HT. HT was manually docked (visualization was done in Pymol 1.2r2) in

the major groove at the crevice formed by the CC mismatch. Since the orientation of HT with respect to RNA was not known, two docking poses rotated 180° with respect to each other were generated. The Parmbsco force field (Perez, Marchan et al. 2007) was used to parameterize the RNA, and the published parameters (Furse, Lindquist et al. 2008) were used for HT.

3.2.2.2 Molecular dynamics simulations

All molecular dynamics (MDs) simulations were performed under periodic boundary conditions using the SANDER module of AMBER10 (Perez, Marchan et al. 2007). The pressure was maintained at 1 atmosphere by isotropic position scaling. A two-step procedure was used to simulate the complex. In the first step, the fully flexible HT was allowed to explore the surface of a restrained TSMC structure and find an optimal docking pose. This was accomplished by harmonically restraining the atoms of TSMC with a force constant of 25 kcal/mol/Å^2, while allowing HT and the solvent atoms to move freely during a production run of 10 ns. The final frame from the A1 step was taken as the initial frame for the second step (referred to as A2) in which induced fit of the HT-TSMC interaction at the CC mismatch was explored. This was accomplished by allowing unrestrained movement of HT, solvent and residues 4–6 and 16–18 of TSMC, i.e. of the CC mismatch and one flanking base pair above and below the mismatch, while the rest of the RNA was harmonically restrained with a force constant of 10 kcal/mol/Å^2, during a production run of 24.6 ns. The conformations sampled during the first 5 ns and the last 500 ps of the A2 run were independently averaged and then minimized using steepest descent and conjugate gradient methods to obtain representative models for HT-TSMC groove binding (Figure 3.06A) and HT-TSMC intercalation (Figure 3.06B) complexes, respectively.

To estimate the binding free energy for the RNA-ligand interaction, continuum solvent calculations were performed using the internal Poisson-Boltzmann Surface Area (PBSA) solver in Sander in the Amber9 package. To compute binding free energies for the groove bound mode, the intercalation mode and the intermediate transition phase, the trajectory from the A2 run was split into segments of 5 ns each (the last segment being 4.6 ns) which were then used independently for the calculations. The structures of unbound TSMC and HT were taken from the simulated complexes, i.e. changes in the conformation of each of the components were neglected for the binding free energy calculations. For reference, the unbound forms of TSMC and of HT were simulated for 9 and 2 ns, respectively, using the same settings as for the A2 run.

3.3 Results

3.3.1 CC mismatch in RNA is the binding site for HT

The prediction of a stem-loop structure for Site I, with a CC mismatch in the stem and the start codon in the loop, was first reported by Chu *et al.* (Chu, Voeller et al. 1993). In agreement with the prediction, a Site I-like RNA construct was reported to have a high melting temperature of 65°C independent of the concentration, suggesting an intra-molecular stem-loop fold (Hasnat, Bichenkova et al. 2007). Another Site I-like RNA construct was also observed to adopt a stem-loop fold by solution fluorescence studies, although an asymmetric dimer of two hybridized RNA strands was obtained in the crystal (Dibrov, McLean et al. 2011). These reports suggest that the Site I indeed adopts a stem-loop structure in solution. Tavares *et al.* (Tavares, Beribisky et al. 2009) have recently reported the NMR structure of a shorter construct, TSMC, and mapped the binding site for the aminoglycoside paromomycin. In TSMC, the three flanking base pairs on either side of the CC mismatch in the Site I stem are preserved and the resulting stem sequence is capped with a stable GNRA (G = guanine, N = any nucleotide, R = purine, A = adenine) tetraloop (Figure 3.01B, leftmost construct).

TSMC was used to study the structural features of the RNA-HT complex by NMR. A 2D ^{1}H, ^{1}H TOCSY experiment, recorded in ^{2}H$_2$O, was used to monitor H6/H5 cross-peaks for the pyrimidine residues (exclusively cytosine residues in the case of TSMC). TSMC H6/H5 resonances were obtained from the assignments performed by Tavares et al (Tavares, Beribisky et al. 2009). These assignments were verified by assigning the aromatic-anomeric portion of the 2D ^{1}H, ^{1}H NOESY of TSMC recorded in 100% ^{2}H$_2$O and transferring those assignments to the TOCSY. An overlay of TOCSY spectra recorded in the presence of increasing concentrations of HT is shown in Figure 3.02. The H6/H5 correlations for C19 and C20 experienced negligible changes in chemical shifts compared with the remaining cytosine residues, suggesting that the binding event occurred at a position distant from the base of the stem. Signal broadening was observed for the H6/H5 correlations of C15, and to a lesser extent, for the joint signal of C16 and C7. Presumably, these residues experience exchange-broadening motions when TSMC is bound to HT. These chemical shift perturbations (CSPs) suggest that the major interaction site between HT and TSMC is the RNA's CC mismatch.
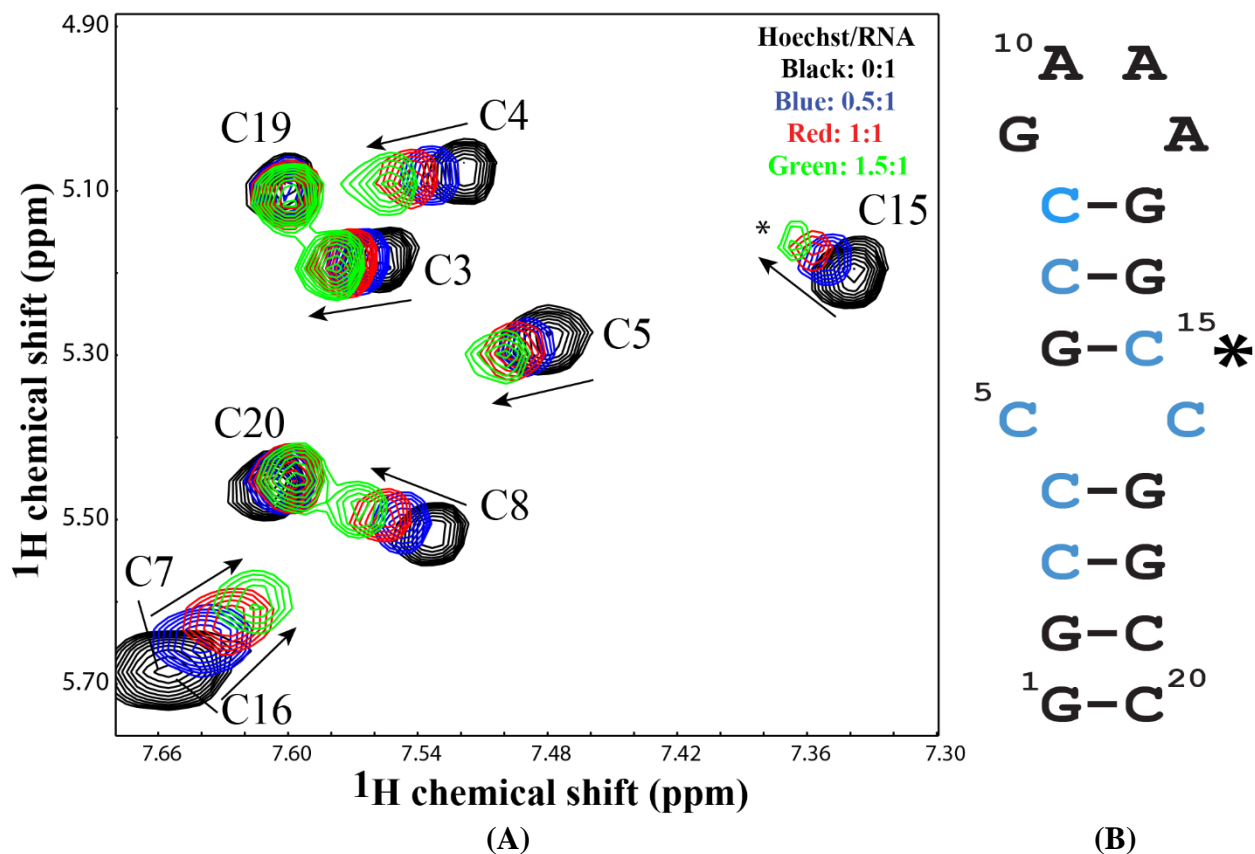
Figure 3.02 – Superposition of 2D ^{1}H, ^{1}H TOCSY spectra of TSMC upon addition of HT recorded in ^{2}H$_2$O. HT/RNA ratios of 0:1, 0.5:1, 1:1, 1.5:1 are marked in black, blue, red and green, respectively. CSPs induced by HT binding are shown by arrows, the corresponding residues are colored in cyan in the RNA sequence. C15, which experiences a strong CSP and line-broadening, is indicated with an asterisk. Adapted from Garg et al., 2013 (Garg, Beribisky et al. 2013).

To further investigate the role of the CC mismatch in HT binding, an additional RNA construct, termed TSGC (Figure 3.01B, middle construct) was also used. In this construct, C5 was replaced with G5, thereby eliminating the mismatch and replacing it with a GC base pair. However, assigning this construct proved to be very challenging due to the very high degree of overlap in the NOESY (Figure 3.03D). A few resonances in this experiment, as well in a TOCSY experiment, were assigned tentatively based on the comparison of the characteristic chemical shifts with the TSMC spectra (Figure 3.03B). For both TSMC and TSGC, chemical shift changes upon HT addition were observed for adenine resonances in 2D NOESY spectra (Figure 3.03C and 3.03D). In addition, the H6/H5 signals of residue C8 in the TOCSY for TSMC (Figure 3.02 and Figure 3.03A) and based on tentative assignments in TSGC (Figure 3.03B) experience significant

chemical shift changes. These changes show that HT also interacts with the GNRA tetraloop in both TSMC and TSGC.
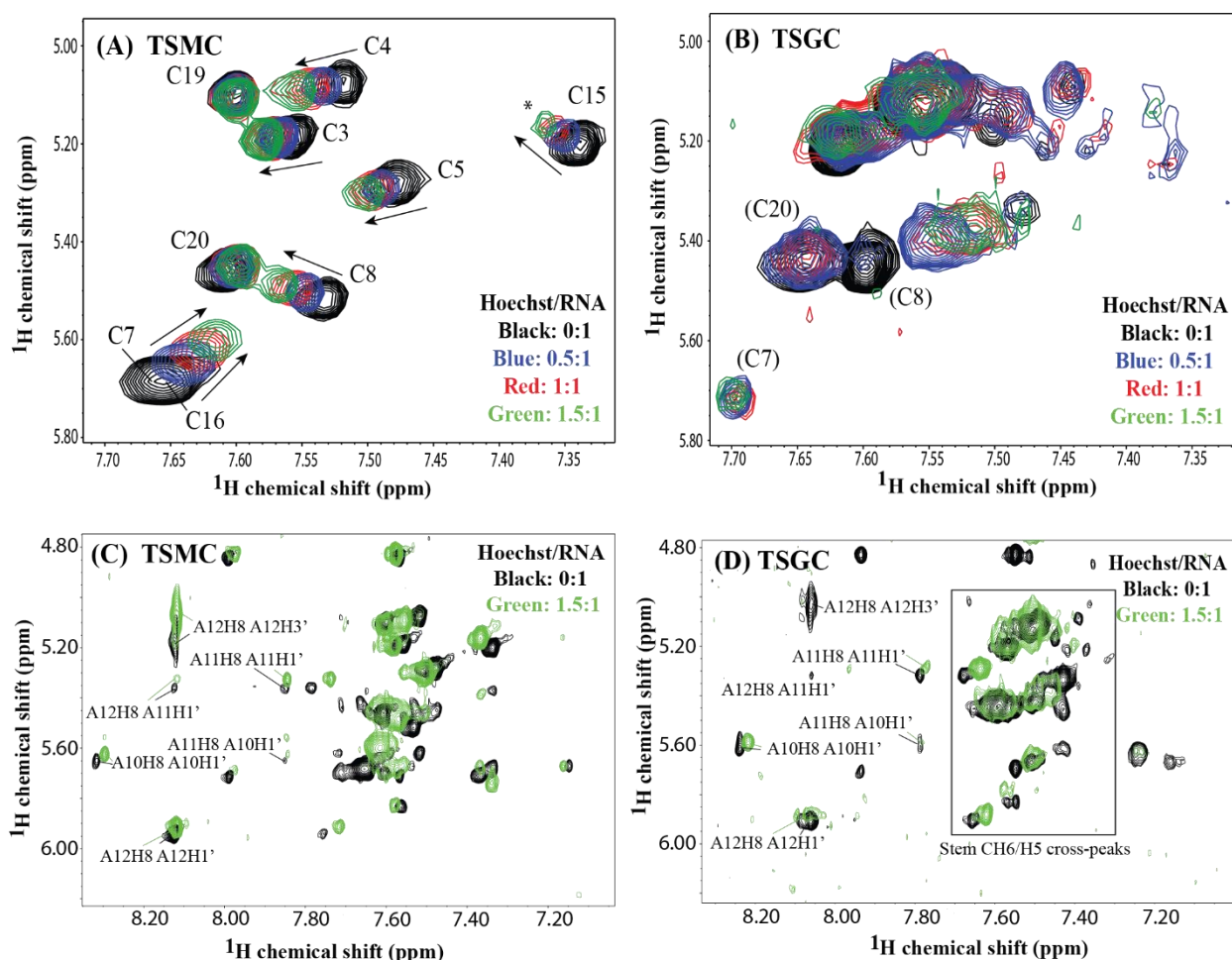
Figure 3.03 – Superposition of 2D ^{1}H TOCSY spectra of TSMC **(A)**; TSGC **(B)**; and 2D ^{1}H NOESY spectra of TSMC **(C)** and TSGC **(D)** recorded in ^{2}H$_2$O, indicating chemical shift changes in the aromatic-anomeric region upon addition of HT. In the TOCSY spectra, HT/RNA ratios of 0:1, 0.5:1, 1:1 and1.5:1 are marked in black, blue, red and green respectively, while in the NOESY spectra, HT/RNA ratios of 0:1 and 1.5:1, are marked in black and green respectively. In **(B)**, the H5/H6 resonances for C7, C8 and C20 are assigned tentatively (indicated by brackets) based on their position in the TOCSY spectrum for TSMC. In the NOESY spectra, the assignments of inter-residue cross-peaks as well as the intra-residue H8/H1' adenine cross-peaks for the GNRA tetraloop are indicated. In **(D)**, the H6/H5 signals for the stem cytosine residues are boxed. (Garg, Beribisky et al. 2013).

In addition, the TSMC-HT interaction was studied using a NOESY experiment recorded in H$_2$O (Figure 3.04). Since every canonical Watson–Crick base pair (except the one at the bottom of the stem) gives rise to a set of cross-peaks in the imino-imino and in the imino-amino regions, chemical shift changes for these protons can be used to determine whether HT binding affects

TSMC base pairing. Chemical shift changes for the TSMC RNA were monitored at HT/RNA ratios of 0:1, 0.5:1, 1:1, 1.5:1 and revealed most significant changes for the G17-C4 base pair. G18-C3 G6-C15 and G14-C7 also seem to be affected; however, a high degree of overlap precludes further evaluation. These changes are consistent with our hypothesis that HT binds TSMC at two sites: the CC mismatch and the GNRA tetraloop. The imino-amino signals from all base pairs are conserved at HT/RNA ratios, indicating that the HT addition does not affect TSMC base pair integrity.
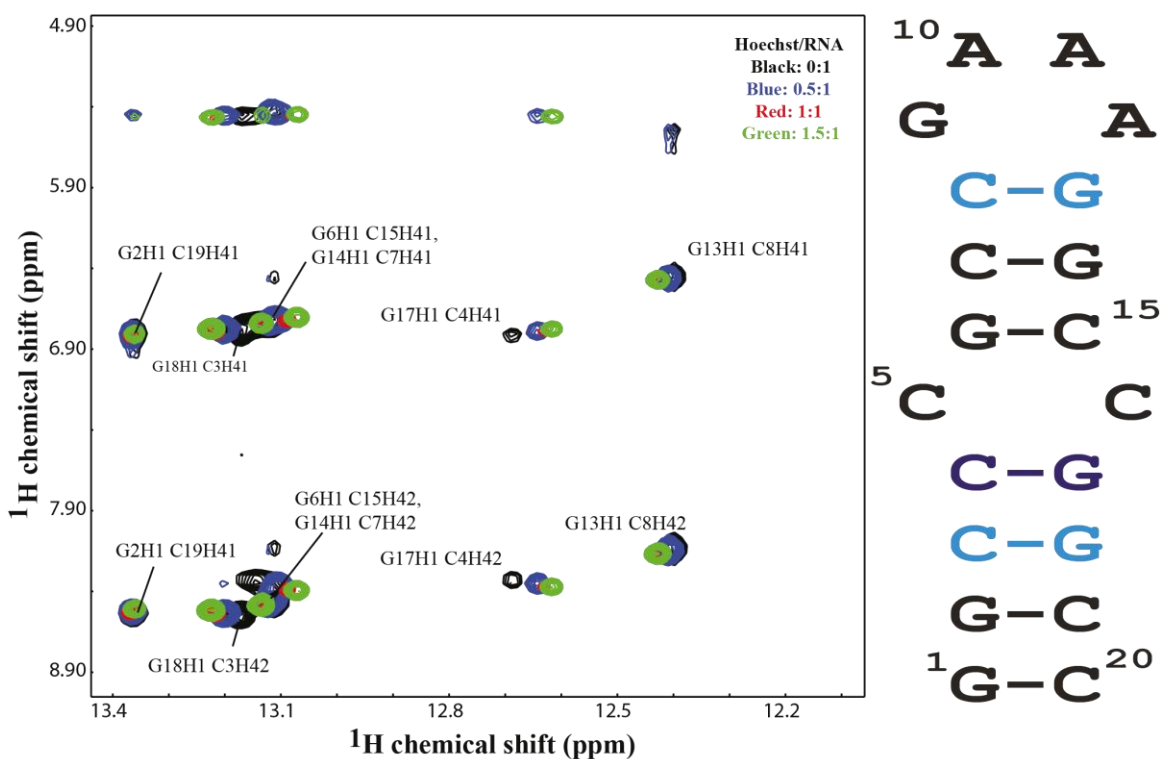
Figure 3.04 – Superposition of 2D water NOESY of TSMC. Chemical shift changes in the imino-amino region, upon addition of HT with HT/RNA ratios of 0:1, 0.5:1, 1:1, 1.5:1 are marked in black, blue, red and green respectively. The base-pairs with a large degree of chemical shift change are colored in cyan on the TSMC RNA sequence on the right. (Garg, Beribisky et al. 2013).

3.3.2 Modelling of HT/RNA interaction

UV-Vis spectroscopy experiments conducted by collaborators have shown that HT binds TSMC by intercalation. However, a lack of intermolecular NOEs between TSMC and HT, precluded the determination of the HT/TSMC complex. A similar problem was observed for the TSMC/paromomycin complex (Tavares et al., 2009). Since there are no available experimentally

determined structures showing such a binding mode, we performed molecular docking and MDs simulations to investigate the mechanism by which HT might intercalate. Docking of HT to the CC mismatch region of TSMC was performed as described in section 2.2.2. Two poses, A and B (Figure 3.05), were generated by docking in which HT lies in the major groove in opposite directions (related by 180° rotation). In the A pose (Figure 3.05A), ring R4 was pointed toward C16 of TSMC, and R1 was oriented toward C5. In the B pose (Figure 3.05B), ring R1 was oriented toward C16, while R4 was toward C5. These two docking poses were used as starting structures for independent MDs simulations.
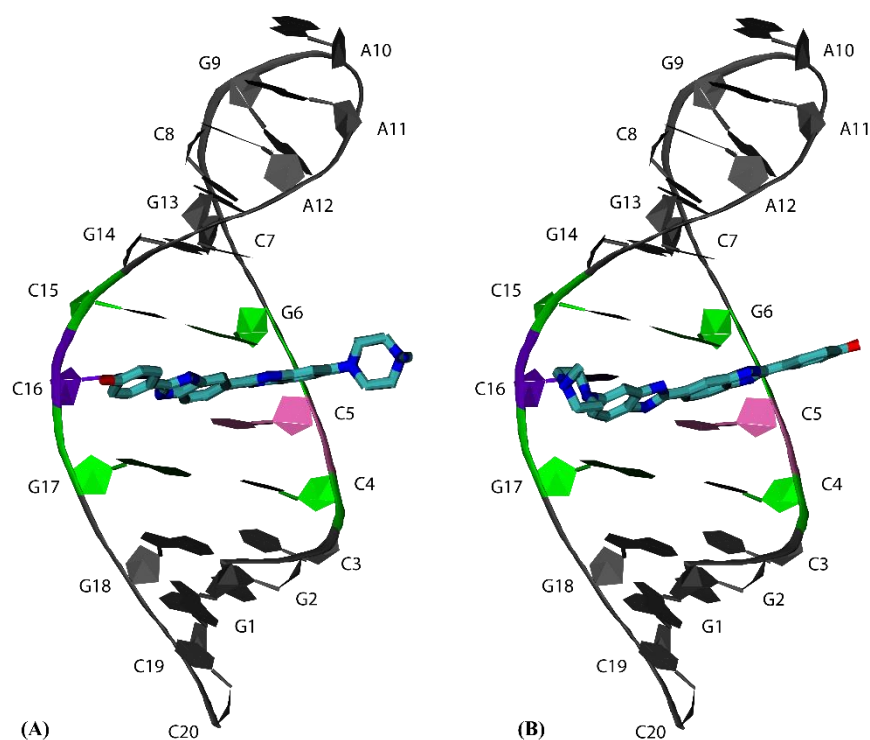
Figure 3.05 – The two docking poses, **(A)** and **(B)**, generated for the HT-TSMC complex and used as starting structures for molecular dynamics simulations. TSMC is represented as a cartoon, where the base-pairs flanking the CC mismatch are colored in green, C5 is colored in pink, C16 in purple. HT is shown in stick representation colored by atom type (C - cyan, N - blue, O - red) (Garg, Beribisky et al. 2013).

During the first step of the two step simulation protocol employed, the RNA was restrained while HT was free to move. In the simulation starting with pose A, HT remained in the major groove in contact with the RNA throughout the simulation and the interactions between TSMC and HT were optimized. In the simulation starting with HT docked in pose B, HT was released into solvent during the first nanosecond of the simulation. Therefore, pose B was discarded, and

the final structure from the simulation with pose A, A1, was used as the starting structure in simulation A2, in which the CC mismatch, one flanking base pair on either side of the mismatch and HT were free while the rest of the RNA was restrained. To our surprise, during simulation A2, HT indeed penetrated through the helix, thus adopting an intercalative binding mode. An interesting chain of events led to this induced fit intercalative binding mode (Figure 3.06B). HT slid over the furrow in the surface formed by the CC mismatch such that HT-H6 of the ring R1 could make an H-bond with the backbone C5-O2P. While the ring R1 was so anchored, a rotation, driven by the possibility of H-bonding between HT-H3 and TSMC-G17-O6, occurred at the γ bond (Figure 3.01B, leftmost construct and Figure 3.01C) between 7 and 9 ns. This was followed by the formation of an H-bond between HT-H1 and the backbone TSMC-G17-O2P at 9.2 ns. Simultaneously, the TSMC-C16 base began to flip out to stack against the ring R2 of HT. Stacking against C16 and H-bonding between HT-N3 and TSMC-G17-N7 were the key interactions that permitted HT to reorient and penetrate through the helix, ring R1 first, at 11.2 ns.
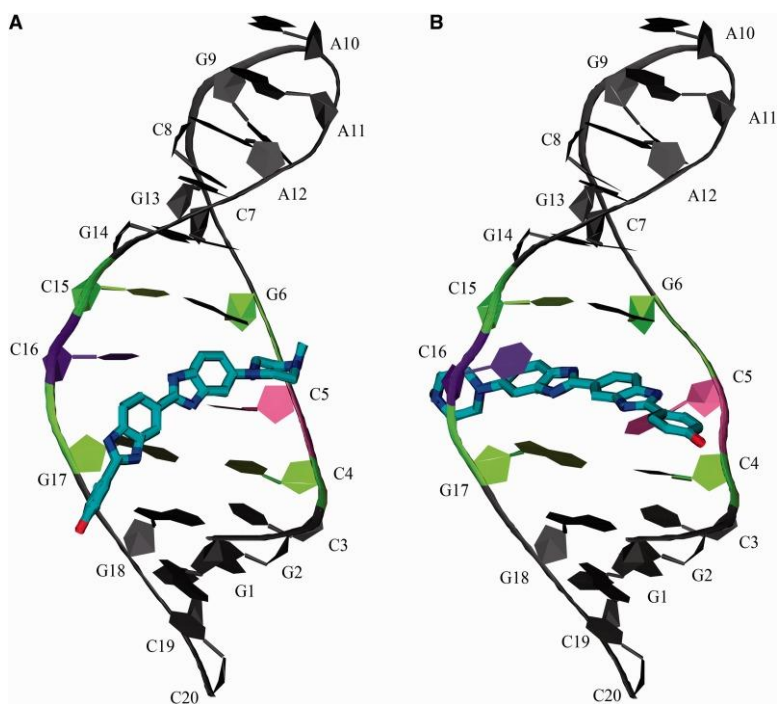
Figure 3.06 – Structural models of **(A)** the groove binding HT-TSMC complex generated by averaging and minimizing the first 5 ns of the simulation A2. TSMC is depicted as a cartoon with the major groove facing the reader, whereas HT is shown in stick representation (C: cyan, N: blue, O: red); and **(B)** the HT-TSMC intercalation complex generated by averaging and minimizing the last 500 ps of the simulation A2 (Garg, Beribisky et al. 2013).

It is imperative to mention that the simulation A2 was run for only 24.6 ns, implying insufficient sampling to establish the binding mode definitively. At the same time, it is difficult to sample completely enough to establish the binding mode. Therefore, the induced-fit mechanism of intercalation discussed here may not be the main method of interaction, rather it represents one possible interaction mode.

3.3.3 The methyl group is not involved in TSMC binding

In the structural model presented here (Figure 3.06B), the methylpiperazine moiety (ring R1) of HT protrudes into the minor groove, whereas the phenolic moiety (ring R4) points toward the major groove of the RNA. It would be logical to expect that a model with the reverse orientation, i.e. ring R4 protruding into the minor groove while the ring R1 extended toward the major groove would be an equally feasible option. In such a structure, HT-H6 could form a hydrogen bond with the backbone phosphate groups of C3 or C4 and ring R4 could shield the C4 of TSMC from solvent. Although such a model could be energetically favorable, our NMR data argue against this mode of binding. Depending on the rate of exchange between the free and bound forms, the ^{1}H NMR signals of HT would shift or broaden out. This effect would be more pronounced for the protons making direct contacts with the RNA. In the NMR spectrum, the protons from aromatic rings R2, R3 and R4 of HT resonate between 6.5 and 7.5 ppm, whereas the aliphatic protons for the methyl group attached to R1 appear at 2.8 ppm. During titrations of HT into RNA, the signals for the methyl protons in R1 built up with increasing concentrations of HT whereas the signals of the aromatic protons remain broadened beyond detection (Figure 3.07A). Similarly, in the TOCSY spectra recorded for free HT in ^{2}H$_2$O and subsequent titrations of TSMC into HT, the signals for the aromatic rings of HT were bleached upon the first instance of RNA addition, with an HT:RNA ratio of 1:0.25, whereas the signals from R1 were neither shifted nor significantly broadened (Figure 3.07B). These observations suggest that ring R1 of HT is relatively free and not involved in interactions with TSMC. Therefore, the orientation of the intercalated HT should be as proposed by our model and not the one flipped by 180°.
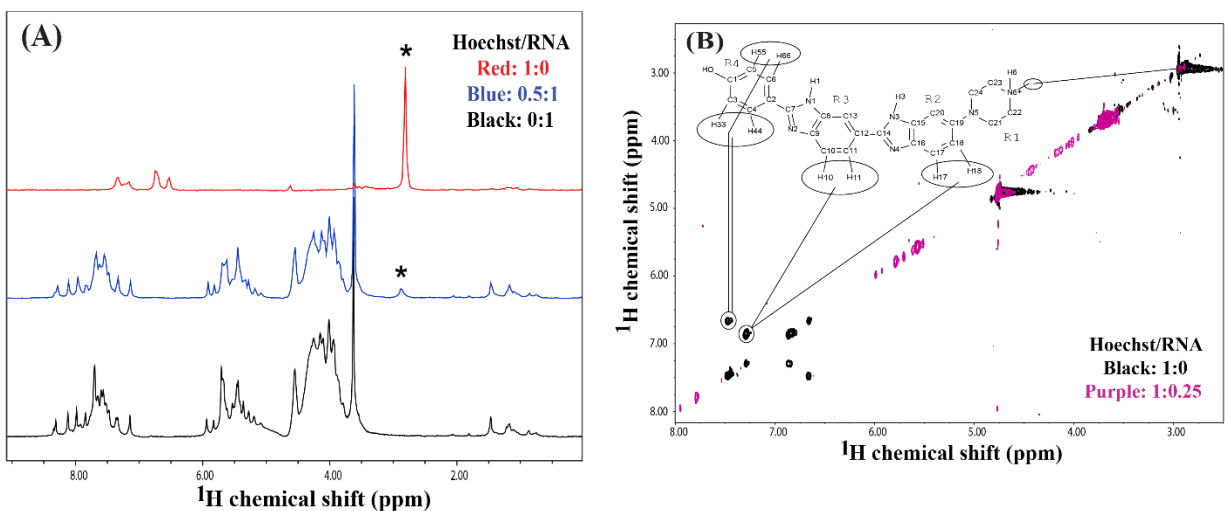
Figure 3.07 – (A) 1D ^{1}H spectra of free TSMC (black), of a sample with a HT/TSMC molar ratio of 0:50:1 (blue) and free HT (red). In the 1D spectra of free HT, and of HT/RNA ratio 0.50:1, the signal for the HT methyl group of methylpiperazine ring (R1) is marked with *. **(B)** Superposition of 2D ^{1}H TOCSY spectra recorded in ^{2}H$_2$O for samples with HT:TSMC RNA molar ratios of 1:0 (black) and 1:0.25 (purple). The assignments for the proton pairs of R4 as well R3 and R2 are shown. The protons from the methyl group in R1 at 2.8 ppm is the only signal observed for HT after addition of RNA, whereas the signals for aromatic protons between 6.5-7.5 ppm were completely bleached (Garg, Beribisky et al. 2013).

3.4 Discussion

3.4.1 HT binding mode to TSMC

HT is a well-known DNA minor groove binder (Vega, Garcia Saez et al. 1994). The binding occurs mostly to AT-rich sequences with the affinity of 60 nM. The HT-DNA complex is suggested to be stabilized by hydrogen bonding and van der Waals interactions with the deep minor groove of B-form DNA helices. HT binding to TSMC via its minor groove is unlikely due to this groove's small size in RNA. However, HT has also shown to bind GC rich DNA via intercalation in the micromolar range (Bailly, Colson et al. 1993, Colson, Houssier et al. 1995, Adhikary, Buschmann et al. 2003). It is thought that Hoechst 33258 partially intercalates one of its benzimidazole rings and the phenol group between two GC base pairs thus placing the second benzimidazole into the major groove with the positively charged N-methyl-piperazine terminal group protruding outside the DNA helix (Bailly, Colson et al. 1993). UV-vis experiments conducted by collaborators have shown that titration of HT with the three RNAs (TSMC, TSGC and TS1), yielded complexes whose spectroscopic and photophysical properties were similar to each other and to those observed upon HT complexation with GC-DNA as opposed to with AT-DNA. This finding strongly suggests that HT binds TS by intercalation.

3.4.2 Other relevant small compound/RNA interactions

3.4.2.1 RNA binding mechanism of paromomycin and other aminoglycosides

The CC mismatch in RNA has been a known binding target for two classes of compounds: Aminoglycosides and aromatic compounds. Various aminoglycosides such as kanamycin (Figure 3.09B) and Neomycin B (Figure 3.09C) have been shown to bind the CC mismatch with affinities from 0.9 μM to 2.7 μM (Tok, Cho et al. 1999, Tok, Cho et al. 2000). Paromomycin (Figure 3.08A) was reported to bind with a slightly higher affinity of 0.5 μM (Tavares, Beribisky et al. 2009).

Aminoglycosides in general, are not very prone to specificity in RNA binding. These compounds are mostly cationic, and bind non-canonical RNA regions via electrostatic interactions and hydrogen bonds (Michael et al., 1988, Tok et al., 1999, Tavares et al., 2009). Paromomycin in particular is thought to interact with the CC mismatch via its rings I (Figure 3.09A) and II (data not shown) which are rich in hydroxyl and amino functionalities which purportedly facilitate this interaction (Tavares, Beribisky et al. 2009). Similar to HT, paromomycin recognizes TSMC at the region where the RNA's helical geometry is distorted and the major groove is widened. In another

similarity to HT/TSMC binding process, a second binding event takes place at higher paromomycin/TSMC ratios (Tavares, Beribisky et al. 2009)
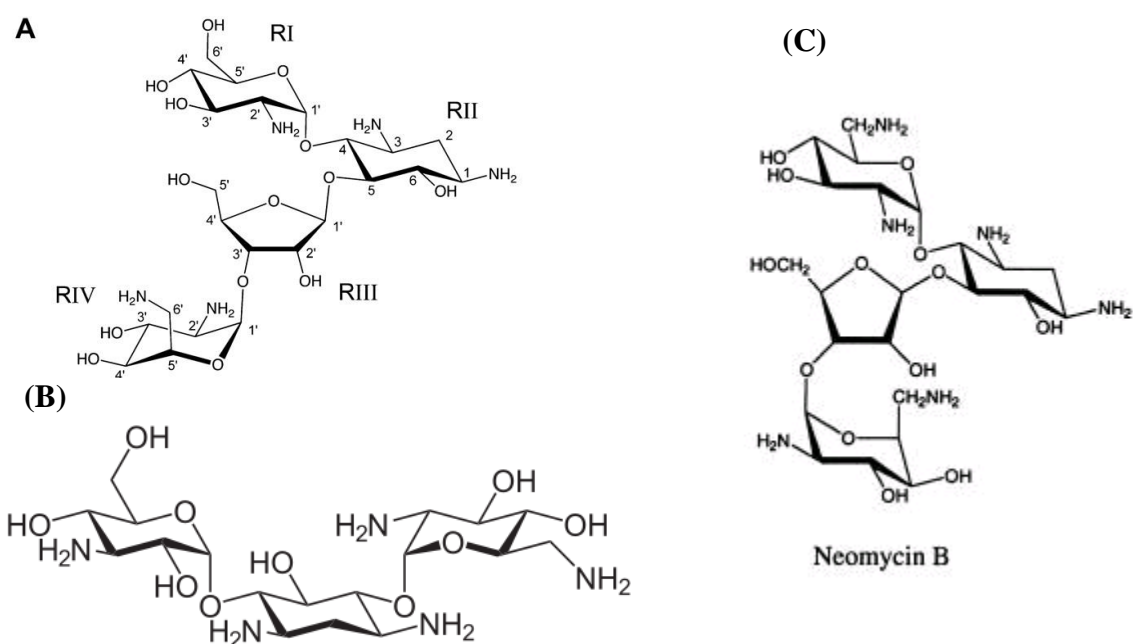
Figure 3.08 – Aminoglycoside compounds shown to bind the CC mismatch RNA **(A)** Paromomycin, **(B)** Kanamycin B, **(C)** Neomycin B. Adapted from Tok et al. and Tavares et al., (Tok, Cho et al. 1999, Beribisky, Tavares et al. 2007).

(A)   **(B)**
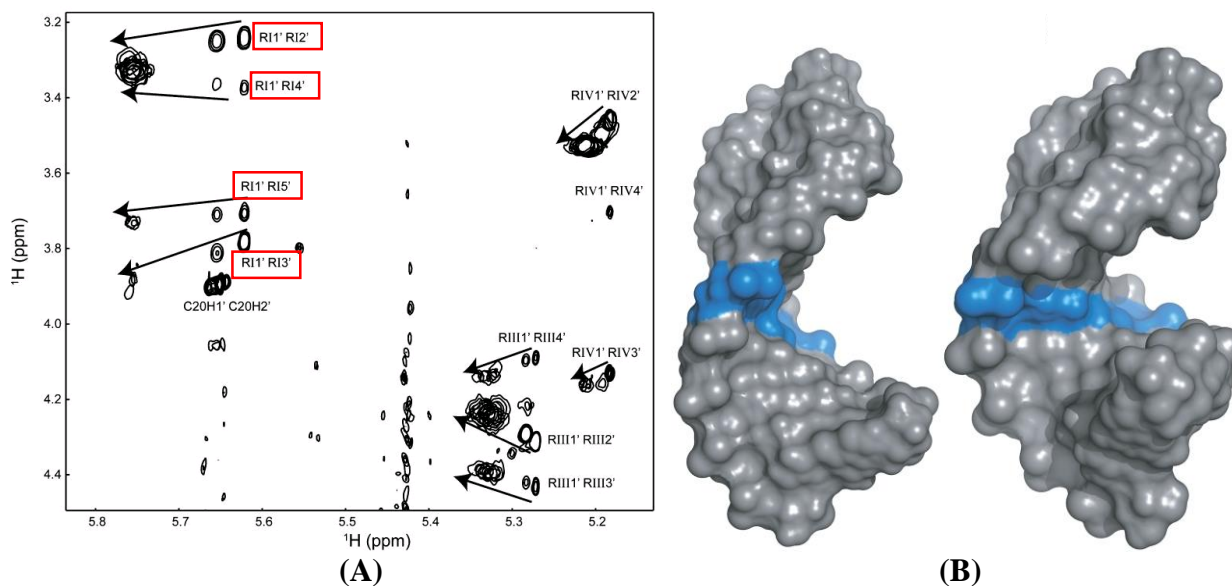
Figure 3.09 – **(A)** A 2D ^{1}H, ^{1}H TOCSY showing changes in paromomycin signals upon addition of TSMC RNA. Largest chemical shift perturbations are observed for resonances for ring I. **(B)** The widening of the major groove in TSMC RNA (on the left) compared to TSGC RNA (on the right) allows for paromomycin binding (Tavares, Beribisky et al. 2009).

Another well-characterized RNA target for small molecule binding, is the non-canonical bulge of the ribosomal A-site (Figure 3.10A). Paromomycin (Figure 3.10B), Kanamycin (Figure 3.10C) as well as Neomycin B (Figure 3.10D) were all shown to interact with this RNA in a manner similar to that of paromomycin and TSMC (Fourmy, Yoshizawa et al. 1998, Vicens and Westhof 2001, Francois, Russell et al. 2005). Rings I and II of these compounds were all shown to be involved. Ring I inserts itself into the RNA helix (against G1491 – bottom right portion ofFigure 3.09B), and forms a putative pseudo base-pair with A1408 (Figure 3.10B). These contacts allow A1492 and A1493 to remain bulged, stabilizing this RNA/small-molecule complex (Francois, Russell et al. 2005). Numerous contacts are mediated by water molecules (Figures 3.10B 3.10C and 3.10D). In cases of neomycin and kanamycin, non-specific binding is also observed, resulting in non-one-to-one stoichometries (Francois, Russell et al. 2005).

(A)   **(B)**

(C)

(D)
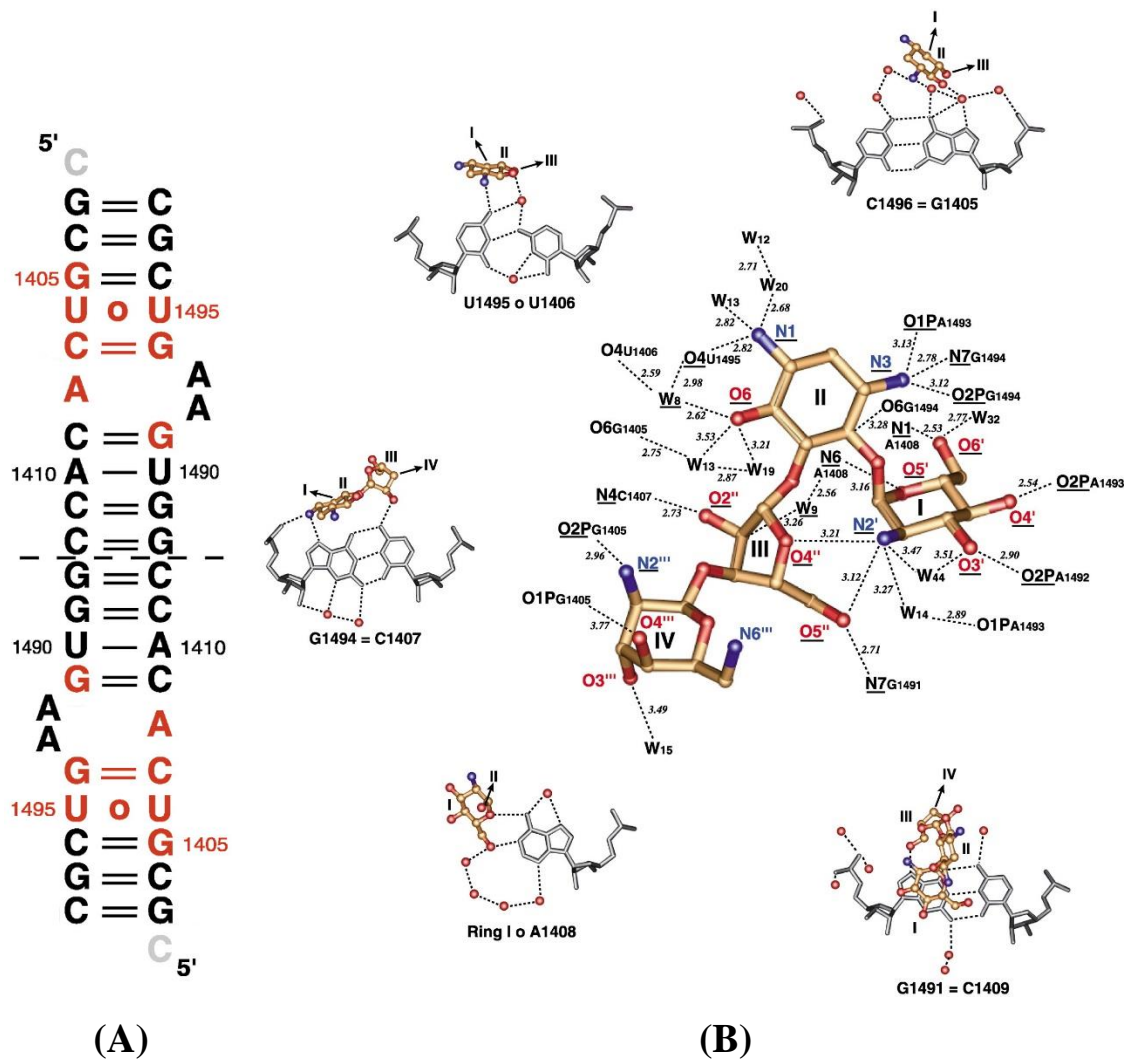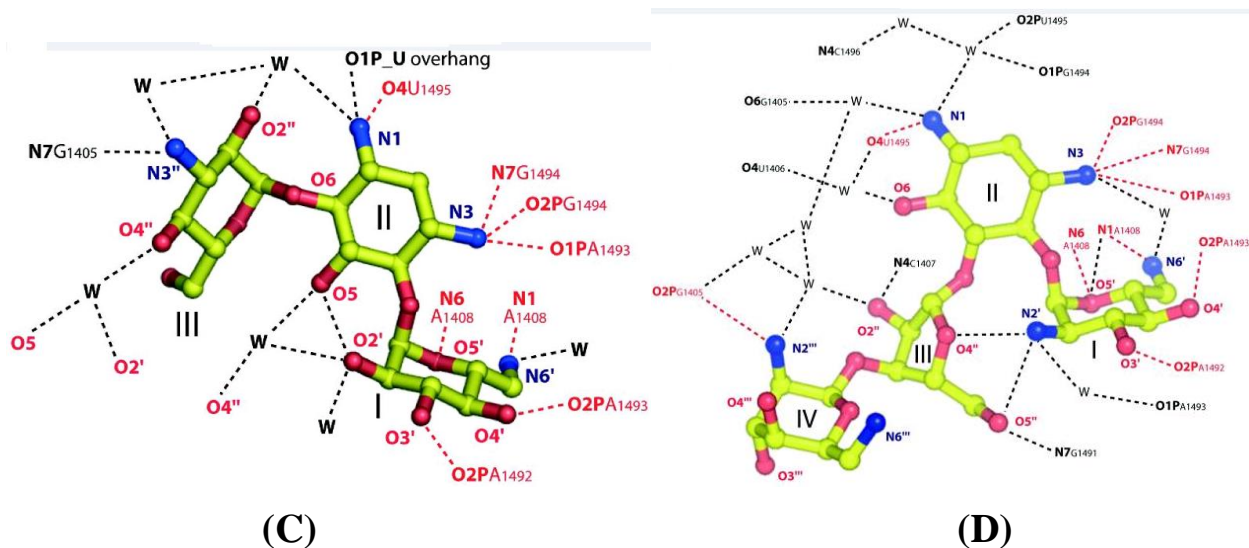
Figure 3.10 – **(A)** Secondary structure representation of A-site RNA. A non-canonical region consisting of three non-base-paired adenine residues, and an adjacent uracil-uracil mismatch is the aminoglycoside binding site. **(B)** In the center, contacts to atoms in A-site RNA (marked in black) made by a molecule of paromomycin (participating atoms are marked in red and blue). Ring numbers are denoted, atomic distances are italicized, and mediating water molecules are marked as "W". The paromomycin is surrounded by smaller figures showing interactions that paromomycin makes to specific residues in the RNA. **(C)** Contacts to atoms in A-site RNA (marked in red and black) made by a molecule of kanamycin (participating atoms are marked in red and blue). Ring numbers are denoted, mediating water molecules are marked as "W". **(D)** Contacts to atoms in A-site RNA (marked in red) made by a molecule of neomycin (atoms are marked in red and blue). Ring numbers are denoted, mediating water molecules are marked as "W" (Vicens and Westhof 2001, Francois, Russell et al. 2005)

Aminoglycoside is also known to interact with the Rev response element (RRE), transcriptional activator RNA sequence in the HIV genome and a binding site for the viral protein Rev (Zapp, Stern et al. 1993). The interaction is known to disrupt Rev function and inhibit viral proliferation (Zapp, Stern et al. 1993). As in the previous examples, aminoglycoside binding takes place in the non-canonical region of the RNA defined by two GG and GA mismatches as well as an extrude uracil residue. Abolition of these motifs and their replacement with canonical elements (Figure 3.11), progressively reduced the affinity of aminoglycosides to this RNA (Cho and Rando 1999).
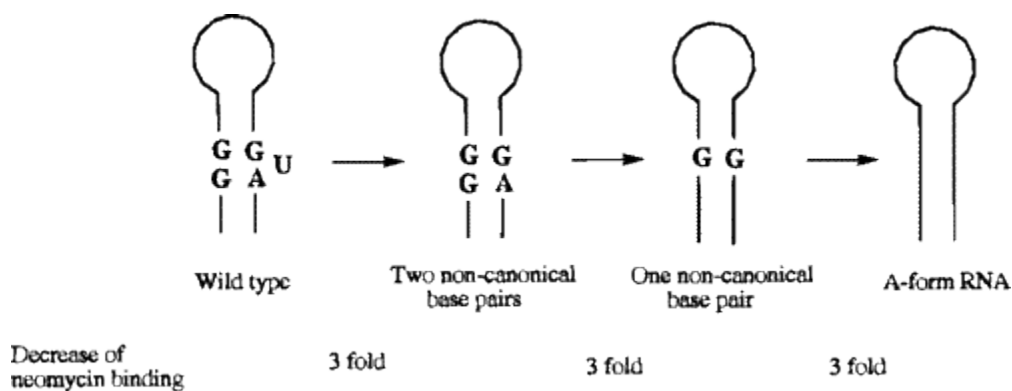
Figure 3.11 – The non-canonical elements in RRE RNA: The GG and GA mismatches and an extruding uracil. As these non-canonical elements are replaced with GC base-pairs, the affinity of aminoglycosides to RRE is reduced (Cho and Rando 1999).

3.4.2.2 Binding mode of aromatic compounds to RNA

Among aromatic compounds (Figure 3.12), the DNA minor groove binder 4',6-diamidino-2-phenylindole (DAPI – Figure 3.12A), the antiprotozoal drug mepacrine (also known as quinacrine – Figure 3.12B), and the bacterial disinfectant proflavine (Figure 3.12C), all bind the CC mismatch with an affinity ranging from 270 nM (DAPI) to 1.17 µM (Proflavine) (Cho and Rando 2000). DAPI's highest affinity can be explained by its structure being the most similar to HT, and therefore being able to extend along the entire length of the CC bulge. Like HT, DAPI has been shown to be able to bind B-DNA in an intercalative manner (Colson, Houssier et al. 1995).
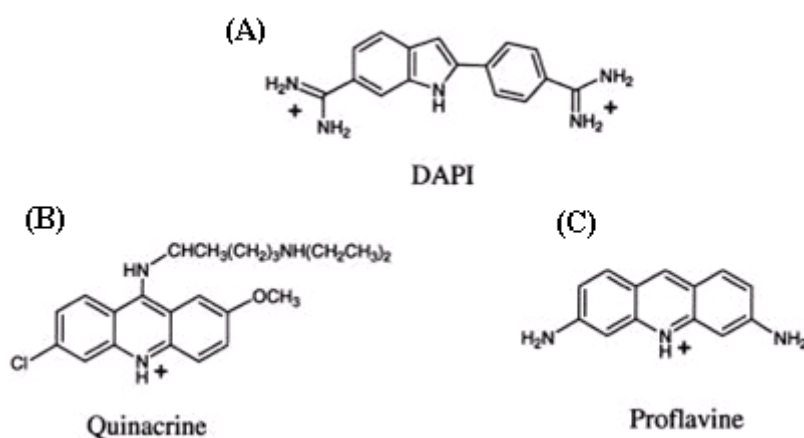
Figure 3.12 – A list of aromatic compounds used in fluorescence anisotropy measurements in binding TS mRNA. **(A)** DAPI, **(B)** Quinacrine, **(C)** Proflavine. Adapted from Cho et al (Cho and Rando 2000).

Similar to HT, proflavine has also been demonstrated to intercalate into a 2',5'-linked GC-rich RNA (Horowitz, Lilavivat et al. 2009). It does so by binding via this RNA's widened major groove between the C3-G6 and G4-C5 base-pairs (that is at two locations in the RNA duplex – Figure 3.13A). A break in the aromatic-anomeric "walk" is observed for these residues, however aromatic proflavine resonances (H2 and H7 – marked with a red box in Figure 3.13B) which show up in a chemical shift range similar to that of H8 and H6 atoms can be used in lieu of them to complete it (Horowitz, Lilavivat et al. 2009).

The interaction between proflavine and the GC-rich RNA duplex (Horowitz and Hud 2006) is weaker than the interaction between HT and TSMC (2.5 μM versus 80 nM). Nevertheless a handful of intermolecular NOEs for the first complex are observed. This is due to this interaction occurring in the slow exchange regime. The presence of this exchange mode is supported by the disappearance and appearance of ^{1}H imino resonances for the free and bound forms of the GC-rich RNA with increasing proflavine/RNA ratios (Horowitz, Lilavivat et al. 2009).

(A)   **(B)**
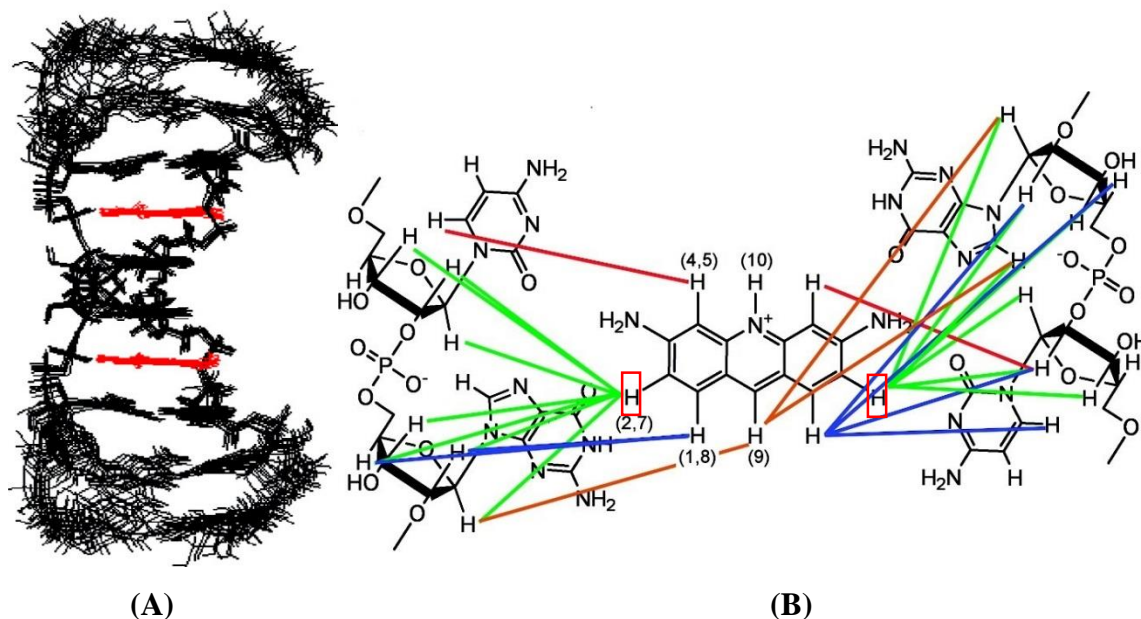
Figure 3.13 – **(A)** Intercalation of proflavine into a 2'5' GC RNA homodimer. **(B)** Contacts to RNA made by proflavine. Proflavine resonances making contacts similar to those made by H8 and H6 are boxed (Horowitz, Lilavivat et al. 2009).

In GC-RNA, C3 and C5 adopt a C2' *endo*, while their neighbouring G4 and G6 residues a C3' *endo* conformations. Alternating sugar puckers also seem to be essential for the proflavine/GC RNA interaction to take place (Horowitz, Lilavivat et al. 2009).

(A)   **(B)**
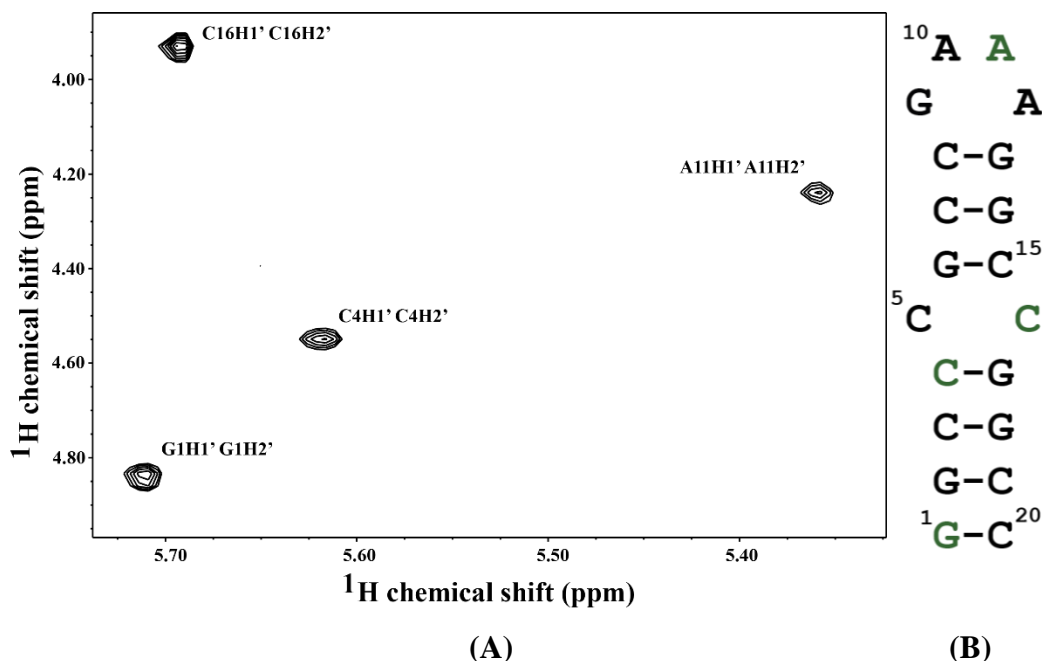
Figure 3.14 – **(A)** 2D^{1}H, ^{1}H TOCSY of TSMC, the anomeric-ribose region. H1'/H2' correlations for A11, C4, G1, C16 indicate that these residues adopt a *C2' endo* conformation. (B) TSMC RNA secondary structure where the residues adopting a *C2' endo* conformation are marked in green while the residues with a *C3' endo* conformation, are marked in black.

In TSMC, the residues C4 and C16 located in the vicinity of the CC mismatch have a *C2' endo* or a mixture between *C2'* and *C3' endo* conformations, while the surrounding residues' sugar pucker are exclusively *C3' endo* (Figure 3.14). Similar as in GC-RNA, the alternating and/or flexible sugar puckering in TSMC may help accommodate HT binding.

3.4.3 Effect of TSMC/HT interaction on TS levels

To explore the effect that HT binding has on TS protein and mRNA levels, cell-based assays were conducted by collaborators to monitor the TS protein and mRNA levels at different administered doses of HT in an ovarian cancer cell line. The TS mRNA levels were essentially unaltered after HT exposure, whereas the TS protein levels were progressively reduced to 25 and 17% of the control on administration of increasing doses of HT for 48 and 72 h, suggesting that HT affects intracellular TS protein levels by translational regulation. Since the expression of the housekeeping gene vinculin (Figure 3.15) remained unaltered upon HT addition, and a control compound did not reduce TS protein levels as much as HT, the observed effect of HT on TS shows specificity.

These results along with our findings above, strongly suggest that HT intercalation into the CC mismatch inhibits TS translation. The exact mechanism of this inhibition is unclear. It is plausible that HT binding stabilizes the site 1 secondary structure not allowing it to become linearized, and thereby preventing it from being supplied into the ribosome for protein synthesis. It is also unclear whether HT affects just the CC mismatch at site 1, or binds additional such bulges in TS mRNA. If it does, what kind of effect would these interactions have on TS protein levels? Additional experiments would be required to address the questions.
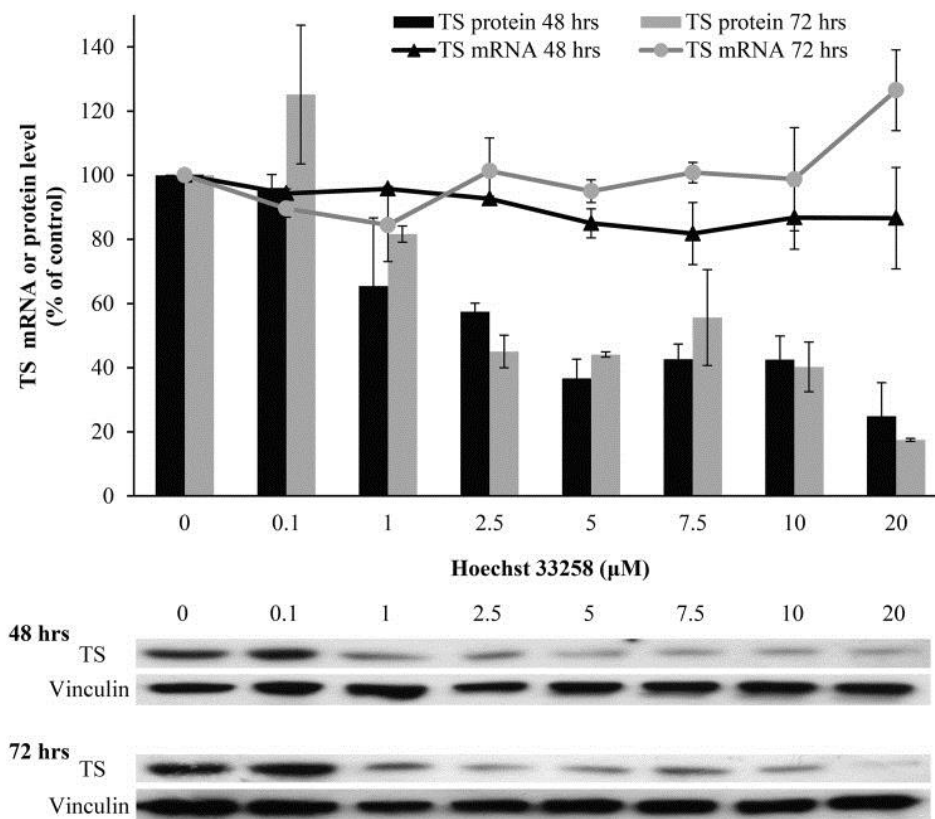
Figure 3.15 – TS protein (in bars) and mRNA levels (marked with a line), monitored with increasing concentrations of HT after 48 (marked in black) and 72 (marked in grey) hours. In the bottom panel, TS protein amounts are compared to those of Vinculin, a housekeeping gene.

As HT acts primarily as a DNA minor groove binder and does so with higher affinity, it cannot by itself be used as a drug candidate. However this study combined with the work on TSMC interaction with paromomycin as well as with other compounds (Tok, Cho et al. 1999, Cho and Rando 2000, Tavares, Beribisky et al. 2009), can be used as a blueprint for future drug design endeavours in this field.

148

References

Adhikary, A., V. Buschmann, C. Muller and M. Sauer (2003). "Ensemble and single-molecule fluorescence spectroscopic study of the binding modes of the bis-benzimidazole derivative Hoechst 33258 with DNA." Nucleic Acids Res **31**(8): 2178-2186.

Alfonzo, J. D., O. Thiemann and L. Simpson (1997). "The mechanism of U insertion/deletion RNA editing in kinetoplastid mitochondria." Nucleic Acids Res **25**(19): 3751-3759.

Arnold, K., L. Bordoli, J. Kopp and T. Schwede (2006). "The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling." Bioinformatics **22**(2): 195-201.

Bagasra, O. and K. R. Prilliman (2004). "RNA interference: the molecular immune system." J Mol Histol **35**(6): 545-553.

Bailly, C., P. Colson, J. P. Henichart and C. Houssier (1993). "The different binding modes of Hoechst 33258 to DNA studied by electric linear dichroism." Nucleic Acids Res **21**(16): 3705-3709.

Bailly, C., P. Colson, C. Houssier and F. Hamy (1996). "The binding mode of drugs to the TAR RNA of HIV-1 studied by electric linear dichroism." Nucleic Acids Res **24**(8): 1460-1464.

Baraldi, P. G., A. Bovero, F. Fruttarolo, D. Preti, M. A. Tabrizi, M. G. Pavani and R. Romagnoli (2004). "DNA minor groove binders as potential antitumor and antimicrobial agents." Med Res Rev **24**(4): 475-528.

Basilio, C., A. J. Wahba, P. Lengyel, J. F. Speyer and S. Ochoa (1962). "Synthetic polynucleotides and the amino acid code. V." Proc Natl Acad Sci U S A **48**: 613-616.

Bass, B. L. (2000). "Double-stranded RNA as a template for gene silencing." Cell **101**(3): 235-238.

Bass, B. L. (2002). "RNA editing by adenosine deaminases that act on RNA." Annu Rev Biochem **71**: 817-846.

Bass, B. L. and H. Weintraub (1988). "An unwinding activity that covalently modifies its double-stranded RNA substrate." Cell **55**(6): 1089-1098.

Bass, B. L., H. Weintraub, R. Cattaneo and M. A. Billeter (1989). "Biased hypermutation of viral RNA genomes could be due to unwinding/modification of double-stranded RNA." Cell **56**(3): 331.

Batey, R. T., M. Inada, E. Kujawinski, J. D. Puglisi and J. R. Williamson (1992). "Preparation of isotopically labeled ribonucleotides for multidimensional NMR spectroscopy of RNA." Nucleic Acids Res. **20**(17): 4515-4523.

Batey, R. T. and J. S. Kieft (2007). "Improved native affinity purification of RNA." RNA **13**(8): 1384-1389.

Becker, E. D. (2000). "High resolution NMR: theory and chemical applications. Third edition.".

Beribisky, A. V. (2008). "Determination of the three-dimensional structure of the Selenocysteine Insertion sequence and analysis of the RNA-binding properties of the Ebola virus transcriptional activator VP30."

Beribisky, A. V., T. J. Tavares, A. N. Amborski, M. Motamed, A. E. Johnson, T. L. Mark and P. E. Johnson (2007). "The three-dimensional structure of the Moorella thermoacetica selenocysteine insertion sequence RNA hairpin and its interaction with the elongation factor SelB." RNA **13**(11): 1948-1956.

Bernstein, E., A. A. Caudy, S. M. Hammond and G. J. Hannon (2001). "Role for a bidentate ribonuclease in the initiation step of RNA interference." Nature **409**(6818): 363-366.

Bhalla, T., J. J. Rosenthal, M. Holmgren and R. Reenan (2004). "Control of human potassium channel inactivation by editing of a small mRNA hairpin." Nat Struct Mol Biol **11**(10): 950-956.

Biswas, R. and M. Sundaralingam (1997). "Crystal structure of r(GUGUGUA)dC with tandem G x U/U x G wobble pairs with strand slippage." J Mol Biol **270**(3): 511-519.

Biswas, R., M. C. Wahl, C. Ban and M. Sundaralingam (1997). "Crystal structure of an alternating octamer r(GUAUGUA)dC with adjacent G x U wobble pairs." J Mol Biol **267**(5): 1149-1156.

Bloomfield, V. A., D. M. Crothers and I. Tinoco (2000). Nucleic acids : structures, properties, and functions. Sausalito, Calif., University Science Books.

Bodenhausen, G. and D. J. Ruben (1980). "Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy." Chemical Physics Letters **69**(1): 185–189.

Burns, C. M., H. Chu, S. M. Rueter, L. K. Hutchinson, H. Canton, E. Sanders-Bush and R. B. Emeson (1997). "Regulation of serotonin-2C receptor G-protein coupling by RNA editing." Nature **387**(6630): 303-308.

Case, D. A., T. E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang and R. J. Woods (2005). "The Amber biomolecular simulation programs." J Comput Chem **26**(16): 1668-1688.

Cattaneo, R., A. Schmid, D. Eschle, K. Baczko, V. ter Meulen and M. A. Billeter (1988). "Biased hypermutation and other genetic changes in defective measles viruses in human brain infections." Cell **55**(2): 255-265.

Caudy, A. A., R. F. Ketting, S. M. Hammond, A. M. Denli, A. M. Bathoorn, B. B. Tops, J. M. Silva, M. M. Myers, G. J. Hannon and R. H. Plasterk (2003). "A micrococcal nuclease homologue in RNAi effector complexes." Nature **425**(6956): 411-414.

Cavanagh, J. (2007). "Protein NMR Spectroscopy, second edition."

Check, E. (2007). "RNA interference: hitting the on switch." Nature **448**(7156): 855-858.

Cheng, J. C., T. B. Moore and K. M. Sakamoto (2003). "RNA interference and human disease." Mol Genet Metab **80**(1-2): 121-128.

Cho, J. and R. R. Rando (1999). "Specificity in the binding of aminoglycosides to HIV-RRE RNA." Biochemistry **38**(26): 8548-8554.

Cho, J. and R. R. Rando (2000). "Specific binding of Hoechst 33258 to site 1 thymidylate synthase mRNA." Nucleic Acids Res **28**(10): 2158-2163.

Chu, E., D. M. Koeller, J. L. Casey, J. C. Drake, B. A. Chabner, P. C. Elwood, S. Zinn and C. J. Allegra (1991). "Autoregulation of human thymidylate synthase messenger RNA translation by thymidylate synthase." Proc Natl Acad Sci U S A **88**(20): 8977-8981.

Chu, E., D. Voeller, D. M. Koeller, J. C. Drake, C. H. Takimoto, G. F. Maley, F. Maley and C. J. Allegra (1993). "Identification of an RNA binding site for human thymidylate synthase." Proc Natl Acad Sci U S A **90**(2): 517-521.

Colson, P., C. Houssier and C. Bailly (1995). "Use of electric linear dichroism and competition experiments with intercalating drugs to investigate the mode of binding of Hoechst 33258, berenil and DAPI to GC sequences." J Biomol Struct Dyn **13**(2): 351-366.

Cotton, F. A., E. E. Hazen, Jr. and M. J. Legg (1979). "Staphylococcal nuclease: proposed mechanism of action based on structure of enzyme-thymidine 3',5'-bisphosphate-calcium ion complex at 1.5-A resolution." Proc Natl Acad Sci U S A **76**(6): 2551-2555.

Das, G. and R. H. D. Lyngdoh (2008). "Can configuration of solitary wobble base pairs determine the specificity and degeneracy of the genetic code? Clues from molecular orbital modelling studies." Journal of Molecular Structure: THEOCHEM **851**: 319-334.

Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Bax (1995). "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." J Biomol NMR **6**(3): 277-293.

Dibrov, S., J. McLean and T. Hermann (2011). "Structure of an RNA dimer of a regulatory element from human thymidylate synthase mRNA." Acta Crystallogr D Biol Crystallogr **67**(Pt 2): 97-104.

Duss, O., C. Maris, C. von Schroetter and F. H. 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 Res **38**(20): e188.

Easton, L. E., Y. Shibata and P. J. Lukavsky (2010). "Rapid, nondenaturing RNA purification using weak anion-exchange fast performance liquid chromatography." RNA **16**(3): 647-653.

Eletsky, A., A. Kienhofer and K. Pervushin (2001). "TROSY NMR with partially deuterated proteins." J Biomol NMR **20**(2): 177-180.

Fernandez, C. and G. Wider (2006). "NMR Spectroscopy of Large Biological Macromolecules in Solution."

Fire, A., S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver and C. C. Mello (1998). "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans." Nature **391**(6669): 806-811.

Fourmy, D., S. Yoshizawa and J. D. Puglisi (1998). "Paromomycin binding induces a local conformational change in the A-site of 16 S rRNA." J Mol Biol **277**(2): 333-345.

Francois, B., R. J. Russell, J. B. Murray, F. Aboul-ela, B. Masquida, Q. Vicens and E. Westhof (2005). "Crystal structures of complexes between aminoglycosides and decoding A site oligonucleotides: role of the number of rings and positive charges in the specific binding leading to miscoding." Nucleic Acids Res **33**(17): 5677-5690.

Friberg, A., L. Corsini, A. Mourao and M. Sattler (2009). "Structure and ligand binding of the extended Tudor domain of D. melanogaster Tudor-SN." J Mol Biol **387**(4): 921-934.

Furse, K. E., B. A. Lindquist and S. A. Corcelli (2008). "Solvation dynamics of Hoechst 33258 in water: an equilibrium and nonequilibrium molecular dynamics study." J Phys Chem B **112**(10): 3231-3239.

Fürtig, B., C. Richter, J. Wöhnert and H. Schwalbe (2003). "NMR spectroscopy of RNA." Chembiochem **4**(10): 936-962.

Gaffney, B. L., P.-P. Kung, C. Wang and R. A. Jones (1995). "Nitrogen-15-labeled oligodeoxynucleotides. 8. Use of 15N NMR to probe Hoogsteen hydrogen bonding at guanine and adenine N7 atoms of a DNA triplex." Journal of the American Chemical Society **117**: 12281-12283.

Gardner, K. H. and L. E. Kay (1998). "The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins." Annu Rev Biophys Biomol Struct. **27**: 357-406.

Garg, D., A. V. Beribisky, G. Ponterini, A. Ligabue, G. Marverti, A. Martello, M. P. Costi, M. Sattler and R. C. Wade (2013). "Translational repression of thymidylate synthase by targeting its mRNA." Nucleic Acids Res **41**(7): 4159-4170.

Garg, D., S. Henrich, O. M. Salo-Ahen, H. Myllykallio, M. P. Costi and R. C. Wade (2010). "Novel approaches for targeting thymidylate synthase to overcome the resistance and toxicity of anticancer drugs." J Med Chem **53**(18): 6539-6549.

Goddard, T. D., D. G. Kneller and "SPARKY 3." University of California, San Francisco.

Gregory, R. I., T. P. Chendrimada and R. Shiekhattar (2006). "MicroRNA biogenesis: isolation and characterization of the microprocessor complex." Methods Mol Biol **342**: 33-47.

Griesinger, C., G. Otting, K. Wüthrich and R. R. Ernst (1988). "Clean TOCSY for 1H spin system identification in macromolecules. ." Journal of American Chemical Society. **110**: 7870-7872.

Guntert, P. (2004). "Automated NMR structure calculation with CYANA." Methods Mol Biol **278**: 353-378.

Hale, S. P., L. B. Poole and J. A. Gerlt (1993). "Mechanism of the reaction catalyzed by staphylococcal nuclease: identification of the rate-determining step." Biochemistry **32**(29): 7479-7487.

Hasnat, A., E. Bichenkova, X. Yu, J. R. Arnold, J. Fisher, O. Fedorova and J. Andrews (2007). "Fluorescence spectroscopic and (19)f NMR studies of human thymidylate synthase with its cognate RNA." J Biomol Struct Dyn **25**(3): 253-270.

Heinrich, E. M., J. Wagner, M. Kruger, D. John, S. Uchida, J. E. Weigand, B. Suess and S. Dimmeler (2013). "Regulation of miR-17-92a cluster processing by the microRNA binding protein SND1." FEBS Lett **587**(15): 2405-2411.

Horowitz, E. D. and N. V. Hud (2006). "Ethidium and proflavine binding to a 2',5'-linked RNA duplex." J Am Chem Soc **128**(48): 15380-15381.

Horowitz, E. D., S. Lilavivat, B. W. Holladay, M. W. Germann and N. V. Hud (2009). "Solution structure and thermodynamics of 2',5' RNA intercalation." J Am Chem Soc **131**(16): 5831-5838.

Inagaki, F., I. Shimada, D. Kohda, A. Suzuki and A. Bax (1989). "Relayed HOHAHA: A useful method for extracting subspectra of individual components of sugar chains. ." J. Magn. Reson. **81**: 186-190

James, T. L. (1998). "Fundamentals of NMR."

Janke, E. M., F. Riechert-Krause and K. Weisz (2011). "Low-temperature NMR studies on inosine wobble base pairs." J Phys Chem B **115**(26): 8569-8574.

Johnson, B. A. and R. A. Blevins (1994). "NMR View: A computer program for the visualization and analysis of NMR data." J Biomol NMR **4**(5): 603-614.

Johnson, P. E. and L. W. Donaldson (2006). "RNA recognition by the Vts1p SAM domain." Nat Struct Mol Biol **13**(2): 177-178.

Jung, Y. S. and M. Zweckstetter (2004). "Mars -- robust automatic backbone assignment of proteins." J Biomol NMR **30**(1): 11-23.

Katiyar-Agarwal, S., R. Morgan, D. Dahlbeck, O. Borsani, A. Villegas, Jr., J. K. Zhu, B. J. Staskawicz and H. Jin (2006). "A pathogen-inducible endogenous siRNA in plant immunity." Proc Natl Acad Sci U S A **103**(47): 18002-18007.

Kawahara, I., K. Haruta, Y. Ashihara, D. Yamanaka, M. Kuriyama, N. Toki, Y. Kondo, K. Teruya, J. Ishikawa, H. Furuta, Y. Ikawa, C. Kojima and Y. Tanaka (2012). "Site-specific isotope labeling of long RNA for structural and mechanistic studies." Nucleic Acids Res **40**(1): e7.

Kawahara, Y., B. Zinshteyn, T. P. Chendrimada, R. Shiekhattar and K. Nishikura (2007). "RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex." EMBO Rep **8**(8): 763-769.

Keeler, J. (2002). "Understanding NMR Spectroscopy, Second Edition."

Kloiber, K., R. Spitzer, M. Tollinger, R. Konrat and C. Kreutz (2011). "Probing RNA dynamics via longitudinal exchange and CPMG relaxation dispersion NMR spectroscopy using a sensitive 13C-methyl label." Nucleic Acids Res **39**(10): 4340-4351.

Komohara, Y., H. Yano, S. Shichijo, K. Shimotohno, K. Itoh and A. Yamada (2006). "High expression of APOBEC3G in patients infected with hepatitis C virus." J Mol Histol **37**(8-9): 327-332.

Kumar, A., R. R. Ernst and K. Wuthrich (1980). "A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules." Biochem Biophys Res Commun **95**(1): 1-6.

Lai, F., R. Drakas and K. Nishikura (1995). "Mutagenic analysis of double-stranded RNA adenosine deaminase, a candidate enzyme for RNA editing of glutamate-gated ion channel transcripts." J Biol Chem **270**(29): 17098-17105.

Lee, R. C., R. L. Feinbaum and V. Ambros (1993). "The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14." Cell **75**(5): 843-854.

Lehmann, K. A. and B. L. Bass (1999). "The importance of internal loops within RNA substrates of ADAR1." J Mol Biol **291**(1): 1-13.

Lehmann, K. A. and B. L. Bass (2000). "Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities." Biochemistry **39**(42): 12875-12884.

Lehninger, A. L., D. L. Nelson and M. M. Cox (2000). Lehninger principles of biochemistry. New York, Worth Publishers.

Leuschner, P. J., S. L. Ameres, S. Kueng and J. Martinez (2006). "Cleavage of the siRNA passenger strand during RISC assembly in human cells." EMBO Rep **7**(3): 314-320.

Leverson, J. D., P. J. Koskinen, F. C. Orrico, E. M. Rainio, K. J. Jalkanen, A. B. Dash, R. N. Eisenman and S. A. Ness (1998). "Pim-1 kinase and p100 cooperate to enhance c-Myb activity." Mol Cell **2**(4): 417-425.

Levitt, M. H. (2008). "Spin Dynamics: Basics of Nuclear Magnetic Resonance, 2nd edition, ."

Li, C. L., W. Z. Yang, Y. P. Chen and H. S. Yuan (2008). "Structural and functional insights into human Tudor-SN, a key component linking RNA interference and editing." Nucleic Acids Res **36**(11): 3579-3589.

Liu, J., M. A. Carmell, F. V. Rivas, C. G. Marsden, J. M. Thomson, J. J. Song, S. M. Hammond, L. Joshua-Tor and G. J. Hannon (2004). "Argonaute2 is the catalytic engine of mammalian RNAi." Science **305**(5689): 1437-1441.

Liu, M., X.-A. Mao, C. Ye, H. Huang, J. K. Nicholson and L. J.C. (1998). "Improved WATERGATE Pulse Sequences for Solvent Suppression in NMR Spectroscopy." J. Magn. Reson. **132**: 125–129.

Liu, Q., T. Rand, S. Kalidas, F. Du, H. Kim, D. Smith and X. Wang (2003). "R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway." Science **301**(5641): 1921-1925.

Liu, Z., M. Reches, I. Groisman and H. Engelberg-Kulka (1998). "The nature of the minimal 'selenocysteine insertion sequence' (SECIS) in Escherichia coli." Nucleic Acids Res **26**(4): 896-902.

Lukavsky, P. J. and J. D. Puglisi (2004). "Large-scale preparation and purification of polyacrylamide-free RNA oligonucleotides." RNA **10**(5): 889-893.

Maas, S., Y. Kawahara, K. M. Tamburro and K. Nishikura (2006). "A-to-I RNA editing and human disease." RNA Biol **3**(1): 1-9.

Macbeth, M. R., A. T. Lingam and B. L. Bass (2004). "Evidence for auto-inhibition by the N terminus of hADAR2 and activation by dsRNA binding." RNA **10**(10): 1563-1571.

Markus, M. A., K. T. Dayie, P. Matsudaira and G. Wagner (1994). "Effect of deuteration on the amide proton relaxation rates in proteins. Heteronuclear NMR experiments on villin 14T." J Magn Reson B **105**(2): 192-195.

Martin, G. E. and A. S. Zektzer (1988). "Two-Dimensional NMR Methods for Establishing Molecular Connectivity."

Mbisa, J. L., R. Barr, J. A. Thomas, N. Vandegraaff, I. J. Dorweiler, E. S. Svarovskaia, W. L. Brown, L. M. Mansky, R. J. Gorelick, R. S. Harris, A. Engelman and V. K. Pathak (2007). "Human

immunodeficiency virus type 1 cDNAs produced in the presence of APOBEC3G exhibit defects in plus-strand DNA transfer and integration." J Virol **81**(13): 7099-7110.

Milligan, J. F. and O. C. Uhlenbeck (1989). "Synthesis of small RNAs using T7 RNA polymerase." Methods Enzymol **180**: 51-62.

Muramatsu, M., K. Kinoshita, S. Fagarasan, S. Yamada, Y. Shinkai and T. Honjo (2000). "Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme." Cell **102**(5): 553-563.

Nakamuta, M., K. Oka, J. Krushkal, K. Kobayashi, M. Yamamoto, W. H. Li and L. Chan (1995). "Alternative mRNA splicing and differential promoter utilization determine tissue-specific expression of the apolipoprotein B mRNA-editing protein (Apobec1) gene in mice. Structure and evolution of Apobec1 and related nucleoside/nucleotide deaminases." J Biol Chem **270**(22): 13042-13056.

Nelissen, F. H., A. J. van Gammeren, M. Tessari, F. C. Girard, H. A. Heus and S. S. Wijmenga (2008). "Multiple segmental and selective isotope labeling of large RNA for NMR structural studies." Nucleic Acids Res **36**(14): e89.

Nikolova, E. N., F. L. Gottardo and H. M. Al-Hashimi (2012). "Probing transient Hoogsteen hydrogen bonds in canonical duplex DNA using NMR relaxation dispersion and single-atom substitution." J Am Chem Soc **134**(8): 3667-3670.

Nikonowicz, E. P. and A. Pardi (1992). "Three-dimensional heteronuclear NMR studies of RNA." Nature **355**(6356): 184-186.

Nishikura, K., C. Yoo, U. Kim, J. M. Murray, P. A. Estes, F. E. Cash and S. A. Liebhaber (1991). "Substrate specificity of the dsRNA unwinding/modifying activity." EMBO J **10**(11): 3523-3532.

Pan, B., S. N. Mitra, L. Sun, D. Hart and M. Sundaralingam (1998). "Crystal structure of an RNA octamer duplex r(CCCIUGGG)2 incorporating tandem I.U wobbles." Nucleic Acids Res **26**(24): 5699-5706.

Parker, G. S., D. M. Eckert and B. L. Bass (2006). "RDE-4 preferentially binds long dsRNA and its dimerization is necessary for cleavage of dsRNA to siRNA." RNA **12**(5): 807-818.

Paukku, K., J. Yang and O. Silvennoinen (2007). "Tudor and nuclease-like domains containing protein p100 function as coactivators for signal transducer and activator of transcriptional 5." Mol. Endocrinol. **17**: 1805–1814.

Perez, A., I. Marchan, D. Svozil, J. Sponer, T. E. Cheatham, 3rd, C. A. Laughton and M. Orozco (2007). "Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers." Biophys J **92**(11): 3817-3829.

Pervushin, K., R. Riek, G. Wider and K. Wüthrich (1997). "Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution." Proc. Natl. Acad. Sci. USA **94**(23): 12366-12371.

Piotto, M., V. Saudek and V. Sklenar (1992). "Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions." J Biomol NMR **2**(6): 661-665.

Polson, A. G., B. L. Bass and J. L. Casey (1996). "RNA editing of hepatitis delta virus antigenome by dsRNA-adenosine deaminase." Nature **380**(6573): 454-456.

Powell, L. M., S. C. Wallis, R. J. Pease, Y. H. Edwards, T. J. Knott and J. Scott (1987). "A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine." Cell **50**(6): 831-840.

Rieping, W., M. Habeck, B. Bardiaux, A. Bernard, T. E. Malliavin and M. Nilges (2007). "ARIA2: automated NOE assignment and data integration in NMR structure calculation." Bioinformatics **23**(3): 381-382.

Rueter, S. M., T. R. Dawson and R. B. Emeson (1999). "Regulation of alternative splicing by RNA editing." Nature **399**(6731): 75-80.

Ryter, J. M. and S. C. Schultz (1998). "Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA." EMBO J **17**(24): 7505-7513.

Salzmann, M., K. Pervushin, G. Wider, H. Senn and K. Wuthrich (1998). "TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins." Proc Natl Acad Sci U S A **95**(23): 13585-13590.

Salzmann, M., K. Pervushin, G. Wider, H. Senn and K. Wuthrich (1999). "[13C]-constant-time [15N,1H]-TROSY-HNCA for sequential assignments of large proteins." J Biomol NMR **14**(1): 85-88.

Sambrook, J., E. F. Fritsch and T. Maniatis (1989). "Molecular Cloning: A Laboratory Manual."

Sanders, J. K. M. and B. K. Hunter (1993). "Modern NMR spectroscopy : a guide for chemists. 2nd edition."

Sattler, M., J. Schleucher and C. Griesinger (1999). "Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients." Prog. NMR Spectrosc. **34**: 93-158.

Saumet, A. and C. H. Lecellier (2006). "Anti-viral RNA silencing: do we look like plants?" Retrovirology **3**: 3.

Scadden, A. D. (2005). "The RISC subunit Tudor-SN binds to hyper-edited double-stranded RNA and promotes its cleavage." Nat Struct Mol Biol **12**(6): 489-496.

Scadden, A. D. and C. W. Smith (1997). "A ribonuclease specific for inosine-containing RNA: a potential role in antiviral defence?" EMBO J **16**(8): 2140-2149.

Scadden, A. D. and C. W. Smith (2001). "RNAi is antagonized by A-->I hyper-editing." EMBO Rep **2**(12): 1107-1111.

Scadden, A. D. and C. W. Smith (2001). "Specific cleavage of hyper-edited dsRNAs." EMBO J **20**(15): 4243-4252.

Schade, M., C. J. Turner, R. Kuhne, P. Schmieder, K. Lowenhaupt, A. Herbert, A. Rich and H. Oschkinat (1999). "The solution structure of the Zalpha domain of the human RNA editing enzyme ADAR1 reveals a prepositioned binding surface for Z-DNA." Proc Natl Acad Sci U S A **96**(22): 12465-12470.

Schultheisz, H. L., B. R. Szymczyna, L. G. Scott and J. R. Williamson (2008). "Pathway engineered enzymatic de novo purine nucleotide synthesis." ACS Chem Biol **3**(8): 499-511.

Schultheisz, H. L., B. R. Szymczyna, L. G. Scott and J. R. Williamson (2011). "Enzymatic de novo pyrimidine nucleotide synthesis." J Am Chem Soc **133**(2): 297-304.

Schwarz, D. S., G. Hutvagner, T. Du, Z. Xu, N. Aronin and P. D. Zamore (2003). "Asymmetry in the assembly of the RNAi enzyme complex." Cell **115**(2): 199-208.

Schwieters, C. D., J. J. Kuszewski, N. Tjandra and G. M. Clore (2003). "The Xplor-NIH NMR molecular structure determination package." J Magn Reson **160**(1): 65-73.

Serra, M. J., P. E. Smolter and E. Westhof (2004). "Pronounced instability of tandem IU base pairs in RNA." Nucleic Acids Res **32**(5): 1824-1828.

Shaw, N., M. Zhao, C. Cheng, H. Xu, J. Saarikettu, Y. Li, Y. Da, Z. Yao, O. Silvennoinen, J. Yang, Z. J. Liu, B. C. Wang and Z. Rao (2007). "The multifunctional human p100 protein 'hooks' methylated ligands." Nat Struct Mol Biol **14**(8): 779-784.

Sheehy, A. M., N. C. Gaddis, J. D. Choi and M. H. Malim (2002). "Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein." Nature **418**(6898): 646-650.

Shen, Y., F. Delaglio, G. Cornilescu and A. Bax (2009). "TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts." J Biomol NMR **44**(4): 213-223.

Simon, B., K. Zanier and M. Sattler (2001). "A TROSY relayed HCCH-COSY experiment for correlating adenine H2/H8 resonances in uniformly 13C-labeled RNA molecules." J. Biomol. NMR **20**(2): 173-176.

Singh, R. and J. Valcarcel (2005). "Building specificity with nonspecific RNA-binding proteins." Nat Struct Mol Biol **12**(8): 645-653.

Sklenar, V., M. Piotto, R. Leppik and V. Saudek (1993). "Gradient-tailored water suppression for 1H-15N HSQC experiments optimized to retain full sensitivity." J. Magn. Reson. Ser. A. **102**: 241–245.

Sundstrom, J. F., A. Vaculova, A. P. Smertenko, E. I. Savenkov, A. Golovko, E. Minina, B. S. Tiwari, S. Rodriguez-Nieto, A. A. Zamyatnin, Jr., T. Valineva, J. Saarikettu, M. J. Frilander, M. F. Suarez, A. Zavialov, U. Stahl, P. J. Hussey, O. Silvennoinen, E. Sundberg, B. Zhivotovsky and P. V. Bozhkov (2009). "Tudor staphylococcal nuclease is an evolutionarily conserved component of the programmed cell death degradome." Nat Cell Biol **11**(11): 1347-1354.

Tavares, T. J., A. V. Beribisky and P. E. Johnson (2009). "Structure of the cytosine-cytosine mismatch in the thymidylate synthase mRNA binding site and analysis of its interaction with the aminoglycoside paromomycin." RNA **15**(5): 911-922.

Theobald, D. L., R. M. Mitton-Fry and D. S. Wuttke (2003). "Nucleic acid recognition by OB-fold proteins." Annu Rev Biophys Biomol Struct **32**: 115-133.

Tjandra, N. and A. Bax (1997). "Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium." Science **278**(5340): 1111-1114.

Tok, J. B., J. Cho and R. R. Rando (1999). "Aminoglycoside antibiotics are able to specifically bind the 5'-untranslated region of thymidylate synthase messenger RNA." Biochemistry **38**(1): 199-206.

Tok, J. B., J. Cho and R. R. Rando (2000). "RNA aptamers that specifically bind to a 16S ribosomal RNA decoding region construct." Nucleic Acids Res **28**(15): 2902-2910.

Tomari, Y., C. Matranga, B. Haley, N. Martinez and P. D. Zamore (2004). "A protein sensor for siRNA asymmetry." Science **306**(5700): 1377-1380.

Tong, X., R. Drapkin, R. Yalamanchili, G. Mosialos and E. Kieff (1995). "The Epstein-Barr virus nuclear protein 2 acidic domain forms a complex with a novel cellular coactivator that can interact with TFIIE." Mol Cell Biol **15**(9): 4735-4744.

Tzakos, A. G., L. E. Easton and P. J. Lukavsky (2007). "Preparation of large RNA oligonucleotides with complementary isotope-labeled segments for NMR structural studies." Nat Protoc **2**(9): 2139-2147.

Valineva, T., J. Yang, R. Palovuori and O. Silvennoinen (2005). "The transcriptional co-activator protein p100 recruits histone acetyltransferase activity to STAT6 and mediates interaction between the CREB-binding protein and STAT6." J Biol Chem **280**(15): 14989-14996.

Valineva, T., J. Yang and O. Silvennoinen (2006). "Characterization of RNA helicase A as component of STAT6-dependent enhanceosome." Nucleic Acids Res **34**(14): 3938-3946.

Varani, G., F. Aboul-ela and F. H.-T. Allain (1996). "NMR investigation of RNA structure." <u>Progr. NMR Spectrosc.</u> **29**: 51-127.

Varani, G. and I. Tinoco, Jr. (1991). "RNA structure and NMR spectroscopy." <u>Q. Rev. Biophys.</u> **24**(4): 479-532.

Vega, M. C., I. Garcia Saez, J. Aymami, R. Eritja, G. A. Van der Marel, J. H. Van Boom, A. Rich and M. Coll (1994). "Three-dimensional crystal structure of the A-tract DNA dodecamer d(CGCAAATTTGCG) complexed with the minor-groove-binding drug Hoechst 33258." <u>Eur J Biochem</u> **222**(3): 721-726.

Vermeulen, A., L. Behlen, A. Reynolds, A. Wolfson, W. S. Marshall, J. Karpilow and A. Khvorova (2005). "The contributions of dsRNA structure to Dicer specificity and efficiency." <u>RNA</u> **11**(5): 674-682.

Vicens, Q. and E. Westhof (2001). "Crystal structure of paromomycin docked into the eubacterial ribosomal decoding A site." <u>Structure</u> **9**(8): 647-658.

Vuister, G. W. and A. Bax (1992). "Measurement of two-bond JCOH alpha coupling constants in proteins uniformly enriched with 13C." <u>J Biomol NMR</u> **2**(4): 401-405.

Wade, L. G. (1999). "Organic Chemistry, fourth edition."

Watkins, N. E., Jr. and J. SantaLucia, Jr. (2005). "Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes." <u>Nucleic Acids Res</u> **33**(19): 6258-6267.

Weik, M., J. Modrof, H. D. Klenk, S. Becker and E. Muhlberger (2002). "Ebola virus VP30-mediated transcription is regulated by RNA secondary structure formation." <u>J Virol</u> **76**(17): 8532-8539.

Wenter, P., L. Reymond, S. D. Auweter, F. H. Allain and S. Pitsch (2006). "Short, synthetic and selectively 13C-labeled RNA sequences for the NMR structure determination of protein-RNA complexes." <u>Nucleic Acids Res</u> **34**(11): e79.

Williamson, M. P. (2009). "Applications of the NOE in Molecular Biology." <u>Annual Reports on NMR Spectroscopy</u>.

Wishart, D. S. and B. D. Sykes (1994). "The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data." <u>J Biomol NMR</u> **4**(2): 171-180.

Wunderlich, C. H., R. Spitzer, T. Santner, K. Fauster, M. Tollinger and C. Kreutz (2012). "Synthesis of (6-(13)C)pyrimidine nucleotides as spin-labels for RNA dynamics." <u>J Am Chem Soc</u> **134**(17): 7558-7569.

Wüthrich, K. (1986). <u>NMR of Proteins and Nucleic Acids</u>. New York, Wiley-Interscience.

Yang, J., T. Valineva, J. Hong, T. Bu, Z. Yao, O. N. Jensen, M. J. Frilander and O. Silvennoinen (2007). "Transcriptional co-activator protein p100 interacts with snRNP proteins and facilitates the assembly of the spliceosome." <u>Nucleic Acids Res</u> **35**(13): 4485-4494.

Yang, W., Q. Wang, K. L. Howell, J. T. Lee, D. S. Cho, J. M. Murray and K. Nishikura (2005). "ADAR1 RNA deaminase limits short interfering RNA efficacy in mammalian cells." <u>J Biol Chem</u> **280**(5): 3946-3953.

Yi, R., Y. Qin, I. G. Macara and B. R. Cullen (2003). "Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs." <u>Genes Dev</u> **17**(24): 3011-3016.

Zambon, R. A., V. N. Vakharia and L. P. Wu (2006). "RNAi is an antiviral immune response against a dsRNA virus in Drosophila melanogaster." <u>Cell Microbiol</u> **8**(5): 880-889.

Zapp, M. L., S. Stern and M. R. Green (1993). "Small molecules that selectively block RNA binding of HIV-1 rev protein inhibit rev function and viral production." <u>Cell</u> **74**(6): 969-978.

Zhang, B., X. Pan, G. P. Cobb and T. A. Anderson (2006). "Plant microRNA: a small regulatory molecule with big impact." <u>Dev Biol</u> **289**(1): 3-16.

Zhang, B., X. Pan, G. P. Cobb and T. A. Anderson (2007). "microRNAs as oncogenes and tumor suppressors." <u>Dev Biol</u> **302**(1): 1-12.

Appendices

Appendix A – Plasmid maps

Appendix A1 – pETZZ-1a
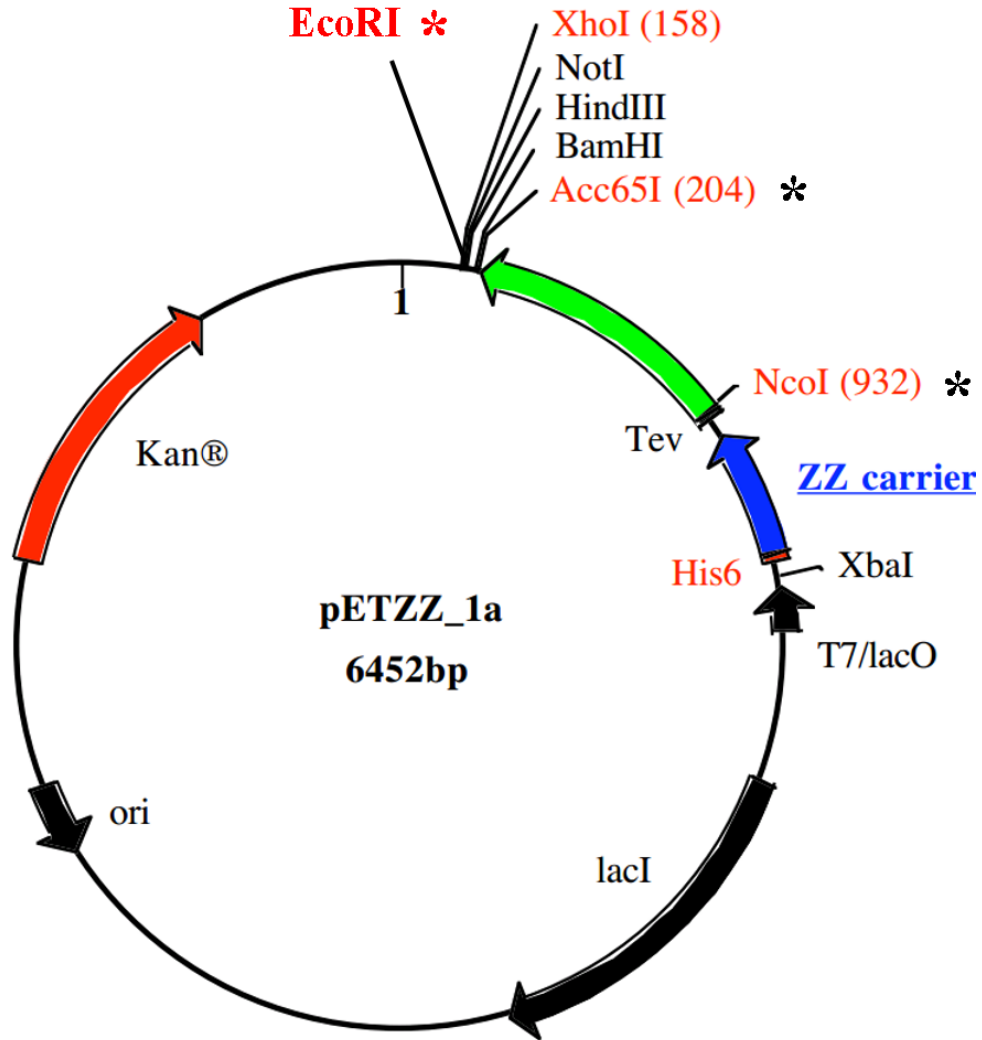
The plasmid map of pETMZZ-a1. The endonuclease cleavage sites used to clone p100HS34 are marked with two black asterisks, while the cleavage sites for p100DM3 and p100DM4 are marked with black and red asterisks.

Appendix A2 – pETM11
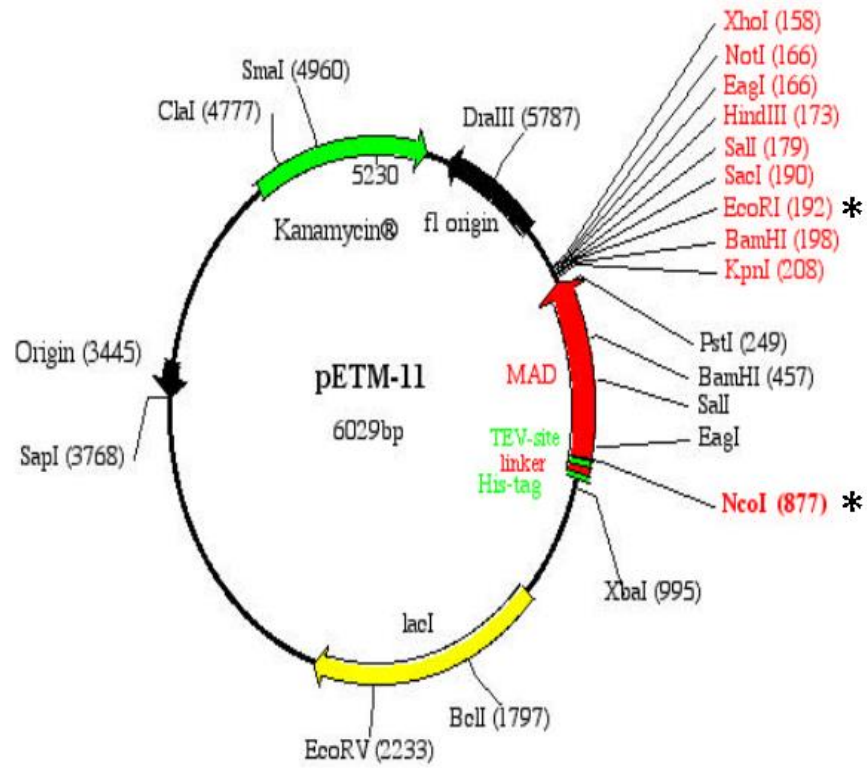
The plasmid map of pETM11. The endonuclease cleavage sites used to clone p100HS34 are marked with an asterisk

Appendix B – Construct information for p100 constructs

Appendix B1 – Construct information for p100DM34

<table>
<tr><td colspan="3" align="center">p100DM34</td></tr>
<tr><td rowspan="6">Construct information</td><td>Protein name</td><td>p100 (also Tudor-SN, SND1)</td></tr>
<tr><td>Organism of origin</td><td>Drosophila melanogaster.</td></tr>
<tr><td>Part(s) encompassed</td><td>Staphylococcal nuclease domains 3 and 4</td></tr>
<tr><td>Residue numbers</td><td>345-687 (343 residues)</td></tr>
<tr><td>MW</td><td>38.5 kDa</td></tr>
<tr><td>Extinction coefficient</td><td>38.4 mM^{-1}cm^{-1}</td></tr>
<tr><td rowspan="5">Plasmid Information</td><td>Plasmid name</td><td>pETZZ-1a</td></tr>
<tr><td>Antibiotic</td><td>Kanamycin</td></tr>
<tr><td>Tag(s)</td><td>N-terminal His-tag followed by double Z-tag (ZZ-tag)</td></tr>
<tr><td>Enzyme for tag cleavage</td><td>TEV protease</td></tr>
<tr><td>Restriction enzymes used</td><td>NcoI and Acc65I</td></tr>
<tr><td colspan="2" align="center">Total number of residues</td><td>343 + 4 (GAMG from TEV cleavage site) = 347</td></tr>
<tr><td colspan="2" rowspan="2" align="center">Primer Sequences</td><td>F:GAGCCCATGGGGGAGAAAGACTTCTCTGGAACAGTG</td></tr>
<tr><td>R:GAGCGGTACCTTATTATCACTCGGTGACAGTTTTTTCT</td></tr>
<tr><td colspan="2" align="center">DNA Sequence</td><td>

```
GGCGCCATGGGGGGAGAAAGACTTCTCTGGAACAGTGGTGGAA
GTCTTCAATGGCGATGCTATTAATGTACGTCTTTCCAATGGA
CAAGTAAAGAAGGTTTTCTTCTCGTCTATCAGGCCGCCGCGT
GACCAACGCGCGGTTGTTGGCACCGATGGCGAGGAAATCGTC
AAGGCACCACCACGCGGAAAAAACTACAGACCACTTTACGAA
ATTCCACACATGTTCGATGCCCGTGAGTTTCTGCGTAAAAAA
CTAATTAATAAAAAGTTCAATGCAACCTGGACTACATCTCA
CCGCCGCGCGAAAACTTCCCTGAGAAATACTGCTATACCGTC
TCTATTGGCGGTCAGAACGTGGCTGAGGCAATGGTAGCCAAA
GGTTTGGCAACATGTGTTCGCTATCGCCAGGACGACGATCAG
AGATCTTCCGCCTATGATCAATTGATTGCCGCCGAACAGCAA
GCCATCAAGGGACTAAAGGGACTGCACGCAAAGAAGGATAAC
GCCACACTGCGTGTTAACGACCTGACTGTTGATCATTCCCGA
ATTAAAGTTCAATACTTGCCATCGTGGCAGCGTGCTTTGCGT
ACTGAAGCCATTGTAGAGTTTGTCGCCAGCGGATCCCGCCTC
CGAATCTTTGTACCGAAGGACAGTTGCCTGGTAACATTCCTG
CTTGCTGGGATTTCATGTCCGCGTTCCTCTCGTCCAGCGCTT
```

</td></tr>
</table>

	AATGGCGTTCCTGCCCAAGAGGGGGAACCATTTGGCGACGAG GCTTTGACCTTTACCAGGGAGCGTGTCTTGCAACGCGACGTT TCTGTTCATATCGATACCACCGACAAGGCAGGATCTTCAGTG ATTGGTTGGCTCTGGACGGACAGCGGCGCCAACTTGTCAGTT GCTCTTGTCGAAGAGGGTCTTGCTGAAGTGCACTTTAGTGCT GAAAAATCTGAATACTACCGCCAGTTGAAGATTGCTGAAGAC CGTGCCAAGGCCGCCAAAAAGAACATCTGGACAAACTACGTA GAAGAAGTGCCTAAAGAAAAAACTGTCACCGAG
Protein sequence	EKDFSGTVVEVFNGDAINVRLSNGQVKKVFFSSIRPPRDQRA VVGTDGEEIVKAPPRGKNYRPLYEIPHMFDAREFLRKKLINK KVQCNLDYISPPRENFPEKYCYTVSIGGQNVAEAMVAKGLAT CVRYRQDDDQRSSAYDQLIAAEQQAIKGLKGLHAKKDNATLR VNDLTVDHSRIKVQYLPSWQRALRTEAIVEFVASGSRLRIFV PKDSCLVTFLLAGISCPRSSRPALNGVPAQEGEPFGDEALTF TRERVLQRDVSVHIDTTDKAGSSVIGWLWTDSGANLSVALVE EGLAEVHFSAEKSEYYRQLKIAEDRAKAAKKNIWTNYVEEVP KEKTVTE
Note 1 :	The NcoI and Acc65I restriction sites on the forward and reverse primer respectively, are marked in bold
Note 2:	The p100DM34 DNA and amino sequences are listed in Appendix B1

Appendix B2 – Construct information for p100HS34

p100HS34		
Construct information	Protein name	p100 (also Tudor-SN, SND1)
	Organism of origin	*Homo sapiens*
	Part(s) encompassed	Staphylococcal nuclease domains 3 and 4
	Residue numbers	345-687 (343 residues)
	MW	38.5 kDa
	Extinction coefficient	28.9 mM^{-1}cm^{-1}
Plasmid Information	Plasmid name	pETM-11
	Antibiotic	Kanamycin
	Tag(s)	N-terminal His-tag
	Enzyme for tag cleavage	TEV protease
	Restriction enzymes used	NcoI and EcoRI

Total number of residues	343 + 4 (GAMG from TEV cleavage site) = 347
Forward and Reverse primer sequences	`F:GAG`**`CCCATG`**`GGGGCCAAGGTGATGCAGGT`
	`R:GAGC`**`GAATTC`**`TTATTACACAAACACGGGCTTGTAGC`
DNA Sequence	GCCAAGGTGATGCAGGTTCTGAATGCTGATGCCATTGTTGTG AAGCTGAACTCAGGCGATTACAAGACGATTCACCTGTCCAGC ATCCGACCACCGAGGCTGGAGGGGGAGAACACCCAGGATAAG AACAAGAAACTGCGTCCCCTGTATGACATTCCTTACATGTTT GAGGCCCGGGAATTTCTTCGAAAAAAGCTTATTGGGAAGAAG GTCAATGTGACGGTGGACTACATTAGACCAGCCAGCCCAGCC ACAGAGACAGTGCCTGCCTTTTCAGAGCGTACCTGTGCCACT GTCACCATTGGAGGAATAAACATTGCTGAGGCTCTTGTCAGC AAAGGTCTAGCCACAGTGATCAGATACCGGCAGGATGATGAC CAGAGATCATCACACTACGATGAACTGCTTGCTGCAGAGGCC AGAGCTATTAAGAATGGCAAAGGATTGCATAGCAAGAAGGAA GTGCCTATCCACCGTGTTGCAGATATATCTGGGGATACCCAA AAAGCAAAGCAGTTCCTGCCTTTTCTTCAGCGGGCAGGTCGT TCTGAAGCTGTGGTGGAATACGTCTTCAGTGGTTCTCGTCTC AAACTCTATTTGCCAAAGGAAACTTGCCTTATCACCTTCTTG CTTGCAGGCATTGAATGCCCCAGAGGAGCCCGAAACCTCCCA GGCTTGGTGCAGGAAGGAGAGCCCTTCAGCGAGGAAGCTACA CTTTTCACCAAGGAACTGGTGCTGCAGCGAGAGGTGGAGGTG GAGGTGGAGAGCATGGACAAGGCCGGCAACTTTATCGGCTGG CTGCACATCGACGGTGCCAACCTGTCCGTCCTGCTGGTGGAG CACGCGCTCTCCAAGGTCCACTTCACCGCCGAACGCAGCTCC TACTACAAGTCCCTGCTGTCTGCCGAGGAGGCCGCAAAGCAG AAGAAAGAGAAGGTCTGGGCCCACTATGAGGAGCAGCCCGTG GAGGAGGTGATGCCAGTGCTGGAGGAGAAGGAGCGATCTGCT AGCTACAAGCCCGTGTTTGTG
Protein Sequence	AKVMQVLNADAIVVKLNSGDYKTIHLSSIRPPRLEGENTQDK NKKLRPLYDIPYMFEAREFLRKKLIGKKVNVTVDYIRPASPA TETVPAFSERTCATVTIGGINIAEALVSKGLATVIRYRQDDD QRSSHYDELLAAEARAIKNGKGLHSKKEVPIHRVADISGDTQ KAKQFLPFLQRAGRSEAVVEYVFSGSRLKLYLPKETCLITFL LAGIECPRGARNLPGLVQEGEPFSEEATLFTKELVLQREVEV EVESMDKAGNFIGWLHIDGANLSVLLVEHALSKVHFTAERSS YYKSLLSAEEAAKQKKEKVWAHYEEQPVEEVMPVLEEKERSA SYKPVFV
Note 1:	The NcoI and EcoRI restriction sites on the forward and reverse primer respectively, are marked in bold
Note 2:	The p100HS34 DNA and amino sequences are listed in Appendix B2

Appendix B3 – Construct information for p100DM3

	p100DM3	
Construct information	Protein name	p100 (also Tudor-SN, SND1)
	Organism of origin	*Drosophila melanogaster.*
	Part(s) encompassed	Staphylococcal nuclease domain 3
	Residue numbers	345-504 (160 residues)
	MW	17.9 kDa
	Extinction coefficient	10.4 mM^{-1}cm^{-1}
Plasmid Information	Plasmid name	pETZZ-1a
	Antibiotic	Kanamycin
	Tag(s)	N-terminal His-tag followed by double Z-tag (ZZ-tag)
	Enzyme for tag cleavage	TEV protease
	Restriction enzymes used to clone	NcoI and EcoRI
Total number of residues		160 + 4 (GAMG from TEV cleavage site) = 164
Primer Sequences		F: GAG**CCCATGG**GGGAGAAAGACTTCTCTGGAACAGTG
		R: GAGC**GAATTC**TTATTATGCGTGCAGTCCCTTTAG
DNA Sequence		GAGAAAGACTTCTCTGGAACAGTGGTGGAAGTCTTCAATGGC GATGCTATTAATGTACGTCTTTCCAATGGACAAGTAAAGAAG GTTTTCTTCTCGTCTATCAGGCCGCCGCGTGACCAACGCGCG GTTGTTGGCACCGATGGCGAGGAAATCGTCAAGGCACCACCA CGCGGAAAAAACTACAGACCACTTTACGAAATTCCACACATG TTCGATGCCCGTGAGTTTCTGCGTAAAAAACTAATTAATAAA AAAGTTCAATGCAACCTGGACTACATCTCACCGCCGCGCGAA AACTTCCCTGAGAAATACTGCTATACCGTCTCTATTGGCGGT CAGAACGTGGCTGAGGCAATGGTAGCCAAAGGTTTGGCAACA TGTGTTCGCTATCGCCAGGACGACGATCAGAGATCTTCCGCC TATGATCAATTGATTGCCGCCGAACAGCAAGCCATCAAGGGA CTAAAGGGACTGCACGCA
Protein Sequence		KDFSGTVVEVFNGDAINVRLSNGQVKKVFFSSIRPPRDQRAV VGTDGEEIVKAPPRGKNYRPLYEIPHMFDAREFLRKKLINKK VQCNLDYISPPRENFPEKYCYTVSIGGQNVAEAMVAKGLATC VRYRQDDDQRSSAYDQLIAAEQQAIKGLKGLHA
Note 1:		The NcoI and EcoRI restriction sites on the forward and reverse primer respectively, are marked in bold.
Note 2:		The p100DM3 DNA and amino sequences are listed in Appendix B3

Appendix B4 – Construct information for p100DM4

p100DM4		
Construct information	Protein name	p100 (also Tudor-SN, SND1)
	Organism of origin	*Drosophila melanogaster.*
	Part(s) encompassed	Staphylococcal nuclease domain 4
	Residue numbers	526-687 (162 residues)
	MW	18.0 kDa
	Extinction coefficient	27.9 mM^{-1}cm^{-1}
Plasmid Information	Plasmid name	pETZZ-1a
	Antibiotic	Kanamycin
	Tag(s)	N-terminal His-tag followed by double Z-tag (ZZ-tag)
	Enzyme for tag cleavage	TEV protease
	Restriction enzymes used to clone	NcoI and EcoRI
Total number of residues		162 + 4 (GAMG from TEV cleavage site) = 166
Primer Sequences		F: GAGC**CCATGG**GGCAATACTTGCCATCGTGGCA
		R: GAGC**GAATTC**TTATTATCACTCGGTGACAGTTTTTTCT
DNA Sequence		CAATACTTGCCATCGTGGCAGCGTGCTTTGCGTACTGAAGCC ATTGTAGAGTTTGTCGCCAGCGGATCCCGCCTCCGAATCTTT GTACCGAAGGACAGTTGCCTGGTAACATTCCTGCTTGCTGGG ATTTCATGTCCGCGTTCCTCTCGTCCAGCGCTTAATGGCGTT CCTGCCCAAGAGGGGGAACCATTTGGCGACGAGGCTTTGACC TTTACCAGGGAGCGTGTCTTGCAACGCGACGTTTCTGTTCAT ATCGATACCACCGACAAGGCAGGATCTTCAGTGATTGGTTGG CTCTGGACGGACAGCGGCGCCAACTTGTCAGTTGCTCTTGTC GAAGAGGGTCTTGCTGAAGTGCACTTTAGTGCTGAAAAATCT GAATACTACCGCCAGTTGAAGATTGCTGAAGACCGTGCCAAG GCCGCCAAAAAGAACATCTGGACAAACTACGTAGAAGAAGTG CCTAAAGAAAAACTGTCACCGAGTGA
Protein Sequence		QYLPSWQRALRTEAIVEFVASGSRLRIFVPKDSCLVTFLLAG ISCPRSSRPALNGVPAQEGEPFGDEALTFTRERVLQRDVSVH IDTTDKAGSSVIGWLWTDSGANLSVALVEEGLAEVHFSAEKS EYYRQLKIAEDRAKAAKKNIWTNYVEEVPKEKTVTE
Note 1:		The NcoI and EcoRI restriction sites on the forward and reverse primer respectively are marked in bold
Note 2:		The p100DM4 DNA and amino sequences are listed in Appendix B4

Appendix C – Primer sequence for p100DM34 mutants

	Oligo name	Sequence (5' ->3')	T_m [°C]
R379E	R379E_sense	AAGGTTTTCTTCTCGTCTATCGAGCCGCCGCGTGAC	74.0
	R379E_antisense	GTCACGCGGCGGCTCGATAGACGAGAAGAAAACCTT	74.0
K447E	K447E_ sense	GCGCGAAAACTTCCCTGAGGAGTACTGCTATACCGTCTCTA	74.4
	K447E_antisense	TAGAGACGGTATAGCAGTACTCCTCAGGGAAGTTTTCGCGC	74.4
S548A	S548A_ sense	TTTGTCGCCAGCGGAGCCCGCCTCC	72.8
	S548A_antisense	GGAGGCGGGCTCCGCTGGCGACAAA	72.8
R572E	R572E_ sense	CTGGGATTTCATGTCCGGAGTCCTCTCGTCCAGCGCT	> 75
	R572E_antisense	AGCGCTGGACGAGAGAGGACTCCGGACATGAAATCCCAG	> 75

Acknowledgements

What can I say? It seems that just yesterday the flight that brought us from Toronto has touched down on the Munich airport runway, and here I am, coming up with a list of people I should thank, after my almost six-year stay here. How time flies…

First of all, of course, my supervisor, Dr. Michael Sattler. Thank you so much for giving me the opportunity to come here, work on all the projects you gave me, for your constant help, support and understanding, and last but not least, for your patience. I know that things were moving much slower than we anticipated, and yet you still stuck with me till the end. Also thanks for your support during my (sometimes trying) post-doc search and of course, for the reference letters.

A great deal of gratitude goes to my Bachelor's and Masters' supervisor, Dr. Philip Johnson, for introducing me to scientific research in general and NMR of RNA in particular, and giving me the initial skills and knowledge in this field, which have greatly benefitted me in my work here.

I would also like to thank Dr. Dierk Niessing, and Dr. Bernd Reif for kindly agreeing to be part of my doctoral examining committee. I would also like to thank Dr. Niessing and Dr. Klaus Klaus Förstemann useful input during our thesis committee meetings.

Throughout my stay in the group, I have also had the opportunity to work with some really bright and talented thesis and practical students. Christina Siedler, Alexei Turochkin, Annica Flemming and Simone Frank, Martin Rübbelke – Thank you for the work that you did, for the effort you put into it, and also for the opportunity to meet and get to know you. I wish you all the best in your future endeavours, whatever they may be.

I would like to thank all my collaborators on our projects. Steffan Schiesser, Felix Gnerlich and their supervisor, Prof. Thomas Carell – a great deal of thanks for the synthesis of the inosine-labelled RNA sample without which, this project could not have moved forward as far as it did. I also would like to thank our collaborators on the TS project in Italy, for experiments which helped us to publish our work. A special thanks goes out to the collaborator on the Zebra/Jun-Fos work, Montse Gustems for not giving up on this challenging project, and her supervisor, Dr. Wolfgang Hammerschmidt.

The group is only as good as its members and I have to say that we had some really great ones. I therefore would like to thank all the past and present members of the Sattler, Reif, Madl,

Lange and Dames groups for all their help and support in the lab and on the spectrometer, for a good and collaborative working atmosphere and finally for some memorable group retreats!

Anders and Andre, thank you so much for welcoming me to the lab and making me feel comfortable from day one. My office mates – Kostas, thank you for your help at the spectrometer, protein assignment and much more, Yun – All the fun times we had at our first Oktoberfest, the group retreats I will not forget.

Malgosia, it was great talking to you about anything, you understood all my frustrations that came with working on RNA. Andre Dallmann – Thanks a lot for working together on the IRD RNA project, all your help with the NMR experiments, and of course the great times we had over at your place with Tina and the kids. I wish you all the best in your postdoc in London and, of course, your future career. Uli – Thank you so much for introducing to the ENB crowd, we had some really good times in all the meetings and retreats! Divita – Thanks for working together on the TS project, no matter how frustrating it was, we managed to pull through. Giambattista, thanks for a lot of great discussions about Italy, Germany and pretty much everything about world culture and history. Lee, it was great to meet you and Eva, I really enjoyed our get-togethers, and wish you all the best in finding a position as soon as possible.

Irina, thank you so much for our lunches and coffee breaks together, our travels both in and out of the lab (especially our Japan trip – BeribiskySAN rocks!), for introducing me to so many great people in Munich, and above all – thank you for your deep friendship! I wish you all the best in your new beginnings, may they always be successful and fruitful, and most importantly make you happy.

Hamed, who would have thought that coming with me to Garching on my second day to help me register, would be first the first step in a journey, which would lead us to the Pizzeria Roma, Mount Fuji, our kitchen, the Cliffs of Moher and the Aran Islands, the plotter room and our almost simultaneous interview in the U.S.? And this is just to name a few… It is really symbolic that we will leave around the same time, because I think one of us cannot exist in this lab without the other. Thank you for all your patience, advice and above all your friendship, I have enjoyed every bit of it (even though sometimes I may say otherwise).

Elke, as with Hamed, these were some truly great times. I will miss our (really brief) coffee breaks, our visits to Eggolsheim, our trip to Ireland… But above all, I will miss our conversations,

your wit, kindness, honesty and care. Thank you for everything. I think a small piece of me will always remain in the plotter room and the kitchen where you, me and Hamed would "sit together".

Outside of the lab, I have also been blessed with some really great friends, both in Munich and elsewhere. Everyone – Thank you so much for some memorable times throughout these past six years, I am sure that there are more to come.

My parents – Thank you so much for your unconditional love and support. At some trying stages of my studies, both in Canada and here, you have never stopped believing in me. By living your life the way you do, you have set a very high example, which I will try very hard to follow. Natashka, my amazing little sister, thank you as well for your love and support all throughout. You have some really great things coming in the future, and I am very happy for you. I also owe a great debt of gratitude to my parents' in-law who also stood by me throughout all my studies.

Masha, my wife, and touchstone. You have stuck with me when I was a student for ages in Toronto. You have stuck with me when I dragged you across the Atlantic to come to Germany, for my studies here. Thank you Mashenka, for everything you do for me every single day for the last eleven years. Without you, all this would not have been possible.

Curriculum Vitae

ALEXANDER V. BERIBISKY

Technical University of Munich
Biomolecular NMR Spectroscopy
Department of Chemistry, Lichtenbergstr. 4
D-85747 Garching, Germany
E-mail: aberibisky@gmail.com

DATE AND PLACE OF BIRTH

07.04.1982, Kharkov, Ukraine

EDUCATION

2008 – Present: **Department of Chemistry, Technical University of Munich, Germany.**
Doctoral studies with Prof. Dr. Michael Sattler.

2005 – 2008: **Department of Chemistry, York University, Toronto, Canada.**
MSc. with Dr. Philip E. Johnson.
Thesis title: Determination of the three dimensional structure of the Selenocysteine Insertion Sequence RNA using NMR spectroscopy.

2000 – 2005: **Department of Chemistry, York University, Toronto, Canada.**
Hon BSc. in Biochemistry.
Thesis title: Determination of the three dimensional structure of the Selenocysteine Insertion Sequence RNA using NMR spectroscopy.
Undergraduate Supervisor: Philip E. Johnson.

PROFESSIONAL EXPERIENCE

2008 – Present: **Department of Chemistry, Technical University of Munich, Germany.**
Teaching Assistant – Second year level organic chemistry, first year level (Master's program) Biomolecular NMR Spectroscopy. Duties involve: Teaching, conducting tests, and marking of reports.

2006 – 2007: **Department of Chemistry, York University, Toronto Canada.**
Teaching Assistant – First year level general chemistry, second and third year level biochemistry. Duties involve: Teaching in undergraduate laboratories, invigilating, marking tests and exams.

2002 – 2008: **Qualified Tutors, Thornhill, Canada.**
Science Tutor – Tutoring of high school and university science students.

PUBLICATIONS IN JOURNALS

Beribisky A.V.[1]., Dallmann A[1]., Gnerlich F., Schiesser S., Carell T., Sattler M. Selective isotope-labeling of inosine phosphoramidites facilitates NMR analysis of inosine-containing RNA. *In preparation.*

Garg D, **Beribisky A.V.,** Ponterini G., Ligabue A., Marverti G., Martello A., Costi M.P., Sattler M, Wade R.C. "Translational repression of thymidylate synthase by targeting its mRNA." *Nucleic Acids Research.* (2013) Vol. **41(7):** 4159-70.

Tavares T.J., **Beribisky A.V.,** Johnson P.E. "Structure of the cytosine-cytosine mismatch in the thymidylate synthase mRNA binding site and analysis of its interaction with the aminoglycoside paromomycin." *RNA.* (2009) Vol. **15(5):** 911-922.

Beribisky A.V.[2], Tavares T.J.[2], Amborski A.N., Motamed M., Johnson A. E., Mark T.L. Johnson P.E. The three-dimensional structure of the *Moorella thermoacetica* selenocysteine insertion sequence RNA hairpin and its interaction with the elongation factor SelB. *RNA.* (2007) Vol. **13 (11)**: 1948–1956.

1,2 – co-first authors

POSTERS AT CONFERENCES

EUROMAR 2012, Dublin, Ireland – July 2012, Sixteenth Annual meeting of the RNA Society, Kyoto, Japan – June 2011.

Beribisky A., Dallmann A., Tripsianes, K., Sattler M. Structural investigation of inosine-edited RNA duplex and its interaction with RNA Induced Silencing complex component p100.

The modern RNA world conference, Murnau, Germany – October 2010

Garg D., Beribisky A., Wade R., Sattler M., Structural investigation of the interaction between CC mismatch in the thymidylate synthase mRNA and Hoechst 33258

Ribo-Club Opening Session, Magog, Quebec, Canada – September 2005, September, 2006.

Beribisky A., Motamed M., Zhao S., Tavares T.J. and Johnson P.E. Progress in Determination of three-dimensional structure of Selenocysteine Insertion Sequence RNA from *Moorella thermoacetica* using NMR spectroscopy.