

Technische Universität München

Lehrstuhl für Biomolekulare NMR-Spektroskopie

Department Chemie

Integrated structural biology of protein complexes in miRNA and peroxisome biogenesis

Hamed Kooshapur

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. Stephan A. Sieber
Prüfer der Dissertation: 1. Univ.-Prof. Dr. Michael Sattler
2. Univ.-Prof. Dr. Klaus Förstermann
Ludwig-Maximilians-Universität München
3. Priv.-Doz. Dr. Dierk Niessing

Die Dissertation wurde am 23.05.2014 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 30.06.2014 angenommen.

Declaration

I hereby declare that parts of this thesis have already been published in scientific journals:

Neuhaus A*, **Kooshapur H***, Wolf J, Meyer NH, Madl T, Saidowsky J, Hambruch E, Lazam A, Jung M, Sattler M, Schliebs W, Erdmann R.

“A novel Pex14 protein-interacting site of human Pex5 is critical for matrix protein import into peroxisomes.”

(2014) **J Biol Chem.** 289(1):437-48 **co-first author*

Dallmann A, Simon B, Duszczuk MM, **Kooshapur H**, Pardi A, Bermel W, Sattler M.

“Efficient detection of hydrogen bonds in dynamic regions of RNA by sensitivity-optimized NMR pulse sequences.”

(2013) **Angew Chem Int Ed Engl.** 52(40), 10487-90

Tavanez JP, Madl T, **Kooshapur H**, Sattler M, Valcárcel J.

“hnRNP A1 proofreads 3' splice site recognition by U2AF.”

(2012) **Molecular Cell.** 45(3), 314-29.

Neufeld C, Filipp FV, Simon B, Neuhaus A, Schüller N, David C, **Kooshapur H**, Madl T, Erdmann R, Schliebs W, Wilmanns M, Sattler M.

“Structural basis for competitive interactions of Pex14 with the import receptors Pex5 and Pex19.”

(2009) **EMBO Journal.** 28(6), 745-54.

Table of contents

Table of contents	3
Summary	7
Zusammenfassung	9
1. Biomolecular NMR	11
1.1 Protein fingerprint spectrum.....	12
1.2 Binding studies by NMR	13
1.3 Dynamic studies by NMR	15
1.4 Structure determination by NMR	16
Part 1	18
Structural basis of regulation of miRNA processing by hnRNP A1	18
2. Introduction	19
2.1 miRNAs.....	19
2.2 miRNA biogenesis	19
2.3 Post-transcriptional regulation of miRNA biogenesis	21
2.3.1 Terminal loop-mediated regulation of miRNA biogenesis.....	22
2.4 The miR-17-92 cluster	24
2.5 hnRNP proteins	26
2.5.1 hnRNP A1.....	27
2.6 Aims and scope of the project	32
3. Materials and methods	33
3.1 Materials	33
3.1.1 Chemicals.....	33
3.1.2 Enzymes.....	33
3.1.3 Bacterial strains and plasmids.....	33
3.1.4 Bacterial growth media and antibiotics.....	34
3.1.5 Buffers	36

3.1.6	Oligonucleotides	37
3.2	Methods	38
3.2.1	Molecular biology	38
3.2.2	Biochemistry	41
3.2.3	NMR spectroscopy	44
3.2.4	Structural biology and biophysics	46
4.	Results	49
4.1	NMR characterization of hnRNP A1	49
4.1.1	Constructs of hnRNP A1	49
4.1.2	NMR analysis of RRM1, RRM2 and UP1	50
4.1.3	NMR relaxation analysis of hnRNP A1 constructs	51
4.2	NMR characterization of pri-miR-18a	54
4.2.1	Constructs of pri-miR-18a	54
4.2.2	Loop 18a RNA	55
4.2.3	Stem 18a RNA	57
4.3	NMR titration experiments	60
4.4	ITC analysis of hnRNP A1–RNA interactions	66
4.5	EMSA and footprint analysis of hnRNP A1-pri-miR-18a interactions	69
4.6	X-ray crystallography	72
4.6.1	X-ray crystallography on the UP1-17-mer complex	72
4.6.2	Crystal structure of the UP1-12-mer complex	75
4.7	Static light scattering (SLS)	82
4.8	The effect of UP1 salt-bridge mutations on RNA binding	84
4.9	Small angle X-ray scattering (SAXS)	85
4.9.1	SAXS studies on UP1	86
4.9.2	SAXS studies on pri-miR-18a	88
4.9.3	SAXS studies on the UP1-18a complex	90
4.10	Functional data	91
5.	Discussion	95
5.1	Solution conformation of UP1	95
5.2	Solution conformation of pri-miR-18a	97
5.3	Recognition of pri-miR-18a by hnRNP A1	98
5.4	Structural basis of interaction of UP1 with the terminal loop of pri-miR-18a	99

5.5	Regulation of miRNA processing: the importance of terminal loops	101
5.6	Conclusions and outlook.....	104
6.	hnRNP A1 in 3' splice site selection.....	105
6.1	Abstract.....	105
6.2	Introduction	105
6.3	Results and discussion	108
Part 2	112
	Structure and dynamics in peroxisomal import	112
7.	Biological background	113
7.1	Peroxisomes	113
7.2	Peroxisomal disorders	113
7.3	Peroxisomal import.....	113
7.4	Aims and scope of the project	115
8.	A novel Pex14 interacting site in human Pex5	116
8.1	Abstract.....	117
8.2	Introduction	117
8.3	Materials and methods	120
8.3.1	Construction of Pex5-fragments expression vectors.....	120
8.3.2	Expression and purification of recombinant proteins	121
8.3.3	Peptide binding.....	121
8.3.4	NMR spectroscopy	122
8.3.5	Structure determination	122
8.3.6	Pex5–Pex14 interaction assays	123
8.3.7	<i>In vivo</i> studies.....	123
8.4	Results.....	124
8.4.1	Identification of a novel Pex14-binding site in human Pex5.....	124
8.4.2	NMR analysis and structure of the Pex14-NTD-Pex5 LVxEF motif complex	127
8.4.3	Equilibrium and kinetic binding properties of Pex14-NTD:Pex5 complexes	132
8.4.4	Functional analysis of the novel Pex14 binding site.....	136
8.5	Discussion.....	139

9. Pex14-Tubulin interactions	143
10. References	146
11. Appendix	166
11.1 Primer sequences	166
11.2 Protein sequences and properties	169
Abbreviations	172
Acknowledgements.....	175
Curriculum Vitae.....	177

Summary

Understanding the interactions between macromolecules is essential for the understanding of biological systems. This thesis describes the characterization of protein-RNA and protein-protein interactions involved in two important biological processes: microRNA biogenesis (discussed in part 1) and peroxisomal import (discussed in part 2) using a multi-disciplinary approach combining different structural biology techniques and biochemical, biophysical and functional assays.

Chapter 1 is an introduction to practical aspects of biomolecular nuclear magnetic resonance (NMR) focusing on the methods applied in this study.

The first part of this thesis, including chapters 2, 3, 4 and 5, describes the structural and biochemical characterization of the interaction between hnRNP A1 and the primary microRNA 18a (pri-miR-18a). The molecular basis of pri-miR-18a recognition by hnRNP A1 has been investigated using an integrated structural biology approach combining NMR, X-ray crystallography and small angle X-ray scattering (SAXS). These analyses reveal how hnRNP A1 specifically binds to the terminal loop of pri-miR-18a. Furthermore, the affinity and stoichiometry of this interaction have been determined by isothermal titration calorimetry (ITC) and light scattering, respectively. Our data are further complemented by mutational analysis and *in vivo* functional assays in collaboration with the group of Dr. Javier Caceres (MRC, Edinburgh). Our results provide important insight into post-transcriptional regulation of microRNA biogenesis. This study also highlights the importance of using complementary structural biology techniques to provide a more precise description of macromolecules and their complexes.

Chapter 5 of this thesis reports a new role for hnRNP A1 in splicing regulation. Using NMR spectroscopy we show that hnRNP A1 forms a ternary complex with the splicing factor U2AF and AG-containing uridine-rich RNAs, while it displaces U2AF from CG-containing uridine-rich RNAs. These results, combined with data from the group of Dr. Juan Valcárcel (CRG, Barcelona) provide novel insight into the proofreading mechanism at the 3' splice site.

The second part of this thesis discusses protein-protein interactions involved in peroxisomal import and motility. Chapter 8 presents a novel protein-binding motif in the peroxisomal import receptor Pex5 which is responsible for interaction with Pex14, a key peroxisomal membrane protein. The NMR-derived solution structure of a conserved domain of Pex14 in complex with a Pex5 peptide reveals the molecular details of this interaction. Complementary biophysical data and cell-based experiments provided by the group of Dr. Ralf Erdmann (Rüch-Universität Bochum) indicate that this novel Pex14-Pex5 interaction has unique features that are essential for peroxisomal import. In addition, chapter 9 summarizes the NMR characterization of a newly identified interaction between Pex14 and tubulin which is implicated in peroxisome motility.

Zusammenfassung

Das Verständnis von Wechselwirkungen zwischen Makromolekülen ist essentiell wichtig für das Verständnis biologischer Systeme. Diese Arbeit beschreibt die Charakterisierung von Protein-RNA- und Protein-Protein-Interaktionen, welche an zwei wichtigen biologischen Prozessen beteiligt sind: microRNA-Biogenese (behandelt in Teil 1) und peroxisomalem Import (behandelt in Teil 2). Hierzu wird ein multidisziplinärer Ansatz, bestehend aus einer Kombination verschiedener Methoden der Strukturbiologie und biochemischen, biophysikalischen und funktionellen Assays herangezogen. Kapitel 1 ist eine Einführung in die praktischen Aspekte der biomolekularen Kernspinresonanz (NMR), und beleuchtet die in dieser Studie angewandten Methoden.

Der erste Teil dieser Arbeit, einschließlich der Kapitel 2, 3, 4 und 5, beschreibt die strukturelle und biochemische Charakterisierung der Wechselwirkung zwischen hnRNP A1 und der Primär 18a microRNA (pri-miR-18a). Die molekularen Grundlagen der pri-miR-18a Bindung an hnRNP A1 wurden mit Hilfe eines integrativen, strukturbiologischen Ansatzes einschließlich NMR, Röntgenstrukturanalyse und Kleinwinkel-Röntgenstreuung (SAXS) untersucht. Diese Analysen zeigen, wie hnRNP A1 spezifisch an die terminale Schleife der pri-miR-18a bindet. Darüber hinaus konnten die Affinität und Stöchiometrie der Wechselwirkung durch isotherme Titrationskalorimetrie (ITC) beziehungsweise Lichtstreuung bestimmt werden. Unsere Daten werden vervollständigt durch Mutationsanalysen und funktionelle *in-vivo*-Tests in Zusammenarbeit mit der Arbeitsgruppe von Dr. Javier Caceres (MRC, Edinburgh). Unsere Ergebnisse liefern wichtige Einblicke in post-transkriptionelle Regulierung der microRNA-Biogenese. Diese Studie hebt zudem die Bedeutung der Verwendung komplementärer, strukturbiologischer Techniken hervor, um eine präzise Beschreibung von Makromolekülen und ihrer Komplexe zu erhalten.

Kapitel 5 dieser Arbeit beschreibt eine neue Rolle von hnRNP A1 in der Spleiß-Regulierung. Mit Hilfe von NMR-Spektroskopie konnte gezeigt werden, dass hnRNP A1 einen ternären Komplex mit dem Spleißfaktor U2AF und AG-haltigen, Uridinreichen RNAs bildet, während es CG-haltige, Uridin-reiche RNAs verdrängt. Diese

Ergebnisse, kombiniert mit Daten der Arbeitsgruppe von Dr. Juan Valcárcel (CRG, Barcelona), liefern neue Einblicke in den Fehlerkorrektur-Mechanismus an der 3'-Spleißstelle.

Der zweite Teil dieser Arbeit behandelt Protein-Protein-Interaktionen, welche am peroxisomalen Import und Motilität beteiligt sind. Kapitel 8 stellt ein neuartiges Protein-Bindungsmotiv des peroxisomalen Importrezeptors Pex5 vor, welches für die Interaktion mit Pex14, ein Hauptmembranprotein des Peroxisoms, verantwortlich ist. Die durch Lösungs-NMR aufgeklärte Struktur einer konservierten Domäne von Pex14 im Komplex mit einem Pex5-Peptid zeigt die molekularen Details dieser Wechselwirkung auf. Ergänzende biophysikalische Daten und zellbasierte Experimente der Arbeitsgruppe von Dr. Ralf Erdmann (Ruhr-Universität Bochum) deuten darauf hin, dass diese neuartige Pex14-Pex5 Interaktion einzigartige Eigenschaften enthält, welche essentiell für den peroxisomalen Import sind. Zusätzlich wird in Kapitel 9 die durch NMR charakterisierte, neu identifizierte Interaktion zwischen Pex14 und Tubulin zusammengefasst, welche an peroxisomaler Motilität beteiligt ist.

1. Biomolecular NMR

Structural biology aims at determining the structures of macromolecules, such as proteins and nucleic acids, and their complexes in order to provide a mechanistic understanding of biological systems.

The two main techniques for obtaining atomic resolution structures of molecules are X-ray crystallography and nuclear magnetic resonance (NMR). Each of these methods has its own strengths and limitations. X-ray crystallography is the most powerful technique for providing high-resolution structures of molecules. However, a major bottleneck in crystallography is obtaining diffraction-quality crystals and this can be challenging or in many cases impossible. An additional challenge for crystallizing macromolecular complexes is that in some cases only one of the components forms crystals. In chapter 4 of this thesis, an example is shown where attempts to crystallize a protein-RNA complex resulted in RNA only crystals. Moreover, structures determined by X-ray crystallography can be influenced by crystallization conditions and crystal-packing forces. Therefore it is important to validate the structures by solution methods such as NMR and SAXS.

NMR spectroscopy is a unique method which allows characterization of the structure and dynamics of macromolecules and their complexes at atomic- and residue-resolution. However, NMR studies are limited by the size of the molecule. Therefore, it is necessary to combine different structural biology methods in an integrated structural biology approach complemented with functional data in order to provide a more precise understanding of biological systems. Chapter 4 reports an integrated structural biology approach applied to an important protein-RNA complex.

In this chapter, I will briefly introduce the biomolecular NMR methods which have been used in this work.

1.1 Protein fingerprint spectrum

One of the most important heteronuclear experiments in biomolecular NMR is the two-dimensional (2D) ^1H , ^{15}N hetero single quantum correlation (HSQC) which correlates ^1H to its attached ^{15}N . It requires ^{15}N -isotope labeling of the protein which can easily be achieved by expressing the protein in cells grown in ^{15}N -labeling media (see chapter 3).

In the ^1H , ^{15}N HSQC spectrum, each amino acid in the protein (with the exception of Proline) gives rise to one peak corresponding to the backbone amide group. Additional peaks from side-chains of residues containing N-H groups, such as Asparagine and Glutamine, are also visible. The 2D ^1H , ^{15}N HSQC is also known as the “fingerprint spectrum” of the protein and can provide important information regarding the folding state and quality of the protein. In well-folded proteins, each amide group has a distinct chemical environment and therefore the HSQC is usually well-dispersed; whereas in unfolded proteins the amides peaks are clustered in the center (around 8 ppm in the ^1H dimension) of the spectrum. This is illustrated in Figure 1.1.

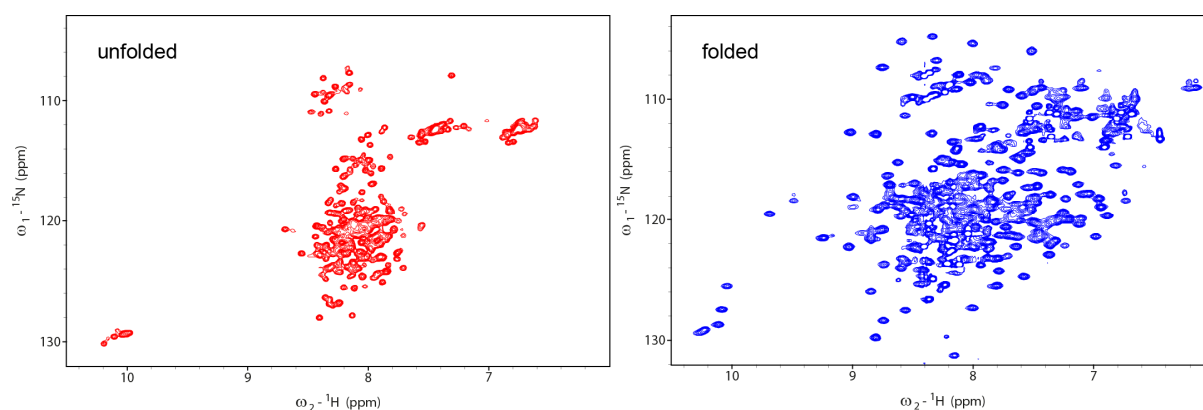


Figure 1.1 Comparison of 2D ^1H , ^{15}N HSQC spectrum of an intrinsically disordered protein (red) and a well folded protein (blue). The unfolded and folded proteins correspond to the N-terminal and C-terminal halves of human Pex5, respectively (see chapter 8).

1.2 Binding studies by NMR

One of the most widely used application of NMR is characterization of molecular interactions. NMR binding studies are based on the fact that the chemical environment of atoms changes upon ligand binding, causing chemical shift perturbations (CSP) which can be monitored by NMR. Because chemical shifts are highly sensitive to even the slightest changes, this methodology is very powerful in detecting weak and transient interactions which might not be detectable by other methods such as gel-shift assays or size-exclusion chromatography. A case is presented in chapter 4, where no binding of a protein to RNA was observed in gel-shift assays while NMR titrations could clearly show the interaction.

The most commonly used approach for studying interactions by NMR is to monitor the chemical shift changes by recording a series of 2D ^1H , ^{15}N HSQC spectra of a ^{15}N -labeled protein at increasing concentrations of an unlabeled ligand. The peaks which are affected by binding report on residues at the binding interface and if the chemical shift assignments are available, this allows mapping of the protein binding interface. In addition, NMR titration experiments can provide information about the affinity and kinetics of the interaction as outlined below.

In general, three exchange regimes are described in NMR binding experiments: fast, intermediate and slow. This is a measure of how the exchange rate (k_{ex}) between the free and bound states of the protein compares with the chemical shift difference between the two states ($\Delta\nu$). For example, if the exchange rate is faster than the chemical shift difference ($k_{\text{ex}} \gg \Delta\nu$) binding is said to be in the fast-exchange on the NMR chemical shift timescale. In the fast exchange regime, the NMR signal is an average of the two states with a population-weighted average chemical shift that changes upon addition of ligand. In intermediate exchange (also known as coalescence) the signal is broadened or disappears due to exchange, but can reappear when saturation is reached. In contrast, in slow exchange two distinct signals are observed corresponding to the free and bound states and the intensity of each signal is determined by the population of that species (Figure 1.2). In fast-exchange binding the NMR signals can be followed during the course of titration; whereas in intermediate and slow exchange the resonances of the bound state have

to be reassigned. It is important to note that the exchange regime can also be shifted by changing the temperature and magnetic field.

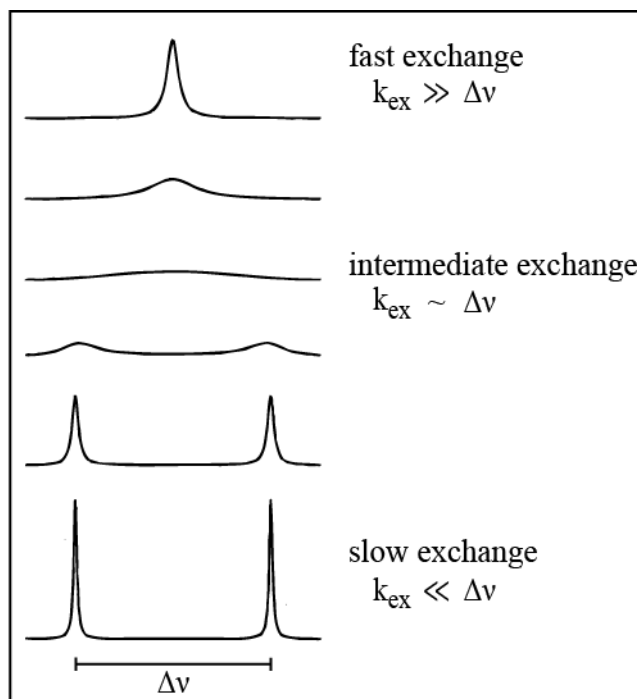


Figure 1.2 Effect of chemical exchange on NMR signals. The exchange regime is also dependent on the temperature and magnetic field. (Adapted from Cavanagh, 2007)

The exchange regime of binding is related to the affinity of the interaction. Fast exchange is characteristic of weak binding (micromolar and higher K_D values), while slow exchange normally represent high affinity interactions (K_D in the nanomolar range). Therefore, the appearance of signals in the NMR titration experiment alone can provide important information regarding the exchange and strength of the interaction. A better estimate of the affinity can be obtained from fitting the titration curve derived from the chemical shift perturbations, especially in the case of fast-exchange binding. The method for extracting dissociation constants from titration analysis is described in section 3.2.3.3, and in addition chapter 9 provides two examples of binding in the fast exchange regime for which this method has been applied.

1.3 Dynamic studies by NMR

The three-dimensional structures of macromolecules are only ‘snapshots’ of dynamic processes in biological systems. NMR is a powerful technique for studying the dynamics of macromolecules at the atomic and residue-specific level. Various NMR experiments have been developed for characterizing the dynamics at different time-scales (reviewed in Kleckner & Foster, 2011).

A commonly used approach for characterizing macromolecular dynamics relies on the spin relaxation phenomenon which is the return of nuclear magnetization to the equilibrium state after perturbation. The magnetization is restored by two processes: T_1 relaxation (longitudinal or spin-lattice relaxation) is the recovery of the magnetization along the static magnetic field and T_2 relaxation (transverse or spin-spin relaxation) which is the loss of phase coherence of spin components in the transverse plane. The time constants T_1 and T_2 (or the relaxation rates $R_1 = 1/T_1$ and $R_2 = 1/T_2$) provide information about the dynamics and are related to the rotational correlation time (τ_c) (Figure 1.3).

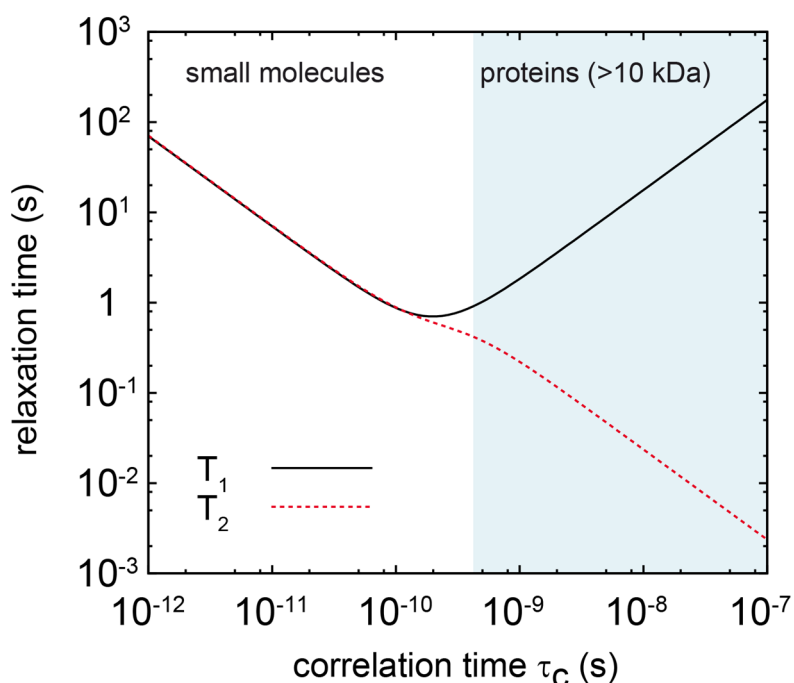


Figure 1.3 Plot of T_1 and T_2 relaxation times as a function of correlation time. The correlation time increases from left to right with increase in molecular size.

The rotational correlation time (τ_c) describes the Brownian motion in solution. For tumbling of a spherical molecule in solution, it is given by the Stoke's law:

$$\tau_c = \frac{4\pi\eta R_H^3}{3k_B T} = \frac{V\eta}{k_B T}$$

where η is the solvent viscosity, R_H is the hydrodynamic radius, k_B is the Boltzmann constant, T is the temperature and V is the volume. The correlation time is proportional to the molecular weight and as a rule of thumb, the rotational correlation time (in nanoseconds) of a protein in aqueous solution at room temperature is about half of its molecular weight in kDa. The correlation time (τ_c) can be approximated from the ^{15}N R_2/R_1 ratio using the equation:

$$\tau_c \cong \frac{1}{2\pi\nu_N} \sqrt{1.5 \times \left(\frac{R_2}{R_1} - \frac{7}{6} \right)}$$

where ν_N is the ^{15}N resonance frequency in Hz.

A rough estimate of the correlation time of a protein can also be obtained using the following equation (Daragan & Mayo, 1997):

$$\tau_c = \left(9.18 \times \frac{10^{-3}}{T} \right) \exp\left(\frac{2416}{T} \right) N^{0.93}$$

where T is the temperature in Kelvin and N is the number of amino acids.

The R_1 and R_2 information can be complemented with data obtained from the $\{^1\text{H}\}$ - ^{15}N heteronuclear NOE experiment which can provide information regarding fast motion of individual N-H bond vectors. This experiment is especially useful for identifying flexible regions of proteins, such as the N- and C- termini or flexible loops.

1.4 Structure determination by NMR

NMR can provide atomic resolution structures of macromolecules in solution. The NMR structure determination process can be summarized as follows: resonance assignment, collection of structural restraints and structure calculation fulfilling the NMR-derived restraints. The first set of structural information is provided by the

chemical shifts. Once the backbone resonances are assigned, the secondary structure can be determined by comparing the chemical shifts to known values for random coil and exploiting the fact that there is a correlation between chemical shifts and protein secondary structures (Wishart et al, 1992; Wishart & Sykes, 1994). For example, the $^{13}\text{C}\alpha$ chemical shift values in α -helices have higher ppm values compared to random coil chemical shifts. In addition, backbone torsion angles (ϕ and ψ) can be predicted from chemical shift by programs such as TALOS+ which compare the chemical shifts with a database of high resolution structures (Shen et al, 2009). In recent years, progress has been made in structure determination using sparse NMR data, for example the Chemical Shift (CS)-Rosetta protocol (Shen et al, 2008) employs only the chemical shift as experimental input for structure determination.

The traditional approach for structure determination by NMR relies heavily on the nuclear Overhauser effect (NOE). The NOE is observed between nuclei which are less than 5-6 Å apart in space and the intensity of the NOE cross-peak is a measure of the distance between the two atoms. For structure determination of macromolecular complexes, it is important to distinguish between inter-molecular and intra-molecular NOE contacts. For this purpose, specific NOE spectroscopy (NOESY) experiments, known as isotope-edited and isotope-filtered, have been developed which allow the detection of selective group of protons. These experiments rely on selective labeling of the components in the complex, for example, in the case of a protein-peptide complex the protein is labeled with ^{13}C , ^{15}N and the peptide is unlabeled (see chapter 8). All the distance restraints from NOEs as well as additional restraints from torsion angles (extracted from J-couplings or predicted based on chemical shifts), hydrogen-bonds (based on hydrogen-deuterium exchange), residual dipolar couplings (RDC) paramagnetic relaxation enhancements (PRE) and small angle scattering (SAS) can be included in the structure calculation protocol. The protocol is based on restrained molecular dynamics (MD) simulations and simulated annealing. The most commonly used programs for structure calculation are XPLOR-NIH (Schwieters et al, 2003), ARIA (Rieping et al, 2007) and CYANA (Guntert, 2004). The final NMR structure is represented by an ensemble of 10-20 lowest-energy structures which are consistent with the NMR data.

Part 1

Structural basis of regulation of miRNA processing by hnRNP A1

2. Introduction

2.1 miRNAs

MicroRNAs (miRNAs) are a class of small (approximately 22 nucleotides) and highly conserved noncoding RNAs. They are important regulators of gene expression in a variety of biological processes including cell growth, proliferation and differentiation. miRNA deregulation can lead to various diseases such as cancer (Ventura & Jacks, 2009). It has been suggested that up to one third of protein coding genes are regulated by miRNAs. Since their first discovery in *C.elegans* by Victor Ambros (Lee et al, 1993b) and Gary Ruvkun (Reinhart et al, 2000), miRNAs have been identified in animals, plants, single cell eukaryotes and viruses. Currently more than 24,000 miRNAs from 206 species have been deposited in the miRNA database (miRBase) (Kozomara & Griffiths-Jones, 2014).

2.2 miRNA biogenesis

The biogenesis of miRNAs comprises two major steps in the nucleus and cytoplasm. miRNA genes are transcribed mainly by RNA polymerase II (Lee et al, 2004) or in some cases by RNA polymerase III (Borchert et al, 2006) producing a transcript known as primary miRNA (pri-miRNA). The pri-miRNA transcript consists of one or more stem-loop structures and is typically capped at the 5'-end and polyadenylated at the 3'-end. Each stem-loop structure of pri-miRNA consists of three parts, i.e. a terminal loop, a partially base-paired stem region and single-stranded regions flanking the stem (Han et al, 2006). The pri-miRNA transcript is cleaved into approximately 70 nucleotide (nt) stem-loop precursor-miRNAs (pre-miRNA) by the nuclear Microprocessor complex consisting of Drosha and DGCR8 (DiGeorge critical region 8, also known as Pasha in *C.elegans* and *D.melanogaster*) (Denli et al, 2004; Gregory et al, 2004). The requirements for recognition and cleavage of pri-miRNA transcripts by the microprocessor complex are not clearly understood, but it seems that the single-stranded regions of pri-miRNAs play an important role in this process (Zeng & Cullen, 2003; Zeng & Cullen, 2005; Zeng et al, 2005; Han et al, 2006; Zhang & Zeng, 2010). Drosha cleaves the pri-miRNA at approximately 2 helical turns away from the terminal loop and approximately 1 helical turn (~11 base-pairs) from the junction of single-stranded and double-stranded RNA at the base of the stem

(Han et al, 2006). Cleavage by Drosha which is a type-III ribonuclease (RNaseIII) yields products with a mono-phosphate group at the 5'-end a 2 nt overhang on the 3'-end (Jinek & Doudna, 2009). The pre-miRNA is then transported to the cytoplasm by the Exportin-5-Ran-GTP system. Exportin-5 recognizes the length of the double-stranded region as well as the 2 nt overhang at the 3'-end of pre-miRNA (Okada et al, 2009).

In the second step of miRNA biogenesis which occurs in the cytoplasm, the RNaseIII enzyme Dicer (Bernstein et al, 2001) cleaves the pre-miRNA into an approximately 22 nt miRNA duplex. The Tar RNA binding protein (TRBP) seems to stabilize Dicer and stimulate its activity (Ma et al, 2008). After cleavage, one strand of the miRNA duplex, known as the guide strand, is selected and loaded together with Argonaute (Ago) proteins into the miRNA-induced silencing complex (miRISC). The other strand, known as the passenger strand or miRNA*, can either be degraded or incorporated into functional Ago complexes (Okamura et al, 2008; Czech et al, 2009; Okamura et al, 2009). Strand selection seems to be determined by different thermodynamic properties of the two strands of the miRNA duplex (Khvorova et al, 2003; Schwarz et al, 2003).

Next, miRISC is guided to the target by base-pairing of miRNA to complementary sequences in the target mRNA transcript. Silencing of target genes occurs by mRNA degradation and/or translational repression, although the exact mechanism of miRNA-mediated gene silencing is not known (Huntzinger & Izaurralde, 2011; Fabian & Sonenberg, 2012).

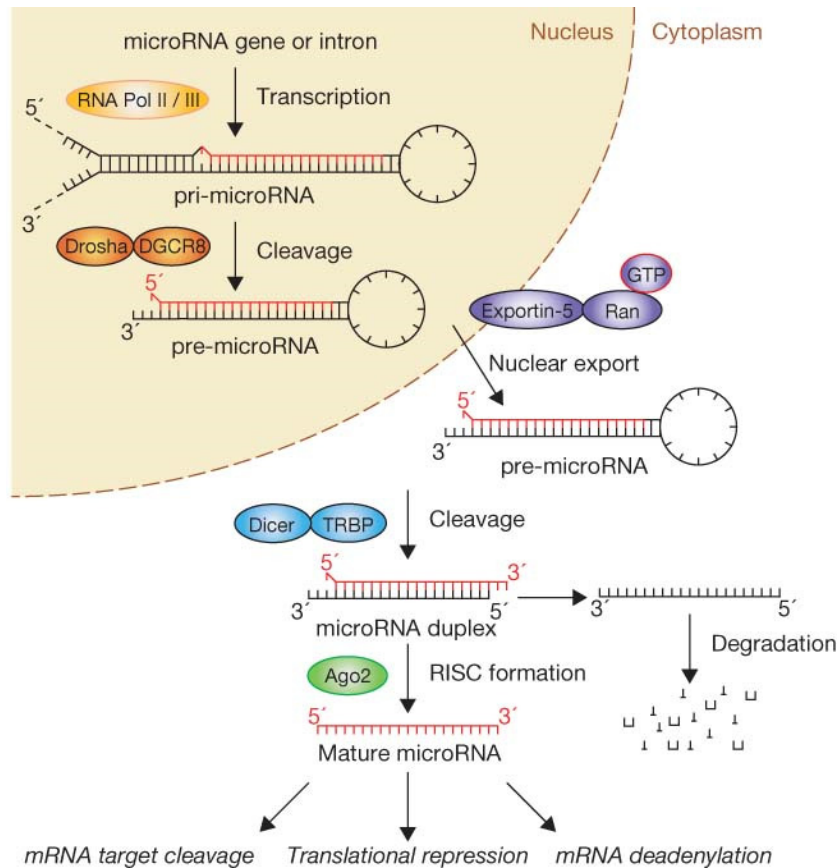


Figure 2.1 Overview of the canonical miRNA biogenesis pathway. It comprises two cleavage steps by type III ribonucleases, Drosha and Dicer. Drosha cleaves the pri-miRNA in the nucleus to generate pre-miRNA which then undergoes a second cleavage by Dicer once it reaches the cytoplasm. (Adapted from Winter et al, 2009)

2.3 Post-transcriptional regulation of miRNA biogenesis

Every step of the miRNA biogenesis pathway is tightly regulated at the transcriptional and post-transcriptional levels (Winter et al, 2009; Krol et al, 2010; Siomi & Siomi, 2010). This results in tissue-specific and developmental-specific expression patterns of miRNAs. Regulation of miRNA biogenesis is critical as aberrant miRNA expression is linked to a variety of human diseases, including cancer (Davis-Dusenbery & Hata, 2010). Post-transcriptional regulation of miRNA biogenesis acts at both the Drosha and Dicer cleavage steps in the nucleus and cytoplasm, respectively.

2.3.1 Terminal loop-mediated regulation of miRNA biogenesis

An important class of miRNA biogenesis regulators are proteins which bind to terminal loops of pri- and pre-miRNAs and either enhance or reduce their expression (reviewed in Choudhury & Michlewski, 2012; Castilla-Llorente et al, 2013). Bioinformatics analyses have shown that approximately 14% of pri-miRNAs have a highly conserved terminal loop and it was postulated that terminal loops can act as a binding platform for trans-acting factors in regulation of miRNA biogenesis. A summary of loop-binding regulators of miRNA biogenesis is listed in Table 2.1. Three of these proteins will be described in more detail below and their proposed mechanism of action is schematically shown in Figure 2.2.

One of the proteins which regulates the processing of miRNAs is Lin28. Lin28 is a negative regulator of let-7 miRNAs and acts at both Drosha and Dicer processing steps (Figure 2.2). Two variants of the protein in mammalian cells, Lin28a and Lin28b, act at different steps of miRNA processing (Piskounova et al, 2011). Lin28a binds to the terminal loop of pre-let7 in the cytoplasm and recruits a terminal uridylyl transferase (TUT) which adds uridines to the 3'-end of the pre-let7 and in turn leads to its degradation (Heo et al, 2008; Hagan et al, 2009; Heo et al, 2009). Lin28b, however, is nuclear and seems to sequester pri-let7 and prevent Drosha cleavage. Structural studies on Lin28 in complex with terminal loops of let-7 have established that the cold-shock domain (CSD) and zinc-fingers of Lin28 specifically recognizes two distinct regions of the RNA (Nam et al, 2011; Loughlin et al, 2012). Moreover, it was reported that Lin28 down-regulates the processing of pre-miR-1 and that MBNL1 (muscleblind-like splicing regulator 1) competes with Lin28 for binding to the loop region of pre-miR-1. Thus, MBNL1 can act as an activator of pre-miR-1 processing (Rau et al, 2011).

Another protein which associates with terminal loops of miRNA precursors and regulates their processing is KSRP (K-homology splicing regulator protein). KSRP has been shown to activate the processing of a subset of miRNAs including let7-a (Trabucchi et al, 2009). Biochemical and structural studies have demonstrated that one of the KH domains of KSRP, KH3, dominates the interaction with the terminal loop of let-7 precursor (Nicastro et al, 2012).

The multifunctional RNA-binding protein hnRNP A1 is another important regulator of miRNA biogenesis. It can act as both an activator and inhibitor of miRNA biogenesis. The role of hnRNP A1 in miRNA biogenesis is described in detail in section 2.5.1.2.

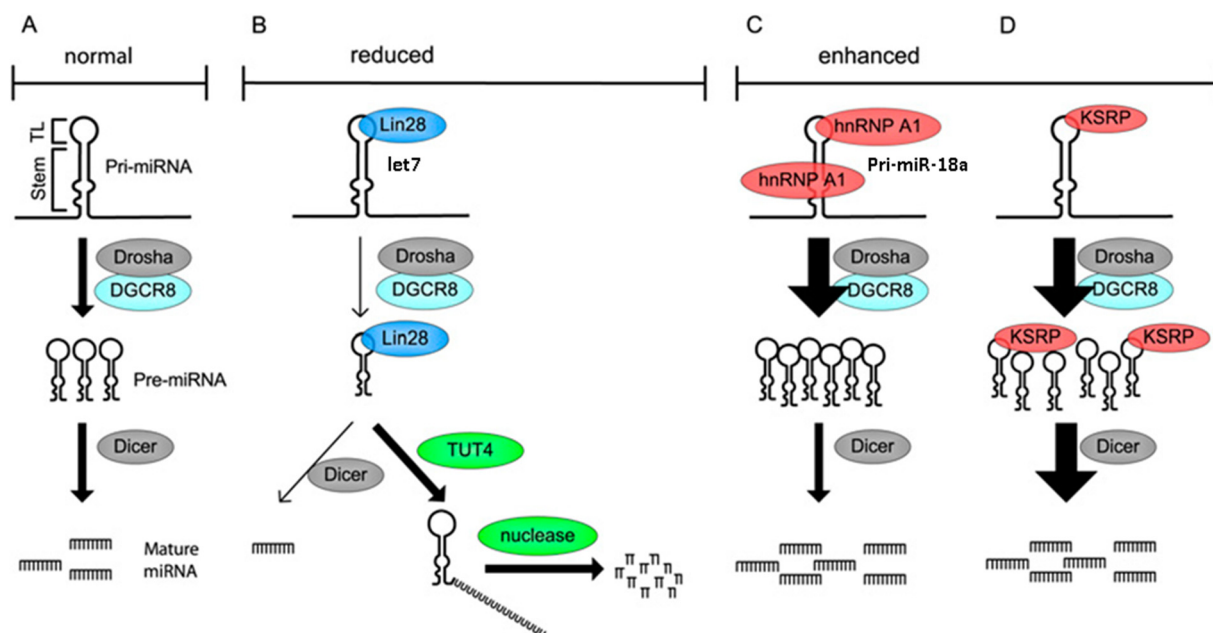


Figure 2.2 Terminal loop-mediated regulation of miRNA processing. (A) Pri-miRNA transcripts contain a terminal loop (TL) and a stem. They are cleaved by the microprocessor complex consisting of Drosha and DGCR8 producing pre-miRNAs which are transported to the cytoplasm and further processed by Dicer. (B) Lin28 binds to the terminal loop of pri- and pre-miRNAs from the let-7 family and down-regulates Drosha and Dicer cleavage. Lin28 also recruits the terminal uridylyl transferase 4 (TUT4) which uridylylates the 3'-end of the RNA and leads to its degradation. (C) hnRNP A1 binds to the terminal loop and stem of pri-miR-18a and facilitates its processing. (D) KSRP binds to the terminal loop of various pri- and pre-miRNAs and up-regulates their processing. (Adapted from Slezak-Prochazka et al, 2010).

Table 2.1 Terminal loop-binding regulators of miRNA biogenesis. (Adapted from Choudhury & Michlewski, 2012)

Protein	Main target	Regulatory mechanism
Positive regulators		
hnRNP A1	Pri-miR-18a	Structural rearrangement
KSRP	Pri-let-7, Pre-let-7	Unknown
MBNL1	Pre-miR-1	Competition with Lin28
Negative regulators		
hnRNP A1	Pri-let-7	Competition with KSRP
Lin28	Pri and pre-let-7	Uridylation by TUT4 and sequestration
MCPIP1	Pre-miR-21 and others	RNA cleavage

2.4 The miR-17-92 cluster

Many miRNAs genes in higher organisms are transcribed together as a cluster (Lau et al, 2001; Lee et al, 2002). A prototypical example is the miR-17-92 cluster which is encoded as an intronic polycistron on chromosome 13 in humans. It encodes for six individual miRNAs (miR-17, miR-18a, miR-19a, miR20a, miR-19b-1, miR-92a-1) which are highly conserved in vertebrates (reviewed in Mendell, 2008). It has been shown that the mir-17-92 cluster acts as an oncogene, hence it is also referred to as OncomiR-1 (He et al, 2005; Esquela-Kerscher & Slack, 2006). Overexpression or knockout of this cluster is associated with tumorigenesis and developmental defects in mouse model systems (He et al, 2005; Ventura et al, 2008).

The miR-17-92 cluster is closely related to the miR-106a-363 cluster located on chromosome X and the miR-106b-25 cluster located on chromosome 7 (Mendell, 2008). It seems that the biogenesis of each miRNA in the cluster is specifically regulated. For example, the biogenesis of two highly similar miRNAs, miR-18a from the miR-17-92 cluster and miR-18b from the miR-106a-363 cluster is different in terms of requirement for an auxiliary factor. It has been shown that the processing of pri-miR-18a requires hnRNP A1, while pri-miR-18b processing is independent of hnRNP A1 (Guil & Caceres, 2007; Michlewski et al, 2008).

Sequence comparison of pri-miR-18a and pri-miR-18b shows that they are 88% identical and only eight nucleotides are different between the two RNAs (Figure 2.3). In terms of secondary structure, differences are seen for the terminal loop as well as the region around the Drosha cleavage site. In the stem region, a UC in pri-miR-18a is replaced by a GU in pri-miR-18b. This results in a tandem UC mismatch in pri-miR-18a, whereas pri-miR-18b features a UC bulge. Interestingly, substitution of UC with GU in the stem region of pri-miR-18a (pri-miR-18a UC→GU) to enforce pri-miR-18b stem conformation, makes its processing efficient and independent of hnRNP A1 (Michlewski et al, 2008).

pri-miR-18a UGUUCUAAGGUGCAUCUAGUGCAGUAGUGAAGUAGAUUAGCAUCUACUGCCCUAAGUGCUCCUUCUGGCA
 pri-miR-18b UGUUGUAAGGUGCAUCUAGUGCAGUUAGUGAAGCAGCUUAGAAUCUACUGCCCUAAUUGCCCUUCUGGCA

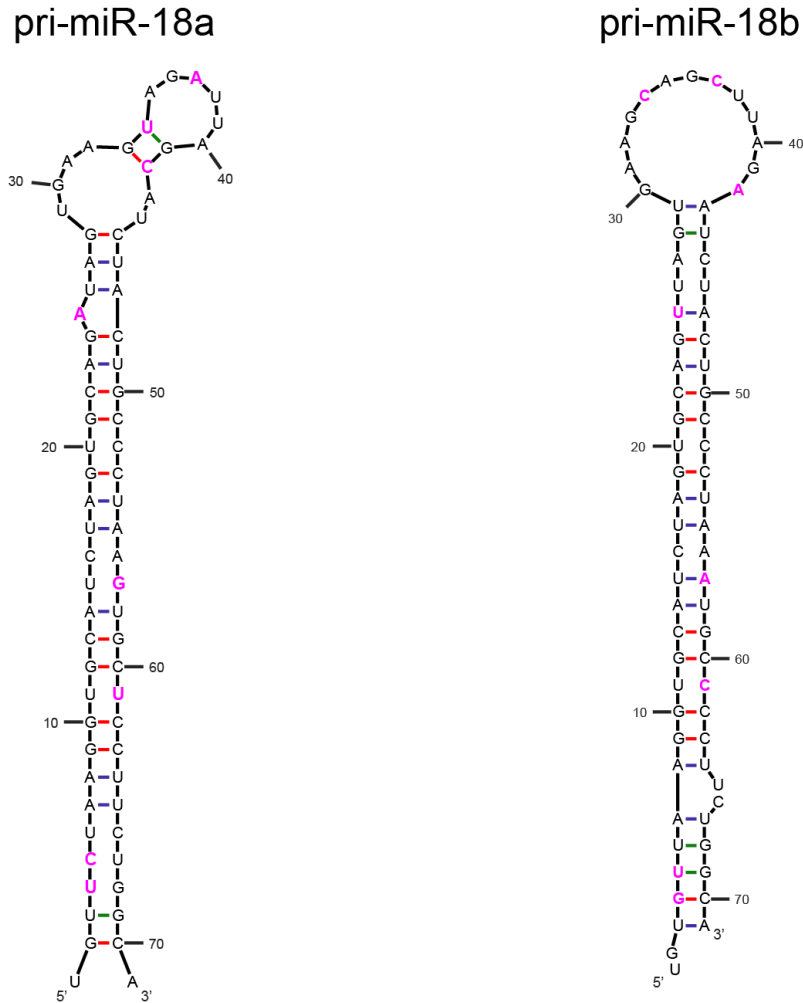


Figure 2.3 Sequence alignment of pri-miR-18a and pri-miR-18b. Mfold (version 3.5) (Zuker, 2003) predicted secondary structures of pri-miR-18a and pri-miR-18b. Nucleotides which differ between the two RNAs are shown in magenta. Sequences were obtained from the miRNA database (miRBase) (Kozomara & Griffiths-Jones, 2014).

Recent studies have shown that the miR-17-92 cluster adopts a compact globular structure that can influence the processing of individual miRNAs within the cluster leading to differential expression levels. These studies suggest that the RNA tertiary structure can also act as a regulator of miRNA biogenesis (Chaulk et al, 2011; Chakraborty et al, 2012; Chaulk et al, 2014).

2.5 hnRNP proteins

Heterogeneous nuclear ribonucleoproteins (hnRNPs) are a family of nuclear RNA-binding proteins which associate with pre-mRNA transcripts (historically known as hnRNA) forming mRNP complexes. In humans, more than 20 proteins belong to the hnRNP family and they are named from A to U. The core hnRNP proteins consist of A1, A2/B1, B2, C1 and C2. HnRNP proteins are involved in various aspects of nucleic acid metabolism in the cell (reviewed in Chaudhury et al, 2010; Han et al, 2010). All hnRNP proteins (with the exception of hnRNP U) have at least one RNA binding domain in the form of RNA recognition motif (RRM) or K-homology (KH) domain. In addition, many hnRNP proteins contain a cluster of Arg-Gly-Gly (RGG) repeats interspersed with aromatic residues which mediate protein and nucleic acid interactions (Thandapani et al, 2013). Although known as nuclear proteins, many hnRNP proteins can shuttle between the nucleus and cytoplasm (Pinol-Roma & Dreyfuss, 1992). The major hnRNP proteins are listed in Table 2.2.

Table 2.2 Major hnRNP proteins and their domain composition. (Adapted from Chaudhury et al, 2010)

hnRNP protein	Domains
A1	2x RRM, RGG
A2/B1	2x RRM, RGG
C1/C2	RRM
D (AUF1)	2x RRM, RGG
E1/E2/E3/E4	3x KH
F	3x RRM
G (RBMX)	RRM, RGG
H	3x RRM
I (PTB)	4x RRM
K	3x KH, RGG
L	4x RRM
M	3x RRM
P2 (FUS)	RRM
Q	3x RRM
R1/R2	RRM
U	RGG

2.5.1 hnRNP A1

Heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) is a highly abundant protein which belongs to the A/B subfamily of hnRNP proteins. The hnRNP A/B subfamily also includes A2/B1, A3 and A0. hnRNP A1 is arguably the most studied hnRNP protein and is implicated in diverse cellular functions including, but not limited to, mRNA export (Izaurralde et al, 1997; Mili et al, 2001), mRNA stability (Henics et al, 1994; Hamilton et al, 1997), telomere maintenance (LaBranche et al, 1998; Zhang et al, 2006), splicing (Okunola & Krainer, 2009; Tavanez et al, 2012), IRES (internal ribosome entry site)-mediated translation (Bonnal et al, 2005; Damiano et al, 2013) and miRNA biogenesis (see section 2.5.1.2). Given its involvement in so many cellular processes, it is not surprising that hnRNP A1 has been linked with several human diseases.

In humans, two main alternatively spliced isoforms of hnRNP A1 have been identified. The shorter isoform (A1-A) is 320 amino-acids long and in most tissues is approximately 20-times more abundant than the larger isoform (A1-B) which has 372 amino acids (Buvoli et al, 1990a). hnRNP A1 comprises two RNA recognition motif (RRM) domains and a flexible glycine-rich tail. A fragment consisting of the tandem RRM1-RRM2 domains is referred to as UP1 (unwinding protein 1). The domain organization of hnRNP A1 is shown in Figure 2.4.

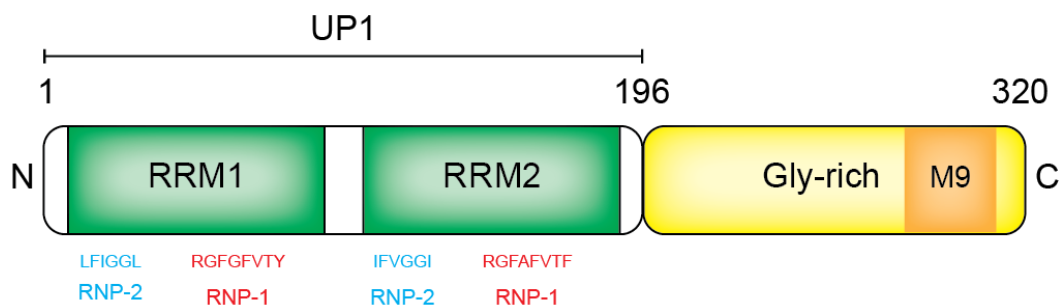


Figure 2.4 Domain organization of human hnRNP A1. A fragment consisting of RRM1 and RRM2 is referred to as UP1. The sequence of conserved RNP-1 and RNP-2 motifs in each RRM domain is shown. M9 is a nuclear import and export signal in the Gly-rich tail of hnRNP A1.

RRMs are the most abundant RNA-binding domains in higher eukaryotes. They typically consist of four β -strands and two α -helices arranged in the order $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$. The RRM fold is composed of a four-stranded antiparallel β -sheet

with two α -helices packed against the β -sheet. Two highly conserved submotifs, RNP-1 octamer (on β 3) and RNP-2 hexamer (on β 1), are the main residues involved in RNA binding (reviewed in Maris et al, 2005; Clery et al, 2008).

The two RRM domains of hnRNP A1 show a high level of similarity (Figure 2.5). They are connected by an approximately 17 residue linker which is conserved both in terms of sequence and length. Despite their high similarity, it has been suggested that the two RRM domains have distinct functions (Mayeda et al, 1998).

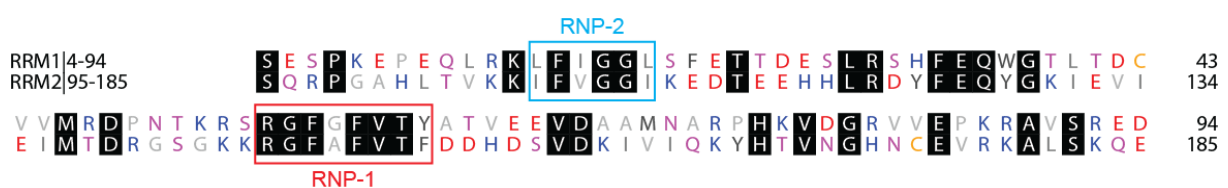


Figure 2.5 Sequence alignment of RRM1 and RRM2 of hnRNP A1. Residues are colored according to amino acid type. Identical residues in the two sequences are shown in white with a black background. RNP-1 and RNP-2 submotifs are indicated by red and cyan boxes, respectively. RRM1 and RRM2 are highly similar (>31% identical and >56% similar with no gaps).

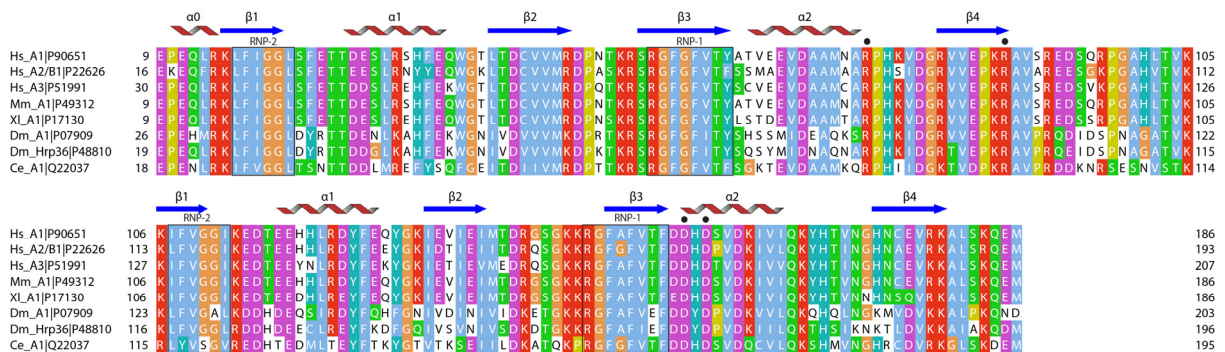


Figure 2.6 Multiple sequence alignment of eukaryotic hnRNP A1 homologs. Sequence alignment was done using MUSCLE (Edgar, 2004) and viewed by Jalview (Waterhouse et al, 2009). The sequence of the Gly-rich tail is not shown. Secondary structure elements (indicated above the sequence) are based on the NMR structure of UP1 (Barraud & Allain, 2013). Conserved residues involved in salt-bridge formation are marked by black circles. Sequences are colored according to the Clustal X coloring scheme (Larkin et al, 2007). The UniProt accession code is indicated for each sequence. (Hs = *Homo sapiens*, Mm = *Mus musculus*, XL = *Xenopus laevis*, Dm = *Drosophila melanogaster*, Ce = *Caenorhabditis elegans*).

Three crystal structures and one NMR structure of UP1 have been reported (Shamoo et al, 1997; Xu et al, 1997; Vitali et al, 2002; Barraud & Allain, 2013). These structures reveal that the two RRM domains adopt the canonical $\beta\alpha\beta\alpha\beta$ topology and they are arranged in an antiparallel manner forming a discontinuous RNA binding platform (Figure 2.7). An important feature of the UP1 structure is the presence of a short 3_{10} helix (α_0) preceding β_1 in RRM1. The two RRM domains are in contact by two conserved Arg-Asp salt-bridges (Arg75-Asp155 and Arg88-Asp157) as well as a small cluster of hydrophobic residues in the interface.

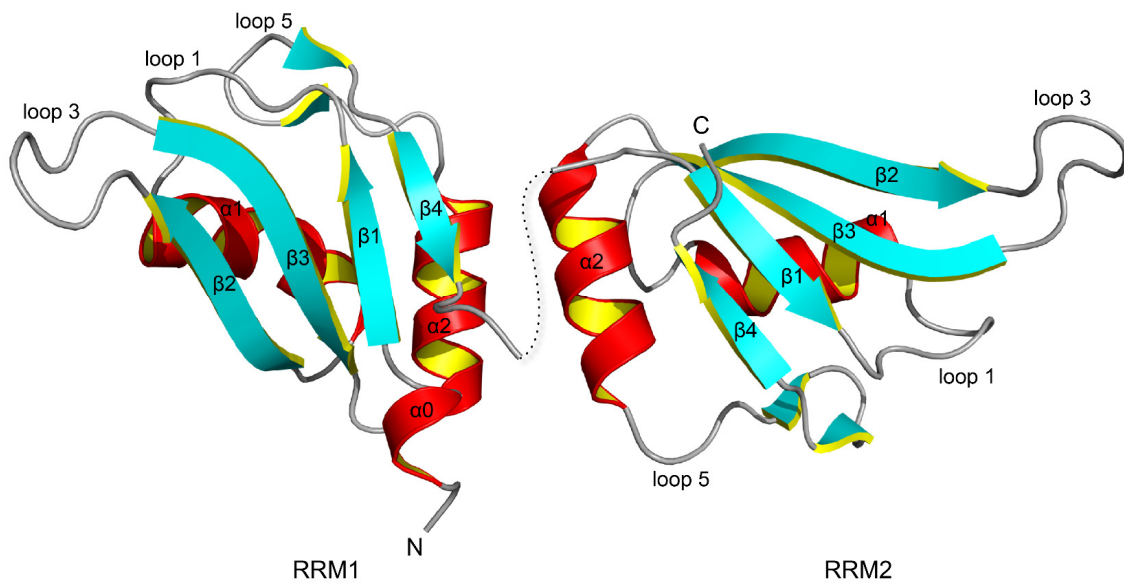


Figure 2.7 Crystal structure of UP1 (PDB code: 1L3K) (Vitali et al, 2002) viewed facing the β -sheets. The two domains are arranged in an antiparallel manner; the order of β -strands from left to right are β_2 - β_3 - β_1 - β_4 for RRM1 and β_4 - β_1 - β_3 - β_2 for RRM2. The linker connecting the two domains, indicated by a dashed line, is not visible in the crystal structure due to flexibility.

Currently, there are no structural information regarding the Gly-rich tail, although it is thought to be flexible. The glycine-rich tail, which consists of 40% glycines, is flexible and mediates interactions with RNAs as well as proteins. It also contains a nuclear localization and export signal known as M9 (Michael et al, 1995; Siomi & Dreyfuss, 1995; Weighardt et al, 1995) which is different from classical nuclear localization signals (NLS). The structural basis of the interaction of M9 peptide (residues 263-289) with Karyopherin β 2 has been revealed (Lee et al, 2006). Furthermore, several studies have recently shown that the Gly-rich tail of hnRNP A1 can form amyloid-like fibrils (Han et al, 2012; Kato et al, 2012; Kim et al, 2013). Importantly, mutations in

this region of hnRNP A1 have been associated with human proteinopathies such as amyotrophic lateral sclerosis (ALS) (Kim et al, 2013).

2.5.1.1 Nucleic-acid binding by hnRNP A1

Nucleic-acid binding properties of hnRNP A1 have been studied extensively (Shamoo et al, 1994; Abdul-Manan et al, 1996; Abdul-Manan & Williams, 1996). These studies have established that hnRNP A1 binds to single-stranded DNA/RNA ligands with a preference for purines over pyrimidines. Based on systematic evolution of ligands by exponential enrichment (SELEX), a high-affinity binding sequence has been identified which contains one or two UAGGGA/U motifs (Burd & Dreyfuss, 1994). It is believed that both UP1 and the Gly-rich tail are involved in nucleic-acid binding; the latter contributing to the binding energy by cooperative binding involving protein-nucleic acid and/or protein-protein interactions.

Although several crystal structures of UP1 bound to DNA have been determined, no structure of UP1 in complex with RNA has yet been reported. The crystal structure of UP1 in complex with a single-stranded telomeric DNA (TTAGGGTTAGGG) revealed the structural basis of nucleic-acid recognition by UP1 (Ding et al, 1999). The structure showed that conserved residues in RNP-1 and RNP-2 submotifs in both RRM domains make specific contacts to DNA nucleotides and that the inter-RRM linker becomes ordered upon nucleic-acid binding. A surprising observation was that UP1 forms a dimer in the crystal and each monomer interacts with a separate DNA strand. A more detailed description of the structure is provided in section 4.6.2.

2.5.1.2 Regulation of miRNA processing by hnRNP A1

In 2007, Guil and Caceres discovered a new function for hnRNP A1. By using an *in vivo* cross-linking and immunoprecipitation protocol (CLIP), they identified hnRNP A1 as an auxiliary factor which facilitates the processing of pri-miR-18a at the Drosha cleavage step (Guil & Caceres, 2007). In a later study, based on footprint analysis performed on pri-miR-18a in the absence and presence of hnRNP A1, it was proposed that hnRNP A1 binds to two regions of pri-miR-18a, i.e. the highly

conserved terminal loop and the stem region (Figure 2.8) (Michlewski et al, 2008). The footprint data suggested that binding of hnRNP A1 induces an opening of the stem region adjacent to the Drosha cleavage site and thus facilitates Drosha-mediated processing. Importantly, a UC→GU mutation in the stem region renders pri-miR-18a processing completely independent of hnRNP A1, suggesting that the native pri-miR-18a sequence or conformation is not optimal for Drosha cleavage. Another interesting finding of this study was that mutating the conserved terminal loop of pri-miR-18a completely abrogates its processing, emphasizing the importance of terminal loops in miRNA processing.

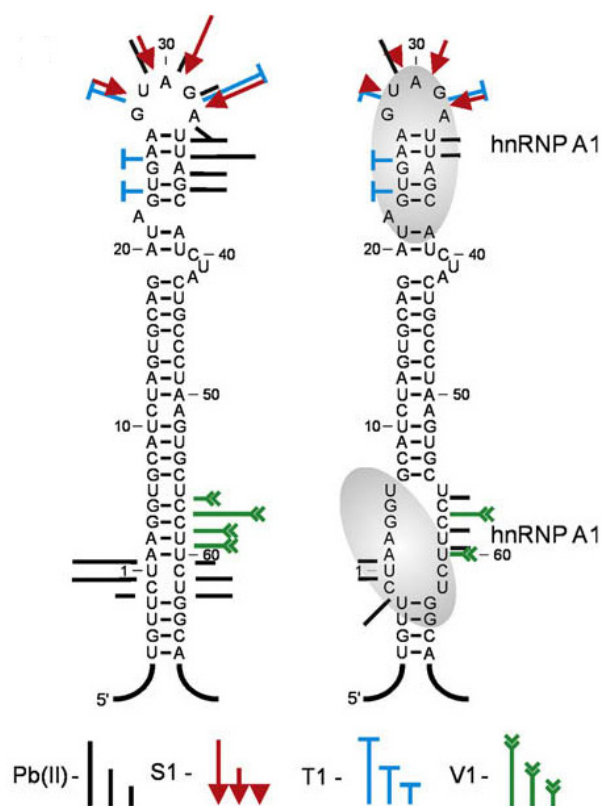


Figure 2.8 Secondary structure of pri-miR-18a in the absence and presence of hnRNP A1 based on footprint analysis. The sites and intensities of cleavage generated by lead(II) ions and RNases S1, T1 and V1 are shown at the hnRNP A1-binding sites. Nucleotides are numbered from the 5'-end of the Drosha cleavage site. (Adapted from Michlewski et al, 2008)

Furthermore, another study by the group of Javier Caceres demonstrated that hnRNP A1 can act as a negative regulator of let-7a processing, presumably by competing with the activator protein KSRP for binding to the pri-let-7a terminal loop (Michlewski & Caceres, 2010).

2.6 Aims and scope of the project

A number of specific RNA-binding proteins have recently emerged as important post-transcriptional regulators of miRNA processing; however, very little is known about their mechanism of action. hnRNP A1 was identified by the group of Javier Caceres (MRC, Edinburgh) as both an activator and inhibitor of specific miRNA precursors. This project focuses on the role of hnRNP A1 as an auxiliary factor required for processing of pri-miR-18a. In a collaboration with the group of Javier Caceres (MRC, Edinburgh), the project aims at understanding the molecular details of the interaction between hnRNP A1 and pri-miR-18a. Specifically, we would like to address the following questions:

Which regions of hnRNP A1 and pri-miR-18a are involved in interaction?

How does hnRNP A1 specifically recognize pri-miR-18a?

What are the structural changes associated with complex formation?

What is the stoichiometry and affinity of this interaction?

How does binding of hnRNP A1 facilitate Drosha-mediated processing of pri-miR-18a?

To address these questions I investigated the molecular basis of the interaction between hnRNP A1 and pri-miR-18a by using a multi-disciplinary approach combining solution NMR spectroscopy, X-ray crystallography, small angle X-ray scattering (SAXS), isothermal titration calorimetry (ITC) and a variety of other biophysical and biochemical methods. Our structural and biochemical data are complemented with functional studies provided from our collaborators. The results of this project will provide insight into the molecular recognition and regulation of miRNA processing by an important RNA-binding protein.

In addition, chapter 6 of this thesis includes a study on the role of hnRNP A1 in splicing. This project is a collaboration with the group of Juan Valcárcel (CRG, Barcelona).

3. Materials and methods

3.1 Materials

3.1.1 Chemicals

Unless otherwise stated, all chemicals were supplied by Sigma-Aldrich, VWR international and SERVA Electrophoresis GmbH (Heidelberg, Germany). Isotopically labeled chemicals were supplied by Cambridge Isotope Laboratories (Massachusetts, USA) and Euriso-Top (Saint-Aubin, France). $^{13}\text{C}/^{15}\text{N}$ -labeled rNTPs (ribonucleoside tri-phosphate) were supplied by Silantes GmbH (Munich, Germany).

3.1.2 Enzymes

All enzymes for molecular biology were purchased from New England Biolabs (NEB) (Massachusetts, USA) and Promega (Wisconsin, USA). They were used with their supplied reaction buffer according to the instructions. T7 RNA polymerase and TEV (Tobacco Etch Virus)-protease were provided by the Protein Expression and Purification Facility (PEPF) of Helmholtz Zentrum München (<http://www.helmholtz-muenchen.de/en/pepf/>).

3.1.3 Bacterial strains and plasmids

Table 3.1 Bacterial strains

strain	supplier
<i>E.coli</i> NovaBlue	Novagen
<i>E.coli</i> DH5 α	Invitrogen
<i>E.coli</i> XL1-Blue	Stratagene
<i>E.coli</i> BL21(DE3)	Novagen

Several bacterial expression vectors for different fragments of human hnRNP A1 have been created (Table 3.2). In the case of full-length hnRNP A1 and Gly-rich tail, various tags were tested for optimal expression and solubility. This includes: Z-tag which is a synthetic IgG-binding domain based on Staphylococcal protein A (Nilsson et al, 1987); MBP (Maltose-binding protein), GST (Glutathione S-transferase) and Trx (Thioredoxin). All tags are removable via TEV cleavage, except for the C-terminal His-tag construct. All generated vectors carry a Kanamycin resistance gene.

Table 3.2 hnRNP A1 expression vectors created in this study

Construct	Gene	Boundaries	Vector	Tag
RRM1	hnRNP A1	1-97	pETM11	N-His ₆
RRM2	hnRNP A1	94-196	pETM11	N-His ₆
UP1	hnRNP A1	1-196	pETM11	N-His ₆
Gly-rich tail	hnRNP A1	195-320	pETM11	N-His ₆
Gly-rich tail	hnRNP A1	195-320	pET Z2-1a	N-His ₆ -Z
Gly-rich tail	hnRNP A1	195-320	pET MBP	N-His ₆ -MBP
Gly-rich tail	hnRNP A1	195-320	pET Trx	N-His ₆ -Trx
Full-length hnRNP A1	hnRNP A1	1-320	pETM11	N-His ₆
Full-length hnRNP A1	hnRNP A1	1-320	pETM30	N-His ₆ -GST
Full-length hnRNP A1	hnRNP A1	1-320	pET ZZ	N-His ₆ -ZZ
Full-length hnRNP A1	hnRNP A1	1-320	pETM13	C-His ₆
Full-length hnRNP A1	hnRNP A1	1-320	pETM13	-

3.1.4 Bacterial growth media and antibiotics

Bacterial growth media were supplemented with appropriate antibiotics based on the resistance of the plasmid. The antibiotics Ampicillin, Kanamycin and Chloramphenicol were used at final concentrations of 100 mg/l, 50 mg/l and 34 mg/l, respectively. All growth media were autoclaved prior to use.

Agar plates were made by adding 1.5 % (w/v) agar to LB-medium. After autoclaving, the medium was allowed to cool to 50-60 °C, appropriate antibiotics were added and the medium was poured into Petri-dishes.

Table 3.3 Media and solutions for bacterial growth

Medium	Component	Amount per liter
LB (lysogeny broth) medium	Tryptone	10 g
	NaCl	10 g
	Yeast extract	5 g
M9 minimal medium (10X)	Na ₂ HPO ₄	60 g
	KH ₂ PO ₄	30 g
	NaCl	5 g
	¹⁵ NH ₄ Cl	5g
Trace elements (100X)	EDTA, pH 7.5	5 g
	FeCl ₃ . 6 H ₂ O	0.83 g
	ZnCl ₂	84 mg
	CuCl ₂ . 2 H ₂ O	13 mg
	CoCl ₂ . 6 H ₂ O	10 mg
	H ₃ BO ₃	10 mg
	MnCl ₂ . 6 H ₂ O	1.6 mg

¹⁵N isotope-labeled proteins were expressed in M9 minimal media supplemented with 0.5 g/litre [¹⁵N]-ammonium chloride as sole nitrogen source (Table 3.4). For ¹³C,¹⁵N isotope-labeling, the unlabeled glucose in the protocol was replaced by 2 g/litre uniformly-labeled [U-¹³C]-glucose.

Table 3.4 Recipe for ¹⁵N-labeling medium

Component	Volume per litre
M9 medium (10X)	100 ml
Trace elements (100X)	10 ml
1 M MgSO ₄	1 ml
1 M CaCl ₂	0.3 ml
Biotin (1 mg/ml)	1 ml
Thiamin (1 mg/ml)	1 ml
Glucose 20% (w/v)	20 ml

3.1.5 Buffers

Table 3.5 Recipe for commonly used buffers

Buffer	pH	Component	Amount (in 1 litre H ₂ O)
10X PBS (Phosphate Buffered Saline)	7.4	NaCl	80 g
		KCl	2 g
		Na ₂ HPO ₄	11.5 g
		KH ₂ PO ₄	2 g
10X TBE (Tris Borate EDTA)	8.3	Tris base	108 g
		Boric acid	55 g
		EDTA (pH 8.0)	40 ml of 0.5M stock
50X TAE (Tris Acetate EDTA)	8.4	Tris base	242 g
		Acetic acid	57.1 ml
		EDTA (pH 8.0)	100 ml of 0.5M stock
10X SDS-PAGE running	8.3	Tris base	30 g
		Glycine	144 g
		SDS	10 g

Table 3.6 Composition of buffers used for protein and RNA biochemistry

Buffer	pH	Component	concentration
Lysis buffer	8.0	Tris-HCl	30 mM
		NaCl	500 mM
		Imidazole	10 mM
		TCEP	1 mM
		Glycerol	5-10%
Wash buffer	8.0	Tris-HCl	30 mM
		NaCl	1000 mM
		Imidazole	20 mM
		TCEP	1 mM
Elution buffer	8.0	Tris-HCl	30 mM
		NaCl	150 mM
		Imidazole	400 mM
		TCEP	1 mM
NMR buffer	6.5	Sodium-phosphate	20 mM
		NaCl	100 mM
		DTT	1 mM
20X transcription buffer	8.0	Tris-HCl	800 mM
		DTT	100 mM
		Spermidine	20 mM
		Triton X-100	0.2%

3.1.6 Oligonucleotides

DNA oligonucleotides were synthesized by Eurofins Genomics (Ebersberg, Germany). A list of the primers used for cloning and mutagenesis are provided in the Appendix. RNA oligonucleotides were purchased from IBA (Göttingen, Germany) and Biospring (Frankfurt, Germany) in double-desalted form and used without further

treatment. The RNA was dissolved in NMR buffer or H₂O at 5-10 mM concentration. To ensure proper folding, RNA samples were heated at 95 °C for 2 minutes and snap-cooled on ice before use.

3.2 Methods

3.2.1 Molecular biology

Basic molecular biology techniques which are not described here have been performed using standard protocols (Sambrook & Russell, 2001).

3.2.1.1 Transformation

A 50 µl aliquot of chemical competent cells was thawed on ice for 5 minutes. 1 µl of plasmid DNA or 10-20 µl of ligation reaction were added and the cells were placed on ice for 30 minutes. Cells were heat-shocked at 42 °C for 55 seconds and placed on ice for 2 minutes. 1 ml of LB medium was added and the cells were incubated at 37 °C for 45 minutes in a thermomixer. The cells were collected by centrifugation (6000 g for 2 minutes), spread on LB agar plates containing the desired antibiotic and incubated at 37 °C overnight.

3.2.1.2 Plasmid purification

All plasmids were purified using the Wizard® *Plus* SV Minipreps DNA Purification System (Promega) and sequenced at GATC Biotech AG (Konstanz, Germany).

3.2.1.3 DNA cloning

To generate constructs of hnRNP A1, DNA fragments were amplified by polymerase chain reaction (PCR) from pET9d-hnRNP A1 (Mayeda & Krainer, 1992) using forward and reverse primers carrying NcoI and Acc65I restriction sites, respectively. The sequence of primers used for DNA cloning are listed in the Appendix.

A typical PCR reaction setup and thermocycling program are shown below.

Table 3.7 PCR reaction setup.

Component	Amount/volume
Template plasmid	50-100 ng
Forward primer	12.5 pmol
Reverse primer	12.5 pmol
<i>Pfu</i> DNA polymerase (Promega)	1.5 U
dNTPs	0.2 mM each
<i>Pfu</i> reaction buffer (10X, Promega)	5 μ l
Final volume	50 μ l

Table 3.8 PCR thermocycling program. The cycling steps are highlighted in blue.

Step	Temperature ($^{\circ}$ C)	Time	Cycles
Initial denaturation	95	1 min	1
Denaturation	95	30 sec	30
Annealing	55	1 min	
Extension	72	2 min	
Final extension	72	10 min	1
Final hold	10	∞	

Following agarose gel electrophoresis of the PCR mixture, the band corresponding to the insert was excised and the DNA was extracted using the Wizard[®] SV Gel and PCR Clean-Up System (Promega). Both insert and vector were digested with *Nco*I and *Acc*65I and the digestion reactions were purified. For ligation, digested vector was combined with a 3-fold molar excess of digested insert and Quick T4 ligase (NEB). The ligation reaction was incubated at room temperature for 30 minutes and cooled on ice before transformation into *E.coli* NovaBlue or *E. coli* XL1-Blue competent cells.

3.2.1.4 Site-directed mutagenesis

Mutations were introduced using the QuickChange® site-directed mutagenesis protocol. Alternatively, a modified procedure for site-directed mutagenesis was used as described elsewhere (Zheng et al, 2004; Liu & Naismith, 2008). Primers containing the desired mutation were used to amplify the template plasmid by PCR. The primers used for mutagenesis are listed in the Appendix. The PCR reaction was essentially as described above. In some cases, addition of 5-10% of DMSO in the PCR reaction was essential. DMSO is known to facilitate denaturation of high GC-content templates. The PCR thermocycling program for site-directed mutagenesis is shown in Table 3.9.

Table 3.9 PCR program for site-directed mutagenesis

Step	Temperature (°C)	Time	Cycles
Initial denaturation	95	1 min	1
Denaturation	95	1 min	
Annealing	55-60	1 min	18
Extension	68	12 min	
Final extension	68	12 min	1
Final hold	10	∞	

Following the PCR reaction, DpnI was added to the mixture in order to digest the template (non-mutated) DNA. DpnI is a restriction enzyme which only digests methylated DNA. The template plasmid purified from a *dam*⁺ *E.coli* strain is methylated and hence destroyed by DpnI while the non-methylated mutated DNA produced by PCR is preserved. After heat inactivation of DpnI at 80 °C for 20 min, 10-20 µl of the digestion mixture was transformed into *E.coli* NovaBlue or *E. coli* XL1-Blue competent cells.

3.2.2 Biochemistry

3.2.2.1 Recombinant protein expression and isotope labeling

All proteins used in this study were recombinantly expressed in *E.coli* BL21 (DE3) cells. A single colony of *E.coli* BL21 (DE3) cells harboring the expression vector was inoculated in 20-40 ml of LB medium containing the appropriate antibiotics at 37 °C for approximately 16 hours. For expression of unlabeled proteins, the starter culture was diluted into 1 litre of LB. For isotope-labeled proteins, the LB was removed by centrifugation and the cells were added to minimal medium. Bacterial cells were grown at 37 °C with vigorous shaking until they reached an OD₆₀₀ of 0.6-0.8. Protein expression was then induced by addition of 0.5 mM IPTG. After approximately 20 hours of expression at 20 °C, the cells were harvested by centrifugation at 6000 g for 20 minutes, washed by 1X PBS buffer, flash frozen in liquid nitrogen and stored at -80 °C until further use.

3.2.2.2 Protein purification

A frozen cell pellet from 1 litre of cell culture was thawed on ice, re-suspended in 20 ml of lysis buffer supplemented with protease inhibitor mix HP (Serva Electrophoresis GmbH) and lysed by sonication. Subsequently, the crude extract was cleared of cellular debris by centrifugation at 20,000 g for 45 minutes at 4 °C. Typically, the protein purification protocol consisted of Ni²⁺-affinity chromatography and size-exclusion chromatography as follows. The cleared lysate was loaded onto Ni²⁺-NTA resin (Qiagen) and after several washing steps with lysis buffer and wash buffer the protein was eluted by 15 ml of elution buffer. The tag was removed by TEV protease cleavage at 4 °C overnight. The next day, the buffer was exchanged to lysis buffer using a PD-10 column (GE Healthcare) and the solution was loaded onto Ni²⁺-NTA resin to remove the His-tagged TEV protease, expression tag and uncleaved protein. The flow-through was concentrated using an Amicon Ultra-15 Centrifugal Filter Unit (Millipore) with the appropriate molecular weight cut-off and applied to a HiLoad 16/600 Superdex 75 prep grade size-exclusion column (GE Healthcare) equilibrated with NMR buffer. Samples for NMR measurements were concentrated to 250-500 µl and 10% D₂O was added for the NMR lock signal. During the purification procedure, protein samples were analyzed by Sodium Dodecyl Sulfate

Polyacrylamide Gel Electrophoresis (SDS-PAGE) according to the Laemmli method (Laemmli, 1970).

3.2.2.3 *In vitro* transcription

In vitro transcription of stem 18a was performed using synthetic DNA templates. The DNA template consists of two single stranded DNA oligonucleotides: top and bottom strands. The top strand comprises the T7 promoter sequence and the bottom strand carries the sequence reverse complimentary to the desired RNA sequence followed by the sequence reverse complimentary to the top strand (Figure 3.1). To increase the transcription yield and homogeneity, the two 5'-end nucleotides of the bottom strand contained ribose C2' methoxy (-OCH₃) groups. It has been shown that 2'-methoxylation of the template DNA can reduce non-templated nucleotide addition by RNA polymerase (Kao et al, 1999).

Stem 18a (24 nt)

T7 promoter

Top strand: 5' TAATACGACTCACTATA**G**
Bottom strand: 3' ATTATGCTGAGTGATATC CAAGATTCCCTTTGGAAGAC**CG** 5'

Figure 3.1 Synthetic DNA templates used for *in vitro* transcription of stem 18a. The first nucleotide to be incorporated into RNA during transcription is colored red. The two 5'-end nucleotides of the bottom strand (shown in green) are 2'-methoxylated.

Analytical scale *in vitro* transcriptions for optimization Mg²⁺ and NTPs concentrations were carried out in 20 µl reactions containing template DNA, 10-50 mM MgCl₂, 4-10 mM of each NTP (ATP,CTP,GTP and UTP) and T7 RNA polymerase in transcription buffer. After 4 hours of incubation at 37 °C, 10 µl of each reaction was analyzed on a denaturing polyacrylamide gel. Large-scale *in vitro* transcriptions were carried out in 5-10 ml reaction volume with optimized MgCl₂ and NTPs concentrations. For stem 18a this corresponds to 28 mM MgCl₂ and 16 mM NTPs. The reaction mixture for a 10 ml *in vitro* transcription reaction of 18a stem is shown in Table 3.10.

Table 3.10 *In vitro* transcription mixture for 18a stem.

Component	Concentration	Volume (μ l)
Top strand	100 μ M	48
Bottom strand	100 μ M	40
MgCl ₂	1 M	20
H ₂ O		1432
PEG 8000	500 mg/ml	1600
MgCl ₂	1 M	260
Transcription buffer	20X	500
NTPs (ATP, CTP, GTP, UTP)	100 mM each	1600 (4 X 400)
H ₂ O		4000
T7 RNA polymerase	~ 10 mg/ml	500
Final volume		10000

In the first step, the top and bottom strands were mixed in the presence of magnesium (highlighted in Table 3.10), incubated at 65 °C for 5 minutes and slowly cooled down to room temperature to allow annealing of the DNA strands. The annealed template was then added to other components of the *in vitro* transcription reaction. After transcription, the magnesium pyrophosphate precipitate produced during the reaction was removed by centrifugation. The supernatant was ethanol precipitated at -20 °C and the resulting pellet was dissolved in RNA gel-loading buffer and loaded onto a 20% (19:1 acrylamide:bisacrylamide) denaturing polyacrylamide gel. For production of ¹³C,¹⁵N-labeled RNA, ¹³C,¹⁵N-labeled rNTPs were used in the reaction.

Following separation of RNAs by polyacrylamide gel electrophoresis, the desired band was visualized by UV shadowing (on a fluorescent thin layer chromatography plate), excised and electroeluted. For electroelution, gel slices were placed in an electroelution cassette (Elutrap system, Whatman) which is then placed in a horizontal electrophoresis chamber containing 0.5X TBE buffer. By applying an electrical field, the RNA migrates out of the gel into a trap area between BT1 and BT2 membranes (Whatman, Schleicher and Schuell) and is then collected. Electroelution was carried out at 4 °C. RNA samples obtained from electroelution

contain TBE buffer, urea, and un-polymerized acrylamide. To remove these contaminants, samples were extensively washed and exchanged to water in an Amicon-Ultra Centrifugal Filter Unit (Millipore) with the appropriate molecular weight cut-off.

Pri-miR-18a was produced by *in vitro* transcription using a linearized plasmid as template. To generate the plasmid encoding pri-miR-18a, the DNA fragment was amplified by PCR and inserted into a pUC19 vector via EcoRI and XbaI restriction sites. The sequence of primers used for cloning is provided in the Appendix. Preparative scale *in vitro* transcription of pri-miR-18a was performed using BbsI-linearized vector in a reaction volume of 10 ml containing 40 mM MgCl₂ and 8 mM of each NTP as described for stem 18a.

The oligomeric state of RNA samples were analyzed by native PAGE. The correct folding of the RNA constructs were examined using 1D and 2D NMR.

3.2.3 NMR spectroscopy

3.2.3.1 Resonance assignments

NMR experiments were recorded on 900, 800, 750, 600 and 500 MHz Bruker Avance NMR spectrometers, all (except the 750 MHz) equipped with cryogenic triple resonance gradient probes. NMR spectra were processed by NMRPipe (Delaglio et al, 1995) and analyzed using Sparky (Goddard & Kneller).

The following samples were used for backbone resonance assignments: Uniformly ¹⁵N,¹³C-labelled RRM2 and uniformly ¹⁵N,¹³C-labelled UP1 with random fractional deuteration, both at 1 mM concentration in NMR buffer. Standard triple resonance experiments HNCA, HNCACB and CBCA(CO)NH (Sattler M et al, 1999) were recorded on a 600 MHz Bruker spectrometer at 298 K.

3.2.3.2 NMR Relaxation analysis

NMR ^{15}N relaxation experiments were recorded on a Bruker 600 MHz spectrometer at 298 K. The $\{^1\text{H}\}$ ^{-15}N heteronuclear NOE values were determined as the ratio of peak intensities in two experiments, with and without proton saturation. ^{15}N T_1 and $T_{1\rho}$ relaxation times were determined from pseudo-3D HSQC-based experiments recorded in an interleaved fashion with 12 different relaxation delays (21.6, 86.4, 162, 248.4, 345.6, 432, 518.4, 669.6, 885.6, 1144.8, 1404 and 1782 ms) for T_1 and 8 different relaxation delays (5, 7, 10, 15, 20, 25, 30 and 40 ms) for $T_{1\rho}$. Two delays in each experiment were recorded in duplicates for error estimation. R_1 and $R_{1\rho}$ rates were extracted by fitting the data to an exponential function using the relaxation module integrated in NMRViewJ (Johnson & Blevins, 1994). ^{15}N R_2 rates were calculated from the relaxation rate constant in the rotating frame ($R_{1\rho}$) according to (Massi et al, 2004):

$$R_2 = \frac{R_{1\rho} - R_1 \cos^2 \theta}{\sin^2 \theta}$$

and

$$\theta = \arctan\left(\frac{\omega_1}{\Omega}\right)$$

where Ω is the resonance offset and ω_1 is the spin-lock frequency.

3.2.3.3 NMR titration analysis

NMR titration were carried out using uniformly ^{15}N -labeled samples (or in the case of UP1 ^{15}N , ^{13}C -labelled UP1 with random fractional deuteration) at 100–200 μM concentration. RNA was added to the protein in a step-wise manner and ^1H , ^{15}N -HSQC spectra were recorded after each addition. Typically, the titration points included RNA to protein molar ratios of: 0.1, 0.2, 0.4, and in steps of 0.2 up to 2-4.

Combined ^{15}N and ^1H chemical shift perturbations (CSP) were calculated according to:

$$CSP = \sqrt{\delta H(\text{ppm})^2 + 0.1 \times \delta N(\text{ppm})^2}$$

Where δH and δN are changes in the chemical shift in the 1H and ^{15}N dimensions of the $^1H,^{15}N$ -HSQC, respectively.

For fast-exchange binding, the dissociation constant (K_D) can be estimated by fitting the CSP for each non-overlapped peak to the following equation:

$$\delta = \delta_{max} \cdot \frac{([L] + [P] + K_D) - \sqrt{([L] + [P] + K_D)^2 - (4[P][L])}}{2[P]}$$

Where δ is the CSP at a given ligand concentration and δ_{max} is the CSP at saturation. $[L]$ and $[P]$ are ligand and protein concentrations, respectively.

3.2.4 Structural biology and biophysics

3.2.4.1 X-ray crystallography

Screening for crystallization conditions was done using commercial screens from QIAGEN and Hampton Research at the MPI crystallization facility and the X-ray crystallography platform at Helmholtz Zentrum München. All crystallization trials were set up at 20 °C.

For the UP1-17-mer complex, UP1 (in 20 mM Hepes pH 7.0, 100 mM NaCl, 50 mM KCl and 1 mM TCEP) was mixed with 17-mer RNA in a final molar ratio of 1:1.3 (0.76 mM protein: 1 mM RNA) and incubated on ice for 1 hour before the crystallization drops were set up. Optimized crystals of UP1-17-mer were grown by the hanging-drop vapor diffusion method by mixing equal volumes of the complex and reservoir (25-30% MPD, 300 mM NaCl and 50 mM HEPES pH 7-7.5) equilibrated over the reservoir. In the case of UP1-12-mer complex, UP1 (in 20 mM Hepes pH 7.0, 100 mM NaCl, and 1 mM TCEP) was mixed with 12-mer RNA in a 1:1 molar ratio (0.95 mM complex concentration) and incubated on ice for 1 hour. Crystallization drops were set up by mixing 100 nl of complex and 100 nl of reservoir in a sitting drop vapor diffusion setup. Crystals with hexagonal plate morphology were obtained in 0.2M sodium citrate and 20% PEG 3350 from the Classics suite II

screen (QIAGEN). The crystals were cryoprotected in a reservoir solution supplemented with 20% (v/v) ethylene glycol and flash frozen in liquid nitrogen. X-ray diffraction data were collected at the European Synchrotron Radiation Facility (ESRF) and Swiss Light Source (SLS).

All diffraction images were processed by XDS (Kabsch, 2010). The CCP4 suite of programs (Winn et al, 2011) was used for all subsequent data analysis. The structure of 17-mer duplex RNA was solved by single-wavelength anomalous diffraction (SAD) exploiting the anomalous signal of Gadolinium ions. The structure of UP1-12-mer RNA was solved by molecular replacement with the program Phaser (McCoy et al, 2007) using the coordinates of UP1 (PDB code:1U1R) (Myers & Shamoo, 2004). Model building was performed manually in Coot (Emsley et al, 2010) and refinement was done using Refmac (Murshudov et al, 1997).

3.2.4.2 Isothermal titration calorimetry (ITC)

ITC measurements were carried out at 25 °C using an iTC200 calorimeter (GE Healthcare). Both protein and ligand were exchanged into NMR buffer without DTT. Typically, protein concentrations in the range of 300-1000 μM (depending on the affinity of the interaction) were injected into the sample cell containing RNA with a concentration of 20-100 μM . Titrations consisted of 20 injections of 2 μl or 26 injections of 1.5 μl with a 3 minute spacing between each injection. After correction for heat of dilution, data were fitted to a one-site binding model using the Microcal Origin 7.0 software. Each measurement was repeated at least three-times.

3.2.4.3 Static light scattering (SLS)

Static light scattering experiment were performed on a S75 10/300 size-exclusion column (GE Healthcare) connected to a Viscotek Tetra Detector Array (TDA) instrument equipped with refractive index (RI), light scattering, viscosity and photo diode array (PDA) detectors (Malvern). The RI, light scattering and viscosity signals provide information about the concentration, absolute molecular weight (MW) and hydrodynamic radius, respectively. Sample volume of 100 μl was injected onto the

column pre-equilibrated with NMR buffer and the flow rate was set to 0.5 ml/min. Calibration was done with BSA (bovine serum albumin) at concentration of 4-5 mg/ml. Data were analyzed by the OmniSEC software using refractive index increment (dn/dc) values of 0.185 mL/g and 0.17 mL/g for protein and RNA samples, respectively (Rambo & Doudna, 2004).

3.2.4.4 Small angle X-ray scattering (SAXS)

SAXS data were recorded at the X33 beamline of the European Molecular Biology Laboratory (EMBL) at Deutsches Elektronen Synchrotron (DESY, Hamburg) at 15 °C. The scattering curves were measured with 120 seconds exposure time (8 frames, 15 seconds each) for concentrations in the range of 1-10 mg/ml. The scattering intensity $I(s)$ was recorded as a function of momentum transfer s ($s = 4\pi \sin\theta/\lambda$, where 2θ is the scattering angle and λ is the wavelength of the incident X-ray beam) over the s range of 0.007 - 0.63 Å⁻¹. Individual frames collected during the exposure time were compared to check for radiation damage and then averaged. Scattering of buffer measured before and after each sample was averaged and subtracted from the scattering of the sample. All SAXS data were processed using the ATSAS software package (Petoukhov et al, 2012). The radius of gyration (R_g) and the maximum dimension (D_{max}) values were obtained from the GNOM program which evaluates the pair-distance distribution function, $P(r)$ (Svergun, 1992). Multiple independent ab initio models were calculated using DAMMIF (Petoukhov & Svergun, 2005) and averaged by DAMAVER (Volkov & Svergun, 2003). For ab initio modeling of protein-RNA complexes, three scattering profiles corresponding to individual components and the complex were used as input for the multiphase modeling program MONSA (Svergun, 1999). The program BUNCH (Petoukhov & Svergun, 2005) was employed for combining high resolution structures with ab initio models of the flexible regions. Theoretical scattering curves were calculated using CRY SOL (Svergun et al, 1995).

4. Results

4.1 NMR characterization of hnRNP A1

4.1.1 Constructs of hnRNP A1

The human hnRNP A1 is a 320 residue long protein and comprises two RRM domains and a flexible glycine-rich tail. A fragment consisting of the tandem RRM1-RRM2 domains is referred to as UP1 (uncleaved protein 1). The domain organization of hnRNP A1 is shown in Figure 4.1. Several constructs of hnRNP A1 were designed for recombinant expression in *E.coli*. These include constructs corresponding to full-length hnRNP A1, RRM1, RRM2, glycine-rich tail and UP1. Unpublished results from the group of Dr. Javier Caceres have shown that UP1 is sufficient to induce processing of pri-miR-18a (see section 4.10). Therefore, in this work I have mainly focused on the RRM domains and the UP1 fragment of hnRNP A1.

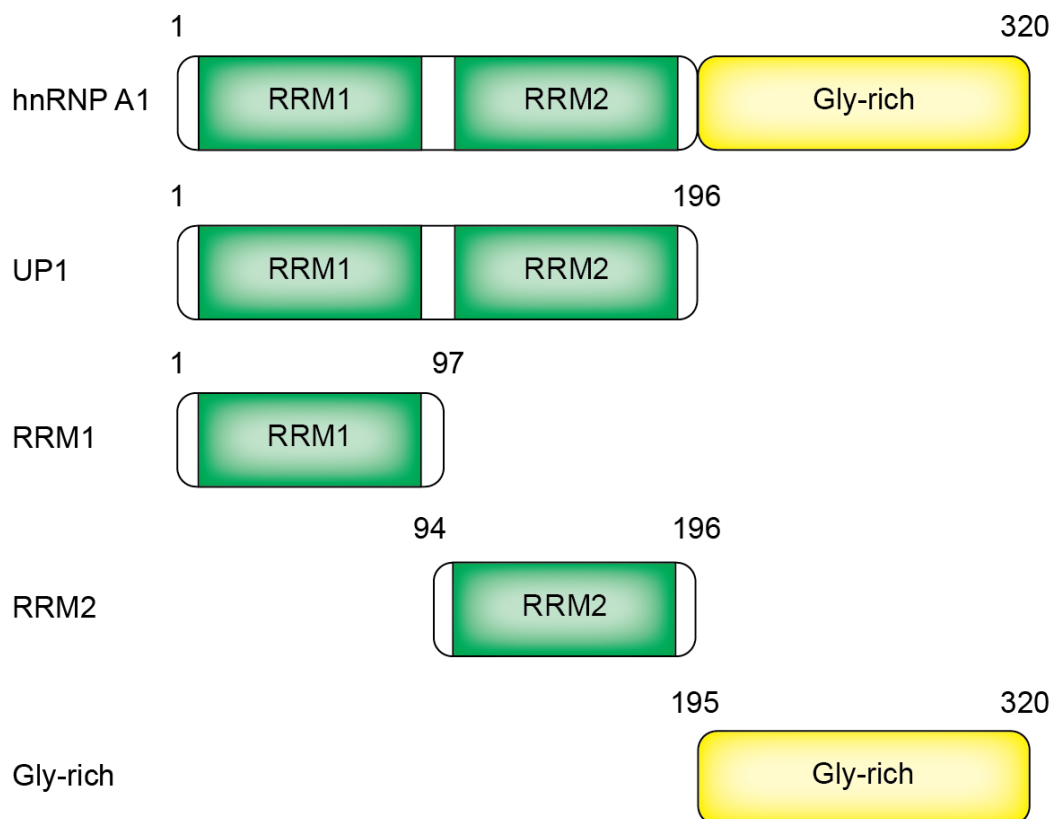


Figure 4.1 Schematic representation of hnRNP A1 constructs created in this study. The boundaries of each construct is indicated.

4.1.2 NMR analysis of RRM1, RRM2 and UP1

RRM1, RRM2 and UP1 were recombinantly expressed, purified and ^1H , ^{15}N HSQC spectra were recorded. The peaks in the spectra are well-dispersed, indicating that the protein samples are folded in the conditions used for NMR measurements. An overlay of the spectra of individual domains resembles the spectrum of UP1, suggesting that the conformation of the domains in the UP1 construct is largely unchanged when compared to the isolated domains (Figure 4.2).

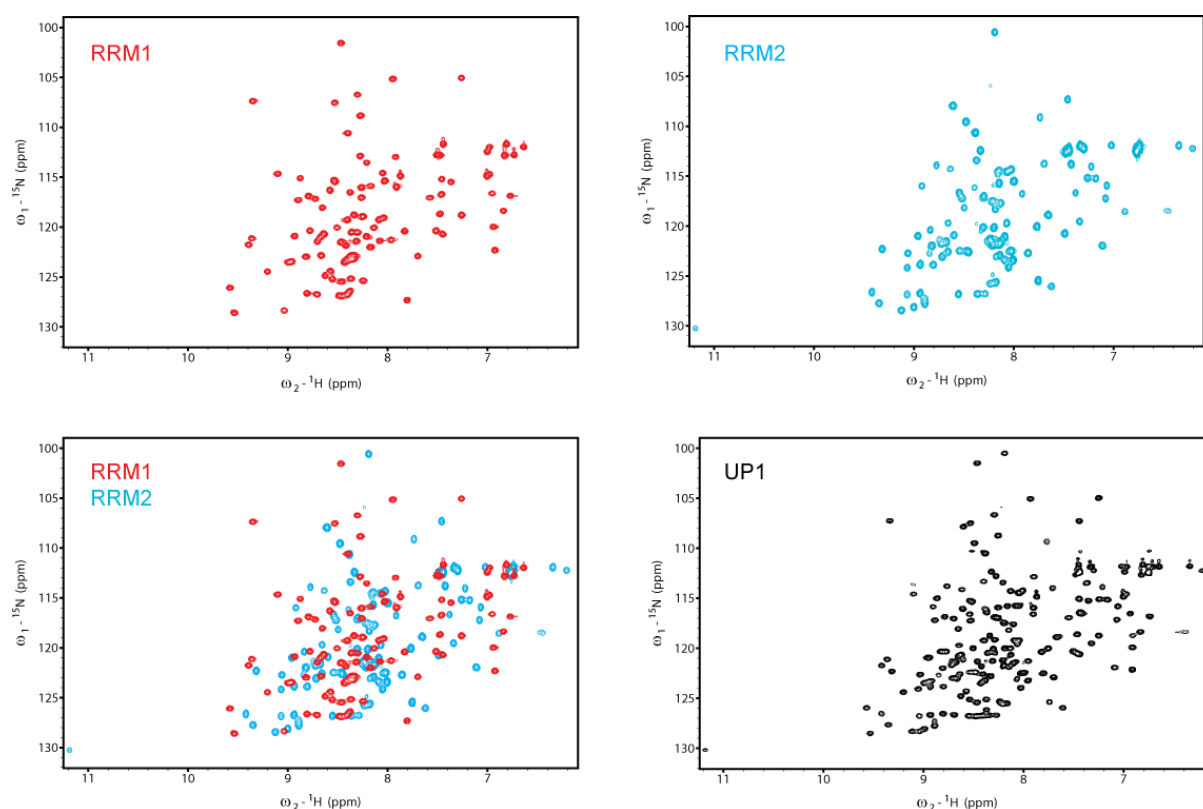


Figure 4.2 ^1H , ^{15}N HSQC spectra of RRM1, RRM2 and UP1 recorded at 298 K on a 750 MHz spectrometer. An overlay of the RRM1 and RRM2 spectra is shown for comparison with the spectrum of UP1.

To assign the backbone chemical shifts of UP1, 3D HNCACB, CBCA(CO)NH and HNCA spectra (Sattler M et al, 1999) were recorded. The assignments were aided by available assignments of RRM1 (Garrett et al, 1992) as well as assignments that were obtained for RRM2. I was able to assign more than 90% of the backbone chemical shifts of UP1. A stretch of five amino acids ($^{112}\text{IKEDT}^{116}$) which corresponds to a loop between β_1 and α_1 (loop 1) in RRM2 were not visible

(probably due to exchange) and could not be assigned. Based on the assigned backbone chemical shifts, I determined the secondary structure of UP1 using TALOS+ (Shen et al, 2009) (Figure 4.3). The results indicate that each domain of UP1 adopts the canonical $\beta\alpha\beta\alpha\beta$ topology known for RRM domains (reviewed in Maris et al, 2005; Clery et al, 2008) with the exception that RRM1 starts with a short α -helix (denoted α_0) and that the C-terminus of RRM2 has helical propensity (denoted α_3). Overall, the secondary structure of UP1 in solution is consistent with the reported crystal structures.

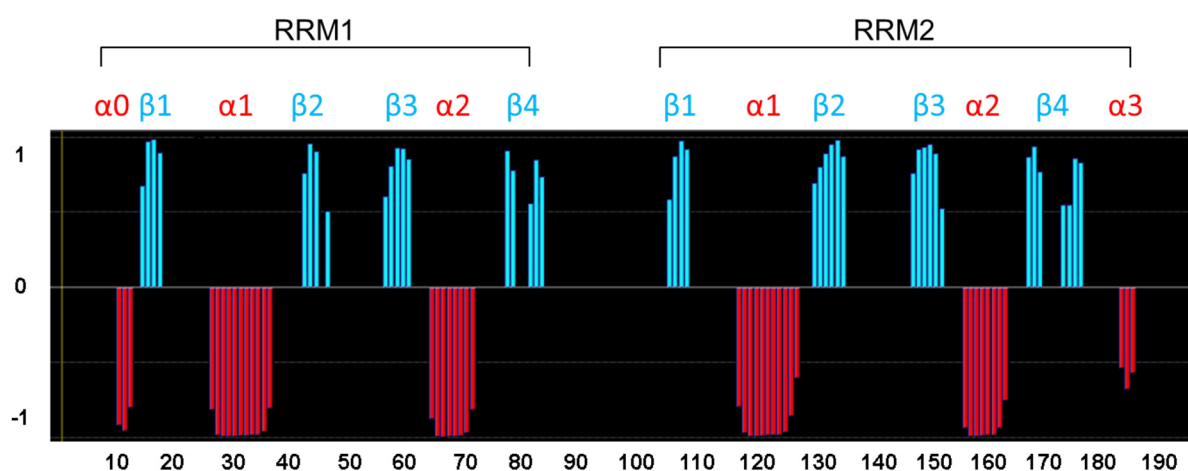


Figure 4.3 Secondary structure prediction of UP1 based on backbone chemical shifts using TALOS+. Negative and positive values represent α -helices and β -strands, respectively. A region of high helical propensity is present at the C-terminus of RRM2 which is absent in the crystal structure of UP1 in the free form.

4.1.3 NMR relaxation analysis of hnRNP A1 constructs

To study the backbone dynamics of RRM1, RRM2 and UP1 we measured ^{15}N longitudinal (R_1) and transverse (R_2) relaxation rates as well as $\{^1\text{H}\} -^{15}\text{N}$ heteronuclear NOE. As expected, all three experiments show that the termini of the constructs are flexible. Interestingly, the linker connecting the two RRM domains of UP1 shows lower than average heteronuclear NOE values and R_2/R_1 ratios, indicating that it is dynamic. A stretch of residues in UP1, Val163, Ile164, Lys166 and Tyr167 (Gln165 was excluded from the analysis due to peak overlap) show increased R_2 values, indicative of microsecond-millisecond timescale dynamics.

The overall tumbling correlation time (τ_c) estimated from the R_2/R_1 ratio indicates that isolated RRM1 and RRM2 domains tumble in solution with a correlation time of ~ 8 ns. The estimated correlation times for RRM1, RRM2 agree well with theoretical values expected for a monomer protein with the respective number of residues (Daragan & Mayo, 1997). In the case of UP1, an average correlation time of ~ 13 ns was obtained which is consistent with a ~ 22 kDa monomer protein and suggests that the two domains tumble together as one unit in solution (Figure 4.4).

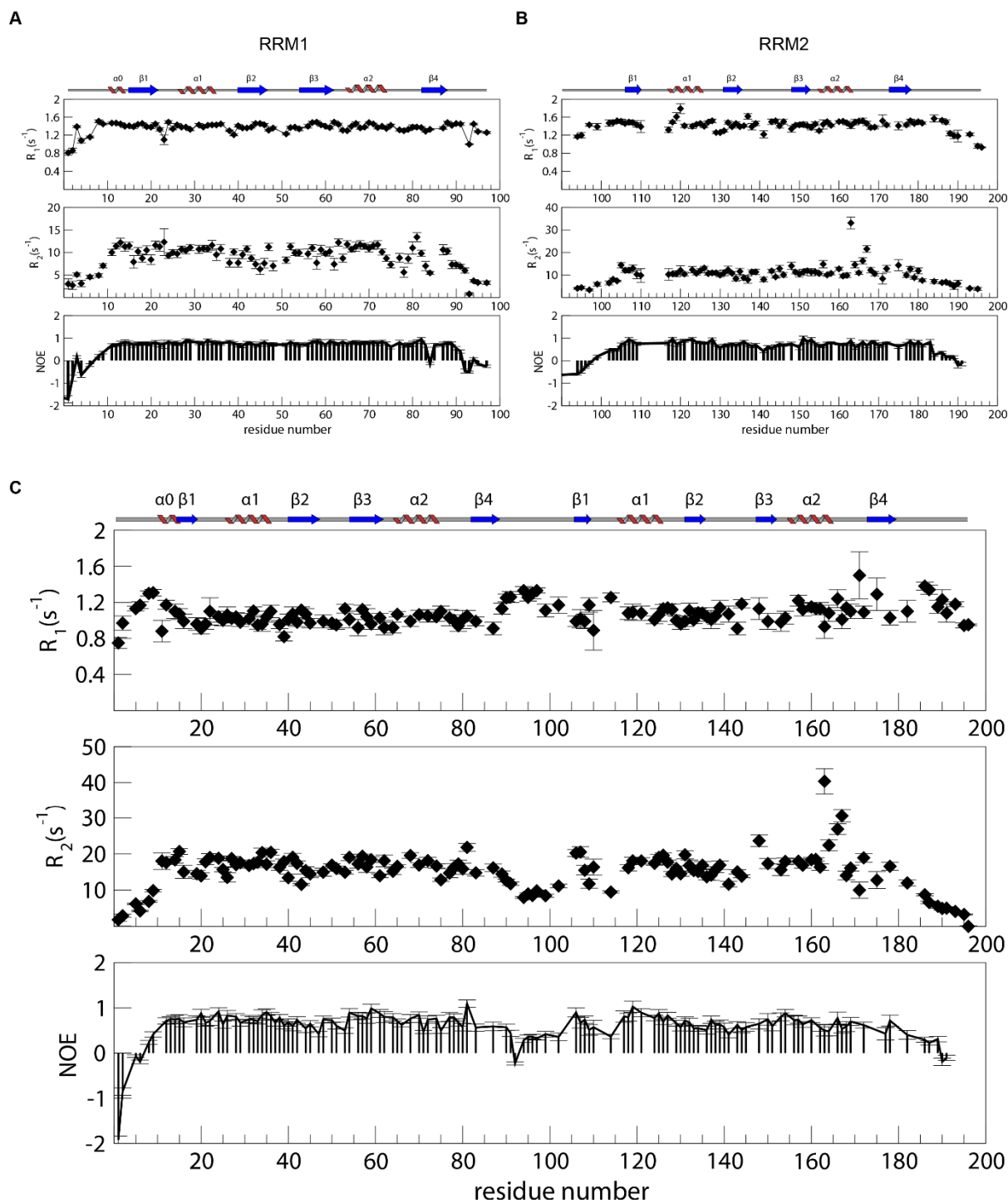


Figure 4.4 Relaxation analysis of (A) RRM1, (B) RRM2 and (C) UP1. ^{15}N R_1 , R_2 (derived from $R_{1\rho}$) and heteronuclear $\{^1H\}$ - ^{15}N NOE values are plotted against the residue number. Measurements were performed at 298 K and 600 MHz proton Larmor frequency. The secondary structure elements are shown above the graphs. Gaps in the data are due to overlapped or unassigned peaks.

4.2 NMR characterization of pri-miR-18a

4.2.1 Constructs of pri-miR-18a

I designed several constructs of human pri-miR-18a based on the published secondary structure and footprint data which have mapped the binding site of hnRNP A1 on the RNA (Michlewski et al, 2008). This corresponds to two hnRNP A1 binding sites in pri-miR-18a, i.e the terminal loop and the lower stem. Three loop constructs used in our studies include a 7-mer oligonucleotide and two 17-mer hairpin constructs. In addition, for the stem part of the pri-miR-18a I designed a construct (stem 18a) in which the two opposite strands are capped with a stable GNRA (N is any ribonucleotide and R is a purine) tetraloop (Figure 4.5).

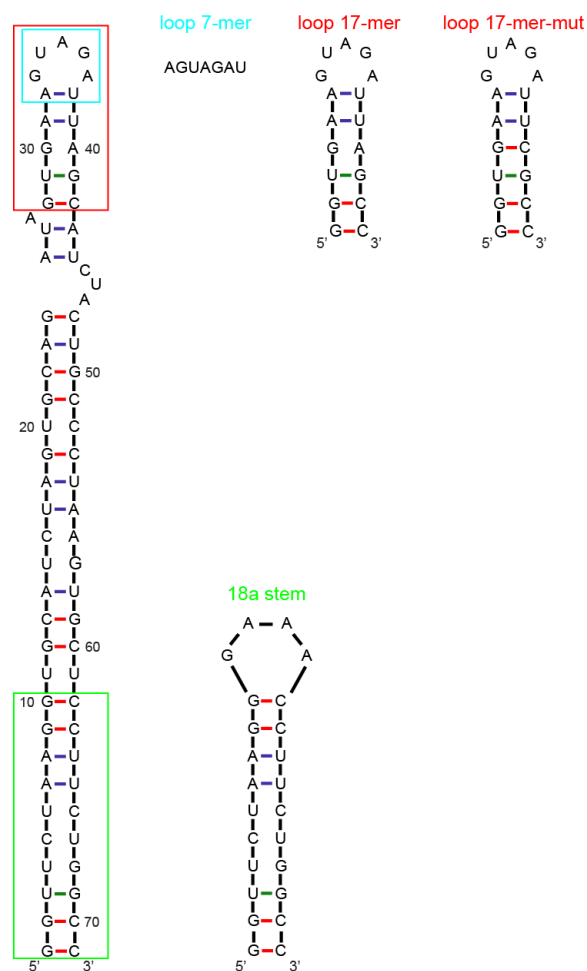


Figure 4.5 RNA constructs used in this study. The secondary structure of full-length pri-miR-18a is based on chemical probing (Michlewski et al, 2008). The stem 18a construct carries an artificial GAAA tetraloop which connects the two opposite strands. With the exception of 7-mer, all constructs contain an additional GC base pair for increased *in vitro* transcription yield and stability.

The full-length pri-miR-18a and stem 18a RNA samples were produced by *in vitro* transcription (as described in chapter 3). As the yield of *in vitro* transcription is sequence-dependent, reaction conditions were optimized for each construct. To this end, small-scale transcriptions were performed in various magnesium and NTPs concentrations and the transcription mixture was analyzed by denaturing polyacrylamide gel electrophoresis (PAGE) (Figure 4.6). For pri-miR-18a the optimum yield was found to be 40 mM Mg²⁺ and 32 mM NTPs.

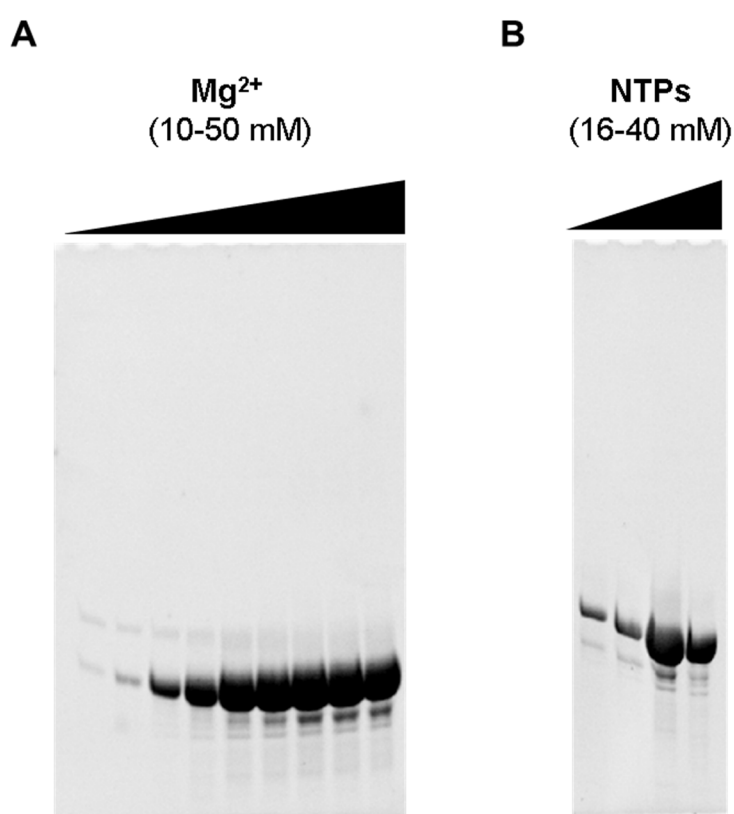


Figure 4.6 (A) Denaturing 12% polyacrylamide gel showing test *in vitro* transcription of pri-miR-18a in varying concentrations of magnesium. Lanes from left to right include 10, 15, 20, 25, 30, 35, 40, 45 and 50 mM Mg²⁺. (B) Denaturing 14% polyacrylamide gel showing test *in vitro* transcription of pri-miR-18a in varying concentrations of NTPs. Lanes from left to right include 16, 24, 32 and 40 mM NTPs.

4.2.2 Loop 18a RNA

The structure of loop RNA fragments was characterized by NMR spectroscopy. The 17-mer loop construct was assigned based on imino 2D-NOESY spectra recorded at 278 K and further confirmed by natural abundance ¹⁵N SOFAST-HMQC (Schanda & Brutscher, 2005) (Figure 4.7). Only four imino peaks were observed which were

assigned to the only G:U and the first two G:C base-pairs in the construct. This indicates that the two predicted A:U base-pairs do not form or only form transiently. This is not surprising as the A:U pairs are flanked by a G:A mismatch and a pentaloop with the sequence GUAGA. Taken together, these data suggest that the upper part of the loop 17-mer hairpin is dynamic and does not form stable base-pairs.

In order to stabilize the 17-mer construct, we designed a second loop construct (referred to as 17-mer-mut) in which the Adenine in the G:A mismatch is mutated to a Cytosine. In contrast to the 17-mer construct, the imino 2D-NOESY spectrum of 17-mer-mut shows three additional imino cross-peaks which can be assigned to the introduced G:C base pair and the two A:U base pairs. Thus, all the expected base-pairs are formed in the 17-mer-mut construct (Figure 4.8).

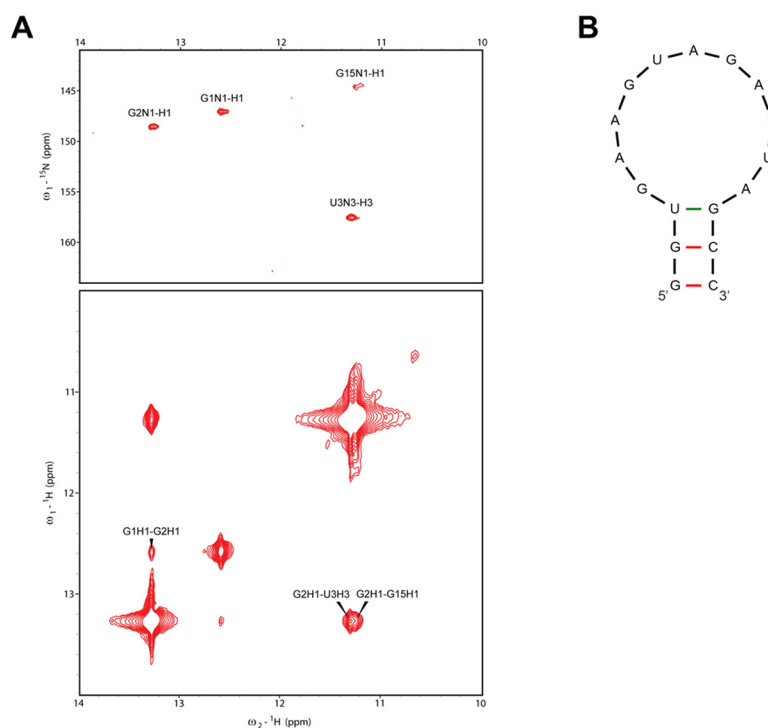


Figure 4.7 (A) Imino region of natural abundance ^1H , ^{15}N SOFAST-HMQC (top) and ^1H , ^1H NOESY (bottom) spectra of loop 17-mer RNA. The assignment for observed cross-peaks is indicated. (B) Secondary structure of loop 17-mer RNA based on the NMR data. The terminal G:C base pair is not part of the natural sequence.

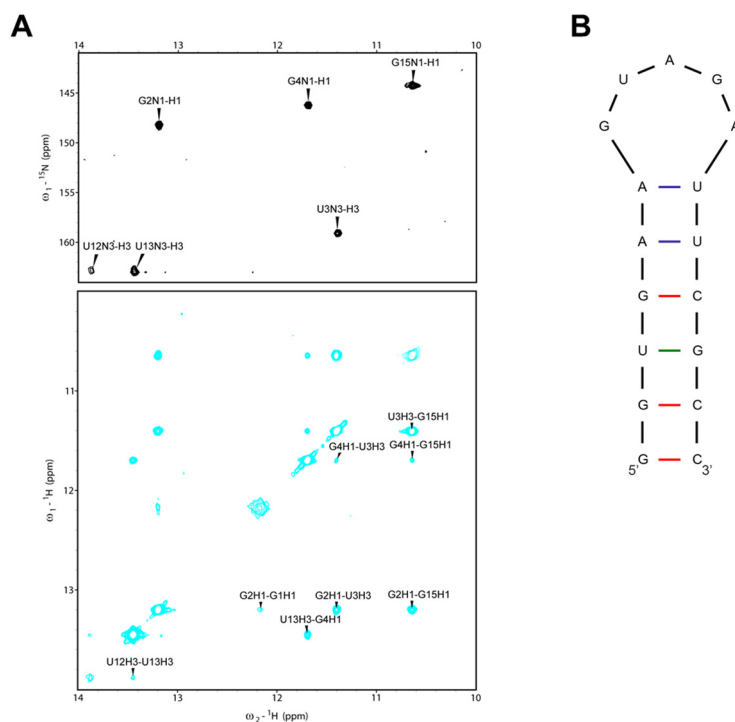


Figure 4.8 (A) Imino region of natural abundance ^1H , ^{15}N SOFAST-HMQC (top) and ^1H , ^1H NOESY (bottom) spectra of loop 17-mer-mut RNA. (B) Secondary structure of loop 17-mer-mut RNA based on the NMR data. The first G:C base pair is not part of the natural sequence.

4.2.3 Stem 18a RNA

I produced unlabeled and $^{13}\text{C}/^{15}\text{N}$ -labeled 18a stem sample for NMR studies. Typical yields were 500-600 nanomoles from a 10 ml *in vitro* transcription reaction. Initial NMR spectra recorded in water showed low thermal stability and inconsistencies, therefore all subsequent experiments were recorded on samples dissolved in phosphate buffer containing sodium chloride. The NOE-correlations from an imino 2D-NOESY experiment indicate the presence of three G:C, one A:U and one U:G base pairs (Figure 4.9). They correspond to base pairs adjacent to the GAAA tetraloop and two base pairs following the terminal G:C. However, from these experiments no information could be obtained regarding the middle part of the RNA which harbors a tandem U:C mismatch.

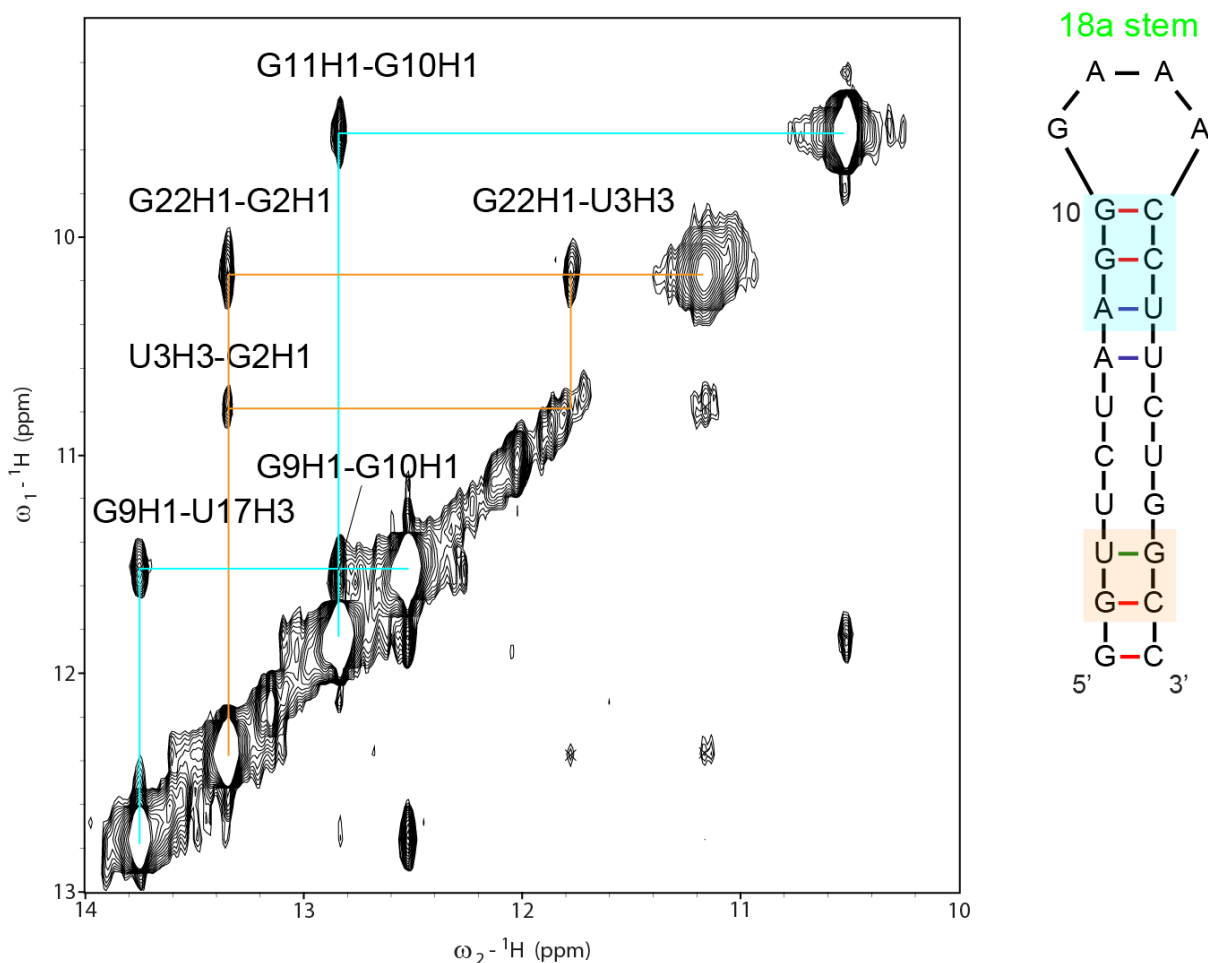


Figure 4.9 Imino region of a 2D NOESY (200 milliseconds mixing time) recorded on 18a stem. The NOESY walk for assignment of imino-imino cross-peaks is shown in cyan and orange for two separate regions of 18a stem. The secondary structure of 18a stem is shown on the right, with the two assigned regions highlighted. The experiment was recorded at 278 K and at 750 MHz proton Larmor frequency.

Imino protons can undergo chemical and conformational exchange, especially in highly dynamic regions such as the middle-part of 18a stem, therefore hampering their detection. To overcome these problems, a new NMR pulse sequence has been developed which detects the non-exchangeable H5 proton of pyrimidines. This experiment, referred to as pyrimidine (Py) H(CC)NN-COSY, is based on out-and-back magnetization transfer from H5 of pyrimidines to C5 to C4 to N3 and across the hydrogen bond (Figure 4.10) (Dallmann et al, 2013). This pulse sequence allows the detection of any base pair involving a pyrimidine and an N-H...N hydrogen bond. The main advantage of the new Py H(CC)NN-COSY compared to the standard

HNN-COSY experiment (Dingley & Grzesiek, 1998) is that it allows the detection of dynamic and transient base pairs in RNA molecules even at higher temperatures. We recorded the (Py) H(CC)NN-COSY experiment on 18a stem RNA at 298 K. As shown in Figure 4.10, all Watson-Crick (4 G:C and 2 A:U) base pairs are observed in the new Py H(CC)NN-COSY experiment, including the terminal G:C base pair. Importantly, in this experiment we could detect a A:U base-pair (between A7 and U18) adjacent to the U:C mismatch. In contrast, only three non-terminal G:C base pairs are observed in the standard HNN-COSY experiment recorded at the same temperature (Figure 4.10).

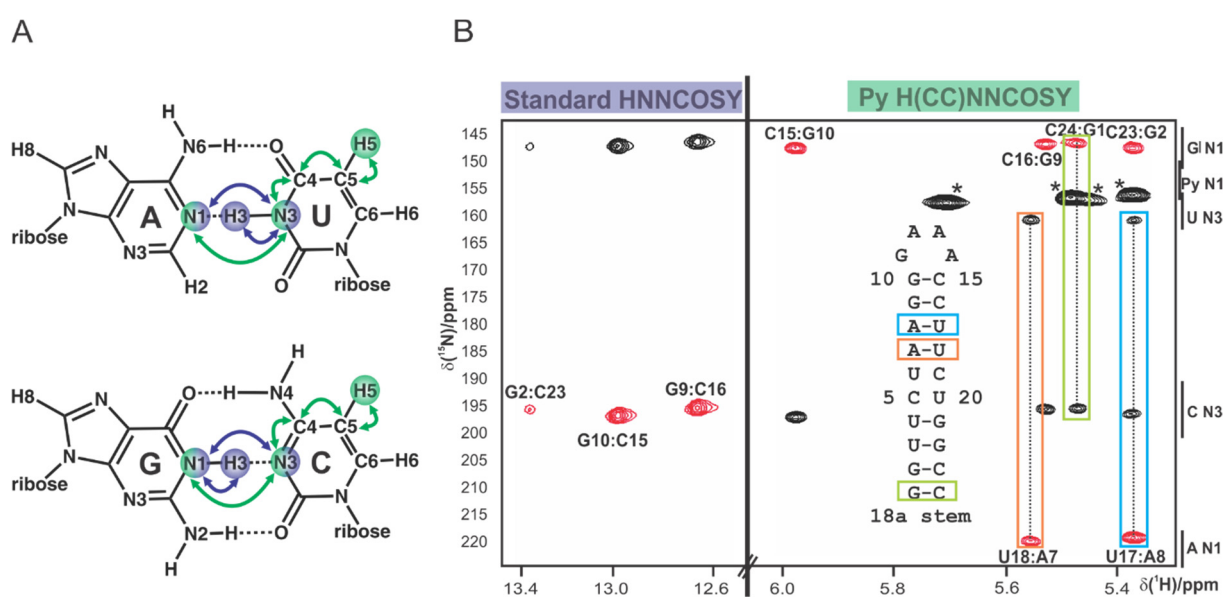


Figure 4.10 (A) Magnetization transfer steps for standard HNN-COSY (purple) and the Py H(CC)NN-COSY (green) experiments illustrated for A:U and G:C base pairs. (B) Imino region of the standard HNN-COSY and the H5 region of the Py H(CC)NN-COSY recorded for 18a stem RNA in H_2O and D_2O , respectively. Base pairs only observed in the Py H(CC)NN-COSY experiment are indicated by a box. As G:U base pairs lack N-H...N-type hydrogen bonds they are not detected. Both experiments were recorded for ~3h at 298 K. Peaks at ~157 ppm (marked with asterix) are intraresidual correlations to the N1 atom of pyrimidine. The secondary structure of 18a stem is shown. (Adapted from Dallmann et al, 2013)

Several spectra for complete assignment of 18a stem have been recorded and structural investigation of this construct is currently in progress.

4.3 NMR titration experiments

The interaction of hnRNP A1 constructs with the RNAs derived from pri-miR-18a was characterized by NMR titration experiments. First, the interaction of the 7-mer RNA (with the sequence 5'-AGUAGAU-3') with RRM1 and RRM2 was investigated. Increasing concentrations of the 7-mer RNA was added to ¹⁵N-labeled protein and the changes in the spectra upon addition of RNA were monitored by ¹H, ¹⁵N HSQC spectra. As shown in Figure 4.11, addition of 7-mer RNA causes extensive chemical shift perturbations (CSP) in RRM1 and RRM2, indicating that the individual domains are able to interact with the 7-mer RNA. For both RRM1 and RRM2, binding is in the intermediate-fast exchange regime on the NMR timescale.

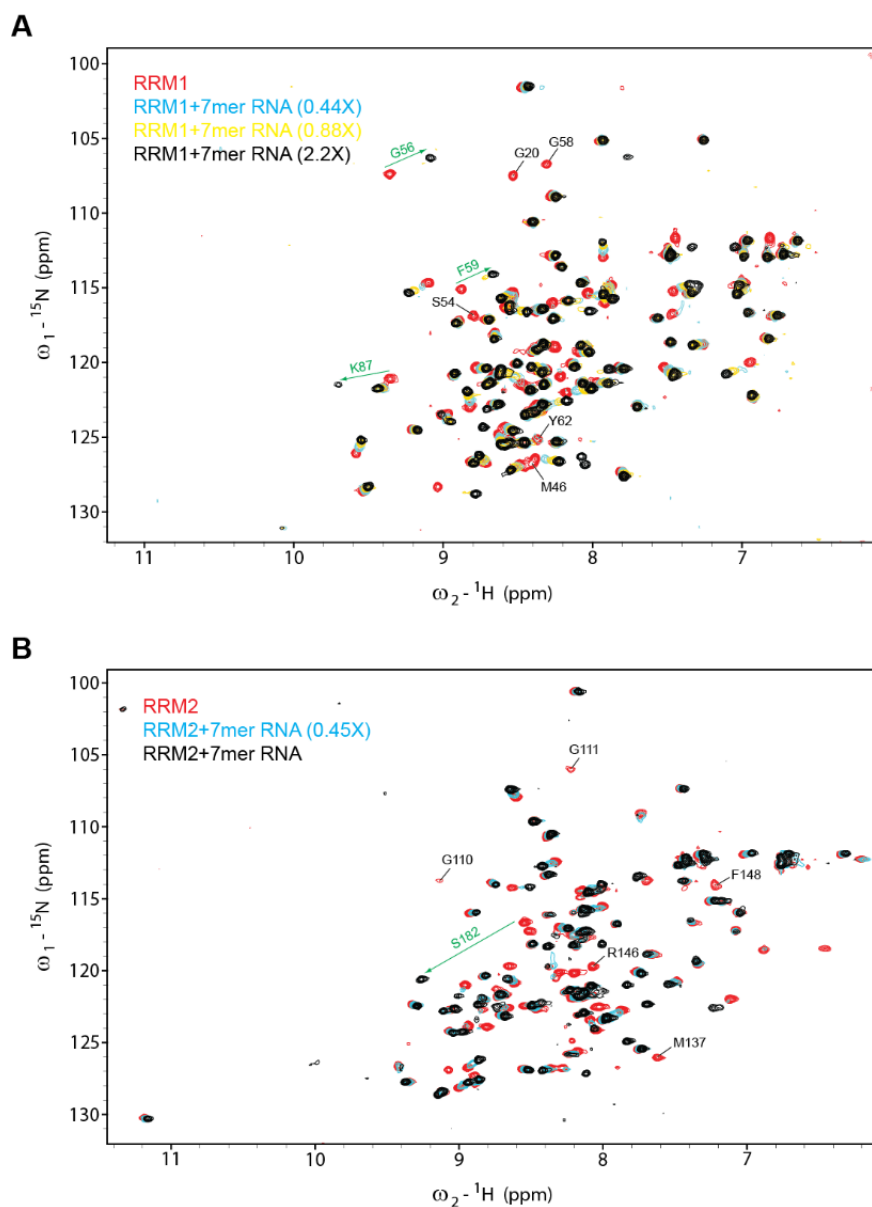


Figure 4.11 NMR titration experiments of RRM1 (A) and RRM2 (B) with 7-mer RNA. Overlay of ^1H , ^{15}N HSQC spectra of free protein (red) and in the presence of increasing concentrations of 7-mer RNA. Selected residues undergoing chemical shift changes and line-broadening upon RNA binding are labeled in green and black, respectively. Both experiments were performed at 298 K on a 750 MHz spectrometer.

Due to intermediate exchange binding of some residues, it was not possible to follow the movement of all peaks during the course of titration, therefore the chemical shifts of the spectrum in the bound form had to be assigned. In the case of RRM2, roughly ten peaks disappear due to line-broadening upon addition of RNA. Among these, six belong to residues in conserved RNP-1 (R146-G147-F148-A149) and RNP-2 (G110-

G111) motifs. For a better representation of the residues affected by RNA binding, the changes in chemical shift are plotted against the residue number (Figure 4.12).

The CSP plot indicates that RNP-1 and RNP-2 residues in both RRM domains are involved in RNA binding. In addition, the C-termini of RRM1 (residues 87-96) and RRM2 (residues 178-190) show large chemical shift perturbations. This is consistent with the crystal structure of UP1 bound to DNA, where it was shown that the inter-RRM linker and the C-terminus of RRM2 make contacts to DNA and becomes ordered (Ding et al, 1999). Specifically, the crystal structure showed hydrogen bonding of backbone amide groups of Leu181 and K183 with two nucleotides in the DNA. In agreement with the crystal structure, we observed large down-field shift for the amides of Leu181, Ser182 and Lys183 in our NMR titration experiments (Figure 4.11). Moreover, comparison of $^{13}\text{C}\alpha$ chemical shifts of RRM2 in the free and RNA-bound form suggests that residues 183-190 form an α -helix (α_3) upon RNA binding (data not shown), consistent with the α -helix observed in the UP1-DNA crystal structure (Ding et al, 1999). As described previously, this region of the protein shows some helical propensity even in the absence of RNA (see section 4.1.2).

Similar to NMR titrations of single domains described above, titration of UP1 with 7-mer RNA causes significant chemical shift changes in residues from both RRM1 and RRM2 (Figure 4.12). In contrast to single domains, saturation of binding is reached with at least two-fold excess of RNA to protein, suggesting that each RRM domain of UP1 binds to one molecule of 7-mer RNA. Many peaks show intermediate-exchange binding, for which the change in chemical shift could not be followed in the titration experiment. Approximately thirty residues in UP1 in complex with 7-mer RNA could not be assigned due to missing peaks (line-broadening) or overlap. The missing assignments are almost exclusively from RRM2 and the inter-RRM linker. They include residues in α_2 at the RRM1-RRM2 interface, in addition to known residues on the β -sheets (as observed for isolated RRM2).

Overall, the chemical shift perturbation pattern obtained for UP1 is very similar to the pattern obtained for individual RRM domains, suggesting that the recognition of 7-mer RNA in isolated RRM domains and in the context of UP1 is highly similar (Figure 4.12).

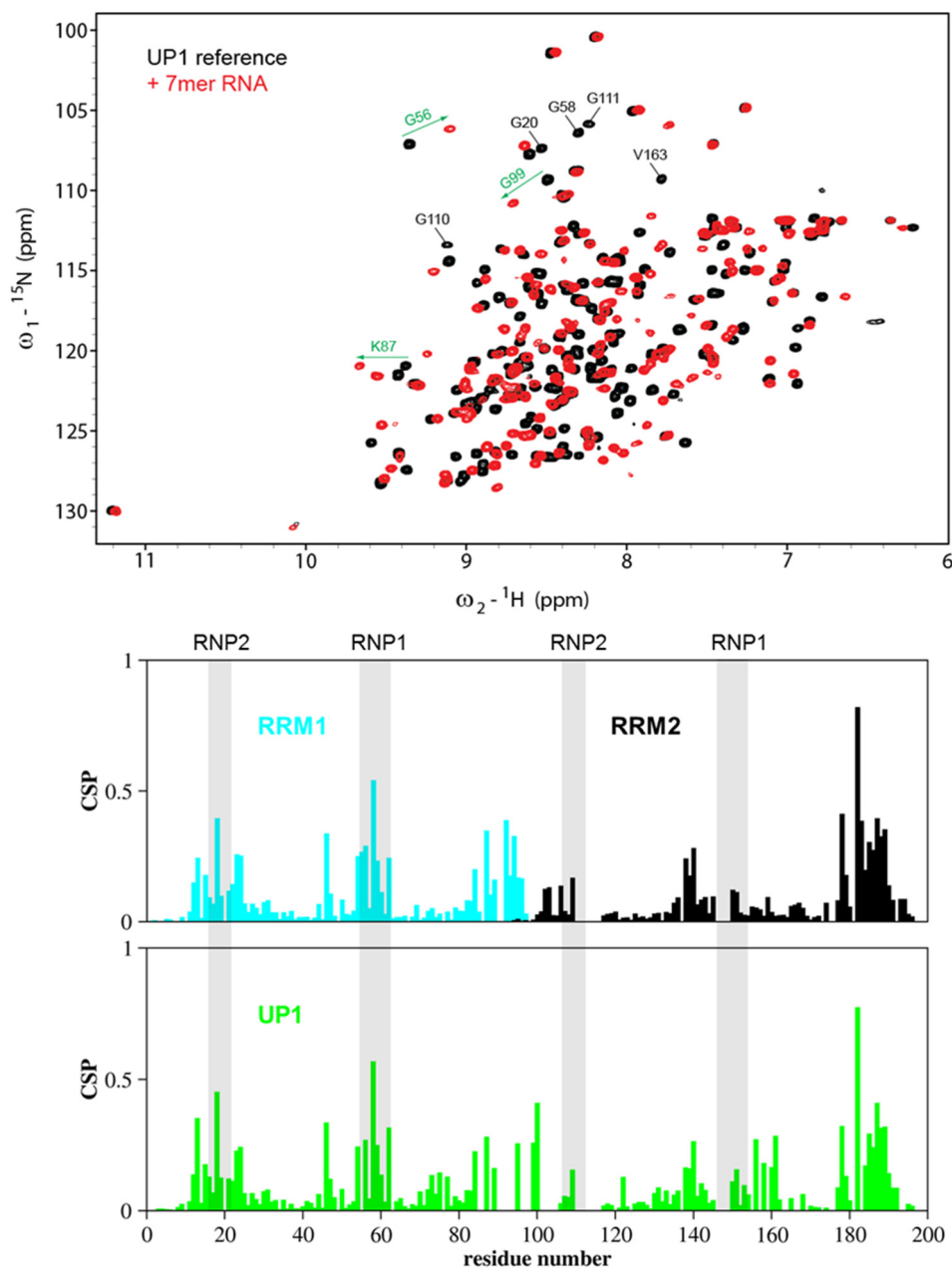


Figure 4.12 (A) NMR titration experiments of UP1 with 7-mer RNA. Overlay of ^1H , ^{15}N HSQC spectra of UP1 in the absence of RNA (black) and in the presence of 4-fold excess of 7-mer RNA (red). Selected residues undergoing chemical shift changes and line-broadening upon RNA binding are labeled in green and black, respectively. (B) Combined ^1H and ^{15}N CSP upon 7-mer RNA binding for RRM1, RRM2 and UP1 plotted against the residue number. CSP for each residue is calculated as $\sqrt{\delta\text{H}(\text{ppm})^2 + 0.1 \times \delta\text{N}(\text{ppm})^2}$. The gaps in the graph are arising from prolines or residues which could not be assigned due to line-broadening. Regions corresponding to conserved RNP motifs are highlighted in grey.

Next, I investigated the binding of UP1 to 17-mer loop RNA by NMR. In contrast to the UP1-7-mer interaction, saturation of binding was reached at 1:1 RNA to protein ratio. A large number of peaks showed line-broadening due to exchange. However, the quality of the spectra were improved by increasing the temperature from 298 K to 308 K and hence shifting the binding kinetics from intermediate to fast exchange (Figure 4.13). A large number of peaks arising mostly from residues in RRM2, are exchange-broadened in the RNA-bound spectrum. The chemical shift perturbations mapped on the structure of UP1 are shown in Figure 4.13. Major perturbations are localized to the β -strands of both RRM domains as well as the short α -helices at the N-terminus (α_0) and C-terminus (α_3).

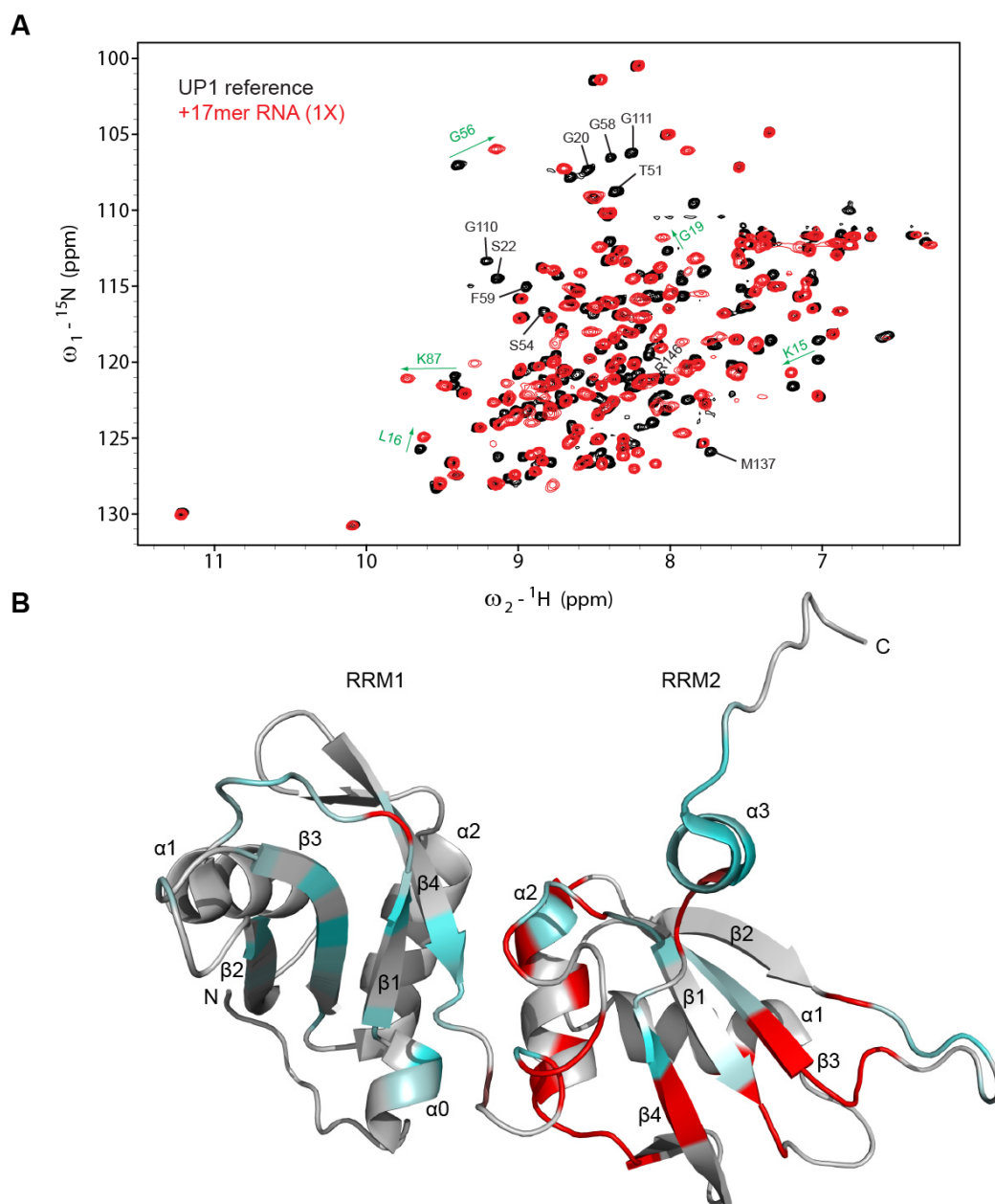


Figure 4.13 (A) NMR titration experiments of UP1 with 17-mer loop RNA. Overlay of ^1H , ^{15}N HSQC spectra of ^{15}N -labelled UP1 in the absence of RNA (black) and in the presence of 17-mer loop RNA (red). Selected residues undergoing chemical shift changes and line-broadening upon RNA binding are labeled in green and black, respectively. The experiment was performed at 600 MHz and at 308 K. (B) Chemical shift perturbations mapped on the ribbon structure of UP1 (gray to cyan, low to high perturbations). Residues corresponding to peaks which disappear upon RNA binding are shown in red.

4.4 ITC analysis of hnRNP A1–RNA interactions

To obtain binding parameters of hnRNP A1-18a interactions, I performed Isothermal Titration Calorimetry (ITC) experiments. A summary of thermodynamic parameters of hnRNPA1-RNA interactions is listed in Table 4.1.

For the interaction of RRM1 and RRM2 with the 7-mer RNA, a dissociation constant (K_D) of $20.40 \pm 1.06 \mu\text{M}$ and $6.85 \pm 0.13 \mu\text{M}$ was obtained, respectively. (Figure 4.14). The higher affinity of RRM2 is in agreement with the NMR titrations which showed mostly intermediate exchange binding. In the case of UP1-7-mer interaction, a K_D of $3.43 \pm 0.12 \mu\text{M}$ and a stoichiometry of 0.75 (protein:RNA) was obtained. Based on our NMR titrations (section 4.3), two 7-mer RNA molecules are required to saturate the binding sites of UP1. Therefore, assuming that both RRM domains of UP1 have the same binding affinity to RNA, a stoichiometry of 0.5 (protein:RNA) is expected. The deviation from the stoichiometry of 0.5 might reflect the slight difference in affinity between the two RRM domains of UP1.

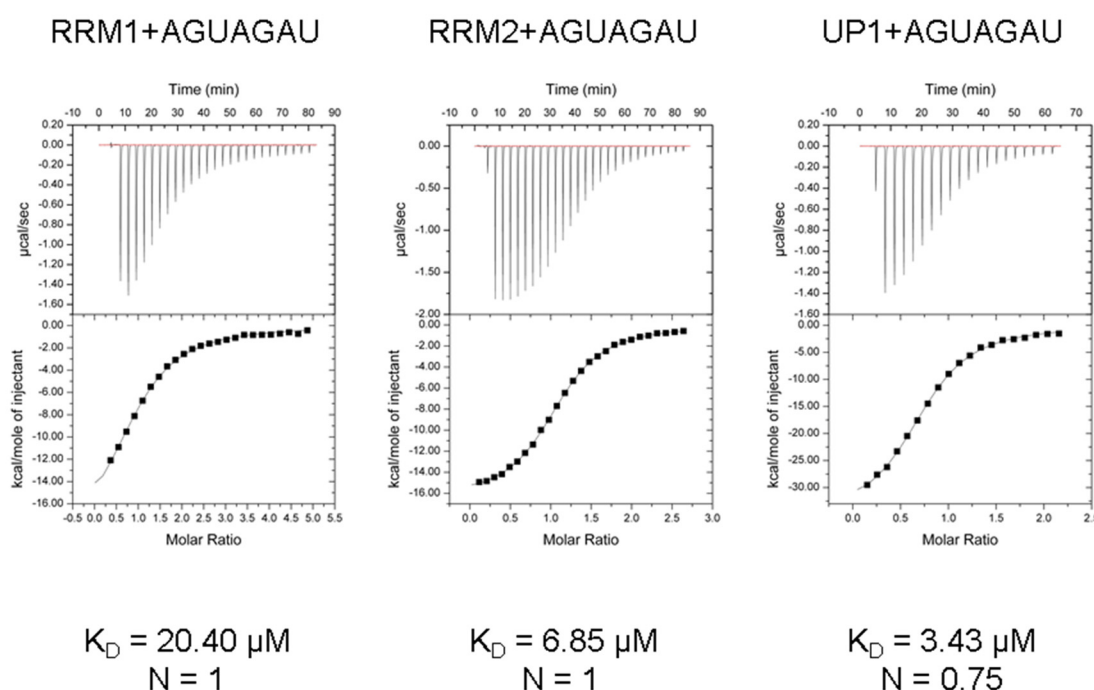


Figure 4.14 ITC binding curves of the interaction of RRM1, RRM2 and UP1 with the 7-mer RNA (5'-AGUAGAU-3'). Raw data (upper panel) were fitted to a one-site binding model. The dissociation constant (K_D) and stoichiometry (N) is indicated for each curve. A summary of thermodynamic parameters is provided in Table 4.1.

In contrast to the UP1+7-mer, the ITC binding curve of UP1+17-mer RNA shows a bimodal shape, suggesting the presence of an additional binding event (Figure 4.15). The UP1+17-mer binding curve could not be fitted using routine fitting procedures. One possible explanation for this observation is that upon UP1 binding, the 17-mer RNA undergoes partial melting therefore exposing a second binding site. Indeed, the 17-mer RNA contains two UAG motifs which are potential UP1 binding sequences and our NMR data suggest that one nucleotide of the second UAG motif in the 17-mer RNA is involved in base-pairing (see section 4.2.2).

To test the effect of stabilizing the stem-loop on UP1 binding, ITC experiments were conducted on the 17-mer-mut construct which carries an A→C mutation. Unlike the 17-mer, UP1 binding to the 17-mer-mut RNA resulted in a binding curve which could be fitted to a one-site binding model (Figure 4.15). Importantly, the K_D obtained for the UP1-17-mer-mut interaction is $3.14 \pm 0.19 \mu\text{M}$, almost identical to the value measured for the UP1-7-mer interaction. However, unlike the UP1-7-mer interaction, UP1 binds to 17-mer-mut with a stoichiometry of 1.

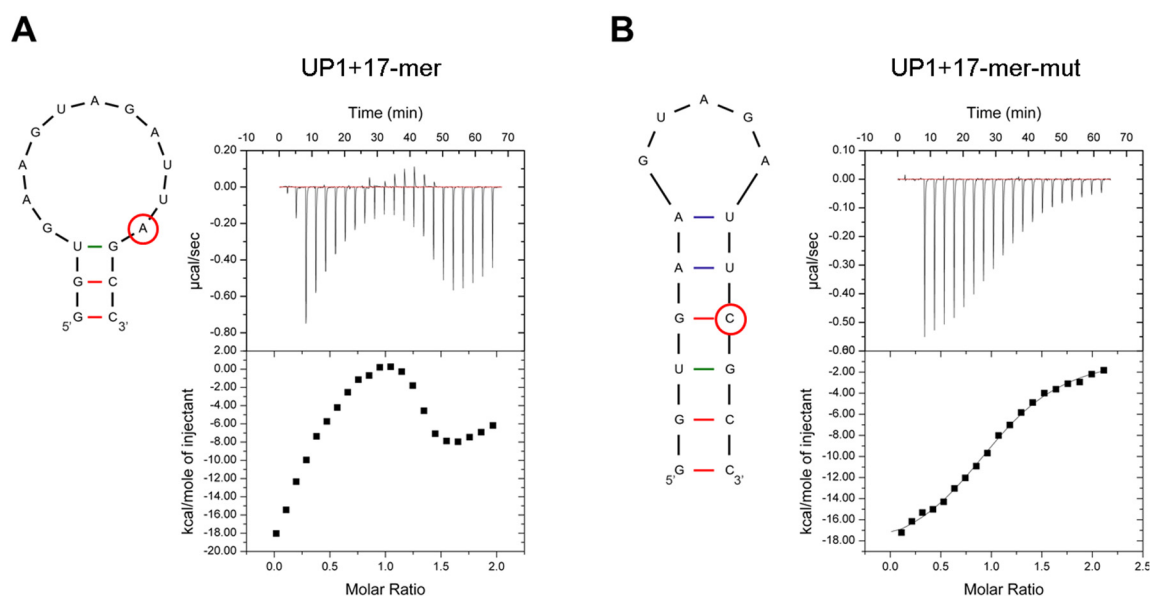


Figure 4.15 Comparison of ITC binding curves of UP1 to (A) 17-mer and (B) 17-mer-mut RNAs. The secondary structure of each RNA is shown. The red circles indicate the difference between the 17-mer and the 17-mer-mut constructs. The UP1-17-mer interaction cannot be explained by a simple one-site binding model.

I also characterized the binding of UP1 to full-length pri-miR-18a by ITC. In comparison to the UP1-7-mer interaction, a ~20-fold stronger binding ($K_D = 147 \pm 10$ nM) was observed (Figure 4.16). This suggests that the 7-mer RNA does not completely represent the binding of UP1 to pri-miR-18a.

Based on the fact that UP1 is known to bind single-stranded nucleic acids and in order to avoid problems associated with RNA secondary structure (as described for the 17-mer construct), we designed a 12-mer RNA construct corresponding to the loop of pri-miR-18a with the sequence 5'-AGUAGAUUAGCA-3'. The 12-mer RNA is predicted to be single-stranded and contains two UAG motifs. ITC analysis showed that the 12-mer RNA binds to UP1 with a high affinity ($K_D = 15.5 \pm 3.4$ nM) (Figure 4.16). Interestingly, the affinity of this construct is higher than the affinity of full-length pri-miR-18a ($K_D = 147 \pm 10$ nM), suggesting that the 12-mer RNA construct represents the main UP1-binding site of pri-miR-18a and that this is not fully accessible in full-length pri-miR-18a.

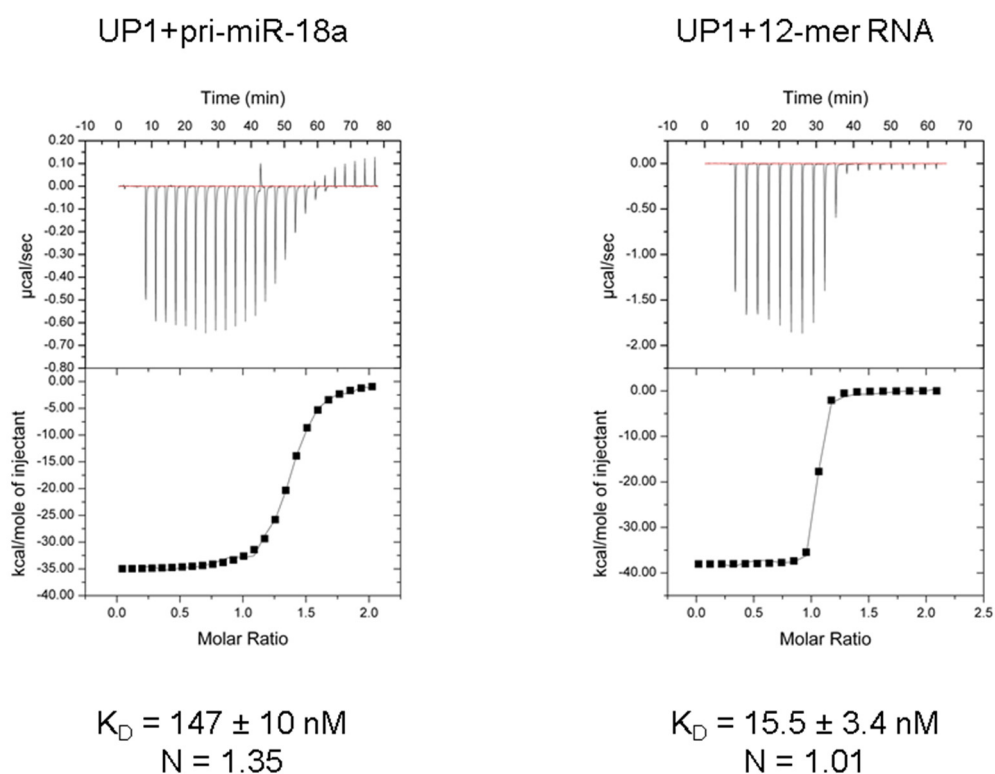


Figure 4.16 Comparison of ITC binding curves obtained for UP1-18a and UP1-12-mer interactions. The dissociation constant (K_D) and stoichiometry (N) is shown below each curve. In the case of pri-miR-18a, the deviation of the stoichiometry from 1 suggests unspecific binding of protein at higher protein to RNA ratios.

Table 4.1 Summary of thermodynamic parameters of hnRNP A1-RNA interactions

Binding reaction	N (protein:RNA)	K _D (μ M)	Δ H (cal/mol)	Δ S (cal/mol/deg)
RRM1 + 7-mer	1.01 \pm 0.02	20.40 \pm 1.06	-1.988 x 10 ⁴ \pm 646.1	-45.2
RRM2 + 7-mer	1.09 \pm 0.00	6.85 \pm 0.13	-1.645 x 10 ⁴ \pm 62.21	-31.6
UP1 + 7-mer	0.750 \pm 0.00	3.43 \pm 0.12	-3.530 x 10 ⁴ \pm 353.6	-93.4
UP1 + 17-mer	-	-	-	-
UP1 + 17-mer-mut	1.07 \pm 0.01	3.14 \pm 0.19	-1.979 x 10 ⁴ \pm 314.7	-41.2
UP1 + 18a	1.35 \pm 0.00	0.147 \pm 0.01	-3.522 x 10 ⁴ \pm 151.2	-86.8
UP1 + 12-mer	1.01 \pm 0.00	0.0155 \pm 0.0034	-3.809 x 10 ⁴ \pm 200.4	-92.0

4.5 EMSA and footprint analysis of hnRNP A1-pri-miR-18a interactions

To test the interaction of different fragments of UP1 and pri-miR-18a we performed electro-mobility shift assays (EMSA). Figure 4.17 shows that single RRM domains do not induce shift in the mobility of the loop as well as the stem RNAs. UP1, on the other hand, forms a stable complex with the loop RNA which results in a shift in the mobility of the bound RNA, whereas no binding to the stem RNA is observed even at higher protein to RNA ratios. Two conclusions can be drawn from these experiments. First, the two RRM domains of UP1 are required for high affinity RNA binding (they bind cooperatively). Second, UP1 binds specifically to the loop RNA and not to the stem RNA. Although no binding to the stem is observed in this assay, it is possible that the UP1-stem interaction is weaker and transient and hence not detectable under the conditions used in this assay.

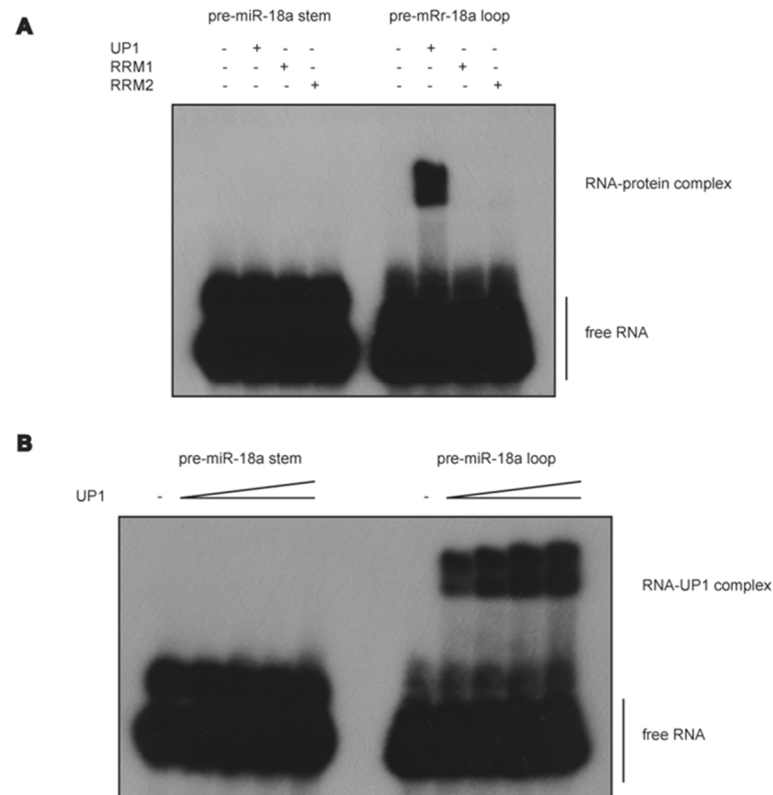


Figure 4.17 EMSA analysis of the interaction between hnRNP A1 and pri-miR-18a constructs. (A) No shift in mobility is observed upon addition of RRM1, RRM2 and UP1 to the stem RNA (left), while UP1 (and not single domains) causes a shift in the mobility of the loop RNA (right). (B) Even at higher protein to RNA ratios no binding is observed to the stem RNA. Position of free RNA and protein-RNA complexes is shown. Experiment was performed by Dr. Gracjan Michlewski (Edinburgh, UK).

In order to map the binding site of UP1 on pri-miR-18a, footprint analysis was performed using lead(II) ions and RNase T1. Lead(II) ions cleave single-stranded regions while RNase T1 cleaves specifically after guanosine residues. It is clearly seen that UP1 protects the region around G25-G31 which corresponds to the terminal loop (Figure 4.18). The cleavage pattern is similar to the reported footprint analysis using full-length hnRNP A1 (Michlewski et al, 2008). However, unlike full-length hnRNP A1 which was shown to bind to the lower stem region of pri-miR-18a, UP1 does not seem to bind efficiently to the lower part of the stem.

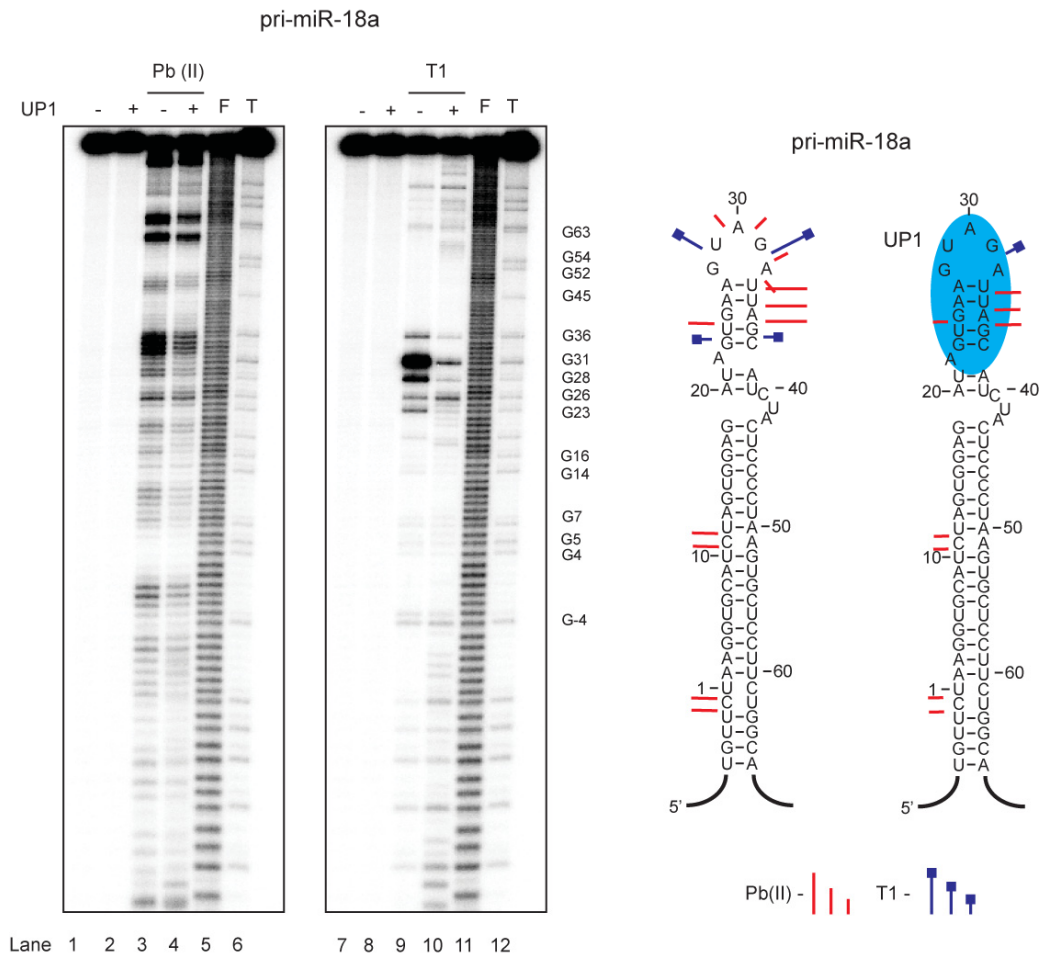


Figure 4.18 Footprint analysis of the interaction of UP1 with pri-miR-18a. Cleavage patterns were obtained for 5'- ³²P-labeled RNA incubated in the absence or presence of recombinant UP1 treated with Pb(II) ions or RNase T1. F and T correspond to partial digest with formamide (cleaves every nucleotide) or RNase T1 (cleaves after guanosine), respectively. Positions of selected guanosine residues are indicated. Proposed secondary structure of free and UP1-bound pri-miR-18a is shown on the right. Numbering is according to the Drosha cleavage site. Experiment was performed by Dr. Gracjan Michlewski (Edinburgh, UK).

Taken together, these data suggest that UP1 specifically recognizes the loop region of pri-miR-18a.

4.6 X-ray crystallography

4.6.1 X-ray crystallography on the UP1-17-mer complex

We aimed to solve the crystal structure of UP1 bound to RNA. Initially, UP1 in complex with 17-mer, 17-mer-mut and full-length pri-miR-18a were subjected to crystallization trials. Diffraction-quality crystals were obtained only for the UP1-17-mer complex. Initial crystals grew in several conditions containing ammonium formate, 2-Methyl-2,4-pentanediol (MPD) or sodium chloride as precipitant (Figure 4.19). The best crystals were obtained in an optimized MPD condition and diffracted to 2.5 Å resolution.

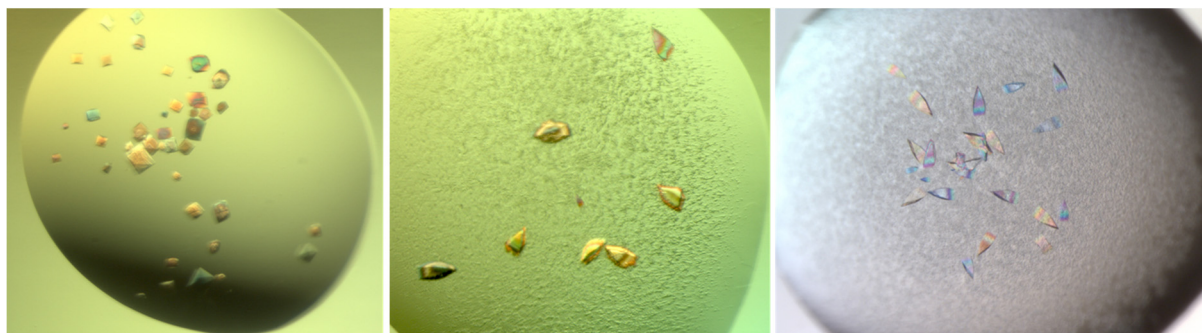


Figure 4.19 Crystals of UP1-17-mer RNA obtained in 3.6 M ammonium formate, 50 mM MES pH 6.0 (left); 30% MPD, 200 mM tri-sodium citrate, 100 mM HEPES pH 7.5 (middle). Optimized crystals were obtained in 25-30% MPD, 300 mM NaCl and 50 mM HEPES pH 7-7.5 (right). Crystals were grown at 20 °C (Courtesy of the MPI-Biochemistry crystallization facility)

Despite the fact that several structures of free UP1 are available in the database, we were not able to solve the phase problem by molecular replacement (MR). Different strategies for phasing were pursued, including seleno-methionine labeling of the protein, 5-bromouridine labeling of the RNA and heavy atom soaking with mercury, platinum, gold and lead (Table 4.2). However, despite extensive efforts, these conventional phasing approaches did not provide sufficient phase information. The only method which proved useful was soaking of the crystals with a Gadolinium compound. This allowed the structure to be solved using the single-wavelength anomalous diffraction (SAD) method (Table 4.3).

Table 4.2 Strategies used to solve the phase problem of the UP1-17-mer complex

Phasing strategy	Outcome
UP1 Se-Met labeling	No anomalous signal
UP1 (L33M, L124M) Se-Met labeling	No anomalous signal
BromoU-labeled RNA (U3)	Weak anomalous signal
BromoU-labeled RNA (U3, U12, U13)	No crystals
Heavy atom soaks (Hg, Pb, Pt, Au, ...)	No diffraction or no anomalous signal
Gadolinium soak	Strong anomalous signal

Table 4.3 Data collection and refinement statistics of the 17-mer duplex RNA structure

	Native	Gd
Data collection		
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	43.88, 101.03, 139.03	44.36, 102.14, 139.50
α , β , γ (°)	90, 90, 90	90, 90, 90
		<i>Peak</i>
Wavelength (Å)	0.99	1.71
Resolution (Å)	81.73 – 2.4 (2.53 – 2.4)*	47.96 – 2.66 (2.94 – 2.79)
R_{merge}	0.043 (1.286)	0.067 (0.805)
$I/\sigma I$	26 (1.3)	18.6 (1.5)
Completeness (%)	99.6 (98)	99.5 (99.5)
Redundancy	5.6 (5.6)	8.2 (6.0)
Refinement		
Resolution (Å)	30.0 – 2.47	
No. reflections	21743	
$R_{\text{work}}/R_{\text{free}}$	0.22/0.25	
No. atoms		
RNA	2920	
Water	19	
B-factors		
RNA	51.73	
Water	43.13	
R.m.s deviations		
Bond lengths (Å)	0.009	
Bond angles (°)	1.815	

*Highest resolution shell is shown in parenthesis.

The structure reveals four RNA duplexes in the asymmetric unit. Two of the duplexes are packed end to end forming a continuous helix with a length of ~ 100.8 Å, with the other two duplexes packing side by side against it (Figure 4.20). The majority of the nucleotides adopt the C3'-endo conformation, suggesting that the RNA duplexes have an A-form geometry. However, there are some differences among the four duplexes in terms of sugar pucker conformation and base-pairing.

Surprisingly, no extra electron density from the protein was observed in the crystal. The presence of protein in the UP1-17-mer crystals was confirmed by SDS-PAGE and also based on our observation that crystals obtained using selenomethionine-labeled UP1 showed characteristic selenium X-ray fluorescence spectrum (data not shown). This suggests that the crystal lattice is based on an RNA scaffold with the protein rather loosely attached to the lattice.

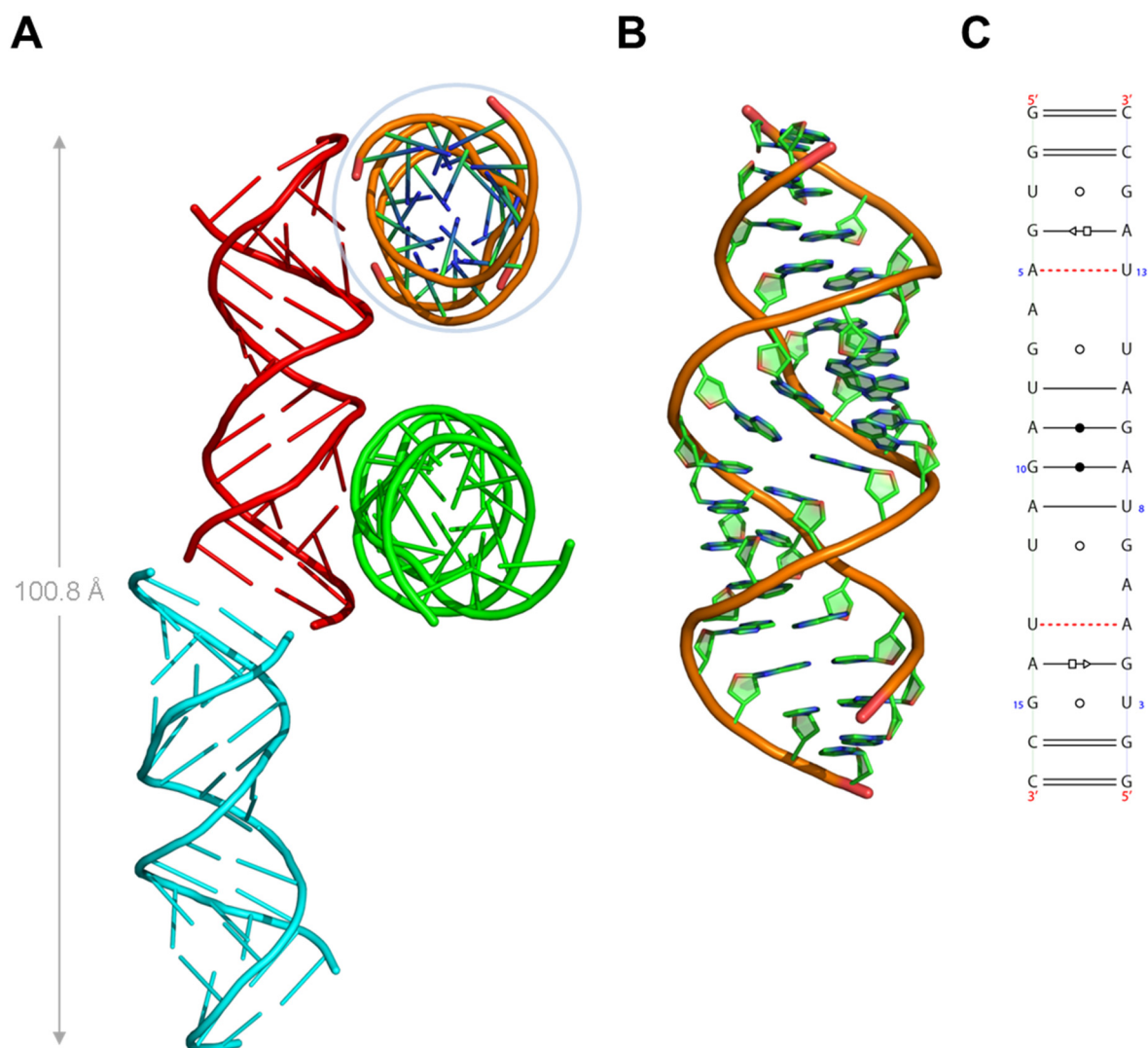


Figure 4.20 Crystal structure of the 17-mer duplex RNA. (A) Packing of four RNA duplexes in the asymmetric unit of the crystals with $P2_12_12_1$ space group. (B) Another view of the A-form RNA duplex encircled in panel A. (C) Secondary structure annotation of the RNA duplex according to Leontis and Westhof's convention (Leontis & Westhof, 2001).

4.6.2 Crystal structure of the UP1-12-mer complex

Because our initial crystallization attempts to crystallize UP1 in complex with stem-loop RNAs lead to RNA crystals, we decided to crystallize UP1 in complex with a single-stranded RNA construct. Crystals of UP1 in complex with the 12-mer RNA were obtained in several conditions containing polyethylene glycol (PEG) or ammonium sulfate as precipitant. Among the tested crystals, plate-like crystals

obtained in 0.2M sodium citrate and 20% PEG 3350 produced good quality diffraction data to 2.5 Å resolution.

We solved the UP1-12-mer structure by molecular replacement using the structure of UP1 bound to modified telomeric DNA (PDB code:1U1R) (Myers & Shamoo, 2004) as search model. The structure reveals two molecules of UP1 and two RNA chains in the asymmetric unit (Figure 4.21). The 5'-end of each RNA strand is recognized by RRM1, whereas the 3'-end binds to RRM2 from a second UP1 molecule. This specific arrangement with a 2:2 (protein:RNA) stoichiometry is similar to the UP1-DNA complex reported previously (Ding et al, 1999).

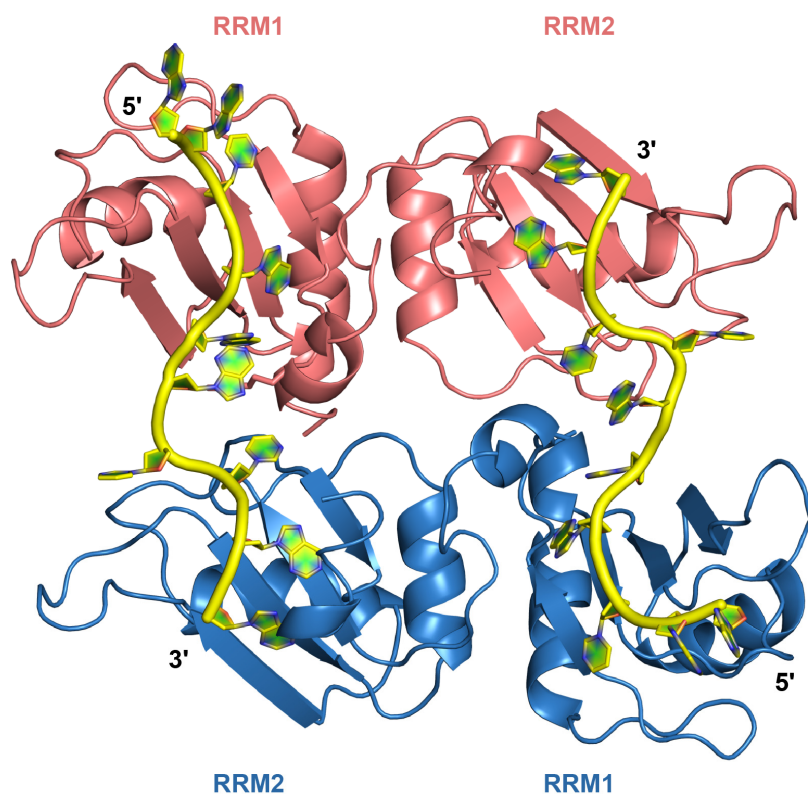


Figure 4.21 Crystal structure of the UP1-12-mer RNA complex at 2.5 Å resolution. The asymmetric unit consists of two UP1 molecules and two RNA strands. The 5'-end of each RNA strand binds to RRM1, whereas the 3'-end is recognized by RRM2. The two RNA strands (colored yellow) run anti-parallel to each other.

In this arrangement, the two UP1 molecules form a dimer mediated by Glu11 in α_0 , Asp94 in the inter-RRM linker and four residues from RRM2 (Ile164, Lys166, Tyr167 and His173) (Figure 4.22). The contacts at the dimer interface are the same as the ones described for the UP1-DNA complex and include hydrogen bonding (Glu11-His173 and Tyr167-Ile164) and charge interaction (Asp94-Lys166) (Ding et al, 1999).

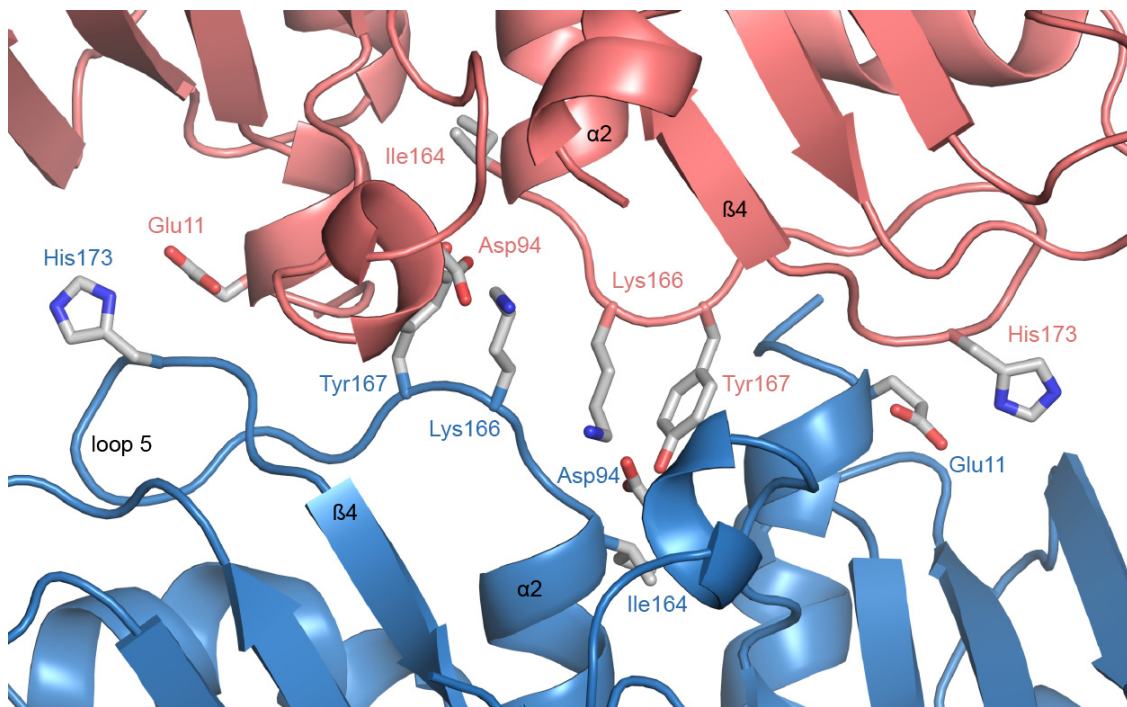


Figure 4.22 Protein-protein interactions at the dimer interface of the UP1-12-mer RNA complex. Side-chain of residues involved in interaction are indicated. The same arrangement is observed in the UP1-telomeric DNA complex (Ding et al, 1999)

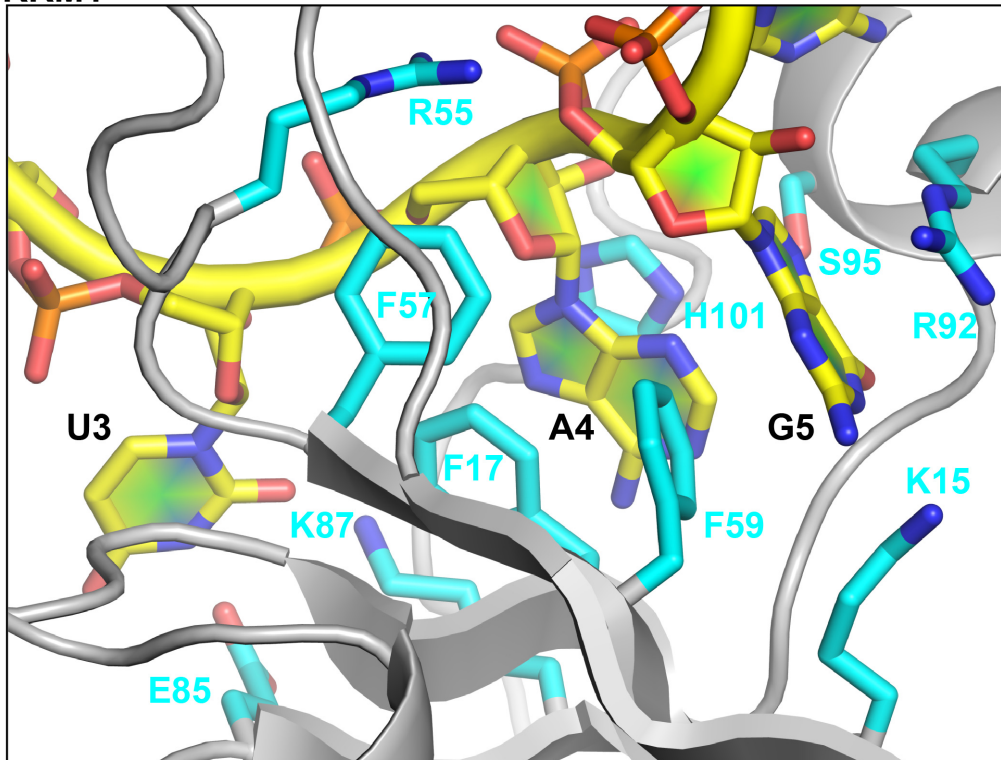
The protein-RNA contacts observed in the UP1-RNA complex are very similar for both RRM domains. Each RRM domain specifically recognizes one UAG motif by making contacts mainly through conserved residues on the β -sheets. In the case of RRM1, two conserved aromatic residues Phe17 (RNP-2 motif residue located on β_1) and Phe59 (RNP-1 motif residue located on β_3) are involved in stacking interactions with the bases of A4 and G5, respectively (Figure 4.23). A third aromatic residue, Phe57 (RNP-1 motif residue located on β_3) interacts with the ribose rings of A4 and G5. Similarly in RRM2, Phe108 (RNP-2 motif residue located on β_1) and Phe150 (RNP-1 motif residue located on β_3) stack with the bases of A9 and G10, while Phe148 (RNP-1 motif residue located on β_3) makes contacts with the sugar rings of

A9 and G10 (Figure 4.23). In each domain, a positively charged residue (Arg55 in RRM1 and Arg146 in RRM2) makes electrostatic interactions with the phosphate backbone of the AG dinucleotide. In addition, two charged residues in each domain (Glu85 and Lys87 in RRM1 and Glu176 and Arg178 in RRM2) make contacts to the Uridines in the UAG motifs (U3 and U8). Another charged residue, Lys15 in RRM1 and Lys106 in RRM2 interact with G5 and G10, respectively.

Importantly, the inter-RRM linker makes a number of interactions with the RNA. Side chains of Arg92 and Ser95 make contacts with G5 while the imidazole ring of His101 stacks with the base of A4. Residues in the C-terminus of RRM2 also contribute to RNA binding, for instance the side chain of Met186 interacts with A9 (Figure 4.23).

Overall, the mode of UAG recognition is highly similar to the TAG recognition described for the UP1-DNA complex (Ding et al, 1999).

RRM1



RRM2

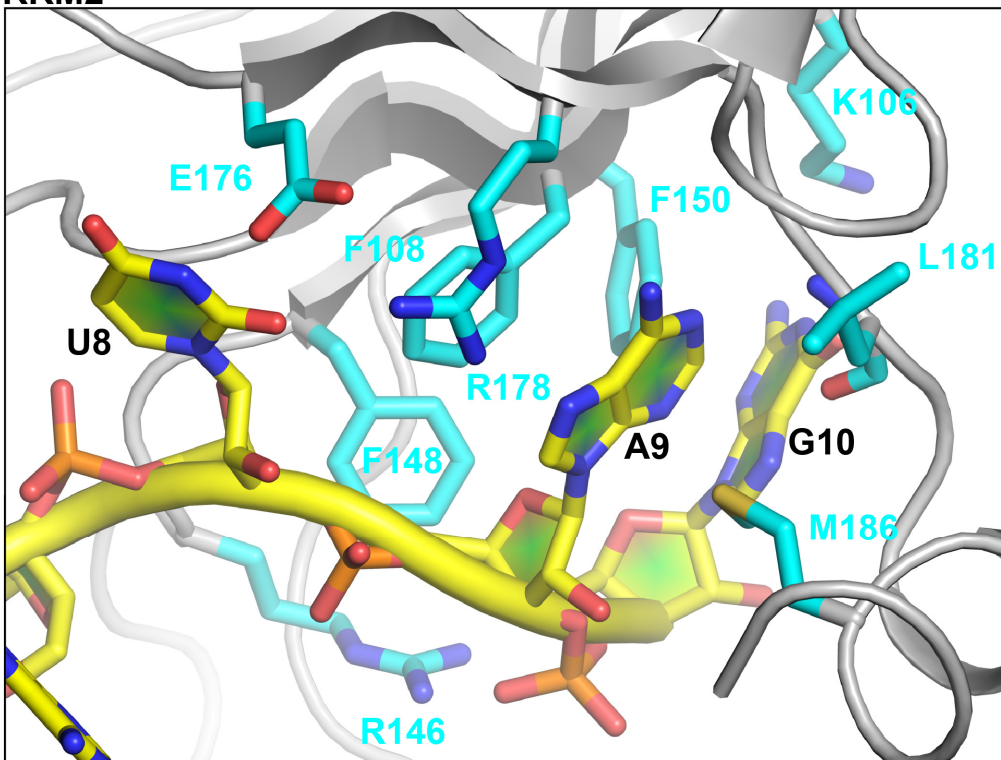


Figure 4.23 Structural details of the recognition of UAG by RRM1 and RRM2 domains of hnRNP A1. The RNA strand is shown in yellow and the nucleotides are labeled in black. Side chain of residues making contacts to RNA are shown as stick model and colored cyan. The backbone of L181 which interacts with RNA is also indicated.

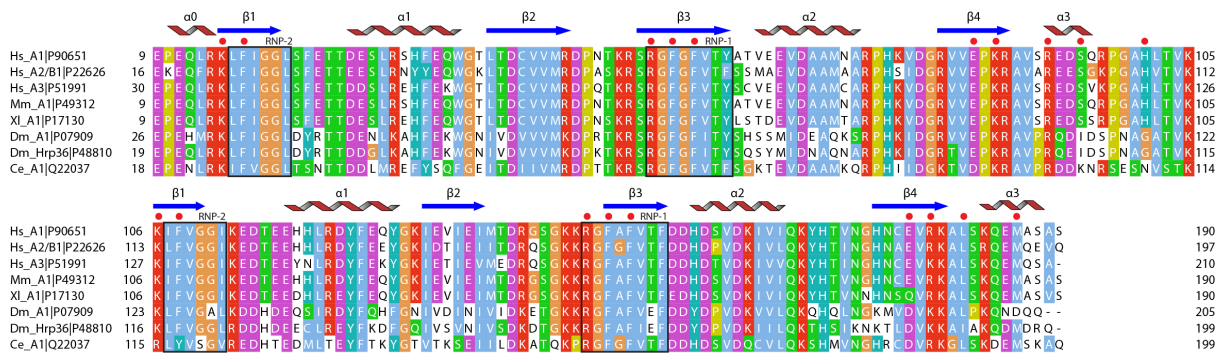


Figure 4.24 Multiple sequence alignment of eukaryotic hnRNP A1 homologs. Sequence alignment was done using MUSCLE (Edgar, 2004) and viewed by Jalview (Waterhouse et al, 2009). The sequence of the Gly-rich tail is not shown. Residues involved in interactions with the 12-mer RNA are marked by red circles. The UniProt accession code is indicated for each sequence. (Hs = Homo sapiens, Mm = Mus musculus, XI = Xenopus laevis, Dm = Drosophila melanogaster, Ce = Caenorhabditis elegans).

The structure of UP1 in the UP1-12-mer RNA complex is almost indistinguishable (backbone root mean square deviation of $\sim 0.5 \text{ \AA}$) from the UP1 structure in complex with DNA (Ding et al, 1999). However, superposition of the 12-mer RNA (5'-AGUAGAUUAGCA-3') and the 12-mer DNA (5'-TTAGGGTTAGGG-3') structures reveals significant differences (Figure 4.25). Whereas in the DNA structure two TAG motifs are connected by three nucleotides (G5-G6-T7), two nucleotides (A6-U7) separate the UAG motifs in the RNA structure. In addition, the main difference in terms of protein contacts is observed in this region where G5 in the DNA sequence is replaced by A6 in the RNA. In the DNA complex, G5 makes two hydrogen bonds with the side-chain of Asp42 in RRM1 $\beta 2$, whereas A6 in the RNA structure is engaged in contacts with Gln96 in the inter-RRM linker (Figure 4.25). Thus, RRM1 in the RNA structure directly binds three nucleotides (UAG), in contrast to the RRM1 in the DNA structure which recognizes four nucleotides (TAGG) (Ding et al, 1999). Consequently, the transition of the nucleic acid chain from one domain to the other is smoother in the UP1-RNA structure compared to the UP1-DNA structure.

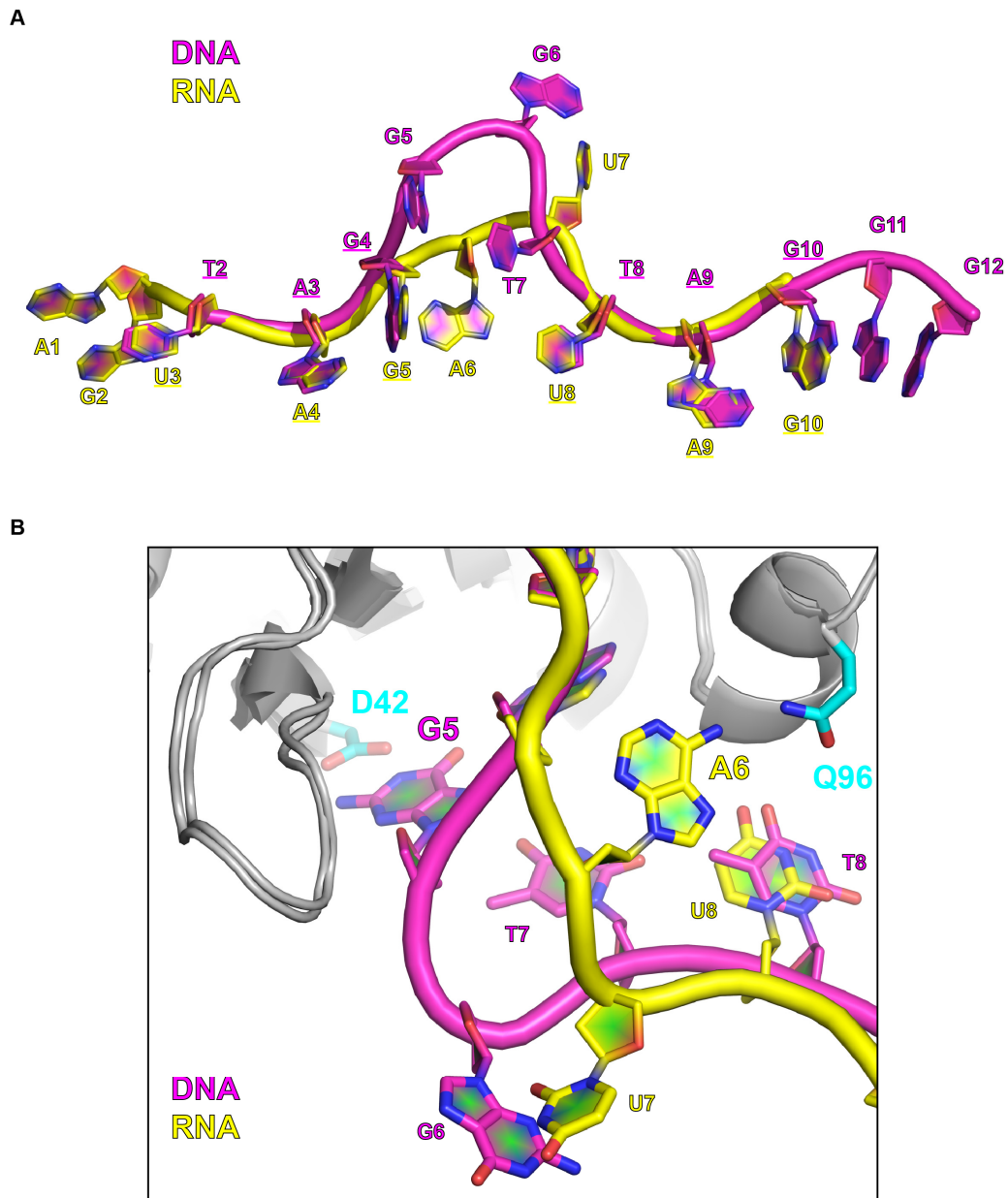


Figure 4.25 (A) Superposition of the structures of telomeric DNA (magenta) (PDB code: 2UP1) and 12-mer RNA derived from pri-miR-18a loop (yellow) both in complex with UP1. Nucleotides C11 and A12 in the RNA structure were not modeled due to lack of clear electron density. The two UAG (TAG in DNA) motifs which are recognized by the RRM domains of UP1 are highly similar (underlined). (B) Zoomed-in view showing the difference in nucleic acid recognition by UP1 in the two structures. Asp42 makes hydrogen-bond contacts to G5 in the DNA structure, whereas in the RNA structure A6 interacts with Gln96 in the inter-RRM linker.

4.7 Static light scattering (SLS)

To characterize the stoichiometry of UP1-RNA complexes in solution, we analyzed the individual components and complexes by multi-angle static light scattering. As seen in figure Figure 4.26, both UP1 and pri-miR-18a run as single species with a molecular weight which is very close to that of a monomer. Importantly, the molecular weight obtained for the UP1+12-mer complex corresponds to a monomer size of 22.4 kDa. This indicates that the stoichiometry of the UP1-12-mer complex in solution is in fact 1:1 (protein:RNA), suggesting that the 2:2 stoichiometry observed in the crystal structure is caused by crystal contacts. In addition, the molecular weight obtained for the UP1+18a complex is in agreement with the expected value for a 1:1 complex. Overall, these data clearly indicate that one UP1 molecule binds to full-length pri-miR-18a. A summary of the data obtained from light scattering is listed in Table 4.4.

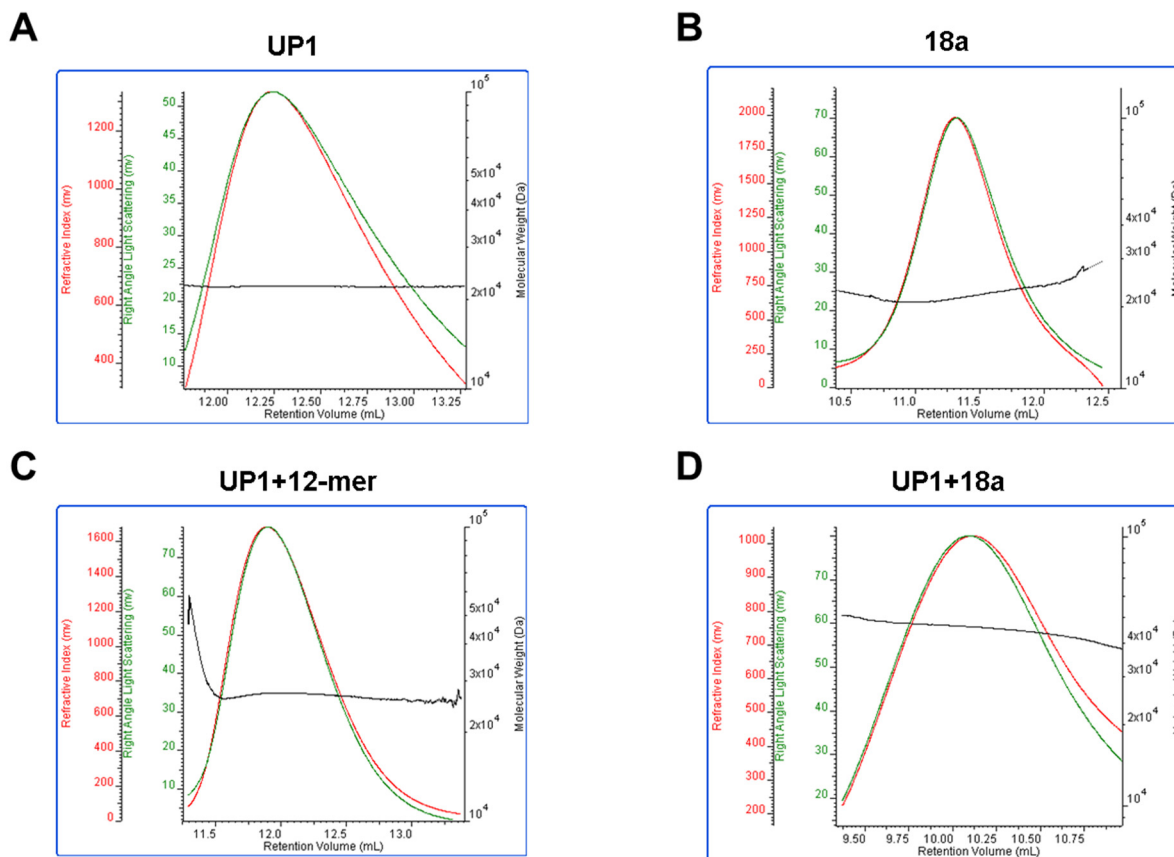


Figure 4.26 Static light scattering profile of (A) UP1, (B) pri-miR-18a, (C) UP1+12-mer RNA and (D) UP1+pri-miR-18a. Refractive Index (RI) and Right Angle Light Scattering (RALS) traces are indicated in red and green, respectively. The black line indicates the distribution of molecular weights along the elution profile.

Table 4.4 Theoretical and experimental molecular masses (derived from light-scattering) of UP1-RNA complexes

	Theoretical mass (kDa)	Measured mass (kDa)
UP1	22.4	22.1
UP1 + 12-mer RNA	26.3 (1:1 complex) 52.6 (2:2 complex)	26.9
18a RNA	22.7	22.6
UP1 + 18a RNA	45.1 (1:1 complex) 95.2 (2:2 complex)	45.2

4.8 The effect of UP1 salt-bridge mutations on RNA binding

In the crystal structure of the UP1-12-mer RNA complex, Arg75 and Arg88 make salt-bridge contacts to Asp155 and Asp157, respectively. These are highly conserved residues (Figure 2.6) and the two salt-bridges have been observed in all the reported structures of UP1 free and in complex with DNA (Figure 4.27).

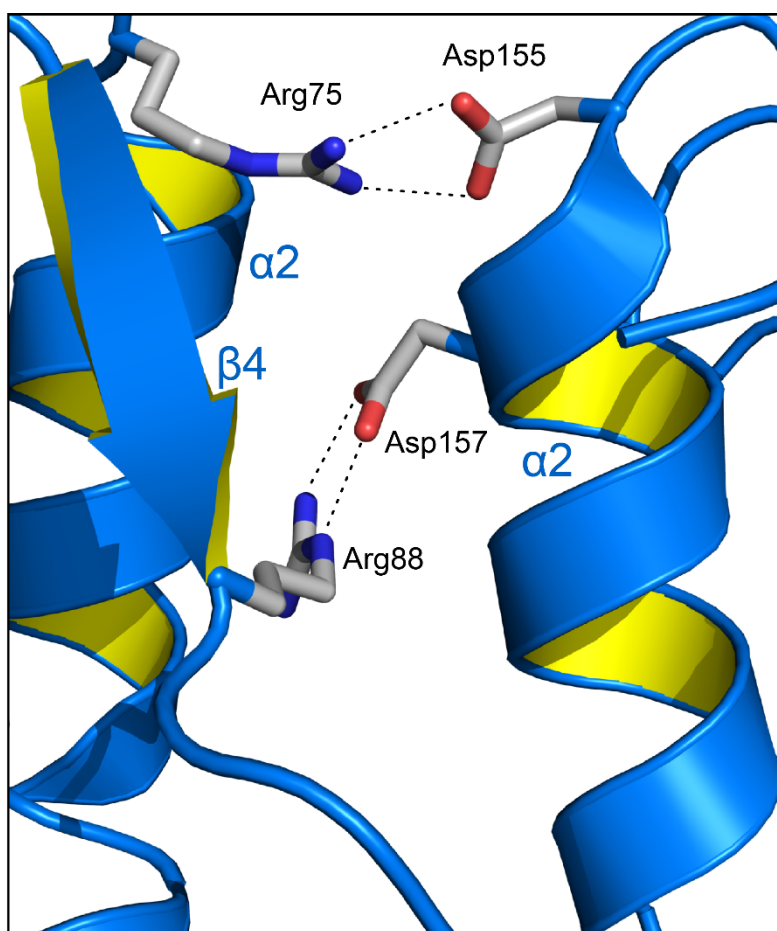


Figure 4.27 Two salt-bridges at the interface of RRM1 and RRM2 domains of UP1. Dotted lines indicate interactions in Arg75-Asp155 and Arg88-Asp157 pairs. (Adapted from PDB code: 1L3K)

In order to investigate the effect of the conserved salt-bridges on RNA binding, we mutated Arg75 and Arg88 either to glutamate to introduce a charge clash or alanine. The correct folding of the UP1 mutants was confirmed by ^1H , ^{15}N HSQC spectra (data not shown). Next, we determined the 12-mer RNA binding affinity of the salt-bridge

mutants using ITC (Figure 4.28). The strongest effect is observed for the double-mutant R75E/R88E where the affinity to 12-mer RNA is decreased by roughly 2.5 fold (K_D of ~ 15.5 nM and ~ 40 nM for UP1 and UP1 R75E/R88E, respectively). These data suggest that the two salt-bridges play an important role in positioning the two RRM domains which can indirectly influence RNA-binding by UP1.

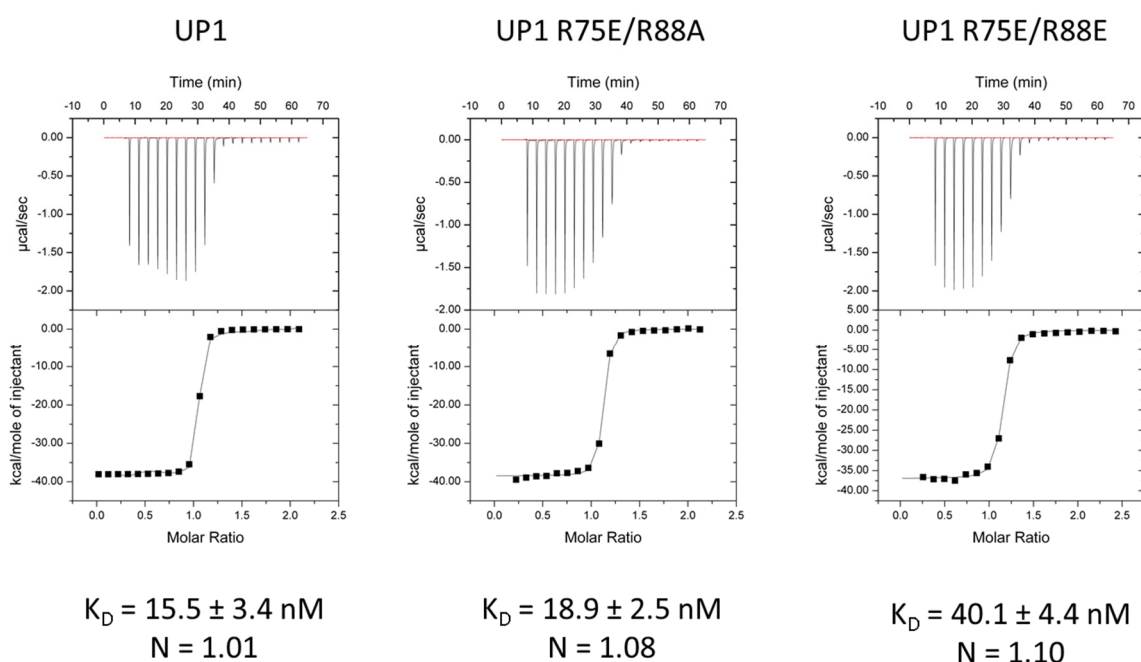


Figure 4.28 The effect of UP1 salt-bridge mutations on RNA binding. ITC binding curves of the interaction of UP1, UP1 R75E/R88A and UP1 R75E/R88E with 12-mer RNA are shown.

4.9 Small angle X-ray scattering (SAXS)

To characterize the shape and topology of hnRNP A1 pri-miR-18a complexes in solution, we used small angle X-ray scattering (SAXS). All SAXS data recorded on complexes involving full-length hnRNP A1 resulted in aggregation and were not analyzed further. Different sample concentrations, temperature and addition of magnesium did not resolve the issue. It has been suggested that the Gly-rich tail of hnRNP A1 is responsible for multimerization and aggregation in the presence of

nucleic acids (Okunola & Krainer, 2009). Therefore we focused on UP1, full-length pri-miR-18a and the UP1-pri-miR-18a complex.

4.9.1 SAXS studies on UP1

For UP1 we recorded SAXS data at multiple concentrations. No sign of inter-particle interference was detected even at the highest protein concentration (10 mg/ml), indicating that the samples are monodisperse in solution. The pair-distance distribution function, $P(r)$, derived from the scattering intensity plot is asymmetric consistent with a non-sphere shape (Figure 4.29). The radius of gyration (R_g) and maximum dimension (D_{max}) calculated using GNOM (Svergun, 1992) are 23.4 Å and 77.5 Å, respectively. We calculated multiple ab initio models of UP1 using the program DAMMIF (Petoukhov & Svergun, 2005). The average ab initio model shows two lobes, presumably corresponding to the two RRM domains, with one lobe slightly more extended (Figure 4.29).

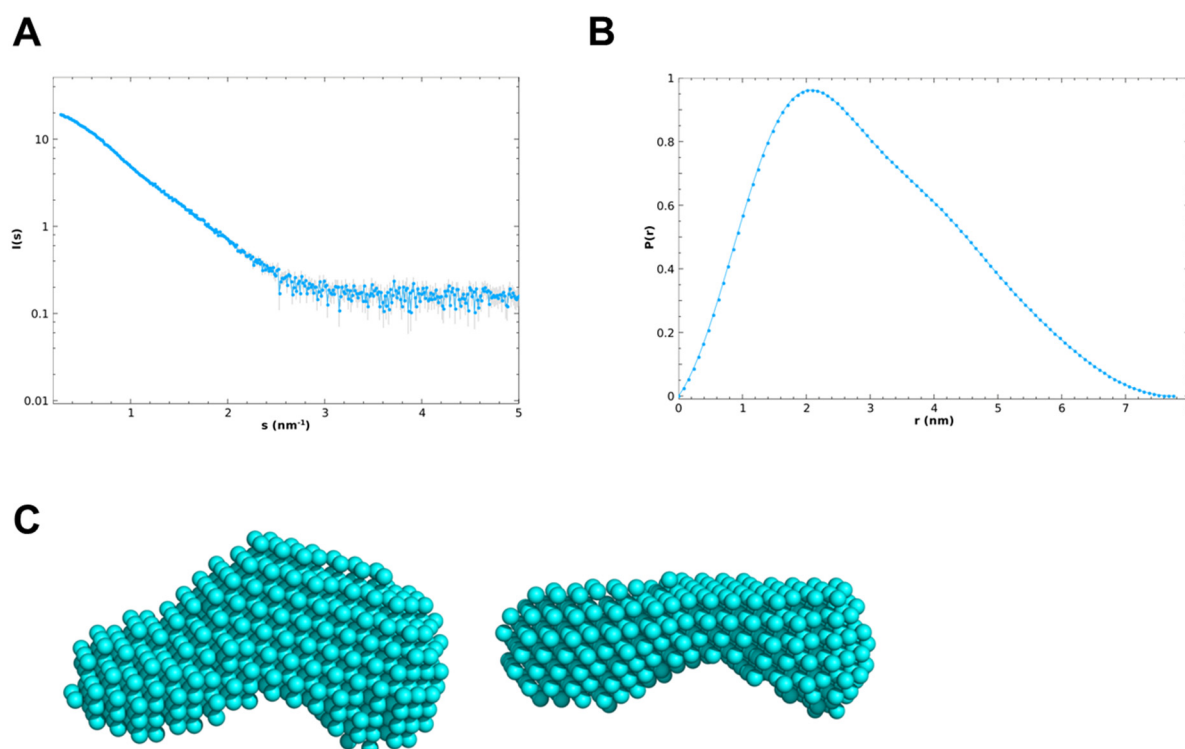


Figure 4.29 SAXS studies on UP1. (A) SAXS intensity plot of UP1 (B) pair-distance distribution function (C) Average ab initio reconstruction of UP1 calculated from SAXS data using DAMMIF. Front and side view are shown.

The scattering curve of UP1 was compared to calculated curves obtained from high resolution structures of free and DNA-bound UP1 using the program CRY SOL (Svergun et al, 1995) (Figure 4.30). The best fit to the experimental curve (Chi value 3.195) was obtained for the 1.1 Å resolution crystal structure of free UP1 (PDB code: 1L3K) (Vitali et al, 2002). The relatively high Chi-values may be due to missing residues in the crystal structure. Surprisingly, the NMR structure of UP1 (PDB code:2LYV) (Barraud & Allain, 2013) provides a worse fit to the scattering curve (Chi value 4.04), despite having the same number of residues as the sample used for our SAXS measurements. One possible explanation for this observation is that the dynamics of the flexible linker connecting RRM1 and RRM2 have not been accounted for in determining the NMR structure. To test this possibility, we have used the program BUNCH which can model multi-domain proteins using a combined rigid body and ab initio modelling approach (Petoukhov & Svergun, 2005). To maintain the domain arrangement of UP1, the two RRM domains were kept fixed while the linker and termini were modelled by dummy atoms. The resulting BUNCH model provides a much better fit to the experimental SAXS curve (Chi value 1.55).

Nevertheless, it is also possible that multiple conformations of UP1 with varying inter-domain distances co-exist in solution. A similar case has been observed for the RRM domains of the splicing factor U2AF65 (Jenkins et al, 2012; Jenkins et al, 2013) where it has been shown that the protein adopts a broad range of conformations in solution and that a single model cannot explain the SAXS data.

Overall, our SAXS data of UP1 suggest that the arrangement and perhaps distance between the two RRM domains in UP1 can show some variations.

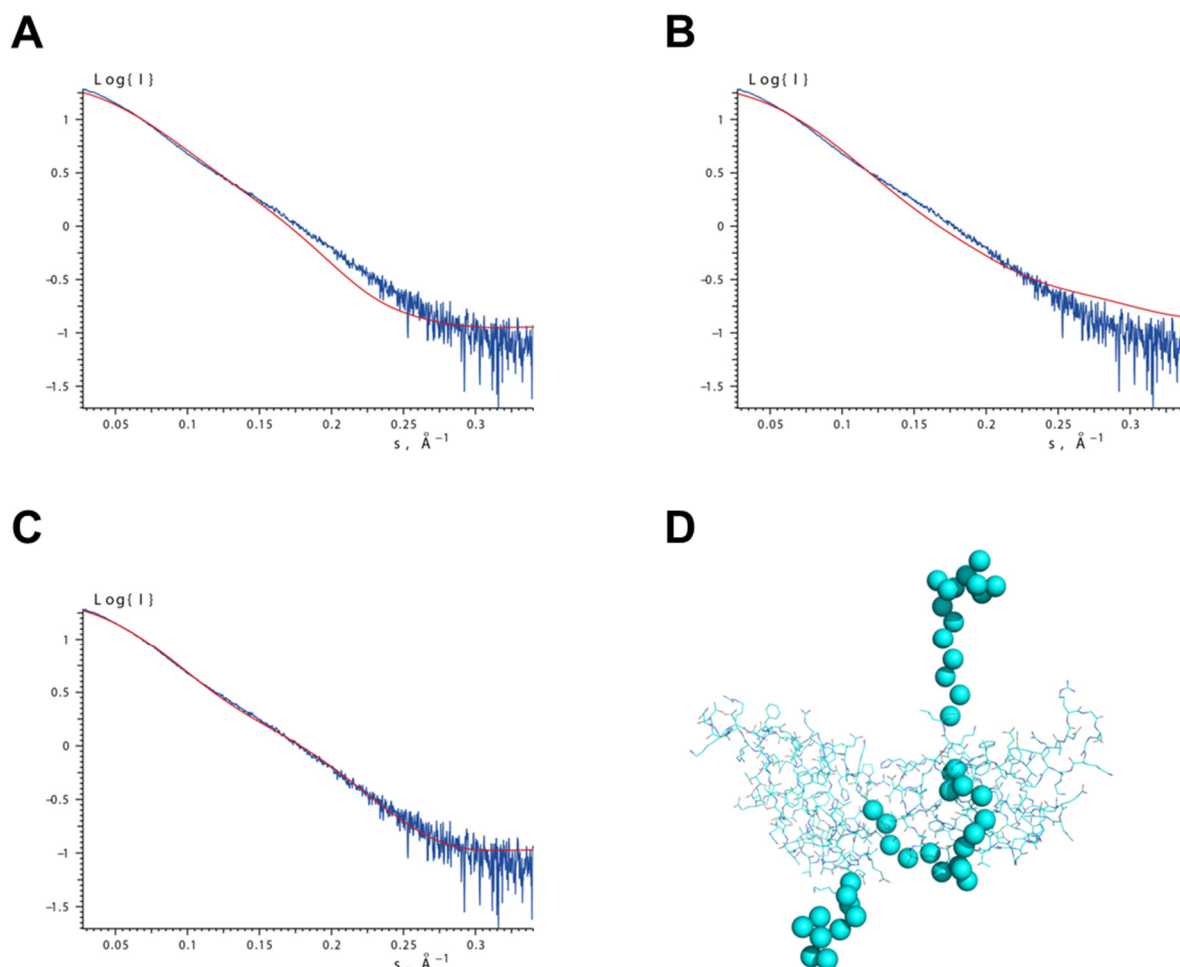


Figure 4.30 (A) CRYSOLO fit of the theoretical scattering curve (red) calculated from the crystal structure (PDB code: 1L3K) to the experimental SAXS curve (blue). Chi value is 3.195 (B) CRYSOLO fit of the theoretical scattering curve (red) calculated from the NMR structure (PDB code: 2LYV) to the experimental SAXS curve (blue). Chi value is 4.045 (C) Fit of the theoretical scattering curve (red) calculated from the BUNCH model of UP1 to the experimental SAXS curve. Chi value is 1.551 (D) BUNCH model of UP1. Flexible regions of the termini and linker are shown as spheres.

4.9.2 SAXS studies on pri-miR-18a

To obtain structural information on the pri-miR-18a, we recorded SAXS data on *in vitro* transcribed RNA. The $P(r)$ function shows an extended tail consistent with an elongated structure. The R_g and D_{max} values derived from the SAXS data are 26.7 Å and 95 Å, respectively. Furthermore, ab-initio calculations using DAMMIF (Petoukhov & Svergun, 2005) resulted in an elongated shape with a slight bent (Figure 4.31).

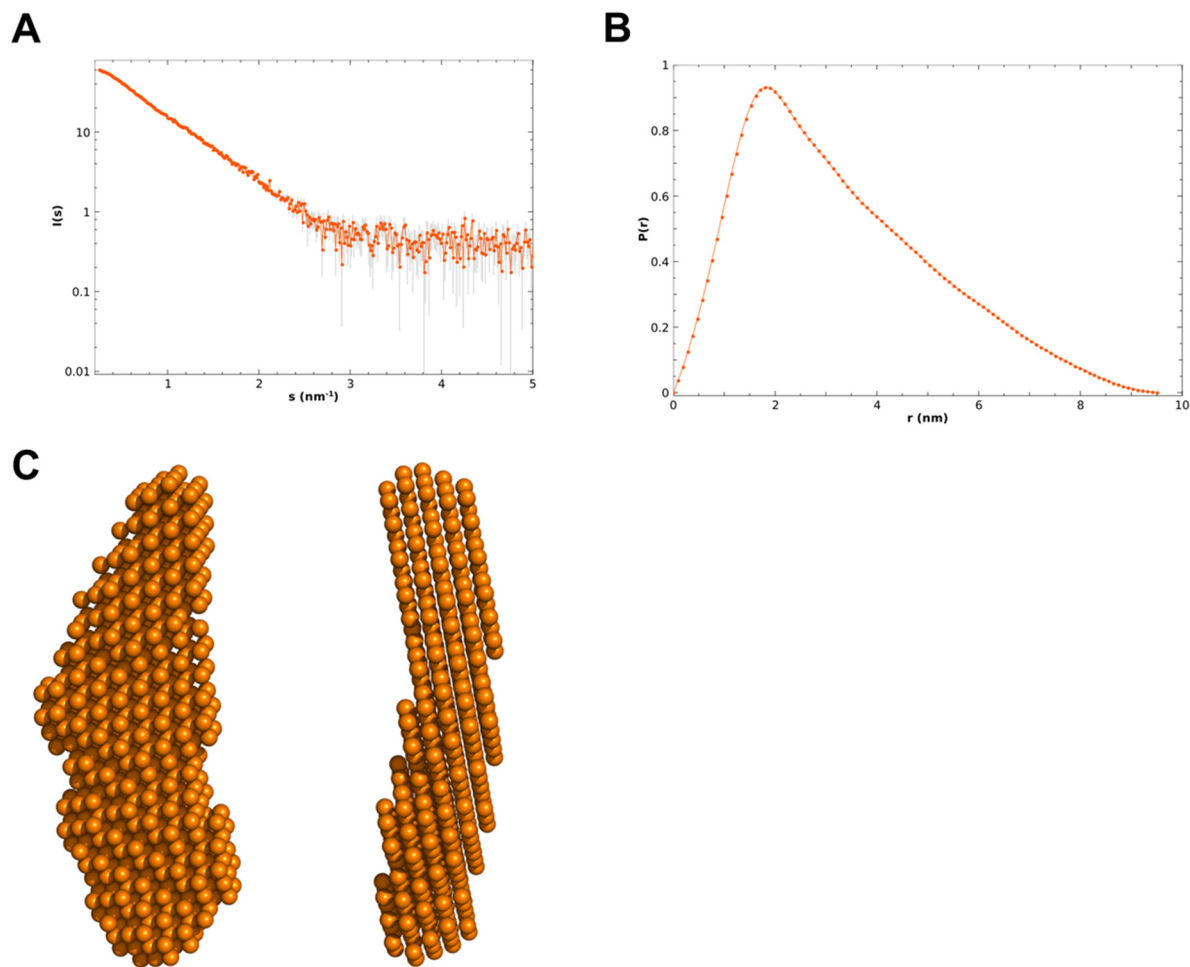


Figure 4.31 SAXS analysis of pri-miR-18a. (A) SAXS intensity plot of 18a (B) pair-distance distribution function (C) Average ab initio reconstruction of pri-miR-18a calculated from SAXS data using DAMMIF.

Since there are no experimental structures available for pri-miR-18a, we calculated 1000 structures using the MC-Sym (Macromolecular Conformations by Symbolic programming) pipeline (Parisien & Major, 2008). We included constraints from the footprinting-derived secondary structure (Michlewski et al, 2008) in the calculation. The theoretical scattering curve of all structures were back-calculated and fitted to the experimental scattering curve using CRY SOL (Svergun et al, 1995). The structures with the best fit (lowest Chi-value) to the SAXS curve were selected and further refined (Figure 4.32).

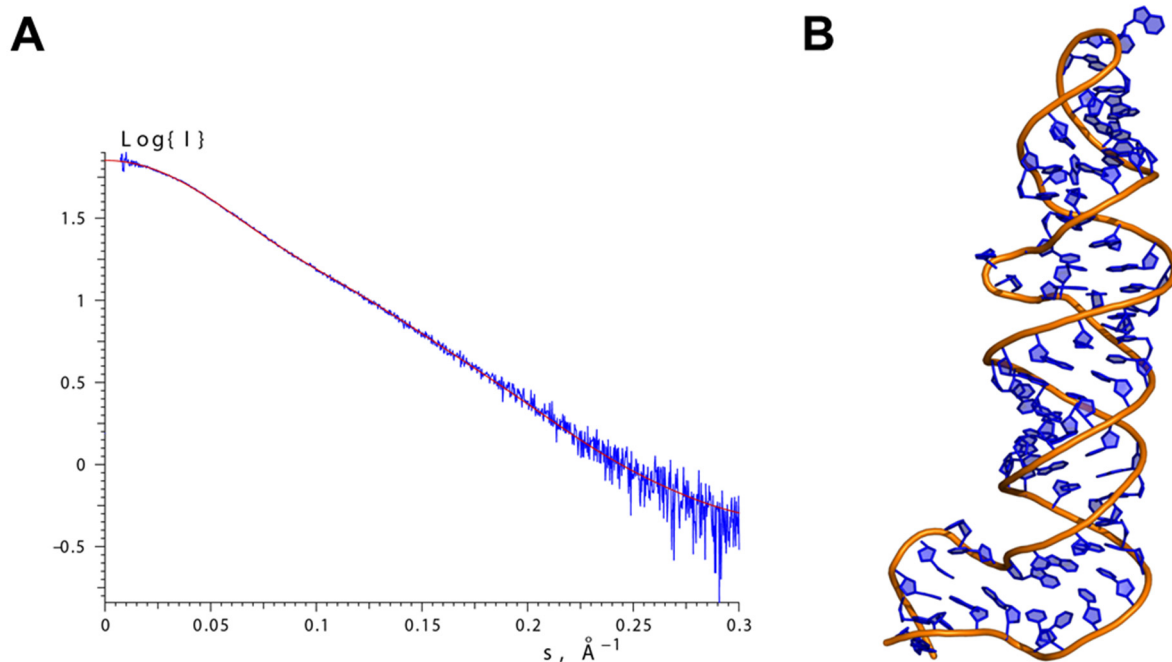


Figure 4.32 A model for pri-miR-18a. (A) CRYSOLOG fit of the theoretical scattering curve (red) to the experimental data (blue). Chi value is 0.9. (B) A representative model of pri-miR-18a calculated by the MC-SYM pipeline.

4.9.3 SAXS studies on the UP1-18a complex

To derive a low resolution model of UP1 bound to pri-miR-18a, we recorded SAXS data on a 1:1 complex of the UP1-18a complex. The R_g and D_{max} values derived from the SAXS data are 37.6 \AA and 125 \AA , respectively. To calculate ab initio models of the protein-RNA complex we used the program MONSA which allows simultaneous fitting of multiple curves (Svergun, 1999). The important assumption for the models obtained using MONSA is that the conformation of the two components do not change upon complex formation. The MONSA-derived model shows that UP1 binds to one end of the elongated pri-miR-18a in a manner in which both domains can make contacts to the RNA (Figure 4.33).

In summary, the SAXS data suggest that one molecule of UP1 binds to the end of pri-miR-18a. This is consistent with the footprint analysis where UP1 was shown to specifically interact with the terminal loop of pri-miR-18a (described in section 4.5).

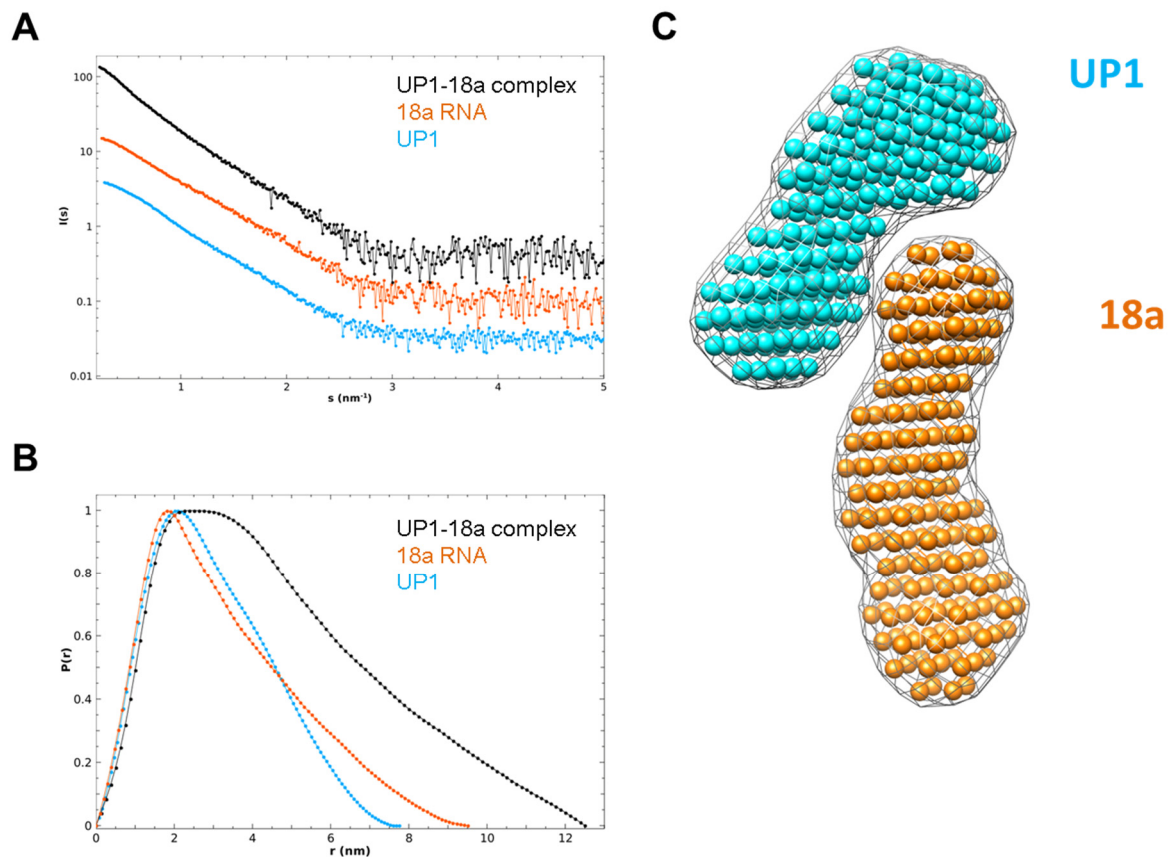


Figure 4.33 SAXS analysis on the UP1-18a complex. (A) SAXS intensity plots of UP1 (cyan), 18a (orange) and UP1-18a complex (black). The curves are shifted arbitrarily. (B) Comparison of pair-distance distribution functions, $P(r)$, of UP1 (cyan), 18a RNA (orange) and UP1-18a complex (black). (C) A model of UP1-18a complex obtained from the multiphase *ab initio* modeling program MONSA (Svergun, 1999).

4.10 Functional data

To test the activity of hnRNP A1 in miRNA processing, we used an *in vivo* assay in collaboration with the group of Dr. Javier Caceres (MRC, Edinburgh). In this assay, different constructs of hnRNP A1 are overexpressed in HeLa cells and the levels of miR-18a are analyzed by qRT-PCR.

We found that overexpression of hnRNP A1 leads to a ~2-fold increase in the levels of miRNA-18a, whereas UP1 has minimum effect. (Figure 4.34) This is most likely due to the fact that UP1 localizes to the cytoplasm. To force the localization of the

protein to the nucleus, we generated constructs in which the target protein is fused to the nuclear import sequence, known as M9 (Siomi & Dreyfuss, 1995). The constructs created for *in vivo* functional assays are listed in Table 4.5.

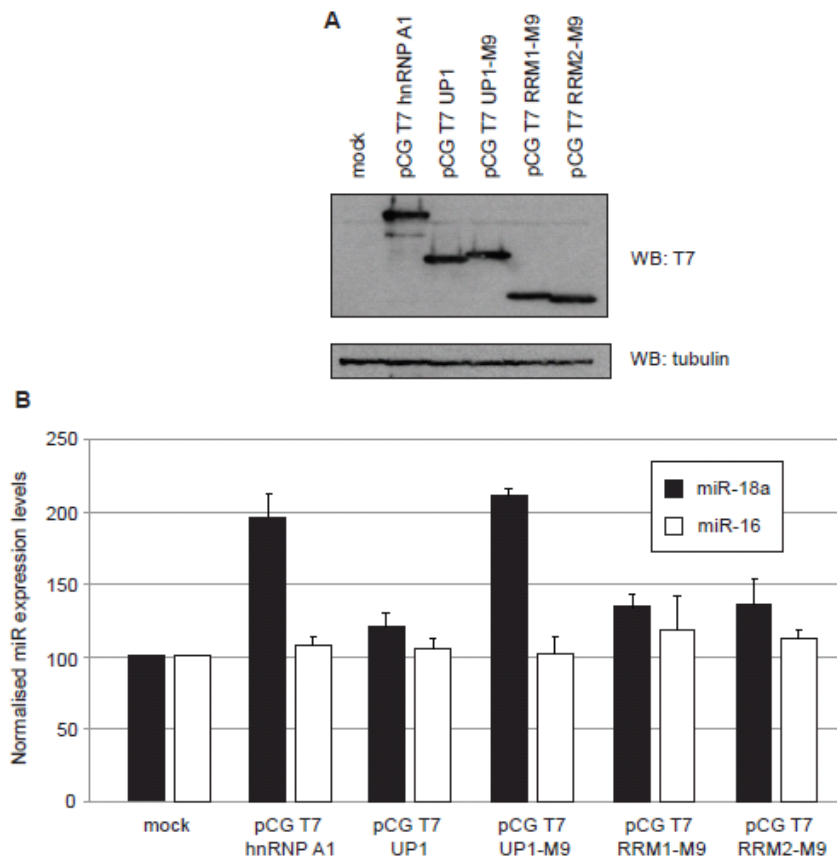


Figure 4.34 (A) Western blot analysis of hnRNP A1 fragments overexpressed in HeLa cells. (B) Graph of the normalized miR-18a expression levels (black bars) upon overexpression of different hnRNP A1 fragments in HeLa cells. The expression levels of miR-16 are shown as control (white bars). Experiments were performed by Dr. Gracjan Michlewski (Edinburgh, UK).

In contrast to UP1, UP1-M9 which localizes exclusively to the nucleus stimulated miR-18a production *in vivo*. However, Nuclear RRM1-M9 and RRM2-M9 did not lead to increase in miR-18a levels. Next, we tested several hnRNP A1 RNA-binding mutants in the functional assay. We created three constructs which carried mutations in the conserved Phe residues of RNP-1 in RRM1 (UP1-M9-FD1), RRM2 (UP1-M9-FD2), or both (UP1-M9-FD12). It has been shown previously that these mutations

affect RNA-binding by hnRNP A1 (Mayeda et al, 1994). In two additional constructs we mutated all Phe residues in conserved RNP-1 and RNP-2 motifs (six Phe residues, four in RNP-1 and two in RNP-2) to Ala or Asp. Immunofluorescence data showed that with the exception of UP1-M9-FD1 (with F57D/F59D mutations in RRM1), all constructs have almost exclusive nuclear localization (Table 4.5).

Table 4.5 Constructs used in the *in vivo* processing assay and their subcellular localization

Construct	Mutations	Subcellular localization
hnRNP A1		Nuclear
RRM1-M9		Mostly nuclear
RRM2-M9		Mostly nuclear
UP1-M9		Nuclear
UP1-M9-FD1	F57D, F59D	Cytoplasmic
UP1-M9-FD2	F148D, F150D	Nuclear
UP1-M9-FD12	F57D, F59D, F148D, F150D	Nuclear
UP1-M9-FD6	F17D, F57D, F59D, F108D, F148D, F150D	Mostly nuclear
UP1-M9-FA6	F17A, F57A, F59A, F108A, F148A, F150A	Nuclear

As seen in Figure 4.35, no increased level of miRNA-18a expression is observed upon overexpression of RNA-binding mutants. This indicates that the RNA-binding activity of hnRNP A1 is essential for its stimulating activity of miRNA-18a production.

Collectively, these experiments clearly show that nuclear localization and RNA-binding are required for stimulating miR-18a production by hnRNP A1.

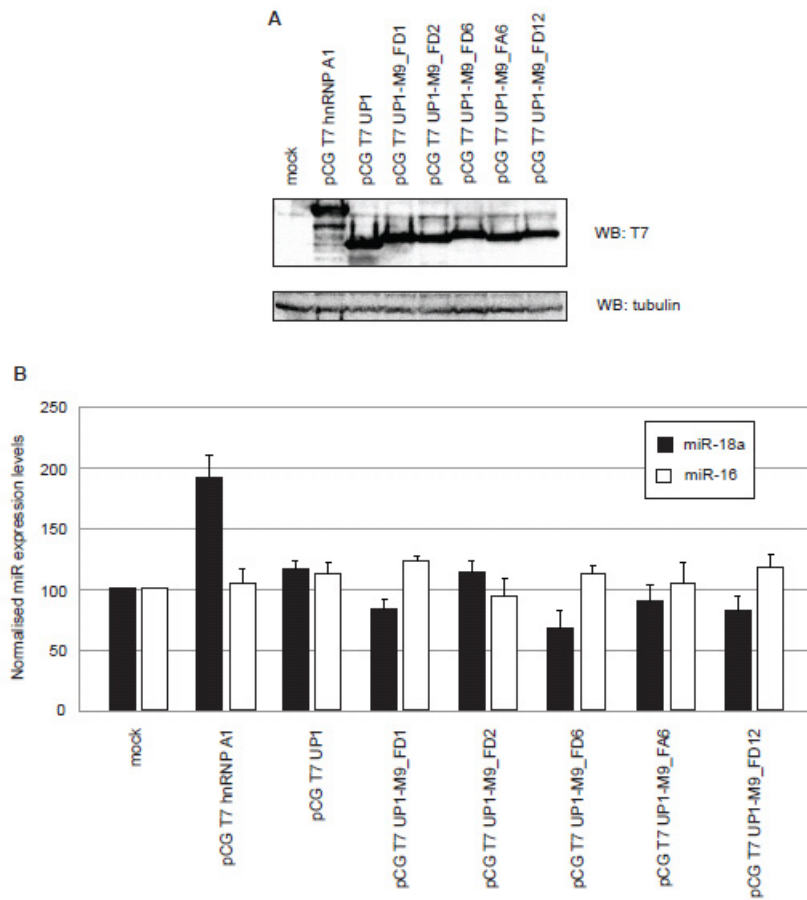


Figure 4.35 (A) Western blot analysis of hnRNP A1 mutants overexpressed in HeLa cells. (B) Graph of the normalized miR-18a expression levels (black bars) upon overexpression of different hnRNP A1 mutants in HeLa cells. The expression levels of miR-16 are shown as control (white bars). Increased production of miR-18a is observed upon transfection of hnRNP A1 and UP1-M9 constructs. Experiments were performed by Dr. Gracjan Michlewski (Edinburgh, UK).

5. Discussion

5.1 Solution conformation of UP1

The NMR data obtained in this study establish the conformation and dynamics of UP1 in solution. Our data indicate that, the two RRM domains of UP1 adopt the canonical $\beta\alpha\beta\beta\alpha\beta$ topology with the exception that two additional α -helices are present at the N- and C- termini (α_0 and α_3 , respectively). Interestingly, both α_0 and α_3 are involved in interactions with RNA. Residues corresponding to α_3 (Ser182-Ser188) were not visible in the crystal structure of UP1 in the free form, but the helix becomes ordered in the DNA-bound form (Ding et al, 1999). We show that in the free form, this region has helical propensity which is increased upon RNA binding. Indeed, the largest chemical shift changes upon RNA binding are localized to the C-terminus in both RRM2 and UP1 constructs. In addition, the relaxation data indicate that the inter-domain linker is flexible, in agreement with the notion that residues 92-102 of the inter-domain linker are not modeled in the crystal structure of free UP1.

Previous crystal structures of free and DNA-bound UP1 revealed that upon DNA binding, there is a $\sim 15^\circ$ rotation of one RRM domain relative to the other resulting in different orientation of the two domains in the free and bound form of UP1 (Ding et al, 1999). Recently, the NMR structure of UP1 determined using a segmental-labeling approach has been reported (Barraud & Allain, 2013) and is fully consistent with our NMR analyses. Overall, the solution structure of UP1 is highly similar to the previously reported crystal structures. However, the relative orientation of the RRM domains in the NMR structure of free UP1 is more similar to the crystal structure of UP1 in the DNA-bound form rather than the unbound conformation (backbone r.m.s.d of 1.8 and 2.8 Å, respectively) (Barraud & Allain, 2013). The authors suggested that, the crystal structure of UP1 in the free form represents a less stable conformation induced by crystal contacts. In addition, the NMR and crystal structures of UP1 reveal small local differences when comparing each domain separately. The most significant difference is seen in residues following α_2 (Lys166 and Tyr167) in RRM2 which could possibly be explained by the involvement of these residues in inter-molecular contacts with neighboring molecules in the crystal. Indeed, both of these residues are located at the dimer interface in our crystal structure of UP1-12-

mer RNA (see section 4.6.2) as well as the previously reported UP1-DNA complex (Ding et al, 1999) where Lys166 and Tyr167 make inter-molecular contacts with Asp94 and Ile164 of a second UP1 molecule, respectively. Interestingly, our NMR relaxation data indicate that residues in this region have increased R_2 values, suggesting that they can undergo exchange in the microsecond-millisecond timescale such as exchange between monomer and dimer states. Although NMR, static light scattering and SAXS data all suggest that UP1 is a monomer in solution, we cannot exclude that the increased R_2 values for certain residues is due to partial dimerization at the high protein concentrations used in our NMR relaxation measurements. To exclude the possibility of dimerization, the NMR relaxation experiments have to be performed at lower protein concentrations. Nevertheless, regardless of whether the differences seen in NMR and crystal structures of UP1 are due to different contacts or dynamics, these analysis suggest that the two RRM domains in UP1 can move relative to each other. Consistent with this idea, static structures of UP1 determined by X-ray crystallography or NMR do not perfectly agree with the SAXS data of UP1. Because SAXS reports on average over an ensemble, it is possible that UP1 samples different conformations in solution.

Despite the differences seen in the domain arrangement, in all the structures of UP1 the two RRM domains are kept close to each other by a pair of Arg-Asp salt-bridges. By mutational analysis we have shown that disruption of the salt-bridges by introducing a charge clash results in ~ 2.5 -fold decrease in affinity for single-stranded 12-mer RNA. This suggests that the inter-domain contact of the two domains is important for RNA binding. The rather small reduction in affinity is likely due to the fact that the two domains are connected by a relatively short (~ 15 residue) flexible linker. Theoretical models propose that, to effectively decouple the binding affinities of individual domains, the two RRM domains must be connected by a flexible linker of 60 residues or longer (Shamoo et al, 1995). In addition to the linker and the two salt-bridges, the RRM domains are in close contact by a number of residues from $\alpha 0$ (RRM1), $\alpha 2$ (RRM1 and RRM2), $\beta 4$ (RRM2) and the inter-RRM linker forming a hydrophobic cluster at the interface. Indeed, 64 inter-domain NOE contacts were used in determining the NMR structure of UP1 (Barraud & Allain, 2013).

Collectively, our data indicate that the two RRM domains of UP1 are kept in close contact in solution forming a single binding platform. We suggest that the relative

domain arrangement in UP1 could undergo small variations which would provide more adaptability in binding to various target RNAs. The inter-domain dynamics can play an important role in recognition of nucleic-acids by hnRNP A1 as it has been shown for several other multi-domain proteins (Mackereth & Sattler, 2012).

5.2 Solution conformation of pri-miR-18a

In this work, we have presented a low resolution model of full-length pri-miR-18a in solution based on SAXS data. The model shows an elongated structure with a length of ~90 Å. More detailed information has been obtained from smaller pri-miR-18a fragments corresponding to the loop and stem regions as described below.

For the 17-mer loop construct, only three stable base-pairs were detected by NMR, one of which is an artificially added G:C base-pair. This suggests that the terminal loop of pri-miR-18a is flexible and does not form stable base pairs. By contrast, the 17-mer-mut construct (carrying an A→C mutation) has a stabilized stem, as all expected base-pairs are visible in the imino region of the NMR spectra. Interestingly, the sequence of the apical GUAGA loop of the 17-mer stem-loop construct resembles the sequence of the GNR(N)A pentaloop family (N is any nucleotide, R is a purine and (N) is any bulged nucleotide) (Legault et al, 1998; Stefl & Allain, 2005). However, lack of imino cross-peaks from this pentaloop in the NMR spectra of 17-mer and 17-mer-mut, suggests that it does not adopt a stable conformation. Notably, the GUAGA pentaloop is part of the hnRNP A1-recognition sequence and its flexibility is likely to be important for protein binding as discussed below.

The stem of pri-miR-18a features a tandem U:C mismatch. A UC→GU mutation in this region to mimic the stem of the highly homologous pri-miR-18b was shown to remove the hnRNP A1 requirement for processing (Michlewski et al, 2008). Using novel NMR experiments, we have determined the secondary structure of the stem 18a RNA. All the base-pairs flanking the tandem U:C mismatches were detected by NMR, suggesting that the two U:C base-pairs are likely to form. In agreement with these data, several crystal structures of tandem U:C mismatches have shown that water mediated contacts as well as a single N-H...O type hydrogen-bond is formed between the amino group of Cytosine (N4) and carbonyl oxygen (O4) of Uracil

(Holbrook et al, 1991; Cruse et al, 1994; Anderson et al, 1999) which is not detectable by the (Py) H(CC)NN-COSY experiment employed in our studies (Dallmann et al, 2013). The formation of U:C base-pairs has also been confirmed in solution (Tanaka et al, 2000).

5.3 Recognition of pri-miR-18a by hnRNP A1

We have performed a detailed investigation of the interaction of hnRNP A1 with RNAs derived from pri-miR-18a using a variety of techniques.

Our EMSA analysis reveal that UP1 specifically binds to RNAs derived from the terminal loop of pri-miR-18a. While we could not detect any binding to the stem 18a RNA by EMSA, it is still possible that in the context of full-length pri-miR-18a, hnRNP A1 makes transient contacts to the stem region, albeit with lower affinity. Moreover, the EMSA data did not show any binding of individual RRM domains to RNA, suggesting that the cooperative action of two RRM domains (as in UP1) is required for efficient binding to RNA. However, we could show by NMR titration analysis and ITC that in fact single RRM domains bind to an RNA loop fragment with low micromolar K_D , which has not been detected by EMSA perhaps due to different experimental conditions. The *in vivo* functional data further confirm that the stimulatory function of hnRNP A1 in processing of pri-miR-18a requires both RRM domains, as mutations that affect RNA binding in one domain (or deletion of one domain) abolish the activity of hnRNP A1.

The complex ITC binding profile of the UP1-17-mer interaction is possibly due to multiple-binding events and/or conformational changes associated with binding. Of the two UAG motifs in the 17-mer construct, the one located in the GUAGA pentaloop is accessible for interaction with UP1, whereas the Guanosine in the second UAG motif is involved in base-pairing with a Uridine. It is noteworthy that, the extra G:C base-pair added at the beginning of the 17-mer RNA can, at least partially, stabilize the region in the vicinity of the second UAG motif. This artificial stabilization might explain the unusual ITC binding curve of the UP1-17-mer interaction. One could imagine that binding of UP1 to the first UAG motif induces melting of the hairpin in order to allow binding to the second UAG motif. Consistent with this

hypothesis, ITC data obtained for interaction of UP1 with 17-mer-mut displays a simple one-site binding profile. Remarkably, UP1 binds to the 17-mer-mut and 7-mer (5'-AGUAGAU-3') RNAs with the same affinity (K_D of $\sim 3 \mu\text{M}$), further supporting the idea that the single-stranded UAG motif of 17-mer-mut RNA is the main recognition site. Although 7-mer and 17-mer-mut RNAs have the same affinity for UP1, the binding stoichiometry is different. UP1 binds to two 7-mer RNAs (one per RRM domain), whereas binding to 17-mer-mut results in a 1:1 stoichiometry which is presumably the average stoichiometry of RRM1 and RRM2 binding to a single UAG motif. This suggests that simultaneous binding of two 17-mer-mut stem-loops to UP1 (one to each RRM domain) would not be possible due to steric hindrance. The single A \rightarrow C mutation in the 17-mer-mut construct not only leads to stabilization of the stem but also converts the second UAG motif to an UCG motif, thereby disrupting the second UP1-binding site. The combined effect should lead to a decrease in the binding affinity for UP1. Although we could not quantify the binding of UP1 to the 17-mer RNA, the affinity should be higher than 17-mer-mut due to the presence of both UAG motifs. Based on these data, we reasoned that for high-affinity binding of UP1, two UAG motifs are required, preferably in a single-stranded form. Consistent with this model, the highest affinity ($K_D \sim 15 \text{ nM}$) was obtained for a single-stranded 12-mer RNA (5'-AGUAGAUUAGCA-3') derived from the terminal loop of pri-miR-18a. Surprisingly, UP1 binds to the 12-mer RNA with ~ 10 -fold higher affinity than the full-length pri-miR-18a ($K_D \sim 150 \text{ nM}$). These results suggest that: (i) the 12-mer RNA is a minimal UP1-binding sequence which represents the binding to pri-miR-18a, (ii) the 12-mer fragment is not fully accessible in full-length pri-miR-18a and this leads to a decrease in affinity for UP1.

5.4 Structural basis of interaction of UP1 with the terminal loop of pri-miR-18a

Our crystal structure of UP1 in complex with the 12-mer loop RNA fragment showed that two UP1 molecules bind to two separate RNA strands in a manner in which each protein is bound to two RNA molecules and each RNA chain is bound to two protein molecules (Figure 5.1). This specific arrangement of UP1 in complex with nucleic acids has been reported previously for the structure of UP1-telomeric DNA

complex (Ding et al, 1999). The 2:2 stoichiometry observed in the crystal structure would require dimerization of UP1 in complex with nucleic acids and raises the question regarding the stoichiometry of the complex in solution. The dimerization of UP1 can be a function of concentration and the sequence of the nucleic acid target. For example, in the case of telomeric DNA it has been suggested that the 2:2 stoichiometry is also valid in solution (Myers et al, 2003). However, our light scattering data indicate that UP1 is a monomer in solution both in the free form and in complex with the 12-mer RNA, suggesting that the dimer observed in the crystal structure is an artifact of crystallization conditions. Based on light scattering and NMR data, we propose that each RRM domain of UP1 recognizes one segment of the 12-mer RNA in a 1:1 stoichiometry (Figure 5.1). A 1:1 complex can be modeled by rearranging the RNA such that it contacts the two RRM domains of the same UP1 molecule. According to this model, RRM1 interacts with the first UAG motif and RRM2 binds to the second UAG in the terminal loop. The 1:1 model is fully consistent with our biochemical and functional data regarding the requirement for two RRM domains and two UAG motifs. More experimental evidence to confirm this model is required, including paramagnetic relaxation enhancements (PRE) and SAXS to define the domain arrangement and inter-domain distance in the UP1-12-mer complex.

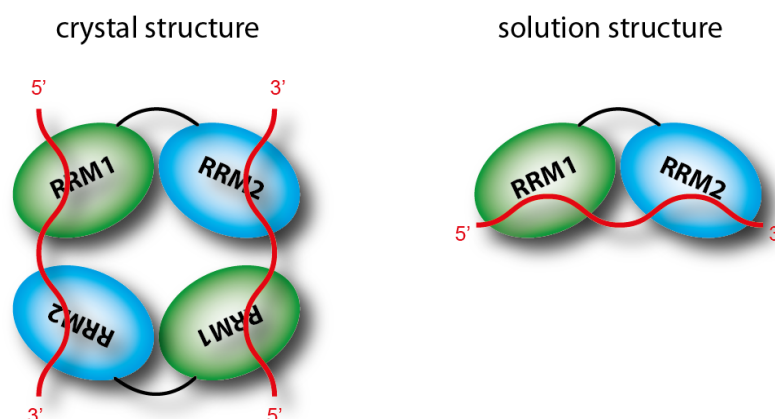


Figure 5.1 Schematic representation of the arrangement of UP1 bound to 12-mer RNA (shown as a red line). The crystal structure shows a 2:2 protein:RNA stoichiometry, while our solution data are consistent with a 1:1 stoichiometry, in which RRM1 and RRM2 bind to the 5'- and 3'-ends of the RNA, respectively.

Besides structural details, another important goal of this study was to determine the stoichiometry of the hnRNP A1 interaction with pri-miR-18a. In all experiments involving full-length hnRNP A1, addition of RNA resulted in extensive aggregation which precluded further analysis (see below). Nevertheless, the stoichiometry of the UP1-pri-miR-18a interaction was determined by static light scattering which clearly showed that UP1 binds to pri-miR-18a in a 1:1 stoichiometry forming a complex with the molecular mass of ~45 kDa. Consistent with these data, SAXS and footprint analysis of the UP1-pri-miR-18a complex also suggest the interaction of one molecule of UP1 with the RNA. These data were further supported by ITC experiments which yielded a stoichiometry of ~1.35 (protein to RNA) for the interaction of UP1 with pri-miR-18a. In this case, the slight deviation of the stoichiometry value from 1 is perhaps due to unspecific binding of UP1 at higher protein to RNA ratios. Indeed, the pri-miR-18a sequence includes other potential hnRNP A1-binding sites including two UAG motifs in the mature miR-18a part (5'-UAAGGUGCAUCUUAGUGCAGAUUAG-3'). Interestingly, a recent study has shown that hnRNP A1 is able to bind to mature miR-18a in the cytoplasm (Fujiya et al, 2013); however, since this region is base-paired in pri-miR-18a it would require melting in order to interact with UP1.

5.5 Regulation of miRNA processing: the importance of terminal loops

The group of Javier Caceres have shown using footprint and mutational analysis, that hnRNP A1 binds to two distinct regions of pri-miR-18a, i.e. the terminal loop and the stem (Michlewski et al, 2008). They reported that mutations in the conserved terminal loop abolishes the processing of pri-miR-18a in cell extracts, highlighting the importance of the terminal loop region in miRNA processing. They also suggested that binding of A1 to the stem region leads to opening of the duplex in the vicinity of the Drosha cleavage site which in turn might facilitate Drosha cleavage. An important finding was that mutating the stem region of pri-miR-18a, to mimic the stem region of pri-miR-18b, leads to its efficient processing even in the absence of hnRNP A1. These data suggest that both the terminal loop and the stem region of pri-miR-18a are important features which can influence the processing of pri-miR-18a. While our

data are consistent with the interaction of UP1 with the terminal loop, we have not detected a strong interaction between UP1 and the stem region of pri-miR-18a. It is worth mentioning that in the study by Michlewski et al, full-length hnRNP A1 was used while we have mainly used UP1 in our assays based on new *in vivo* data from the group of Javier Caceres which suggest that the UP1 fragment is sufficient to stimulate the processing of pri-miR-18a (unpublished data). Although we have shown that the main hnRNP A1-binding site in pri-miR-18a is the terminal loop, we cannot exclude that at higher protein to RNA ratios, hnRNP A1 binds to the stem region. Indeed, it has been shown that hnRNP A1 can multimerize and spread along RNA molecules in a preferred 3' to 5' direction (Okunola & Krainer, 2009). It has also been proposed that the first hnRNP A1 molecule binds sequence-specifically and then additional copies are recruited to the RNA in a cooperative manner, presumably through the Gly-rich tail of hnRNP A1 (Zhu et al, 2001). However, since UP1 has the same functional activity as hnRNP A1 in activating pri-miRNA processing, the important question is: How does binding of UP1 stimulate cleavage of pri-miR-18a by Drosha?

One possibility is that binding of UP1 leads to partial opening/melting of the terminal loop which can lead to destabilization of the stem region and thus render it more accessible for processing by Drosha. Indeed, UP1 (unwinding protein 1), as the name suggests, can unwind secondary and higher order structures of DNA and RNA (Herrick & Alberts, 1976; Fukuda et al, 2002; Zhang et al, 2006). A similar mechanism has been suggested based on the crystal structure of Lin28 in complex with let-7 precursors. It was proposed that binding of Lin28 to the terminal loop of pre-let-7 would lead to partial melting in the upper part of the stem which can inhibit processing by both Drosha and Dicer (Nam et al, 2011). In both the hnRNP A1-18a and Lin28-let7 systems, it is unclear how binding of a regulator to the terminal loop can affect Drosha cleavage at the opposite end of the RNA. Whether unwinding of the terminal loop can be transferred to the stem region remains to be shown experimentally. It is also possible that binding of hnRNP A1 to the terminal loop, even without affecting the stem region of the pri-miRNA, can potentially influence Drosha processing. Due to lack of high resolution structures of Drosha in complex with pri-miRNAs, the exact sequence and structural requirements for recognition and cleavage of pri-miRNAs by Drosha are not known. While it is believed that the stem

of pri-miRNA and single-stranded regions flanking the stem are important for Drosha processing, the importance of the terminal loop in miRNA processing has been debated. Several studies have shown that a large flexible terminal loop enhances miRNA production (Zeng & Cullen, 2003; Zeng & Cullen, 2005; Zeng et al, 2005; Zhang & Zeng, 2010), while one group has suggested that the terminal loop is dispensable for miRNA processing, at least for one class of miRNA precursors (Han et al, 2006). However, recent studies from the group of Javier Caceres and others have established that terminal loops are important regulatory elements in miRNA processing (reviewed in Choudhury & Michlewski, 2012; Castilla-Llorente et al, 2013).

It has been shown that the processing efficiency of pri-miR-18a is context-dependent, suggesting that the sequence and/or structure of 18a as part of the miR-17-92 cluster is not optimal for Drosha processing (Guil & Caceres, 2007). Interestingly, several studies have recently shown that the miR-17-92 cluster adopts a compact tertiary structure in which individual miRNAs have different expression levels depending on whether they are located on the surface or buried inside the core (Chaulk et al, 2011; Chakraborty et al, 2012; Chaulk et al, 2014). Notably, SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension) analysis revealed that GU and GA nucleotides in the terminal GUAGA pentaloop of 18a in the cluster are solvent inaccessible (Chakraborty et al, 2012). The GUAGA pentaloop corresponds to the sequence that we have identified as the main hnRNP A1-binding site in the terminal loop of pri-miR-18a, therefore it is tempting to speculate that binding of hnRNP A1 to the miR-17-92 cluster is associated with a conformational change in the RNA which can facilitate Drosha cleavage. It is also possible that hnRNP A1 recruits other, as yet unknown, proteins which can increase the activity of the Drosha complex.

We propose that the tertiary structure of pri-miR-18a in the context of the miR-17-92 cluster as well as sequences of the stem and loop region are important determinants of miRNA processing by Drosha. This process is regulated by the trans-acting factor hnRNP A1 which primarily interacts with the conserved terminal loop of pri-miR-18a.

5.6 Conclusions and outlook

In this study we have used a multi-disciplinary approach to understand the molecular mechanism by which hnRNP A1 binds to pri-miR-18a and facilitates its processing. Our results establish that hnRNP A1 specifically binds to the terminal loop region of pri-miR-18a through interactions involving both RRM domains and a fragment in the terminal loop of pri-miR-18a which harbors two UAG motifs. The crystal structure of the UP1-loop RNA complex reveals how each RRM domain specifically recognizes one UAG motif. Furthermore, we have characterized the structure of full-length pri-miR-18a in complex with UP1 using SAXS, static light scattering and footprint analysis which show how UP1 specifically associates with the terminal loop of pri-miR-18a. Our biochemical and structural data are complemented with mutagenesis and *in vivo* functional assays. This study showcases the importance of combining different structural biology techniques together with functional data in order to provide a more comprehensive understanding of biological systems.

An important direction for further studies will be to understand the molecular details of post-transcriptional regulation of miRNA processing in the miR-17-92 cluster. To this end, the tertiary structures of the miRNA cluster in the presence and absence of hnRNP A1 have to be determined by an integrated structural biology approach as described in this thesis. Due to the large size of the complex, more advanced NMR methods and segmental isotope-labeling approaches would be advantageous. In addition, it would be interesting to characterize the molecular details of the interaction of hnRNP A1 with pri-let-7a where it acts as a negative regulator. Another interesting topic to pursue is to explain why hnRNP A1 is dispensable for the processing of pri-miR-18b, which is highly homologous to pri-miR-18a. In order to understand the differences between the two RNAs, the three-dimensional structures of both pri-miR-18a and pri-miR-18b have to be determined. Here, protocols for high yield production of 18a stem RNA have been established and the secondary structure of 18a stem has been determined which will be a good starting point for future investigations. The main focus should be on comparing the stem region of 18a and 18b in terms of structure as well as dynamics by NMR spectroscopy and SAXS.

6. hnRNP A1 in 3' splice site selection

This chapter is published under:

Tavanez JP, Madl T, Kooshapur H, Sattler M, Valcarcel J (2012) hnRNP A1 proofreads 3' splice site recognition by U2AF. *Molecular Cell* 45: 314-329

6.1 Abstract

One of the earliest steps in metazoan pre-mRNA splicing involves binding of U2 snRNP auxiliary factor (U2AF) 65 kDa subunit (U2AF65) to the polypyrimidine (Py) tract and of the 35 kDa subunit (U2AF35) to the invariant AG dinucleotide at the intron 3'-end. However, the mechanism by which the splicing machinery discriminates between bona fide splice sites and other related sequences is not clearly understood. The group of Juan Valcárcel have identified hnRNP A1 as a factor which allows U2AF to discriminate between pyrimidine-rich RNA sequences followed or not by a 3' splice site AG. Biochemical data indicate that hnRNP A1 forms a ternary complex with the U2AF heterodimer on AG-containing uridine-rich RNAs, while it displaces U2AF from non-AG-containing uridine-rich RNAs, an activity that requires the glycine-rich domain of hnRNP A1. The aim of this project was to complement these studies by characterizing the molecular interactions and binding interfaces involving the minimal U2AF heterodimer, hnRNP A1, and the AG- and CG-containing pyrimidine-rich RNAs using NMR spectroscopy.

6.2 Introduction

Splicing of mRNA precursors is an important step in eukaryotic gene expression whereby noncoding intronic sequences are excised and exonic sequences are spliced together to generate mature mRNAs (reviewed by Wahl et al, 2009). The process offers multiple opportunities for gene regulation, which are abundantly exploited during the development and cell differentiation of multicellular organisms (reviewed by Nilsen & Graveley, 2010). While the sequences defining the intron/exon boundaries are well defined in yeast, splice site sequences are much more degenerate in higher eukaryotes (reviewed by Keren et al, 2010), which may be related to the higher prevalence of alternative splicing in these organisms (Pan et al,

2008; Wang et al, 2008). The mechanisms by which the splicing machinery discriminates between bona fide splice sites and many other related sequences are therefore of great interest to reconcile the accuracy required to maintain gene expression with the versatility necessary to allow alternative splicing regulation.

3' splice sites in higher eukaryotes are composed of three main sequence elements. The branch site is an invariable adenosine residue which forms a 2'-5' phosphodiester bond with the 5' end of the intron after the first catalytic step of the splicing reaction. The branchpoint adenosine is flanked by sequences that establish base-pairing interactions with U2 snRNA, the RNA component of the U2 snRNP ribonucleoprotein, during formation of prespliceosomal (A) complex (Parker et al, 1987; Nelson & Green, 1989; Wu & Manley, 1989; Zhuang & Weiner, 1989; Zhang, 1998). The 3' end of the intron contains an invariable AG dinucleotide within the consensus CAG/G ("/' indicates the intron-exon boundary). A pyrimidine-rich stretch (the polypyrimidine [Py] tract) is often located between the branch site and the 3' splice site AG, although recent results in fission yeast indicate that other sequence arrangements can also be functional (Sridharan & Singh, 2007; Sridharan et al, 2011). The length and pyrimidine richness of the Py tract often correlate with the efficiency of 3' splice site recognition (Zamore et al, 1992; Singh et al, 1995).

The branch site sequence is recognized initially by the branchpoint binding protein (BBP/SF1) (Berglund et al, 1997; Liu et al, 2001) and subsequently through base-pairing interactions with U2. The Py tract and 3' splice site AG are recognized by the 65 kDa and 35 kDa subunits of the U2 snRNP auxiliary factor (U2AF), respectively (Zamore et al, 1992; Merendino et al, 1999; Wu et al, 1999; Zorio & Blumenthal, 1999). An interaction between the amino-terminal region of BBP/SF1 and the carboxy-terminal U2AF homology motif (UHM) of U2AF65 facilitates cooperative binding of these proteins to their relatively loosely defined cognate sites (Berglund et al, 1998; Selenko et al, 2003). Similarly, the tight interaction between the two subunits of the U2AF heterodimer stabilizes U2AF65 binding to the Py tract through recognition of the 3' splice site AG by U2AF35 (Zamore & Green, 1989; Zhang et al, 1992; Merendino et al, 1999; Wu et al, 1999; Zorio & Blumenthal, 1999). The arrangement of protein and RNA contacts in the initial complex is believed to enforce a particular RNA conformation in the 3' end of the intron that serves as a trigger for subsequent events in the splicing process (Kent et al, 2003).

U2AF65 contains two RNA recognition motifs (RRM1 and RRM2) that interact with Py tracts (Zamore et al, 1992; Singh et al, 1995; Banerjee et al, 2003; Banerjee et al, 2004), an amino-terminal arginine/serine-rich (RS) domain that facilitates base pairing between the branch site sequences and U2 snRNA (Lee et al, 1993a; Valcarcel et al, 1996) and a carboxy-terminal UHM, a RRM-like domain that engages in interactions with the SF3b155 protein component of U2 snRNP to facilitate U2 snRNP recruitment (Gozani et al, 1998). The domain structure of U2AF35 (Zhang et al, 1992) includes a UHM domain that has been shown to bind to the 3' splice site AG dinucleotide (Merendino et al, 1999; Wu et al, 1999; Zorio & Blumenthal, 1999). This interaction becomes essential for spliceosome assembly on AG-dependent introns characterized by relatively uridine-poor Py tracts that can only weakly recruit U2AF65 in the absence of the U2AF35/AG interaction (Merendino et al, 1999; Wu et al, 1999; Guth et al, 2001). It has been previously shown that purified U2AF heterodimer fails to provide discrimination between uridine-rich RNAs followed by a bona fide 3' splice site and a similar RNA containing a 3' splice site mutation, while proofreading activities present in nuclear extracts allow U2AF to discriminate between the two RNAs. The protein DEK is necessary, but not sufficient, for this discriminatory activity (Soares et al, 2006).

Multiple hnRNP proteins associate with nascent pre-mRNA molecules and influence RNA metabolism from transcription to RNA decay (Reed & Maniatis, 1986; McAfee et al, 1996; Krecic & Swanson, 1999; Singh & Valcarcel, 2005). A classical function of hnRNP A1 in splicing is to antagonize SR proteins to promote the use of distal splice sites and exon skipping (Mayeda & Krainer, 1992; Caceres et al, 1994; Mayeda et al, 1994). One proposed mechanism relies on glycine-rich domain-mediated intermolecular contacts between hnRNP A1 molecules, which loop out the intervening RNA (Blanchette & Chabot, 1999). In other instances, hnRNP A1 binding to specific splicing silencer elements interferes with spliceosome assembly (Tange et al, 2001; Kashima & Manley, 2003; Cartegni et al, 2006).

A more general role for hnRNP A1 in the splicing reaction has also been proposed. Selection of RNA sequences from a random pool (SELEX) identified high-affinity hnRNP A1 binding sites composed of one or more copies of the motif UAGGG(A/U) which has similarities with 3' splice site sequences (Swanson & Dreyfuss, 1988). hnRNP A1 has been proposed to display binding preference for both 5' and 3' splice

sites (Swanson & Dreyfuss, 1988; Buvoli et al, 1990b), and one study identified clustering of high-affinity hnRNP A1 binding sites at 3' splice sites, with mutations at the 3' splice site AG dinucleotide affecting hnRNP A1 binding (Burd & Dreyfuss, 1994). Interestingly, it has also been reported that hnRNP A1 interacts with U2 snRNP, therefore suggesting a role for hnRNP A1 in early stages of spliceosome assembly (Buvoli et al, 1992).

6.3 Results and discussion

Protein-specific isotope-labeled complexes were prepared and chemical shifts were analyzed using $^1\text{H},^{15}\text{N}$ correlation spectra (Figure 6.1). Large chemical shift changes were observed when comparing NMR spectra of isotope-labeled U2AF65 (assembled in the minimal heterodimer) alone and bound to AG-RNA or CG-RNA (Figure 6.1B and 6.1C). The RNA-bound states (black spectra in Figures 6.1B and 6.1C) show identical NMR spectra, indicating that -as expected- U2AF65 contacts the Py tract region and not the AG or CG nucleotides of these RNAs. Upon addition of hnRNP A1 and in the presence of CG-RNA, the NMR spectrum of U2AF65-labeled heterodimer is very similar to the free heterodimer (Figure 6.1B, magenta spectrum; notice the more extensive signal overlap between the green [free U2AF65] and the magenta [U2AF65 in the presence of CG-RNA and hnRNP A1] than with the black [U2AF65 bound to CG-RNA] spectra within the circled residues). This is consistent with the notion that A1 blocks the U2AF heterodimer from binding to CG-RNAs, as supported by the results of crosslinking and electrophoretic mobility shift experiments. In contrast, a higher-order complex is formed with AG-RNA, as indicated by line-broadening and disappearance of particular signals in the U2AF65-RNA binding interface (Figure 6.1C, magenta spectrum; notice, for example, the reduction of magenta signals for F202, G154, G301, or T296). The overall increased line widths are consistent with the formation of a higher molecular weight complex involving U2AF, hnRNP A1, and the AG-RNA. The chemical shifts of observable amide signals for a number of residues surrounding the RNA binding interface resemble the RNA-bound state of U2AF65, as expected (I160, L270 in Figure 6.1C).

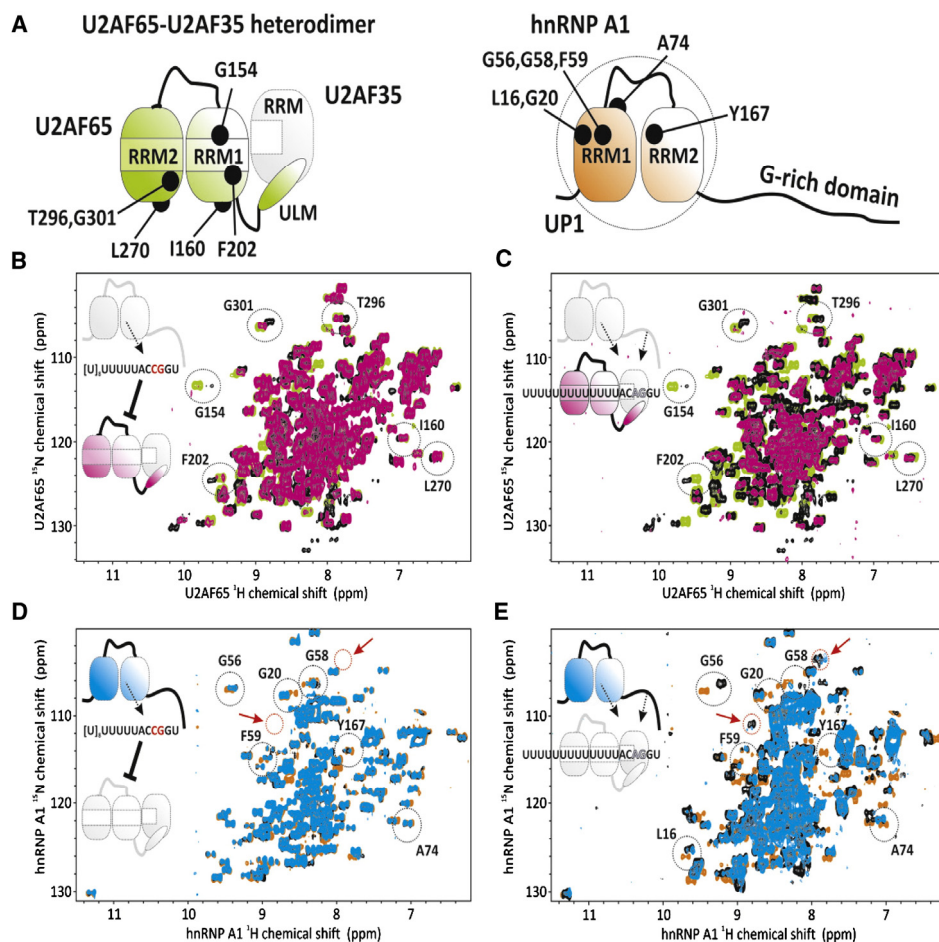


Figure 6.1 NMR Analysis of the U2AF/hnRNP A1 complexes with CG and AG RNA. (A) Diagram of the U2AF heterodimer representing domains of U2AF65 and U2AF35 (left) and of hnRNP A1 (right). Characteristic domains (RRM, RNA recognition motif; ULM, U2AF homology domain [UHM] ligand motif) and selected residues highlighted in the spectra are labeled. U2AF65 and hnRNP A1 are colored green and orange, respectively, with the same color code for the RNA-unbound proteins being used in the NMR spectra in (B)–(E). (B) Superposition of ^1H , ^{15}N CRINEPT-HMQC NMR spectra of the U2AF65-labeled U2AF heterodimer free (green), when bound to a 19 nucleotide CG-containing RNA (black) or in the presence of RNA and unlabeled hnRNP A1 (magenta). (C) Superposition of ^1H , ^{15}N CRINEPT-HMQC NMR spectra of U2AF65-labeled U2AF heterodimer free (green), when bound to a 19 nucleotide AG-containing RNA (black) or in the presence of RNA and unlabeled hnRNP A1 (magenta). (D) Superposition of ^1H , ^{15}N CRINEPT-HMQC NMR spectra of isotope-labeled hnRNP A1 free (orange), bound to the CG-containing RNA (black) and in the presence of RNA and unlabeled U2AF heterodimer (blue). (E) Superposition of ^1H , ^{15}N CRINEPT-HMQC NMR spectra of isotope-labeled hnRNP A1 free (orange) bound to the AG RNA (black) or in the presence of RNA and unlabeled U2AF heterodimer (blue). In all spectra, some characteristic residues affected by the binding are encircled and labeled. Two additional NMR signals which can be tentatively assigned to the glycine-rich tail of hnRNP A1 are encircled in red and marked by red arrows. Protein concentrations used for the NMR experiments were 50 μM (green, black, orange spectra) or 25 μM (magenta spectra), respectively. The CG- and AG-containing RNAs were added in a 1:1 heterodimer:RNA ratio. (Adapted from Tavanetz et al, 2012)

These conclusions are further supported by analysis of NMR spectra of hnRNP A1 (Figures 6.1D and 6.1E). Addition of CG-RNA to hnRNP A1 shows chemical shift perturbations consistent with formation of hnRNP A1 contacts with the RNA (Figure 6.1D, black spectrum). NMR spectra of isotope-labeled hnRNP A1 with CG-RNA are virtually identical in the presence or absence of U2AF (Figure 6.1D, compare black and blue spectra), confirming that hnRNP A1 binds to the CG-RNA independently of U2AF. The line widths and overall good quality of the NMR spectra are consistent with the formation of only a binary complex involving hnRNP A1 and CG-RNA, without the involvement of U2AF. NMR spectra of hnRNP A1 upon addition of AG RNA show large chemical shift perturbations, demonstrating the formation of a protein-RNA complex (Figure 6.1E, compare black and orange spectra). As observed with the NMR data of the U2AF65-labeled heterodimer, the spectra of hnRNP A1 bound to the AG-RNA in the presence of U2AF show line broadening and partial disappearance of NMR signals (Figure 6.1E, compare black and blue spectra), consistent with the formation of a higher-order complex involving hnRNP A1, U2AF, and the AG-RNA.

NMR spectra also provide evidence for the requirement of full-length hnRNP A1 for discriminating CG-RNAs versus AG-RNAs. Both hnRNP A1 and UP1 (which lacks the C-terminal glycine-rich tail) show significant and comparable chemical shift changes upon binding to AG or CG RNAs. However, some additional NMR signals are observed only in the complex with AG-RNA and full-length hnRNP A1 (red arrows/circles in Figures 6.1D and 6.1E). Based on their characteristic chemical shift, these signals can be tentatively assigned to glycines in the C-terminal tail of hnRNP A1. This suggests that the glycine-rich tail in hnRNP A1 mediates additional contacts with the RNA or protein components of the ternary complex, which are required for CG versus AG discrimination.

In summary, the NMR data are consistent with the biochemical assays and confirm the displacement of U2AF from CG RNAs by hnRNP A1, and the formation of a ternary complex containing hnRNP A1 and the U2AF heterodimer on AG RNAs. Collectively, the biochemical and structural analyses reported in this manuscript reveal that initial recognition of 3' splice sites by U2AF is subject to proofreading by hnRNP A1 such that U2AF binding to Py tracts not followed by a bona fide AG 3' splice site is destabilized and displaced. The presence of an AG accommodates both

subunits of U2AF and hnRNP A1 in a ternary complex that is compatible with U2AF recognition of the Py tract and with subsequent events in spliceosome assembly.

Part 2

Structure and dynamics in peroxisomal import

7. Biological background

7.1 Peroxisomes

Peroxisomes are single-membrane organelles found in all eukaryotes. They are involved in multiple metabolic pathways, such as β -oxidation of fatty acids and decomposition of toxic hydrogen peroxide (reviewed in Smith & Aitchison, 2013). Other functions of peroxisomes include purine and polyamine metabolism as well as alcohol oxidation and biosynthesis of plasmalogens. In addition, plant peroxisomes are involved in photorespiration and in the form of glyoxysomes they function in the glyoxylate cycle in germinating seeds. In *Trypanosoma*, specialized peroxisomes termed glycosomes harbor the first steps of glycolysis. At least 32 genes (named *PEX*) associated with peroxisome biogenesis and maintenance have been identified.

7.2 Peroxisomal disorders

Because peroxisomes are associated with a variety of metabolic functions, genetic defects in the peroxisome function can lead to a number of diseases. These can be subdivided into two categories: single-enzyme defects and peroxisome biogenesis disorders. The Zellweger syndrome is the most common peroxisome biogenesis disorder and Zellweger patients often die within their first year of life (reviewed in Fujiki et al, 2000).

7.3 Peroxisomal import

As peroxisomes lack genetic material, all of their proteins need to be imported from the cytosol. Unlike protein import into other organelles such as mitochondria and chloroplasts, peroxisomes can import folded and even oligomeric proteins. Interestingly, gold particles with a diameter of 9 nm conjugated to peroxisomal targeting signal (PTS) were shown to be imported into peroxisomes (Walton et al, 1995). However, the putative peroxisomal translocon has not been observed and a transient pore model has been proposed (Erdmann & Schliebs, 2005). Soluble peroxisomal matrix proteins are directed to the peroxisome by two distinct

peroxisomal import receptors, Pex5 and Pex7, which recognize two type of peroxisomal targeting signals, PTS1 and PTS2, respectively. The peroxisomal import pathway can be divided into four main steps: cargo recognition, receptor-cargo complex docking, cargo release and receptor recycling (Figure 7.1) (reviewed in Rucktaschel et al, 2011).

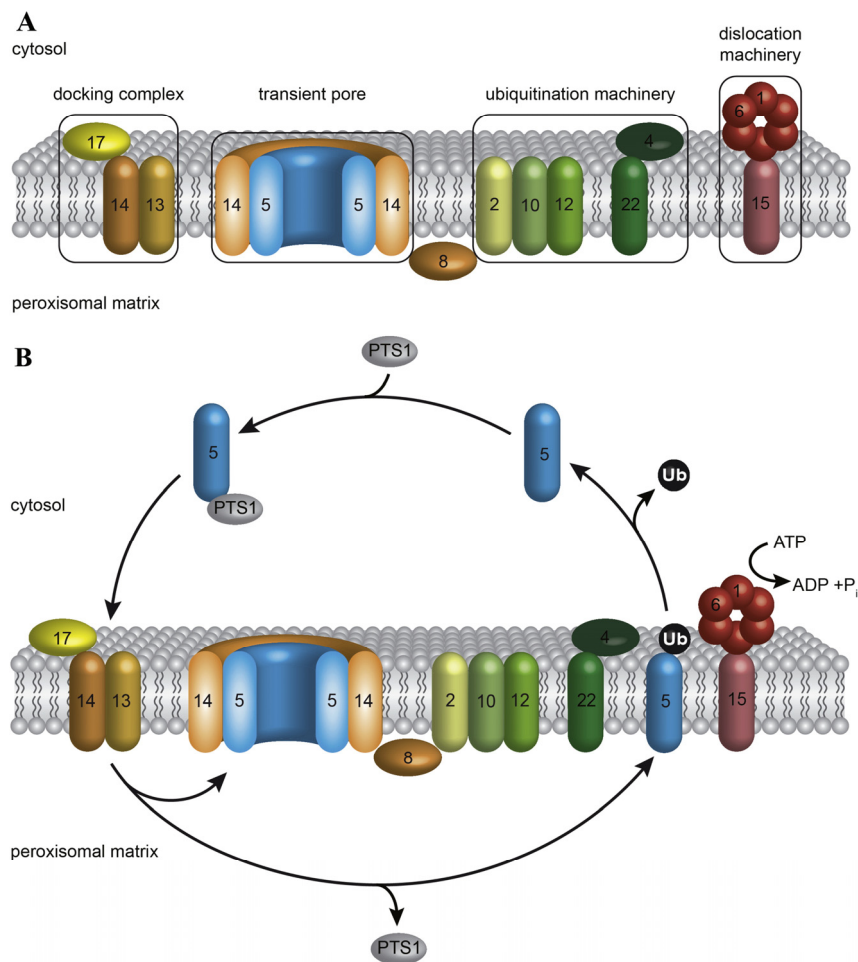


Figure 7.1 Peroxisomal matrix protein import. (A) Membrane complexes of the peroxisomal protein import machinery. (B) Model of the cycling PTS1 import receptor pathway. Proteins harboring PTS1 are recognized by the import receptor Pex5 in the cytosol. The cargo-loaded receptor is directed to the peroxisomal membrane and binds to the docking complex (Pex13/Pex14/Pex17). A transient pore is somehow formed and the cargo proteins are translocated across the peroxisomal membrane. Cargo release might involve the function of Pex8. The import receptor is monoubiquitinated at a conserved cysteine residue by the E2-enzyme complex Pex4/Pex22 and the RING-complex (Pex2, Pex10, Pex12) and is released from the peroxisomal membrane by the function of the AAA-peroxins Pex1 and Pex6, which are anchored via Pex15. The ubiquitin moiety is removed and the receptor may initiate the next import cycle. (Adapted from Platta et al, 2013)

Recently, it was shown that a complex consisting mainly of the PTS1-receptor Pex5 and its docking partner Pex14 forms a gated ion-conducting membrane pore that can be activated by cytosolic receptor-cargo complexes (Meinecke et al, 2010).

7.4 Aims and scope of the project

This part of the thesis is about protein-protein interactions involved in peroxisomal import. Protein import into peroxisomes is fundamentally different from other import machineries. It involves the association of the import receptor with components at the peroxisomal membrane to form a transient pore, through which the cargo can be translocated. The mechanism by which proteins are imported from the cytoplasm into peroxisomes is largely unknown and there are no available structural information on the peroxisomal translocon.

This project, which is a collaboration with the group of Ralf Erdmann (Bochum, Germany), aims at understanding the molecular details of the interaction between two central components of the peroxisomal import machinery, i.e the import receptor Pex5 and the docking protein Pex14. In addition, chapter 9 of this thesis includes a study on the interactions of Pex14 with tubulin which plays an important role in microtubule-based peroxisome motility.

8. A novel Pex14 interacting site in human Pex5

This chapter is a reprint of the following article:

Neuhaus A¹, Kooshapur H¹, Wolf J, Meyer NH, Madl T, Saidowsky J, Hambruch E, Lazam A, Jung M, Sattler M, Schliebs W, Erdmann R (2014) A novel Pex14 protein-interacting site of human Pex5 is critical for matrix protein import into peroxisomes.

J. Biol. Chem. 2014, 289: 437-448.²

¹ Both authors contributed equally to this work.

² The atomic coordinates have been deposited in the Protein Data Bank under accession code 4BXU. The NMR chemical shifts and restraints have been deposited in the Biological Magnetic Resonance Data Bank under accession number 19368.

8.1 Abstract

Protein import into peroxisomes relies on the import receptor Pex5, which recognizes proteins with a peroxisomal targeting signal 1 (PTS1) in the cytosol and directs them to a docking complex at the peroxisomal membrane. Receptor-cargo docking occurs at the membrane-associated protein Pex14. In human cells, this interaction is mediated by seven conserved diaromatic pentapeptide motifs (WxxxF/Y-motifs) in the N-terminal half of Pex5 and the N-terminal domain (NTD) of Pex14. A systematic screening of a Pex5-peptide library by ligand blot analysis revealed a novel Pex5-Pex14 interaction site of Pex5. The novel motif comprises the sequence LVAEF with the evolutionary conserved consensus sequence LVxEF. Replacement of the amino acids LVAEF sequence by alanines strongly affects matrix protein import into peroxisomes *in vivo*. The NMR structure of a complex of Pex5-(57-71) with Pex14-NTD showed that the novel motif binds in a similar α -helical orientation as the WxxxF/Y motif but that the tryptophan pocket is now occupied by a leucine residue. Surface plasmon resonance analyses revealed 33-times faster dissociation rates for the LVxEF ligand when compared to a WxxxF/Y motif. Surprisingly, substitution of the novel motif with the higher affinity WxxxF/Y motif impairs protein import into peroxisomes. These data indicate that the distinct kinetic properties of the novel Pex14-binding site in Pex5 are important for processing of the PTS1 receptor at the peroxisomal membrane. The novel Pex14 binding site may represent the initial tethering site of Pex5 from which the cargo-loaded receptor is further processed in a sequential manner.

8.2 Introduction

Import of peroxisomal matrix proteins is a cycling multi-step process comprising the post-translational recognition of fully-synthesized and folded peroxisomal proteins in the cytosol, transport to the peroxisome, docking of the receptor cargo complex at the organelle surface and receptor dependent formation of a transient protein-conducting channel (Ma et al, 2011; Nuttall et al, 2011). After translocation of the matrix proteins across the membrane, the receptors are released into the cytosol to initiate a further round of import (Platta et al, 2012).

For the cytosolic cargo recognition, folded and even oligomerized peroxisomal matrix proteins interact with peroxisome targeting signal (PTS) receptors Pex5 or Pex7 via cognate targeting sequences PTS1 or PTS2, respectively (Rucktaschel et al, 2011). Only a minor class of peroxisomal proteins contains the PTS2 signal, which corresponds to a nonapeptide sequence near the N-terminus. The majority of matrix proteins harbor PTS1 motifs which consist of the tripeptide -SKL (or a conserved variant thereof) at the extreme C-terminus of the protein. The PTS1 motif interacts with the tetratricopeptide repeat (TPR) domain in the C-terminal half of Pex5. The intrinsically disordered N-terminal half of Pex5 *per se* is capable to perform all transport steps of the receptor cycle, including docking, pore-formation and dislocation from the peroxisomal membrane (Gouveia et al, 2003; Schafer et al, 2004).

A peroxisomal membrane complex consisting of Pex13 and Pex14 (and Pex17 in yeast) provides the primary docking site for the cargo-loaded receptors (Albertini et al, 1997; Bottger et al, 2000). Pex14 fulfills additional important functions in cargo translocation across the peroxisomal membrane. For the model organism *Saccharomyces cerevisiae*, it was shown that Pex14 together with Pex5 constitutes a large dynamic channel at the peroxisomal membrane, supposed to act as a protein conducting pore (Meinecke et al, 2010). Recent data suggest that the Pex14-Pex5 interaction also plays a critical role for release of the cargo protein into the lumen of the peroxisome (Freitas et al, 2011).

Pex5 interacts with Pex14 via diaromatic pentapeptide motifs which were originally defined by the signature WxxxF/Y and are located in the N-terminal halves of all Pex5 proteins. The number of the conserved WxxxF/Y motifs varies depending on the species with for example two in yeast or seven in human Pex5 (Schliebs et al, 1999). It was shown that each of the seven WxxxF/Y motifs of the human PTS1 receptor can interact with the highly-conserved N-terminal domain (NTD) of Pex14 with equilibrium dissociation constants in the nanomolar range (Otera et al, 2000; Saidowsky et al, 2001). A recent biophysical study showed that the presence of multiple WxxxF/Y motifs in the N-terminus of Pex5 allows the formation of higher order complexes with the Pex14-NTD *in vitro*, although it is not clear whether this also reflects the situation in peroxisomes *in vivo* (Stanley et al, 2004; Shiozawa et al, 2009). Interestingly, human Pex19 which is supposed to act as an import

receptor/chaperone for peroxisomal membrane proteins, binds competitively to the same surface in Pex14-NTD via an amphipathic helix forming pentapeptide sequence (Neufeld et al, 2009). In contrast to typical Pex5 WxxxF/Y motifs, the identified FFxxxF of Pex19 binds with opposite directionality. The mode of binding is essentially similar as the ligand adopts an amphipathic α -helical conformation from which the protruding conserved aromatic side-chains are positioned into hydrophobic pockets formed by Pex14-NTD.

A crystallographic study (Su et al, 2009) and NMR analysis (Neufeld et al, 2009) of the free Pex14-NTD revealed that the Pex5 binding domain of Pex14 adopts a three helical fold. *Saccharomyces cerevisiae* Pex5 binds the N-terminal Pex14 domain via a short segment containing the reverse motif FxxxW (Kerssen et al, 2006), while the Pex14 binding site of *Leishmania* Pex5 has been mapped to a region without any sequence similarity to WxxxF/Y motifs (Madrid & Jardim, 2005). Recently, it was reported that Pex14 interacts with tubulin and that this interaction also depends on the Pex14-NTD (Bharti et al, 2011).

Taken together, these data demonstrate that the Pex14-NTD is a small structured domain that mediates versatile protein-protein interactions linked to peroxisomal function. For binding to the N-terminal Pex14 domain, ligands harbor the well-studied WxxxF/Y motif or variants thereof. However, not all conserved WxxxF/Y motifs form stable complexes with Pex14, indicating that the consensus sequence might be essential but not sufficient for Pex14 binding. Prominent examples are two typical WxxxF/Y motifs of *S. cerevisiae* Pex5 (Williams et al, 2005; Kerssen et al, 2006) or one of the three WxxxF/Y motifs of *Trypanosoma brucei* Pex5 (Choe et al, 2003). These sequences may mediate the interaction with other peroxins, as shown for human and yeast Pex5 motifs, which both can interact with the SH3 domain of Pex13 (Bottger et al, 2000; Douangamath et al, 2002; Otera et al, 2002; Pires et al, 2003).

In this study, we identified a novel LVxEF binding motif for human Pex14 in the N-terminal half of human Pex5. The new motif binds to Pex14-NTD with a comparable affinity but with notably distinct binding kinetics than the well-characterized di-aromatic WxxxF/Y motifs. NMR data and structural analysis revealed that the novel motif binds in a similar orientation as the WxxxF/Y motif but that the tryptophan

pocket is now occupied by a Leu residue. Interestingly, replacement of the new motif by a canonical WxxxF/Y motif with different binding properties impairs Pex5-dependent protein import. Our data suggest that the novel Pex14 binding site represents the initial tethering site of Pex5 from which the cargo-loaded receptor is further processed in a sequential manner.

8.3 Materials and methods

8.3.1 Construction of Pex5-fragments expression vectors

For DNA cloning, *E. coli* strain DH5 α was used. The human His-tagged Pex5 fragments Pex5-(1-113) and Pex5-(1-131) were amplified by PCR from pET9d-His₆-Pex5L (Schliebs et al, 1999) using the sense T7 promotor primer (Novagen) and the antisense primers KU582 (Pex5-(1-131)) and KU644 (Pex5-(1-113)) and subcloned into NcoI/SalI digested pET9d-His₆-Pex5L. The essential tryptophan in the first WxxxF/Y-motif (W118) was replaced by an alanine using the Quickchange XL Site-directed Mutagenesis kit (Stratagene) using primer pair KU690/KU1278. Pex5-(1-110) was amplified by PCR from pET9d-His₆-Pex5L (Schliebs et al, 1999) using primer pair 1-110 F and 1-110 R. The resulting PCR fragment was subcloned into a pETM11 vector (EMBL, Heidelberg) using NcoI and Acc65I restriction sites. Mutagenic disruption of the novel Pex14 binding site was achieved by substitution of the amino acid sequence LVAEF by AAAAA. Mutagenesis was carried out by overlapping PCR using the primer pairs (sense: T7 promotor primer / antisense: KU1102 and KU1101 / antisense T7 termination). As templates both pET9d-His₆-Pex5L and pET9d-His₆-Pex5-(1-131) were used.

Amplification products were subcloned in a modified *E. coli* expression plasmid pET24d, allowing N-terminal fusion of a His₆-tag (modified by G. Stier, EMBL, Heidelberg) by using XbaI and NotI restriction endonuclease recognition sites. In addition, the full length PCR product was restricted with NcoI and blunted before it was digested by NotI for subcloning into EcoRV/NotI digested mammalian expression vector pcDNA3.1-zeo+. All point mutations were verified by DNA-sequencing. The sequences of the primers used are listed in the Appendix.

8.3.2 Expression and purification of recombinant proteins

The genes coding for recombinant proteins were all expressed in *E. coli* strain BL21(DE3). Expression of the genes encoding His₆-tagged Pex5 proteins and GST-Pex14-(1-78) were carried out as described previously (Schliebs et al, 1999). Purifications of His₆-tagged Pex5-fragments using Ni-NTA agarose and ResourceQ (GE Lifescience) anion exchange chromatography and GST-Pex14-(1-78) were carried out as described previously (Schliebs et al, 1999).

Pex14-(16-80) and Pex5-(1-110) for NMR studies were expressed as fusion proteins with a His₆-tag followed by a TEV cleavage site as described previously for Pex14-(16-80) (Neufeld et al, 2009). ¹³C, ¹⁵N- or ¹⁵N-labeled proteins were expressed in minimal M9 medium supplemented with 2 g/l [¹³C]-glucose and/or 1 g/l [¹⁵N]-ammonium chloride, respectively. Protein samples for NMR studies were purified by Ni²⁺ affinity chromatography followed by TEV cleavage and removal of the His₆-tag by a second Ni²⁺ affinity chromatography step. The final step of purification consisted of size-exclusion chromatography on a HiLoad 16/60 Superdex75 column (GE Healthcare) in 20 mM sodium-phosphate buffer (pH 6.5), 100 mM NaCl and 1 mM DTT.

8.3.3 Peptide binding

Synthetic peptides were purchased from Peptide Specialty laboratories (Heidelberg, Germany) and dialyzed extensively against water before use in NMR studies. Immobilized peptides were synthesized by the Fmoc method of solid phase peptide chemistry on a Milligen/Biosearch 9050 Pep-synthesizer. Cleavage of the peptides from the resin and removal of the protecting groups was achieved by treating the peptide with 90% trifluoroacetic acid, 2.5% phenol, 2.5%, 1,2-ethanedithiol, 1% triisopropylsilane (all from Sigma) and 5% water (Hilpert et al, 2007). A peptide library of the protein sequences of human Pex5-(1-343) and a combinatorial library to analyze amino acid substitutions of Pex5-(59-67) were generated by the Spots technique according to the manufacturers protocol (ABIMED, Auto-Spot-Robot ASP 222) as described previously (Hilpert et al, 2007). The library representing Pex5-(1-343) consisted of 15-mer peptides with a 13 amino acid overlapping region which were immobilized on a cellulose membrane. The library was incubated with purified

His₆-tagged Pex14-(1-78) as described previously (Neufeld et al, 2009). Bound Pex14-(1-78) was immunodetected using monoclonal Anti-His Antibodies (Qiagen).

8.3.4 NMR spectroscopy

NMR experiments were recorded at 25 °C on 900, 600 or 500 MHz Bruker Avance NMR spectrometers equipped with cryogenic triple resonance gradient probes. All spectra were processed with NMRPipe (Delaglio et al, 1995) and analyzed in Sparky (T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco). NMR spectra for assignment and structure determination of the Pex14-(16-80)-Pex5-(57-71) complex were acquired on a sample containing ¹³C,¹⁵N-labeled Pex14-(16-80) bound to unlabeled Pex5-(57-71). Protein backbone and side-chain assignments were obtained using standard triple-resonance experiments (Sattler M et al, 1999). ¹H chemical shifts of unlabeled Pex5-(57-71) in the bound form were assigned based on ω1-filtered TOCSY and NOESY spectra. Inter-molecular NOEs between Pex14-(16-80) and Pex5-(57-71) were obtained from a ω1-filtered NOESY spectrum (Sattler M et al, 1999) with a mixing time of 80 ms. Paramagnetic relaxation enhancements (PREs) were recorded on Pex14-(16-80) in complex with either of two 16-mer Pex5 peptides (CASEDELVAEFLQDQN or ASEDELVAEFLQDQNC). The additional cysteine residue at the N- or C-terminus of each peptide was used for covalent attachment of the iodacetamido-proxyl spin-label (IPSL). PRE data were recorded and analyzed as described elsewhere (Simon et al, 2010).

8.3.5 Structure determination

The solution structure of the Pex14-(16-80)-Pex5-(57-71) complex was calculated with the previously established protocol using a modified version of Aria1.2/CNS (Simon et al, 2010). The structure of free Pex14-(16-80) (pdb code: 3FF5) (Su et al, 2009) was used as input for structure calculation and maintained using non-crystallographic symmetry restraints. Distance restraints were derived from inter (Pex14-Pex5)- and intra (Pex5)-molecular NOEs. Backbone torsion angle restraints were obtained from chemical shifts of Pex5-(1-110) in complex with Pex14-(16-80)

using TALOS+ (Shen et al, 2009). All restraints were employed during a molecular dynamics and simulated annealing run. The final structures were refined in a shell of water molecules (Linge et al, 2003) and validated using the iCing web interface (<http://nmr.cmbi.ru.nl/icing/>) (Doreleijers et al, 2012). Molecular images were generated with PyMol (Schrödinger, LLC).

8.3.6 Pex5–Pex14 interaction assays

Isothermal titration calorimetry (ITC) measurements were carried out at 25 °C using an iTC200 calorimeter (GE Healthcare). Pex14-(16-80) at concentration of ~ 600 µM in buffer A (20 mM sodium phosphate, 100 mM NaCl, pH 6.5) was injected in the cell containing Pex5-(1-110) at concentration of ~ 35 µM in buffer A. After correction for heat of dilution, the data were fitted to a one-site binding model using the Microcal Origin 7.0 software.

The interaction between Pex5 proteins and GST-Pex14-(1-78) was also studied using surface plasmon resonance spectroscopy with a BIAcore 2000 instrument (BIAcore AB) as described previously (Schliebs et al, 1999). Data were evaluated with the BIAevaluation software version 4.1.1 (BIAcore AB) using the 1:1 Langmuir binding fit. Interaction assays using size exclusion chromatography were performed on a Superose 6 PC 3.2/30 column (GE Healthcare) as described previously (Schliebs et al, 1999).

8.3.7 *In vivo* studies

Pex14-binding sites in full-length Pex5 were disrupted either individually or simultaneously by substituting amino acids 62 to 66 by penta-alanine (LVAEF→AAAAA) or by replacing essential tryptophans of respective WxxxF/Y-motifs, W118 (W118A) and W140 (W140A) by alanine. Mutagenesis of LVAEF was carried out by overlapping PCR using mutagenic primers KU1101 and KU1102 and pcDNA3-PEX5S (Dodt et al, 1995) as a template. Substitution of the tryptophans of the first and the second WxxxF/Y-motif with alanines was achieved by using Quickchange XL Site-directed Mutagenesis kit (Stratagene) and primer pairs KU690/KU1278 and KU1255/KU1279, respectively. For substitution of LVAEF by

WAQEF, primers RE2142/RE2143 were used. A bicistronic expression plasmid coding for EGFP-SKL and non-tagged Pex14 (pMF120; (Huybrechts et al, 2008)) was kindly provided by Marc Fransen (Leuven, Belgium). pIRES2-EGFP-SKL was constructed by replacing the open reading frame of PEX14 with the original multiple cloning site from pIRES2-EGFP. The *PEX5S* open reading frames with and without mutations were amplified using primer pair RE3533/3534 and subcloned into pIRES-EGFP-SKL using BglIII and Sall restriction sites.

The bicistronic expression vectors coding for Pex5 variants and EGFP-PTS1 were transfected into the human cell-line Δ PEX5T which was derived from Pex5-deficient Zellweger patient fibroblasts (Bharti et al, 2011). Forty-eight hours after transfection, cells were subjected to immunofluorescence microscopy using polyclonal rabbit antiserum against human Pex14 and monoclonal antibody against AFP (MP Biomedicals, Heidelberg). For quantitative analysis, three to six independent transfections of each pIRES-Pex5-EGFP-SKL expression plasmid were monitored. Based on the appearance of EGFP-SKL fluorescence pattern, about 100 cells of each experiment were visually categorized in full or partial complementation of peroxisomal import as visualized by punctuate staining pattern or no complementation indicated by diffuse non-punctuate cytosolic background staining. All micrographs were recorded on a Zeiss Axioplan 2 microscope with a Zeiss Plan-Apochromat 63 \times /1.4 oil objective and an Axiocam MR digital camera and were processed with AxioVision 4.6 software (Zeiss, Jena, Germany). The steady-state level of Pex5 expression was assessed by immunoblot analyses of transfected cells using polyclonal rabbit antibodies against HsPex5. Anti-prohibitin antibodies (Abcam) were used as loading control.

8.4 Results

8.4.1 Identification of a novel Pex14-binding site in the N-terminal domain of human Pex5

In order to identify the Pex14 binding sites of human Pex5, we performed a peptide scan analysis. The membrane array consisted of synthetic 15-mer peptides that sequentially overlap by 13 residues and represented Pex5-(1-343) comprising the

entire sequence of the Pex14-interacting region (Schliebs et al, 1999). The cellulose-bound peptides were incubated with His-tagged Pex14-NTD and bound protein was visualized by immunodetection with anti-His antibodies. Peptides containing one of the seven known WxxxF/Y-motifs did bind the Pex14 probe resulting in a staining pattern of clusters of four to six successive peptides (Figure 8.1A, white boxes). In accordance with previously reported dissociation constants (K_D) for WxxxF/Y-binding (Saidowsky et al, 2001), the strongest immunological labeling was obtained with peptides containing either the first WxxxF/Y sequence (W1, amino acids 118-122, spots 56-60) or the fifth WxxxF/Y motif (W5, amino acids 243-247, spots 117-122). Surprisingly, the peptide scan revealed, in addition to the seven known Pex14 binding sites, staining for five consecutive peptides which do not contain a WxxxF/Y motif (Figure 8.1). The overlapping region of the new binding site represents 7 amino acids with the sequence ELVAEFL. The core sequence LVAEF resembles Pex5 WxxxF/Y motifs with the major difference that tryptophan is replaced by the hydrophobic amino acid leucine (Figure 8.1A).

To further characterize the sequence requirements of the novel Pex14-binding motif, we carried out a peptide scan substitution analysis. To this end, 15-mer peptides corresponding to the Pex5 sequence ⁵⁹EDELVAEFL⁶⁷ were synthesized to harbor a single amino acid exchange at each position of the minimal binding sequence (Figure 8.1B). The major difference between the novel LVAEF and canonical WxxxF/Y motifs concerns the two first residues. Whereas in WxxxF/Y motifs only the tryptophan is invariant for interaction with Pex14 (Saidowsky et al, 2001; Neufeld et al, 2009), in the mono-aromatic ⁶²LVAEF⁶⁶ binding sequence, two hydrophobic residues Leu62 and Val63 are critical for Pex14 binding. However, the leucine residue in the new motif is interchangeable with the aromatic residues tryptophan, phenylalanine and tyrosine (Figure 8.1B). Interestingly, the peptide containing tryptophan instead of leucine seems to bind to Pex14-NTD with a higher affinity than the original sequence. Common to all Pex14 binding motifs of human Pex5 are the invariant position of an aromatic residue, either phenylalanine or tyrosine at position 5 of the pentapeptide as well as the preference for a glutamate in position 4.

Sequence alignment of known PTS1 receptors revealed that a LVxEF sequence motif preceding the first canonical WxxxF/Y motif is conserved in many metazoan (Figure 8.1C). In *Caenorhabditis elegans*, the residues in positions 1 and 2 of the

corresponding sequence are methionine and alanine, respectively, both of which would interfere with binding of the peptide to human Pex14. Remarkably, the LVxEF sequence motif is absent in Pex5 from plants and yeasts but present in Pex5 of several filamentous fungi, including *Aspergillus fumigatus*.

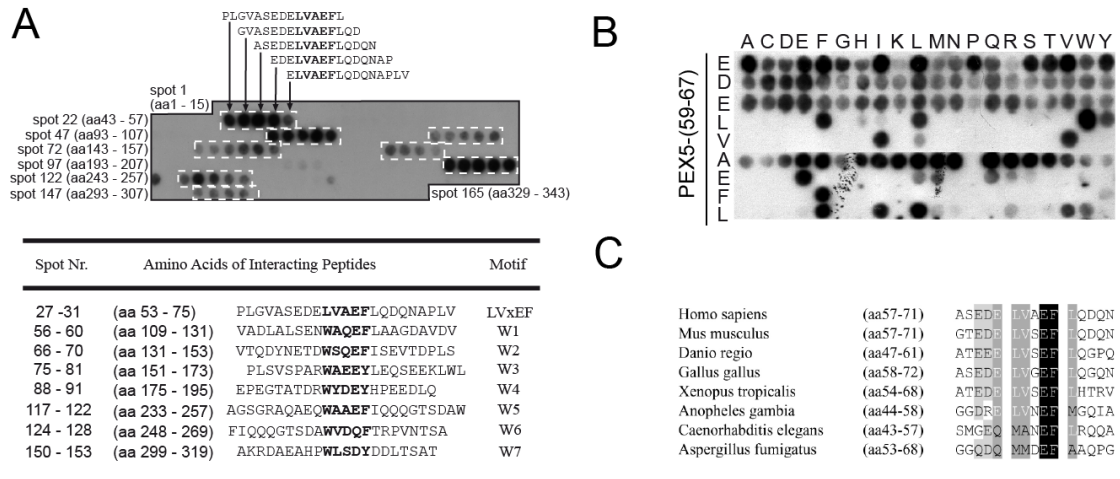


Figure 8.1 Identification of a novel Pex14-binding motif in the N-terminal domain of the human PTS1 receptor. (A) Peptide scan analysis of Pex5 for Pex14-binding sites. 15-mer peptides with two amino acids shifts between successive peptides representing amino acids 1 to 343 of Pex5, were synthesized on a cellulose membrane. The first spot of each row is indicated by numbers. After incubation with purified His-tagged Pex14-NTD, interacting peptide spots were detected with anti-His antibodies. Sequences of interacting peptide arrays (dashed boxes), including the novel binding motif (spots 27 to 31) are shown, and the overlapping amino acids of all peptides within each array are highlighted by grey colour, core sequences with highest structural similarity are indicated by bold letters. (B) Substitution analysis of the conserved residues of the novel Pex14-binding region of human Pex5. Each row represents a variant of human Pex5-(59-67) sequence, EDELVAEFL in which the indicated core amino acid was replaced with any of the 20 proteinogenic amino acids. His-tagged Pex14-NTD interacting peptides were detected by immunoblot analysis using anti-His antibodies. (C) LVxEF motifs within the N-terminal domain of Pex5p are conserved among several groups of metazoans. Representative sequences are shown. Similar sequences with the consensus MxxEF were detected in Pex5 sequences of *Caenorhabditis elegans* and filamentous fungi represented here by *Aspergillus fumigatus*.

8.4.2 NMR analysis and structure of the complex of Pex14-NTD with the novel Pex5 LVxEF motif

To further characterize the binding of Pex5 to Pex14-NTD using NMR spectroscopy, we studied a construct comprising only the first 110 residues of human Pex5, which lacks any of the known WxxxF/Y motifs. NMR fingerprint spectra of ^{15}N -labeled Pex5-(1-110) show very little dispersion, characteristic of an unstructured protein. However, upon addition of unlabeled Pex14-NTD the spectrum changes significantly, indicating a strong interaction between Pex5-(1-110) and Pex14-NTD (Figure 8.2A). Analysis of $^{13}\text{C}\alpha/\beta$ secondary chemical shifts revealed that the region comprising residues E59 to Q70 of human Pex5 ($^{59}\text{EDELVAEFLQDQ}^{70}$) forms an α -helix upon interacting with Pex14-NTD (Figure 8.2B). The identified α -helix contains the core sequence LVAEF, matching the sequence obtained from our peptide scan analysis.

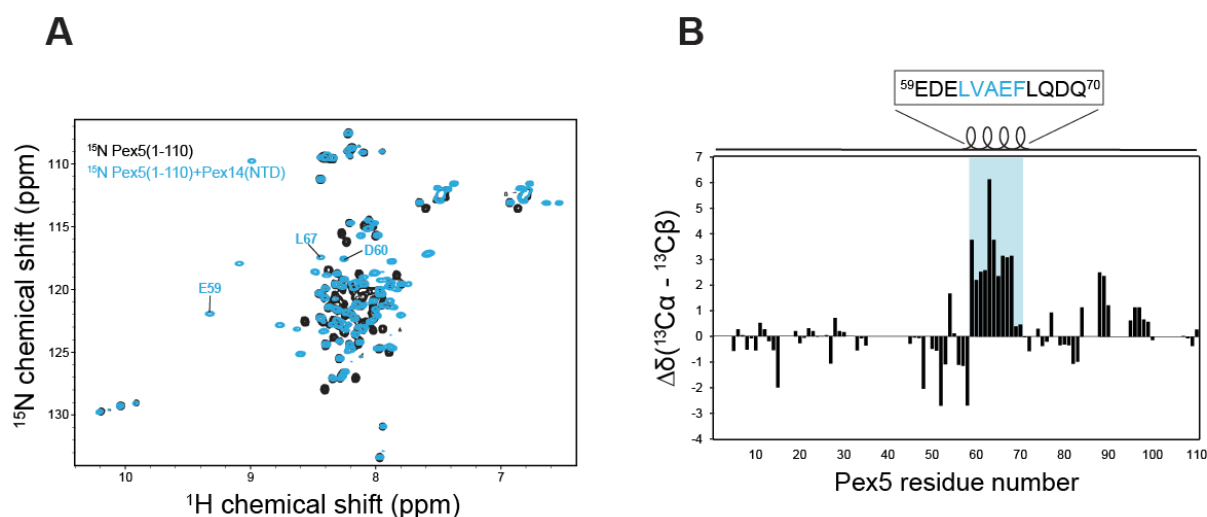


Figure 8.2 Residues E59 to Q70 of Pex5 form an α -helix upon interacting with Pex14-NTD. (A) ^1H , ^{15}N HSQC spectra of ^{15}N -Pex5(1-110), free (black) and in complex with Pex14-NTD (cyan). Few residues undergoing large chemical shift perturbations are labeled. (B) Plot of $^{13}\text{C}\alpha/\beta$ secondary chemical shifts of Pex5-(1-110) in complex with Pex14. Positive values indicate an α -helical conformation for residues E59-Q70 of Pex5 which harbor the novel LVxEF motif (highlighted in cyan).

The binding site of the LVxEF motif on the structure of Pex14-NTD was mapped based on chemical shift perturbations seen upon addition of a 16-mer Pex5 peptide containing the LVAEF sequence to ^{15}N -labeled Pex14-NTD (Figure 8.3A). The

chemical shift perturbations (Figure 8.3A) involve the same set of residues that have been previously shown to bind to a WxxxF/Y motif of Pex5, indicating that the new motif interacts with the same surface of Pex14-NTD (Neufeld et al, 2009). In order to elucidate molecular details for the recognition of the novel Pex5 motif by Pex14-NTD, we determined the solution structure of the complex. The structure of the complex was determined based on the previously reported structure of Pex14-NTD (Neufeld et al, 2009; Su et al, 2009), using a protocol that we have recently established (Simon et al, 2010) and included 47 intermolecular NOE-derived distance restraints between Pex14-NTD and Pex5-(57-71) (Table 8.1). The Pex14-NTD adopts a three-helical bundle in which helices $\alpha 1$ and $\alpha 2$ constitute the main binding interface. The Pex5 LVxEF motif binds in an α -helical conformation to the hydrophobic surface of Pex14-NTD (Figure 8.3B). Three conserved residues (Leu62, Val63 and Phe66) of the LVxEF motif are involved in hydrophobic interactions with Pex14-NTD. Phe66 is inserted into a hydrophobic pocket in Pex14 and engages contacts with Pex14 Phe35 and Phe52 on helices $\alpha 1$ and $\alpha 2$, respectively. Leu62 is not deeply buried in the second hydrophobic pocket of Pex14 but interacts with the aliphatic side chain of the conserved Lys56 in Pex14. In addition, Pex5 Val63 is involved in hydrophobic interactions with Pex14 Val41 and Phe35. The conserved Val41 is centrally located in the short 3_{10} -helix of Pex14-NTD between helices $\alpha 1$ and $\alpha 2$ and makes contacts to both Val63 and Phe66 in the LVxEF motif of Pex5 (Figure 8.3C). Thus, similar to the previously reported structures, where hydrophobic pockets in Pex14-NTD accommodate aromatic side chains of WxxxF/Y motifs or the inversely oriented FFxxxF motif in Pex5 and Pex19, respectively (Neufeld et al, 2009), the LVxEF motif uses the same pocket for recognition of phenylalanine. In contrast, recognition of the LV sequence in the newly identified LVxEF motif is distinct from the interactions with Phe or Trp in canonical WxxxF/Y motifs (Figure 8.3C).

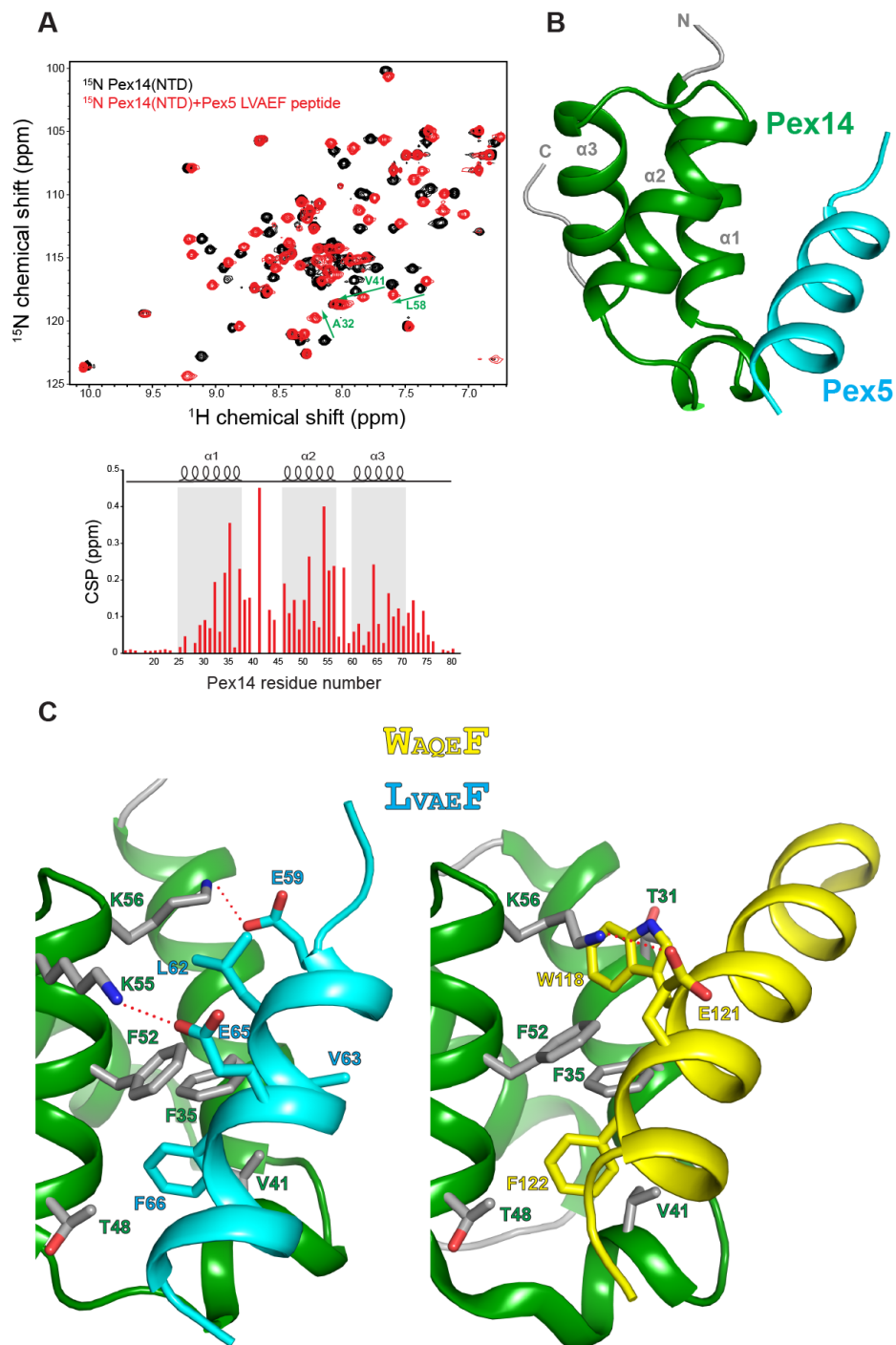


Figure 8.3 (A) ^1H , ^{15}N HSQC spectra of ^{15}N -Pex14-NTD, free (black) and titrated with two-fold excess of Pex5-(57-71) (red). Plot of chemical shift perturbations (CSP) is shown. Shaded regions correspond to helices $\alpha 1$, $\alpha 2$ and $\alpha 3$. (B) Ribbon representation of the lowest-energy NMR structure of the Pex14-NTD-Pex5-(57-71) complex. (C) Structural comparison of Pex14 interaction with LVAEF (left) and WAQEF (right) peptide ligands of Pex5. Side chains of residues involved in intermolecular interactions are shown as sticks. The recognition of Phe (in LVAEF and WAQEF) by Pex14 is similar in the two structures; however, the main difference lies in the recognition of LV (in LVAEF) and W (in WAQEF) residues. Plausible salt-bridges between Glu and Lys side chains are indicated by red dashed lines in both structures.

Table 8.1 NMR and structural statistics

NMR restraints	
Total NOE distance restraints	173
Inter (Pex14-Pex5 peptide)	47
Intra (Pex5 peptide)	126
PRE-derived distance restraints	113
Hydrogen-bonds	16
$\Phi+\psi$ torsion angle restraints**	30
Restraint violations	
Distance violations >0.3 Å	0.0
Dihedral angle violations >5°	0.0
r.m.s.d. from experimental distance restraints (Å)	0.031 ± 0.007
r.m.s.d. from experimental torsion angle restraints (°)	0.698 ± 0.273
Deviations from idealized geometry	
Bonds (Å)	0.005 ± 0.000
Angles (°)	0.668 ± 0.007
Impropers (°)	0.965 ± 0.083
Coordinate precision r.m.s.d. ***	
Backbone (Å)	0.56 ± 0.16
Heavy atom (Å)	1.17 ± 0.07
Ramachandran plot analysis ***	
Residues in most favoured regions	97.5%
Residues in additionally allowed regions	2.5%
Residues in generously allowed regions	0%
Residues in disallowed regions	0%

* Statistics are reported for the 10 lowest-energy structures from a total of 100 calculated.

** Dihedral angle restraints were derived from TALOS+.

*** Applies to residues 25-70 of Pex14 and 57-71 of Pex5.

Electrostatic interactions also play an important role in ligand binding by Pex14-NTD. As described previously, the binding interface of Pex14-NTD forms a highly positively charged surface complementary to the negatively charged Pex5 and Pex19 ligands (Neufeld et al, 2009). The structure of the Pex14-Pex5 LVxEF motif suggests that Lys55 and Lys56 of Pex14 form salt-bridges with Glu65 and Glu59 of Pex5, respectively (Figure 8.3C). According to our peptide scan analysis, Glu65 in the Pex5 LVxEF motif is critical for Pex14 binding. Furthermore, charge-reversal substitution of Glu59 located C-terminally to the LVxEF motif leads to loss of Pex14 binding (Figure 8.3B). Thus, electrostatic interactions make important contributions to the recognition of the Pex5 LVxEF motif by Pex14.

Our previous structural studies have shown that Pex14-NTD can interact with helical motifs of Pex5 and Pex19 in opposite orientations. The Pex5 LVAEF peptide adopts the same directionality as the previously reported Pex5 WxxxF/Y peptide (Neufeld et al, 2009). To further validate the orientation of the novel Pex5 LVxEF motif helix, we used PRE data for two spin-labeled peptide ligands with the spin-label attached either to the N- or C-terminus. Addition of a spin-labeled peptide to Pex14-NTD leads to line-broadening of amide protons which are in the vicinity of the spin-label. Analysis of the PRE data for both peptides show that experimentally observed line-broadening effects are in agreement with the theoretically calculated ones based on our NMR structure (Figure 8.4). Thus, the PRE data independently confirm that the Pex5 LVxEF motif binds to Pex14 in the same orientation as the Pex5 WxxxF/Y peptide.

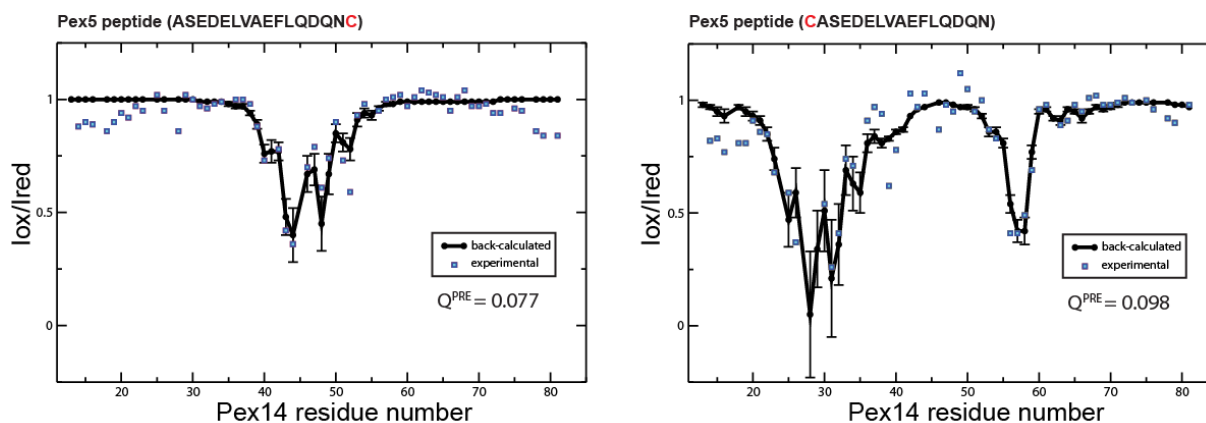


Figure 8.4 Paramagnetic relaxation enhancement (PRE) experiments confirm the orientation of the Pex5-(57-71) peptide bound to Pex14-NTD observed in the NMR structure. Comparison of experimental (blue squares) and back-calculated PRE data (black circles, error bars indicate variations across the NMR ensemble) obtained for Pex14-NTD in complex with C-terminally (left panel) and N-terminally (right panel) spin-labeled Pex5 peptides. In both cases, there is an excellent agreement between experimental and back-calculated values, as judged from the low value of the quality factor Q^{PRE} . The paramagnetic spin-label was attached to a terminal cysteine residue added to the Pex5 peptide (shown in red).

8.4.3 Equilibrium and kinetic binding properties of Pex14-NTD:Pex5 complexes

We determined the binding affinities and kinetic rate constants for the complex of Pex14-NTD with the novel Pex5 motif using isothermal titration calorimetry (ITC) and surface plasmon resonance (SPR) experiments. The ITC data show that Pex5-(1-110) binds to Pex14-NTD with a dissociation constant (K_D) of 157 ± 9 nM (Figure 8.5A), demonstrating high affinity binding, similar to the WxxxF/Y motifs (Saidowsky et al, 2001; Neufeld et al, 2009). The stoichiometry of binding was 1:1, confirming that the identified binding sequence represents the only accessible interacting site within the N-terminal 110 residues of Pex5.

For analysis by surface plasmon resonance (SPR) experiments, GST-Pex14-NTD was coupled to the surface of sensor chips and incubated with increasing concentrations of N-terminal Pex5 fragment. The LVxEF containing Pex5-(1-117) interacted with Pex14-NTD with a K_D of 87 nM (Table 8.2). Interestingly, a longer N-terminal fragment, Pex5-(1-131) which harbors both the LVxEF and an additional WxxxF/Y motif (W1) exhibited much higher binding affinity with a K_D of 4.6 nM.

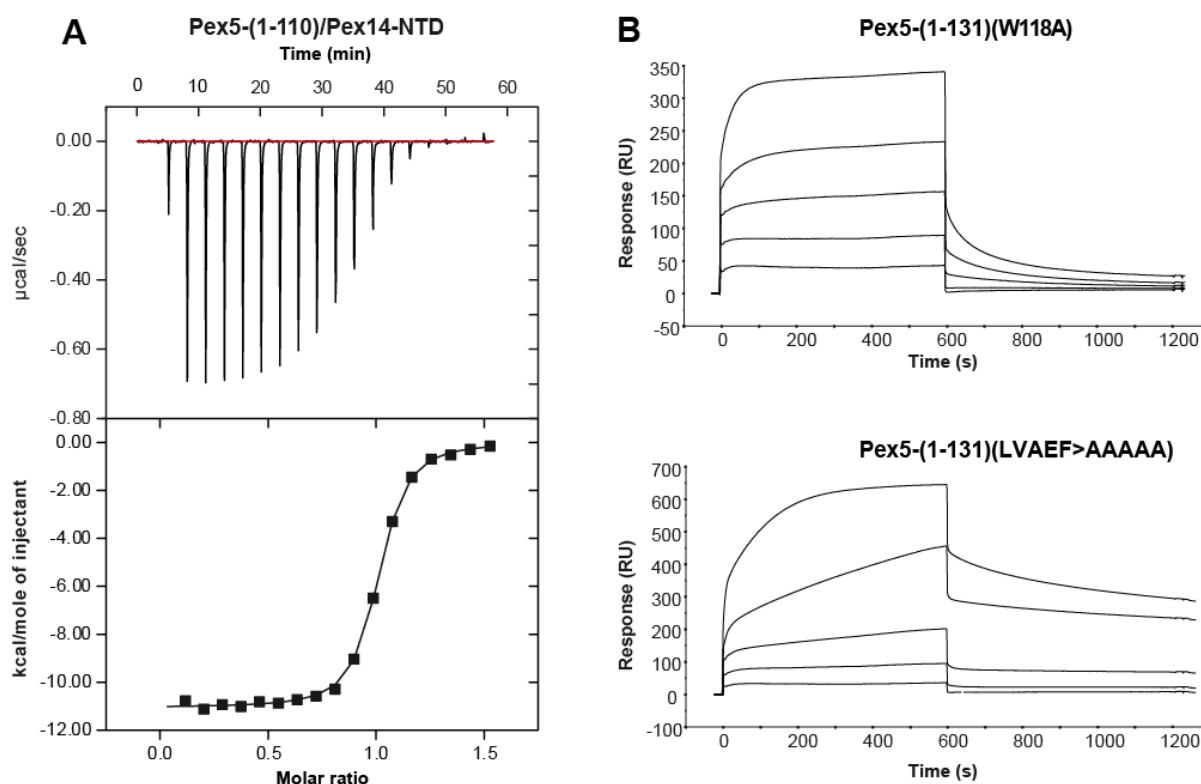


Figure 8.5 Rate and equilibrium binding constants of the interaction of regions of Pex5 and Pex14. (A) Isothermal titration calorimetry (ITC) curve of the interaction between Pex5-(1-110) and the N-terminal domain of Pex14 (Pex14-NTD). Pex14-NTD was injected into the cell, containing Pex5-(1-110) and the heat of binding was measured per mole of injectant. The raw data were fitted to a one-site binding model, which yielded a K_D of 157 ± 9 nM with a stoichiometry of 0.97 ± 0.002 . (B) For surface plasmon resonance spectroscopy analyses, 700 RU of Purified GST-Pex14-NTD were immobilized on an anti-GST sensor chip surface and varying concentrations of purified Pex5-(1-131)(W118A) and Pex5-(1-131)(LVAEF \rightarrow AAAAA) were applied. The concentrations used for both Pex5-(1-131) constructs were: 1 nM, 4 nM, 15 nM 60 nM and 240 nM (graphs from bottom to top). Analyses were performed as described under Material and Methods, and the results are summarized in Table 8.2.

To analyze and compare the kinetic properties for the interaction of the novel and canonical Pex14 binding motifs, fragments containing mutations disrupting either the novel Pex14-binding site or the WxxxF/Y motif W1 (WAQEF) were analyzed (Figure 8.5B). The kinetic binding constants of monovalent complex formation were assessed by fitting the surface plasmon resonance binding curves assuming the 1:1 Langmuir binding model.

Inactivation of the Pex14-binding W1 motif was achieved by substitution of the tryptophan by alanine (Schliebs et al, 1999). The LVxEF-containing Pex5-(1-131) (with inactive W1) exhibited similar binding kinetics as the shorter fragment Pex5-(1-117) (Table 8.2). To study the kinetics of W1-containing Pex5-(1-131), the newly identified LVxEF motif was replaced by penta-alanine. As shown by *in vitro* binding assays, this mutation abolishes the monovalent interaction between N-terminal fragments of Pex5 and Pex14 (Figure 8.6).

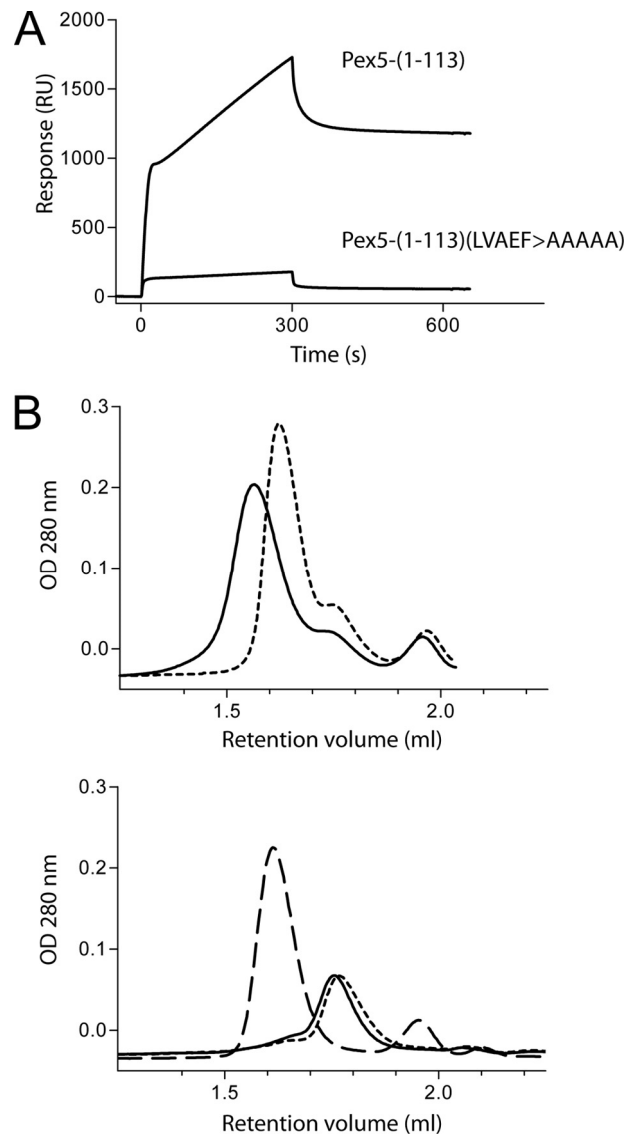


Figure 8.6 Substitution of LVxEF motif with AAAAA inhibits *in vitro* binding of an N-terminal fragment of Pex5 to the N-terminal domain of Pex14. (A) Surface plasmon resonance spectroscopy shows that Pex5-(1-113) binds to immobilized GST-tagged N-terminal domain of Pex14 (GST-Pex14-NTD), while Pex5-(1-113)(LVAEF→AAAAA) lacking the LVxEF motif does not. For the binding experiments, concentration of the Pex5 fragments was 50 nM. (B) The isolated N-terminal domain of Pex14 (Pex14-NTD) was pre-incubated with wild type or the mutant Pex5-(1-113), which lacks the LVxEF motif and the mixtures were analyzed by size exclusion chromatography. The curves in the upper panel were recorded after applying mixtures of GST-Pex14-NTD either with Pex5-(1-113) (solid line) or with Pex5-(1-113)(LVAEF→AAAAA) (dotted line). The data show that incubation of GST-Pex14-NTD with the Pex5-(1-113) results in a shift to higher molecular weight, indicative of complex formation. This shift is not observed with the mutated Pex5 fragment Pex5-(1-113)(LVAEF→AAAAA). As control, the lower panel displays gel-filtration profiles of the purified fragments in the absence of binding partners: GST-Pex14-NTD (dashed line), Pex5-(1-113) (solid line) and Pex5-(1-113)(LVAEF→AAAAA) (dotted line). Optical density, OD, of the column effluent was monitored at 280 nm.

The fragment containing the functional WxxxF/Y motif displayed a 3-fold slower association rate (k_{on}) and a 33-fold reduced dissociation rate (k_{off}) compared to the same fragment containing only the functional LVxEF motif (Table 8.2). Thus, the novel LVxEF motif exhibits significantly faster binding kinetics with Pex14 than the adjacent WxxxF/Y motif.

Table 8.2 Rate and equilibrium binding constants of the interaction of immobilized GST-Pex14-NTD and Pex5 proteins. Association rate constants, k_{on} , dissociation rate constants, k_{off} , and equilibrium binding constants, K_D , were determined by surface plasmon resonance spectroscopy, using GST-Pex14-NTD immobilized on an anti-GST surface as ligand and the various Pex5 proteins as analytes. Disruption of Pex14-binding sites, either of the new motif (LVAEF) or the N-terminal WxxxF/Y motifs (W118 and W140) by alanine replacements are indicated. For each Pex5 construct, the number of remaining Pex14-binding sites is shown.

Pex5-	Pex14 binding site	k_{on} [$M^{-1}s^{-1}$]	k_{off} [s^{-1}]	K_D [M]
1-117 (LVAEF)	1	7.0×10^4	6.1×10^{-3}	87×10^{-9}
1-131 (LVAEF, ΔW1)	1	7.5×10^4	6.6×10^{-3}	87×10^{-9}
1-131 (ΔLVAEF, W1)	1	2.6×10^4	0.2×10^{-3}	9.2×10^{-9}
1-131 (LVAEF, W1)	2	31×10^4	1.4×10^{-3}	4.5×10^{-9}
full-length	8	55×10^4	1.2×10^{-3}	2.2×10^{-9}
full-length (ΔLVAEF)	7	67×10^4	1.7×10^{-3}	2.5×10^{-9}
full-length (ΔLVAEF ΔW1 ΔW2)	5	65×10^4	2.8×10^{-3}	4.3×10^{-9}

8.4.4 Functional analysis of the novel Pex14 binding site

To study the physiological relevance of the LVxEF motif, we replaced these amino acids by five alanines in the context of full-length Pex5 and tested its ability to rescue the peroxisome import defect of Pex5-deficient fibroblast cells. For this purpose, a pIRES vector was constructed which encodes a bicistronic mRNA for full-length Pex5 harboring the respective point mutations and EGFP fused with the peroxisomal matrix protein targeting sequence –SKL (EGFP-PTS1) at its carboxy terminus. 48 hours after transfection, cells were inspected by fluorescence microscopy. A peroxisomal punctate pattern indicates either partial or full complementation of the import phenotype (Figure 8.7A). Wild-type Pex5 rescued the PTS1 import defect of

most of the transfected cells (82%) whereas the LVAEF→AAAAA mutant restored peroxisomal protein import in only 43% of the cells. Thus, most cells exhibited no punctuate staining, indicative for a complete mislocalization of the peroxisomal marker protein to the cytosol and thus a severe import defect. Moreover, almost all cells that still exhibited a punctuate staining upon expression of the LVAEF→AAAAA mutant displayed a partial import defect with strong cytosolic background (Figure 8.7A, lower panel). The data indicate that inactivation of the novel Pex14-binding motif strongly affects the efficiency of peroxisomal protein. Importantly, the inactivation of the LVxEF motif (LVAEF→AAAAA) showed a more severe effect on matrix protein import than mutations of the adjacent first WxxxF/Y motif (W118A) or the second WxxxF/Y motif (W140A). When these mutations were combined with mutation of the LVAEF sequence (LVAEF→AAAAA, W118A, W140A), the complementing activity of the triple mutant protein was almost completely abolished (Figure 8.7A). Remarkably, the severe import defect of the triple mutation does not correlate with a notable loss of affinity with Pex14 as indicated by SPR binding experiments with purified proteins (Table 8.2). Figure 6B shows that the different complementing activity of wild-type and mutant Pex5 is not due to different steady-state levels of the proteins.

We next asked whether a canonical WxxxF/Y sequence can functionally replace the novel mono-aromatic LVxEF motif. To this end, the residues Leu62, Val63 and Ala64 of the LVAEF sequence were substituted by tryptophan, alanine and glutamine. The corresponding penta-peptide sequence (⁶²WAQEF⁶⁶) matches exactly the sequence of the N-terminal WxxxF/Y motif. Remarkably, this mutant displays a similar phenotype as the Pex5 variant which lacks a functional Pex14-binding site at this position (Figure 8.7A). Most of the cells transfected with the substitution mutant (LVAEF → WAQEF) exhibit a severe import defect. Apparently, a WxxxF/Y motif cannot fully account for the LVxEF motif, indicating that not only the presence of a Pex14-binding site at this N-terminal position determines import efficiency but that also the kinetic properties of the binding site play a critical role. Replacement of the first WxxxF/Y motif by LVAEF did not affect the complementing activity of full-length Pex5 or of the variants Pex5-(LVAEF→AAAAA) and Pex5-(LVAEF→WAQEF) (data not shown). This further supports the view that the fast on/off rates of Pex14 binding are only critical at the N-terminal position.

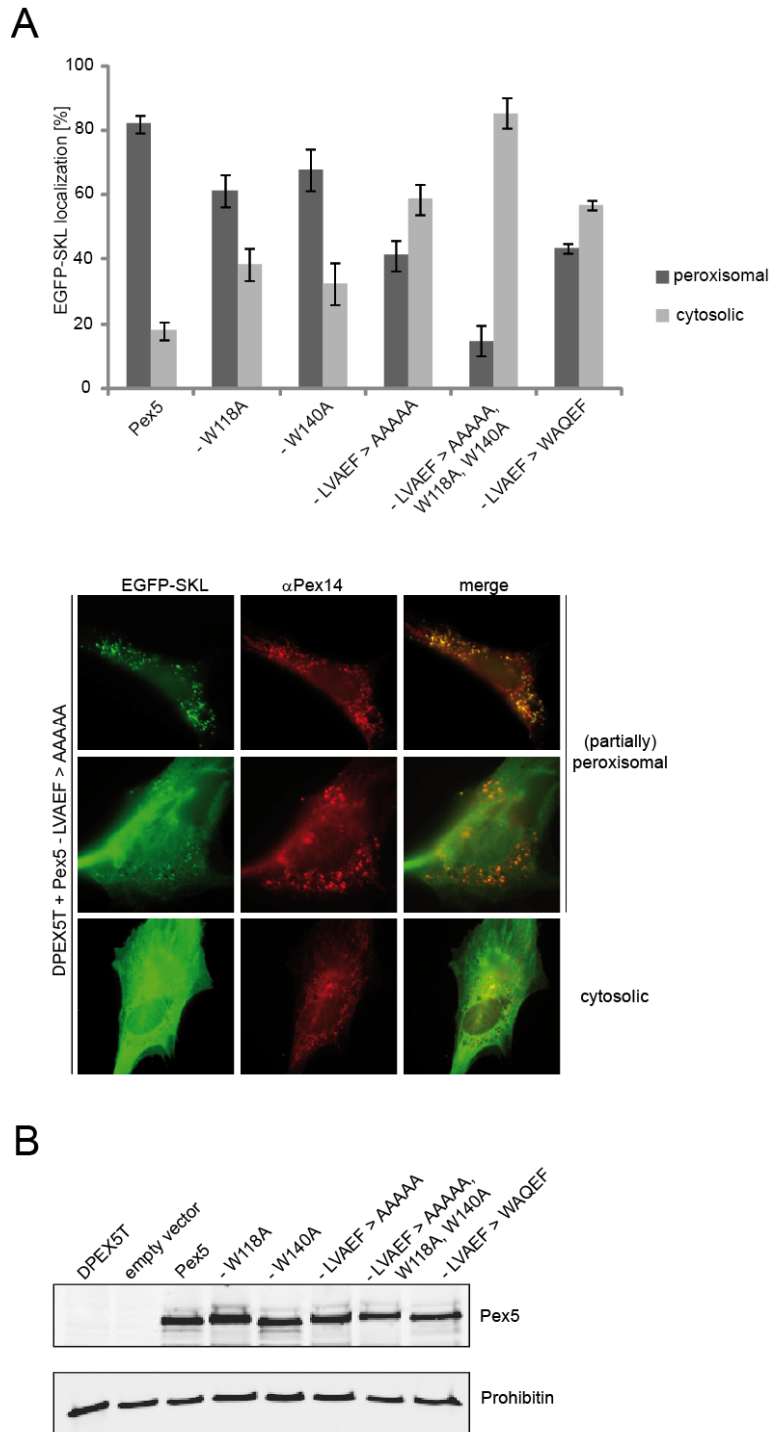


Figure 8.7 Functional complementation of PEX5-deficient fibroblasts by Pex5 carrying mutations in distinct Pex14 binding sites. Full-length Pex5 was mutagenized either in single Pex14-binding sites by alanine substitution of the LVxEF motif (LVAEF→AAAAA), the first (W118A) or the second (W140A) WxxxF/Y motif, or in combination (LVAEF→AAAAA, W118A,W140A). Furthermore, the LVxEF motif was substituted by the first WxxxF/Y motif, indicated by LVAEF→WAQEF. All constructs were expressed in Pex5-deficient fibroblasts (Δ PEX5T) from bicistronic expression vectors coding for the full-length Pex5 variants as indicated and EGFP fused to a PTS1 (EGFP-SKL). (A) Localization of the peroxisomal matrix marker protein EGFP-SKL was monitored by immunofluorescence microscopy

using antibodies against EGFP. Peroxisomes were labeled in red by immunofluorescence microscopy of the peroxisomal membrane marker Pex14. Representative cells expressing the Pex5-LVAEF→AAAAA mutant are shown in the lower panel. A congruent punctate red and green fluorescent pattern indicates partial or complete restoration of peroxisomal protein transport, while diffuse cytosolic staining of EGFP-SKL shows that the protein is at least partially mislocalized to the cytosol, indicating a low complementation activity of transfected Pex5 variant. The graph shows a quantitative analysis of mutant phenotype complementation by the individual Pex5 variants. Values were obtained from three or six independent transfection experiments. For each experiment, 100 transfected cells were categorized into those enabling EGFP-PTS1 import (peroxisomal) and those without complementation of the import defect (cytosolic). (B) Immunoblot analysis shows that all Pex5 variants are present at the same steady-state level as the plasmid-encoded full-length wild-type Pex5. Negative controls show lysates of cells either non-transfected or transfected with empty vector (pIRES2). The analysis of prohibitin served as loading control.

8.5 Discussion

The highly conserved N-terminal domain of Pex14 possesses a much broader binding specificity for peptide ligands as anticipated. Originally, the sequence signature of the core binding region has been defined as W-x-x-[QED]-[FY] based on the analysis of human Pex5-Pex14 interaction (Saidowsky et al, 2001). However, the existence of 2 to 9 WxxxF/Y motifs in each of the known Pex5 sequences can neither predict that the regions containing the consensus sequences indeed interact with Pex14 nor does it exclude that additional Pex14 binding sites exist. By using a peptide scan approach, we here identified all linear Pex14-binding sites of human Pex5 experimentally. Besides the seven WxxxF/Y-motifs, we detected a novel site that comprises Pex5 residues 62 to 66 with the sequence LVAEF. NMR analysis showed that similar to WxxxF/Y motifs, the novel penta-peptide motif adopts an amphipathic α -helix where the phenylalanine residue is recognized by one of the two hydrophobic pockets in Pex14-NTD. However, compared to canonical WxxxF/Y ligands, the Leu and Val residues do not completely occupy the corresponding pocket but form alternative hydrophobic interactions with residues in the proximity. Furthermore, our structural and mutational analyses suggest that electrostatic interactions may play a more important role in binding of Pex14 to LVxEF compared to the canonical WxxxF/Y motif. The length and helical conformation show that the novel LVxEF motif is a variant of the WxxxF/Y motif that binds to the same surface of

Pex14. The LVxEF ligand has a comparable affinity to human Pex14 (K_D around 100 nM) as the seven WxxxF/Y motifs of Pex5 (K_D values ranging from 3 to 150 nM) (Saidowsky et al, 2001; Neufeld et al, 2009).

Interestingly, comparison of the newly identified LVxEF motif with the first WxxxF/Y motif at amino acids position 118 to 122 of human Pex5 revealed distinct binding kinetics. Most remarkably, the LVxEF motif exhibits a 33-times higher dissociation rate when compared with a di-aromatic Pex14 ligand (Figure 8.5B and Table 8.2). One possible explanation for these differences could be that the hydrophobic pocket in the binding surface of Pex14 is better suited to accommodate a single tryptophan residue as in the WxxxF/Y motif than the two amino acids L and V of the LVAEF sequence.

Alanine substitution of the hydrophobic residues or the whole penta-peptide sequence of the LVxEF motif disrupts interaction with Pex14-NTD. However, site-directed mutagenesis of the new motif in full-length constructs of Pex5 does not significantly disturb Pex14-NTD interaction in terms of affinity and kinetic properties (Table 8.2). These *in vitro* results could be expected since each of the remaining WxxxF/Y motifs per se can act as high-affinity ligand for Pex14. Thus, a triple mutant of full-length Pex5 lacking the novel motif and two adjacent WxxxF/Y motifs retains full affinity with Pex14-NTD (Table 8.2). In contrast, the same triple mutant expressed in Pex5 deficient cells almost completely failed to rescue the import defect (Figure 8.7A). In fact, comparison of the complementing activities of single site mutants revealed that mutagenesis of the LVxEF motif has the most severe effect on the protein import into peroxisomes.

The *in vivo* results indicate a critical role for the new Pex14-binding site, which cannot be efficiently substituted by other Pex5 WxxxF/Y motifs even with higher affinity for Pex14. A possible mechanistic role for the new motif could be derived from the fast kinetics of LVxEF interaction with Pex14. In comparison with WxxxF/Y motifs, the LVxEF ligand associates three times faster and the dissociation rate is even 33 times faster. Accordingly, it is tempting to speculate that the novel Pex14 interacting site also *in vivo* establishes the first contact between the PTS1 receptor and its docking protein Pex14. On the basis of the relative high dissociation rate of the LVxEF-Pex14 complex, it seems possible that after initial contact, Pex14 is then

transferred to other WxxxF/Y-motifs with higher affinity for further processing of the PTS1-receptor at the peroxisomal membrane. The hypothesis that the LVxEF sequence serves as initial docking and loading site is supported by *in vivo* studies. Intriguingly, replacement of the novel motif in human Pex5 with a high-affinity WxxxF/Y motif impairs the function of Pex5 in the same way as its substitution with a sequence which does not interact with Pex14 at all. This further suggests that the structural and kinetic features of the LVxEF motif but not simply the presence of a Pex14 binding site at the N-terminus of Pex5 are important for the function of the receptor in matrix protein import.

In conclusion, our data clearly demonstrate that human Pex5 contains eight Pex14 binding sites and each of these motifs *per se* is a high-affinity ligand for the peroxisomal docking protein Pex14. An intriguing question is whether these eight motifs can interact with Pex14 in a sequential manner or whether simultaneous binding of these motifs would form higher order complexes of Pex14 and Pex5. The presence of multiple, partially redundant binding sites could be advantageous for docking of the PTS1-receptor by increasing the avidity of Pex5-Pex14 interaction. However, this work and other recent data suggest distinct roles of the different Pex14-binding sites in Pex5 during protein import.

Although still speculative, we propose a mechanistic model enabling sequential binding of the receptor during the protein import cascade (Figure 8.8). In this model, the new motif serves as a preferred docking site for cargo-loaded PTS1-receptor. Dissociation and sliding along the Pex5 polypeptide chain could initiate further binding of Pex14 to WxxxF/Y motifs 1 to 5 resulting in the formation of a multimeric Pex5/Pex14 complex which presumably represents the protein-conducting channel enabling matrix protein translocation across the peroxisomal membrane. Interestingly, a stable subcomplex consisting of Pex5 and Pex14 with a 1:5 stoichiometry could be isolated from rat liver peroxisomes (Gouveia et al, 2003). Finally, Pex14 binding to the downstream-located WxxxF/Y motifs six and seven triggers cargo-release (Freitas et al, 2011). Pex13 might also play a critical role in pore assembly or disassembly since it was shown that WxxxF/Y motifs 2, 3 and 4 of Pex5 are able to interact with this peroxin (Otera et al, 2002). To test the model, future studies need to be initiated to characterize the membrane topology of Pex5 mutants defective in one or more Pex14-binding sites.

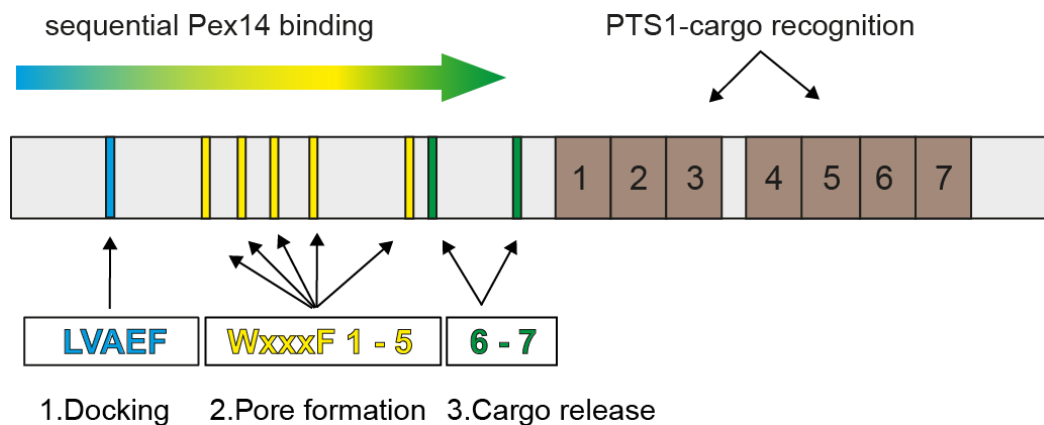


Figure 8.8 Sliding model of the Pex5/Pex14 interaction. While PTS1-containing peroxisomal matrix proteins bind to the C-terminal seven TPR-domains of Pex5 (brown boxes, 1-7), Pex14- interacts via its N-terminal domain with eight penta-peptide motifs located in the N-terminal half of the long isoform of human Pex5 (cyan, yellow and green bars). 1) The N-terminal LVxEF motif (LVAEF, cyan) is supposed to make the first contact to Pex14 and serves as the initial membrane docking site of receptor-cargo complex. 2) Upon dissociation, Pex14 can associate with one of the adjacent WxxxY motifs W1 to W5 and the novel LVxEF motif can recruit another Pex14. Sequential uploading of W1 to W5 correlates with structural changes of the Pex5-cargo/Pex14 complex and may result in the formation of a transient protein conducting channel in the peroxisomal membrane. 3) Finally, interaction of the WxxxY motifs W6 and W7 with Pex14-NTD enables cargo-release into the peroxisomal matrix.

9. Pex14-Tubulin interactions

The group of Ralf Erdmann (Bochum, Germany) have recently shown that human Pex14, in addition to its role in peroxisomal import, is required for microtubule-based motility of peroxisomes (Bharti et al, 2011). They have mapped the tubulin binding site to the conserved N-terminal domain of Pex14 (Pex14-NTD). Furthermore, by peptide scan analysis they have identified two peptides in the C-terminus of human tubulin β -chain (TUBB) which interacts with Pex14-NTD (unpublished data). The sequences of the identified peptides are MNDLVSEYQQYQDAT and RKAFLHWYTGEGMDE, hereafter referred to as LVSEY and FLHWY, respectively. The aim of this project was to characterize the interaction of human Pex14 with the peptides derived from tubulin using NMR spectroscopy.

NMR titration experiments were performed using ^{15}N -labeled Pex14-NTD and two unlabeled 15-mer tubulin peptides. In both cases, clear chemical shift perturbations were observed upon addition of the peptide to Pex14-NTD, indicating that each peptide can interact with Pex14-NTD (Figure 9.1). Binding of the LVSEY peptide is mostly in the fast-exchange regime on the NMR chemical shift timescale as judged from the gradual shift in NMR peak positions with increasing peptide concentrations. Few residues, such as Leu58, exhibit intermediate-exchange binding as the peaks disappear and reappear at higher peptide concentrations. In the case of the FLHWY peptide, however, binding is exclusively in the fast-exchange regime and less residues experience chemical shift perturbations, indicative of a weaker interaction (Figure 9.1). Based on the chemical shift perturbations, K_D values of $5.03 \pm 1.55 \mu\text{M}$ and $282.33 \pm 31.33 \mu\text{M}$ were estimated for the LVSEY and FLHWY peptides, respectively (Figure 9.1). These values are at least 50-fold weaker than the K_D of $\sim 100 \text{ nM}$ reported for the interaction of Pex14-NTD with the first WxxxF/Y motif in human Pex5 (Neufeld et al, 2009), suggesting that Pex5 can compete with tubulin for binding to Pex14-NTD.

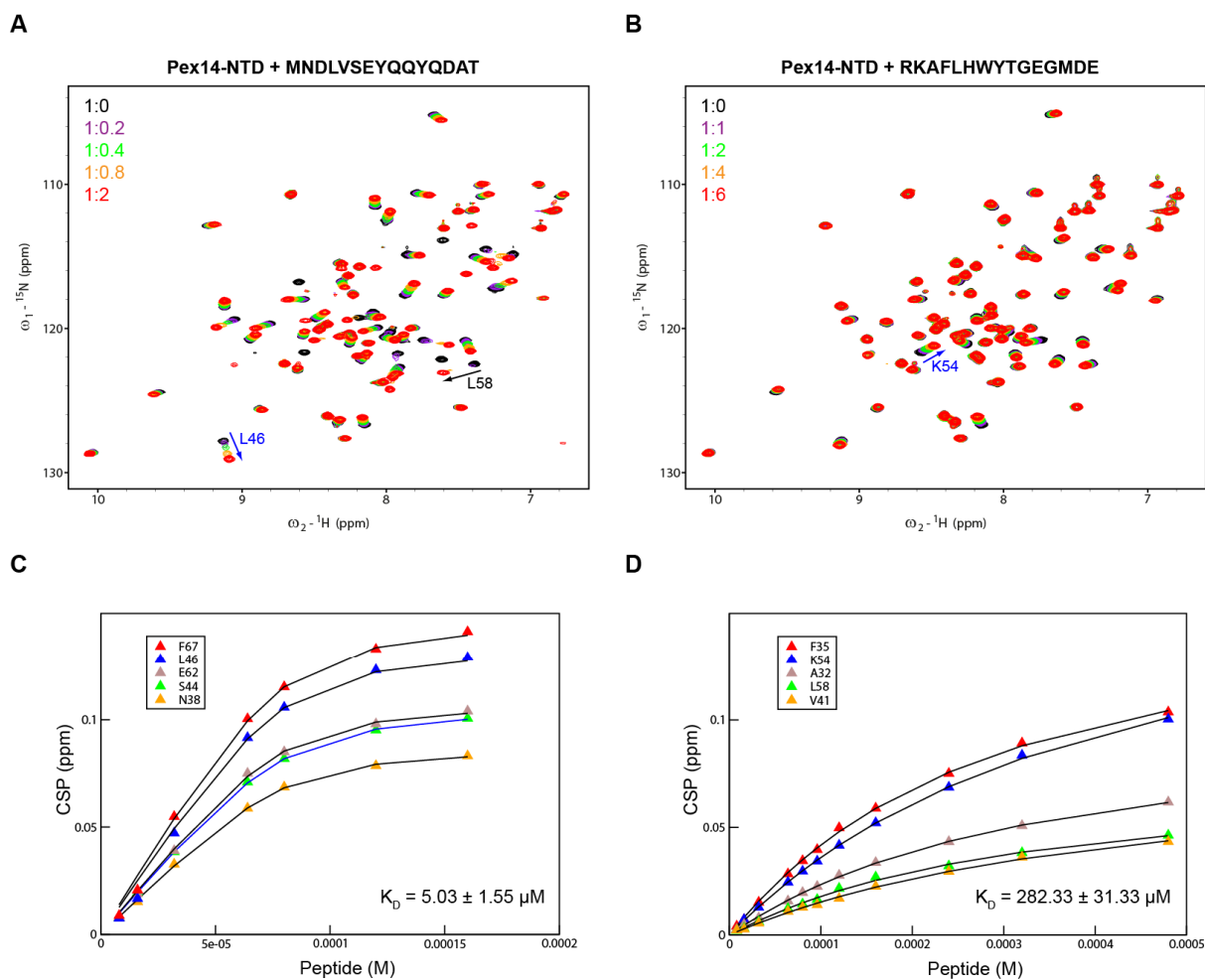


Figure 9.1 NMR titration of Pex14-NTD with (A) LVSEY and (B) FLHWY peptides derived from human tubulin β -chain. Overlay of ^1H , ^{15}N HSQC spectra of free ^{15}N -labelled Pex14-NTD (black) and in the presence of increasing concentrations of tubulin peptide (magenta, green, orange and red). Selected residues undergoing chemical shift perturbation are indicated. Five residues undergoing fast-exchange binding to (C) LVSEY and (D) FLHWY peptides were selected and the K_D was obtained by fitting the CSP to a one site-binding equation. K_D values are reported as the average K_D for five residues together with the largest fitting error.

To test whether the first WxxxF/Y motif in Pex5 can compete with the Pex14-tubulin interaction, an NMR double-titration experiment was performed, in which unlabeled LVSEY peptide was added to saturation to ^{15}N -labeled Pex14-NTD followed by addition of Pex5 peptide corresponding to the first WxxxF/Y motif with the sequence ALSENWAQEFLAAGDA (referred to as the WAQEF peptide). As shown in Figure 9.2, addition of only 0.5 molar equivalent of WAQEF peptide to the Pex14-LVSEY complex resulted in drop in intensity of peaks which had shifted upon LVSEY binding

and appearance of new signals corresponding to the Pex14-WAQEF complex. This is indicative of slow exchange binding of the WAQEF peptide, consistent with the nanomolar K_D value and also shows that both LVSEY and WAQEF bind to the same binding site on Pex14. Further addition of WAQEF peptide completely shifts the spectrum of Pex14-NTD to the WAQEF-bound form, indicating that the WAQEF peptide can replace the LVSEY peptide (Figure 9.2).

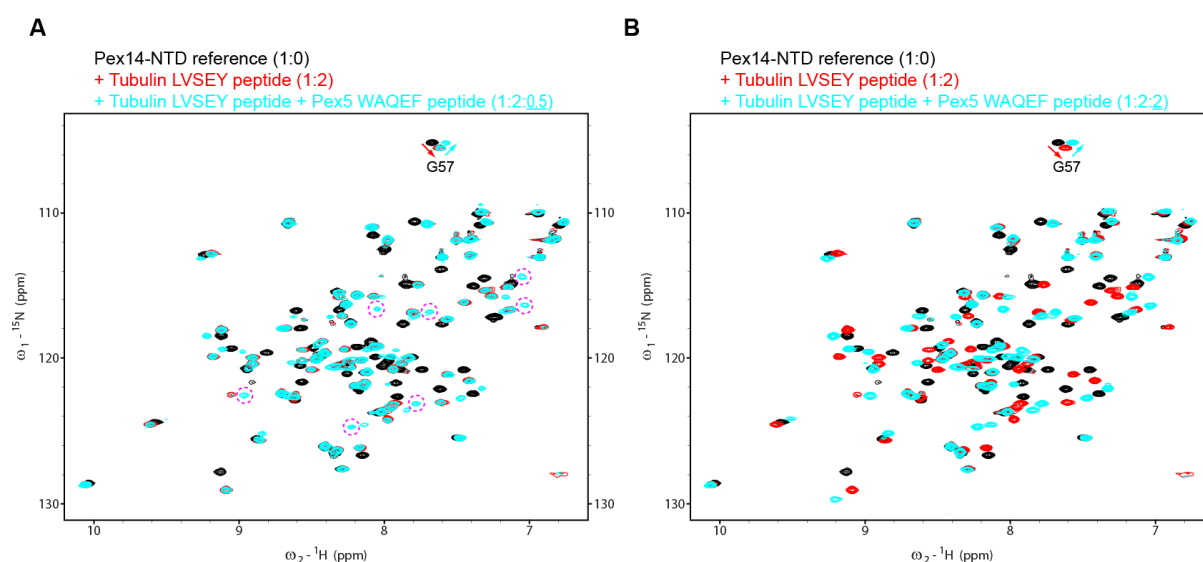


Figure 9.2 ^1H , ^{15}N HSQC spectra of free ^{15}N -labelled Pex14-NTD (black), in the presence of 2-fold molar excess of LVSEY peptide (red) and in the presence of both LVSEY and WAQEF peptides (cyan). (A) Addition of 0.5 molar equivalent of WAQEF peptide to the Pex14-LVSEY complex results in appearance of new signals (selected peaks are encircled) corresponding to the Pex14-WAQEF complex. (B) At equal concentrations, the WAQEF peptide completely replaces the LVSEY peptide, thus shifting the spectrum of Pex14-NTD to the WAQEF-bound form. The peak corresponding to Gly57 is annotated.

In summary, these data indicate that human Pex14-NTD interacts with a region comprising two binding motifs in the C-terminus of tubulin β -chain. Interestingly, one of the motifs (LVSEY), is highly similar to the newly identified LVxEF motif in Pex5 (see chapter 8) and is expected to have the same mode of binding to Pex14.

The presence of two binding motifs in tubulin suggests that two molecules of Pex14 are able to bind to tubulin. Moreover, our data clearly show that a WxxxF/Y peptide from Pex5 can compete with tubulin for binding to Pex14-NTD. It would be interesting to understand the functional significance of this competition in peroxisome motility and function.

10. References

- Abdul-Manan N, O'Malley SM, Williams KR (1996) Origins of binding specificity of the A1 heterogeneous nuclear ribonucleoprotein. *Biochemistry* **35**: 3545-3554
- Abdul-Manan N, Williams KR (1996) hnRNP A1 binds promiscuously to oligoribonucleotides: utilization of random and homo-oligonucleotides to discriminate sequence from base-specific binding. *Nucleic Acids Res* **24**: 4063-4070
- Albertini M, Rehling P, Erdmann R, Girzalsky W, Kiel JAKW, Veenhuis M, Kunau W-H (1997) Pex14p, a peroxisomal membrane protein binding both receptors of the two PTS-dependent import pathways. *Cell* **89**: 83-92
- Anderson AC, O'Neil RH, Filman DJ, Frederick CA (1999) Crystal structure of a brominated RNA helix with four mismatched base pairs: An investigation into RNA conformational variability. *Biochemistry* **38**: 12577-12585
- Banerjee H, Rahn A, Davis W, Singh R (2003) Sex lethal and U2 small nuclear ribonucleoprotein auxiliary factor (U2AF65) recognize polypyrimidine tracts using multiple modes of binding. *RNA* **9**: 88-99
- Banerjee H, Rahn A, Gawande B, Guth S, Valcarcel J, Singh R (2004) The conserved RNA recognition motif 3 of U2 snRNA auxiliary factor (U2AF 65) is essential in vivo but dispensable for activity in vitro. *RNA* **10**: 240-253
- Barraud P, Allain FH (2013) Solution structure of the two RNA recognition motifs of hnRNP A1 using segmental isotope labeling: how the relative orientation between RRM3 influences the nucleic acid binding topology. *J Biomol NMR* **55**: 119-138
- Berglund JA, Abovich N, Rosbash M (1998) A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev* **12**: 858-867
- Berglund JA, Chua K, Abovich N, Reed R, Rosbash M (1997) The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* **89**: 781-787
- Bernstein E, Caudy AA, Hammond SM, Hannon GJ (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**: 363-366

Bharti P, Schliebs W, Schievelbusch T, Neuhaus A, David C, Kock K, Herrmann C, Meyer HE, Wiese S, Warscheid B, Theiss C, Erdmann R (2011) PEX14 is required for microtubule-based peroxisome motility in human cells. *J Cell Sci* **124**: 1759-1768

Blanchette M, Chabot B (1999) Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J* **18**: 1939-1952

Bonnal S, Pileur F, Orsini C, Parker F, Pujol F, Prats AC, Vagner S (2005) Heterogeneous nuclear ribonucleoprotein A1 is a novel internal ribosome entry site trans-acting factor that modulates alternative initiation of translation of the fibroblast growth factor 2 mRNA. *J Biol Chem* **280**: 4144-4153

Borchert GM, Lanier W, Davidson BL (2006) RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol* **13**: 1097-1101

Bottger G, Barnett P, Klein AT, Kragt A, Tabak HF, Distel B (2000) *Saccharomyces cerevisiae* PTS1 receptor Pex5p interacts with the SH3 domain of the peroxisomal membrane protein Pex13p in an unconventional, non-PXXP-related manner. *Mol Biol Cell* **11**: 3963-3976.

Burd CG, Dreyfuss G (1994) RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J* **13**: 1197-1204

Buvoli M, Cobianchi F, Bestagno MG, Mangiarotti A, Bassi MT, Biamonti G, Riva S (1990a) Alternative splicing in the human gene for the core protein A1 generates another hnRNP protein. *EMBO J* **9**: 1229-1235

Buvoli M, Cobianchi F, Biamonti G, Riva S (1990b) Recombinant hnRNP protein A1 and its N-terminal domain show preferential affinity for oligodeoxynucleotides homologous to intron/exon acceptor sites. *Nucleic Acids Res* **18**: 6595-6600

Buvoli M, Cobianchi F, Riva S (1992) Interaction of hnRNP A1 with snRNPs and pre-mRNAs: evidence for a possible role of A1 RNA annealing activity in the first steps of spliceosome assembly. *Nucleic Acids Res* **20**: 5017-5025

Caceres JF, Stamm S, Helfman DM, Krainer AR (1994) Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science* **265**: 1706-1709

Cartegni L, Hastings ML, Calarco JA, de Stanchina E, Krainer AR (2006) Determinants of exon 7 splicing in the spinal muscular atrophy genes, SMN1 and SMN2. *Am J Hum Genet* **78**: 63-77

Castilla-Llorente V, Nicastro G, Ramos A (2013) Terminal loop-mediated regulation of miRNA biogenesis: selectivity and mechanisms. *Biochem Soc Trans* **41**: 861-865

Cavanagh J (2007) *Protein NMR spectroscopy : principles and practice*, 2nd edn. Amsterdam ; Boston: Academic Press.

Chakraborty S, Mehtab S, Patwardhan A, Krishnan Y (2012) Pri-miR-17-92a transcript folds into a tertiary structure and autoregulates its processing. *RNA* **18**: 1014-1028

Chaudhury A, Chander P, Howe PH (2010) Heterogeneous nuclear ribonucleoproteins (hnRNPs) in cellular processes: Focus on hnRNP E1's multifunctional regulatory roles. *RNA* **16**: 1449-1462

Chaulk SG, Thede GL, Kent OA, Xu Z, Gesner EM, Veldhoen RA, Khanna SK, Goping IS, MacMillan AM, Mendell JT, Young HS, Fahlman RP, Glover JN (2011) Role of pri-miRNA tertiary structure in miR-17~92 miRNA biogenesis. *RNA Biol* **8**: 1105-1114

Chaulk SG, Xu Z, Glover MJ, Fahlman RP (2014) MicroRNA miR-92a-1 biogenesis and mRNA targeting is modulated by a tertiary contact within the miR-17~92 microRNA cluster. *Nucleic Acids Res*

Choe J, Moyersoer J, Roach C, Carter TL, Fan E, Michels PA, Hol WG (2003) Analysis of the sequence motifs responsible for the interactions of peroxins 14 and 5, which are involved in glycosome biogenesis in *Trypanosoma brucei*. *Biochemistry* **42**: 10915-10922

Choudhury NR, Michlewski G (2012) Terminal loop-mediated control of microRNA biogenesis. *Biochem Soc Trans* **40**: 789-793

Clery A, Blatter M, Allain FH (2008) RNA recognition motifs: boring? Not quite. *Curr Opin Struct Biol* **18**: 290-298

Cruse WB, Saludjian P, Biala E, Strazewski P, Prange T, Kennard O (1994) Structure of a mispaired RNA double helix at 1.6-Å resolution and implications for the prediction of RNA secondary structure. *Proc Natl Acad Sci U S A* **91**: 4160-4164

Czech B, Zhou R, Erlich Y, Brennecke J, Binari R, Villalta C, Gordon A, Perrimon N, Hannon GJ (2009) Hierarchical rules for Argonaute loading in *Drosophila*. *Mol Cell* **36**: 445-456

Dallmann A, Simon B, Duszczyc MM, Kooshapur H, Pardi A, Bermel W, Sattler M (2013) Efficient Detection of Hydrogen Bonds in Dynamic Regions of RNA by Sensitivity-Optimized NMR Pulse Sequences. *Angew Chem Int Ed Engl*

Damiano F, Rochira A, Tocci R, Alemanno S, Gnoni A, Siculella L (2013) hnRNP A1 mediates the activation of the IRES-dependent SREBP-1a mRNA translation in response to endoplasmic reticulum stress. *Biochem J* **449**: 543-553

Daragan VA, Mayo KH (1997) Motional Model Analyses of Protein and Peptide Dynamics Using ¹³C and ¹⁵N NMR Relaxation. *Progress in Nuclear Magnetic Resonance Spectroscopy* **31**: 63-105

Davis-Dusenbery BN, Hata A (2010) MicroRNA in Cancer: The Involvement of Aberrant MicroRNA Biogenesis Regulatory Pathways. *Genes & cancer* **1**: 1100-1114

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

Denli AM, Tops BB, Plasterk RH, Ketting RF, Hannon GJ (2004) Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**: 231-235

Ding J, Hayashi MK, Zhang Y, Manche L, Krainer AR, Xu RM (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev* **13**: 1102-1115

Dingley AJ, Grzesiek S (1998) Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide 2JNN couplings. *J Am Chem Soc* **120**: 8293-8297

Dotz G, Braverman N, Wong C, Moser A, Moser HW, Watkins P, Valle D, Gould SJ (1995) Mutations in the PTS1 receptor gene, PXR1, define complementation group 2 of the peroxisome biogenesis disorders. *Nat Genet* **9**: 115-125

Doreleijers JF, Sousa da Silva AW, Krieger E, Nabuurs SB, Spronk CA, Stevens TJ, Vranken WF, Vriend G, Vuister GW (2012) CING: an integrated residue-based structure validation program suite. *J Biomol NMR* **54**: 267-283

Douangamath A, Filipp FV, Klein AT, Barnett P, Zou P, Voorn-Brouwer T, Vega MC, Mayans OM, Sattler M, Distel B, Wilmanns M (2002) Topography for independent binding of alpha-helical and PPII-helical ligands to a peroxisomal SH3 domain. *Mol Cell* **10**: 1007-1017

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797

Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**: 486-501

Erdmann R, Schliebs W (2005) Peroxisomal matrix protein import: the transient pore model. *Nat Rev Mol Cell Biol* **6**: 738-742

Esquela-Kerscher A, Slack FJ (2006) Oncomirs - microRNAs with a role in cancer. *Nature reviews Cancer* **6**: 259-269

Fabian MR, Sonenberg N (2012) The mechanics of miRNA-mediated gene silencing: a look under the hood of miRISC. *Nat Struct Mol Biol* **19**: 586-593

Freitas MO, Francisco T, Rodrigues TA, Alencastre IS, Pinto MP, Grou CP, Carvalho AF, Fransen M, Sa-Miranda C, Azevedo JE (2011) PEX5 protein binds monomeric catalase blocking its tetramerization and releases it upon binding the N-terminal domain of PEX14. *J Biol Chem* **286**: 40509-40519

Fujiki Y, Okumoto K, Otera H, Tamura S (2000) Peroxisome biogenesis and molecular defects in peroxisome assembly disorders. *Cell biochemistry and biophysics* **32 Spring**: 155-164

Fujiya M, Konishi H, Mohamed Kamel MK, Ueno N, Inaba Y, Moriichi K, Tanabe H, Ikuta K, Ohtake T, Kohgo Y (2013) microRNA-18a induces apoptosis in colon cancer cells via the autophagolysosomal degradation of oncogenic heterogeneous nuclear ribonucleoprotein A1. *Oncogene*

Fukuda H, Katahira M, Tsuchiya N, Enokizono Y, Sugimura T, Nagao M, Nakagama H (2002) Unfolding of quadruplex structure in the G-rich strand of the minisatellite repeat by the binding protein UP1. *Proc Natl Acad Sci U S A* **99**: 12685-12690

Garrett DS, Powers R, March CJ, Frieden EA, Clore GM, Gronenborn AM (1992) Determination of the secondary structure and folding topology of human interleukin-4 using three-dimensional heteronuclear magnetic resonance spectroscopy. *Biochemistry* **31**: 4347-4353

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, Bairoch A (2005) Protein Identification and Analysis Tools on the ExPASy Server. In *The Proteomics Protocols Handbook*, Walker JM (ed), pp 571-607. Humana Press

Goddard TD, Kneller DG. SPARKY 3. University of California, San Francisco

Gouveia AM, Guimaraes CP, Oliveira ME, Sa-Miranda C, Azevedo JE (2003) Insertion of Pex5p into the peroxisomal membrane is cargo protein-dependent. *J Biol Chem* **278**: 4389-4392

Gozani O, Potashkin J, Reed R (1998) A potential role for U2AF-SAP 155 interactions in recruiting U2 snRNP to the branch site. *Mol Cell Biol* **18**: 4752-4760

Gregory RI, Yan KP, Amuthan G, Chendrimada T, Doratotaj B, Cooch N, Shiekhattar R (2004) The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**: 235-240

Guil S, Caceres JF (2007) The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nat Struct Mol Biol* **14**: 591-596

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

Guth S, Tange TO, Kellenberger E, Valcarcel J (2001) Dual function for U2AF(35) in AG-dependent pre-mRNA splicing. *Mol Cell Biol* **21**: 7673-7681

Hagan JP, Piskounova E, Gregory RI (2009) Lin28 recruits the TUTase Zcchc11 to inhibit let-7 maturation in mouse embryonic stem cells. *Nat Struct Mol Biol* **16**: 1021-1025

Hamilton BJ, Burns CM, Nichols RC, Rigby WF (1997) Modulation of AUUUA response element binding by heterogeneous nuclear ribonucleoprotein A1 in human T lymphocytes. The roles of cytoplasmic location, transcription, and phosphorylation. *J Biol Chem* **272**: 28732-28741

Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, Sohn SY, Cho Y, Zhang BT, Kim VN (2006) Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. *Cell* **125**: 887-901

Han SP, Tang YH, Smith R (2010) Functional diversity of the hnRNPs: past, present and perspectives. *Biochem J* **430**: 379-392

Han TW, Kato M, Xie S, Wu LC, Mirzaei H, Pei J, Chen M, Xie Y, Allen J, Xiao G, McKnight SL (2012) Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell* **149**: 768-779

He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, Powers S, Cordon-Cardo C, Lowe SW, Hannon GJ, Hammond SM (2005) A microRNA polycistron as a potential human oncogene. *Nature* **435**: 828-833

Henics T, Sanfridson A, Hamilton BJ, Nagy E, Rigby WF (1994) Enhanced stability of interleukin-2 mRNA in MLA 144 cells. Possible role of cytoplasmic AU-rich sequence-binding proteins. *J Biol Chem* **269**: 5377-5383

Heo I, Joo C, Cho J, Ha M, Han J, Kim VN (2008) Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol Cell* **32**: 276-284

Heo I, Joo C, Kim YK, Ha M, Yoon MJ, Cho J, Yeom KH, Han J, Kim VN (2009) TUT4 in concert with Lin28 suppresses microRNA biogenesis through pre-microRNA uridylation. *Cell* **138**: 696-708

Herrick G, Alberts B (1976) Purification and physical characterization of nucleic acid helix-unwinding proteins from calf thymus. *J Biol Chem* **251**: 2124-2132

Hilpert K, Winkler DF, Hancock RE (2007) Peptide arrays on cellulose support: SPOT synthesis, a time and cost efficient method for synthesis of large numbers of peptides in a parallel and addressable fashion. *Nature protocols* **2**: 1333-1349

Holbrook SR, Cheong C, Tinoco I, Jr., Kim SH (1991) Crystal structure of an RNA double helix incorporating a track of non-Watson-Crick base pairs. *Nature* **353**: 579-581

Huntzinger E, Izaurralde E (2011) Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nature reviews Genetics* **12**: 99-110

Huybrechts SJ, Van Veldhoven PP, Hoffman I, Zeevaert R, de Vos R, Demaerel P, Brams M, Jaeken J, Franssen M, Cassiman D (2008) Identification of a novel PEX14 mutation in Zellweger syndrome. *J Med Genet* **45**: 376-383

Izaurralde E, Jarmolowski A, Beisel C, Mattaj IW, Dreyfuss G, Fischer U (1997) A role for the M9 transport signal of hnRNP A1 in mRNA nuclear export. *The Journal of cell biology* **137**: 27-35

Jenkins JL, Agrawal AA, Gupta A, Green MR, Kielkopf CL (2013) U2AF65 adapts to diverse pre-mRNA splice sites through conformational selection of specific and promiscuous RNA recognition motifs. *Nucleic Acids Res* **41**: 3859-3873

Jenkins JL, Laird KM, Kielkopf CL (2012) A Broad range of conformations contribute to the solution ensemble of the essential splicing factor U2AF(65). *Biochemistry* **51**: 5223-5225

Jinek M, Doudna JA (2009) A three-dimensional view of the molecular machinery of RNA interference. *Nature* **457**: 405-412

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

Kabsch W (2010) Xds. *Acta Crystallogr D Biol Crystallogr* **66**: 125-132

Kao C, Zheng M, Rudisser S (1999) A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA* **5**: 1268-1272

Kashima T, Manley JL (2003) A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. *Nat Genet* **34**: 460-463

Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, Goldsmith EJ, Longgood J, Pei J, Grishin NV, Frantz DE, Schneider JW, Chen S, Li L, Sawaya MR, Eisenberg D, Tycko R, McKnight SL (2012) Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**: 753-767

Kent OA, Reayi A, Foong L, Chilibeck KA, MacMillan AM (2003) Structuring of the 3' splice site by U2AF65. *J Biol Chem* **278**: 50572-50577

Keren H, Lev-Maor G, Ast G (2010) Alternative splicing and evolution: diversification, exon definition and function. *Nature reviews Genetics* **11**: 345-355

Kerssen D, Hambruch E, Klaas W, Platta HW, de Kruijff B, Erdmann R, Kunau WH, Schliebs W (2006) Membrane association of the cycling peroxisome import receptor Pex5p. *J Biol Chem* **281**: 27003-27015

Khvorova A, Reynolds A, Jayasena SD (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**: 209-216

Kim HJ, Kim NC, Wang YD, Scarborough EA, Moore J, Diaz Z, MacLea KS, Freibaum B, Li S, Molliex A, Kanagaraj AP, Carter R, Boylan KB, Wojtas AM, Rademakers R, Pinkus JL, Greenberg SA, Trojanowski JQ, Traynor BJ, Smith BN, Topp S, Gkazi AS, Miller J, Shaw CE, Kottlors M, Kirschner J, Pestronk A, Li YR, Ford AF, Gitler AD, Benatar M, King OD, Kimonis VE, Ross ED, Wehl CC, Shorter J, Taylor JP (2013) Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS. *Nature* **495**: 467-473

Kleckner IR, Foster MP (2011) An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et biophysica acta* **1814**: 942-968

Kozomara A, Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42**: D68-73

Krecic AM, Swanson MS (1999) hnRNP complexes: composition, structure, and function. *Current opinion in cell biology* **11**: 363-371

Krol J, Loedige I, Filipowicz W (2010) The widespread regulation of microRNA biogenesis, function and decay. *Nature reviews Genetics* **11**: 597-610

LaBranche H, Dupuis S, Ben-David Y, Bani MR, Wellinger RJ, Chabot B (1998) Telomere elongation by hnRNP A1 and a derivative that interacts with telomeric repeats and telomerase. *Nat Genet* **19**: 199-202

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680-685

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948

Lau NC, Lim LP, Weinstein EG, Bartel DP (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**: 858-862

Lee BJ, Cansizoglu AE, Suel KE, Louis TH, Zhang Z, Chook YM (2006) Rules for nuclear localization sequence recognition by karyopherin beta 2. *Cell* **126**: 543-558

Lee CG, Zamore PD, Green MR, Hurwitz J (1993a) RNA annealing activity is intrinsically associated with U2AF. *J Biol Chem* **268**: 13472-13478

Lee RC, Feinbaum RL, Ambros V (1993b) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**: 843-854

Lee Y, Jeon K, Lee JT, Kim S, Kim VN (2002) MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J* **21**: 4663-4670

Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, Kim VN (2004) MicroRNA genes are transcribed by RNA polymerase II. *EMBO J* **23**: 4051-4060

Legault P, Li J, Mogridge J, Kay LE, Greenblatt J (1998) NMR structure of the bacteriophage lambda N peptide/boxB RNA complex: recognition of a GNRA fold by an arginine-rich motif. *Cell* **93**: 289-299

Leontis NB, Westhof E (2001) Geometric nomenclature and classification of RNA base pairs. *RNA* **7**: 499-512

Linge JP, Williams MA, Spronk CA, Bonvin AM, Nilges M (2003) Refinement of protein structures in explicit solvent. *Proteins* **50**: 496-506

Liu H, Naismith JH (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC biotechnology* **8**: 91

Liu Z, Luyten I, Bottomley MJ, Messias AC, Houngninou-Molango S, Sprangers R, Zanier K, Kramer A, Sattler M (2001) Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* **294**: 1098-1102

Loughlin FE, Gebert LF, Towbin H, Brunschweiler A, Hall J, Allain FH (2012) Structural basis of pre-let-7 miRNA recognition by the zinc knuckles of pluripotency factor Lin28. *Nat Struct Mol Biol* **19**: 84-89

Ma C, Agrawal G, Subramani S (2011) Peroxisome assembly: matrix and membrane protein biogenesis. *The Journal of cell biology* **193**: 7-16

Ma E, MacRae IJ, Kirsch JF, Doudna JA (2008) Autoinhibition of human dicer by its internal helicase domain. *J Mol Biol* **380**: 237-243

Mackereth CD, Sattler M (2012) Dynamics in multi-domain protein recognition of RNA. *Curr Opin Struct Biol* **22**: 287-296

Madrid KP, Jardim A (2005) Peroxin 5-peroxin 14 association in the protozoan *Leishmania donovani* involves a novel protein-protein interaction motif. *Biochem J* **391**: 105-114

Maris C, Dominguez C, Allain FH (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J* **272**: 2118-2131

Massi F, Johnson E, Wang C, Rance M, Palmer AG, 3rd (2004) NMR R1 rho rotating-frame relaxation with weak radio frequency fields. *J Am Chem Soc* **126**: 2247-2256

Mayeda A, Krainer AR (1992) Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* **68**: 365-375

Mayeda A, Munroe SH, Caceres JF, Krainer AR (1994) Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *EMBO J* **13**: 5483-5495

Mayeda A, Munroe SH, Xu RM, Krainer AR (1998) Distinct functions of the closely related tandem RNA-recognition motifs of hnRNP A1. *RNA* **4**: 1111-1123

McAfee JG, Soltaninassab SR, Lindsay ME, LeSturgeon WM (1996) Proteins C1 and C2 of heterogeneous nuclear ribonucleoprotein complexes bind RNA in a highly cooperative fashion: support for their contiguous deposition on pre-mRNA during transcription. *Biochemistry* **35**: 1212-1222

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**: 658-674

Meinecke M, Cizmowski C, Schliebs W, Kruger V, Beck S, Wagner R, Erdmann R (2010) The peroxisomal importomer constitutes a large and highly dynamic pore. *Nat Cell Biol* **12**: 273-277

Mendell JT (2008) miRiad roles for the miR-17-92 cluster in development and disease. *Cell* **133**: 217-222

Merendino L, Guth S, Bilbao D, Martinez C, Valcarcel J (1999) Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature* **402**: 838-841

Michael WM, Choi M, Dreyfuss G (1995) A nuclear export signal in hnRNP A1: a signal-mediated, temperature-dependent nuclear protein export pathway. *Cell* **83**: 415-422

Michlewski G, Caceres JF (2010) Antagonistic role of hnRNP A1 and KSRP in the regulation of let-7a biogenesis. *Nat Struct Mol Biol* **17**: 1011-1018

Michlewski G, Guil S, Semple CA, Caceres JF (2008) Posttranscriptional regulation of miRNAs harboring conserved terminal loops. *Mol Cell* **32**: 383-393

Mili S, Shu HJ, Zhao Y, Pinol-Roma S (2001) Distinct RNP complexes of shuttling hnRNP proteins with pre-mRNA and mRNA: candidate intermediates in formation and export of mRNA. *Mol Cell Biol* **21**: 7307-7319

Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**: 240-255

Myers JC, Moore SA, Shamoo Y (2003) Structure-based incorporation of 6-methyl-8-(2-deoxy-beta-ribofuranosyl)isoxanthopterin into the human telomeric repeat DNA as a probe for UP1 binding and destabilization of G-tetrad structures. *J Biol Chem* **278**: 42300-42306

Myers JC, Shamoo Y (2004) Human UP1 as a model for understanding purine recognition in the family of proteins containing the RNA recognition motif (RRM). *J Mol Biol* **342**: 743-756

Nam Y, Chen C, Gregory RI, Chou JJ, Sliz P (2011) Molecular basis for interaction of let-7 microRNAs with Lin28. *Cell* **147**: 1080-1091

Nelson KK, Green MR (1989) Mammalian U2 snRNP has a sequence-specific RNA-binding activity. *Genes Dev* **3**: 1562-1571

Neufeld C, Filipp FV, Simon B, Neuhaus A, Schuller N, David C, Kooshapur H, Madl T, Erdmann R, Schliebs W, Wilmanns M, Sattler M (2009) Structural basis for competitive interactions of Pex14 with the import receptors Pex5 and Pex19. *EMBO J* **28**: 745-754

Nicastro G, Garcia-Mayoral MF, Hollingworth D, Kelly G, Martin SR, Briata P, Gherzi R, Ramos A (2012) Noncanonical G recognition mediates KSRP regulation of let-7 biogenesis. *Nat Struct Mol Biol* **19**: 1282-1286

Nilsen TW, Graveley BR (2010) Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463**: 457-463

Nilsson B, Moks T, Jansson B, Abrahmsen L, Elmlblad A, Holmgren E, Henrichson C, Jones TA, Uhlen M (1987) A synthetic IgG-binding domain based on staphylococcal protein A. *Protein engineering* **1**: 107-113

Nuttall JM, Motley A, Hettema EH (2011) Peroxisome biogenesis: recent advances. *Current opinion in cell biology* **23**: 421-426

Okada C, Yamashita E, Lee SJ, Shibata S, Katahira J, Nakagawa A, Yoneda Y, Tsukihara T (2009) A high-resolution structure of the pre-microRNA nuclear export machinery. *Science* **326**: 1275-1279

Okamura K, Liu N, Lai EC (2009) Distinct mechanisms for microRNA strand selection by *Drosophila* Argonautes. *Mol Cell* **36**: 431-444

Okamura K, Phillips MD, Tyler DM, Duan H, Chou YT, Lai EC (2008) The regulatory activity of microRNA* species has substantial influence on microRNA and 3' UTR evolution. *Nat Struct Mol Biol* **15**: 354-363

Okunola HL, Krainer AR (2009) Cooperative-binding and splicing-repressive properties of hnRNP A1. *Mol Cell Biol* **29**: 5620-5631

Otera H, Harano T, Honsho M, Ghaedi K, Mukai S, Tanaka A, Kawai A, Shimizu N, Fujiki Y (2000) The mammalian peroxin Pex5pL, the longer isoform of the mobile peroxisome targeting signal (PTS) type 1 transporter, translocates the Pex7p.PTS2 protein complex into peroxisomes via its initial docking site, Pex14p. *J Biol Chem* **275**: 21703-21714

Otera H, Setoguchi K, Hamasaki M, Kumashiro T, Shimizu N, Fujiki Y (2002) Peroxisomal targeting signal receptor Pex5p interacts with cargoes and import machinery components in a spatiotemporally differentiated manner: conserved Pex5p WXXXF/Y motifs are critical for matrix protein import. *Mol Cell Biol* **22**: 1639-1655

Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**: 1413-1415

Parisien M, Major F (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**: 51-55

Parker R, Siliciano PG, Guthrie C (1987) Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell* **49**: 229-239

Petoukhov MV, Franke D, Shkumatov AV, Tria G, Kikhney AG, Gajda M, Gorba C, Mertens HDT, Konarev PV, Svergun DI (2012) New developments in the ATSAS

program package for small-angle scattering data analysis. *Journal of Applied Crystallography* **45**: 342-350

Petoukhov MV, Svergun DI (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys J* **89**: 1237-1250

Pinol-Roma S, Dreyfuss G (1992) Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature* **355**: 730-732

Pires JR, Hong X, Brockmann C, Volkmer-Engert R, Schneider-Mergener J, Oschkinat H, Erdmann R (2003) The ScPex13p SH3 domain exposes two distinct binding sites for Pex5p and Pex14p. *J Mol Biol* **326**: 1427-1435

Piskounova E, Polytarchou C, Thornton JE, LaPierre RJ, Pothoulakis C, Hagan JP, Iliopoulos D, Gregory RI (2011) Lin28A and Lin28B inhibit let-7 microRNA biogenesis by distinct mechanisms. *Cell* **147**: 1066-1079

Platta HW, Hagen S, Erdmann R (2012) The exportomer: the peroxisomal receptor export machinery. *Cell Mol Life Sci*

Platta HW, Hagen S, Reidick C, Erdmann R (2013) The peroxisomal receptor dislocation pathway: To the exportomer and beyond. *Biochimie*

Rau F, Freyermuth F, Fugier C, Villemin JP, Fischer MC, Jost B, Dembele D, Gourdon G, Nicole A, Duboc D, Wahbi K, Day JW, Fujimura H, Takahashi MP, Auboeuf D, Dreumont N, Furling D, Charlet-Berguerand N (2011) Misregulation of miR-1 processing is associated with heart defects in myotonic dystrophy. *Nat Struct Mol Biol* **18**: 840-845

Reed R, Maniatis T (1986) A role for exon sequences and splice-site proximity in splice-site selection. *Cell* **46**: 681-690

Reinhart BJ, Slack FJ, Basson M, Pasquinelli AE, Bettinger JC, Rougvie AE, Horvitz HR, Ruvkun G (2000) The 21-nucleotide let-7 RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**: 901-906

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

Rucktaschel R, Girzalsky W, Erdmann R (2011) Protein import machineries of peroxisomes. *Biochimica et biophysica acta* **1808**: 892-900

Saidowsky J, Dodt G, Kirchberg K, Wegner A, Nastainczyk W, Kunau WH, Schliebs W (2001) The di-aromatic pentapeptide repeats of the human peroxisome import receptor PEX5 are separate high affinity binding sites for the peroxisomal membrane protein PEX14. *J Biol Chem* **276**: 34524-34529

Sambrook J, Russell D (2001) *Molecular cloning : a laboratory manual*, Third edn.: Cold Spring Harbor Laboratory Press.

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

Schafer A, Kerssen D, Veenhuis M, Kunau WH, Schliebs W (2004) Functional similarity between the peroxisomal PTS2 receptor binding protein Pex18p and the N-terminal half of the PTS1 receptor Pex5p. *Mol Cell Biol* **24**: 8895-8906

Schanda P, Brutscher B (2005) Very Fast Two-Dimensional NMR Spectroscopy for Real-Time Investigation of Dynamic Events in Proteins on the Time Scale of Seconds. *Journal of the American Chemical Society* **127**: 8014-8015

Schliebs W, Saidowsky J, Agianian B, Dodt G, Herberg FW, Kunau WH (1999) Recombinant human peroxisomal targeting signal receptor PEX5. Structural basis for interaction of PEX5 with PEX14. *J Biol Chem* **274**: 5666-5673

Schwarz DS, Hutvagner G, Du T, Xu Z, Aronin N, Zamore PD (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**: 199-208

Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* **160**: 65-73

Selenko P, Gregorovic G, Sprangers R, Stier G, Rhani Z, Kramer A, Sattler M (2003) Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol Cell* **11**: 965-976

Shamoo Y, Abdul-Manan N, Patten AM, Crawford JK, Pellegrini MC, Williams KR (1994) Both RNA-binding domains in heterogenous nuclear ribonucleoprotein A1 contribute toward single-stranded-RNA binding. *Biochemistry* **33**: 8272-8281

Shamoo Y, Abdul-Manan N, Williams KR (1995) Multiple RNA binding domains (RBDs) just don't add up. *Nucleic Acids Res* **23**: 725-728

Shamoo Y, Krueger U, Rice LM, Williams KR, Steitz TA (1997) Crystal structure of the two RNA binding domains of human hnRNP A1 at 1.75 Å resolution. *Nat Struct Biol* **4**: 215-222

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

Shen Y, Lange O, Delaglio F, Rossi P, Aramini JM, Liu G, Eletsky A, Wu Y, Singarapu KK, Lemak A, Ignatchenko A, Arrowsmith CH, Szyperski T, Montelione GT, Baker D, Bax A (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proc Natl Acad Sci U S A* **105**: 4685-4690

Shiozawa K, Konarev PV, Neufeld C, Wilmanns M, Svergun DI (2009) Solution structure of human Pex5.Pex14.PTS1 protein complexes obtained by small angle X-ray scattering. *J Biol Chem* **284**: 25334-25342

Simon B, Madl T, Mackereth CD, Nilges M, Sattler M (2010) An efficient protocol for NMR-spectroscopy-based structure determination of protein complexes in solution. *Angew Chem Int Ed Engl* **49**: 1967-1970

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

Singh R, Valcarcel J, Green MR (1995) Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science* **268**: 1173-1176

Siomi H, Dreyfuss G (1995) A nuclear localization domain in the hnRNP A1 protein. *The Journal of cell biology* **129**: 551-560

Siomi H, Siomi MC (2010) Posttranscriptional regulation of microRNA biogenesis in animals. *Mol Cell* **38**: 323-332

Slezak-Prochazka I, Durmus S, Kroesen BJ, van den Berg A (2010) MicroRNAs, macrocontrol: regulation of miRNA processing. *RNA* **16**: 1087-1095

Smith JJ, Aitchison JD (2013) Peroxisomes take shape. *Nat Rev Mol Cell Biol* **14**: 803-817

Soares LM, Zanier K, Mackereth C, Sattler M, Valcarcel J (2006) Intron removal requires proofreading of U2AF/3' splice site recognition by DEK. *Science* **312**: 1961-1965

Sridharan V, Heimiller J, Singh R (2011) Genomic mRNA profiling reveals compensatory mechanisms for the requirement of the essential splicing factor U2AF. *Mol Cell Biol* **31**: 652-661

Sridharan V, Singh R (2007) A conditional role of U2AF in splicing of introns with unconventional polypyrimidine tracts. *Mol Cell Biol* **27**: 7334-7344

Stanley WA, Sokolova A, Brown A, Clarke DT, Wilmanns M, Svergun DI (2004) Synergistic use of synchrotron radiation techniques for biological samples in solution: a case study on protein-ligand recognition by the peroxisomal import receptor Pex5p. *Journal of synchrotron radiation* **11**: 490-496

Steffl R, Allain FH (2005) A novel RNA pentaloop fold involved in targeting ADAR2. *RNA* **11**: 592-597

Su JR, Takeda K, Tamura S, Fujiki Y, Miki K (2009) Crystal structure of the conserved N-terminal domain of the peroxisomal matrix protein import receptor, Pex14p. *Proc Natl Acad Sci U S A* **106**: 417-421

Svergun D (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography* **25**: 495-503

Svergun D, Barberato C, Koch MHJ (1995) CRY SOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography* **28**: 768-773

Svergun DI (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* **76**: 2879-2886

Swanson MS, Dreyfuss G (1988) RNA binding specificity of hnRNP proteins: a subset bind to the 3' end of introns. *EMBO J* **7**: 3519-3529

Tanaka Y, Kojima C, Yamazaki T, Kodama TS, Yasuno K, Miyashita S, Ono A, Ono A, Kainosho M, Kyogoku Y (2000) Solution structure of an RNA duplex including a C-U base pair. *Biochemistry* **39**: 7074-7080

Tange TO, Damgaard CK, Guth S, Valcarcel J, Kjems J (2001) The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. *EMBO J* **20**: 5748-5758

Tavanez JP, Madl T, Kooshapur H, Sattler M, Valcarcel J (2012) hnRNP A1 proofreads 3' splice site recognition by U2AF. *Mol Cell* **45**: 314-329

Thandapani P, O'Connor TR, Bailey TL, Richard S (2013) Defining the RGG/RG motif. *Mol Cell* **50**: 613-623

Trabucchi M, Briata P, Garcia-Mayoral M, Haase AD, Filipowicz W, Ramos A, Gherzi R, Rosenfeld MG (2009) The RNA-binding protein KSRP promotes the biogenesis of a subset of microRNAs. *Nature* **459**: 1010-1014

Valcarcel J, Gaur RK, Singh R, Green MR (1996) Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. *Science* **273**: 1706-1709

Ventura A, Jacks T (2009) MicroRNAs and cancer: short RNAs go a long way. *Cell* **136**: 586-591

Ventura A, Young AG, Winslow MM, Lintault L, Meissner A, Erkeland SJ, Newman J, Bronson RT, Crowley D, Stone JR, Jaenisch R, Sharp PA, Jacks T (2008) Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters. *Cell* **132**: 875-886

Vitali J, Ding J, Jiang J, Zhang Y, Krainer AR, Xu RM (2002) Correlated alternative side chain conformations in the RNA-recognition motif of heterogeneous nuclear ribonucleoprotein A1. *Nucleic Acids Res* **30**: 1531-1538

Volkov VV, Svergun DI (2003) Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography* **36**: 860-864

Wahl MC, Will CL, Luhrmann R (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**: 701-718

Walton PA, Hill PE, Subramani S (1995) Import of stably folded proteins into peroxisomes. *Mol Biol Cell* **6**: 675-683

Wang ET, Sandberg R, Luo S, Khrebtkova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**: 470-476

Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189-1191

Weighardt F, Biamonti G, Riva S (1995) Nucleo-cytoplasmic distribution of human hnRNP proteins: a search for the targeting domains in hnRNP A1. *J Cell Sci* **108** (Pt 2): 545-555

Williams C, van den Berg M, Distel B (2005) *Saccharomyces cerevisiae* Pex14p contains two independent Pex5p binding sites, which are both essential for PTS1 protein import. *FEBS Lett* **579**: 3416-3420

Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**: 235-242

Winter J, Jung S, Keller S, Gregory RI, Diederichs S (2009) Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat Cell Biol* **11**: 228-234

Wishart DS, Sykes BD (1994) The ¹³C chemical-shift index: a simple method for the identification of protein secondary structure using ¹³C chemical-shift data. *J Biomol NMR* **4**: 171-180

Wishart DS, Sykes BD, Richards FM (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**: 1647-1651

Wu J, Manley JL (1989) Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. *Genes Dev* **3**: 1553-1561

Wu S, Romfo CM, Nilsen TW, Green MR (1999) Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* **402**: 832-835

Xu RM, Jokhan L, Cheng X, Mayeda A, Krainer AR (1997) Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs. *Structure* **5**: 559-570

Zamore PD, Green MR (1989) Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc Natl Acad Sci U S A* **86**: 9243-9247

Zamore PD, Patton JG, Green MR (1992) Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**: 609-614

Zeng Y, Cullen BR (2003) Sequence requirements for micro RNA processing and function in human cells. *RNA* **9**: 112-123

Zeng Y, Cullen BR (2005) Efficient processing of primary microRNA hairpins by Drosha requires flanking nonstructured RNA sequences. *J Biol Chem* **280**: 27595-27603

Zeng Y, Yi R, Cullen BR (2005) Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *EMBO J* **24**: 138-148

Zhang M, Zamore PD, Carmo-Fonseca M, Lamond AI, Green MR (1992) Cloning and intracellular localization of the U2 small nuclear ribonucleoprotein auxiliary factor small subunit. *Proc Natl Acad Sci U S A* **89**: 8769-8773

Zhang MQ (1998) Statistical features of human exons and their flanking regions. *Hum Mol Genet* **7**: 919-932

Zhang QS, Manche L, Xu RM, Krainer AR (2006) hnRNP A1 associates with telomere ends and stimulates telomerase activity. *RNA* **12**: 1116-1128

Zhang X, Zeng Y (2010) The terminal loop region controls microRNA processing by Drosha and Dicer. *Nucleic Acids Res* **38**: 7689-7697

Zheng L, Baumann U, Reymond JL (2004) An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Res* **32**: e115

Zhu J, Mayeda A, Krainer AR (2001) Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol Cell* **8**: 1351-1361

Zhuang Y, Weiner AM (1989) A compensatory base change in human U2 snRNA can suppress a branch site mutation. *Genes Dev* **3**: 1545-1552

Zorio DA, Blumenthal T (1999) Both subunits of U2AF recognize the 3' splice site in *Caenorhabditis elegans*. *Nature* **402**: 835-838

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**: 3406-3415

11. Appendix

11.1 Primer sequences

Table 11.1 Primers used for cloning of hnRNP A1 constructs

Primer	Sequence (5' to 3')
A1_1_for	TGATCCATGGGCTCTAAGTCAGAGTCTCCTAAAGAGC
A1_94_for	TGATCCATGGATTCTCAAAGACCAGGTGCCAC
A1_195_for	TGATCCATGGGTCTGAAGTGGTTCTGGAACTTTGG
A1_97_rev	TCATGGTACCTTATCTTTGAGAATCTTCTCTGGAG
A1_196_rev	TCATGGTACCTTATCGACCTCTTTGGCTGGATGAAGC
A1_320_rev	TCATGGTACCTTAAAATCTTCTGCCACTGCCATAGC
A1_320_rev (XhoI)	TACTCTCGAGAAATCTTCTGCCACTGCCATAGC

Table 11.2 Primers used for cloning of the pUC-pri-miR-18a construct

Primer	Sequence (5' to 3')
18a_for	AATTGAATTCTAATACGACTCACTATAGGTTCTAAGGTGC ATCTAGTGC
18a_rev	GAAGTCTAGAGAAGACAAGGCCAGAAGGAGCACTTAGG

Table 11.3 List of primers used for site-directed mutagenesis of hnRNP A1 constructs

Primer	Sequence (5' to 3')
L33M_for	GAAACAACCTGATGAGAGCATGAGGAGCCATTTTGAGC
L33M_rev	GCTCAAATGGCTCCTCATGCTCTCATCAGTTGTTTC
L124M_for	GAAGACACTGAAGAACATCACATGAGAGATTATTTGAACAGTATG
L124M_rev	CATACTGTTCAAATAATCTCTCATGTGATGTTCTTCAGTGTCTTC
C43A_for	GGGGAACGCTCACGGACGCTGTGGTAATGAGAGATC
C43A_rev	GATCTCTCATTACCACAGCGTCCGTGAGCGTTCCCC
C175S_for	CTGTGAATGGCCACAACAGTGAAGTTAGAAAAGCC
C175S_rev	GGCTTTTCTAACTTCAGTGTGGCCATTCACA
E66C_for	GCCACTGTGTGCGAGGTGGATGCAGCTATGAATGCAAGGCCACAC
E66C_rev	CACCTCGCACACAGTGGCATATGTGACAAACCCAAAGCCCCTGG
N73C_for	GCAGCTATGTGTGCAAGGCCACACAAGGTGGATGGAAGAG
N73C_rev	CCTTGCACACATAGCTGCATCCACCTCCTCCACAGTGGC
Q127C_for	GATTATTTTGAATGCTATGGAAAATTGAAGTGATTGAAATCATGAC
Q127C_rev	TCCATAGCATTCAAATAATCTCTTAGGTGATGTTCTTCAGTGTC
Q165C_for	AAGATTGTCATTTGCAAATACCATACTGTGAATGGCCACAAC
Q165C_rev	GTATTTGCAAATGACAATCTTATCCACGGAGTCATGGTCTG
R75E_for	ATGAATGCAGAGCCACACAAGGTGGATGGAAGAGTTGTGG
R75E_rev	GTGTGGCTCTGCATTCATAGCTGCATCCACCTCCTCCAC
R88E_for	GGAAGAGTTGTGGAACCAAAGGAGGCTGTCTCCAGAGAAGATTCT
R88E_rev	AGAATCTTCTCTGGAGACAGCCTCCTTTGGTTCCACAACCTCTTC
R88A_for	AAGAGTTGTGGAACCAAAGGCAGCTGTCTCCAGAGAAGAT
R88A_rev	ATCTTCTCTGGAGACAGCTGCCTTTGGTTCCACAACCTCTT

Table 11.4 List of primers used in chapter 8

Primer	Sequence (5' to 3')
T7 promoter	TAATACGACTCACTATAG
T7 terminator	GCTAGTTATTGCTCAGCGG
KU582	CTAGGTCGACTATACATCCACAGCATCTCCAGC
KU644	CTAGGTCGACCTAGGCCAAGTCTGCCACACC
KU690	TGGCCTTGTCTGAGAATGCGGCCAGGAGTTTCTT
KU1101	GGAGTAGCTTCTGAAGATGAGGCGGCGGCGGCGCTGCAGGACCAGAAT GCACCC
KU1102	GGGTGCATTCTGGTCCTGCAGCGCCGCCGCCGCCCTCATCTTCAGAAGCT ACTCC
KU1255	TATAATGAGACTGACGCGTCCCAAGAATTCATCTCTG
KU1278	AAGAACTCCTGGGCCGATTCTCAGACAAGGCCA
KU1279	CAGAGATGAATTCTTGGGACGCGTCAGTCTCATTATA
RE2142	GAGTAGCTTCTGAAGATGAGTGGGCGCAGGAGTTCCTGCAGGACCAGAATGC
RE2143	GCATTCTGGTCCTGCAGGAACTCCTGCGCCCACTCATCTTCAGAAGCTACTC
RE3533	GATCAGATCTATGGCAATGCGGGAGCTGG
RE3534	GATCGTCGACTCACTGGGGCAGGCCAAAC
1-110_F	TAATCCATGGCAATGCGGGAGCTGGTGG
1-110_R	TATTGGTACCTTATGCCACACCAGGGGCTCTCTGG

11.2 Protein sequences and properties

Sequence of protein constructs and their properties calculated using the ProtParam tool (Gasteiger et al, 2005). Residues which are not part of the native sequence are underlined.

RRM1 (1-97)

GAMGSKSESPKEPEQLRKLFIGGLSFETTDESLRSHFEQWGTLTDCVVMRDPNTK
RSRGFGFV TYATVEEVDAAMNARPHKVDGRVVEPKRAVSREDSQR

Number of amino acids: 100

Molecular weight: 11230.5 Da

Theoretical pI: 6.95

Extinction coefficient (at 280 nm): 6990 M⁻¹ cm⁻¹ and Abs 0.1% (=1 g/l): 0.622

RRM2 (94-196)

GAMDSQRPGAHLTVKKIFVGGIKEDTEEHHLRDYFEQYGKIEVIEIMTDRGSGKKR
GFAFVTFDDHDSVDKIVIQKYHTVNGHNCEVRKALSKQEMASASSSQRGR

Number of amino acids: 106

Molecular weight: 11981.4 Da

Theoretical pI: 8.08

Extinction coefficient (at 280 nm): 4470 M⁻¹ cm⁻¹ and Abs 0.1% (=1 g/l): 0.373

UP1 (1-196)

GAMGSKSESPKEPEQLRKLFIGGLSFETTDESLRSHFEQWGTLTDCVVMRDPNTK
RSRGFGFV TYATVEEVDAAMNARPHKVDGRVVEPKRAVSREDSQRPGAHLTVKKI

FVGGIKEDTEEHHLRDYFEQYGKIEVIEIMDRGSGKKRGFAFVTFDDHDSVDKIVIQ
KYHTVNGHNCEVRKALSKQEMASASSSQRGR

Number of amino acids: 199

Molecular weight: 22448.1 Da

Theoretical pI: 7.96

Extinction coefficient (at 280 nm): 11585 M⁻¹ cm⁻¹ and Abs 0.1% (=1 g/l): 0.516
(assuming all Cys residues appear as half cystines)

Extinction coefficient (at 280 nm): 11460 M⁻¹ cm⁻¹ and Abs 0.1% (=1 g/l): 0.511
(assuming no Cys residues appear as half cystines)

Gly-rich tail (195-320)

GAMGRSGSGNFGGGRGGGFGGNDNFGRGGNFSGRGGFGGSRGGGGYGGSGD
GYNGFGNDGSNFGGGGSYNDFGNYNQSSNFGPMKGGNFGGRSSGPYGGGGQ
YFAKPRNQGGYGGSSSSSSSYGSGRRF

Number of amino acids: 129

Molecular weight: 12423.7 Da

Theoretical pI: 10.15

Extinction coefficient (at 280 nm): 11920 M⁻¹ cm⁻¹ and Abs 0.1% (=1 g/l): 0.959

Full-length hnRNP A1 (1-320)

GAMGSKSESPKEPEQLRKLFIGGLSFETTDESLRSHFEQWGTLTDCVVMRDPNTK
RSRGFGFVYATVEEVDAAMNARPHKVDGRVVEPKRAVSREDSQRPGAHLTVKKI
FVGGIKEDTEEHHLRDYFEQYGKIEVIEIMDRGSGKKRGFAFVTFDDHDSVDKIVIQ
KYHTVNGHNCEVRKALSKQEMASASSSQRGRSGSGNFGGGRGGGFGGNDNFG
RGGNFSGRGGFGGSRGGGGYGGSGDGYNGFGNDGSNFGGGGSYNDFGNYNQ
QSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGYGGSSSSSSSYGSGRRF

Number of amino acids: 323

Molecular weight: 34381.3 Da

Theoretical pI: 9.27

Extinction coefficient (at 280 nm): $23505 \text{ M}^{-1} \text{ cm}^{-1}$ and Abs 0.1% (=1 g/l): 0.684
(assuming all Cys residues appear as half cystines)

Extinction coefficient (at 280 nm): $23380 \text{ M}^{-1} \text{ cm}^{-1}$ and Abs 0.1% (=1 g/l): 0.680
(assuming no Cys residues appear as half cystines)

Human Pex5 (1-110)

GAMAMRELVEAECGGANPLMKLAGHFTQDKALRQEGLRPGPWPPGAPASEAASK
PLGVASEDELVAEFLQDQNAPLVSRAPQTFKMDLLAEMQQIEQSNFRQAPQRAP
GVA

Number of amino acids: 112

Molecular weight: 11981.5 Da

Theoretical pI: 4.77

Extinction coefficient (at 280 nm): $5500 \text{ M}^{-1} \text{ cm}^{-1}$ and Abs 0.1% (=1 g/l) = 0.459

Human Pex14 (16-80W)

GAMATPGSENVLPREPLIATAVKFLQNSRVRQSPLATRRRAFLKKKGLTDEEIDMAF
QQSGTAADEPSSLW

Number of amino acids: 70

Molecular weight: 7630.6 Da

Theoretical pI: 8.43

Extinction coefficient (at 280 nm): $5500 \text{ M}^{-1} \text{ cm}^{-1}$ and Abs 0.1% (=1 g/l) = 0.753

Abbreviations

1D, 2D, 3D	one-, two-, three-dimensional
Å	Angstrom (10^{-10} m)
ALS	Amyotrophic lateral scelerosis
ATP	Adenosine tri-phosphate
BMRB	Biological Magnetic Resonance Bank
BSA	Bovine serum albumin
CD	Circular Dichroism
CSP	Chemical shift perturbation
CTP	Cytidine tri-phosphate
Da	Dalton
Dmax	Maximum dimension
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DTT	Dithiothreitol
E.coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic Mobility Shift Assay
g	Gravitational acceleration
GTP	Guanosine Triphosphate
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMQC	Heteronuclear Multiple Quantum Coherence
hnRNP	heterogenous nuclear ribonucleoprotein
HSQC	heteronuclear single quantum correlation
Hz	Hertz
IPTG	Isopropyl β -D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
K _D	Dissociation constant
LB	Lysogeny broth
MAD	Multi-wavelength anomalous diffraction
MES	2-(<i>N</i> -morpholino)ethanesulfonic acid

min	Minute
MPD	2-Methyl-2,4-pentanediol
Ni-NTA	Nickel-Nitrilotriacetic acid
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect spectroscopy
ns	Nanoseconds
nt	Nucleotide
NTP	Nucleoside Triphosphate
OD ₆₀₀	Optical density at 600 nm wavelength
PAGE	Polyacrylamide Gel Electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank
Pex	Peroxisomal protein
pI	Isoelectric point
ppm	parts per million
PRE	Paramagnetic relaxation enhancement
pre-miRNA	precursor micro-RNA
pre-mRNA	precursor messenger RNA
pri-miR	primary micro-RNA
PTS	Peroxisomal targeting signal
qRT-PCR	quantitative reverse-transcription PCR
RDC	residual dipolar coupling
R _g	Radius of gyration
rmsd	Root mean square deviation
RNA	Ribonucleic Acid
RNAse	Ribonuclease
RRM	RNA recognition motif
SAD	Single anomalous diffraction
SANS	Small angle neutron scattering
SAXS	Small angle x-ray scattering
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis

SHAPE	selective 2'-hydroxyl acylation analyzed by primer extension
SOFAST	band-selective optimized flip-angle short-transient
TEV	tobacco etch virus
TOCSY	total correlation spectroscopy
Tris	Tris(hydroxymethyl)aminomethane
TROSY	transverse relaxation optimized spectroscopy
U2AF	U2 snRNP auxiliary factor
UHM	U2AF homology motifs
ULM	UHM ligand motifs
UTP	Uridine tri-phosphate

Acknowledgements

I would like to express my sincere gratitude to my supervisor Dr. Michael Sattler for giving me the opportunity to work on many interesting projects in his group and the freedom to follow my own ideas. I am very grateful for his encouragement and support over the years.

I would like to acknowledge the members of my thesis advisory committee Dr. Dierk Niessing and Dr. Gunter Meister for their valuable suggestions. I am very grateful to all my collaborators in particular Dr. Javier Caceres, Dr. Gracjan Michlewski and Dr. Wolfgang Schliebs for sharing unpublished data and stimulating discussions.

Many thanks to Dr. Jerome Basquin at the MPI crystallization facility for all his efforts in setting up crystallization trials and collecting data at the synchrotron and Dr. Andreas Bracher for his help and advice. I also would like to thank Dr. Robert Janowski at the Helmholtz Zentrum München X-ray crystallography platform and Dr. Arie Geerlof at PEPF for their excellent support.

I am deeply grateful to Dr. Fatiha Kateb for her help with NMR analysis. In addition, I would like to thank Dr. Andre Dallmann, Dr. Tobias Madl and Dr. Helge Meyer for collaboration and insightful discussions on a number of projects.

I have been fortunate to work with three talented students: Philipp Wortmann, Martin Rübhelke and Johannes Günther. I thank them for their contributions and wish them all the best in their future careers.

Of course, thanks to the past and present members of the Sattler group, Kessler group, Reif group and Madl group for day-to-day life in the lab and all the fun.

A big thank you to all the people who have shared the famous big office with me: Fatiha Kateb, Divita Garg, Iren Wang, Irina Anosova, Giambattista Guaitoli, Ashish Kawale, Carolina Sánchez Rico, Johannes Günther and Mohan Gopalswamy for creating such a nice and unique working atmosphere.

My special thanks goes to Alex Beribisky and Elke Prade for their friendship and all the memorable times we have spent together, the coffee-breaks, late night stays in the lab, trips to conferences and much more!

Last but not least, I owe my deepest gratitude to my family and in particular my parents for their continuous love, support and patience.

Curriculum Vitae

HAMED KOOSHAPUR

Department Chemie, Technische Universität München
Lichtenbergstr. 4,
85747 Garching, Germany
E-mail: hamed.kooshapur@tum.de

EDUCATION

Technical University of Munich (TUM), Germany	2008-2014
PhD studies	
University of Heidelberg, Germany	2005-2007
M.Sc in Molecular and Cellular Biology	
University of Tehran, Iran	2000-2005
B.Sc in Molecular and Cellular Biology	

RESEARCH EXPERIENCE

PhD thesis	2008-2014
Technical University of Munich and Institute of Structural Biology, Helmholtz Zentrum München Advisor: Prof. Dr. Michael Sattler	
Research scientist	2007-2008
University of Heidelberg Biochemistry Center (BZH)	
M.Sc thesis	2006-2007
University of Heidelberg Biochemistry Center (BZH) Title: "Biochemical and structural characterization of the human proliferation-associated 2G4 protein (PA2G4)" Advisor: Prof. Dr. Irmgard Sinning	

TEACHING EXPERIENCE

Department of Chemistry, Technical University of Munich, Germany	
Teacher in Biomolecular NMR Spectroscopy practical course	2012-2009
Co-supervisor of a master thesis project (duration 6-7 months)	2012
Supervisor of a bachelor thesis project (duration 3 months)	2011
Teacher in organic chemistry practical course	2010
Supervisor of a bachelor thesis project (duration 3 months)	2009

PUBLICATIONS

Neuhaus A*, **Kooshapur H***, Wolf J, Meyer NH, Madl T, Saidowsky J, Hambruch E, Lazam A, Jung M, Sattler M, Schliebs W, Erdmann R.

“A novel Pex14 protein-interacting site of human Pex5 is critical for matrix protein import into peroxisomes.”

(2014) **J Biol Chem.** 289(1), 437-48 (**JBC paper of the week**) **contributed equally*

Dallmann A, Simon B, Duszczyk MM, **Kooshapur H**, Pardi A, Bermel W, Sattler M.

“Efficient detection of hydrogen bonds in dynamic regions of RNA by sensitivity-optimized NMR pulse sequences.”

(2013) **Angew Chem Int Ed Engl.** 52(40), 10487-90

Tavanez JP, Madl T, **Kooshapur H**, Sattler M, Valcárcel J.

“hnRNP A1 proofreads 3' splice site recognition by U2AF.”

(2012) **Molecular Cell.** 45(3), 314-29

Neufeld C, Filipp FV, Simon B, Neuhaus A, Schüller N, David C, **Kooshapur H**, Madl T, Erdmann R, Schliebs W, Wilmanns M, Sattler M.

“Structural basis for competitive interactions of Pex14 with the import receptors Pex5 and Pex19.”

(2009) **EMBO Journal.** 28(6), 745-54

CONFERENCES AND WORKSHOPS

EUROMAR (2012, Dublin, Ireland)

Sixteenth Annual meeting of the RNA Society (2011, Kyoto, Japan)

EMBO practical course: Structure, dynamics and function of biomacromolecules by solution NMR (2011, Garching, Germany)

EMBO practical course: Small Angle Neutron & X-ray Scattering from proteins in solution (2011, Grenoble, France)

Murnau Conference 2010 Structural Biology of the RNA world (2010, Murnau, Germany)

Open European Peroxisome Meeting (2010, Lunteren, The Netherlands)

Microcal Advanced ITC course (2010, Munich, Germany)

NMR-Life Workshop: Protein expression and isotope labelling for structural biology (2008, Munich, Germany)