

How to study the structure and dynamics of protein-RNA complexes by NMR spectroscopy

Mario SCHUBERT^a, Cyril DOMINGUEZ^a, Olivier DUSS^a, Sapna RAVINDRANATHAN^b and Frédéric H.-T. ALLAIN^{a,1}

^a*Institute for Molecular Biology and Biophysics, ETH Zürich, CH-8093 Zürich, Switzerland*

^b*Central NMR Facility, National Chemical Laboratory, Pune 411008, India*

Abstract. The role of protein-RNA recognition is fundamental to many biological processes, protein-RNA interaction being at the heart of every molecular mechanisms controlling post-transcriptional gene expression. Deciphering the 3D structures of protein-RNA complexes is therefore of high significance. RNA binding proteins are very abundant in all kingdoms of life and often embed one to several small RNA binding domains. Since these domains often act as independent units, NMR spectroscopy is ideally suited to study the structure and dynamics of such domains in complex with their RNA targets. We review here how NMR spectroscopy has been used to solve the structure of more than fifty protein-RNA complexes and to understand for a few their dynamics.

Keywords. Protein-RNA complex, NMR spectroscopy, structure determination, dynamics

Introduction

In March 2010, 141 structures of protein-RNA complexes with a molecular weight lower than 40 kDa have been deposited in the Protein Data Bank (PDB). Of these structures, 52 were determined using classical NMR methodology, resulting in a contribution of 37%. This figure alone illustrates the important role of NMR spectroscopy in elucidating structures of protein-RNA complexes within this molecular weight range. NMR spectroscopy is expected to become increasingly prominent in structural and RNA biology, considering the growing understanding of the importance of protein-RNA interactions in the regulation of gene expression.

The first NMR structure of a peptide-RNA complex was determined in 1995 and consisted of the structure of a small peptide (14 amino acids) of the viral Tat protein bound to the 26-nucleotide RNA stem-loop of TAR [1,2]. One year later, in 1996, the first structure of a protein-RNA complex was solved by NMR spectroscopy. This was the structure of the N-terminal RNA recognition motif (RRM) of the U1A protein (100 amino acids) in complex with a 30-nt stem-loop RNA [3]. The 52 structures of protein-RNA/peptide-RNA complex structures available today provide us with the opportunity

¹ Corresponding Author: Frédéric H.-T. Allain, Institute for Molecular Biology and Biophysics, ETH Zürich, CH-8093 Zürich, Switzerland; Email: allain@mol.biol.ethz.ch

to review how those structures were determined and what we learned from them. The first part of the review (1.1-1.2) describes what makes a protein-RNA complex amenable for NMR structure determination and how the appropriate solution conditions can be obtained. The second part focuses on the NMR spectroscopy of these complexes (2.1-2.43) and how, from the NMR spectra, one can derive a precise structure of a protein-RNA complex (2.4) including a discussion on the validation of the resulting structures (2.5). The third part describes what we can learn from the few dynamics studies of protein-RNA complexes performed using NMR. This review ends with a brief description of a few examples of protein-RNA complexes determined by NMR spectroscopy and a discussion on how these have impacted the field of structural and RNA biology.

1. How to get a Protein-RNA Complex Sample for NMR Spectroscopy

1.1. Finding Optimal Protein and RNA Constructs for NMR Studies of Protein-RNA Complexes

Before starting an NMR study of a protein-RNA complex, biological and biochemical knowledge of the complex is crucial. Most RNA binding proteins (RBPs) are easily identifiable since they often contain well-known RNA binding domains (RBDs). However, finding the RNA sequence that is recognized by the protein of interest is often not trivial for several reasons. First, RBPs or RBDs can recognize and bind RNA in a shape-specific, sequence-specific or even non-specific manner. Second, RNA molecules can form a variety of secondary and tertiary structures that might be crucial for protein recognition. For studying protein-RNA complexes by NMR, it is of particular importance to understand the specificity of the interaction. The main questions that should be addressed concerning NMR studies of a protein-RNA complex are:

- What is the minimum protein domain necessary for RNA binding? Are RBDs sufficient for efficient RNA binding?
- Does the protein or the RBD bind single-stranded RNA (ssRNA) or double-stranded RNA (dsRNA)?
- Is the interaction specific? If yes, is it shape-specific or sequence-specific?
- If shape-specific, how does the RNA structure influence the binding?
- If sequence-specific, which RNA sequence is specifically recognized?

The major challenge consists of identifying a small biological relevant RNA sequence that is bound both specifically and with sufficient affinity by the protein. There are many techniques and methods to identify protein-RNA interactions. Some of these techniques, such as protein-RNA cross-linking, immunoprecipitation or affinity purification, aim at the identification of natural RNA sequences specifically bound by RNA binding proteins. Other techniques allow the definition of RNA aptamers that are bound with high affinity by RNA binding proteins but those might not be natural sequences.

1.1.1. Finding Minimum Natural RNA Substrates Bound by RNA Binding Proteins

Natural RNA sequences used for NMR structure determination of protein-RNA complexes have usually been mainly derived from two different strategies: the use of

footprinting techniques that were initially developed for protein-DNA complexes [4], or the use of RNA truncation and mutagenesis combined with protein-RNA binding assays, e.g. Electrophoretic Mobility Shift Assay (EMSA) [5, 6]. For example, footprinting experiments have been performed to define the RNA region of the 5S rRNA bound by the protein L25 [7] which was then used to solve the NMR structure of the L25-5S rRNA complex [8]. RNA truncation and mutagenesis together with binding assays have, for example, been used to define the minimal RNA sequence recognized by the HIV Rev protein [9] which was then used for the structure determination of the HIV-1 Rev-RRE complex [10]. Recently, a novel *in vivo* method called CLIP (UV Cross-Linking and ImmunoPrecipitation assay) has been developed to identify natural RNA targets of RNA binding proteins using high-throughput technologies [11, 12]. The RNA sequence retrieved from CLIP experiments with the protein Fox-2 [13] corresponds to the sequence identified by SELEX (see next section) and by the NMR structure of the complex [14, 15]. This indicates that CLIP has a high potential for identifying natural RNA binding sequences suitable for NMR investigation of protein-RNA complexes.

1.1.2. Finding High-Affinity RNA Aptamers Bound by RNA Binding Proteins

The main technique used to identify non-natural RNA (aptamers) bound by RNA binding proteins is the Systematic Evolution of Ligands by Exponential Enrichment (SELEX) approach [16, 17]. An RNA binding sequence identified by SELEX provides a very good starting point for the NMR investigation of a protein-RNA complex as SELEX experiments select sequences with high affinity for the RNA binding protein. Numerous protein-RNA complexes have been solved using RNA sequences initially derived from SELEX that also match perfectly natural sequences [15, 18, 19]. Nonetheless, SELEX results can also differ from natural RNA sequences. This can be due to the fact that either the natural RNA sequence was not represented in the pool of random RNA sequences used for the SELEX procedure, or that the SELEX-derived RNA aptamer has a stronger affinity for the protein than the natural RNA [20-22].

1.1.3. Optimizing the RNA Target of a Protein-RNA Complex for its NMR Study

To optimize a protein-RNA complex for NMR studies, three main aspects are of particular importance: the stability of the complex, the quality of the NMR spectra, and the presence of intermolecular NOEs that are essential for the structure calculation of the complex (see sections 2.1 and 2.2). When the protein binds stem-loop structures, the RNA can be stabilized by extending the RNA stem by additional G-C base pairs in order to improve the quality of the spectra by minimizing fraying [23]. In cases where the protein is known to bind the stem but not the loop, the loop sequence can be optimized to stabilize the fold of the RNA, e.g. in the case of the L25-5S rRNA complex, the natural apical loop was replaced by a highly stable UUCG tetraloop in order to stabilize the structure of the RNA [8].

In the case of RNA binding proteins or domains that specifically bind ssRNA, often only a few nucleotides are specifically recognized; flanking nucleotides might influence and increase the stability of the complex. One way to optimize the length of the RNA is to perform chemical shift perturbation experiments with RNAs of different lengths. The quality of the NMR spectra can be quite sensitive to the RNA length, in particular when the RNA sequence is degenerate, e.g. the RRM of PTB bind to the RNA target CUCUCU in two registers because the RNA contains two UCU motifs,

leading to line broadening in the RNA resonances and two sets of intermolecular NOEs [24, 25]. Therefore, shorter RNA sequences with lower affinity for the protein were tested and a CUCU RNA gave intense intermolecular NOEs that corresponded to one single conformation of the complex. Sometimes SELEX consensus sequences are degenerate, e.g. the consensus sequence (^{A/U})C(^{A/U})(^{A/U})C was found for the protein SRp20 [26]. To solve the NMR structure of SRp20 in complex with RNA, a total of 13 different RNA sequences were tested by NMR ¹⁵N-HSQC and ¹H-¹H-TOCSY spectra [18]. This analysis identified the RNA sequence CAUC as the optimal sequence for obtaining good NMR spectra for both components of the complex and was therefore used for the structure determination of the complex [18]. A systematic approach to elucidate the binding specificity of protein-RNA complexes by NMR at the single nucleotide level, the Scaffold-Independent Analysis (SIA), has been developed in 2007 by Ramos and coworkers [27] and is based on synthetic randomized RNA sequences. The main advantage of this method is that the analysis is directly performed by NMR and therefore, the quality of the spectra is directly assessed for each protein-RNA complex. This approach has been recently used for the structure determination of the RRM2 of Prp24 in complex with the optimized RNA sequence AGAGAU [28].

1.2. Complex Formation

The protein component is typically expressed in *E. coli* or by using *in vitro* expression (see chapter 1), whereas the RNA is obtained by *in vitro* transcription or chemical synthesis (see Chapters 2, 11 and 12). Once the RNA binding protein or domain of interest and the RNA have been produced in sufficient amounts for NMR analysis, both components are mixed together in order to form the protein-RNA complex.

1.2.1. Ribonuclease Activity

Since the protein of interest will be studied in complex with RNA, it is very important, to first test whether the protein sample possesses ribonuclease (RNase) activity. The effect of RNases is highly dependent on the RNA sequence and structure, the affinity of the complex and the nature of interaction. Generally, structured RNAs, such as those forming stem-loop structures, are less prone for degradation than small ssRNAs. Furthermore, ssRNAs that bind proteins with high affinity are generally less prone to degradation than RNAs that bind proteins with low affinity. Therefore, traces of RNases in the sample solution do not necessarily hamper the NMR study.

Sometimes it is difficult to remove RNases by protein purifications, e.g. in the case of LicT-RNA complex, RNase activity could not be eliminated and NMR samples were therefore only stable for few days in the NMR spectrometer [40]. To slow down the degradation of the RNA during NMR measurements, RNase inhibitors can be added into the final buffer [21, 29]. Our experience showed that additional protein purification steps are sometimes very effective in eliminating the RNase activity of the sample, e.g. in the case of the SRp20-RNA complex, three consecutive Ni-NTA purification steps were necessary to eliminate RNase activity [17]. An additional extensive washing step with a high salt buffer was used to eliminate RNase activity during the purification of the protein RsmE by Ni-NTA chromatography [23].

To test the presence of RNase activity in the sample, RNase activity tests are commercially available using a cleavable fluorescent-labeled RNase substrate. RNase activity can also be measured by monitoring the appearance of degradation signals over

time using 2D ^1H - ^1H -TOCSY spectra (generally by following the pyrimidines H5-H6 cross-peaks).

1.2.2. Monitoring Complex Formation by NMR Titration Experiments

Complex formation can easily be monitored by NMR spectroscopy using chemical shift perturbation mapping upon titration experiments. The complex can be formed by titrating the RNA into the protein or vice versa. Chemical shift perturbations of the protein are typically monitored by ^{15}N -HSQC spectra in the absence or presence of increasing amounts of RNA. Similarly, RNA chemical shift perturbations are followed using the H5-H6 cross-peaks of pyrimidines in 2D (^1H - ^1H) TOCSY spectra by adding increasing amounts of protein. Additionally, imino protons can be monitored by 1D spectra or if labeled RNA is available by ^{15}N -HSQC spectra. Non-exchangeable protons can be monitored by ^{13}C -HSQC spectra.

The addition of RNA into proteins can lead to irreversible precipitation of the sample which can sometimes be prevented by the addition of the protein into the RNA. Precipitation can often be avoided by performing the titration at low concentrations and subsequently concentrating the protein-RNA complex by ultrafiltration using an appropriate molecular weight cut-off membrane [19].

When the protein-RNA complex is formed, the protein and the RNA are in equilibrium between their free and bound states. This equilibrium is described by the dissociation constant (K_d). During titration experiments of a complex, chemical shifts of nuclei that are at the interface experience a different environment and are perturbed. There are three main exchange regimes that can be observed by NMR and these are largely governed by two parameters: the exchange rate of the complex formation, k_{ex} , and the difference in resonance frequency of a nucleus between the free, ν_A , and the bound states, ν_B . The three main exchange regimes are denoted slow exchange that occurs when k_{ex} is much smaller than $2\pi(\nu_A - \nu_B)$, fast exchange when k_{ex} is much larger than $2\pi(\nu_A - \nu_B)$, and the intermediate exchange regime when k_{ex} is similar to $2\pi(\nu_A - \nu_B)$.

In the slow exchange regime, when a component (for example the RNA) is gradually added to the other component (the protein), two sets of signals are observed, one corresponding to the protein free state and the other one corresponding to the protein bound state, as was observed in the case of the Fox-1-RNA complex [15] (Fig. 1A). The integral of each signal is linearly dependent on the population of the two states and is directly correlated to the molar ratio of both components. Slow exchange regimes were reported for protein-RNA complexes with high affinity corresponding to dissociation constants ranging from 0.5 (Fox-1-UGCAUGU) [15] to 250 nM (protein NC-AACAGU) [29].

In the fast exchange regime, only one NMR signal is visible and corresponds to the weighted average of the signals of the free and the bound states. Upon gradual addition of the RNA to the protein, the signals of the protein gradually shift from the free state towards the bound state, as was observed for the PTB-RNA complex [24] (Fig. 1B). When further addition of RNA no longer affects the chemical shift position of the signal, the NMR signal corresponds to the bound state. Fast exchange regimes were reported for protein-RNA complexes with dissociation constants higher than 15–20 μM (PTB-CUCU [24, 25] or SRp20-CAUC [18]).

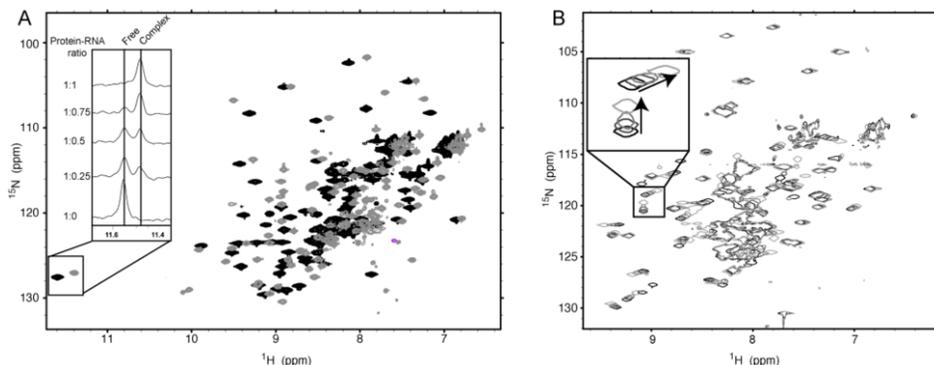


Figure 1. (A) The complex formation of the Fox-1-RNA complex [15] is in slow exchange on the NMR time scale. ^{15}N -HSQC spectra of Fox-1 free (black) or in complex with a UGCAUGU RNA at a 1:1 molar ratio (grey). Insert: 1D spectrum of the signal at 11.5 ppm is shown as a function of the protein:RNA molar ratio. (B) The PTB-RNA complex [24] formation is in fast exchange. ^{15}N -HSQC spectra of PTB RRM1 free (black) or in complex with CUCUCU at protein:RNA molar ratios of 1:0.2, 1:0.4, and 1:3 (decreasing grey scale).

In the intermediate exchange, the NMR signals of the free state undergo line broadening upon addition of the partner, often beyond detection, until more than half the stoichiometry is reached and then the linewidth of the signal corresponding to the bound state sharpens and becomes visible when the stoichiometry of the complex is about to be reached, as was observed for the Staufen-RNA complex [30]. Intermediate exchange regimes were reported for protein-RNA complexes with dissociation constants ranging from 400 nM (hnRNP F-AGGGAU) [31] to 2 μM (CUG-BP1 RRM3-CUGCUG) [32]. However, in some cases, signals of the bound state do not sharpen and are therefore invisible even in excess of the partner component [32, 33]. In this case, an optimization of the conditions should be performed in order to make the resonance of the complex visible [31].

Since the exchange regime depends on the difference between the resonance frequencies of the free and the bound state, it is common to observe different exchange regimes for different signals during a titration experiment. Since the difference of resonance frequency is dependent on the magnetic field, it is also possible to modify the exchange regime of certain signals by recording NMR experiments at different magnetic fields.

In addition, other phenomena can sometimes add to the complexity of the structural analysis. For example, if the protein binds the RNA in multiple registers, especially in the case of proteins binding ssRNAs, additional exchange phenomena can arise. This occurred in the presence of repetitive RNA sequences such as CUCUCU that is bound by the protein PTB [25] as described in section 1.1.3.

2. NMR Spectroscopy and Structure Determination of Protein-RNA Complexes

2.1. NMR Methodology and Resonance Assignment

2.1.1. NMR Titration Experiments are Crucial Initial Steps in Evaluating the Quality of Spectra of Protein-RNA Complexes

Following complex formation by NMR represents a crucial step towards the determination of a protein-RNA complex structure by NMR spectroscopy and gives insights into the stoichiometry of the complex, the exchange regime, the quality of the NMR spectra, the binding interface and an estimation of the binding constants. The most challenging step is to identify both an optimal RNA target and conditions that result in good quality spectra.

In the case of the fast exchange regime, a plot of the chemical shift perturbations as a function of the protein:RNA ratio gives a good indication of the stoichiometry of the complex. For example, in cases where the protein forms a dimer, it was possible to determine if a protein dimer binds one or two RNA molecules [23, 34, 35].

Titration experiments are also used for the optimization of the RNA sequence and the buffer and temperature conditions. For example, initial NMR studies of the complex between the protein hnRNP F and a single-stranded RNA indicated a complex formation in intermediate exchange on the NMR time scale [33] leading to the loss of many amide signals even in excess of RNA. Using ^{15}N -HSQC spectra, different buffer and temperature conditions were tested and optimal conditions could be determined where all amide signals corresponding to the bound form of the protein were present in the spectra [31]. In a second example, out of 13 tested RNA sequences binding the protein SRp20, only one resulted in sharp NMR signals for both the protein and the RNA and was subsequently used for the structure determination of the complex [18].

2.1.2. Further Experiments to Evaluate the Quality of Spectra

After initial NMR titration experiments and the first optimization of conditions, resonance assignments of the bound protein and RNA are the next steps toward solving the structure. However, before proceeding it has to be judged whether a structure determination can be made under the current conditions. Two main criteria have to be fulfilled: first, all or at least nearly all resonances visible in the free state should be visible in the bound state and second, a sufficient number of intermolecular NOEs should be observed in order to solve the complex structure.

All expected amino signals of the protein should be visible in a ^{15}N -HSQC and accordingly all H6-H5 correlations of cytosines and uracils of the RNA in a 2D ^1H - ^1H TOCSY (Fig. 2A). Appearance of imino signals in 1D ^1H spectra upon complex formation is an indication of intermolecular hydrogen bonds like G11 H1 bound to RsmE (Fig. 2B).

Intermolecular NOEs between imino protons and aliphatic protons of the protein can be detected in a 2D NOESY spectrum recorded in H_2O and typically at low temperatures (5–15 °C). Any NOE from an imino to the aliphatic region <2.5 ppm is most probably an intermolecular NOE (Fig. 3A). Additionally, 2D NOESY spectra measured in D_2O can be used to estimate the dispersion of RNA signals and the amount of intermolecular NOEs. Signals in the region between 5 and 6 ppm typically originate from the RNA, in particular H1' and H5 nuclei whereas protein H α signals are rarely found in this region and amide signals are mostly absent in D_2O . Cross-peaks between

resonances at 5–6 ppm and aliphatic protein signals, e.g. upfield of 2 ppm are likely intermolecular NOEs (Fig. 3B). The quality and amount of intermolecular NOEs can be judged by those spectra. Note that for those experiments no isotope labeling is required. Eventually, conditions need to be further optimized or constructs changed if necessary.

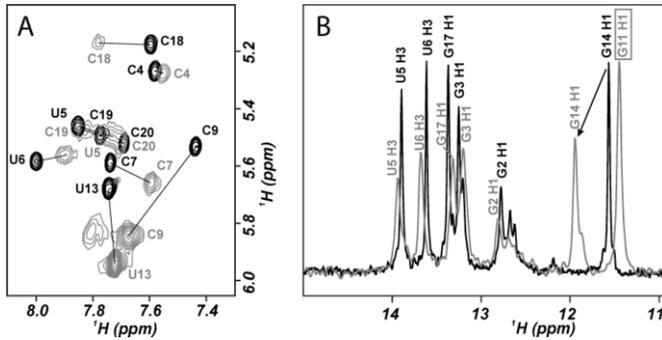


Figure 2. Comparison of 2D ^1H - ^1H TOCSY (A) and 1D NMR spectra (B) of the 20nt Shine-Dalgarno sequence of the *hcnA* mRNA in the free state (black) and complexed with the protein RsmE (grey) measured at 310 K [23]. The new imino signal formed upon complex formation is boxed.

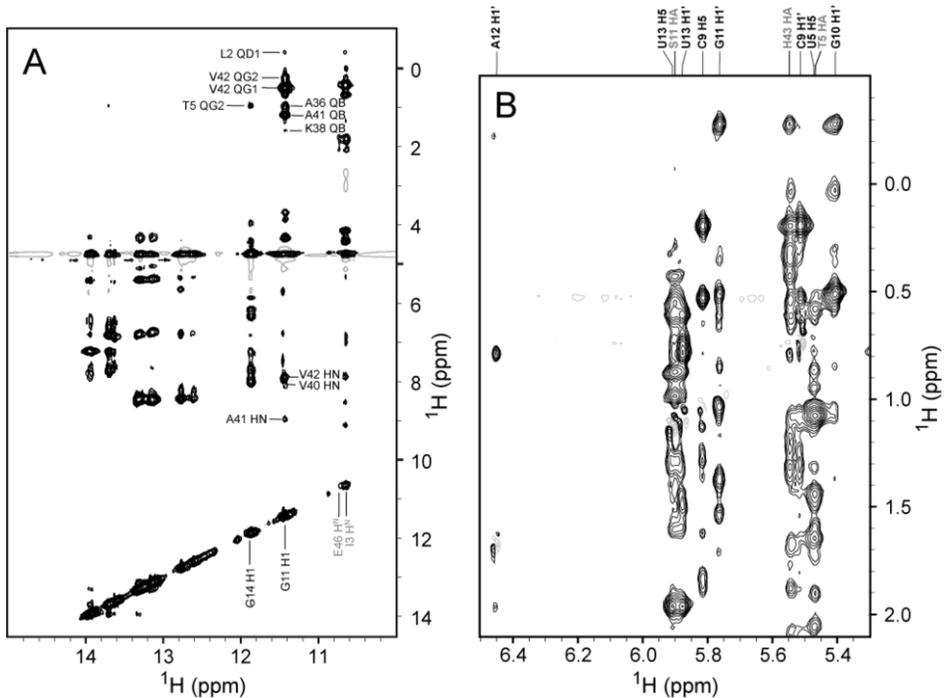


Figure 3. 2D NOESY spectra of RsmE in complex with the Shine-Dalgarno sequence of the *hcnA* mRNA [23]. (A) Imino region of a 2D NOESY in H_2O . The intermolecular NOEs are labeled in black. (B) H1/H5 region of a 2D NOESY in D_2O . RNA resonances are indicated with black labels on the top. Three downfield $\text{H}\alpha$ chemical shifts of the protein are labeled in grey. All cross peaks in this region represent intermolecular NOEs except for the ones involving the three protein $\text{H}\alpha$ resonances of Thr5, Ser11 and His43.

2.1.3. Temperature, Ionic Strength and Solvent

To obtain the best signal to noise and line shape in NMR experiments, factors like temperature and salt concentration have to be optimized in the range that the sample stability allows. With the widely used cryogenic probes, the signal-to-noise ratio decreases significantly in the presence of salt. The ionic strength should therefore be as low as possible, such as buffers without any NaCl or KCl are used, e.g., solely 50 mM Na-phosphate [18]. One promising buffer is arginine/glutamate not only because it can increase protein stability and solubility [18, 36] but it also results in better S/N ratios on cryogenic probes due to the lower ionic strength compared to other buffers [37]. However, the disadvantage is that this buffer causes baseline distortions due to its strong NMR signals; however, these signals can be eliminated using deuterated arginine and glutamate.

After testing the temperature range in which the sample remains in solution, the spectral quality can be tested within this temperature range. Tightly bound RNA can change the stability of a protein significantly and often the complex becomes very stable even at elevated temperatures. For example the complex between RRM3/RRM4 of PTB and CUCUCU RNA could be studied at 40 °C [24] whereas the free form precipitated at 40 °C and was therefore measured at 30 °C [38]. For larger complexes, elevated temperatures such as 40 or 50 °C have the advantage of significantly decreasing the line widths due to faster molecular tumbling rates.

2.1.4. Typical Samples for NMR Measurements of Protein-RNA Complexes

Isotope labeling is absolutely required for resonance and NOESY assignment of protein-RNA complexes. Uniform ^{15}N - and $^{15}\text{N}/^{13}\text{C}$ labeled proteins are usually used in complexes with unlabeled RNA to assign the backbone and side-chain resonances of the protein in complex and to obtain distance restraints within the protein. These samples can also be used to obtain intermolecular distance restraints to the RNA and restraints within the bound RNA using filtered NOESY experiments (section 2.2). If possible, *in vitro* transcribed $^{15}\text{N}/^{13}\text{C}$ labeled RNAs are used for complexes with either unlabeled or ^{15}N labeled proteins. Although uniform $^{15}\text{N}/^{13}\text{C}$ labeling of RNA is most often used, nucleotide-type specific labeling schemes can provide certain advantages.

The structure determination of several recently determined protein-RNA complexes using *in vitro*-transcribed RNA were made with four samples: one containing ^{15}N -labeled protein and unlabeled RNA, one containing $^{15}\text{N}/^{13}\text{C}$ -labeled protein and unlabeled RNA and two samples with nucleotide specific $^{15}\text{N}/^{13}\text{C}$ -labeled RNA in complex with ^{15}N -labeled protein. Two samples of combined nucleotide specific $^{15}\text{N}/^{13}\text{C}$ labeling proved to be most useful, e.g. one sample containing labeled A+C and another sample labeled G+U [23, 39] or alternatively one sample labeled A+U and another sample labeled G+C [22, 40]. Sometimes four complex samples of single nucleotide specific $^{15}\text{N}/^{13}\text{C}$ -labeled RNA [41, 42] or a combination of single and double-nucleotide specific $^{15}\text{N}/^{13}\text{C}$ -labeled RNA were necessary to resolve degeneracies [21, 43].

Since short ssRNA cannot be produced by *in vitro* transcription, chemical synthesis is an alternative to introduce ^{13}C labeling at specific nucleotide positions that greatly improves the structure determination of a protein-RNA complex [44]. Isotope labeling can be very useful in this context, especially if long stretches of the same nucleotide types are present [24, 25, 31]. Residue specific or alternating labeling

obtained by chemical synthesis also proved to be very beneficial for complexes with an RNA stem-loop [39, 40]. However, such RNAs are not commercially available.

2.1.5. Resonance Assignment of Proteins in Complex with RNA

Resonance assignment of proteins in complex with RNA is in principal identical to the procedures used for isolated proteins. Standard triple resonance experiments [45,46] are applied on samples containing uniformly $^{15}\text{N}/^{13}\text{C}$ labeled proteins. The unlabeled RNA component is invisible in those experiments.

Experiments that are not widely used in protein NMR are briefly discussed. Since positively charged arginine and lysine residues often play a crucial role in protein-RNA recognition, their side-chain assignment is necessary for obtaining useful intra- but more importantly inter-molecular distance restraints. Therefore, experiments optimized for the Arg $^{15}\text{N}\epsilon$ and $^{15}\text{N}\eta$ with adjusted offsets, delay lengths and sometimes flip-back pulses and ^{15}N selective pulses have been used for some complexes [47, 48]. For example a 2D Arg- $^{15}\text{N}\epsilon\eta$ -edited HSQC-NOESY [47], a 2D Arg-(H)C(C)TOCSY- $\text{N}\epsilon\text{H}\epsilon$ and Arg-H(CC)TOCSY- $\text{N}\epsilon\text{H}\epsilon$ for correlating arginine $\text{H}\epsilon$ to side chain carbons and protons [49] and Arg-H η ($\text{N}\eta\text{C}\zeta$ $\text{N}\epsilon$) $\text{H}\epsilon$ correlating arginine $\text{H}\eta$ and $\text{H}\epsilon$ [48] have been successfully used to assign arginine and lysine side-chains in protein-RNA complexes.

With increasing molecular size, fast relaxation resulting in line broadening becomes a major obstacle [50]. In addition, the complexity of the spectra increases with increasing number of resonances. Deuteration together with TROSY experiments have to be used for larger complexes as discussed in Chapters 4 and 12. Higher temperature also decreases line broadening, e.g. three protein-RNA complexes with a size of ~28 kDa were studied at higher temperatures such as 40, 45 and 50 °C which was sufficient for complete resonance assignment and structure determination without the need of deuteration or TROSY triple resonance experiments [21, 23, 24]. For the protein backbone assignment of the 28 kDa protein-RNA complex consisting of an RNA stem-loop and three Zinc-fingers [41], deuteration combined with TROSY versions of HNCA, HN(CA)CO and HNCO experiments [51, 52] were used in combination with a 3D ^{15}N -NOESY-HSQC.

2.1.6. Resonance Assignment of RNA in Small to Medium Size Complexes

The assignment of an RNA bound to a protein follows similar strategies as for assigning a free RNA. We refer readers to Chapters 11 and 12. The main difficulty during the assignment procedure of RNA is the small chemical shift dispersion found in the RNA sugars resonances (Fig. 4A). Fortunately, the part of highest interest, namely the RNA at the binding interface, often experiences large chemical shift changes, leading to a larger dispersion of the RNA resonances (Fig. 4B) that helps for the assignment process. Furthermore, RNA nucleotides distant to the binding site typically retain the same conformation as in the free form and therefore do not experience chemical shift changes upon binding. For those nucleotides the resonance assignment of the free RNA can then be transferred to the bound RNA.

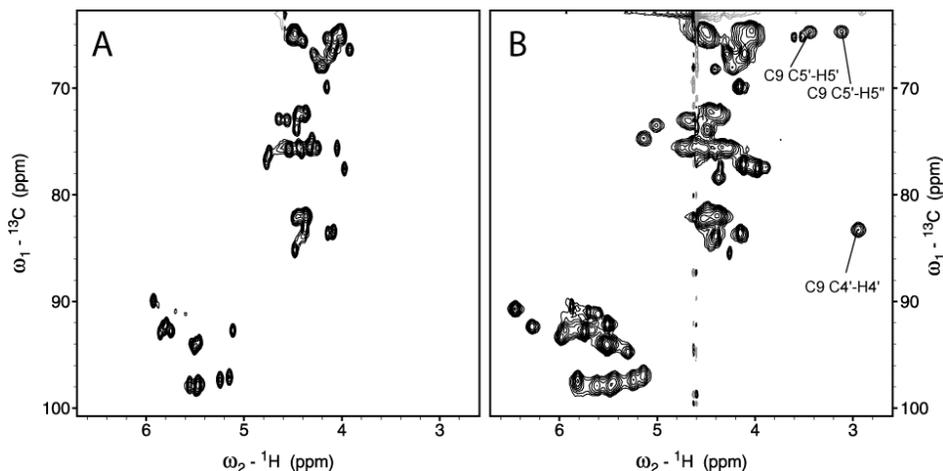


Figure 4. ¹³C-HSQC spectra of a ¹³C/¹⁵N A+C labeled hcnA mRNA (20nt) [23] in its free state (A) and bound to the protein RsmE (B). The resonances of C9 experience a large upfield chemical shift because they are located above the following guanine base G10.

Nucleotide specific ¹³C/¹⁵N labeling simplifies NOESY spectra, but also reduces overlap in crowded regions of the ¹³C-HSQC such as C6/H6 and C1'/H1' correlations, e.g. C6-H6 signals of cytosines and uracils are separated when two samples with A+C and G+U specific labeling are used [23, 39, 40]. Labeling with one single nucleotide type at a time has also been used for the assignment of a 28 kDa protein-RNA complex [41]. The drawback is that four different samples need to be prepared and 3D and 4D spectra need to be recorded for all those samples.

For larger complexes, methods to overcome line broadening and an increased number of resonances have been also developed for RNA such as specific ²H labeling schemes and segmental isotope labeling as discussed in Chapters 2, 11 and 12. The usefulness of such labeling schemes has been demonstrated for two protein-RNA complexes with molecular weights of 31 and 39 kDa [53, 54].

2.2. Defining the Protein-RNA Interface

In order to define the intermolecular protein-RNA recognition interface, a sufficient number of intermolecular NOEs are required. These include sugar-aliphatic, aromatic-aliphatic, aromatic-aromatic and NH-aliphatic/aromatic NOEs. In principle highly-resolved 2D NOESY spectra measured in D₂O and H₂O contain all these NOEs but often peaks cannot be assigned unambiguously due to severe overlap. Therefore, specific NMR experiments have been developed that use editing and filtering elements. To unambiguously identify intermolecular NOEs, samples with opposite labeling of the two components are used (either the RNA unlabeled and the protein ¹⁵N/¹³C or vice versa) together with 2D and 3D NOESY spectra that select, for example, protons attached to ¹³C in one dimension and to ¹²C in another dimension. Selecting protons covalently attached to NMR active nuclei (e.g. ¹H[¹³C] or ¹H[¹⁵N]) is called editing, whereas suppressing such resonances and thus indirectly selecting ¹H[¹²C/¹⁴N] represents filtering [55, 56].

2.2.1. 2D and 3D Filtered/Edited NOESY Experiments

A variety of 2D and 3D experiments based on filtering and editing elements have been used for the structure determination of protein-RNA complexes. Filtered and edited NOESY spectra are generally recorded in D₂O in order to improve sensitivity due to the decrease in noise and the ability to use a higher receiver gain. In addition, signals around 4.7 ppm such as RNA sugar protons are not obscured by the water signal or artifacts from water suppression techniques. A 2D ¹³C 1F-filtered 2F-filtered NOESY is generally used to derive NOEs within the unlabeled smaller molecule (peptide or RNA) in the presence of the ¹³C labeled larger molecule. Depending on line broadening, sensitivity and requirements on the filtering efficiency, either only one purge element in F1 and one in F2 is applied using tuned adiabatic pulses [57] or double purge elements according to Peterson et al. [58] are used.

To detect intermolecular NOEs, a variety of 2D and 3D filtered-edited NOESY experiments have been developed. The 3D ¹³C F1-filtered F3-edited NOESY-HSQC [59] detects the ¹H[¹³C] resonance in the direct dimension. This way unambiguous identification of the ¹H[¹³C] resonances is achieved by the high resolution of F3 together with the ¹³C chemical shift in F2. However, for the identification of the ¹H[¹²C] resonance in F1 only low resolution is available. The 3D ¹³C F1-edited F3-filtered HMQC-NOESY [60] detects the ¹H[¹²C] resonances in the well-resolved direct dimension (Fig. 5A). This way, even intermolecular NOEs between three nucleotides and a methyl group could be extracted as shown with the spectrum of the RsmE-hcnA protein-RNA complex [23]. The importance of those restraints is illustrated in the 3D complex structure (Fig. 6A). The identity of the ¹H of the labeled part can usually be determined via the ¹H-¹³C correlations in F1 and F2 even if these dimensions have lower resolution (Fig. 5B). The observed intermolecular NOEs between an H1' and four protein side chains are illustrated by the distances in the 3D structure (Fig. 6B).

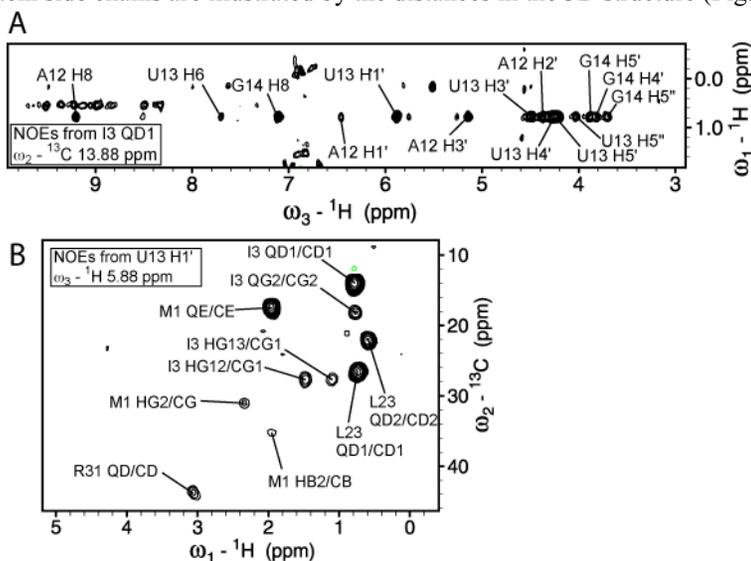


Figure 5. 3D ¹³C F1-edited F3-filtered HMQC-NOESY [60] measured with a sample of ¹³C/¹⁵N-labeled RsmE in complex with an unlabeled 20 nt hcnA mRNA [23]. (A) The F1-F3 plane of the 3D spectrum at the ¹³C chemical shift of the CD1 of Ile3 showing intermolecular NOEs between QD1 of Ile3 and RNA resonances in the direct dimension. (B) The F1-F2 plane of the 3D spectrum at the ¹H chemical shift of U13 H1' showing intermolecular NOEs.

The 2D ^{13}C F1-filtered F2-edited NOESY [58] based on double tuned purge elements is used largely for assignment and NOE extraction of nucleotide specific labeled RNA. Note that the filter and editing elements are rather long (for ^{13}C : tuned purge 2.5–4.0 ms, double tuned purge 6–8.0 ms; for ^{15}N : tuned purge ~5.4 ms) and can therefore lead to signal loss due to relaxation effects. Some intermolecular NOEs are only visible in the very sensitive 2D NOESY in D_2O but not in the 2D or 3D filtered and edited NOESY.

^{15}N editing and filtering can be used in H_2O combined with or without ^{13}C editing and filtering, which can be useful to detect intermolecular NOEs or intra-RNA NOEs between $^1\text{H}[^{15}\text{N}]$ and $^1\text{H}[^{12}\text{C}/^{14}\text{N}]$. 2D $^{13}\text{C}/^{15}\text{N}$ 1F-filtered 2F-filtered NOESY measured in H_2O is used to assign HN resonances of unlabeled peptides in the presence of $^{13}\text{C}/^{15}\text{N}$ labeled RNA. For example a 3D $^{13}\text{C}/^{15}\text{N}$ -filtered ^{15}N -edited NOESY-HSQC was used to solve the P22 N peptide-box B complex in order to obtain 15 amide-RNA NOEs out of 81 intermolecular NOEs involving arginine $\text{N}\epsilon\text{H}\epsilon$ and $\text{N}\eta\text{H}_2$ [61].

2.2.2. Hydrogen Bond Restraints

Amide protons that are protected from H to D exchange are typically involved in hydrogen bonds. Similarly, the presence of imino signals indicates a hydrogen bond (H-bond). However, the H-bond acceptor cannot be derived from such exchange data. Intermolecular H-bonds typically lead to the largest NH and C' chemical shift deviations between the free and bound protein state. This is illustrated using the RsmE-hcnA mRNA complex in Fig. 7. The two largest NH chemical shift deviations (I3 and T5 amides) are forming intermolecular H-bonds to two adenines (A8 and A12, respectively). Typically the amide ^1H chemical shifts are downfield shifted upon the formation of H-bonds, most pronounced for H-bonds to RNA bases. When chemical shift deviations of NH and C' are observed, these can then be used as support for an intermolecular H-bond. Note that the H-bond partner is not identified this way. Typically, potential H-bonding partners are derived from initial structure calculations without any H-bond constraints. An iterative, step-wise introduction of such H-bonds into the refinement of the structure is typically used and results in better convergence of the calculations.

Recently a variety of experiments that directly identify unambiguously H-bond partners have been developed [62] that are based on small scalar couplings across H-bonds which are in the range of 5–11 Hz ($^{\text{h}2}\text{J}_{\text{NN}}$) for $\text{N}-\text{H}\cdots\text{N}$ bonds and only -0.1 to -0.9 Hz for $\text{N}-\text{H}\cdots\text{O}=\text{C}$ bonds. However, intermolecular H-bond restraints derived from such experiments have not been used so far for structure calculation of protein-RNA complexes but it would be very valuable to have direct experimental evidence for such H-bonds. The first direct NMR detection of intermolecular H-bonds between an RNA and a protein has been observed between an arginine side chain $\text{N}\eta\text{H}\eta$ and a guanine N7 atom [63] of the HTLV-1 Rex peptide and a 33-mer RNA aptamer [48] after the structure had been already published.

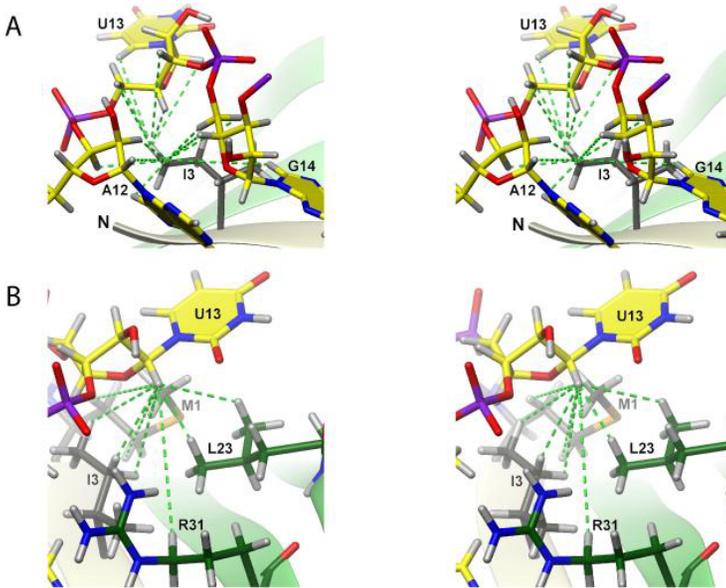


Figure 6. Illustration of the observed intermolecular NOEs measured with RsmE in complex with a 20 nt hcnA mRNA [23] as displayed in Figure 5. (A) 3D structure of the RsmE/hcnA complex illustrating the observed NOEs to QD1 of Ile3. (B) 3D structure of the RsmE/hcnA complex illustrating the observed NOEs to U13 H1'. Figures were generated with molmol [64]

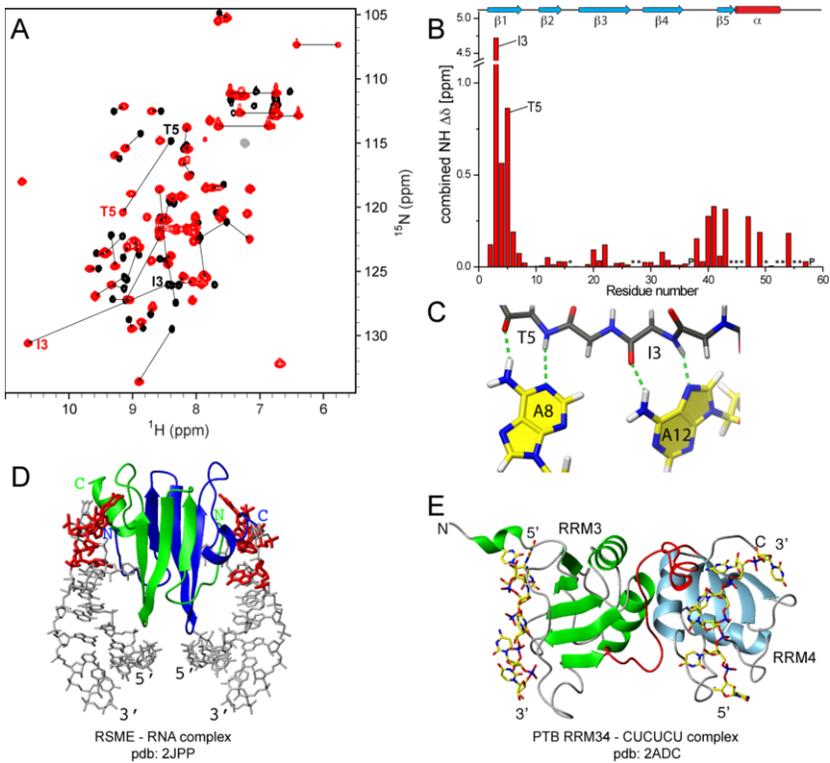


Figure 7. Illustration of chemical shift deviations caused by intermolecular H-bonds in the complex of the RsmE protein bound to the hcnA mRNA [23] and overall structures of the RsmE-hcnA and the PTB RRM34 complexes. (A) Overlay of ^{15}N -HSQC of the free RsmE protein (black) and in complex with the RNA (red). (B) Combined chemical shift perturbation of amide ^1H and ^{15}N upon complex formation. The chemical shift deviations were combined using $(\Delta\delta = [\Delta\delta_{\text{HN}}^2 + (\Delta\delta_{\text{N}}/R_{\text{scale}})^2]^{1/2}$, $R_{\text{scale}} = \gamma_{\text{H}}/\gamma_{\text{N}} = 9.85$). An asterisk denotes residues not assigned in free RsmE due to the absence or very low intensity of signals. A “P” denotes prolines. (C) Three dimensional structure of the intermolecular H-bonds between Ile3, Thr5 and the RNA bases A8 and A12. (D) Structure of the RsmE protein bound to the hcnA mRNA. The protein dimer is colored blue and green. The nucleotides of the Shine-Dalgarno sequence are colored red. (E) Structure of PTB RRM34 in complex with two pyrimidine tracts. RRM3 and RRM4 are colored green and blue, respectively. The interdomain linker is colored red. The RNA molecules are colored yellow. 3D figures were generated with molmol [64].

2.3. Long Range Restraints – Global Orientation

For most protein-RNA complexes, the structure determination has been based primarily on intermolecular NOEs to derive intermolecular distance restraints. However, several protein-RNA complexes have poor specificities and affinities (in the higher micromolar range). Dynamic interfaces and chemical exchange can hamper the detection of enough intermolecular NOEs to properly define the orientation of both macromolecules relative to each other. Furthermore, elongated structures like nucleic acid stems in protein-RNA complexes have poor global precision and accuracy due to the short-range nature of the NOE restraint ($< 6 \text{ \AA}$). With increasing molecular size, spectral crowding and increased relaxation can prevent collection of enough NOE restraints to determine the structure. The introduction of long-range orientational restraints such as residual dipolar couplings (RDCs) or long-range translational restraints like paramagnetic relaxation enhancement (PRE) has enabled the study of large protein-RNA complexes by NMR spectroscopy.

When a macromolecule is partially aligned in the magnetic field either by its own magnetic susceptibility or by the use of an external orienting medium, the dipolar coupling between two nuclei is not averaged to zero [65, 66]. This residual dipolar coupling is correlated to the orientation of the inter-nuclear bond vector with respect to the magnetic field or the global alignment tensor of the macromolecule. Therefore, RDCs contain long-range orientational information that is not present in conventional NOE restraints. Several papers and reviews have discussed how to measure RDCs and compared the different approaches to determine the alignment tensor and how to include the measured RDCs into the structure calculation and refinement [66, 67]. RDCs measured in protein-RNA complexes are very powerful in defining the global orientation of several protein or RNA domains in a macromolecular complex and in docking the RNA onto the protein, especially if only few intermolecular NOEs are observed. As an example, by measuring RDCs for both the protein and the RNA in a phospholipid solution, the orientation of a double-stranded RNA with respect to dsRBD3 of the protein Staufen could be established and a structural ensemble generated despite the fact that only ten intermolecular NOEs could be observed [30]. Additionally, orientational information obtained from RDCs can not only refine the global structure of an elongated part like a RNA helical stem, but also contribute to increasing the quality of the local structure [40].

RDCs yielding long-range orientational information do not contain any translational information. Although two domains or two macromolecules in a complex can be oriented with respect to each other, their inter-domain or inter-molecular translational displacements cannot be obtained with RDCs. Paramagnetic relaxation

enhancement (PRE) yields long-range distance information and can be very useful in complementing the long-range orientational information obtained from RDCs [68,69]. The paramagnetic relaxation enhancement of a nucleus by a biochemically introduced paramagnetic center can be directly correlated to a distance between the nucleus and the paramagnetic center. Paramagnetic probes can be introduced either on the protein or on the RNA molecules to measure long-range distances in protein-RNA complexes [70]. For example, the structure of the 38 kDa trimolecular complex between two U1A proteins and the PIE RNA, was solved by introducing single cysteine mutations and attaching a nitroxide spin-label at three different positions on an unlabeled U1A protein [71]. By mixing with another equivalent of ^{15}N labeled U1A protein and unlabeled PIE RNA, Varani and coworkers could measure 30 unambiguous intermolecular long-range distance constraints.

Although not applied yet for protein-RNA complexes, the simultaneous use of both RDC and PRE restraints or in combination with other non-NMR-based methods such as small angle X-ray scattering (SAXS) has been shown to be very powerful and should gain importance in studying systems of increasing molecular weight [69, 72]. SAXS yields information on the overall shape and dimensionality of a complex and can be used optimally in combination with orientational restraints such as residual dipolar couplings or residual chemical shift anisotropy (rCSA) [72].

Utilization of different short and long-range NMR restraints yielding complementary orientational (i.e. RDC or rCSA) or translational (i.e. PRE) information in combination with other techniques such as small angle X-ray scattering, will allow the study of protein-RNA complexes by NMR to enter into new dimensions of space (high molecular weight complexes) and time (transient interactions, dynamics).

2.4. Structure Determination

2.4.1. Experimental Restraints Used in Structure Calculations

NMR spectroscopy provides numerous sources of structural information that can be used for the structure calculation of a macromolecule or a macromolecular complex. These are distance restraints, H-bond and dihedral angle restraints, long-range orientational restraints (RDCs and rCSA) and long range distance restraints (PRE) as described in sections 2.1–2.3.

Distance restraints have been the major source of experimental restraints that are used to solve the NMR structures of protein-RNA complexes. A general approach for deriving NOE-derived distance restraints is to classify peak volumes into categories, such as weak, medium, strong and then define an upper distance limit for each class of NOEs based on known proton-proton distances. Intramolecular H-bonds have been widely used for structure determination of protein-dsRNA complexes for both the protein and the RNA. Intermolecular hydrogen bonds are typically implemented in the later stages of the structure refinement leading to a better convergence of the ensemble [23]. How H-bond restraints are derived was discussed in section 2.2.2. The most commonly used dihedral angle restraints are for the protein and the RNA backbone. Dihedral angle restraints can be directly derived from the measurement of J-couplings, or indirectly by the analysis of COSY or TOCSY spectra for RNA (see Chapters 11 and 12) or derived from chemical shifts [73]. The use long-range restraints such as RDC and PRE data have been already described in section 2.3.

2.4.2. Structure Determination Protocols

Most protein-RNA structure calculations have been performed using simulated annealing (SA) protocols within the software programs of Xplor [74], CNS [75], Xplor-NIH [76, 77], DYANA [78] or CYANA [79].

Two different procedures for calculating the structures of protein-RNA complexes have been described. In one case, NMR restraints are directly used to fold the complex starting from an extended conformation of the complex using a simulated annealing protocol that includes either a potential term (X-PLOR) or a target function (DYANA). Potential terms or target functions are generally very simple and mainly consist of a term that reflects how well a structural model is consistent with experimental NMR data and a term that prevents close contacts between atoms. In the other case, the structure of the complex is calculated by starting directly with both components in an extended conformation and separated by a flexible artificial linker, while in the other case, structure calculation is performed on each component separately and then both structures are docked using intermolecular restraints.

2.4.3. Structure Refinement

Structure refinement of protein-RNA complexes, and more generally protein-nucleic acid complexes is less straightforward than structure refinement of each component alone. Nucleic acids are highly charged molecules and RNA structures are generally not globular, in contrast to most small proteins, or protein domains. Two force fields are most commonly used for protein-RNA complexes, AMBER [80, 81] and CHARMM [82, 83]. These force fields are highly optimized for both proteins and nucleic acids and are therefore highly suitable for the refinement of protein-RNA structures solved by NMR. Generally, the refinement of a protein-RNA complex is performed by a simulated annealing protocol that can be preceded and/or followed by an energy minimization step.

Since protein-RNA interactions are often driven by electrostatic interactions, the electrostatic non-bonded energy term of the structures has to be optimized which is achieved by the use of solvation models during the refinement of NMR structures [84, 85]. Two main approaches to introduce solvent in the refinement procedure have been used: explicit solvent representation, where the structure of the complex is refined in a box of water molecules [86], or implicitly using a continuum solvent model based on the Generalized-Born model [87]. The most accurate but also most computational expensive approach is the use of explicit solvent. The implicit solvent model is mostly used for protein-RNA complexes.

2.4.4. An Example of a Structure Calculation and Validation Protocol

Fig. 8 shows a flowchart describing the procedure for structure calculation, refinement, and validation used in our laboratory to solve the structures of 11 protein-RNA complexes [15, 18, 23, 24, 31, 39, 40]. Intra-protein distance restraints were generated with the software ATNOS/CANDID [79, 88] using NOESY spectra and a list of protein chemical shifts. In addition, a list of intra-protein H-bond restraints was often used. Intra-RNA and intermolecular distance restraints were generated by manual NOEs assignment. These intra-protein, intra-RNA and intermolecular distance restraints were then combined and used together with H-bond and torsion angle restraints to generate preliminary structures of the complex using the program CYANA

[88]. Typically, between 200 and 500 structures were calculated and the 20 to 50 structures with lowest target functions were analyzed in terms of convergence and NOE violations. This analysis was used to refine the distance restraints, including the unambiguous assignment of additional NOEs that were previously ambiguous and therefore not included, and the modification of upper bound limits, especially in the cases of overlapping NOE cross-peaks whose intensities correspond to the contributions of more than one proton-proton distance. New structure calculations were then performed using these new sets of distance restraints until a final ensemble of solutions was satisfactory in terms of structure precision and NOE violations. This final ensemble of structures was then subjected to a structural refinement procedure.

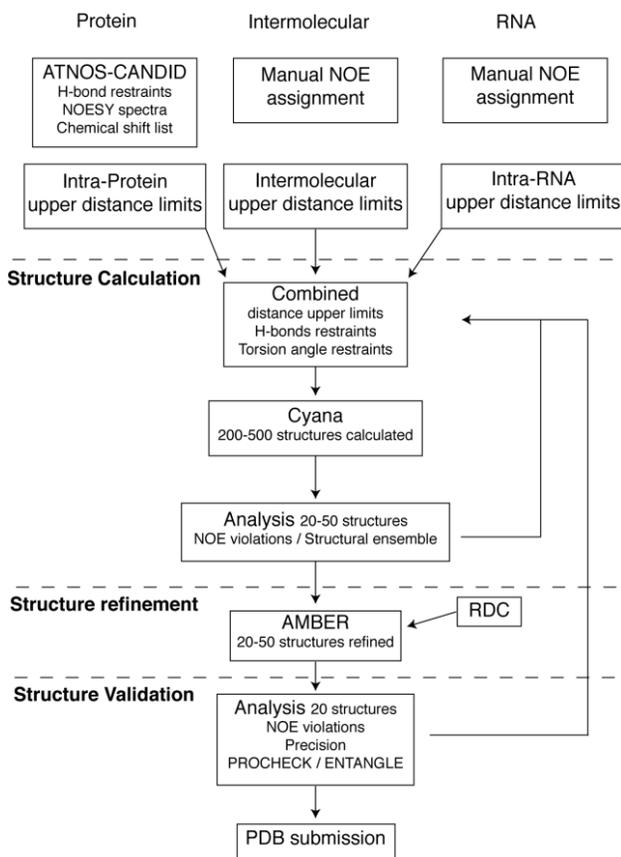


Figure 8. Example of the structure calculation protocol used to determine structures of protein-RNA complexes in our laboratory.

Structure refinement was performed using the AMBER software [89]. Inputs for structure refinements consisted of the 20 to 50 structures derived from CYANA, the distance, hydrogen-bond, and torsion angle restraints and, when available, the RDC restraints. Structure refinement was performed using the AMBER ff99 or ff94 force fields [80, 81] in combination with a generalized Born (GB) solvation model [87]. The refinement procedure consisted of a simulated annealing protocol that was optimized for nucleic acids [90] after a short energy minimization of the starting structure.

Following the refinement procedure, structures were analyzed in terms of energy, NOE violations, and structure precision. As for structure calculations, analysis of the refined structures, especially the NOE violations, could be used to refine the distance restraints and restart another cycle of structure calculations and refinement.

2.5. Validation of Protein-RNA Complexes

Validation of protein-RNA NMR structures includes three main aspects: the validation of the structural ensemble against experimental restraints, the validation of the structural ensemble based on geometrical and structural characteristics, and the confirmation of the intermolecular contacts based on biochemical or biophysical experiments.

2.5.1. Validation of Structures Against Experimental Restraints

In structure determination of protein-RNA complexes, three types of NMR-derived restraints have mainly been used, namely distance, torsion angle and orientational restraints. The most common approach to validate distance and dihedral angle restraints consists of generating a list of violations and analyzing the number and magnitude of these violations. Most NMR structure calculations and refinement programs (XPLOR, CYANA, AMBER) include validation routines that create a list of distance and dihedral angle violations for a detailed analysis. Validation of orientational restraints is generally achieved by determining an RMSD between the experimental and the back-calculated RDCs.

An additional, independent way of validating structures of protein-RNA complexes resides in the agreement between the experimental chemical shifts and those back-calculated from the ensemble of structures. Chemical shifts are highly sensitive to the local electronic environment of nuclei, and are among the most accurate quantities that can be measured by NMR spectroscopy. For proteins, programs have been developed that predict nitrogen, carbon and proton chemical shifts from coordinate files [91, 92]. However, no programs for RNA chemical shift prediction have been developed so far. Nevertheless, it was shown that proton chemical shifts of RNAs could be predicted with a good accuracy and precision from a coordinate file [93]. Therefore, it should be possible to use chemical shift predictions as a tool for the validation of protein-RNA structures. Manual inspection of unusual chemical shifts in protein-RNA complexes sometimes explain structural features (Fig. 4 and 7).

2.5.2. Validation of Structures Based on Geometrical and Structural Characteristics

Geometrical and structural properties are important criteria in the validation of structures. The quality of NMR structures in terms of geometrical and structural properties is generally driven by the force field used during the structure calculation and refinement procedures. Many structure validation software packages that analyze a structural ensemble and report on the quality of the geometrical and structural properties are available. The most widely used packages are PROCHECK_NMR [118] and WHAT IF [119] for proteins as well as the module NUCHECK that is part of the Nucleic-acids Database [120] and MOLPROBITY for nucleic acids [121]. In addition, a validation tool is available on the Protein Data Bank website (<http://deposit.rcsb.org/>) that checks the quality of a coordinate file using all the software mentioned above and

presents a summary of structure quality as well as the reports from PROCHECK, NUCHECK, and MOLPROBITY.

2.5.3. Confirmation and Quantification of Intermolecular Contacts

The goal of solving macromolecular complexes is to understand the molecular basis that governs the specificity of the complex formation. In protein-RNA complexes, this specificity is largely governed by H-bonds between the protein and the nucleic acid bases. Additionally, protein-RNA interactions are often stabilized by electrostatic interactions and by stacking interactions involving aromatic amino acids and RNA bases.

In order to confirm and quantify the intermolecular contacts observed in the structure, site-directed mutagenesis combined with binding assays are often performed by mutating specific amino acids or nucleotides that are involved in intermolecular contacts. An interesting example that illustrates the power of combining site-directed mutagenesis with NMR and Surface Plasmon Resonance binding assays involved the complex between the Fox-1 protein and its RNA target [15]. In the NMR structure, six nucleotides are specifically recognized by the protein through an extensive network of intermolecular and intra-RNA hydrogen-bonds. The contribution of each intermolecular and intra-RNA H-bonds was evaluated using mutant RNAs. This analysis showed that the loss of free binding energy in mutant RNAs is directly correlated to the number of H-bonds that are lost in the complex based on the NMR structure.

2.5.4. Structure-Function Relationship of Protein-RNA Complexes

Finally, because the fundamental aim of solving structures of protein-RNA complexes is to provide the molecular basis with which to understand their biological functions, mutations that affect the complex formation can be tested using functional assays. The structure of the third double-stranded RNA binding domain of the protein Staufen in complex with RNA has been used to design a quintuple mutant that disrupts RNA binding [30]. This mutant was then tested by *in vivo* mRNA localization assays in order to demonstrate that the RNA binding properties of the protein Staufen are crucial for the proper localization of specific mRNAs. Based on the structure of the viral protein NC in complex with its RNA target, RNA mutants that disrupt the binding of the protein were designed and tested by *in vivo* reverse transcriptase assays to assess the effect of the NC-RNA interaction on the virus infectivity [54]. Finally, based on the structure of the protein RsmE in complex with its target RNA, specific mutants that disrupt the interaction were designed and *in vivo* translation assays of these mutants showed that the RNA binding properties of RsmE were crucial for its function in translation repression [23].

3. Dynamics of Protein-RNA Complexes

Detailed investigations of molecular motions in protein-RNA recognition has been motivated by the question of how dynamics and conformational changes within the interacting partners influence the binding process and specificity in those interactions. Despite the impressive number of structural studies on protein-RNA complexes, relatively few studies have been dedicated to a quantitative analysis of molecular

motions of protein and RNA in these systems. Relaxation rate measurements by NMR provide residue specific information on dynamics in both protein and RNA over a range of different time-scales [94-96]. As discussed in chapter 20, fast motions on the pico-second (ps) to nano-second (ns) timescales are typically characterized by the measurement of longitudinal (R_1), transverse (R_2) or rotating frame ($R_{1\rho}$) and nuclear Overhauser effect (NOE) relaxation rates for ^{15}N and ^{13}C nuclei. Relaxation data are mapped to molecular dynamics through the well established model-free formalism [97, 98]. Slow motions on the micro-second (μs) to milli-second (ms) time-scales can be analyzed quantitatively by Carr-Purcell-Meiboom-Gill (R_2) and spin-lock ($R_{1\rho}$) relaxation dispersion experiments [99, 100] (see Chapter 10). Along with relaxation rates, measurements of RDCs, extend the range of accessible motional time-scales to the sub-micro/milli second regime thus allowing a wide spectrum of dynamical processes to be examined [66, 101].

3.1. Changes in Dynamics Upon Protein-RNA Complex Formation

The relatively few reports of dynamics studies have revealed significant implications for molecular mobility in governing the mechanism, specificity and thermodynamical aspects of the interaction thereby underscoring the importance of complementing structural studies with investigations of molecular dynamics.

Comparison of molecular motions in the free and RNA bound states of the human U1A protein has provided significant insights into the role of dynamics in the recognition process [102]. U1A protein binds the 3'UTR-RNA with high specificity and the interaction involves an induced-fit mechanism. Analysis of backbone ^{15}N relaxation rates in the free U1A indicates that several residues at the RNA binding surface undergo slow conformational fluctuations in the μs -ms time-scale. On binding RNA, these motions are significantly reduced. In addition, analysis of ^2H relaxation in side-chain methyl groups indicates loss of μs -ms motions upon interaction with RNA. ^{13}C relaxation studies of the 3'UTR of the mRNA coding for U1A in the free state indicates the presence of fast and slow motions particularly in the loop forming the binding surface on the RNA [103]. Interaction with the U1A protein quenches motions in the binding loop making it more ordered [104]. Conformational flexibility of protein and RNA in the free-state allows various conformations to be sampled in order to obtain an optimal conformation at the binding interface in the complex. Reduced molecular mobility results from the formation of a tightly packed interface where multiple intermolecular interactions ensure high specificity. This high specificity accompanied by a loss of flexibility is achieved at a large entropic cost that is compensated by an increase in flexibility at other sites located away from the binding interface. Another example of binding by an induced fit mechanism is the interaction of the two domain protein L11 with 23S-rRNA [105]. The free L11 protein shows considerable flexibility in the ps-ns timescales for the residues in the RNA binding region that becomes rigid on interaction with RNA.

The interaction of the 64 kDa subunit of the Cstf-64 protein with RNA occurs with a rather diffuse specificity since Cstf-64 does not recognize a specific RNA sequence or consensus but binds many GU rich RNA sequences. The dynamics profile determined by ^{15}N relaxation studies is also markedly different from those observed in U1A and L11 which show highly specific interaction with RNA [106]. In the free state the protein is mostly uniformly rigid in the ps-ns and μs -ms time scales. On binding to RNA; however, there is an overall increase in fast and slow time-scale motions,

particularly at the RNA binding surface and a high degree of mobility is retained at the binding interface in the complex.

The interaction of the VTS1p-SAM domain with SRE-RNA is another example of binding occurring without any conformational changes in the protein and RNA. The interaction is mostly a shape specific recognition and combines elements of sequence-specificity and of non sequence-specificity [40]. VTS1p-SAM domain recognizes tetra- or penta-loop RNA with a general consensus sequence of the form XNGY(N) for the loop, where N is any nucleotide and X and Y form a Watson-Crick base pair. Only the central G nucleotide and the shape of the RNA fold induced by the base-pairing are specifically recognized. ^{15}N relaxation studies of the backbone dynamics indicate that in its free state, the VTS1p-SAM domain is mostly rigid with no significant motions in the fast and slow time-scales [107]. On interaction with the CUGGC loop of the SRE-RNA, the residues of the binding surface which are associated with the specific recognition of the central G nucleotide in the RNA loop become more rigid. In contrast, the residues of the binding surface which interact in a non-specific manner with other nucleotides in the RNA loop show increased flexibility in the bound state. ^{13}C relaxation studies of dynamics in the free RNA show that the nucleotides which form a base-pair within the loop, have fast motion amplitudes very similar to that observed in the stem region while the other nucleotides show considerable flexibility in the fast and slow time-scales [108]. The central G nucleotide which is specifically recognized by the protein does not undergo μs -ms motions but has fast motion amplitude that is intermediate between that of the base-paired nucleotides and the two highly flexible loop residues. Binding to the VTS1p-SAM domain reduces the flexibility of the nucleotide base of the central G nucleotide which makes maximum contacts with the protein through the base moiety. The base-paired nucleotides which contact the protein through the sugar-phosphate backbone, show reduced amplitudes of fast motion at the anomeric sites and a slight increase in flexibility at the aromatic sites. The two loop nucleotides with high flexibility in the free-state have much more restricted motions in the bound state since they also make contacts with the protein. The dynamics changes in the protein and RNA at the binding interface clearly indicates that sequence-specificity of recognition is accompanied by increased rigidity whereas the parts interacting in a non sequence-specific manner attain increased flexibility on binding.

3.2. Insights and Future Directions

The nature of molecular motions, particularly at the binding interface differs depending on the mechanism of protein-RNA interaction. The free states of protein and RNA, which bind by an induced fit method, are characterized by highly mobile binding surfaces which become ordered on complex formation. Increased mobility in the free-states allows different conformations to be sampled so that an optimal arrangement of the binding surfaces can be achieved so as to maximize the interface contacts. In shape specific recognition on the other hand, the free-states have relatively limited mobility, thus providing a conformationally pre-ordered surface for binding.

Regions of the binding interface associated with high specificity of interaction become more rigid on binding while regions with non-sequence-specificity attain increased flexibility. This is perhaps functionally relevant since non-sequence-specific interaction requires nucleotides of different sizes and hydrogen bonding strengths to be accommodated at the binding surface of the protein.

Interestingly, complex formation sometimes results in increased mobility at regions located away from the binding interface. This seems to have implications for the thermodynamics of the interaction. Highly specific interactions and the accompanying rigidity result in a high entropic cost in complex formation. Flexibility gain at regions away from the binding interface helps to offset the entropic cost resulting from restricted mobility at the binding site.

Almost all of the protein dynamics studies have focused on dynamics at the amide ^{15}N sites along the backbone. More work needs to be undertaken to examine dynamics in protein side-chains since much of the intermolecular contacts are established through the side-chains. There have been significant advances in the experimental characterization of slow motions in proteins and RNA involving a variety of nuclei as probes [100, 109-111]. Application of these methods to protein-RNA complexes can provide far more insights into the complex conformational dynamics at protein-RNA interfaces. In addition, measurements of RDCs in protein-RNA complexes will extend the dynamical time scales that can be probed, by including the intermediate regime not readily accessed by the techniques that probe fast and slow motions. RDC measurements offer the possibility of examining large scale motions such as domain orientation changes in multi-domain proteins and RNAs with more complex structures involving several helical stems and internal loops [112].

4. NMR Structures of Protein-RNA Complexes: What Did We Learn From Them

The protein-RNA structures solved by NMR provide significant structural insights to understand important biological processes at the molecular level. Many of these structures were fundamental in deciphering the role of these interactions and to guide further studies and characterizations of biological functions. Furthermore, many of these interactions play an important role in disease related processes and protein-RNA structures provided templates for structure-based drug design. In the following are summarized some examples where NMR structures of protein-RNA complexes gave important insights into biological functions.

In retroviruses, such as HIV-1, the NC protein plays a critical role in viral replication and participates in genome recognition and encapsidation. The NC protein specifically binds a portion of the viral RNA, called the Ψ -site, which undergoes a monomer-dimer transition important for encapsidation. The structures of NC proteins in complex with RNAs from different retroviruses [29, 53, 54, 113, 114] could demonstrate that the NC protein binds specifically single-stranded portions of the RNA that are accessible only in the dimeric conformation. A mechanism was therefore proposed where genome packaging is regulated by a structural RNA switch, in which NC binding sites are sequestered by base-pairing in the monomeric form of the RNA and become exposed upon dimerization to promote the encapsidation of the RNA.

In bacteria, translation repression can occur when the Ribosome binding site (RBS) of the mRNA is masked by bound proteins. The structure of RsmE in complex (Fig. 7D) with an RNA containing the RBS showed that the RNA adopts a stem-loop structure and that the RBS is sequestered by the protein, therefore preventing ribosome binding to the mRNA [23].

NMR studies of protein-RNA complexes also provided insights into the mechanisms that control eukaryotic post-transcriptional gene regulation. For example, PTB is a general splicing repressor that contains four RNA binding domains of the

RRM (RNA recognition motif) type and recognizes specifically RNA sequences rich in pyrimidines. The structures of each RRM of PTB in complex with a CUCUCU RNA were solved [24, 25]. The structure of PTB RRM34 in complex with RNA showed that two molecules of RNA are bound and are located on opposite sides of the structure (Fig. 7E) indicating that PTB RRM34 can bind a single RNA sequence only if two pyrimidine tracts are separated by a linker of at least 15 nucleotides. This unprecedented structural feature suggested that PTB could repress alternative splicing by looping out specific exons. This model was recently confirmed using a combination of fluorescence resonance energy transfer (FRET) and NMR [115].

Finally, NMR studies of protein-RNA complexes also provided insights into the understanding of large molecular assemblies in eukaryotes. For example, NMR studies of the protein U1A in complex with its target RNA led to a model for polyadenylation repression. U1A is a spliceosomal protein that prevents the formation of the poly(A) tail through binding an RNA sequence called the polyadenylation inhibition element (PIE). Two molecules of U1A bind cooperatively the PIE element via their N-terminal RNA Recognition Motif (RRM). Inhibition of polyadenylation is achieved by a repressive interaction between U1A and the poly(A) polymerase (PAP). Residues of U1A important for PAP binding are adjacent to the N-terminal RRM domain. Two NMR structures of U1A N-terminal RRM in complex with the PIE RNA led to a model for U1A-PAP complex formation [3, 71]. The structure of the trimolecular complex between the PIE RNA and two molecules of U1A showed that upon RNA binding, U1A homodimerizes through an α -helix located at the C-terminus of the RRM, bringing the PAP interacting regions in close proximity and on the same side of the structure [71]. The conformation of the protein dimer in the RNA-protein complex structure is optimal to bind PAP and therefore to repress polyadenylation.

5. Conclusions

In this review we describe the achievements in the field of NMR structure determination of protein-RNA complexes over the last fifteen years. Although, NMR is certainly today a mature method to solve the structures of most protein-RNA complexes below 20 kDa, every structure determination of such a complex is still a challenge in itself. If this field has benefited tremendously from the technological development in the field of NMR like high sensitivity, RDCs, fast computing, or possibility semi-automated protein structure determination, solving a protein-RNA complex by NMR still requires high manual intervention and to master the spectroscopy of both the protein and the RNA components. However, it is very clear that NMR spectroscopy has now demonstrated to be a very competitive method for investigating the structures of protein-RNA complexes. Not only more protein-RNA complex structures will be determined in the near future, but NMR will be particularly useful to investigate how several RNA binding proteins assemble or compete for binding RNA. Since protein-RNA interactions are at the heart of every molecular mechanism controlling post-transcriptional gene expression, we have in front of us, as biomolecular NMR spectroscopists, an infinite and very attractive field of study for the decades to come.

Acknowledgements

We would like to thank F. Oberstrass for his input in the initial stage of this review, S. Auweter for providing spectra of the Fox-1-RNA complex, F. Loughlin for carefully reading the manuscript and all members of the laboratory for fruitful discussions. This investigation was supported by grants from the Swiss National Science Foundation Structural Biology National Center of Competence in Research (SNF-NCCR), the SNF grant (Nr. 3100A0-118118) to F.H.T.A and the Indo Swiss Joint Research Programme (ISJRP) to F.H.T.A and to S.R.

References

- [1] J.D. Puglisi, L. Chen, S. Blanchard, A.D. Frankel, Solution structure of a bovine immunodeficiency virus Tat-TAR peptide-RNA complex, *Science* **270** (1995), 1200–1203.
- [2] X. Ye, R.A. Kumar, D.J. Patel, Molecular recognition in the bovine immunodeficiency virus Tat peptide-TAR RNA complex, *Chem. Biol.* **2** (1995), 827–840.
- [3] F.H. Allain, C.C. Gubser, P.W. Howe, K. Nagai, D. Neuhaus, G. Varani, Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation, *Nature* **380** (1996), 646–650.
- [4] D.J. Galas, A. Schmitz, DNase footprinting: a simple method for the detection of protein-DNA binding specificity, *Nucleic Acids Res.* **5** (1978), 3157–3170.
- [5] M.M. Garner, A. Revzin, A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system, *Nucleic Acids Res.* **9** (1981), 3047–3060.
- [6] M. Fried, D.M. Crothers, Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis, *Nucleic Acids Res.* **9** (1981), 6505–6525.
- [7] S. Douthwaite, A. Christensen, R.A. Garrett, Binding site of ribosomal proteins on prokaryotic 5S ribonucleic acids: a study with ribonucleases, *Biochemistry* **21** (1982), 2313–2320.
- [8] M. Stoldt, J. Wohnert, O. Ohlenschlager, M. Gorlach, L.R. Brown, The NMR structure of the 5S rRNA E-domain-protein L25 complex shows preformed and induced recognition, *EMBO J.* **18** (1999), 6508–6521.
- [9] K.S. Cook, G.J. Fisk, J. Hauber, N. Usman, T.J. Daly, J.R. Rusche, Characterization of HIV-1 REV protein: binding stoichiometry and minimal RNA substrate, *Nucleic Acids Res.* **19** (1991), 1577–1583.
- [10] J.L. Battiste, H. Mao, N.S. Rao, R. Tan, D.R. Muhandiram, L.E. Kay, A.D. Frankel, J.R. Williamson, Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex, *Science* **273** (1996), 1547–1551.
- [11] J. Ule, K.B. Jensen, M. Ruggiu, A. Mele, A. Ule, R.B. Darnell, CLIP identifies Nova-regulated RNA networks in the brain, *Science* **302** (2003), 1212–1215.
- [12] J. Ule, K. Jensen, A. Mele, R.B. Darnell, CLIP: a method for identifying protein-RNA interaction sites in living cells, *Methods* **37** (2005), 376–386.
- [13] G.W. Yeo, N.G. Coufal, T.Y. Liang, G.E. Peng, X.D. Fu, F.H. Gage, An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells, *Nat. Struct. Mol. Biol.* **16** (2009), 130–137.
- [14] Y. Jin, H. Suzuki, S. Maegawa, H. Endo, S. Sugano, K. Hashimoto, K. Yasuda, K. Inoue, A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG, *EMBO J.* **22** (2003), 905–912.
- [15] S.D. Auweter, R. Fasan, L. Reymond, J.G. Underwood, D.L. Black, S. Pitsch, F.H. Allain, Molecular basis of RNA recognition by the human alternative splicing factor Fox-1, *EMBO J.* **25** (2006), 163–173.
- [16] A.D. Ellington, J.W. Szostak, In vitro selection of RNA molecules that bind specific ligands, *Nature* **346** (1990), 818–822.
- [17] C. Tuerk, L. Gold, Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase, *Science* **249** (1990), 505–510.
- [18] Y. Hargous, G.M. Hautbergue, A.M. Tintaru, L. Skrisovska, A.P. Golovanov, J. Stevenin, L.Y. Lian, S.A. Wilson, F.H. Allain, Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8, *EMBO J.* **25** (2006), 5126–5137.

- [19] B.P. Hudson, M.A. Martinez-Yamout, H.J. Dyson, P.E. Wright, Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d, *Nat. Struct. Mol. Biol.* **11** (2004), 257–264.
- [20] L. Ghisolfi-Nieto, G. Joseph, F. Puvion-Dutilleul, F. Amalric, P. Bouvet, Nucleolin is a sequence-specific RNA-binding protein: characterization of targets on pre-ribosomal RNA, *J. Mol. Biol.* **260** (1996), 34–53.
- [21] F.H. Allain, P. Bouvet, T. Dieckmann, J. Feigon, Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin, *EMBO J.* **19** (2000), 6870–6881.
- [22] C. Johansson, L.D. Finger, L. Trantirek, T.D. Mueller, S. Kim, I.A. Laird-Offringa, J. Feigon, Solution structure of the complex formed by the two N-terminal RNA-binding domains of nucleolin and a pre-rRNA target, *J. Mol. Biol.* **337** (2004), 799–816.
- [23] M. Schubert, K. Lapouge, O. Duss, F.C. Oberstrass, I. Jelesarov, D. Haas, F.H. Allain, Molecular basis of messenger RNA recognition by the specific bacterial repressing clamp RsmA/CsrA, *Nat. Struct. Mol. Biol.* **14** (2007), 807–813.
- [24] F.C. Oberstrass, S.D. Auweter, M. Erat, Y. Hargous, A. Henning, P. Wenter, L. Reymond, B. Amir-Ahmady, S. Pitsch, D.L. Black et al., Structure of PTB bound to RNA: specific binding and implications for splicing regulation, *Science* **309** (2005), 2054–2057.
- [25] S.D. Auweter, F.C. Oberstrass, F.H.T. Allain, Solving the structure of PTB in complex with pyrimidine tracts: An NMR study of protein-RNA complexes of weak affinities, *J. Mol. Biol.* **367** (2007), 174–186.
- [26] Y. Cavaloc, C.F. Bourgeois, L. Kister, J. Stevenin, The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers, *RNA* **5** (1999), 468–483.
- [27] B. Beuth, M.F. Garcia-Mayoral, I.A. Taylor, A. Ramos, Scaffold-independent analysis of RNA-protein interactions: the Nova-1 KH3-RNA complex, *J. Am. Chem. Soc.* **129** (2007), 10205–10210.
- [28] S. Martin-Tumasz, N.J. Reiter, D.A. Brow, S.E. Butcher, Structure and functional implications of a complex containing a segment of U6 RNA bound by a domain of Prp24, *RNA* **16** (2010), 792–804.
- [29] A. Dey, D. York, A. Smalls-Mantey, M.F. Summers, Composition and sequence-dependent binding of RNA to the nucleocapsid protein of Moloney murine leukemia virus, *Biochemistry* **44** (2005), 3735–3744.
- [30] A. Ramos, S. Grunert, J. Adams, D.R. Micklem, M.R. Proctor, S. Freund, M. Bycroft, D. St Johnston, G. Varani, RNA recognition by a Staufen double-stranded RNA-binding domain, *EMBO J.* **19** (2000), 997–1009.
- [31] C. Dominguez, J.F. Fiset, B. Chabot, F.H.-T. Allain, Structural basis of G-tract recognition and engaging by hnRNP F quasi RRM, *Nat. Struct. Mol. Biol.* **17** (2010), 853–861.
- [32] K. Tsuda, K. Kuwasako, M. Takahashi, T. Someya, M. Inoue, T. Terada, N. Kobayashi, M. Shirouzu, T. Kigawa, A. Tanaka et al., Structural basis for the sequence-specific RNA-recognition mechanism of human CUG-BP1 RRM3, *Nucleic Acids Res.* **37** (2009), 5151–5166.
- [33] C. Dominguez, F.H.T. Allain, NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: a novel mode of RNA recognition, *Nucleic Acids Res.* **34** (2006), 3634–3645.
- [34] M.B. Kamphuis, A.M. Bonvin, M.C. Monti, M. Lemonnier, A. Munoz-Gomez, R.H. van den Heuvel, R. Diaz-Orejas, R. Boelens, Model for RNA binding and the catalytic site of the RNase Kid of the bacterial parD toxin-antitoxin system, *J. Mol. Biol.* **357** (2006), 115–126.
- [35] Y. Yang, N. Declerck, X. Manival, S. Aymerich, M. Kochoyan, Solution structure of the LicT-RNA antitermination complex: CAT clamping RAT, *EMBO J.* **21** (2002), 1987–1997.
- [36] A.P. Golovanov, G.M. Hautbergue, S.A. Wilson, L.Y. Lian, A simple method for improving protein solubility and long-term stability, *J. Am. Chem. Soc.* **126** (2004), 8933–8939.
- [37] G.M. Hautbergue, A.P. Golovanov, Increasing the sensitivity of cryoprobe protein NMR experiments by using the sole low-conductivity arginine glutamate salt, *J. Magn. Reson.* **191** (2008), 335–339.
- [38] F. Vitali, A. Henning, F.C. Oberstrass, Y. Hargous, S.D. Auweter, M. Erat, F.H.T. Allain, Structure of the two most C-terminal RNA recognition motifs of PTB using segmental isotope labeling, *EMBO J.* **25** (2006), 150–162.
- [39] L. Skrisovska, C.F. Bourgeois, R. Stefl, S.N. Grellescheid, L. Kister, P. Wenter, D.J. Elliott, J. Stevenin, F.H. Allain, The testis-specific human protein RBMY recognizes RNA through a novel mode of interaction, *EMBO Rep.* **8** (2007), 372–379.
- [40] F.C. Oberstrass, A. Lee, R. Stefl, M. Janis, G. Chanfreau, F.H. Allain, Shape-specific recognition in the structure of the Vts1p SAM domain with RNA, *Nat. Struct. Mol. Biol.* **13** (2006), 160–167.
- [41] B.M. Lee, J. Xu, B.K. Clarkson, M.A. Martinez-Yamout, H.J. Dyson, D.A. Case, J.M. Gottesfeld, P.E. Wright, Induced fit and "lock and key" recognition of 5S RNA by zinc fingers of transcription factor IIIA, *J. Mol. Biol.* **357** (2006), 275–291.

- [42] H. Wu, A. Henras, G. Chanfreau, J. Feigon, Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III, *Proc. Natl. Acad. Sci. U. S. A.* **101** (2004), 8307–8312.
- [43] P.E. Johnson, L.W. Donaldson, RNA recognition by the Vts1p SAM domain, *Nat. Struct. Mol. Biol.* **13** (2006), 177–178.
- [44] P. Wenter, L. Reymond, S.D. Auweter, F.H. Allain, S. Pitsch, Short, synthetic and selectively ¹³C-labeled RNA sequences for the NMR structure determination of protein-RNA complexes, *Nucleic Acids Res.* **34** (2006), e79.
- [45] M. Sattler, J. Schleucher, C. Griesinger, Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, *Prog. Nucl. Magn. Reson. Spectrosc.* **34** (1999), 93–158.
- [46] J. Cavanagh, W.J. Fairbrother, A.G. Palmer III, M. Rance, N.J. Skelton, *Protein NMR Spectroscopy: Principles and Practice* 2nd ed. Academic Press Inc., San Diego, 2006.
- [47] Z. Cai, A. Gorin, R. Frederick, X. Ye, W. Hu, A. Majumdar, A. Kettani, D.J. Patel, Solution structure of P22 transcriptional antitermination N peptide-boxB RNA complex, *Nat. Struct. Biol.* **5** (1998), 203–212.
- [48] F. Jiang, A. Gorin, W. Hu, A. Majumdar, S. Baskerville, W. Xu, A. Ellington, D.J. Patel, Anchoring an extended HTLV-1 Rex peptide within an RNA major groove containing junctional base triples, *Structure* **7** (1999), 1461–1472.
- [49] N.S. Rao, P. Legault, D.R. Muhandiram, J. Greenblatt, J.L. Battiste, J.R. Williamson, L.E. Kay, NMR pulse schemes for the sequential assignment of arginine side-chain H epsilon protons, *J. Magn. Reson. B* **113** (1996), 272–276.
- [50] R.A. Venters, C.C. Huang, B.T. Farmer, 2nd, R. Trolard, L.D. Spicer, C.A. Fierke, High-level ²H/¹³C/¹⁵N labeling of proteins for NMR studies, *J. Biomol. NMR* **5** (1995), 339–344.
- [51] M. Salzmann, G. Wider, K. Pervushin, K. Wuthrich, Improved sensitivity and coherence selection for [¹⁵N, ¹H]-TROSY elements in triple resonance experiments, *J. Biomol. NMR* **15** (1999), 181–184.
- [52] M. Salzmann, K. Pervushin, G. Wider, H. Senn, K. Wuthrich, TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins, *Proc. Natl. Acad. Sci. U. S. A.* **95** (1998), 13585–13590.
- [53] V. D'Souza, M.F. Summers, Structural basis for packaging the dimeric genome of Moloney murine leukaemia virus, *Nature* **431** (2004), 586–590.
- [54] J. Zhou, R.L. Bean, V.M. Vogt, M. Summers, Solution structure of the Rous sarcoma virus nucleocapsid protein: muPsi RNA packaging signal complex, *J. Mol. Biol.* **365** (2007), 453–467.
- [55] A.L. Breeze, Isotope-filtered NMR methods for the study of biomolecular structure and interactions, *Prog. Nucl. Magn. Reson. Spectrosc.* **36** (2000), 323–372.
- [56] G. Otting, K. Wuthrich, Heteronuclear Filters in 2-Dimensional [¹H-1, H-1] Nmr-Spectroscopy - Combined Use with Isotope Labeling for Studies of Macromolecular Conformation and Intermolecular Interactions, *Q. Rev. Biophys.* **23** (1990), 39–96.
- [57] J. Iwahara, J.M. Wojciak, R.T. Clubb, Improved NMR spectra of a protein-DNA complex through rational mutagenesis and the application of a sensitivity optimized isotope-filtered NOESY experiment, *J. Biomol. NMR* **19** (2001), 231–241.
- [58] R.D. Peterson, C.A. Theimer, H. Wu, J. Feigon, New applications of 2D filtered/edited NOESY for assignment and structure elucidation of RNA and RNA-protein complexes, *J. Biomol. NMR* **28** (2004), 59–67.
- [59] C. Zwahlen, P. Legault, S.J.F. Vincent, J. Greenblatt, R. Konrat, L.E. Kay, Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: Application to a bacteriophage lambda N-peptide/boxB RNA complex, *J. Am. Chem. Soc.* **119** (1997), 6711–6721.
- [60] W. Lee, M.J. Revington, C. Arrowsmith, L.E. Kay, A pulsed field gradient isotope-filtered 3D ¹³C HMQC-NOESY experiment for extracting intermolecular NOE contacts in molecular complexes, *FEBS Lett.* **350** (1994), 87–90.
- [61] S. Cai, L.Y. Zhu, Z.M. Zhang, Y. Chen, Determination of the three-dimensional structure of the Mrf2-DNA complex using paramagnetic spin labeling, *Biochemistry* **46** (2007), 4943–4950.
- [62] S. Grzesiek, F. Cordier, A.J. Dingley, Scalar couplings across hydrogen bonds, *Methods Enzymol.* **338** (2001), 111–133.
- [63] A.Z. Liu, A. Majumdar, F. Jiang, N. Chernichenko, E. Skripkin, D.J. Patel, NMR detection of intermolecular N-H center dot center dot center dot N hydrogen bonds in the human T cell leukemia virus-1 Rex peptide-RNA aptamer complex, *J. Am. Chem. Soc.* **122** (2000), 11226–11227.
- [64] R. Koradi, M. Billeter, K. Wüthrich, MOLMOL: a program for display and analysis of macromolecular structures, *J. Mol. Graph.* **14** (1996), 51–55.

- [65] J.R. Tolman, J.M. Flanagan, M.A. Kennedy, J.H. Prestegard, Nuclear Magnetic Dipole Interactions in Field-Oriented Proteins - Information for Structure Determination in Solution, *Proc. Natl. Acad. Sci. U. S. A.* **92** (1995), 9279–9283.
- [66] J.R. Tolman, K. Ruan, NMR residual dipolar couplings as probes of biomolecular dynamics, *Chem. Rev.* **106** (2006), 1720–1736.
- [67] A. Bax, G. Kontaxis, N. Tjandra, Dipolar couplings in macromolecular structure determination, *Methods Enzymol.* **339** (2001), 127–174.
- [68] J.L. Battiste, G. Wagner, Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data, *Biochemistry* **39** (2000), 5355–5365.
- [69] B. Simon, T. Madl, C.D. Mackereth, M. Nilges, M. Sattler, An efficient protocol for NMR-spectroscopy-based structure determination of protein complexes in solution, *Angew. Chem. Int. Ed.* **49** (2010), 1967–1970.
- [70] X.C. Su, G. Otting, Paramagnetic labelling of proteins and oligonucleotides for NMR, *J. Biomol. NMR* **46** (2010), 101–112.
- [71] L. Varani, S.I. Gunderson, I.W. Mattaj, L.E. Kay, D. Neuhaus, G. Varani, The NMR structure of the 38 kDa U1A protein - PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein, *Nat. Struct. Biol.* **7** (2000), 329–335.
- [72] A. Grishaev, V. Tugarinov, L.E. Kay, J. Trehwella, A. Bax, Refined solution structure of the 82-kDa enzyme malate synthase G from joint NMR and synchrotron SAXS restraints, *J. Biomol. NMR* **40** (2008), 95–106.
- [73] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS plus : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts, *J. Biomol. NMR* **44** (2009), 213–223.
- [74] A.T. Brünger, *XPLOR Manual Version 3.1*. Yale University Press, New Haven, 1993.
- [75] A.T. Brünger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, J.S. Jiang, J. Kuszewski, M. Nilges, N.S. Pannu et al., Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. D Biol. Crystallogr.* **54** (1998), 905–921.
- [76] C.D. Schwieters, J.J. Kuszewski, N. Tjandra, G.M. Clore, The Xplor-NIH NMR molecular structure determination package, *J. Magn. Reson.* **160** (2003), 65–73.
- [77] C.D. Schwieters, J.J. Kuszewski, G.M. Clore, Using Xplor-NIH for NMR molecular structure determination, *Prog. Nucl. Magn. Reson. Spectrosc.* **48** (2006), 47–62.
- [78] P. Guntert, C. Mumenthaler, K. Wuthrich, Torsion angle dynamics for NMR structure calculation with the new program DYANA, *J. Mol. Biol.* **273** (1997), 283–298.
- [79] T. Herrmann, P. Guntert, K. Wuthrich, Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS, *J. Biomol. NMR* **24** (2002), 171–189.
- [80] J. Wang, P. Cieplak, P.A. Kollman, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?, *J. Comp. Chem.* **21** (2000), 1049–1074.
- [81] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A 2nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules, *J. Am. Chem. Soc.* **117** (1995), 5179–5197.
- [82] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha et al., All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem. B* **102** (1998), 3586–3616.
- [83] A.D. MacKerell, N. Banavali, N. Foloppe, Development and current status of the CHARMM force field for nucleic acids, *Biopolymers* **56** (2000), 257–265.
- [84] B. Xia, V. Tsui, D.A. Case, H.J. Dyson, P.E. Wright, Comparison of protein solution structures refined by molecular dynamics simulation in vacuum, with a generalized Born model, and with explicit water, *J. Biomol. NMR* **22** (2002), 317–331.
- [85] J.P. Linge, M.A. Williams, C.A.E.M. Spronk, A.M.J.J. Bonvin, M. Nilges, Refinement of protein structures in explicit solvent, *Proteins* **50** (2003), 496–506.
- [86] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* **79** (1992), 926–935.
- [87] V. Tsui, D.A. Case, Molecular dynamics simulations of nucleic acids with a generalized born solvation model, *J. Am. Chem. Soc.* **122** (2000), 2489–2498.
- [88] T. Herrmann, P. Guntert, K. Wuthrich, Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA, *J. Mol. Biol.* **319** (2002), 209–227.

- [89] D.A. Case, T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, The Amber biomolecular simulation programs, *J. Comput. Chem.* **26** (2005), 1668–1688.
- [90] P. Padrta, R. Stefl, L. Kralik, L. Zidek, V. Sklenar, Refinement of d(GCGAAGC) hairpin structure using one- and two-bond residual dipolar couplings, *J. Biomol. NMR* **24** (2002), 1–14.
- [91] X.P. Xu, D.A. Case, Automated prediction of N-15, C-13(alpha), C-13(beta) and C-13 ' chemical shifts in proteins using a density functional database, *J. Biomol. NMR* **21** (2001), 321–333.
- [92] S. Neal, A.M. Nip, H.Y. Zhang, D.S. Wishart, Rapid and accurate calculation of protein H-1, C-13 and N-15 chemical shifts, *J. Biomol. NMR* **26** (2003), 215–240.
- [93] J.A.M.T.C. Cromsig, C.W. Hilbers, S.S. Wijmenga, Prediction of proton chemical shifts in RNA - Their use in structure refinement and validation, *J. Biomol. NMR* **21** (2001), 11–29.
- [94] A. Mittermaier, L.E. Kay, New tools provide new insights in NMR studies of protein dynamics, *Science* **312** (2006), 224–228.
- [95] Z. Shajani, G. Varani, NMR studies of dynamics in RNA and DNA by ¹³C relaxation, *Biopolymers* **86** (2007), 348–359.
- [96] M. Getz, X. Sun, A. Casiano-Negroni, Q. Zhang, H.M. Al-Hashimi, NMR studies of RNA dynamics and structural plasticity using NMR residual dipolar couplings, *Biopolymers* **86** (2007), 384–402.
- [97] G. Lipari, A. Szabo, Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity, *J. Am. Chem. Soc.* **104** (1982), 4546–4559.
- [98] G. Lipari, A. Szabo, Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .2. Analysis of Experimental Results, *J. Am. Chem. Soc.* **104** (1982), 4559–4570.
- [99] D.M. Korzhnev, L.E. Kay, Probing invisible, low-populated states of protein molecules by relaxation dispersion NMR spectroscopy: An application to protein folding, *Acc. Chem. Res.* **41** (2008), 442–451.
- [100] A.G. Palmer, F. Massi, Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy, *Chem. Rev.* **106** (2006), 1700–1719.
- [101] A.T. Frank, A.C. Stelzer, H.M. Al-Hashimi, I. Andricioaei, Constructing RNA dynamical ensembles by combining MD and motionally decoupled NMR RDCs: new insights into RNA dynamics and adaptive ligand recognition, *Nucleic Acids Res.* **37** (2009), 3670–3679.
- [102] A. Mittermaier, L. Varani, D.R. Muhandiram, L.E. Kay, G. Varani, Changes in side-chain and backbone dynamics identify determinants of specificity in RNA recognition by human U1A protein, *J. Mol. Biol.* **294** (1999), 967–979.
- [103] Z. Shajani, G. Varani, ¹³C NMR relaxation studies of RNA base and ribose nuclei reveal a complex pattern of motions in the RNA binding site for human U1A protein, *J. Mol. Biol.* **349** (2005), 699–715.
- [104] Z. Shajani, G. Drobny, G. Varani, Binding of U1A protein changes RNA dynamics as observed by ¹³C NMR relaxation studies, *Biochemistry* **46** (2007), 5875–5883.
- [105] H.R. Jonker, S. Ilin, S.K. Grimm, J. Wohnert, H. Schwalbe, L11 domain rearrangement upon binding to RNA and thioesteron studied by NMR spectroscopy, *Nucleic Acids Res.* **35** (2007), 441–454.
- [106] P. Deka, P.K. Rajan, J.M. Perez-Canadillas, G. Varani, Protein and RNA dynamics play key roles in determining the specific recognition of GU-rich polyadenylation regulatory elements by human Cstf-64 protein, *J. Mol. Biol.* **347** (2005), 719–733.
- [107] S. Ravindranathan, F.C. Oberstrass, F.H. Allain, Increase in Backbone Mobility of the VTS1p-SAM Domain on Binding to SRE-RNA, *J. Mol. Biol.* **396** (2010), 732–746.
- [108] F.C. Oberstrass, F.H. Allain, S. Ravindranathan, Changes in dynamics of SRE-RNA on binding to the VTS1p-SAM domain studied by ¹³C NMR relaxation, *J. Am. Chem. Soc.* **130** (2008), 12007–12020.
- [109] D.M. Korzhnev, P. Neudecker, A. Mittermaier, V.Y. Orekhov, L.E. Kay, Multiple-site exchange in proteins studied with a suite of six NMR relaxation dispersion experiments: an application to the folding of a Fyn SH3 domain mutant, *J. Am. Chem. Soc.* **127** (2005), 15602–15611.
- [110] P. Neudecker, A. Zarrine-Afsar, W.Y. Choy, D.R. Muhandiram, A.R. Davidson, L.E. Kay, Identification of a collapsed intermediate with non-native long-range interactions on the folding pathway of a pair of Fyn SH3 domain mutants by NMR relaxation dispersion spectroscopy, *J. Mol. Biol.* **363** (2006), 958–976.
- [111] A.L. Hansen, E.N. Nikolova, A. Casiano-Negroni, H.M. Al-Hashimi, Extending the Range of Microsecond-to-Millisecond Chemical Exchange Detected in Labeled and Unlabeled Nucleic Acids by Selective Carbon R-1 rho NMR Spectroscopy, *J. Am. Chem. Soc.* **131** (2009), 3818–3819.
- [112] Q. Zhang, A.C. Stelzer, C.K. Fisher, H.M. Al-Hashimi, Visualizing spatially correlated dynamics that directs RNA conformational transitions, *Nature* **450** (2007), 1263–1267.
- [113] G.K. Amarasinghe, R.N. De Guzman, R.B. Turner, K.J. Chancellor, Z.R. Wu, M.F. Summers, NMR structure of the HIV-1 nucleocapsid protein bound to stem-loop SL2 of the psi-RNA packaging signal. Implications for genome recognition, *J. Mol. Biol.* **301** (2000), 491–511.

- [114] R.N. De Guzman, Z.R. Wu, C.C. Stalling, L. Pappalardo, P.N. Borer, M.F. Summers, Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element, *Science* **279** (1998), 384–388.
- [115] R. Lamichhane, G.M. Daubner, J. Thomas-Crusells, S.D. Auweter, C. Manatschal, K.S. Austin, O. Valniuk, F.H. Allain, D. Rueda, RNA looping by PTB: Evidence using FRET and NMR spectroscopy for a role in splicing repression, *Proc. Natl. Acad. Sci. U. S. A.* **107** (2010), 4105–4110.