



Structure determination and dynamics of protein–RNA complexes by NMR spectroscopy

Cyril Dominguez^{a,1}, Mario Schubert^{a,1}, Olivier Duss^{a,1}, Sapna Ravindranathan^{b,1}, 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

ARTICLE INFO

Article history:

Received 7 April 2010

Accepted 24 April 2010

Available online 4 November 2010

Keywords:

Protein–RNA complex

NMR spectroscopy

Structure determination

Dynamics

Contents

1. Introduction	2
2. How to get a protein–RNA complex sample for NMR spectroscopy	3
2.1. Finding optimal protein and RNA constructs for NMR studies of protein–RNA complexes	3
2.1.1. Finding natural RNA substrate bound by RNA binding proteins	4
2.1.2. Finding high-affinity RNA aptamers bound by RNA binding proteins	5
2.1.3. Optimizing the RNA target of a protein–RNA complex for its NMR study	6
2.2. Protein production	7
2.2.1. Domain boundaries	7
2.2.2. Cloning and expression of RNA binding proteins	8
2.2.3. The cell-free <i>in vitro</i> system	9
2.2.4. Purification of RNA binding proteins	9
2.2.5. Ribonuclease activity	10
2.3. RNA production	10
2.3.1. RNA synthesis	10
2.3.2. RNA purification	12
2.3.3. Segmental isotope labeling for larger RNAs	12
2.4. Complex formation	13
2.4.1. Estimation of the protein and RNA concentrations	13
2.4.2. Monitoring complex formation by NMR titration experiments	13
2.4.3. NMR titration experiments are crucial steps for defining a “good” protein–RNA complex	15
3. NMR spectroscopy and structure determination of protein–RNA complexes	17
3.1. NMR methodology and resonance assignment	17
3.1.1. Preliminary NMR experiments	17
3.1.2. Temperature, ionic strength, and solvent	18
3.1.3. Typical samples for NMR measurements of protein–RNA complexes	19
3.1.4. Typical samples for NMR measurements of peptide–RNA complexes	19

* Corresponding author. Tel.: +41 (0) 44 63 33940; fax: +41 (0) 44 63 31294.

E-mail address: allain@mol.biol.ethz.ch (F.H.-T. Allain).

¹ All authors contributed equally to this work.

3.1.5.	Resonance assignment of proteins and peptides in complex with RNA	19
3.1.6.	Resonance assignment of proteins in large protein–RNA complexes	20
3.1.7.	Resonance assignment of RNA in small to medium size complexes	21
3.1.8.	Resonance assignment of RNA in large complexes	23
3.2.	Restraints for the protein–RNA interface	24
3.2.1.	Editing and filtering building blocks	24
3.2.2.	2D and 3D filtered/edited NOESY experiments	26
3.2.3.	Editing or filtering of NH	27
3.2.4.	Alternative methods used to derive intermolecular NOEs	28
3.2.5.	Special case: protein dimer	28
3.3.	Hydrogen bond and angle restraints	28
3.3.1.	Experiments for obtaining hydrogen bond restraints	28
3.3.2.	Experiments for obtaining angle restraints	30
3.4.	Long range restraints – global orientation	31
3.4.1.	Residual dipolar couplings	31
3.4.2.	Paramagnetic relaxation enhancement	34
3.4.3.	Combining several methods to reach complexes of ever increasing size	36
3.5.	Structure determination	37
3.5.1.	Experimental restraints	37
3.5.2.	Structure determination protocols	38
3.5.3.	Structure refinement	39
3.5.4.	Using solvation models in structure refinement	39
3.5.5.	An example of structure calculation and refinement protocol	39
3.6.	Validation, precision and accuracy	40
3.6.1.	Accuracy and precision of the structure ensemble	40
3.6.2.	Validation of the structures against experimental restraints	42
3.6.3.	Validation of the structures based on geometrical and structural characteristics	43
3.6.4.	Confirmation and quantification of intermolecular contacts	43
3.6.5.	Structure–function relationship of protein–RNA complexes	44
3.7.	Dynamics of protein–RNA complexes	44
3.7.1.	Protein dynamics	45
3.7.2.	RNA dynamics	47
3.7.3.	Insights from dynamics studies	48
3.7.4.	Future directions	48
4.	NMR structures of protein–RNA complexes: what did we learn from them?	48
4.1.	Introduction	48
4.2.	Retroviral and bacteriophage protein–RNA complexes	48
4.2.1.	Viral NC–RNA complexes	49
4.2.2.	Viral Rev–RRE, Tat–TAR and bacteriophage N protein–BoxB RNA complexes	49
4.3.	Prokaryotic gene regulation	49
4.4.	Eukaryotic post-transcriptional gene regulation	51
4.4.1.	RNA binding domains	51
4.4.2.	Mechanistic insights provided by protein–RNA structures	54
4.4.3.	Structural insights into macromolecular assemblies	56
5.	Conclusion	56
6.	Note added in proof	56
	Acknowledgements	56
	References	56

1. Introduction

As of 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 [1–45] (Fig. 1) and two consist of structural models generated by docking using sparse NMR data [46,47]. This results in an estimated weight for the contribution of NMR in the structure elucidation of protein–RNA complexes to approach 40%. This by itself illustrates the important role taken by NMR spectroscopy in elucidating structures of protein–RNA complexes in this molecular weight range. This promises an even more important role for NMR in structural biology in the future considering the growing role played by protein–RNA interactions in regulating gene expression.

Historically, 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) bound to a 26-nucleotide (26-nt) RNA stem-loop [31,43]. One year later, in 1996, the first structure of a

protein–RNA complex was solved by NMR. 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 [2]. Since then a total of 52 structures of peptide–RNA and protein–RNA complexes have been determined using NMR spectroscopy (Fig. 1). This provides us with the opportunity to review what are those structures, how they were determined and what did we learn from them. The first part of the review (Sections 2.1–2.4) describes what makes a good protein–RNA complex amenable for NMR structure determination and how the appropriate solution conditions can be obtained. The second part focuses more on the NMR spectroscopy of these complexes (Sections 3.1–3.4) and how, from the NMR spectra, one can derive a precise structure of a protein–RNA complex (Section 3.5). This second part ends with a discussion on the precision and accuracy of the resulting structures (Section 3.6) and with a section on what can we learn from the few dynamics studies of protein–RNA complexes performed using NMR (Section 3.7). This review ends with a brief description of this large ensemble of NMR structures and a discussion on how these

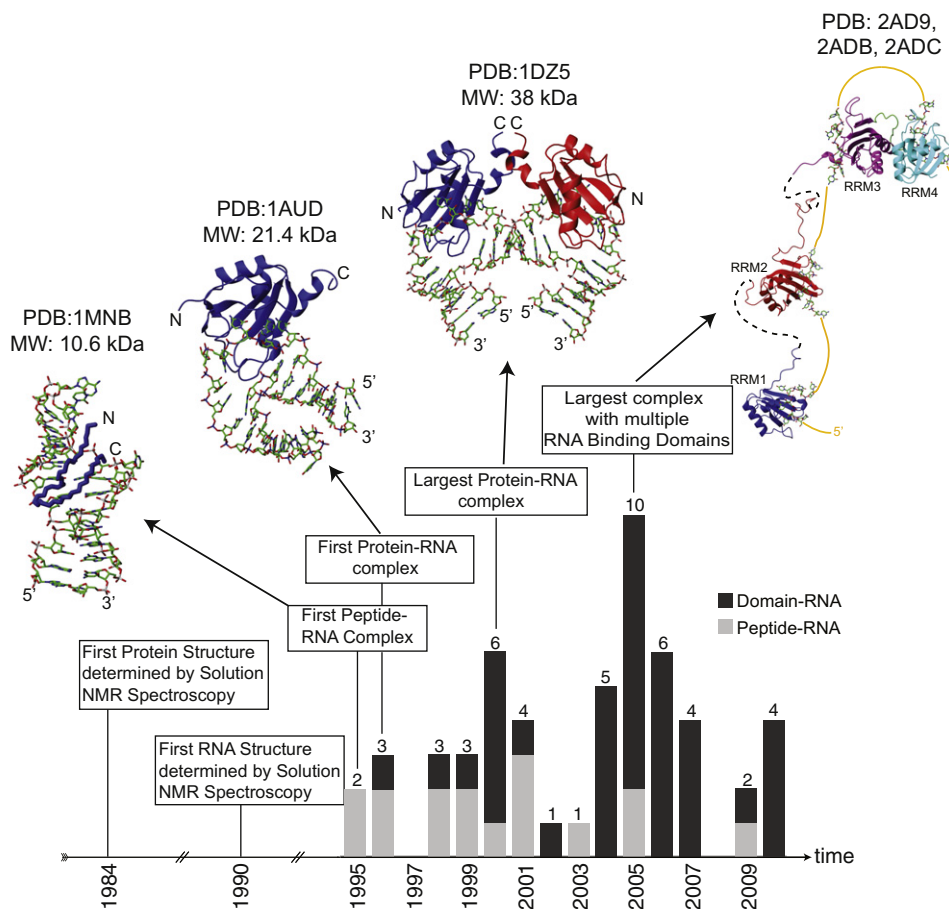


Fig. 1. Historic diagram of protein–RNA complexes solved by NMR. Diagram displaying the number of NMR structures of peptide–RNA (grey) or protein–RNA (black) complexes per year. The NMR structures of the first peptide–RNA [31], the first protein–RNA [2,3], the largest molecular weight protein–RNA complexes [38], and the largest protein–RNA complex with multiple RBDs [28] are depicted. RNA is depicted as stick structure and colored green, peptide or proteins are shown as ribbon structures and colored blue, red, magenta, or cyan. Figures were generated with molmol [470]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

NMR structures of protein–RNA complexes have impacted on the field of structural and RNA biology (Sections 4.1–4.4).

2. How to get a protein–RNA complex sample for NMR spectroscopy

2.1. 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). Finding the RNA sequence that is specifically recognized by the protein of interest, however, is often not trivial for several reasons. First, RBPs or RBDs can recognize and bind RNA in a shape-specific, in a sequence-specific or even in a non-specific manner. Second, RNA molecules are composed of only four different nucleotides and can form a variety of secondary and tertiary structures that can be crucial for protein recognition. Therefore, to study a protein–RNA complex by NMR, it is of particular importance to understand the specificity of the complex formation. The major challenge consists of identifying a suitable RNA sequence that is bound both specifically and with sufficient affinity by the protein. Although this identification does not necessarily involve NMR or structural biology techniques, we think that it is worth mentioning it because the success of the NMR study

strongly depends on this prior biological knowledge. The main questions that one should address 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?
- In the case of a shape-specific complex, how does the RNA structure influence the binding?
- In the case of a sequence-specific complex, which RNA sequence is specifically recognized?

There are many techniques and methods used to identify protein–RNA complexes. 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 small RNA sequences selected from a large random pool of sequences denoted aptamers that are bound with high affinity by RNA binding proteins. In this case, the RNA sequence might not be found in natural RNA targets (see Section 2.1.2) but this method can provide a suitable sequence for structural analysis. Most NMR structures of protein–RNA complexes have been solved using natural RNA sequences, while in few other complexes, the RNA

sequences used were aptamers derived from *in vitro* methods that did not necessarily represent a natural RNA sequence (Table 1). Finally, in some cases, structures of a protein in complex with both a natural target and an aptamer allow an understanding of the molecular basis of the specificity and the affinity of the protein–RNA complexes (see below).

2.1.1. Finding natural RNA substrate bound by RNA binding proteins

There are many strategies and techniques that allow the definition of a relatively short natural RNA sequence specifically recognized by an RNA binding protein. Natural RNA sequences used for NMR structure determination of protein–RNA complexes have been mainly derived from two different strategies: either the use of footprinting techniques that were initially developed for pro-

tein–DNA complexes [48], or the use of RNA truncation and mutagenesis combined with protein–RNA binding assays. The most commonly used binding assay in studies of protein–RNA complexes is the Electrophoretic Mobility Shift Assay (EMSA) [49,50].

The footprinting technique applied to protein–RNA complexes is a protection assay based on the ability of an RNA binding protein to protect RNA from cleavage by different ribonucleases (RNase) or chemical cleavage agents. This technique generally allows for the identification of the RNA binding site of a protein. The use of different ribonucleases that specifically cleave nucleotides located in single-stranded or double-stranded regions also allow the identification of the secondary structures of the RNA bound by the protein. For the NMR structure determinations of 15 protein–RNA complexes, the RNA binding site or sequence was derived from

Table 1
Overview of the protein–RNA complexes solved by NMR.

Year	PDB code	Protein (aa)	RNA (nt)	MW (kDa)	REF	Type of complex ^a	Source of RNA	Method to derive the RNA sequence ^b	
1995	1MNB	14	24	10.6	[31]	Pep–RNA	Natural	Footprinting [414]	
	1BIV	17	24	10.9	[43]	Pep–RNA	Natural	Footprinting [414]	
1996	1ETG	22	39	14	[6]	Pep–RNA	Natural	RNA truncation + EMSA [53,54]	
	1ULL	17	35	13.7	[41]	Pep–RNA	Aptamer	SELEX [415,416]	
	1AUD	100	28	21.4	[2,3]	Prot–dsRNA	Natural	Footprinting + EMSA [413]	
1998	1A4T	20	15	7.2	[7]	Pep–RNA	Natural	Footprinting [417]	
	–	22	19	8.9	[24]	Pep–RNA	Natural	Footprinting [417]	
	1A1T	55	17	13.7	[12]	Prot–dsRNA	Natural	Truncation + EMSA [418]	
1999	1EXY	16	32	12.8	[19]	Pep–RNA	Aptamer	SELEX [419]	
	484D	17	27	11.1	[42]	Pep–RNA	Aptamer	SELEX [420]	
	1D6K	94	37	22.7	[36]	Prot–dsRNA	Natural	Footprinting [52]	
2000	1QFQ	36	15	8.9	[33]	Pep–RNA	Natural	Footprinting [417]	
	1FNX	174	10	22.3	–	Prot–ssRNA	–	–	
	1EKZ	68	30	19.5	[32]	Prot–dsRNA	Unnatural	Designed [32]	
	1F6U	55	19	12.6	[4]	Prot–dsRNA	Natural	Truncation + EMSA [418]	
	1FJE	158	24	28	[1]	Prot–dsRNA	Aptamer	SELEX [68]	
	1DZ5	100 (×2)	44	38	[38]	Prot–dsRNA	Natural	Footprinting + EMSA [413]	
2001	1HJI	15	15	8.1	[15]	Pep–RNA	Natural	Footprinting [417]	
	1G70	21	32	12.9	[16]	Pep–RNA	Natural	RNA truncation + EMSA [53,54]	
	1I9F	19	39	13.1	[44]	Pep–RNA	Natural	RNA truncation + EMSA [53,54]	
	1K1G	127	11	18.2	[26]	Prot–ssRNA	Natural	UV crosslinking + EMSA [421]	
2002	1L1C	55 (×2)	31	21.5	[40]	Prot–dsRNA	Natural	Mutagenesis + Genetic assays [422,423]	
2003	1NYB	22	24	10.3	[9]	Pep–RNA	Natural	Footprinting [417]	
2004	1RGO	70	9	11.1	[18]	Prot–ssRNA	Aptamer	SELEX [67]	
	1T2R	119	5	15.4	[25]	Prot–ssRNA	Unnatural	NMR + EMSA [403,404]	
	1RKJ	158	21	28	[20]	Prot–dsRNA	Natural	Footprinting [68]	
	1T4L	88	32	20.1	[39]	Prot–dsRNA	Natural	Footprinting [398]	
	1U6P	56	101	39.1	[10]	Prot–dsRNA	Natural	RNA Truncation + EMSA [55]	
	1ZBN	17	26	11	[8]	Pep–RNA	Natural	Footprinting [414]	
	2A9X	14	29	10.7	[23]	Pep–RNA	Natural	Footprinting [414]	
2005	1WWD	56	6	8.3	[13]	Prot–ssRNA	Natural	Chemical probing [385,424]	
	1WWE	56	7	8.5	[13]	Prot–ssRNA	Natural	Chemical probing [385,424]	
	1WWF	56	7	8.5	[13]	Prot–ssRNA	Natural	Chemical probing [385,424]	
	1WWG	56	6	8.3	[13]	Prot–ssRNA	Natural	Chemical probing [385,424]	
	2AD9	115	6	14.9	[28]	Prot–ssRNA	Natural	Mutagenesis + EMSA [56,425]	
	2ADB	145	6	18.2	[28]	Prot–ssRNA	Natural	Mutagenesis + EMSA [56,425]	
	2ADC	208	6	27.1	[28]	Prot–ssRNA	Natural	Mutagenesis + EMSA [56,425]	
	2C06	110 (×2)	5	13.4	[47]	Prot–ssRNA	Natural	Cleavage of natural RNAs [426]	
	2006	2CJK	166	9	21.6	[30]	Prot–ssRNA	Natural	Mutation + UV cross-linking [427]
		2ERR	100	7	14.6	[5]	Prot–ssRNA	Natural	Splicing + SELEX [61]
2I2Y		87	4	18.1	[17]	Prot–ssRNA	Aptamer	SELEX [66]	
2ESE		81	23	18.9	[29]	Prot–dsRNA	Natural	EMSA + Fluorescence Polarization [402,428]	
2B6G		117	19	19.1	[21]	Prot–dsRNA	Natural	EMSA + Fluorescence Polarization [402,428]	
2HGH		87	55	27.8	[22]	Prot–dsRNA	Natural	Truncation + EMSA [429]	
2FY1		108	21	19.5	[35]	Prot–dsRNA	Aptamer	SELEX [35]	
2007	2IHX	61	75	31	[45]	Prot–dsRNA	Natural	Mutagenesis + EMSA + ITC [430]	
	2JPP	70 (×2)	20	28	[34]	Prot–dsRNA	Natural	EMSA / SELEX [431,432]	
	2JQ7	141	58	36	[46]	Prot–dsRNA	Natural	Footprinting [51]	
	2KDQ	14	27	11.1	[11]	Pep–RNA	Natural	Footprinting [414]	
2009	2RQ4	99	6	14.2	[37]	Prot–ssRNA	Natural	EMSA [433]	
	2KH9	83	6	10.8	[27]	Prot–ssRNA	Natural	SIA [27]	
2010	2KFY	102	6	13.4	[14]	Prot–ssRNA	Natural	RNA affinity assays [73]	
	2KG0	92	6	12.2	[14]	Prot–ssRNA	Natural	RNA affinity assays [73]	
	2KG1	105	6	13.5	[14]	Prot–ssRNA	Natural	RNA affinity assays [73]	

^a Pep–RNA: peptide–RNA complex, Prot–dsRNA: protein–double-stranded RNA complex, Prot–ssRNA: protein–single-stranded RNA complex.

^b SELEX: Systematic Evolution of Ligands by Exponential Enrichment, EMSA: ElectroMobility Shift Assay, SIA: Scaffold-Independent Analysis.

footprinting experiments (Table 1). For example, footprinting experiments have been performed to define the RNA region of the 23S ribosomal RNA (rRNA) bound by the protein L11 [51]. This study showed that the addition of protein L11 induced the protection from RNase digestion of a small RNA fragment of 58 nucleotides among the 2904 nucleotides of the 23S rRNA. This fragment was then used to determine the structure of L11–23SRNA complex [46]. Similarly, a 37-nt RNA fragment of the 5S rRNA (115 nucleotides) was shown to be protected by the protein L25 [52]. This RNA fragment was then used to solve the NMR structure of the L25–5S rRNA complex [36].

The second strategy to identify natural RNA sequences recognized by proteins uses RNA truncation and mutagenesis together with binding assays. In this case, a long natural RNA sequence is truncated into various, sometimes overlapping, small fragments. Each fragment is then tested for binding to the protein of interest. Such an approach has been used to define the minimal RNA sequence recognized by the human immunodeficiency virus (HIV) Rev (regulator of viral expression) protein [53]. Rev binds an RNA segment of 367 nucleotides, termed Rev Response Element (RRE) that is thought to adopt a highly ordered structure composed of stem-loops [54]. By truncating this RNA fragment into 13 different sub-fragments and testing the binding of the Rev protein to each of them using EMSA, a small 40-nucleotide fragment that is sufficient for Rev binding could be identified [53]. This fragment was then used to solve the NMR structure of the HIV Rev–RRE complex [6]. The minimal region of the moloney murine leukemia virus (MoMuLV) Ψ -RNA bound by the nucleocapsid (NC) protein was also determined by RNA truncation experiments combined with EMSA [55]. This RNA is 350-nucleotides long and contains a central portion of 102 nucleotides that forms three stem loops connected by short linkers. The 102-nucleotide segment was truncated into 8 different and overlapping fragments and their binding to the NC protein was tested by EMSA. In this case, however, the study showed that the full 102-nucleotide fragment is necessary for high affinity binding to the protein NC [55]. This full RNA fragment was therefore used to solve the structure of the NC– Ψ RNA complex [10]. Creating mutants of the RNA sequence combined with binding assays also allows one to gain insights into the specificity of the interaction and therefore helps in defining a suitable RNA sequence for NMR analysis. EMSA and ultraviolet (UV) crosslinking experiments have shown that the protein PTB (Polypyrimidine Tract Binding protein) binds RNA sequences containing a CUCUCU motif and that binding is abolished when this motif is mutated [56]. The CUCUCU RNA was therefore used to solve the NMR structure of PTB in complex [28].

Recently, a novel *in vivo* method has been developed to identify natural RNA targets of RNA binding proteins using high-throughput technologies [57,58]. This method, called CLIP (UV Cross-Linking and Immunoprecipitation assay), is based on the ability of UV irradiations to induce the formation of covalent bonds between proteins and RNAs when interacting ([59] and references therein). UV cross-linking can be performed on cellular or nuclear extracts to identify natural sequences bound by RNA binding proteins. The sequences obtained are then compared to define the RNA binding sequence motif of the protein of interest. CLIP experiments have been performed to identify RNA substrates of many RNA binding proteins and in particular of Fox-2 [60]. The RNA sequences retrieved from CLIP experiments identified a consensus sequence, UGCAUG, that corresponds exactly to the sequence identified by SELEX (see next Section) and was used to solve the NMR structure of the complex [5,61]. This indicates that CLIP has a high potential for identifying natural RNA binding sequences suitable for NMR investigation of protein–RNA complexes.

2.1.2. Finding high-affinity RNA aptamers bound by RNA binding proteins

The main technique to identify RNA aptamers bound by RNA binding proteins is the Systematic Evolution of Ligands by Exponential Enrichment (SELEX) approach. SELEX is an *in vitro* method that was developed in 1990 in two independent laboratories [62,63]. It allows the identification of small DNA or RNA sequences that bind with high affinity to nucleic acid binding proteins. The main principles as well as the advantages, limitations and applications of this method have been recently reviewed [64,65]. In brief, the SELEX method applied to protein–RNA complexes consists of creating a random RNA oligonucleotide library that possesses a central random sequence (between 20 and 80 nucleotides). The binding of the protein of interest to this pool of RNA molecules is performed in an iterative manner until a small number of RNA sequences bound by the protein can be identified. Generally, between 5 and 20 rounds of selection are needed to identify a good RNA binding consensus sequence. Commonly, about 50 sequences are analyzed using sequence alignment and secondary structure prediction algorithms. The sequence alignment allows the definition of a high affinity RNA binding motif for the protein of interest, while the secondary structure prediction defines if the protein binds preferentially RNAs embedded in particular secondary structures. Further studies, such as RNA binding assays are then performed in order to quantify the SELEX results.

The SELEX method is a powerful tool for identifying RNA sequences specifically recognized by proteins or protein domains. Therefore, an RNA binding sequence identified by SELEX provides a good starting point for the NMR investigation of a protein–RNA complex. RNA sequences derived from SELEX experiments are also useful to identify natural RNA targets of a specific RNA binding protein. Numerous protein–RNA complexes have been solved using RNA sequences initially derived from SELEX that also match perfectly natural sequences. For example, SELEX experiments have been performed to define a RNA binding consensus sequence for the splicing factors 9G8 and SRp20 and also to identify natural targets for both proteins [66]. This information was used to solve the NMR structure of SRp20 in complex with RNA [17]. Similarly, SELEX experiments helped identify RNA binding sequences for Tis11d that matched perfectly natural sequences identified previously [67] and could be used to solve the NMR structure of Tis11d in complex with RNA [18]. Again, the structure of Fox-1 in complex with RNA was solved using an RNA sequence initially derived from SELEX [5]. Subsequently, both the SELEX results and the structural work [61] allowed the identification of many natural pre-messenger RNA (pre-mRNA) targets of Fox-1 [60].

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. Three NMR structures of the HIV Rev peptide in complex with RNA have been solved using either a natural RNA target [6] or two different RNA aptamers derived from SELEX experiments [41,42]. The three RNAs adopt a stem-loop structure but have different nucleotide sequences and structural features (Fig. 2A). A striking difference in these structures is that the Rev peptide adopts an α -helical conformation when bound to the natural or the Class I SELEX sequences but adopts an elongated conformation when bound to a Class II RNA sequence. These structural differences were explained by the ability of arginine-rich peptides to undergo adaptive folding transitions correlated to the structural properties of the bound RNA [42]. Another interesting example illustrating the difference between natural RNA sequence and aptamers derived from SELEX concerns the protein nucleolin. In this case, natural RNA sequences bound by nucleolin were identi-

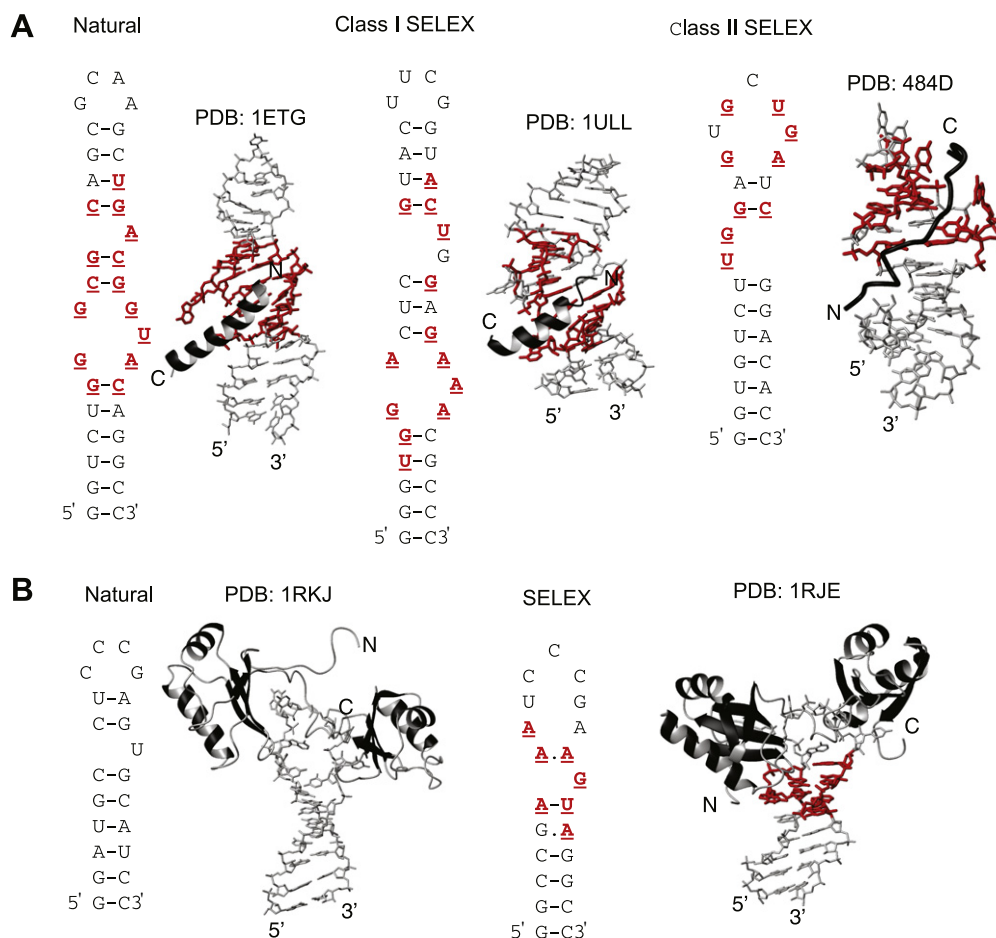


Fig. 2. Example of discrepancy between natural RNA sequences and SELEX-derived RNA sequences. (A) RNA sequences used to solve the NMR structures of the HIV Rev-RRE complexes [6,41,42]. Nucleotides in red, bold and underlined are contacting the peptide. (B) RNA sequences used to solve the NMR structures of the nucleolin-RNA complexes [1,20]. Nucleotides in red, bold and underlined adopt a loop E conformation. Figures were generated with molmol [470].

fied by footprinting experiments and aptamer RNAs were derived from SELEX experiments [68]. Both approaches identified RNAs adopting a stem-loop structure with a highly conserved loop sequence. The nucleotides forming the stem, however, differ between the two RNAs (Fig. 2B). The affinity of nucleolin for these two RNAs is very different (dissociation constants of 1.9 nM for the SELEX RNA and of 1.1 μ M for the natural RNA). The structures of the two N-terminal RRM of nucleolin bound to the SELEX RNA [1] and to the natural RNA sequence [20] were solved by NMR. The overall structural features in both complexes are the same and the intermolecular contacts are also almost identical. However, differences were observed in the contacts of nucleolin to the top of the RNA stem. In the SELEX RNA, the RNA structure adopts a loop E motif that is recognized by the protein. This structural feature is absent in the natural RNA. This structural difference of the two RNAs explains the large difference of affinity between these two related complexes.

2.1.3. Optimizing the RNA target of a protein-RNA complex for its NMR study

The methods described above are very powerful to identify RNA sequences that are bound both specifically and with high affinity by an RNA binding protein. However, these two methods generate an ensemble of RNA sequences and, in many cases, the derived consensus motif is degenerate. One example of such degeneracy was observed for the protein ASF/SF2 (alternative splicing factor/splicing factor 2) that contains two RNA binding domains. SE-

LEX experiments using a protein construct containing both RBDs identified two consensus sequences, RGAAGAAC and AGGACR-RAGC (where R indicates a purine), while SELEX experiments using a protein construct containing only the first RBD of ASF/SF2 identified a different RNA sequence, ACGCGCA [69]. The RNA consensus recognized by the protein NOVA is another example of a degenerate sequence that can be identified by such methods. In this case, both SELEX and CLIP methods identified an RNA consensus sequence recognized by the RNA binding protein NOVA as YCAYY, where Y indicates a pyrimidine [57,70]. In addition, these methods generate rather long RNA sequences containing a generally short consensus motif. For NMR studies, it is important to identify the minimum RNA sequence that provides both high affinity and specificity to the protein of interest. Furthermore, based on our experience, high affinity RNA sequences do not necessarily provide NMR spectra of the best quality. Therefore, the study of protein-RNA complexes by NMR often needs optimization in order to define the optimal RNA binding sequence for structural investigation.

2.1.3.1. Scaffold-Independent Analysis (SIA) of RNA-protein interactions.

The scaffold-Independent Analysis of RNA-protein interactions was developed in 2007 by Ramos and coworkers [71]. This method aims at elucidating the binding specificity of protein-RNA complexes by NMR at the single nucleotide level and makes use of synthetic randomized RNA sequences. Ramos and coworkers successfully used this method to decipher the RNA binding specificity of the KH (heterogeneous nuclear ribonucleoproteins

(hnRNP) K homology) domain of the protein Nova [71]. Assuming a RBD that binds a tetranucleotide RNA, 16 randomized RNAs are synthesized. For each RNA, one position (1–4) is occupied by A, G, C or U and the three other positions are occupied by a randomized mixture of the four bases (for example position 1 analysis was composed of four RNAs with sequences ANNN, GNNN, CNNN and UNNN, where N is a mixture of the four nucleotides). Using ^{15}N - ^1H heteronuclear single quantum coherence (HSQC) NMR spectra, each of these RNAs was tested for binding to the protein of interest by following chemical shift perturbations of the protein amide NH resonances upon RNA titration up to a protein:RNA ratio of 1:4. The amplitude of the chemical shift changes as a function of the RNA sequence was then used to rank the preference of the protein for a specific nucleobase at a specific position. 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 complexes. The drawbacks, however are that the exact number of nucleotides necessary for protein binding must be guessed prior to the SIA analysis and that this method requires the synthesis of a large number of degenerate RNA sequences and can be costly. In addition, this approach can be used only for very small RNA sequences since the number of RNAs to be tested is four times higher than the number of nucleotide positions under investigation. This approach has been very recently used for the determination of a protein–RNA complex [27]. The protein Prp24 (precursor RNA processing 24) possesses 3 RNA recognition motifs and specifically recognizes the U6 ribosomal RNA. Using SIA, Butcher and coworkers identified the optimal sequence recognized by Prp24 RRM2 as GAGA, a sequence that is naturally present in the U6 RNA, and therefore solved the structure of Prp24 RRM2 in complex with AGAGAU [27].

2.1.3.2. Further refinement of protein–RNA complexes. 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 Nuclear Overhauser Effects (NOEs) that are essential for the structure calculation of the complex (see Sections 2.4 and 3.2).

When the protein binds double-stranded RNA, the stability of the RNA structure is crucial. Therefore, both the length and the base composition of the RNA need to be optimized. Often, RNA binding proteins or domains specifically bind an RNA sequence that forms a stem-loop structure corresponding to a loop surrounded by sequences that self-complement and form an A-form helix. In this case, an optimization of the sequence and the length of the stem structure can increase its thermal melting temperature and lead to a better stability of the complex. For example, initial NMR studies of the complex between the protein RsmE (regulator of secondary metabolism E) and its 12-nucleotide natural RNA target showed that this RNA sequence does not form a free stable stem-loop structure in solution [34]. Therefore, the RNA sequence was extended by four G–C base-pairs in order to enforce an initial stable stem-loop structure [34]. In cases where the protein is known to bind the stem but not the loop, the loop sequence was also often optimized to stabilize the fold of the RNA. This strategy has been used to solve the NMR structure of the L25–5S rRNA complex [36]. In this case, it was known that the protein L25 binds an internal loop within a stem region but not the apical loop [52]. The natural apical loop was therefore replaced by a highly stable UUCG tetraloop in order to stabilize the structure of the RNA [36]. In another study, the third double-stranded RNA binding domain (dsRBD3) of the protein Staufen was assumed to bind double-stranded RNA in a non-specific manner. Therefore, an optimal RNA sequence containing a highly stable UUCG tetraloop was designed in order to solve the structure of the complex [32]. Quite

surprisingly, contacts of the dsRBD to the loop were then observed in the structure.

In the case of RNA binding proteins or domains that specifically bind single-stranded RNA, different lengths of RNA can be tested. Although the consensus sequence is often only a few nucleotides long, flanking nucleotides that are not specifically recognized, 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 length. The length of the RNA can also, in certain cases, modify the NMR spectral quality of a protein–RNA complex. For example, in the NMR study of the complex between the RRM of PTB and their CUCUCU RNA target, the RNA was bound by the RRM in two registers due to the fact that the RNA contains two UCU motifs leading to line broadening in the RNA resonances and two sets of intermolecular NOEs that could not be accommodated by a single structural model [28]. 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 [72]. Furthermore, consensus RNA sequences derived from SELEX experiments are often degenerate. Therefore, different sequences should be tested to optimize the NMR spectral quality. For example, SELEX experiments using the protein SRp20 defined a consensus sequence (A/U)C(A/U)(A/U)C [66]. To solve the NMR structure of SRp20 in complex with RNA, a total of 13 different RNA sequences were tested by NMR ^{15}N - ^1H HSQC for the protein and ^1H - ^1H -TOCSY (Total Correlation Spectroscopy) for the RNA [17]. This analysis identified the RNA sequence CAUC as the optimal sequence for obtaining good NMR spectral quality for both components of the complex and was therefore used for the structure determination of the complex [17]. We have often observed that a single nucleotide modification directly at the binding interface with the protein can improve the spectral quality of the complex, without necessarily modifying the affinity of the complex. For example, in the case of the protein hnRNP F that specifically binds G-tract RNA [73], we observed many intermolecular NOEs between the RRM of hnRNP F and the sequence AGGGAU that could not be observed with the sequence CGGGAU, even though this first nucleotide is not specifically recognized by the protein, possibly because purines have better stacking properties than pyrimidines (Fig. 3) [14].

2.2. Protein production

NMR is an insensitive technique that requires milligram amounts of macromolecules leading to NMR samples at high concentration (optimally in the millimolar range). RNA binding proteins or domains must therefore be obtained in large amounts with the possibility of using isotope labeling. Many reviews have addressed this issue in detail (see for example [74,75]). Here, we will briefly discuss the main steps and issues that have been used in the case of RNA binding proteins or domains.

2.2.1. Domain boundaries

Most RNA-binding proteins consist of modules called RNA binding domains together with other modules specific to their functions (protein–protein interaction module, catalytic module, etc.). Due to the size limitation for successful NMR and the fact that RBDs are often sufficient for the RNA binding ability, it is common to use a modular approach consisting of studying isolated RBDs in complex with RNA instead of the full-length protein. There are many known RBDs, such as the RRM, the dsRBD, or the KH domain (for reviews, see [76,77]). In order to study these domains in isolation, a crucial step is to identify the boundaries of the domain, that is, defining the minimum protein sequence that can adopt a proper folded structure and that is able to bind RNA.

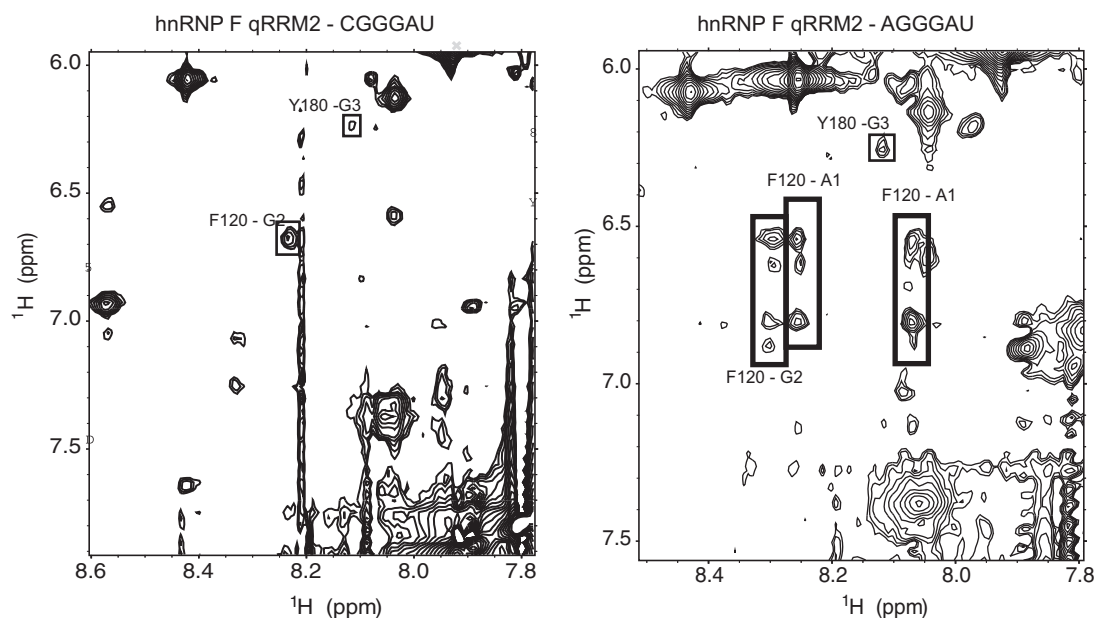


Fig. 3. The RNA sequence can influence the quality of NOESY spectra. 2D NOESY spectra at 900 MHz of hnRNP F qRRM2 in complex with a CGGGAU (left) or an AGGGAU (right) RNA recorded in D_2O in the same buffer and temperature [14]. Intermolecular NOEs are boxed and labeled.

If the structure of such a domain is not known, bioinformatics tools can be used to identify the domain boundaries. The most frequently used tools are the search of amino acid conservation through multiple sequence alignment, the prediction of secondary structures and the analysis of the hydrophobicity profile of the amino acid sequence. Currently, many methods to predict domain boundaries are available (for more details, refer to [78]). The accuracy of the predictions highly depends on the prior knowledge of the domain under investigation. Another approach for defining domain boundaries consists of creating numerous protein constructs with different length and then testing of their RNA binding properties using an RNA binding assay such as EMSA, isothermal titration calorimetry (ITC), surface plasmon resonance (SPR), or NMR titration experiments.

Previously solved structures of similar RNA binding domains can be very helpful for defining the domain boundaries using sequence alignments tools between the protein of interest and the protein for which the structure was solved. However, in some cases, it was shown that small RNA binding domains contain additional structural features that are important for the proper folding of the domain or for the RNA binding capability that could not be predicted based on previously solved structures. The RRM is composed of a well-known $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$ fold that could be used to predict the domain boundaries for NMR studies of novel RRMs. However, it was later shown that RRMs can contain supplementary secondary structures that are important for proper folding or RNA binding. For example, the structure of the U1A RRM in complex with RNA showed that this domain possesses an additional helix α_3 at its C-terminus that interacts with the β -sheet surface of the RRM in its free form but rotates away from the β -sheet when the RRM is in complex with RNA [2]. Although this helix is not directly involved in RNA binding, it was proposed that its repositioning allows the formation of a hydrophobic core that can stabilize the domain. Later, the NMR structure of U1A RRM in complex with a natural RNA containing two U1A binding sites showed that this additional helix induces the dimerization of two U1A RRMs and is crucial for stabilizing the ternary complex [38]. Similarly, the protein PTB contains four RRM domains. NMR structures of these domains both free and in complex with RNA showed that two RRMs of PTB (RRM2 and 3) possess an additional β -strand (β_5) at

their C-terminus extending the β -sheet surface [28,79,80]. Interestingly, this additional β -strand in PTB RRM3 is involved in RNA recognition allowing the binding of two additional nucleotides [28]. The dsRBD adopts a canonical $\alpha\beta\beta\alpha$ fold [32,81,82]. The NMR structure of the dsRBD of the protein Rnt1p in complex with RNA identified an additional α -helix at the C-terminus that is important for stabilizing the fold of Rnt1p dsRBD [39,83]. Therefore, the definition of domain boundaries based on previously solved structures of similar domains might lead to truncated domains that are insoluble or do not bind RNA. In these cases, bioinformatic approaches and RNA binding assays can be used.

2.2.2. Cloning and expression of RNA binding proteins

In the cases of small RNA binding peptides, chemical synthesis has been used to obtain high amounts of peptides [8,11,23,31,41,43]. This method, however, makes obtaining isotopically labeled peptides rather expensive. Therefore, the most common way of producing peptides and proteins for NMR studies is the use of recombinant DNA technology and bacterial expression. For NMR purposes, this strategy has a major advantage because it allows the production of isotopically labeled molecules by growing bacteria in a medium free of natural nitrogen and carbon sources but supplemented with isotopically labeled chemical compounds (generally, ^{15}N labeled ammonium chloride, and $^{13}C_6$ -glucose). In addition, it is possible to obtain deuterated proteins by growing bacteria in a medium containing D_2O instead of H_2O .

When expressing proteins in bacteria, especially eukaryotic proteins, one needs to take account of certain aspects in order to obtain proteins in high quantities and in soluble form. The codon usage in eukaryotes and prokaryotes is different. Prokaryotes contain less transfer RNAs (tRNAs) than eukaryotes and therefore do not recognize all codons. These specific codons are termed “rare codons” and can be identified using bioinformatic tools. To overcome this problem, two main approaches have been used. First, site-directed mutagenesis can be used to replace rare codons by usual codons that encode the same amino acid but that are recognized by the bacterial translation machinery. This strategy was applied for determining the structure of the Moloney Murine Leukemia virus NC protein in complex with different RNA targets [10,13]. Another approach is to supplement bacterial strains with

tRNAs that recognize these rare codons. Currently, there are commercially available bacterial strains that contain the genes of tRNAs that are specific for rare codons. This second approach is used most frequently and often leads to a significant improvement of the protein expression yield. Many RNA binding proteins have been expressed in such bacterial strains [17,29,34,39,45].

Protein solubility is another important issue that can cause problems. It is not unusual that eukaryotic proteins become insoluble when expressed in isolation in bacteria. This problem can be due to many reasons. Most proteins are never isolated in cells but always in complex with their partner, such as RNA for RNA binding proteins. Consequently, the isolation of the protein might expose hydrophobic residues that are otherwise buried in the complex. Some proteins can also be harmful for the bacteria and therefore, as a defense mechanism, bacteria often incorporate the overexpressed protein into inclusion bodies, an aggregation of proteins. There are many ways to improve the solubility of a protein during bacterial expression. The simplest approach is to grow bacteria at lower temperature, which slows down the production of proteins but often leads to a higher amount of soluble proteins. Another approach is to express the protein of interest as a fusion with a solubility enhancement tag (SET). Many SETs have been described and consist of very soluble protein domains (for more details, see [84]). For example, two commonly used SETs are the maltose binding protein (MBP) and the streptococcal B1 immunoglobulin-binding domain of protein G (GB1) [85]. The DNA sequence coding for the SET is inserted in the plasmid between the promoter and the DNA coding for the protein of interest resulting in the expression of a fusion protein where the SET and the protein of interest are separated by a short linker. These tags are highly soluble and in many cases, fusion proteins containing SETs become soluble during bacterial expression and purification. A GB1 tag has been used, for example, to obtain a soluble fraction of the protein SRp20 that was otherwise highly insoluble [17]. The drawback of using solubility enhancement tags is that, since they often consist of protein domains, additional NMR signals are present in the NMR spectra, adding to the complexity of spectral analysis [85]. To circumvent this problem, a protease cleavage site can be inserted between the SET and the protein of interest allowing the separation of the protein from the tag. However, removing the SET often results in a loss of solubility of the protein of interest [17]. An alternative is to use a segmental isotope labeling approach [86] that makes use of the properties of inteins to ligate two polypeptide chains (for more details, see [87]). Using this technique, it was possible to attach an “NMR invisible” solubility enhancement tag to a protein of interest. Since the protein is not soluble without SET, the strategy adopted was to express the protein of interest as a fusion protein with a SET attached at its C-terminus, the two domains being linked by a protease cleavage site. Segmental isotope labeling was then used to ligate an unlabeled SET at the N-terminus of the isotopically labeled protein. This was then followed by cleaving the labeled C-terminus SET using a protease and separating the protein of interest fused to the unlabeled SET from the cleaved labeled SET by further purification steps [86].

Another method to obtain soluble proteins is to purify and refold insoluble proteins from inclusion bodies, using purification procedures under denaturing conditions followed by refolding steps. This method has been used to solubilize the RNA binding protein L25 [36]. Similarly, a protein construct containing the two N-terminal RRM domains of the protein nucleolin was purified under denaturing conditions and refolded at a later stage [1]. Another similar approach was used to express the protein TIS11d that contains two zinc-fingers [18]. In this case, the protein expression was induced by adding isopropyl β -D-thiogalactopyranoside (IPTG) supplemented with ZnSO₄. The protein was then purified under denaturing conditions, lyophilized in a non-denaturing buffer and

refolded by titrating ZnSO₄ and monitoring the refolding by circular dichroism [18].

An interesting example for the expression of RNA binding protein is the case of the protein Kid. Kid is a bacterial toxin that is harmful for bacteria. In order to obtain this protein in *Escherichia coli* (*E. coli*), the protein Kid was co-expressed with its partner, the antitoxin Kis protein that neutralizes the toxic effect of Kid. After cell lysis, the proteins were purified and separated [47].

2.2.3. The cell-free *in vitro* system

The bacterial expression system is the simplest system and the most common way used to produce proteins in large quantities. However, certain proteins or domains cannot easily be obtained soluble and in high amount using this system. An alternative of the bacterial expression is to use a cell-free *in vitro* system that was optimized for obtaining high amounts of proteins by Yokoyama and coworkers [88,89]. Detailed descriptions and advantages of this method have been previously reviewed (for example, see [90,91]). This *in vitro* method of protein production uses a coupled transcription/translation reaction. As for *in vivo* expression using *E. coli*, the DNA encoding the protein of interest is cloned into an expression vector. In this case, however, the reaction consists of mixing the plasmid DNA with a cell extract (generally from *E. coli*), a suitable RNA polymerase, magnesium chloride, creatine phosphate, creatine kinase, the four different nucleotides, and the 20 different amino acids. The use of cell-free *in vitro* protein synthesis, often allows the expression of soluble proteins that are either not expressed or are insoluble in the *E. coli* expression system. Another advantage of this method is that it allows for specific labeling of certain amino acids of the protein. This can be very advantageous in cases of large protein–RNA complexes. The use of cell-free *in vitro* protein synthesis was used very recently to help solving the NMR structure of the third RRM domain of the CUG binding protein 1 (CUG-BP1) in complex with RNA [37].

2.2.4. Purification of RNA binding proteins

Purification strategies for RNA binding proteins depend on the properties of the proteins such as their isoelectric point and their molecular weight. Many RNA binding proteins have been purified using a combination of ion exchange chromatography and size-exclusion chromatography [2,12,18,22,36,46].

Alternatively, purification tags have commonly been used in the purification of RNA binding proteins from the bacterial proteome. In this case, the DNA coding for the protein of interest was cloned into a plasmid containing an N-terminal or a C-terminal purification tag. The most commonly used tags have been the poly-histidine tag [1,5–7,9,15–17,19,25,26,28,29,33–35,37,42,44,47] and the glutathione S-transferase (GST) tag [10,13,21,24,32,39,40,45]. The fusion protein overexpressed in *E. coli* could then be purified by affinity chromatography using resins containing covalently bound metal ions (nickel or cobalt) in the case of poly-histidine tags or glutathione affinity matrix in the case of GST tags. After affinity purification, fairly pure fractions of fusion proteins are generally obtained. Because the poly-histidine tag is rather small (a few amino acids), it is not crucial to cleave off the tag as long as the tag does not influence the RNA binding properties of the protein [1,5,17,28,29,34,35]. Alternatively, the affinity purification tag can be separated from the protein of interest. In the case of small peptides, removal of the affinity purification tag was often achieved using cyanogen bromide that specifically hydrolyzes peptide bonds C-terminal of methionines [7,9,15,16,19,33,42,44]. In the case of protein domains that often contain internal methionines, other methods for cleavage are generally used and involve specific proteases. Three proteases have been mainly used in the case of RNA binding proteins. These are thrombin that specifically recognizes a Leu-Val-Pro-Arg-Gly-Ser segment and cleaves the

peptide bond between the arginine and the glycine [24,30,39,40], the tobacco etch virus (Tev) protease that specifically recognizes a Glu-Asn-Leu-Tyr-Phe-Gln-Gly and cleaves between the glutamine and the glycine [21,25,26,37], and the PreScission protease that specifically recognizes a Leu-Glu-Val-Leu-Phe-Gln-Gly-Pro segment and cleaves between the glutamine and the glycine [10,13,45]. In these cases, during cloning, the DNA sequence encoding for a thrombin, a Tev or a PreScission cleavage site was inserted between the fused tag and the peptide of interest. After cleavage, the protein of interest was then separated from the tag using again an affinity chromatography (in this case, only the tag is retained by the resin) or further purification steps such as ion exchange or size-exclusion chromatography. Finally, the protein can be dialyzed against a buffer suitable for NMR analysis and concentrated.

2.2.5. Ribonuclease activity

Since the protein of interest will be studied in complex with RNA, it is very important, before mixing the protein and the RNA, to test whether the protein sample possesses RNase activity. The effect of RNases is highly dependent on the RNA sequence and structure used in the study, the protective effect of the protein for the RNA, and the affinity of the protein for the RNA. Generally, structured RNAs, such as those forming stem-loop structures, are less prone to degradation than small single-stranded RNAs. Furthermore, single-stranded RNAs 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.

Bacteria used to overexpress the protein of interest contain many RNases that sometimes cannot be removed by protein purifications. For example, it was reported that for the NMR study of the complex between the protein LicT and its target RNA, RNase activity could not be eliminated and NMR samples were therefore stable for only a few days in the NMR spectrometer [40]. RNase inhibitors can be added into the final buffer to slow down the degradation of the RNA during NMR measurements [21,29]. Our experience also showed that additional protein purification steps are sometimes very effective for eliminating the RNase activity of the sample. For example, in the NMR study of the SRp20-RNA complex, three consecutive NiNTA purification steps were necessary in order to eliminate RNase activity [17]. 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 NMR. In this case, the formation of degradation products over time can be assessed using 2D ^1H - ^1H -TOCSY spectra (generally by following the pyrimidines H5-H6 cross-peaks).

2.3. RNA production

2.3.1. RNA synthesis

RNA can be synthesized in three different ways depending on the length and the requirements for isotope labeling: chemical synthesis, *in vitro* enzymatic transcription and *in vivo* production of RNA (see Fig. 4A). Chemical synthesis is the method of choice for preparing small RNAs, as *in vitro* enzymatic synthesis of RNAs smaller than 10 nucleotides has been reported not to be successful [18], except in one case [30]. Chemical synthesis of RNA is reported for RNAs up to 80 nucleotides [92–94]. However, low yields and high costs for larger RNAs make the chemical synthesis suitable only for short RNAs (<20nt). A unique advantage of chemical synthesis is the possibility of introducing modified nucleotides at desired positions. For example, introduction of thiouridines at specific positions in a RNA allows the attachment of nitroxide spin-labels for measuring paramagnetic relaxation enhancement (PRE) [95]. Also a protocol for synthesizing short RNAs, that are

selectively ^{13}C labeled on sugar carbons has been developed [96] and used to solve the structure of several protein–RNA complexes [5,14,28,29,35], but unfortunately isotope labeled phosphoramidites are not yet commercially available.

In vitro enzymatic transcription using SP6, T3 or T7 RNA polymerases is the most widely used method for the production of RNAs larger than 12 nucleotides [97–100]. The possibility of incorporating commercially available ^{13}C , ^{15}N or even partially ^2H labeled nucleoside triphosphates (NTPs) allows production of RNAs suitable for heteronuclear multidimensional NMR [101–105]. Wijmenga and coworkers produced enzymatically NTPs, which are stereo-specifically deuterated on the 1', 3', 4' and 5'' positions and ^{13}C labeled on all sugar positions. This approach allowed for the stereo-specific assignment of the H5' resonances, a reduction of spectral crowding and resulted in line narrowing compared with spectra of ^{13}C labeled non-deuterated RNA [103].

T7 RNA polymerase can be obtained commercially or produced in-house by overexpressing a His-tagged T7 RNA polymerase in *E. coli* [106]. Transcription reactions must be first optimized on small scale reactions by changing concentrations of MgCl_2 , DNA, NTPs and T7 RNA polymerase and testing the influence of the addition of pyrophosphatase and/or guanine monophosphate (GMP). The best condition can be scaled-up to a large scale reaction of for example 10 ml, which yields typically around 500 nmol of RNA. Transcription using T7 RNA polymerase can be performed from chemically synthesized double-stranded DNA templates or from linearized plasmids. Since only the 18 nt T7 promoter on the top-strand is sufficient for transcription, the same top-strand can be used for any transcription. However, it has been observed that higher yields are obtained when fully double-stranded DNA is used [107]. The first nucleotide, which is incorporated, must be a guanine. Transcription efficiency is highly dependent on the starting six nucleotides. Excellent starting sequences are GGGAGA, GGGAUC, GGCAAC or GGCGCU [99]. Besides the 5' sequence requirements, another drawback of T7 *in vitro* transcription is the 3' and 5' inhomogeneity. More than 30% of untemplated 5' nucleotides have been observed for sequences starting with 4–5 consecutive guanines, whereas a sequence starting with GCG showed no detectable 5' inhomogeneity [108]. More severe is the 3' inhomogeneity, where up to 6 additional nucleotides can be added. An overview of several methods to overcome 5' and 3' inhomogeneity is presented in Fig. 4B.

The problem of 3' and 5' inhomogeneity can be circumvented by incorporation of a ribozyme sequence in *cis*, which cleaves co-transcriptionally leading to an homogenous 5'-hydroxyl or a 2',3'-cyclic phosphate end [105,109]. Concerning 5'-inhomogeneity, hammerhead ribozymes are interesting because they have no sequence requirements [100]. When placed 5' to the RNA of interest, they allow cleavage of the RNA with MgCl_2 as cofactor almost to completion [100,110]. Concerning 3'-inhomogeneity, the hepatitis delta virus (HDV) RNA ribozyme, that has no sequence requirements [111,112], or the *Neurospora* Varkud satellite (VS) ribozyme that has minimal sequence requirements (VS will cut efficiently after any nucleotide other than cytosine) can be efficiently used [110,113]. It has been shown that hammerhead ribozymes [114] and VS ribozymes [110] can be added in *trans* saving isotope labeled NTPs that otherwise would be used to produce the ribozyme incorporated in *cis*.

In addition to ribozymes, DNAzymes have been developed by *in vitro* evolution as engineering tools [115–117]. The 10–23 family of DNAzymes cleaves between a purine and a pyrimidine, which is the only sequence requirement. Cleavage results in a 5'-hydroxyl group and a 2',3'-cyclic phosphate similarly to small ribozymes. Moreover, it has been shown that RNAs can be cleaved sequence-specifically by RNase H, when the RNA of interest is hybridized with a 2'-O-methyl-RNA/DNA chimera [118,119]. In contrast to ribozyme and DNAzyme mediated RNA cleavage, RNase H pro-

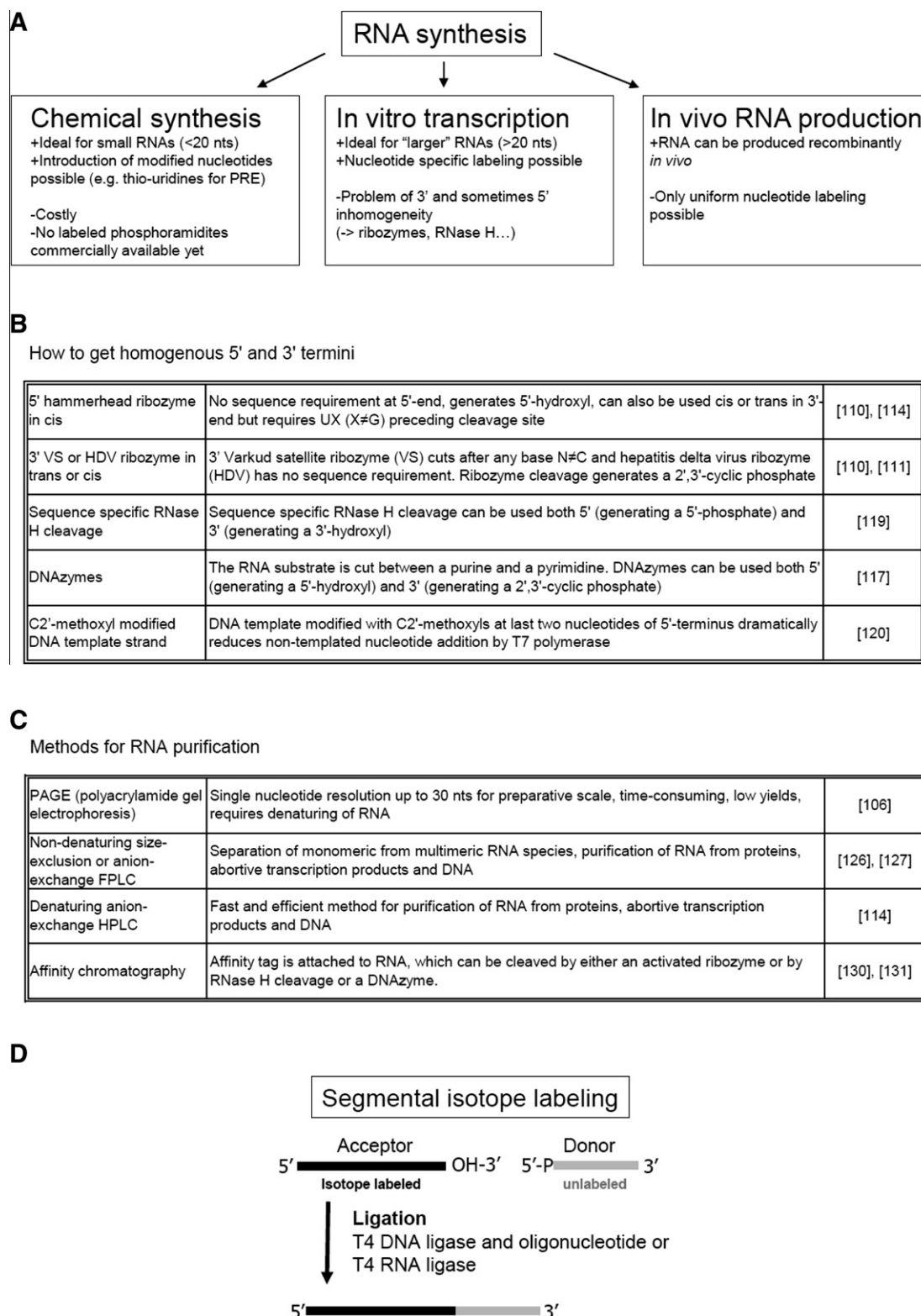


Fig. 4. RNA production. (A) RNA synthesis can be performed either by chemical synthesis, *in vitro* transcription or *in vivo* RNA production. (B) Methods for generating homogenous 5'- and 3'-termini. (C) Methods for preparative RNA purification. (D) Segmental isotope labeling of RNA: for proper ligation, a 3'-hydroxyl on the acceptor fragment and a 5'-monophosphate on the donor fragment are required. Ligation can be performed with either T4 DNA ligase and a DNA or 2'-O-methyl RNA/DNA chimera or with T4 RNA ligase.

duces 5'-monophosphates and 3'-hydroxyl groups. Another approach is the use of a DNA template strand for transcription, in which the two 5' nucleotides are modified with C2'-methoxyls. This dramatically reduced 3'-end inhomogeneities [120].

A third method used to produce RNA for NMR studies was recently developed by Dardel and coworkers by producing RNA *in vivo* using a tRNA scaffold to protect the RNA from cellular RNases [121,122]. The tRNA scaffold can be removed either by DNA-

zymes or by sequence-specific RNase H cleavage [115,117–119]. Using this method, a reasonable amount of RNA for NMR studies (0.8 μ moles) was obtained from 2 l of *E. coli* culture grown in $^{15}\text{N}/^{13}\text{C}$ labeled medium.

2.3.2. RNA purification

RNA obtained by *in vitro* enzymatic transcription or *in vivo* must be purified from proteins (T7 RNA polymerase, pyrophosphatase) and abortive transcription products (a large number of smaller oligoribonucleotides of 2–6 nucleotides in length are generated during transcription due to abortive initiation events) as well as from unused NTPs. In addition, RNAs with one or two additional nucleotides arising from untemplated nucleotide addition must be removed, when a homogenous RNA is required. An overview of different purification methods is presented in Fig. 4C.

The most commonly used purification method for large quantities of RNA needed for NMR spectroscopy is denaturing polyacrylamide gel electrophoresis (PAGE). Single nucleotide resolution for preparative scales are typically achieved for RNAs up to 30 nucleotides. However, this procedure is laborious and suffers from low recovery yields. Additionally, PAGE requires the RNA to be denatured and refolded, which might lead to aggregation and dimerization of the RNA [123]. Furthermore, the RNA is not free of low-molecular-weight acrylamide contaminants, which might interact with RNA and also compromise NMR spectral analysis [124]. Therefore, different chromatographic methods have been developed to purify RNA. Frederick and coworkers proposed purifying RNA by non-denaturing anion-exchange chromatography [125]. Depending on the salt type of the elution buffer (NaCl, CsCl or MgCl_2) they could separate RNAs with different conformations. Very recently, Lukavsky and coworkers showed that weak anion-exchange fast protein liquid chromatography (FPLC) under non-denaturing conditions can be used to separate the desired RNA product from the T7 RNA polymerase, unincorporated NTPs, small abortive transcripts and the plasmid DNA template [126]. Rapid purification of homogeneous RNAs can be achieved by using trans-acting hammerhead ribozymes in combination with anion-exchange high performance liquid chromatography (HPLC) at high temperature (90 °C) [114]. Certain biologically relevant RNAs might fold into different conformations or might form multimers, which can be separated by purifying them under non-denaturing conditions using size-exclusion FPLC [123,124,127]. In addition to reverse-phase HPLC [128], the use of affinity chromatography has been described [129–132]. Batey and Kieft developed a sophisticated approach, where an affinity tag is attached to the 3'-end of the RNA by a glucosamine-6-phosphate activated (*glmS*) ribozyme [130]. The affinity tag is based on two RNA stem-loops having high affinity for the MS2 coat protein fused to a 6xHis-tagged MBP, which binds to a Ni^{2+} -affinity column. Elution of the RNA can be achieved by activating the ribozyme with addition of GlcN6P. Affinity purification based on aptamer tags binding Sephadex or Streptavidin have also been proposed [122,131].

Depending on the purification method, the RNA can either be eluted directly into NMR buffer or needs to be exchanged into a suitable buffer. Buffer exchange can be performed by dialysis or by washing and concentrating the RNA with ultracentrifugation with an appropriate molecular weight cut-off (MWCO). Dialysis bags and ultracentrifugation filter devices with 1000 MWCO are commercially available and are appropriate for RNAs produced by *in vitro* transcription. The RNAs can be lyophilized and subsequently resuspended into NMR buffer. Typical NMR buffers for RNA are 10–50 mM sodium phosphate at pH = 5.5–6.

2.3.3. Segmental isotope labeling for larger RNAs

In recent years, NMR methodology has been developed to study larger macromolecular systems but increasing relaxation rates and

spectral overlap for larger biologically relevant RNAs make them difficult to be studied by NMR without specific isotope labeling of the RNA. One can for example ligate a small isotopically labeled fragment produced by chemical synthesis or *in vitro* transcription to a larger unlabeled fragment. Using this approach, Puglisi and coworkers could show that a small RNA adopts the same structure in isolation as that found in the context of the entire 100 kDa natural RNA [133,134].

Two excellent reviews were recently published on isotope labeling strategies of RNA [105,109]. Here, we will only briefly discuss the different approaches available. RNA ligation can be performed by T4 RNA or T4 DNA ligase [135–137] or by using a deoxyribozyme that catalyzes RNA ligation [138]. Both T4 RNA and T4 DNA ligases require a 5'-monophosphate on the donor fragment and 3'-hydroxyl on the acceptor fragment at the site of ligation (see Fig. 4D), whereas the deoxyribozyme catalyzes a ligation reaction with a 5'-triphosphate on the donor fragment with a 3'-hydroxyl on the acceptor fragment. Ligation with DNA ligase, which recognizes a nicked double-stranded substrate, is performed by annealing a DNA oligonucleotide or a 2'-*O*-methyl-RNA/DNA chimera to the site of ligation [137]. Unlike DNA ligase, RNA ligase requires a single-stranded site of ligation. Preferentially, the acceptor and the donor are brought together by base-pairing such that the site of ligation is in a hairpin loop [136,139]. However, it has been shown that RNA ligase can also be used in combination with a DNA oligonucleotide annealing with the site of ligation designed to mimic the natural substrate of RNA ligase [140]. To prevent self-ligation or ligation of the fragments in the improper sequential order (especially using T4 RNA ligase), the acceptor fragment should contain a hydroxyl group both at its 5' and 3' ends, whereas the donor fragment should have a monophosphate at the ligation site and a monophosphate or a 2',3'-cyclic phosphate at the 3'-end. As discussed in Section 3.3.1, RNAs obtained by *in vitro* transcription contain a 5'-terminus with a tri-phosphate and an inhomogenous 3'-hydroxyl terminus. Ribozymes engineered at the 3'-end producing homogenous 2',3'-cyclic phosphates can thus be used to generate 3'-ends of both acceptor and donor fragment, whereupon the acceptor 3'-end has to be further dephosphorylated using T4 polynucleotide kinase, which has 3'-phosphatase activity [105,141,142]. Hammerhead ribozymes located 5' to the acceptor fragment generate the correct 5'-hydroxyl end, whereas the 5' donor end generated by a hammerhead ribozyme has to be phosphorylated by T4 polynucleotide kinase (PNK). To generate two samples, in which only one segment is labeled, two unlabeled and two labeled transcription reactions have to be performed. In addition, the use of T4 PNK is an additional costly step, which also requires an additional purification. Two elegant approaches requiring only one labeled and one unlabeled transcription reaction have been proposed [134,143–145]. Lukavsky and coworkers used a plasmid encoding the 3' donor fragment followed by a hammerhead ribozyme, which is connected by a flexible linker to a second hammerhead ribozyme preceding the 5' acceptor fragment yielding a terminal 3'-hydroxyl after transcription [134,143,144]. If the transcription reaction is primed with GMP, both fragments are correctly protected for ligation with T4 RNA ligase, the 5' acceptor fragment being protected both 5' and 3' with hydroxyl groups and the 3' donor fragment being protected by a 5'-phosphate and a 2',3'-cyclic phosphate. The only drawback of this method is that a G is required at the 3' of the ligation site and that transcription can potentially generate an inhomogenous 3'-end of the acceptor fragment leading to possible incorporation of additional nucleotides at the site of ligation, especially when using RNA ligase. A second approach described by Crothers and coworkers showed that sequence specific RNase H cleavage of an unlabeled and a labeled RNA can be followed by direct cross re-ligation of a labeled with an unlabeled fragment using T4 DNA ligase [145].

Finally, two groups presented another approach that combines the use of both T4 RNA and T4 DNA ligase in order to obtain multiple segmental isotopically labeled RNAs (i.e: three RNA fragments ligation) [139,146,147].

Segmental isotope labeling of RNA is expected to become increasingly important for the study of larger RNAs in isolation or in complex with proteins, especially in combination with measurements of residual dipolar couplings and paramagnetic relaxation enhancement (see Section 3.4).

2.4. Complex formation

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.

2.4.1. Estimation of the protein and RNA concentrations

Before mixing the protein and the RNA, it is important to measure the concentration of both components since this information will allow accessing the stoichiometry of the complex. Generally, protein and RNA concentrations are estimated by measuring the optical density (OD) at 280 and 260 nm, respectively. In that case, the molar extinction coefficients of the protein and the RNA are estimated, generally using bioinformatic tools that accurately predict the extinction coefficient of unfolded proteins or nucleic acids based on their primary sequences. Therefore, protein and nucleic acid concentrations can be derived from measuring the OD under denaturing conditions (high temperature or in presence of denaturing agents such as guanidinium chloride). This method works well for measuring the concentration of nucleic acids. In proteins, however, only tryptophan and tyrosine are significantly contributing to the extinction coefficient. Therefore, concentration of proteins can only be estimated by optical density if a certain amount of these two residues are present in the protein. Otherwise, other methods to determine the protein concentration must be used, such as measuring the difference of absorbance of proteins at 215 and 225 nm [148] or using colorimetric assays [149]. NMR spectroscopy is another alternative method for determining protein concentrations using the PULCON method that has been developed by Wider and coworkers [150].

2.4.2. Monitoring complex formation by NMR titration experiments

Complex formation can easily be monitored by NMR spectroscopy using the so-called chemical shift perturbation mapping or titration experiments. The chemical shift of a nucleus is highly dependent on its chemical environment and can be accurately measured. Changes in the environment of the nucleus results in changes of its chemical shift and these changes can be measured to identify the interface of a macromolecular complex (for reviews on this topic, see [151,152]). There are many possible ways of following the complex formation of a protein–RNA complex by NMR. The complex can be formed by adding the protein into the RNA or vice versa. Furthermore, chemical shifts of different nuclei can be used. Chemical shift perturbations of the protein can be followed by measuring ^{15}N – ^1H HSQC spectra in absence or in presence of increasing amounts of unlabeled RNA. Similarly, RNA chemical shift perturbations can be followed by adding increasing amount of protein into the RNA. If the RNA is unlabeled, ^1H chemical shift perturbations of the H5–H6 cross-peaks of pyrimidines can be monitored using 2D ^1H – ^1H TOCSY spectra. Alternatively, chemical shift perturbations of imino protons can be monitored by 1D NMR if they are visible. If ^{15}N - and/or ^{13}C labeled RNA samples are available, chemical shift perturbations of imino or non-exchangeable protons can be followed using ^{15}N – ^1H HSQC or ^{13}C – ^1H HSQC spectra, respectively.

Initial chemical shift perturbation experiments are typically performed by adding the RNA solution to the protein solution. Nonetheless, the addition of RNA into proteins can sometimes lead to precipitation of the sample and this phenomenon can be irreversible. In this case, the addition of the protein into RNA can sometimes avoid the precipitation. In addition, some RNA binding proteins are not very soluble and cannot be concentrated in absence of RNA. In this case, chemical shift perturbation experiments can be performed at low concentration since 2D NMR experiments, such as ^{15}N – ^1H HSQC or ^1H – ^1H TOCSY, are highly sensitive and the protein–RNA complex can be subsequently concentrated by ultrafiltration using an appropriate molecular weight cut-off membrane [18].

When the protein–RNA complex is formed, the protein and the RNA are in equilibrium between their free and bound states. This equilibrium is mainly described by the affinity of both components in the complex, which is expressed as 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 mainly 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 occur when k_{ex} is much smaller than $2\pi(\nu_A - \nu_B)$, the fast exchange regime when k_{ex} is much larger than $2\pi(\nu_A - \nu_B)$, and the intermediate exchange regime when k_{ex} is similar to $2\pi(\nu_A - \nu_B)$. The exchange regime governs the behavior of the NMR signals during the titration experiments.

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 [5] (Fig. 5). 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. Therefore, while gradually adding the RNA, the signal of the protein corresponding to the free state decreases and the signal corresponding to the bound state increases. Slow exchange regimes were reported for protein–RNA complexes with high affinity corresponding to dissociation constants ranging from 0.5 (Fox-1–UGCAUGU) [5] to 250 nM (protein NC–AACAGU) [13].

In the fast exchange regime, only one NMR signal is visible and corresponds to the weighted average of the signals corresponding to 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 [28] (Fig. 6). When further addition of RNA no longer affects the chemical shift position of the signal, the NMR signal corresponds to that of the bound state. Fast exchange regimes were reported for protein–RNA complexes with dissociation constants higher than 15–20 μM (PTB–CUCU [28,72] or SRp20–CAUC [17]).

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 as the stoichiometry of the complex is about to be reached, as was observed for the Staufen–RNA complex [32] (Fig. 7). Intermediate exchange regimes were reported for protein–RNA complexes with dissociation constants ranging from 400 nM (hnRNP F–AGGGAU) [14] to 2 μM (CUG–BP1 RRM3–CUGCUG) [37]. However, in some cases, signals of the bound state do not sharpen and are therefore invisible even in ex-

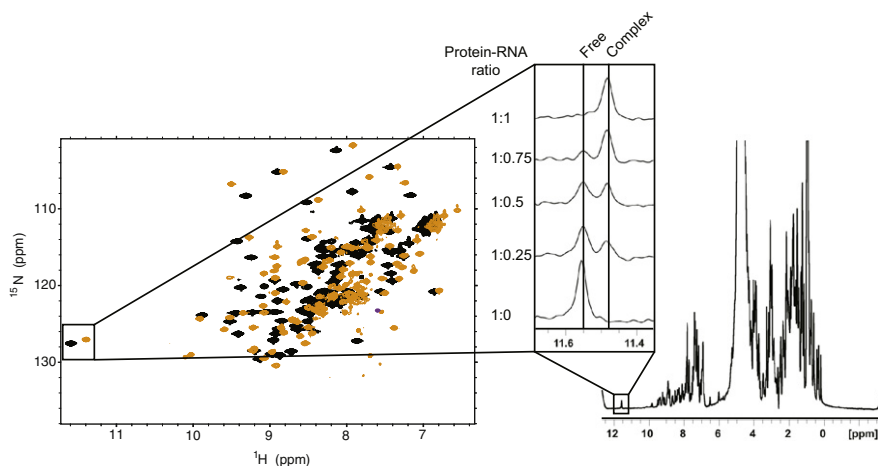


Fig. 5. Example of a protein–RNA complex in the slow exchange regime. NMR spectra at 900 MHz of the Fox-1–RNA complex [5]. Left: ^{15}N – ^1H HSQC spectra of Fox-1 free (black) or in complex with a UGCAUGU RNA at a 1:1 molar ratio (orange). Right: 1D spectrum of Fox-1 free. The signal at 11.5 ppm is shown as a function of the protein:RNA molar ratio. The complex formation is in slow exchange on the NMR time scale. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

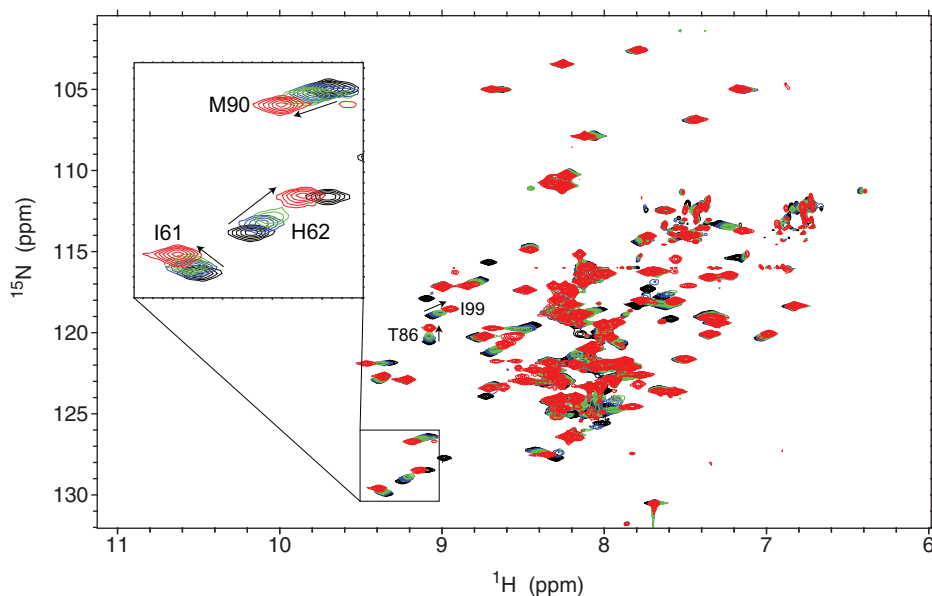


Fig. 6. Example of a protein–RNA complex in the fast exchange regime. NMR spectra at 600 MHz of the PTB–RNA complex [28]. ^{15}N – ^1H HSQC spectra of PTB RRM1 free (black) or in complex with CUCUCU at a protein:RNA molar ratio of 1:0.2 (blue), 1:0.4 (green), and 1:3 (red). The complex formation is in fast exchange on the NMR time scale. Reproduced with permission from Ref. [72]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

cess of the partner component [37,153]. In this case, an optimization of the conditions should be performed in order to make the resonance of the complex visible (see Section 2.4.3) [14].

Therefore, in protein–RNA complexes, the affinity between the two molecules plays an important role in defining the exchange regime of the system. For many protein–RNA complexes studied by NMR, the affinities between the two partners have been measured and range from 0.5 nM [5] to about 20 μM [17]. For all protein–RNA complexes having an affinity below 250 nM, NMR titration experiments indicate a slow exchange regime [1,5,11,13,20,24,29,36,40], while for protein–RNA complexes having an affinity above 400 nM, the exchange regime is intermediate to fast [14,17,28,32,37].

Since the exchange regime depends on the magnitude of 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 titration experiments. Because the difference between the resonance frequencies 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 field strengths.

The exchange regime of the complex formation is an important parameter for consideration in the structure determination of a protein–RNA complex. Distance restraints extracted from Nuclear Overhauser Effect Spectroscopy (NOESY) spectra recorded on a complex might correspond to a weighted average of the free and the bound states. Furthermore, the intensities of intermolecular NOEs depend not only on the distance between the two atoms but also on the exchange rate between the free and the bound states. Therefore, for structural analysis of protein–RNA complexes, systems in slow exchange regime are generally most suitable. The main advantage is that the signals arising from the bound and the free states are

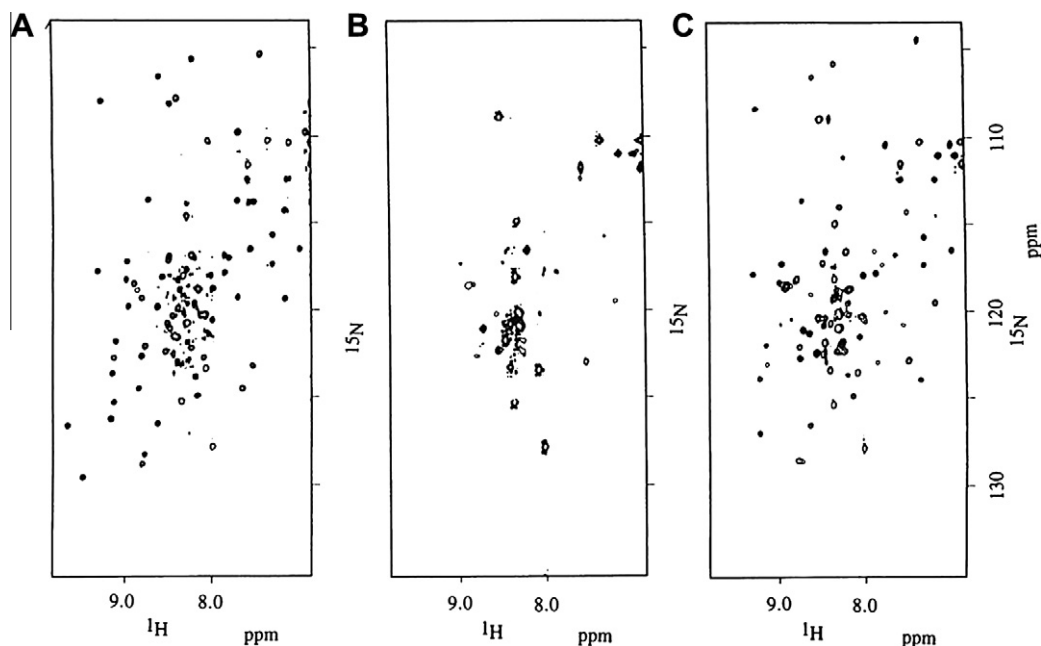


Fig. 7. Example of a protein–RNA complex in the intermediate exchange regime. NMR spectra at 600 MHz of the Staufen–RNA complex [32]. (A) ^{15}N – ^1H HSQC spectrum of free Staufen. (B) ^{15}N – ^1H HSQC spectrum of Staufen in complex with RNA at a protein:RNA molar ratio of 1:0.5. (C) ^{15}N – ^1H HSQC spectrum of Staufen in complex with RNA at a protein:RNA molar ratio of 1:1. Reproduced with permission from Ref. [32].

distinct. Therefore, distance restraints that reflect exclusively the bound state can be separated from those reflecting the free state. However, the exchange regime is not a crucial parameter since structure determination of protein–RNA complexes by NMR can also be performed for systems in intermediate and fast exchange.

In addition, other phenomena can sometimes add to the complexity of the structural analysis. For example, if the protein binds to the RNA in multiple registers, especially in the case of proteins binding single-stranded RNAs, additional exchange phenomena can arise. This occurs for example when the repetitive RNA sequence CUCUCU is bound by the protein PTB [56]. Such repetitive sequences provide a good affinity for the complex formation but lead to two or more complex sub-populations indicating that the protein binds the RNA in two or more different registers. In such cases, shortening the RNA results in a lower affinity for the complex formation but more importantly the protein binds the RNA in only one register. As a consequence, NMR line widths can become sharper and unambiguous intermolecular NOEs can be observed. This strategy of RNA shortening was successfully used to solve the NMR structure of PTB in its complex with a polypyrimidine-tract RNA [28].

2.4.3. NMR titration experiments are crucial steps for defining a “good” protein–RNA complex

Following the complex formation by NMR is a crucial step in the structure determination of a protein–RNA complex by NMR spectroscopy and provides many valuable insights into the complex formation. The main types of information that can be derived from these simple NMR experiments are the determination of the stoichiometry of the complex, the evaluation of the exchange regime, the quality of the NMR spectra, the identification of the binding interface, and an estimation of the binding constant.

NMR titration experiments can define the stoichiometry of the complex. Titration experiments are performed by adding one component into the other one. In the slow exchange regime, titration of one component into the other allows observation of new signals corresponding to the bound form and disappearance of the signals

corresponding to the free form. In the fast exchange regime, one component is added to the other one until no further chemical shift perturbations occur, which indicates a saturation of the complex formation. A plot of the chemical shift perturbations as a function of the protein:RNA ratio in the sample gives a good indication of the stoichiometry of the complex. For example, in the case where the protein forms a dimer, it was possible to determine if the protein dimer binds one or two RNA molecules. Two structures and one structural model of protein–RNA complexes involving protein dimers have been solved to date [34,40,47]. In two cases, one RNA molecule is bound by one dimer while in the other case, the dimer binds two RNA molecules. In this case, the stoichiometry of the complexes is different and could be assessed by NMR titration experiments. The Co-Antiterminator (CAT) domain is an RNA binding domain that folds into a symmetrical homodimer [154,155]. Upon RNA titration, the complex formation was in the slow exchange regime and most amide signals of the domain split into two components indicating that the symmetry of the dimer is broken and the maximum intensity of the bound signals was reached at a protein–RNA ratio of 2:1 indicating that one protein dimer bound one RNA molecule [40]. Similarly, the RNA binding protein Kid adopts a symmetrical homodimer fold [156]. In this case, the complex formation was in the fast exchange regime. NMR titration experiments indicated that one dimer binds a single RNA molecule and the stoichiometry of the complex was further confirmed by native mass spectrometry [47]. The RNA binding protein RsmE also folds into a symmetrical homodimer [157,158]. In this case, the complex formation was in the slow exchange regime and NMR titration experiments indicated a protein–RNA molar ratio of 1:1 resulting in two RNA molecules binding to one protein dimer [34]. In another NMR study, an RNA molecule contained two symmetrical binding sites for the protein U1A leading to a complex with a protein:RNA molar ratio of 2:1 [38]. Another example for which NMR titration can be very important for determining the stoichiometry of a protein–RNA complex is when the protein under study consists of two or more RNA binding domains. In this case, NMR titration experiments can show whether each RNA binding

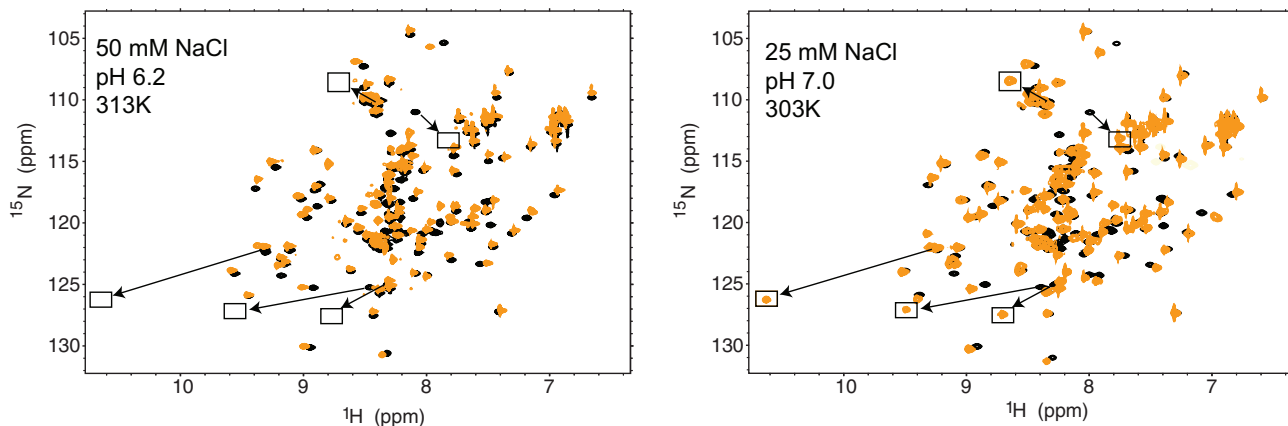


Fig. 8. Example of condition optimization using NMR chemical shift perturbation experiments. ^{15}N - ^1H HSQC spectra at 500 MHz of hnRNP F qRRM2 free (black) or bound to a CGGGAU RNA at a molar ratio of 1:1 (orange) [14]. The complex formation is in intermediate exchange. In the initial buffer conditions, many signals corresponding to the bound states are not visible (left spectrum). After optimization of the buffer and temperature conditions, all signals corresponding to the bound state are visible (squared signals in the right spectrum). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

domain binds an RNA molecule independently at a protein–RNA ratio of 1:2 or if both domains bind together a single RNA molecule at a protein–RNA ratio of 1:1. There are four NMR structures of a protein–RNA complex in which two or three RNA binding domains bind one RNA molecule at a 1:1 ratio [1,18,22,30]. The molar ratio was obtained from NMR titration experiments. In contrast, the two C-terminal RNA binding domains of the protein PTB interact with each other bringing the two RNA binding surfaces on opposite sides of the structure. In this case, NMR titration experiments indicated that each domain binds independently one RNA molecule resulting in a protein:RNA molar ratio of 1:2 [72].

Titration experiments also provide a good indication of the exchange regime of the complex formation and therefore offer a rapid evaluation of the spectral quality of the complex, which is a crucial parameter for the determination of the structure. Therefore, already at this stage, an optimization of the RNA sequence, the buffer and temperature conditions can be performed. 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 [153] leading to the loss of many amide signals even in the presence of excess RNA. Using ^{15}N - ^1H HSQC experiments, different buffer and temperature conditions were therefore tested and optimal conditions were determined where all signals corresponding to the bound form of the protein

are present in the spectra (Fig. 8) [14]. Another striking example of using NMR titration experiments to assess the quality of NMR spectra and to define the optimal conditions for the structure determination of a protein–RNA complex is the study of the protein SRp20 in its complex with RNA [17]. In this case, a total of 13 different RNA sequences, derived from natural binding sequences and from SELEX experiments, were used. For each RNA, the quality of the NMR signal linewidths was evaluated for the protein using ^{15}N - ^1H HSQC spectra and for the RNA using 2D ^1H - ^1H TOCSY spectra. Depending on the RNA sequence, the complex formation leads to broad signals for both the RNA and the protein, sharp signals for the protein but broad signals for the RNA or vice-versa. Only one RNA sequence resulted in sharp NMR signals for both the protein and the RNA. This RNA sequence was therefore used for the structure determination of the complex [17].

NMR titration experiments are also useful for comparing the RNA binding properties of an RNA binding domain with the RNA binding properties of the full-length protein. By comparing titration experiments using a full-length protein or a single RNA binding domain, it is possible to estimate if the RNA binding domain is necessary and sufficient for RNA binding or if additional regions of the protein play a role in the affinity and the specificity of the interaction. In the case of RNA binding proteins that contain multiple RNA binding domains, chemical shift perturbation experiments al-

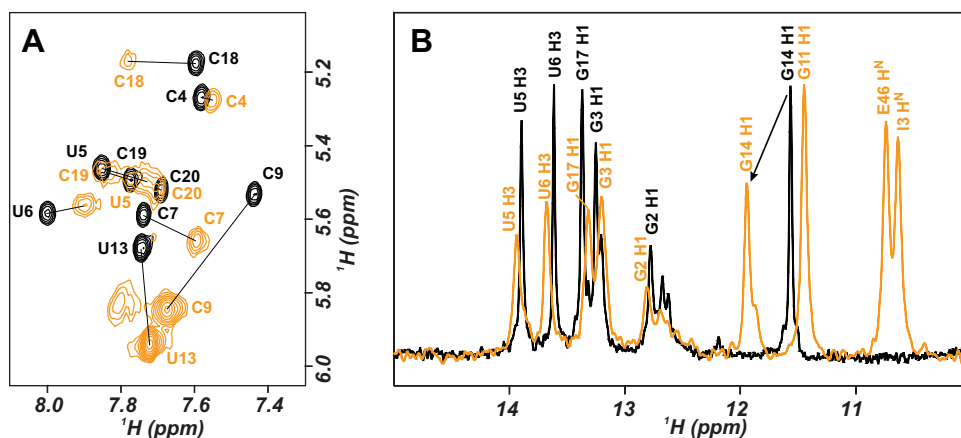


Fig. 9. Comparison of 2D and 1D NMR spectra of the 20 nt Shine–Dalgarno sequence of the *hcnA* mRNA in the free state (black) and complexed with the protein RsmE (orange) measured at 310 K [34]. (A) H5–H6 correlations in 2D TOCSY spectra at 500 MHz (free) and 900 MHz (bound) display large chemical shift changes between the free and bound RNA. (B) 1D spectra with 3–9–19 water suppression [161] at 900 MHz (free) and 600 MHz (bound) display one additional imino signal of G11 that forms a hydrogen bond with the protein, a large deviation for the G14 imino proton (signal arrowed) that is close to the binding interface and some minor changes of imino signals from the stem. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

low one to determine which RNA binding domains are able to bind RNA, if they are independent or if they bind RNA cooperatively. For instance, the protein PTB contains four RNA binding domains. RNA binding by the full-length protein has been tested by chemical shift perturbation experiments. Saturation was reached when four equivalents of the RNA were added to the protein suggesting that each RNA binding domain binds one RNA sequence. NMR titration experiments were then performed on each individual domain of the protein showing that each RNA binding domain binds RNA independently and a comparison with the NMR spectra of the full-length protein showed that the RNA binding is similar in both contexts [28].

Finally, if the resonances of the protein and/or the RNA have been previously assigned, chemical shift perturbation experiments allow the identification of the residues directly or indirectly involved in the complex formation. It is therefore possible at this early stage of the study to identify the binding interface of each component. When the structure of the individual components are known, structural models of the complex can already be generated solely based on chemical shift perturbation analysis [159]. This approach has been used to derive the structural models of two protein–RNA complexes that have been deposited in the protein data bank [46,47]. Although these models are less accurate to describe the intermolecular interactions, they can provide useful information for further biochemical analysis of these complexes.

Altogether, initial NMR studies of a protein–RNA complex formation are crucial steps that provide rapid information useful for the determination of the structure of the complex. Since structure determination of protein–RNA complexes is a time and resource consuming process, we believe that it is very important to optimize the conditions at this early stage of the NMR study in order to obtain NMR spectra of good quality.

3. NMR spectroscopy and structure determination of protein–RNA complexes

3.1. NMR methodology and resonance assignment

3.1.1. Preliminary NMR experiments

After initial NMR titration experiments and a first optimization of conditions as described in Sections 2.4.2 and 2.4.3, obtaining the resonance assignments of the bound protein and RNA is the next step towards a structure determination. However, before proceeding it has to be judged whether or not 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 also be visible in the bound state and second, a sufficient number of intermolecular NOEs should be observed in order to solve the complex structure.

In addition to the initial HSQC experiments that are used to monitor chemical shift changes upon complex formation, additional experiments are needed to evaluate the quality of the complex under specific conditions. A 2D ^1H – ^1H TOCSY is generally used to monitor the state of the RNA by analysing the H6–H5 correlations of cytosines and uracils (see Fig. 9A). All expected signals should be visible with a good line shape and no additional signals, such as two sets of signals, should appear. This monitoring is of course not possible for RNAs lacking cytosines and uracils. In addition, the imino region is analyzed by 1D jump-and-return echo [160] or 1D WATERGATE spectra [161]. The appearance of imino-signals upon complex formation is an indication of intermolecular hydrogen bonds like G11 H1 bound to RsmE (Fig. 9B). More importantly, a 2D NOESY in H_2O optimized for the imino region is typically recorded at low temperature. To suppress the water signal, the jump-and-return echo method [160] usually gives better signal

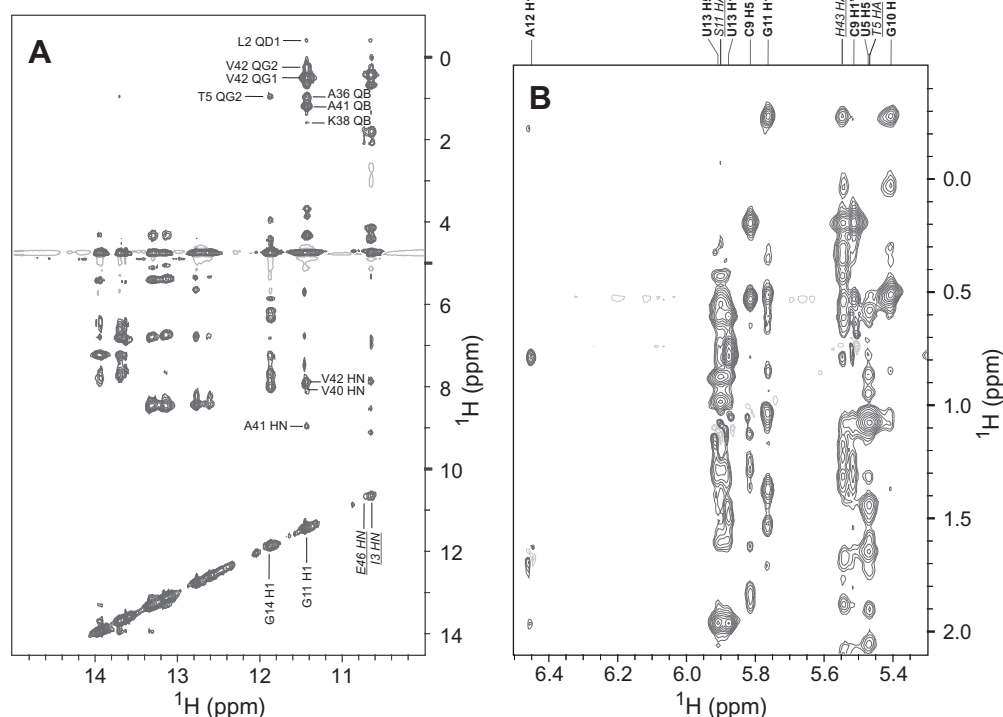


Fig. 10. 2D NOESY spectra of RsmE in complex with the Shine–Dalgarno sequence of the *hcnA* mRNA [34]. (A) Imino region of a 2D NOESY at 900 MHz in H_2O . The intermolecular NOEs are labeled. (B) H1'/H5 region of a 2D NOESY in D_2O . RNA resonances are indicated with black labels at the top. Three downfield Hz chemical shifts of the protein are italic and underlined. All cross peaks in this region represent intermolecular NOEs except for the ones involving the three protein Hz resonances of Thr5, Ser11 and His43.

to noise than the WATERGATE technique [161] but leads to more baseline distortions. Any NOE from an imino proton to the aliphatic region <2.5 ppm is most probably an intermolecular NOE (see Fig. 10A). The quality and number of intermolecular NOEs can be judged by such a 2D NOESY. However, if no imino signals are observable, sufficient intermolecular NOEs might still be observed using other experiments, as seen, for example, for the complexes of SRp20 [17] and PTB [28]. Then a 2D NOESY spectrum measured in D_2O can be used to estimate the dispersion of RNA signals and the number of intermolecular NOEs. Signals in the region between 5 and 6 ppm typically originate from the RNA, in particular from H1' and H5 nuclei. Protein signals of $H\alpha$ and aromatic residues are rarely found in this region and amide signals are mostly absent in D_2O . Correlations between the region of 5–6 and 7.2–8.2 ppm include H8/H6–H1', H2–H1' and H6–H5 correlations. These can be used to estimate the chemical shift dispersion of the RNA and for initial assignment attempts. Cross peaks between resonances at 5–6 ppm and aliphatic protein signals, e.g. upfield of 2 ppm are likely to be intermolecular NOEs (Fig. 10B). Eventually, conditions might need to be further optimized or constructs changed. Note that for all those experiments no isotope labeling is required.

3.1.2. Temperature, ionic strength, and solvent

To obtain the best signal to noise and line shape in NMR experiments, factors such as 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. Buffers without any NaCl or KCl have been used, e.g. solely 50 mM Na-phosphate [17]. Arginine/glutamate buffer is a promising buffer not only because it can increase protein stability and solubility [17,162] but it also results in better signal to noise ratios in cryogenic probes due to its lower ionic strength compared to other buffers [163]. However, the disadvantage is that this buffer causes baseline distortions due to its strong NMR signals. To prevent this, the buffer needs to be prepared with deuterated arginine and glutamate.

After testing with a few μ l amounts of material (using a water bath or a Polymerase Chain reaction (PCR) machine) the temperature range over 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 [28] whereas the free form precipitated at 40 °C and was therefore measured at 30 °C [164]. The most challenging step is to identify both an optimal RNA target and conditions that result in good quality spectra. For larger complexes, elevated temperatures such as 40 or 50 °C have the advantage of significantly decreasing the line widths because of faster molecular tumbling. However, many factors militate against high or low temperatures such as solubility, RNA stability in regard to degradation, protein unfolding, thermodynamics of the binding equilibrium, the binding kinetics and exchange broadening of NMR signals. Some protein–RNA complexes can precipitate reversibly at low temperatures (unpublished work). At temperatures just below the melting

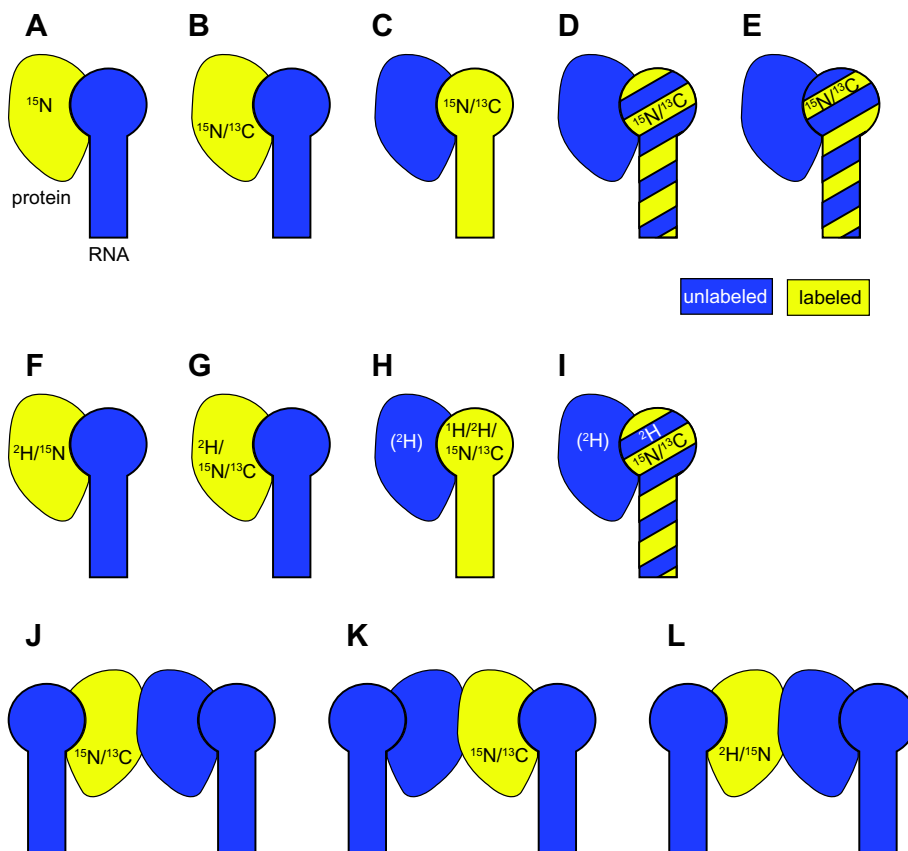


Fig. 11. Schematic presentation of the different labeling schemes used for the structure determination of protein–RNA complexes. (A) ^{15}N labeled protein and unlabeled RNA. (B) $^{13}C/^{15}N$ labeled protein and unlabeled RNA. (C) Unlabeled protein and uniformly $^{13}C/^{15}N$ labeled RNA. (D) and (E) Unlabeled protein and different versions of nucleotide-specific $^{13}C/^{15}N$ labeled RNA. (F) Deuterated and ^{15}N labeled protein and unlabeled RNA. (G) deuterated and $^{13}C/^{15}N$ labeled protein and unlabeled RNA. (H) Unlabeled or deuterated protein and site-specific deuterated $^{13}C/^{15}N$ labeled RNA. (I) Unlabeled or deuterated protein and nucleotide-specific unlabeled and otherwise perdeuterated RNA. (J) and (K) Complex of a protein dimer consisting of a $^{13}C/^{15}N$ labeled and an unlabeled monomer together with two unlabeled RNA molecules. (L) Complex of a protein dimer consisting of a perdeuterated ^{15}N labeled and an unlabeled monomer together with two unlabeled RNA molecules.

temperature, a small part of the protein can denature and irreversibly form aggregates. Therefore also the long-term stability should be checked at each temperature to be used in the study.

Often the structure determination of protein–RNA complexes requires the collection of spectra in D_2O and in H_2O . Typically the samples can be lyophilized and thus the solvent exchanged between D_2O and H_2O . However, some buffers cannot be easily lyophilized or need to be readjusted after lyophilization. For example in the case of commonly used acetate buffer [4,12,17,23] the acetic acid evaporates during lyophilization.

3.1.3. Typical samples for NMR measurements of protein–RNA complexes

Isotope labeling is absolutely essential for resonance and NOESY assignment of protein–RNA complexes. Uniform ^{15}N - and $^{15}N/^{13}C$ labeled proteins are usually used in complexes with unlabeled RNA (Fig. 11A and B) to assign the backbone and side-chain resonances of the protein in the 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 3.2). If possible *in vitro* transcribed $^{15}N/^{13}C$ labeled RNAs are used for complexes with either unlabeled or ^{15}N labeled proteins (Fig. 11C). Although uniform $^{15}N/^{13}C$ labeling of RNA is most often used, nucleotide-type specific labeling schemes (Fig. 11D and E) can provide certain advantages as discussed subsequently for RNA assignment (Section 3.1.7). For example two samples, one containing $^{15}N/^{13}C$ labeled adenines and cytosines and another one with $^{15}N/^{13}C$ labeled guanines and uracils, can be helpful. The use of ^{15}N only labeled nucleotides is not very beneficial considering the small difference of price between ^{15}N NTPs compared to $^{15}N/^{13}C$ NTPs. The advantage of using ^{15}N labeled protein in complex with $^{15}N/^{13}C$ labeled RNA is that the complex formation can be easily monitored by ^{15}N - 1H HSQC spectra.

For example, the structure determination of several recently determined protein–RNA complexes using *in vitro*-transcribed RNA were made with four samples: one containing ^{15}N -protein

and unlabeled RNA, one containing $^{15}N/^{13}C$ -protein and unlabeled RNA and two samples with nucleotide specific $^{15}N/^{13}C$ labeled RNA in complex with ^{15}N -protein. Two samples of combined nucleotide specific $^{15}N/^{13}C$ labeling proved to be most useful, e.g. one sample containing labeled A + C and another sample labeled G + U [34,35] or alternatively one sample labeled A + U and another sample labeled G + C [20,29]. Sometimes four complex samples of single nucleotide specific $^{15}N/^{13}C$ labeled RNA [22,39] or a combination of single and double-nucleotide specific $^{15}N/^{13}C$ labeled RNA were necessary to resolve degeneracies [1,21].

To determine the structure of a protein in its complex with a short ssRNA two samples are often used: one with ^{15}N protein–unlabeled RNA and one with $^{15}N/^{13}C$ protein–unlabeled RNA, because the target RNA cannot be transcribed *in vitro* (see Section 2.3.1). However, specific positions in an RNA can be $^{15}N/^{13}C$ labeled by chemical synthesis and greatly improve the quality of the structural determination of a protein–RNA complex [96]. Isotope labeling can be very beneficial in this context, especially if long stretches of the same nucleotide types are present [14,28]. The chemical shifts of three consecutive guanines were clearly distinguished and assigned by using such an approach as shown in Fig. 12. Chemical synthesis of RNA in which nucleotides are labeled at specific positions, for example at each alternating nucleotide, also proved to be very beneficial for complexes with an RNA stem-loop [29,35].

3.1.4. Typical samples for NMR measurements of peptide–RNA complexes

The samples and methodologies for peptide–RNA complexes are almost identical to those for protein–RNA complexes with the exception that peptides are often chemically synthesized [7,8,11,19,23,31,41–44]. This offers the possibility to introduce site-specific isotope labeling at a certain residue. For example single ^{15}N [Gly] labels were introduced into a 14-residue peptide to facilitate the assignments of three glycine residues [31]. However, bacterial expression is also frequently used to generate uniform ^{15}N or $^{15}N/^{13}C$ labeled peptides [6,7,9,19,33,42,44,165].

3.1.5. Resonance assignment of proteins and peptides in complex with RNA

Resonance assignment of proteins in complex with RNA is in principle identical to the procedures used for isolated proteins. Standard triple resonance experiments [166,167] are applied on samples containing uniformly $^{15}N/^{13}C$ labeled proteins. The unlabeled RNA component is invisible in these experiments.

The types of experiments that are typically recorded and the required samples are illustrated in Table 2. As an example, this table lists all the NMR experiments that were recorded on the protein RsmE/hcnA protein–RNA complex [34]. The measuring time for all experiments amounted to a total of 6 weeks if run on a single NMR spectrometer. The experiments for protein assignment are all standard experiments for the protein NMR community and will not be further discussed here. The experiments for assigning RNA are further described in Section 3.1.7 and filtered and/or edited NOESY experiments are discussed in Section 3.2.

Experiments that are not widely used in protein NMR are discussed in more detail in this Section. Since positively charged arginine and lysine residues play often a crucial role in protein–RNA recognition, their side-chain assignments are necessary for obtaining useful intra- but more importantly inter-molecular distance restraints. Although arginine N_ϵ - H_ϵ can normally be observed in standard ^{15}N - 1H HSQC type correlations including in 3D ^{15}N -edited NOESY-HSQC spectra, the typical ^{15}N offsets are not optimal for arginine N_ϵ - H_ϵ resulting in sensitivity losses. Therefore, experiments optimized for the Arg $^{15}N_\epsilon$ and $^{15}N_\eta$ with adjusted offsets, delay lengths and sometimes flip-back pulses and ^{15}N selective

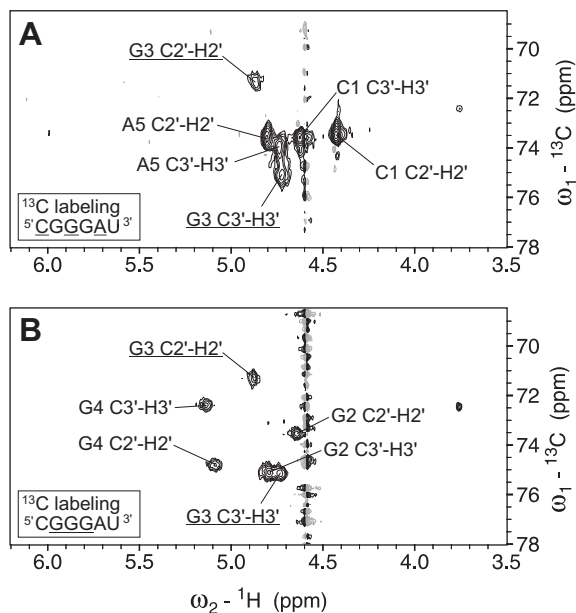


Fig. 12. ^{13}C - 1H HSQC spectra of a guanine-rich single stranded RNA (6 nt) obtained by chemical synthesis in complex with the RRM1 of hnRNPF [14]. (A) ^{13}C labeled ribose was introduced at positions 1, 3 and 5. (B) ^{13}C labeled ribose was introduced at positions 2, 3 and 4. The $C2'$ - $H2'$ and $C3'$ - $H3'$ correlations (underlined) present in both ^{13}C - 1H HSQC spectra at 600 MHz (A) and 900 MHz (B) can be assigned to G3.

Table 2
Experiments used for the assignment and structure determination of the complex of RsmE and the Shine–Dalgarno sequence of the *hcnA* mRNA [34]. Experiments in bold are discussed in detail in Section 3.2.

Protein	RNA	Solvent	experiments	purpose
¹⁵ N	Non labeled	H ₂ O	2D NOESY (low T) 2D TOCSY 2D ¹⁵ N– ¹ H HSQC	RNA imino assignment, NOE restraints for identifying NH ₂ groups in RNA reference spectrum
¹⁵ N	Non labeled	D ₂ O	2D NOESY, 2D TOCSY 2D ¹⁵ N– ¹ H HSQC 2D ¹³ C– ¹ H HSQC	extracting protein–protein and protein–RNA NOEs extracting NOEs for whole complex (high resolution in both dimensions) ¹ H assignment of aromatic rings (protein and RNA) identifying 2' endo ribose pucker identifying protected HN groups natural abundance ¹³ C– ¹ H HSQC of RNA (no <i>J</i> _{CC} coupling)
¹³ C/ ¹⁵ N	Non labeled	H ₂ O	2D ¹⁵ N– ¹ H HSQC 3D HNCA 3D HNCACB 3D CBCA(CO)NH 3D HNCO 3D HN(CA)CO 3D ¹³ C ^{aro} -ed. NOESY 3D ¹³ C ^{ali} -ed. NOESY	seed and reference spectrum protein backbone assignment protein backbone assignment protein backbone assignment protein backbone assignment protein backbone assignment extracting NOEs mainly for protein but also for whole complex extracting NOEs mainly for protein but also for whole complex
¹³ C/ ¹⁵ N	Non labeled	D ₂ O	¹³ C– ¹ H HSQC (H)CCH-TOCSY 3D F1eF3f NOESY^a 2D F1fF2e NOESY^b 2D F1fF2f NOESY^c CB(CGCD)HD CB(CGCDCE)HE	protein seed and reference spectrum protein side chain assignment extracting protein–RNA intermolecular NOEs extracting protein–RNA intermolecular NOEs extracting intra-RNA NOEs, assignment of aromatic side chains assignment of aromatic side chains
¹⁵ N	¹³ C/ ¹⁵ N Ade/Cyt	H ₂ O	¹⁵ N– ¹ H HSQC	RNA Amino seed and reference spectrum
¹⁵ N	¹³ C/ ¹⁵ N Gua/Ura	H ₂ O	¹⁵ N– ¹ H HSQC	RNA Imino seed and reference spectrum
¹⁵ N	¹³ C/ ¹⁵ N Ade/Cyt	D ₂ O	¹³ C– ¹ H HSQC HC(C)H-TOCSY 2D F1fF2f NOESY^c 3D F1eF3f NOESY^a ¹³ C ^{ali} -ed. NOESY	RNA seed and reference spectrum ribose assignment extracting RNA intra-NOEs within the unlabeled part extracting protein–RNA intermolecular NOEs and intra RNA NOEs between ¹³ C and unlabeled residues extracting intra and intermolecular NOEs
¹⁵ N	¹³ C/ ¹⁵ N Gua/Ura	D ₂ O	¹³ C– ¹ H HSQC HC(C)H-TOCSY 2D F1fF2f NOESY^c 3D F1eF3f NOESY^a ¹³ C ^{ali} -ed. NOESY	RNA seed and reference spectrum ribose assignment extracting RNA intra NOEs within the unlabeled part extracting protein–RNA intermolecular NOEs and intra RNA NOEs between ¹³ C and unlabeled residues extracting intra- and intermolecular NOEs

^a 3D ¹³C F1-edited F3-filtered HMQC-NOESY.

^b 2D ¹³C F1-filtered F2-edited NOESY.

^c 2D ¹³C F1-filtered F2-filtered NOESY.

pulses have been used for some complexes [7,19]. For example a 2D Arg-¹⁵N_{εη}-edited HSQC-NOESY [7], a 2D Arg-(H)C(C)TOCSY-N_εH_ε and Arg-H(CC)TOCSY-N_εH_ε for correlating arginine H_ε to side chain carbons and protons [168] and Arg-H_η(N_ηC_εN_ε)H_ε correlating arginine H_η and H_ε [19] have been successfully used to assign arginine and lysine side-chains in protein–RNA complexes.

Since protein–RNA complexes are prone to chemical exchange phenomena due either to conformational exchange or to exchange between the free and bound form, Carr–Purcell–Meiboom–Gill (CPMG) type magnetization transfers can sometimes refocus such effects leading to a better signal to noise ratio. The disadvantage is that additional heating is introduced into the sample. The ¹⁵N–¹H CPMG-HSQC proved to be very beneficial for the detection of Arg N_η–H_η signals as was shown in a protein–DNA complex [169] and could be very useful for better defining protein–RNA complexes. In the case of the HIV-1 nucleocapsid protein bound to tRNA^{Lys} D hairpin where the protein signals display fast to intermediate exchange, some of the exchange-broadened amide signals became better visible with the ¹⁵N–¹H CPMG-HSQC [170].

Generally, backbone triple resonance experiments should be measured preferably at lower fields (500–700 MHz) because more uniform excitation of ¹³C can be achieved and the effect of chemical shift anisotropy, especially of C', is not as severe as at higher fields [171,172]. In contrast, NOESY spectra should be recorded at the highest fields (700–1000 MHz) to obtain maximal resolution and sensitivity. Nevertheless, the field strength is also a factor influencing the exchange regime and spectral quality might be different at different field strengths.

The resonance assignment of peptides in peptide–RNA complexes follows either the same strategy as for proteins if the peptide can be ¹⁵N/¹³C or ¹⁵N labeled or with a different strategy if the peptide cannot be labeled. The unlabeled peptide is assigned in the presence of uniformly ¹⁵N/¹³C labeled RNA using 2D ¹H–¹H NOESY and TOCSY experiments that eliminate the RNA signals by filtering out protons attached to ¹³C (see Section 3.2) [11,23,31,41,43]. Alternatively, perdeuterated RNA can be used to eliminate the RNA signals [11]. In addition, a natural abundance ¹³C–¹H HSQC can be used to assign ¹³C chemical shifts of the peptide, because the carbon resonances of the peptide are usually separated from the RNA resonances.

3.1.6. Resonance assignment of proteins in large protein–RNA complexes

With increasing molecular size, fast relaxation resulting in line broadening becomes a major obstacle [173]. In addition, the complexity of the spectra increases with increasing number of resonances. The effect of slower tumbling on the signal to noise ratio depends on the type of NMR experiments. While the HNCA and HNCO experiments usually provide signals for most residues even for fairly large complexes, the CBCA(CO)NH, HNCACB, HN(CO)CA, and HN(CA)CO experiments are more sensitive to the molecular size and therefore higher temperature or/and deuteration together with Transverse Relaxation Optimized Spectroscopy (TROSY) experiments have to be used for larger complexes. For example three protein–RNA complexes with a size of ~28 kDa were studied at higher temperatures such as 40, 45 and 50 °C which was suffi-

cient for complete resonance assignment and structure determination without the need of deuteration or TROSY triple resonance experiments [1,28,34]. However, in other cases deuterated samples were required [22]. Deuteration significantly increases the T_2 relaxation times of the ^{13}C nuclei and therefore magnetization transfer from ^{13}C to neighboring ^{13}C and ^{15}N nuclei is efficient even in larger molecules or complexes [173–175]. For the protein backbone assignment of the 28 kDa protein–RNA complex consisting of an RNA stem-loop and three zinc-fingers [22], deuteration combined with TROSY versions of HNCA, HN(CA)CO and HNCOC experiments [176,177] were used in combination with a ^{15}N -edited NOESY-HSQC.

For side chain assignment based on TOCSY transfers, different ^2H labeling schemes can be applied, for example, almost 100% or ~70% deuteration. In 70% deuterated proteins, the remaining 30% ^1H nuclei still contribute to ^{13}C relaxation and also cause small ^{13}C chemical shift degeneracies because of isotope effects of the different isotopomers (e.g. CD_2 versus CHD). However, 70% deuteration also offers advantages because the remaining ^1H nuclei can be used for NMR experiments that require protons. For example 3D H(CCCO)NH-TOCSY and a 3D HBHA(CO)NH experiments have been successfully recorded on a 19 kDa protein–RNA complex with 70% deuteration of the protein while 100% deuteration was most effective for a 3D CC(CO)NH-TOCSY experiment [20]. However, side-chain assignment in most large complexes was achieved by NOE based approaches and by comparison of spectra from the free and bound protein form because often HCCH-TOCSY experiments become too insensitive due to enhanced relaxation [22].

The use of complete deuteration has the disadvantage that the observable NOEs are restricted to NOEs originating from the exchangeable amide protons, which is not sufficient to obtain precise structural ensembles. Random 70% deuteration does not solve this problem since the probability of having two neighboring hydrogens is only 9% and in addition frequency degeneracies due to different isotopomers broaden the signals. Selective re-introduction of protons into an otherwise deuterated protein is a good method for obtaining additional NOE distance restraints. This is increasingly being used to study large proteins but has not yet been applied to protein–RNA complexes. Detailed protocols for introducing protons into methyl groups have been published recently [178]. For example one of the two methyl groups of Leu and Val is selectively protonated ($^{13}\text{CH}_3$) in an otherwise uniformly- $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ labeled protein that is then ideally suited for

detection of through-bond methyl-NH correlations in order to assign the methyl groups [179]. The selective introduction of one $^{13}\text{CH}_3$ group in Ile, Leu and Val into an otherwise uniformly deuterated and non-carbon labeled sample results in very good line shapes in methyl-TROSY experiments [180] and could potentially be used for detecting NOEs in protein–RNA complexes.

The stereo-array isotope labeling (SAIL) method developed by the Kainosho group uses stereo-selective deuteration optimized for structure calculations of large proteins [181]. The method is based on cell-free protein synthesis using chemically synthesized amino acids with stereospecifically introduced ^2H and ^{13}C isotopes [182]. The protons are diluted to 50–60%, resulting in reduced ^{13}C relaxation, and the remaining ^1H are well distributed over the protein to yield sufficient NOE restraints. The synthesized amino acids are commercially available by SAIL Technologies Inc.. The drawback of the method is that the costs are significantly higher than for bacterial protein expression. SAIL has not yet been used for the study of protein–RNA complexes but would be very powerful for studying large proteins in complex with RNA.

Segmental isotope labeling can also be used to reduce the spectral complexity of large proteins resulting in a number of resonances comparable to those of small proteins. Different techniques and applications have already been mentioned in Section 2.2.2 and are found in recent reviews [87,183]. Segmental isotope labeling has been used for solving the structure of the two interacting RNA binding domains RRM3 and RRM4 of PTB in the free state [164]. One domain was $^{15}\text{N}/^{13}\text{C}$ labeled and the other domain unlabeled and vice versa. Filtered and edited NOESY spectra (see Section 2.3) were used to extract more than a hundred inter-domain NOE restraints.

3.1.7. Resonance assignment of RNA in small to medium size complexes

The assignment of an RNA bound to a protein follows similar strategies to those used for assigning a free RNA. We refer to excellent reviews describing the chemical shift assignment of RNAs [184–188]. The main difficulty during the assignment procedure of RNA is the small chemical shift dispersion found in the RNA sugar ^1H resonances (Fig. 13A). 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 ^1H resonances (Fig. 13B) that helps the assignment process. Furthermore, RNA nucleotides distant from the binding site typi-

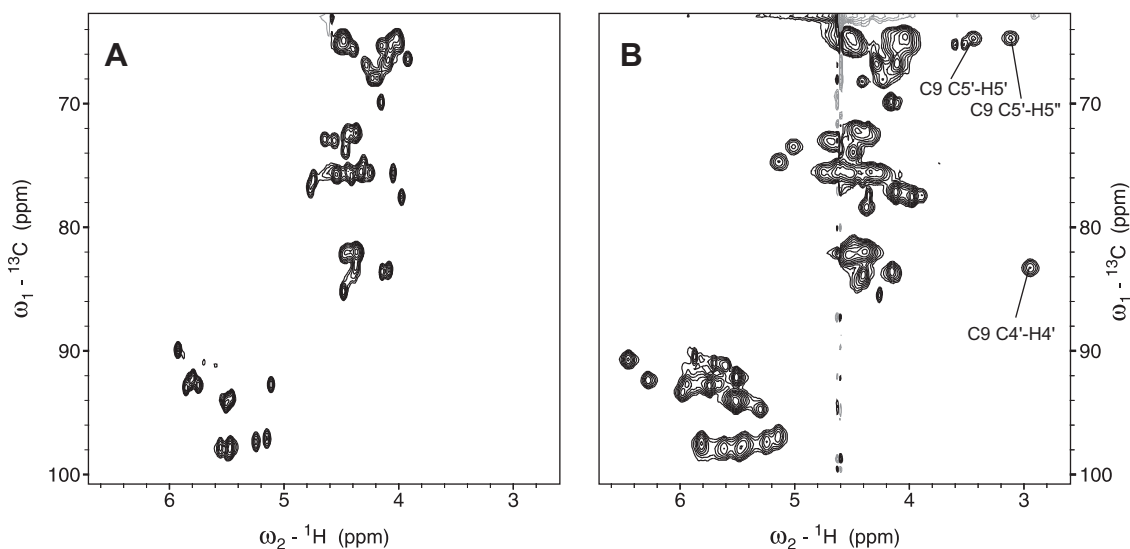


Fig. 13. ^{13}C - ^1H HSQC spectra at 500 MHz of a $^{13}\text{C}/^{15}\text{N}$ Ade/Cyt labeled Shine–Dalgarno sequence of the *hcnA* mRNA (20 nt) [34] in its free state (A) and bound to the protein RsmE (B). The resonances of C9 experience a large upfield chemical shift because they are located above the following guanine base G10 in the complex.

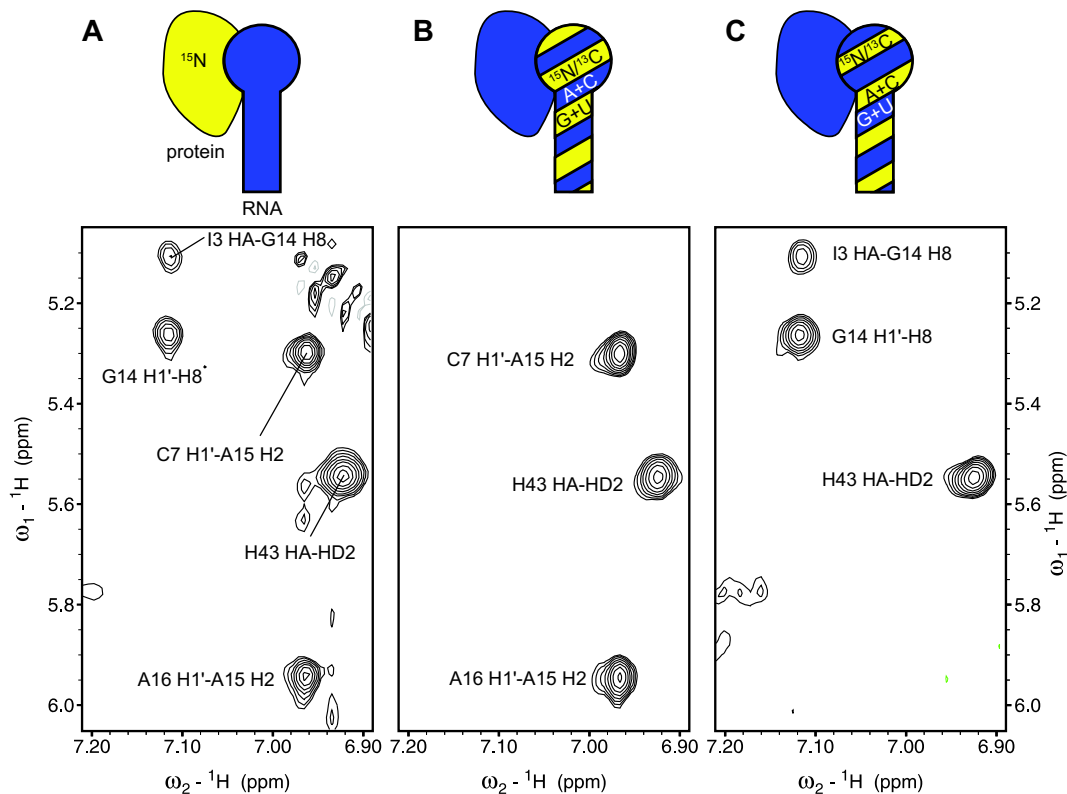


Fig. 15. Illustration of the use of nucleotide specific $^{13}\text{C}/^{15}\text{N}$ labeling for the simplification of H8/H6/H2–H1' correlations in the spectra of the protein–RNA complex of RsmE and the Shine Dalgarno sequence of the *hcnA* mRNA [34]. (A) Section of a 2D NOESY spectrum of unlabeled RNA and ^{15}N labeled RsmE measured in D_2O displaying NOEs from H1' and Hx resonances. (B) Same region in a 2D F1-filtered-F2-filtered NOESY spectrum of $^{13}\text{C}/^{15}\text{N}$ GUA/URA labeled RNA in complex with RsmE (^{15}N labeled). All NOEs involving guanine and uracil resonances are suppressed. (C) Same region in a 2D F1-filtered-F2-filtered NOESY spectrum of $^{13}\text{C}/^{15}\text{N}$ ADE/CYT labeled RNA in complex with RsmE (^{15}N labeled). All NOEs involving adenine and cytosine resonances are suppressed. Spectra were recorded on a 900 MHz spectrometer.

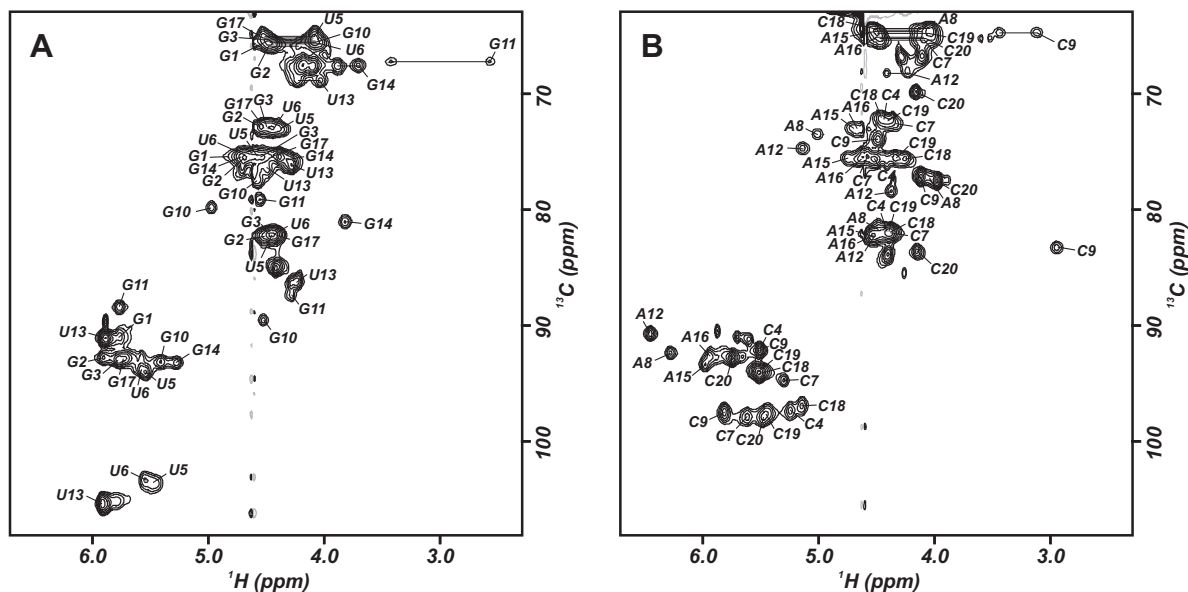


Fig. 16. ^{13}C – ^1H HSQC at 500 MHz of the Shine–Dalgarno sequence of *hcnA* mRNA bound to the protein RsmE [34] using (A) $^{13}\text{C}/^{15}\text{N}$ labeling of guanines and uracils or (B) $^{13}\text{C}/^{15}\text{N}$ labeling of adenines and cytosines.

imately the upper size limit at which the HCP experiment can be efficiently used due to the large ^{31}P chemical shift anisotropy relaxation effects and the short transverse ^{13}C relaxation times.

3.1.8. Resonance assignment of RNA in large complexes

The NMR structure determination of large protein–RNA complexes faces two major difficulties, line broadening and an in-

creased number of overlapping resonances. Methods to overcome these two effects on RNA are the use of specific ^2H labeling schemes and segmental isotopic labeling of RNA as discussed in Section 2.3.3.

Although the line broadening of the RNA base proton signals does not increase to the same extent as protein signals for a molecule of the same molecular weight, the ribose resonances broaden as much as in proteins. Decreasing the ^1H density by different strategies of specific deuteration reduces the line broadening and improves the quality of the spectra. Instead of randomly introducing ^2H , selectively deuterated RNA nucleotides prove to be most effective for assisting in the sequential assignment of RNA. Specific ^2H and ^{13}C labeling schemes have been developed by Williamson and coworkers [104]. For example deuterium labels are introduced at carbons 3', 4', 5' resulting in non-exchangeable protons at H1', H2' and at the base positions H6, H8, H2 [203]. This can also be combined with ^{13}C labeling of the entire ribose [204].

^1H labeling of a single nucleotide type in an otherwise deuterated RNA was introduced by Summers and coworkers [197]. In combination with nucleotide specific $^{13}\text{C}/^{15}\text{N}$ labeling as discussed previously, the authors developed a strategy for assigning large RNAs, for example the 101 nucleotide m Ψ RNA. Nucleotide specific protonated samples with an otherwise deuterated ($\sim 90\%$) background were analyzed using 2D NOESY spectra. The incomplete deuteration was an advantage since weak internucleotide NOE cross-peaks were observed whereas intra-nucleotide NOEs within the deuterated nucleotides were absent. The line widths of these spectra were much sharper compared to spectra recorded with $^{13}\text{C}/^{15}\text{N}$ labeled samples due to the absence or reduction of ^1H - ^{13}C and ^1H - ^1H dipole induced relaxation. The usefulness of such a labeling scheme has been demonstrated for two protein-RNA complexes with molecular weights of 31 and 39 kDa [10,45].

3.2. Restraints for the protein–RNA interface

Intermolecular NOEs between imino and protein protons can already be observed in 2D NOESY spectra recorded in H_2O (Fig. 10A). In order to fully define the intermolecular protein–RNA recognition interface significantly more 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 D_2O and H_2O contain all these NOEs but often peaks cannot be assigned unambiguously because of the severe signal overlap. Therefore, specific NMR experiments have been developed that use editing and filtering elements to select for intermolecular NOEs.

3.2.1. Editing and filtering building blocks

To unambiguously identify intermolecular NOEs, samples with opposite labeling of the two components are used (either the RNA unlabeled and the protein $^{15}\text{N}/^{13}\text{C}$ labeled or vice versa, see Figs. 11B–E) together with 2D and 3D NOESY spectra that select, for example, protons attached to ^{13}C in one dimension and to ^{12}C in another dimension. Selecting protons covalently attached to NMR active nuclei (e.g. $^1\text{H}[^{13}\text{C}]$ or $^1\text{H}[^{15}\text{N}]$) is called editing, whereas suppressing such resonances and thus indirectly selecting $^1\text{H}[^{12}\text{C}/^{14}\text{N}]$ is called filtering. Editing is widely used, e.g. in a ^{15}N -edited NOESY–HSQC and can be applied with or without chemical shift evolution of the heteronucleus. Editing can be achieved without breakthrough of $^1\text{H}[^{12}\text{C}/^{14}\text{N}]$ signals by phase cycling the pulses on the heteronucleus and thus either selecting for multiple quantum (HMQC-type) or hetero-nuclear single quantum (HSQC-type) coherences that cannot be formed by $^1\text{H}[^{12}\text{C}/^{14}\text{N}]$ nuclei. Note that slightly mistuned delays and imperfections of pulses only reduce the selected signal intensity but have no influence on the perfect elimination of $^1\text{H}[^{12}\text{C}/^{14}\text{N}]$.

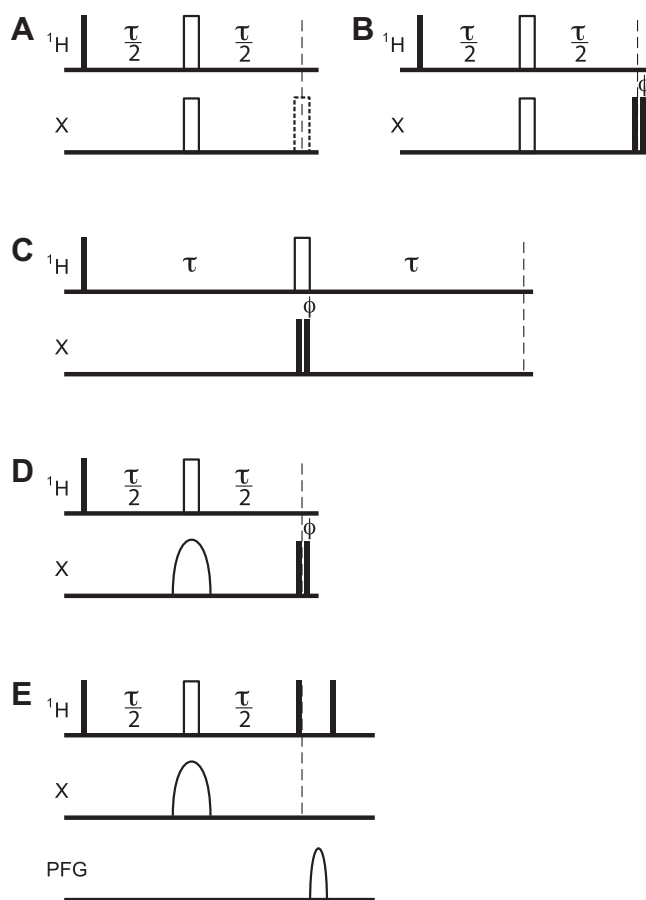


Fig. 17. Pulse sequence elements for filtering. (A) original X-half-filter. (B) Modified X-half-filter. (C) Modified X-half-filter with refocusing delay. (D) X-half-filter with shaped 180° pulse. (E) z-Filter. Unless indicated otherwise pulses are applied with phase x . The delay τ is set to $1/2J$. Narrow filled and wide unfilled rectangles correspond to 90° and 180° pulses, respectively. Magnetic field gradients as well as adiabatic ^{13}C pulses are represented by sine shapes.

Filtering is based on eliminating coherences formed by ^1H and the attached heteronucleus which can be antiphase-coherence e.g. $2\text{H}_x\text{C}_z$ or $2\text{H}_z\text{C}_y$, or multiple quantum coherences. The reader is referred to reviews that provide more information about the different experimental techniques employed [205,206]. The most important filter elements are shown in Fig. 17. The initial X half-filter shown in Fig. 17A [207] consists of a spin-echo element of the duration τ with two simultaneously applied 180° pulses on both the ^1H and the X channel. The delay is set to $1/2J$ resulting in antiphase magnetization for a ^1H -X spin system. A second 180° pulse is applied on the heteronuclear channel (dashed) only every second scan leading to a sign change of the antiphase magnetization and thus cancellation upon summation. In an improved version this 180° pulse is replaced by two consecutive 90° pulses as shown in Fig. 17B [208]. Both 90° pulses applied with phase x result in a 180° pulse whereas the effect of the first 90° pulse is cancelled if the second pulse is applied with phase x resulting in “no pulse”. The advantage is that less artifacts are generated due to similar offset dependencies of the $90^\circ_x 90^\circ_x$ and the $90^\circ_x 90^\circ_{-x}$ pulse pairs in comparison to one 180° pulse every second scan. A modified half-filter with a refocusing time τ is shown in Fig. 17C [208]. The disadvantage of the modified X half-filters is that the inversion profile of such a 180° pulse consisting of two 90° pulses is not perfect and the elimination of $^1\text{H}[X]$ is not complete as discussed later. To improve the inversion profile of the first 180° pulse, shaped inversion pulses and later on also adiabatic pulses were applied

as shown in Fig. 17D [165]. To circumvent any 90° pulse on the heteronucleus channel, the z-filter was developed as shown in Fig. 17E [209]. Here the antiphase magnetization $2H_xX_z$ is untouched by a 90°_x pulse in the 1H channel and subsequently eliminated by a purge gradient whereas the desired H_y term is converted by the 90°_x pulse to H_z that is not affected by the gradient pulse. The initially used hyperbolic secant inversion pulses are best replaced by adiabatic pulses [165,196].

In order to completely eliminate $^1H[^{13}C]$ or $^1H[^{15}N]$ resonances, it is critical that all coherences of these nuclei are in a form that can be purged by pulses, pulsed field gradients and/or phase cycling. In general, the antiphase term $2H_xC_z$ is formed from initial H_y magnetization after a delay $\tau = 1/2J_{HC}$ during which scalar coupling is active. However, since the scalar couplings vary depending on the type of proton, the delay should be slightly different for the elimination of each proton. This can not be fully achieved experimen-

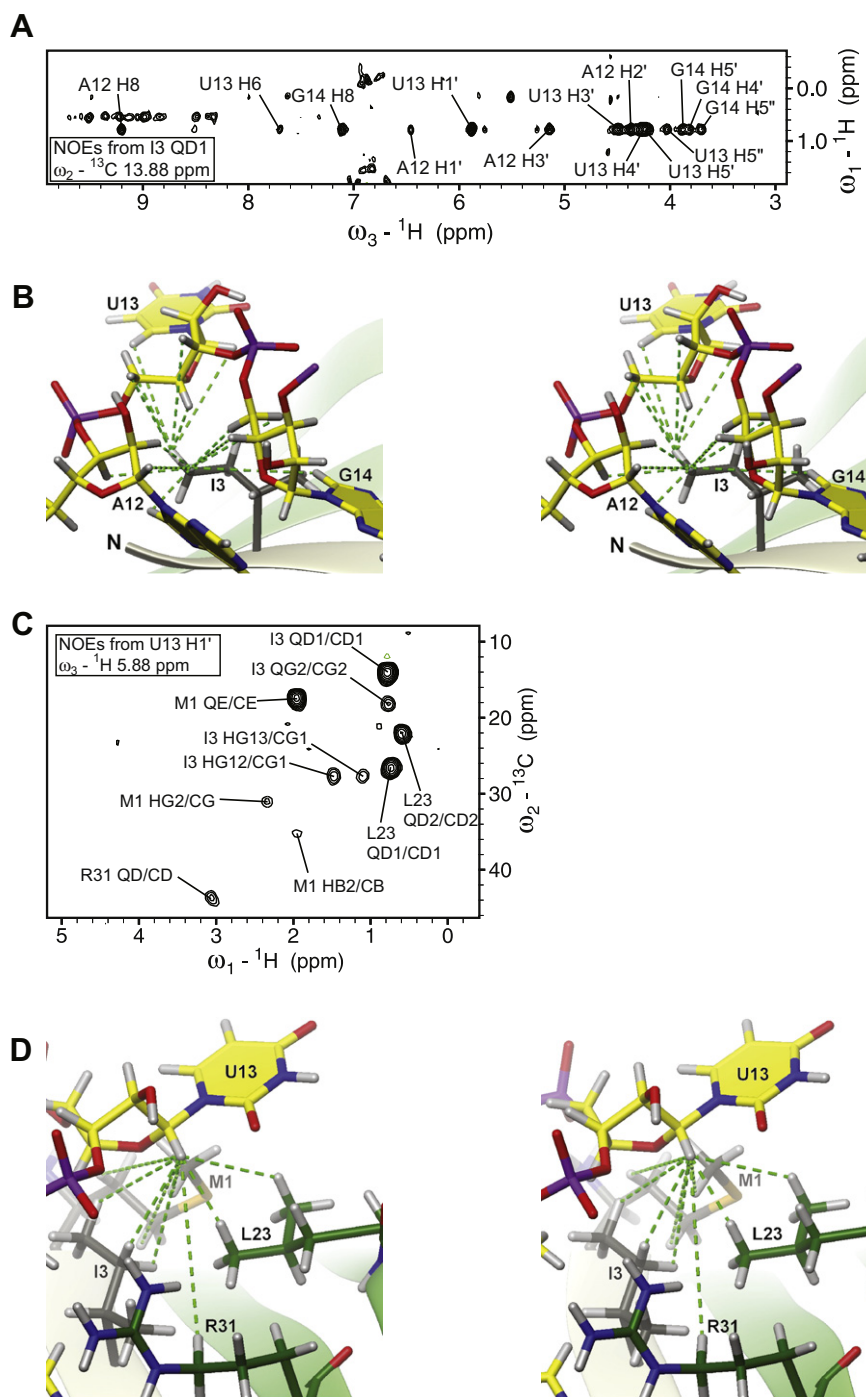


Fig. 18. 3D ^{13}C F1-edited F3-filtered HMQC-NOESY [212] at 600 MHz measured with a sample of $^{13}C/^{15}N$ labeled RsmE in complex with unlabeled 20 nt *hcnA* mRNA [34]. The spectrum was recorded exclusively with hard pulses using $n_s = 16$, $t_{1max} = 12.2$ ms, $t_{2max} = 38.7$ ms, a recycle delay of 1.3 s and a total measurement time of 86 h. (A) F1–F3 plane of the 3D spectrum at the ^{13}C chemical shift of the CD1 of Ile3 showing intermolecular NOEs between QD1 of Ile3 and RNA resonances in the direct dimension. (B) Three-dimensional structure of the RsmE–*hcnA* complex illustrating the observed NOEs to QD1 of Ile3. (C) F1–F2 plane of the 3D spectrum at the 1H chemical shift of U13 H1' showing intermolecular NOEs. (D) Three-dimensional structure of the RsmE–*hcnA* complex illustrating the observed NOEs to U13 H1'. Figures were generated with molmol [470].

tally and a compromise value for the delay has to be used. Therefore small breakthrough signals of the undesired $^1\text{H}[^{13}\text{C}]$ or $^1\text{H}[^{15}\text{N}]$ nuclei can be present in such spectra. Another reason for breakthrough signals are imperfections of radiofrequency pulses in particular of ^{13}C inversion pulses that have to cover a large frequency range. Either only aliphatic carbons (proteins: 5–75 ppm, RNA: 60–110 (including C5) or aromatic carbons (proteins: 110–150 ppm, RNA: 130–160 ppm) are filtered. Alternatively, the entire ^{13}C frequency range (proteins: 5–150 ppm, RNA: 60–160 ppm) is filtered which means nearly perfect ^{13}C pulses are required covering this range. Adiabatic pulses like WURST [210] are well-suited for such purposes. In contrast to such improved 180° pulses the broadband excitation profiles of 90° pulses are not perfect and therefore filters without 90° pulses on the heteronucleus such as z-filters (Fig. 17E) should be used in order to prevent further breakthrough signals.

To improve filtering, two consecutive filter elements are often used with either the same or different delays. In a double-tuned purge, the first element can be tuned e.g. for RNA to $^1J_{\text{CH}} = 200$ Hz, which is optimal for suppressing base proton signals, and the second element to $^1J_{\text{CH}} = 145$ Hz, which is optimal to suppress signals from ribose protons [196]. An approximately linear correlation between the ^{13}C chemical shift and the scalar coupling constant $^1J_{\text{CH}}$ was found by Kay and coworkers [165]. The time of active scalar coupling can then be adjusted to the $^1J_{\text{CH}}$ scalar coupling just by tuning it to the ^{13}C chemical shift. This was achieved in a very elegant way by using a carefully tuned adiabatic WURST pulse that sweeps from high to lowfield and inverts first upfield shifted ^{13}C resonances (methyls) and at the end downfield shifted resonances (aromatic region) [165]. For methyl groups with a small $^1J_{\text{CH}}$, a longer period during which scalar coupling evolves is active compared to aromatic ring CH with larger $^1J_{\text{CH}}$ values for which a shorter active coupling delay is active. Combining two such filter elements gives excellent results for filtering either ^{13}C labeled RNA or ^{13}C labeled protein signals. Slightly asymmetrical double purges (delay and pulse of the first element slightly different than in the second element) gave better results for filtering ^{13}C labeled

protein signals but not for filtering RNA signals [165]. Experimentally, the method of Kay and coworkers and the method of a double tuned purge are equally effective when two consecutive purge elements are used but the latter is slightly more sensitive due to the shorter delays [196].

3.2.2. 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 D_2O in order to improve sensitivity because of the increase in signal to noise and the ability to use a higher receiver gain. In addition, signals around 4.7 ppm are not obscured by the water signal or artifacts from water suppression techniques. A 2D ^{13}C F1-filtered F2-filtered NOESY is used to derive NOEs within the unlabeled smaller molecule (peptide or RNA) in the presence of the ^{13}C 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 [211] or by using double purge elements according to Feigon and coworkers [196]. Since 2D ^{13}C F1-filtered F2-filtered NOESY spectra are normally recorded in D_2O , most amide protons are absent because they are exchanged to deuterons. For the most sensitive and cleanest ^{13}C filters, either no ^{15}N pulses are applied (then the remaining amide protons appear as doublets in the spectrum) or ^{15}N is decoupled during t_1 and/or t_2 . Alternatively, $^1\text{H}[^{15}\text{N}]$ is filtered in addition to $^1\text{H}[^{13}\text{C}]$ but since the ^{15}N filter requires longer delays than the ^{13}C filter, the pulse sequence is either prolonged or one ^{15}N filter element is combined with two ^{13}C filter elements [165]. In either case the performance of the ^{13}C filtering is compromised either by signal decrease due to relaxation effects or by the appearance of more breakthrough signals.

To detect intermolecular NOEs, a variety of 2D and 3D filtered-edited NOESY experiments have been developed. The 3D ^{13}C F1-filtered F3-edited NOESY-HSQC [165] detects the $^1\text{H}[^{13}\text{C}]$ resonance in the direct dimension. This way an unambiguous identification

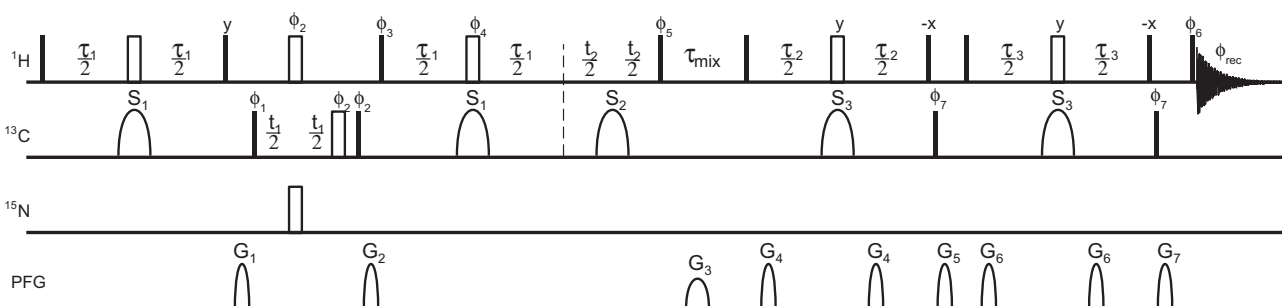


Fig. 19. Pulse sequence of the 3D ^{13}C F1-edited F3-filtered HSQC-NOESY used in our laboratory. Narrow filled and wide unfilled rectangles correspond to 90° and 180° pulses, respectively. Magnetic field gradients as well as adiabatic ^{13}C pulses are represented by sine shapes. Proton hard 90° pulses are typically applied with 25-kHz field strength. The phase cycling employed is as follows: $\phi_1 = (x, -x)$; $\phi_2 = 4(x), 4(-x)$; $\phi_3 = 2(y), 2(-y)$; $\phi_4 = 2(-x), 2(x)$; $\phi_5 = 16(x), 16(-x)$; $\phi_6 = 8(x), 8(-x)$; $\phi_7 = (x, y, -x, -y)$; $\phi_{\text{rec}} = (x, -x, -x, x), 2(-x, x, x, -x), (x, -x, -x, x), (-x, x, x, -x), 2(x, -x, -x, x), (-x, x, x, -x)$. The duration and strength of the gradients are $G_1 = 1$ ms (6 G/cm), $G_2 = 1$ ms (-3 G/cm), $G_3 = 1$ ms (-12 G/cm), $G_4 = 500$ μs (1.8 G/cm), $G_5 = 1$ ms (6.6 G/cm), $G_6 = 500$ μs (1.2 G/cm), $G_7 = 1$ ms (4.2 G/cm). The gradients were applied as a sinusoidal function from 0 to π . As adiabatic pulses 500 μs CHIRP pulses (80kHz linear sweep at 700 MHz, upfield to downfield) are used. For the optimized delays, pulses and offsets that depend on the application see Table 3. A presaturation period can be applied if the remaining water signal needs to be suppressed.

Table 3

Appropriate shaped pulses and tuned delays for the ^{13}C F1-edited F3-filtered HSQC-NOESY spectra used in our laboratory at 700 MHz.

	S1 Shaped pulse	S2 Shaped pulse	S3 Shaped pulse	τ_1 (corresp. J)	τ_2 (corresp. J)	τ_3 (corresp. J)
^{13}C labeled RNA	adiabatic at 85 ppm	adiabatic at 85 ppm	adiabatic at 85 ppm	2.94 ms (170 Hz)	3.57 ms (140 Hz)	2.27 ms (220 Hz)
^{13}C labeled protein	adiabatic at 67 ppm	adiabatic at 67 ppm	adiabatic at 67 ppm	3.45 ms (145 Hz)	4.00 ms (125 Hz)	2.78 ms (180 Hz)
^{13}C labeled protein, aliphatic region	180 μs Q3 at 40 ppm	hard rectangular	adiabatic at 67 ppm	3.45 ms (145 Hz)	4.00 ms (125 Hz)	2.78 ms (180 Hz)
^{13}C labeled protein, aromatic region	360 μs Q3 at 120 ppm	hard rectangular	adiabatic at 67 ppm	2.78 ms (180 Hz)	4.00 ms (125 Hz)	2.78 ms (180 Hz)

of the $^1\text{H}[^{13}\text{C}]$ resonance is achieved by the high resolution of F3 together with the ^{13}C chemical shift in F2. However, for the identification of the $^1\text{H}[^{12}\text{C}]$ resonance in F1 only low resolution is available. The 3D ^{13}C F1-edited F3-filtered HMQC-NOESY [212] detects the $^1\text{H}[^{12}\text{C}]$ resonances in the well-resolved direct dimension (Fig. 18A). 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 [34]. The importance of these restraints is illustrated in the 3D complex structure (Fig. 18B). The identity of the ^1H resonance of the labeled part can usually be determined via the ^1H – ^{13}C correlations in F1 and F2 even if these dimensions have lower resolution (Fig. 18C). The observed intermolecular NOEs between an H1' and four protein side chains are illustrated by the distances in the 3D structure (Fig. 18D). However, since the original F1-edited F3-filtered HMQC-NOESY uses only hard pulses that leads to break-through and artifacts with phase distortions, we now often use a 3D ^{13}C F1-edited F3-filtered HSQC-NOESY derived from the Bruker standard pulse sequence “hsqcgpnowx33d” that is based on a double tuned purge filter using adiabatic pulses (Fig. 19). We prefer to detect the $^1\text{H}[^{12}\text{C}]$ nuclei in the direct dimension, because resonances of the unlabeled molecules can be more easily identified with high resolution. Although the adiabatic pulses lead to satisfactory filtering over the entire carbon range, we prefer to record two separate experiments with complexes containing ^{13}C labeled protein, one optimized for the aromatic region, and one for the aliphatic region because of better resolution (Table 3). In both experiments, we select the region of interest by applying a selective 180° pulse and choose spectral windows and delays optimized for either aliphatic or aromatic residues.

The 2D ^{13}C F1-filtered F2-edited NOESY [196] based on double tuned purge elements is mainly used for assignment and NOE

extraction of nucleotide specific labeled RNA as shown in Fig. 20. However, no additional NOEs are expected compared to the aforementioned 3D experiments. 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 visible only in the very sensitive 2D NOESY in D_2O but not in the 2D or 3D filtered and edited NOESY. This is illustrated by a comparison of a 2D NOESY (Fig. 20 A) and a 2D ^{13}C F1-filtered F2-filtered NOESY (Fig. 20B) of the RsmE-*hcnA* protein–RNA complex [34]. However, signal overlap could be resolved by two 2D ^{13}C F1-filtered F2-filtered NOESY spectra using two differently RNA-labeled complex samples (Fig. 20 C and E). An alternative method for detecting intermolecular NOEs is to compare a F2 ^{13}C -filtered NOESY [211], in which the F1 dimension contains signals of both $^1\text{H}[^{12}\text{C}]$ and $^1\text{H}[^{13}\text{C}]$ with a 2D ^{13}C F1-filtered F2-filtered NOESY [20]. Using this approach, there is even a chance to see intermolecular NOEs from fast relaxing ^1H – ^{13}C groups that do not show up in ^{13}C -filtered-edited 2D or 3D NOESY spectra.

3.2.3. Editing or filtering of NH

^{15}N editing and filtering can be used in H_2O combined with or without ^{13}C editing and filtering, which can be useful for detecting intermolecular NOEs or intra-RNA NOEs between $^1\text{H}[^{15}\text{N}]$ and $^1\text{H}[^{12}\text{C}/^{14}\text{N}]$. 2D $^{15}\text{N}/^{13}\text{C}$ F1-filtered F2-filtered NOESY measured in H_2O is used to assign HN resonances of unlabeled peptides in the presence of $^{15}\text{N}/^{13}\text{C}$ labeled RNA. A 3D F1 ^{13}C -edited F3 $^{13}\text{C}/^{15}\text{N}$ -filtered HSQC-NOESY with WATERGATE, named hsqcgpnowx33d and a 3D F1 $^{13}\text{C}/^{15}\text{N}$ -filtered F3 ^{15}N -edited NOESY-HSQC with WATERGATE, named noesyhsqc3gpwngx13d, are part of the Bruker pulse sequence library and can be used for the detection of intermolecular NOEs to amide groups. For example

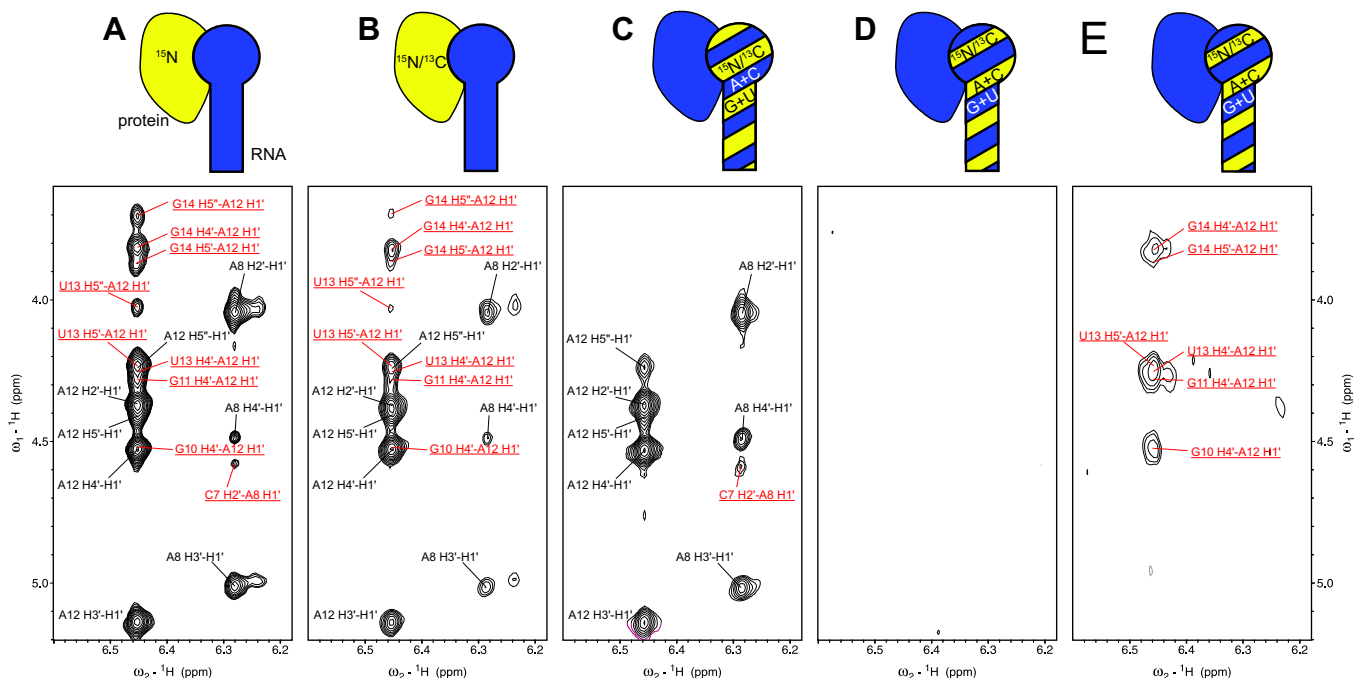


Fig. 20. Spectra simplification of 2D NOESY spectra using F1-filtered F2-filtered and F1-filtered F2-edited versions as demonstrated with the complex of RsmE with the Shine-Dalgarno sequence of the *hcnA* mRNA [34]. (A) 2D NOESY spectrum of the unlabeled complex showing all NOEs of the A8 and A12 H1' resonances to other ribose protons. Intranucleotide NOEs are labeled black, internucleotide NOEs labels are underlined and colored in red. (B) F1-filtered F2-filtered NOESY of the complex with ^{13}C labeled protein displaying only NOEs of the unlabeled RNA. In principle all signals as in the regular 2D NOESY should appear but due to relaxation during the filter elements the intensity is reduced and some NOEs might not be present. This spectrum is typically used to filter out NOE cross-peaks from the protein. (C) F1-filtered F2-filtered NOESY of the complex with $^{13}\text{C}/^{15}\text{N}$ Gua/Ura labeled RNA displaying only NOEs within the adenine and cytosine nucleotides of the RNA. This way inter-nucleotide NOEs can be distinguished from intra-nucleotide NOEs since only the later ones are visible. (D) F1-filtered F2-filtered NOESY of the complex with $^{13}\text{C}/^{15}\text{N}$ Ade/Cyt labeled RNA displaying no NOE cross peaks since the A8 and A12 H1' resonances are suppressed. (E) F1-filtered F2-edited NOESY of the complex with $^{13}\text{C}/^{15}\text{N}$ Ade/Cyt labeled RNA displaying only internucleotide NOEs to the A12 H1' resonance. Such a spectrum reveals very crucial sequential and long range NOEs and is much better dispersed than a regular 2D NOESY where those NOEs might be overlapped. Spectra were recorded on a 900 MHz spectrometer.

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 [7] in order to obtain 15 amide-RNA NOEs out of 81 intermolecular NOEs involving arginine $\text{N}_\epsilon\text{H}_\epsilon$ and N_ηH_2 . Since arginine side chains often play important roles at the protein–RNA interface the intermolecular NOE restraints from arginine $\text{N}_\epsilon\text{H}_\epsilon$ and N_ηH_2 are crucial for defining this interaction. A 2D $^{15}\text{N}_{\text{en}}$ -edited HSQC-NOESY recorded in H_2O was very useful for obtaining such important intermolecular NOEs. An additional filter resulted in a 2D $^{15}\text{N}_{\text{en}}$ -edited, $^{13}\text{C}/^{15}\text{N}$ -filtered HSQC-NOESY that yielded 43 intermolecular NOEs in the Rex peptide–RNA aptamer complex involving arginine $\text{N}_\epsilon\text{H}_\epsilon$ and N_ηH_2 [19].

3.2.4. Alternative methods used to derive intermolecular NOEs

A sensitive method used to detect amide to aliphatic intermolecular NOEs is based on high (>98%) deuteration of one component that is also ^{15}N labeled [213] in complex with an unlabeled component. Any detected NH-aliphatic NOE is then an intermolecular NOE. Typically, a 3D ^{15}N -edited NOESY-HSQC or NOESY-TROSY is recorded. Note that the method does not require filter elements and is therefore more sensitive than the previously described methods based on ^{13}C labeling. In addition deuteration slows down the relaxation so that also large complexes can be analyzed. This is a very promising approach for studying large protein–RNA complexes but it has not yet been used for such complexes.

3.2.5. Special case: protein dimer

If the protein is a dimer, then several interfaces are present, such as protein–protein, and protein–RNA interfaces. To identify intermolecular protein–protein NOEs at the dimer interface, a mixed isotopically labeled protein sample is crucial. Typically, a 50% unlabeled and 50% $^{13}\text{C}/^{15}\text{N}$ labeled protein sample is mixed under denaturing conditions and refolded. For this procedure it is crucial to show that the proteins are fully unfolded upon mixing and that they can be refolded. Finally 25% of the dimers will be unlabeled–unlabeled, another 25% labeled–labeled and 50% will have the desired unlabeled–labeled pattern (Fig. 11J). A 3D F1-filtered F3-edited ^{13}C - ^1H HSQC-NOESY or a 2D F1-filtered F2-edited NOESY can then be used to identify intermolecular NOEs at the dimer interface [38]. Another method used to detect inter-monomer NOEs is based on recording a 3D ^{13}C -edited NOESY spectra without ^{13}C decoupling in the indirect ^1H dimension yielding doublet signals for intra-monomer NOEs but singlet signals for inter-monomer NOEs [157]. An alternative to the mixture of 50% unlabeled and 50% $^{13}\text{C}/^{15}\text{N}$ labeled protein, is a mixture of >98% deuterated ^{15}N protein and unlabeled protein [213]. The presence of aliphatic-NH NOEs in a ^{15}N -edited (TROSY) NOESY indicate intermolecular NOEs. The advantage is that it is less sensitive to the

molecular size and thus large complexes can be studied. However, only intermolecular NOEs from exchangeable amide protons are visible.

If the dimer interface of a homodimer overlaps with the RNA binding interface, then protein–RNA intermolecular NOEs become ambiguous. This is illustrated with the symmetric protein dimer RsmE that binds two RNA stem loops [34]. The two protein monomers are labeled A and B and the RNA (molecule C) binds at the dimer interface between A and B, then observed intermolecular NOEs to the RNA can originate from either protein molecule A or B (Fig. 21). However, chemical shift mapping on the known 3D structure of the free protein dimer resulted in two symmetry-related binding sites that are clearly separate. Then it became apparent that residues from β -strand β_1 from monomer A and residues from β -strands β_3 , β_4 and β_5 from monomer B form one RNA binding site. Assigning the intermolecular NOEs to the two protein and two RNA molecules was only possible with the knowledge of the free protein structure.

The aforementioned examples presented symmetric dimers bound to either two identical RNAs or to one symmetrical RNA resulting in C_2 symmetry of the whole complex. In contrast, the symmetry of the homodimer LicT-CAT is broken when bound to a non-palindromic RNA stem-loop and most amino-acid resonances split into two components [40]. This break in symmetry led to two separate sets of resonances for each monomer and protein–protein inter- and intra-molecular NOEs could then be distinguished. Likewise protein–RNA intermolecular NOEs could be unambiguously assigned to each monomer.

3.3. Hydrogen bond and angle restraints

3.3.1. Experiments for obtaining hydrogen bond restraints

Amide protons that are protected from H to D exchange are typically involved in a hydrogen bond. However, the hydrogen bond acceptor cannot be derived from such exchange data. Before discussing recently developed experiments that identify unambiguously the hydrogen bond acceptors, we will discuss other indicators of potential intermolecular hydrogen bonds.

Intermolecular hydrogen 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. 22. The two largest NH chemical shift deviations (I3 and T5 amides) result from formation of intermolecular hydrogen bonds to two adenines (A8 and A12, respectively). Typically the amide ^1H resonances are downfield shifted upon the formation of hydrogen bonds, most pronounced for hydrogen bonds to RNA bases. This downfield shift is likely dominated by the ring-current effect

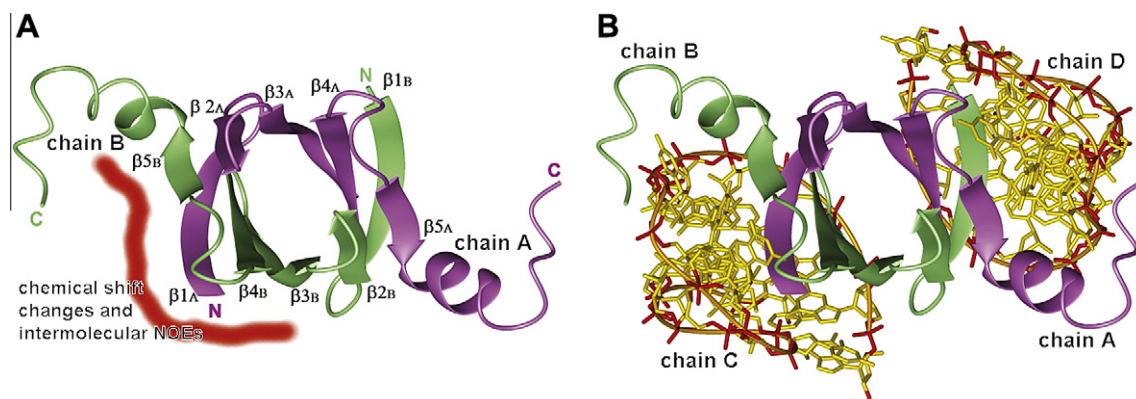


Fig. 21. Illustration of distinguishing chains in a dimeric protein–RNA complex [34]. (A) Dimer of the protein RsmE with chain A and B. Regions where chemical shift changes and intermolecular NOEs were observed are schematically indicated in red on one of the two symmetric binding sites. (B) One complex structure of the final ensemble with two RNA molecules (chains C and D). Figures were generated with molmol [470].

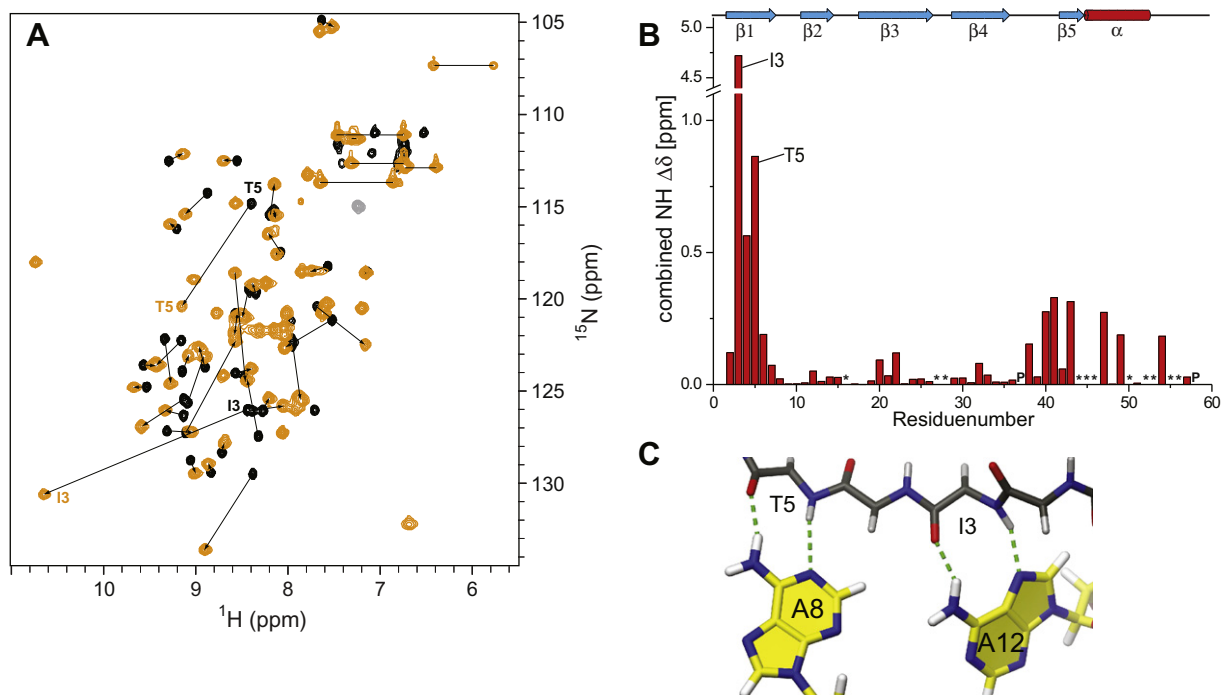


Fig. 22. Illustration of chemical shift deviations caused by intermolecular hydrogen bonds in the complex of the RsmE protein bound to the *hcnA* mRNA [34]. (A) Overlay of ^{15}N - ^1H HSQC spectra of the free RsmE protein (black) and in complex with the RNA (orange). (B) Combined chemical shift perturbation 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 hydrogen bonds between Ile3, Thr5 and the RNA bases A8 and A12. Spectra were recorded at 600 (free) and 500 (bound) MHz. Figures were generated with molmol [470]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

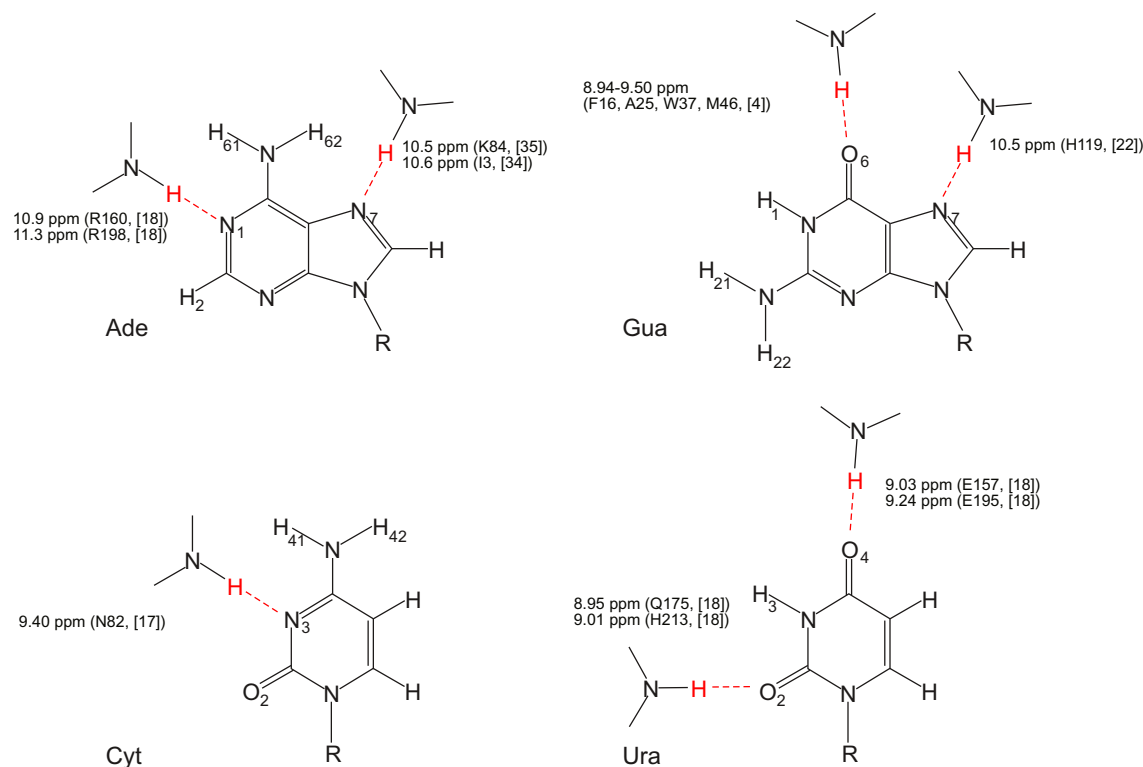


Fig. 23. Typical amide ^1H chemical shifts in intermolecular hydrogen bonds to RNA bases collected from the few chemical shift data available. The residue number from the original publication and the reference is given.

of the base since most downfield chemical shifts (>10 ppm) are observed in intermolecular hydrogen bonds to purine N7 and adenine

N1 where the protons are very close to the strong ring current of purines (Fig. 23). Amide ^1H Chemical shifts of 8.9–9.5 ppm are typ-

ical for intermolecular hydrogen bonds to carbonyls of the RNA bases. In all those cases large NH chemical shift deviations occurred. There is no up- or downfield trend for C' chemical shifts upon hydrogen bond formation but typically chemical shift deviations in the range of 0.5–3 ppm occur. When such chemical shift deviations of NH and C' are observed, these can then be used as an indication for an intermolecular hydrogen-bond. Note that the hydrogen-bond partner is not identified this way. Typically, potential hydrogen bonding partners are derived from initial structure calculations without any hydrogen-bond constraints. An iterative, step-wise introduction of such hydrogen-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 hydrogen-bond partners have been developed [214]. These are based on detecting the small scalar couplings across hydrogen-bonds which are in the range of 5–11 Hz ($^hJ_{NN}$) for N–H...N bonds and only –0.1 to –0.9 Hz for N–H...O=C bonds.

The fairly sensitive HNN-COSY experiment [215,216] is increasingly being used to detect hydrogen bonds in RNA base pairs by correlating the imino ^1H with the ^{15}N frequency of the hydrogen bond acceptor. The experiment uses TROSY to reduce relaxation effects so that larger molecules can also give sufficient signals to be studied. The HNN-COSY has been used for the structure determination of protein–RNA complexes ranging from 10 to 28 kDa [9,16,22,36]. A detailed protocol for the HNN-COSY TROSY was recently published [217]. Commonly, if a signal based on $^hJ_{NN}$ is detected in a Watson–Crick base-pair, then all Watson–Crick hydrogen bonds are used as restraints. Modified versions exist to detect also amino N–H...N hydrogen bonds as demonstrated with a DNA helix and quadruplex [218,219] and an RNA pseudoknot [220] but these HNN-COSY versions have not been used so far for protein–RNA complexes. To be measured, such hydrogen bonds would be required to have both components labeled. In the future, an HNN-COSY with adiabatic pulses [220] will be the method of choice for measuring intermolecular amide N–H...N hydrogen bonds, e.g. to N7 of adenine or guanine. Detecting N–H...O=C hydrogen bonds would be also possible but the H(N)CO experiments [221,222] or H(N)CO TROSY experiments [223,224] are very insensitive due to the small $^hJ_{NC}$ coupling constant of –0.1 to –0.9 Hz. Although the use of the H(N)CO TROSY for identifying hydrogen bond acceptors has been demonstrated on a 30 kDa deuterated protein at 0.7 mM measured for approximately four days on a cryogenic probe [223], it has not yet been used for protein–RNA complexes.

The most valuable hydrogen bond constraints for protein–RNA complex structures are intermolecular hydrogen-bonds. Such data, measured via scalar couplings, have not been used for structure calculation so far but it would be very valuable to have direct experimental evidence for such hydrogen-bonds. The first direct NMR detection of intermolecular hydrogen 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 [225] of the human T-cell leukemia virus (HTLV)-1 Rex peptide and a 33-mer RNA aptamer [19] after the structure had already been published. There might even be the possibility of detecting hydrogen bonds between backbone or side-chain amides with RNA phosphates oxygens because the scalar couplings $^hJ_{NP}$ are in the order of 1.7–4.6 Hz as observed for N–H...O=P hydrogen bonds in a flavodoxin containing a riboflavin 5' monophosphate [226] and in a Ras(Q61L)–GDP complex [227]. Such experiments require a probe tunable to ^{15}N and ^{31}P , for example a $^1\text{H}/^{13}\text{C}/^{15}\text{N}/^{31}\text{P}$ quadruple probe head.

3.3.2. Experiments for obtaining angle restraints

Scalar couplings can be used to extract angle restraints for structure calculations of proteins and of RNA since J couplings

are related to dihedral angles by Karplus type equations. Experiments for the measurement of J -couplings in proteins were developed at an early stage of protein NMR. Use of J -coupling for structure calculation in proteins have been reviewed by Bax and coworkers [228].

Scalar couplings $^3J_{\text{HNH}\alpha}$ are measured using the diagonal-peak to cross-peak intensity ratio in a 3D ^{15}N -separated quantitative J -correlation spectrum HNHA [229]. $^3J_{\text{HNH}\alpha}$ is used to derive ϕ backbone angle restraints or can be directly implemented in structure calculation programs, e.g. X-PLOR, XPLOR-NIH and CNS. In protein–RNA complexes, measurement of $^3J_{\text{HNH}\alpha}$ has been regularly used [6,7,9,16,18,19,24,26,33,37]. Backbone angles ϕ were restrained for example to $\phi = -120 \pm 45^\circ$ if $^3J_{\text{HNH}\alpha} > 7.5$ Hz [18]. For small couplings a constants ($^3J_{\text{HNH}\alpha} < 6.0$ Hz) sometimes ϕ is restrained to $-50 \pm 45^\circ$ [18], but such a constraint ignores residues in the α_L region of the Ramachandran plot (with $\phi = 60 \pm 45^\circ$ and $^3J_{\text{HNH}\alpha} = 4\text{--}7.5$ Hz). Such constraints should therefore be either omitted or only included with great care in structure calculations.

To derive other backbone and side-chain angles, a combination of several J coupling constants need to be measured. Scalar couplings of $\text{H}\beta$ to amide ($^3J_{\text{H}\beta\text{N}}$) and carbonyl ($^3J_{\text{H}\beta\text{C}}$) are measured from 3D HNHB [230] and 3D HN(CO)HB spectra [231] using signal intensity ratios. The $^3J_{\text{H}\beta\text{N}}$ depends on the angles ψ and χ_1 whereas $^3J_{\text{H}\beta\text{C}}$ depends on ϕ and χ_1 . $^3J_{\text{H}\beta\text{H}\beta}$ depends solely on χ_1 and can be measured using a 3D HACAHB-COSY experiment for which the coupling constant is also based on diagonal-peak to cross-peak intensity ratios [232]. For the aliphatic side chains of Thr, Ile and Val, the scalar couplings $^3J_{\text{NC}\gamma}$ and $^3J_{\text{CC}\gamma}$ are measured with spin echo difference constant time HSQC spectra [233,234]. Here the signal is detected on the methyl protons. $^3J_{\text{C}\gamma\text{N}}$ and $^3J_{\text{C}\gamma\text{C}}$ directly depend on χ_1 . For valines, stereo-specific assignment of the methyl groups is required. In order to measure $^3J_{\text{NC}\gamma}$ and $^3J_{\text{CC}\gamma}$ of all residues containing a $\text{C}\gamma$ (not restricted to Thr, Val, Ile), 3D HNCOC γ and 3D HNC γ experiments were developed [235]. In protein–RNA complexes, χ_1 angles have been derived from $^3J_{\text{CC}\gamma}$ and $^3J_{\text{NC}\gamma}$ scalar couplings [18,24,37]. Alternatively chemical shifts are often used to derive the backbone angles ϕ and ψ for protein–RNA complexes using the program TALOS [236] or the recently released TALOS+ [237].

Several techniques for deriving scalar coupling restraints for RNA structure determination have been reviewed [185,186,188,238,239]. Whereas conventional methods based on J -couplings can be applied to small peptide–RNA complexes (<15 kDa) to derive RNA torsion angles only a few methods are suitable for larger protein–RNA complexes (>15 kDa). Therefore only C2' endo and C3' endo ribose sugar pucker conformations have so far been identified using 2D TOCSY and 2D double quantum filter (DQF)-COSY spectra for protein–RNA complexes. The presence of a strong H1'–H2' signal in a DQF-COSY spectrum indicates a C2' endo sugar pucker while the absence of such a signal indicates a C3' endo pucker. In a 2D TOCSY experiment a strong signal for the H1'–H2' correlation and a weak signal for the H1'–H3' correlation are observed for a C2' endo sugar pucker while the absence of these signals indicates a C3' endo sugar pucker (Fig. 24A). C2' endo conformations are typically found in loops and in single stranded RNA whereas in A-form RNA the ribose pucker typically adopts a C3' endo conformation. The δ angle is then restrained to $130\text{--}190^\circ$ or $50\text{--}110^\circ$ for the C2' endo or C3' endo pucker, respectively. The C3' endo pucker is only restrained when the absence of the H1'–H2' and H1'–H3' cross peaks is not caused by line broadening effects which can be checked in other spectra such as 2D NOESY (Fig. 24B). For the case where the 2D spectra are too crowded, a 3D ^{13}C -edited TOCSY-HSQC can also be applied to extract the same information [23].

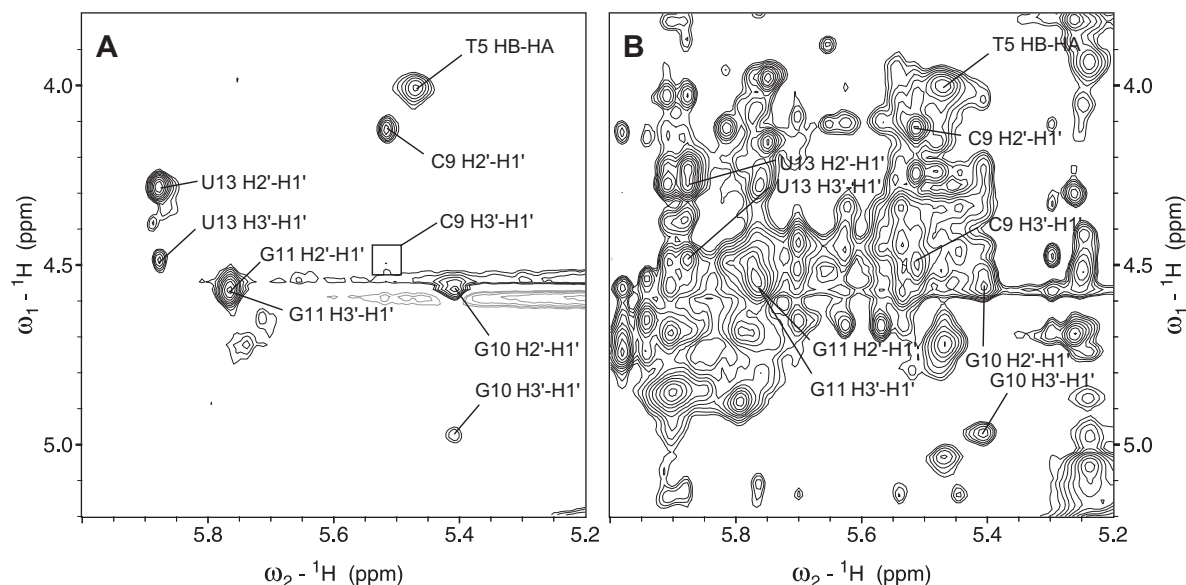


Fig. 24. 2D spectra of ^{15}N RsmE in complex with a 20 nt fragment of *hcnA* mRNA [34]. (A) 2D TOCSY spectrum recorded with four scans, a recycle delay of 1.3 s and $t_{1\text{max}} = 27$ ms in 52 min. Both, the $\text{H}1' - \text{H}2'$ and the $\text{H}1' - \text{H}3'$ correlation are observed for nucleotides G10 and U13 and were translated into $2'$ endo conformational restraints ($\delta = 130 - 190^\circ$). The signals of G10 are weaker probably due to some exchange phenomena, which are also broadening some other resonances of nuclei in the vicinity. The chemical shifts of $\text{H}2'$ and $\text{H}3'$ of G11 coincided leading to one strong signal consisting of both, $\text{H}1' - \text{H}2'$ and $\text{H}1' - \text{H}3'$ correlations. Thus $2'$ endo conformational restraints were also applied to G11. For C9 only the $\text{H}1' - \text{H}2'$ correlation but not the $\text{H}1' - \text{H}3'$ correlation was detected. In this case a loose constraint allowing for both $2'$ and $3'$ endo pucker was applied ($\delta = 50 - 190^\circ$). (B) 2D NOESY spectrum for comparison. Spectra were recorded at 900 MHz.

3.4. Long range restraints – global orientation

For most protein–RNA complexes, the structure determination has been primarily based only on the observation of intermolecular NOEs to derive intermolecular distance constraints. However, several protein–RNA complexes have poor specificities and affinities (in the higher micro–molar range). Dynamic interfaces and chemical exchange might hamper the collection 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 Å). When the molecular size of the complex increases, spectral crowding and increased relaxation might prevent collection of enough NOE restraints to determine the structure. Therefore, the introduction of long-range orientational restraints such as residual dipolar couplings (RDCs) or long-range translational restraints such as paramagnetic relaxation enhancement (PRE) has enabled the successful study of large protein–RNA complexes by NMR spectroscopy (see Fig. 25).

3.4.1. Residual dipolar couplings

In isotropic solution, the dipolar coupling between two spins is averaged to zero because of fast isotropic molecular tumbling on the NMR timescale. However, molecules having a strong anisotropy in their magnetic susceptibility can partially orient in a magnetic field. In this case the dipolar coupling is not averaged to zero. The residual dipolar coupling depends on the strength of the magnetic alignment of the macromolecule, on the distance between both nuclei and their identity, and most importantly, on the orientation of the inter-nuclear bond vector with respect to the magnetic field [240]. Therefore, residual dipolar couplings contain long-range orientational information that is not present in conventional NOE restraints. The alignment of the macromolecule in the magnetic field can be induced either by its own anisotropic magnetic susceptibility or by the use of an external orienting medium [241,242]. The elongated structure of nucleic acids can even allow RDC measurements without the need of an alignment medium.

The first example, where RDCs were used to refine an NMR structure, was provided by Bax and coworkers on a nucleic acid–protein complex [243]. RDCs could be obtained by magnetic field induced alignment of a protein in complex with a 16-nucleotide DNA. A high value of magnetic anisotropy is observed, when coplanar stacking of a significant number of aromatic rings occurs like in a helix of a nucleic acid. However, several protein–RNA complexes do not possess large enough magnetic anisotropic tensors to yield RDCs of useful magnitude. Therefore, all protein–RNA complexes published to date have required the addition of an external alignment medium, in order to yield residual dipolar couplings of useful magnitude. The alignment medium should not interact specifically with the macromolecules (to prevent severe line-broadening and/or sample precipitation) and should be stable under a broad range of buffer conditions. Several alignment media such as bicelles, filamentous phages, cetylpyridinium-based media, purple membrane fragments, cellulose crystallites, alkyl poly(ethylene glycol) and polyacrylamide gels have been developed over the years (reviewed in [240,244]). The negatively charged Pf1 filamentous bacteriophage is well suited for aligning nucleic acids and nucleic acid–protein complexes, because of repulsive interactions between the alignment medium and the macromolecules [245]. Therefore, the macromolecule is usually aligned by weak steric contacts with the medium [246]. Due to its very convenient handling and its excellent properties, Pf1 filamentous bacteriophage is the most commonly used alignment medium for protein–RNA complexes. It can simply be added to the macromolecular sample (The buffer containing the Pf1 phage might be exchanged by ultracentrifugation and resuspended in the appropriate buffer before mixing with the sample). It is fully aligned even at low magnetic fields (300 MHz) and its alignment is constant over useful ranges of temperature (15–45 °C in general, down to 5 °C in certain cases), pH (5.5–8.0) and buffer (0–100 mM NaCl, 10–50 mM Na-phosphate or Tris) [245,247]. The degree of alignment of the macromolecule is simply tuned by modifying the phage concentration. However, in some cases, the use of another alignment medium was required because of interactions of Pf1 with the macromolecule of interest. Bicelles, which are prepared by mixing dihexanoyl-phosphatidyl-

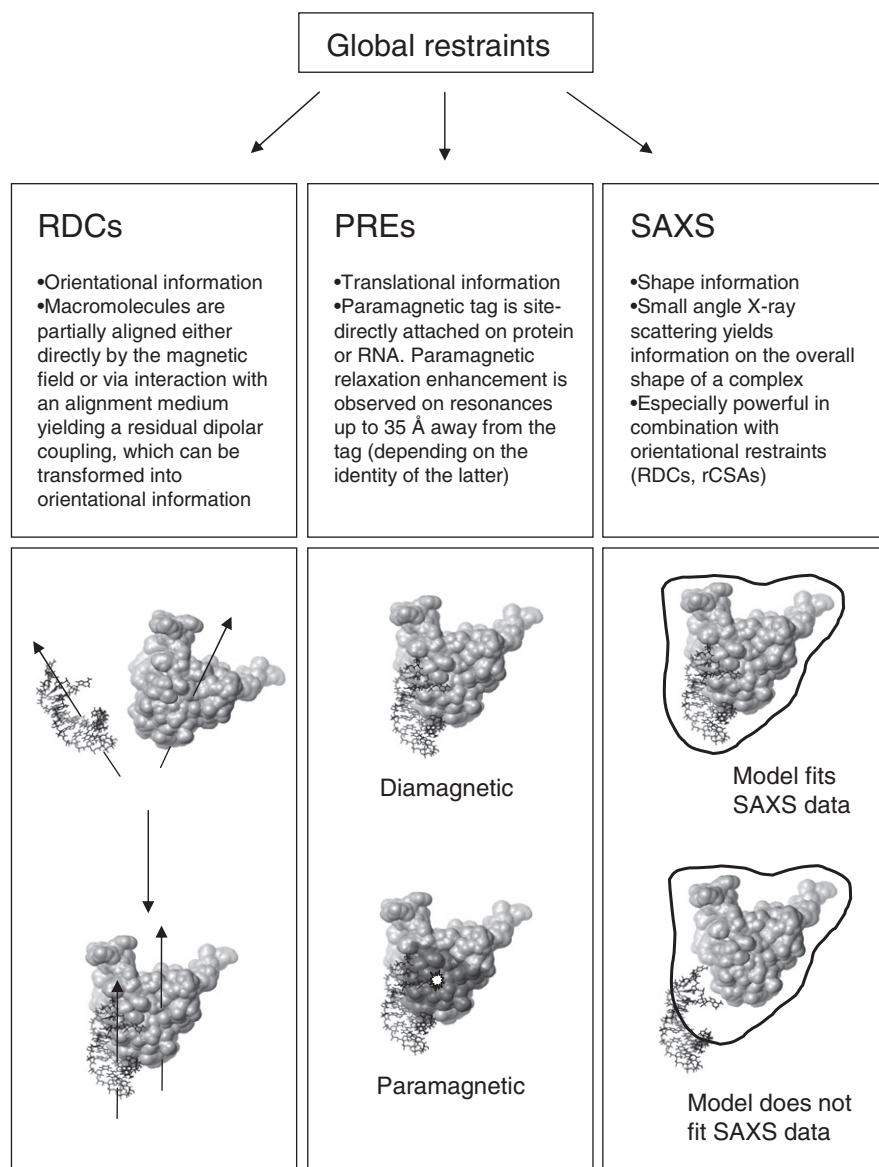


Fig. 25. Global restraints. RDCs (residual dipolar couplings) yield long-range orientational information. The relative orientation of the RNA and the protein in a protein–RNA complex can be determined using RDCs. PRE (paramagnetic relaxation enhancement) yields long-range translational information. In this example the paramagnetic tag (white star) is attached to the RNA. Resonances on the labeled protein, which are affected by the paramagnetic tag (vanishing or decreasing in intensity), are colored in black. SAXS (Small angle X-ray scattering) yields information on the overall shape of the complex.

choline/dimyristoyl-phosphatidylcholine (DHPC/DMPC) phospholipids [32,242,248], have been successfully used as well as C12E6/hexanol, which is applicable over a wide pH- and temperature-range (10–40 °C) [20,39].

In principle, RDCs can be measured between two arbitrary spins, which are close in space. As the dipolar couplings do not scale with r^{-6} but with r^{-3} , longer ^1H – ^1H interaction distances can be observed as compared to those extracted from NOEs. ^1H – ^1H long-range through-space dipolar interactions of protons separated by 7.4 Å in a 16mer DNA duplex [249] or ^1H – ^{31}P RDCs in the 30mer HIV-2 TAR (trans-activator response element) RNA [250] were observed. In addition, several one- and two-bond ^{15}N – ^1H , ^{13}C – ^1H , ^{13}C – ^{13}C and ^{13}C – ^{15}N couplings can be measured both on nucleic acids and proteins [240,251,252]. Because of the often large linewidths found in protein–RNA complexes, resulting from increased molecular size and exchange broadening, only one-bond ^{15}N – ^1H , ^{13}C – ^1H and ^{13}C – ^{13}C dipolar couplings of large magnitude have been used in the structure determination of protein–RNA complexes [10,11,18,20,22,23,27,29,32,35,39]. In addition,

the one-bond RDCs are more easily interpretable since both the inter-nuclear angle and bond length are well defined [240]. In nucleic acids, the best resolved one-bond correlations are the imino ^{15}N – ^1H , the aromatic ^{13}C – ^1H and the sugar $^{13}\text{C}1'$ – $^1\text{H}1'$ pairs [185], which are therefore best suited for extracting RDCs. On the protein side, mainly amide ^{15}N – ^1H and sometimes ^{13}C – ^{13}C or ^{13}C – ^1H couplings have been measured for this purpose [18].

One-bond ^{15}N – ^1H and ^{13}C – ^1H coupling constants can simply be extracted from a $t1$ -coupled and/or $t2$ -coupled HSQC spectrum [20,39], measured both in isotropic solution (yielding the scalar coupling constant J) and in the presence of magnetic alignment (which yields the sum of scalar coupling J and residual dipolar coupling D). In higher molecular weight systems, measuring splittings in coupled HSQC spectra is no longer possible because of spectral overlap and dramatic line-broadening of the upfield TROSY component [253]. To overcome spectral overlap, the upfield or downfield component can be extracted separately by adding or subtracting the in-phase and anti-phase doublet recorded from two different spectra [254,255]. This experiment is called IPAP for in-phase/

anti-phase. Furthermore, if the faster relaxing TROSY upfield component is too broad to be accurately measured, the coupling constants can be determined by measuring the frequency difference in Hz between a resonance in a ^{15}N - ^1H TROSY and a decoupled ^{15}N - ^1H HSQC spectrum yielding half the coupling constant [240]. One-bond amide ^{15}N - ^1H couplings can be recorded in very high molecular weight systems when they are measured on perdeuterated proteins. It is important to record the TROSY spectrum with heat-compensation to account for temperature differences, which might be generated during ^{15}N -decoupling during acquisition in a decoupled ^{15}N - ^1H HSQC. Coupling constants of one-bond aromatic ^{13}C - ^1H pairs in large systems can be extracted from both ^{13}C -lowfield components in a ^1H - ^{13}C TROSY spectrum [256,257]. Because of the large homonuclear ^{13}C - ^{13}C couplings, well-resolved $^{13}\text{C5}$ - $^1\text{H5}$ and $^{13}\text{C6}$ - $^1\text{H6}$ couplings must be acquired using a constant time version of the ^1H - ^{13}C TROSY experiment. An alternative approach has been presented, where ^{13}C - ^{13}C homonuclear couplings were removed by using a fractionally 15% ^{13}C labeled 33-mer RNA [258].

Once the RDCs are measured, they can be included into the structure calculation. Several papers and reviews have discussed and compared the different approaches for determining the alignment tensor and how to include the measured RDCs (both local RDCs of separate domains and global RDCs of the whole complex) into the structure calculation and refinement [240,259,260]. Most methods for the determination of the alignment tensor usually do not work well for the RNA component, where the set of RDCs is often limited and the predicted RDCs from initial structures do not fit well to the experimental RDCs [260]. If enough bond vectors are measured and if they are evenly distributed all over the orientational space, a histogram of the ensemble of normalized RDCs will approximate to a powder pattern, from which the magnitude of the axial and rhombic components can be extracted in the absence of any structural information [261]. In studying two TAR RNA-peptide complexes, a uniform distribution of bond-vectors all over the RNA allowed one to get an initial estimation of the magnitude and the rhombicity with the histogram of RDCs measured on the RNA [11,23]. However, the highly regular structure of an RNA double helix gives rise to a non-random distribution of orientations of the inter-nuclear vectors, preventing the use of a histogram of the bond vectors for the determination of the tensor components. An alternative method has therefore been developed by Varani and coworkers [248]. Initial values for the magnitude of the alignment tensor D_a were estimated with the histogram of the protein amide NH RDCs followed by a grid-search procedure searching over a small range of D_a and the rhombic component R to minimize the difference between the measured RDCs and those predicted from the structure calculated with the RDCs included [11,32,248,262]. Other groups have used a direct refinement against the measured RDCs in AMBER using a dipolar energy term [18,29,35,263]. The optimal alignment tensor for a fixed structure was found by optimizing the five parameters of the alignment tensor minimizing the dipolar energy term [263]. Using this approach, it is important to use a small value for the dipolar force constant and to increase the angle and the torsion force constant at the early steps of structure refinement to prevent high violations of local geometries [263]. Subsequently, during the refinement, the dipolar energy term is then increased.

The power of RDCs to refine the structures of protein-RNA complexes was first presented by Varani and coworkers [248]. Although the structure of the first protein-RNA complex solved by NMR, U1A in complex with the polyadenylation inhibitory element (PIE) RNA, was determined with high precision at the protein-RNA interface [2], the relative position of the two extruding elongated double helical stems were not well defined in this structure calculated without RDCs. One-bond ^{15}N - ^1H and ^{13}C - ^1H RDCs

were measured on this 23 kDa complex for both the protein amides and RNA imino-, base- and sugar resonances. Inclusion of RDCs did not change the precision of the structure but changed the global conformation of the RNA significantly. On the other hand, measurements of ^{13}C RDCs on the TAR RNA in a complex with a peptide significantly improved the global precision of the RNA structure from 2 Å to 1 Å root mean square deviation (RMSD) (all heavy atoms) when RDCs were included [23].

RDCs have also been used to improve the relative orientation of different protein or RNA domains in protein-RNA complexes [18,20,22]. The ensemble of structures of a 9-nucleotide single-stranded RNA in complex with the two zinc finger domains of Tis11d separated by an 18 amino-acid linker calculated from NOE and torsion angle restraints only were poorly superimposable due to the elongated shape of the complex [18]. Although the local structure was well defined, the relative orientation of both zinc finger motifs was poorly defined. However, the overall shape of the complex was defined upon inclusion of RDCs measured on the protein. To investigate the possibility of domain motions, the structures were also refined using independent alignment tensors for each zinc finger. The alignment tensors of both domains were very similar indicating highly restricted inter-domain motions. In another study, by comparing the relative protein domain orientations in the binary L11-RNA and the ternary L11-RNA-thiostrepton complex using RDCs, it was observed that the relative domain orientation of both L11 ribosomal protein domains changes upon binding to the RNA and the thiostrepton antibiotic [46]. The antibiotic thiostrepton locks the L11 conformation in a more rigid inhibitory state and rearrangement of the N-terminal domain occurs upon binding to the RNA and thiostrepton.

A sparse number of intermolecular NOEs might prevent a proper docking of the RNA onto the protein. By measuring RDCs on both the protein and the RNA in a phospholipid solution, the orientation of a double-stranded RNA with respect to dsRBD3 of Staufen could be established despite the fact that only a few intermolecular NOEs were observed [32,264]. Negative values for the amide ^{15}N - ^1H RDCs in two α -helices (where ^{15}N - ^1H bond vectors point along the α -helical axis) and positive couplings for both imino ^{15}N - ^1H and base ^{13}C - ^1H of the RNA double-helix (where base protons and imino protons point perpendicular to the double-helical axis) indicated that Staufen is bound to the dsRNA with the α -helices approximately parallel to the RNA double-helix. Following this pioneering work by the Varani lab, other protein-RNA complex structures were refined using RDCs measured on both protein and RNA to improve their mutual orientation in the complex [22,29,32,35].

Except for the PIE RNA-U1A complex [248], the power of orienting different RNA helical stems could unfortunately not be used so far for other protein-RNA complexes. Because of precipitation in presence of Pf1 phage and severe line-broadening using 5% polyacrylamide gel, high-quality ^1H - ^{13}C RDCs could be obtained only for the three short stem-loops in isolation but not for the protein-RNA complex between the nucleocapsid domain of the retroviral Gag polyprotein and the 101 nucleotide core encapsidation segment of the MoMuLV ψ site (ψ RNA) [10]. The same problem occurred in the structure determination of the 82-nucleotide segment of the 5'UTR $\mu\psi$ RNA packaging signal and the NC protein [45]. However, orienting different helical domains in nucleic acids has already been shown several times for medium size to large RNA molecules. The global structure of tRNA^{Val} [265], the hammerhead ribozyme [266] or of a branched nucleic acid Holliday Junction [267] has been determined using RDCs. In the future, it is expected that RDCs will play an important role in orienting different RNA helical segments also in large protein-RNA complexes.

RDCs can be used to gain information on the global structure but can also improve significantly the local structure. Measuring amide ^{15}N - ^1H and $^1\text{H}\alpha$ - $^{13}\text{C}\alpha$ one-bond couplings on a protein in a

complex with a 16-nucleotide DNA oligonucleotide improved the quality of the protein backbone structure significantly [243]. The percentage of residues lying in the most favorable region in the Ramachandran plot increased from 62% using no RDCs to 71% (including only ^{15}N amide RDCs) and to 79% if both ^{15}N - ^1H and ^{13}C - ^1H one-bond RDCs were included [243]. Inclusion of RDCs on the RNA part is even more important for increasing the local precision and accuracy, which has been shown on RNA or DNA [133,268,269], and protein–RNA complex structures [22,23,29,35,45,197]. Compared to proteins, RNAs yield only a limited number of NOEs, which are insufficient to obtain a precise local structure. For example, a looped-out guanine base, in which the imino resonance is not protected and therefore not observable, contains only one proton (H8), that can give NOEs. Use of RDCs in addition might therefore help dramatically to define the orientation of the base. A study refining the solution structure of the iron-responsive element RNA with RDC has discussed very extensively the impact of RDCs on both the local and global structure [260]. Molecular dynamic calculations with simulated constraints derived from two DNA duplex molecules have also shown that RDCs improve the local structure while also dramatically improving the global structure [268,269].

In summary, RDCs measured in protein–RNA complexes are very powerful for defining the global orientation of several protein or RNA domains in a macromolecular complex and for docking the RNA onto the protein, especially if only few intermolecular NOEs are observed. Additionally, orientational information obtained from RDCs can not only refine the global structure of an elongated part like a RNA helical stem, but can also contribute to increased quality of the local structure. Aiming at solving the structure of macromolecular complexes of ever increasing size, the use of RDCs will gain more and more importance and become indispensable in the future.

3.4.2. Paramagnetic relaxation enhancement

RDCs yield long-range orientational information but 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. Furthermore, measuring RDCs normally requires the addition of an alignment medium that should not interact with the macromolecule in order to prevent line-broadening or precipitation, which has already been observed [10,45,197]. In addition, RDC data have orientational degeneracy, leading to four different possible inter-domain orientations, when measured in a single alignment medium [270]. In the absence of intermolecular NOEs or other information such as mutational data, RDC data are generally not sufficient to provide an unambiguous relative orientation of two domains or components in a complex.

PRE yields long-range distance information and can be very useful in complementing the long-range orientational information obtained from RDCs [271,272]. The presence of a paramagnetic center enhances relaxation of nuclei within a radius of up to 35 Å from the paramagnetic center dependent on the identity of the latter (see Fig. 25 as an illustration) [272]. 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. This distance can be determined by measuring the intensity ratio between correlations found in the ^{15}N - ^1H or ^{13}C - ^1H HSQCs of the paramagnetic and the corresponding diamagnetic sample, which is obtained by reducing the paramagnetic center [271,272].

Paramagnetic probes can be introduced both on the protein and on the RNA molecules in order to measure long-range distances in protein–RNA complexes. Site-directed spin labeling on the protein can be performed by engineering a single cysteine mutation at dif-

ferent sites on a protein [271]. If more than one cysteine is natively present, they should be mutated into alanine. Most commonly, a nitroxide spin-label is introduced by the reagent MTSL ((1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl)-methanethiosulfonate) to a cysteine yielding a disulfide bond [38,271,273]. Other types of approaches have been the use of small peptides attached to the N-terminus of the protein or tags attached to cysteine residues, which bind paramagnetic ions such as lanthanides (reviewed recently in [274]). It is important that the cysteine mutation and the introduced tag neither disturb the protein expression or folding nor interfere with the protein–RNA interaction. This can be checked by comparing the ^{15}N - ^1H or ^{13}C - ^1H HSQC spectra of both native and spin-labeled protein in the diamagnetic state. Furthermore, it should be mentioned that removal of unreacted nitroxide label is crucial, because of the non-specific electron–proton relaxation that might occur due to the binding of unreacted nitroxide label to exposed hydrophobic patches on the protein [38]. Removal of unreacted nitroxide can be achieved by extensive dialysis or by additional purification steps.

Paramagnetic tags on the RNA side have mostly been introduced by incorporating a thiouridine base, to which a proxyl radical can be attached [95,275,276]. It has been shown, that thiouridines can also be attached to base-paired sites in a RNA helix, without disturbing the secondary structure [277]. Another promising method is to introduce the tag on the phosphate backbone by linking a proxyl radical to a modified thiophosphate either at the 5' of the RNA [278] or in an internal position [279]. The potential problem of the inherent flexibility of the tag as well as the existence of two diastereomers in the phosphorothioate linkage can be overcome by sampling the distribution of the tag orientations allowing an accurate distance measurement [279]. For a small 21 nt RNA, it was possible to separate both diastereomers by reverse-phase or anion-exchange HPLC [280]. It has to be mentioned that a deoxynucleotide has to precede the thiophosphate in order to prevent strand scission upon labeling of the phosphorothioate linkage [281]. More recently, Edwards and Sigurdsson have proposed a fast and efficient method for incorporating a proxyl label into the sugar 2'-position, which does not significantly disrupt the RNA structure and where the linkage provides moderate restriction on the motion of the probe [282,283]. Several other methods for introducing a label into RNAs like at the 5'- or 3'-end, into the 5-position of pyrimidines or the 2-position of adenines have been recently reviewed [284].

Paramagnetic relaxation enhancement has been used only in three studies of protein–RNA complexes so far [27,38,95,264]. 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 [38]. By mixing with another equivalent of each ^{15}N labeled U1A protein and unlabeled PIE RNA, Varani and coworkers measured 30 unambiguous intermolecular long-range distance constraints, which were treated very conservatively by applying an upper limit of 25 Å. The ensembles obtained with and without PREs were superimposable. However, the structural statistics improved by 10–15% upon inclusion of PRE restraints. The same group introduced a 3-(2-iodoacetamidoproxy) to a single 4-thiouridine at two different positions on a stem-loop RNA to solve the structure of the Staufen–dsRNA complex [95,264]. The modified thiouridines were attached either at the 5'-terminus or in the loop, in order to prevent the tag being too close to the interaction site of the protein, which could affect the binding. Several resonances became sharper upon reduction of the paramagnetic group with sodium hydrosulfite. However, the results were not in agreement with a single binding mode for the protein–RNA complex. The authors proposed that the higher sensitivity of the paramagnetic data allowed the identification of

Table 4
NMR-derived restraints used for the structure determination of protein–RNA complexes.

PDB	Nb struct	RNA restraints				Protein restraints				Intermolecular		Total	RMSD (Å)
		Distances	TA	Hbonds	RDC	Distances	TA	Hbonds	RDC	Distances	Hbonds		
<i>Peptide–double stranded RNA</i>													
1MNB [31]	20	301	68	30	–	57	8	–	–	26	–	490	1.49
1BIV [43]	4	691	132	22	–	259	–	–	–	102	–	1206	1.16
1ETG [6]	19	394	94	86	–	303	17	34	–	61	–	989	1.05
1ULL [41]	7	528	204	33	–	138	–	–	–	92	–	995	1.18
1A4T [7]	20	181	105	28	–	214	14	–	–	81	–	623	1.27
– ^a [24]	24	899	36	40	–	910	15	–	–	135	–	2035	0.96
1EXY [19]	12	508	161	66	–	276	11	–	–	189	–	1211	0.98
484D [42]	12	204	^b	^b	–	107	–	–	–	49	–	^b	2.79
1QFQ [33]	29	571	90	32	–	490	11	–	–	167	–	1361	0.82
1HJI [15]	20	^c	^c	^c	–	191	–	–	–	30	–	221	2.4
1G7O [16]	14	530	174	30	–	330	14	–	–	106	–	1184	1.14
1I9F [44]	7	620	^d	^d	–	99	90	–	–	8	–	^d	2.12
1NYB [9]	14	614	138	48	–	350	29	–	–	52	–	1231	1.16
1ZBN [8]	10	226	61	61	–	293	–	–	–	181	–	822	1.21
2A9X [23]	10	348	165	64	55	78	–	–	–	107	–	817	0.79
2KDQ [11]	10	464	165	78	25	204	28	–	–	143	–	1107	0.86
<i>Protein–single-stranded RNA</i>													
1K1G [26]	10	175	–	–	–	2382	169	76	–	298	–	3100	1.37
1RGO [18]	20	158	–	–	–	1480	174	0	270	196	–	2278	1.09
1T2R [25]	10	19	34	–	–	3074	194	31	–	49	–	3401	0.91
1WWD [13]	20	33	–	–	–	161	–	10	–	37	3	244	0.98
1WWE [13]	20	43	–	–	–	161	–	10	–	37	3	253	1.71
1WWF [13]	20	39	–	–	–	161	–	10	–	35	3	248	1.72
1WWG [13]	20	25	–	–	–	161	–	10	–	42	12	250	2.77
2AD9 [28]	20	34	4	–	–	1494	–	29	–	46	–	1607	0.96
2ADB [28]	20	49 ^e	6	–	–	1862 ^e	–	30	–	53	–	1986	1.07
2ADC [28]	20	88 ^f	12	–	–	4414 ^f	–	60	–	118	–	4659	1.99
2CJK [30]	25	49	8	–	–	3080	–	45	–	117	–	3299	1.41
2ERR [5]	30	119	6	–	–	1192	–	29	–	149	–	1495	0.9
2I2 Y[17]	20	32	4	–	–	692	–	47	–	40	–	815	2
2RQ4 [37]	20	40	4	–	–	1419	62	23	–	90	–	1638	0.81
2KH9 [27]	10	84	37	–	–	1407	86	38	130	34 ^g	5	1821	1.26
2KPY [14]	20	82	6	–	–	1312	–	28	–	56	–	1484	1.54
2KGO [14]	20	101	6	–	–	2180	–	26	–	63	–	2376	0.9
2KG1 [14]	20	140	6	–	–	982	–	28	–	60	–	1216	1.42
<i>Protein–double-stranded RNA</i>													
1NUM [2]	25	600	110	25	–	1737	–	21	–	96	–	2589	2.52
1AUD [3]	31	591	110	25	–	1710	–	21	–	124	–	2581	2.03
1AIT [12]	25	^h	^h	–	–	^h	–	–	–	59	–	^h	0.93
1D6K[36]	20	436	98	43	–	2623	–	28	–	121	–	3349	1.2
1EKZ [32]	36	588	127	35	5	665	–	24	54	10	–	1508	2.56
1FGU [4]	20	196	–	20	–	299	–	25	–	65	6	611	1.91
1FJE [1]	19	267	31	19	–	2593	–	76	–	157 ⁱ	8	3144	1.84
1DZ5 [38]	13	826	154	90	–	3346 ^j	–	84	–	174	–	4674 ^k	2.56
1L1C [40]	20	262	32	21	–	1485 ^l	13	–	–	120	–	1934 ^m	–
1RKJ [20]	14	424	98	38	–	2679	158	122	80	134 ⁿ	18	3743	2.18
1T4L [39]	15	699	225	34	–	1614	–	80	43	46	–	2741	1.04
1U6P [10]	20	1061	403	139	158	161	–	10	–	42	22	1996	11.81
2ESE[29]	20	426	91	31	42	1680	–	25	69	48	–	2412	1.48
2B6G [21]	10	60	93	44	–	1890	140	80	–	20	–	2327	2.22
2HGH [22]	20	434 ^o	352	58	21	867	93	18	77	97	–	2017	0.83
2FYI [35]	17	350	116	21	6	1392	81	60	48	124	–	2198	1.13
2IHX [45]	20	^p	^p	^p	–	^p	^p	^p	–	31	^p	^p	11.72
2JPP [34]	10	770	160	38	–	972 ^q	–	46 ^q	–	440	6	2432	1.2

Nb struct: number of structures used for the analysis and RMSD calculation.

RMSD: RMSD was calculated using the heavy atoms of the structured part of the protein and the RNA.

TA: Torsion angle restraints.

^a Structures were not deposited at the Protein Data Bank.^b RNA torsion angle and hydrogen bond restraints were used to restraint the duplex region as a standard A-form helix. The number of restraints used is not indicated.^c The structure of the RNA was previously solved and was kept as a fixed template during the structure calculation procedure.^d RNA torsion angle and hydrogen bond restraints were used to restraint the duplex region as a standard A-form helix and to constraint the GCAA tetraloop. The number of restraints used is not indicated.^e Including 2 intra-RNA and 12 intra-protein repulsive NOEs.^f Including 6 intra-RNA and 27 intra-protein repulsive NOEs.^g Intermolecular restraints included 13 unambiguous restraints derived from intermolecular NOEs, 18 ambiguous restraints derived from chemical shift perturbation data, and 3 long-range restraints derived from paramagnetic relaxation enhancement experiments.^h 719 distance restraints were used for the structure calculation including 59 intermolecular restraints. The number of intraRNA and intraProtein restraints are not indicated. Torsion angle restraints were used for the RNA but their number is not indicated.ⁱ Including 7 repulsive NOEs.^j Including 90 intermolecular protein–protein distance restraints to define the dimer interface.

(continued on next page)

Table 4 (continued)

^k Including 30 long-range distance restraints derived from paramagnetic relaxation enhancement experiments.

^l Including 69 intermolecular protein–protein distance restraints to define the dimer interface.

^m RMSD for the protein–RNA complex was not reported. Only one structure was deposited at the Protein Data Bank.

ⁿ Including 8 repulsive NOEs.

^o 129 distance restraints were derived from NOESY spectra. 305 additional intra-RNA restraints were used based on previously determined structures of certain regions of the RNA.

^p The ensemble of structures was calculated using a total of 680 distance restraints (including 31 intermolecular), 608 hydrogen bond restraints, 239 dihedral angle restraints and 170 inter-phosphate restraints. The number of restraints used as intra-RNA or intra-protein are not indicated.

^q Including 132 intermolecular protein–protein distance and 10 hydrogen-bond restraints to define the dimer interface.

a minor conformer. Therefore, the paramagnetic data was not used for further structure calculation of the complex. Very recently, Butcher and coworkers attached a 3-(2-iodoacetamidopropyl) to a single 4-thiouridine on a six nucleotide RNA [27]. They included 3 spin-label derived restraints for docking the RNA onto the protein in their structure calculation.

Although only two NMR studies of protein–RNA complexes have used PRE data, the power of PRE has been recognized widely in the field of nucleic acids and DNA–protein, protein–protein and protein–ligand complexes. In addition to assisting global protein fold determination [273,285], paramagnetic relaxation enhancement has been widely used for positioning components in macromolecular complexes relative to one another, especially when no or only a sparse number of intermolecular NOEs or RDC measurements were available or not sufficient [275,286–288]. Clore and coworkers in particular made very interesting usage of PRE measurements for elucidating transient macromolecular interactions in protein–protein and protein–DNA complexes [272,289].

A major drawback of using PRE as a precise constraint is the inherent flexibility of the paramagnetic tag. Most groups have therefore used large error boundaries of ± 4 Å [273] or only upper limits of such as 25 Å [38]. A theoretical and computational strategy, which treats the intrinsically flexible paramagnetic label in a multiple conformer representation, has been used successfully to refine a protein–DNA complex (the target function was incorporated into the molecular structure determination package X-plor-NIH and can be downloaded from <http://www.nmr.cit.nih.gov/xplor-nih>) [290]. The structure could be refined with similar precision and accuracy using either PRE or RDC restraints in addition to a single intermolecular NOE. Including both PRE and RDC restraints further increased the accuracy, which highlights the complementary information content provided by both constraints. It should be mentioned, that using a single-conformer representation for the spin-label resulted in a good agreement between the observed and calculated PRE, but at the expense of coordinate accuracy [290]. To exploit the full repertoire of PRE in structure refinement, it is therefore crucial to treat the tag in a multiple conformation representation.

The power of using PRE as an independent long-range distance constraint to obtain translational and orientational information of two or several domains or macromolecules in respect to one another or in structure refinement has been appreciated both on the fields of protein–protein and protein–DNA complexes and is expected to play an important role also in the future for protein–RNA complexes.

3.4.3. Combining several methods to reach complexes of ever increasing size

The effect of global RDC or PRE restraints in supplementing short-range NOE data has been highlighted in the previous paragraphs. However, a 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 even more powerful and to gain in importance for studying sys-

tems of increasing molecular weight, although such combinations of techniques have not been used yet for protein–RNA complexes.

Using a combination of both RDCs and PREs, a 58 kDa protein–protein complex could be docked [291]. Seven long-range PRE distance constraints, which were distributed over a large portion of the binding interface, were sufficient to constrain both proteins translationally and RDCs were then used to overcome the ambiguity problem in orienting the two domains or macromolecules relative to each other [259,270]. The RDCs were measured in a single alignment medium.

Small angle X-ray scattering 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), which is the difference between the chemical shift anisotropy measured in isotropic and aligned media [292,293]. SAXS is particularly useful for nucleic acids having low proton spin density and therefore lacking global translational information despite the measurement of RDCs or rCSAs. SAXS data are particularly suited for larger macromolecules or complexes due to the quadratic dependence of the signal intensity on the molecular weight [294]. However, SAXS is also very sensitive to small amounts of aggregation. The first combination of SAXS with NMR data was used to build the overall shape of a protein–protein complex (see Fig. 25 as an illustration) [292]. The measured RDC data served to reduce the angular degrees of freedom, whereas the SAXS data constrained the translational degrees of freedom. So far, SAXS data has not been used in combination with NMR data for protein–RNA complexes. However, it has been successfully applied to refine the global structure of multi-domain proteins [292,295], nucleic acids [293], RNA–RNA [296], and protein–DNA [297] complexes. The interface and the global structure of an RNA–RNA complex could be determined with SAXS in the absence of any intermolecular NOE [296]. The SAXS refined structures had an increased precision. The introduction of SAXS data improved the backbone RMSD from 1.8 to 1.3 Å. The application of SAXS data on a two-domain protein and inclusion of RDCs supplemented by a small number of HN–HN and CH₃–CH₃ NOEs, resulted in an increase of accuracy and in a tighter packing of both domains with respect to each other, probably because of the lack of inter-domain NMR restraints [298]. However, SAXS data alone was insufficient for independent structure determination. The inclusion of at least one set of RDC data was crucial for correct positioning of both domains [298]. A low resolution structure of the multi-domain RNA-binding protein PTB was determined by fitting the high-resolution NMR structures of individual RRM1 and RRM2 with the scattering data [299]. The SAXS data, for example, confirmed that domains 3 and 4 of PTB form a compact structure [164], whereas RRM1 and RRM2 have loose contacts to the rest of the protein. Small angle scattering is expected also to support the structure determination of large protein–RNA complexes by NMR in the near future.

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 now allow NMR of pro-

tein–RNA complexes to enter into new dimensions of space (high molecular weight complexes) and time (transient interactions, dynamics).

3.5. Structure determination

3.5.1. Experimental restraints

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 (detailed in Section 3.2), hydrogen-bond and dihedral angle restraints (Section 3.3), orientational restraints (Section 3.4.1) and long range distance restraints (Section 3.4.2). Table 4 displays the number and types of restraints that have been used for the structure determination of protein–RNA complexes by NMR.

Distance restraints have been the major source of experimental restraints used to solve the NMR structures of protein–RNA complexes. The distance restraints can be subdivided into intramolecular protein–protein, intramolecular RNA–RNA and intermolecular protein–RNA restraints (see Section 3.2). These restraints provide essential information required to define the secondary and tertiary structures of the protein and the RNA and to define the interface between the protein and the RNA. The conversion of NOE volumes into distance restraints is, however, not straightforward. The distance between two protons is directly correlated to the NOE intensity/volume extracted from the NOESY spectra [300]. However, many additional factors, such as spin diffusion and conformational exchange, can influence the NOE signal intensity, and it is therefore very difficult to define precise distances based on the NOE intensity. In macromolecular complexes, intermolecular NOEs are even more difficult to interpret in terms of distances, especially when the complex formation is in fast or intermediate exchange on the NMR timescale because the intensity of the intermolecular NOEs might partly be influenced by the exchange rate. In addition, in the case of flexibility at the interface, NOEs might reflect multiple conformations and the intensity of such NOEs can therefore not be easily translated into precise distances. Therefore, a general approach for obtaining NOE-derived distance restraints is to classify peak volumes into categories, such as weak, medium, or strong and then to define an upper distance limit for each class of NOEs. For example, distances derived from strong, medium and weak NOEs are below 3, 4.5 and 6 Å, respectively. In this case, the lower bound limit for all NOEs is the van der Waals radius between two protons (1.8 Å). In order to perform a classification of the distances derived from NOEs, known proton–proton distances are generally used. For example, in proteins, the distance between two sequential amide protons is 2.8 Å in helices and the intra-residue distance between a H α and the amide proton is 2.2 Å in β -sheets [300]. Similarly, intramolecular RNA distances can be used. For example, the intra-residue distance between the H5 and the H6 protons of pyrimidines is 2.4 Å and the intramolecular distance between the H1' and H8 protons of purines in the trans conformation is 3.5 Å. Concerning the intramolecular protein–protein distance restraints, methods have been developed to iteratively assign and calibrate NOEs from NOESY spectra, convert these NOEs into distance restraints and then calculate structures using these distance restraints. The two main methods that allow for “automated assignment” of NOEs in protein structure calculations are ARIA [301,302] and CYANA or ATNOS/CANDID [303,304]. These programs generate a list of intramolecular protein distance restraints that can be used in the structure calculation of the protein–RNA complex. Concerning the intramolecular RNA–RNA and the intermolecular NOEs, such programs are less suited, mainly because of the low density of protons present in RNAs, and the NOE assignment and calibration is therefore generally performed manually by analysis of the NOESY spectra.

Hydrogen-bond restraints can be derived from the direct detection of scalar couplings across the hydrogen bond by NMR or indirectly derived from NMR experiments such as hydrogen–deuterium exchange experiments in the case of proteins or by analysis of the presence of imino protons in the case of RNA (see Section 3.3.1). To define an hydrogen bond restraint, a distance between the donor hydrogen and the acceptor heavy atom is generally set between 1.8 and 2.4 Å, and a distance between the donor and acceptor heavy atoms is set between 2.4 and 3.4 Å. Intramolecular RNA hydrogen-bonds have been widely used for structure determination of peptide–double-stranded RNA and protein–double-stranded RNA complexes (Table 4). In most cases, these restraints were generated by the identification of imino protons and from NOE patterns in order to enforce base-pairing interactions (see Section 3.3.1) [1–4,6–8,10,19–24,29,31–35,38–44]. However, in some cases, these intra-RNA hydrogen bonds were measured using HNN COSY experiments [9,16,22,36]. Intramolecular protein hydrogen bond restraints have also been widely used for the structure determination of protein–RNA complexes [1–6,10,13,14,17,20–22,25,26,28–30,32,34–39]. In this case, these hydrogen-bonds were derived from hydrogen–deuterium exchange experiments. Finally, in some cases, intermolecular hydrogen-bonds have been used for the final structure calculations [1,4,10,13,20,34,40]. In these cases, preliminary structures were generated without these hydrogen-bond restraints and at a latter stage, when a possible hydrogen-bond was observed in the preliminary structures and found to be consistent with observed chemical shift changes, these restraints were then included in the structure calculation.

Dihedral or torsion angle restraints can be directly derived from the measurement of J -coupling constants by NMR, or indirectly from the analysis of COSY or TOCSY spectra (see Section 3.3.2). The J -coupling constants are translated into torsion angles using the appropriate Karplus relation and can be used as constraints, for example, for the backbone conformation of the protein and/or the RNA (see Section 3.3.2). In almost all peptide–dsRNA and protein–dsRNA complexes, dihedral angle restraints were derived from COSY and TOCSY spectra analysis and used to restrain the stem region of the RNA [1,2,4,6–10,12,16,19–24,29,31–36,38–45]. In addition, some dihedral angles derived from the analysis of COSY and TOCSY spectra were also used in protein–ssRNA complexes to constrain the backbone RNA to C2' or C3' endo conformations [5,14,17,25,28,30,37]. For most peptide–RNA complexes and three protein–RNA complexes, J -couplings were measured, mainly using HNHA experiments (see Section 3.3.2) and used to restrain backbone dihedral angles of the peptide [6,7,9,16,19,24,33,44] or the protein [18,26,37]. In other cases, backbone dihedral angles were predicted using the program TALOS [305] based on backbone chemical shifts [11,20–22,25,35,40].

Orientational restraints derived from residual dipolar coupling constant measurements are used to define the relative orientation of the protein and the RNA in the complex and to refine the local geometry of the complex (see Section 3.4.1). These restraints are generally only incorporated in the final stages of the structure calculation or in the structure refinement steps. Ten NMR structures of peptide–RNA or protein–RNA complexes have been solved with the use of orientational restraints [10,11,18,20,22,23,29,32,35,39]. In some cases, residual dipolar couplings were measured only on the RNA [10,11,23] or only on the protein [18,20,39]. In four cases, residual dipolar couplings were measured on both the protein and the RNA allowing the definition of the relative orientation of both components in the complex [22,29,32,35].

In addition, the structures of two protein–RNA complex were solved using long-range distance restraints derived from paramagnetic relaxation enhancement measurements (see Section 3.4.2) [27,38].

experimental NMR data and a term that prevents close contacts between atoms. It should be noted that, in the case of nucleic acids, DYANA parameter sets do not contain a bond between the C3' and the O3' atoms in order to allow the sugar ring to remain flexible [310]. Therefore, when using DYANA or CYANA, artificial distance restraints between these two atoms (1.44 Å) for each sugar ring of the nucleic acid molecule need to be added.

Two different procedures for calculating the structures of protein–RNA complexes have been described. In one case, the structure of the complex is calculated directly starting with both components in an extended conformation and separated by a flexible artificial linker [1,2,4–8,10–13,16,17,19–26,28–39,41,42,44,45], while in the other case, structure calculation is performed on each component separately and then both structures are docked using intermolecular restraints [9,15,18,40,43]. Docking approaches have also been used to generate structural models of the complex between the protein Kid and a short RNA and of the complex between protein L11 and the 23S rRNA [46,47].

3.5.3. 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. The force fields used to refine the structure of protein–RNA complexes need to be properly balanced to accommodate the properties of both types of molecules (RNA and protein) [312]. Two force fields are most commonly used for protein–RNA complexes: AMBER and CHARMM. Both force fields have been optimized for treating nucleic acids. The AMBER force field commonly used in protein–RNA structure refinement is the ff99 force field [313] that is derived from the original ff94 force field [314]. Concerning CHARMM, two force fields can be combined. These are the CHARMM22 protein force field [315] and the CHARMM27 nucleic acid force field [316]. All these force fields are well optimized for both proteins and nucleic acids and are therefore highly suitable for the refinement of protein–RNA structures solved by NMR. Refinements of the structures using these force fields were performed using the SANDER (simulated annealing with NMR-derived energy restraints) module of the molecular dynamics simulation package AMBER [317,318], or using XPLOR and XPLOR-NIH that include parameter sets derived from AMBER and CHARMM force fields. Both the SANDER module of AMBER and XPLOR can incorporate most NMR-derived restraints such as unambiguous and ambiguous distance restraints, angle and torsional restraints, pseudo-contact shift restraints, RDC restraints, and CSA restraints. Additionally, both packages include different solvent models for refinement.

Generally, the refinement of a protein–RNA complex is performed by restrained molecular dynamics that consists of a simulated annealing protocol that can be preceded and/or followed by an energy minimization step. The refinement is also generally performed in explicit solvent representation or implicitly using the Generalized-Born solvation model (see following Section). For further information on molecular dynamic simulations and force fields, we recommend consideration of more specialized reviews [312,319–321].

3.5.4. Using solvation models in structure refinement

Nucleic acids are highly charged molecules and protein–RNA interactions are often driven by electrostatic interactions. Furthermore, the specificity of the recognition is often determined by intermolecular hydrogen-bonds. In order to optimize the electrostatic non-bonded energy term of the structures, the use of solvation models during the refinement of NMR structures is therefore of particular importance [322,323]. There are various approaches to introducing solvents (generally water) into the refinement proce-

dure. The most accurate approach is the use of explicit solvent where the structure of the complex is refined in a box of water molecules [324]. The main drawback of using explicit solvent, however, is that refinement protocols are time consuming, and most of the calculation time is spent on the computation of solvent–solvent interaction and not on the critical protein–RNA, protein–water or RNA–water interactions. To overcome this, models have been developed where water molecules are not incorporated in a box but are added in a shell of 5–8 Å around the biomolecules, reducing the number of water molecules in the refinement and allowing for a significant reduction of the computational time [325]. This model is implemented, for example, in the water refinement protocol of CNS that uses the OPLS force field [326] and has been applied for the refinement of several protein–RNA structures [25,46,47].

Another approach consists of using a continuum solvent model, where the water is treated implicitly. In this case, generally, the protein–RNA complex degrees of freedom are treated explicitly while the solvent degrees of freedom of the solvent are not. There are different methods used to treat solvent implicitly in which the high dielectric solvent is approximated by continuum electrostatics models that interact with charges at the surface of the molecule or the complex. The Generalized Born (GB) model [327], which is an approximation of the Poisson–Boltzman equation, has been commonly used for the structure refinement of protein–RNA complexes in implicit solvent (see Table 5) [5,9,17,18,22,28,29,34,35,37]. GB solvation models are implemented in both AMBER and XPLOR packages.

3.5.5. An example of structure calculation and refinement protocol

Fig. 26 shows a flowchart describing the procedure for structure calculation, refinement, and validation used in our laboratory to

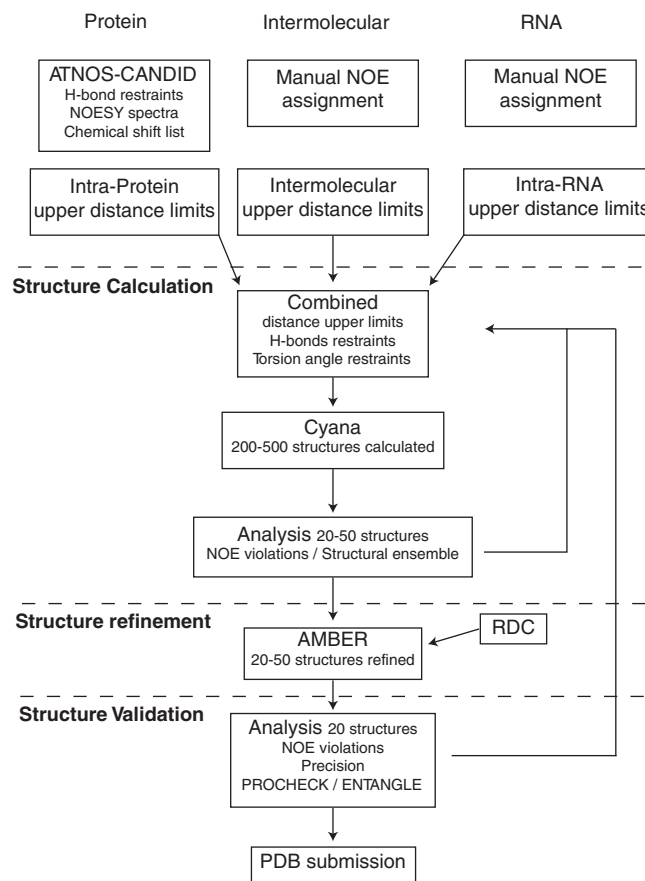


Fig. 26. Example of structure calculation protocol used to determine structures of protein–RNA complexes in our laboratory.

solve the structures of eleven protein–RNA complexes [5,14,17,28,29,34,35]. Intra-protein distance restraints were generated with the software ATNOSCANDID [303,304] using NOESY spectra and a list of protein chemical shifts. In addition, a list of intra-protein hydrogen-bond restraints, based on hydrogen-deuterium exchange experiments and the analysis of preliminary structures, was often used. Intra-RNA and intermolecular distance restraints were generated by manual assignment of the NOEs. These intra-protein, intra-RNA and intermolecular distance restraints were then combined and used, together with hydrogen-bond and torsion angle restraints, to generate preliminary structures of the complex using the program CYANA [304]. Typically, between 200 and 500 structures were calculated and the 20–50 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 structure was then subjected to a structural refinement procedure.

Structure refinement was performed using the SANDER module of the AMBER software [317,318]. Inputs for structure refinements consisted of the 20–50 structures derived from CYANA, the distance, hydrogen-bond, and torsion angle restraints and, when available, the residual dipolar coupling restraints. Structure refinement were performed using the ff99 or ff94 force fields [313,314] in combination with a GB solvation model [327]. The refinement procedure consisted of a simulated annealing protocol that was optimized for nucleic acids [252] followed by a short energy minimization. Following the refinement procedure, structures were analyzed in terms of energy, NOE violations, and structure precision. As for structure calculation, analysis of the refined structures, especially the NOE violations, could be used to refine the distance restraints and restart another cycle of structure calculation and refinement.

For a complex below 15 kDa, one round of 500 structure calculations (about one minute per structure) followed by 50 structure refinements (about two hours per structure) was typically achieved in 2–3 days using one CPU. However, most structure calculation and refinement packages can be used on a multi-processor cluster reducing considerably the calculation time.

3.6. Validation, precision and accuracy

Calculation and refinement of NMR-derived structures implicitly incorporate and reconcile two different types of criteria. On one hand, structures calculated should agree with the structural restraints that are derived from the NMR experiments. On the other hand, structures should fulfill geometrical and structural requirements that are typically driven by the force field used during the calculation and refinement of the structures.

In contrast to X-ray structures, NMR structures are presented as an ensemble of structures (generally between 10 and 20) that match best the experimental data. Generally, a large number of structures are generated during the structure calculation and refinement steps and various selection criteria are then applied to select the ensemble of “best structures”. Ensembles of structures for protein–RNA complexes have been selected based on different criteria such as the agreements with the NMR experimental restraints, the total force field energy, or the geometrical quality of each structure.

Recommended validation criteria for NMR structure determination have been described [328]. 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. A variety of validation protocols and software have been developed for proteins (for specialized reviews, see [329,330]), whereas for nucleic acids, only a few validation programs are available and are mentioned in Section 3.6.3.

3.6.1. Accuracy and precision of the structure ensemble

An important aspect for the analysis of NMR structure ensembles is the uncertainty of the molecular coordinates within the ensemble. Two main aspects concerning the uncertainty are the precision and the accuracy of the ensemble of structures. The precision of an NMR ensemble reflects the convergence of the different structural models. Accuracy, on the other hand, is the measure of the deviation between the calculated structures and the “real” structure.

The accuracy of a structure is generally difficult to assess, mainly because the definition of a “real” structure is not straightforward. In some cases, a comparison of the structure ensemble can be compared with similar structures solved by X-ray crystallography or NMR and this comparison can provide an estimate of the accuracy. For example, the NMR structure of the complex between the protein U1A and an RNA containing an internal loop [2,3] was compared with the X-ray structure of the complex between the same protein and a hairpin RNA [331]. Although the two RNAs were different, the single-stranded RNA sequence recognized by the protein was identical in both complexes and a comparison of both structures showed that the intermolecular contacts observed in both complexes are very similar. Similarly, a comparison of the NMR structure of the complex between the transcription factor IIIA (TFIIIA) and the 5S RNA [22] with the same structure solved by X-ray crystallography [332] showed similar intermolecular contacts. Another example illustrating the accuracy of NMR structures is the structure of the protein Vts1 in its complex with RNA. Three structures of this complex were solved independently and at the same time, two by NMR [21,29] and one by X-ray crystallography at 2.0 Å resolution [333] (Fig. 27A). In all structures, Vts1 binds a conserved pentaloop (CUGGC in the NMR structures and UUGAC in the X-ray structure). However, striking differences can be observed when comparing these structures. One NMR structure [29] is very similar to the X-ray structure (Fig. 27B, left). The backbone RMSD for the pentaloop between the two structures after fitting on the protein backbone is 1.44 Å. Most intermolecular contacts are similar in both complexes. The other NMR structure [21], however, differs significantly from the X-ray structure (Fig. 27B, right). The backbone RMSD for the pentaloop between the two structures after fitting on the protein backbone is 3.53 Å and the intermolecular contacts are different from those found in the other two structures. One explanation of these differences is that the NMR structure (pdb:2ESE) that is similar to the X-ray structure was solved using 48 intermolecular NOEs, as well as 69 and 42 RDC restraints for the protein and the RNA, respectively, while the NMR structure (pdb:1B6G) that differs from the other two was solved with only 20 intermolecular NOEs and no RDC restraints. In another example, the structure of the complex between the L30 ribosomal protein and an RNA was solved by NMR [334]. Later, two X-ray structures of homologous proteins in complex with RNA were solved [335,336] and showed that although the global orientation was similar to the NMR structure, there were significant differences in hydrogen bond and stacking interactions. The structure of the L30 protein in complex with

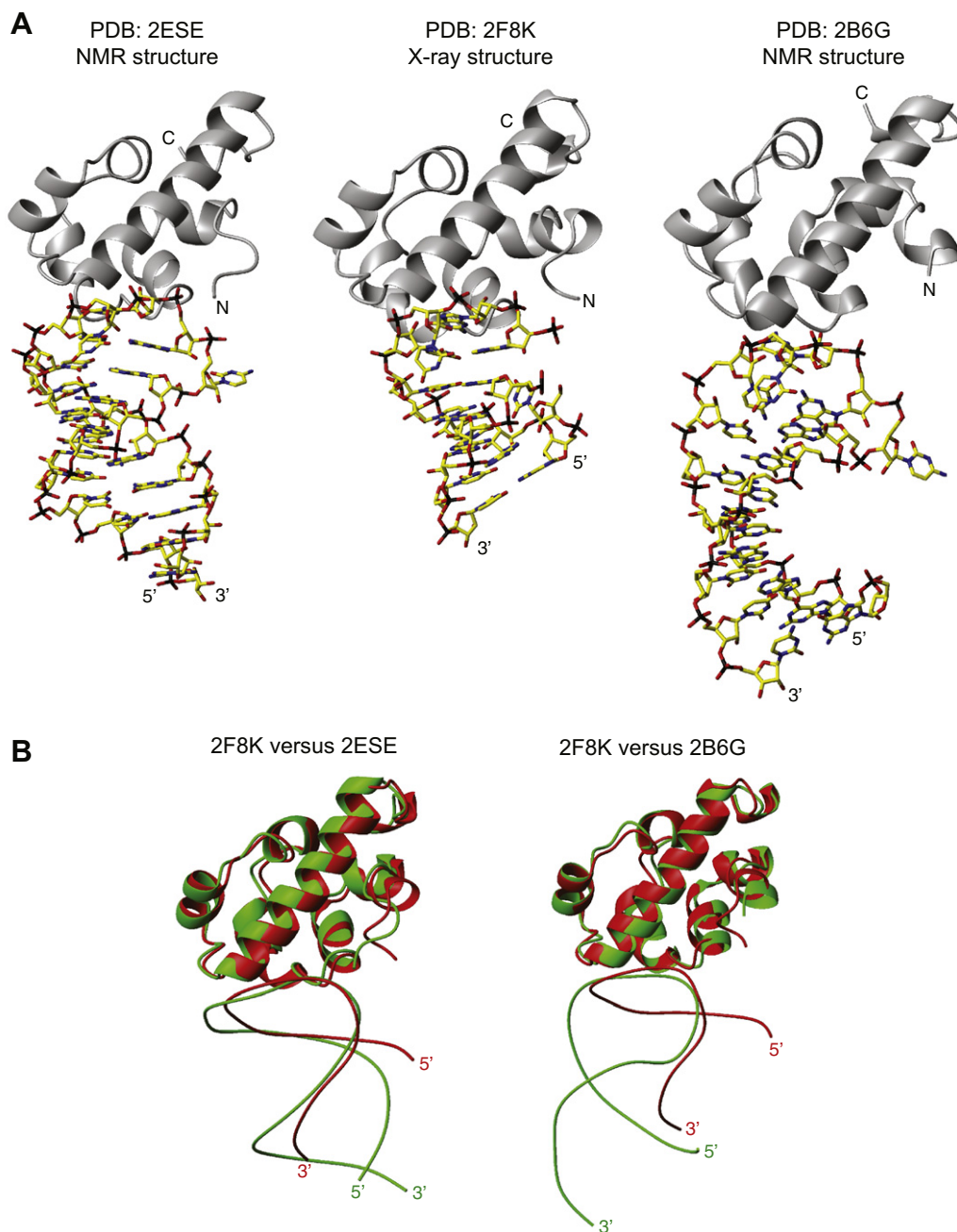


Fig. 27. Comparison of the NMR and X-ray structures of Vts1p in complex with RNA. (A) Structures of the different Vts1p–RNA complexes [21,29,333]. Vts1 is represented as a ribbon structure in grey, RNA is represented as a stick structure in yellow. (B) Superposition of the X-ray structure (red) [333] with the NMR structures (green) (left: 2ESE [29], and right: 2B6G [21]). Fitting was performed on the protein backbone. Figures were generated with molmol [470].

RNA was therefore solved by a combination of X-ray and NMR refinement and showed that the original NMR structure was incorrect due to a single misassigned imino proton [337].

The precision of NMR ensembles has generally been estimated by calculating the root mean square deviations (RMSD) of the ensemble of structures where low RMSD values reflect a precise ensemble of structures (Table 4). The precision of NMR structures mainly depends on the size of the system studied, the number of distance restraints used in the calculation and the intrinsic dynamics of the molecules in the complex. In the case of protein–RNA complexes, different RMSDs were reported. Most commonly, RMSDs were calculated for each component individually, reflecting the precision of the structure for each component, and for both components together, reflecting the precision of the complex

structure. Some structures of protein–RNA complexes are very precise (heavy atom RMSD below 1 Å) because of use of a large number of NOEs, especially intermolecular NOEs. For example, the structure of the N peptide of the bacteriophage λ in complex with its BoxB RNA target is among the most precise NMR structure of a protein–RNA complex determined to date [33]. This complex has a molecular weight of 8.9 kDa and consists of a 36-residue peptide bound to a 15-nucleotide RNA. A total of 1361 restraints were used for the structure determination, including 167 intermolecular distance restraints, that were derived from NOE cross-peaks (Table 4, PDB: 1QFQ). In addition, 11 intra-protein and 90 intra-RNA torsion angle restraints were derived from *J*-coupling measurements (peptide) or deduced from typical NOE patterns (RNA). Finally, 32 intra-RNA hydrogen-bond restraints were derived from hydrogen–deu-

terium exchange experiments on RNA imino protons. Structure calculation was performed with X-PLOR [306] and an ensemble of 29 structures was deposited in the Protein Data Bank (PDB code: 1QFQ). The precision of the structure ensemble was assessed by measuring various RMSDs. The values of the RMSDs of the peptide or the RNA heavy atoms are 0.76 and 0.67 Å, respectively, while the heavy atom RMSD of the complex is 0.82 Å. Another precise structure of a protein–RNA complex determined by NMR is the complex between the protein Fox-1 and its target RNA, UGCAUGU [5]. The RRM domain of Fox-1 is composed of 100 amino acids and the RNA is seven nucleotides long. A total number of 1495 restraints, including 149 intermolecular distance restraints, 29 protein hydrogen-bond restraints and six RNA torsion angle restraints were used for the structure calculation (Table 4, PDB: 2ERR). The structure was calculated with CYANA [304] and refined with AMBER [318] using a GB solvation model [327], and an ensemble of 30 structures was deposited at the PDB (PDB code: 2ERR). The heavy atom RMSD for the ensemble is 0.95, 0.55, and 0.90 Å for the protein, the RNA and the protein–RNA complex, respectively.

In contrast, lower precision structure ensembles were also obtained using NMR restraints. This is the case of the protein–RNA complex between the protein Staufen and an aptamer RNA [32]. The structure was calculated using 1508 restraints (distance, hydrogen-bond, torsion angle and RDC restraints) but only 10 intermolecular NOEs could be unambiguously assigned and were very weak in NOESY spectra. This low number of intermolecular NOEs was due to the fact that the complex formation involves mainly functional groups at the end of long side chains, which are difficult to assign and generally more flexible, and the RNA phosphodiester backbone where resonances are sparse and overlapped. Furthermore, a related X-ray structure shows that many intermolecular contacts are mediated by water molecules [338]. Therefore, the heavy atom RMSD for the ensemble of 46 structures was determined to be 2.56 Å. However, even at this precision, the molecular basis for RNA recognition by Staufen could be derived and a comparison with the related X-ray structure showed that the main features of the protein–RNA recognition are present in the NMR structure.

Sometimes, an interface RMSD was reported to assess the precision of the protein–RNA interface, although both components might not have a refined relative orientation [10,45]. This was the case for two protein–RNA complexes, the complex between the MoMuLV NC protein in complex with the MoMuLV 101-nucleotides Ψ RNA [10], and the complex between the Rous Sarcoma virus (RSV) NC protein in complex with the RSV 75-nucleotides Ψ RNA [45]. In both cases, the RNA is composed of three stem-loops connected by flexible linkers and the global RMSD of the protein–RNA complex is therefore very high (heavy atom RMSD of 10.8 and 10.7 Å, respectively). However, in the case of the MoMuLV complex, when considering only the atoms at the interface (22 amino acids of the protein and four nucleotides of the RNA) the heavy atom RMSD becomes 1.07 Å indicating that the interface of the complex is precise. In the case of the RSV complex, the precision of the structure was more difficult to assess because the RSV NC protein is composed of two independent domains separated by a flexible linker. In this case, the precision of the interface could be assessed by measuring two different RMSDs: one including the first domain of the NC protein (22 amino acids) and the nucleotides bound by this domain (five nucleotides) giving an heavy atom RMSD of 0.89 Å, and another RMSD including the second domain of the NC protein (17 amino acids) and the nucleotides bound by this domain (four nucleotides) giving an heavy atom interface RMSD of 3.45 Å.

3.6.2. Validation of the structures against experimental restraints

In structure determinations of protein–RNA complexes, three types of NMR-derived restraints have mainly been used, namely the distance, torsion angle and orientational restraints (see Sec-

tion 4.4.1). To date, the NOE-derived distance restraints have been the most important source of experimental information for the structure determination of protein–RNA complexes by NMR. The most common approach to validate distance restraints consists of generating a list of violations and analyzing the number and magnitude of these violations. For each structure of the ensemble, distances are measured and compared with the upper bound limit of the restraints. A distance cut-off is then set to evaluate whether a distance restraint is violated. Generally, maximal distance violations are not larger than 0.3 or 0.5 Å. Most NMR structure calculation and refinement programs (XPLOR, CYANA, AMBER, etc.) include validation routines and create a list of distance violations for a detailed analysis.

Similar to distance restraints, dihedral angle restraints are generally validated by comparing the angles defined by the restraints with the angles calculated in each structure of the ensemble. A list of dihedral angle violations is generally provided together with the list of distance restraint violations by most structure calculation and refinement softwares.

When orientational restraints derived from RDC measurements are used during the structure calculation, a validation of the structures against these restraints is also necessary. This validation is generally achieved by determining a RMSD between the experimental and the back-calculated RDCs, which reflects the agreement between the orientational restraints and the ensemble of structures [10,11,18,20,22,23,29,32,35,39,46]. In addition, programs, such as MODULE [339] or REDCAT [340] are available that validate structural ensembles against experimental orientational restraints. So far, the results of these programs have not been reported for the structures of protein–RNA complexes determined using RDC restraints.

An additional, more indirect 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. In recent years, much effort has been expended in developing software that predicts chemical shifts of proteins solely based on 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. In macromolecules, chemical shifts are dependent on the structure of the molecule through many factors, such as ring current effects, hydrogen bond effects, electrostatics, etc. There are several programs that predict chemical shifts from a protein structure, such as SHIFTS [341,342], SHIFTX [343], PROSHIFT [344], or SPARTA [305]. These programs have been developed for proteins and predict nitrogen, carbon and proton chemical shifts from coordinate files. These predictions can then be compared with the experimental chemical shifts and used as a validation for structure quality. All these programs, however, are restricted to the prediction of protein chemical shifts and not nucleic acid chemical shifts. Nonetheless, using a dataset of 28 RNA structures, it was shown that proton chemical shifts of RNAs possessing different structures (stem-loop, bulges, pseudoknot, base-pair mismatches or quadruplexes) can be predicted with a good accuracy and precision from a coordinate file [345]. Therefore, it should be possible to use chemical shift prediction as a tool for the validation of protein–RNA structures. With the growing number of protein–RNA complexes deposited in the protein data bank, it would be interesting to investigate the validity of these prediction methods on such complexes. Of particular importance, it would be very useful to test whether chemical shifts of amino acids and nucleic acids that are at the interface and involved in intermolecular contacts can be accurately predicted. To our knowledge, such a study has not been performed yet. However, a manual analysis of the relation between chemical shift and structure can be performed to validate or invalidate the structure. For example, when unusual chemical shifts are

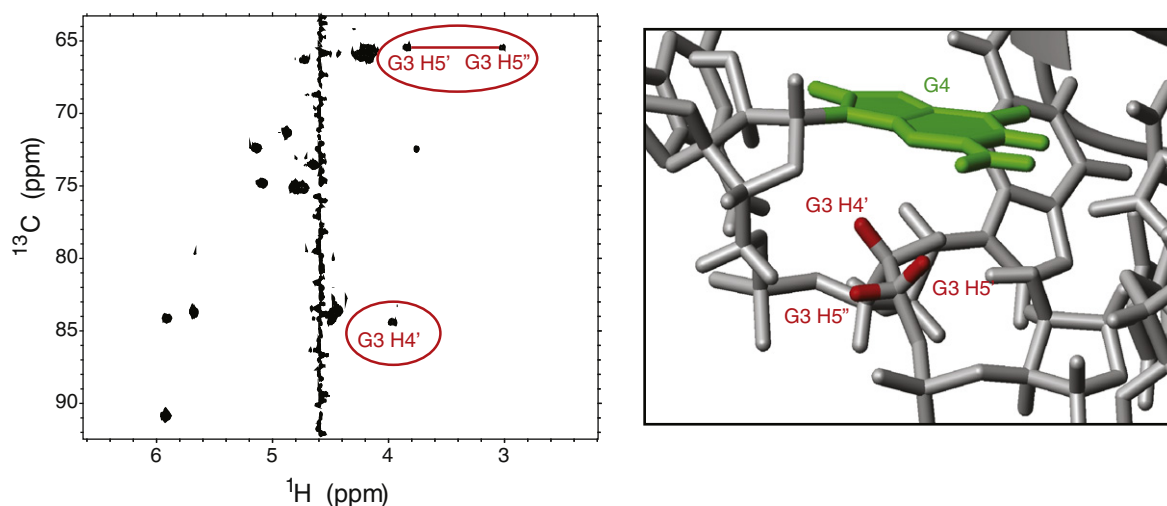


Fig. 28. Example of correlation between unusual chemical shift and structural feature. Left: ^{13}C - ^1H HSQC spectrum of a sugar labeled CGGGAU RNA in complex with hnRNP F qRRM1 [14]. Peaks corresponding to G3 H4', H5' and H5'' are labeled and circled. Right: Structure of the hnRNP F qRRM1-AGGGAU complex showing that G3 H4', H5' and H5'' (red) are experiencing the ring current effect of G4 base (green). Spectra were recorded on a 900 MHz spectrometer. Figures were generated with molmol [470].

observed in the NMR spectra of a complex, an explanation should become apparent from the analysis of the structure. Very recently, we performed an NMR study of the complex between the protein hnRNP F and a G-tract RNA that revealed unusual chemical shifts for some RNA sugar protons. A ^{13}C - ^1H HSQC spectrum of a sugar ^{13}C labeled CGGGAU in complex with hnRNP F quasi RNA recognition motif (qRRM) 1 indicated that the protons H4', H5' and H5'' of G3 had an unusual upfield chemical shift (Fig. 28). The NMR structure of the complex between hnRNP F qRRM1 and an AGGGAU RNA showed that these protons are located immediately below the base of G4 experiencing the ring current effect of this base [14]. This correlation between chemical shift and structural feature also indirectly validated the conformation of the G4 base in the structure.

3.6.3. Validation of the structures based on geometrical and structural characteristics

Geometrical and structural properties are important criteria used for the validation of structures. Geometrical properties refer to the covalent geometry of the structure, such as bond angles and bond lengths. On the other hand, structural properties refer to non-bonded interactions, such as electrostatics, close contacts, etc. For more information, refer to specialized reviews on these topics ([329] and [330]).

The quality of NMR structures in terms of geometrical and structural properties is generally driven by the force field used during 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 [346] and WHAT IF [347] for proteins as well as the module NUCHECK that is part of the Nucleic-acids Database [348] and MOLPROBITY for nucleic acids [349]. In addition, a validation tool is available on the Protein Data Bank website (<http://www.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.

An important aspect of protein-RNA structures is the analysis of intermolecular contacts, such as hydrogen-bonds and stacking interactions. In most cases, specific intermolecular recognition between a protein and an RNA is driven by intermolecular hydrogen-bonds. Since, in most cases, intermolecular hydrogen-bond re-

straints are not included in structure calculations (see Section 3.3.1), the intermolecular hydrogen-bonding network highly depends on the protocol used during structure calculation and/or refinement. Most refinement protocols use force fields that include electrostatic terms. Therefore, structure refinement steps are crucial for optimizing the hydrogen-bond network of the structural ensemble. Additionally, the use of solvation models during the refinement further optimizes this network, especially at the protein-RNA interface. Intermolecular contacts present in protein-RNA complexes can be assessed by specific programs such as NUCPLOT [350] or ENTANGLE [351]. Both programs read a coordinate file of a protein-nucleic acid complex in PDB format and identify intermolecular interactions, such as hydrogen-bonding and stacking interactions.

3.6.4. Confirmation and quantification of intermolecular contacts

The quality of a structure and the definition of the protein-RNA interface largely depend on the number of intermolecular NOEs that is directly correlated to the quality of the NMR spectra. The fundamental aim of solving macromolecular complexes is to understand the molecular basis that governs the specificity of the complex formation. In protein-RNA complexes, the specificity of the interactions is mainly governed by electrostatic interactions and hydrogen bonds between the protein and the nucleic acid bases. In most cases, however, it is not possible to measure directly such hydrogen bonds by NMR spectroscopy and the network of hydrogen bonds is therefore generally indirectly derived from the NMR experimental restraints and driven by the force field used during the structure calculation or refinement procedures. Additionally, protein-RNA interactions are often stabilized 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 have often been performed on the protein and/or the RNA by mutating specific amino acids or nucleotides that are involved in intermolecular contacts. Three main techniques have been used to quantify the effect of a mutation on the affinity of a protein-RNA complex, namely, EMSA [1,20,26,35,352], SPR [5], and ITC [34,37]. For a description of these techniques, we refer the reader to specialized reviews [353–356]. These techniques allow one to derive the dissociation constant of the complex and therefore, the effect of a mutation can be quanti-

fied in terms of affinity lost. In addition, SPR and ITC allow the obtaining of kinetic and thermodynamic parameters, respectively, that can be used to estimate the energetic contribution of individual interactions. An important aspect when using site-directed mutagenesis is to verify that the mutation does not affect the fold of the protein or of the RNA in the case of folded RNA. This can be checked by comparing NMR spectra, such as 1D, ^{15}N - ^1H HSQC or ^{13}C - ^1H HSQC spectra, of the mutant with the spectra of the wild-type.

An interesting example illustrating the power of combining site-directed mutagenesis and binding assays is the NMR and SPR study of the complex between the Fox-1 protein and its RNA target [5]. Fox-1 contains a single RRM and binds specifically to UGCAUG RNAs [61]. The NMR structure of Fox-1 in complex with this RNA gives structural insights into the specificity of recognition [5]. The first six nucleotides are specifically recognized by the protein through an extensive network of intermolecular and intra-RNA hydrogen-bonds. In total, 3 intra-RNA and 10 intermolecular specific hydrogen-bonds involving RNA bases are observed in the structure. In addition, four stacking interactions involving three aromatic residues of the protein contribute to the affinity of the complex. Using SPR, it was shown that the affinity of the protein for binding to this RNA is very high ($K_d = 0.49$ nM). The importance of electrostatic interactions in the complex formation was assessed by SPR at different salt concentrations and showed that both association and dissociation rate constants are affected by the salt concentration. Furthermore, the contribution of each intermolecular and intra-RNA hydrogen-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 hydrogen-bonds that are lost in the complex based on the NMR structure. Therefore, in this case, all intermolecular and intra-RNA hydrogen bonds observed in the structure could be confirmed and quantified.

3.6.5. 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 for understanding 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 Staufin in its complex with RNA has been used to design a quintuple mutant that disrupts RNA binding [32]. This mutant was then tested by *in vivo* mRNA localization assays in order to demonstrate that the RNA binding properties of the protein Staufin 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 [45]. Also based on the structure of the protein Rnt1p in complex with its RNA target, specific mutants of the protein that disrupt binding were designed and used to show the importance of certain residues of Rnt1p in the processing of certain precursor RNAs [39]. 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 [34].

3.7. Dynamics of protein–RNA complexes

The large number of three dimensional structures of protein–RNA complexes which has been determined in recent years has provided unprecedented insights into how proteins and RNA recognize each other. Structural studies indicate two types of interactions, one in which there are conformational changes in either or

both partners and the other where the two partners are structurally pre-organized resulting in shape specific recognition. A majority of protein–RNA structures studied so far seem to fall in the former category where the interaction occurs by an induced fit mechanism involving structural rearrangements. Recent studies of intermolecular interactions however indicate that conformational rearrangements which occur in an induced fit probably follow an initial binding process which occurs through conformational selection [357]. Also, in several RRMs even though the same type of protein surface is involved in RNA binding, each protein seems to achieve sequence specificity slightly differently. The interest in undertaking detailed investigations of the role of molecular motions in protein–RNA recognition has been motivated by the question of how dynamical processes in the interacting partners govern conformational changes and influence the binding process and specificity in protein–RNA interactions.

In spite of 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 measurement by NMR has the unique ability to provide residue specific information on dynamics in both protein and RNA over a range of different time-scales [358–360]. Fast motions on the pico-second (ps) to nano-second (ns) time-scales influence spin relaxation through the modulation of various spin interactions and 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. Slow motions on the micro-second (μs) to milli-second (ms) timescale influence transverse relaxation as a result of modulation of isotropic chemical shifts and can be observed as an additional contribution to the R_2 and $R_{1\rho}$ relaxation rates. Relaxation data are mapped to molecular dynamics through the well established model-free formalism which quantifies fast motions in terms of an order parameter S^2 reflecting the amplitude of the ps time scale motions and the parameter R_{ex} which is an indicator of slow motions in the μs –ms timescale [361,362]. A detailed quantitative analysis of slow exchange processes in the μs –ms time scale is possible by CPMG (R_2) and spin-lock ($R_{1\rho}$) relaxation dispersion experiments [363,364]. These methods quantify the conformation exchange rate constant, relative equilibrium conformer populations and difference in chemical shifts for the different conformations. Techniques for measurement and analysis of spin relaxation rate constants used in order to extract information on molecular motions have been extensively reviewed [365]. Table 6 gives a summary of the different nuclei which can be employed as dynamics probes in proteins and RNA. Even though several experiments involving different nuclei widely distributed across protein and RNA have been developed, only a few of these have been applied to dynamics studies of protein–RNA complexes as indicated in the table.

Along with relaxation rates, measurements of RDCs, extend the range of motional time-scales to the sub-micro/millisecond regime [244,366,367]. Together these techniques allow a wide spectrum of dynamical processes to be examined in conjunction with 3D structures in order to obtain a comprehensive picture of protein–RNA interactions.

Dynamics studies are typically undertaken after the three dimensional structures have been determined since the atomic coordinates are required for determining parameters relating to overall molecular motion in cases where it is not isotropic. The same sample conditions employed in the structural studies are therefore retained while carrying out the dynamics investigations. Most of the experiments for dynamics studies were initially developed for application to the ^1H - ^{15}N spin system in proteins. These are readily adapted to isolated ^1H - ^{13}C spin systems in proteins and RNA. When applying to fully labeled systems, modifications involving the use of shaped pulses for selective excitation and

Table 6

Overview of relaxation based experiments for dynamics studies of proteins and RNA.

Experiment	Purpose	Target nucleus ^a	Labeling scheme	Reference
R ₁ experiment (inversion recovery)	Quantitative analysis of fast motions in ps–ns timescales at the target nucleus site by model-free analysis or spectral density mapping	¹⁵N backbone amide	¹⁵ N labeled protein	[365,436–438]
		¹⁵ N Asn, Gln sidechains	¹⁵ N labeled protein	[439,440]
R ₂ experiment (CPMG fast pulsing)		¹³ C ^α backbone	¹⁵ N/ ¹³ C labeled or ¹⁵ N/ ¹³ C isolated or fractional	[441,442] [443,444]
		¹³ C ^γ backbone	labeled protein	
R _{1ρ} experiment (high spin-lock field)		¹³ C ^γ Asp, Asn	¹⁵ N/ ¹³ C/ ² H(50%) labeled protein	[445]
		¹³ C ^δ Glu, Gln sidechains	¹³ C fractional or ¹³ C alternating/ ² H fractional labeled protein	[446]
		¹³ C methyl sidechain		
Heteronuclear NOE experiment		² H ^α backbone	¹⁵ N/ ¹³ C/ ² H labeled protein	[447]
		²H methyl, methylene, methine sidechain	¹³ C/ ² H random fractional labeled protein	[446,448,449]
R ₂ dispersion experiment (variation of CPMG frequency)	Quantitative analysis of slow motions in μs–ms timescales at the target nucleus site using analytical expressions for various exchange models or by numerical integration of Bloch–McConnell equations	<u>RNA</u>		
		¹⁵N imino	¹⁵ N labeled RNA	[436–438,450]
		¹³C aromatic, aliphatic	¹⁵ N/ ¹³ C labeled RNA	[368,369,374,451,452]
		²H aromatic, aliphatic	¹³ C/ ² H random fractional labeled RNA	[453]
R _{1ρ} dispersion experiment (variation of spin- lock field strength, offset or both)		³¹ P phosphodiester backbone	perdeuterated RNA	[454]
		<u>Protein</u>		
		¹⁵N backbone amide	¹⁵ N labeled protein	[358,364,376,378, 455–458]
		¹⁵ N Asn, Gln sidechains	¹⁵ N labeled protein	[459]
		¹³ C ^α backbone	¹⁵ N/isolated ¹³ C labeled protein	[364,460,461]
		¹³ C ^γ backbone	¹⁵ N/ ¹³ C labeled protein	[462]
		¹³ C methyl sidechain	¹³ C methyl labeled or ¹⁵ N/isolated ¹³ C labeled protein	[459,461,463]
		¹ H backbone amide	¹⁵ N labeled or ¹⁵ N/ ² H (high levels) labeled protein	[464–466]
		<u>RNA</u>		
		¹³C aromatic, aliphatic	¹⁵ N/ ¹³ C labeled or ¹³ C alternate site labeled RNA	[368,369,380,467,468]
		³¹ P phosphodiester backbone	unlabeled	[469]

^a Nuclei which have been employed as dynamics probes in protein–RNA complexes are indicated in bold type.

the use of constant time chemical shift evolution periods are introduced. In some cases, it becomes necessary to undertake special ¹³C labeling strategies, which avoid the presence of adjacent ¹³C labeled sites. Other considerations in experiments for dynamics studies involve accurate temperature control. Most of the experimental schemes involve the application of 180° pulses during the relaxation interval which can be considerably long. Transverse relaxation rate measurements involve the application of spin-lock fields and CPMG pulse trains of long durations. These result in considerable sample heating that can affect the dynamics measurements since the rates are temperature dependent. Thus temperature regulation throughout the measurement has to be ensured by employing compensating cycles which are introduced during the interscan delay between free induction decays (FIDs). In addition to temperature control, careful adjustment of water suppression methods is necessary in order to avoid saturation transfer from the water ¹H spins. This necessitates the use of water flip-back techniques and in some cases use of selective ¹H 180° pulses in the relaxation interval.

The significance and role of molecular motions in intermolecular recognition is best revealed from a comparison of protein and RNA dynamics in the free and bound states. Detailed analysis of relaxation rates in both protein and RNA in their free and bound states, along with dynamics changes following binding have been reported for two systems, human U1A protein interaction with the 3' untranslated region (UTR) of its own pre-mRNA and VTS1p sterile alpha motif (SAM) domain interaction with the *smaug* recognition element (SRE) stem-loop RNA [368–372]. Other studies

have examined dynamics in the free and RNA bound states of the cleavage stimulation factor (Cstf-64) and the ribosomal protein L11 [46,373]. Systematic characterization of RNA dynamics changes that occur on binding of a ligand has been reported for HIV-1 TAR RNA and HIV-2 TAR RNA binding to arginineamide [374,375]. The findings from these studies reveal 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.

3.7.1. Protein 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 [370]. U1A protein binds the 3' untranslated region (UTR)-RNA with high specificity and the interaction involves an induced-fit mechanism. Analysis of backbone ¹⁵N relaxation rates in the free U1A indicates several residues at the RNA binding surface which have significant *R_{ex}* contribution to the transverse relaxation rates. On binding to the RNA, the conformational fluctuations in the μs–ms timescale at the binding surface are reduced significantly. In addition, analysis of ²H relaxation in side-chain methyl groups indicate loss of μs–ms motions on interaction with RNA. Conformational flexibility 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 inter-

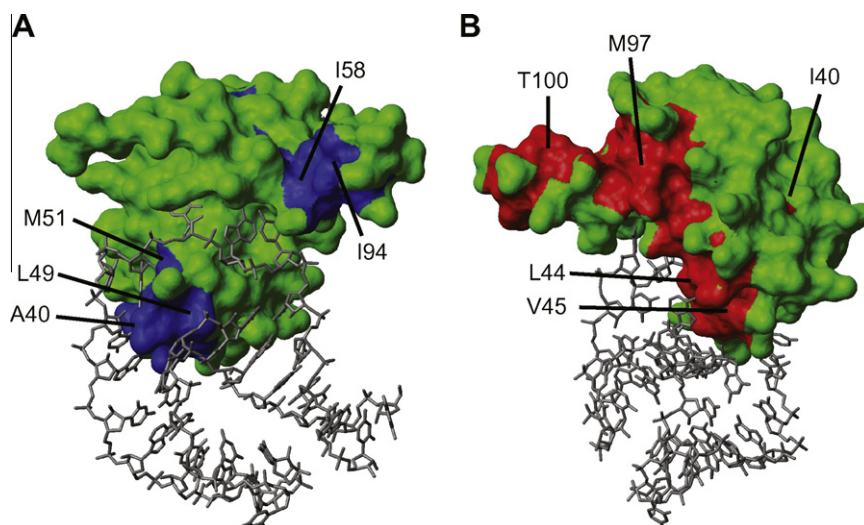


Fig. 29. Surface representation of U1A protein bound to RNA. U1A protein surface is colored in green and the RNA is in stick and colored in grey [23]. (A) Residues of U1A protein [370] that become significantly more rigid in the complex (blue) cluster in two distinctive patches, at the intermolecular interface and the hydrophobic patch that positions helix C. (B) Residues gaining significant flexibility in the complex (red) are found in the solvent exposed surface of helix C and in a solvent exposed patch on the edge of the binding interface. Figures were generated with molmol [470]. Adapted from Ref. [370].

actions ensure high specificity (Fig. 29). Clearly, this high specificity accompanied by a loss of flexibility is achieved at a large entropic cost. Interestingly, some residues at the edge of the protein–RNA interface which are partially solvent exposed retain their conformational flexibility in the bound state. This preservation of flexibility in regions of the interface which are less critical for specificity, compensates to some extent the entropic penalty associated with the loss of flexibility in residues which are crucial for specificity.

The interaction of the two domain protein L11 with 23S-rRNA is similar to U1A in that it occurs by an induced-fit mechanism [46]. The free protein shows considerable flexibility in the ps–ns time-scales as indicated by the low S^2 values observed for the residues in the RNA binding loop region. Upon interaction with the RNA, the binding region becomes more rigid and the dynamics becomes similar to the rest of the C-terminal domain which carries the RNA binding site. Also, RNA binding influences overall motion of the two domains differently. While the domains tumble as a rigid unit in the free state, RNA binding to the C-terminal domain results in more freedom of movement in the N-terminal domain. This is in agreement with the observation of different possible conformations in the N-terminal domain.

The interaction of the 64 kDa subunit of the Cstf-64 protein with GU rich RNA sequences is in contrast to the above examples. The interaction has 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 the above examples of highly specific interaction [373]. 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. In particular there is a significant decrease in S^2 and large R_{ex} contributions observed at the RNA binding surface. Also there is a structural rearrangement involving unfolding of the C-terminal helix. In this case, the binding interface retains a high degree of mobility in the complex. A mobile interface maybe intrinsic to the functional requirement of Cstf-64 which binds many GU rich RNA sequences and yet discriminates against other RNAs.

Our studies on VTS1p SAM domain interaction 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 [29]. VTS1p SAM domain recognizes a general consensus sequence of the form XNGY(N) for the RNA 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. Deletion of the nucleotide indicated in parentheses does not alter binding affinity indicating that penta- or tetra loop RNAs can bind to the VTS1p SAM domain. ^{15}N relaxation studies of the backbone dynamics indicate that in its free state the VTS1p SAM domain is mostly rigid with no significant motion in the fast and slow time scales [372]. This is consistent with the idea of a conformationally pre-organized binding surface on the protein which can accommodate the RNA loop. On interaction with the CUGGC loop of the SRE RNA, there is a decrease in S^2 for a majority of the residues indicating increased flexibility in the bound state. The only residues which show an increase (or a negligible decrease) in S^2 values all belong to the binding surface and are associated with the specific recognition of the central G nucleotide in the RNA loop. In contrast to the increased rigidity of residues involved in specific interaction with the central G nucleotide, those protein residues involved in non-specific interaction with other nucleotides in the RNA loop show lower S^2 values corresponding to increased flexibility in the bound state (Fig. 30).

As in the case of ps–ns time scale motions, interaction with RNA also results in an overall increase in μs –ms motions in most residues of the protein. The only exceptions are the residues in the binding region which are involved in specific recognition of the RNA which show none or negligible R_{ex} contribution to relaxation. The interface dynamics thus 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. Thus molecular motions play a role in modulating the binding affinity for different combinations of loop nucleotides allowing a general consensus sequence of the form XNGY(N) for the RNA loop for the recognition of SRE RNA by the VTS1p SAM domain.

Interaction by shape recognition has also been reported in the binding of dsRNA to the Staufen protein [32]. This is a case of non-sequence-specific recognition with no significant structural

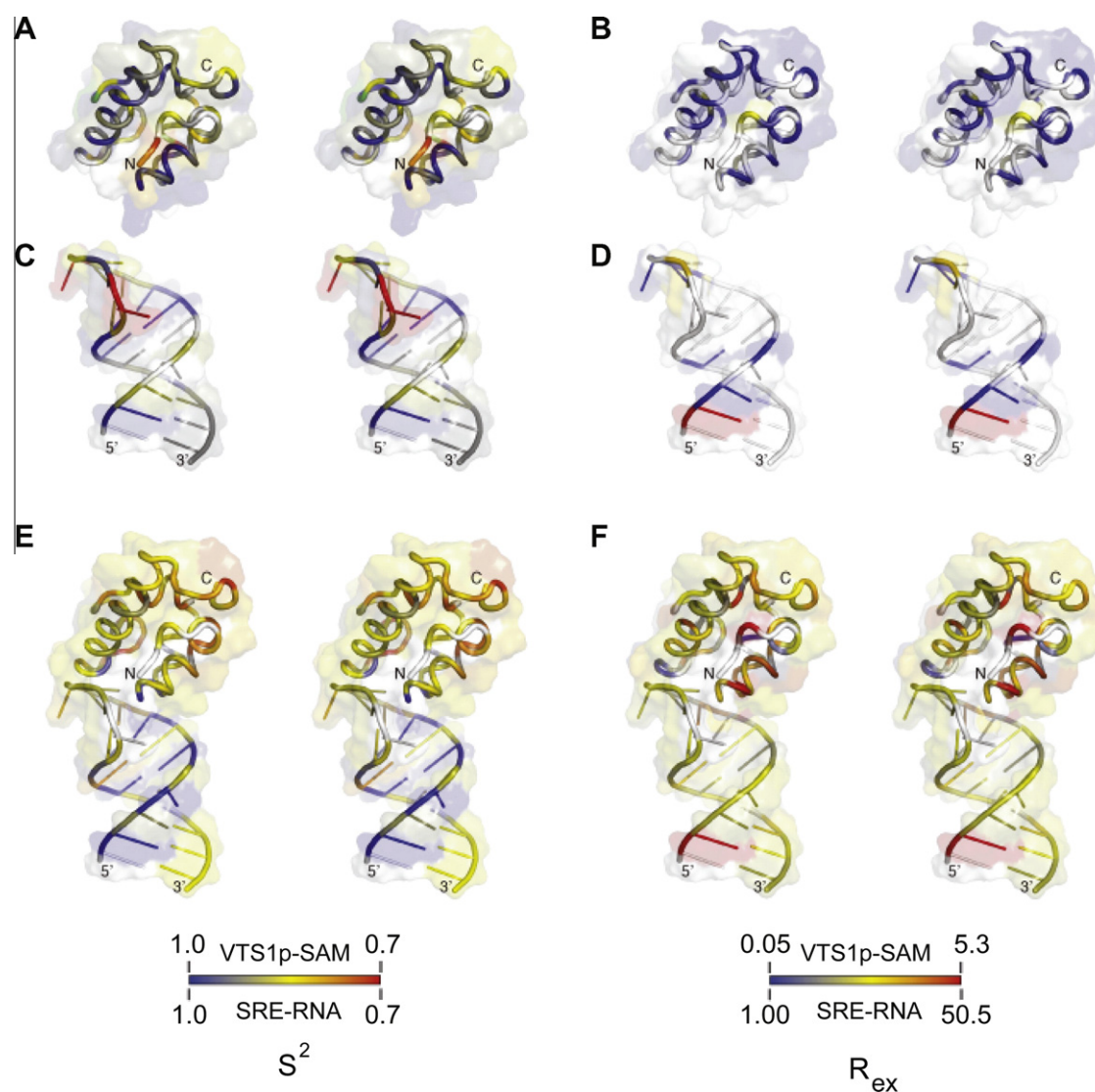


Fig. 30. Representation of internal motion parameters for backbone ^{15}N sites of VTS1p SAM [372] and aromatic ^{13}C sites of SRE RNA [369] in the free and bound states. (A) S^2 , (B) R_{ex} for VTS1p SAM and (C) S^2 , (D) R_{ex} for SRE RNA in the free state. (E) S^2 and (F) R_{ex} for both components in the bound state. Different scales are used since measurements for VTS1p SAM domain and SRE RNA were carried out at 288 and 303 K respectively. Reproduced with permission from Ref. [372].

reorganizations in the protein or the RNA upon binding. Studies of the 3D structures and ^1H - ^{15}N NOE measurements in the free and bound states of the protein indicate that the RNA binding surface is highly mobile in the free form and this flexibility is also retained in the RNA bound state.

3.7.2. RNA dynamics

The interaction of the human U1A protein with the 3' UTR RNA is one of the first examples where the role of RNA dynamics in protein-RNA recognition has been examined by relaxation studies [368,371]. The RNA has two helical domains with an apical loop and a seven nucleotide internal loop which defines the protein binding surface. ^{13}C relaxation studies at the aromatic and anomeric sites in the different nucleotides clearly shows considerable mobility in the free state of the RNA. Low S^2 values indicating high flexibility in the ps-ns time scale were observed in the loop residues, particularly those of the binding region. Some residues in the loops have large R_{ex} contributions to relaxation indicating the presence of μs -ms timescale motions, which were analyzed quantitatively from the spin-lock field dependence of $R_{1\rho}$ rates. The only

nucleotide in the binding region which is relatively rigid is the one which has a stacking interaction with the closing base-pair of the upper helix. The binding loop of the RNA undergoes slow motion at the hinges which connects it to the double helical stems. The residues which bind most strongly with the protein undergo fast motion whereas the remaining residues in the loop which lack stacking interactions exhibit motional freedom in fast and slow timescales.

On binding to the protein, a collective motion of the upper helical domain at a faster timescale relative to the overall motion of the binding loop and the lower helix becomes apparent. This collective motion is normally masked in the free RNA since it occurs on a similar timescale to that of the overall tumbling of a small RNA. Interaction with the protein quenches the fast and slow motions in the binding loop making it almost as ordered as the lower helix. The exception is the closing base-pair of the upper helix and the adjacent nucleotide in the binding loop which has stacking interactions with it. The latter acts as a hinge point between the two domains with different overall motions.

The interaction of the stem-loop SRE RNA with the VTS1p SAM domain occurs through the CUGGC pentaloop. ^{13}C relaxation stud-

ies indicate that in the free RNA, the nucleotides which form a base-pair within the loop, have small fast motion amplitudes very similar to those observed in the stem region while the other nucleotides show considerable flexibility with large amplitude motions in the ps–ns timescale [369]. The central G nucleotide which is specifically recognized by the protein has a fast motion amplitude that is intermediate between that of the base-paired nucleotides and the two highly flexible loop residues. The loop thus adopts a well defined shape defined by the base-pair and aids in the shape specific recognition by the protein. The loop nucleotides and those of the flanking base-pair also undergo slow motions in the μ s–ms timescale as indicated by the spin-lock field dependence of $R_{1\rho}$ rates and the R_{ex} parameters. The striking exception is the lack of slow motions at the central G nucleotide, which is unusual for one that is located in a loop.

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, shows 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 (Fig. 30). The stem region on the other hand indicates an overall increase in fast motions, especially at the aromatic sites in the bound state. Also there seems to be a net slow motion affecting the entire RNA in the bound state suggesting the possibility of a collective motion within the binding cavity of the protein.

3.7.3. Insights from dynamics studies

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, 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. The observation of μ s–ms motions in the free states, particularly at the binding surfaces, favors the conformational selection mechanism in which binding occurs between selected conformers from among several conformational substates which exist in a dynamic equilibrium. In a conformational selection process, binding is followed by a population distribution favoring the bound state conformations and possible conformational rearrangements which would constitute an induced fit process [357].

In shape specific recognition on the other hand, the free states have relatively limited mobility, thus providing a conformationally ordered surface for binding. This may also be viewed as a limiting case of conformational selection, in which binding occurs between the lowest energy conformations of the two interacting partners.

Regions of the binding interface associated with high specificity of interaction become more rigid on binding while regions with non sequence-specificity attains 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 results in a high entropic cost in complex formation. Flexibility gain at other locations helps to offset the entropic cost resulting from restricted mobility at the binding interface.

3.7.4. Future directions

From the few examples reported so far it is abundantly clear that dynamics studies reveal important aspects of recognition that would not have been accessible from structures alone. Almost all of the protein dynamics publications have focused on the dynamics at the amide ^{15}N sites along the backbone. As revealed from the side-chain dynamics studies in the U1A protein, more work needs to be undertaken to examine dynamics in the protein side-chains. This is especially relevant considering that much of the intermolecular contacts are established through the side-chains.

Slower motions on the μ s–ms timescales are biologically very important because they are close to the time scales with which docking, folding, allosteric transitions, product release etc. take place and are thus associated with functional processes. There have been significant advances in the experimental characterization of slow motions in proteins involving a variety of nuclei as probes [364,376,377]. In addition, new strategies for analysis of relaxation dispersion data have been reported [378,379]. More recently, the application of similar techniques to RNA have been reported [380]. Application of these methods to protein–RNA complexes can provide far more insights into the complex conformational dynamics at protein–RNA interfaces which goes beyond a qualitative identification of the presence of these motions inferred from the R_{ex} parameter.

The measurements of RDCs in protein–RNA complexes to extract dynamics information will extend the dynamical time scales which can be probed, by including the intermediate regime not readily accessed by the techniques which probe fast and slow motions. While RDCs have been invaluable in defining molecular structures, they also offer the possibility of examining domain orientation changes induced on RNA binding in multi-domain RNA binding proteins as evidenced from the studies in of the L11 protein. Interesting details of RNA dynamics are also accessible from RDC measurements as shown in the work of Al-Hashimi [367,381]. Most RNAs which interact with proteins have more complex structures involving several helical stems and internal loops. For instance, large amplitude motions of the helical domains in HIV-1 TAR RNA which allows binding of diverse targets in the bulge between helices have been revealed by RDC measurements [382]. The bound state of the 3' UTR RNA has revealed domain motions following RNA binding [371] and newer methodologies based on RDCs are most suited to examining these large scale motions.

4. NMR structures of protein–RNA complexes: what did we learn from them?

4.1. Introduction

The protein–RNA structures solved by NMR provided significant structural insights for understanding important biological processes at the molecular level. Many of these structures were fundamental to deciphering the role of these interactions and to guiding further studies and characterization of biological functions. Furthermore, many of these interactions play an important role in disease related processes and thus protein–RNA structures can provide templates for structure-based drug design. In particular, NMR structures of protein–RNA complexes are fundamental to understanding the molecular basis of the interaction between viral proteins and RNAs, certain gene regulation mechanisms in prokaryotes, and many post-transcriptional gene regulation events in eukaryotes.

4.2. Retroviral and bacteriophage protein–RNA complexes

Retroviruses include the HIV, the RSV, or the MoMuLV. The life cycle of these viruses involves numerous protein–RNA interac-

tions, especially the interaction between viral proteins and viral RNAs. Structures of these complexes are crucial for understanding the molecular basis of these interactions and hence for the structure-based design of drugs necessary for the development of anti-retroviral therapies [383]. NMR has provided a major contribution towards the understanding of viral protein–RNA interactions. Indeed, 19 out of 22 structures of viral protein–RNA complexes were solved by NMR [4,6,8,10–13,16,19,23,31,41–45]. Three types of viral protein–RNA complexes were studied: the complexes between the viral protein Rev and its RNA target, the RRE, involved in RNA export [6,16,19,41,42,44]; the complexes between the viral protein Tat (trans-activator) protein and its RNA target, the TAR (trans-activator response element), involved in transcription regulation [8,11,23,31,43]; the complexes between the viral protein NC and its RNA target, the Ψ -site, involved in RNA packaging [4,10,12,13,45].

In addition, five NMR structures of a bacteriophage N protein–BoxB RNA have been solved by NMR [7,9,15,24,33].

4.2.1. Viral NC–RNA complexes

The NC protein plays a critical role in viral replication and participates in genome recognition and encapsidation. The recognition of the viral genome is directed by the interaction between the NC protein and a region of the unspliced viral RNA termed the Ψ -site. In most retroviruses, this Ψ -site RNA is highly structured and is composed of stem-loop structures (Fig. 31). The NC protein contains one or two zinc–knuckle domains that are stabilized by one atom of zinc and are responsible for RNA binding. The laboratory of Summers solved eight NMR structures of a NC zinc–knuckle in complex with RNA, four with dsRNA [4,10,12,45] and four with ssRNA [13], that provide significant insights into the RNA genome recognition and encapsidation of three different viruses, the Moloney Murine Leukemia Virus, the Human Immunodeficiency Virus, and the Rous Sarcoma Virus.

The MoMuLV Ψ -site is a \sim 370-nucleotide RNA fragment that consists of a series of closely spaced stem-loops (Fig. 31, top). This RNA fragment undergoes a monomer–dimer transition that is important for encapsidation [384]. The structure of the MoMuLV NC protein in its complex with a modified portion of this RNA that remains monomeric but retains the dimeric base-pairing was solved by NMR [10]. This structure shows that the NC protein binds with high affinity and specificity to a UAUCUG sequence located in a linker connecting two stem-loops. This UAUCUG sequence is involved in base-pairing in the monomeric form of the RNA and cannot bind the NC protein. 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. Chemical accessibility mapping showed that other segments of the Ψ -site similar to the UAUCUG segment are sequestered in the monomeric form and become exposed in the dimeric form, including AACAGU, CCUCCGU, and UUUUGCU [385]. The structures of the MoMuLV NC protein with these three RNA segments were then solved [13] indicating, together with isothermal titration calorimetry data, that these three segments bind the NC protein with affinities similar to the UAUCUG fragment. These structures allowed the definition of a general mechanism for MoMuLV NC–RNA interaction.

In the case of HIV-1 packaging, the Ψ -site is composed of a \sim 120-nucleotide segment that contains four stem-loops (SL1–SL4) (Fig. 31, bottom). The structures of the HIV-1 NC protein in its complexes with the SL2 and the SL3 stem-loops were solved by NMR [4,12] and show that the HIV-1 NC protein specifically recognize the GGUG (SL2) and GGAG (SL3) segment of the RNA loop. However, in HIV-1, the structures of the NC-binding SL2 and SL3 stem-loops appear to be present both in the monomeric and the dimeric form

of the RNA [384]. In this case, a model was proposed in which the stem-loops recognized by the NC protein would serve as cooperative packaging elements. In addition, the structure of the RSV NC protein in complex with the RSV 82-nucleotide Ψ -site was solved by NMR and it was proposed that NC binding could potentially stabilize an RNA structure that is favorable for encapsidation [45].

4.2.2. Viral Rev–RRE, Tat–TAR and bacteriophage N protein–BoxB RNA complexes

The viral proteins Rev and Tat possess an arginine-rich motif (ARM) responsible for RNA binding. In addition, similar ARMs are found in bacteriophage proteins and are also responsible for RNA binding. The N proteins of bacteriophages possess an N-terminal ARM that binds its target RNA, the boxB RNA, and regulates transcriptional anti-termination.

Sixteen structures of an ARM–RNA complex were solved by NMR [6–9,11,15,16,19,23,24,31,33,41–44]. In all complexes, the ARM peptide specifically binds the stem region of the RNA that adopts a stem-loop structure. Although the amino acid sequences of the ARMs are quite similar in all complexes, the ARM peptide can adopt different structures when binding to the RNA, that is a α -helical [6,7,9,15,16,24,33,41,44], a β -hairpin [8,11,23,31,43] or an extended conformation [19,42] (Fig. 32). In all cases, the ARM binds the major groove of the RNA stem but structural differences could be observed (reviewed in [386]). The structural features of the RNA stem drives the conformation adopted by the peptide. In the cases of Tat–TAR complexes, the stem of the TAR RNA contains a U-(A-U) base triple that induces a bend in the stem and a slight widening of the major groove. This particular structural feature is crucial for the binding of the peptide that adopts a β -hairpin structure and penetrates deeply into the major groove of the RNA (Fig. 32, left) [8,11,23,31,43]. In contrast, the RRE RNA major groove is largely widened by the presence of two purine mismatches. This widening is suitable for binding the Rev peptide that adopts an α -helical conformation and that also deeply penetrates into the major groove of the RNA (Fig. 32, middle) [6,16,41,44]. Finally, BoxB RNAs from bacteriophages adopt stem-loop structures with regular A-form helix stems. In these cases, the major groove is not widened and the ARM peptides bind at the surface of the major groove (Fig. 32, right). In contrast to retroviral complexes, the specific recognition of the BoxB RNAs by the N peptides is not driven by the stem structure but by specific contacts to the nucleotides of the loop that adopts a GNRA stable tetraloop structure or a GNRA-like structure [7,9,15,24,33]. In all complexes, the arginine residues play a crucial role in RNA binding through hydrogen bonds to the phosphate backbone atoms of the RNA or to the O6 and N7 of guanines in the major groove.

4.3. Prokaryotic gene regulation

Bacteria have developed original systems for regulating their gene expression that are not present in eukaryotes. These specific regulations can occur at the transcriptional and the translational level. Understanding these regulations at a structural level is important in order to develop new drugs for anti-bacterial treatments. Out of 50 structures of bacterial protein–RNA complexes (excluding ribosomal structures), four were solved by NMR and helped in the understanding of bacterial gene regulation [34,36,40,47].

In bacteria, genes are often organized in operons, a cluster of genes under the control of a single promoter. The genes included in an operon are then transcribed together into an mRNA that is in turn translated into different proteins. Transcription of operons is highly regulated. One regulatory system consists of a termination/antitermination system (reviewed in [387]). Terminators are specific structures present in the 3' end of mRNAs. A terminator consists of a palindromic sequence that forms a stable stem-loop

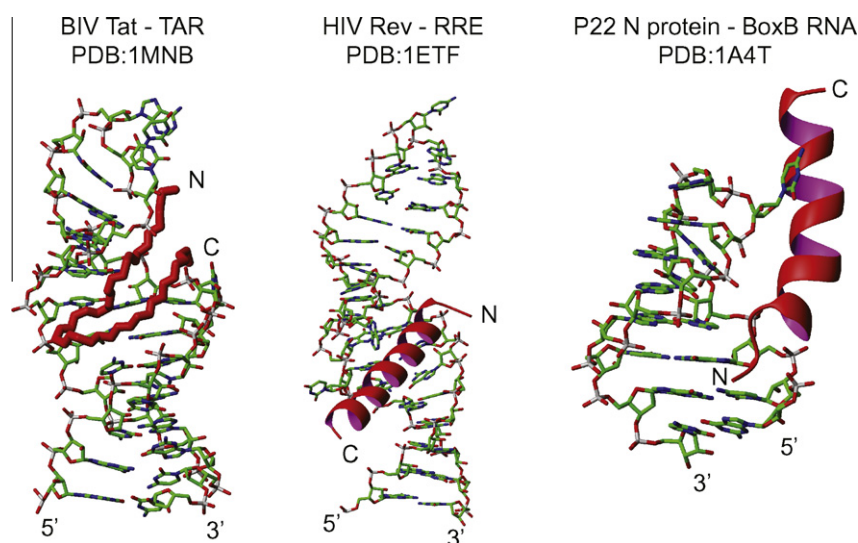


Fig. 32. Structures of viral peptide–RNA complexes. Left: the Tat–TAR complex from the bovine immunodeficiency virus (BIV) [31]. Middle: the Rev–RRE complex from the HIV [6]. Right: the N–BoxB complex from the bacteriophage P22 [7]. The RNA is shown as a stick structure and colored green and the proteins are colored red and magenta with their N- and C-termini labeled. Figures were generated with molmol [470].

nator site can adopt different structures, the terminator structure suitable for transcription termination, and the anti-terminator structure that prevents transcription termination. The anti-terminator structure is driven by a small RNA sequence called RAT (ribonucleic anti-terminator) that overlap the terminator palindromic sequence and is stabilized by a family of bacterial proteins, the transcriptional anti-terminator (AT) proteins that possess an RNA binding domain called CAT (co-antiterminator). The structure of a complex between the CAT domain of the LicT protein and a RAT RNA was determined by NMR (Fig. 33, top) [40]. This structure explains how the CAT domain binds the RNA stem-loop and act as a protein clamp that stabilizes the anti-terminator RNA hairpin and prevents formation of the terminator hairpin, hence allowing the transcription to proceed further through the following coding sequence of the operon.

Another interesting system found in bacteria for gene regulation is a toxin/antitoxin system. Bacterial genomes often contain operons that encode a toxin and an antitoxin (reviewed in [388]). These toxin–antitoxin modules play an important role in cell growth arrest and cell death upon bacterial stress. A subfamily of bacterial toxins possesses RNase activity and cleaves mRNAs, thus preventing protein translation. In contrast to most ribonucleases, toxins are highly specific for their RNA targets. For example, the toxin Kid cleaves specifically at the 5' side of an adenine in single-stranded RNAs containing an UA(A/C) sequence. Using NMR combined with docking approaches, a structural model of the Kid–RNA complex has been proposed [47]. This structural model allowed the definition of the Kid active site and a model for Kid RNase activity was derived that is similar to other structurally unrelated RNases.

At the translational level, bacteria also developed specific regulatory mechanisms. One example of translation repression in bacteria is when the ribosome binding site (RBS) that contains a Shine–Dalgarno sequence on the mRNA is occluded by proteins or base-pairing with non-coding RNAs (ncRNA) [389,390]. An illustration of this mechanism concerns a family of RNA binding proteins, the regulator of secondary metabolism (RsmA)/carbon storage regulator (CsrA) that binds mRNAs at the RBS by specifically recognizing ANNGA sequences. Translation repression is then released by ncRNAs containing multiple ANNGA motifs that bind with high affinity to the RsmA/CsrA proteins and sequester them

away from the mRNA. The structure of RsmE, a member of the RsmA/CsrA family, in complex with an RNA containing the Shine–Dalgarno sequence was solved by NMR and provided structural insights into the regulation of bacterial translation initiation (Fig. 33, bottom) [34]. The structure shows that the RNA adopts a stem-loop structure and that the Shine–Dalgarno sequence is sequestered by the protein, therefore preventing ribosome binding to the mRNA.

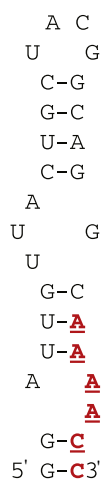
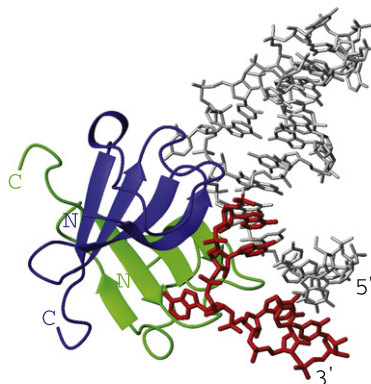
4.4. Eukaryotic post-transcriptional gene regulation

RNA binding proteins are very abundant in eukaryotes (more than 2% of the genome encode for RNA binding proteins). These proteins are involved in a wide range of biological functions, notably in post-transcriptional gene regulation that include constitutive and alternative splicing, polyadenylation, RNA editing, mRNA export, mRNA stability, and translation regulation. All these biological processes are highly regulated and misregulations of protein–RNA interactions often lead to various diseases [391,392]. Out of 54 structures of eukaryotic protein–RNA complexes below 40 kDa, 20 were solved by NMR. These complexes can be subdivided depending on their biological implications in alternative splicing [5,17,26,28,35,37], mRNA stability [2,18,21,29,30,38,39], RNA localization [32], RNA export [22], ribosome biogenesis [1,20], and microRNA biogenesis [25].

4.4.1. RNA binding domains

In contrast to viral or prokaryotic RNA binding proteins, most eukaryotic RNA binding proteins are multi-module proteins containing more than one RNA binding domain. In most cases, a eukaryotic RNA binding protein possesses multiple copies of the same RNA binding domain. The most common RNA binding domains found in eukaryotes are the RNA Recognition Motif (RRM) [393,394], the heterogeneous nuclear ribonucleoprotein K homology (KH) domain [395], the double-stranded RNA binding domain (dsRBD) [396], and the zinc-finger domain [397]. These RNA binding domains are generally rather small (below 20 kDa) and therefore highly suitable for NMR structure determination. For this reason, NMR has contributed significantly to understanding how these domains specifically bind to their RNA targets. Thirteen RRM–RNA complexes [1,2,5,17,20,28,30,35,37,38], two dsRBD–

Termination/antitermination transcription regulation

LicT CAT - RNA complex
pdb: 1L1C

Regulation of translation initiation

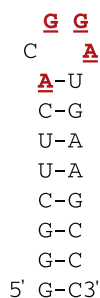
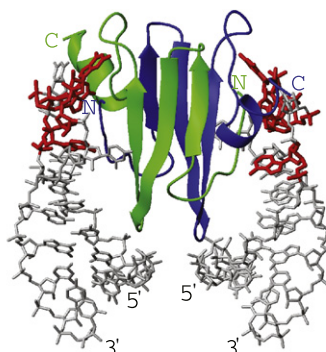
RSME - RNA complex
pdb: 2JPP

Fig. 33. Bacterial protein–RNA complexes. Top: Structure of LicT in complex with an antiterminator stem-loop showing that the terminator sequence is sequestered by the protein [40]. Left: secondary structure of the RNA used in the NMR study. The terminator sequence is bold, underlined and colored red. Right: Structure of the LicT CAT–RNA complex. The protein dimer is colored blue and green. The nucleotides of the terminator sequence are colored red. Bottom: Structure of RSME in its complex with RNA showing that the ribosome binding site is sequestered by the protein [34]. Left: secondary structure of the RNA used in the NMR study. The Shine–Dalgarno sequence is bold, underlined and colored in red. Right: Structure of the RSME–RNA complex. The protein dimer is colored blue and green. The nucleotides of the Shine–Dalgarno sequence are colored red. Figures were generated with molmol [470].

RNA complexes [32,39], two zinc-finger–RNA complexes [18,22], and one KH–RNA complex [26] structure have been solved by NMR. In addition, three protein–RNA complexes involving other less abundant RNA binding domains, such as the Sam domain [21,29] and the PAZ domain [25], have been solved by NMR.

The RRM is a very abundant domain often found in multiple copies in RNA binding proteins [393,394]. This domain is composed of about 100 amino acids and adopts a $\beta\alpha\beta\beta\alpha\beta$ fold organized in a four-stranded β -sheet packed with two α -helices on one side. The different structures of RRM–RNA complexes show that RRMs bind RNA mainly through their solvent accessible β -sheet surface. These structures explain the structural basis of RNA specific recognition by this domain and provide important insights into the versatility of this domain for RNA binding [1,2,5,17,20,28,30,35,37,38]. RRMs can bind single-stranded RNA sequences (Fig. 34A) [5,17,28,30,37], RNA embedded in stem-loop structures (Fig. 34B) [1,20,35] or stems containing an internal loop (Fig. 34C) [2,38]. In the case of dsRNA binding, RRMs often bind solely and specifically to the loop sequence [1,2,20,38] using their β -sheet surface. In the case of the RBMY (RNA binding motif gene on chromosome Y) protein, however, it was shown that, in addition to

the sequence-specific recognition of the loop by the β -sheet, a loop of the RRM also inserts into the major groove of the stem in a shape-specific manner (Fig. 34D) [35]. Generally, RRMs can recognize specifically between two and four nucleotides on their β -sheet surface. In some cases, RRMs also use additional regions (mainly loops) to bind specifically more nucleotides. For example, the RRM of the protein Fox-1 binds specifically seven nucleotides, three nucleotides being bound by the canonical β -sheet surface and four nucleotides being bound by three loops (Fig. 34A) [5]. In addition, a sub-family of RRMs, the qRRMs (quasi RNA recognition motifs) recognizes specifically its RNA target using solely three loops while the β -sheet surface is not involved in binding (Fig. 34E) [14]. Finally, in the cases of proteins containing multiple RRMs, it was shown that two consecutive RRMs are able to bind cooperatively one RNA molecule, creating a molecular clamp around the RNA (Fig. 34B and F) [1,20,30] (Inoue et al., unpublished). Altogether, these structures demonstrate that RRMs bind RNA in a sequence-specific fashion by recognizing mainly functional groups of the bases.

The dsRBD is also a common RNA binding domain that binds double-stranded regions of RNAs [396]. This domain adopts an

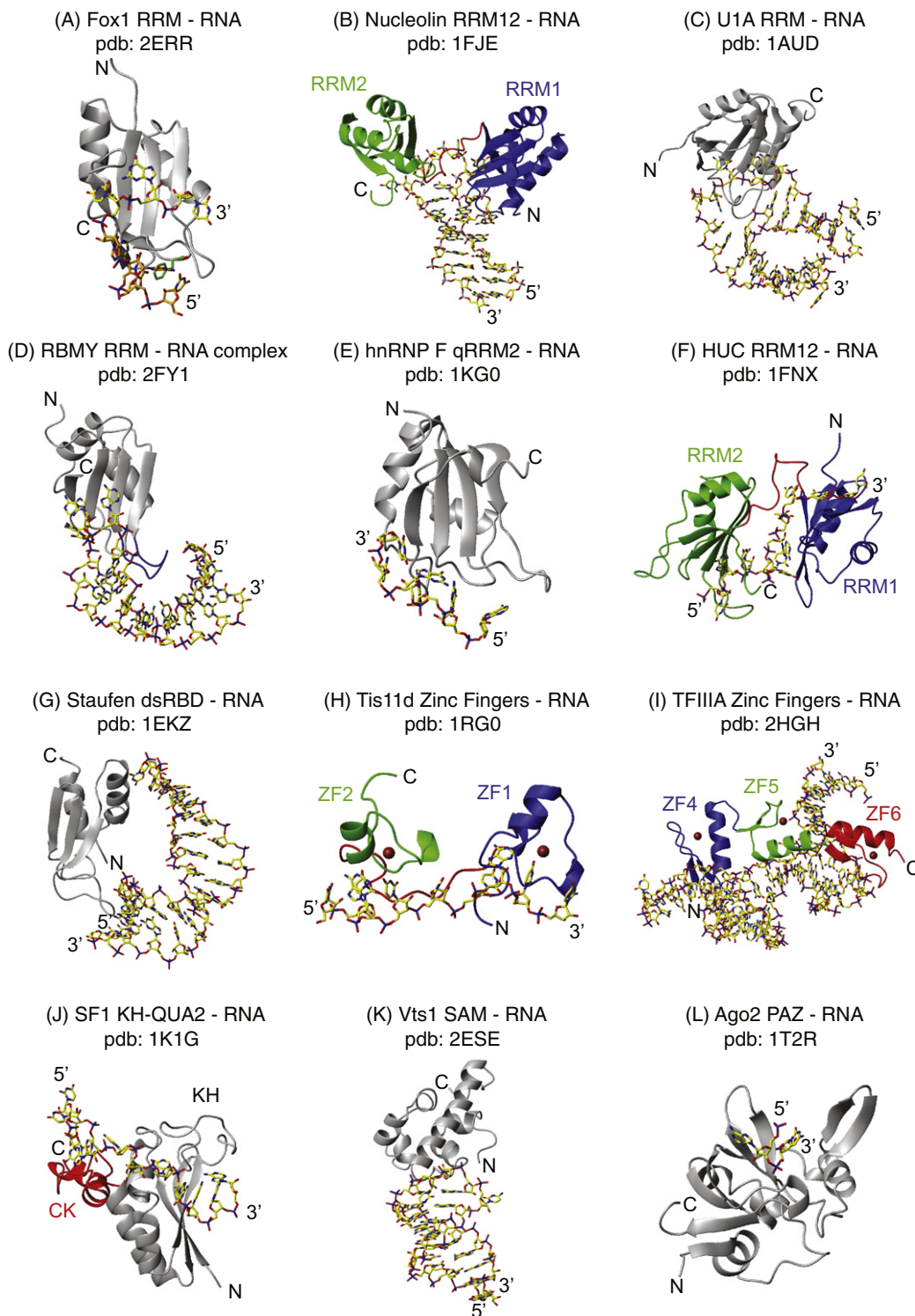


Fig. 34. Structures of eukaryotic protein domain-RNA complexes. (A) Fox-1-RNA complex [5]. (B) Nucleolin RRM12-RNA complex [1]. (C) U1A-RNA complex [2,3]. (D) RBMY-RNA complex [35]. (E) hnRNP F qRRM2-RNA complex [14]. (F) HUC RRM12-RNA complex (Inoue et al., unpublished). (G) Staufen dsRBD3-RNA complex [32]. (H) Tis11d zinc fingers12-RNA complex [18]. (I) TFIIIA zinc fingers456-RNA complex [22]. (J) SF1 KH-QUA2-RNA complex [26]. (K) VTS1 SAM-RNA complex [29]. (L) Argonaute 2 PAZ-RNA complex [25]. Protein domains are colored grey, blue, green or red and RNA is colored yellow. Figures were generated with molmol [470]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$\alpha\beta\beta\alpha$ fold with two α -helices packing on one side of a three-stranded β -sheet. Two structures of a dsRBD in complex with dsRNA were solved by NMR and provided structural insights into the RNA recognition by dsRBDs (Fig. 34G) [32,39]. The first NMR structure of a dsRBD in complex with RNA was solved in 2000

[32]. Interestingly, since there is no specificity of interaction, the dsRNA was artificially designed. A highly stable and well-characterized loop, which is not involved in binding, was used to stabilize the double-helical region and the nucleotide composition of the stem was designed to be fully symmetrical to simplify the NMR

spectral analysis. For the first time, orientational restraints derived from RDCs, measured on both protein and RNA, were used in the structure determination of a protein–RNA complex. This structure confirms the features of the dsRBD–dsRNA interaction that were initially identified by X-ray crystallography [338]. In both cases, an artificial dsRNA was used for the structure determination and the recognition was found not to be sequence specific. In 2004, another dsRBD–dsRNA structure was solved by NMR using for the first time a natural RNA sequence [39]. In this case, the dsRBD of the Rnt1p protein recognizes specifically a stem-loop structure containing an AGNN tetraloop [398]. The structure shows that the fold and not the sequence of the AGNN tetraloop is specifically recognized by the dsRBD.

The zinc-finger domain is a well-known domain present in transcription factors that is involved in DNA binding. However, a subclass of zinc-fingers has been shown to bind specifically to RNA molecules [397]. The zinc-finger domain is a small domain of approximately 30 amino acids that is stabilized by one atom of zinc coordinated by four amino acids of the protein (cysteines or histidines) and adopts a $\beta\beta\alpha$ -fold. Two NMR structures of a zinc-finger–RNA complex were solved by NMR and provided insights into the specific RNA recognition by this domain (Fig. 34H and I) [18,22]. The protein TIS11d possesses a TZF (tandem zinc-finger) composed of two CCCH zinc-fingers and involved in binding AU-rich element (ARE) RNA. The structure of the TIS11d TZF in its complex with a 5'-UUAUUAUU-3' RNA was solved by NMR [18]. This structure provided the first structural insights into single-stranded RNA recognition by zinc-finger domains. The structure of a portion of TFIIA in complex with a dsRNA was also solved by NMR [22]. TFIIA contains nine zinc-fingers but zinc-fingers 4–6 are sufficient to bind a 55-nucleotide portion of the 5S ribosomal RNA. The NMR structure is similar to the X-ray structure of the same complex solved previously [332].

The KH domain is also a common RNA binding domain in eukaryotes. It consists of approximately 70 amino acids, adopts a $\beta\alpha\alpha\beta\beta\alpha$ -fold in eukaryotes, and specifically binds single-stranded RNAs [395]. The STAR (signal transduction and activation of RNA) family of proteins are unique among the RNA binding proteins because they possess an extended KH fold, comprising a central KH domain flanked by two conserved sequences, the NK (N-terminal of KH) and the CK (C-terminal of KH) sequences [399]. The protein SF1 is a member of the STAR family of proteins and possess a KH domain extended by a conserved CK sequence necessary for RNA binding. The structure of SF1 KH-CK in complex with a 5'-UUAUACUACAA-3' RNA was solved by NMR (Fig. 34J) [26]. Although a structure of a KH domain in complex with RNA was solved previously [400], this NMR structure is the first one, and is still the only one, describing how an extended KH domain from a STAR protein binds to RNA. The CK domain was shown to participate in the RNA binding, increasing the affinity of the protein for the RNA.

The SAM (sterile alpha motif) domain is a protein domain that is generally involved in protein–protein interactions [401]. However, it was shown that in certain proteins, such as *Saccharomyces cerevisiae* Vts1p or *Drosophila melanogaster* Smaug, SAM domains can also bind RNA [402]. The RNA sequence specifically recognized by Vts1p and Smaug was called SRE (Smaug Recognition Element) and consists of a stem-loop. The NMR structure of Vts1p SAM domain in complex with SRE RNA was solved independently and simultaneously by two groups (Fig. 34K) [21,29], together with an X-ray structure [402]. In addition the structures of the free SAM domain and of the free SRE RNA were solved by NMR [29]. A comparison of the structures free and bound showed that binding occurs via a rigid body fit mechanism. These structures also show that the SAM domain recognizes RNA in a shape-specific rather than sequence specific manner.

The PAZ (named after the proteins Piwi Argonaute and Zwiille) domain is an RNA binding domain found in Argonaute and some Dicer proteins involved in small interfering RNA (siRNA) biogenesis, adopts a $\beta\alpha\alpha\beta\beta\beta\alpha\beta\beta$ -fold and binds single-stranded RNA or RNA duplexes with a single-stranded 3'-overhang [403,404]. The NMR structure of the Argonaute 2 PAZ domain in complex with a single-stranded RNA (Fig. 34L) [25], together with the X-ray structure of the Argonaute 1 PAZ domain in complex with an RNA duplex containing a 3'-overhang [405] showed that the PAZ domain interacts solely with the 3' overhang and that the recognition of the 3' end of the RNA is achieved by steric exclusion because the structure cannot accommodate an extension of the phosphate backbone.

All the NMR structures of eukaryotic protein–RNA complexes were crucial to understand how proteins or protein domains specifically recognize their RNA targets, in a shape-specific or in a sequence-specific manner. For many years, RNA molecules, especially mRNAs were considered as passive molecules. More recently, it has been demonstrated that post-transcriptional gene regulations are at as important and possibly more important than transcription regulations. However, the main pathways governing these regulations remain to be elucidated. It is now clear that numerous RNA binding proteins play an important role in many cellular functions.

Structures of protein–RNA complexes provide the structural template required to understand the RNA recognition code that could be used to predict the RNA sequence bound by a certain protein, solely based on its amino acid sequence. The definition of such a code could be used to predict the functions of the numerous RNA binding proteins with unknown functions, and to design specific RNA binders to inhibit protein–RNA interactions. Several surveys of protein–RNA structures have attempted to understand the RNA recognition code [77,406,407]. These surveys identified numerous general features governing protein–RNA interactions, but also highlighted how this code seems to be highly complex and influenced by numerous parameters, such as the RNA secondary and tertiary structures, the cooperative binding of multiple RNA binding domains from a single RNA binding protein, or the competition of different RNA binding proteins for a same RNA binding site. Therefore, deciphering a code for protein–RNA recognition and its implication in post-transcriptional gene regulation will require further structural studies of such complexes, and recent advances in RNA and protein labeling (see Sections 2.2 and 2.3) and NMR methodologies (see Section 3.0) will be very useful, especially for studying molecular protein assembly onto RNA molecules that will require the study of multi-molecular complexes of high molecular weight.

4.4.2. Mechanistic insights provided by protein–RNA structures

Structural studies of protein–RNA complexes have provided insights into the mechanisms that control post-transcriptional gene regulation. For example, PTB is a general splicing repressor that contains four RNA binding domains of the RRM type and recognizes specifically RNA sequences rich in pyrimidines. NMR studies of PTB in complex with a CUCUCU RNA showed that each RRM is capable of binding RNA, and the structures of each RRM in complex with CUCUCU was solved [28,72]. Interestingly, it was shown that, while PTB RRM1 and RRM2 are independent in solution, PTB RRM3 and RRM4 interact with each other and have a fixed orientation relative to one another. 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 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 (Fig. 35). This

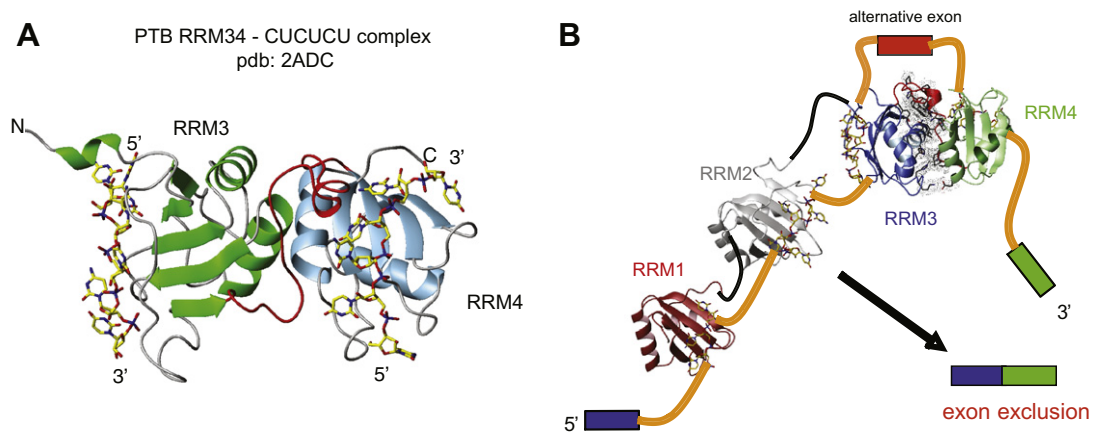


Fig. 35. Model representing how PTB represses alternative exon inclusion [28]. (A) Structure of PTB RRM34 in complex with two pyrimidine tracts. RRM3 and RRM4 are colored green and blue, respectively. The interdomain linker is colored red. The RNA molecules are colored yellow. (B) Structural model representing how PTB loops out RNA and represses alternative exon inclusion. RRM1–4 of PTB are colored red, grey, blue and green, respectively. Constitutive exons included are colored blue and green. The excluded alternative exon is colored red and labeled. The intronic regions are colored orange. Figures were generated with molmol [470].

unprecedented structural feature suggested that PTB can repress alternative splicing by looping out specific exons. PTB RRM1, RRM2 and RRM3 can bind a polypyrimidine tract upstream of an alternative exon while RRM4 can bind a polypyrimidine tract located downstream of this exon. The domain organization of PTB RRM34 will therefore loop out the alternative exon and induce its exclusion. This proposed model is consistent with biochemical data of various alternative splicing events, such as the GABA- γ 2 exon 9 repression [408] or the *c-src* N1 exon [409]. This model was recently confirmed using a combination of fluorescence resonance energy transfer (FRET) and NMR [410].

Another example for which NMR studies of protein–RNA has provided mechanistic functional insights concerns the NMR structure of hnRNP F in its complex with G-tract RNA [14]. This protein contains three RNA binding domains denoted qRRM and specifically recognizes single-stranded RNA containing a G-tract (3 or more consecutive guanines) [73,411]. The NMR structures of the three qRRMs of hnRNP F were recently solved and they explain how three consecutive guanines are specifically recognized by the qRRM [14]. This NMR study also demonstrated that G-tract RNAs are often structured in solution and that qRRM binding prevents RNA structure formation. Together with alternative splicing

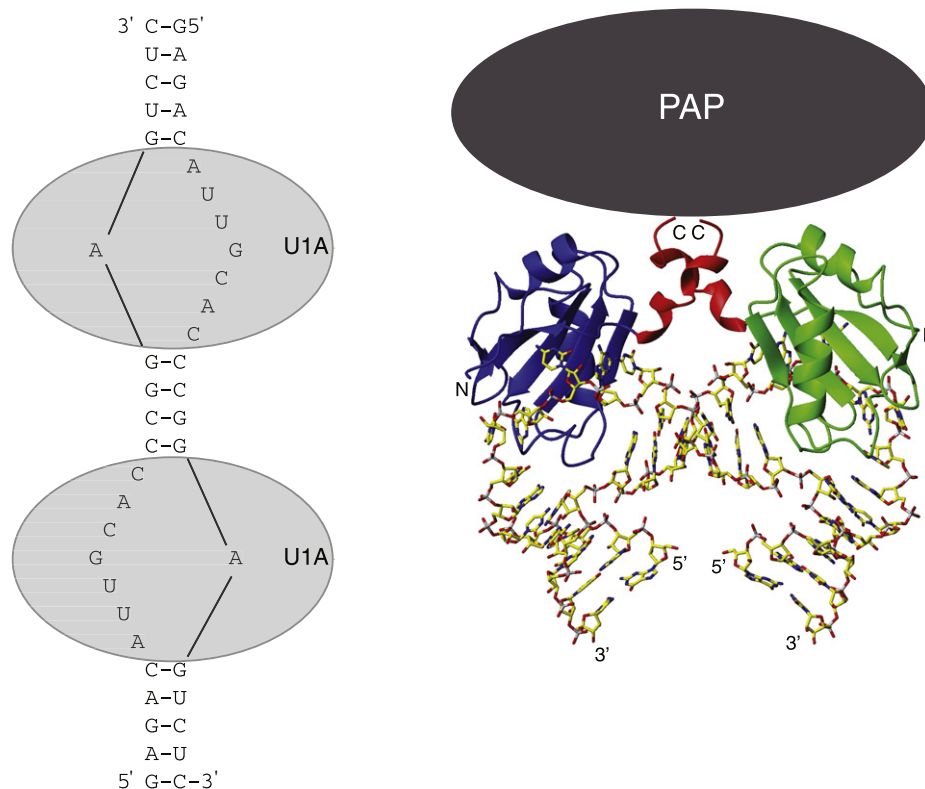


Fig. 36. Model of polyadenylation repression by U1A [38]. Left: schematic diagram of the RNA used in the NMR study. The U1A binding sites are circled. Right: structure of the trimolecular complex between the PIE RNA (stick structure colored yellow) and two U1A molecules (ribbon structure colored blue and green). The dimerizing C-terminal helix (colored red) recruits PAP. Figures were generated with molmol [470].

assays on a natural substrate showing that a single qRRM is sufficient for splicing regulation, a mechanistic model was proposed in which hnRNP F could regulate alternative splicing by remodeling RNA structures [14].

4.4.3. Structural insights into macromolecular assemblies

The molecular mechanisms that govern most post-transcriptional gene regulations are still poorly understood. These regulations involve mainly the binding of RNA binding proteins (trans-acting factors) that specifically recognize short RNA sequences (cis-acting elements). Some proteins act generally as post-transcriptional activators, others act as repressors, while some can act both as repressor and activator depending on the mRNA target. An mRNA contains numerous cis-acting elements bound by many proteins and the fate of the mRNA highly depends on which proteins bind cis-acting elements in specific tissues or at certain developmental stages. Current views suggest that numerous RNA binding proteins may compete to enhance or inhibit the use of a specific cis-acting element. Post-transcriptional gene regulation is therefore often modulated by the relative concentrations of these proteins in the nucleus. This modulation is achieved in different manner. Some proteins may be expressed only at certain stages of the development or only in specific tissues. Alternatively, the nuclear localization of these proteins can be regulated by post-translational modifications, such as phosphorylation, that affect the distribution of these factors between the nucleus and the cytoplasm. Finally, post-transcriptional gene regulation is often controlled by the binding of trans-acting factors that recruit other proteins, such as spliceosomal components in the case of alternative splicing or enzymes in the case of mRNA polyadenylation. These protein molecular assemblies onto the pre-mRNA are very important and it is therefore crucial to understand them at a structural level. The major difficulty in studying such assemblies by NMR is that it involves many molecules interacting with each other thus increasing significantly the molecular weight of such macromolecular complexes. However, recent developments in NMR methodologies (see Section 3.4.3) should allow such studies in the future.

In some cases, NMR structures of protein–RNA complexes have provided valuable insights into the understanding of such molecular assemblies onto the mRNA. 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 is also involved in polyadenylation regulation by preventing the formation of the poly(A) tail of its own mRNA through binding an RNA sequence in the 3' UTR, called the polyadenylation inhibition element (PIE), located at a conserved distance from the polyadenylation site [412,413]. Two molecules of U1A bind cooperatively the PIE element via their N-terminal 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 its complex with the PIE RNA led to a model for U1A–PAP complex formation [2,38]. 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 (Fig. 36) [38]. Using EMSA, the authors showed that the spacing between the two U1A binding sites is optimal in the natural RNA sequence for inducing the dimerization of U1A. The dimerization involving the C-terminal helix, brings the PAP interacting regions into close proximity and on the same side of the structure. The conformation of the proteins observed in the NMR structure is optimal for binding PAP and therefore to repress polyadenylation. Similar to this study of the binding of two U1A molecules to its own 3' UTR, more examples of assembly of multiple proteins on RNA are

likely to be studied in the future in order to understand post-transcriptional regulatory processes with NMR as the primary method of investigation.

5. Conclusion

We now arrive at the end of this extensive review that describes the achievement over the last fifteen years in the field of NMR structure determination of protein–RNA complexes. Unlike other structural biology areas, the number of structures has increased only linearly when one might have expected an exponential increase. Although, NMR is certainly today a mature method for solving the structures of most protein–RNA complexes below 20 kDa, every structure determination of such a complex is still a challenge in itself. Although this field has benefited tremendously from the technological development in the field of NMR such as high sensitivity, RDCs, fast computing, or semi-automated protein structure determination, solving a protein–RNA complex by NMR still requires significant manual intervention and the need to master the spectroscopy of both the protein and the RNA components. However, as hopefully shown here, it is very clear that NMR spectroscopy has now been shown to be a very competitive method for investigating the structures of protein–RNA complexes. Not only will more protein–RNA complex structures be determined in the near future but NMR will be particularly useful for investigating how several RNA binding proteins assemble or compete for binding RNA. Protein–RNA interactions being at the heart of every molecular mechanisms controlling post-transcriptional gene expression, we have in front of us as biomolecular NMR spectroscopists an infinite and very attractive field of study for decades to come.

6. Note added in proof

After the revision process of this review, three additional protein–RNA complexes were published. The 34 kDa ternary complex of the two RRM domains of Hrp1 and the RRM of Rna15 bound to RNA provide insights into 3'-processing of mRNA in yeast [471]. The structure of the human PHAX-RBD in complex with a 4 nt RNA revealed a novel RNA binding motif and lead to a model of snRNA export [472]. A structure of HIV TAR RNA in complex with an extended designed cyclic peptide with improved affinity displays a larger interaction site and gives clues for further improvement of those antiviral leads [473]. In addition, a fast, efficient and sequence-independent method for multiple segmental isotope labeling of RNA has been published [474]. In contrast to previously reported methods, there are no sequence requirements and up to 10-fold higher yields can be obtained.

Acknowledgements

We would like to thank F. Oberstrass for his involvement in the initial stage of this review, S. Auweter for providing spectra of the Fox-1-RNA complex, F. Loughlin 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] F.H. Allain, P. Bouvet, T. Dieckmann, J. Feigon, EMBO J. 19 (2000) 6870.
- [2] F.H. Allain, C.C. Gubser, P.W. Howe, K. Nagai, D. Neuhaus, G. Varani, Nature 380 (1996) 646.
- [3] F.H. Allain, P.W. Howe, D. Neuhaus, G. Varani, EMBO J. 16 (1997) 5764.
- [4] G.K. Amarasinghe, R.N. De Guzman, R.B. Turner, K.J. Chancellor, Z.R. Wu, M.F. Summers, J. Mol. Biol. 301 (2000) 491.

- [5] S.D. Auweter, R. Fasan, L. Reymond, J.G. Underwood, D.L. Black, S. Pitsch, F.H. Allain, *EMBO J.* 25 (2006) 163.
- [6] J.L. Battiste, H. Mao, N.S. Rao, R. Tan, D.R. Muhandiram, L.E. Kay, A.D. Frankel, J.R. Williamson, *Science* 273 (1996) 1547.
- [7] Z. Cai, A. Gorin, R. Frederick, X. Ye, W. Hu, A. Majumdar, A. Kettani, D.J. Patel, *Nat. Struct. Biol.* 5 (1998) 203.
- [8] V. Calabro, M.D. Daugherty, A.D. Frankel, *Proc. Natl. Acad. Sci. USA* 102 (2005) 6849.
- [9] C.D. Cille, J.R. Williamson, *RNA* 9 (2003) 663.
- [10] V. D'Souza, M.F. Summers, *Nature* 431 (2004) 586.
- [11] A. Davidson, T.C. Leeper, Z. Athanassiou, K. Patora-Komisarska, J. Karn, J.A. Robinson, G. Varani, *Proc. Natl. Acad. Sci. USA* 106 (2009) 11931.
- [12] R.N. De Guzman, Z.R. Wu, C.C. Stalling, L. Pappalardo, P.N. Borer, M.F. Summers, *Science* 279 (1998) 384.
- [13] A. Dey, D. York, A. Smalls-Mantey, M.F. Summers, *Biochemistry* 44 (2005) 3735.
- [14] C. Dominguez, J.F. Fisette, B. Chabot, F.H.-T. Allain, *Nat. Struct. Mol. Biol.* 17 (2010) 853.
- [15] C. Faber, M. Scharpf, T. Becker, H. Sticht, P. Rosch, *J. Biol. Chem.* 276 (2001) 32064.
- [16] Y. Gosser, T. Hermann, A. Majumdar, W. Hu, R. Frederick, F. Jiang, W. Xu, D.J. Patel, *Nat. Struct. Biol.* 8 (2001) 146.
- [17] Y. Hargous, G.M. Hautbergue, A.M. Tintaru, L. Skrisovska, A.P. Golovanov, J. Stevenin, L.Y. Lian, S.A. Wilson, F.H. Allain, *EMBO J.* 25 (2006) 5126.
- [18] B.P. Hudson, M.A. Martinez-Yamout, H.J. Dyson, P.E. Wright, *Nat. Struct. Mol. Biol.* 11 (2004) 257.
- [19] F. Jiang, A. Gorin, W. Hu, A. Majumdar, S. Baskerville, W. Xu, A. Ellington, D.J. Patel, *Structure* 7 (1999) 1461.
- [20] C. Johansson, L.D. Finger, L. Trantirek, T.D. Mueller, S. Kim, I.A. Laird-Offringa, J. Feigon, *J. Mol. Biol.* 337 (2004) 799.
- [21] P.E. Johnson, L.W. Donaldson, *Nat. Struct. Mol. Biol.* 13 (2006) 177.
- [22] B.M. Lee, J. Xu, B.K. Clarkson, M.A. Martinez-Yamout, H.J. Dyson, D.A. Case, J.M. Gottesfeld, P.E. Wright, *J. Mol. Biol.* 357 (2006) 275.
- [23] T.C. Leeper, Z. Athanassiou, R.L. Dias, J.A. Robinson, G. Varani, *Biochemistry* 44 (2005) 12362.
- [24] P. Legault, J. Li, J. Mogridge, L.E. Kay, J. Greenblatt, *Cell* 93 (1998) 289.
- [25] A. Lingel, B. Simon, E. Izaurralde, M. Sattler, *Nat. Struct. Mol. Biol.* 11 (2004) 576.
- [26] Z. Liu, I. Luyten, M.J. Bottomley, A.C. Messias, S. Houngrinou-Molango, R. Sprangers, K. Zanier, A. Kramer, M. Sattler, *Science* 294 (2001) 1098.
- [27] S. Martin-Tomasz, N.J. Reiter, D.A. Brow, S.E. Butcher, *RNA* 16 (2010) 792.
- [28] F.C. Oberstrass, S.D. Auweter, M. Erat, Y. Hargous, A. Henning, P. Wenter, L. Reymond, B. Amir-Ahmady, S. Pitsch, D.L. Black, F.H. Allain, *Science* 309 (2005) 2054.
- [29] F.C. Oberstrass, A. Lee, R. Steffl, M. Janis, G. Chanfreau, F.H. Allain, *Nat. Struct. Mol. Biol.* 13 (2006) 160.
- [30] J.M. Perez-Canadillas, *EMBO J.* 25 (2006) 3167.
- [31] J.D. Puglisi, L. Chen, S. Blanchard, A.D. Frankel, *Science* 270 (1995) 1200.
- [32] A. Ramos, S. Grunert, J. Adams, D.R. Micklem, M.R. Proctor, S. Freund, M. Bycroft, D. St Johnston, G. Varani, *EMBO J.* 19 (2000) 997.
- [33] M. Schärpf, H. Sticht, K. Schweimer, M. Boehm, S. Hoffmann, P. Rösch, *Eur. J. Biochem.* 267 (2000) 2397.
- [34] M. Schubert, K. Lapouge, O. Duss, F.C. Oberstrass, I. Jelesarov, D. Haas, F.H. Allain, *Nat. Struct. Mol. Biol.* 14 (2007) 807.
- [35] L. Skrisovska, C.F. Bourgeois, R. Steffl, S.N. Grellescheid, L. Kister, P. Wenter, D.J. Elliott, J. Stevenin, F.H. Allain, *EMBO Rep.* 8 (2007) 372.
- [36] M. Stoldt, J. Wohnert, O. Ohlenschlager, M. Grolach, L.R. Brown, *EMBO J.* 18 (1999) 6508.
- [37] K. Tsuda, K. Kuwasako, M. Takahashi, T. Someya, M. Inoue, T. Terada, N. Kobayashi, M. Shirouzu, T. Kigawa, A. Tanaka, S. Sugano, P. Guntert, Y. Muto, S. Yokoyama, *Nucleic Acids Res.* 37 (2009) 5151.
- [38] L. Varani, S.I. Gunderson, I.W. Mattaj, L.E. Kay, D. Neuhau, G. Varani, *Nat. Struct. Biol.* 7 (2000) 329.
- [39] H. Wu, A. Henras, G. Chanfreau, J. Feigon, *Proc. Natl. Acad. Sci. USA* 101 (2004) 8307.
- [40] Y. Yang, N. Declerck, X. Manival, S. Aymerich, M. Kochoyan, *EMBO J.* 21 (2002) 1987.
- [41] X. Ye, A. Gorin, A.D. Ellington, D.J. Patel, *Nat. Struct. Biol.* 3 (1996) 1026.
- [42] X. Ye, A. Gorin, R. Frederick, W. Hu, A. Majumdar, W. Xu, G. McLendon, A. Ellington, D.J. Patel, *Chem. Biol.* 6 (1999) 657.
- [43] X. Ye, R.A. Kumar, D.J. Patel, *Chem. Biol.* 2 (1995) 827.
- [44] Q. Zhang, K. Harada, H.S. Cho, A.D. Frankel, D.E. Wemmer, *Chem. Biol.* 8 (2001) 511.
- [45] J. Zhou, R.L. Bean, V.M. Vogt, M. Summers, *J. Mol. Biol.* 365 (2007) 453.
- [46] H.R. Jonker, S. Ilin, S.K. Grimm, J. Wohnert, H. Schwalbe, *Nucleic Acids Res.* 35 (2007) 441.
- [47] M.B. Kamphuis, A.M. Bonvin, M.C. Monti, M. Lemonnier, A. Munoz-Gomez, R.H. van den Heuvel, R. Diaz-Orejas, R. Boelens, *J. Mol. Biol.* 357 (2006) 115.
- [48] D.J. Galas, A. Schmitz, *Nucleic Acids Res.* 5 (1978) 3157.
- [49] M.M. Garner, A. Revzin, *Nucleic Acids Res.* 9 (1981) 3047.
- [50] M. Fried, D.M. Crothers, *Nucleic Acids Res.* 9 (1981) 6505.
- [51] F.J. Schmidt, J. Thompson, K. Lee, J. Dijk, E. Cundliffe, *J. Biol. Chem.* 256 (1981) 12301.
- [52] S. Douthwaite, A. Christensen, R.A. Garrett, *Biochemistry* 21 (1982) 2313.
- [53] K.S. Cook, G.J. Fisk, J. Hauber, N. Usman, T.J. Daly, J.R. Rusche, *Nucleic Acids Res.* 19 (1991) 1577.
- [54] T.J. Daly, K.S. Cook, G.S. Gray, T.E. Maione, J.R. Rusche, *Nature* 342 (1989) 816.
- [55] V. D'Souza, J. Melamed, D. Habib, K. Pullen, K. Wallace, M.F. Summers, *J. Mol. Biol.* 314 (2001) 217.
- [56] R.C. Chan, D.L. Black, *Mol. Cell. Biol.* 17 (1997) 4667.
- [57] J. Ule, K.B. Jensen, M. Ruggiu, A. Mele, A. Ule, R.B. Darnell, *Science* 302 (2003) 1212.
- [58] J. Ule, K. Jensen, A. Mele, R.B. Darnell, *Methods* 37 (2005) 376.
- [59] J.W. Hockensmith, W.L. Kubasek, W.R. Vorachek, P.H. von Hippel, *J. Biol. Chem.* 261 (1986) 3512.
- [60] G.W. Yeo, N.G. Coufal, T.Y. Liang, G.E. Peng, X.D. Fu, F.H. Gage, *Nat. Struct. Mol. Biol.* 16 (2009) 130.
- [61] Y. Jin, H. Suzuki, S. Maegawa, H. Endo, S. Sugano, K. Hashimoto, K. Yasuda, K. Inoue, *EMBO J.* 22 (2003) 905.
- [62] A.D. Ellington, J.W. Szostak, *Nature* 346 (1990) 818.
- [63] C. Tuerk, L. Gold, *Science* 249 (1990) 505.
- [64] M. Djordjevic, *Biomol. Eng.* 24 (2007) 179.
- [65] R. Stoltenburg, C. Reinemann, B. Strehlitz, *Biomol. Eng.* 24 (2007) 381.
- [66] Y. Cavaloc, C.F. Bourgeois, L. Kister, J. Stevenin, *RNA* 5 (1999) 468.
- [67] M.T. Worthington, J.W. Pelo, M.A. Sachedina, J.L. Applegate, K.O. Arsenau, T.T. Pizarro, *J. Biol. Chem.* 277 (2002) 48558.
- [68] L. Ghisolfi-Nieto, G. Joseph, F. Puvion-Dutilleul, F. Amalric, P. Bouvet, *J. Mol. Biol.* 260 (1996) 34.
- [69] R. Tacke, J.L. Manley, *EMBO J.* 14 (1995) 3540.
- [70] K.B. Jensen, K. Musunuru, H.A. Lewis, S.K. Burley, R.B. Darnell, *Proc. Natl. Acad. Sci. USA* 97 (2000) 5740.
- [71] B. Beuth, M.F. Garcia-Mayoral, I.A. Taylor, A. Ramos, *J. Am. Chem. Soc.* 129 (2007) 10205.
- [72] S.D. Auweter, F.C. Oberstrass, F.H.T. Allain, *J. Mol. Biol.* 367 (2007) 174.
- [73] M. Caputi, A.M. Zahler, *J. Biol. Chem.* 276 (2001) 43850.
- [74] A.M. Edwards, C.H. Arrowsmith, D. Christendat, A. Dharamsi, J.D. Friesen, J.F. Greenblatt, M. Vedadi, *Nat. Struct. Biol.* 7 (Suppl) (2000) 970.
- [75] S. Yokoyama, *Curr. Opin. Chem. Biol.* 7 (2003) 39.
- [76] R. Steffl, L. Skrisovska, F.H.T. Allain, *EMBO Rep.* 6 (2005) 33.
- [77] B.M. Lunde, C. Moore, G. Varani, *Nat. Rev. Mol. Cell. Biol.* 8 (2007) 479.
- [78] L. Kong, S. Ranganathan, *Brief. Bioinform.* 5 (2004) 179.
- [79] M.R. Conte, T. Grune, J. Ghuman, G. Kelly, A. Ladas, S. Matthews, S. Curry, *EMBO J.* 19 (2000) 3132.
- [80] P.J. Simpson, T.P. Monie, A. Szendroi, N. Davydova, J.K. Tyzack, M.R. Conte, C.M. Read, P.D. Cary, D.I. Svergun, P.V. Konarev, S. Curry, S. Matthews, *Structure* 12 (2004) 1631.
- [81] M. Bycroft, S. Grunert, A.G. Murzin, M. Proctor, D. Johnston, *EMBO J.* 14 (1995) 3563.
- [82] A. Kharrat, M.J. Macias, T.J. Gibson, M. Nilges, A. Pastore, *EMBO J.* 14 (1995) 3572.
- [83] N. Leulliot, S. Quevillon-Cheruel, M. Graille, H. van Tilbeurgh, T.C. Leeper, K.S. Godin, T.E. Edwards, S.T. Sigurdsson, N. Rozenkrants, R.J. Nagel, M. Ares, G. Varani, *EMBO J.* 23 (2004) 2468.
- [84] P. Zhou, G. Wagner, *J. Biomol. NMR* (2009).
- [85] P. Zhou, A.A. Lugovskoy, G. Wagner, *J. Biomol. NMR* 20 (2001) 11.
- [86] Y. Kobashigawa, H. Kumeta, K. Ogura, F. Inagaki, *J. Biomol. NMR* 43 (2009) 145.
- [87] L. Skrisovska, M. Schubert, F.H. Allain, *J. Biomol. NMR* 46 (2010) 51.
- [88] T. Kigawa, Y. Muto, S. Yokoyama, *J. Biomol. NMR* 6 (1995) 129.
- [89] T. Kigawa, T. Yabuki, Y. Yoshida, M. Tsutsui, Y. Ito, T. Shibata, S. Yokoyama, *FEBS Lett.* 442 (1999) 15.
- [90] A.S. Spirin, *Trends Biotechnol.* 22 (2004) 538.
- [91] K. Ozawa, N.E. Dixon, G. Otting, *IUBMB Life* 57 (2005) 615.
- [92] S. Pitsch, P.A. Weiss, *Curr. Protoc. Nucleic Acid Chem.* Chapter 3 (2002) Unit 3.8.
- [93] S.A. Scaringe, F.E. Wincott, M.H. Caruthers, *J. Am. Chem. Soc.* 120 (1998) 11820.
- [94] T. Ohtsuki, R. Vinayak, Y. Watanabe, K. Kita, G. Kawai, K. Watanabe, *J. Biochem.* 120 (1996) 1070.
- [95] A. Ramos, G. Varani, *J. Am. Chem. Soc.* 120 (1998) 10992.
- [96] P. Wenter, L. Reymond, S.D. Auweter, F.H. Allain, S. Pitsch, *Nucleic Acids Res.* 34 (2006) e79.
- [97] V.V. Gurevich, I.D. Pokrovskaya, T.A. Obukhova, S.A. Zozulya, *Anal. Biochem.* 195 (1991) 207.
- [98] I.D. Pokrovskaya, V.V. Gurevich, *Anal. Biochem.* 220 (1994) 420.
- [99] J.F. Milligan, D.R. Groebe, G.W. Witherell, O.C. Uhlenbeck, *Nucleic Acids Res.* 15 (1987) 8783.
- [100] S.R. Price, N. Ito, C. Oubridge, J.M. Avis, K. Nagai, *J. Mol. Biol.* 249 (1995) 398.
- [101] E.P. Nikonowicz, A. Sirr, P. Legault, F.M. Jucker, L.M. Baer, A. Pardi, *Nucleic Acids Res.* 20 (1992) 4507.
- [102] R.T. Batey, J.L. Battiste, J.R. Williamson, *Methods Enzymol.* 261 (1995) 300.
- [103] J. Cromsigt, J. Schleucher, T. Gustafsson, J. Kihlberg, S. Wijmenga, *Nucleic Acids Res.* 30 (2002) 1639.
- [104] L.G. Scott, T.J. Tolbert, J.R. Williamson, *Methods Enzymol.* 317 (2000) 18.
- [105] K. Lu, Y. Miyazaki, M.F. Summers, *J. Biomol. NMR* (2009).
- [106] R.P. Price, C. Oubridge, G. Varani, K. Nagai, in: C.W.J. Smith (Ed.), *RNA-Protein Interactions: A Practical Approach*, Oxford University Press, 1998.
- [107] S. Gallo, M. Furler, R.K.O. Sigel, *Chimia* 59 (2005) 812.
- [108] J.A. Pleiss, M.L. Derrick, O.C. Uhlenbeck, *RNA* 4 (1998) 1313.
- [109] K.T. Dayie, *Int. J. Mol. Sci.* 9 (2008) 1214.
- [110] A.R. Ferre-D'Amare, J.A. Doudna, *Nucleic Acids Res.* 24 (1996) 977.
- [111] S.C. Walker, J.M. Avis, G.L. Conn, *Nucleic Acids Res.* 31 (2003).

- [112] H. Schurer, K. Lang, J. Schuster, M. Morl, *Nucleic Acids Res.* 30 (2002).
- [113] H.C. Guo, R.A. Collins, *EMBO J.* 14 (1995) 368.
- [114] T.P. Shields, E. Molloy, L.S. Marie, M.R. Hansen, A. Pardi, *RNA* 5 (1999) 1259.
- [115] S.W. Santoro, G.F. Joyce, *Proc. Natl. Acad. Sci. USA* 94 (1997) 4262.
- [116] S.W. Santoro, G.F. Joyce, *Biochemistry* 37 (1998) 13330.
- [117] A.M. Pyle, V.T. Chu, E. Jankowsky, M. Boudvillain, *Methods Enzymol.* 317 (2000) 140.
- [118] H. Inoue, Y. Hayase, S. Iwai, E. Ohtsuka, *FEBS Lett.* 215 (1987) 327.
- [119] J. Lapham, D.M. Crothers, *RNA* 2 (1996) 289.
- [120] C. Kao, M. Zheng, S. Rudisser, *RNA* 5 (1999) 1268.
- [121] L. Ponchon, G. Beauvais, S. Nonin-Lecomte, F. Dardel, *Nat. Protoc.* 4 (2009) 947.
- [122] L. Ponchon, F. Dardel, *Nat. Methods* 4 (2007) 571.
- [123] I. Kim, S.A. McKenna, E.V. Puglisi, J.D. Puglisi, *RNA* 13 (2007) 289.
- [124] P.J. Lukavsky, J.D. Puglisi, *RNA* 10 (2004) 889.
- [125] A.C. Anderson, S.A. Scaringe, B.E. Earp, C.A. Frederick, *RNA* 2 (1996) 110.
- [126] L.E. Easton, Y. Shibata, P.J. Lukavsky, *RNA* 16 (2010) 647.
- [127] S.A. McKenna, I. Kim, E.V. Puglisi, D.A. Lindhout, C.E. Aitken, R.A. Marshall, J.D. Puglisi, *Nat. Protoc.* 2 (2007) 370.
- [128] J.B. Murray, A.K. Collier, J.R.P. Arnold, *Anal. Biochem.* 218 (1994) 177.
- [129] J.S. Kieft, R.T. Batey, *RNA* 10 (2004) 988.
- [130] R.T. Batey, J.S. Kieft, *RNA* 13 (2007) 1384.
- [131] S.C. Walker, F.H. Scott, C. Srisawat, D.R. Engelke, *Methods Mol. Biol.* 488 (2008) 23.
- [132] H.K. Cheong, E. Hwang, C. Lee, B.S. Choi, C. Cheong, *Nucleic Acids Res.* 32 (2004) e84.
- [133] P.J. Lukavsky, I. Kim, G.A. Otto, J.D. Puglisi, *Nat. Struct. Biol.* 10 (2003) 1033.
- [134] I. Kim, P.J. Lukavsky, J.D. Puglisi, *J. Am. Chem. Soc.* 124 (2002) 9338.
- [135] M.J. Frilander, J.J. Turunen, in: R.K. Hartmann et al. (Eds.), *Handbook of RNA Biochemistry*, WILEY-VCH Verlag GmbH & Co., Weinheim, 2005.
- [136] T. Persson, D.K. Willkomm, R.K. Hartmann, in: R.K. Hartmann, et al. (Eds.), *Handbook of RNA Biochemistry*, WILEY-VCH Verlag GmbH & Co., Weinheim, (2005).
- [137] M.J. Moore, C.C. Query, *Methods Enzymol.* 317 (2000) 109.
- [138] W.E. Purtha, R.L. Coppins, M.K. Smalley, S.K. Silverman, *J. Am. Chem. Soc.* 127 (2005) 13124.
- [139] K. Lang, R. Micura, *Nat. Protoc.* 3 (2008) 1457.
- [140] M.R. Stark, J.A. Pleiss, M. Deras, S.A. Scaringe, S.D. Rader, *RNA* 12 (2006) 2014.
- [141] T. Ohtsuki, G. Kawai, K. Watanabe, *J. Biochem.* 124 (1998) 28.
- [142] T. Ohtsuki, G. Kawai, Y. Watanabe, K. Kita, K. Nishikawa, K. Watanabe, *Nucleic Acids Res.* 24 (1996) 662.
- [143] A.G. Tzakos, L.E. Easton, P.J. Lukavsky, *J. Am. Chem. Soc.* 128 (2006) 13344.
- [144] A.G. Tzakos, L.E. Easton, P.J. Lukavsky, *Nat. Protoc.* 2 (2007) 2139.
- [145] J. Xu, J. Lapham, D.M. Crothers, *Proc. Natl. Acad. Sci. USA* 93 (1996) 44.
- [146] C. Hobartner, R. Rieder, C. Kreutz, B. Puffer, K. Lang, A. Polonskaia, A. Serganov, R. Micura, *J. Am. Chem. Soc.* 127 (2005) 12035.
- [147] F.H. Nelissen, A.J. van Gammeren, M. Tessari, F.C. Girard, H.A. Heus, S.S. Wijmenga, *Nucleic Acids Res.* 36 (2008) e89.
- [148] W.J. Waddell, *J. Lab. Clin. Med.* 48 (1956) 311.
- [149] M.M. Bradford, *Anal. Biochem.* 72 (1976) 248.
- [150] G. Wider, L. Dreier, *J. Am. Chem. Soc.* 128 (2006) 2571.
- [151] A. Marintchev, D. Frueh, G. Wagner, *Methods Enzymol.* 430 (2007) 283.
- [152] E.R.P. Zuiderweg, *Biochemistry* 41 (2002) 1.
- [153] C. Dominguez, F.H.T. Allain, *Nucleic Acids Res.* 34 (2006) 3634.
- [154] X. Manival, Y.S. Yang, M.P. Strub, M. Kochoyan, M. Steinmetz, S. Aymerich, *EMBO J.* 16 (1997) 5019.
- [155] H. van Tilbeurgh, X. Manival, S. Aymerich, J.M. Lhoste, C. Dumas, M. Kochoyan, *EMBO J.* 16 (1997) 5030.
- [156] D. Hargreaves, S. Santos-Sierra, R. Giraldo, R. Sabariego-Jareno, G. de la Cueva-Mendez, R. Boelens, R. Diaz-Orejas, J.B. Rafferty, *Structure* 10 (2002) 1425.
- [157] P. Gutierrez, Y. Li, M.J. Osborne, E. Pomerantseva, Q. Liu, K. Gehring, *J. Bacteriol.* 187 (2005) 3496.
- [158] C. Rife, R. Schwarzenbacher, D. McMullan, P. Abdubek, E. Ambing, H. Axelrod, T. Biorac, J.M. Canaves, H.J. Chiu, A.M. Deacon, M. DiDonato, M.A. Elslinger, A. Godzik, C. Grittini, S.K. Grzechnik, J. Hale, E. Hampton, G.W. Han, J. Haugen, M. Hornsby, L. Jaroszewski, H.E. Klock, E. Koesema, A. Kreusch, P. Kuhn, S.A. Lesley, M.D. Miller, K. Moy, E. Nigoghossian, J. Paulsen, K. Quijano, R. Reyes, E. Sims, G. Spraggon, R.C. Stevens, H. van den Bedem, J. Velasquez, J. Vincent, A. White, G. Wolf, Q.P. Xu, K.O. Hodgson, J. Wooley, I.A. Wilson, *Proteins* 61 (2005) 449.
- [159] C. Dominguez, R. Boelens, A.M.J.J. Bonvin, *J. Am. Chem. Soc.* 125 (2003) 1731.
- [160] V. Sklenar, A. Bax, *J. Magn. Reson.* 74 (1987) 469.
- [161] V. Sklenar, M. Piotto, R. Leppik, V. Saudek, *J. Magn. Reson. A* 102 (1993) 241.
- [162] A.P. Golovanov, G.M. Hautbergue, S.A. Wilson, L.Y. Lian, *J. Am. Chem. Soc.* 126 (2004) 8933.
- [163] G.M. Hautbergue, A.P. Golovanov, *J. Magn. Reson.* 191 (2008) 335.
- [164] F. Vitali, A. Henning, F.C. Oberstrass, Y. Hargous, S.D. Auweter, M. Erat, F.H.T. Allain, *EMBO J.* 25 (2006) 150.
- [165] C. Zwaalen, P. Legault, S.J.F. Vincent, J. Greenblatt, R. Konrat, L.E. Kay, *J. Am. Chem. Soc.* 119 (1997) 6711.
- [166] M. Sattler, J. Schleucher, C. Griesinger, *Prog. Nucl. Magn. Reson. Spectrosc.* 34 (1999) 93.
- [167] J. Cavanagh, W.J. Fairbrother, A.G. Palmer, M. Rance, N.J. Skelton, *Protein NMR Spectroscopy: Principles and Practice*, Academic Press Inc., San Diego, 2006.
- [168] N.S. Rao, P. Legault, D.R. Muhandiram, J. Greenblatt, J.L. Battiste, J.R. Williamson, L.E. Kay, *J. Magn. Reson. B* 113 (1996) 272.
- [169] F.A.A. Mulder, C.A.E.M. Spronk, M. Slijper, R. Kaptein, R. Boelens, *J. Biomol. NMR* 8 (1996) 223.
- [170] P. Barraud, F. Dardel, C. Tisne, *C.R. Chim.* 11 (2008) 474.
- [171] J. Engelke, H. Ruterjans, *J. Biomol. NMR* 9 (1997) 63.
- [172] K.T. Dayie, G. Wagner, *J. Am. Chem. Soc.* 119 (1997) 7797.
- [173] R.A. Venters, C.C. Huang, B.T. Farmer 2nd, R. Trolard, L.D. Spicer, C.A. Fierke, *J. Biomol. NMR* 5 (1995) 339.
- [174] K.H. Gardner, L.E. Kay, *Annu. Rev. Biophys. Biomol. Struct.* 27 (1998) 357.
- [175] R.A. Venters, R. Thompson, J. Cavanagh, *J. Mol. Struct.* 602 (2002) 275.
- [176] M. Salzmann, G. Wider, K. Pervushin, K. Wuthrich, *J. Biomol. NMR* 15 (1999) 181.
- [177] M. Salzmann, K. Pervushin, G. Wider, H. Senn, K. Wuthrich, *Proc. Natl. Acad. Sci. USA* 95 (1998) 13585.
- [178] V. Tugarinov, V. Kanelis, L.E. Kay, *Nat. Protoc.* 1 (2006) 749.
- [179] V. Tugarinov, L.E. Kay, *J. Am. Chem. Soc.* 125 (2003) 13868.
- [180] V. Tugarinov, L.E. Kay, *J. Biomol. NMR* 28 (2004) 165.
- [181] M. Kainosho, T. Torizawa, Y. Iwashita, T. Terauchi, A. Mei Ono, P. Guntert, *Nature* 440 (2006) 52.
- [182] T. Torizawa, M. Shimizu, M. Taoka, H. Miyano, M. Kainosho, *J. Biomol. NMR* 30 (2004) 311.
- [183] V. Muralidharan, T.W. Muir, *Nat. Methods* 3 (2006) 429.
- [184] J. Crossigt, B. van Buuren, J. Schleucher, S. Wijmenga, *Methods Enzymol.* 338 (2001) 371.
- [185] G. Varani, F. Aboulela, F.H.T. Allain, *Prog. Nucl. Magn. Reson. Spectrosc.* 29 (1996) 51.
- [186] S.S. Wijmenga, B.N.M. van Buuren, *Prog. Nucl. Magn. Reson. Spectrosc.* 32 (1998) 287.
- [187] B. Furtig, C. Richter, J. Wohnert, H. Schwalbe, *Chembiochem* 4 (2003) 936.
- [188] L.G. Scott, M. Hennig, *Methods Mol. Biol.* 452 (2008) 29.
- [189] V. Sklenar, R.D. Peterson, M.R. Rejante, J. Feigon, *J. Biomol. NMR* 3 (1993) 721.
- [190] R. Fiala, F. Jiang, V. Sklenar, *J. Biomol. NMR* 12 (1998) 373.
- [191] L. Zidek, R. Stefl, V. Sklenar, *Curr. Opin. Struct. Biol.* 11 (2001) 275.
- [192] R. Fiala, J. Czernek, V. Sklenar, *J. Biomol. NMR* 16 (2000) 291.
- [193] W. Hu, L. Jiang, Y.Q. Gosser, *J. Magn. Reson.* 145 (2000) 147.
- [194] R. Riek, K. Pervushin, C. Fernandez, M. Kainosho, K. Wuthrich, *J. Am. Chem. Soc.* 123 (2001) 658.
- [195] T. Dieckmann, J. Feigon, *J. Biomol. NMR* 9 (1997) 259.
- [196] R.D. Peterson, C.A. Theimer, H. Wu, J. Feigon, *J. Biomol. NMR* 28 (2004) 59.
- [197] V. D'Souza, A. Dey, D. Habib, M.F. Summers, *J. Mol. Biol.* 337 (2004) 427.
- [198] J.P. Marino, H. Schwalbe, C. Anklin, W. Bermel, D.M. Crothers, C. Griesinger, *J. Am. Chem. Soc.* 116 (1994) 6472.
- [199] H.A. Heus, S.S. Wijmenga, F.J.M. Vandeven, C.W. Hilbers, *J. Am. Chem. Soc.* 116 (1994) 4983.
- [200] G. Varani, F. Aboul-ela, F. Allain, C.C. Gubser, *J. Biomol. NMR* 5 (1995) 315.
- [201] F. Aboul-ela, J. Karn, G. Varani, *J. Mol. Biol.* 253 (1995) 313.
- [202] M.H. Kolk, S.S. Wijmenga, H.A. Heus, C.W. Hilbers, *J. Biomol. NMR* 12 (1998) 423.
- [203] T.J. Tolbert, J.R. Williamson, *J. Am. Chem. Soc.* 118 (1996) 7929.
- [204] T.J. Tolbert, J.R. Williamson, *J. Am. Chem. Soc.* 119 (1997) 12100.
- [205] A.L. Breeze, *Prog. Nucl. Magn. Reson. Spectrosc.* 36 (2000) 323.
- [206] G. Otting, K. Wuthrich, *Quart. Rev. Biophys.* 23 (1990) 39.
- [207] G. Otting, H. Senn, G. Wagner, K. Wuthrich, *J. Magn. Reson.* 70 (1986) 500.
- [208] G. Otting, K. Wuthrich, *J. Magn. Reson.* 85 (1989) 586.
- [209] K. Ogura, H. Terasawa, F. Inagaki, *J. Biomol. NMR* 8 (1996) 492.
- [210] E. Kupce, R. Freeman, *J. Magn. Reson. A* 115 (1995) 273.
- [211] J. Iwahara, J.M. Wojciak, R.T. Clubb, *J. Biomol. NMR* 19 (2001) 231.
- [212] W. Lee, M.J. Revington, C. Arrowsmith, L.E. Kay, *FEBS Lett.* 350 (1994) 87.
- [213] K.J. Walters, A.E. Ferentz, B.J. Hare, P. Hidalgo, A. Jasanoff, H. Matsuo, G. Wagner, *Methods Enzymol.* 339 (2001) 238.
- [214] S. Grzesiek, F. Cordier, A.J. Dingley, *Methods Enzymol.* 338 (2001) 111.
- [215] A.J. Dingley, S. Grzesiek, *J. Am. Chem. Soc.* 120 (1998) 8293.
- [216] K. Pervushin, A. Ono, C. Fernandez, T. Szyperki, M. Kainosho, K. Wuthrich, *Proc. Natl. Acad. Sci. USA* 95 (1998) 14147.
- [217] A.J. Dingley, L. Nisius, F. Cordier, S. Grzesiek, *Nat. Protoc.* 3 (2008) 242.
- [218] A. Majumdar, A. Kettani, E. Skripkin, *J. Biomol. NMR* 14 (1999) 67.
- [219] A.J. Dingley, J.E. Masse, J. Feigon, S. Grzesiek, *J. Biomol. NMR* 16 (2000) 279.
- [220] P.V. Cornish, D.P. Giedroc, M. Hennig, *J. Biomol. NMR* 35 (2006) 209.
- [221] F. Cordier, S. Grzesiek, *J. Am. Chem. Soc.* 121 (1999) 1601.
- [222] G. Cornilescu, J.S. Hu, A. Bax, *J. Am. Chem. Soc.* 121 (1999) 2949.
- [223] Y.X. Wang, J. Jacob, F. Cordier, P. Wingfield, S.J. Stahl, S. Lee-Huang, D. Torchia, S. Grzesiek, A. Bax, *J. Biomol. NMR* 14 (1999) 181.
- [224] F. Cordier, L. Nisius, A.J. Dingley, S. Grzesiek, *Nat. Protoc.* 3 (2008) 235.
- [225] A.Z. Liu, A. Majumdar, F. Jiang, N. Chernichenko, E. Skripkin, D.J. Patel, *J. Am. Chem. Soc.* 122 (2000) 11226.
- [226] F. Lohr, S.G. Mayhew, H. Ruterjans, *J. Am. Chem. Soc.* 122 (2000) 9289.
- [227] M. Mishima, M. Hatanaka, S. Yokoyama, T. Ikegami, M. Walchli, Y. Ito, M. Shirakawa, *J. Am. Chem. Soc.* 122 (2000) 5883.
- [228] A. Bax, G.W. Vuister, S. Grzesiek, F. Delaglio, A.C. Wang, R. Tschudin, G. Zhu, *Methods Enzymol.* 239 (1994) 79.
- [229] G.W. Vuister, A. Bax, *J. Am. Chem. Soc.* 115 (1993) 7772.
- [230] S.J. Archer, M. Ikura, D.A. Torchia, A. Bax, *J. Magn. Reson.* 95 (1991) 636.
- [231] S. Grzesiek, M. Ikura, G.M. Clore, A.M. Gronenborn, A. Bax, *J. Magn. Reson.* 96 (1992) 215.

- [232] S. Grzesiek, H. Kuboniwa, A.P. Hinck, A. Bax, *J. Am. Chem. Soc.* 117 (1995) 5312.
- [233] S. Grzesiek, G.W. Vuister, A. Bax, *J. Biomol. NMR* 3 (1993) 487.
- [234] G.W. Vuister, A.C. Wang, A. Bax, *J. Am. Chem. Soc.* 115 (1993) 5334.
- [235] R. Konrat, D.R. Muhandiram, N.A. Farrow, L.E. Kay, *J. Biomol. NMR* 9 (1997) 409.
- [236] G. Cornilescu, F. Delaglio, A. Bax, *J. Biomol. NMR* 13 (1999) 289.
- [237] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, *J. Biomol. NMR* 44 (2009) 213.
- [238] J.P. Marino, H. Schwalbe, C. Griesinger, *Acc. Chem. Res.* 32 (1999) 614.
- [239] C. Griesinger, M. Hennig, J.P. Marino, B. Reft, C. Richter, H. Schwalbe, *Biol. Magn. Reson.* 16 (1999) 259.
- [240] A. Bax, G. Kontaxis, N. Tjandra, *Methods Enzymol.* 339 (2001) 127.
- [241] J.R. Tolman, J.M. Flanagan, M.A. Kennedy, J.H. Prestegard, *Proc. Natl. Acad. Sci. USA* 92 (1995) 9279.
- [242] N. Tjandra, A. Bax, *Science* 278 (1997) 1111.
- [243] N. Tjandra, J.G. Omichinski, A.M. Gronenborn, G.M. Clore, A. Bax, *Nat. Struct. Biol.* 4 (1997) 732.
- [244] J.R. Tolman, K. Ruan, *Chem. Rev.* 106 (2006) 1720.
- [245] M.R. Hansen, P. Hanson, A. Pardi, *Methods Enzymol.* 317 (2000) 220.
- [246] M. Zweckstetter, A. Bax, *J. Am. Chem. Soc.* 122 (2000) 3791.
- [247] M.R. Hansen, L. Mueller, A. Pardi, *Nat. Struct. Biol.* 5 (1998) 1065.
- [248] P. Bayer, L. Varani, G. Varani, *J. Biomol. NMR* 14 (1999) 149.
- [249] M.R. Hansen, M. Rance, A. Pardi, *J. Am. Chem. Soc.* 120 (1998) 11210.
- [250] M. Hennig, T. Carlomagno, J.R. Williamson, *J. Am. Chem. Soc.* 123 (2001) 3395.
- [251] L. Zidek, H.H. Wu, J. Feigon, V. Sklenar, *J. Biomol. NMR* 21 (2001) 153.
- [252] P. Padra, R. Stefl, L. Kralik, L. Zidek, V. Sklenar, *J. Biomol. NMR* 24 (2002) 1.
- [253] K. Pervushin, R. Riek, G. Wider, K. Wuthrich, *Proc. Natl. Acad. Sci. USA* 94 (1997) 12366.
- [254] M. Ottiger, F. Delaglio, A. Bax, *J. Magn. Reson.* 131 (1998) 373.
- [255] F. Cordier, A.J. Dingley, S. Grzesiek, *J. Biomol. NMR* 13 (1999) 175.
- [256] J. Weigelt, *J. Am. Chem. Soc.* 120 (1998) 10778.
- [257] P.J. Lukavsky, J.D. Puglisi, *Methods Enzymol.* 394 (2005) 399.
- [258] B. Brutscher, J. Boisbouvier, A. Pardi, D. Marion, J.P. Simorre, *J. Am. Chem. Soc.* 120 (1998) 11845.
- [259] R.S. Lipsitz, N. Tjandra, *Ann. Rev. Biophys. Biomol. Struct.* 33 (2004) 387.
- [260] S.A. McCallum, A. Pardi, *J. Mol. Biol.* 326 (2003) 1037.
- [261] G.M. Clore, A.M. Gronenborn, A. Bax, *J. Magn. Reson.* 133 (1998) 216.
- [262] G.M. Clore, A.M. Gronenborn, N. Tjandra, *J. Magn. Reson.* 131 (1998) 159.
- [263] V. Tsui, L.M. Zhu, T.H. Huang, P.E. Wright, *J. Biomol. NMR* 16 (2000) 9.
- [264] A. Ramos, P. Bayer, G. Varani, *Biopolymers* 52 (2001) 181.
- [265] E.T. Mollova, M.R. Hansen, A. Pardi, *J. Am. Chem. Soc.* 122 (2000) 11561.
- [266] K. Bondensgaard, E.T. Mollova, A. Pardi, *Biochemistry* 41 (2002) 11532.
- [267] B.N.M. van Buuren, J. Schleucher, V. Wittmann, C. Griesinger, H. Schwalbe, S.S. Wijmenga, *Angew. Chem. Intl. Ed.* 43 (2004) 187.
- [268] A. Vermeulen, H.J. Zhou, A. Pardi, *J. Am. Chem. Soc.* 122 (2000) 9638.
- [269] H.J. Zhou, A. Vermeulen, F.M. Jucker, A. Pardi, *Biopolymers* 52 (1999) 168.
- [270] D. Fushman, R. Varadan, M. Assfalg, O. Walker, *Prog. Nucl. Magn. Reson. Spectrosc.* 44 (2004) 189.
- [271] J.L. Battiste, G. Wagner, *Biochemistry* 39 (2000) 5355.
- [272] G.M. Clore, C. Tang, *J. Iwahara, Curr. Opin. Struct. Biol.* 17 (2007) 603.
- [273] T.P. Roosild, J. Greenwald, M. Vega, S. Castronovo, R. Riek, S. Choe, *Science* 307 (2005) 1317.
- [274] X.C. Su, G. Otting, *J. Biomol. NMR* (2009).
- [275] S. Cai, L.Y. Zhu, Z.M. Zhang, Y. Chen, *Biochemistry* 46 (2007) 4943.
- [276] Y.T. Yu, *Methods Enzymol.* 318 (2000) 71.
- [277] P.Z. Qin, K. Hideg, J. Feigon, W.L. Hubbell, *Biochemistry* 42 (2003) 6772.
- [278] G.P.G. Grant, P.Z. Qin, *Nucleic Acids Res.* 35 (2007).
- [279] Q. Cai, A.K. Kusnetzow, W.L. Hubbell, I.S. Haworth, G.P.C. Gacho, N. Van Eps, K. Hideg, E.J. Chambers, P.Z. Qin, *Nucleic Acids Res.* 34 (2006) 4722.
- [280] D. Dertinger, L.S. Behlen, O.C. Uhlenbeck, *Biochemistry* 39 (2000) 55.
- [281] P.Z. Qin, S.E. Butcher, J. Feigon, W.L. Hubbell, *Biochemistry* 40 (2001) 6929.
- [282] T.E. Edwards, T.M. Okonogi, B.H. Robinson, S.T. Sigurdsson, *J. Am. Chem. Soc.* 123 (2001) 1527.
- [283] T.E. Edwards, S.T. Sigurdsson, *Nat. Protoc.* 2 (2007) 1954.
- [284] G.Z. Sowa, P.Z. Qin, *Prog. Nucleic Acid Res. Mol. Biol.* 82 (2008) 147.
- [285] J.C. Hus, D. Marion, M. Blackledge, *J. Mol. Biol.* 298 (2000) 927.
- [286] M.D. Vlasie, R. Fernandez-Busnadiego, M. Prudencio, M. Ubbink, *J. Mol. Biol.* 375 (2008) 1405.
- [287] P.E. Johnson, E. Brun, L.F. MacKenzie, S.G. Withers, L.P. McIntosh, *J. Mol. Biol.* 287 (1999) 609.
- [288] T. Ueda, A. Kato, Y. Ogawa, T. Torizawa, S. Kuramitsu, S. Iwai, H. Terasawa, I. Shimada, *J. Biol. Chem.* 279 (2004) 52574.
- [289] G.M. Clore, J. Iwahara, *Chem. Rev.* 109 (2009) 4108.
- [290] J. Iwahara, C.D. Schwieters, G.M. Clore, *J. Am. Chem. Soc.* 126 (2004) 5879.
- [291] W. Zhang, S.S. Pochapsky, T.C. Pochapsky, N.U. Jain, *J. Mol. Biol.* 384 (2008) 349.
- [292] M.L. Mattinen, K. Paakkonen, T. Ikonen, J. Craven, T. Drakenberg, R. Serimaa, J. Waltho, A. Annala, *Biophys. J.* 83 (2002) 1177.
- [293] A. Grishaev, J. Ying, M.D. Canny, A. Pardi, A. Bax, *J. Biomol. NMR* 42 (2008) 99.
- [294] J. Lipfert, S. Doniach, *Ann. Rev. Biophys. Biomol. Struct.* 36 (2007) 307.
- [295] F. Gabel, B. Simon, M. Nilges, M. Petoukhov, D. Svergun, M. Sattler, *J. Biomol. NMR* 41 (2008) 199.
- [296] X.B. Zuo, J.B. Wang, T.R. Foster, C.D. Schwieters, D.M. Tiede, S.E. Butcher, Y.X. Wang, *J. Am. Chem. Soc.* 130 (2008) 3292.
- [297] L.M.T.R. Lima, Y. Cordeiro, L.W. Tinoco, A.F. Marques, C.L.P. Oliveira, S. Sampath, R. Kodali, G. Choi, D. Foguel, I. Torriani, B. Caughey, J.L. Silva, *Biochemistry* 45 (2006) 9180.
- [298] A. Grishaev, J. Wu, J. Trewella, A. Bax, *J. Am. Chem. Soc.* 127 (2005) 16621.
- [299] M.V. Petoukhov, T.P. Monie, F.H.T. Allain, S. Matthews, S. Curry, D.I. Svergun, *Structure* 14 (2006) 1021.
- [300] K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1986.
- [301] J.P. Linge, M. Habeck, W. Rieping, M. Nilges, *Bioinformatics* 19 (2003) 315.
- [302] W. Rieping, M. Habeck, B. Bardiaux, A. Bernard, T.E. Malliavin, M. Nilges, *Bioinformatics* 23 (2007) 381.
- [303] T. Herrmann, P. Guntert, K. Wüthrich, *J. Biomol. NMR* 24 (2002) 171.
- [304] T. Herrmann, P. Guntert, K. Wüthrich, *J. Mol. Biol.* 319 (2002) 209.
- [305] Y. Shen, A. Bax, *J. Biomol. NMR* 38 (2007) 289.
- [306] A.T. Brünger, *XPLOR Manual Version 3.1*, Yale University Press, New Haven, 1993.
- [307] 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, R.J. Read, L.M. Rice, T. Simonson, G.L. Warren, *Acta Crystallogr. D Biol. Crystallogr.* 54 (1998) 905.
- [308] C.D. Schwieters, J.J. Kuszewski, N. Tjandra, G.M. Clore, *J. Magn. Reson.* 160 (2003) 65.
- [309] C.D. Schwieters, J.J. Kuszewski, G.M. Clore, *Prog. Nucl. Magn. Reson. Spectrosc.* 48 (2006) 47.
- [310] P. Guntert, C. Mumenthaler, K. Wüthrich, *J. Mol. Biol.* 273 (1997) 283.
- [311] E.G. Stein, L.M. Rice, A.T. Brünger, *J. Magn. Reson.* 124 (1997) 154.
- [312] A.D. Mackerell Jr., L. Nilsson, *Curr. Opin. Struct. Biol.* 18 (2008) 194.
- [313] J. Wang, P. Cieplak, P.A. Kollman, *J. Comput. Chem.* 21 (2000) 1049.
- [314] 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, *J. Am. Chem. Soc.* 117 (1995) 5179.
- [315] A.D. Mackerell, D. Bashford, M. Bellott, R.L. Dunbrick, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* 102 (1998) 3586.
- [316] A.D. Mackerell, N. Banavali, N. Foloppe, *Biopolymers* 56 (2000) 257.
- [317] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.S. Ross, T.E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel, P. Kollman, *Comput. Phys. Commun.* 91 (1995) 1.
- [318] D.A. Case, T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, R.J. Woods, *J. Comput. Chem.* 26 (2005) 1668.
- [319] T.E. Cheatham 3rd, M.A. Young, *Biopolymers* 56 (2000) 232.
- [320] O. Guvench, A.D. Mackerell Jr., *Methods Mol. Biol.* 443 (2008) 63.
- [321] J.W. Ponder, D.A. Case, *Protein Simulations* 66 (2003) 27.
- [322] B. Xia, V. Tsui, D.A. Case, H.J. Dyson, P.E. Wright, *J. Biomol. NMR* 22 (2002) 317.
- [323] J.P. Linge, M.A. Williams, C.A.E.M. Spronk, A.M.J.J. Bonvin, M. Nilges, *Proteins* 50 (2003) 496.
- [324] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, *J. Chem. Phys.* 79 (1992) 926.
- [325] A.K. Mazur, *J. Am. Chem. Soc.* 120 (1998) 10928.
- [326] W.L. Jorgensen, J. Tirado-Rives, *J. Am. Chem. Soc.* 110 (1988) 1657.
- [327] V. Tsui, D.A. Case, *J. Am. Chem. Soc.* 122 (2000) 2489.
- [328] J.L. Markley, A. Bax, Y. Arata, C.W. Hilbers, R. Kaptein, B.D. Sykes, P.E. Wright, K. Wüthrich, *Pure Appl. Chem.* 70 (1998) 117.
- [329] S.B. Nabuurs, C.A.E.M. Spronk, G. Vriend, G.W. Vuister, *Concepts Magn. Res. Part A* 22A (2004) 90.
- [330] C. Spronk, S.B. Nabuurs, E. Krieger, G. Vriend, G.W. Vuister, *Prog. Nucl. Magn. Reson. Spectrosc.* 45 (2004) 315.
- [331] C. Oubridge, N. Ito, P.R. Evans, C.H. Teo, K. Nagai, *Nature* 372 (1994) 432.
- [332] D. Lu, M.A. Searles, A. Klug, *Nature* 426 (2003) 96.
- [333] T. Aviv, Z. Lin, G. Ben-Ari, C.A. Smibert, F. Sicheri, *Nat. Struct. Mol. Biol.* 13 (2006) 168.
- [334] H. Mao, S.A. White, J.R. Williamson, *Structure* 6 (1999) 1139.
- [335] N. Ban, P. Nissen, J. Hansen, P.B. Moore, T.A. Steitz, *Science* 289 (2000) 905.
- [336] I. Vidovic, S. Nottrott, K. Hartmuth, R. Luhrmann, R. Ficner, *Mol. Cell* 6 (2000) 1331.
- [337] J.A. Chao, J.R. Williamson, *Structure* 12 (2004) 1165.
- [338] J.M. Ryter, S.C. Schultz, *EMBO J.* 17 (1998) 7505.
- [339] P. Dossat, J.C. Hus, D. Marion, M. Blackledge, *J. Biomol. NMR* 20 (2001) 223.
- [340] H. Valafar, J.H. Prestegard, *J. Magn. Reson.* 167 (2004) 228.
- [341] X.P. Xu, D.A. Case, *J. Biomol. NMR* 21 (2001) 321.
- [342] X.P. Xu, D.A. Case, *Biopolymers* 65 (2002) 408.
- [343] S. Neal, A.M. Nip, H.Y. Zhang, D.S. Wishart, *J. Biomol. NMR* 26 (2003) 215.
- [344] J. Meiler, *J. Biomol. NMR* 26 (2003) 25.
- [345] J.A.M.T.C. Crossigt, C.W. Hilbers, S.S. Wijmenga, *J. Biomol. NMR* 21 (2001) 11.
- [346] R.A. Laskowski, J.A.C. Rullmann, M.W. MacArthur, R. Kaptein, J.M. Thornton, *J. Biomol. NMR* 8 (1996) 477.
- [347] G. Vriend, *J. Mol. Graph.* 8 (1990) 52.
- [348] H.M. Berman, W.K. Olson, D.L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.H. Hsieh, A.R. Srinivasan, B. Schneider, *Biophys. J.* 63 (1992) 751.
- [349] I.W. Davis, L.W. Murray, J.S. Richardson, D.C. Richardson, *Nucleic Acids Res.* 32 (2004) W615.
- [350] N.M. Luscombe, R.A. Laskowski, J.M. Thornton, *Nucleic Acids Res.* 25 (1997) 4940.
- [351] N. Morozova, J. Allers, J. Myers, Y. Shamoo, *Bioinformatics* 22 (2006) 2746.
- [352] P. Bouvet, F.H. Allain, L.D. Finger, T. Dieckmann, J. Feigon, *J. Mol. Biol.* 309 (2001) 763.

- [353] R. Stolarski, *Acta Biochim. Pol.* 50 (2003) 297.
- [354] M. Gaudreault, M.E. Gingras, M. Lessard, S. Leclerc, S.L. Guerin, *Methods Mol. Biol.* 543 (2009) 15.
- [355] P.S. Katsamba, S. Park, I.A. Laird-Offringa, *Methods* 26 (2002) 95.
- [356] N.N. Salim, A.L. Feig, *Methods* 47 (2009) 198.
- [357] D.D. Boehr, R. Nussinov, P.E. Wright, *Nat. Chem. Biol.* 5 (2009) 789.
- [358] A. Mittermaier, L.E. Kay, *Science* 312 (2006) 224.
- [359] Z. Shajani, G. Varani, *Biopolymers* 86 (2007) 348.
- [360] M. Getz, X.Y. Sun, A. Casiano-Negroni, Q. Zhang, H.M. Al-Hashimi, *Biopolymers* 86 (2007) 384.
- [361] G. Lipari, A. Szabo, *J. Am. Chem. Soc.* 104 (1982) 4546.
- [362] G. Lipari, A. Szabo, *J. Am. Chem. Soc.* 104 (1982) 4559.
- [363] D.M. Korzhnev, L.E. Kay, *Acc. Chem. Res.* 41 (2008) 442.
- [364] A.G. Palmer, F. Massi, *Chem. Rev.* 106 (2006) 1700.
- [365] A.G. Palmer, *Chem. Rev.* 104 (2004) 3623.
- [366] J. Meiler, W. Peti, C. Griesinger, *J. Am. Chem. Soc.* 125 (2003) 8072.
- [367] A.T. Frank, A.C. Stelzer, H.M. Al-Hashimi, I. Andricioaei, *Nucleic Acids Res.* 37 (2009) 3670.
- [368] Z. Shajani, G. Varani, *J. Mol. Biol.* 349 (2005) 699.
- [369] F.C. Oberstrass, F.H. Allain, S. Ravindranathan, *J. Am. Chem. Soc.* 130 (2008) 12007.
- [370] A. Mittermaier, L. Varani, D.R. Muhandiram, L.E. Kay, G. Varani, *J. Mol. Biol.* 294 (1999) 967.
- [371] Z. Shajani, G. Drobny, G. Varani, *Biochemistry* 46 (2007) 5875.
- [372] S. Ravindranathan, F.C. Oberstrass, F.H. Allain, *J. Mol. Biol.* 396 (2010) 732.
- [373] P. Deka, P.K. Rajan, J.M. Perez-Canadillas, G. Varani, *J. Mol. Biol.* 347 (2005) 719.
- [374] A.L. Hansen, H.M. Al-Hashimi, *J. Am. Chem. Soc.* 129 (2007) 16072.
- [375] K.T. Dayie, A.S. Brodsky, J.R. Williamson, *J. Mol. Biol.* 317 (2002) 263.
- [376] D.M. Korzhnev, P. Neudecker, A. Mittermaier, V.Y. Orekhov, L.E. Kay, *J. Am. Chem. Soc.* 127 (2005) 15602.
- [377] P. Neudecker, A. Zarrine-Afsar, W.Y. Choy, D.R. Muhandiram, A.R. Davidson, L.E. Kay, *J. Mol. Biol.* 363 (2006) 958.
- [378] M. Tollinger, N.R. Skrynnikov, F.A. Mulder, J.D. Forman-Kay, L.E. Kay, *J. Am. Chem. Soc.* 123 (2001) 11341.
- [379] O. Trott, D. Abergel, A.G. Palmer, *Mol. Phys.* 101 (2003) 753.
- [380] A.L. Hansen, E.N. Nikolova, A. Casiano-Negroni, H.M. Al-Hashimi, *J. Am. Chem. Soc.* 131 (2009) 3818.
- [381] M.H. Bailor, C. Musselman, A.L. Hansen, K. Gulati, D.J. Patel, H.M. Al-Hashimi, *Nat. Protoc.* 2 (2007) 1536.
- [382] Q. Zhang, A.C. Stelzer, C.K. Fisher, H.M. Al-Hashimi, *Nature* 450 (2007) 1263.
- [383] S. Bhattacharya, H. Osman, *J. Infect.* 59 (2009) 377.
- [384] V. D'Souza, M.F. Summers, *Nat. Rev. Microbiol.* 3 (2005) 643.
- [385] M. Mouguel, N. Tounekti, J.L. Darlix, J. Paoletti, B. Ehresmann, C. Ehresmann, *Nucleic Acids Res.* 21 (1993) 4677.
- [386] D.J. Patel, *Curr. Opin. Struct. Biol.* 9 (1999) 74.
- [387] M. Naville, D. Gautheret, *Brief. Funct. Genomic Proteomic* 8 (2009) 482.
- [388] L. Buts, J. Lah, M.H. Dao-Thi, L. Wyns, R. Loris, *Trends Biochem. Sci.* 30 (2005) 672.
- [389] P. Babitzke, C.S. Baker, T. Romeo, *Annu. Rev. Microbiol.* 63 (2009) 27.
- [390] G. Storz, J.A. Opdyke, A. Zhang, *Curr. Opin. Microbiol.* 7 (2004) 140.
- [391] T.A. Cooper, L. Wan, G. Dreyfuss, *Cell* 136 (2009) 777.
- [392] K.E. Lukong, K.W. Chang, E.W. Khandjian, S. Richard, *Trends Genet.* 24 (2008) 416.
- [393] A. Clery, M. Blatter, F.H. Allain, *Curr. Opin. Struct. Biol.* 18 (2008) 290.
- [394] C. Maris, C. Dominguez, F.H. Allain, *FEBS J.* 272 (2005) 2118.
- [395] R. Valverde, L. Edwards, L. Regan, *FEBS J.* 275 (2008) 2712.
- [396] M. Doyle, M.F. Jantsch, *J. Struct. Biol.* 140 (2002) 147.
- [397] T.M. Hall, *Curr. Opin. Struct. Biol.* 15 (2005) 367.
- [398] G. Chanfreau, M. Buckle, A. Jacquier, *Proc. Natl. Acad. Sci. USA* 97 (2000) 3142.
- [399] C. Vernet, K. Artzt, *Trends Genet.* 13 (1997) 479.
- [400] H.A. Lewis, K. Musunuru, K.B. Jensen, C. Edo, H. Chen, R.B. Darnell, S.K. Burley, *Cell* 100 (2000) 323.
- [401] F. Qiao, J.U. Bowie, *Sci. STKE* 2005 (2005) re7.
- [402] T. Aviv, Z. Lin, S. Lau, L.M. Rendl, F. Sicheri, C.A. Smibert, *Nat. Struct. Biol.* 10 (2003) 614.
- [403] A. Lingel, B. Simon, E. Izaurralde, M. Sattler, *Nature* 426 (2003) 465.
- [404] K.S. Yan, S. Yan, A. Farooq, A. Han, L. Zeng, M.M. Zhou, *Nature* 426 (2003) 468.
- [405] J.B. Ma, K. Ye, D.J. Patel, *Nature* 429 (2004) 318.
- [406] S.D. Auweter, F.C. Oberstrass, F.H. Allain, *Nucleic Acids Res.* 34 (2006) 4943.
- [407] A. Serganov, D.J. Patel, *Curr. Opin. Struct. Biol.* 18 (2008) 120.
- [408] H. Liu, W. Zhang, R.B. Reed, W. Liu, P.J. Grabowski, *RNA* 8 (2002) 137.
- [409] M.Y. Chou, J.G. Underwood, J. Nikolic, M.H. Luu, D.L. Black, *Mol. Cell* 5 (2000) 949.
- [410] R. Lamichhane, G.M. Daubner, J. Thomas-Crusells, S.D. Auweter, C. Manatschal, K.S. Austin, O. Valniuk, F.H. Allain, D. Rueda, *Proc. Natl. Acad. Sci. USA* 107 (2010) 4105.
- [411] B. Honoré, H.H. Rasmussen, H. Vorum, K. Dejgaard, X. Liu, P. Gromov, P. Madsen, B. Gesser, N. Tommerup, J.E. Celis, *J. Biol. Chem.* 270 (1995) 28780.
- [412] W.C. Boelens, E.J. Jansen, W.J. van Venrooi, R. Stripecke, I.W. Mattaj, S.I. Gunderson, *Cell* 72 (1993) 881.
- [413] C.W. van Gelder, S.I. Gunderson, E.J. Jansen, W.C. Boelens, M. Polycarpou-Schwarz, I.W. Mattaj, W.J. van Venrooi, *EMBO J.* 12 (1993) 5191.
- [414] L. Chen, A.D. Frankel, *Biochemistry* 33 (1994) 2708.
- [415] D.P. Bartel, M.L. Zapp, M.R. Green, J.W. Szostak, *Cell* 67 (1991) 529.
- [416] L. Giver, D. Bartel, M. Zapp, A. Pawul, M. Green, A.D. Ellington, *Nucleic Acids Res.* 21 (1993) 5509.
- [417] S. Chattopadhyay, J. Garcia-Mena, J. DeVito, K. Wolska, A. Das, *Proc. Natl. Acad. Sci. USA* 92 (1995) 4061.
- [418] K. Sakaguchi, N. Zambrano, E.T. Baldwin, B.A. Shapiro, J.W. Erickson, J.G. Omichinski, G.M. Clore, A.M. Gronenborn, E. Appella, *Proc. Natl. Acad. Sci. USA* 90 (1993) 5219.
- [419] S. Baskerville, M. Zapp, A.D. Ellington, *J. Virol.* 73 (1999) 4962.
- [420] W. Xu, A.D. Ellington, *Proc. Natl. Acad. Sci. USA* 93 (1996) 7475.
- [421] J.A. Berglund, K. Chua, N. Abovich, R. Reed, M. Rosbash, *Cell* 89 (1997) 781.
- [422] F. Houtman, M.R. Diaz-Torres, A. Wright, *Cell* 62 (1990) 1153.
- [423] S. Aymerich, M. Steinmetz, *Proc. Natl. Acad. Sci. USA* 89 (1992) 10410.
- [424] N. Tounekti, M. Mouguel, C. Roy, R. Marquet, J.L. Darlix, J. Paoletti, B. Ehresmann, C. Ehresmann, *J. Mol. Biol.* 223 (1992) 205.
- [425] R.C. Chan, D.L. Black, *Mol. Cell. Biol.* 15 (1995) 6377.
- [426] A.J. Munoz-Gomez, M. Lemonnier, S. Santos-Sierra, A. Berzal-Herranz, R. Diaz-Orejas, *J. Bacteriol.* 187 (2005) 3151.
- [427] M.M. Kessler, M.F. Henry, E. Shen, J. Zhao, S. Gross, P.A. Silver, C.L. Moore, *Genes Dev.* 11 (1997) 2545.
- [428] A. Dahanukar, J.A. Walker, R.P. Wharton, *Mol. Cell* 4 (1999) 209.
- [429] L.S. Neely, B.M. Lee, J. Xu, P.E. Wright, J.M. Gottesfeld, *J. Mol. Biol.* 291 (1999) 549.
- [430] J. Zhou, J.K. McAllen, Y. Taylor, M.F. Summers, *J. Mol. Biol.* 349 (2005) 976.
- [431] C. Valverde, M. Lindell, E.G. Wagner, D. Haas, *J. Biol. Chem.* 279 (2004) 25066.
- [432] A.K. Dubey, C.S. Baker, T. Romeo, P. Babitzke, *RNA* 11 (2005) 1579.
- [433] L.T. Timchenko, J.W. Miller, N.A. Timchenko, D.R. DeVore, K.V. Datar, L. Lin, R. Roberts, C.T. Caskey, M.S. Swanson, *Nucleic Acids Res.* 24 (1996) 4407.
- [434] P. Lugnbuhl, P. Gunter, M. Billeter, K. Wüthrich, *J. Biomol. NMR* 8 (1996) 136.
- [435] M. Nilges, S.I. O'Donoghue, *Prog. Nucl. Magn. Reson. Spectrosc.* 32 (1998) 107.
- [436] J.W. Peng, G. Wagner, *J. Magn. Reson.* 98 (1992) 308.
- [437] L.E. Kay, L.K. Nicholson, F. Delaglio, A. Bax, D.A. Torchia, *J. Magn. Reson.* 97 (1992) 359.
- [438] N.A. Farrow, R. Muhandiram, A.U. Singer, S.M. Pascal, C.M. Kay, G. Gish, S.E. Shoelson, T. Pawson, J.D. Forman-Kay, L.E. Kay, *Biochemistry* 33 (1994) 5984.
- [439] J. Boyd, *J. Magn. Reson. B* 107 (1995) 279.
- [440] M. Buck, J. Boyd, C. Redfield, D.A. Mackenzie, D.J. Jeenes, D.B. Archer, C.M. Dobson, *Biochemistry* 34 (1995) 4041.
- [441] T. Yamazaki, R. Muhandiram, L.E. Kay, *J. Am. Chem. Soc.* 116 (1994) 8266.
- [442] R.B. Hill, C. Bracken, W.F. DeGrado, A.G. Palmer, *J. Am. Chem. Soc.* 122 (2000) 11610.
- [443] L. Zeng, M.W.F. Fischer, E.R.P. Zuiderweg, *J. Biomol. NMR* 7 (1996) 157.
- [444] S.F. Lienin, T. Bremi, B. Brutscher, R. Bruschweiler, R.R. Ernst, *J. Am. Chem. Soc.* 120 (1998) 9870.
- [445] R. Paquin, F. Ferrage, F.A. Mulder, M. Akke, G. Bodenhausen, *J. Am. Chem. Soc.* 130 (2008) 15805.
- [446] T.I. Igumenova, K.K. Frederick, A.J. Wand, *Chem. Rev.* 106 (2006) 1672.
- [447] D. Sheppard, D.W. Li, R. Bruschweiler, V. Tugarinov, *J. Am. Chem. Soc.* 131 (2009) 15853.
- [448] V. Tugarinov, L.E. Kay, *Chembiochem* 6 (2005) 1567.
- [449] O. Millet, D.R. Muhandiram, N.R. Skrynnikov, L.E. Kay, *J. Am. Chem. Soc.* 124 (2002) 6439.
- [450] M. Akke, R. Fiala, F. Jiang, D. Patel, A.G. Palmer, *RNA* 3 (1997) 702.
- [451] J. Boisbouvier, Z. Wu, A. Ono, M. Kainosho, A. Bax, *J. Biomol. NMR* 27 (2003) 133.
- [452] E. Duchardt, H. Schwalbe, *J. Biomol. NMR* 32 (2005) 295.
- [453] P. Vallurupalli, L.E. Kay, *J. Am. Chem. Soc.* 127 (2005) 6893.
- [454] J. Rinnenthal, C. Richter, S. Nozinovic, B. Furtig, J.J. Lopez, C. Glaubitz, H. Schwalbe, *J. Biomol. NMR* 45 (2009) 143.
- [455] F.A.A. Mulder, R.A. de Graaf, R. Kaptein, R. Boelens, *J. Magn. Reson.* 131 (1998) 351.
- [456] A.G. Palmer, M.J. Grey, C. Wang, *Methods Enzymol.* 394 (2005) 430.
- [457] D.M. Korzhnev, N.R. Skrynnikov, O. Millet, D.A. Torchia, L.E. Kay, *J. Am. Chem. Soc.* 124 (2002) 10743.
- [458] D.M. Korzhnev, V.Y. Orekhov, L.E. Kay, *J. Am. Chem. Soc.* 127 (2005) 713.
- [459] F.A. Mulder, N.R. Skrynnikov, B. Hon, F.W. Dahlquist, L.E. Kay, *J. Am. Chem. Soc.* 123 (2001) 967.
- [460] P. Lundstrom, M. Akke, *Chembiochem* 6 (2005) 1685.
- [461] D.M. Korzhnev, T.L. Religa, P. Lundstrom, A.R. Fersht, L.E. Kay, *J. Mol. Biol.* 372 (2007) 497.
- [462] F.A.A. Mulder, M. Akke, *Magn. Res. Chem.* 41 (2003) 853.
- [463] F.A.A. Mulder, B. Hon, A. Mittermaier, F.W. Dahlquist, L.E. Kay, *J. Am. Chem. Soc.* 124 (2002) 1443.
- [464] R. Ishima, P.T. Wingfield, S.J. Stahl, J.D. Kaufman, D.A. Torchia, *J. Am. Chem. Soc.* 120 (1998) 10534.
- [465] R. Ishima, D.A. Torchia, *J. Biomol. NMR* 25 (2003) 243.
- [466] P. Lundstrom, M. Akke, *J. Biomol. NMR* 32 (2005) 163.
- [467] C.G. Hoogstraten, J.R. Wank, A. Pardi, *Biochemistry* 39 (2000) 9951.
- [468] J.E. Johnson Jr., C.G. Hoogstraten, *J. Am. Chem. Soc.* 130 (2008) 16757.
- [469] L.J. Catoire, *J. Biomol. NMR* 28 (2004) 179.
- [470] R. Koradi, M. Billeter, K. Wüthrich, *J. Mol. Graph.* 14 (1996) 51.
- [471] T.C. Leeper, X. Qu, C. Lu, C. Moore, G. Varani, *J. Mol. Biol.* 401 (2010) 334.
- [472] A. Mourão, A. Varrot, C.D. Mackereth, S. Cusack, M. Sattler, *RNA* (2010) 1205.
- [473] A. Davidson, K. Patora-Komisarska, J.A. Robinson, G. Varani, *Nucleic Acids Res.* (2010), doi:10.1093/nar/gkq713.

[474] O. Duss, C. Maris, C. von Schroetter, F.H.-T. Allain, *Nucleic Acids Res.* (2010), doi:10.1093/nar/gkq756.

Glossary of terms as used in this paper

3'UTR: 3' untranslated region

AMBER: assisted model building with energy refinement

ARE: AU-rich element

ARM: arginine-rich motif

ASF/SF2: alternative splicing factor/splicing factor 2

ATNOS: automated NOESY peak picking

CANDID: combined automated NOE assignment and structure determination

CAT: co-antiterminator

CHARMM: chemistry at Harvard macromolecular mechanics

CLIP UV cross-linking and immuno-precipitation assay

CNS: crystallography and NMR system

COSY: correlation spectroscopy

CPMG: Carr-Purcell-Meiboom-Gill

CsrA: carbon storage regulator A

Csf: cleavage stimulation factor

CUG-BP1: CUG binding protein 1

CYANA: combined assignment and dynamics algorithm for NMR applications

DHPC: dihexanoyl-phosphatidylcholine

DMPC: dimyristoyl-phosphatidylcholine

DNA: deoxyribonucleic acid

DQF-COSY: double quantum filter COSY

dsRBD: double-stranded RNA binding domain

dsRNA: double-stranded RNA

DYANA: dynamics algorithm for NMR applications

E. coli: Escherichia coli

EMSA: electrophoretic mobility shift assay

FID: free induction decay

FPLC: fast protein liquid chromatography

FRET: fluorescence resonance energy transfer

GB: generalized-Born

GB1: streptococcal B1 immunoglobulin-binding domain of protein G

Glms ribozyme: glucosamine-6-phosphate activated ribozyme

GMP: guanine monophosphate

GST: glutathione S-transferase

HADDOCK: high ambiguity driven docking

hcnA: gene encoding for hydrogen cyanide synthase subunit A

HDV: hepatitis delta virus

HETCOR: heteronuclear correlation

HIV: human immunodeficiency virus

HMBC: heteronuclear multiple bond correlation

HMQC: heteronuclear multiple quantum correlation

hnRNP: heterogeneous nuclear ribonucleoprotein

HPLC: high performance liquid chromatography

HSQC: heteronuclear single quantum correlation

HTLV: human T-cell leukemia virus

IPAP: in-phase/anti-phase

IPTG: isopropyl β -D-thiogalactopyranoside

ITC: isothermal titration calorimetry

Kd: dissociation constant

kDa: kilodalton

kex: exchange rate

KH: heterogeneous nuclear ribonucleoprotein (hnRNP) K homology

MBP: maltose binding protein

MMDG: metric matrix distance geometry

MoMuLV: Moloney murine leukemia virus

MQ: multiple quantum

mRNA: messenger RNA

MTSL: (1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl)-methanethiosulfonate

MWCO: molecular weight cut-off

NC: nucleocapsid

NMR: nuclear magnetic resonance

NOE: nuclear Overhauser effect

NOESY: nuclear Overhauser effect spectroscopy

nt: nucleotide

NTP: nucleoside triphosphate

OD: optical density

OPLS: optimized potentials for liquid simulations

PAGE: polyacrylamide gel electrophoresis

PAP: poly(A) polymerase

PAZ: Piwi Argonaut and Zwiille

PCR: polymerase chain reaction

PDB: protein data bank

PIE: polyadenylation inhibitory element

PNK: polynucleotide kinase

ppm: part per millions

PRE: paramagnetic relaxation enhancement

Prp24: precursor RNA processing 24

PTB: polypyrimidine tract binding protein

qRRM: quasi RNA recognition motif

RAT: ribonucleic anti-terminator

RBD: RNA binding domain

RBMY: RNA-binding motif gene on Y chromosome

RBP: RNA binding protein

RBS: ribosome binding site

rCSA: residual chemical shift anisotropy

RDC: residual dipolar coupling

Rev: regulation of viral expression

RMSD: root mean square deviation

RNA: ribonucleic acid

RNase: ribonuclease

RRE: Rev response element

RRM: RNA recognition motif

rRNA: ribosomal RNA

RsmE: regulator of secondary metabolism E

RSV: rous sarcoma virus

SA: simulated annealing

SAIL: tereo-array isotope labeling

SAM: sterile alpha motif

SANDER: simulated annealing with NMR-derived energy restraints

SAXS: small angle X-ray scattering

SELEX: systematic evolution of ligands by exponential enrichment

SET: solubility enhancement tag

SF1: splicing factor 1

SIA: scaffold-independent analysis

siRNA: small interfering RNA

SPR: surface plasmon resonance

SRE: smaug recognition element

ssRNA: single-stranded RNA

STAR: signal transduction and activation or RNA

TAD: torsion angle dynamics

TAR: trans-activator response element

Tat: trans-activator

Tev: tobacco etch virus

TFIIIA: transcription factor IIIA

TOCSY: total correlation spectroscopy

tRNA: transfer RNA

TROSY: transverse relaxation optimized spectroscopy

TZF: tandem zinc-finger

UV: ultraviolet

VS: neurospora varkud satellite