NMR Spectroscopy

Structure of a Protein–RNA Complex by Solid-State NMR Spectroscopy

Mumdooh Ahmed, Alexander Marchanka,* and Teresa Carlomagno*

Abstract: *Solid-state NMR (ssNMR) is applicable to high molecular-weight (MW) protein assemblies in a non-amorphous precipitate. The technique yields atomic resolution structural information on both soluble and insoluble particles without limitations of MW or requirement of crystals. Herein, we propose and demonstrate an approach that yields the structure of protein–RNA complexes (RNP) solely from ssNMR data. Instead of using low-sensitivity magnetization transfer steps between heteronuclei of the protein and the RNA, we measure paramagnetic relaxation enhancement effects elicited on the RNA by a paramagnetic tag coupled to the protein. We demonstrate that this data, together with chemical-shift-perturbation data, yields an accurate structure of an RNP complex, starting from the bound structures of its components. The possibility of characterizing protein–RNA interactions by ssNMR may enable applications to large RNP complexes, whose structures are not accessible by other methods.*

Introduction

RNA acts either in isolation or in complex with proteins in a multitude of cellular processes, such as gene expression, and immune and stress response. Our understanding of the regulatory roles that RNA exerts in both prokaryotic and eukaryotic cells grows every year and demands a mechanistic and structural description of new ribonucleoprotein complexes (RNP). X-ray crystallography and electron microscopy are the most popular structural determination methods in the biological community, because of their straightforward implementation and ability to study high molecular-weight (MW) complexes. On the other hand, nuclear magnetic resonance spectroscopy (NMR) is the method of choice to study particles containing large flexible parts as well as transiently forming complexes; in addition, NMR can yield both structural and dynamic information. RNA molecules are flexible and capable of adopting different folds in dependence of both environmental conditions and binding partners; consequently, NMR has played and will continue to play an important role in studying RNA structure and dynamics.[1]

Next to solution-state NMR, solid-state NMR (ssNMR) has developed into a powerful structural biology tool to study insoluble particles, such as membrane proteins or amyloid fibrils.[2] The advantage of ssNMR resides in the fact that the line width, in spite of being intrinsically larger than in solution-state NMR, does not depend on the MW. Furthermore, ssNMR does not require crystallization and is applicable to particles with disordered or flexible regions, such as many RNP complexes. Thus, with high-sensitivity ssNMR instrumentation, the structure of large RNP particles should be in reach.

Despite these advantages, the application of ssNMR to RNA-containing particles remains limited. ssNMR studies of RNA are challenged by the small chemical-shift dispersion of both sugar and base resonances, which is aggravated by the relatively broad ssNMR line width. In the past few years, we have proposed a route to achieve the assignment of RNA resonances as well as to obtain structural restraints for RNA-structure determination by ssNMR.[3] The strategy relies on nucleotide-type selective $^{13}C,^{15}N$-labeling.[4] In more recent work, we have extended ssNMR proton detection at fast magic angle spinning (MAS) rates to RNA and demonstrated the assignment of ribose resonances in a uniformly labeled sample.[5] These experiments, together with those developed by several other laboratories to yield the structure of proteins,[6] permit the structure determination of the RNA and protein parts of an RNP complex by ssNMR. For RNPs containing long RNAs, segmental isotope labeling of the RNA may become necessary to alleviate resonance overlap.

Once the structures of the protein and RNA parts of the RNP complex are available, intermolecular contacts should yield the structure of the complex. The determination of protein–RNA distances by ssNMR is challenged by the considerable overlap of the RNA resonances typically in contact with the protein (backbone phosphorus, C3′ and C4′ as well as minor groove base atoms), the paucity of hydrogen atoms that facilitate transfer of magnetization between heteroatoms, and the relatively long distance between heteroatoms (≥ 3.5 Å), which makes the direct transfer of magnetization weak because of dipolar truncation.[7]

Here, we develop a strategy to determine the structure of RNP complexes by ssNMR starting from the structures of the protein(s) and RNA in the complex. To overcome the

[*] Dr. M. Ahmed, Prof. Dr. A. Marchanka, Prof. Dr. T. Carlomagno
Centre for Biomolecular Drug Research and
Institute of Organic Chemistry, Leibniz University Hannover
Schneiderberg 38, 30167 Hannover (Germany)
E-mail: alexander.marchanka@oci.uni-hannover.de
teresa.carlomagno@oci.uni-hannover.de

Prof. Dr. T. Carlomagno
Group of NMR-based Structural Chemistry,
Helmholtz Centre for Infection Research
Inhoffenstrasse 7, 38124 Braunschweig (Germany)

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:
https://doi.org/10.1002/anie.201915465.

limitations described above, our method does not rely on dipolar-coupling-mediated transfer of magnetization to measure intermolecular distances, but uses chemical-shift perturbation (CSP) and paramagnetic relaxation enhancement (PRE) data. We demonstrate the method by determining the structure of the L7Ae—26mer Box C/D RNA complex from *Pyrococcus furiosus* (*Pf*), which is part of a large enzyme performing 2′-O-methylation of ribose in ribosomal RNA (Figure 1).

PRE effects are used extensively in solution-state NMR to determine distances in the absence of NOEs.[8] In ssNMR, PREs have been first measured on metalloproteins[9] and have been used in pioneering work by the Jaroniec laboratory to determine protein structure.[10] Here, we measure PRE effects elicited on the RNA resonances by paramagnetic tags coupled to L7Ae and CSPs produced in the NCACX and $^{13}C,^{13}C$ DARR spectra of L7Ae by RNA binding. We demonstrate that the combination of the CSP- and PRE-derived restraints guides the docking of individual protein and RNA structures to yield a precise and accurate complex structure. We foresee that this method will contribute substantially to promoting the application of ssNMR to large RNP complexes that cannot be studied in solution.

Results and Discussion

Measurement of PREs in ssNMR

PRE effects are based on the strong interaction between nuclei and unpaired electrons, which have a gyromagnetic ratio approximately 600-times larger than the ^1H nucleus. A detailed description of the theory of nuclear relaxation in the presence of unpaired electrons is beyond the scope of this manuscript; the reader may refer to a recent comprehensive review by Pell et al.[11] In the most simplified, semi-classical description by Solomon,[12] neglecting the Fermi-contact term,[13] the paramagnetic contributions Γ_1^{para} and Γ_2^{para} to the longitudinal and transversal nuclear relaxation rates, respectively, are described as follows:
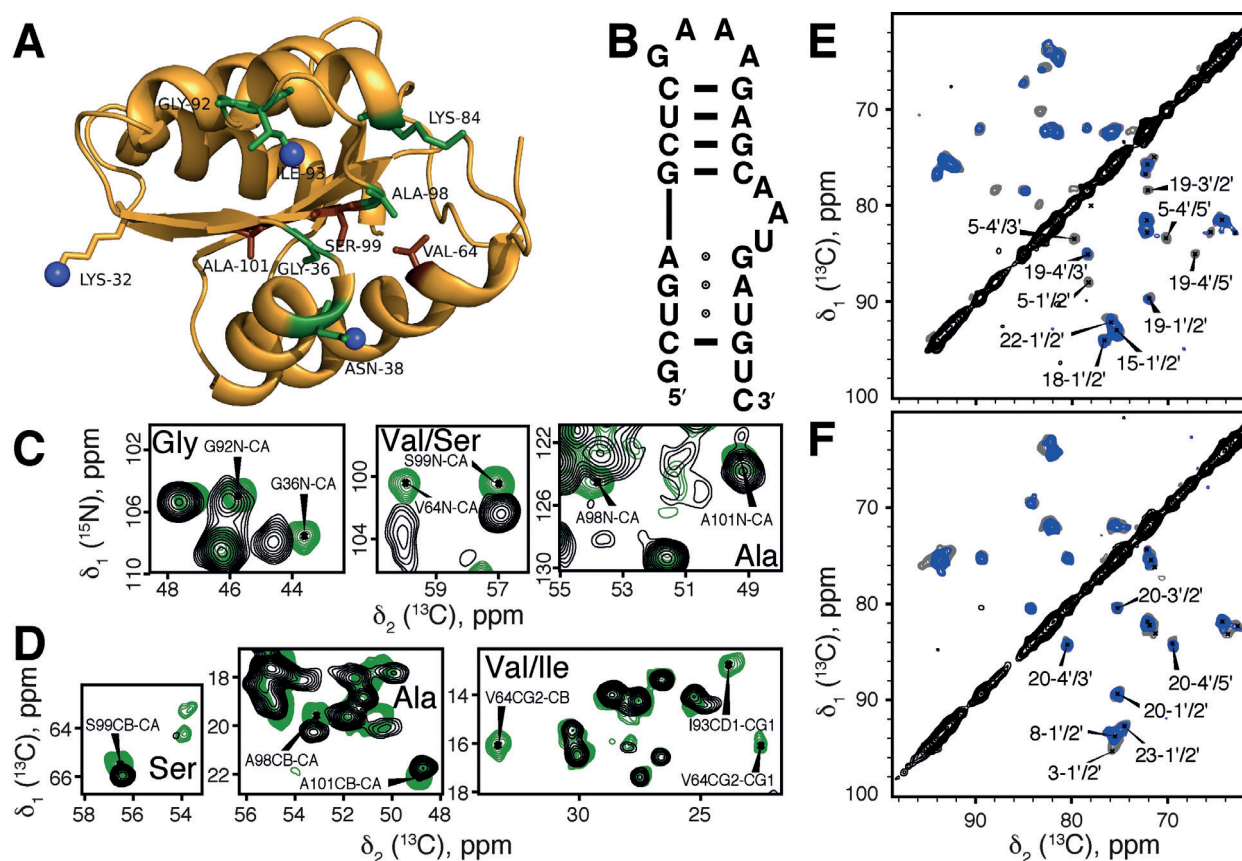
Figure 1. Binding of 26mer RNA to L7Ae. A) Cartoon representation of the L7Ae crystal structure[16] showing the amino acids with strong CSPs (> 0.5 ppm) upon binding to the 26mer Box C/D RNA (green and brown sticks, corresponding to the solvent exposed and buried residues, respectively) as well as the sites selected for spin-labeling (blue spheres). B) Sequence of the RNA used in this study. C) Overlay of 2D NCACX spectra of RNA-bound (green) and unbound (black) L7Ae. The spectra show N-C$^\alpha$ correlations for Gly, Val, Ser, and Ala. The peaks displaying noticeable CSPs upon RNA binding are marked with their assignment. D) Same as in (C) for 2D $^{13}C,^{13}C$ DARR spectra showing the side-chain correlations of Ser, Ala, Val, and Ile. E) Overlay of the 2D $^{13}C,^{13}C$ spectra of the SL-L7Ae-K32C–A^{lab}-26mer RNA complex with the spin-label in the diamagnetic (gray) and paramagnetic (blue) state. F) Same as (E) for the SL-L7Ae-K32C–U^{lab}-26mer RNA complex. Acquisition and processing parameters of all spectra are given in Tables S3 and S4 in the Supporting Information, respectively.

$$\Gamma_1^{para} \approx \frac{2}{15}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_I^2 g_e^2 \mu_B}{r^6}\left(\frac{7\tau_c}{1+\omega_I^2\tau_c^2}+\frac{3\tau_c}{1+\omega_e^2\tau_c^2}\right) \quad (1)$$

$$\Gamma_2^{para} \approx \Gamma_{1\rho}^{para} \approx \frac{1}{15}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\gamma_I^2 g_e^2 \mu_B}{r^6}\left(4\tau_c+\frac{3\tau_c}{1+\omega_I^2\tau_c^2}+\frac{13\tau_c}{1+\omega_e^2\tau_c^2}\right) \quad (2)$$

where r is the electron–nucleus distance, γ_I is the gyromagnetic ratio of the nucleus, g_e is the free electron g-factor, μ_0 is the vacuum permeability, μ_B is the Bohr magneton, τ_c is the electron–nucleus correlation time, ω_I and ω_e are the nucleus and electron frequency, respectively. As both Γ_1^{para} and Γ_2^{para} depend on the electron–nucleus distance, either of the two can be measured to obtain r. In general, Γ_1^{para} and Γ_2^{para} can be converted into long-range intra- or inter-molecular distance restraints, reaching up to approximately 30 Å.

In solution-state NMR and in the absence of a native paramagnetic center, Γ_2^{para} is usually measured instead of Γ_1^{para}, due to its convenient large value in the presence of a nitroxide-based paramagnetic tag, which is easily coupled to single cysteine residues engineered at specific positions. In ssNMR, both the longitudinal relaxation rate Γ_1 and the transversal relaxation rate $\Gamma_{1\rho}$ have been measured in the presence of paramagnetic metals to extract structural restraints;[10a,b,14] in these works the method has been applied to amide ^{15}Ns and carbonyl ^{13}Cs of proteins. ^{15}N Γ_1 can be measured at moderate MAS rates (< 20 kHz); in contrast, the measurement of ^{13}C Γ_1 or ^{15}N $\Gamma_{1\rho}$ requires a better decoupling of the proton bath to avoid the contamination of coherent effects (MAS rates > 40 kHz).

In this work, we want to apply PRE to measure distances between proteins and RNA. In our experimental design, the protein carries a paramagnetic tag and the relaxation rates are estimated for the RNA peaks. Unlike proteins, RNA does not have well-resolved and isolated ^{15}N or ^{13}C resonances. The RNA ^{15}N spins are usually too overlapped to provide residue resolution in 2D experiments (with the exception of pyrimidine N1 and purine N9 in non-canonical structural elements), while the carbons with an acceptable resolution (C1′, C8 of purines, C6 of pyrimidines) carry a directly bound proton, which triggers the requirement of higher MAS rates for a reliable measurement of relaxation rates (> 60 kHz). Thus, measuring either Γ_1 or Γ_2 of RNA sites in ssNMR is technically challenging.

Previous work on proteins focused on the measurement of Γ_1^{para} rates, due to the lower technical requirements (for example, no proton decoupling is needed) and better signal-to-noise ratio.[10] Sizeable Γ_1^{para} rates are elicited by paramagnetic metals bound by an EDTA-like tag coupled to an engineered cysteine residue.[10b] Unfortunately, this strategy is not generally applicable to RNP complexes, as RNA folding depends on the presence of divalent ions and may be incompatible with the presence of EDTA-like tags. In alternative to divalent-ion-binding tags, nitroxide-based tags can be used to induce PREs; however, the Γ_1^{para} elicited by these paramagnetic tags is minuscule, resulting in the necessity to measure either Γ_2^{para} or $\Gamma_{1\rho}^{para}$ to estimate protein–RNA distances in RNP complexes.

In principle, the values of ^{13}C $\Gamma_{1\rho}$ can be measured for the spin-labeled RNP (SL-RNP, that is, the RNP coupled with the paramagnetic tag) in both the paramagnetic and diamagnetic states in series of experiments with variable relaxation delays; the ^{13}C $\Gamma_{1\rho}^{para}$ is then obtained as the difference of the two $\Gamma_{1\rho}$ values. In the case of protein–RNA PREs, there are several technical problems linked to this approach. First, in our previous work[5] we measured bulk coherence lifetimes (T_2') of pyrimidine C6 and purine C8 at different MAS rates and obtained values of 5–6 ms and 11 ms at MAS rates of 40 and 100 kHz, respectively. These short values suggest that the measurement of transversal relaxation rates for these sites with acceptable signal-to-noise ratio would require long measurement times.[10b] Second, all well-resolved RNA carbons are directly attached to a proton and thus tightly coupled with the proton bath. Reliable decoupling from the proton bath requires high MAS rates (> 60 kHz), which in turn is linked to a reduction of the sample quantities (a 0.7 mm rotor is required to reach a MAS rate of 100 kHz versus a 3.2 mm rotor for an MAS rate of 20 kHz). Third, the conversion of PRE effects into distances demands that the PREs are measured for isotope-labeled SL-RNP particles diluted in at least three times as many unlabeled RNP particles, to avoid interparticle PRE effects. Altogether, these restrictions would result in extraordinarily long measurement times (we estimated about 1 month per sample), making the direct measurement of transversal relaxation rates unfeasible.

Here, to avoid the challenging task of measuring $\Gamma_{1\rho}$ rates in low-sensitivity RNA samples, we borrowed the approach used in solution-state NMR[8b] and we sought to quantify PRE effects from two 2D correlation spectra acquired for the SL-RNP in the paramagnetic and diamagnetic states. To estimate the PRE effects elicited on the RNA by a paramagnetic tag coupled to the protein, we acquired 2D ^{13}C-^{13}C homonuclear correlations experiments using the SPC5$_3$ mixing scheme[15] and an RNP containing SL-L7Ae and isotope-labeled RNA (Figure 1 E,F). The ratios of the peaks' volumes in the spectra of the SL-RNP in the paramagnetic and diamagnetic states (V^{para}/V^{dia}) report indirectly on the size of the PREs. However, while in solution-state NMR the V^{para}/V^{dia} ratios can be translated into accurate Γ_2^{para} values,[8b] in ssNMR the conversion of V^{para}/V^{dia} ratios into distances is less straightforward. Here, we develop a semi-quantitative analysis based on an empirical PRE-to-distance converter parameterized on the basis of intramolecular PREs, as explained in the following sections.

Choice of the Paramagnetic Tag Positions

To develop and test the methodology we used the complex of the 26mer Box C/D RNA and the protein L7Ae from *Pyrococcus furiosus* (*Pf*). The details of sample preparation are given in the Supporting Information (Supporting Information, Figure S1). First, we assigned the ^{13}C and ^{15}N resonances of L7Aa in complex with unlabeled 26mer RNA using standard ssNMR protocols[17] (Supporting Information, Table S1 and Figures S2 and S3). The spectra were of good quality, with ^{13}C and ^{15}N line widths down to 0.55 and 1.4 ppm, respectively. All backbone resonances (C$^\alpha$, CO, N) were assigned for 98 of the 121 residues of L7Ae, while the

assignment of the C^β and of most of the side chains was achieved for 60% of the residues. Most of the unassigned resonances belong to amino acids in loop regions or to the long side chains of surface residues; these sites are more prone to polymorphism, due to crystal-packing effects or different microcrystalline forms.

When comparing the NCACX and $^{13}C,^{13}C$ DARR[18] spectra of L7Ae in its RNA-bound and free states, we observe significant CSPs for amino acids G36, N38, V64, K84, G92, I93, A98, S99, and A101 (Figure 1 A,C,D). These residues map to one surface of L7Ae, which is thus identified as the RNA-binding surface.

Ideally, the sites of L7Ae to be coupled with the spin label should be close to the RNA but the paramagnetic tag should not interfere with complex formation. Guided by the CSPs (Figure 1 C,D), we selected sites at the borders of the RNA-interacting surface of L7Ae. For all chosen sites (K32C, N38C, E61C, K84C, and I93C), we compared NMR spectra of wild-type and spin-labeled (SL) L7Ae in the RNP to verify that the cysteine mutation as well as the coupling of the paramagnetic tag did not perturb either protein structure or complex formation. Of the five chosen sites, E61C and K84C were subsequently discarded, because of low accessibility to the reducing agent in the L7Ae–RNA complex and inefficient coupling with the paramagnetic tag (3-(2-Iodoacetamido)-PROXYL), respectively. In the final sample, the isotope-labeled SL-RNPs were diluted with unlabeled RNP in a 1:3 molar ratio, in order to avoid PRE effects generated from interparticle contacts in the microcrystals (see the Supporting Information and Table S2 in the Supporting Information for a list of samples).

Measurement of Protein–RNA PREs

We measured six samples corresponding to SL-L7Ae-K32C–A^{lab}-RNA, SL-L7Ae-K32C–U^{lab}-RNA, SL-L7Ae-N38C–A^{lab}-RNA, SL-L7Ae-N38C–U^{lab}-RNA, SL-L7Ae-I93C–A^{lab}-RNA, and SL-L7Ae-I93C–U^{lab}-RNA complexes (in the A^{lab}- and U^{lab}-RNAs only adenosines and uridines are $^{13}C,^{15}N$-labelled, respectively) packed in a 3.2 mm rotor. The 2D ^{13}C-^{13}C correlation experiments used the $SPC5_3$ mixing scheme at moderate MAS rates. The ^{13}C magnetization was prepared by a 75 µs-long ^1H-^{13}C cross-polarization (CP) period.[19] This short CP time was optimized for the samples carrying the paramagnetic tag, which have short coherence lifetimes, and was used for the measurement of both paramagnetic and diamagnetic states of the spin-labeled complexes. Such a short CP time ensures the selective transfer of magnetization from the proton to the directly attached carbon. The $SPC5_3$ mixing scheme with a short mixing time of 1.3 ms was chosen to restrict polarization transfer to direct neighboring nuclei, and thus avoid the dilution of the PRE effect through space via long-range magnetization transfer. A first analysis of the spectra reveals a sizeable reduction of the peaks' intensities in the spectra of both the SL-L7Ae-K32C–A^{lab}-RNA and SL-L7Ae-K32C–U^{lab}-RNA complexes in the paramagnetic state with respect to the spectra for the same complexes in the diamagnetic state (Figure 1 E,F). This

observation demonstrates that these measurements contain information on protein–RNA distances.

However, the conversion of the V^{para}/V^{dia} ratios to Γ_2^{para} values and then to intermolecular distances is not straightforward. First, our experiments do not measure the PRE effect solely on the two carbons involved in the $SPC5_3$ mixing, as partial coupling to the proton bath is still present. Second, the effect of relaxation during the CP and $SPC5_3$ mixing cannot be described by analytical equations. This effect could be estimated from simulations, for example with the SPINACH package[20] using the built-in extended T_1/T_2 approximation model; however, these simulations are unfeasible for large spin-systems and the outcome depends on the exact geometry of the system. Third, the conversion of V^{para}/V^{dia} ratios to Γ_2^{para} values requires knowledge of the T_1 and/or T_2 values of each measured nucleus in the RNA, and these are in general not available.

Given all the above, we decided to interpret the V^{para}/V^{dia} ratios in a semi-quantitative manner and use them to derive broad distance ranges, rather than specific distances. We reasoned that, if a linear correlation between the V^{para}/V^{dia} ratios and the electron–nucleus distances exists, this correlation should be evident from intramolecular PRE measurements conducted on L7Ae carbon atoms, whose distances from the paramagnetic center are known. From this correlation, we should be able to extract the parameters of the linear function that links the electron–nucleus distances r to the V^{para}/V^{dia} ratios.

$$r = a\frac{V^{para}}{V^{dia}} + b \tag{3}$$

We then further assumed that the coefficient a and b of the linear function optimized for the conversion of intramolecular PREs into distances can be used also for the interpretation of the intermolecular PREs. The validity of this assumption depends on the difference in spin dynamics, as well as in the distribution of ^1H atoms, in proteins and RNA. When designing the approach, we reasoned that the large error bars resulting from the crude linear fitting of intramolecular V^{para}/V^{dia} ratios to interatomic distances would compensate for neglecting these differences. This assumption turned out to be correct, as demonstrated in the following.

Determination of the Correlation Between V^para/V^dia Values and Distances

First, for each of the three spin-labeled complexes containing either L7Ae-K32C, L7Ae-N38C, or L7Ae-I93C, we measured intramolecular V^{para}/V^{dia} ratios on RNP complexes assembled with unlabeled RNA and $^{13}C,^{15}N$-labeled protein (Supporting Information, Table S2). Excerpts from the ^{13}C-^{13}C correlations of the SL-L7Ae-K32C–26mer RNA complex in either the paramagnetic or diamagnetic state are shown in Figure 2 A: the bleaching of several peaks in the spectrum of L7Ae in the complex paramagnetic state is evident. From these spectra we selected all non-overlapped peaks and quantified the experimental V^{para}/V^{dia} (Figure 2 B,
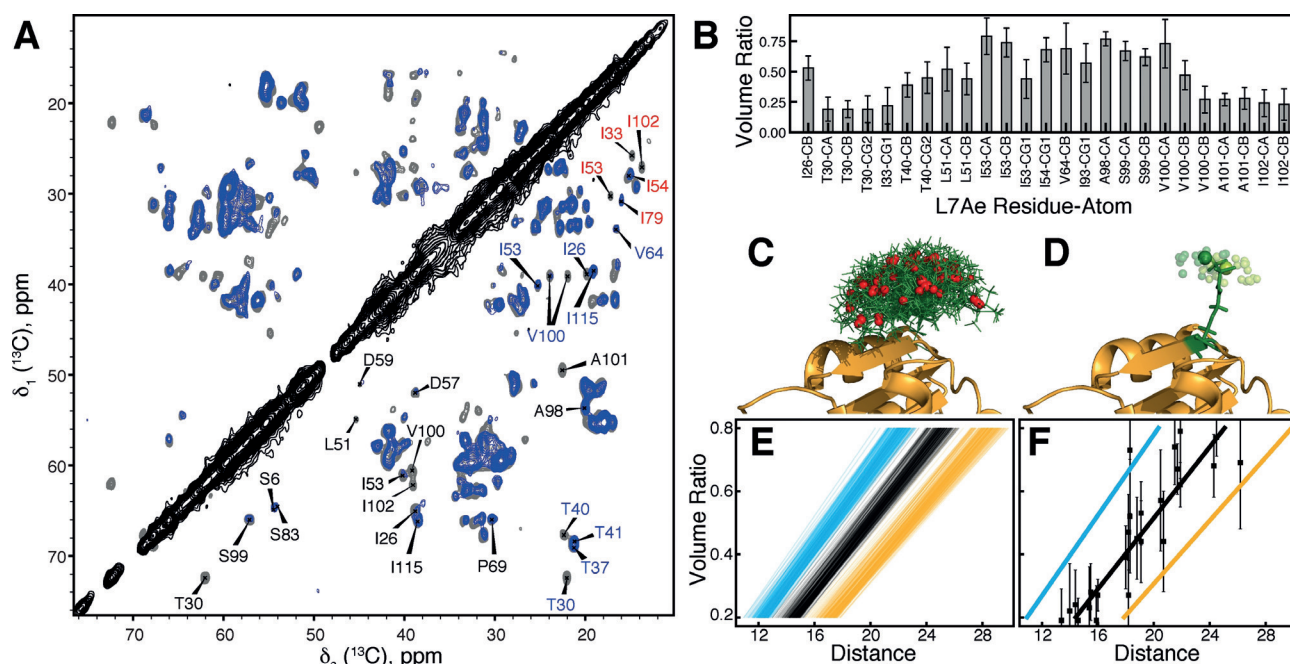
Figure 2. Analysis of intramolecular protein PREs and setup of the PRE-to-distance converter for the SL-L7Ae-K32C–RNA complex. A) Overlay of ^{13}C-^{13}C SPC5$_3$ spectra of the paramagnetic (blue) and diamagnetic (gray) states of the SL-L7Ae-K32C–RNA complex containing ^{13}C,^{15}N-labeled L7Ae. Peak labels are color-coded as follows: black: C$^\alpha$-C$^\beta$; blue: C$^\beta$-C$^\gamma$; red: C$^\gamma$-C$^\delta$. Acquisition and processing details are given in Tables S3 and S4 in the Supporting Information, respectively. B) Intramolecular protein V^{para}/V^{dia} for well-resolved peaks. Only peaks with V^{para}/V^{dia} in the range 0.2–0.8 were considered. C) Representation of the 100 random conformations generated for the paramagnetic tag. The nitroxide nitrogen is marked by a red sphere. D) The cluster of tag conformations that best fit the intramolecular protein PREs. The cluster consists of groups of two conformations (represented in light and dark green). The position of the nitroxide nitrogen atoms in each group is shown in transparent spheres; the cluster averaged positions of the nitroxide nitrogen atoms are shown by solid spheres. E) Individual lines of best fit (LBF, black; LBF^{max}, orange; LBF^{min}, cyan) for each of the groups contained in the cluster of panel (D). LBF^{min} (LBF^{max}) is the fit line with linear parameters a−Δa and b−Δb (a+Δa, b+Δb), where a, b, Δa, and Δb are the parameters of the linear LBF and the corresponding fit errors. F) PRE-to-distance converter. The black line LBF^{clust} is the average of all LBFs of (E). The orange and cyan lines are the LBF^{max} and LBF^{min} of (E) furthest from the LBF^{clust}; these lines are defined as LBF^{max,clust} and LBF^{min,clust}. The points represent the experimental intramolecular V^{para}/V^{dia} ratios (with their experimental errors) versus the distances calculated from each carbon atom to the averaged position of the nitroxide nitrogen atoms of the cluster, as shown in (D). The errors on V^{para} or V^{dia} were defined as the root-mean-square of the noise of the spectrum in an integration area of the same size. The error of V^{para}/V^{dia} was calculated according to the error propagation theory. The L7Ae ssNMR chemical-shift table has been deposited to the Biological Magnetic Resonance Bank (BMRB, accession code 34465).

see the Supporting Information for details on peak selection and volume quantification). To ensure a linear dependence of the V^{para}/V^{dia} ratios on the distances, we considered only V^{para}/V^{dia} values in the range 0.2–0.8.

As demonstrated in solution-state NMR, the conversion of PRE data to electron–nucleus distances benefits from considering the conformational flexibility of the paramagnetic tag: this is done by describing the tag with 2–5 conformations, which represent the conformational space accessible to it.[21] These conformations are found by optimizing the fit of intramolecular PREs to the corresponding electron–nucleus distances. Here, we followed the same strategy and sought for the conformations that best represent the positions accessible to the nitroxide nitrogen of the paramagnetic tag in each of the three protein mutants (SL-L7Ae-K32C, SL-L7Ae-N38C, and SL-L7Ae-I93C). Using the Xplor-NIH[22] script of Nick Anthis, we generate 100 different conformations of the tag (Figure 2C and Supporting Information, Figures S4 and S5) and combined them to yield groups of tag conformations containing either one (100 groups) or two (4950 groups) or three (161 700 groups) conformations. All groups were

clustered as described in the Supporting Information, to generate ensembles of groups with similar tag conformations. For each cluster (Figure 2D), we determined the average positions for the nitroxide nitrogen atoms (one, two or three average positions for the clusters of the groups containing one, two or three tag conformations, respectively). Distances were then calculated between protein carbon atoms and the positions of the tag nitrogen atoms in each group or cluster. Groups with more than one tag conformation were associated with r^{-6} averaged distances between the protein carbon atoms and the nitrogen of the two or three tag conformations present in each group; similarly, clusters were associated with distances between the protein carbon atoms and the averaged nitrogen positions representing each cluster.

For each group, we generated a line of best fit (LBF) of the V^{para}/V^{dia} values versus the measured distances as well as LBF^{min} and LBF^{max}, which are the LBFs obtained after subtraction or addition of the fit errors to the coefficients a and b (Figure 2E). From these linear fits, we calculated the LBF^{clust} as the average of the LBFs of all groups belonging to the cluster, as well as the LBF^{min,clust} and the LBF^{max,clust},

defined as the group LBF^{min} and LBF^{max} furthest from the LBF^{clust}.

After associating each cluster of tag conformations with a LBF^{clust}, a LBF^{min,clust} and a LBF^{max,clust}, the clusters were ranked according to the χ^2 values between the distances back-predicted from the V^{para}/V^{dia} ratios and the distances measured between the carbon atoms and the average nitroxide nitrogen positions in each cluster. The procedure to convert the V^{para}/V^{dia} ratios into distances and obtain their errors is explained in Figure S6 in the Supporting Information. The cluster with the best χ^2 is chosen and its LBF^{clust}, LBF^{min,clust} and LBF^{max,clust} are used as a converter for the intermolecular protein–RNA PREs measured with the same tag. Figure 2 and Figures S4 and S5 in the Supporting Information show the fit of the intramolecular protein PREs for the SL-L7Ae-K32C–26mer RNA, SL-L7Ae-N38C–26mer RNA, and SL-L7Ae-I93C–26mer RNA complexes, respectively. For all three complexes, a cluster of groups of two conformations resulted in the best fit of the intramolecular PRE data with the lowest number of degrees of freedom.

Intermolecular Distances from V^{para}/V^{dia} Ratios

From the 2D $^{13}C,^{13}C$ SPC5$_3$ spectra of the RNP complexes assembled with non-isotope labeled SL-L7Ae and nucleotide-type specific labeled[4] RNA (Figure 1 E,F and Supporting Information, Figure S7), we obtained 72 V^{para}/V^{dia} values (24 for each of the SL-L7Ae-K32C–RNA, SL-L7Ae-N38C–RNA, and SL-L7Ae-I93C–RNA complexes). These values were converted into distance ranges using the PRE-to-distance converter established from the intramolecular PRE data and the procedure shown in Figure S6 in the Supporting Information. The V^{para}/V^{dia} ratios and the corresponding distance ranges are shown in Figure 3 and Table S5 in the Supporting Information. Of the 72 restraints, 64 yielded defined distances with defined error ranges; 3 V^{para}/V^{dia} ratios with lowest limit larger than 0.8 were converted into distances $D_{min}(0.8) < d < 50$ Å (see Figure S6 in the Supporting Information); 5 V^{para}/V^{dia} ratios with upper limit less than 0.2 were converted into distances 2 Å $< d < D_{max}(0.2)$ (Supporting Information, Figure S6).

Calculation of the RNP Structure from ssNMR Data

Next we used the L7Ae CSP data and the PRE-derived protein–RNA distance restraints to calculate the structure of the RNP complex. As starting structure we used the protein-bound structure of the 26mer RNA determined previously (PDB ID 2N0R)[3b] and the crystallographic structure of L7Ae bound to a similar RNA structural element (Box C/D K-turn RNA, PDB ID 3NMU).[16] Because both the protein and the RNA starting structure were determined in the bound form, we did not consider any possible changes in the structure of the individual subunit upon complex formation. The structure of L7Ae could have been determined by ssNMR as well. However, we did not consider it necessary given the existence of several very similar crystallographic
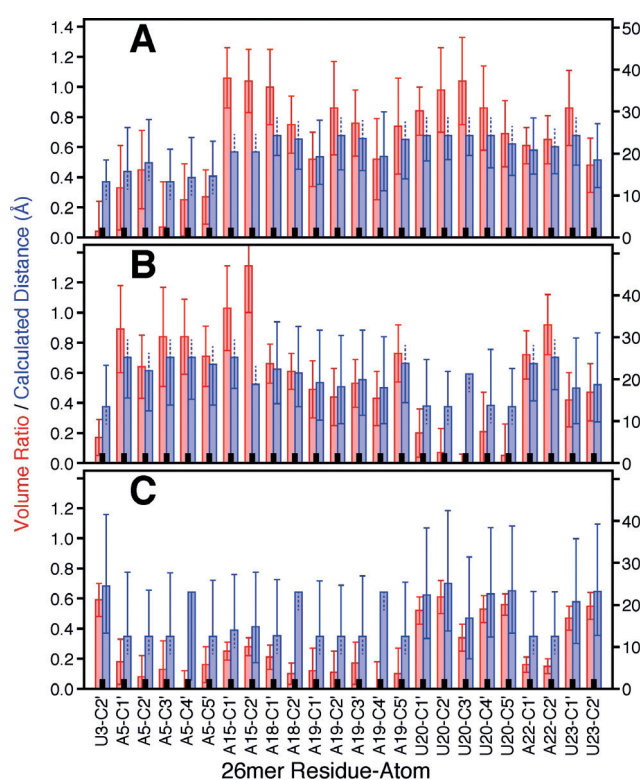
Figure 3. Protein–RNA distance restraints derived from intermolecular PREs. V^{para}/V^{dia} ratios (red bars) measured for the SL-L7Ae-K32C–RNA complex (A), SL-L7Ae-N38C–RNA complex (B), and SL-L7Ae-I93C–RNA complex (C) assembled with either A^{lab}- or U^{lab}-RNA. The errors on the V^{para}/V^{dia} ratios were calculated as explained in the legend to Figure 2. The distances derived from each V^{para}/V^{dia} value are shown as blue bars with their error ranges calculated as shown in Figure S6 in the Supporting Information. Dashed positive/negative error bars indicate an upper/lower limit of 50/2 Å, respectively.

structures of L7Ae in complex with K-turn RNAs of different sequence.

The protein-bound structure of the 26mer RNA (PDB ID 2N0R) and the RNA-bound structure of L7Ae (PDB ID 3NMU) were assembled together using the ssNMR-derived restraints and the program Haddock.[23] The PRE restraints were imposed as r^{-6} averaged distances between RNA carbon atoms and the two dummy atoms representing the cluster fitting best the intramolecular PRE data of each spin-labeled complex (cluster of groups of two tag conformations, Supporting Information, Table S5). L7Ae CSPs were implemented as ambiguous restraints (AIRs), as recommended in Haddock. Out of the nine L7Ae amino acids showing strong CSPs upon RNA binding, six amino acids were found to be solvent accessible (G36, N38, K84, G92, I93, A98) in unbound L7Ae and were used to define the RNA-interacting surface. As partner of this surface, we conservatively defined the entire RNA, as we did not measure CSPs for the RNA upon L7Ae binding. This choice was dictated by the fact that the 26mer RNA does not have a well-defined structure in isolation; thus the CSPs would report on both the folding of the RNA into the structure that binds the protein (K-turn motif) and on the intermolecular interactions. These CSPs could not be used as AIRs in the docking protocol. During the

structure calculation process, the AIRs to G92 were eliminated from the restraints list, as this residue, despite being solvent accessible, is far from the protein surface and its AIRs were constantly violated.

The docking protocol yielded 200 structures separated into three main clusters (Figure 4 A): one cluster was ranked much better than the other two by the standard Haddock score[23] (blue circles, Figure 4 A) and was chosen as the solution. The ten structures of this cluster with the lowest Haddock score were selected and the structure closest to the average was subjected to 15.5 ns of molecular dynamics (MD) refinement in explicit water. The structure closest to the average structure from the last 8 ns of MD is shown in Figure 4 B and is referred to as the ssNMR-derived structure of the L7Ae–26mer RNA complex.

To verify the accuracy of the method, we compared the ssNMR-derived structure with a structure of an orthologous L7Ae–RNA complex from *Archaeoglobus fulgidus* (PDB ID 4BW0).[24] To evaluate the similarity of the position of the RNA with respect to the protein, we aligned the L7Ae of the ssNMR-derived RNP structure with the L7Ae of PDB 4BW0 and calculated the root-mean-square-deviation (RMSD) of the RNA backbone atoms C3′, C4′, C5′, P of nucleotides 2–8 and 15–24, which are present in both structures. The RMSD is only 1.7 Å, indicating that the docking protocol guided by the CSP- and PRE-derived restraints is able to accurately reproduce the RNP structure. The RMSD value of 1.7 Å is particularly low, considering that the RNA of PDB ID 4BW0 differs from our 26mer RNA both in length and sequence. If we restrict the RMSD to the nucleotides that are in direct contact with the protein (nucleotides 4–6 and 17–22), we obtain a value as low as 1.3 Å. Furthermore, a comparison of the intermolecular interactions present in our structure with

those seen in PDB ID 4BW0 confirms that the structure-determination method presented here reproduces accurately the RNA–protein contacts at the intermolecular interface (Supporting Information, Figure S8)

Next, we tested the effect of different restraints sets on the accuracy of the ssNMR-derived RNP structure. Eight additional docking runs were performed excluding either the CSP-derived restraints and/or the PRE restraints derived from individual spin-labeled complexes (Supporting Information, Table S6, and Figures S9 and S10). Both CSP-derived and PRE-derived restraints are necessary to obtain an accurate structure. In the docking run guided by CSP-derived restraints only, the RNA differs by an approximately 180° rotation with respect to the RNA in the crystal structure (PDB ID 4BW0, Supporting Information, Figure S10 B), while in the docking run guided by the PRE-derived restraints only, the RNA is in the correct orientation but too far from the protein (Supporting Information, Figure S10 C). The docking runs performed with the CSP restraints and the PRE restraints derived from one spin-labeled complex yield variable results. The PRE restraints measured with either the SL-L7Ae-N38C–RNA or the SL-L7Ae-I93C–RNA complexes together with the CSPs are not sufficient to identify one best cluster unambiguously (Supporting Information, Figure S9 E,F). In contrast, the PRE restraints derived from the SL-L7Ae-K32C–RNA complex together with the CSP allow the unambiguous identification of the best cluster and yield a similar structure to the one obtained with the full restraint set (Supporting Information, Figures S9 D and S10 D). Similarly, the combination of PRE restraints from the SL-L7Ae-K32C–RNA complex with either the SL-L7Ae-N38C–RNA or the SL-L7Ae-I93C–RNA complex and the CSP-derived restraints yields the correct structure (Supporting Information, Table S6 and Figure S9 G,H). In our experience, the minimum number of spin-labeled complexes necessary to obtain an accurate structure depends on the structure itself. In general, we advise to measure at least three to four spin-labeled samples, with the paramagnetic tags distributed all around the RNA binding surface identified by the CSPs. The first requirement is that the CSP and PRE restraints yield docked structures converging to one low-score cluster. Once such a set of paramagnetic tags has been found, we advise to design and measure one or two additional spin-labeled complexes to validate the accuracy of the calculated structure.
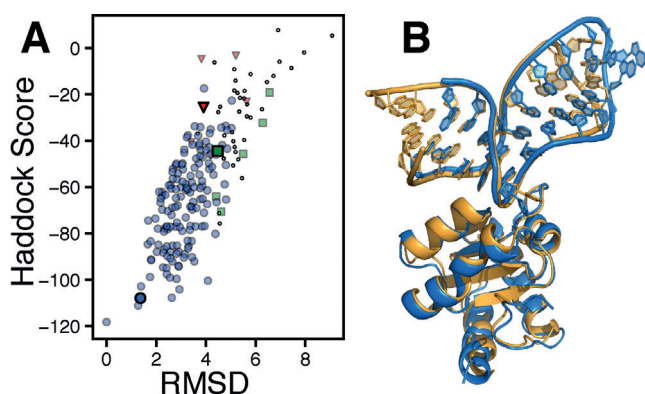
Figure 4. Molecular docking using CSP- and PRE-derived restraints yields an accurate structure of the L7Ae–26mer RNP. A) Plots of Haddock score versus RMSD from the structure with the lowest score of the 200 structures obtained by molecular docking. The RMSD was calculated for all atoms in the RNA apart from hydrogen excluding the loop region (nucleotides 10–13), after aligning the L7Ae molecules in all structures to that of the structure with the lowest score. The structures are separated in three clusters indicated by blue circles, green squares, and red triangles. The best structure of each cluster is shown with a larger symbol. B) Overlay of the MD-refined best docking structure (blue, PDB ID 6TPH) with the crystallographic structure of the orthologous L7Ae–RNA complex from *Af* (gold, PDB ID 4BW0).[24]

Conclusion

We have demonstrated that the structure of RNP complexes is accessible by ssNMR using a combination of CSP- and PRE-derived restraints and the pre-determined structures of the individual components. The PRE data is analyzed in a semi-quantitative manner and the distances derived from it are given generous error bars to compensate for the lack of an accurate treatment of spin relaxation. Despite these approximations, the combination of PRE-derived restraints generated from two spin-labeled complexes and the CSP-derived restraints is sufficient to obtain an accurate structure of the RNP complex. Because ssNMR can

be applied to complexes of any MW, this methodology, in combination with segmental RNA labeling, may give unprecedented access to structural information on specific protein–RNA interactions in large RNP complexes.

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

[1] A. Marchanka, T. Carlomagno, *eMagRes* **2014**, *3*, 110–128.

[2] a) V. S. Mandala, J. K. Williams, M. Hong, *Annu. Rev. Biophys.* **2018**, *47*, 201–222; b) A. Loquet, N. El Mammeri, J. Stanek, M. Berbon, B. Bardiaux, G. Pintacuda, B. Habenstein, *Methods* **2018**, *138–139*, 26–38.

[3] a) A. Marchanka, B. Simon, T. Carlomagno, *Angew. Chem. Int. Ed.* **2013**, *52*, 9996–10001; *Angew. Chem.* **2013**, *125*, 10180–10185; b) A. Marchanka, B. Simon, G. Althoff-Ospelt, T. Carlomagno, *Nat. Commun.* **2015**, *6*, 7024.

[4] A. Marchanka, C. Kreutz, T. Carlomagno, *J. Biomol. NMR* **2018**, *71*, 151–164.

[5] A. Marchanka, J. Stanek, G. Pintacuda, T. Carlomagno, *Chem. Commun.* **2018**, *54*, 8972–8975.

[6] a) D. Mance, M. Weingarth, M. Baldus in *Modern Magnetic Resonance* (Ed.: G. A. Webb), Springer International Publishing, Cham, **2016**, pp. 1–17; b) G. Comellas, C. M. Rienstra, *Annu. Rev. Biophys.* **2013**, *42*, 515–536; c) S. J. Opella, *Biomed. Spectrosc. Imaging* **2014**, *3*, 81–105; d) S. Wang, V. Ladizhansky, *Prog. Nucl. Magn. Reson. Spectrosc.* **2014**, *82*, 1–26.

[7] M. J. Bayro, M. Huber, R. Ramachandran, T. C. Davenport, B. H. Meier, M. Ernst, R. G. Griffin, *J. Chem. Phys.* **2009**, *130*, 114506.

[8] a) A. Lapinaite, B. Simon, L. Skjaerven, M. Rakwalska-Bange, F. Gabel, T. Carlomagno, *Nature* **2013**, *502*, 519–523; b) J. L. Battiste, G. Wagner, *Biochemistry* **2000**, *39*, 5355–5365; c) N. Danilenko, L. Lercher, J. Kirkpatrick, F. Gabel, L. Codutti, T. Carlomagno, *Nat. Commun.* **2019**, *10*, 3435.

[9] a) G. Pintacuda, N. Giraud, R. Pierattelli, A. Bockmann, I. Bertini, L. Emsley, *Angew. Chem. Int. Ed.* **2007**, *46*, 1079–1082; *Angew. Chem.* **2007**, *119*, 1097–1100; b) S. Balayssac, I. Bertini, M. Lelli, C. Luchinat, M. Maletta, *J. Am. Chem. Soc.* **2007**, *129*, 2218–2219; c) S. Balayssac, I. Bertini, K. Falber, M. Fragai, S. Jehle, M. Lelli, C. Luchinat, H. Oschkinat, K. J. Yeo, *ChemBioChem* **2007**, *8*, 486–489.

[10] a) P. S. Nadaud, J. J. Helmus, N. Hofer, C. P. Jaroniec, *J. Am. Chem. Soc.* **2007**, *129*, 7502–7503; b) P. S. Nadaud, J. J. Helmus, S. L. Kall, C. P. Jaroniec, *J. Am. Chem. Soc.* **2009**, *131*, 8108–8120; c) I. Sengupta, P. S. Nadaud, J. J. Helmus, C. D. Schwieters, C. P. Jaroniec, *Nat. Chem.* **2012**, *4*, 410–417.

[11] A. J. Pell, G. Pintacuda, C. P. Grey, *Prog. Nucl. Magn. Reson. Spectrosc.* **2019**, *111*, 1–271.

[12] I. Solomon, *Phys. Rev.* **1955**, *99*, 559–565.

[13] N. Bloembergen, *J. Chem. Phys.* **1957**, *27*, 572–573.

[14] a) M. J. Knight, A. J. Pell, I. Bertini, I. C. Felli, L. Gonnelli, R. Pierattelli, T. Herrmann, L. Emsley, G. Pintacuda, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 11095–11100; b) C. P. Jaroniec, *J. Magn. Reson.* **2015**, *253*, 50–59.

[15] a) M. Hohwy, C. M. Rienstra, C. P. Jaroniec, R. G. Griffin, *J. Chem. Phys.* **1999**, *110*, 7983–7992; b) M. Hohwy, C. M. Rienstra, R. G. Griffin, *J. Chem. Phys.* **2002**, *117*, 4973–4987.

[16] S. Xue, R. Y. Wang, F. P. Yang, R. M. Terns, M. P. Terns, X. X. Zhang, E. S. Maxwel, H. Li, *Mol. Cell* **2010**, *39*, 939–949.

[17] a) A. Schuetz, C. Wasmer, B. Habenstein, R. Verel, J. Greenwald, R. Riek, A. Bockmann, B. H. Meier, *ChemBioChem* **2010**, *11*, 1543–1551; b) L. Shi, V. Ladizhansky in *Intrinsically Disordered Protein Analysis*, *Vol. 1* (Eds.: V. N. Uversky, A. K. Dunker), Humana Press, Totowa, **2012**, pp. 153–165.

[18] K. Takegoshi, S. Nakamura, T. Terao, *Chem. Phys. Lett.* **2001**, *344*, 631–637.

[19] A. Pines, M. G. Gibby, J. S. Waugh, *J. Chem. Phys.* **1973**, *59*, 569–590.

[20] H. J. Hogben, M. Krzystyniak, G. T. P. Charnock, P. J. Hore, I. Kuprov, *J. Magn. Reson.* **2011**, *208*, 179–194.

[21] J. Iwahara, C. D. Schwieters, G. M. Clore, *J. Am. Chem. Soc.* **2004**, *126*, 5879–5896.

[22] C. D. Schwieters, J. J. Kuszewski, N. Tjandra, G. M. Clore, *J. Magn. Reson.* **2003**, *160*, 65–73.

[23] C. Dominguez, R. Boelens, A. M. J. J. Bonvin, *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737.

[24] L. Huang, D. M. Lilley, *RNA* **2013**, *19*, 1703–1710.