J Biomol NMR (2013) 56:17–30
DOI 10.1007/s10858-013-9719-9

ARTICLE

Combining NMR and small angle X-ray and neutron scattering in the structural analysis of a ternary protein-RNA complex

Janosch Hennig · Iren Wang · Miriam Sonntag · Frank Gabel · Michael Sattler

Received: 2 December 2012 / Accepted: 19 February 2013 / Published online: 3 March 2013

Abstract Many processes in the regulation of gene expression and signaling involve the formation of protein complexes involving multi-domain proteins. Individual domains that mediate protein-protein and protein-nucleic acid interactions are typically connected by flexible linkers, which contribute to conformational dynamics and enable the formation of complexes with distinct binding partners. Solution techniques are therefore required for structural analysis and to characterize potential conformational dynamics. Nuclear magnetic resonance spectroscopy (NMR) provides such information but often only sparse data are obtained with increasing molecular weight of the complexes. It is therefore beneficial to combine NMR data with additional structural restraints from complementary solution techniques. Small angle X-ray/neutron scattering (SAXS/SANS) data can be efficiently combined with NMR-derived information, either for validation or by providing additional restraints for structural analysis. Here, we show that the combination of SAXS and SANS data can help to refine structural models obtained from data-driven docking using HADDOCK based on sparse NMR data. The approach is demonstrated with the ternary protein-protein-RNA complex involving two RNA recognition motif (RRM) domains of Sex-lethal, the N-terminal cold shock domain of Upstream-to-N-Ras, and msl-2 mRNA. Based on chemical shift perturbations we have mapped protein-protein and protein-RNA interfaces and complemented this NMR-derived information with SAXS data, as well as SANS measurements on subunit-selectively deuterated samples of the ternary complex. Our results show that, while the use of SAXS data is beneficial, the additional combination with contrast variation in SANS data resolves remaining ambiguities and improves the docking based on chemical shift perturbations of the ternary protein-RNA complex.

Keywords NMR · SANS · SAXS · Protein-RNA complexes · Integrated structural biology

Electronic supplementary material The online version of this article (doi:10.1007/s10858-013-9719-9) contains supplementary material, which is available to authorized users.

J. Hennig · I. Wang · M. Sonntag · M. Sattler (✉)
Institute of Structural Biology, Helmholtz Zentrum München, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany
e-mail: sattler@helmholtz-muenchen.de

J. Hennig · I. Wang · M. Sonntag · M. Sattler
Center for Integrated Protein Science Munich at Chair Biomolecular NMR Spectroscopy, Department Chemie, Technische Universität München, Lichtenbergstr. 4, 85747 Garching, Germany

F. Gabel
Extremophiles and Large Molecular Assemblies Group (ELMA), Institut de Biologie Structurale (IBS) CEA-CNRS-UJF, 41, rue Jules Horowitz, 38027 Grenoble, France

Introduction

Protein-RNA complexes play key roles in transcriptional and post-transcriptional regulation of gene expression (Licatalosi and Darnell 2010; Nilsen and Graveley 2010; van Kouwenhove et al. 2011; Hoskins and Moore 2012). Various distinct processes contribute to the maturation, function and degradation of mRNA after its transcription and involve large protein assemblies that control post-transcriptional regulation (Wahl et al. 2009; Huntzinger and Izaurralde 2011). Structural biology provides insight into the molecular mechanisms underlying these processes. Despite recent successes in the structure determination of protein-RNA complexes (Mackereth and Sattler 2012), the

number of protein-RNA complexes in the protein data bank (PDB) remains scarce (Dominguez et al. 2011). NMR is a powerful method to investigate dynamic protein-RNA interactions, but its application to large protein-RNA assemblies is challenging. In the past decade optimized sample preparation, isotope-labeling and deuteration protocols (Tugarinov et al. 2006) as well as relaxation-optimized pulse sequences (Pervushin et al. 1997; Tugarinov et al. 2003; Tzakos et al. 2006) have been introduced to increase the sensitivity of NMR experiments. Residual dipolar couplings and paramagnetic restraints (Tolman et al. 1995; Tjandra and Bax 1997; Battiste and Wagner 2000; Bax et al. 2001; de Alba and Tjandra 2002; Prestegard et al. 2004; Blackledge 2005; Su et al. 2008; Clore and Iwahara 2009; Madl et al. 2010; Simon et al. 2010) provide long-range structural information to define domain arrangements and interfaces in multi-domain protein complexes. Nevertheless, the NMR data that can be obtained are often sparse and do not allow high resolution structure determination. Therefore, it is important to combine the NMR-derived information with complementary data from other solution techniques.

Within the last decade small angle scattering (SAS) has become an important complementary technique in structural biology due to improvements in sensitivity, instrumentation and access, and its suitability to combine with NMR data (Mattinen et al. 2002; Grishaev et al. 2005; Gabel et al. 2006; Mareuil et al. 2007; Gabel et al. 2008; Wang et al. 2009; Madl et al. 2011a; Takayama et al. 2011; Lange et al. 2012). Here, we explore how the combination of Small Angle X-ray and Neutron Scattering (SAXS/SANS) data with sparse NMR data can resolve ambiguities in the structural analysis of protein-RNA complexes. Our aim is to explore the impact of SAS data in NMR-data driven docking calculations when combined with very limited information from NMR data, i.e. only considering information derived from chemical shift perturbations that report on protein–protein and protein-RNA interfaces. The sparse NMR data are combined with protein-RNA distance restraints that are derived from available crystal structures or homology models of subunits of the complex. Ambiguities in the resulting models can be resolved by SAXS data, which yield information about the overall shape of the protein-RNA complex. SANS combined with contrast matching using subunit-selectively deuterated samples provides valuable additional information between subunits within the complex. The combination of sparse NMR data and SAS is especially advantageous in the structure determination process of large multi-protein-RNA complexes. Efficient structure calculation protocols, using data-driven docking are available (Ubbink et al. 1998; Clore 2000; Fahmy and Wagner 2002; Dominguez et al. 2003; Matsuda et al. 2004; de Vries et al. 2010b; Simon et al. 2010; van Dijk and Bonvin 2010). Here we have combined HADDOCK (Dominguez et al. 2003; de Vries et al. 2010a) for data-driven docking with SAXS and SANS data. HADDOCK docking is driven by experimentally derived knowledge about interfaces within a complex from various sources, e.g. mutagenesis (Dominguez et al. 2004), cross-linking (Vlach et al. 2009), limited proteolysis (Hennig et al. 2005; Hennig et al. 2008; Hennig et al. 2012) or a variety of NMR experiments (e.g. Dominguez et al. 2004; Schreiner et al. 2008). Although data-driven docking with HADDOCK is very useful as solutions are restricted a priori to be in agreement with experimental information it is often difficult to decide which of the resulting structural clusters represents the correct structure. Solvent PREs (Madl et al. 2011b) and SAXS data (Arthur et al. 2011) have been shown to be able to greatly enhance the scoring and validate solutions obtained from HADDOCK.

Here, we show that SANS data can be particularly useful to discriminate between NMR-derived structural models, where SAXS data alone fail to resolve ambiguities. We explore the following flow scheme for combining NMR and SAS data (Fig. 1): In a first step acquisition of sparse NMR data and analysis of input structures is

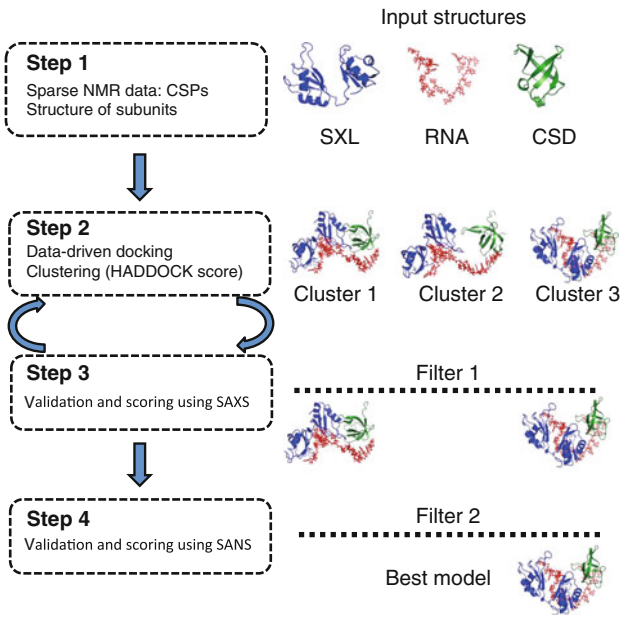


Fig. 1 Illustration of the workflow employed in this study to model a ternary protein–protein-RNA complex consisting of the two RRM motifs RRM1 and RRM2 of Sex-lethal (SXL), the cold shock domain 1 of UNR (CSD) and an 18-mer RNA binding to both proteins. Step 1 involves the NMR data acquisition (e.g. chemical shift perturbations (CSPs)) and retrieval of input structures (shown to the *right*). In step 2 HADDOCK is used to obtain structural complex models, which are grouped into a certain number of clusters. All complex models in each cluster are then validated and scored using SAXS data, which enables clustering refinement and/or removal of clusters which do not fit SAXS data. If ambiguities remain, a fourth step is used and involves further scoring and validation of remaining clusters using SANS data

performed. This is followed by HADDOCK docking and clustering of the output complex structures (step 2). Step 3 and 4 are the iterative validation and scoring of structures based on SAXS and SANS data, respectively, further filtering the output structures to obtain a reasonable structural model.

We test this approach with the ternary protein–protein-RNA complex comprising SXL, UNR and msl-2 mRNA, which plays an essential role in translational repression of *MSL2* in female fruit flies (Gebauer et al. 2003; Grskovic et al. 2003; Beckmann et al. 2005; Duncan et al. 2006; Graindorge et al. 2011). The complex consists of the two RNA recognition motifs (RRMs) of SXL, the first of five cold shock domains of UNR (CSD) shown to bind SXL (Abaza et al. 2006; Abaza and Gebauer 2008) and an 18 nucleotide stretch of msl-2 mRNA which is recognized by SXL and CSD. Our data suggest that the use of SAXS and SANS data improves docking based on chemical shift perturbations of ternary protein–protein RNA complexes.

Materials and methods

Protein expression, purification, and sample preparation

The plasmids pTRX-SXL (corresponding to residues 123–294 of *Drosophila melanogaster* SXL, with an additional N-terminal G-A-M-A left after TEV cleavage) and pTRX-CSD (corresponding to residues 181–252 of *D. melanogaster* UNR with an additional N-terminal G-A-M-A left after TEV cleavage) were transformed into the *Escherichia coli* strain BL21(DE3). Expression was induced with 0.8 mM IPTG for 16 h at 20 °C after an optical density at 600 nm of 1.0 was reached. After harvesting, cell pellets were reconstituted in lysis buffer (50 mM NaP, 300 mM NaCl, pH 8) and sonicated for 5 min (1 s intervals) on ice. After centrifugation for 30 min at 8,500 rpm and 4 °C, the supernatant was filtered and purified on a Ni–NTA column. The column with bound Trx-SXL or Trx-CSD was then washed with 10 column volumes (CV) lysis buffer, 10 CV wash buffer A (50 mM NaP, 300 mM NaCl, 10 mM imidazole, pH 8), and 10 CV wash buffer B (50 mM NaP, 300 mM NaCl, 20 mM imidazole, pH 8), before eluting the proteins with elution buffer (50 mM NaP, 300 mM NaCl, 120 mM imidazole, pH 8). The His_6-thioredoxin solubility tag was then cleaved by TEV protease in presence of β-mercaptoethanol during dialysis over night against wash buffer B. SXL or CSD was then separated from the His_6-thioredoxin tag, using a second Ni–NTA column equilibrated in wash buffer B. The flow through, containing SXL or CSD was further purified on an S75 gelfiltration column (running buffer (10 mM KP, 50 mM NaCl, pH 6) after dialysis overnight against NMR buffer (10 mM KP, 50 mM NaCl, 10 mM DTT, pH 6). SXL or CSD containing fractions were pooled and concentrated to desired concentrations after adding 10 mM DTT and 0.02 % sodium azide. Unlabeled proteins were expressed in LB medium, whereas ^{15}N-, or ^{15}N, ^{13}C labeled proteins for NMR spectroscopy were expressed in M9 minimal medium supplemented with ^{15}N-NH_4Cl (Sigma), or ^{15}N-NH_4Cl and ^{13}C-Glucose (Sigma). Perdeuterated proteins for SANS measurements were expressed in M9 minimal medium supplemented with fully deuterated Glucose (Sigma) in 100 % D_2O (Sigma). In order to obtain the complex, the three components, SXL, CSD, and 18-mer RNA (synthesized, IBA GmbH, Göttingen, Germany) were mixed in a 1:1:1 molar ratio and then subjected to gel filtration to separate possible excess of single components. The complex peak was identified with SDS-PAGE and static light scattering and was confirmed to be a 1:1:1 complex (33.5 kDa). All fractions containing the complex were then concentrated to concentrations desired for NMR and SAXS measurements. For SANS measurements, three different complexes were produced: i) all components protonated, ii) CSD perdeuterated, SXL and RNA protonated, and iii) SXL perdeuterated, CSD and RNA protonated. The latter two were exchanged to NMR buffer in 42 and 70 % D_2O. Oligonucleotides corresponding to U9 and the 18-mer RNA were purchased from Biospring and IBA, respectively.

NMR spectroscopy

The backbone resonance assignments for free SXL and RNA-bound SXL were taken from the Biological Magnetic Resonance Data Bank (BMRB, Accession codes 4029 and 4028, respectively (Ulrich et al. 2008)) and confirmed by 3D HNCACB and CBCACONH experiments (Sattler et al. 1999). Binding of SXL to RNA is in the slow exchange regime on the NMR chemical shift time scale. Therefore, RNA-bound SXL was formed by adding a U9 oligoribonucleotide in a 1:1 ratio. Backbone assignment for CSD was achieved by measuring 3D HNCA, HNCACB, and CBCACONH experiments in the free form and when bound to RNA (Sattler et al. 1999). RNA binding of CSD to the 18-mer RNA was fast at the NMR chemical shift timescale and was monitored by titrating RNA stepwise until saturation and no further chemical shift perturbations (CSPs) were observed. Additional CSPs upon formation of the ternary SXL:CSD:18-mer complex compared to SXL:RNA or CSD:RNA were monitored in ^{1}H,^{15}N-HSQC experiments. Binding kinetics upon ternary complex formation was in the slow exchange regime. Therefore, chemical shifts of both proteins in the ternary complex with RNA were assigned by TROSY HNCA experiments (Salzmann et al. 1999) with deuterium decoupling, measured on samples with the

observed protein ^{2}H,^{15}N,^{13}C-labeled and the binding partner deuterated only. All NMR measurements were carried out at 298 K on Bruker Avance III spectrometers at field strengths corresponding to proton Larmor frequencies of 750 MHz and 800 MHz, equipped with a TXI room temperature probe and TCI cryo-probe head, respectively. Spectra were processed with NMRPipe (Delaglio et al. 1995) and analyzed with CARA (http://cara.nmr.ch) and Sparky (Goddard and Kneller).

Small angle X-ray scattering experiments

30 µl of Sxl-CSD-18-mer complex sample and buffer were measured at 25 °C at the BioSAXS beamline BM29 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France, using a 2D Pilatus detector. Ten frames with 2 s exposure time per frame were recorded for each complex and buffer sample, using an X-ray wavelength of $\lambda = 1.008$ Å. Measurements were performed in flow mode where samples were pushed through the capillary at a constant flow rate to minimize radiation damage. Frames showing radiation damage were removed prior to data analysis.

Small angle neutron scattering experiments

200 µl of all samples (including buffers and water) were measured in Hellma® quartz cells 100QS with 1 mm optical path length at the large dynamic range diffractometer D22 at the Institut Laue-Langevin (ILL Grenoble, France). Scattering data from all samples were recorded at a 2 m/2 m instrumental detector/collimator configuration (centered detector) at a neutron wavelength $\lambda = 6$ Å. H_2O/D_2O buffers, an H_2O water reference sample, the empty beam, an empty quartz cell, as well as a boron sample (electronic background) were measured for data reduction purposes. Exposure times varied between 15 min (boron) and 45 min (empty cell, H_2O and all buffers and samples). Transmissions were measured for 1 min for each sample and the empty beam. All measurements were done at 25 °C.

SANS/SAXS data reduction and integral parameters

The raw SANS data were reduced (accounting for detector efficiency, electronic background, angular averaging) using a standard ILL software package (Gosh et al. 2006). For SAXS data collection and processing, the dedicated beamline software BsxCuBE was used in an automated fashion. The one-dimensional scattering intensities of samples and buffers were expressed as a function of the modulus of the scattering vector $Q = (4\pi/\lambda)\sin\theta$ with 2θ being the scattering angle and λ the neutron/X-ray wavelength. Buffer intensities were subtracted from the respective sample intensities using PRIMUS (Konarev et al. 2003). The radii of gyration R_g of all samples were extracted by the Guinier approximation with the same program. The validity of the Guinier approximation, R_g for $Q < 1.3$, was verified and fulfilled in each case.

Structure calculations

Structure calculations were performed at the HADDOCK webserver. Ambiguous distance restraints based on chemical shift perturbations were used to drive the docking, while unambiguous distance restraints were used to define protein-RNA contacts based on available input structures. As input structures, the crystal structure of SXL bound to transformer RNA was used (PDB: 1B7F, (Handa et al. 1999)) after removing the RNA from the coordinate file and addition of non-polar hydrogens (using MOLMOL (Koradi et al. 1996)). For CSD, a homology model was calculated with MODELLER 9v4 (Sali and Blundell 1993), based on the crystal structure of cold shock protein cspB from *Bacillus subtilis* as template (PDB: 3PF5, (Sachs et al. 2012)). We also considered using a CS-ROSETTA (Lange et al. 2012) model for CSD, as complete backbone chemical shift assignments are available. Although the CS-ROSETTA model adopted the fold typical of a cold shock domain, the backbone RMSD to our NOE-based experimental solution structure (to be published elsewhere) was 3.5 Å, compared to the homology model, which exhibits a backbone RMSD of 1.2 Å. Therefore, the homology model has been used for the complex modeling. For the 18-mer RNA, a model was generated by Maestro (Schrödinger, LLC, New York, NY, 2011), which provides a helical conformation for the RNA. To randomize the RNA conformation during docking calculations all nucleotides were declared as fully flexible during simulated annealing and water refinement. During rigid-body docking, 4000 structures were calculated, and 400 during both simulated annealing and water refinement. For each calculation 50 % of active residues were randomly removed to allow for increased sampling of possible protein–protein orientations. The possibility to remove non-polar hydrogens during structure calculations was not enabled because their presence in the models is crucial for fitting of SAS data. This is especially important for SANS data fitting in order to calculate the scattering density for the natural abundance and perdeuterated proteins within the complex sample correctly. All 400 water-refined structures were analyzed, and cut-off for clustering was 7.5 Å (interface RMSD), with at least four structures per cluster.

Scoring of HADDOCK models against SANS/SAXS data using CRYSON/CRYSOL

We used CRYSON/CRYSOL (Svergun et al. 1995, 1998) to back-calculate and fit the SAXS and SANS curves of the

differently labeled complexes and contrast conditions. Since the absolute values of χ^2 depend on the signal-to-noise level and are not directly comparable quantitatively, we chose a color code to illustrate good fits as green, medium quality fits as orange and bad fits as red in all supplementary tables.

Results

Structural analysis of the ternary SXL, UNR and msl-2 mRNA complex was based on available input structures of the two protein subunits, i.e. the crystal structure of SXL bound to transformer RNA (Handa et al. 1999), and a homology model of CSD based on the cold shock protein cspB from *B. subtilis* (PDB: 3PF5 (Sachs et al. 2012), Fig. 1). The two proteins bind to the RNA individually but do not interact with each other in the absence of RNA. In the presence of the RNA the ternary complex is readily formed with high affinity ($K_D = 20$ nM, data not shown). We performed NMR titrations to monitor chemical shift perturbations (CSPs) upon complex formation (Fig. 2). These CSP data were then used as ambiguous interaction restraints (AIRs) in data-driven docking calculations using HADDOCK. SXL-RNA restraints were derived from the crystal structure (PDB: 1B7F), and CSD-RNA restraints from the homologous crystal structure of cspB (PDB: 3PF5) and defined as unambiguous interaction restraints.

We considered the following combination of experimental data and restraints for the assembly of the ternary complex (Fig. 2e): NMR CSP data of SXL upon CSD and RNA binding, CSP data of CSD upon RNA binding, (- note, that CSPs of CSD upon SXL binding are not considered), unambiguous SXL-RNA and CSD-RNA restraints, as well as SAXS and SANS data for validation and scoring. Based on this data set, referred to as "CSP + interface", active residues for HADDOCK docking were derived from chemical shift perturbation between SXL and CSD, CSD and 18-mer, and SXL and 18-mer (Fig. 3). These data were supplemented with 973 unambiguous distance restraints that define the protein-RNA interface between SXL and the first 10 nucleotides of 18-mer RNA and were derived from the crystal structure of SXL bound to transformer RNA (PDB accession: 1B7F (Handa et al. 1999)). The binding interface was confirmed by titrating U9 RNA to SXL (Fig. 3c). The resulting chemical shift perturbations are virtually identical to previously published data (Lee et al. 1997). CSD-RNA distance restraints (347) were included, which were derived from a homologous cold shock domain bound to a uridine 6-mer (PDB accession: 3PF5, (Sachs et al. 2012). Active residues of CSD towards RNA were also derived from the homology model and confirmed by chemical shift perturbations (Fig. 3a) and used as ambiguous interaction restraints (AIRs) in HADDOCK calculations. The corresponding residues (K193, F200, C204, L210, F211, F212, H213, Y236, D237, and K242) are only active towards the last 6 nucleotides of 18-mer RNA, based on additional NMR titrations comparing CSPs for SXL and CSD (data not shown). The SXL-CSD interface was identified based on chemical shift perturbations, by comparing NMR spectra of RNA-bound SXL, RNA-bound CSD with the corresponding amide chemical shifts in the ternary SXL-CSD-RNA complex (Fig. 3b, d). Residues, which experienced additional chemical shift perturbations in the ternary complex were defined as active residues (N126, L127, T137, D138, L141, Y142, R146, T153, I156, G163, Y164, V171, and N193 on SXL, Fig. 3d). CSD residues identified by CSPs to be affected by SXL binding (L194, R239, I245, and V249 on CSD, Fig. 3b) were not included in structure calculations with the CSP + interface dataset. However, these data were considered in a second data set ("CSP_only", see supplement), where we explore the impact of unambiguous protein-RNA restraints derived from existing structural information of binary interfaces and to what extent this information could be replaced by complete, but ambiguous information derived from CSP data.

The sparse NMR data were used as input for the multibody docking protocol on the HADDOCK webserver (see methods). The resulting 400 structures after water refinement were clustered into 7 clusters (using a cut-off of 7.5 Å); with the largest cluster also showing the lowest HADDOCK score. SAXS data were then used to score the 7 HADDOCK clusters. Supplementary Table 1 shows the χ^2 of CRYSOL fits (see methods) of the lowest energy structure ("best") and of the ten lowest energy structures ("average") to SAXS data. In addition to SAXS data, SANS measurements of subunit-selectively perdeuterated complexes were used to score the HADDOCK output (Fig. 4).

The largest cluster is the cluster with the lowest average HADDOCK score. In the absence of SAXS/SANS data, this cluster would be chosen to represent the best structures. However, the ten lowest energy structures in this cluster exhibit the worst χ^2 values compared to other clusters with respect to the SAXS data (Supplementary Table 1, Fig. 5). Clearly, cluster 1, although exhibiting the lowest HADDOCK score, has to be discarded, as this cluster (as well as clusters 4 and 6) show poor χ^2 values for the SAXS data (Figs. 5, 6a). On the other hand, the SAXS data alone are not providing sufficient information to discriminate and score the remaining clusters as they have similar HADDOCK scores and χ^2 values.

We therefore considered the use of SANS data on subunit-selectively perdeuterated complexes combined with contrast variation (Jacrot 1976; Zaccai and Jacrot 1983; Petoukhov and Svergun 2006; Neylon 2008; Heller 2010;

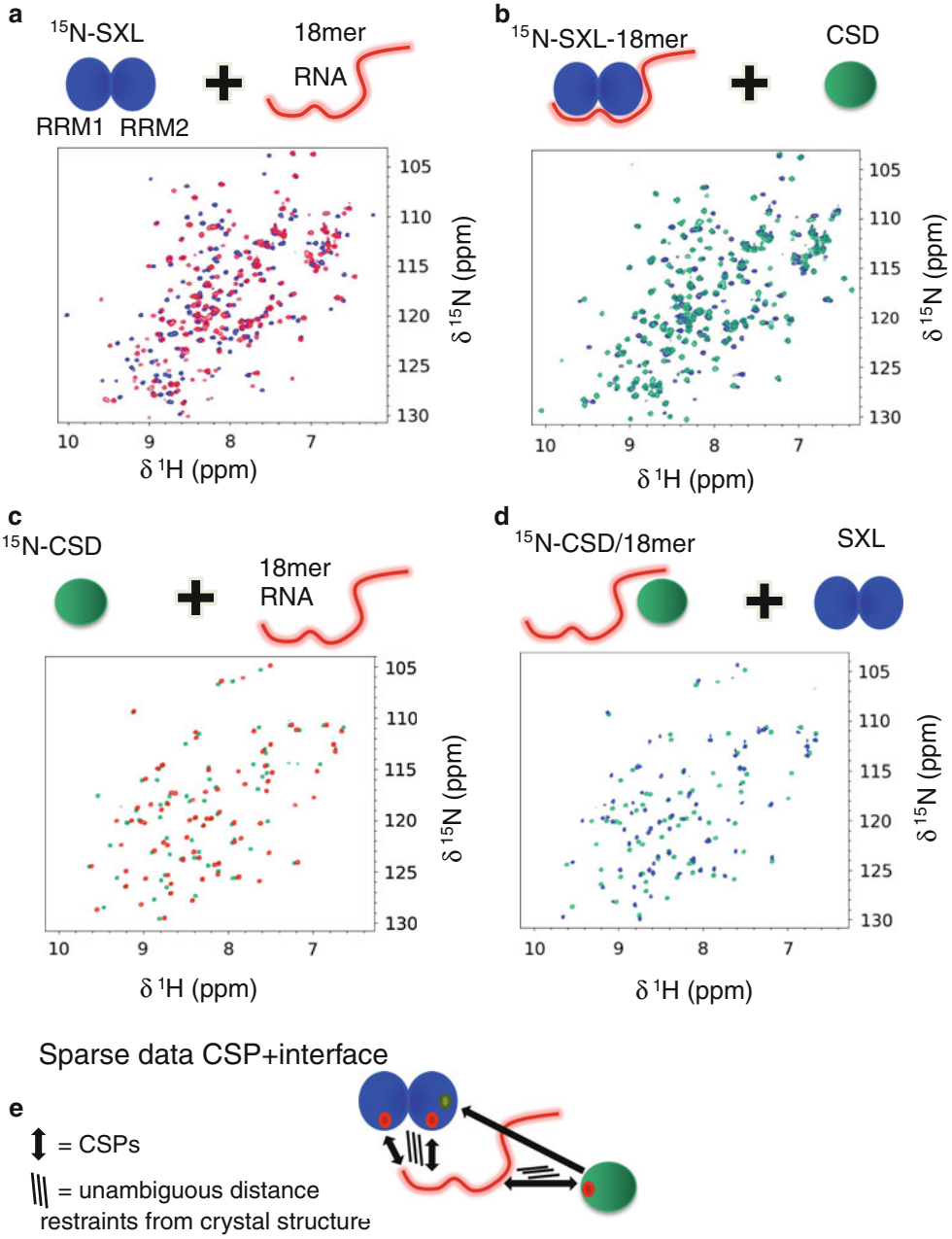


Fig. 2 Overlay of ^{1}H, ^{15}N-HSQC spectra to monitor chemical shift perturbations of residues involved in binding upon titration of 18-mer RNA to SXL **a** CSD to SXL-18-mer RNA **b** 18-mer RNA to CSD **c** and SXL to CSD-18-mer RNA **d**. **e** sparse NMR data considered for the docking calculations are indicated: Chemical shift perturbations of both proteins upon RNA binding, chemical shift perturbations of SXL upon CSD binding, unambiguous distance restraints between SXL and 18-mer RNA (973), and unambiguous distance restraints between CSD and RNA (347)

Jacques and Trewhella 2010; Madl et al. 2011a) to provide additional information for validation and discrimination of the remaining HADDOCK clusters (clusters 2, 3, 5, and 7, Fig. 6b). We prepared samples in 42 and 70 % D_2O buffer with one of the subunits perdeuterated. As can be seen in Fig. 4a, at 42 % D_2O concentration the scattering density of protonated proteins matches the density of water, whereas at 70 % D_2O, the scattering density of the RNA matches the density of water. In addition, perdeuterated proteins can be easily distinguished from their hydrogenated counterparts at both 42 and 70 % D_2O. Therefore, SANS, in contrast to SAXS, can distinguish between subunits of the complex based on the different scattering of perdeuterated and non-deuterated proteins and RNA. SANS data recorded on complex samples with CSD or SXL subunits perdeuterated in 42 % D_2O solution yields information about the relative arrangement of CSD or SXL with respect to the RNA (Fig. 5). At 70 % D_2O, the scattering of the RNA is matched and the SANS data report on the relative protein component arrangement within the complex. Perdeuterated protein has positive contrast while protonated protein gives negative contrast at 70 % D_2O (Fig. 5), yielding inter-subunit restraints between the two protein partners within the ternary complex.

We rescored the ten lowest energy structures of the four best clusters, concerning the χ^2 of CRYSOL fits towards

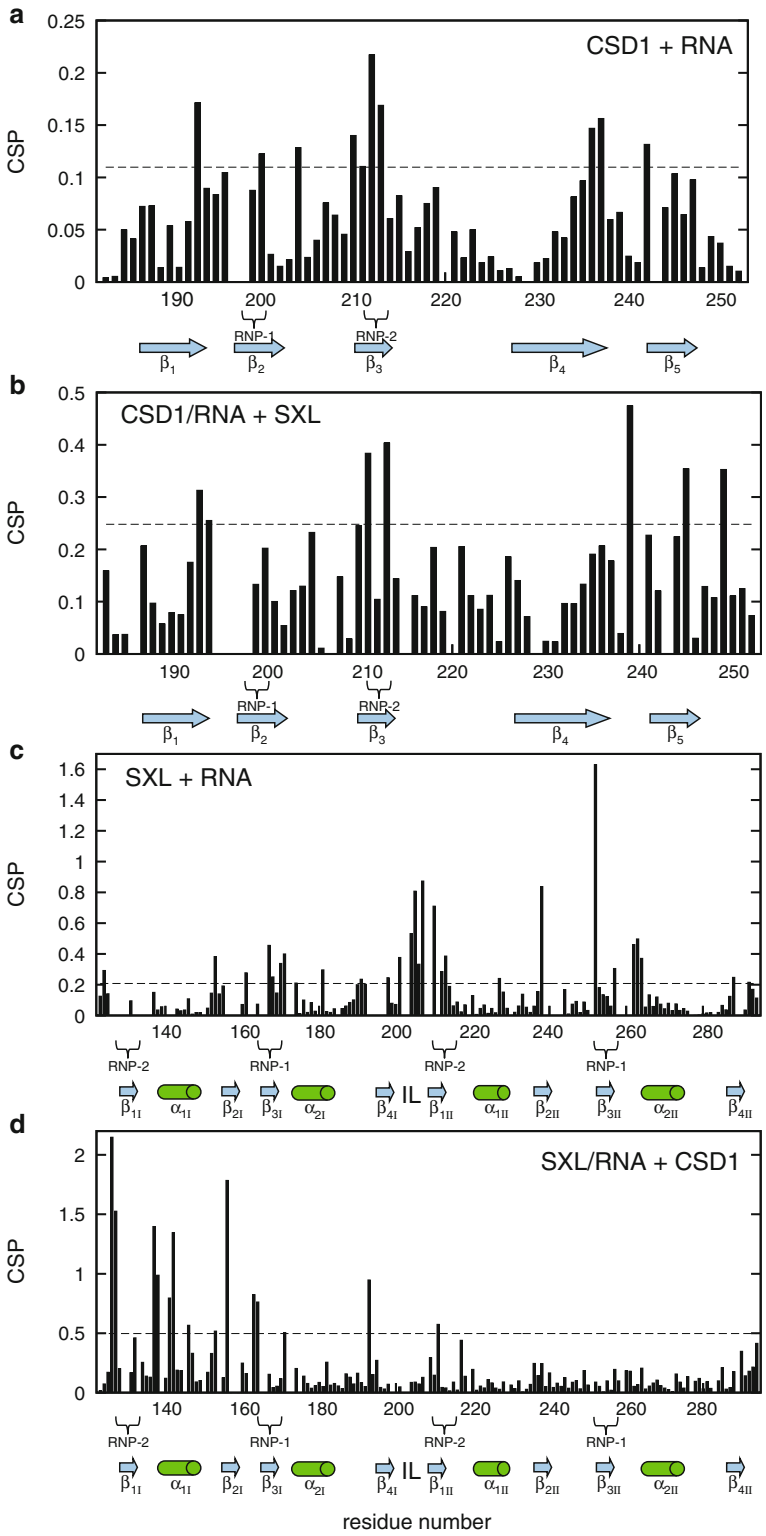


Fig. 3 Chemical shift perturbations (CSPs) of CSD and SXL upon RNA titration and complex formation. **a** CSPs of CSD upon 18-mer RNA titration. **b** CSPs of CSD upon complex formation with 18-mer RNA and SXL. **c** CSPs of SXL upon U9 titration. **d** CSPs of SXL upon complex formation with 18-mer RNA and CSD. The dashed line in each graph resembles the cut-off above which residues are considered to interact with the titrant. The cut-off was calculated as described in Schumann et al. (2007). Below each graph, the secondary structure motifs are shown and the location of RNPs is indicated, as well as the interdomain linker of SXL (IL)

SAXS data (cluster 2, 3, 5, and 7, with the latter three having very similar HADDOCK scores), by comparing their back-calculated SANS data (CRYSON) with the experimental SANS data (Supplementary Table 2, Fig. 5), SANS curves are shown in Fig. 4b and c, where experimental data are fitted with back-calculated curves from the best model from cluster 7. At 42 % D_2O, χ^2 values do not vary significantly and are similar within standard deviations as the large number of unambiguous distance restraints between the protein components and RNA enforces tight conformational ensembles

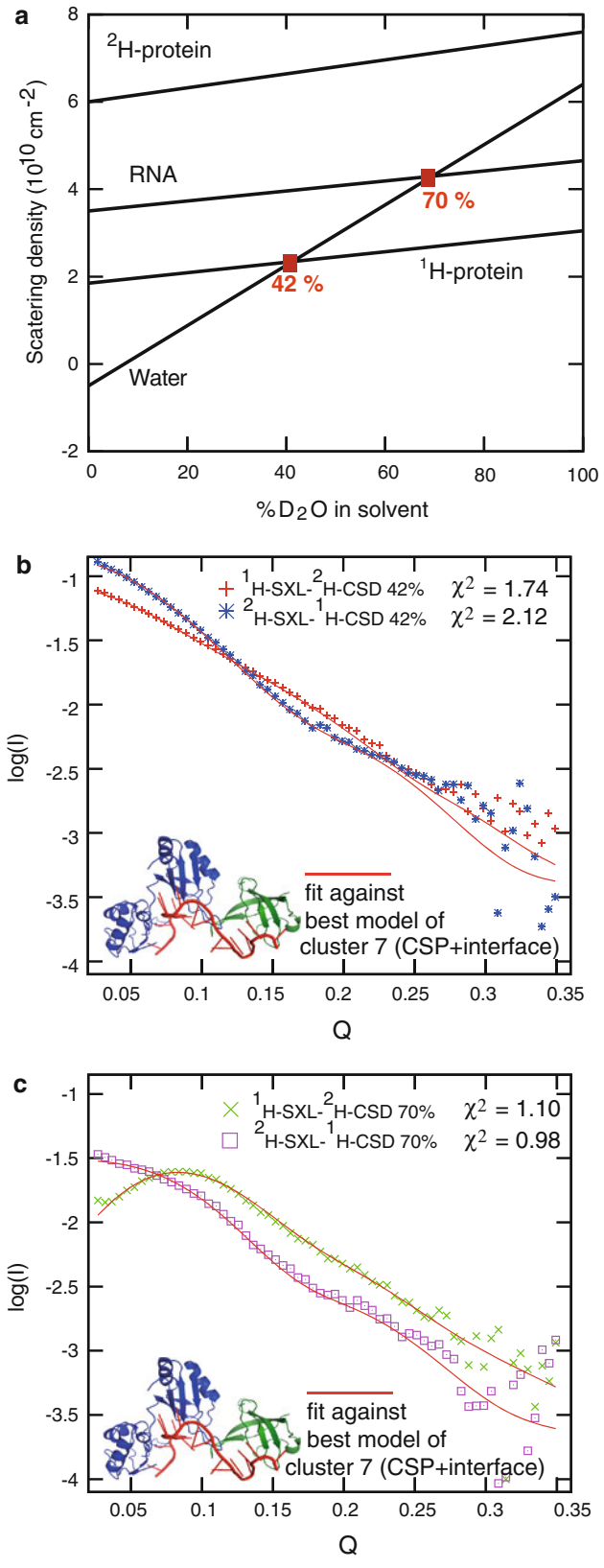


Fig. 4 **a** Scattering density versus % D_2O content in the solvent (according to Jacrot 1976). This graph illustrates how the contrast changes upon increase of D_2O in the solvent for water, protonated protein, RNA, and perdeuterated protein. Contrast match points of protonated protein (42 %), and RNA (70 %) are indicated. **b** Experimental SANS data of differently perdeuterated SXL-CSD-18-mer complexes at 42 % D_2O: protonated SXL/perdeuterated CSD (*red* cross), and perdeuterated SXL/protonated CSD (*blue* star). The red lines are the back-calculated curves from structures (inlaid) with the best SAXS χ^2 from CRYSOL. **c** Experimental SANS data of differently perdeuterated SXL-CSD-18-mer complexes at 70 % D_2O: protonated SXL/perdeuterated CSD (*green* cross), and perdeuterated SXL/protonated CSD (*purple* square). The red lines are the back-calculated curves from structures (inlaid) with the best SAXS χ^2 from CRYSOL. χ^2 values of each fit are indicated within each plot

(Fig. 5b, d). Cluster 3 and cluster 7, however, have slightly better average χ^2 values than 2 and 5 when CSD is perdeuterated (Fig. 5b). In the case of perdeuterated SXL, cluster 2 and cluster 3 show slightly better fits (Fig. 5d). At 70 % D_2O, where RNA is invisible and scattering data report only on the protein components, cluster 2 and 7 outperform cluster 3 and 5 with lower χ^2 values (Fig. 5e). Based on these four SANS data sets, cluster 3 and 5 can be discarded (Fig. 6). Although cluster 3 has slightly better χ^2 values at 42 % D_2O and a good HADDOCK score, it shows much worse fits at 70 % D_2O, which indicates that protein–protein distances and orientation in the model are not consistent with the SANS data. The remaining two clusters (2 and 7, Fig. 6c) have very similar χ^2 towards SAXS and SANS data but exhibit very different HADDOCK scores and other docking statistics. This indicates that only cluster 7 is simultaneously consistent with the NMR, SAXS and SANS data (Table 1). Analysis of structures in cluster 2 indicate that the different HADDOCK scores result from CSP-derived restraint violations. The lowest energy structure of cluster 2 shows non-native RNA contacts, where the RNA is detached from the β-sheets of the N-terminal RRM domain of SXL thus pushing away CSD from SXL (Fig. 6c). The lowest energy structure of cluster 7, on the other hand, has much lower restraint violation energy, with native-like protein-RNA contacts.

These results show that very sparse NMR data, only considering CSPs of binding interfaces, when combined with unambiguous distance restraints derived from available structural information for binary protein-RNA interfaces, and with SAXS and SANS data can resolve ambiguities in HADDOCK models and identify a single ensemble of structures. To explore the impact of the unambiguous interface restraints derived from structures of subcomplexes, we tested another sparse data docking calculation using exclusively CSP-derived restraints, i.e. without considering unambiguous distance restraints to define binary protein-RNA interfaces (CSP_only, Supplementary Information). In these calculations the combination of NMR and SAXS was already sufficient to identify a unique solution (Supplementary, and Supplementary Figs. 1–3, Tables 3–5). Notably, the lowest energy structures obtained from these calculations are not identical to

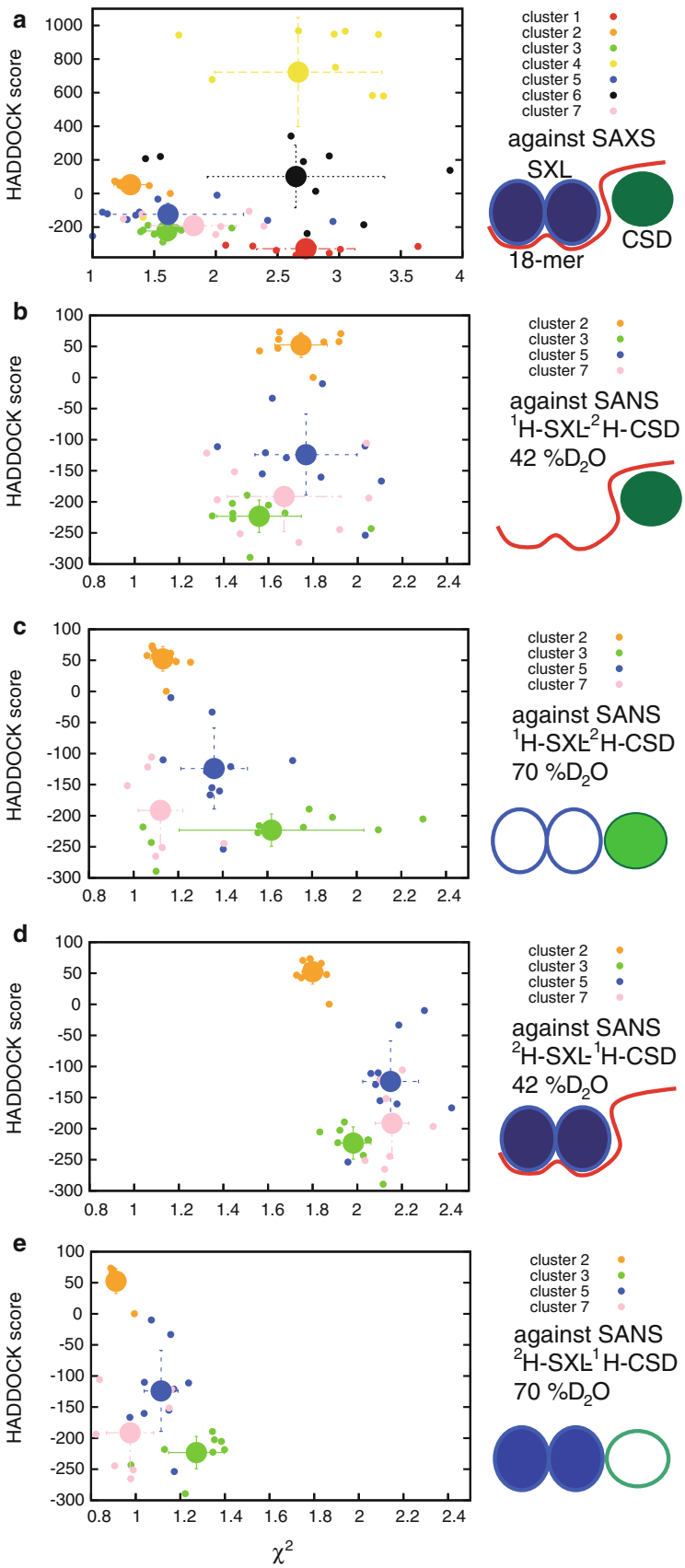
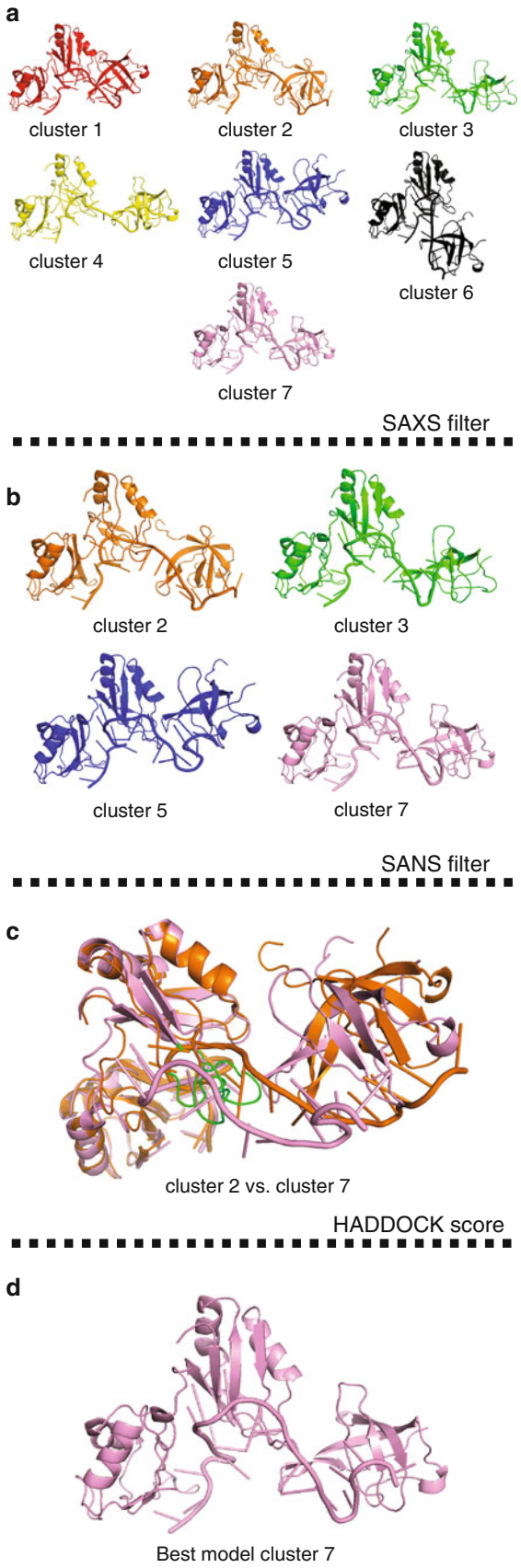


Fig. 5 **a** HADDOCK score of the ten lowest energy structures in clusters consisting of at least ten structures are plotted versus the χ^2 between back-calculated SAXS data from the structures using CRYSOL and experimental SAXS data. **b–e** Corresponding plots, where the structures are fitted against experimental SANS data from samples with different subunit deuteration and D_2O content in the sample buffers (as indicated to the *right* of each plot). The larger filled circles are the average values over the ten lowest energy structures and error bars indicate the standard deviation in HADDOCK score (*vertical*) and χ^2 (*horizontal*). To the right of each plot a schematic explanation about contrast matching of the complexes' components at 42 and 70 % D_2O are shown. SAXS cannot readily distinguish between protein and RNA and yields information about the entire complex, whereas SANS data at 42 % D_2O yield additional information about the conformation of the perdeuterated protein and RNA (red). At 70 % D_2O, information can be obtained about the protein–protein conformation, since the RNA is invisible, whereas the perdeuterated component has a positive contrast (colored *blue* and *green* for SXL and CSD respectively) and the protonated a negative contrast (open, *colored circles*). The contrast differences are indicated by different color shading



◀**Fig. 6** **a** Lowest energy structures of the initial clusters obtained using HADDOCK are shown. **b** SAXS scoring filters out cluster 1, 4, and 6. **c** SANS scoring further filters out clusters 3, and 5. The different HADDOCK scores between clusters 2 and 7 (Table 1) finally discriminates between the two final clusters. **d** The lowest energy structure of cluster 7 is shown, which fulfills best all experimental NMR, SAXS and SANS restraints

the structures obtained based on the CSP + interface data, suggesting that specific interface restraints are important for the structural analysis. This demonstrates that a minimal set of experimental restraints and prior structural knowledge on the subunits is required for unambiguous structural refinement of these complexes.

Discussion

Structural analysis of protein-RNA complexes is challenging and often requires the combination of different structural biology techniques. Flexible linkers and conformational dynamics of these complexes can prevent crystallization. While solution state NMR spectroscopy can be applied to such systems, *de novo* high-resolution structure determination becomes exceedingly difficult with increasing molecular weight (above 20 kDa). However, the structures of individual domains or subunits of protein complexes are often available and less demanding experiments can be conducted to determine the domain arrangements and binding interfaces.

Here, we explored how a very limited set of experimental NMR data, mainly consisting of chemical shift perturbations to locate protein–protein and protein-RNA binding interfaces, can be complemented by small angle scattering data. We took advantage of available structural data for subunits of the complex, which were used to provide distance restraints between protein and RNA and thus enable semi-rigid body docking of the protein-RNA subunits. In the absence of experimental structures of subunits, homology or, preferably, CS-ROSETTA-derived models can be used. HADDOCK was used for structural modeling as it employs data-driven docking. As expected, when considering only a very limited number of experimentally derived restraints HADDOCK yielded several clusters with rather similar scores, which did not allow identifying a unique best structural ensemble. Obviously, the success depends on the number of active residues. If only one residue is identified on each component of the complex as active interactions, the number of different orientations, all of which fulfill the experimental restraints, is rather large. Moreover, erroneous interfaces might be obtained as false positives if conformational changes in regions remote from the site of interaction induce significant chemical shift perturbations.

Table 1 Statistical comparison of cluster 2 and 7 of the HADDOCK run using sparse data set"CSP + interface"

HADDOCK statistics		
Parameter	Cluster 2	Cluster 7
HADDOCK score	34.5	−239.3
Cluster size	61	8
RMSD lowest[a]	6.0 Å	4.2 Å
VdW[b]	−90.2 kcal/mol	−203.1 kcal/mol
Elec[c]	−768.2 kcal/mol	−820 kcal/mol
Violation[d]	2111.2 kcal/mol	645.9 kcal/mol
BSA[e]	5307.7 Å^2	5571.6 Å^2
χ^2 SAXS[f]	1.31 ± 0.13	1.82 ± 0.39
χ^2 SANS (^{1}H-SXL-^{2}H-CSD, 42 %)	1.75 ± 0.12	1.67 ± 0.28
χ^2 SANS (^{1}H-SXL-^{2}H-CSD, 70 %)	1.13 ± 0.06	1.12 ± 0.12
χ^2 SANS (^{2}H-SXL-^{1}H-CSD, 42 %)	1.80 ± 0.05	2.16 ± 0.08
χ^2 SANS (^{2}H-SXL-^{1}H-CSD, 70 %)	0.91 ± 0.03	0.98 ± 0.12

[a] Root mean square deviation of the backbone 10 Å away from interface residues compared with the overall lowest energy structure

[b] Van der Waals contribution to intermolecular energies

[c] Electrostatic contribution to intermolecular energies

[d] Restraints violation energies

[e] Buried surface area

[f] average χ^2 of the ten lowest energy structures

To overcome these limitations we tested the utility of SAXS and SANS data in providing additional scores. We demonstrate that SANS data using contrast matching with subunit-selectively perdeuterated complex samples can improve convergence of HADDOCK derived models. HADDOCK structures obtained with sparse NMR data (CSP + interface), where protein-RNA interfaces were well defined by unambiguous distance restraints but ambiguous information was considered for the protein–protein contacts, yielded a number of clusters, some with similar scores, which did not identify a unique best cluster. Here, the additional SANS data enabled the discrimination between structures, for which SAXS data alone could not resolve ambiguities (Fig. 5). Both the HADDOCK score and the agreement with SAXS data did not discriminate between some of these clusters. SANS data provided the additional information needed to select the best cluster. SANS data containing information about the protein-RNA components (42 % D_2O) agreed equally well with structures of all clusters, which also fulfilled the SAXS restraints (Fig. 5). However, SANS data at 70 % D_2O that provide strong restraints for the respective arrangement of deuterated and non-deuterated protein components enabled the identification of a single cluster.

When only CSP data were considered for the docking calculations (CSP_only, Supplement), SAXS was sufficient to validate the best cluster already chosen by HADDOCK. However, this cluster yielded different structures than those derived from the CSP + interface data. This suggests that the inclusion of unambiguous interface restraints is important to obtain unique and accurate structures from HADDOCK calculations and, not surprisingly, that with too sparse experimental data even the inclusion of SAXS and SANS data cannot compensate for the lack of unambiguous structural restraints. In general, it is always advisable to obtain as much experimental information as possible to avoid ambiguities in structure calculations based on sparse data. Most powerful restraints are obtained from NOE measurements, which however are often difficult to obtain or to assign unambiguously in large protein complexes. Instead, paramagnetic NMR can provide long-range distance and/or orientational restraints from paramagnetic relaxation enhancements (PREs), and pseudo contact shifts (PCS) respectively (Peters et al. 1996; Battiste and Wagner 2000; Bertini et al. 2002; Pintacuda et al. 2007; Clore and Iwahara 2009; Simon et al. 2010; Su and Otting 2010; Mackereth et al. 2011; Madl et al. 2011a; Madl et al. 2011b). Although these methods require the introduction of a paramagnetic center by spin labels or metal binding tags (Battiste and Wagner 2000; Gaponenko et al. 2000; Clore and Iwahara 2009; Simon et al. 2010; Su and Otting 2010) they provide powerful unambiguous structural restraints. RDC data can provide important information about relative domain orientations. For example, the structures obtained with the CSP + interface and CSP_only data show similar position for the center-of-masses of individual domains but different relative domain orientations for the CSD domain. Here, RDC data are expected to greatly improve the docking results when combined with CSP_only data.

In this study, we have deliberately explored the impact of SAS data in the presence of a very limited set of NMR data, i.e. exclusively based on chemical shift perturbations that report on protein–protein and protein-RNA interfaces, as such information should be accessible also for very high molecular weight protein complexes. Clearly, in this context the combination of SAXS and SANS data has a significant impact on identifying a unique structural solution. Nevertheless, with such sparse data only low-resolution structural information is obtained and details of the complex structure need to be defined based on additional experimental data, such as using paramagnetic restraints (vide infra).

It should be noted that SAS data will not provide structural information beyond inter-subunit distances within a complex if the subunits have approximate spherical symmetry (Gabel et al. 2006). In this case, SAS data cannot distinguish between different orientations, as the scattering shape is identical. This is mirrored in the difference of orientations of the CSD towards SXL (although the same space volume is occupied) between results of different data sets used. On the other hand, if various complementary NMR data are available, for example from residual dipolar couplings defining relative domain orientations, inter-subunit NOE data or restraints from paramagnetic data, additional structural information from SAS will provide useful long-range restraints (e.g. between distant subunits or flexible parts) and validate the NMR-derived structures. It should also be considered that NMR and SAS data are rather complementary and report on different conformational features. For example, a potential minor population of open structures where multiple domains connected by flexible linkers do not interact may not be detected by NMR data but would clearly contribute in a SAS experiment. In this case, the combination of NMR and SAS data will provide a complementary view of the structure and dynamics of the protein complex in solution. Similarly, the presence of a small fraction of specific oligomers or unspecific aggregates may not be visible by NMR, but will induce strong scattering in SAS experiments, even to the point that SAS data cannot be analyzed. Although this can affect the utility of SAXS and SANS data sample conditions can be checked and optimized to minimize such effects. Whether the additional effort in acquiring SAXS or SANS data is justified depends on the availability of other data, the molecular weight and complexity of the system studied. Considering, that for large complexes subunit-selectively deuterated samples have to be used for NMR studies in any case, the acquisition of SANS data comes at no additional costs.

In conclusion, we have shown that sparse NMR data derived from chemical shift perturbations can be efficiently combined with SAXS and SANS data in data-driven modeling of a large protein-RNA complex. The use of SANS contrast variations with subunit-selectively samples was able to resolve ambiguities of structural models obtained from NMR-data alone. Nevertheless, it is strongly recommended to acquire as much unambiguous experimental data as possible, including restraints derived from PREs, NOEs, and RDCs to ensure that the resulting structures are accurate and not biased by insufficient experimental input.

Acknowledgments We are grateful to the ILL for BAG SANS beamtime on D22 (local contact Dr. Anne Martel) and the ESRF for beamtime on BM29 (local contact Dr. Louiza Zerrad). We thank Dr. Fatima Gebauer (CRG Barcelona, Spain) for plasmids and helpful discussions. Janosch Hennig acknowledges postdoctoral fellowships by the Swedish Research Council (VR, Vetenskapsrådet) and the European Molecular Biology Organization (EMBO, ALTF 276-2010). This work was supported by the Deutsche Forschungsgemeinschaft (SFB1035, GRK1721 to M.S.) and the European Commission (FP7 NMI3 project, to M.S.). The authors declare that they have no conflict of interest.

References

Abaza I, Gebauer F (2008) Functional domains of Drosophila UNR in translational control. RNA 14(3):482–490. doi:10.1261/rna.802908

Abaza I, Coll O, Patalano S, Gebauer F (2006) Drosophila UNR is required for translational repression of male-specific lethal 2 mRNA during regulation of X-chromosome dosage compensation. Genes Dev 20(3):380–389. doi:10.1101/gad.371906

Arthur DC, Edwards RA, Tsutakawa S, Tainer JA, Frost LS, Glover JN (2011) Mapping interactions between the RNA chaperone FinO and its RNA targets. Nucleic Acids Res 39(10):4450–4463. doi:10.1093/nar/gkr025

Battiste JL, Wagner G (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. Biochemistry 39(18):5355–5365

Bax A, Kontaxis G, Tjandra N (2001) Dipolar couplings in macromolecular structure determination. Methods Enzymol 339: 127–174

Beckmann K, Grskovic M, Gebauer F, Hentze MW (2005) A dual inhibitory mechanism restricts msl-2 mRNA translation for dosage compensation in Drosophila. Cell 122(4):529–540. doi: 10.1016/j.cell.2005.06.011

Bertini I, Luchinat C, Parigi G (2002) Magnetic susceptibility in paramagnetic NMR. Prog NMR Spectrosc 40:249–273

Blackledge M (2005) Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings. Prog NMR Spectrosc 46:23–61

Clore GM (2000) Accurate and rapid docking of protein–protein complexes on the basis of intermolecular nuclear overhauser enhancement data and dipolar couplings by rigid body minimization. Proc Natl Acad Sci USA 97(16):9021–9025

Clore GM, Iwahara J (2009) Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. Chem Rev 109(9):4108–4139. doi:10.1021/cr900033p

de Alba E, Tjandra N (2002) NMR dipolar couplings for the structure determination of biopolymers in solution. Prog Nucl Magn Reson Spectrosc 40(2):175–197

de Vries SJ, Melquiond AS, Kastritis PL, Karaca E, Bordogna A, van Dijk M, Rodrigues JP, Bonvin AM (2010a) Strengths and weaknesses of data-driven docking in critical assessment of prediction of interactions. Proteins 78(15):3242–3249. doi:10.1002/prot.22814

de Vries SJ, van Dijk M, Bonvin AM (2010b) The HADDOCK web server for data-driven biomolecular docking. Nat Protoc 5(5):883–897. doi:10.1038/nprot.2010.32

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

Dominguez C, Boelens R, Bonvin AM (2003) HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. J Am Chem Soc 125(7):1731–1737. doi:10.1021/ja026939x

Dominguez C, Bonvin AM, Winkler GS, van Schaik FM, Timmers HT, Boelens R (2004) Structural model of the UbcH5B/CNOT4

complex revealed by combining NMR, mutagenesis, and docking approaches. Structure 12(4):633–644. doi:10.1016/j.str.2004.03.004

Dominguez C, Schubert M, Duss O, Ravindranathan S, Allain FH (2011) Structure determination and dynamics of protein-RNA complexes by NMR spectroscopy. Prog Nucl Magn Reson Spectrosc 58(1–2):1–61. doi:10.1016/j.pnmrs.2010.10.001

Duncan K, Grskovic M, Strein C, Beckmann K, Niggeweg R, Abaza I, Gebauer F, Wilm M, Hentze MW (2006) Sex-lethal imparts a sex-specific function to UNR by recruiting it to the msl-2 mRNA 3′ UTR: translational repression for dosage compensation. Genes Dev 20(3):368–379. doi:10.1101/gad.371406

Fahmy A, Wagner G (2002) TreeDock: a tool for protein docking based on minimizing van der Waals energies. J Am Chem Soc 124(7):1241–1250

Gabel F, Simon B, Sattler M (2006) A target function for quaternary structural refinement from small angle scattering and NMR orientational restraints. Eur Biophys J 35(4):313–327

Gabel F, Simon B, Nilges M, Petoukhov M, Svergun D, Sattler M (2008) A structure refinement protocol combining NMR residual dipolar couplings and small angle scattering restraints. J Biomol NMR 41(4):199–208. doi:10.1007/s10858-008-9258-y

Gaponenko V, Howarth JW, Columbus L, Gasmi-Seabrook G, Yuan J, Hubbell WL, Rosevear PR (2000) Protein global fold determination using site-directed spin and isotope labeling. Protein Sci 9(2):302–309

Gebauer F, Grskovic M, Hentze MW (2003) Drosophila sex-lethal inhibits the stable association of the 40S ribosomal subunit with msl-2 mRNA. Mol Cell 11(5):1397–1404

Gosh RE, Egelhaaf SU, a.l. e (2006) A computing guide for small-angle scattering. ILL Technical Report ILL06GH05T

Graindorge A, Militti C, Gebauer F (2011) Posttranscriptional control of X-chromosome dosage compensation. Wiley interdisciplinary reviews RNA 2(4):534–545. doi:10.1002/wrna.75

Grishaev A, Wu J, Trewhella J, Bax A (2005) Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data. J Am Chem Soc 127(47):16621–16628. doi:10.1021/ja054342m

Grskovic M, Hentze MW, Gebauer F (2003) A co-repressor assembly nucleated by Sex-lethal in the 3′UTR mediates translational control of Drosophila msl-2 mRNA. The EMBO journal 22(20):5571–5581. doi:10.1093/emboj/cdg539

Handa N, Nureki O, Kurimoto K, Kim I, Sakamoto H, Shimura Y, Muto Y, Yokoyama S (1999) Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. Nature 398(6728):579–585. doi:10.1038/19242

Heller WT (2010) Small-angle neutron scattering and contrast variation: a powerful combination for studying biological structures. Acta Crystallogr D Biol Crystallogr 66(Pt 11):1213–1217. doi:10.1107/S0907444910017658

Hennig J, Ottosson L, Andresen C, Horvath L, Kuchroo VK, Broo K, Wahren-Herlenius M, Sunnerhagen M (2005) Structural organization and Zn2 + -dependent subdomain interactions involving autoantigenic epitopes in the Ring-B-box-coiled-coil (RBCC) region of Ro52. The Journal of biological chemistry 280(39):33250–33261. doi:10.1074/jbc.M503066200

Hennig J, Hennig KD, Sunnerhagen M (2008) MTMDAT: automated analysis and visualization of mass spectrometry data for tertiary and quaternary structure probing of proteins. Bioinformatics 24(10):1310–1312. doi:10.1093/bioinformatics/btn116

Hennig J, de Vries S, Hennig KD, Randles L, Walters KJ, Sunnerhagen M, Bonvin AM (2012) MTMDAT-HADDOCK: high-throughput, protein complex structure modeling based on limited proteolysis and mass spectrometry. BMC Struct Biol 12(1):29. doi:10.1186/1472-6807-12-29

Hoskins AA, Moore MJ (2012) The spliceosome: a flexible, reversible macromolecular machine. Trends Biochem Sci 37(5):179–188. doi:10.1016/j.tibs.2012.02.009

Huntzinger E, Izaurralde E (2011) Gene silencing by microRNAs: contributions of translational repression and mRNA decay. Nat Rev Genet 12(2):99–110. doi:10.1038/nrg2936

Jacques DA, Trewhella J (2010) Small-angle scattering for structural biology–expanding the frontier while avoiding the pitfalls. Protein Sci 19(4):642–657. doi:10.1002/pro.351

Jacrot B (1976) The study of biological structures by neutron scattering from solution. Rep Prog Phys 39:911–953

Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) PRIMUS - a Windows-PC based system for small-angle scattering data analysis. J Appl Cryst 36:1277–1282

Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. Journal of molecular graphics 14 (1):51-55, 29-32

Lange OF, Rossi P, Sgourakis NG, Song Y, Lee HW, Aramini JM, Ertekin A, Xiao R, Acton TB, Montelione GT, Baker D (2012) Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples. Proc Natl Acad Sci USA 109(27):10873–10878. doi:10.1073/pnas.1203013109

Lee AL, Volkman BF, Robertson SA, Rudner DZ, Barbash DA, Cline TW, Kanaar R, Rio DC, Wemmer DE (1997) Chemical shift mapping of the RNA-binding interface of the multiple-RBD protein sex-lethal. Biochemistry 36(47):14306–14317. doi:10.1021/bi970830y

Licatalosi DD, Darnell RB (2010) RNA processing and its regulation: global insights into biological networks. Nat Rev Genet 11(1):75–87. doi:10.1038/nrg2673

Mackereth CD, Sattler M (2012) Dynamics in multi-domain protein recognition of RNA. Curr Opin Struct Biol 22(3):287–296. doi:10.1016/j.sbi.2012.03.013

Mackereth CD, Madl T, Bonnal S, Simon B, Zanier K, Gasch A, Rybin V, Valcarcel J, Sattler M (2011) Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. Nature 475(7356):408–411. doi:10.1038/nature10171

Madl T, Felli IC, Bertini I, Sattler M (2010) Structural analysis of protein interfaces from 13C direct-detected paramagnetic relaxation enhancements. J Am Chem Soc 132(21):7285–7287. doi:10.1021/ja1014508

Madl T, Gabel F, Sattler M (2011a) NMR and small-angle scattering-based structural analysis of protein complexes in solution. J Struct Biol 173(3):472–482. doi:10.1016/j.jsb.2010.11.004

Madl T, Guttler T, Gorlich D, Sattler M (2011b) Structural analysis of large protein complexes using solvent paramagnetic relaxation enhancements. Angew Chem 50(17):3993–3997. doi:10.1002/anie.201007168

Mareuil F, Sizun C, Perez J, Schoenauer M, Lallemand JY, Bontems F (2007) A simple genetic algorithm for the optimization of multidomain protein homology models driven by NMR residual dipolar coupling and small angle X-ray scattering data. Eur Biophys J 37(1):95–104

Matsuda T, Ikegami T, Nakajima N, Yamazaki T, Nakamura H (2004) Model building of a protein–protein complexed structure using saturation transfer and residual dipolar coupling without paired intermolecular NOE. J Biomol NMR 29(3):325–338. doi:10.1023/B:JNMR.0000032613.05864.87

Mattinen ML, Paakkonen K, Ikonen T, Craven J, Drakenberg T, Serimaa R, Waltho J, Annila A (2002) Quaternary structure built from subunits combining NMR and small-angle x-ray scattering data. Biophys J 83(2):1177–1183

Neylon C (2008) Small angle neutron and X-ray scattering in structural biology: recent examples from the literature. Eur Biophys J 37(5):531–541. doi:10.1007/s00249-008-0259-2

Nilsen TW, Graveley BR (2010) Expansion of the eukaryotic proteome by alternative splicing. Nature 463(7280):457–463. doi:10.1038/nature08909

Pervushin K, Riek R, Wider G, Wuthrich K (1997) Attenuated T-2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc Natl Acad Sci USA 94(23):12366–12371

Peters JA, Huskens J, Raber DJ (1996) Lanthanide induced shifts and relaxation rate enhancements. Prog Nucl Magn Reson Spectrosc 28:283–350

Petoukhov MV, Svergun DI (2006) Joint use of small-angle X-ray and neutron scattering to study biological macromolecules in solution. Eur Biophys J 35(7):567–576. doi:10.1007/s00249-006-0063-9

Pintacuda G, John M, Su XC, Otting G (2007) NMR structure determination of protein-ligand complexes by lanthanide labeling. Accounts Chem Res 40(3):206–212. doi:10.1021/ar050087z

Prestegard JH, Bougault CM, Kishore AI (2004) Residual dipolar couplings in structure determination of biomolecules. Chem Rev 104(8):3519–3540

Sachs R, Max KE, Heinemann U, Balbach J (2012) RNA single strands bind to a conserved surface of the major cold shock protein in crystals and solution. RNA 18(1):65–76. doi:10.1261/rna.02809212

Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234(3):779–815. doi:10.1006/jmbi.1993.1626

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

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

Schreiner P, Chen X, Husnjak K, Randles L, Zhang N, Elsasser S, Finley D, Dikic I, Walters KJ, Groll M (2008) Ubiquitin docking at the proteasome through a novel pleckstrin-homology domain interaction. Nature 453(7194):548–552. doi:10.1038/nature06924

Schumann FH, Riepl H, Maurer T, Gronwald W, Neidig KP, Kalbitzer HR (2007) Combined chemical shift changes and amino acid specific chemical shift mapping of protein–protein interactions. J Biomol NMR 39(4):275–289. doi:10.1007/s10858-007-9197-z

Simon B, Madl T, Mackereth CD, Nilges M, Sattler M (2010) An efficient protocol for NMR-spectroscopy-based structure determination of protein complexes in solution. Angew Chem 49(11):1967–1970. doi:10.1002/anie.200906147

Su XC, Otting G (2010) Paramagnetic labelling of proteins and oligonucleotides for NMR. J Biomol NMR 46(1):101–112. doi:10.1007/s10858-009-9331-1

Su XC, McAndrew K, Huber T, Otting G (2008) Lanthanide-binding peptides for NMR measurements of residual dipolar couplings and paramagnetic effects from multiple angles. J Am Chem Soc 130(5):1681–1687

Svergun DI, Barberato C, Koch MHJ (1995) CRYSOL—a program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. J Appl Cryst 28:768–773

Svergun DI, Richard S, Koch MH, Sayers Z, Kuprin S, Zaccai G (1998) Protein hydration in solution: experimental observation by x-ray and neutron scattering. Proc Natl Acad Sci USA 95(5):2267–2272

Takayama Y, Schwieters CD, Grishaev A, Ghirlando R, Clore GM (2011) Combined Use of Residual Dipolar Couplings and Solution X-ray Scattering To Rapidly Probe Rigid-Body Conformational Transitions in a Non-phosphorylatable Active-Site Mutant of the 128 kDa Enzyme I Dimer. J Am Chem Soc 133(3):424–427. doi:10.1021/ja109866w

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

Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH (1995) Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. Proc Natl Acad Sci 92:9279

Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE (2003) Cross-correlated relaxation enhanced 1H[bond]13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. J Am Chem Soc 125(34):10420–10428. doi:10.1021/ja030153x

Tugarinov V, Kanelis V, Kay LE (2006) Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. Nat Protoc 1(2):749–754. doi:10.1038/nprot.2006.101

Tzakos AG, Grace CRR, Lukavsky PJ, Riek R (2006) NMR techniques for very large proteins and RNAs in solution. Annu Rev Biophys Biomol Struct 35:319–342

Ubbink M, Ejdeback M, Karlsson BG, Bendall DS (1998) The structure of the complex of plastocyanin and cytochrome f, determined by paramagnetic NMR and restrained rigid-body molecular dynamics. Structure 6(3):323–335

Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Kent Wenger R, Yao H, Markley JL (2008) BioMagResBank. Nucleic acids research 36 (Database issue):D402-408. doi:10.1093/nar/gkm957

van Dijk M, Bonvin AM (2010) Pushing the limits of what is achievable in protein-DNA docking: benchmarking HADDOCK's performance. Nucleic Acids Res 38(17):5634–5647. doi:10.1093/nar/gkq222

van Kouwenhove M, Kedde M, Agami R (2011) MicroRNA regulation by RNA-binding proteins and its implications for cancer. Nat Rev Cancer 11(9):644–656. doi:10.1038/nrc3107

Vlach J, Srb P, Prchal J, Grocky M, Lang J, Ruml T, Hrabal R (2009) Nonmyristoylated matrix protein from the Mason-Pfizer monkey virus forms oligomers. J Mol Biol 390(5):967–980. doi:10.1016/j.jmb.2009.05.063

Wahl MC, Will CL, Luhrmann R (2009) The spliceosome: design principles of a dynamic RNP machine. Cell 136(4):701–718. doi:10.1016/j.cell.2009.02.009

Wang J, Zuo X, Yu P, Xu H, Starich MR, Tiede DM, Shapiro BA, Schwieters CD, Wang YX (2009) A Method for Helical RNA Global Structure Determination in Solution Using Small-Angle X-Ray Scattering and NMR Measurements. J Mol Biol. doi:10.1016/j.jmb.2009.08.001

Goddard TD, Kneller DG SPARKY 3. University of California

Zaccai G, Jacrot B (1983) Small angle neutron scattering. Annu Rev Biophys Bioeng 12:139–157. doi:10.1146/annurev.bb.12.060183.001035