Structural Biology

DOI: 10.1002/anie.200906147

An Efficient Protocol for NMR-Spectroscopy-Based Structure Determination of Protein Complexes in Solution**

*Bernd Simon, Tobias Madl, Cameron D. Mackereth, Michael Nilges, and Michael Sattler**

Dedicated to Professor Horst Kessler on the occasion of his 70th birthday and to Professor Christian Griesinger on the occasion of his 50th birthday

Eukaryotic proteins typically have a modular architecture, characterized by multiple structural domains that are connected by flexible linkers. Regulation of cellular processes depends on an interaction network between these individual modules and the formation of a quaternary structure. Dynamic rearrangement, often coupled to ligand binding, is a common feature of these multicomponent systems. While compact and rigid complexes can be efficiently studied using X-ray crystallography, protein complexes or multidomain proteins that involve weak and transient domain interactions should be preferably investigated using solution techniques. A number of NMR spectroscopy studies of protein complexes have been reported in recent years.[1–11] However, a general protocol is not available, and many applications still rely on NOE-based interdomain distance restraints, which are difficult to obtain and to assign in high-molecular-weight systems.

Herein, we present a general and robust protocol for the structural analysis of multidomain proteins and protein complexes in solution. We rely primarily on the type of NMR spectroscopy data that can be obtained for complexes with molecular weights well above 50 kDa. Starting from the three-dimensional structures of isolated domains, residual dipolar coupling (RDC) data and paramagnetic relaxation enhancements (PREs) are combined to define the relative domain orientation and interdomain distances, respectively.

A salient feature of our calculation protocol is the semi-rigid-body assembly of structural domains. The reasons for this are: 1) domains are typically compact, predominantly rigid, and of a small size that is amenable to standard structure determination by X-ray crystallography or NMR spectroscopy; 2) atomic coordinates of isolated domains or subunits from a complex of interest are often available in the Protein Data Bank; 3) in the absence of a precalculated domain structure, homology modeling, combined with chemical shift information,[12,13] may be used to generate a reasonable initial domain structure. The generalized protocol consists of four steps (Figure 1): 1) local refinement of the template domains using RDCs and backbone dihedral angles derived from ^{13}C secondary chemical shifts; 2) randomization of regions for which no prior structural knowledge is available (e.g. linkers) and global domain orientation from RDC data; 3) application of domain–domain distance restraints from PREs; 4) structure calculation using the RDC and PRE data.

Our approach is demonstrated with the structure determination of the tandem RNA recognition motif (RRM) domains (RRM12) of the human splicing factor U2AF65 bound to a nine-uridine (U9) RNA oligonucleotide. U2AF65, together with U2AF35 and Splicing Factor 1, recognizes the 3′ splice site (Figure 1) during the assembly of the spliceosome.[14] NMR ^{15}N relaxation data of the U2AF65 RRM12/U9 RNA complex indicate that the two RRMs and the RNA tumble together. Thus, the two domains interact even though they are connected by a 20-residue flexible linker (Mackereth et al., in preparation).

The first step of the protocol is to evaluate whether the input domain structures available in the PDB[15,16] resemble those in the complex. This comparison is efficiently achieved by comparing secondary chemical shifts and experimental RDC data measured for the complex with those predicted from the available three-dimensional domain structures. If small conformational differences are detected, a local refinement of the backbone conformation of the template structures can be performed using torsion-angle restraints from secondary chemical shifts and the RDC data.[7,17–19] For the RRM12/U9 complex, we recorded H^N–N and N–C′ RDCs in two alignment media, Pf1–Phage[20] and PEG–hexanol (PEG = poly(ethylene glycol)).[21] The RDC data, together with backbone torsion angle restraints obtained from TALOS[22] and hydrogen-bond-derived distance restraints, were used for local refinement of the RRM1 and RRM2

[*] Dr. T. Madl, Prof. Dr. M. Sattler
Institut für Strukturbiologie, Helmholtz Zentrum München
and
Munich Center for Integrated Protein Science at
Department Chemie, Technische Universität München
Lichtenbergstrasse 4, 85747 Garching (Germany)
Fax: (+49) 89-289-13867
E-mail: sattler@helmholtz-muenchen.de
Homepage: http://www.helmholtz-muenchen.de/stb
Dr. B. Simon
European Molecular Biology Laboratory, Heidelberg (Germany)
Dr. C. D. Mackereth
Institut Européen de Chimie et Biologie, Pessac (France)
Prof. Dr. M. Nilges
Institut Pasteur, NRS URA 2185, Paris (France)

[**] We thank Frank Gabel for discussions. C.D.M. and T.M. acknowledge support by EMBO Long Term Fellowships. T.M. thanks the Austrian Science Fund (FWF) for a Schrödinger Fellowship. We thank the EU NMR LSF, Frankfurt and the Bavarian NMR Centre (BNMRZ), Garching for NMR measurement time. This work was supported by the European Commission, grants 3D Repertoire and FSG-V-RNA (M.S.) and Deutsche Forschungsgemeinschaft (Sa 823/3).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200906147.

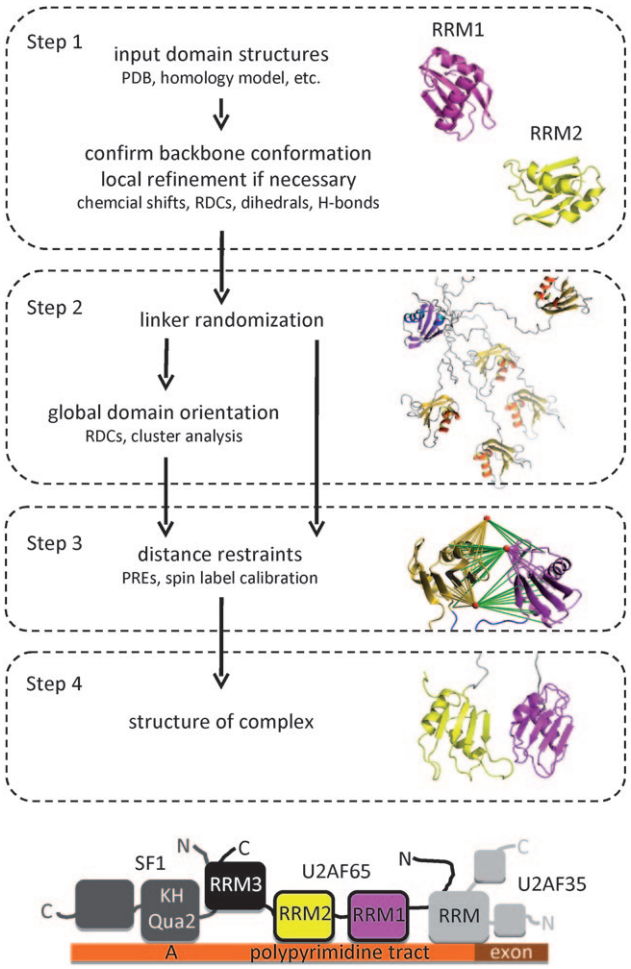

Figure 1. General overview of the protocol, illustrated with the example of the U2AF65 tandem RRM domains (RRM12), which bind to the polypyrimidine tract RNA in introns as part of the 3′-splice-site recognition complex (bottom).

templates. We have developed an efficient protocol for such a local refinement calculation, implemented in ARIA/CNS.[23] For each of the domains, a copy of the available structure is used as reference template. For local refinement, which is done separately for the individual domains, the actively calculated domain structure is coupled to the template by a harmonic potential in such a way that overall translations and rotations are possible (see the Supporting Information).

The results and analysis of the local refinement of the two RRM domains, using different forces for coupling the refined structure to the template, are summarized in Table 1. The refinement using the strongest coupling to the reference (10000 kcal mol^{-1} Å^{-2}) corresponds effectively to a fitting of RDC data to the rigid input structures. The structures are unaltered, as indicated by a low coordinate root mean square deviation (RMSD). Only minor local rearrangements occur during the refinement, and the Ramachandran plot is essentially the same as for the template structures. Calculations with lower coupling force lead to some deviations of the refined structures from the template structures, as indicated by larger backbone coordinate RMSDs. The

Table 1: Local refinement statistics.

Coupling force[a]	RDC *Q* factor[b]		Ramachandran[c]		BB RMSD[d]	
	RRM1	RRM2	RRM1	RRM2	RRM1	RRM2
input	0.57	0.58	94/6/0/0	91/9/0/0		
10000	0.50	0.57	95/5/0/0	91/9/0/0	0.00	0.01
10	0.41	0.46	99/1/0/0	94/6/0/0	0.11	0.19

[a] Energy constant (kcal mol^{-1} Å^{-2}) for coupling the active and the template structures, applied uniformly to all atoms. [b] RDC *Q* factors are calculated as described.[24] [c] Ramachandran plot distribution from PROCHECK:[25] most favored/additionally allowed/generously allowed/disallowed regions (rounded values). [d] Average backbone coordinate RMSD in Å of the ten lowest-energy structures (out of 25 calculated) with respect to the template structures of the individual domains taken from the crystal structure (PDB code 2G4B).

strength of the coupling to the reference structure is optimized by simultaneous monitoring of the coordinate RMSD and the quality of the backbone geometry in the Ramachandran plot. The local refinement should not deteriorate the local backbone geometry and thus give a similar or improved Ramachandran plot distribution. This is the case for RRM1 and RRM2 using a coupling constant of 10 kcal mol^{-1} Å^{-2}.

After the local refinement, the relative domain orientation is obtained from the RDC data in the second step. The locally refined template structures of the two RRM domains are now restrained individually with a high coupling force (effectively treating them as rigid bodies), while the 20-residue linker connecting the two domains is randomized. The ensemble of structures after simulated annealing with the experimentally derived RDC data is analyzed by clustering molecules with similar orientations (see the Supporting Information). If RDC data from a single alignment medium are used, four clusters are expected, for which the relative domain orientations are related by 180° rotations around the axes of the alignment tensor (Figure S1 in the Supporting Information).[7] If RDC data from two distinct alignment tensors are available, the four-fold degeneracy is lifted, and a unique relative domain orientation is obtained. A clear energy gap is observed between the structures with the domain orientation seen for the lowest-energy conformations and the structures with 180° rotations (Figure 2a). It is important to employ a two-step protocol, that is, local refinement and subsequent rigid-body domain orientation for the unique identification of the correct domain arrangement. The precision and accuracy of the domain orientation protocol depends on the quality of the template structures and the convergence of the domain orientation calculation (Figure 2b and the Supporting Information).

The RDC data provide orientational but no translational information. Even though the relative domain orientations may be well-defined, the quaternary structure is not obtained. Therefore, interdomain long-range distance restraints are utilized in the third step to define the relative domain arrangement. For this purpose, it is advantageous to exploit paramagnetic effects on the nuclear spins, such as paramagnetic relaxation enhancements (PREs), since the effects

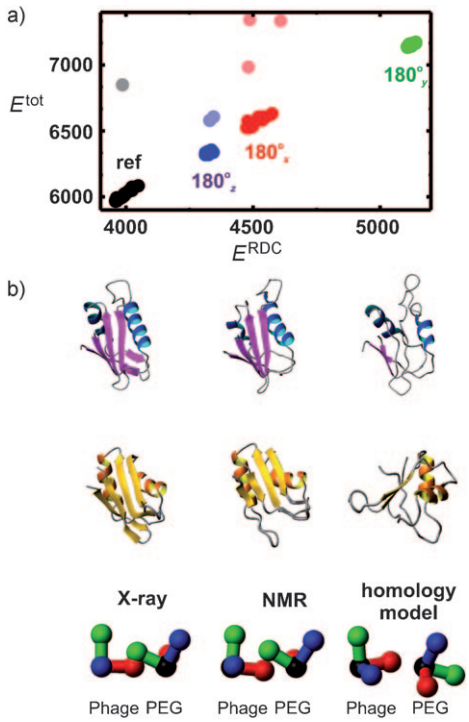

Figure 2. Domain orientation of the U2AF65 tandem RRM domains from RDC data recorded in two alignment media. a) Clustering and separation of relative domain orientations based on total and RDC energy in kcal mol^{-1} using the NMR-derived domain structures as input. b) Relative domain orientation obtained using different template domain structures, that is, based on crystal[16] and NMR[15] structures and homology models of RRM1 (magenta) and RRM2 (yellow). At the bottom, the principal axes of the two alignment tensors along *x*,*y*,*z* are shown in red, green, blue, respectively.

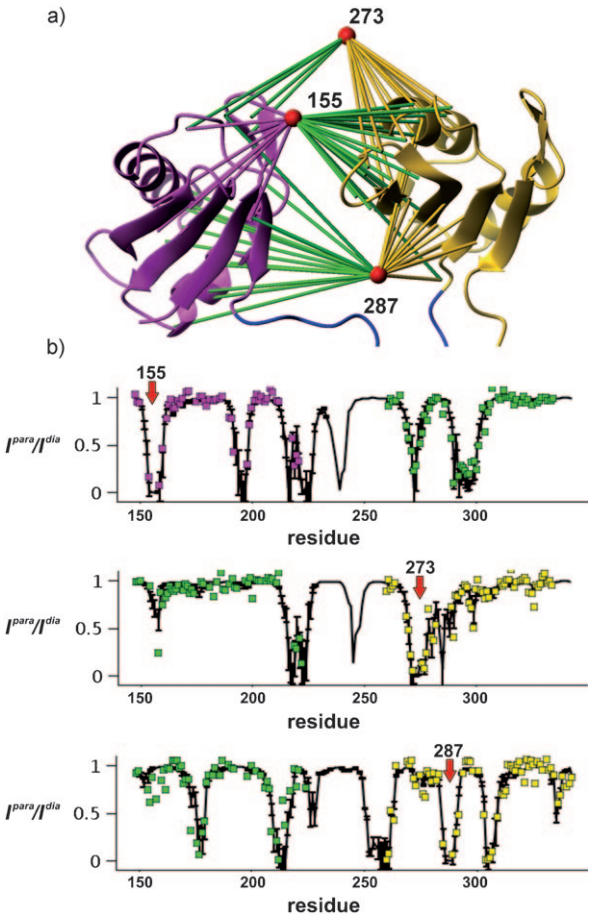

Figure 3. a) Distance restraints derived from paramagnetic relaxation enhancements of three spin labels (SL) at residues 155, 273, and 287 in U2AF65 RRM12 bound to U9 RNA. Interdomain distance restraints are colored in green, intradomain distance restraints are colored in magenta (SL155) and yellow (SL273, SL287). b) Experimental (colored points, colors as defined in (a)) and back-calculated intensity ratios I^{para}/I^{dia} (black) of amide proton signals in the three spin-labeled samples.

of electron spins on the nuclear spins are typically strong and long-range.[8,26–30] In addition to PREs, paramagnetic effects can also give rise to pseudo contact shifts and magnetic alignment (allowing for the measurement of RDCs).[31–33] In this case, we use long-range (up to 20 Å) distance restraints from PRE data to define the overall architecture of the RRM12/RNA complex. PREs are obtained by introducing nitroxyl spin labels at various positions in the protein. The spin labels can be readily prepared using standard molecular biology techniques by engineering monocysteine variants of the protein and reaction with iodoacetamido-proxyl.[26] The PREs were measured for transverse relaxation (R_2) of the proton spins as intensity ratios in 2D $^{1}H,^{15}N$ or 3D HNCO correlation spectra with oxidized (paramagnetic) and reduced (diamagnetic) spin labels on the same sample. Intensity ratios were transferred into distance restraints for the structure calculations as described[27] (see the Supporting Information).

In the final step, the RDCs and PRE data from ten spin-labeled samples were employed to calculate the domain arrangement of the two RRM domains of U2AF65 bound to U9 RNA. The structure ensemble is well-defined by the data (Figure 3 and the Supporting Information). Notably, the relative domain orientation obtained with our protocol in solution is very different to the arrangement seen in the crystal structure.[16] This difference is most likely linked to crystal packing effects (Mackereth et al., in preparation), thus underlining the importance of studying such systems using solution techniques.

The good convergence and the clustering of the ensemble of structures indicate the precision of the resulting domain arrangement and correlate with the accuracy of the structures as long as sufficient RDC and PRE data are employed. Care has to be taken with the number and the positioning of the spin labels. The experimental PREs should comprehensively cover the available conformational space to ensure the uniqueness of the solution. At the same time, the spin label groups must not interfere with domain contacts or ligand binding. It is also important to utilize a sufficient number of both attractive and repelling distance restraints from the PRE data.

We note that the overall structure of the RRM12/U9 RNA complex derived from PRE data only is similar to the structure derived by combining RDC and PRE data. Nevertheless, inclusion of RDC data is important for improved precision and accuracy of the resulting structures. The

combination of PREs without or with the inclusion of H^N–N RDCs from one or two alignment media yields structures with coordinate RMSDs of 2.1, 1.7, and 1.0 Å, respectively, compared to using the full set of RDCs (Figure S7 in the Supporting Information). This is advantageous if only a single set of RDC data is available and for application of the protocol to high-molecular-weight systems (i.e. above 50 kDa), for which it will be more difficult to obtain a sufficient amount of high-quality RDC data.

The protocol is implemented in the ARIA/CNS setup for structure calculation[23] and is therefore highly flexible. Any type of NMR data or additional restraints can be easily incorporated into the calculation, such as the inclusion of ambiguous distance restraints[4,34] derived from chemical shift perturbations, from mutational analysis, or from NOEs. The structure calculation protocol can be readily combined with other solution data, that is, from small-angle scattering experiments, which provide complementary structural information.[35–37]

The approach presented herein is not limited to structural analysis of two-domain systems. By defining rigid and flexible regions—depending on the available pre-existing structural information—the protocol can be generally applied to define the quaternary structure of multidomain proteins and multisubunit complexes in solution.

Received: November 1, 2009
Revised: December 21, 2009
Published online: February 10, 2010

Keywords: NMR spectroscopy · paramagnetic relaxation enhancement · protein structures · residual dipolar coupling · structural biology

[1] P. J. Bolon, H. M. Al-Hashimi, J. H. Prestegard, *J. Mol. Biol.* **1999**, *293*, 107.
[2] G. M. Clore, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 9021.
[3] J. D. Gross, N. J. Moerke, T. von der Haar, A. A. Lugovskoy, A. B. Sachs, J. E. McCarthy, G. Wagner, *Cell* **2003**, *115*, 739.
[4] C. Dominguez, R. Boelens, A. M. Bonvin, *J. Am. Chem. Soc.* **2003**, *125*, 1731.
[5] G. M. Clore, C. D. Schwieters, *J. Am. Chem. Soc.* **2003**, *125*, 2902.
[6] A. D. van Dijk, D. Fushman, A. M. Bonvin, *Proteins Struct. Funct. Genet.* **2005**, *60*, 367.
[7] M. Blackledge, *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *46*, 23.
[8] A. N. Volkov, J. A. Worrall, E. Holtzmann, M. Ubbink, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 18945.
[9] C. Tang, J. Iwahara, G. M. Clore, *Nature* **2006**, *444*, 383.
[10] P. Li, J. Kirkpatrick, T. Carlomagno, *J. Mol. Biol.* **2009**, *388*, 283.
[11] I. Bertini, P. Kursula, C. Luchinat, G. Parigi, J. Vahokoski, M. Wilmanns, J. Yuan, *J. Am. Chem. Soc.* **2009**, *131*, 5134.
[12] A. Cavalli, X. Salvatella, C. M. Dobson, M. Vendruscolo, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 9615.
[13] Y. Shen, O. Lange, F. Delaglio, P. Rossi, J. M. Aramini, G. Liu, A. Eletsky, Y. Wu, K. K. Singarapu, A. Lemak, A. Ignatchenko, C. H. Arrowsmith, T. Szyperski, G. T. Montelione, D. Baker, A. Bax, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 4685.
[14] C. D. Mackereth, B. Simon, M. Sattler, *ChemBioChem* **2005**, *6*, 1578.
[15] T. Ito, Y. Muto, M. R. Green, S. Yokoyama, *EMBO J.* **1999**, *18*, 4523.
[16] E. A. Sickmier, K. E. Frato, H. Shen, S. R. Paranawithana, M. R. Green, C. L. Kielkopf, *Mol. Cell* **2006**, *23*, 49.
[17] N. Sibille, A. Pardi, J. P. Simorre, M. Blackledge, *J. Am. Chem. Soc.* **2001**, *123*, 12135.
[18] J. J. Chou, S. Li, A. Bax, *J. Biomol. NMR* **2000**, *18*, 217.
[19] J. H. Prestegard, C. M. Bougault, A. I. Kishore, *Chem. Rev.* **2004**, *104*, 3519.
[20] M. R. Hansen, P. Hanson, A. Pardi, *Methods Enzymol.* **2000**, *317*, 220.
[21] M. Rückert, G. Otting, *J. Am. Chem. Soc.* **2000**, *122*, 7793.
[22] G. Cornilescu, F. Delaglio, A. Bax, *J. Biomol. NMR* **1999**, *13*, 289.
[23] J. P. Linge, M. Habeck, W. Rieping, M. Nilges, *Bioinformatics* **2003**, *19*, 315.
[24] G. Cornilescu, J. L. Marquardt, M. Ottiger, A. Bax, *J. Am. Chem. Soc.* **1998**, *120*, 6836.
[25] R. A. Laskowski, J. A. Rullmannn, M. W. MacArthur, R. Kaptein, J. M. Thornton, *J. Biomol. NMR* **1996**, *8*, 477.
[26] J. R. Gillespie, D. Shortle, *J. Mol. Biol.* **1997**, *268*, 170.
[27] J. L. Battiste, G. Wagner, *Biochemistry* **2000**, *39*, 5355.
[28] V. Gaponenko, J. W. Howarth, L. Columbus, G. Gasmi-Seabrook, J. Yuan, W. L. Hubbell, P. R. Rosevear, *Protein Sci.* **2000**, *9*, 302.
[29] B. Liang, J. H. Bushweller, L. K. Tamm, *J. Am. Chem. Soc.* **2006**, *128*, 4389.
[30] G. A. Bermejo, M. P. Strub, C. Ho, N. Tjandra, *J. Am. Chem. Soc.* **2009**, *131*, 9532.
[31] I. Bertini, C. Luchinat, G. Parigi, *Prog. Nucl. Magn. Reson. Spectrosc.* **2002**, *40*, 249.
[32] G. Otting, *J. Biomol. NMR* **2008**, *42*, 1.
[33] G. M. Clore, J. Iwahara, *Chem. Rev.* **2009**, *109*, 4108.
[34] M. Nilges, S. I. O'Donoghue, *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *32*, 107.
[35] A. Grishaev, J. Wu, J. Trewhella, A. Bax, *J. Am. Chem. Soc.* **2005**, *127*, 16621.
[36] F. Mareuil, C. Sizun, J. Perez, M. Schoenauer, J. Y. Lallemand, F. Bontems, *Eur. Biophys. J.* **2007**, *37*, 95.
[37] F. Gabel, B. Simon, M. Nilges, M. Petoukhov, D. Svergun, M. Sattler, *J. Biomol. NMR* **2008**, *41*, 199.

