NMR Spectroscopy

DOI: 10.1002/ange.201007168

Structural Analysis of Large Protein Complexes Using Solvent Paramagnetic Relaxation Enhancements**

Tobias Madl, Thomas Güttler, Dirk Görlich, and Michael Sattler*

Understanding the function of biomolecular complexes requires their structural analysis at atomic resolution. To solve high-resolution structures by ab initio calculations typically data from NMR spectroscopy or X-ray crystallography are employed. In the latter approach, intrinsic flexibility and dynamics may prevent crystallization or introduce artificial conformations linked to crystal packing. Solution NMR spectroscopy does not suffer from such limitations, but is demanding because the adverse relaxation properties of large complexes may lead to extensive signal broadening and severe spectral overlap. Consequently, only sparse restraints can be obtained from such complexes by NMR experiments. Provided that the structures of the individual components of the complex (i.e. proteins, DNA/RNA) are available and that no large-scale conformational changes occur upon complex formation, experimental and computational approaches can be used to obtain the quaternary arrangement of complexes.

The assembly of protein complexes by (semi-)rigid-body/ torsion-angle dynamics protocols is widely used and can yield accurate and precise structural models.^[1] However, for high-molecular-weight complexes, conventional approaches become highly ambiguous and often cannot distinguish between several possible arrangements of subunits in the complex. Different types of NMR data can provide powerful complementary information for restraining the interface and orientation of the complex^[2] and thereby resolve these ambiguities. One technique that has gained popularity in recent years is the site-specific incorporation of paramagnetic spin labels or lanthanide binding tags.^[3] These labels are

covalently attached to single cysteine residues and provide a rich source of distance (paramagnetic relaxation enhancements) and orientation information (pseudocontact shifts, PCS; residual dipolar couplings, RDCs). [3a,b,d] A potential drawback of this approach is the requirement of a single accessible cysteine residue for cross-linking with the paramagnetic tag. This requires removal of native cysteines by site-directed mutagenesis which can be difficult for large proteins (that may contain many cysteines).

Here, we present an efficient, generally applicable, and robust strategy for improving the precision and accuracy of (semi-)rigid-body/torsion-angle dynamics protocols based on paramagnetic relaxation enhancements (PREs) derived from the soluble paramagnetic agent Gd(DTPA-BMA) (DTPA: diethylenetriamine pentaacetic acid, BMA: bismethylamide).[4] This chemically inert compound can simply be titrated to the sample, does not require any covalent modifications, and can be easily removed by dialysis. [4a,b,5] Dipolar interactions with the unpaired electron(s) of the chelated lanthanide ion (Gd³⁺) lead to a concentrationdependent increase of nuclear relaxation rates (PRE), which result, for example, in line broadening for NMR signals of nuclear spins. [4b,5a,6] The PRE can be translated into direct distance information that reflects the solvent accessibility or, in more quantitative terms, the (minimal) distance to the closest point of the molecular surface (Figure 1).[4b]

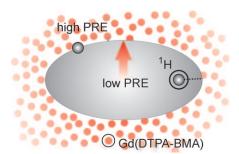


Figure 1. The general principle of solvent PREs. Paramagnetic centers of Gd(DTPA-BMA) are shown as red spheres. The red arrow indicates increasing PREs from the interior to solvent-accessible areas of the biomolecule (i.e. the surface, flexible loops/linkers).

biomolecule (i.e. the surface, flexible loops/linkers).

Our protocol uses solvent PREs in two critical steps

Our protocol uses solvent PREs in two critical steps during a semirigid-body/torsion-angle dynamics calculation, namely for 1) scoring of initial structures and 2) direct refinement against solvent PRE restraints (Figure 2). The overall approach for the structure determination is outlined in Figure 2. First, an ensemble of structures is generated by using either standard structure calculation approaches, semirigid-

[*] Dr. T. Madl, Prof. Dr. M. Sattler

Institute of Structural Biology, Helmholtz Zentrum München Biomolecular NMR and Center for Integrated Protein Science Munich

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. T. Güttler, Prof. Dr. D. Görlich

Max-Planck-Institut für biophysikalische Chemie Am Fassberg 11, 37077 Göttingen (Germany)

[**] We thank the Bavarian NMR Center (BNMRZ) for NMR measurement time. This study was supported by the EMBO (fellowship to T.M.), the Austrian Science Fund (FWF, Schrödinger fellowship to T.M.), the Max-Planck-Gesellschaft, the Boehringer Ingelheim Fonds, the Alfried Krupp von Bohlen und Halbach Foundation (fellowships to T.G.), and the European Commission contracts 3D Repertoire (LSHG-CT-2005-512028), NIM3 (No. 226507), and EUNMR (No. RII3-026145).



Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201007168.

Zuschriften

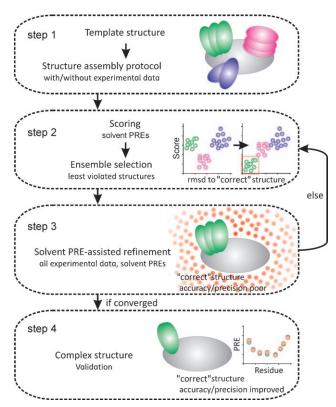


Figure 2. General overview of the solvent PRE assisted structure calculation protocol, illustrated for a schematic binary protein complex. The target structure is shown in gray; positions of the different clusters of ligand conformations are colored green, magenta, or blue.

body assembly of structural domains, [la,c,n] or a docking program (step 1). [lf,m,q] If structural restraints are available, they can be readily included at the initial stage by using the corresponding protocols. [la-d,g-j,k,l,n-p] In step 2 we use solvent PRE data to score the ensemble against experimentally determined solvent PREs and total energy. This is motivated by the principal difficulty of conventional protocols to discriminate between different solutions and to identify correct structures by appropriate quality criteria (i.e. energy scores).

By using the solvent PRE score, the most accurate cluster is identified and the ambiguities are resolved. Although the correct cluster can be identified with confidence, the corresponding ensemble of structures might still deviate significantly from the "correct" solution. To further improve the accuracy and convergence of the calculation, the best scoring ensemble is selected for a direct refinement against all experimental restraints, including solvent PREs in step 3 (Figure 2), using an extended version of a protocol described previously. [4b] Succinctly, a cloud of dummy atoms is placed around the ensemble of structures (red spheres in Figures 1 and 2), and refinement is carried out using simulated annealing/molecular dynamics in ARIA/CNS.[7] PREs are converted into distances between the nuclear spins for which a PRE was measured, and the cloud of dummy atoms represents the paramagnetic cosolvent (see the Supporting Information). A high PRE reflects a high exposure of the spin to the soluble spin label (equivalent to solvent exposure) and thus a short minimal distance to the cloud of dummy atoms. Conversely, spins with a low PRE are less surface-exposed and therefore also located far from the cloud of dummy atoms. Typically, distance restraints between 3.5–15 Å are obtained. As long as significant structural changes to the preceding iteration step are observed (i.e. backbone root mean square deviation (rmsd) > 1.0 Å), the lowest-energy subensemble of structures is selected, the dummy atoms are placed around this ensemble, and the energy minimization is repeated. The final ensemble of structures is validated by comparing experimental and back-calculated solvent PREs.

We applied this protocol to analyze a 150 kDa ternary nuclear export complex, which comprises the 123 kDa nuclear export receptor CRM1 (also known as Exportin 1 or Xpo1p), the 20 kDa guanine nucleotide-binding protein Ran (in its GTP-bound state), and a peptide with the prototypic nuclear export signal (NES) of the protein kinase A inhibitor (PKI; see Ref.[8] for a review on CRM1). NESs are the simplest CRM1-dependent nuclear export determinants. They comprise five characteristically spaced hydrophobic residues (denoted Φ^0 , Φ^1 - Φ^4) that follow the consensus $\Phi^0(x)_2$ - Φ^1 - $(x)_3-\Phi^2-(x)_{2\cdot3}-\Phi^3-x-\Phi^4$ for PKI-type NESs ("x" is an amino acid that is preferentially charged, polar, or small^[9] with the sequence around Φ^0 showing a preference for acidic residues^[10]). The recent crystal structure RanGTP·CRM1·PKI NES complex shows that the NES peptide docks its Φ residues into five rigid pockets of RanGTP-bound CRM1.[10] With conventional NMR approaches we obtained only sparse experimental data for the export complex. Therefore, the Ran-CRM1-PKI NES complex was an excellent target to test the protocol outlined above. Indeed, the model obtained by our approach validates the crystal structure of a related complex and vice versa.^[10]

The unbound PKI NES peptide is largely unfolded in solution but shows a propensity to form an α -helical secondary structure.^[10] The structural analysis of the complex required optimized isotope labeling schemes (see the Supporting Information). For the PKI NES peptide, we used methyl-protonated, linearized ¹³C side-chain labeling combined with ²H and ¹⁵N labeling. Furthermore, complete deuteration of CRM1 was needed, which we achieved by an optimized deuterium labeling protocol. As a consequence of the high molecular weight of the complex, the TROSY-HSQC spectra (TROSY = transverse relaxation-optimized NMR spectroscopy, HSQC = heteronuclear single quantum coherence) did not yield all expected cross-peaks (see Figure S1 in the Supporting Information). We therefore applied ¹H, ¹⁵N CRINEPT-HMQC experiments (CRINEPT = cross-correlated relaxation-enhanced polarization). Upon binding to CRM1, the chemical shifts in the central (hydrophobic) Φ region of the PKI NES peptide were significantly perturbed (Figure 3 B,C). The standard NMR triple-resonance approaches[11] for sequential assignment failed because of the large size of the complex. We therefore used a combination of ¹H- and ¹³C-detected through-bond and NOE-correlated NMR experiments for chemical shift assignment and derivation of structural restraints of the NES peptide (see the Supporting Information).

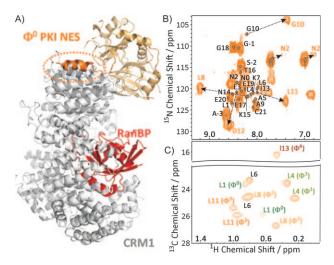


Figure 3. NMR spectra of the PKI NES peptide in the 150 kDa ternary complex with CRM1 and RanGTP. A) Ribbon representation of the crystal structure of the related nuclear export complex (PDB-ID: 3NBY), where the snurportin fusion domain (yellow) was used to facilitate crystallization of the complex. [10] B) 1 H, 15 N CRINEPT-HMQC spectrum of the NES peptide when free (black) or bound to the CRM1-RanGTP complex (orange). Chemical shift changes of selected residues upon binding are indicated by dashed arrows. C) 1 H, 13 C methyl TROSY spectrum. The Φ residues and the non-Φ Leu6 are shown in green and orange, respectively.

The conventional approach for determining the structure of the complex would rely on chemical shift assignments of all subunits and the measurement of intermolecular restraints such as NOEs. However, while NMR data can be readily obtained for the PKI NES peptide, detailed NMR analysis of isotope-labeled CRM1 is challenging because of its size (123 kDa), which would result in severe overlap of NMR signals. An alternative method to derive structural restraints involves spin labeling of otherwise unlabeled CRM1 and detection of intermolecular paramagnetic effects (i.e. PRE) on the isotope-labeled PKI NES peptide. For CRM1 this is difficult as the protein contains a large number of solventaccessible cysteines.[1p,n,3b-d,12] We therefore performed structural analysis of the 150 kDa complex with few restraints defining the interaction between the PKI NES peptide and CRM1: 1) ambiguous NOE-based distance restraints between the Φ residues of the PKI NES and amide protons of CRM1 and 2) unambiguous NOEs between PKI NES residues and Cys528 Hy (see the Supporting Information). These NOEs could be unambiguously assigned to Cys528. [10]

Several initial calculations were carried out for step 1 (Figure 2) using 1) a modified version of ARIA/CNS^[7] (see the Supporting Information), 2) HADDOCK^[1g,i,l,o], and 3) the docking program ClusPro.^[13] The modified version of ARIA/CNS allows semi-rigid-body assembly of the CRM1·NES complex by simulated annealing using experimental restraints. ClusPro was chosen as a the best-performing^[1q] representative docking program. HADDOCK allows the incorporation of experimental restraints whereas in ClusPro only residues located in the binding interface can be specified. The results for the calculations are shown in Figure 4 and Table 1. Notably, both the accuracy and the

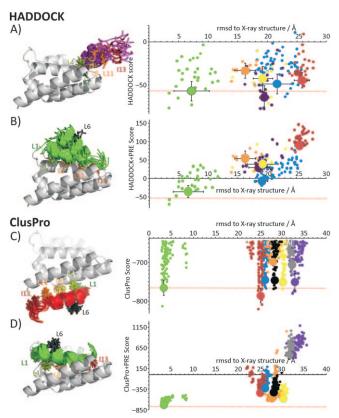


Figure 4. Impact of PRE scoring on the accuracy of the docking approaches. The 10 lowest-energy structures (left) and clusters (right) as derived from HADDOCK (A,B) and ClusPro (C,D), respectively, are shown, either before (A,C) or after (B,D) solvent PRE scoring. Residues involved in the interaction between the PKI NES peptide and CRM1 were used as restraints for calculations (see the Supporting Information). Clusters with increasing size are colored in blue, red, green, magenta, yellow, orange, black, and gray. The HADDOCK and ClusPro scores before the PRE scoring procedure (A,C) are marked by red dotted lines. Hydrophobic Φ residues and Leu6 are color-coded as in Figure 3.

precision of the best-scoring ensemble of structures are poor when compared to the crystal structure of the PKI export complex, which we used as a reference structure.^[10]

In docking calculations, the scoring of the resulting structures is critical, and it is assumed that the lowest-energy structures are the most accurate ones. However, when the restraint density is sparse, there often is no significant correlation between the accuracy and the scoring function. Furthermore, different weighting of individual terms of the scoring function can yield ambiguous results (see the Supporting Information). In the present example, both ClusPro and HADDOCK would yield an incorrect structural model of the NES complex.

We then scored the structure obtained by ClusPro and HADDOCK against the experimental solvent PRE data. Scoring was carried out by calculating the sum of the violations of $\langle r^{-6} \rangle$ averaged distance restraints between observed NMR nuclei and a grid representing the paramagnetic cosolvent. The average PRE score for all structures is scaled to the size of the average scoring energy used in

Zuschriften

Table 1: Statistics of the structural calculation protocol.

	Precision ensemble rmsd [Å] ^[a]	Accuracy rmsd to reference [Å]
Step 1		
ARIA/CNS	8.43 ± 7.61	$\textbf{9.21} \pm \textbf{2.82}$
HADDOCK ^[b]	2.81 ± 1.22	19.45 ± 0.61
HADDOCK ^[c]	$\textbf{2.25} \pm \textbf{0.98}$	3.07 ± 0.71
ClusPro	1.49 ± 0.45	25.15 ± 0.40
Step 2		
ARIA/CNS	$\textbf{7.77} \pm \textbf{2.98}$	6.90 ± 1.76
HADDOCK ^[b]	4.34 ± 1.85	$\textbf{6.58} \pm \textbf{2.58}$
HADDOCK ^[c]	2.14 ± 0.98	3.49 ± 1.09
ClusPro	1.06 ± 0.41	$\textbf{3.35} \pm \textbf{0.32}$
Step 3 ^[d]		
	0.28 ± 0.08	$\textbf{1.22} \pm \textbf{0.04}$

[a] Pairwise backbone rmsd values are calculated for residues 1–13 for the 10 lowest-energy structures from the ensemble. [b] Restraints: Φ residues and Cys528 were defined as active residues. [c] Restraints: ambiguous NOEs between Φ residues and CRM1 amide protons; unambiguous NOEs between hydrophobic Φ residues $(\Phi^3,\,\Phi^4),\, Asp12$ H^N and Cys528 H γ . [d] Structures were calculated with a modified version of ARIA/CNS^[7] as detailed in the Supporting Information.

step 1 and combined to yield a modified score. The results of step 2 show a clear correlation between accuracy and scoring energy (Figure 4; Table 1). The lowest-energy structures are closest to the "correct" reference structure. Thus, the scoring against PREs resolves the ambiguities of the conventional docking programs.

However, the resulting ensemble of structures still shows poor convergence and accuracy as indicated by a high-coordinate rmsd to the reference structure. In step 3 (Figure 2) we therefore selected the 10 best structures obtained with the combined score in step 2 and carried out a direct refinement against the PRE data in ARIA/CNS. Both the accuracy and the precision of the resulting structures improve significantly throughout the iterative refinement (Table 1, Figure 5). The final ensemble of structures was validated by comparing experimental and back-calculated solvent PREs (step 4) which show excellent agreement (Figure 5).

A closer look at the initial docking models provides an explanation as to why the solvent PRE data significantly improve the structural quality. PRE data were available only for methyl groups of (iso)leucine residues in the hydrophobic positions of the NES (Φ^{0-4} , and the non- Φ Leu6). In the initial docking models, many of these residues show considerable solvent exposure, whereas in the final complex structure they are in close contact with CRM1 and thus shielded from the paramagnetic cosolvent. The quantitative information derived from PREs defines the distance of the methyl groups from the molecular surface and thus restrains those residues that do not face the solvent into the binding pocket. On the other hand, solvent-exposed residues are moved towards the solvent. Thus, solvent PREs provide precise information about the solvent exposure of a spin and are well suited for quantitatively defining the interaction interfaces within a complex—provided that the lifetime of the complex is sufficiently long. In the NES complex studied here, K_D is 7 nm, $^{[14,10]}$ corresponding to an off-rate in the order of 10^{-2} s⁻¹

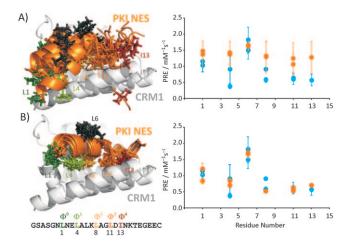


Figure 5. Impact of PRE refinement on a representative ensemble of low accuracy and precision (ARIA/CNS of Table 1, step 2). Left: Structures A) before and B) after the PRE refinement (step 3) are shown in ribbon representation. Right: Experimental solvent PREs and values back-calculated from the ensemble are shown in blue and orange, respectively. The PKI NES peptide is depicted in orange, CRM1 in gray. Hydrophobic Φ residues and Leu6 are shown in green and orange depending on their location in the sequence of the NES peptide. The amino acid sequence of the PKI NES is shown at the bottom

assuming a diffusion-controlled on-rate $(10^7 \, \text{m}^{-1} \, \text{s}^{-1})$. Even in the case of weakly associating complexes (with K_D values in the high μm range) solvent PREs can provide at least qualitative information about the binding interface. [5a]

The structural analysis of the 150 kDa NES complex demonstrates that solvent PREs are an excellent indicator of the quality of structural models and can define a valuable score for structure assembly of macromolecular complexes. Moreover, a direct refinement against solvent PRE data improves the accuracy and precision of the structural models. This approach can be readily combined with various types of experimental data and represents a powerful strategy for NMR-based structural analysis of large complexes in solution.

Received: November 15, 2010 Published online: March 25, 2011

Keywords: NMR spectroscopy · paramagnetic relaxation enhancement · proteins · structural biology

a) M. Ubbink, M. Ejdeback, B. G. Karlsson, D. S. Bendall, Structure 1998, 6, 323; b) D. S. Garrett, Y. J. Seok, A. Peterkofsky, A. M. Gronenborn, G. M. Clore, Nat. Struct. Biol. 1999, 6, 166; c) G. M. Clore, Proc. Natl. Acad. Sci. USA 2000, 97, 9021; d) A. Fahmy, G. Wagner, J. Am. Chem. Soc. 2002, 124, 1241; e) G. M. Clore, J. Kuszewski, J. Am. Chem. Soc. 2003, 125, 1518; f) J. Janin, K. Henrick, J. Moult, L. T. Eyck, M. J. Sternberg, S. Vajda, I. Vakser, S. J. Wodak, Proteins Struct. Funct. Genet. 2003, 52, 2; g) C. Dominguez, R. Boelens, A. Bonvin, J. Am. Chem. Soc. 2003, 125, 1731; h) T. Matsuda, T. Ikegami, N. Nakajima, T. Yamazaki, H. Nakamura, J. Biomol. NMR 2004, 29, 325;

- i) A. D. J. van Dijk, R. Kaptein, R. Boelens, A. Bonvin, J. Biomol. NMR 2006, 34, 237; j) M. van Dijk, A. D. J. van Dijk, V. Hsu, R. Boelens, A. Bonvin, Nucleic Acids Res. 2006, 34, 3317; k) A. M. Wu, T. Singh, J. H. Liu, M. Krzeminski, R. Russwurm, H. C. Siebert, A. Bonvin, S. Andre, H. J. Gabius, Glycobiology 2007, 17, 165; l) S. J. De Vries, A. D. J. van Dijk, M. Krzeminski, M. van Dijk, A. Thureau, V. Hsu, T. Wassenaar, A. Bonvin, Proteins Struct. Funct. Genet. 2007, 69, 726; m) S. J. Wodak, Proteins Struct. Funct. Genet. 2007, 69, 697; n) B. Simon, T. Madl, C. D. Mackereth, M. Nilges, M. Sattler, Angew. Chem. 2010, 122, 2011; Angew. Chem. Int. Ed. 2010, 49, 1967; o) S. J. de Vries, M. van Dijk, A. M. Bonvin, Nat. Protoc. 2010, 5, 883; p) T. Madl, I. C. Felli, I. Bertini, M. Sattler, J. Am. Chem. Soc. 2010, 132, 7285; q) J. Fernandez-Recio, M. J. Sternberg, *Proteins Struct*. Funct. Genet. 2010, 78, 3065.
- [2] a) G. C. K. Roberts, NMR of macromolecules: a practical approach, IRL Press at Oxford University Press, Oxford, 1993: b) A. Bax, G. Kontaxis, N. Tjandra in Nuclear Magnetic Resonance of Biological Macromolecules, Pt B, Vol. 339, Academic Press, San Diego, 2001, p. 127; c) E. R. P. Zuiderweg, Biochemistry 2002, 41, 1; d) E. de Alba, N. Tjandra, Prog. Nucl. Magn. Reson. Spectrosc. 2002, 40, 175; e) J. H. Prestegard, C. M. Bougault, A. I. Kishore, Chem. Rev. 2004, 104, 3519; f) P. J. Lukavsky, J. D. Puglisi in Nuclear Magnetic Resonance of Biological Macromolecules, Part C, Vol. 394, Elsevier Academic Press, San Diego, 2005, pp. 399; g) M. Blackledge, Prog. Nucl. Magn. Reson. Spectrosc. 2005, 46, 23; h) L. Fielding, Prog. Nucl. Magn. Reson. Spectrosc. 2007, 51, 219.
- [3] a) I. Bertini, C. Luchinat, G. Parigi, Prog. Nucl. Magn. Reson. Spectrosc. 2002, 40, 249; b) G. M. Clore, J. Iwahara, Chem. Rev. 2009, 109, 4108; c) X. Xu, P. H. Keizers, W. Reinle, F. Hannemann, R. Bernhardt, M. Ubbink, J. Biomol. NMR 2009, 43, 247; d) X. C. Su, G. Otting, J. Biomol. NMR 2010, 46, 101.

- [4] a) G. Pintacuda, G. Otting, J. Am. Chem. Soc. 2002, 124, 372; b) T. Madl, W. Bermel, K. Zangger, Angew. Chem. 2009, 121, 8409; Angew. Chem. Int. Ed. 2009, 48, 8259.
- [5] a) T. Madl, L. Van Melderen, N. Mine, M. Respondek, M. Oberer, W. Keller, L. Khatai, K. Zangger, J. Mol. Biol. 2006, 364, 170; b) A. Bernini, V. Venditti, O. Spiga, N. Niccolai, Prog. Nucl. Magn. Reson. Spectrosc. 2009, 54, 278.
- [6] a) I. Solomon, Phys. Rev. 1955, 99, 559; b) I. Solomon, N. Bloembergen, J. Chem. Phys. 1956, 25, 261; c) N. Bloembergen, J. Chem. Phys. 1957, 27, 572; d) N. Bloembergen, L. O. Morgan, J. Chem. Phys. 1961, 34, 842.
- [7] J. P. Linge, M. Habeck, W. Rieping, M. Nilges, Bioinformatics
- [8] S. Hutten, R. H. Kehlenbach, Trends Cell Biol. 2007, 17, 193.
- [9] U. Kutay, S. Guttinger, Trends Cell Biol. 2005, 15, 121.
- [10] T. Guttler, T. Madl, P. Neumann, D. Deichsel, L. Corsini, T. Monecke, R. Ficner, M. Sattler, D. Gorlich, Nat. Struct. Mol. Biol. 2010, 17, 1367.
- [11] M. Sattler, J. Schleucher, C. Griesinger, Prog. Nucl. Magn. Reson. Spectrosc. 1999, 34, 93.
- [12] a) J. L. Battiste, G. Wagner, Biochemistry 2000, 39, 5355; b) V. Gaponenko, J. W. Howarth, L. Columbus, G. Gasmi-Seabrook, J. Yuan, W. L. Hubbell, P. R. Rosevear, Protein Sci. 2000, 9, 302.
- [13] a) S. R. Comeau, D. W. Gatchell, S. Vajda, C. J. Camacho, Nucleic Acids Res. 2004, 32, W96; b) S. R. Comeau, D. W. Gatchell, S. Vajda, C. J. Camacho, Bioinformatics 2004, 20, 45; c) D. Kozakov, R. Brenke, S. R. Comeau, S. Vajda, Proteins Struct. Funct. Genet. 2006, 65, 392; d) S. R. Comeau, D. Kozakov, R. Brenke, Y. Shen, D. Beglov, S. Vajda, Proteins Struct. Funct. Genet. 2007, 69, 781.
- [14] E. Paraskeva, E. Izaurralde, F. R. Bischoff, J. Huber, U. Kutay, E. Hartmann, R. Luhrmann, D. Gorlich, J. Cell Biol. 1999, 145, 255.

4083