A Journal of the Gesellschaft Deutscher Chemiker

Angewandte International Edition Chemie

GDCh

www.angewandte.org

Accepted Article

Title: Segmental, domain-selective perdeuteration and small angle neutron scattering for structural analysis of multi-domain proteins

Authors: Miriam Sonntag, Pravin Jagtap, Bernd Simon, Marie-Sousai Appavou, Arie Geerlof, Ralf Stehle, Frank Gabel, Janosch Hennig, and Michael Sattler

To be cited as: *Angew. Chem. Int. Ed.* 10.1002/anie.201702904
Angew. Chem. 10.1002/ange.201702904

Link to VoR: http://dx.doi.org/10.1002/anie.201702904
http://dx.doi.org/10.1002/ange.201702904

WILEY-VCH

10.1002/anie.201702904

WILEY-VCH

COMMUNICATION

Segmental, domain-selective perdeuteration and small angle neutron scattering for structural analysis of multi-domain proteins

Miriam Sonntag[+], Pravin Kumar Ankush Jagtap[+], Bernd Simon, Marie-Sousai Appavou, Arie Geerlof, Ralf Stehle, Frank Gabel, Janosch Hennig* and Michael Sattler *

Abstract: Multi-domain proteins play critical roles in fine-tuning essential processes in cellular signaling and gene regulation. Typically, multiple globular domains that are connected by flexible linkers undergo dynamic re-arrangements upon binding to protein, DNA or RNA ligands. RNA binding proteins (RBPs) represent an important class of multi-domain proteins, which regulate gene expression by recognizing linear or structured RNA sequence motifs. Here, we employ segmental perdeuteration of the three RNA recognition motif (RRM) domains in the RBP TIA-1 using Sortase A-mediated protein ligation. We show that domain-selective perdeuteration combined with contrast-matched small-angle neutron scattering (SANS), SAXS and computational modelling provides valuable information to precisely define relative domain arrangements. The approach is generally applicable to study conformational arrangements of individual domains in multi-domain proteins and changes induced by ligand binding.

Most eukaryotic proteins are composed of multiple globular domains connected by flexible intrinsically disordered linkers.[1] The domain arrangements and changes induced by ligand binding play important roles for the molecular function of these proteins, and can be regulated by dynamic population shifts or cooperative assembly of the domains.[2] Structural analysis of such multi-domain proteins requires the use of solution techniques to assess the conformational dynamics. Integrated structural biology, especially the combination of solution techniques such as NMR and small-angle scattering experiments, has been shown to be highly useful to characterize the structure and dynamics of large protein complexes in solution.[3] NMR is usually well suited to obtain structural information of multi-domain proteins featuring disordered regions and flexible linkers. However, with increasing molecular weight NMR analysis of such protein complexes is challenging and only sparse NMR data, such as chemical shift perturbations, paramagnetic relaxation enhancements, and residual dipolar couplings are accessible. Most often a divide and conquer approach is employed and high-resolution structural information obtained for domains or subunits of a larger complex can then be combined with information from complementary techniques.

The combination of solution NMR with small angle scattering is advantageous as NMR provides information on binding interfaces, domain conformations and dynamics, while small angle scattering experiments yield information about overall shapes.[4] Of special interest is the use of contrast matching[5] in SANS, where the location of individual subunits in subunit-selectively perdeuterated complexes can be determined. While this approach is well established for reconstituted protein complexes,[3c, 3i, 6] SANS analysis of domain arrangements within a single-chain multi-domain protein has not been reported, as this requires the preparation of segmentally, i.e. domain-selectively perdeuterated proteins. Recently, native chemical ligation was combined with deuteration in a neutron reflectometry study[7]. However, SANS requires protein amounts in the milligram range, requiring an efficient, robust and generally applicable segmental isotope labeling method.

Here, we have used an efficient protocol for Sortase A-mediated ligation to generate domain-selective perdeuterated samples of the multi-domain RBP TIA-1, which harbors three RRMs, RRM123.[8] This method has been used previously to enable segmental isotope labeling for NMR studies.[9] The Sortase A-catalyzed reaction requires only the presence of the enzyme recognition site[10] and yields milligram amounts of ligated product with little need for optimization of the ligation protocol. Preparation of segmentally perdeuterated proteins for SANS requires efficient expression and high final yields of purified protein (> 2 mg/l) to ensure sufficient signal-to-noise.

The alternative splicing factor TIA-1 features three RRM domains followed by an unstructured Q-rich domain at its C-terminus, all connected by flexible linkers (Figure 1a).[11] TIA-1

M. Sonntag [+], Dr. P. K. A. Jagtap [+], Dr. A. Geerlof, Dr. J. Hennig, Prof. Dr. M. Sattler
Institute of Structural Biology, Helmholtz Zentrum München
Ingolstädter Landstr. 1, 85764 Neuherberg, Germany
E-mail: sattler@helmholtz-muenchen.de

M. Sonntag, Dr. P. K. A. Jagtap, Dr. R. Stehle, Prof. Dr. M. Sattler
Center for Integrated Protein Science Munich at Chair Biomolecular NMR Spectroscopy
Department Chemie, Technische Universität München
Lichtenbergstr. 4, 85747 Garching, Germany

Dr. B. Simon, Dr. J. Hennig
Structural and Computational Biology Unit
European Molecular Biology Laboratory (EMBL) Heidelberg
69117 Heidelberg, Germany. E-mail: janosch.hennig@embl.de

Dr. M.-S. Appavou
Jülich Centre for Neutron Science (JCNS) at Heinz Maier-Leibnitz Zentrum (MLZ), Forschungszentrum Jülich GmbH
Lichtenbergstr. 1, 85748 Garching, Germany

Dr. F. Gabel
Univ. Grenoble Alpes, CEA, CNRS, IBS, F-38000 Grenoble
Institut Laue-Langevin (ILL)
Avenue des Martyrs, 38042 Grenoble, France

[+] These authors contributed equally to this work.

Supporting information for this article is given via a link at the end of the document

10.1002/anie.201702904

COMMUNICATION

has been described to regulate splicing of cancer-related *fas* pre-mRNA by binding to an intronic splicing enhancer site, which promotes the recruitment of the spliceosomal U1 snRNP to initiate splicing.[12] A current model for the molecular functions of TIA-1 suggests that RRM2 and RRM3 bind to pre-mRNA, whereas RRM1 and the Q-rich domain interact with the U1C protein; a component of the spliceosomal U1 snRNP complex (Figure 1b).[13] Other studies suggested that RRM1 might contribute to RNA binding.[14] Efforts to determine high-resolution structures of tandem RRM23 or RRM123 bound to RNA using crystallography or NOE-based NMR methods have so far been unsuccessful.

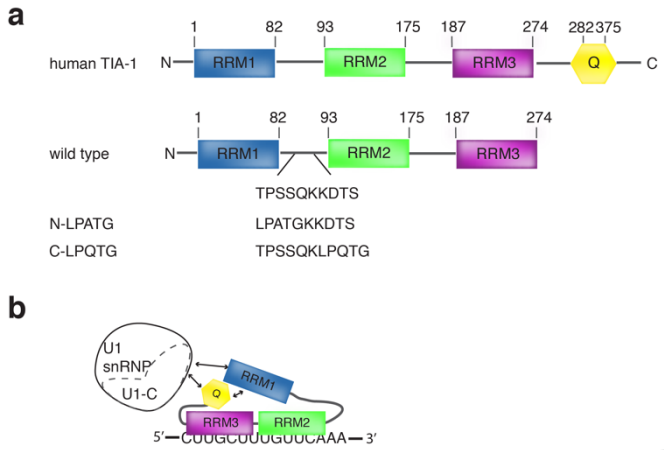


Figure 1. Multi-domain architecture of TIA-1: a) Schematic representations of the TIA-1 constructs used in this study. The RRM1-RRM2 linker sequences for wild type and two Sortase A ligation sites are given. b) Model of the domain arrangement of TIA-1 upon binding to fas mRNA and the spliceosomal U1 snRNP complex.

To analyze the structure of TIA-1 RRM123 free and bound to RNA and assess the role of RRM1 in solution we introduced the Sortase A recognition motif in the RRM1-RRM2 linker (Figure 1a). To ensure that the small modifications introduced in the linker sequence do not affect the overall conformation of TIA-1 we designed two different constructs where the Sortase A recognition site was placed either in the N-terminal half (N-LPATG) or C-terminal half (C-LPQTG) of the linker (Figure 1a). ^{1}H-^{15}N HSQC experiments, SAXS in the absence and presence of RNA and static light scattering (SLS) measurements were performed to compare both Sortase A - ligated samples with the wild type TIA-1 RRM123 protein. Comparison of NMR spectra of segmentally ^{15}N-isotope labeled and wild type RRM123 (Figure S1) do not show any significant differences. Experimental SAXS data and pairwise distribution functions (Figure S2a, b, Table S1) are also very similar for various constructs tested in the free and RNA bound form. As expected, the RNA-bound proteins are significantly more compact compared to the free forms.[11] SDS-PAGE and SLS experiments confirm the molecular weight of the proteins and the formation of a protein-RNA complex with equimolar 1:1 stoichiometry (Figure S2c, d).

Due to the absence of significant differences between both constructs, we chose the C-LPQTG construct for further structural analysis by contrast-matched SANS experiments. To investigate the role of RRM1 upon RNA binding, we prepared protonated and perdeuterated RRM1 as well as protonated and perdeuterated RRM23 for ligation (Figure 2, the deuteration level has been confirmed by SANS and NMR to be close to 100 %, see Figure S3).

The neutron scattering data of domain-selective perdeuterated TIA-1 RRM123 is summarized in Figure 3 (see Table S2 for statistics). In the absence of RNA and at 0 % D_2O in the solvent (positive scattering contrast for all components, Figure S4), the pairwise distance distribution, P(r), shows two peaks indicating the presence of two domain arrangements (Figure 3a, also visible in the SAXS data, Figure S2a right panel).

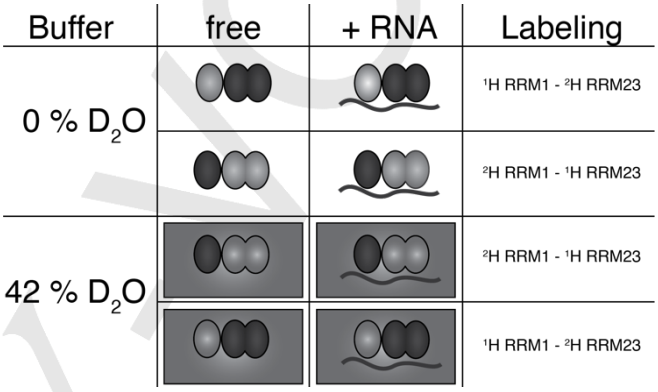


Figure 2. Contrast matching and the segmentally deuterated TIA-1 proteins for SANS measurements.

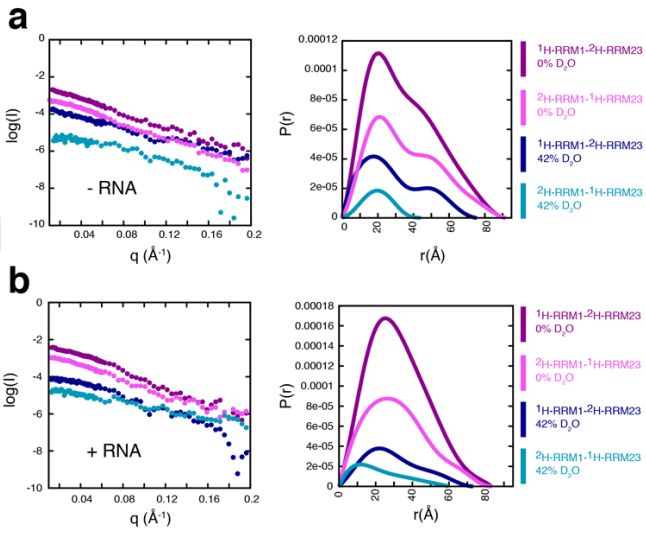


Figure 3. Small-angle neutron scattering using contrast variation of free a) and RNA bound b) segmentally deuterated TIA-1 RRM123. The scattering curves are shown on the left-hand side and the pair-wise distribution functions on the right-hand side.

The P(r) functions of the complex in the absence of RNA at 42% D_2O (Figure 3a) nicely validate the successful implementation of segmental labeling: at this contrast the protonated domains are invisible and only deuterated ones contribute to the SANS signal (Figure S4). Consequently, only one peak is visible for ^{2}H-RRM1-^{1}H-RRM23 (cyan), representing the single deuterated RRM1 domain. For ^{1}H-RRM1-^{2}H-RRM23

(blue) two domains contribute to the scattering, resulting in the presence of two peaks and a larger dimension D_{max}.

Note, that in general, SANS data recorded at D_2O concentrations of 70% and 100% provide further valuable structural information. However, in the constructs used here, the theoretical matching points are close to both D_2O concentrations and therefore yield too low S/N ratios (Figure S3).

We performed structural analysis of the protein-RNA complex based on experimental SAXS and SANS data and structures of individual domains of TIA-1 (Figure 4; Supporting Information, Figure S5-S7, Table S3, S4). We first determined the NMR solution structure of RRM1 (PDB ID: 5O2V), which shows a canonical RRM fold (Figure S5). However, the RNP1 motif exhibits two negatively charged residues that could account for poor RNA binding affinity of RRM1 (Figure S6).[11, 13a] Next we determined a crystal structure of RRM2 bound to a UU RNA motif (PDB ID: 5O3J), providing insight in the RNA recognition of this domain (Figure S7). Finally, the solution structure of RRM3 reported previously[11] exhibits an additional helix α_0 that precedes a canonical RRM fold.

These structures were used for rigid body modelling of the three-domain arrangement. Details of the restrained molecular dynamics protocol are provided in the supplement. In brief 5000 initial structural models were generated and scored against the experimental SAXS and SANS data (Figure S8 illustrates the total conformational space sampled). A representative ensemble of 100 initial models (Figure 4a) illustrates the large possible conformational space sampled by randomizing the linker sequences connecting the three RRMs. For these structures SAXS data were back-calculated using CRYSOL[15] and scored against the experimental SAXS data of the RRM123-U15 and RRM23-10mer. From the large range of χ^2-values (Figure 4d) we kept only those structures that exhibit a χ^2 below a cut-off of 1.5x the smallest χ^2 (SAXS filter, Figure 4a).[3e]

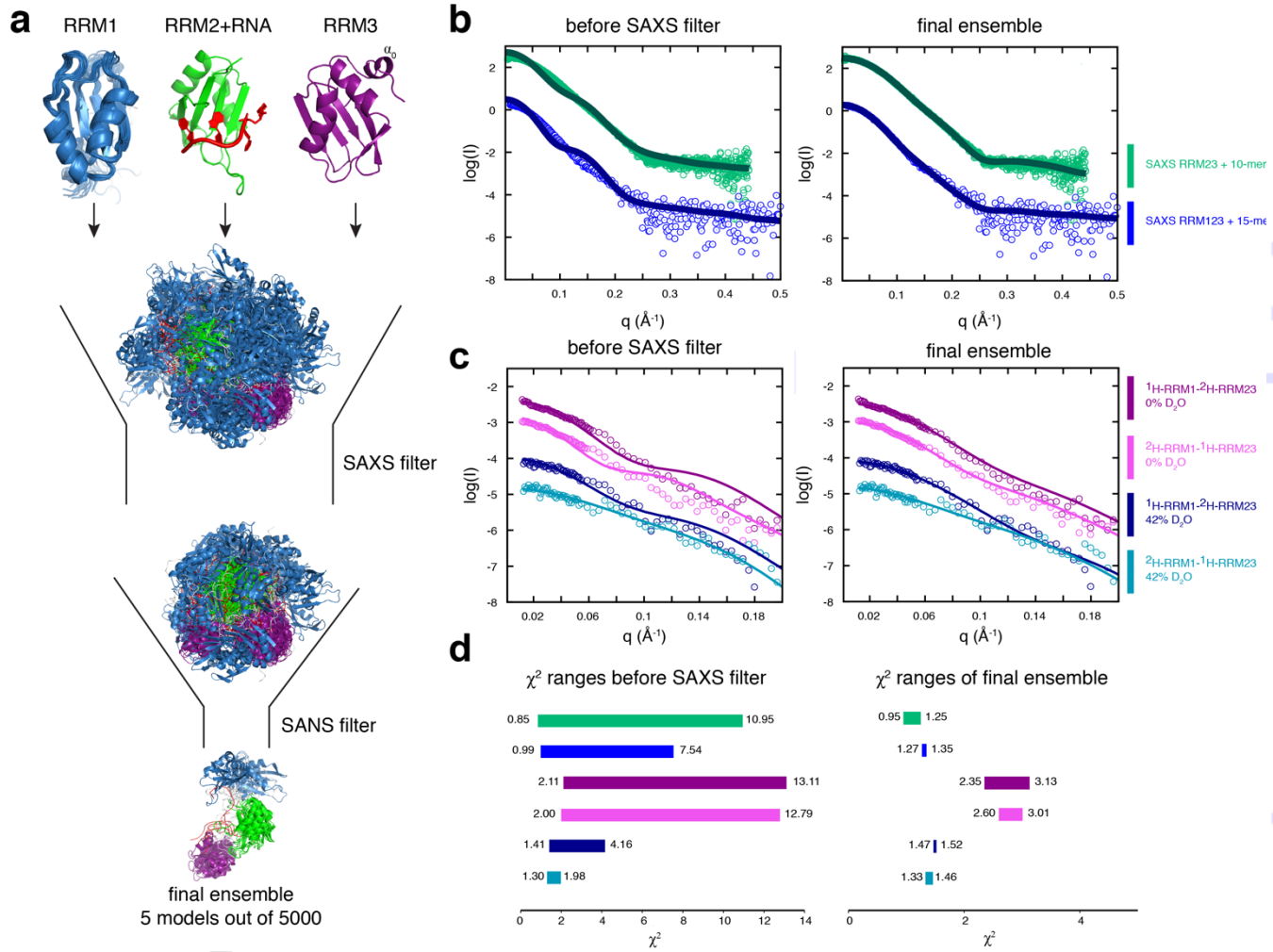


Figure 4 Structural modelling of TIA-1 RRM123 domain arrangements bound to U15 RNA based on SAXS and SANS data. a) The three available high-resolution input structures of individual domains are used to calculate 5000 models of the RRM123-U15 RNA complex, where domain linkers were randomized. 100 structures chosen from the pool of 5000 are shown. After application of a SAXS filter an ensemble of 67 structures is left, which is further reduced to five structures after scoring against the SANS data. b) Upper left: SAXS data and fits of back-calculated scattering densities of an arbitrarily chosen structure from the pool of 5000. Upper right: Back-calculated scattering densities of the five best structures. c) The same is shown for SANS data and back-calculated scattering densities. d) The χ^2 range of the ensemble (left: all 5000, right: best five) is given for each curve.

To further refine the domain arrangements in the RRM123-RNA complex we scored the ensemble of 67 structures that passed the SAXS filter against the experimental SANS data using CRYSON[16] with the same cut-off criteria. This reduced the number of structures to five and significantly restricted the allowed conformational space of domain arrangements (Figure 4a). The average RMSD for the center-of-mass of each RRM to the mean decreases from 6.6 +/- 3.2 Å for the initial 4500 structures, to 4.5 +/- 2.3 Å for the SAXS selected ensemble and finally to 2.0 +/- 1.2 Å for the final five SAXS/SANS selected structures. A second calculation with a different random seed produced a final ensemble with very similar statistics, thus confirming that the protocol efficiently samples conformational space and converges (Table S5). Note, that the SANS data for ^{1}H-RRM1-^{2}H-RRM23 at 0 % and 42 % and for ^{2}H-RRM1-^{1}H-RRM23 at 0 % have a much larger discriminative power with χ^2 values ranging from 1.41-13.11 than for ^{2}H-RRM1-^{1}H-RRM23 at 42 %, which has a much narrower range of χ^2 values from 1.30-1.98 and barely contributes to the filtering process (Figure 4d). The final ensemble provides unique structural insight into the architecture of the RRM123-RNA complex. i) RRM123 adopts a compact domain arrangement upon binding to U15 RNA with an elongated L-shape. ii) RRM23 are closely packed against each other while RRM1 is spatially more separated. iii) RRM1 is detached from RRM23 and the RNA, consistent with a minor role of RRM1 in RNA binding.

In conclusion, segmental, domain-selective perdeuteration enables efficient rigid body refinement of multi-domain proteins by combining SAXS and SANS data. The utility of contrast-matched SANS data based on segmental perdeuteration is illustrated by the excellent performance of the SANS filter. The structural ensemble provides unique insight into domain arrangements. Additional experimental data, e.g. from NMR data[4, 17], can be included to provide high-resolution structural information, i.e. for domain orientations and residue-level distance information. In general SANS measurements at 70% and 100% D_2O concentrations provide additional discriminative power for multi-domain proteins in general, further enhancing the utility of this protocol. Taken together, contrast-matched SANS combined with segmental domain-selective perdeuteration provides an efficient tool for structural analysis of multi-domain proteins and is expected to be widely applicable.

Acknowledgements

We are thankful to Robert Janowski for helpful discussions. SANS experiments were performed at the KWS-1 and KWS-2 beamlines operated by JCNS at the Heinz Maier-Leibnitz Zentrum (MLZ), Garching, Germany, and at the D22 beamline at the ILL, Grenoble, France, where we gratefully acknowledge local contact Anne Martel. SAXS data were recorded in-house at the SAXS instrument provided by SFB 1035 at TUM Garching and at the ESRF, BM29 beamline in Grenoble. This work is supported by the Deutsche Forschungsgemeinschaft grants SFB1035 and GRK1721 (to M.S). J.H. acknowledges support from the European Molecular Biology Laboratory (EMBL). PKAJ acknowledges Boehringer Ingelheim Fonds for a doctoral fellowship and funding from GRK1721.

Keywords: protein ligation • segmental perdeuteration • small angle neutron scattering •X-ray diffraction • NMR • structural biology

[1] a) B. Ma, C. J. Tsai, T. Haliloglu, R. Nussinov, *Structure* **2011**, *19*, 907-917; b) Y. Shamoo, N. Abdul-Manan, K. R. Williams, *Nucleic acids research* **1995**, *23*, 725-728.

[2] a) C. D. Mackereth, T. Madl, S. Bonnal, B. Simon, K. Zanier, A. Gasch, V. Rybin, J. Valcarcel, M. Sattler, *Nature* **2011**, *475*, 408-411; b) E. Delaforge, S. Milles, J. R. Huang, D. Bouvier, M. R. Jensen, M. Sattler, D. J. Hart, M. Blackledge, *Frontiers in molecular biosciences* **2016**, *3*, 54; c) P. Bernado, M. Blackledge, *Nature* **2010**, *468*, 1046-1048.

[3] a) B. Antonsson, R. Leberman, B. Jacrot, G. Zaccai, *Biochemistry* **1986**, *25*, 3655-3659; b) M. Falb, I. Amata, F. Gabel, B. Simon, T. Carlomagno, *Nucleic acids research* **2010**, *38*, 6274-6285; c) J. Hennig, C. Militti, G. M. Popowicz, I. Wang, M. Sonntag, A. Geerlof, F. Gabel, F. Gebauer, M. Sattler, *Nature* **2014**, *515*, 287-290; d) J. Hennig, I. Wang, M. Sonntag, F. Gabel, M. Sattler, *Journal of biomolecular NMR* **2013**, *56*, 17-30; e) A. Lapinaite, B. Simon, L. Skjaerven, M. Rakwalska-Bange, F. Gabel, T. Carlomagno, *Nature* **2013**, *502*, 519-523; f) A. G. Kikhney, D. I. Svergun, *FEBS letters* **2015**, *589*, 2570-2577; g) T. Madl, F. Gabel, M. Sattler, *Journal of structural biology* **2011**, *173*, 472-482; h) D. A. Jacques, J. Trewhella, *Protein science : a publication of the Protein Society* **2010**, *19*, 642-657; i) T. Carlomagno, *Journal of magnetic resonance* **2014**, *241*, 126-136.

[4] J. Hennig, M. Sattler, *Protein science : a publication of the Protein Society* **2014**, *23*, 669-682.

[5] G. Zaccai, B. Jacrot, *Annu Rev Biophys Bioeng* **1983**, *12*, 139-157.

[6] V. Ramakrishnan, *Science* **1986**, *231*, 1562-1564.

[7] Z. Jiang, F. Heinrich, R. P. McGlinchey, J. M. Gruschus, J. C. Lee, *J Phys Chem Lett* **2017**, *8*, 29-34.

[8] L. Freiburger, M. Sonntag, J. Hennig, J. Li, P. Zou, M. Sattler, *Journal of biomolecular NMR* **2015**, *63*, 1-8.

[9] a) O. R. Lorenz, L. Freiburger, D. A. Rutz, M. Krause, B. K. Zierer, S. Alvira, J. Cuellar, J. M. Valpuesta, T. Madl, M. Sattler, J. Buchner, *Mol Cell* **2014**, *53*, 941-953; b) F. P. Williams, A. G. Milbradt, K. J. Embrey, R. Bobby, *PloS one* **2016**, *11*, e0154607.

[10] a) H. Mao, S. A. Hart, A. Schink, B. A. Pollok, *Journal of the American Chemical Society* **2004**, *126*, 2670-2671; b) T. Proft, *Biotechnol Lett* **2010**, *32*, 1-10; cJ. M. Antos, M. C. Truttmann, H. L. Ploegh, *Curr Opin Struct Biol* **2016**, *38*, 111-118.

[11] I. Wang, J. Hennig, P. K. Jagtap, M. Sonntag, J. Valcarcel, M. Sattler, *Nucleic acids research* **2014**, *42*, 5949-5966.

[12] a) P. Forch, O. Puig, N. Kedersha, C. Martinez, S. Granneman, B. Seraphin, P. Anderson, J. Valcarcel, *Mol Cell* **2000**, *6*, 1089-1098; b) P. Forch, J. Valcarcel, *Apoptosis* **2001**, *6*, 463-468; c) J. M. Izquierdo, *Biochem Biophys Res Commun* **2010**, *402*, 324-328; d) J. M. Izquierdo, N. Majos, S. Bonnal, C. Martinez, R. Castelo, R. Guigo, D. Bilbao, J. Valcarcel, *Mol Cell* **2005**, *19*, 475-484.

[13] a) L. M. Dember, N. D. Kim, K. Q. Liu, P. Anderson, *J Biol Chem* **1996**, *271*, 2783-2788; b) P. Forch, O. Puig, C. Martinez, B. Seraphin, J. Valcarcel, *EMBO J* **2002**, *21*, 6882-6892.

[14] W. J. Bauer, J. Heath, J. L. Jenkins, C. L. Kielkopf, *Journal of molecular biology* **2012**, *415*, 727-740.

[15] D. Svergun, C. Barberato, M. H. J. Koch, *Journal of Applied Crystallography* **1995**, *28*, 768-773.

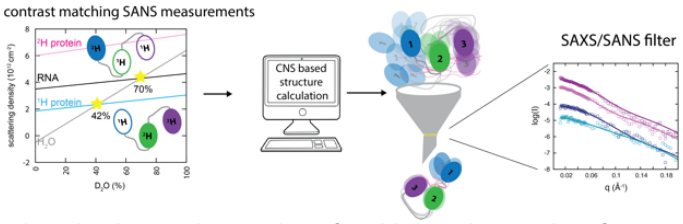
[16] D. I. Svergun, S. Richard, M. H. Koch, Z. Sayers, S. Kuprin, G. Zaccai, *Proc Natl Acad Sci U S A* **1998**, *95*, 2267-2272.

[17] C. Gobl, T. Madl, B. Simon, M. Sattler, *Progress in nuclear magnetic resonance spectroscopy* **2014**, *80*, 26-63.

10.1002/anie.201702904

WILEY-VCH

COMMUNICATION



Domain-selective perdeuteration of multi-domain proteins: Segmental perdeuterated multi-domain proteins are obtained by Sortase A mediated ligation. Using contrast matching SANS experiments on different domain-selectively perdeuterated samples of the same multi-domain protein gives unique insight into single domain arrangement upon ligand binding and can help to discriminate between different structural models.

Miriam Sonntag, Pravin Kumar Ankush Jagtap, Bernd Simon, Marie-Sousai Appavou, Arie Geerlof, Ralf Stehle, Frank Gabel, Janosch Hennig and Michael Sattler **

Page No. – Page No.

Segmental, domain-selective perdeuteration and small angle neutron scattering for structural analysis of multi-domain proteins