



LEGO-NMR Spectroscopy: A Method to Visualize Individual Subunits in Large Heteromeric Complexes**

Markus Mund, Jan H. Overbeck, Janina Ullmann, and Remco Sprangers*

NMR spectroscopy is unique as it provides a means to study bio-molecules with atomic resolution in a near natural environment. Traditionally, NMR spectroscopic analysis of structures, interactions, and dynamics has been reserved for molecular complexes that are smaller than 20 kDa. However, in recent years, the introduction of TROSY techniques,^[1] protein deuteration,^[2] and selective methyl-group labeling^[3] has significantly extended this molecular weight limit.^[4] Indeed, systems far over 100 kDa have been analyzed in great detail, revealing unique functional aspects of large molecular machines.^[5]

Many NMR spectroscopic studies on large systems have been performed on highly symmetric complexes, as these assemblies are relatively easy to prepare and result in simple NMR spectra in which the resonance signals from all the subunits are identical.^[5a] For large asymmetric assemblies that can be produced in *E. coli* by co-expression of all the components,^[6] spectral crowding will lead to NMR spectra that can no longer be analyzed in detail. In a limited number of cases this crowding could be circumvented by in vitro reconstitution of the complex from separately expressed NMR active and NMR inactive subunits.^[5b,c,h] This strategy is, however, not generally applicable. As a result, most eukaryotic systems that are much more complex than their bacterial or archaeal counterparts will remain inaccessible to high-resolution NMR studies.

Herein, we introduce a sequential co-expression method for the preparation of large asymmetric complexes that combines the advantages of in vivo reconstitution and the benefits of partial NMR isotope labeling to reduce NMR spectral complexity. We transform *E. coli* cells with two

plasmids carrying different promoters so that protein expression can be induced independently. In this manner, it is possible to induce protein synthesis for one set of proteins in an NMR active medium (stage 1), whereas a second set of proteins can be produced in an NMR invisible medium (stage 2; Figure 1 A). As all expressed proteins are present in a single *E. coli* cell, the final complex can assemble in a cellular environment preventing the aggregation of subunits that are otherwise instable in isolation. We refer to our method to “label, express, and generate oligomers” for NMR as “LEGO-NMR”.

The LEGO method requires tightly controlled individual DNA promoters such that the promoter that induces protein expression in stage 1 is completely switched off in stage 2, whereas the promoter for stage 2 is not active in stage 1. In LEGO methods A1 and A2 (Supporting Information, Figure S1), protein production is induced from an araBAD promoter using arabinose in stage 1 and from a T7 promoter using IPTG in the stage 2.^[7] In this case, the glucose that is present in stage 2 efficiently turns off the araBAD promoter.^[8] In LEGO method B, we introduce a three-promoter system, where protein production is induced from a T7 promoter in stage 1, and from an araBAD promoter in stage 2. In this case, the T7 promoter is actively switched off by the expression of T7 lysozyme in between stage 1 and stage 2 from a third plasmid that contains a rhamnose inducible promoter.^[9] This inhibition is required as T7 expression would otherwise continue for over 4 h after the removal of IPTG from the growth medium.^[7a]

To establish the LEGO-NMR methodology, we use two different LSm complexes that play a role in mRNA degradation and pre-mRNA splicing. The LSm1–7 complex^[10] (containing the LSm1 to LSm7 proteins) and LSm2–8 complex^[11] (containing the LSm2 to LSm8 proteins) contain seven different protein chains that are arranged in a unique order.^[11b,12] As most LSm proteins are insoluble in isolation, neither the LSm1–7 nor the LSm2–8 complex can be efficiently reconstituted in vitro from separately expressed proteins.^[13] On the other hand, co-expression of the different LSm proteins yields homogeneous NMR samples (Figure 1 B), showing that in-cell reconstitution functions efficiently. However, owing to the large number of unique resonances (649 expected backbone amide signals) the resulting NMR spectra suffer significantly from spectral overlap (Figure 1 B), preventing an accurate analysis. LSm complexes are thus a good example of eukaryotic protein complexes that are currently not accessible for detailed high-resolution NMR spectroscopic techniques.

To reduce the spectral overlap for the LSm2–8 complex by a factor of approximately two, we labeled the LSm5, LSm6,

[*] M. Mund,^[†] J. H. Overbeck, J. Ullmann, Dr. R. Sprangers
Max Planck Institute for Developmental Biology
Spemannstrasse 35, 72076 Tübingen (Germany)
E-mail: remco.sprangers@tuebingen.mpg.de

[†] Present address: European Molecular Biology Laboratory
Meyerhofstrasse 1, 69117 Heidelberg (Germany)

[**] We acknowledge Silke Wiesner for advice and suggestions to the manuscript. We thank Ancilla Neu, Lisa Strittmatter, and other group members for discussions. This study was supported by the Max Planck Society and a Marie Curie Reintegration Grant (FP7/2007-2013, grant agreement no. 239164) to R.S.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201304914>.

© 2013 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of Creative Commons the Attribution Non-Commercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

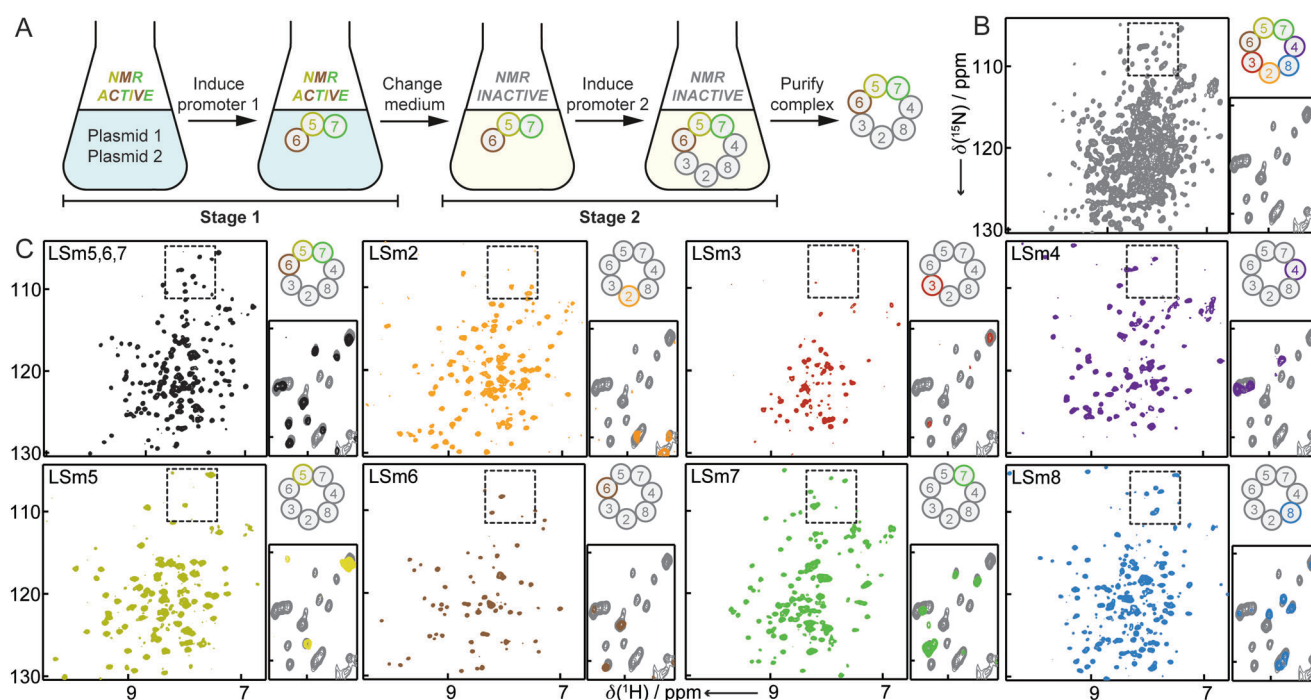


Figure 1. Principle of LEGO-NMR spectroscopy: A) shown is method A1 (Table S2, Figure S1). In stage 1 the *E. coli* cells are grown in NMR active medium and protein synthesis is induced from plasmid 1, resulting in the production of NMR active LSm5, LSm6, and LSm7. In stage 2, the cells are transferred to an NMR inactive medium, where protein synthesis from plasmid 2 is induced, resulting in the production of NMR invisible LSm2, LSm3, LSm4, and LSm8. By addition of a purification tag to a single subunit of the complex, the intact complex can be straightforwardly isolated and purified to homogeneity. NMR active subunits are colored, whereas NMR invisible subunits are in gray. B) ^1H - ^{15}N TROSY NMR spectrum (gray) of the uniformly NMR active LSm2–8. Especially the central region suffers from severe resonance overlap, complicating spectral analysis significantly. C) Top left: LEGO ^1H - ^{15}N NMR spectrum (black) of the LSm2–8 complex, in which LSm5, LSm6, and LSm7 are NMR active and LSm2, LSm3, LSm4, and LSm8 are NMR invisible. Other panels: LEGO ^1H - ^{15}N NMR spectra of the LSm2–8 complex in which individual LSm subunits are NMR active in an otherwise NMR inactive background. The resonance signals in the spectra display a subset of the resonance signals observed in the fully labeled complex (see enlargements).

and LSm7 proteins with ^{15}N in stage 1, whereas the LSm2, LSm3, LSm4, and LSm8 proteins were produced in an NMR inactive medium in stage 2. The resulting spectrum of the LSm2–8 complex that only displays LSm5, LSm6, and LSm7 is significantly simplified (Figure 1C, top left). Importantly, a very good overlay of a subset of the resonances of the fully NMR active LSm2–8 complex is observed, as we intended to achieve. This situation clearly allows for the identification of the resonances in the LSm2–8 complex that result from the LSm5, LSm6, and LSm7 proteins.

To establish the power of the sequential co-expression methodology further, we produced seven different NMR samples of the LSm2–8 complex, in which only a single LSm protein was ^{15}N -labeled in the stage 1, whereas the remaining six LSm proteins were expressed in an NMR inactive form in stage 2 (Figure 1C). The seven spectra of the LSm2–8 complex allow for the unambiguous identification of the resonance signals that result from each individual LSm protein in the LSm2–8 complex. In this manner, a simplification of 89% can be achieved (74 expected amide signals in the LEGO LSm6 spectrum). Our approach is thus able to deconvolute the complicated spectrum of the hetero-heptameric complex into seven significantly simplified sub-spectra. At the same time, the overlay of the seven NMR spectra of

the complexes that contain a single labeled LSm protein yields the spectrum of the uniformly labeled LSm2–8 complex (Figure S2). Note that the proteins that are produced in a deuterated form in stage 1 are efficiently re-protonated at the beginning of stage 2 before the individual subunits are incorporated in the final complex. This eliminates the need of (refolding) methods to re-protonate backbone amides (Figure S3) in the LSm2–8 complex.

The LSm2–8 complex is part of the U6 snRNP, where it interacts with the 3' end of the U6 snRNA.^[11a,14] To establish which subunits in the LSm2–8 complex contact the RNA substrate, we performed NMR titration experiments with LSm2–8 LEGO complexes that either contained NMR active LSm2, LSm3, LSm4, and LSm8 (Figure S4A) or that contained NMR active LSm5, LSm6, and LSm7 (Figure S4B) in an otherwise NMR inactive background. In both complexes, we observed significant chemical shift perturbations upon complex formation with the RNA. Importantly, the single subunit LEGO spectra (Figure 1C) establish that all seven LSm proteins are involved in RNA binding (Figure S4A,B) as resonance signals for all the LSm proteins experience chemical shift changes upon interaction with the RNA. To resolve the remaining spectral overlap, we performed an RNA titration experiment with an LSm2–8 complex that was

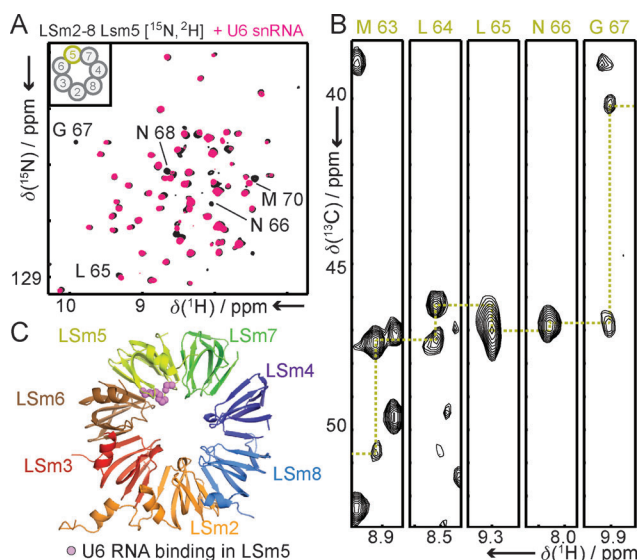


Figure 2. Assignment of LSm5 in the LSm2–8 complex. A) LSm5 in the LSm2–8 complex before (black) and after (pink) the addition of the 3' end of the U6 snRNA B) LSm5 resonance strips from an HNCA experiment that was recorded on a fully labeled LSm2–8 complex. The resonance strips could be selected based on the LSm5 LEGO-NMR spectrum (A, and Figure 1 C). C) LSm5 residues that experience the largest chemical shift changes upon addition of the U6 snRNA are marked as pink spheres on a model of the LSm2–8 complex.

labeled at LSm5 only (Figure 2 A). We then combined information from the previously assigned LSm657 complex,^[13a] the LSm5 LEGO spectrum (Figure 1 C) and an HNCA spectrum of a fully ^2H , ^{13}C , ^{15}N -labeled LSm2–8 complex (Figure 2 B) to assign the LSm5 residues in the LSm2–8 complex that contact the RNA. In this case, we exploited the fact that we were able to select the LSm5 resonance signals in the HNCA spectrum of the fully labeled LSm2–8 complex, thus reducing the number of expected resonance signals from 649 to 77, which significantly simplified the assignment process. This approach revealed that the

residues that experience large chemical shift perturbations upon interaction with the U6 snRNA are located in loop 5 of the LSm5 protein (Figure 2 C). This loop connects β -strands 4 and 5 in the LSm fold and lines the central pore of the LSm ring. As the RNA we used for the interaction experiments contains only nine bases and as all the LSm proteins are involved in the RNA interaction (Figure S4) our data suggests that the RNA binding site in LSm2–8 is at the central pore. Additional information to support this observation can be obtained from the assignment of the other LSm proteins in the LSm2–8 complex in an analogous manner. Interestingly, the eukaryotic Sm complex,^[15] the archaeal LSm complex,^[16] and Hfq^[17] have all been shown to use this region to interact with substrate RNA indicating that this binding site is conserved in the eukaryotic LSm complexes.

Methyl TROSY spectroscopy has been shown to be highly suitable for the study of supramolecular complexes that are inaccessible to backbone-directed TROSY spectroscopy. To establish ϵ - ^1H - ^{13}C methyl labeling of methionine residues in concert with LEGO-NMR, we used the hetero-heptameric LSm1–7 complex whose ^1H - ^{15}N TROSY spectra are of lower quality compared to those of the LSm2–8 complex (Figure S5). Methionine methyl TROSY spectra of the LSm1–7 complex, where all proteins are fully methionine labeled, display a large number of well resolved methyl resonances in addition to a region that suffers from significant spectral overlap (Figure 3 A, top left). To resolve the spectral overlap and to assign the well-resolved resonances to specific LSm proteins, we prepared seven different LEGO NMR samples of the LSm1–7 ring. In each of these samples a single LSm protein was methionine labeled, whereas the other six LSm proteins were NMR invisible. Methyl TROSY spectra of these hetero-heptameric complexes allowed for the unambiguous assignment of the methionine methyl groups to individual LSm proteins (Figure 3 A). Site-specific assignment of these methyl groups can be made using a mutational approach.^[5a,18] In addition, the “singly labeled” LSm1–7 rings significantly resolved the spectral overlap of the

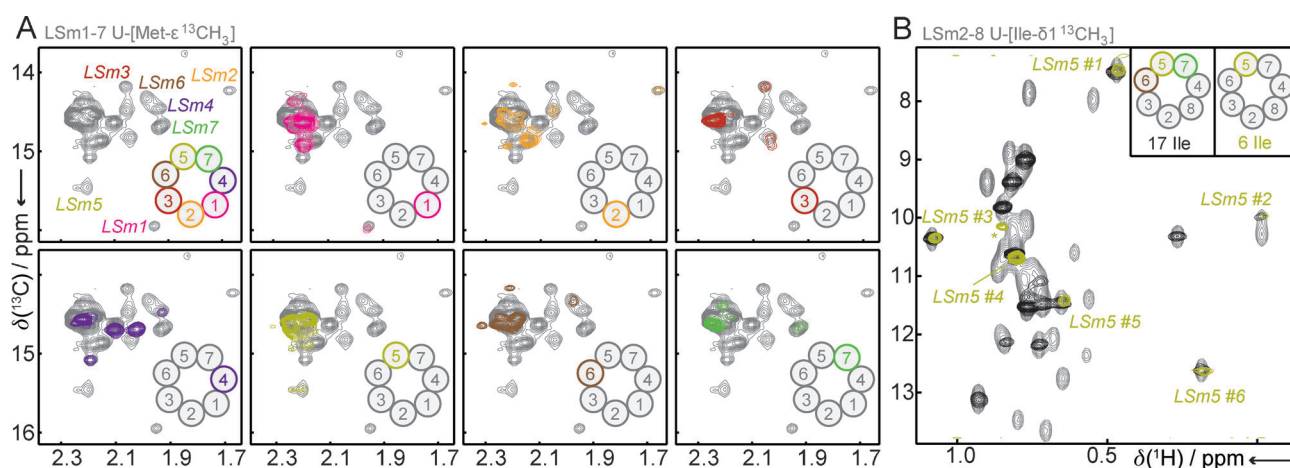


Figure 3. Methyl-group labeling. A) Methionine methyl-group spectra of the LSm1–7 complex. Top left: the methyl TROSY spectrum of an LSm1–7 complex in which all the LSm proteins are NMR active. This spectrum can be deconvoluted into seven simplified spectra that contain only a singly NMR-active LSm protein (other panels). B) Methyl TROSY spectra of Ile- δ 1 labeled LSm2–8 complexes. The spectrum of the fully isoleucine labeled spectrum (gray) is simplified by labeling only the LSm5, LSm6, and LSm7 proteins (black) or only the LSm5 protein (olive).

spectrum. Methionine methyl TROSY spectroscopy is thus fully compatible with the LEGO-NMR methodology and can provide high-resolution spectra for complexes that are not amenable to ^1H , ^{15}N -based TROSY spectroscopy. Note that it has been shown recently that methionine methyl groups are excellent probes to study molecular interactions.^[19]

In addition to methionine methyl groups, methyl TROSY spectroscopy is often performed in concert with labeled methyl groups of isoleucine, leucine, valine,^[20] or alanine^[21] residues. As opposed to methionine labeling, these amino acids are incorporated into the protein through *E. coli* metabolization of specifically labeled precursor molecules. For isoleucine residues this is only possible in the presence of glucose as that induces catabolite repression that inhibits metabolic pathways that would otherwise degrade α -ketobutyric acid.^[22] Stage 1 in method A1, that we used for methionine and nitrogen labeling (Figure S1, Table S2), uses glycerol as a carbon source and can thus not be used for isoleucine labeling. To label selected subunits with isoleucine methyl groups we thus use method A2 (where the NMR labeling is moved from stage 1 to stage 2) or method B (where an arabinose-inducible vector is used in stage 2; Figure S1).

The high quality of the spectrum of the fully isoleucine- $\delta 1$ labeled spectrum reflects the strength of methyl-group labeling for high molecular-weight complexes (Figure 3B, gray). We then used method B to prepare a “half-labeled” LSm2–8 LEGO complex that contains NMR active isoleucine- $\delta 1$ methyl groups in LSm5, LSm6, and LSm7 (Figure 3B, black). As observed for the H,N-based spectra (Figure 1C), a subset of the resonances that result from the labeled proteins can be readily identified. It is worth noting that the LSm5, LSm6, and LSm7 proteins contain 17 isoleucine residues, 16 of which yield well-dispersed resonance signals in the spectrum. To extend the LEGO approach one step further, we used method A2 to prepare an LSm2–8 complex, in which only LSm5 is NMR active (Figure 3B, olive). The resulting HMQC spectrum displays six distinct resonance signals that result from the six isoleucine residues that are present in the LSm5 protein.

In the examples shown above, we ensured that isotope labeling was restricted to a subset of the subunits in the complex, whereas the remaining subunits were NMR invisible. Interestingly, it is also possible to distribute different labeling schemes over the different subunits. We demonstrate this approach with an LSm2–8 complex that is uniformly ^{15}N labeled, LSm2, LSm3, LSm4, and LSm8 methionine labeled, and LSm5, LSm6, and LSm7 isoleucine labeled (Figure 4). Owing to the spectral separation of methionine and isoleucine methyl groups this approach allows for the independent and simultaneous monitoring of NMR parameters from different parts of a large complex.

NMR spectroscopic studies of large and asymmetric protein complexes suffer from significant challenges related to sample preparation and from spectral crowding owing to a high number of unique resonances. We have introduced a sequential co-expression strategy that tackles both issues simultaneously. Using the LSm1–7 and LSm2–8 complexes, we show that highly homogeneous samples that contain only one NMR active subunit can be readily prepared. Importantly,

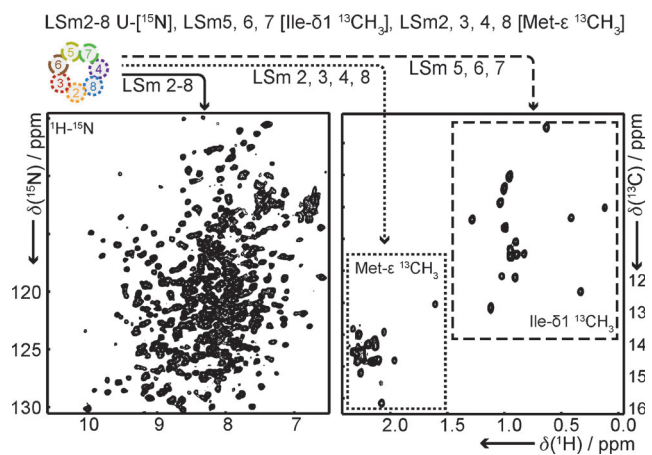


Figure 4. Mixed ^{15}N , ϵ - ^{13}C -methyl methionine and $\delta 1$ - ^{13}C -methyl isoleucine labeling of LSm2–8. The ^1H - ^{15}N TROSY spectrum (left) displays all subunits in the complex (see Figure 1B). LSm2, LSm3, LSm4, and LSm8 are labeled with methionine residues (see Figure 3A, that displays the methionine methyl groups for LSm1–7). LSm5, LSm6, and LSm7 are labeled with isoleucine residues (see Figure 3B).

tantly, our strategy is compatible with backbone and methyl group side-chain labeling. LEGO-NMR is thus suitable for the study of large asymmetric complexes including eukaryotic systems that are currently inaccessible to detailed NMR analysis. Interestingly, around 50% of the assemblies in the protein data bank (PDB) that contain three or more unique chains have been prepared in *E. coli*, indicating that our method is applicable to a wide variety of complexes.

Received: June 7, 2013

Published online: August 14, 2013

Keywords: isotopic labeling · LSm complex · methyl TROSY · NMR spectroscopy · protein structures

- [1] a) K. Pervushin, R. Riek, G. Wider, K. Wuthrich, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12366–12371; b) V. Tugarinov, P. M. Hwang, J. E. Ollerenshaw, L. E. Kay, *J. Am. Chem. Soc.* **2003**, *125*, 10420–10428.
- [2] M. Sattler, S. W. Fesik, *Structure* **1996**, *4*, 1245–1249.
- [3] V. Tugarinov, V. Kanelis, L. E. Kay, *Nat. Protoc.* **2006**, *1*, 749–754.
- [4] R. Sprangers, A. Velyvis, L. E. Kay, *Nat. Methods* **2007**, *4*, 697–703.
- [5] a) R. Sprangers, A. Gribun, P. M. Hwang, W. A. Houry, L. E. Kay, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 16678–16683; b) R. Sprangers, L. E. Kay, *Nature* **2007**, *445*, 618–622; c) H. Kato, H. van Ingen, B. R. Zhou, H. Feng, M. Bustin, L. E. Kay, Y. Bai, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 12283–12288; d) T. L. Religa, R. Sprangers, L. E. Kay, *Science* **2010**, *328*, 98–102; e) A. Velyvis, Y. R. Yang, H. K. Schachman, L. E. Kay, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 8815–8820; f) I. Gelis, A. M. Bonvin, D. Keramisanou, M. Koukaki, G. Gouridis, S. Karamanou, A. Economou, C. G. Kalodimos, *Cell* **2007**, *131*, 756–769; g) R. Rosenzweig, S. Moradi, A. Zarrine-Afsar, J. R. Glover, L. E. Kay, *Science* **2013**, *339*, 1080–1083; h) M. Audin, G. Dorn, S. A. Fromm, K. Reiss, S. Schütz, M. K. Vorländer, R. Sprangers,

- Angew. Chem.* **2013**, *125*, 8470–8474; *Angew. Chem. Int. Ed.* **2013**, *52*, 8312–8316.
- [6] W. Selleck, R. Howley, Q. Fang, V. Podolny, M. G. Fried, S. Buratowski, S. Tan, *Nat. Struct. Biol.* **2001**, *8*, 695–700.
- [7] a) D. S. Burz, K. Dutta, D. Cowburn, A. Shekhtman, *Nat. Methods* **2006**, *3*, 91–93; b) S. Zuger, H. Iwai, *Nat. Biotechnol.* **2005**, *23*, 736–740; c) M. Muona, A. S. Aranko, V. Raulinaitis, H. Iwai, *Nat. Protoc.* **2010**, *5*, 574–587; d) R. H. Kimura, E. R. Steenblock, J. A. Camarero, *Anal. Biochem.* **2007**, *369*, 60–70.
- [8] a) C. G. Miyada, L. Stoltzfus, G. Wilcox, *Proc. Natl. Acad. Sci. USA* **1984**, *81*, 4120–4124; b) L. M. Guzman, D. Belin, M. J. Carson, J. Beckwith, *J. Bacteriol.* **1995**, *177*, 4121–4130.
- [9] S. Wagner, M. M. Klepsch, S. Schlegel, A. Appel, R. Draheim, M. Tarry, M. Hogbom, K. J. van Wijk, D. J. Slotboom, J. O. Persson, J. W. de Gier, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 14371–14376.
- [10] S. Tharun, W. He, A. E. Mayes, P. Lennertz, J. D. Beggs, R. Parker, *Nature* **2000**, *404*, 515–518.
- [11] a) T. Achsel, H. Brahm, B. Kastner, A. Bachi, M. Wilm, R. Luhrmann, *Embo J.* **1999**, *18*, 5789–5802; b) J. Salgado-Garrido, E. Bragado-Nilsson, S. Kandels-Lewis, B. Seraphin, *Embo J.* **1999**, *18*, 3451–3462.
- [12] C. J. Wilusz, J. Wilusz, *Nat. Struct. Mol. Biol.* **2005**, *12*, 1031–1036.
- [13] a) M. Mund, A. Neu, J. Ullmann, U. Neu, R. Sprangers, *J. Mol. Biol.* **2011**, *414*, 165–176; b) B. Zaric, M. Chami, H. Remigy, A. Engel, K. Ballmer-Hofer, F. K. Winkler, C. Kambach, *J. Biol. Chem.* **2005**, *280*, 16066–16075.
- [14] A. E. Mayes, L. Verdone, P. Legrain, J. D. Beggs, *Embo J.* **1999**, *18*, 4321–4331.
- [15] A. K. Leung, K. Nagai, J. Li, *Nature* **2011**, *473*, 536–539.
- [16] I. Toro, S. Thore, C. Mayer, J. Basquin, B. Seraphin, D. Suck, *Embo J.* **2001**, *20*, 2293–2303.
- [17] M. A. Schumacher, R. F. Pearson, T. Moller, P. Valentin-Hansen, R. G. Brennan, *Embo J.* **2002**, *21*, 3546–3556.
- [18] C. Amero, M. Asuncion Dura, M. Noirclerc-Savoye, A. Perollier, B. Gallet, M. J. Plevin, T. Vernet, B. Franzetti, J. Boisbouvier, *J. Biomol. NMR* **2011**, *50*, 229–236.
- [19] M. C. Stoffregen, M. M. Schwer, F. A. Renschler, S. Wiesner, *Structure* **2012**, *20*, 573–581.
- [20] N. K. Goto, K. H. Gardner, G. A. Mueller, R. C. Willis, L. E. Kay, *J. Biomol. NMR* **1999**, *13*, 369–374.
- [21] I. Ayala, R. Sounier, N. Use, P. Gans, J. Boisbouvier, *J. Biomol. NMR* **2009**, *43*, 111–119.
- [22] C. Hesslinger, S. A. Fairhurst, G. Sawers, *Mol. Microbiol.* **1998**, *27*, 477–492.