Structural Biology

Deutsche Ausgabe: DOI: 10.1002/ange.201709274
Internationale Ausgabe: DOI: 10.1002/anie.201709274

Characterization of the Conjugation Pattern in Large Polysaccharide–Protein Conjugates by NMR Spectroscopy

Stefano Giuntini, Evita Balducci, Linda Cerofolini, Enrico Ravera, Marco Fragai, Francesco Berti,* and Claudio Luchinat**

Abstract: *Carbohydrate-based vaccines are among the safest and most effective vaccines and represent potent tools for prevention of life-threatening bacterial infectious diseases, like meningitis and pneumonia. The chemical conjugation of a weak antigen to protein as a source of T-cell epitopes generates a glycoconjugate vaccine that results more immunogenic. Several methods have been used so far to characterize the resulting polysaccharide–protein conjugates. However, a reduced number of methodologies has been proposed for measuring the degree of saccharide conjugation at the possible protein sites. Here we show that detailed information on large proteins conjugated with large polysaccharides can be achieved by a combination of solution and solid-state NMR spectroscopy. As a test case, a large protein assembly,* L*-asparaginase II, has been conjugated with Neisseria meningitidis serogroup C capsular polysaccharide and the pattern and degree of conjugation were determined.*

Encapsulated bacteria, such as *Neisseria meningitidis*, *Streptoccoccus pneumoniae* and *Haemophilus influenzae*, have polysaccharides on their exterior.[1] The capsular polysaccharides protect the microorganism from adverse environmental conditions, prevent the activation of the host immune system and improve adhesion to invade the host.[2–4] Capsular polysaccharides of pathogens have been used to produce vaccines that are effective in adults,[5,6] but result ineffective in children below 2 years of age:[7–9] being T-cell-independent antigens, they do not induce B-cell memory and do not stimulate the production of immunoglobulin G (IgG). However, a T-cell-dependent response (thus inducing B-cell memory) can be obtained by chemical coupling of polysaccharides to immunogenic proteins. Polysaccharide–protein conjugates are commonly obtained by covalent linkage of reactive precursors to the ε-amino groups of surface-exposed lysine residues or other reactive amino acids.[10] This conjugation strategy may lead to a highly random derivatization: the protein can be glycosylated at different sites and to different extents at each site (conjugation pattern); this in turn leads to a polydispersity in the resulting protein conjugates (Figure 1). To overcome this limitation, dedicated synthetic strategies are being developed.[11]

Together with the conjugation pattern of the polysaccharide, the preservation of the native protein structure may influence the immunogenicity and efficacy of the glycoconjugate vaccines.[12–14] Therefore, the structural characterization of the protein component, the identification of the conjugation sites, and the evaluation of the conjugation degree at different lysine sites are a compelling requirement towards structure-based development[15–18] of glycoconjugate vaccines. The presence of polysaccharide chains prevents the crystallization of the protein conjugates, and extends the molecular weight of the system beyond the limits of solution NMR spectroscopy,[19] similarly to what happens upon PEGylation.[20–22] Solution NMR spectroscopy is currently used to characterize both the polysaccharide[23,24] and the protein components separately before the conjugation.[25] Several solution-based analytical methods have been used so far to characterize the polysaccharide–protein conjugates.[13,26,27] SDS-PAGE and HPLC-SEC are often used to determine the size and hydrodynamic volume of the conjugates, and to estimate the number, but not the position, of the polysaccharide chains covalently bound to the protein. Also small-angle X-ray/neutron scattering and light scattering provide information on the size and shape of the coated proteins, but cannot map the conjugation sites.[28,29] More detailed information on the conjugation sites along the primary sequence of the protein can be achieved by mass spectrometry.[27] However, also for mass spectrometry the evaluation of the conjugation degree at the different sites in polydisperse populations of isomers is a challenge. Recently, we have shown that structural information on large PEGylated proteins, such as *E. coli* L-Asparaginase-II (138 kDa, ANSII, see the Supporting Information), can be obtained by solid-state NMR spectroscopy.[30] Several excellent reports on capsular polysaccharides, bacterial cell walls and their

[*] S. Giuntini, Dr. E. Ravera, Prof. M. Fragai, Prof. C. Luchinat
Department of Chemistry, University of Florence
Via della Lastruccia 3, 50019 Sesto Fiorentino (Italy)
E-mail: fragai@cerm.unifi.it
claudioluchinat@cerm.unifi.it
S. Giuntini, Dr. L. Cerofolini, Dr. E. Ravera, Prof. M. Fragai, Prof. C. Luchinat
Magnetic Resonance Center (CERM)
University of Florence and Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP)
Via L. Sacconi 6, 50019 Sesto Fiorentino (Italy)
Dr. E. Balducci, Dr. F. Berti
GSK Vaccines, Preclinical R&D
Via Fiorentina 1, 53100 Siena (Italy)
E-mail: francesco.x.berti@gsk.com
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.201709274.

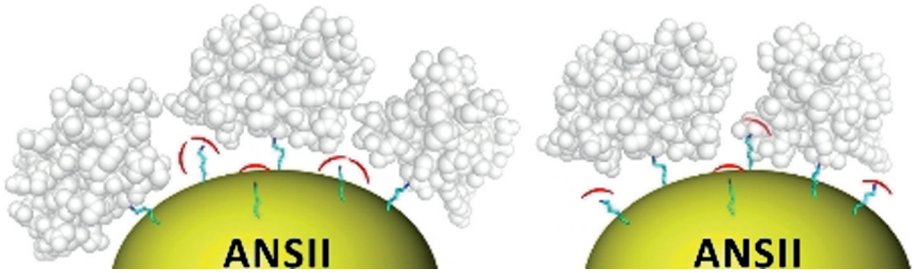

Figure 1. Schematic representation of the polydispersity in the population of ANSII-MenC conjugates. The steric hindrance (represented by red curves) of the already bound polysaccharide chains prevents further conjugation of the surrounding lysine residues.

maturation have been obtained by the same technique.[31–33] ANSII is not used as carrier protein in glycoconjugate vaccines, however, it yields highly resolved SSNMR spectra,[30] and approximately 74% of the resonances have been assigned (unpublished data). We have thus selected ANSII as a model system for this investigation. In the present work, ANSII has been conjugated to the capsular polysaccharide obtained from the meningococcal serogroup C (MenC). The polysaccharide is a homopolymer of α2-9-linked *N*-acetylneuraminic acid (NeuNAc) with partial O-acetylation at position C_7 and C_8, and has been coupled to lysine residues through an adipic acid linker. This work shows for the first time that highly glycosylated proteins are now an affordable target for structural characterization by SSNMR spectroscopy, and the pattern and the degree of conjugation can be identified.

ANSII-MenC was investigated through solution and solid-state NMR spectroscopy. The increase in hydrodynamic volume of the ANSII after the conjugation with MenC prevents the detection of the H^N signals of the backbone by standard solution ^{1}H-^{15}N TROSY-HSQC spectra. However, such a spectrum of ANSII-MenC [U- ^{2}H ^{13}C ^{15}N] displays about 10 new H^N signals in the region between 125 and 128 ppm on the ^{15}N dimension. As already observed in the case of PEGylation,[30] these signals correspond to the amide groups generated by functionalization of the ε-amino group of lysines (H_ζ-N_ζ) with the MenC polysaccharide moieties. The resonances of the non-functionalized lysines would appear at very low ^{15}N chemical shift values (around 30 ppm), but are usually wiped out by exchange. On the contrary, when they become amides, exchange is slowed down and the intrinsic flexibility of the sidechain justifies the appearance of the H_ζ-N_ζ cross-peaks. These signals are markers that can provide direct information on the number of lysine residues involved in the formation of new bonds with the MenC polysaccharide. However, the assignment of each H_ζ-N_ζ couple to the corresponding H-N of the backbone is a difficult task because of the high overlap between H_ζ-N_ζ signals, and the lack of connections between the two H-N couples across the side chain. For the solid-state NMR investigation, ANSII (either free or its MenC conjugate) was freeze-dried, packed into the magic-angle spinning (MAS) rotors, then rehydrated to restore the resolution.[30,34–36] The spectra collected on the pelleted ANSII-MenC conjugate (Figure 2) are largely superimposable to those of the native and PEGylated forms of the enzyme (see Figure S2 in the Supporting Information) making possible the assessment of the preservation of the protein fold after conjugation with the MenC polysaccharide.[37–40] The resonances of the protein residues were easily reassigned starting from the assignment of the native form of ANSII, by using a carbon-detection approach (Figure S3).[41] However, the spectral overlap due to the large number of protein resonances does not allow for straightforward identification of the lysine resonances. Therefore, selective labeling with ^{1}H, ^{13}C, ^{15}N lysine residues was applied in an otherwise U-^{2}H, ^{NA}C, ^{NA}N sample of ANSII, hereon [U-^{2}H, Lys-^{1}H, ^{13}C, ^{15}N]. Assignment of 19 out of 22 lysine sidechains of free and functionalized ANSII was achieved through such a combination of solution and solid-state NMR spectra (see the Supporting Information for details). The three unassigned lysine residues (K44, K51 and K229) are all solvent-exposed. Two spin systems could be identified in the SSNMR spectra, while the third is probably lost because of overlap.

The evaluation of the conjugation pattern was achieved by integrating solution and solid-state NMR data. The H_ζ-N_ζ cross-peaks corresponding to seven lysine residues could be assigned in the 2D ^{1}H-^{15}N TROSY-HSQC (Figure 3b, top). The intensity of the signals was used as a parameter to evaluate the functionalization degree, although the signal overlapping and the different dynamics of each lysine residue do not make possible a precise quantitative measurement. In this respect, the high signal intensity of K251, K101 and K273 is consistent with a high degree of conjugation, while for K208, K218, K273 and K310, the lower signal intensity indicates a lower derivatization. Complementary information has been obtained from the analysis of the changes in the

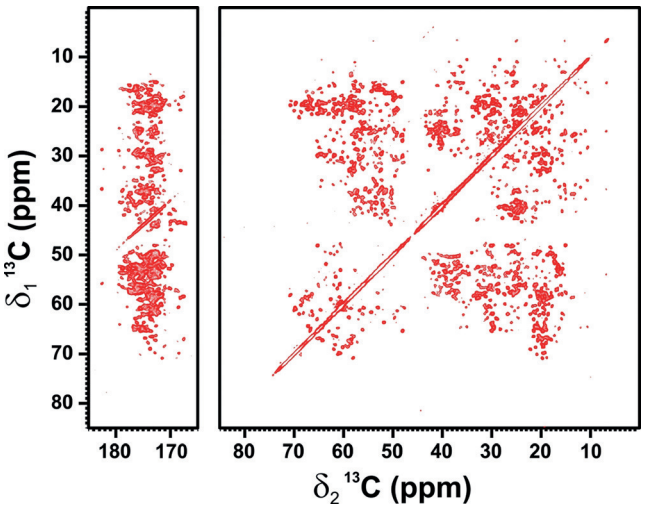

Figure 2. 2D ^{13}C-^{13}C DARR solid-state NMR spectra of ANSII-MenC acquired at 850 MHz and 290 K (mixing time 50 ms).

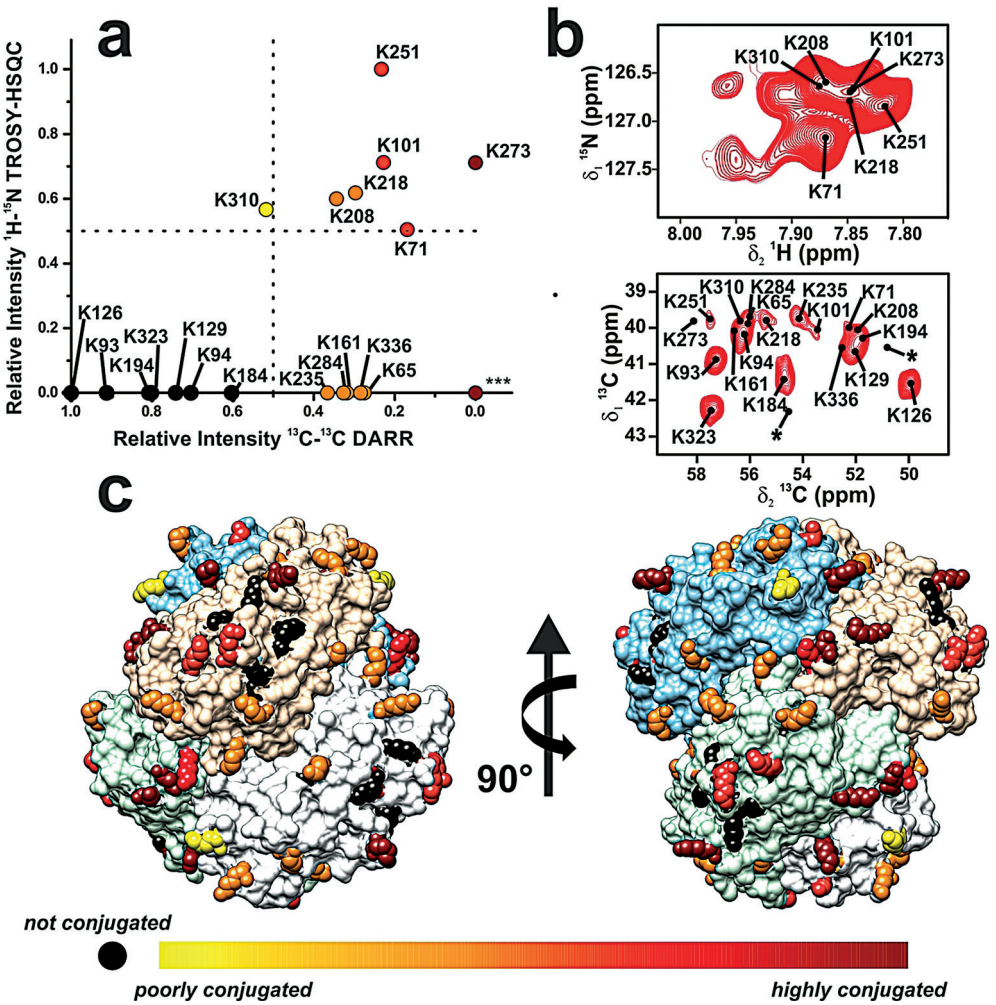

Figure 3. a) Relative signal intensities of H_ξ-N_ξ cross-peaks in the 2D ^{1}H-^{15}N TROSY-HSQC versus the relative signal intensities of the C_α-C_ε cross-peaks in 2D ^{13}C-^{13}C DARR of ANSII-MenC. The H_ξ-N_ξ cross-peak intensities have been related to the intensity of the H_ξ-N_ξ of K251, that is the most intense, while the C_α- C_ε peak intensities have been related to the intensity of the C_α- C_ε cross-peaks of K126 that is the most intense and isolated. The stars indicate the two spin systems for which the assignment is missing. b) Enlarged region of the 2D ^{1}H-^{15}N TROSY-HSQC of ANSII-MenC acquired at 950 MHz, 310 K (top). Enlargement of the region of the 2D ^{13}C-^{13}C DARR spectrum of the ANSII-MenC displaying the correlation between the C_α and the C_ε of the lysine sidechains (bottom). The assignment of each lysine is reported in the Figure. The stars indicate the two spin systems for which the assignment is missing. c) Surface representation of ANSII (PDB code: 3ECA) with the lysine residue color-coded according to the degree of conjugation, from dark red (highly conjugated) to yellow (poorly conjugated). The non-conjugated lysines are represented as black spheres.

relative intensity of the cross-peaks correlating the C_α and the C_ε of the lysine side chains in the 2D ^{13}C-^{13}C dipolar-aided rotary resonance (DARR) of the free and conjugated protein (Figure 3b, bottom). In particular, we have observed remarkable differences in the signal intensity among different groups of resonances in the spectrum of ANSII-MenC (Figure 3b, bottom). The disappearance of the C_α-C_ε cross-peaks in the spectra of the conjugated protein can be taken as a proof of an extensive conjugation of the lysines. A complete disappearance of the C_α-C_ε cross-peak was observed for two solvent-exposed unassigned lysines (belonging to K44 and/or K51 and/or K229), and for K273. Also, the signals of residues K65, K71, K101, K161, K208, K218, K235, K251, K284 and K336 exhibit a decrease in signal intensity. However, the corresponding cross-peaks are still visible in the spectrum of the conjugated protein, indicating the presence of a significant amount of unreacted residue (Figure 3b). Conversely, for the residues K93, K94, K126, K129, K184, K194, K323 bearing no solvent exposed N_ξ the decrease in intensity is negligible as expected for non-conjugated lysines.

For better evaluation of the conjugation degree, the information provided by solution and solid-state NMR spectroscopy is summarized in a simple graph reporting the relative signal intensity in the ^{13}C-^{13}C DARR and ^{1}H-^{15}N TROSY-HSQC (Figure 3a). The analysis of the conjugation degree obtained by combining the solution and solid-state data is summarized in Figure 3c where the lysines are represented as spheres and color-coded from dark red (highly conjugated) to yellow (poorly conjugated). The non-conjugated lysines are represented as black spheres.

In conclusion, solid-state NMR spectra of glycosylated proteins showing an extremely high quality were obtained,

making it possible to confirm the preservation of the protein fold and to achieve a residue-specific assignment of the spectral perturbations that occur upon glycosylation. A simple comparison of patterns of the H_ζ-N_ζ and Cα-Cε cross-peaks, respectively in solution and solid state, can be used to assess the preservation of the saccharide conjugation degree at the protein sites, for instance after major changes in the production process of the glycoconjugate vaccines. We have here applied a strategy that is based on a combination of solution and solid-state NMR experiments, which permits the identification of the residues bearing the polysaccharide chain and provides semi-quantitative information on the conjugation degree obtained by random chemical coupling. The traditional strategy based on the generation of several single point mutants of a protein to assign a specific amino acid may be risky under this circumstance to assign the single lysines and to evaluate their degree of conjugation, since the functionalization of a lysine may sterically hinder nearby lysine residues (Figure 1). The present approach allowed us to assign 19 lysines, to identify two additional spin systems while one is still missing or in overlap, without altering the reaction propensity of any of the lysines. The characterization of large polysaccharide–protein conjugates and highly glycosylated proteins is still a challenge in structural biology and an important goal for the development of vaccines and biologics, and this protocol has the possibility to provide key information for rational design of novel products with improved efficacy.

Experimental Section

Preparation of ANSII-MenC conjugates is described in the Supporting Information.

Solution NMR spectra were recorded at 310 K on Bruker AVANCEIII-HD NMR spectrometers operating at 900 and 950 MHz ^{1}H larmor frequency, equipped with triple resonance cryoprobes (see the Supporting Information). Protein samples were in water buffered solution [20 mM sodium phosphate, pH 7.5, 0.02% NaN_3, 0.1 mg mL^{-1} protease inhibitors (Pefabloc)].

SSNMR spectra were acquired on a Bruker AvanceIII spectrometer operating at 800 MHz (Bruker 3.2 mm Efree NCH probehead) or on a Bruker Avance III spectrometer operating at 850 MHz (Bruker 3.2 mm NCH DVT probehead). All spectra were recorded at 14 kHz MAS frequency and the sample temperature was kept at 290 K.

Samples of freeze-dried free ANSII (U-^{2}H, Lys-^{1}H, ^{13}C, ^{15}N ca. 10 mg of material) and ANSII-MenC (U-^{13}C-^{15}N ca. 14 mg of material; U-^{2}H, Lys-^{1}H, ^{13}C, ^{15}N ca. 5 mg of material) were packed into Bruker 3.2 mm zirconia rotors, then rehydrated until the resolution of the ^{13}C 1D SSNMR spectrum stopped changing.[30,34–36,42–44] Standard ^{13}C-detected two-dimensional spectra [^{13}C-^{13}C DARR,[45,46] ^{15}N-^{13}C NCA,[47] ^{15}N-^{13}C NCO[48]] were acquired according to reference.[49] These experiments last around 5 and 10 days for samples of free ANSII and ANSII-MenC, respectively. ^{1}H decoupling was set to 80 kHz using the swept-frequency two-pulse phase modulation scheme[50] and the optimal interscan delay was found to be 2.2 s.

Acknowledgements

This work has been supported by Fondazione Cassa di Risparmio di Firenze, EC Contracts pNMR number 317127; EU ESFRI Instruct (Core Centre CERM Italy); H2020-INFRAIA iNEXT number 653706; MIUR PRIN 2012SK7ASN; Regione Toscana (CERM-TT and BioEnable).

Conflict of interest

Francesco Berti is an employee of the GSK group of companies.

Keywords: carbohydrate-based vaccines · drug design · protein modifications · solid-state NMR spectroscopy · structural biology

How to cite: *Angew. Chem. Int. Ed.* **2017**, *56*, 14997–15001
Angew. Chem. **2017**, *129*, 15193–15197

[1] C. Schwechheimer, M. J. Kuehn, *Nat. Rev. Microbiol.* **2015**, *13*, 605–619.
[2] S. L. Reckseidler-Zenteno, **2012**, DOI: https://doi.org/10.5772/50116.
[3] I. S. Roberts, *Annu. Rev. Microbiol.* **1996**, *50*, 285–315.
[4] C. Whitfield, *Can. J. Microbiol.* **1988**, *34*, 415–420.
[5] P. H. Mäkelä, *Infection* **1984**, *12*, S72-S75.
[6] J. B. Robbins, R. Schneerson, G. Xie, L. Å. Hanson, M. A. Miller, *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 17871–17875.
[7] G. B. Lesinski, M. A. Westerink, *J. Microbiol. Methods* **2001**, *47*, 135–149.
[8] N. E. MacDonald, S. A. Halperin, B. J. Law, B. Forrest, L. E. Danzig, D. M. Granoff, *JAMA J. Am. Med. Assoc.* **1998**, *280*, 1685–1689.
[9] E. E. Adderson, *Springer Semin. Immunopathol.* **2001**, *23*, 387–403.
[10] P. Costantino, R. Rappuoli, F. Berti, *Expert Opin. Drug Discovery* **2011**, *6*, 1045–1066.
[11] C. Bello, C. F. W. Becker, *Bioorg. Med. Chem.* **2017**, https://doi.org/10.1016/j.bmc.2017.05.046.
[12] C. Jones, *An. Acad. Bras. Cienc.* **2005**, *77*, 293–324.
[13] N. Ravenscroft, G. Averani, A. Bartoloni, S. Berti, M. Bigio, V. Carinci, P. Costantino, S. D'Ascenzi, A. Giannozzi, F. Norelli, et al., *Vaccine* **1999**, *17*, 2802–2816.
[14] S. Crotti, H. Zhai, J. Zhou, M. Allan, D. Proietti, W. Pansegrau, Q.-Y. Hu, F. Berti, R. Adamo, *ChemBioChem* **2014**, *15*, 836–843.
[15] M. Scarselli, B. Aricò, B. Brunelli, S. Savino, F. Di Marcello, E. Palumbo, D. Veggi, L. Ciucchi, E. Cartocci, M. J. Bottomley, et al., *Sci. Transl. Med.* **2011**, *3*, 91ra62.
[16] F. Y. Avci, D. L. Kasper, *Annu. Rev. Immunol.* **2010**, *28*, 107–130.
[17] Z. Lai, J. R. Schreiber, *Vaccine* **2009**, *27*, 3137–3144.
[18] A. Kotecha, Q. Wang, X. Dong, S. L. Ilca, M. Ondiviela, R. Zihe, J. Seago, B. Charleston, E. E. Fry, N. G. A. Abrescia, et al., *Nat. Commun.* **2017**, *8*, 15408.
[19] B. Plesner, P. Westh, S. Hvidt, A. D. Nielsen, *J. Pharm. Biomed. Anal.* **2011**, *55*, 597–602.
[20] G. Cattani, L. Vogeley, P. B. Crowley, *Nat. Chem.* **2015**, *7*, 823–828.
[21] D. J. Hodgson, Y. Aubin, *J. Pharm. Biomed. Anal.* **2017**, *138*, 351–356.

[22] G. Digilio, L. Barbero, C. Bracco, D. Corpillo, P. Esposito, G. Piquet, S. Traversa, S. Aime, *J. Am. Chem. Soc.* **2003**, *125*, 3458 – 3470.
[23] N. Ravenscroft, S. D'Ascenzi, D. Proietti, F. Norelli, P. Costantino, *Dev. Biol.* **2000**, *103*, 35 – 47.
[24] C. Jones, *J. Pharm. Biomed. Anal.* **2005**, *38*, 840 – 850.
[25] C. Göbl, T. Madl, B. Simon, M. Sattler, *Prog. Nucl. Magn. Reson. Spectrosc.* **2014**, *80*, 26 – 63.
[26] N. Ravenscroft, J. X. Wheeler, C. Jones, *Bioanalysis* **2010**, *2*, 343 – 361.
[27] A. Bardotti, G. Averani, F. Berti, S. Berti, V. Carinci, S. D'Ascenzi, B. Fabbri, S. Giannini, A. Giannozzi, C. Magagnoli, et al., *Vaccine* **2008**, *26*, 2284 – 2296.
[28] M. Guttman, P. Weinkam, A. Sali, K. K. Lee, *Struct. Lond. Engl. 1993* **2013**, *21*, 321 – 331.
[29] Y. Liang, M. Guttman, T. M. Davenport, S.-L. Hu, K. K. Lee, *Biochemistry* **2016**, *55*, 2197 – 2213.
[30] E. Ravera, S. Ciambellotti, L. Cerofolini, T. Martelli, T. Kozyreva, C. Bernacchioni, S. Giuntini, M. Fragai, P. Turano, C. Luchinat, *Angew. Chem. Int. Ed.* **2016**, *55*, 2446 – 2449; *Angew. Chem.* **2016**, *128*, 2492 – 2495.
[31] P. Schanda, S. Triboulet, C. Laguri, C. M. Bougault, I. Ayala, M. Callon, M. Arthur, J.-P. Simorre, *J. Am. Chem. Soc.* **2014**, *136*, 17852 – 17860.
[32] H. Takahashi, I. Ayala, M. Bardet, G. De Paëpe, J.-P. Simorre, S. Hediger, *J. Am. Chem. Soc.* **2013**, *135*, 5105 – 5110.
[33] G. Zandomeneghi, K. Ilg, M. Aebi, B. H. Meier, *J. Am. Chem. Soc.* **2012**, *134*, 17513 – 17519.
[34] S. D. Kennedy, R. G. Bryant, *Biopolymers* **1990**, *29*, 1801 – 1806.
[35] K. Seidel, M. Etzkorn, H. Heise, S. Becker, M. Baldus, *ChemBioChem* **2005**, *6*, 1638 – 1647.
[36] M. Fragai, C. Luchinat, G. Parigi, E. Ravera, *J. Biomol. NMR* **2013**, *57*, 155 – 166.
[37] S. Balayssac, I. Bertini, K. Fälber, M. Fragai, S. Jehle, M. Lelli, C. Luchinat, H. Oschkinat, K. J. Yeo, *ChemBioChem* **2007**, *8*, 486 – 489.
[38] A. Loquet, L. Bousset, C. Gardiennet, Y. Sourigues, C. Wasmer, B. Habenstein, A. Schütz, B. H. Meier, R. Melki, A. Böckmann, *J. Mol. Biol.* **2009**, *394*, 108 – 118.
[39] D. Marulanda, M. L. Tasayco, A. McDermott, M. Cataldi, V. Arriaran, T. Polenova, *J. Am. Chem. Soc.* **2004**, *126*, 16608 – 16620.
[40] P. Turano, D. Lalli, I. C. Felli, E. C. Theil, I. Bertini, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 545 – 550.
[41] T. Wiegand, C. Gardiennet, R. Cadalbert, D. Lacabanne, B. Kunert, L. Terradot, A. Böckmann, B. H. Meier, *J. Biomol. NMR* **2016**, *65*, 79 – 86.
[42] D. L. Jakeman, D. J. Mitchell, W. A. Shuttleworth, J. N. S. Evans, *J. Biomol. NMR* **1998**, *12*, 417 – 421.
[43] R. B. Gregory, M. Gangoda, R. K. Gilpin, W. Su, *Biopolymers* **1993**, *33*, 513 – 519.
[44] J. Pauli, B. van Rossum, H. Förster, H. J. de Groot, H. Oschkinat, *J. Magn. Reson.* **2000**, *143*, 411 – 416.
[45] K. Takegoshi, S. Nakamura, T. Terao, *Chem. Phys. Lett.* **2001**, *344*, 631 – 637.
[46] K. Takegoshi, S. Nakamura, T. Terao, *J. Chem. Phys.* **2003**, *118*, 2325 – 2341.
[47] N. M. Loening, M. Bjerring, N. C. Nielsen, H. Oschkinat, *J. Magn. Reson.* **2012**, *214*, 81 – 90.
[48] M. Baldus, A. T. Petkova, J. Herzfeld, R. G. Griffin, *Mol. Phys.* **1998**, *95*, 1197 – 1207.
[49] A. Schuetz, C. Wasmer, B. Habenstein, R. Verel, J. Greenwald, R. Riek, A. Böckmann, B. H. Meier, *ChemBioChem* **2010**, *11*, 1543 – 1551.
[50] R. S. Thakur, N. D. Kurur, P. K. Madhu, *Chem. Phys. Lett.* **2006**, *426*, 459 – 463.

Manuscript received: September 7, 2017
Accepted manuscript online: October 10, 2017
Version of record online: October 19, 2017

