

VIP Very Important Paper



# Deriving Ligand Orientation in Weak Protein–Ligand Complexes by DEEP-STD NMR Spectroscopy in the Absence of Protein Chemical-Shift Assignment

Ridvan Nepravishta<sup>+</sup>,<sup>[a]</sup> Samuel Walpole<sup>+</sup>,<sup>[a]</sup> Louise Tailford,<sup>[b]</sup> Nathalie Juge,<sup>[b]</sup> and Jesus Angulo<sup>\*[a]</sup>

Differential epitope mapping saturation transfer difference (DEEP-STD) NMR spectroscopy is a recently developed powerful approach for elucidating the structure and pharmacophore of weak protein–ligand interactions, as it reports key information on the orientation of the ligand and the architecture of the binding pocket.<sup>[1]</sup> The method relies on selective saturation of protein residues in the binding site and the generation of a differential epitope map by observing the ligand, which depicts the nature of the protein residues making contact with the ligand in the bound state. Selective saturation requires knowledge of the chemical-shift assignment of the protein residues, which can be obtained either experimentally by NMR spectroscopy or predicted from 3D structures. Herein, we propose a simple experimental procedure to expand the DEEP-STD NMR methodology to protein–ligand cases in which the spectral assignment of the protein is not available. This is achieved by experimentally identifying the chemical shifts of the residues present in binding hot-spots on the surface of the receptor protein by using 2D NMR experiments combined with a paramagnetic probe.

The 3D structure of a small bioactive molecule in complex with its receptor gives atomic information that is essential for understanding the biological effects triggered by biomolecular recognition processes, as well as for the discovery and design of new drugs. Several techniques are used to achieve this aim, with X-ray crystallography, NMR spectroscopy and, more recently, cryo-electron microscopy being the most relevant approaches.

Recently, we developed the Differential EPitope mapping saturation transfer difference (DEEP-STD) NMR methodology for weak protein–ligand interactions,<sup>[1]</sup> as an extension of the general STD NMR method.<sup>[2]</sup> The DEEP-STD NMR technique allows the orientation of the ligand to be derived through differential selective saturation of different sets of key protein residues in the binding site. Namely, two STD NMR experiments are carried out, each one saturating different sets of protein residues, and the difference between the resulting spectra is quantified and mapped onto the ligand structure (differential epitope map). In order to perform the DEEP-STD NMR experiment accurately, it is of paramount importance to know beforehand the chemical shifts of the residues present in the binding site, in order to identify which set of residues to target, that is, choosing the irradiation frequencies to ensure that the selective saturation is applied on residues that are present in the binding site. Experimental chemical shifts can be obtained by NMR spectroscopy or derived from a 3D structure obtained by X-ray diffraction or by homology modelling.<sup>[3]</sup> The experimental DEEP-STD factors can be further combined with molecular docking and STD intensity predictions by CORCEMA-ST<sup>[4]</sup> in order to select the docking model that best fits the experimental data.

For when there is no chemical-shift assignment of the receptor protein, we here propose a general approach to experimentally identify the chemical shifts of those binding pocket resonances that relies on identifying ligand binding hot-spots on the surface of the protein (a ligand-binding hot-spot is a site on the surface of the protein that has a high probability for interaction with a ligand<sup>[5]</sup>) by using 2D NMR spectroscopy. This approach is compatible with the STD NMR technique, inexpensive and relatively fast, all of which should allow broad applicability.

The slow molecular tumbling of large proteins in solution is characterized by an overall correlation time expected to be in the range of 10<sup>-8</sup> seconds. However, the internal correlation time of the surface residues of globular proteins might be significantly shorter. As a result, residues in the core of the protein follow a slow-motion regime due to their low flexibility; this causes the signals to become broadened beyond detection. Conversely, the greater flexibility of surface residues causes them to follow a fast-motion regime that leads to cross-peaks in 2D NMR experiments that will be narrower and, hence, detectable. Therefore, these spectra are more likely to display signals from residues exposed on the surface or in very flexible regions of the protein.<sup>[6]</sup>

[a] Dr. R. Nepravishta,<sup>+</sup> S. Walpole,<sup>+</sup> Dr. J. Angulo  
School of Pharmacy, University of East Anglia  
Norwich Research Park, Norwich, Norfolk NR4 7TJ (UK)  
E-mail: j.angulo@uea.ac.uk

[b] Dr. L. Tailford, Dr. N. Juge  
The Gut Microbes and Health Institute Strategic Program  
Quadram Institute of Bioscience  
Norwich Research Park, Norwich, Norfolk NR4 7UA (UK)

[\*] These authors contributed equally to this work.

Supporting information and the ORCID identification numbers for the authors of this article can be found under <http://doi.org/10.1002/cbic.201800568>.

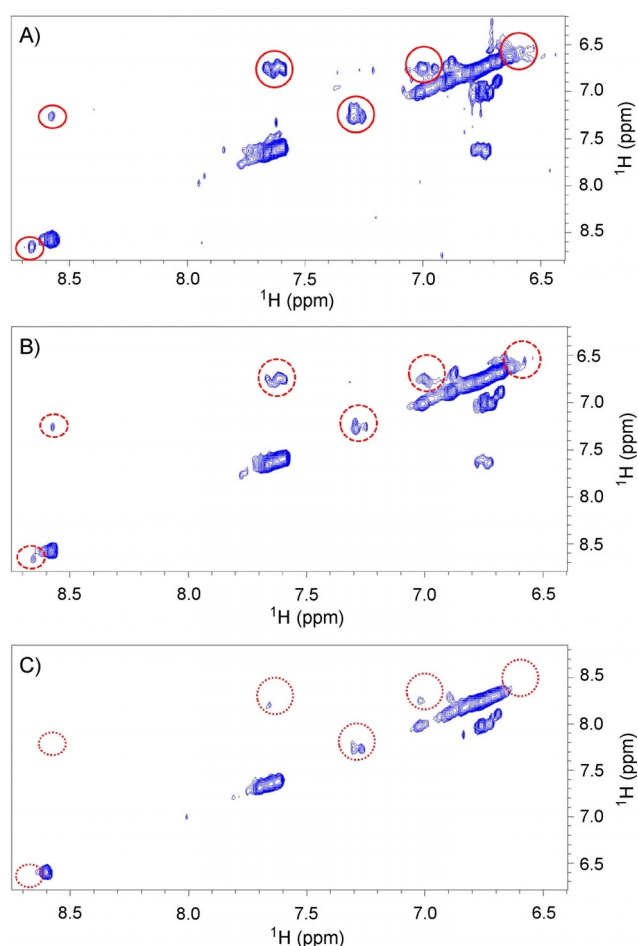
© 2019 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

This article is part of the young researchers' issue ChemBioTalents. To view the complete issue, visit: <http://chembiochem.org/chembiotalents>

By using 2D  $^1\text{H},^1\text{H}$  TOCSY experiments, hot-spots can be readily mapped by adding paramagnetic probes such as 4-hydroxy-2,2,6,6-tetramethylpiperidin-1-oxyl (TEMPOL) to a protein sample.<sup>[7]</sup> A decrease in the intensity of specific protein TOCSY crosspeaks, compared with the spectrum recorded without the paramagnetic probe, allows those residues interacting with the probe to be easily identified because they are affected by the paramagnetic relaxation enhancement (PRE) effect.<sup>[8,9]</sup> Previous PRE studies with TEMPOL have demonstrated the greater accessibility of this probe to proteins' specific binding sites rather than surface regions.<sup>[7,10,11]</sup> As a result of these experiments, the identified hot-spot resonances can then be considered as input frequencies for the DEEP-STD NMR experiments.

Binding hot-spots on the surface of proteins have previously been identified by NMR spectroscopy using paramagnetic probes along with classical molecular dynamics (MD) as well as by mixed MD using 5–50% probe/water mixtures.<sup>[7,12,13]</sup> For the development of our protocol, we combined this approach with DEEP-STD NMR using the structurally characterized catalytic domain (belonging to glycoside hydrolase family 33, GH33) of the intramolecular *trans*-sialidase (IT-sialidase) from the human gut symbiont *Ruminococcus gnavus*, RgNanH-GH33, in complex with 2,7-anhydro-Neu5Ac (PDB ID: 4X4A) as a benchmark.<sup>[14]</sup> RgNanH-GH33 is a 489-residue domain that can be considered out of the typical range for swift assignment and structure determination by NMR spectroscopy and was previously used to develop the DEEP-STD NMR approach.<sup>[1]</sup>

We first performed 2D homonuclear  $^1\text{H},^1\text{H}$  TOCSY experiments on RgNanH-GH33. The protein was exchanged in 10 mM  $[\text{D}_{11}]\text{Tris D}_2\text{O}$  buffer (pH 7.8) with 100 mM NaCl and used at a concentration of 1.2 mM. First, a 2D  $^1\text{H},^1\text{H}$  TOCSY reference spectrum of the protein was acquired, then two spectra in the presence of 2 and 12 mM of TEMPOL. The spectra obtained in the absence or in the presence of increasing concentrations of TEMPOL showed that the probe selectively interacts with some residues of the protein, as only some resonances in the spectra were significantly affected, as seen by a decreased intensity (Figure 1). The chemical shifts most affected by the presence of TEMPOL were at 0.6, 0.74, 1.06, 1.15, 1.26, 6.6, 6.74, 7.04,



**Figure 1.** Expansion of the aromatic spectral region of the TOCSY spectra of 1.2 mM RgNanH-GH33 A) alone or in the presence of B) 2 or C) 12 mM TEMPOL. Red circles highlight some resonances affected by the presence of TEMPOL (see also Figure S1).

7.57, 8.56 ppm (Figure S1 in the Supporting Information). These resonances, although lacking a specific assignment, are typical of aliphatic and aromatic amino acids, and we can exclude the presence of the NH resonances in the spectra as the protein was solvated in a  $\text{D}_2\text{O}$  buffer. The identified resonances from the TEMPOL-attenuated TOCSY spectra of RgNanH-GH33 were indeed in very good agreement with the predicted chemical shifts of key aliphatic and aromatic residues in the binding pocket of the enzyme (Ile258, Ile338, Val502, Thr557, Tyr525, Tyr677 and Trp698).<sup>[1]</sup>

To further validate our approach and exclude false positives (i.e., binding hot-spots outside the binding site), we carried out MD simulations, an approach successfully used in the past to identify ligand binding pockets for the development of small-molecule inhibitors.<sup>[15]</sup> Here, MD simulations were used to confirm the accessibility of TEMPOL to the specific binding pocket of RgNanH-GH33. To efficiently explore the configurational space of the RgNanH-GH33-TEMPOL system, three different MD approaches were considered: 1) long MD (1.0  $\mu\text{s}$ ) with a low concentration of TEMPOL (10 mM) starting from a random configuration of the system, 2) 16 independent repli-

Jesús Angulo obtained his doctorate at the Instituto de Investigaciones Químicas (CSIC/University of Seville) in 2002 under the supervision of Dr. Pedro Nieto. He moved to the group of Prof. Thomas Peters (University of Lübeck, Germany) to specialize in ligand-based NMR spectroscopy. He rejoined the IIQ in 2008 before joining the School of Pharmacy at the University of East Anglia in 2013, where he is a Senior Lecturer in NMR Spectroscopy.

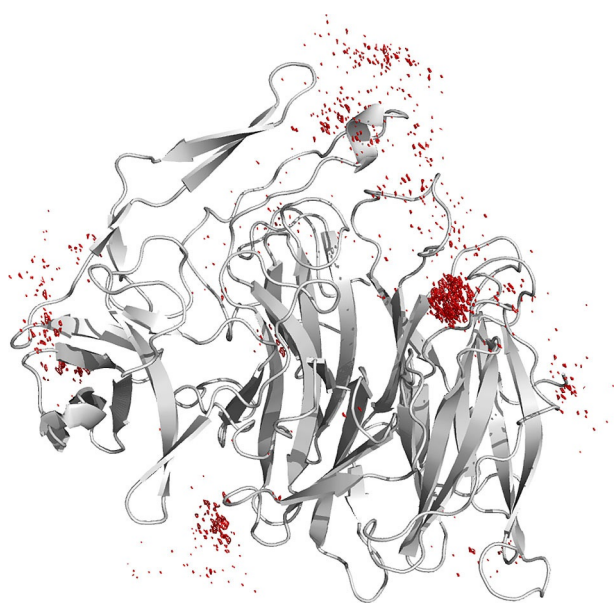


His research group focuses on structural and dynamics characterization of biologically active molecules and their complexes, with a particular focus on glycans, by using NMR spectroscopy and computational techniques.

cas of short MD (0.8  $\mu$ s, total 10 mM of TEMPOL), and 3) 16 independent short replica MD simulations with a high concentration (50% w/w) of TEMPOL in water, known as the MixMD approach.<sup>[12]</sup>

In each case, we first analysed the backbone RMSD of *RgNanH-GH33* for each trajectory, and showed that the presence of TEMPOL did not affect the structure of the protein, even for MixMD, as the average backbone RMSD was only approximately 1 Å (Figure S5). In the case of the long MD and the 16 short replicas, for which there were relatively few molecules of TEMPOL in the simulation, the interaction between TEMPOL and *RgNanH-GH33* was analysed by computing the contacts between TEMPOL and each residue in *RgNanH-GH33* over the course of each trajectory, in order to construct a fractional occupancy map for each residue. In the case of MixMD, in which there were 381 molecules of TEMPOL in the bounding box, the occupancy was measured by using a 0.5 Å grid to create bins for each TEMPOL molecule in each frame of the trajectory. The resulting 3D histograms were then visualized by means of the isomesh feature in PyMOL<sup>[16]</sup> by using a structure averaged over the whole trajectory for *RgNanH-GH33* (Figure 2).

Firstly, the long MD simulation containing a low concentration of TEMPOL did map several binding hot-spots, including the area of the binding site, but the outcome was dependent on the starting coordinates of the system. To overcome this issue, the same experiment was repeated with 16 different, independent short replicas of 50 ns each, according to a previous protocol.<sup>[7]</sup> In this case, although the mapping of the binding hot-spots was clear, the extension of sampling of the surface was not complete (Figure S5). In the MixMD approach, high concentrations of the probe enabled most of the protein surface to be mapped in a short time. In order to avoid biasing



**Figure 2.** Distribution of TEMPOL (red) around *RgNanH-GH33* (grey) as determined by MixMD. The distribution reveals a single large hot-spot at the known 2,7-anhydro-Neu5Ac binding site, thereby revealing that TEMPOL can act as a probe to selectively target residues of the binding site.

the system by the starting coordinates, 12 independent trajectories were run starting from different initial random configurations of the system. Figure 2 displays the average structure of the protein together with the occupancy grid, showing that the area of the known binding site of the 2,7-anhydro-Neu5Ac ligand is the major site for the interaction with the paramagnetic probe.

This result clearly excludes the presence of false positives and, more importantly, confirms that TEMPOL is selective for the binding site. Together with the TEMPOL/*RgNanH-GH33* interaction TOCSY experiment, these MD data build a solid argument for the use of TEMPOL-based TOCSY experiments to identify specific chemical shifts from residues in the binding pocket in order to perform the DEEP-STD NMR experiments.

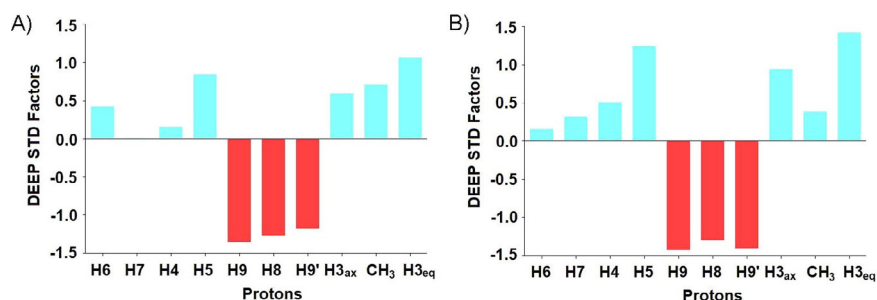
We then carried out the DEEP-STD NMR study with the frequencies identified by the TEMPOL approach. *RgNanH-GH33* (50  $\mu$ M) in the presence of 2,7-anhydro-Neu5Ac (1 mM) in 10 mM [D<sub>11</sub>]Tris D<sub>2</sub>O buffer (pH 7.8) with 100 mM NaCl at 298 K was saturated with a train of Gaussian pulses of 50 ms each for 0.75 s, centred on the chemical shifts of the binding hot-spots 0.6, 0.74, 1.06, 1.15, 1.26, 6.6, 6.74, 7.04, 7.57, 8.56 ppm.

As the absence of chemical-shift assignment of the protein prevents the irradiation frequencies to specific protons in the binding pocket from being known, we propose here a novel approach: instead of using a single pair of frequencies to determine the differential epitope map of the ligand (DEEP-STD map),<sup>[11]</sup> an averaging approach should be followed. First, the DEEP-STD factors for each experiment resulting from all the possible pairs of aliphatic and aromatic frequencies experimentally identified before are calculated (Figure S3). In our case, this resulted in 25 differential epitope maps. Secondly, all the obtained DEEP-STD factors are averaged to obtain a unique DEEP-STD map. This approach produces a more accurate depiction of the orientation and the nature of the amino acids surrounding the ligand in the binding pocket, particularly when no chemical shifts from the protein are available.

Figure 3A shows the experimental average DEEP-STD map of 2,7-anhydro-Neu5Ac binding to *RgNanH-GH33*.

The map highlights that CH<sub>3</sub>, H<sub>3<sub>ax</sub></sub>, H<sub>3<sub>eq</sub></sub>, H5 are oriented toward aliphatic residue protons, whereas H6, H7, H4 present little to no preferred orientation, and H9, H9' and H8 protons are oriented toward aromatic residues. This result is in excellent agreement with the crystal structure of the complex between 2,7-anhydro-Neu5Ac and *RgNanH-GH33*.<sup>[11]</sup> To confirm that our average DEEP-STD map is a reliable representation of the architecture of the binding pocket, we compared theoretical predictions of the average DEEP-STD map by using the CORCEMA-ST approach (see the Supporting Information).<sup>[4]</sup> The average DEEP-STD factors calculated by using CORCEMA-ST (Figure 3B) are in excellent agreement with the experimentally obtained ones. This result further validates our approach, demonstrating that the TEMPOL-based TOCSY approach is a reliable and powerful approach for identifying the suitable set of saturating frequencies to carry out DEEP-STD NMR studies in the absence of protein chemical-shift assignment.

Although here we have applied this approach to an enzyme with a polar binding pocket that favours H-bond interactions



**Figure 3.** A) Experimental average DEEP-STD factors of the binding of 2,7-anhydro-Neu5Ac to RgNanH-GH33 obtained from 25 differential epitope maps through the use of selective saturation at pairs of frequencies from the two sets experimentally determined by the TOCSY + TEMPOL experimental approach; set 1 (0.6, 0.74, 1.06, 1.15, 1.26 ppm) and set 2 (6.6, 6.74, 7.04, 7.57, 8.56 ppm; Figure S3). B) Theoretical average DEEP-STD factors calculated by using CORCE-MA-ST with ranges of saturation frequencies encompassing the experimental values in the calculation.

with TEMPOL, it has been previously shown that the interaction between proteins and TEMPOL can involve weak van der Waals forces, hydrogen bonding and hydrophobic interactions. Several authors have described interactions of TEMPOL with proteins such as ubiquitin, lysozyme, tendamistat, Sso7d, cyclophyllin, and BTPI,<sup>[10,17–21]</sup> which present different hydrophobicity/hydrophilicity profiles in their binding sites; this makes us confident of the general applicability of the new protocol to different types of protein target. Nonetheless, some expected limitations are that TEMPOL must bind the protein with low affinity so as to allow an easy interpretation of the spectra in the absence or in the presence of the paramagnetic agent, and it should not induce changes in the conformation of the protein upon binding, which would lead to misinterpretation of resonances to consider for DEEP-STD NMR or to conformational instability of the protein. The competition of TEMPOL with water tightly bound to the protein is also worth noting; in unfavourable cases this might prevent the probe from approaching the protein surface.<sup>[18]</sup>

In summary, we have developed a simple experimental procedure to expand the field of application of the DEEP-STD NMR methodology for deriving ligand orientation to protein–ligand cases in where the spectral assignment of the protein is not available, that is, when 1) a full NMR assignment is not possible, 2) the predicted chemical shifts from the structure are not in line with the experimental data (e.g., due to the dynamics of the protein, not accounted for in calculations on a static X-ray structure) or 3) chemical-shift assignments are lacking. Combining 2D TOCSY experiments in the absence/presence of a paramagnetic probe with the determination of an average DEEP-STD map by saturation at all the experimentally determined frequencies has been demonstrated to be a powerful approach to determine the type of protein residues most likely to interact with the ligand. The obtained information on the orientation of the ligand in the binding pocket of the protein opens several interesting applications of the DEEP-STD NMR methodology, for example in the hit-to-lead stage of drug discovery as in 3D-QSAR studies. Further, if combined with the  $K_D$  of the complex, the experimentally obtained averaged DEEP-STD factors could be used as descriptors to evaluate success or failure of hit modifications during the hit-to-lead stage.

## Acknowledgements

This work was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) through a New Investigator grant (BB/P010660/1) awarded to J.A. N.J. and L.E.T. acknowledge funding by the BBSRC Institute Strategic Programmes for Gut Health and Food Safety (BB/J004529/1) and Gut Microbes and Health (BB/R012490/1). We also acknowledge access to UEA's Faculty of Science Research Facilities. Data supporting this article are available as electronic supplementary files at <https://bit.ly/2BTwSER>, or upon request to the corresponding author.

## Conflict of Interest

The authors declare no conflict of interest.

**Keywords:** DEEP-STD • mixed molecular dynamics • NMR spectroscopy • TEMPOL

- [1] S. Monaco, L. E. Tailford, N. Juge, J. Angulo, *Angew. Chem. Int. Ed.* **2017**, *56*, 15289–15293; *Angew. Chem.* **2017**, *129*, 15491–15495.
- [2] a) M. Mayer, B. Meyer, *Angew. Chem. Int. Ed.* **1999**, *38*, 1784–1788; *Angew. Chem.* **1999**, *111*, 1902–1906; b) M. Mayer, B. Meyer, *J. Am. Chem. Soc.* **2001**, *123*, 6108–6117; c) A. Viegas, J. Manso, F. L. Nobrega, E. J. Cabrita, *J. Chem. Educ.* **2011**, *88*, 990–994; d) J. Angulo, P. M. Nieto, *Eur. Biophys. J.* **2011**, *40*, 1357–1369; e) A. Bhunia, S. Bhattacharjya, S. Chatterjee, *Drug Discovery Today* **2012**, *17*, 505–513.
- [3] B. Han, Y. Liu, S. Ginzinger, D. Wishart, *J. Biomol. NMR* **2011**, *50*, 43–57.
- [4] V. Jayalakshmi, N. R. J. Krishna, *J. Magn. Reson.* **2004**, *168*, 36–45.
- [5] B. S. Zerbe, D. R. Hall, S. Vajda, A. Whitty, D. Kozakov, *J. Chem. Inf. Model.* **2012**, *52*, 2236–2244.
- [6] M. A. Keniry, J. A. Carver in *Annual Reports on NMR Spectroscopy*, Vol. 48 (Ed.: G. A. Webb), Academic Press, London, **2002**, pp. 32–63.
- [7] N. Niccolai, E. Morandi, S. Gardini, V. Costabile, R. Spadaccini, O. Crescenzi, D. Picone, O. Spiga, A. Bernini, *Biochim. Biophys. Acta Proteins Proteomics* **2017**, *1865*, 201–207.
- [8] I. Solomon, *Phys. Rev.* **1955**, *99*, 559–566.
- [9] G. M. Clore, J. Iwahara, *Chem. Rev.* **2009**, *109*, 4108–4139.
- [10] S. W. Fesik, G. Gemmecker, E. T. Olejniczak, A. M. Petros, *J. Am. Chem. Soc.* **1991**, *113*, 7080–7082.
- [11] A. M. Petros, L. Mueller, K. D. Kopple, *Biochemistry* **1990**, *29*, 10041–10048.
- [12] P. M. Ung, P. Ghanakota, S. E. Graham, K. W. Lexa, H. A. Carlson, *Biopolymers* **2016**, *105*, 21–34.
- [13] K. W. Lexa, H. A. Carlson, *J. Chem. Inf. Model.* **2013**, *53*, 391–402.

- [14] L. E. Tailford, C. D. Owen, J. Walshaw, E. H. Crost, J. Hardy-Goddard, G. Le Gall, W. M. de Vos, G. L. Taylor, N. Juge, *Nat. Commun.* **2015**, *6*, 7624.
- [15] J. Seco, F. J. Luque, X. Barril, *J. Med. Chem.* **2009**, *52*, 2363–2371.
- [16] *The PyMOL Molecular Graphics System*, Version 1.7, Schrödinger, LLC.
- [17] N. Niccolai, O. Spiga, A. Bernini, M. Scarselli, A. Ciutti, I. Fiaschi, S. Chiellini, H. Molinari, P. A. Temussi, *J. Mol. Biol.* **2003**, *332*, 437–447.
- [18] M. Scarselli, A. Bernini, C. Segoni, H. Molinari, G. Esposito, A. M. Lesk, F. Laschi, P. Temussi, N. Niccolai, *J. Biomol. NMR* **1999**, *15*, 125–133.
- [19] N. Niccolai, R. Spadaccini, M. Scarselli, A. Bernini, O. Crescenzi, O. Spiga, A. Ciutti, D. Di Maro, L. Bracci, C. Dalvit, P. A. Temussi, *Protein Sci.* **2001**, *10*, 1498–1507.
- [20] A. Bernini, V. Venditti, O. Spiga, A. Ciutti, F. Prischi, R. Consonni, L. Zetta, I. Arosio, P. Fusi, A. Guagliardi, N. Niccolai, *Biophys. Chem.* **2008**, *137*, 71–75.
- [21] G. Pintacuda, G. Otting, *J. Am. Chem. Soc.* **2002**, *124*, 372–373.

---

Manuscript received: September 21, 2018

Accepted manuscript online: October 31, 2018

Version of record online: December 13, 2018