

## Assessing protein conformational landscapes: integration of DEER data in Maximum Occurrence analysis

Lucia Gigli,<sup>a</sup> Witold Andrałojć,<sup>ab</sup> Arina Dalaloyan,<sup>c</sup> Giacomo Parigi,<sup>a</sup> Enrico Ravera,<sup>a</sup> Daniella Goldfarb,<sup>c</sup> Claudio Luchinat<sup>\*a</sup>

a. CERM and Department of Chemistry “Ugo Schiff”, University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino (FI), Italy.

b. Current address: Institute of Bioorganic Chemistry, Polish Academy of Sciences, Z. Noskowskiego 12/14 61-704 Poznań, Poland

c. Department of Chemical and Biological Physics, Weizmann Institute of Science, Rehovot 76100, Israel.

**Understanding the properties of the conformational landscape of a biomolecule is of capital importance to understand its function. It is widely accepted that a statistical ensemble is far more representative than a single structure. The experimental data are the only handle on the conformational variability that the system is experiencing, but they are either time or ensemble averages. Since the available number of conformations largely outnumber the (independent) available experimental data, the latter can be equally well reproduced by a variety of ensembles. We have proposed the Maximum Occurrence (MaxOcc) approach to provide an upper bound of the statistical weight of each conformation. This method is expected to converge towards the true statistical weights by increasing the number of independent experimental datasets. In this paper we explore the ability of DEER (Double Electron Electron Resonance) data, which report on the distance distribution between two spin labels attached to a biomolecule, to restrain the MaxOcc values and its complementarity to previously introduced experimental techniques such as NMR and Small-Angle X-ray Scattering. We here study the case of Ca<sup>2+</sup> bound calmodulin (CaM) as a test case.**

### Introduction

The high level of complexity that is achieved in biological systems can be attributed, at least partly, to the considerable flexibility of biological macromolecules, such as nucleic acids and proteins. These molecules often perform their tasks by reorganizing their 3D structures, and such rearrangements can range from very small to very large, collectively defining the conformational space of the system under investigation.<sup>1-3</sup> Intuitively, the dimensionality of the conformational space depends on the number of degrees of freedom that the system can explore. Most often, at physiological pressure, temperature, pH, and salinity, a system can experience quite an extensive conformational variability, rather than being confined into a single state.<sup>4-6</sup> Therefore, a structural ensemble, i.e. a statistical ensemble of conformations together with their statistical weights, is more representative than a single structure. Pathological alterations of conditions, including mutations, may significantly alter the available conformational space.<sup>7, 8</sup>

The only handle on the conformational heterogeneity experienced by a biological system is through experiment. However, a sample comprises a number of macromolecules of the order of the Avogadro's constant, each occurring in a different state (ensemble average) and undergoing

several state transitions during the timeframe of the experiment (time average), with the overall result that a single experimental dataset is the outcome of the averaging over a potentially infinite number of conformations. Several ensemble reconstruction methods that combine experimental data and molecular models have been proposed to recover the conformational ensemble from the averaged experimental data.<sup>9,10</sup> Ensemble reconstruction is an ill-posed inverse problem: infinite different ensembles can reproduce the experimental results equally well. The proposed methods to circumvent the problem can be divided into two classes: those determining a large ensemble, following the maximum entropy principle, and those determining a minimal set of conformations that fit the data, in the spirit of the maximum parsimony principle. We have developed the Maximum Occurrence (MaxOcc) approach, where the MaxOcc of a given conformation is defined as the highest fraction of time the system can spend in that conformation.<sup>11-15</sup> This method provides an upper bound of the conformation's statistical weight and converges (decreases) towards the true statistical weight by increasing the number of independent experimental restraints. As such the MaxOcc method does not yield a single ensemble solution to the problem. MaxOcc can be calculated for any sterically possible conformation, without imposing limits on how many conformations are sampled.<sup>16</sup> We have observed that conformations with large MaxOcc are more frequently sampled in MD simulations,<sup>17</sup> and we have also found that conformations with high MaxOcc are likely selected by sample-and-select methods.<sup>18</sup>

As a test case, we have selected calmodulin (CaM), a small two domains protein, which is known to experience a sizeable interdomain mobility.<sup>19-21</sup> We have previously described its conformational heterogeneity through paramagnetic Residual Dipolar Couplings (pRDCs)<sup>13</sup> and Pseudocontact Shifts (PCSs) derived by NMR measurements,<sup>12</sup> and later included Small-Angle X-ray Scattering<sup>22</sup> (SAXS) and Paramagnetic Relaxation Enhancements (PREs) obtained from pNMR.<sup>23</sup>

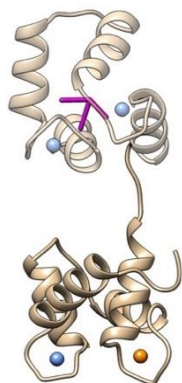
In this work, we demonstrate the effect of including multiple distance distributions obtained by DEER (Double Electron-Electron Resonance, called also Pulsed ELeCtron DOuble Resonance (PELDOR)) measurements in the MaxOcc approach. The DEER data are recorded as modulation traces that depend on distance distributions in the 1.6-8 nm range (and up to 16 nm for deuterated proteins).<sup>24</sup> They are usually carried out on frozen solutions at low temperatures and between pairs of usually identical spin labels that are attached at two different sites of a biomacromolecule.<sup>25</sup> Nitroxide spin labels are employed most frequently, but paramagnetic metal ions such as Cu(II),<sup>26, 27</sup> Mn(II) and Gd(III)<sup>28, 29</sup> can also be used. In this work we used DEER Gd(III)-Gd(III) distance measurements, which are best carried out at W-band (~95 GHz) where the Gd(III) EPR signals are intense,<sup>30-38</sup> and fit the DEER traces directly, not extracting the distances. Three different mutants were used, N53C-T110C, N53C-A103C, and T34C-T117C, where one Gd(III) label is situated in the C-terminal domain and the other in the N-terminal domain, specifically designed to probe the inter-domain distance distributions. We found that DEER is effective in restricting the allowed conformational space for calmodulin.

## Experimental

**Graphical representation of the model system.** In order to simplify the visualization of several conformations within the same figure, each conformation is schematized as follows: the N-terminal domain, considered fixed in an arbitrary reference frame,<sup>16</sup> is represented as a ribbon. The C-terminal domain of each conformation is represented as a triad of axes centered in its center of mass and color-coded according to its MaxOcc value. The orientation of the triad of axes is with respect to a reference position, which is arbitrarily selected. In the reference position the triad of axes corresponds to the axes of the PDB file (i.e.: (1,0,0;0,1,0;0,0,1)) if the C-terminal is in the extended conformation as in the 1CLL structure.<sup>39</sup> The calcium ions are shown as light blue spheres and the lanthanide ion replacing the calcium ion to obtain PCSs and RDCs as an orange sphere. Fig. 1 shows the representation of a single conformation, with the C-terminal domain shown both explicitly as a ribbon and implicitly as a triad of axis.

**Experimental data.** As experimental restraints we used NMR, SAXS and DEER data.

The paramagnetic NMR data (PCS and RDC for Tb<sup>3+</sup>, Tm<sup>3+</sup>, and Dy<sup>3+</sup>) have been taken from Bertini et al.<sup>12</sup> and SAXS data from Bertini et al.,<sup>22</sup> where CaM was dissolved in 20 mM MES and 400 mM KCl, pH 6.5.



**Fig. 1** Explanation of the representation of the different conformations: the N-terminal is represented as a solid ribbon, the C-terminal is represented both as a transparent ribbon and as a triad of axes placed in its center of mass and indicating the domain orientation. The paramagnetic metal ion is shown as an orange sphere and the calcium ions as light blue spheres.

The DEER datasets were acquired on samples of the mutants of human calmodulin N53C-T110C, N53C-A103C and T34C-T117C, whose expression and purification will be described elsewhere. The labeling of the mutants with Gd<sup>3+</sup> *DOTA-maleimide* (*DOTA-M*) was carried out as described earlier.<sup>40</sup> The final CaM concentration was 50  $\mu$ M (in a 20 mM phosphate buffer, pH 6.8) in a mixture of D<sub>2</sub>O/glycerol-d<sub>8</sub> (8/2, v:v). DEER traces were recorded on a home-built W-band spectrometer operating at 94.9 GHz at 10 K.<sup>41,42</sup> The dead-time free 4-pulse DEER sequence,  $(\pi/2)_{mw1}-\tau-(\pi)_{mw1}-t-\tau-(\pi)_{mw2}-T+\tau-t-(\pi)_{mw1}-T$ -echo was used. The observed pulse  $\pi/2$  and  $\pi$  microwave lengths were 15 ns and 30 ns, respectively, the pump pulse 15 ns,  $\tau$  350 ns and the repetition time 0.8 ms. The value of T varied depending on the experiment. In order to maximize the modulation depth, the DEER experiments were carried out by setting the pump pulse frequency on the maximum of the Gd<sup>3+</sup> spectrum. The detection pulses frequency was 100 MHz higher. As an estimate of the uncertainty in each DEER datapoint, for the N53C-T110C and N53C-A103C datasets we used the standard deviation of the consecutive differences between 60 and 160 ns. The data for T34C-T117C were obtained as average over three different measurements and the uncertainty was in this case taken as the largest difference between the three curves.

**Calculations.** We made use of a pre-defined pool of conformations over which the experimental observables were calculated. For the present work we have used a pool of sterically allowed conformations<sup>16, 22</sup> obtained by rigidly moving the C-terminal domain with respect to the N-terminal domain, around the backbone bonds of the residues 78-81 of the linker (identified as mobile by NMR relaxation measurements<sup>19</sup>). For each conformation we predicted and stored the values of the experimental observables that would have been measured if the system assumed solely this conformation. Paramagnetic NMR data (PCSs and RDCs) arise from the magnetic susceptibility anisotropy of terbium(III), dysprosium(III), and thulium(III) ions that were substituted for one of the calcium(II) ions in the N-terminal domain.

PCSs provide information on the coordinates of any given proton in the reference frame provided by the magnetic susceptibility anisotropy tensor  $\mathbb{X}$  (hence on the orientation of the C-terminal domain and its distance from the N-terminal domain) and are calculated as:

$$A_{ij} = \frac{1}{12\pi r_{(H_jM)_i}^5} \mathbf{r}_{(H_jM)_i}^T \mathbb{X} \mathbf{r}_{(H_jM)_i}$$

where  $\mathbf{r}_{(H_jM)_i}$  is the distance vector between the observed  $j$ -th nucleus  $H$  and the metal ion  $M$  for the  $i$ -th conformation.

RDCs provide information about the orientation of each  $NH$  pair in the reference frame provided by the magnetic susceptibility anisotropy (hence reflecting the orientation of the C-terminal domain) and are calculated as:

$$B_{ij} = \frac{k_{RDC}}{r_{(H_jN_j)_i}^5} \mathbf{r}_{(H_jN_j)_i}^T \mathbb{X} \mathbf{r}_{(H_jN_j)_i}$$

where the  $\mathbf{r}_{(H_jN_j)_i}$  is the distance vector between the two coupled nuclei  $N_j$  and  $H_j$ ,  $\mathbb{X}$  is the magnetic susceptibility tensor, and  $k_{RDC}$  is the constant (which depends on the residual alignment, in turn proportional to the square of the field):

$$k_{RDC} = -\frac{B_0^2 \gamma_N \gamma_H \hbar}{40\pi^2 k_B T}$$

where  $B_0$  is the magnetic field and  $T$  is the absolute temperature.

SAXS profiles provide information on the overall shape of the molecule and are calculated as:<sup>43</sup>

$$I_{ij} = \left\langle \left| A_{1i}(s_j) - \rho_0 A_{2i}(s_j) + (\rho_b - \rho_0) A_{3i}(s_j) \right|^2 \right\rangle_{\Omega}$$

where  $A_{1i}$  is the scattering amplitude from the  $i$ -th conformation in vacuum,  $A_{2i}$  from the excluded volume by the  $i$ -th conformation, and  $A_{3i}$  from the border layer by the  $i$ -th conformation;  $\Omega$  represents the solid angle in the reciprocal space,  $s_j$  is each increment in the momentum transfer,  $\rho_0$  and  $\rho_b$  are the average scattering densities of the solvent and of the hydration shell, respectively.

For predicting the DEER data, we have to explicitly consider the intrinsic mobility of the DOTA-M tag. To do so, we have modelled it using X-plor-NIH<sup>44,45</sup> as described in reference 46: we have generated 100 tag conformations for the N-terminal domain and 100 for the C-terminal domain in its reference position. For each pre-determined conformation, we rotated rigidly all the 100 tags conformations together with the C-terminal domain. Then we calculated the distance distribution  $p(r)$  (100 tags in the N-terminal x 100 tags in the C-terminal = 10000 distances) that pertains to each individual conformation, excluding all the tags with significant steric clashes.

For the sake of simplicity, we have assumed an ideal model for the DEER dipolar evolution, including the assumption that the dipolar couplings are in the weak coupling limit. With this assumption, the predicted distance distribution for the  $i$ -th conformer  $p(r_{ik})$  is transformed into the prediction matrix  $D_{ij}$ :<sup>47</sup>

$$D_{ij} = \sum_k p(r_{ik}) \int_0^{\pi/2} \cos \left[ \frac{1}{r_{ik}^3} \frac{\mu_0}{4\pi\hbar} g_e^2 \mu_B^2 (3 \cos^2 \theta - 1) t_j \right] \sin \theta d\theta$$

This corresponds to assuming that the modulation depth  $\lambda$  is 1 (vide infra).

The application of a more rigorous treatment which takes into account the pseudo-secular terms of the dipolar interaction along with the zero-field splitting<sup>48</sup> would yield a better description at short distances, hence would most likely reduce the MaxOcc values for closed conformations, which are already strongly discouraged by the PCSs (see considerations in reference 23); therefore, this is not expected to impact substantially on the MaxOcc values.

As previously mentioned, the determination of the MaxOcc of a single conformation corresponds to finding the largest weight that a selected conformer can have and still ensure a good agreement with the experimental data:<sup>5</sup>

$$\underset{\mathbf{w}}{\operatorname{argmin}}[\|\mathbb{P}\mathbf{w} - \mathbf{d}\|_2^2 + \xi_1(\|\mathbf{w}\|_1 - 1)^2 - \xi_2 w_c], s. t. \mathbf{w} \geq 0 \quad (1)$$

where  $\mathbb{P}$  is the predicted data matrix,  $\mathbf{w}$  is the weights vector,  $\mathbf{d}$  is the experimental data vector, and the values of the multipliers  $\xi_1$  and  $\xi_2$  are adjusted by an L-curve-like approach.<sup>49</sup> The numerical value of  $w_c$  is the Maximum Occurrence of the  $c$ -th conformation out of the entire conformational space. In this case MaxOcc is implemented as a Matlab script.

Since we are combining multiple datasets, we have used literature approaches to select the relative weight of each dataset in the least squares problems. For quantifying the agreement for paramagnetic NMR data, we have used the  $q$ -factor<sup>50</sup> which is defined as:

$$q = \frac{\|\mathbb{P}\mathbf{w} - \mathbf{d}\|_2}{\|\mathbf{d}\|_2}$$

where  $\mathbb{P}$  is the matrix of the predicted data ( $\mathbb{A}$  for PCSs and  $\mathbb{B}$  for RDCs),  $\mathbf{w}$  is the weights vector, and  $\mathbf{d}$  is the vector of the experimental data (for either PCSs or RDCs). Whereas for SAXS and DEER we have used the reduced  $\chi^2$ , defined as:<sup>43</sup>

$$\chi^2 = \frac{1}{N} \|(c\mathbb{P}\mathbf{w} - \mathbf{d}) \oslash \sigma\|_2^2$$

where  $N$  is the number of data points,  $\oslash$  is the Hadamard division,  $c$  is a scaling factor,  $\mathbf{w}$  is the weights vector,  $\mathbb{P}$  and  $\mathbf{d}$  are, respectively, the predicted matrix ( $\mathbb{I}$  for SAXS,  $\mathbb{D}$  for DEER) and the experimental data vector of the observable of interest, and  $\sigma$  is the standard deviation of the points of the curve. The scaling factor  $c$  is needed because, while paramagnetic NMR data magnitudes are fixed by the magnitude of the magnetic susceptibility anisotropy tensor and thus  $\mathbb{P}\mathbf{w}$  is directly comparable to  $\mathbf{d}$ , SAXS and DEER have arbitrary scales.

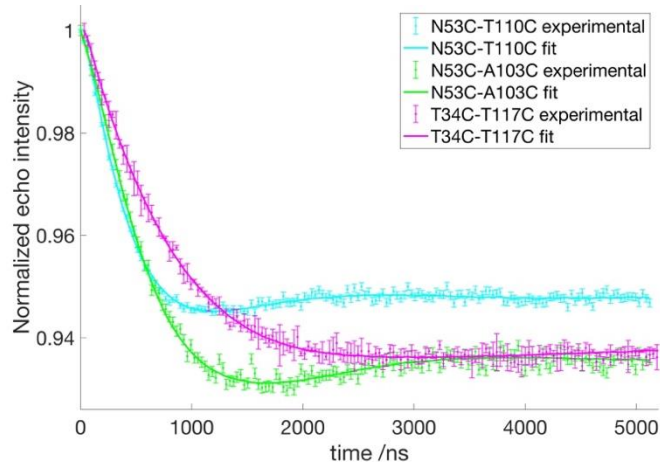
For SAXS  $c$  is calculated as:<sup>43</sup>

$$c = \frac{\|\mathbb{P}\mathbf{w}_2 \circ \mathbf{d}\|_1}{\|\mathbb{P}\mathbf{w}_2 \circ \mathbb{P}\mathbf{w}_2\|_1}$$

where  $\mathbb{P}$  and  $\mathbf{d}$  are, respectively, the predicted matrix and the experimental data vector of SAXS, and  $\circ$  is the Hadamard product. This formula corresponds to the zero in the first derivative in  $c$  of  $\|c\mathbb{P}\mathbf{w}_2 - \mathbf{d}\|_2$ .

In DEER an arbitrary scaling is needed to match the experimental modulation depth. To do so, we use the same approach used for SAXS, with the following modification: first we calculate  $\mathbb{M} = \mathbb{P} - \mathbb{J}$  and  $\mathbf{m} = \mathbf{d} - \mathbf{1}$  ( $\mathbb{J}$  and  $\mathbf{1}$  are, respectively, a matrix and a vector of ones), then the modulation depth, is calculated according to:  $\lambda = \frac{\|\mathbb{M}\mathbf{w}_2 \circ \mathbf{m}\|_1}{\|\mathbb{M}\mathbf{w}_2 \circ \mathbb{M}\mathbf{w}_2\|_1}$ .

This ensures a good quality fit (see Fig. 2).



**Fig. 2** W-band DEER data after background removal and their best fit (solid line) obtained with equation (1), imposing  $\xi_2 = 0$ .

## Results and discussion

A brief description of the MaxOcc approach and how it is implemented in practice is given in the calculations section. We have calculated the MaxOcc values for 1000 calmodulin conformations using different combinations of different experimental datasets, as listed in Table 1, and summarized the results in Fig. 3-5 and S1-S7. All panels (a) in Fig. 3-5 and S1-S7 show the MaxOcc values for the 1000 selected conformations, sorted according to the MaxOcc values calculated with dataset A only (MaxOcc<sub>A</sub>).

**Table 1** Combinations of datasets used for the calculations.

<i>Dataset</i>	<i>Id</i>	<i>Fig.</i>
NMR (3 datasets) = PCS + RDC	A	3
NMR + SAXS + DEER (3 datasets)	B	4
NMR + DEER	C	5,S8
SAXS	D	S1
DEER	E	S2
SAXS + DEER	F	S3
NMR + SAXS	G	S4
NMR + N53C-T110C	H	S5,S8
NMR + N53C-A103C	I	S6
NMR + T34C-T117C	L	S7
NMR + N53C-T110C + N53C-A103C	M	S8

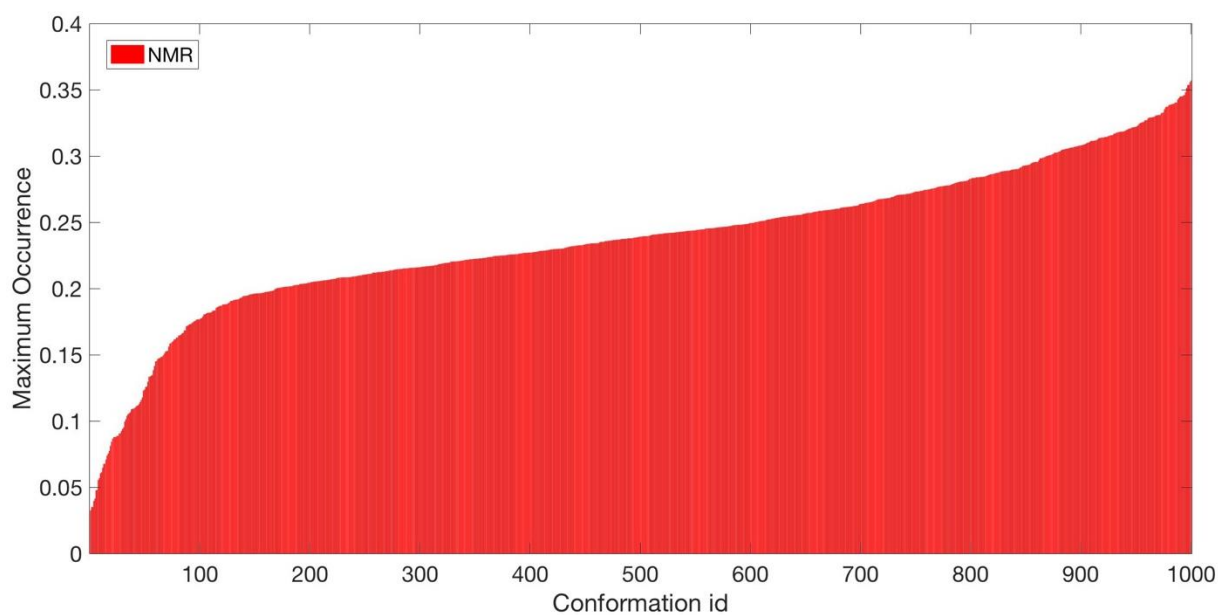
MaxOcc is defined as the maximum statistical weight that can be associated with any conformation and still preserve the agreement with the experimental restraints (see the Experimental section). Since this method does not provide the statistical weights themselves but only an upper bound, the addition of further restraints is expected to be reflected by a general decrease of Maximum Occurrence. Indeed, our results clearly indicate that the larger the number and the complementarity of the different experimental restraints, the higher is the decrease of the MaxOcc values. Fig. 4 shows that DEER data are extremely effective in differentiating the MaxOcc values for different conformers. In particular, our calculations indicate that the mutant N53C-A103C (Fig. S6) yields the best discrimination, followed by the N53C-T110C (Fig. S5), then by the T34C-T117C (Fig. S7), however it is important to stress that the more mutants are simultaneously used, the more the MaxOcc values decrease (Fig. S8). The improved discrimination of DEER over SAXS behavior is most likely linked to the fact that, as the tag is not positioned in the

center of the domain, the DEER information also depends on the orientational degrees of freedom. For most of the conformations there is a significant decrease of the MaxOcc value, whereas those that have high MaxOcc<sub>A</sub> values are relatively less affected (Fig. 4). This can be observed considering, for example, the conformations with MaxOcc values  $\geq 25\%$  (396) and  $30\%$  (131) calculated with only NMR restraints (Table 2). The 396 conformations with MaxOcc values  $\geq 25\%$  become 161 and 68 when either SAXS or DEER data are included and 50 when both SAXS and DEER data are added. Analogously, the 131 conformations with MaxOcc values  $\geq 30\%$  become 24 or 18 when either SAXS or DEER data are included and 9 (explicitly represented in Fig. S9) when both SAXS and DEER data are added.

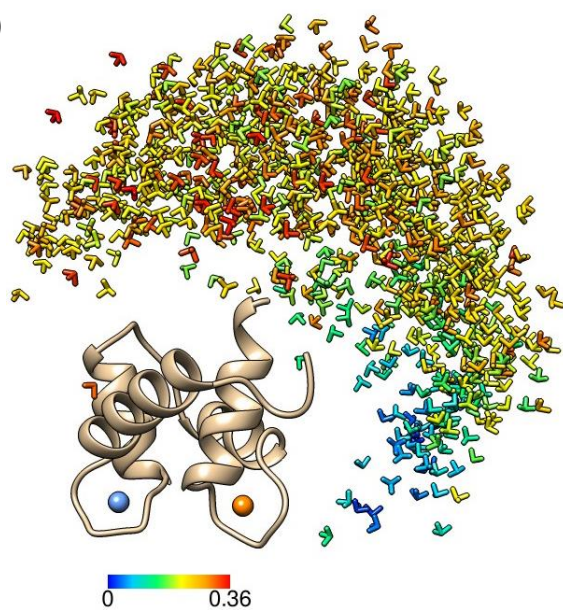
**Table 2** Number of conformations (out of 1000) with MaxOcc values  $\geq 25\%$  and  $30\%$ .

<i>Dataset</i>	<i>Id</i>	<i>MaxOcc</i> $\geq 25$ %	<i>MaxOcc</i> $\geq 30$ %
NMR	A	396	131
NMR + SAXS	G	161	24
NMR + DEER	C	68	18
NMR + SAXS + DEER	B	50	9

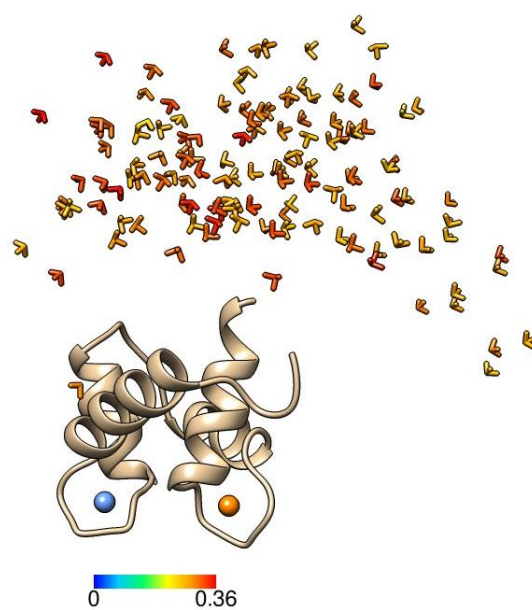
(a)



(b)



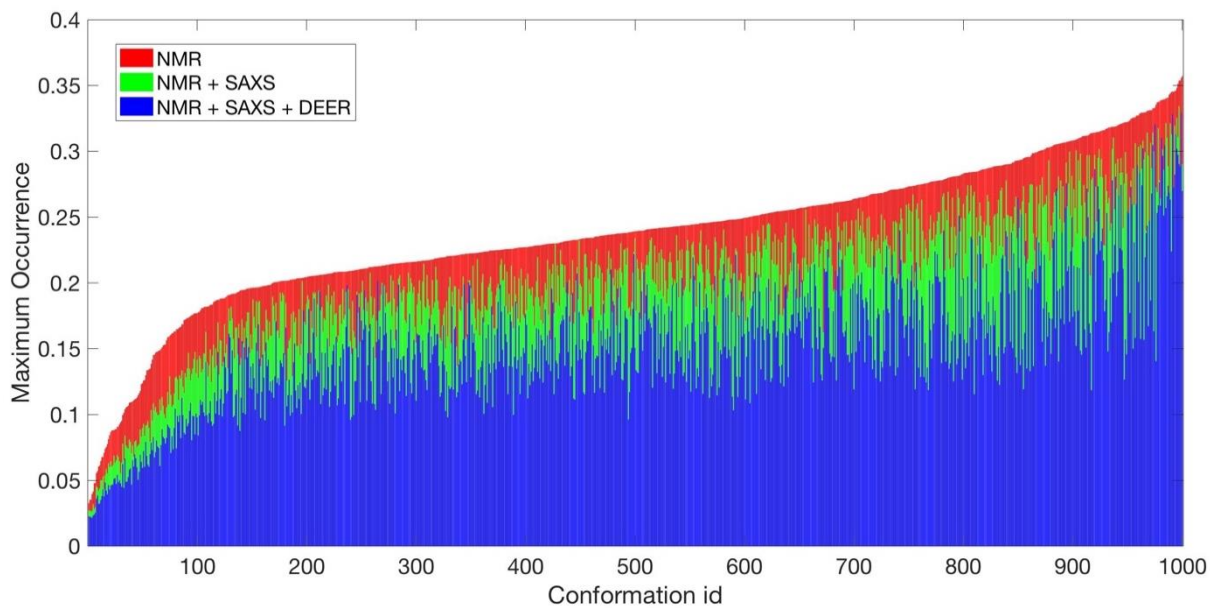
(c)



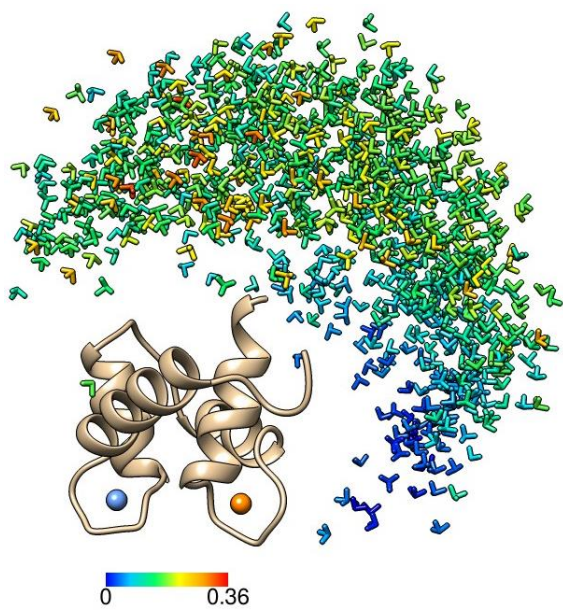
**Fig. 3** (a) MaxOcc<sub>A</sub> values for 1000 conformations, sorted by ascending value. This order will be used for all the other MaxOcc calculations to allow for the comparison between the different datasets; (b) representation of the 1000 MaxOcc<sub>A</sub> values of the selected conformations. Each conformation is color coded according to the MaxOcc<sub>A</sub> from low (< 5%, blue) to high (> 34%, red) MaxOcc values; (c) conformations with MaxOcc<sub>A</sub> ≥ 30%.



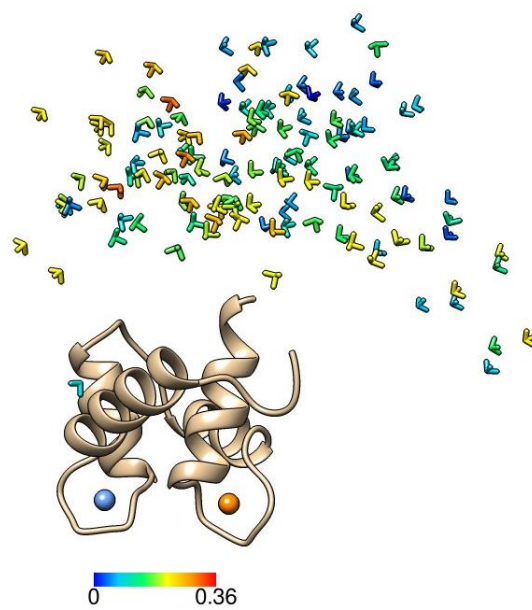
(a)



(b)

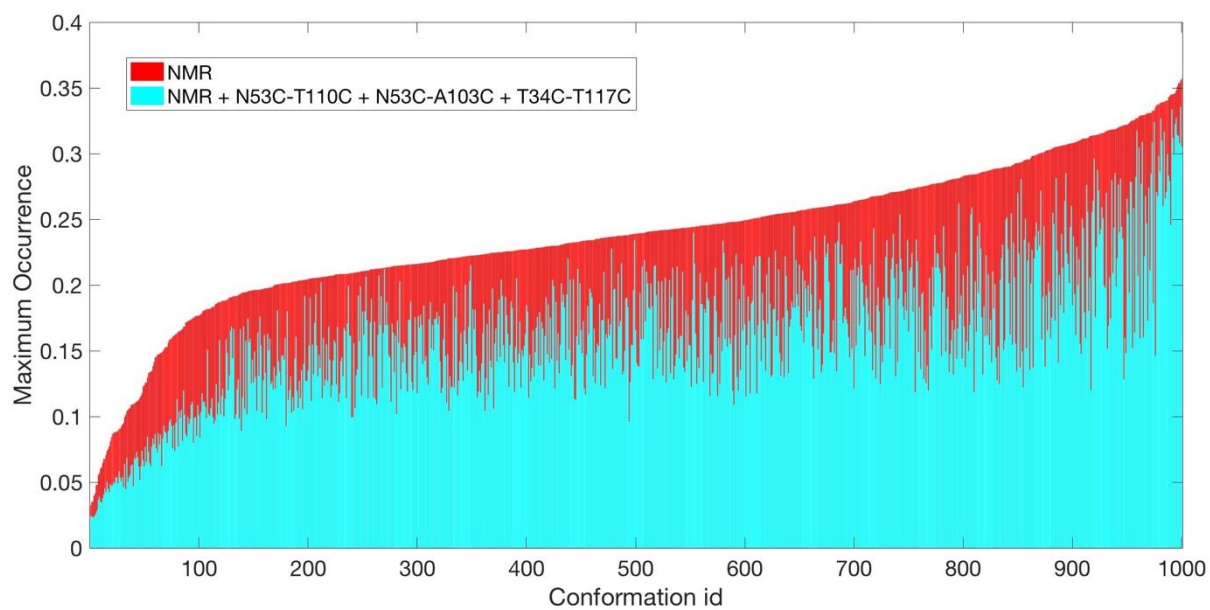


(c)

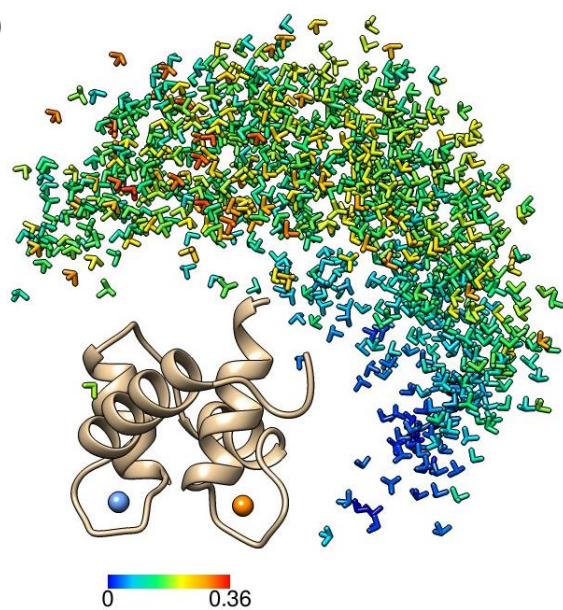


**Fig. 4** (a) MaxOcc<sub>A</sub>, MaxOcc<sub>G</sub>, and MaxOcc<sub>B</sub> values for 1000 conformations, sorted by ascending MaxOcc<sub>A</sub>; (b) representation of the 1000 MaxOcc<sub>B</sub> values of the selected conformations (color coding as in Fig. 3); (c) conformations with MaxOcc<sub>A</sub> ≥ 30% color-coded according to MaxOcc<sub>B</sub>.

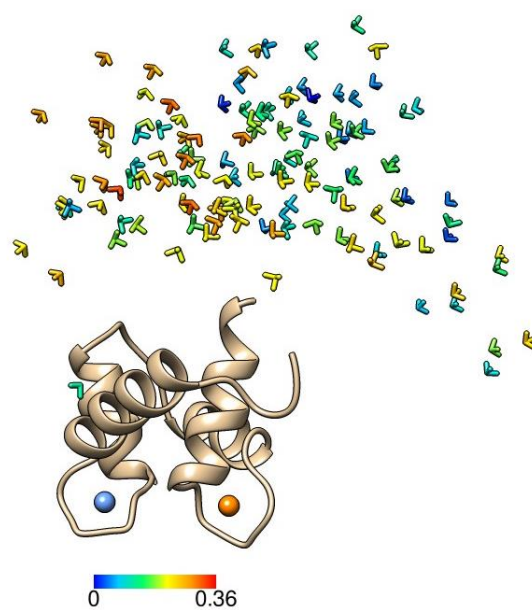
(a)



(b)

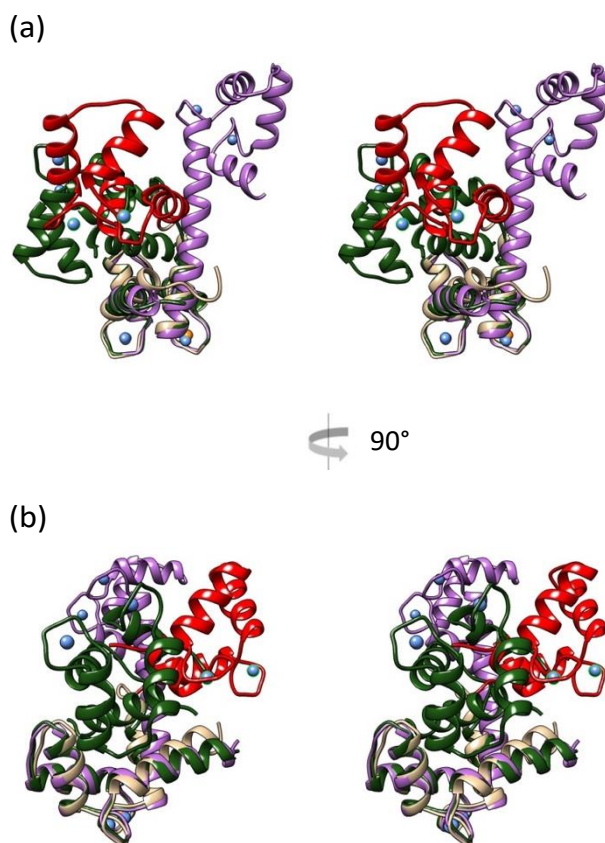


(c)



**Fig. 5** (a) MaxOcc<sub>C</sub> values for 1000 conformations, sorted by ascending MaxOcc<sub>A</sub>; (b) representation of the 1000 MaxOcc<sub>C</sub> values of the selected conformations (color coding as in Fig. 3); (c) conformations with MaxOcc<sub>A</sub> ≥ 30% color-coded according to MaxOcc<sub>C</sub>

One representative conformation is shown in Fig. 6, compared with two X-ray determined structures: the extended, dumbbell conformation (1CLL)<sup>39</sup> and one CaM-peptide complex closed in the canonical orientation (1YR5).<sup>51</sup> The conformations that retain the highest MaxOcc values are, in general, open, with the C-terminal bent towards the first helix of the N-terminal domain. Although open, they do not correspond to the classical elongated, dumbbell-shaped conformation, nor to the closed canonical conformations. Furthermore, they do not correspond to averages between the two X-ray extremes.



**Fig. 6** Comparison (cross-eye stereo) of the conformation with the highest MaxOcc value (red) with two X-ray determined structures (1CLL in violet and 1YR5 in green). The paramagnetic metal ion is shown as an orange sphere and the calcium ions as light blue spheres.

## Conclusions

In conclusion, our results confirm that DEER is highly complementary to NMR and SAXS in better defining the conformations that can be sampled for longer times, compatibly with the experimental data. As expected, the MaxOcc values are decreased by the addition of further experimental dataset, but it is remarkable to observe how the different datasets consistently favour a subset of conformations that share common domain arrangements.

## Acknowledgements

The authors thank Gottfried Otting for many helpful discussions. This work has been supported by the EC contracts #317127 (pNMR), #653706 (iNext) and #675858 (West-Life) and the support and the use of resources of Instruct-ERIC. D. G. acknowledges the support of the Israel Science Foundation (ISF grant No. 334/14). D. G. holds the Erich Klieger Professorial Chair in Chemical Physics.

## References

1. E. Z. Eisenmesser, D. A. Bosco, M. Akke and D. Kern, *Science*, 2002, **295**, 1520-1523.
2. Y. J. Huang and G. T. Montelione, *Nature*, 2005, **438**, 36-37.
3. M. Fragai, C. Luchinat and G. Parigi, *Acc. Chem. Res.*, 2006, **39**, 909-917.
4. N. J. Anthis and G. M. Clore, *Q Rev Biophys*, 2015, **48**, 35-116.
5. J. Schilder and M. Ubbink, *Curr. Opin. Struct. Biol.*, 2013, **23**, 911-918.
6. G. Schreiber and A. E. Keating, *Curr. Opin. Struct. Biol.*, 2011, **21**, 50-61.
7. D. A. Ragland, E. A. Nalivaika, M. N. Nalam, K. L. Prachanronarong, H. Cao, R. M. Bandaranayake, Y. Cai, N. Kurt-Yilmaz and C. A. Schiffer, *J. Am. Chem. Soc.*, 2014, **136**, 11956-11963.
8. M. Halvorsen, J. S. Martin, S. Broadaway and A. Laederach, *PLoS Genet.*, 2010, **6**, e1001074.
9. M. Bonomi, G. T. Heller, C. Camilloni and M. Vendruscolo, *Curr. Opin. Struct. Biol.*, 2017, **42**, 106-116.
10. E. Ravera, L. Sgheri, G. Parigi and C. Luchinat, *Phys. Chem. Chem. Phys.*, 2016, **18**, 5686-5701.
11. L. Sgheri, *Inv. Probl.*, 2010, **26**, 035003-035003-035019.
12. I. Bertini, Y. K. Gupta, C. Luchinat, G. Parigi, M. Peana, L. Sgheri and J. Yuan, *J. Am. Chem. Soc.*, 2007, **129**, 12786-12794.
13. M. Longinetti, C. Luchinat, G. Parigi and L. Sgheri, *Inv. Probl.*, 2006, **22**, 1485-1502.
14. R. J. Gardner, M. Longinetti and L. Sgheri, *Inv. Probl.*, 2005, **21**, 879-898.
15. I. Bertini, M. Longinetti, C. Luchinat, G. Parigi and L. Sgheri, *J. Biomol. NMR*, 2002, **22**, 123-136.
16. I. Bertini, L. Ferella, C. Luchinat, G. Parigi, M. V. Petoukhov, E. Ravera, A. Rosato and D. I. Svergun, *J. Biomol. NMR*, 2012, **53**, 271-280.
17. W. Andralojc, E. Ravera, L. Salmon, G. Parigi, H. M. Al-Hashimi and C. Luchinat, *Phys. Chem. Chem. Phys.*, 2016, **18**, 5743-5752.
18. W. Andralojc, K. Berlin, D. Fushman, C. Luchinat, G. Parigi, E. Ravera and L. Sgheri, *J. Biomol. NMR*, 2015, **62**, 353-371.
19. G. Barbato, M. Ikura, L. E. Kay, R. W. Pastor and A. Bax, *Biochemistry*, 1992, **31**, 5269-5278.
20. I. Bertini, C. Del Bianco, I. Gelis, N. Katsaros, C. Luchinat, G. Parigi, M. Peana, A. Provenzani and M. A. Zoroddu, *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 6841-6846.
21. J. J. Chou, S. Li, C. B. Klee and A. Bax, *Nat. Struct. Biol.*, 2001, **8**, 990-997.
22. I. Bertini, A. Giachetti, C. Luchinat, G. Parigi, M. V. Petoukhov, R. Pierattelli, E. Ravera and D. I. Svergun, *J. Am. Chem. Soc.*, 2010, **132**, 13553-13558.
23. I. Bertini, C. Luchinat, M. Nagulapalli, G. Parigi and E. Ravera, *Phys. Chem. Chem. Phys.*, 2012, **14**, 9149-9156.
24. T. Schmidt, M. A. Walti, J. L. Baber, E. J. Hustedt and G. M. Clore, *Angew. Chem. Int. Ed.*, 2016, **55**, 15905-15909.
25. G. Jeschke and Y. Polyhach, *Phys. Chem. Chem. Phys.*, 2007, **9**, 1895-1910.
26. Z. Y. Yang, D. Kise and S. Saxena, *J. Phys. Chem. B*, 2010, **114**, 6165-6174.
27. T. F. Cunningham, M. R. Putterman, A. Desai, W. S. Horne and S. Saxena, *Angew. Chem. Int. Ed.*, 2015, **54**, 6330-6334.
28. A. Feintuch, G. Otting and D. Goldfarb, *Methods Enzymol.*, 2015, **563**, 415-457.
29. D. Goldfarb, *Phys. Chem. Chem. Phys.*, 2014, **16**, 9685.

30. A. P. Welegedara, Y. Yang, M. D. Lee, J. D. Swarbrick, T. Huber, B. Graham, D. Goldfarb and G. Otting, *Chem. Eur. J.*, 2017, **23**, 11694-11702.
31. E. H. Abdelkader, M. D. Lee, A. Feintuch, M. R. Cohen, J. D. Swarbrick, G. Otting, B. Graham and D. Goldfarb, *J. Phys. Chem. Lett.*, 2015, **6**, 5016-5021.
32. D. T. Edwards, T. Huber, S. Hussain, K. M. Stone, M. Kinnebrew, I. Kaminker, E. Matalon, M. S. Sherwin, D. Goldfarb and S. I. Han, *Structure*, 2014, **22**, 1677-1686.
33. A. Potapov, H. Yagi, T. Huber, S. Jergic, N. E. Dixon, G. Otting and D. Goldfarb, *J. Am. Chem. Soc.*, 2010, **132**, 9040-9048.
34. H. Yagi, D. Banerjee, B. Graham, T. Huber, D. Goldfarb and G. Otting, *J. Am. Chem. Soc.*, 2011, **133**, 10418-10421.
35. E. Matalon, T. Huber, G. Hagelueken, B. Graham, V. Frydman, A. Feintuch, G. Otting and D. Goldfarb, *Angew. Chem. Int. Ed.*, 2013, **52**, 11831-11834.
36. M. Gordon-Grossman, I. Kaminker, Y. Gofman, Y. Shai and D. Goldfarb, *Phys. Chem. Chem. Phys.*, 2011, **13**, 10771-10780.
37. E. H. Abdelkader, A. Feintuch, X. Yao, L. A. Adams, L. Aurelio, B. Graham, D. Goldfarb and G. Otting, *Chem. Commun.*, 2015, **51**, 15898-15901.
38. N. Manukovsky, V. Frydman and D. Goldfarb, *J. Phys. Chem. B*, 2015, **119**, 13732-13741.
39. R. Chattopadhyaya, W. E. Meador, A. R. Means and F. A. Quijoco, *J. Mol. Biol.*, 1992, **228**, 1177-1192.
40. A. Martorana, G. Bellapadrona, A. Feintuch, G. E. Di, S. Aime and D. Goldfarb, *J. Am. Chem. Soc.*, 2014, **136**, 13458-13465.
41. D. Goldfarb, Y. Lipkin, A. Potapov, Y. Gorodetsky, B. Epel, A. M. Raitsimring, M. Radoul and I. Kaminker, *J. Magn. Reson.*, 2008, **194**, 8-15.
42. F. Mentink-Vigier, A. Collauto, A. Feintuch, I. Kaminker, V. T. Le and D. Goldfarb *J. Magn. Reson.*, 2013, **236**, 117-125.
43. D. I. Svergun, C. Barberato and M. H. J. Koch, *J. Appl. Crystallogr.*, 1995, **28**, 768-773.
44. L. Banci, I. Bertini, G. Cavallaro, A. Giachetti, C. Luchinat and G. Parigi, *J. Biomol. NMR*, 2004, **28**, 249-261.
45. C. D. Schwieters, J. Kuszewski, N. Tjandra and G. M. Clore, *J. Magn. Reson.*, 2003, **160**, 65-73.
46. N. J. Anthis, M. Doucleff and G. M. Clore, *J. Am. Chem. Soc.*, 2011, **133**, 18966-18974.
47. G. Jeschke, *Annu. Rev. Phys. Chem.*, 2012, **63**, 419-446.
48. N. Manukovsky, A. Feintuch, I. Kuprov and D. Goldfarb, *J. Chem. Phys.*, 2017, **147**, 044201.
49. W. Andralojc, Y. Hiruma, W.-M. Liu, E. Ravera, M. Nojiri, G. Parigi, C. Luchinat and M. Ubbink, *Proc. Natl. Acad. Sci. USA*, 2017, **114**, E1840-E1847.
50. G. Cornilescu, J. Marquardt, M. Ottiger and A. Bax, *J. Am. Chem. Soc.*, 1998, **120**, 6836-6837.
51. I. Bertini, P. Kursula, C. Luchinat, G. Parigi, J. Vahokoski, M. Wilmanns and J. Yuan, *J. Am. Chem. Soc.*, 2009, **131**, 5134-5144.