Rotamer Modelling of Cu(II) Spin Labels Based on the Double-Histidine Motif

ORIGINAL PAPER

CrossMark

Rotamer Modelling of Cu(II) Spin Labels Based on the Double-Histidine Motif

Shreya Ghosh[1] · Sunil Saxena[1] · Gunnar Jeschke[2]

Abstract

Spin labels attached to two residues of a protein chain have less conformational flexibility than those attached to a single residue and thus lead to a narrower spatial distribution of the unpaired electron. The case of Cu^{2+} labels based on the double-histidine (dHis) motif is of particular interest, as it combines the advantage of precise localization of the unpaired electron with a labelling scheme orthogonal to the more common cysteine-based labelling. Here, we introduce an approach for in silico spin labelling of a protein by dHis motifs and Cu^{2+} complexes of iminodiacetic acid or nitrilotriacetic acid. We discuss a computerized scan for native histidine pairs that might be prone to bind such Cu^{2+} complexes and spin-labelling site pair scans that can identify suitable double mutants for labelling. Predicted distance distributions between two Cu^{2+} labels are compared to experimental distance distributions. We also test the hypothesis that elastic network modelling of conformational transitions with Cu^{2+}-dHis labels can provide more accurate structural models than with nitroxide labels.

1 Introduction

The combination of pulsed dipolar spectroscopy techniques, such as double electron electron resonance (DEER), also known as PELDOR [1, 2], double-quantum coherence (DQC) [3, 4], relaxation-induced dipolar modulation enhancement (RIDME) [5, 6], or the single-frequency techniques for refocusing (SIFTER) dipolar couplings [7, 8], with site-directed spin labelling [9, 10] has developed into an important tool in structural biology [11–14], especially for proteins and protein complexes with

✉ Sunil Saxena
   sksaxena@pitt.edu

✉ Gunnar Jeschke
   gjeschke@ethz.ch

1   Department of Chemistry, University of Pittsburgh, Pittsburgh, PA 15260, USA

2   Department of Chemistry and Applied Biosciences, ETH Zürich, Vladimir-Prelog-Weg 2, 8093 Zurich, Switzerland

intermediate order of backbone conformation [15]. Most applications use nitroxide spin labels attached to a single cysteine residue, such as the methanethiosulfonate spin label (MTSSL), which has a moderately flexible linker between the backbone and the nitroxide moiety. Bifunctional attachment to two cysteine residues, as in the Rx spin label that features two methanethiosulfonate groups [16], leads to a much narrower spatial distribution of the unpaired electron as demonstrated by molecular dynamic simulations [17]. However, binding to both residues is not always achieved. Furthermore, given that the removal of all native cysteines in a protein is not always feasible, alternative site-directed labelling approaches are in demand. For some problems in structure determination, spectroscopically orthogonal labelling by two different paramagnetic centres is required [18], which has also led to an interest in metal-ion based labels.

In this context, pulsed electron paramagnetic resonance (EPR) spectroscopic techniques involving Cu^{2+} have been developed [19–28]. Strategies involving Cu^{2+} bound to native binding sites in the protein have proven to be an important technique for structural determination of proteins [19, 29–33]. In cases where proteins do not have endogenous Cu^{2+}-binding sites, several spin labelling techniques have been developed which site-specifically incorporate Cu^{2+} to a protein using chelating tags [34]. While these tags show high affinity to Cu^{2+} and require only a single mutation in the protein (specifically, cysteine), they are flexible and can provide distributions that are at best the same as MTSSL. Another alternative labelling method for Cu^{2+} involves the use of two histidine residues, placed strategically to bind Cu^{2+} [35] known as the double histidine (dHis) motif. In addition this motif uses an $i, i+4$ placement of histidine residues on an α-helix or $i, i+2$ for a β-sheet. Even though each dHis-Cu^{2+} requires mutation of two residues, the motif shows almost five times narrower distance distribution than MTSSL, indicating that it is a much rigid spin label. However, a major limitation of this motif when using just Cu^{2+} is that Cu^{2+} shows poor selectivity to the dHis. Thus, Cu^{2+} may coordinate unspecifically to other residues in the protein.

To overcome this limitation, the Cu^{2+} ion was chelated by ligands such as iminodiacetic acid (IDA) [35, 36] or nitrilotriacetic acid (NTA) [37]. Using Cu^{2+}–IDA or Cu^{2+}–NTA increases the binding specificity of the complex to the dHis site only, while reducing unspecific binding. The dHis motif, being highly rigid, provides a distance distribution which can be almost five times narrower than that obtained using standard nitroxide spin labels such as methanethiosulfonate spin label [35, 37]. This development is important because side chain flexibility is a major limitation for the precision and accuracy of structural models that can be obtained with restraints from MTSSL-based spin labelling. Because of the high flexibility of MTSSL, the distance distribution is dominated by the motion of the spin label rather than the actual backbone fluctuations [38] unless backbone disorder is substantial [15].

To further develop this methodology, systematic methods are required that use the Cu^{2+}-based distance restraints to provide protein structure and conformational dynamics. So far, nitroxide-based distance restraints along with the multiscale modeling of macromolecular systems (MMM) software [39, 40] have been used to provide such information. MMM generates a library of rotamers based on the conformational space of the spin label. Distance restraints along with restraints generated

from the elastic network model (ENM) in MMM [41, 42] can then be used to create models of protein structure and conformational dynamics. Due to the high rigidity, distance restraints obtained with the dHis motif with Cu^{2+}–IDA/NTA are expected to improve the accuracy of such models as compared to nitroxide spin labels, as the initial study had shown that the flexible linker leads to an accuracy loss [41]. Hence, there is a need to incorporate dHis-based distance restraints into the MMM software. Such an implementation should also allow to test whether native pairs of histidine residues may be prone to binding Cu^{2+}–IDA or Cu^{2+}–NTA and should be able to suggest site pairs for labelling in the spirit of the site scan feature [43] for nitroxide labels attached to a single cysteine residue. To this end, in this work we have generated histidine rotamer libraries and structural models for the Cu^{2+} complexes formed with a dHis motif and either of the two chelating ligands, IDA or NTA. Experimental distance constraints obtained on dHis-Cu^{2+}–IDA/NTA along with MMM simulations can thus provide direct insight into the protein backbone structure and conformation.

This paper is structured as follows: In Sect. 1, we discuss the general approach for modelling attachment of a bifunctional label to a pair of residues. In Sect. 2, we explain our hypothesis on the structure of the formed complexes, which is based on available crystal structures, general knowledge of coordination preferences of Cu^{2+}, and density functional theory computations. In Sect. 3, we describe the implementation of labelling, visualization, and site scans into MMM. In Sect. 4, we compare distance distributions predicted by our approach to those measured previously on the B1 immunoglobulin-binding domain of protein G (GB1) [17–19] and on human glutathione *S*-transferase A1-1 protein (huGSTA1-1) [37, 44]. In Sects. 5 and 6, we test by simulations for a set of known structural transitions whether restraints based on dHis-Cu^{2+} labelling can indeed be expected to improve on the accuracy of structural models as compared to restraints obtained with MTSSL.

2 Modelling Approach

2.1 Attachment of Bifunctional Labels

A bifunctional label attaches to the sidechains of two amino acid residues if these are in a spatial arrangement that satisfies the steric requirements of the label. Usually, these requirements will be rather strict since low conformational variability of the label is a desired property. It is thus necessary for the two sidechains to adopt a rather well-defined binding pose. This requirement may still allow for some variability in the C^{α}–C^{α} distance and relative local frame orientation of the residues, since even native sidechains such as the ones of cysteine or histidine can adopt different rotameric states. The first step of modelling the attachment thus consists of testing whether any rotamer pairs of the binding residues fulfil the spatial requirements for label attachment. We shall define these requirements for dHis-Cu^{2+} coordination in the next Section.

The problem of finding suitable rotamer combinations for the two histidine residues can be solved with the same rotamer library approach that we have previously

introduced for modelling single-site labels [40]. For generation of the histidine rotamer library, we employed our newer approach based on a Monte Carlo sampling of conformations and hierarchical clustering [42], except that no softening of the Lennard–Jones potentials was required and an ensemble of 50,000 low-energy conformations was generated and clustered. The initial structure of histidine was optimized with the MMFF94 force field in Chem3D (Perkin Elmer). The histidine sidechain has only two rotatable bonds with three energy minima for the torsion angle χ_1 and four minima for the torsion angle χ_2 (see Fig. 1), giving 12 rotamers. Hence, for each site pair only 144 rotamer pairs need to be tested, which makes site scans feasible even for large proteins. The histidine rotamer library is included with the name HCU_298K_UFF_12_r5 in MMM 2018.

2.2 Modelling of Copper Coordination

The complex of Cu^{2+} with a dHis site and iminodiacetic acid (IDA) was originally expected to have an approximately octahedral geometry with two equatorial nitrogen ligands from the dHis site, an equatorial nitrogen and an equatorial oxygen ligand from IDA and an axial oxygen ligand from IDA [35]. The final axial position is thought to be occupied by water. A corresponding crystal structure [45] was downloaded from the Cambridge Structural Database (CSD) as a CIF file (accession number 650681) and converted with Mercury to a Sybyl MOL2 file. In this structure (Fig. 1), the coordinated nitrogen atoms of the two imidazole ligands have distances of 1.987 and 2.001 Å from the Cu^{2+} ion, whereas the nitrogen atom of the IDA ligand has a distance of 2.085 Å and the equatorial oxygen atom of the IDA ligand has a distance of 1.965 Å. The axial oxygen atom of IDA has a distance of 2.225 Å. The N^δ–Cu–$N^{\delta'}$ angle for the two imidazole ligands is 92.1° and the N^δ–$N^{\delta'}$ distance is 2.872 Å. Because of the two very similar Cu–N distances, we term this coordination mode of IDA 'symmetric'. It is modelled by the conformer library dHis_Cu_IDA_sym that can be accessed with the label name IDS.

The structure was edited in Accelrys Discovery Studio to the *cis*-[iminodiacetato-κ^3N,O,O′]bis(1H-imidazole-κN^3) copper(II) complex (N^3 is N$^\delta$) and the MOL2 file
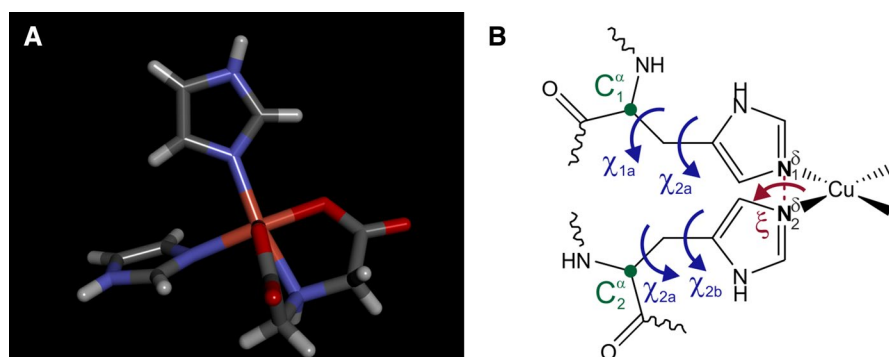

Fig. 1 **a** Stick model of the crystal structure of the bis(imidazole)–IDA Cu^{2+} complex. **b** Definition of torsion angles for the two histidine residues and for copper orientation with respect to their two N^δ atoms

of the edited structure was converted to a Cartesian coordinate file using Chem3D (Perkin-Elmer). It was then optimized on restricted Kohn–Sham level with the BP86 density functional, the def2-SVP basis set for all atoms, and the conductor-like polarizable continuum solvent model for water in ORCA 4.0.1 [46]. In the optimized structure (Fig. 2), we find lengths of the Cu–N dative bonds to the two imidazole nitrogens of 2.012 and 2.216 Å, respectively, an N^δ–Cu–$N^{\delta'}$ angle of 101.4°, and an N^δ–$N^{\delta'}$ distance of 3.273 Å between the two imidazole nitrogens. In this structure, one of the two imidazoles acts as an axial ligand, whereas the square plane is of a 2N2O type with distances of 2.011 and 2.014 Å to the two coordinated oxygen atoms and 2.057 Å to the coordinated nitrogen atom of IDA. We term this coordination mode 'asymmetric'. It is modelled by the conformer library dHis_Cu_IDA_asym that can be accessed with the label name IDA. This is the default library for the IDA ligand that we use throughout this paper.

The complex of Cu^{2+} with a dHis site and nitrilotriacetic acid (NTA) was originally expected to have an approximately octahedral geometry with two equatorial nitrogen ligands from the dHis site, an equatorial nitrogen and an equatorial oxygen ligand from NTA and two axial oxygen ligands from NTA [37]. A crystal structure with the same number and type of coordinated atoms and at least similar geometry [47] has accession code 927203 in the CSD and features a 2,2′-bipyridine ligand instead of two imidazole ligands. Indeed, the two nitrogen atoms of the bipyridine form strong dative bonds with Cu–N distances of 1.989 and 2.005 Å. Ligand geometry restrains the N–N distance to 2.597 Å, leading to a N–Cu–N angle of 81.1°. One of the coordinating oxygen atoms of the NTA ligand lies in the N–Cu–N plane and forms a strong dative bond with a length of 1.936 Å. All other coordinating atoms of NTA are out of plane.

The ligand was edited in Accelrys Discovery Studio using the imidazole bond lengths and angles observed in the DFT-optimized structure of the IDA complex. The monoanionic NTA complex was then optimized on restricted Kohn–Sham level with the BP86 density functional, the def2-SVP basis set for all atoms, and the
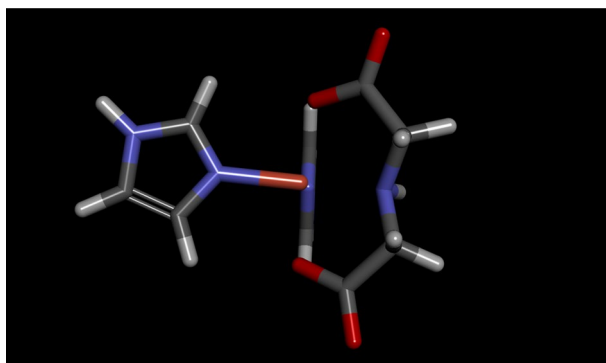

Fig. 2 Stick model of the DFT-optimized structure of the bis(imidazole)–IDA Cu^{2+} complex. View perpendicular to the equatorial plane. The second imidazole ligand in the axial position is perpendicular to the image plane and situated below the Cu^{2+} atom. Differences in the Cu^{2+} position compared to symmetric binding are minor

conductor-like polarizable continuum solvent model for water in ORCA 4.0.1 [46]. In the optimized structure, only one of the imidazole ligands was found to be coordinated. This result was reproduced for two somewhat different initial geometries.

We then decided to link the two imidazole ligands, as they are linked in a protein context. To that end we edited them to 4-methylene-imidazoles and inserted a –CH_2–CH=CH–CH=CH– link between the methylene carbons. To overcome difficulties with self-consistent field convergence in the vicinity of our initial geometry, we pre-optimized the structure without invoking the conductor-like polarizable continuum solvent model and with Fermi-like occupation number smearing at a temperature of 5000 K. Even in this case, we observed dissociation of one of the imidazole ligands from the complex. In fact, the bipyridine complex, which is part of a coordination polymer in the crystal structure, does not retain its octahedral coordination geometry when DFT optimized in isolation. The optimized geometry is a square pyramid with the two nitrogen atoms of the bipyridine ligand and two oxygens of NTA in equatorial positions and the third oxygen in the axial position at a distance of 2.122 Å. The nitrogen atom of NTA is not coordinated (2.592 Å, not in the other axial position).

Therefore, we gave up on the hypothesis that the dHis motif and NTA together occupy all six coordination positions of Cu^{2+} and based optimization on the complex of a keto-bridged bismethylimidazole-NTA ligand with Cu^{2+} with CSD accession number 167295 [48]. Coordination geometry is again a square pyramid with the two methylimidazole nitrogen atoms and two oxygen atoms of NTA taking the equatorial positions and the nitrogen of NTA taking the axial position (Fig. 3). The third carboxylic group of NTA is not coordinated. We first optimized the complex, then severed the keto bridge and rotated the methylimidazole rings out of plane and optimized again. In the final structure, we find lengths of the Cu–N dative bonds to the two imidazole nitrogens of 2.033 and 2.034 Å, and N–Cu–N angle of 93.2°, and an N–N distance of 2.954 Å between the two imidazole nitrogens. The equatorial oxygen atoms are at a distances of 1.986 and 2.027 Å and the axial nitrogen is at a distance of 2.360 Å. This is again
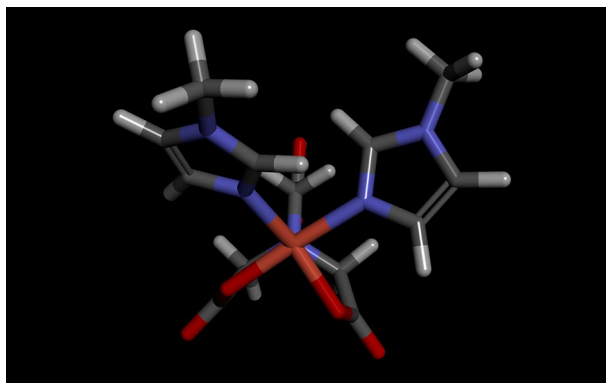

Fig. 3 Stick model of the DFT-optimized structure of the bis(methylimidazole)–NTA Cu^{2+} complex. View perpendicular to the equatorial plane. The nitrogen atom of the NTA ligand in the axial position is situated below the Cu^{2+} atom

a symmetric coordination mode, which we model by the conformer library dHis_Cu_NTA_sym that can be accessed with the label name NTA.

2.3 Modelling of the Cu²⁺ Site

We assume that a dHis motif will bind Cu^{2+}–IDA or Cu^{2+}–NTA if a combination of histidine rotamers exists that poses the N^3 (N^δ) atoms of the two histidine sidechains at a distance allowing for coordination. The Cu^{2+} coordination polyhedron is rather plastic [49, 50] as is also apparent from the results of the DFT computations described above. These results suggest that N^δ–$N^{\delta'}$ distances between 2.6 and 3.2 Å can be accommodated. We apply this range for all three ligand conformer libraries. In the first step, we select all combinations of histidine rotamers at the two sites whose N^δ–$N^{\delta'}$ distance d_{NN} falls within this range.

We further assume that the polar (IDA) or charged (NTA) Cu^{2+} complexes will tend to point away from the protein surface. If we choose the x-axis of a protein-fixed local frame along the N^3–$N^{3'}$ vector and the centre of mass of the protein in the xy-plane with a coordinate $y < 0$, our first guess for the Cu^{2+} position is thus in the xy-plane with a coordinate $y > 0$. As the origin of the frame we take the centre of the N^δ–$N^{\delta'}$. This local frame is also defined for the complex structure, so that the initial guess for the Cu^{2+} and ligand (IDA or NTA) position can be obtained by superimposition of the two local frames.

The geometry of the Cu^{2+} site is thus fixed, except for rotation of all non-protein atoms about the N^δ–$N^{\delta'}$ axis. We assume an empirical cosine potential for this rotation with an energy minimum at rotation angle $\xi = 0$ of the geometry described above, where the Cu^{2+} ion is farthest away from the centre of mass of the protein. The barrier height V_ξ of this torsion potential, which has a maximum at $\xi = 180°$, can be fitted to reproduce experimentally observed distance distributions. To the energy of this potential, we add the Lennard–Jones interaction energy of the non-protein atoms with the protein atoms as in our rotamer library approach [40]. For a soft torsion potential with a barrier height $V_\xi < 1$ kJ/mol, the distribution of ξ is mainly restrained by clashes of the ligand with protein atoms. For a hard torsion potential with $V_\xi > 100$ kJ/mol, coordination induces significant strain on local protein structure. Given the fact that at a 1:1 ratio of Cu^{2+}–IDA to dHis sites complete attachment is not observed [36], we consider it as unlikely that such coordination can induce substantial strain.

We discretize the ξ torsion about the N^3–$N^{3'}$ axis to keep in line with our approach of modelling a label by a small number of conformations. Discretization in steps of 5° ensures that the Cu^{2+} coordinates of the ξ rotamers are sufficiently close spaced.

3 Implementation and Usage in MMM

3.1 Site Scans

In spin-labelling site scans [43] we consider only residue pairs within the same chain. Such site scans can still correspond to a considerable computational effort,

since for a chain with n residues we need to test $n(n-1)/2$ site pairs. This quadratic scaling of the computational expense of a site scan with peptide chain length reduces to a nearly linear scaling by considering that N^δ–$N^{\delta'}$ distances between 2.6 and 3.2 Å can be realized only if the C^α–C^α distance is shorter than 10 Å. Although such distances can be realized for residue pairs $(i, i + \Delta i)$ with large Δi, we consider only residues whose numbers differ by $1 \leq \Delta i \leq 5$. As explained above, canonical site pairs have $\Delta i = 4$ for two sites in an α helix and $\Delta i = 2$ for two sites in a β sheet. The site scan report is stored as a text file and specifies, for each site pair that is expected to coordinate Cu^{2+}, the two potential coordinating residues, their C^α–C^α distance, the number of significantly populated conformers, the partition function, the root-mean-square deviation of the Cu^{2+} position from its mean, and the Cu-backbone distance, defined as the distance of the Cu^{2+} ion from the midpoint of the C^α–C^α vector of the two residues. The report is automatically opened in the MMM report editor.

For finding pairs of native histidine residues that are susceptible to binding a Cu^{2+}–IDA or Cu^{2+}–NTA complex, we allow for any sequence distance Δi within the same peptide chain. This functionality was tested on a few crystal structures and an unexpected potential-binding motif was found between the remote histidines 106 and 203 in bacterial L-asparaginase crystal structure 1JSL [51]. Whether this protein indeed binds to Cu^{2+}–IDA was not checked experimentally, but the possibility of such coordination should be checked if a structure of a protein is available.

3.2 Visualization

In contrast to our previous implementation for nitroxide spin labels, the implementation for Cu^{2+}-dHis labels stores information on labelling separately from the structure, i.e., the internal representation of the protein is not changed. While this has some advantages, for instance, when testing alternative labels in the same MMM session, it prevents visualization of all-atom models of the attached Cu^{2+} complex. Instead, we opted for visualization of the distribution of Cu^{2+} positions, which leads to a clearer representation (Fig. 4).

3.3 Usage

In MMM 2018, Cu^{2+}-dHis site scans, labelling, and visualization are accessible only via the command line. After the Cu^{2+} position distributions have been predicted, distance distributions and DEER time-domain data can be computed as for nitroxide labels in the DEER window. The use of Cu^{2+}-dHis restraints in elastic network modelling of conformational transitions is explained in Sect. 5.

For any protein of small or moderate size, we recommend working with the *blscan* command, which has the syntax

blscan *address label outfile* [*native*] [*torsionpot*],

where address denotes a chain address (Example: (*A*) for chain A of the current structure), *label* can be *IDA* for the asymmetric attachment of the IDA ligand (default), *IDS* for the symmetric attachment of the IDA ligand, or *NTA* for the NTA
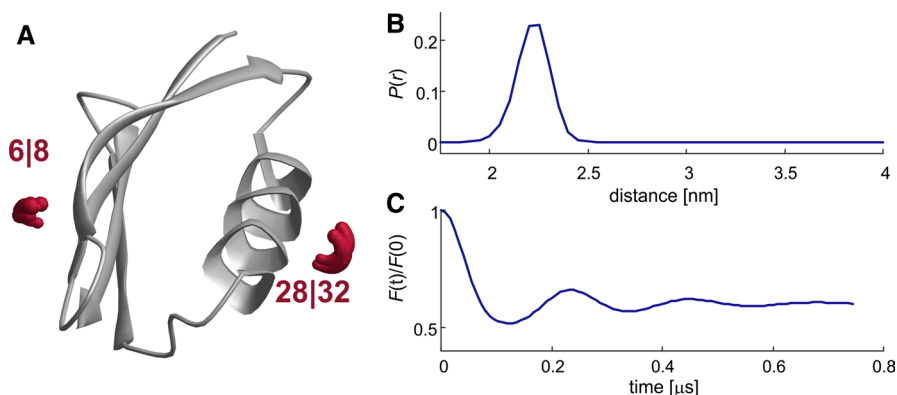
Fig. 4 Predicted spatial distribution of Cu^{2+}–IDA in the GB1 double pair mutant 6H/8H//28H/32H based on crystal structure PDB 1PGA [52]. **a** Visualization of the Cu^{2+} distribution. **b** Distance distribution. **c** DEER form factor

ligand. The parameter *outfile* is the file name for the site scan report (default: *bilabel_site_scan*), which is stored in the current directory, but can be saved to any other directory from the report editor window. The parameter *native* can be *native*, *true*, or *false*, where the first two choices instruct the program to search for pairs of native histidine pairs that can coordinate Cu^{2+}, irrespective of the sequence distance. This parameter defaults to *false*. The parameter *torsionpot* denotes the torsion potential in kJ/mol, is optional and defaults to 10 kJ/mol (see Sect. 4).

If the dHis sites of interest are already known and the protein is large, it may be more convenient to label individual dHis sites with the *bilabel* command with the syntax

bilabel *address*1 *address*2 *label* [*torsionpot*].

The parameters are analogous, except that *address1* and *address2* are addresses of the two residues that together make up the dHis site. These residues can be of any type, mutation to histidine is automatically computed. Although the command allows for the two residues to be situated in different peptide chains, we discourage such labelling except for methodological work.

Finally, the Cu^{2+} positions can be visualized with the standard *show* command with modified address syntax

show *address*1|*address*2 label *color*.

Again, *address1* and *address2* are addresses of the two residues that together make up the dHis site, which are separated without space by the vertical line separator. The argument 'label' must be written verbatim, it does not denote the type of label. The *color* argument is optional and can be either a scalable vector graphics color name (*crimson* was used for Fig. 4) or a red–green–blue (RGB) triple of values in the range from 0 to 1.

Other standard functionality of MMM is available after either the *blscan* or *bilabel* command has been performed. In particular, Cu^{2+}-dHis labels can be selected in

the DEER window of MMM and distance distributions (Fig. 4b) as well as DEER form factors (Fig. 4c) can be predicted and compared to experimental data that were processed with the DeerAnalysis software.

The new features are implemented into MMM from version 2018 onwards. MMM is freely available as an open-source package for Matlab (The MathWorks Inc, Natick, MA, USA) at http://www.epr.ethz.ch/software.html.

4 Comparison to Experimental Data

4.1 B1 Immunoglobulin Binding Domain of Protein G (GB1)

GB1 is a 56-amino acid B1 immunoglobulin-binding domain of protein G. GB1 is a simple model system used in protein-folding studies and other applications because of its high-thermal stability and well resolved structures obtained from X-ray crystallography, NMR, EPR and other methods [20, 53, 54]. Hence, using the dHis-Cu^{2+} motif on GB1 helps to predict the accuracy of the motif for determining structural constraints and conformational changes in the protein in solution. In this work, we have compared data from MMM simulations with the experimental data previously obtained on two different mutant constructs of GB1 [36, 37]. For both the tetramutant GB1 constructs, the dHis site at the α-helix was positioned at 28H/32H. For the β-sheet, one of the construct had the dHis site at 6H/8H, while the other had it at positions 15H/17H. While Cu^{2+}–IDA was used only for 6H/8H/28H/32H [36], Cu^{2+}–NTA was used with both the mutants [37]. The 6H/8H/28H/32H tetramutant GB1 is shown in Fig. 5a.

The experimental most probable distance for the tetramutant 6H/8H/28H/32H-GB1 spin labeled with Cu^{2+}–IDA was 2.47 nm with a standard deviation of 0.08 nm [36]. For performing MMM simulations, we considered five different PDB files as well as different torsion potentials on the spin label (Fig. 6). We considered the coordination mode of IDA to be 'symmetric' such that the two nitrogen atoms of the imidazole ligands are in the equatorial plane of Cu^{2+}. The default torsion potential used by MMM is 10 kJ/mol. An increase in the torsion potential leads to a narrower distance distribution while a broader distance distribution is favored for a lower torsion potential. For most cases, a torsion potential of 0.1 kJ/mol leads to a larger distribution width than the experimental one and the most probable distance was short (~2.16 nm). On increasing the potential to 10 kJ/mol, the distribution width was comparable to that of the experiment with the most probable distance still being ~2.5 Å shorter than the experimental one. Further increase in the torsion potential to 100 kJ/mol only leads to a marginal improvement. Keeping in mind that the crystal structures 4WH4, 1PGA and 2QMT have a resolution of 2.2 Å, 2.07 Å and 1.05 Å, respectively, we can consider the simulated most probable distance to be within error.

We see similar agreement between simulated and experimental results for the tetramutants 6H/8H/28H/32H-GB1 and 15H/17H/28H/32H-GB1 with Cu^{2+}–NTA (Figs. 7, 8). A discrepancy was, however, observed when spin labeling the PDB files 4WH4 and 1PGA at sites 6, 8 (for a potential of 100 kJ/mol) and 15, 17,
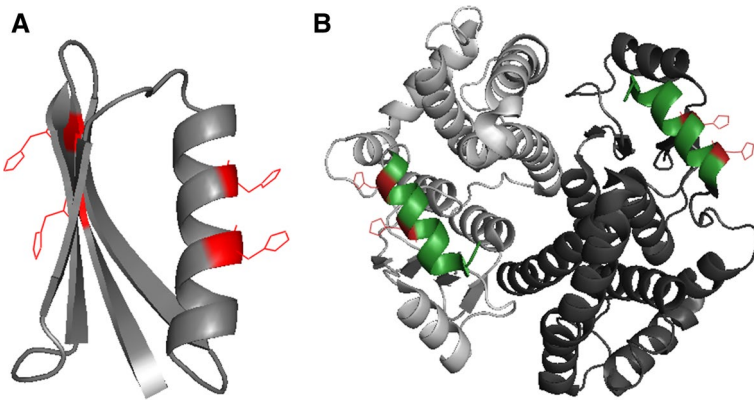
Fig. 5 **a** Tetramutant GB1 with dHis at sites 6H/8H in the β-sheet and 28H/32H in the α-helix. The residues in red indicate the histidine of the dHis. **b** Dimeric huGSTA1-1 with a dHis site (211H/215H) at each subunit. The two subunits are shown by light and dark gray while the Histidine residues are shown in red. The helices in green represent α9 (color figure online)
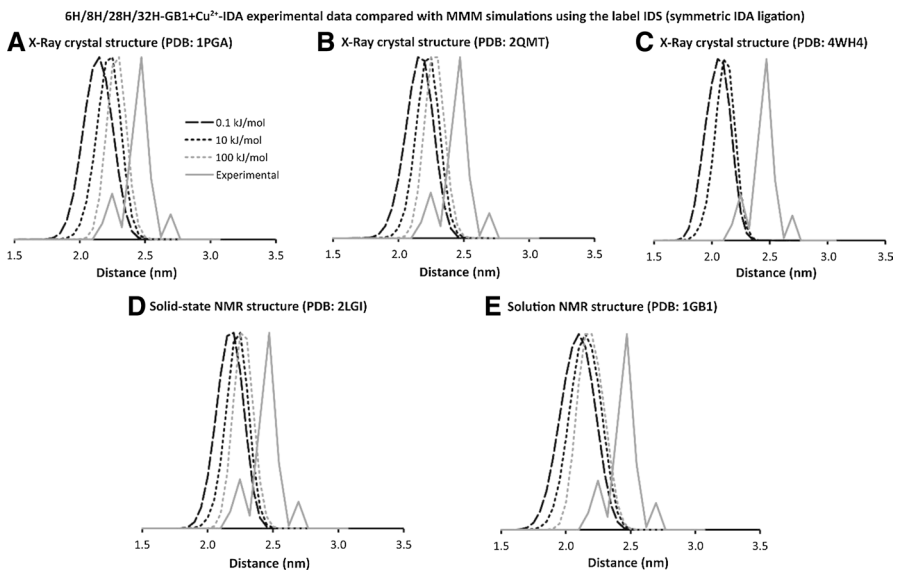

Fig. 6 MMM simulations compared to experimental data of 6H/8H/28H/32H-GB1 with Cu^{2+}–IDA for five different PDB structures. Simulations were also performed at different torsion potentials V_ξ (0.1, 10 and 100 kJ/mol). For most cases, while a potential of 0.1 kJ/mol leads to a larger distribution width, a potential of 10 kJ/mol leads to a distribution width similar to that of the experimental. Further increase to 100 kJ/mol did not lead to any significant change

respectively. We were not able to spin label the abovementioned sites with Cu^{2+}–NTA in silico. The likely reason is that the ligand clashes with other protein atoms at the minimum of the assumed torsion potential.
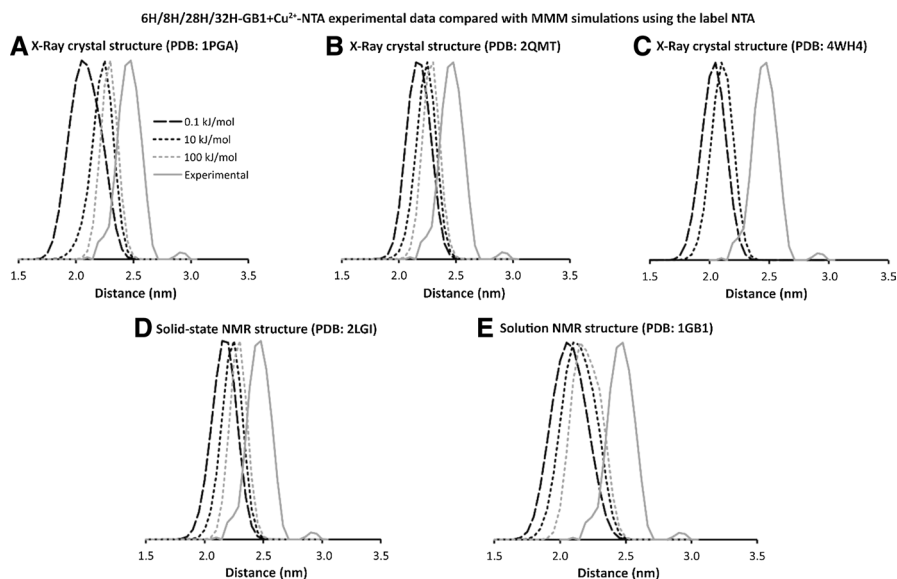
Fig. 7 MMM simulations and DEER data for 6H/8H/28H/32H-GB1 + Cu^{2+}–NTA showed similar agreement as mentioned for Cu^{2+}–IDA
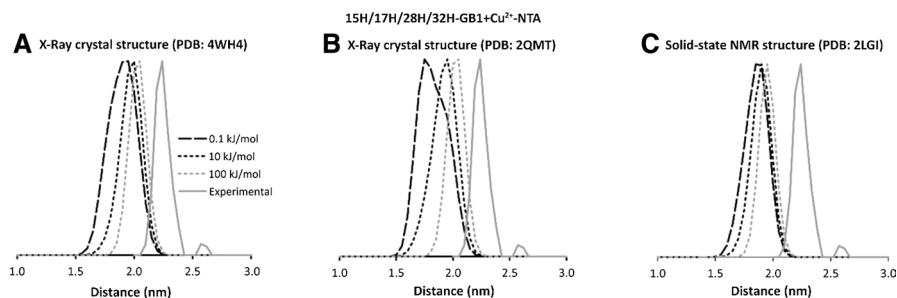

Fig. 8 Comparing MMM simulations with DEER data on 15H/17H/28H/32H-GB1 + Cu^{2+}–NTA showed results similar to that of 6H/8H/28H/32H with Cu^{2+}–NTA

In general, the most probable distance appears to be within the resolution of the PDB structures. That said, the simulated most probable distances are systematically shorter than the experimental ones for both IDA and NTA labels in GB1. Variation of the predicted mean distance 6H/8H/28H/32H-GB1 among the 60 conformations in the NMR structural ensemble does not exceed 0.9 Å. The remaining discrepancy between predictions and experimental data may then indicate that the Cu^{2+} ion is slightly further away from the backbone than suggested by our histidine rotamer library and our complex structures. This issue needs to be re-examined once more data from proteins with known structure become available.

4.2 Human Glutathione *S*-transferase A1-1 Protein (huGSTA1-1)

Glutathione *S*-transferase (GST) superfamily consists of a class of proteins that are diverse in their amino acid sequence. All GSTs play an important role in the cellular defence mechanism for a large number of species. The vital function of GSTs is their ability to catalyse the nucleophilic attack by glutathione (GSH) on the electrophilic atoms of the xenobiotic substrates [55]. The electrophiles then have a deactivated reactive center which prevents the xenobiotic substrates from reacting with crucial cellular proteins and nucleic acids.

The human α-GST (huGSTA1-1) is a 222-residue homodimer protein that is believed to exist in two major conformations in the ligand-free state [56]. Conformations of the huGSTA1-1 in the ligand-bound state has been well resolved using X-ray crystallography and NMR [57–59]. However, in the ligand-free state, the conformation of a critical 11-residue helix at the C-terminal, called α9, has only recently been identified [44]. Recently, DEER was performed on the unliganded huGSTA1-1 [44]. Distance measurements using MTSSL spin labels on the α9 helix displayed a bimodal distribution which was consistent with two distinct conformations of the protein. These distances along with MMM simulations were used to generate a model of the α9 conformation in the unliganded state. In addition, distance measurements were performed by incorporating a dHis site (K211H/E215H) in each subunit within the dimeric protein (cf. Fig. 5b). The resultant DEER data using Cu^{2+}–IDA showed bimodal distance distribution, which is consistent with the presence of two conformations of the α9 helix in ligand-free huGSTA1-1.

Figure 9 shows the dHis-based distance distribution obtained in the absence of the ligand. Previously, we identified the larger distance as being consistent with the α9 conformation observed in the crystal structure of the ligand-bound enzyme (PDB: 1K3L). Indeed, labeling this structure (PDB: 1K3L) by dHis in MMM generates a distance distribution that closely matches the experimental distance. For the shorter distance we used the structure obtained by a combination of MTSSL distances and MMM simulations [44]. On performing MMM simulations on the same conformer but using dHis-Cu^{2+}, we observed the most probable distance to be around 4.0 nm for a 0.1 kJ/mol torsion potential and 4.1 nm for torsion potentials of 10 and 100 kJ/mol. Given that the structure of this conformer was generated using only MTSSL-based distance constraints, we expect the simulated most probable dHis-based distance to differ from that of experimental (~3.6 nm). Nevertheless, the modeled distance is within the experimental distribution.

In the case of huGSTA1-1, the experimental distance distributions are much broader than the ones predicted by our modelling. This cannot be traced back to an underestimate of the spatial distribution of the label with respect to the backbone in our approach, since even with a flat torsion potential, the experimentally observed widths cannot be matched. The broad distance distributions for huGSTA1-1 thus indicate some conformational flexibility of the peptide backbone in solution that is suppressed by packing effects upon crystallization. While preliminary, these results underscore the potential of dHis-based distances to measure backbone flexibility.

Cu^{2+}–Cu^{2+} distance distributions simulated for a fixed backbone conformation have full widths at half height widths around 2 Å that vary little and agree well with
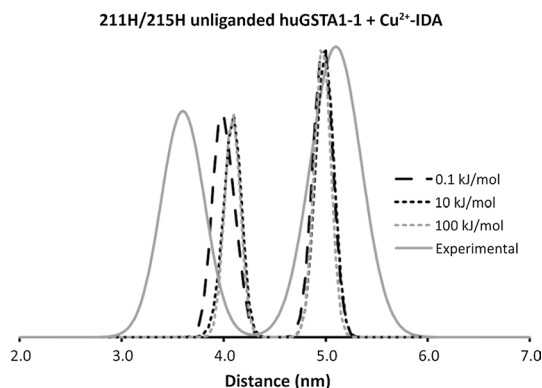
211H/215H unliganded huGSTA1-1 + Cu²⁺-IDA

Fig. 9 Experimental data of unliganded 211H/215H huGSTA1-1 with Cu²⁺–IDA compared to MMM simulations showing both the possible conformations. MMM simulations using dHis-Cu²⁺ were performed on the liganded GST structure (PDB: 1K3L) and showed good agreement with the longer distance of the unliganded protein. Simulations were also performed on the previously obtained model which was based on MTSSL distances and X-ray crystal structure constraints [44]. The simulations resulted in a most probable distance of ~ 4 nm compared to the experimental 3.6 nm. This difference can be attributed to the fact that the model used for simulation was roughly based on sparse experimental MTSSL data

the experimental widths measured on GB1. Backbone flexibility beyond this width should thus be visible in experimental data. Given a sufficient number of restraints, it can be simulated with the approach described in [60] which is implemented in MMM and has been adapted for Cu²⁺-dHis labels.

5 Accuracy of Modelling Conformational Change by an Elastic Network Model Approach

5.1 Basic Considerations and Implementation

Here, we consider the modelling of a conformational transition of a protein based on a structure of the initial state and a moderate number of distance restraints for the final state, supported by an anisotropic elastic network model (ENM) [61]. Previous research has demonstrated that with distances between spin labels, the accuracy of the model for the final state, quantified by a fractional coverage f of the conformational change, is worse than when C^α–C^α distances for the same site pairs are used [41]. This loss in accuracy presumably results from the conformational flexibility of the spin label sidechain. Since this flexibility is smaller for Cu²⁺-dHis labelling, we hypothesized that a higher fractional coverage can be obtained with Cu²⁺-dHis labels as compared to MTSSL. This hypothesis is best tested by simulations on state pairs where both the initial and final structure are experimentally known [41, 61]. From the set of 18 such pairs used in our previous studies [41, 62], we focused on cases where a fractional coverage $f > 0.5$ was found with 20 optimal C^α–C^α distance restraints, indicating that the ENM

reasonably well models the transition. We use the optimized algorithm based on equipartitioning of energy between the normal modes of the ENM with 20 restraints, an initial active space dimension $B_0 = 20$, and the edENM parametrization of the ENM [62]. For the two label cases, we determined the mean fractional coverage for an ensemble of 20 models of the final structure computed with different distance restraint sets where each individual restraint was set by a normally distributed random number with the mean and standard deviation of the simulated distance distribution for this label pair.

The choice of labelling sites is based on optimization for MTSSL sites [54]. Following site scans for Cu^{2+}-dHis-IDA, we selected histidine pairs that overlapped with the optimal MTSSL sites or had a maximum sequence distance of two residues from them. This strategy excluded four potential test cases (1GGG/1WDN, 1BNC/1DV2, 1B7T/1DFK, 1N0V/1N0U), since in these cases corresponding dHis sites could not be found for all MTSSL sites in both the initial and final structure. We still preferred this strategy to independent optimization of the sites for the individual labels, since for the latter strategy the results are influenced not only by conformational variability of the label but also by differences in the linear combination of the normal modes of the ENM.

The results for the remaining six test cases are summarized in Table 1. In five cases, fractional coverage indeed improves when going from MTSSL to a dHis-based label. In the case of 1L5B/1L5E we observe a slight decrease. We note that the mean coverage for an ensemble obtained with varying restraint sets has an uncertainty. For the smallest protein, which incidentally is the 1L5B/1L5E case, we computed six 20-membered ensembles each for MTSSL and Cu^{2+}-dHis-IDA. The mean fractional coverage for all ensembles is 0.570 for MTSSL and 0.562 for Cu^{2+}-dHis-IDA, with standard deviations of 0.062 and 0.016, respectively. The larger standard deviation for MTSSL is expected because of the broader distance distributions. For the test case 1AON/1OEL the optimal set of MTSSL restraint sites performs poorly compared to an optimal set of C^α restraint sites ($f = 0.566$), even if C^α restraints are used at the optimal MTSSL sites. Altogether, we can conclude that the expected improvement in accuracy when using the more rigid dHis-based labels is borne out in the simulations for most cases.

Table 1 Fractional mean coverage of conformational transitions in elastic network modelling with simulated restraints

	1L5B/1L5E	2LAO/1LST	1OMP/1ANF	1EPS/1G6S	1AON/1OEL	1SU4/1T5S
$C\alpha$	0.783	0.793	0.468	0.780	0.162	0.696
MTSSL	0.613	0.518	0.224	0.444	0.190	0.498
Cu^{2+}-dHis-IDA	0.575	0.578	0.320	0.613	0.214	0.602

For MTSSL and Cu^{2+}-dHis-IDA mean values for an ensemble of 20 models are given

6 Conclusion

In-silico spin labelling for labels that attach to two residues is feasible with a computational effort that allows for site scans of large proteins. The approach is based on a rotamer library for the sidechains of the native attachment residues and a conformer library for the actual label. It was implemented for Cu^{2+}-dHis labelling with the auxiliary ligands IDA and NTA. Despite some uncertainty on the structure of the complexes, the approach is robust for $i,i+4$ surface-exposed dHis sites in α-helices and for $i,i+2$ surface-exposed dHis sites in β-sheets, since for given histidine coordinates, the Cu^{2+} coordinates depend only weakly on exact complex structure.

Tests of the prediction of most probable interspin distance suggest an accuracy of about 2–3 Å, not much worse than the resolution of the crystal structures on which the predictions were based. While this accuracy is similar to that observed for the best MTSSL rotamer libraries, prediction of the width and shape of the distribution is much more accurate for Cu^{2+}-dHis-based labels than for MTSSL. Furthermore, the more rigid Cu^{2+} labels lead to a better accuracy in modeling of conformational transitions based on simulated label-to-label distance restraints and an elastic network model. Although we did not test by either experiments or simulations whether such improvement is also found for other types of modelling structure from label–label distance distributions restraints, this appears likely. For this reason, we believe that further experimental and computational research into Cu^{2+}-dHis labels is desirable. Such research will profit from the in-silico labelling approach introduced here.

References

1. A.D. Milov, K.M. Salikhov, M.D. Shchirov, Fiz. Tverd. Tela **23**, 975 (1981)
2. M. Pannier, S. Veit, A. Godt, G. Jeschke, H.W. Spiess, J. Magn. Reson. **142**, 331 (2000)
3. S. Saxena, J.H. Freed, J. Chem. Phys. **107**, 1317 (1997)
4. P.P. Borbat, J.H. Freed, Chem. Phys. Lett. **313**, 145 (1999)
5. L.V. Kulik, S.A. Dzuba, I.A. Grigoryev, Y.D. Tsvetkov, Chem. Phys. Lett. **343**, 315 (2001)
6. S. Milikisyants, F. Scarpelli, M.G. Finiguerra, M. Ubbink, M. Huber, J. Magn. Reson. **201**, 48 (2009)
7. G. Jeschke, M. Pannier, A. Godt, H.W. Spiess, Chem. Phys. Lett. **331**, 243 (2000)
8. P. Schöps, P.E. Spindler, A. Marko, T.F. Prisner, J. Magn. Reson. **250**, 55 (2015)
9. W.L. Hubbell, D.S. Cafiso, C. Altenbach, Nat. Struct. Biol. **7**, 735 (2000)
10. S.A. Shelke, S.T. Sigurdsson, Eur. J. Org. Chem. **2012**, 2291 (2012)
11. O. Schiemann, T.F. Prisner, Q. Rev. Biophys. **40**, 1 (2007)
12. P.P. Borbat, J.H. Freed, Methods Enzymol. **423**, 52 (2007)

13. H.S. Mchaourab, P.R. Steed, K. Kazmier, Structure **19**, 1549 (2011)
14. G. Jeschke, Annu. Rev. Phys. Chem. **63**, 419 (2012)
15. G. Jeschke, Emerg. Top. Life Sci. **2**, 9 (2018)
16. M.R. Fleissner, M.D. Bridges, E.K. Brooks, D. Cascio, T. Kálai, K. Hideg, W.L. Hubbell, Proc. Natl. Acad. Sci. USA **108**, 16241 (2011)
17. S.M. Islam, B. Roux, J. Phys. Chem. B **119**, 3901 (2015)
18. B. Joseph, V.M. Korkhov, M. Yulikov, G. Jeschke, E. Bordignon, J. Biol. Chem. **289**, 3176 (2014)
19. M. Ji, S. Ruthstein, S. Saxena, Acc. Chem. Res. **47**, 688 (2014)
20. Z. Yang, M. Ji, T.F. Cunningham, S. Saxena, Methods Enzymol. **563**, 459 (2015)
21. I.M.C. van Amsterdam, M. Ubbink, G.W. Canters, M. Huber, Angew. Chem. Int. Ed. **42**, 62 (2003)
22. E. Narr, A. Godt, G. Jeschke, Angew. Chem. Int. Ed. **41**, 3907 (2002)
23. Z. Yang, J. Becker, S. Saxena, J. Magn. Reson. **188**, 337 (2007)
24. B.E. Bode, J. Plackmeyer, T.F. Prisner, O. Schiemann, J. Phys. Chem. A **112**, 5064 (2008)
25. J.E. Lovett, A.M. Bowen, C.R. Timmel, M.W. Jones, J.R. Dilworth, D. Caprotti, S.G. Bell, L.L. Wong, J. Harmer, Phys. Chem. Chem. Phys. **11**, 6840 (2009)
26. Z. Yang, D. Kise, S. Saxena, J. Phys. Chem. B **114**, 6165 (2010)
27. J.S. Becker, S. Saxena, Chem. Phys. Lett. **414**, 248 (2005)
28. S. Ruthstein, M. Ji, P. Mehta, L. Jen-Jacobson, S. Saxena, J. Phys. Chem. B **117**, 6227 (2013)
29. G.E. Merz, P.P. Borbat, A.J. Pratt, E.D. Getzoff, J.H. Freed, B.R. Crane, Biophys. J. **107**, 1669 (2014)
30. J.H. van Wonderen, D.N. Kostrz, C. Dennison, F. MacMillan, Angew. Chem. Int. Ed. **52**, 1990 (2013)
31. Z. Yang, M.R. Kurpiewski, M. Ji, J.E. Townsend, P. Mehta, L. Jen-Jacobson, S. Saxena, Proc. Natl. Acad. Sci. USA **109**, E993 (2012)
32. E.G.B. Evans, M.J. Pushie, K.A. Markham, H.-W. Lee, G.L. Millhauser, Structure **24**, 1057 (2016)
33. H.E.M. Christopher, W.M. Kay, A. Richard Cammack, R.W. Evans, J. Am. Chem. Soc. **129**, 4868 (2007)
34. T.F. Cunningham, M.D. Shannon, M.R. Putterman, R.J. Arachchige, I. Sengupta, M. Gao, C.P. Jaroniec, S. Saxena, J. Phys. Chem. B **119**, 2839 (2015)
35. T.F. Cunningham, M.R. Putterman, A. Desai, W.S. Horne, S. Saxena, Angew. Chem. Int. Ed. **54**, 6330 (2015)
36. M.J. Lawless, S. Ghosh, T.F. Cunningham, A. Shimshi, S. Saxena, Phys. Chem. Chem. Phys. **19**, 20959 (2017)
37. S. Ghosh, M.J. Lawless, G.S. Rule, S. Saxena, J. Magn. Reson. **286**, 163 (2018)
38. J.L. Sarver, J.E. Townsend, G. Rajapakse, L. Jen-Jacobson, S. Saxena, J. Phys. Chem. B **116**, 4024 (2012)
39. G. Jeschke, Protein Sci. **27**, 76 (2018)
40. Y. Polyhach, E. Bordignon, G. Jeschke, Phys. Chem. Chem. Phys. **13**, 2356 (2011)
41. G. Jeschke, J. Chem. Theory Comput. **8**, 3854 (2012)
42. E.A. Suturina, D. Häussinger, K. Zimmermann, L. Garbuio, M. Yulikov, G. Jeschke, I. Kuprov, Chem. Sci. **8**, 2751 (2017)
43. Y. Polyhach, G. Jeschke, Spectroscopy **24**, 651 (2010)
44. M.J. Lawless, J.R. Pettersson, G.S. Rule, F. Lanni, S. Saxena, Biophys. J. **114**, 592 (2018)
45. M.P. Brandi-Blanco, M.M. de Benavides-Giménez, J.M. González-Pérez, D. Choquesillo-Lazarte, Acta Crystallogr. Sect. E **63**, m1678 (2007)
46. F. Neese, Wiley Interdiscip. Rev. Comput. Mol. Sci. **2**, 73 (2012)
47. N. Kumari, M.A. Hasan, B.D. Ward, L. Mishra, Ind. Eng. Chem. Res. **52**, 15007 (2013)
48. C.J. Burns, L.D. Field, T.W. Hambley, T. Lin, D.D. Ridley, P. Turner, M.P. Wilkinson, ARKIVOC **vii**, 157 (2001)
49. J. Gaǎzo, I.B. Bersuker, J. Garaj, M. Kabešová, J. Kohout, H. Langfelderová, M. Melník, M. Serator, F. Valach, Coord. Chem. Rev. **19**, 253 (1976)
50. B. Murphy, B. Hathaway, Coord. Chem. Rev. **243**, 237 (2003)
51. K. Aghaiypour, A. Wlodawer, J. Lubkowski, Biochim. Biophys. Acta Protein Struct. Mol. Enzymol. **1550**, 117 (2001)
52. T. Gallagher, P. Alexander, P. Bryan, G.L. Gilliland, Biochemistry **33**, 4721 (1994)

53. I. Sengupta, P.S. Nadaud, J.J. Helmus, C.D. Schwieters, C.P. Jaroniec, Nat. Chem. **4**, 410 (2012)
54. V. Muñoz, P.A. Thompson, J. Hofrichter, W.A. Eaton, Nature **390**, 196 (1997)
55. R.N. Armstrong, Chem. Res. Toxicol. **10**, 2 (1997)
56. M.T. Honaker, M. Acchione, W. Zhang, B. Mannervik, W.M. Atkins, J. Biol. Chem. **288**, 18599 (2013)
57. Y. Zhan, G.S. Rule, Biochemistry **43**, 7244 (2004)
58. E. Grahn, M. Novotny, E. Jakobsson, A. Gustafsson, L. Grehn, B. Olin, D. Madsen, M. Wahlberg, B. Mannervik, G.J. Kleywegt, Acta Crystallogr. Sect. D Biol. Crystallogr. **62**, 197 (2006)
59. I. Le Trong, R.E. Stenkamp, C. Ibarra, W.M. Atkins, E.T. Adman, Proteins Struct. Funct. Bioinform. **48**, 618 (2002)
60. G. Jeschke, Proteins Struct. Funct. Bioinform. **84**, 544 (2016)
61. W. Zheng, B.R. Brooks, Biophys. J. **90**, 4327 (2006)
62. G. Jeschke, Z. Phys. Chem. **226**, 1395 (2012)