3D protein structure determination using pseudocontact shifts of backbone amide protons generated by double-histidine Co^{2+} -binding motifs at multiple sites

Alireza Bahramzadeh, Thomas Huber*, and Gottfried Otting*

Research School of Chemistry, Australian National University, Canberra, ACT 2601
(Australia)

KEYWORDS: double-histidine- Co^{2+} motif, ERp29, GPS-Rosetta, pseudocontact shifts

ABSTRACT: Pseudocontact shifts (PCS) generated by paramagnetic metal ions contribute highly informative long-range structure restraints that can be measured in solution and are ideally suited to guide structure prediction algorithms in determining global protein folds. We recently demonstrated that PCSs, that are relatively small but of high quality, can be generated by a double histidine (dHis) motif in an α -helix, which provides a well-defined binding site for a single Co^{2+} ion. Here we show that PCSs of backbone amide protons generated by dHis- Co^{2+} motifs positioned in four different α -helices of a protein deliver excellent restraints to determine the three-dimensional (3D) structure of a protein in a way akin to the global positioning system (GPS). We demonstrate the approach with GPS-Rosetta calculations of the 3D structure of the C-terminal domain of the chaperone ERp29 (ERp29-C). Despite the relatively small size of the PCSs generated by the dHis- Co^{2+} motifs, the structure calculations converged readily. Generating PCSs by the dHis- Co^{2+} motif thus presents an excellent alternative to the use of lanthanide tags.

Introduction

Nuclear magnetic resonance (NMR) spectroscopy is a well-established tool to determine the three-dimensional (3D) structure of proteins in solution. While the first structure determinations primarily relied on nuclear Overhauser effects (NOE),¹ a variety of alternative structure restraints can be obtained from NMR experiments to assist with 3D structure determination, including homo- and heteronuclear scalar couplings,² residual dipolar couplings (RDC),³ chemical shifts,⁴ cross-correlated relaxation effects,⁵ paramagnetic relaxation enhancements (PRE),⁶ and pseudocontact shifts.⁷ With the advent of algorithms and data bases that allow modeling of the 3D structure of small proteins from their amino acid sequences with increasing accuracy,⁸ experimental restraints increasingly serve to confirm the correctness of the modeled protein fold rather than as fundamental prerequisites of 3D structure determination. As a result, sparse experimental data become valuable parameters to confirm the accuracy of a 3D structure model. For best computational efficiency, the sparse data need to be provided to the modeling software from the beginning, to aid in the identification of correctly folded polypeptide segments and allow discarding misfolded decoys prior to time-consuming structure refinement. Examples of this approach are CS-Rosetta, where experimental chemical shifts bias the selection of suitable 9-residue peptide segments,^{9, 10} CS-RDC-Rosetta, where RDCs and a small number of backbone NOEs provide additional restraints to guide the folding of larger proteins,^{11,12} and PCS-Rosetta, where long-range restraints are derived from experimental PCSs.¹³

Backbone resonance assignments are key to the analysis of proteins by NMR spectroscopy and, therefore, backbone chemical shifts are a readily available source of experimental structure restraints. While backbone chemical shifts are indicative of secondary structure, they carry much less information about tertiary folds. In this situation, PCSs present

an almost ideal set of complementary restraints, because they are long-range (15 Å and longer)¹⁴ and derived from chemical shifts, which can be measured with more sensitive NMR experiments than NOEs, small scalar couplings, RDCs, or quantitative relaxation rates. In recognition of the benefit of PCSs, PCS-Rosetta was first extended to GPS-Rosetta to integrate PCSs from paramagnetic centers at different sites,¹⁵ and subsequently supplemented by iterative resampling protocols to accelerate convergence of the calculations for larger proteins.^{16, 17}

In the present work, we used PCSs of backbone amide protons only, measured as the difference in chemical shift observed in [¹H,¹⁵N]-HSQC spectra of paramagnetic and diamagnetic samples. The PCS of a nuclear spin can be described by

$$\Delta\delta^{\text{PCS}} = \frac{1}{12\pi r^3} [\Delta\chi_{\text{ax}}(3\cos^2\theta - 1) + 1.5\Delta\chi_{\text{rh}}\sin^2\theta\cos 2\phi] \quad (1)$$

where $\Delta\delta^{\text{PCS}}$ is the PCS (measured in ppm), r , θ , and ϕ are the polar coordinates of the nuclear spin in the frame of the magnetic susceptibility anisotropy tensor $\Delta\chi$, which is described by the axial and rhombic components $\Delta\chi_{\text{ax}}$ and $\Delta\chi_{\text{rh}}$ respectively.¹⁸

To generate PCSs requires a paramagnetic metal ion in a specific location of the protein. Many metal ions with unpaired electrons generate PCSs¹⁹ but the largest and most varied PCSs are generated by lanthanide ions.²⁰ Despite many different approaches developed to attach lanthanide ions site-specifically to proteins,²¹⁻²³ however, site-specific attachment with minimal residual mobility of the metal ion relative to the protein remains difficult. In principle, any translational motions of the metal ion on the protein compromise the validity of equation 1, requiring more than a single $\Delta\chi$ tensor to describe the PCSs. As eight parameters are needed to characterize a single $\Delta\chi$ tensor ($\Delta\chi_{\text{ax}}$, $\Delta\chi_{\text{rh}}$, three coordinates of the paramagnetic centre, three Euler angles describing the tensor orientation), multi- $\Delta\chi$ tensor descriptions are difficult

and the accepted approach is to interpret the PCSs by a single “effective” $\Delta\chi$ tensor as a first approximation.²⁴ The problem of metal mobility is particularly critical when metal tags are ligated to unnatural amino acids, which generates long linkers between metal ion and protein backbone.²⁵⁻²⁷ We recently showed that this problem can be overcome by installing a double-histidine- Co^{2+} (dHis- Co^{2+}) motif in an α -helix, where histidine residues in positions i and $i+4$ of the helix cooperate to bind a single Co^{2+} ion. Although Co^{2+} ions generate smaller PCSs than some of the lanthanide ions, the dHis- Co^{2+} motif features a number of advantages, including exceptional localization of the metal ion in relatively close proximity of the protein backbone, compatibility with naturally occurring cysteine residues in the protein, and straightforward sample preparation.²⁸

To evaluate the performance of the dHis- Co^{2+} motif for PCS-aided 3D structure determination, we used the C-terminal domain of the rat protein ERp29 (ERp29-C) as a model system. ERp29 is a chaperone associated with the endoplasmic reticulum.²⁹ In previous work of ERp29-C, we showed that CS-Rosetta calculations fail to identify the correct protein fold by Rosetta energy despite the provision of chemical shifts and used four different single-cysteine mutants of ERp29-C to demonstrate the performance of GPS-Rosetta with PCSs elicited by IDA-SH^{30, 31} or C1 tags,³² which were attached at four different sites via disulfide bonds and loaded with Tb^{3+} or Tm^{3+} ions.¹⁵ The 3D structure obtained differed significantly from the crystal structure of the human homologue.³³ In the present work, we installed dHis- Co^{2+} motifs at four different sites of rat ERp29-C, measured PCSs using the corresponding Zn^{2+} complexes as diamagnetic reference, and performed GPS-Rosetta calculations. The 3D structure of the previous model determined with PCSs from lanthanide tags was reproduced with a backbone RMSD as low as 1.3 Å.

Experimental Procedures

Construct Design. The codon-optimized gene of wild-type rat ERp29-C was ordered as gBlock from Integrated DNA Technologies (USA), amplified by PCR, and cloned into the T7 expression vector pETMCSI³⁴ between the *NdeI* and *EcoRI* sites using RQ-SLIC.³⁵ The plasmids were transformed into *E. coli* BL21(DE3)pLysS cells for overexpression. The ERp29-C double mutants A163H/Q167H, A218H/A222H, A179H/Q183H, and G207H/D211H, referred to in the following as Mu1, Mu2, Mu3, and Mu4, respectively, were prepared by point mutagenesis using two overlapping primers.³⁵ An initial construct of ERp29-C, which comprised residues M154 to L260, produced only low protein yields. Therefore, the final constructs were prepared with the addition of a N-terminal His₆-tag followed by a TEV cleavage site, which added a glycine residue N-terminal of M154 but significantly increased the protein yield.

Protein Expression and Purification. Cells were grown at 37 °C in LB medium (10 g/L tryptone, 5 g/L yeast extract, and 5 g/L NaCl) containing 100 mg/L ampicillin and 34 mg/L chloramphenicol. Following cell growth to OD₆₀₀ = 0.8–1.0, the cells were centrifuged and the pellet resuspended in M9 medium (6 g/L Na₂HPO₄, 3 g/L KH₂PO₄, 0.5 g/L NaCl) supplied with 1 g/L ¹⁵NH₄Cl and 3 g/L ¹³C-glucose for uniform ¹⁵N/¹³C-labeling. Following further growth for 1 h, isopropyl-β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM to induce protein expression. After induction, the cells were incubated overnight at room temperature. The cell culture was centrifuged at 5,000 g for 15 minutes at 4 °C and the cells were lysed in lysis buffer (10% glycerol, 1 mM PMSF, 2 mM 2-mercaptoethanol, 150 mM NaCl, 50 mM sodium phosphate, pH 6.8) using sonication. The supernatant was applied to a 5 mL Ni-NTA column (GE Healthcare, USA) and bound protein was eluted with a gradient of 10–500 mM imidazole in 150 mM NaCl and 50 mM sodium

phosphate, pH 6.8. TEV protease (MHT238Δ)³⁶ was added to purified protein and the mixture was dialysed overnight at 4 °C against 50 mM Tris-HCl, 300 mM NaCl, and 1 mM 2-mercaptoethanol. The cleavage yield was about two thirds as estimated from SDS-PAGE. The mixture was loaded onto a 5 mL Ni-NTA column to separate the final product from the His₇-tagged TEV protease. Following elution, the purified proteins were treated with EDTA in 5-fold excess to remove any divalent metal ions bound to the dHis motif and the buffer was exchanged to NMR buffer (50 mM NaCl, 20 mM MES, pH = 6.5) using an Amicon Ultra-15 centrifugal filter unit with Ultracel-3 membrane (Merck Millipore). D₂O was added to a final ratio of 10%D₂O/90%H₂O and tris(2-carboxyethyl)phosphine (TCEP) was added to a final concentration of 1 mM to prevent oxidation of C157. Protein yields were about 12 mg of purified protein per litre of medium. SDS-PAGE indicated >95% purity.

NMR Spectroscopy. All NMR data were recorded at 25 °C on a Bruker Avance II 800 MHz NMR spectrometer equipped with a TCI cryoprobe. [¹⁵N,¹H]-HSQC spectra were recorded using $t_{1\max} = 40$ ms, $t_{2\max} = 106$ ms, and total recording times of about 40 min. BEST-TROSY versions of HNCO, HNCACO, HNCA, HNCACB, and HNCOCACB spectra were recorded of wild-type ERp29-C and all four mutants without metal ion using the IBS pulse sequence tools,³⁷ and the backbone resonances were assigned using CCPNMR.³⁸ The measurements were performed using between 0.14 and 0.94 mM solutions of uniformly ¹⁵N/¹³C-labeled proteins.

PCS Measurements. PCSs were measured as the ¹H chemical shift observed in [¹⁵N,¹H]-HSQC spectra with CoCl₂ minus the chemical shift measured for a sample containing ZnCl₂. Co²⁺ salts are known to change the ¹H chemical shift of the water resonance.³⁹ Therefore, the calibration of the ¹H carrier frequency (water resonance) was corrected by 0.0078 ppm per 1 mM Co²⁺ ion concentration.

Structure Calculations. The GPS-Rosetta algorithm is an extension of the CS-Rosetta algorithm. It includes PCS data and $\Delta\chi$ -tensor fits early in the Rosetta calculations to restrict the sampling space and discard ill-fitting structures before any further structure refinement.^{13,15} GPS-Rosetta was supplied with the backbone chemical shifts of the protein ($^1\text{H}\text{N}$, N , C^α , C^β , C') for initial selection of fragments from the Rosetta library of proteins, which was modified to exclude any structures homologous to ERp29-C. PCSs of backbone amide protons (Table S2) were used in the fragment assembly step, where fits of $\Delta\chi$ tensors allowed discrimination between successful and unsuccessful models.¹⁵ At the fragment assembly step, the $\Delta\chi$ -tensor fits explored optimal metal positions within a grid centered between the C^α atom of residue i and the C^β atom of residue $i+4$ of the dHis motif, followed by an off-grid tensor optimization starting from the best grid position. In addition, the final models calculated of the whole protein were screened for good fits between back-calculated and experimental PCSs following restriction of the metal coordinates to the position expected for the α -helical dHis motif.

Results

Protein Expression and Purification. In principle, a dHis motif alone is sufficient to allow purification by affinity chromatography on a Ni-NTA column,^{28, 40} but the expression levels of the ERp29-C samples without His₆ tag were insufficient for facile purification on a Ni-NTA column. With an additional His₆ tag, the ERp29-C samples with dHis motif bound very tightly, requiring 500 mM imidazole to elute the protein from the column. The presence of the His₆-tag resulted in significantly better final protein yields.

dHis motifs are straightforward to implement by positioning two histidyl residues in the hydrophilic face of an amphiphilic α -helix with a spacing of three residues. The following

ten mutants of ERp29-C were designed with dHis motif: A163H/Q167H (Mu1), Q178H/K182H, A179H/Q183H (Mu3), K197H/Q201H, S200H/K204H, G207H/D211H (Mu4), A218H/A222H (Mu2), G235H/E239H, Q241H/N245H, and L244H/T248H, but acceptable expression yields were obtained only with the mutants Mu1–Mu4.

PCS Measurements. Only the four mutants Mu1–Mu4 displayed PCSs following titration with CoCl₂, suggesting that the dHis-Co²⁺ motif did not always form. When samples were titrated with ZnCl₂ instead of CoCl₂ to obtain the chemical shifts of corresponding diamagnetic references, some pronounced precipitation occurred over a period of hours, which was not observed with CoCl₂. The intrinsic widths of the NMR signals of nuclear spins located far from the metal center did not change during titration with CoCl₂ or ZnCl₂, indicating that the protein remained monomeric. In the absence, but not the presence, of TCEP, additional cross-peaks appeared with time, indicating sensitivity of C157 to oxygen. Therefore, all spectra were recorded in the presence of 1 mM TCEP.

The ERp29-C samples with dHis motif were titrated with metal ions in equimolar ratio except for Mu4, where a five-fold excess of CoCl₂ was needed to observe PCSs. There was no evidence for alternative binding sites for Co²⁺ ions in ERp29, as PREs were evident only in the vicinity of the dHis motif. Although the PCSs observed were relatively small, PCSs were nonetheless observable for most of the protein (Figure 1).

Figure 1. PCSs induced by a dHis-Co²⁺ motif at four different sites of ERp29-C. (a) Superimpositions of [¹⁵N,¹H]-HSQC spectra of a 260 μ M solution of uniformly ¹⁵N-labeled ERp29-C A163H/Q167H (Mu1) in NMR buffer (20 mM MES, 50 mM NaCl, pH 6.5) with either ZnCl₂ (diamagnetic reference, red cross-peaks) or CoCl₂ (paramagnetic spectrum, blue cross-peaks) added in equimolar ratio. The spectra were recorded at 25 °C. Assignments are shown for cross-peaks of backbone amides with significant PCSs. (b) Same as (a), but for a 140 μ M solution of ERp29-C A218H/A222H (Mu2). (c) Same as (a), but for a 150 μ M solution of ERp29-C A179H/Q183H (Mu3). (d) Same as (a), but for a 320 μ M solution of ERp29-C G207H/D211H (Mu4).

As the PCSs were small, most of the cross-peaks in the paramagnetic state were readily assigned by comparison with the spectrum recorded with ZnCl_2 . To complete the resonance assignments of the paramagnetic samples, the PCSs measured from the resolved cross-peaks were used to fit $\Delta\chi$ tensors to the published coordinates of ERp29-C (PDB ID: 2M66¹⁵) using the program Numbat,⁴¹ which allowed prediction of all remaining PCSs and resolution of ambiguous cross-peak assignments in the paramagnetic state. The $\Delta\chi$ tensors and PCSs are listed in Tables S1 and S2, respectively.

In total, the dHis- Co^{2+} motif in the mutants Mu1–Mu4 delivered 286 PCSs, whereas only 212 PCSs had been assigned in our previous work using two different lanthanide ions at four different sites.¹⁵ Figure 2 shows that at least one PCS was obtained for all non-proline residues, two PCSs were observed for over 90% of residues 161–256 (i.e. excluding the flexible N- and C-terminal segments) and three PCSs were observed for 65% of the protein.

Figure 2. Summary of the experimental amide proton PCSs. The names of the different mutants are indicated on the left. Arrows identify the residues mutated to histidine in the different mutants. Red, blue, green, and yellow bars identify the residues for which the dHis- Co^{2+} motif delivered significant PCSs for the backbone amide protons in the mutants Mu1, Mu2, Mu3, and Mu4, respectively. The residue numbering is shown above the amino acid sequence. The PCS data are listed in Table S2.

Structure Computation. The PCSs were used as input into the program GPS-Rosetta to calculate the 3D structure of the protein. 20,000 all-atom models were calculated using a non-homologous fragment library that explicitly excluded any fragments from previously determined structures of ERp29. As expected for good convergence in the Rosetta calculations,

the lowest combined Rosetta + PCS energies were observed for models similar to the target structure 2M66 (Figure 3a). As the metal position in the dHis-Co²⁺ motif can be predicted accurately, the models were screened for the subset of models with low PCS deviations, where the metal coordinates at each of the four dHis sites fulfilled the expected geometry of the dHis-Co²⁺ motif (Figure 3b). Selecting the models, for which the sum of the squared deviation (SD) of experimental and back-calculated PCSs was less than 0.244 ppm² (the value obtained by fitting $\Delta\chi$ tensors to the first conformer of the structure 2M66), we obtained about 1,000 models, which were consistently similar to the target structure (red and blue points in Figure 3a and referred to in the following as Top1k). Finally, we applied the lowest combined PCS + Rosetta score as an additional selection criterion to identify the ten best models, referred to in the following as Top10 (blue points in Figure 3a). Although the input had not used any experimental restraints for the amino acid side chains, the Top10 conformers displayed well-defined conformations for some of the buried amino acid side chains (Figure S2).

Using PyMOL⁴² for optimal structure alignment, the conformer top1 delivered a C ^{α} RMSD value of about 1.4 Å to the first conformer of 2M66 for 96 aligned residues. All ten conformers displayed a lower C ^{α} RMSD with respect to the GPS-Rosetta structure 2M66 than to the crystal structure 2QC7 of human ERp29 (Figure S3),³³ but this difference is less pronounced when the comparison includes all C-terminal residues for which the crystal structure reports electron density (Figure S4).

Figure 3. PCS and Rosetta energy scores of 20,000 GPS-Rosetta structure calculations of ERp29-C using PCSs from dHis-Co²⁺ motifs, and distance restraints used. (a) Combined PCS + Rosetta energy scores of the structural models calculated in the present work plotted versus the C^α RMSD to the first conformer of the structure determined previously using PCSs from lanthanide tags (PDB ID: 2M66). The alignment excluded the disordered C-terminal residues 256–260. Black points mark the result for all 20,000 structures obtained from GPS-Rosetta. Blue and red points mark the approximately 1000 structures fulfilling the PCS restraints with an SD value of less than 0.244 ppm², when the metal positions were defined by the distance restraints specified in (b) (structures referred to as Top1k). Blue points identify the 10 structures (Top10) with the lowest combined PCS + Rosetta score among the Top1k structures selected for detailed analysis. (b) C^α-Co²⁺ and C^β-Co²⁺ distance restraints used to determine a metal position for each model, which is in agreement with the location expected for the α -helical dHis-Co²⁺ motif, where the histidine residues are located in positions *i* and *i*+4.²⁸ For each model, metal coordinates were determined by a least squares fit to match the restraints and these coordinates were used for $\Delta\chi$ tensor fits. The restraints correspond to the average

distances found in the ubiquitin E24H/A28H and A28H/D32H mutants, using the Co^{2+} ion position determined from PCSs and modeling the histidine side chain conformations.²⁸

To assess the contribution of the PCS data to the structure calculation, the performance of each PCS dataset was analysed individually. All four PCS datasets contributed to the convergence of the GPS-Rosetta modelling (Figure S5). As expected for good convergence, the $\Delta\chi$ tensors fitted to the individual conformers of the Top10 ensemble were all very similar to each other and to the $\Delta\chi$ tensor fitted to the first conformer of 2M66 (Table S3). The structural similarity between the Top10 ensemble and 2M66 is also apparent when the Top10 conformers are individually aligned with the first conformer of 2M66 in a global superimposition (Figure S6). Clearly, helices 5 to 7 in all top10 structures are highly similar to 2M66, whereas helix 8 and, in particular, helix 9 are more variable. Notably, helix 9 is straight in 5 of 11 conformers in the structure 2M66, but has a kink in the remaining 6 conformers.

Figure 4 illustrates the structural similarity between the first conformer of the 2M66 structure and the model with the lowest PCS + Rosetta Energy (in the following referred to as top1) determined in the present work. The PCS SD value for this structure is only 0.184 ppm² (corresponding to an RMSD of 0.025 ppm) and the correlation between back-calculated and experimental PCSs, shown in Figure 5a, shows no sign of systematic deviations. The quality factors associated with the $\Delta\chi$ tensor fits to top1 were 0.18, 0.29, 0.19, and 0.35 for Mu1 – Mu4, respectively, and thus better than the Q -factors obtained when the same PCSs were used to fit $\Delta\chi$ tensors to the first conformer of 2M66, which were 0.32, 0.28, 0.37, and 0.35, respectively. The improved fit of the mutants Mu1 and Mu3 compared to the 2M66 structure is also apparent in the correlation plot of back-calculated versus experimental PCSs (Figure 5b).

Figure 4. Locations of Co^{2+} ions relative to the 3D structure of ERp29-C with the lowest combined PCS + Rosetta energy (top1) and comparison of top1 with the first conformer of the structure 2M66¹⁵ previously determined using PCSs generated by lanthanide tags. (a) Ribbon representation of the structural model top1 marking the positions of the histidyl residues of the dHis- Co^{2+} motifs of the different mutants in cyan and the Co^{2+} ions in blue. (b) Superimposition of the model top1 with the structure 2M66. The SD value of the PCSs of top1 is 0.18 ppm² and the C $^{\alpha}$ RMSD of top1 to 2M66 is 1.4 Å for 95 atoms aligned.

Figure 5. Correlations between back-calculated and experimental PCSs generated by dHis- Co^{2+} motifs in rat ERp29-C. The $\Delta\chi$ tensors were fitted with the metal position fixed to the locations reported in Table S3. (a) Correlation plots following $\Delta\chi$ tensor fits to top1. (b) Correlation plots following $\Delta\chi$ tensor fits to the first conformer of 2M66.

Conversely, the PCSs determined with Tm^{3+} and Tb^{3+} tags in previous work¹⁵ fit somewhat better to the structure 2M66 than to the top1 conformer, highlighting the influence of PCSs on the final structure. Nonetheless, the Q -factors obtained from the fits to top1 were of comparable quality as those obtained with the original structure 2M66 (Table 1).

Table 1. Quality factors obtained from fitting $\Delta\chi$ tensors to the structure top1, using the PCSs reported previously for single-cysteine mutants of ERp29-C labeled with C1- Tb^{3+} and C1- Tm^{3+} tags.^{15,a}

ERp29-C mutant ^b	top1		2M66 ^c	
	Tb	Tm	Tb	Tm
C157	0.13	0.13	0.10	0.08
C200	0.23	0.24	0.22	0.20
C218	0.15	0.15	0.08	0.13
C241	0.08	0.19	0.06	0.12

^a PCSs are of backbone amide protons as reported previously.¹⁵ The C1- Tb^{3+} and C1- Tm^{3+} tags were attached to cysteine residues at four different sites. The program Numbat⁴¹ was used to fit the $\Delta\chi$ tensors. Fits included the optimization of metal coordinates. Quality factors of the fits were calculated as the root-mean-square (RMS) of the differences between experimental and back-calculated PCSs divided by the RMS of the experimental PCSs.

^b The respective single-cysteine mutants are named by the position of the cysteine residue in ERp29-C.

^c Using the first conformer of 2M66.

Discussion

The C-terminal domain of rat ERp29 is a 12 kDa protein comprising 106 residues. The fold of ERp29-C comprises five α -helices and therefore is ideally suited for assessing the performance

of the dHis-Co²⁺ motif in generating PCSs, as the dHis-Co²⁺ motif is more readily installed in α -helices than in other secondary structure elements.²⁸ In the present work, the dHis motifs were installed in four different helices, in analogy to previous GPS-Rosetta calculations, where lanthanide tags were installed in different helices.¹⁵ In contrast to the previous work, where lanthanide tags were attached to single cysteinyl residues and thus required mutation of the naturally occurring cysteine residue (C157) to serine, C157 was retained in all dHis mutants of the present work.

Like any paramagnetic metal ion, Co²⁺ ions generate PREs which broaden the NMR signals of nearby nuclear spins beyond detection. In the present study, no [¹⁵N,¹H]-HSQC cross-peaks could be observed for residues three positions prior to or following the histidyl residues of the dHis-Co²⁺ motif (Table S2). To measure PCSs for the entire length of the amino acid sequence, it is thus important to position dHis motifs at least at two different sites. Quite generally, the GPS principle works best, if a nuclear spin displays PCSs in four different measurements conducted with paramagnetic centers at different locations, as the intersection between two, three, or four PCS isosurfaces constitutes a line, a choice of two points, or a single point, respectively.⁴³ Although the $\Delta\chi$ tensors associated with dHis-Co²⁺ motifs are relatively small (Table S3), the coverage obtained for ERp29-C (Figure 2) was better than in our previous work, even though that work had employed two different paramagnetic lanthanides with different $\Delta\chi$ tensor magnitudes and orientations (Tm³⁺ and Tb³⁺ tags) at each site. The present work thus demonstrates that PCSs from a single paramagnetic ion can be sufficient to guide GPS-Rosetta calculations to determine the 3D structure of the protein from the PCSs of backbone amide protons only.

Availability of samples with two different paramagnetic metal ions that generate PCSs of opposite sign helps in assigning the NMR spectra of the paramagnetic states.⁴⁴ As only a

single set of PCSs is available with a dHis-Co²⁺ motif, the resonance assignment of the paramagnetic state is less straightforward for cross-peaks with large PCSs. In the present study, we used $\Delta\chi$ tensor fits to the known target structure to assist with the assignment of these cross-peaks. Alternatively, the assignments could have been supported by the preparation of selectively isotope-labeled samples to improve the resolution of the [¹⁵N,¹H]-HSQC spectrum. Preparation of multiple selectively isotope-labeled samples with dHis-Co²⁺ motifs is relatively economical, because no chemical reactions with expensive tags are required. Unfortunately, I₂S_z-exchange experiments, which we had successfully conducted with ubiquitin samples containing a dHis motif,²⁸ did not reveal exchange cross-peaks of significant intensity in the case of the ERp29-C mutants.

The Rosetta algorithm is one of the most successful methods for *de novo* protein structure prediction. For small proteins (<100 residues), it has been shown to deliver correct structures with atomic level accuracy in blind prediction challenges.⁴⁵ For larger proteins, however, Rosetta becomes increasingly less efficient and more reliant on experimental data to guide the structural search towards the global minimum. In particular, restraints from residual dipolar couplings and long-range NOEs have been shown to assist convergence of Rosetta calculations,^{11,12} but such data are considerably more difficult to measure than the chemical shift changes arising from PCSs. In addition, the long-range nature of PCSs is particularly helpful to guide Rosetta calculations. The present work shows that the relatively small PCSs generated by the dHis-Co²⁺ motif are fit for this purpose, even if they are measured only for backbone amide protons. We have previously shown that CS-Rosetta calculations of ERp29-C without PCSs fail to identify the correct protein fold.¹⁵

While chemical lanthanide tags invariably produce flexible tethers between lanthanide ion and protein,⁴⁶ dHis motifs are capable of holding the Co²⁺ ion in a uniquely well-defined

position relative to the protein backbone.²⁸ As a result, one might expect lower quality factors for the $\Delta\chi$ tensor fits than observed in the present work (Table S3). Quite generally, however, it is easier to obtain small Q -factors with large PCSs, as the denominator in the Q -factor definition becomes larger and the numerator less sensitive to experimental uncertainties caused by small differences in sample preparation. Indeed, the final structures yielded similarly good Q -factors for the Top10 conformers as for the first conformer of the 2M66 structure, irrespective of fitting $\Delta\chi$ tensors with PCSs from lanthanide ions (Table 1) or PCSs from dHis-Co²⁺ motifs (Table S3). Unfortunately, the absence of any experimental structure restraints for buried amino acid side chains creates a sampling problem that is not easily overcome by increasing the weight of the PCS versus the Rosetta energy.

Conclusion

PCSs of backbone amide protons generated by dHis-Co²⁺ motifs at multiple sites open a new and practical approach to harness the power of PCS restraints for protein fold determination. Unexpectedly, it proved easier to achieve a high PCS coverage of the protein with the dHis mutants than with lanthanide tags, despite the smaller magnitude of PCSs arising from a Co²⁺ ion than from Tm³⁺ or Tb³⁺ ions. In view of the economic advantage of generating PCSs with natural amino acids rather than synthetic lanthanide tags and the ease with which the solvent-exposed surface of an amphipathic helix can be identified from the amino acid sequence, the α -helical dHis-Co²⁺ motif carries great promise for the determination of the 3D fold of proteins, for which it is difficult to measure NMR parameters other than backbone amide chemical shifts.

PDB accession code

The coordinates and structure factors of the Top10 models have been deposited in the Protein Data Bank under the entry code 6O6I. The chemical shifts of wild-type ERp29-C were deposited in the BioMagResBank⁴⁷ under entry ID 30583.

Associated Content

Supporting Information

$\Delta\chi$ tensor parameters of ERp29-C fitted to the structure 2M66 (Table S1), Correlation between back-calculated and experimental PCSs for the four different mutants of ERp29-C (Figure S1), Experimental PCSs of the backbone amide protons of ERp29-C with a dHis-Co²⁺ motif at four different sites (Table S2), $\Delta\chi$ tensor parameters and cobalt ion coordinates relative to the first conformer of 2M66 and the individual conformers of the Top10 ensemble (Table S3), Comparison of the Top10 conformers to the first conformer of the structure 2M66 and the crystal structure 2QC7 (Figure S2), Performance of the PCS data set from each dHis-Co²⁺ motif in the GPS-Rosetta calculations (Figure S3), Comparison of the individual α -helices of the Top10 conformers with the first conformer of 2M66 (Figure S4)

Authors information

Corresponding Author

*E-mail: t.huber@anu.edu.au

*E-mail: gottfried.otting@anu.edu.au

Phone: +61 2 61256507

ORCID

Alireza Bahramzadeh: 0000-0002-6857-0961

Gottfried Otting: 0000-0002-0563-0146

Thomas Huber: 0000-0002-3680-8699

Funding

Financial support by the Australian Research Council, including a Discovery Project (DP170100162) and Laureate Fellowship to G. O. (FL170100019), is gratefully acknowledged.

Abbreviations

ERp29-C, C-terminal domain of the endoplasmic reticulum protein p29; dHis-Co²⁺, double histidine motif with a bound Co²⁺ ion; NMR, nuclear magnetic resonance; PCS, pseudocontact shifts.

References

- (1) Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.

- (2) Vuister, G. W., Tessari, M., Karimi-Nejad, Y., and Whitehead, B. (2002) Pulse sequences for measuring coupling constants. In *Modern Techniques in Protein NMR* (Krishna N.R., B. L. J., Ed.), pp 195–257, Springer, Boston, MA.
- (3) Tjandra, N. and Bax, A. (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* 278, 1111–1114.
- (4) Wishart, D. S. (2011) Interpreting protein chemical shift data. *Prog. NMR Spectrosc.* 58, 62–87.
- (5) Reif, B., Hennig, M., and Griesinger, C. (1997) Direct measurement of angles between bond vectors in high-resolution NMR. *Science* 276, 1230–1233.
- (6) Battiste, J. L. and Wagner, G. (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear Overhauser effect data. *Biochemistry* 39, 5355–5365.
- (7) Bertini, I., Luchinat, C., and Parigi, G. (2002) Paramagnetic constraints: an aid for quick solution structure determination of paramagnetic metalloproteins. *Concepts Magn. Reson.* 14, 259–286.
- (8) Das, R. and Baker, D. (2008) Macromolecular modeling with Rosetta. *Annu. Rev. Biochem.* 77, 363–382.
- (9) Shen, Y., Vernon, R., Baker, D., and Bax, A. (2009) De novo protein structure generation from incomplete chemical shift assignments. *J. Biomol. NMR* 43, 63–78.
- (10) Shen, Y., Lange, O., Delaglio, F., Rossi, P., Aramini, J. M., Liu, G., Eletsky, A., Wu, Y., Singarapu, K. K., and Lemak, A. (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci.* 105, 4685–4690.
- (11) Raman, S., Lange, O. F., Rossi, P., Tyka, M., Wang, X., Aramini, J., Liu, G., Ramelot, T. A., Eletsky, A., Szyperski, T., Kennedy, M. A., Prestegard, J., Montelione, G. T., and Baker,

D. (2010) NMR structure determination for larger proteins using backbone-only data. *Science* 327, 1014–1018.

(12) Lange, O. F. and Baker, D. (2012) Resolution-adapted recombination of structural features significantly improves sampling in restraint-guided structure calculation. *Proteins* 80, 884–895.

(13) Schmitz, C., Vernon, R., Otting, G., Baker, D., and Huber, T. (2012) Protein structure determination from pseudocontact shifts using Rosetta. *J. Mol. Biol.* 416, 668–677.

(14) Allegrozzi, M., Bertini, I., Janik, M. B., Lee, Y.-M., Liu, G., and Luchinat, C. (2000) Lanthanide-induced pseudocontact shifts for solution structure refinements of macromolecules in shells up to 40 Å from the metal ion. *J. Am. Chem. Soc.* 122, 4154–4161.

(15) Yagi, H., Pilla, K. B., Maleckis, A., Graham, B., Huber, T., and Otting, G. (2013) Three-dimensional protein fold determination from backbone amide pseudocontact shifts generated by lanthanide tags at multiple sites. *Structure* 21, 883–890.

(16) Pilla, K. B., Otting, G., and Huber, T. (2016) Pseudocontact shift-driven iterative resampling for 3D structure determinations of large proteins. *J. Mol. Biol.* 428, 522–532.

(17) Pilla, K. B., Otting, G., and Huber, T. (2017) Protein structure determination by assembling super-secondary structure motifs using pseudocontact shifts. *Structure* 25, 559–568.

(18) Bertini, I., Luchinat, C., and Parigi, G. (2002) Magnetic susceptibility in paramagnetic NMR. *Prog. NMR Spectrosc.* 40, 249–273.

(19) Bertini, I. and Luchinat, C. (1986) *NMR of Paramagnetic Molecules in Biological Systems*, Benjamin-Cummings.

(20) Bertini, I., Janik, M. B., Lee, Y.-M., Luchinat, C., and Rosato, A. (2001) Magnetic susceptibility tensor anisotropies for a lanthanide ion series in a fixed protein matrix. *J. Am. Chem. Soc.* 123, 4181–4188.

- (21) Su, X.-C. and Otting, G. (2010) Paramagnetic labelling of proteins and oligonucleotides for NMR. *J. Biomol. NMR* 46, 101–112.
- (22) Keizers, P. H. and Ubbink, M. (2011) Paramagnetic tagging for protein structure and dynamics analysis. *Prog. NMR Spectrosc.* 58, 88–96.
- (23) Nitsche, C. and Otting, G. (2017) Pseudocontact shifts in biomolecular NMR using paramagnetic metal tags. *Prog. NMR Spectrosc.* 98, 20–49.
- (24) Shishmarev, D. and Otting, G. (2013) How reliable are pseudocontact shifts induced in proteins and ligands by mobile paramagnetic metal tags? A modelling study. *J. Biomol. NMR* 56, 203–216.
- (25) Loh, C. T., Ozawa, K., Tuck, K. L., Barlow, N., Huber, T., Otting, G., and Graham, B. (2013) Lanthanide tags for site-specific ligation to an unnatural amino acid and generation of pseudocontact shifts in proteins. *Bioconjugate Chem.* 24, 260–268.
- (26) Loh, C.-T., Graham, B., Abdelkader, E. H., Tuck, K. L., and Otting, G. (2015) Generation of pseudocontact shifts in proteins with lanthanides using small “clickable” nitrilotriacetic acid and iminodiacetic acid tags. *Chem. Eur. J.* 21, 5084–5092.
- (27) Abdelkader, E. H., Yao, X., Feintuch, A., Adams, L. A., Aurelio, L., Graham, B., Goldfarb, D., and Otting, G. (2016) Pulse EPR-enabled interpretation of scarce pseudocontact shifts induced by lanthanide binding tags. *J. Biomol. NMR* 64, 39–51.
- (28) Bahramzadeh, A., Jiang, H., Huber, T., and Otting, G. (2018) Two histidines in an α -helix: a rigid Co^{2+} -binding motif for PCS measurements by NMR spectroscopy. *Angew. Chem. Int. Ed.* 57, 6226–6229.
- (29) Mkrtchian, S. and Sandalova, T. (2006) ERp29, an unusual redox-inactive member of the thioredoxin family. *Antioxidants* 8, 325–337.

- (30) Swarbrick, J. D., Ung, P., Chhabra, S., and Graham, B. (2011) An iminodiacetic acid based lanthanide binding tag for paramagnetic exchange NMR spectroscopy. *Angew. Chem. Int. Ed.* 50, 4403–4406.
- (31) Yagi, H., Maleckis, A., and Otting, G. (2013) A systematic study of labelling an α -helix in a protein with a lanthanide using IDA-SH or NTA-SH tags. *J. Biomol. NMR* 55, 157–166.
- (32) Graham, B., Loh, C. T., Swarbrick, J. D., Ung, P., Shin, J., Yagi, H., Jia, X., Chhabra, S., Barlow, N., Pintacuda, G., Huber, T., and Otting, G. (2011) DOTA-amide lanthanide tag for reliable generation of pseudocontact shifts in protein NMR spectra. *Bioconjugate Chem.* 22, 2118–2125.
- (33) Barak, N. N., Neumann, P., Sevvana, M., Schutkowski, M., Naumann, K., Malešević, M., Reichardt, H., Fischer, G., Stubbs, M. T., and Ferrari, D. M. (2009) Crystal structure and functional analysis of the protein disulfide isomerase-related protein ERp29. *J. Mol. Biol.* 385, 1630–1642.
- (34) Neylon, C., Brown, S. E., Kralicek, A. V., Miles, C. S., Love, C. A., and Dixon, N. E. (2000) Interaction of the *Escherichia coli* replication terminator protein (Tus) with DNA: a model derived from DNA-binding studies of mutant proteins by surface plasmon resonance. *Biochemistry* 39, 11989–11999.
- (35) Qi, R. and Otting, G. (2019) Mutant T4 DNA polymerase for easy cloning and mutagenesis. *PLOS One* 14, e0211065.
- (36) Blommel, P. G. and Fox, B. G. (2007) A combined approach to improving large-scale production of tobacco etch virus protease. *Protein Expr. Purif.* 55, 53–68.
- (37) Solyom, Z., Schwarten, M., Geist, L., Konrat, R., Willbold, D., and Brutscher, B. (2013) BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *J. Biomol. NMR* 55, 311–321.

- (38) Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59, 687–696.
- (39) Luz, Z. and Shulman, R. (1965) Proton magnetic resonance shifts in aqueous solutions of paramagnetic metal ions. *J. Chem. Phys.* 43, 3750–3756.
- (40) Arnold, F. H. and Haymore, B. L. (1991) Engineered metal-binding proteins: purification to protein folding. *Science* 252, 1796–1798.
- (41) Schmitz, C., Stanton-Cook, M. J., Su, X.-C., Otting, G., and Huber, T. (2008) Numbat: an interactive software tool for fitting $\Delta\chi$ -tensors to molecular coordinates using pseudocontact shifts. *J. Biomol. NMR* 41, 179–189.
- (42) The PyMOL Molecular Graphics system, Version 2.0 Schrödinger, LLC.
- (43) Otting, G. (2008) Prospects for lanthanides in structural biology by NMR. *J. Biomol. NMR* 42, 1–9.
- (44) John, M. and Otting, G. (2007) Strategies for measurements of pseudocontact shifts in protein NMR spectroscopy. *ChemPhysChem* 8, 2309–2313.
- (45) Raman, S., Vernon, R., Thompson, J., Tyka, M., Sadreyev, R., Pei, J., Kim, D., Kellogg, E., DiMaio, F., Lange, O., Kinch, L., Sheffler, W., Kim, B.-H., Das, R., Grishin, N. V., and Baker, D. (2009) Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins* 77, 89–99.
- (46) Welegedara, A. P., Yang, Y., Lee, M. D., Swarbrick, J. D., Huber, T., Graham, B., Goldfarb, D., and Otting, G. (2017) Double-arm lanthanide tags deliver narrow Gd^{3+} – Gd^{3+} distance distributions in double electron–electron resonance (DEER) measurements. *Chem. Eur. J.* 23, 11694–11702.

(47) Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C. F., Tolmie, D. E., Wenger, K. R., Yao, H., and Markley, J. L. (2007) BioMagResBank. *Nucleic Acids Res.* 36, D402–D408.

For Table of Contents Only