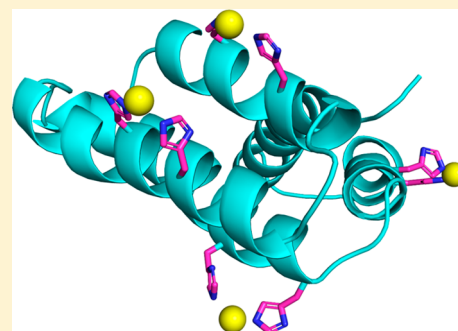# Three-Dimensional Protein Structure Determination Using Pseudocontact Shifts of Backbone Amide Protons Generated by Double-Histidine Co²⁺-Binding Motifs at Multiple Sites

Alireza Bahramzadeh, Thomas Huber,* and Gottfried Otting*

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

S *Supporting Information*

ABSTRACT: Pseudocontact shifts (PCSs) 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, which 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 Co²⁺ ion. Here we show that PCSs of backbone amide protons generated by dHis-Co²⁺ 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-Co²⁺ motifs, the structure calculations converged readily. Generating PCSs by the dHis-Co²⁺ motif thus presents an excellent alternative to the use of lanthanide tags.

Nuclear magnetic resonance (NMR) spectroscopy is a well-established tool for determining the three-dimensional (3D) structure of proteins in solution. While the first structure determinations primarily relied on nuclear Overhauser effects (NOEs),[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 (RDCs),[3] chemical shifts,[4] cross-correlated relaxation effects,[5] paramagnetic relaxation enhancements (PREs),[6] and pseudocontact shifts (PCSs).[7] With the advent of algorithms and databases 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 for confirming the accuracy of a 3D structure model. For the 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 of misfolded decoys prior to time-consuming structure refinement. Examples of this approach are CS-Rosetta, in which experimental chemical shifts bias the selection of suitable nine-residue peptide segments,[9,10] CS-RDC-Rosetta, in which RDCs and a small number of backbone NOEs provide additional restraints to guide the folding of larger proteins,[11,12] and PCS-Rosetta, in which 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)[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 work presented here, we used PCSs of backbone amide protons only, measured as the difference in the chemical shift observed in ¹H−¹⁵N HSQC spectra of paramagnetic and diamagnetic samples. The PCS of a nuclear spin can be described by

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

(1)

where $\Delta\delta^{PCS}$ is the PCS (measured in parts per million) and 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 $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$, respectively.[18]

Generating PCSs requires a paramagnetic metal ion in a specific location of the protein. Many metal ions with unpaired electrons generate PCSs,[19] 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 eq 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 ($\Delta\chi_{ax}$, $\Delta\chi_{rh}$, three coordinates of the paramagnetic center, and 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 the 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 PCSs smaller than those of some of the lanthanide ions, the dHis-Co^{2+} motif features a number of advantages, including exceptional localization of the metal ion relatively close to 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 rat protein ERp29 (ERp29-C) as a model system. ERp29 is a chaperone associated with the endoplasmic reticulum.[29] In previous work with 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 was similar to the crystal structure of the human homologue with a C^{α} root-mean-square deviation (RMSD) of 2−2.9 Å.[15,33] In the work presented here, we installed dHis-Co^{2+} motifs at four different sites of rat ERp29-C, measured PCSs using the corresponding Zn^{2+} complexes as a 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, amplified by polymerase chain reaction, and cloned into T7 expression vector pETMCSI[34] between the *Nde*I and *Eco*RI sites using RQ-SLIC.[35] The plasmids were transformed into *Escherichia coli* BL21(DE3)pLysS cells for overexpression. The ERp29-C double mutants A163H/Q167H, A218H/A222H, A179H/Q183H, and G207H/D211H, herein termed Mu1−Mu4, respectively, were prepared

by point mutagenesis using two overlapping primers.[35] An initial construct of ERp29-C, which comprised residues M154−L260, produced only low protein yields. Therefore, the final constructs were prepared with the addition of an N-terminal His$_6$ tag followed by a TEV cleavage site, which added a glycine residue to the N-terminal side 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 an OD_{600} of 0.8−1.0, the cells were centrifuged and the pellet was resuspended in M9 medium (6 g/L Na_2HPO_4, 3 g/L KH_2PO_4, and 0.5 g/L NaCl) supplied with 1 g/L $^{15}NH_4Cl$ and 3 g/L [^{13}C]glucose for uniform ^{15}N and ^{13}C 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 5000g for 15 min at 4 °C, and the cells were lysed in lysis buffer [10% glycerol, 1 mM phenylmethanesulfonyl fluoride, 2 mM 2-mercaptoethanol, 150 mM NaCl, and 50 mM sodium phosphate (pH 6.8)] using sonication. The supernatant was applied to a 5 mL Ni-NTA column (GE Healthcare), and bound protein was eluted with a gradient from 10 to 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 dialyzed overnight at 4 °C against 50 mM Tris-HCl, 300 mM NaCl, and 1 mM 2-mercaptoethanol. The cleavage yield was approximately two-thirds as estimated by sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE). The mixture was loaded onto a 5 mL Ni-NTA column to separate the final product from the His$_7$-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 with NMR buffer [50 mM NaCl and 20 mM MES (pH 6.5)] using an Amicon Ultra-15 centrifugal filter unit with an Ultracel-3 membrane (Merck Millipore). D_2O was added to a final 10% D_2O/90% H_2O ratio, and tris(2-carboxyethyl)phosphine (TCEP) was added to a final concentration of 1 mM to prevent oxidation of C157. Protein yields were ∼12 mg of purified protein per liter 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. $^{15}N−^{1}H$ HSQC spectra were recorded using a t_{1max} of 40 ms, a t_{2max} of 106 ms, and total recording times of ∼40 min. BEST-TROSY versions of HNCO, HNCACO, HNCA, HNCACB, and HNCOCACB spectra were recorded for wild-type ERp29-C and all four mutants without a metal ion using the IBS pulse sequence tools,[37] and the backbone resonances were assigned using CcpNmr.[38] The measurements were performed using 0.14−0.94 mM solutions of uniformly ^{15}N- and ^{13}C-labeled proteins.

PCS Measurements. PCSs were measured as the ^{1}H chemical shift observed in $^{15}N−^{1}H$ HSQC spectra with $CoCl_2$ minus the chemical shift measured for a sample containing $ZnCl_2$. Co^{2+} salts are known to change the ^{1}H chemical shift of the water resonance.[39] Therefore, the calibration of the ^{1}H carrier frequency (water resonance) was corrected by 0.0078 ppm per 1 mM Co^{2+} ion concentration.
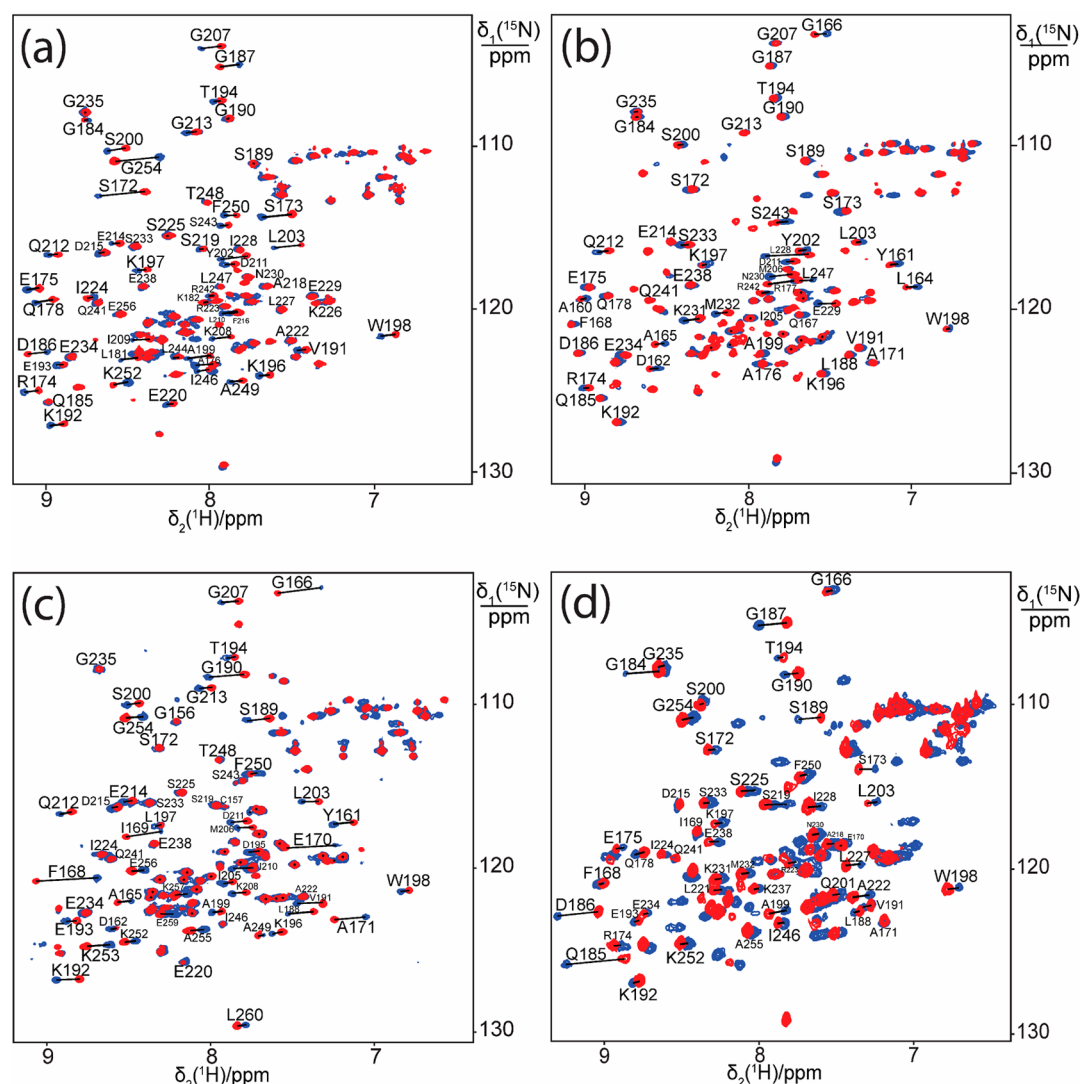
Figure 1. PCSs induced by a dHis-Co²⁺ motif at four different sites of ERp29-C. (a) Superimpositions of ^{15}N−^{1}H HSQC spectra of a 260 μM solution of uniformly ^{15}N-labeled ERp29-C A163H/Q167H (Mu1) in NMR buffer [20 mm MES and 50 mM NaCl (pH 6.5)] with either ZnCl$_2$ (diamagnetic reference, red cross-peaks) or CoCl$_2$ (paramagnetic spectrum, blue cross-peaks) added in an equimolar ratio. The spectra were recorded at 25 °C. Assignments are shown for cross-peaks of backbone amides with significant PCSs. (b) Same as panel a, but for a 140 μM solution of ERp29-C A218H/A222H (Mu2). (c) Same as panel a, but for a 150 μM solution of ERp29-C A179H/Q183H (Mu3). (d) Same as panel a, but for a 320 μM solution of ERp29-C G207H/D211H (Mu4).

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 (^{H}N, N, C$^\alpha$, C$^\beta$, and C′) for the 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$^\alpha$ atom of residue i and the C$^\beta$ 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 calculated final models 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 a His$_6$ tag were insufficient for facile purification on a Ni-NTA column. With an additional His$_6$ tag, the ERp29-C samples with the dHis motif bound very tightly, requiring 500 mM imidazole to elute the protein from the column. The presence of the His$_6$ 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. Ten mutants of ERp29-C were designed with the 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 mutants Mu1−Mu4.

PCS Measurements. Only mutants Mu1−Mu4 displayed
PCSs following titration with $CoCl_2$, suggesting that the dHis-
Co^{2+} motif did not always form. When samples were titrated
with $ZnCl_2$ instead of $CoCl_2$ to obtain the chemical shifts of
the corresponding diamagnetic references, some pronounced
precipitation occurred over a period of hours, which was not
observed with $CoCl_2$. The intrinsic widths of the NMR signals
of nuclear spins located far from the metal center did not
change during titration with $CoCl_2$ or $ZnCl_2$, 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 the dHis motif were titrated
with metal ions in an equimolar ratio except for Mu4, where a
5-fold excess of $CoCl_2$ was needed to observe PCSs. There was
no evidence of alternative binding sites for Co^{2+} ions in ERp29,
as PREs were evident only in the vicinity of the dHis motif.
Although the PCSs observed were relatively small, PCSs
nonetheless could be observed for most of the protein (Figure
1).

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 [Protein Data Bank
(PDB) entry 2M66[15]] 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 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 >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.

Structure Computation. The PCSs were used as input
into the program GPS-Rosetta to calculate the 3D structure of
the protein. Twenty thousand all-atom models were calculated
using a nonhomologous 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 target structure 2M66 (Figure
3a). As the metal position in the dHis-Co^{2+} 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-Co^{2+} motif (Figure 3b). Selecting the
models, for which the sum of the squared deviation (SD) of
experimental and back-calculated PCSs was <0.244 ppm² (the
value obtained by fitting $\Delta\chi$ tensors to the first conformer of
structure 2M66), we obtained ~1000 models, which were
consistently similar to the target structure (red and blue points
in Figure 3a and hereafter termed Top1k). Finally, we applied
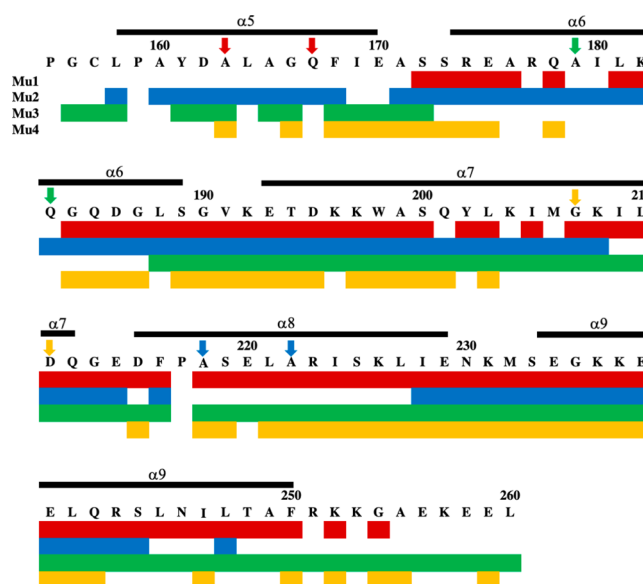

Figure 2. Summary of the experimental amide proton PCSs. The
names of the different mutants are indicated on the left. Arrows
identify the residues mutated to histidine in the different mutants.
Red, blue, green, and yellow bars identify the residues for which the
dHis-Co^{2+} motif delivered significant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

the lowest combined PCS + Rosetta score as an additional
selection criterion to identify the 10 best models, hereafter
termed 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 PyMOL[42] for optimal structure alignment, conformer
top1 delivered a C^{α} RMSD of ~1.4 Å to the first conformer of
2M66 for 96 aligned residues. All 10 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). For
comparison, the Top10 conformers display an average pairwise
C^{α} RMSD of 1.8 Å for residues 156−255.

To assess the contribution of the PCS data to the structure
calculation, the performance of each PCS data set was analyzed
individually. All four PCS data sets contributed to the
convergence of the GPS-Rosetta modeling (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 5−7 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
2M66 but has a kink in the six remaining conformers.

Figure 4 illustrates the structural similarity between the first
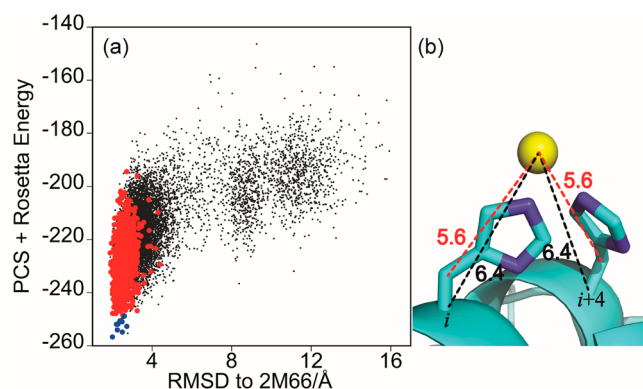conformer of the 2M66 structure and the model with the

Figure 3. PCS and Rosetta energy scores of 20000 GPS-Rosetta structure calculations of ERp29-C using PCSs from dHis-Co^{2+} motifs and the distance restraints used. (a) Combined PCS + Rosetta energy scores of the structural models calculated in this work plotted vs the C$^\alpha$ RMSD to the first conformer of the structure determined previously using PCSs from lanthanide tags (PDB entry 2M66). The alignment excluded disordered C-terminal residues 256−260. Black points mark the results for all 20000 structures obtained from GPS-Rosetta. Blue and red points mark the approximately 1000 structures fulfilling the PCS restraints with an SD value of <0.244 ppm^2, when the metal positions were defined by the distance restraints specified in panel b (structures termed 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$^\alpha$−Co^{2+} and C$^\beta$−Co^{2+} distance restraints used to determine a metal position for each model, which is in agreement with the location expected for the α-helical dHis-Co^{2+} motif, where the histidine residues are located at 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]
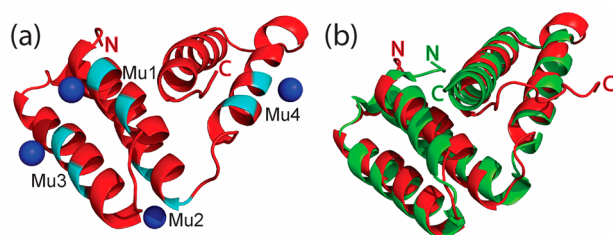

Figure 4. Locations of Co^{2+} ions relative to the 3D structure of ERp29-C with the lowest combined PCS + Rosetta energy (top1) and comparison of top1 with the first conformer of structure 2M66[15] previously determined using PCSs generated by lanthanide tags. (a) Ribbon representation of the structural model of 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 of top1 with the structure 2M66. The SD value of the PCSs of top1 is 0.18 ppm^2, and the C$^\alpha$ RMSD of top1 to 2M66 is 1.4 Å for 95 aligned atoms.

lowest PCS + Rosetta energy (hereafter termed top1) determined in this work. The PCS SD value for this structure is only 0.184 ppm^2 (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 fits of the $\Delta\chi$ tensor 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 mutants Mu1 and Mu3 compared to that of the 2M66 structure is also apparent in the correlation plot of back-calculated versus experimental PCSs (Figure 5b).

Conversely, the PCSs determined with Tm^{3+} and Tb^{3+} tags in our previous work[15] fit somewhat better to structure 2M66 than to the top1 conformer, highlighting the influence of PCSs on the final structure. Nonetheless, the quality of the Q factors obtained from the fits to top1 was comparable to the quality of those obtained with the original structure 2M66 (Table 1).

■ 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-Co^{2+} motif in generating PCSs, as the dHis-Co^{2+} motif is more readily installed in α-helices than in other secondary structure elements.[28] In the work presented here, 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 this work.

Like any paramagnetic metal ion, Co^{2+} ions generate PREs that broaden the NMR signals of nearby nuclear spins beyond detection. In this study, no ^{15}N−^{1}H HSQC cross-peaks could be observed for residues three positions prior to or following the histidyl residues of the dHis-Co^{2+} motif (Table S2). To measure PCSs for the entire length of the amino acid sequence, it is thus important to position dHis motifs at two different sites at least. 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-Co^{2+} 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 (Tm^{3+} and Tb^{3+} tags) at each site. This 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.

The 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-Co^{2+} motif, the resonance assignment of the paramagnetic state is less straightforward for cross-peaks with large PCSs. In the study presented here, we used fits of the $\Delta\chi$ tensor 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 ^{15}N−^{1}H HSQC spectrum. The preparation of multiple selectively isotope-labeled samples with dHis-Co^{2+} motifs is relatively economical,
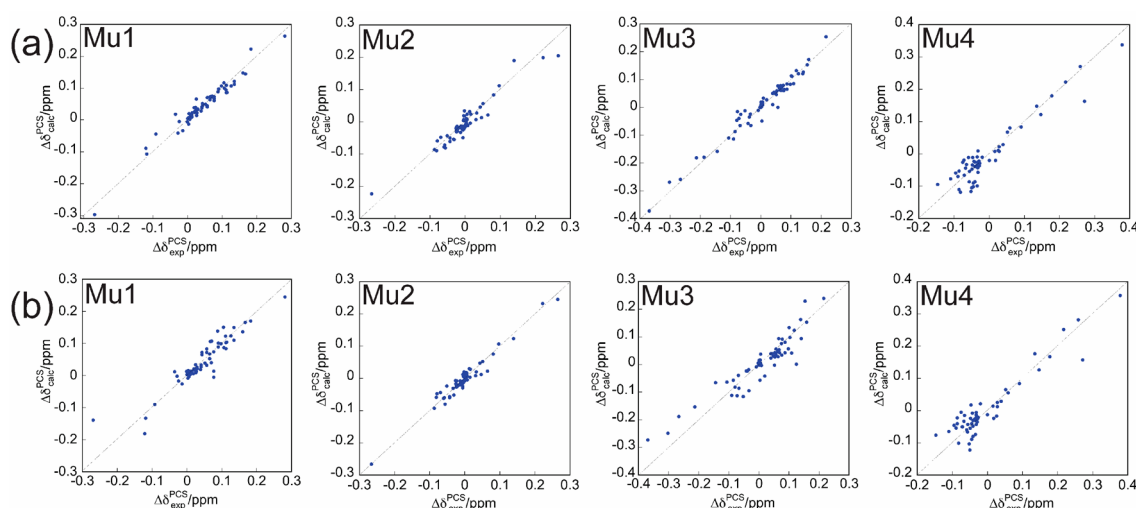
Figure 5. Correlations between back-calculated and experimental PCSs generated by dHis-Co^{2+} motifs in rat ERp29-C. The $\Delta\chi$ tensors were fitted with the metal position fixed to the locations reported in Table S3. (a) Correlation plots following fits of the $\Delta\chi$ tensor to top1. (b) Correlation plots following fits of the $\Delta\chi$ tensor to the first conformer of 2M66.

Table 1. Quality Factors Obtained from Fitting $\Delta\chi$ Tensors to the Structure of 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]

	top1		2M66[c]	
ERp29-C mutant[b]	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 Numbat[41] 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.

because no chemical reactions with expensive tags are required. Unfortunately, I$_z$S$_z$-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 toward 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 for guiding Rosetta calculations. This work shows that the relatively small PCSs generated by the dHis-Co^{2+} 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 the lanthanide ion and protein,[46] dHis motifs are capable of holding the Co^{2+} ion in a uniquely well-defined position relative to the protein backbone.[28] As a result, one might expect quality factors for the $\Delta\chi$ tensor fits lower than those observed in the work presented here (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-Co^{2+} 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-Co^{2+} motifs at multiple sites open a new and practical approach for harnessing the power of PCS restraints for protein fold determination. Unexpectedly, it proved to be 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 Co^{2+} ion than from Tm^{3+} or Tb^{3+} 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-Co^{2+} 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.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.biochem.9b00404.

Δχ tensor parameters of ERp29-C fitted to 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), Δχ tensor parameters and cobalt ion coordinates relative to the first conformer of 2M66 and the individual conformers of the Top10 ensemble (Table S3), side chain conformations determined by the GPS-Rosetta calculations (Figure S2), comparison of the Top10 conformers to the first conformer of structure 2M66 and crystal structure 2QC7 (Figure S3), combined PCS + Rosetta energy scores of the GPS-Rosetta structural models plotted versus the C^α RMSD to the crystal structure of human ERp29 (Figure S4), performance of the PCS data set from each dHis-Co²⁺ motif in the GPS-Rosetta calculations (Figure S5), and comparison of the individual α-helices of the Top10 conformers with the first conformer of 2M66 (Figure S6) (PDF)

Accession Codes

The coordinates and structure factors of the Top10 models have been deposited in the Protein Data Bank as entry 6O6I. The chemical shifts of wild-type ERp29-C were deposited in the BioMagResBank[47] as entry 30583.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: t.huber@anu.edu.au.
*E-mail: gottfried.otting@anu.edu.au. Phone: +61 2 61256507.

ORCID Ⓘ

Gottfried Otting: 0000-0002-0563-0146

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS

ERp29-C, C-terminal domain of 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., and Berliner, L. J., Eds.) pp 195−257, Springer, Boston.

(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. Nucl. Magn. Reson. 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., Lemak, A., et al. (2008) Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl. Acad. Sci. U. S. A. 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: Struct., Funct., Genet. 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. Nucl. Magn. Reson. 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. Nucl. Magn. Reson. Spectrosc. 58*, 88−96.

(23) Nitsche, C., and Otting, G. (2017) Pseudocontact shifts in biomolecular NMR using paramagnetic metal tags. *Prog. Nucl. Magn. Reson. 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 g-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. *Antioxid. Redox Signaling* 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, No. e0211065.

(36) Blommel, P. G., and Fox, B. G. (2007) A combined approach to improving large-scale production of tobacco etch virus protease. *Protein Expression 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: Struct., Funct., Genet.* 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) *PyMOL Molecular Graphics system*, version 2.0, Schrödinger, LLC (2007).

(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: Struct., Funct., Genet.* 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) Bio Mag Res Bank. *Nucleic Acids Res.* 36, D402−D408.