Structure
Technical Advance

Three-Dimensional Protein Fold Determination from Backbone Amide Pseudocontact Shifts Generated by Lanthanide Tags at Multiple Sites

Hiromasa Yagi,[1,3] Kala Bharath Pilla,[1,3] Ansis Maleckis,[1] Bim Graham,[2] Thomas Huber,[1,*] and Gottfried Otting[1,*]
[1]Research School of Chemistry, Australian National University, Canberra, ACT 0200, Australia
[2]Medicinal Chemistry and Drug Action, Monash Institute of Pharmaceutical Sciences, Parkville VIC 3052, Australia
[3]These authors contributed equally to this work
*Correspondence: t.huber@anu.edu.au (T.H.), gottfried.otting@anu.edu.au (G.O.)
http://dx.doi.org/10.1016/j.str.2013.04.001

SUMMARY

Site-specific attachment of paramagnetic lanthanide ions to a protein generates pseudocontact shifts (PCS) in the nuclear magnetic resonance (NMR) spectra of the protein that are easily measured as changes in chemical shifts. By labeling the protein with lanthanide tags at four different sites, PCSs are observed for most amide protons and accurate information is obtained about their coordinates in three-dimensional space. The approach is demonstrated with the chaperone ERp29, for which large differences have been reported between X-ray and NMR structures of the C-terminal domain, ERp29-C. The results unambiguously show that the structure of rat ERp29-C in solution is similar to the crystal structure of human ERp29-C. PCSs of backbone amides were the only structural restraints required. Because these can be measured for more dilute protein solutions than other NMR restraints, the approach greatly widens the range of proteins amenable to structural studies in solution.

INTRODUCTION

Conventional three-dimensional (3D) structure determinations of proteins with nuclear magnetic resonance (NMR) spectroscopy rely on a very large number of nuclear Overhauser effects (NOE) to provide a dense network of short-range distance restraints (Wüthrich, 1986). Additional NMR parameters, such as scalar and residual dipolar couplings (RDC), cross-correlation effects, and paramagnetic relaxation enhancements (PRE), provide powerful restraints for further structure refinements but none of these parameters is sufficient on its own to determine 3D structures of proteins, except for RDCs, where this has been achieved for small proteins with a large number of experimental RDCs (Hus et al., 2001; Tian et al., 2001; Wang et al., 2007; Rasia et al., 2011; Kontaxis, 2012). The measurement of small heteronuclear RDCs, however, becomes increasingly difficult for proteins of limited solubility or proteins yielding broad NMR line widths due to mobility effects or increased molecular weight. Here, we present an alternative approach to 3D structure determinations of proteins that relies on pseudocontact shifts of backbone amide protons induced by paramagnetic lanthanide ions attached to multiple different sites of the target protein. PCS data can be measured with much greater sensitivity than NOEs or RDCs, opening the door to detailed structural studies of a much greater range of proteins in solution. Furthermore, we show that PCSs of backbone amides are sufficient for 3D fold determination. Combined with the remarkable capability of the program Rosetta to predict correct amino acid side chain conformations even if experimental data are only available for backbone atoms (Raman et al., 2010), complete protein structures can be obtained also without cumbersome analysis of poorly resolved NMR resonances of amino acid side chains.

PCSs stand out for their long-range effects (up to 40 Å for lanthanides; Allegrozzi et al., 2000). PCSs are generated by many paramagnetic metal ions (Bertini and Luchinat, 1986) and have long been used successfully for structure refinement of metalloproteins (Bertini et al., 2001, 2008; Arnesano et al., 2005; Otting, 2010) and interaction studies (Pintacuda et al., 2007; Keizers and Ubbink, 2011). Arguably, lanthanide ions are most suited as paramagnetic centers for generating PCSs because many of them are highly paramagnetic and once a lanthanide binding site has been established, it will bind any lanthanide similarly well, with yttrium as a suitable diamagnetic reference (Otting, 2008). With the advent of efficient protocols for site-specific attachment of lanthanide ions (for reviews, see Rodriguez-Castañeda et al., 2006; Su and Otting, 2010, 2011; Koehler and Meiler, 2011), including methods that do not depend on single cysteine residues (Wöhnert et al., 2003; Su et al., 2009; Barthelmes et al., 2011; Loh et al., 2013), large PCSs can be routinely generated in proteins amenable to genetic engineering.

The PCS $\Delta\delta^{PCS}$ of a nuclear spin is measured in ppm as the difference in chemical shift observed between a sample with a paramagnetic lanthanide and a reference sample with a chemically similar diamagnetic metal (e.g., Y^{3+}). It depends on the polar coordinates r, θ and ϕ of the nuclear spin with respect to the principal axes of the $\Delta\chi$ tensor that describes the pertinent paramagnetic properties of the lanthanide ion:

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

where $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ denote, respectively, the axial and rhombic components of the magnetic susceptibility tensor χ (Bertini et al.,

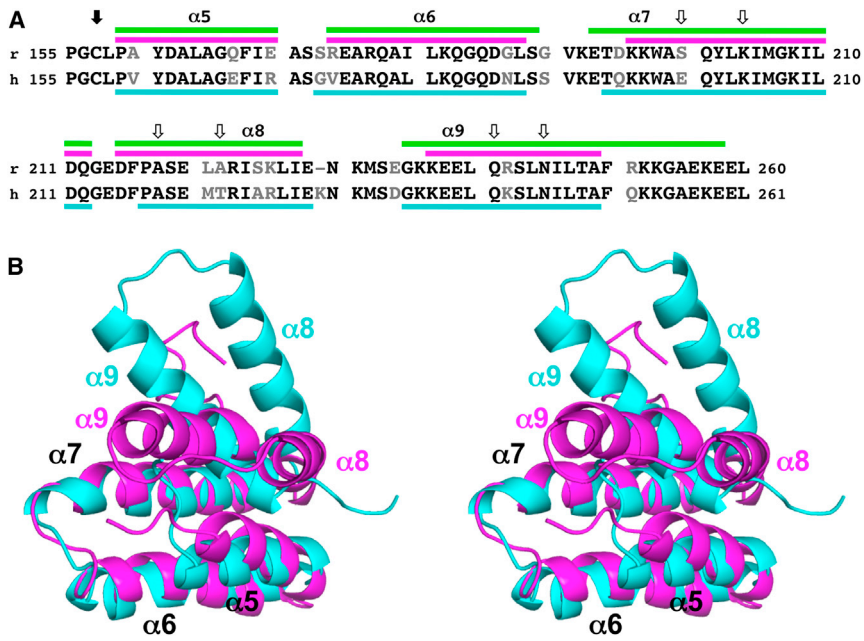

Figure 1. Comparison between the NOE Structure of Rat ERp29-C and the Crystal Structure of Human ERp29-C

(A) Amino acid sequence alignment of rat (r) and human (h) ERp29-C. Sequence numbers are shown at the start and end of the sequences. Identical residues are highlighted in black. The α-helical segments in the NOE structure of rat ERp29-C (PDB ID 1G7D; Liepinsh et al., 2001) and in the crystal structure of human ERp29-C (PDB ID 2QC7; Barak et al., 2009) are indicated by pink and blue lines above and below the respective sequences, and the green line at the top delineates the helical segments found in the GPS-Rosetta structure determined in the present work. Open arrows identify the sites of the Cys/Asp double mutations for the IDA-SH tag and the filled arrow identifies the tagging site for the C1 tag.

(B) Stereo view of a superimposition of the NOE structure of the rat protein (pink) and the crystal structure of the human protein (blue). Helices 5–7 were superimposed for minimal rmsd.

2002a) and where the $\Delta\chi$ tensor is defined as the χ tensor minus its isotropic component. Therefore, if the $\Delta\chi$ tensor and its location and orientation with respect to the protein are known, the PCS of a nuclear spin provides useful information about its location in the coordinate frame of the $\Delta\chi$ tensor.

It is instructive to visualize the PCSs described by Equation 1 as isosurfaces of constant PCS, where each isosurface depicts all coordinates at which nuclear spins would display a certain PCS value. While the PCS of a nuclear spin ties its location to a certain isosurface, a second PCS measured for a sample with a lanthanide attached at a different site confines the nuclear spin further to lie on the line defined by the intersection of the respective PCS isosurfaces. A third PCS value from a sample with a lanthanide attached at yet another site would pin down the position of the nuclear spin even more.

Using PCSs in this manner for 3D structure determination poses the dilemma that the $\Delta\chi$ tensor parameters of a metal ion are usually not known a priori. Without 3D coordinates of the protein, however, it is not possible to fit a $\Delta\chi$ tensor that reproduces the experimentally observed PCSs. To circumvent this conundrum, we used the Rosetta algorithm (Qian et al., 2007) to model the protein structure to fit initial $\Delta\chi$ tensors. For PCSs from a single site, this strategy has been implemented at the fragment assembly step in the program PCS-Rosetta, which discards ill-fitting structures before any further time-consuming refinement to improve the final yields of correct structures (Schmitz et al., 2012).

To avoid the ambiguities associated with PCSs from a paramagnetic tag at a single site and to achieve complete coverage of the polypeptide chain with PCSs, we expanded the concept of PCS-Rosetta to multiple tagging sites. The results show that PCSs generated by lanthanide tags attached at several different sites offer a robust route to 3D fold determinations of proteins that can be more accurate than structure determinations by NOEs, even if only PCSs of backbone amide protons are considered.

The approach was applied to the C-terminal domain of the rat protein ERp29, ERp29-C (106 residues, 12 kDa). ERp29 is a chaperone expressed in the endoplasmic reticulum, where it facilitates the processing and transport of proteins. Its chaperone activity underpins numerous biological functions, including roles in virus infections, cystic fibrosis, and tumor suppression and treatment (Rainey-Barger et al., 2009; Das et al., 2009; Farmaki et al., 2011; Suaud et al., 2011; Zhang and Richardson, 2011; Bambang et al., 2013; Wu et al., 2012). The 3D structure of the C-terminal domain of rat ERp29 was first determined by NMR spectroscopy using the conventional NOE approach. This structure is henceforth referred to as the NOE structure of ERp29-C (Liepinsh et al., 2001). Subsequently, the 3D structure of the full-length human protein was determined by X-ray crystallography (Barak et al., 2009), revealing a significantly different structure of ERp29-C, with a C^{α} root mean square deviation (rmsd) of 4.5 Å to the NOE structure (residues 158–244). Specifically, helices 8 and 9 of the crystal structure form fewer contacts with the rest of the protein, giving it a less globular appearance (Figure 1). The present work thus not only establishes a method for 3D structure determinations of proteins in solution, but also reassesses the structure of the rat ERp29-C domain.

RESULTS

Sample Preparation

To generate PCSs from multiple metal positions and to obtain PCSs for most of the backbone amides, ERp29-C was tagged at four different sites with lanthanide ions. Wild-type rat ERp29-C contains a single cysteine residue, Cys157, which was ligated with the C1 tag (Graham et al., 2011) to generate PCSs. To create additional lanthanide binding sites, Cys157 was mutated to serine and the three additional double mutations S200C/K204D, A218C/A222D, and Q241C/N245D

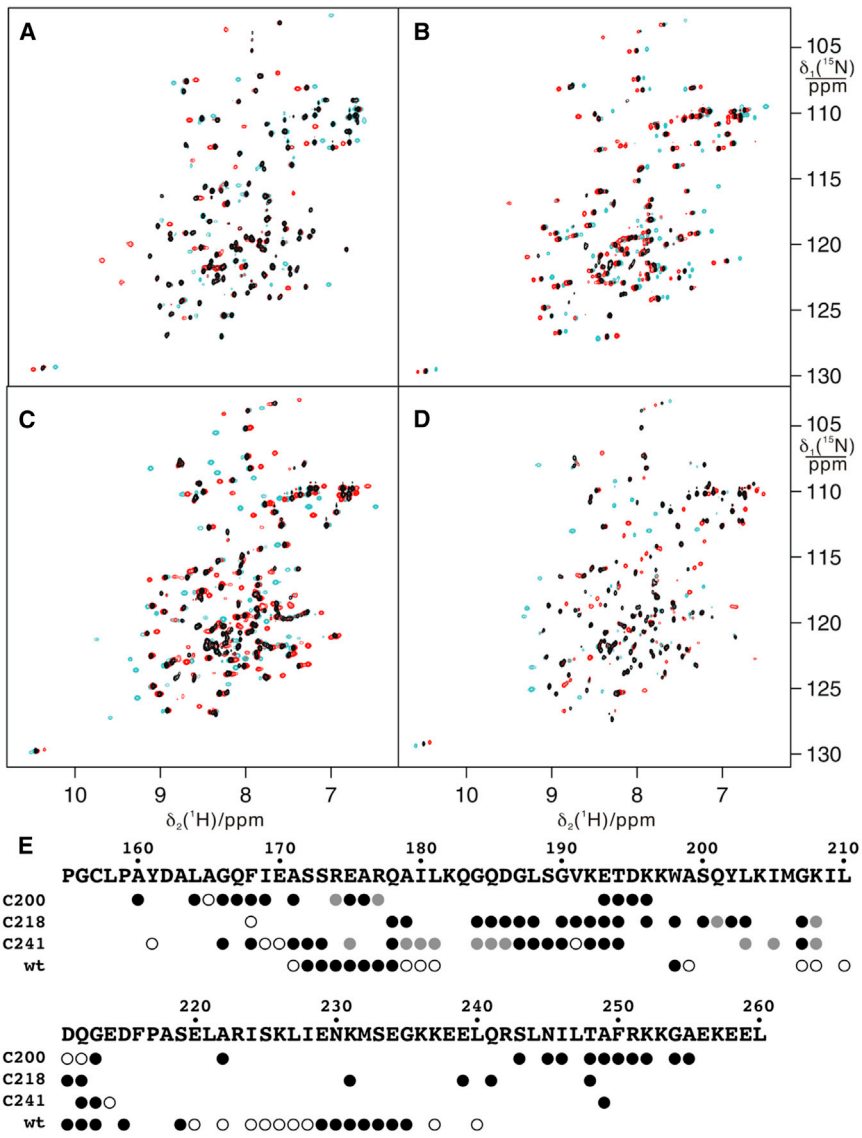

Figure 2. Pseudocontact Shifts of Amide Protons in ERp29-C

(A–D) Superimposition of ^{15}N-HSQC spectra of uniformly ^{15}N-labeled rat ERp29-C, labeled with an IDA-SH or C1 tag loaded with Y^{3+} (black), Tm^{3+} (red), or Tb^{3+} (cyan). All spectra were recorded at 31°C at a ^{1}H NMR frequency of 800 MHz. C200 mutant with IDA-SH tag (A), C218 mutant with IDA-SH tag (B), C241 mutant with IDA-SH tag (C), and wild-type protein with C1 tag (D). Each of the mutants with an IDA-SH tag also contained the C157S mutation and a mutation of the residue in position *i*+4 to aspartate.

(E) Summary of the experimental amide proton PCSs. The names of the different mutants are indicated on the left. The IDA-SH tag was used for the mutants and the C1 tag was used for the wild-type protein. Open, gray, and black circles identify the residues for which PCSs were observed for backbone amide protons with Tm^{3+}, Tb^{3+}, or both Tm^{3+} and Tb^{3+}, respectively. The residue numbering is shown at the top of the amino acid sequence. The PCS data are listed in Table S1.

were introduced for attachment of the IDA-SH tag via a disulfide bond. (For simplicity, these triple mutants are referred to in the following as C200, C218, and C241.) All double mutation sites were in α helices (Figure 1A). The aspartate residue in position *i*+4 served to create a specific lanthanide binding site in which the metal is immobilized by coordination to the IDA-SH tag and by additional coordination to the aspartate residue (Swarbrick et al., 2011; Yagi et al., 2013). Samples of the mutants with the IDA-SH tag loaded with Y^{3+}, Tb^{3+}, or Tm^{3+} were prepared by titration with the respective lanthanide chlorides. Corresponding samples of the wild-type protein with a C1 tag were prepared by ligation with C1 tags preloaded with the respective metal ions because high temperatures are required to install lanthanides in the C1 tag.

NMR Measurements

All four constructs displayed significant PCSs in ^{15}N-HSQC spectra (Figures 2A–2D). The cross-peaks in the paramagnetic states were assigned using the fact that the PCSs of the ^{1}H and ^{15}N spins of any single amide group are very similar because their coordinates are close compared with their distance from the paramagnetic center (cross-peaks from amides close to the metal are broadened beyond detection due to PREs; Allegrozzi et al., 2000; John and Otting, 2007). 3D HNCO spectra recorded of ^{13}C/^{15}N-labeled samples of the wild-type protein and the C200 mutant, both labeled with paramagnetic lanthanides, confirmed the reliability of the resonance assignments obtained from the two-dimensional (2D) ^{15}N-HSQC spectra.

A total of 212 PCSs were measured for the backbone amide protons from the 2D ^{15}N-HSQC spectra, using Tm^{3+} and Tb^{3+} (Table S1 available online). Figure 2E shows that at least one PCS was obtained for almost 90% of residues 161–256, i.e. excluding the flexible N- and C-terminal segments (Liepinsh et al., 2001), and two or more PCSs were observed for over half of the protein.

GPS-Rosetta Calculations

The PCSs served as input for 3D structure calculations using a new version of the PCS-Rosetta algorithm (Schmitz et al., 2012). Like CS-Rosetta (Shen et al., 2008), PCS-Rosetta uses the backbone chemical shifts of the protein to assist the initial fragment selection from the Rosetta library of proteins. The PCSs are used only in the subsequent fragment assembly step to fit $\Delta\chi$ tensors and thus probe the agreement of the emerging model with the PCS data. The long-range nature of the PCSs effectively discriminates between successful and unsuccessful models in this early stage of calculations. Our version of the PCS-Rosetta program, called GPS-Rosetta, uses the same

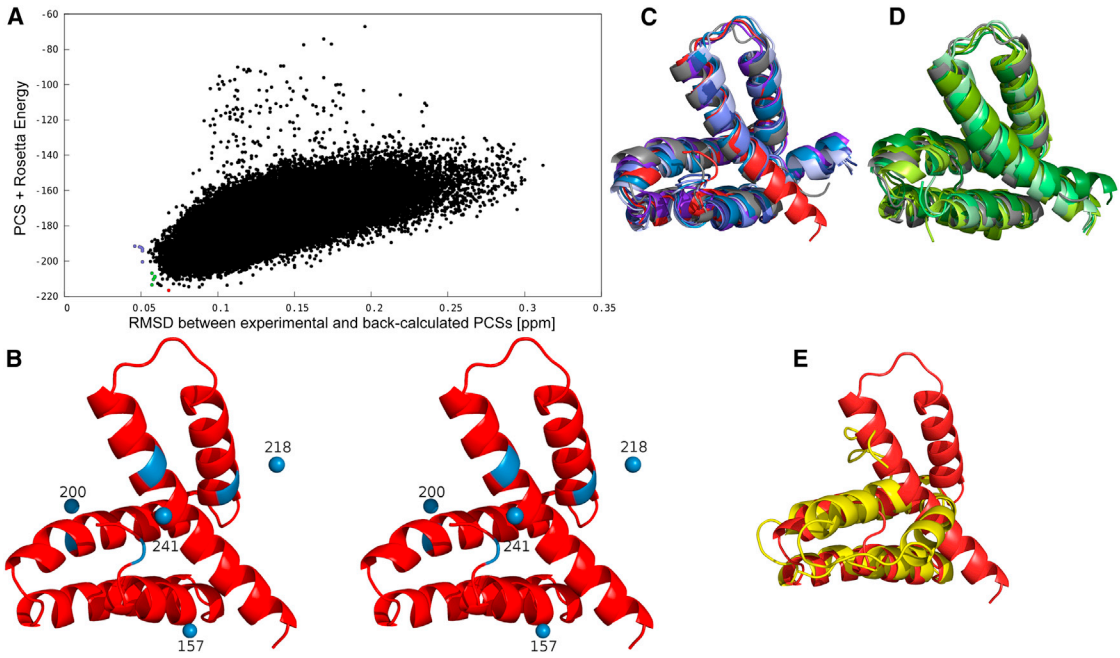

Figure 3. The Solution Structure of Rat ERp29-C Is More Similar to the Crystal Structure of Human ERp29-C than to the NOE Structure of Rat ERp29-C

(A) Combined score of weighted PCS + Rosetta energy versus the rmsd between experimental and back-calculated PCSs. The final selected structure (red point) has the lowest combined score and is referred to as "the GPS-Rosetta structure". Blue points represent the structures with the lowest PCS rmsd values and green points represent structures combining a low combined PCS + Rosetta energy with a low PCS rmsd value. A comparison with the performance of CS-Rosetta is given in the Supplemental Experimental Procedures and Figure S1.

(B–E) Comparisons between GPS-Rosetta structures, the crystal structure, and the NOE structure. Superimpositions of ribbon representations of the top structures calculated with the GPS-Rosetta protocol. The colors of the structures correspond to those used in (A). The crystal structure (PDB ID 2QC7) is shown in gray and the GPS-Rosetta structure with the lowest combined PCS + Rosetta energy is shown in red. The stereo view of the GPS-Rosetta structure of rat ERp29-C, shown in (B), also marks the four positions of the lanthanide ions by blue spheres labeled with the residue number of the cysteine they are attached to and highlights the sites of the respective cysteine residues on the ribbon in blue. The C^{α} rmsd between the GPS-Rosetta structure and the crystal structure of human ERp29-C is 2.4 Å [where the rmsd calculation included residues 158–244, while omitting Lys229, which is inserted in the human protein (Figure 1A), and excluding the C-terminal segment that forms an extended helix in the GPS-Rosetta structure although residues 254–260 are known to be highly flexible; Liepinsh et al., 2001]. The superimposition in (C) also shows the five structures with the lowest PCS rmsd values, represented in shades of blue. Their C^{α} rmsd to the crystal structure (residues 158–228 and 230–244) is 2.0–2.9 Å. A comparison of the side chain conformations of conserved amino acids between the GPS-Rosetta and crystal structures is given in Figure S2. Structures combining a low combined PCS + Rosetta energy with low PCS rmsd values are represented in shades of green (D). They have a C^{α} rmsd range of 2.2–2.6 Å to the crystal structure. The NOE structure shown in yellow (PDB ID 1G7D) has a C^{α} rmsd of 6 Å (residues 158–244) to the GPS-Rosetta structure shown in red (E).

approach but is no longer limited to a single paramagnetic center, accepting PCSs generated by multiple lanthanides at different sites and optimizing their respective $\Delta\chi$ tensors during the structure calculation.

About 100,000 full-atom models were generated with GPS-Rosetta. The structure identified by the red point in Figure 3A was selected for its low Rosetta energy and high quality of back-calculated versus experimental PCSs. This structure is shown in Figure 3B and in the following referred to as the GPS-Rosetta structure. Most notably, in this structure, as in all structures with good PCS fits, the orientation of helices 8 and 9 with respect to the rest of the structure is similar as in the crystal structure of human ERp29-C. Figure 3C illustrates the similarity between the GPS-Rosetta structure, the crystal structure, and the structures with the lowest PCS rmsd values (blue points in Figure 3A and blue structures in Figure 3C). The structures with low combined PCS + Rosetta energy and small PCS rmsd values (green points in Figure 3A) are also similar to the crystal structure (Figure 3D). In contrast, the GPS-Rosetta structure is markedly different from the original NOE structure (Figure 3E). The disagreement with the NOE structure is highlighted by the correlation between the back-calculated and experimental PCSs, which is much better for the GPS-Rosetta structure than for the NOE structure (Figure 4). Cross-validation with PCSs not used in the structure calculation (measured with C1-Yb and C1-Dy tags attached to C157) confirm this conclusion (Figure S3 and Tables S2–S4). The results thus clearly show that the fold of rat ERp29-C in solution is similar to the crystal structure of the human protein, superseding the original NOE structure.

To check whether the Rosetta program would have found the same structure without the help of PCSs, we repeated the same calculations with CS-Rosetta (Shen et al., 2008; see Supplemental Experimental Procedures). Much fewer models were in agreement with the experimental PCSs (Figure S1) and the two models with the lowest Rosetta energy displayed very different folds, with C^{α} rmsd values to the GPS-Rosetta structure greater

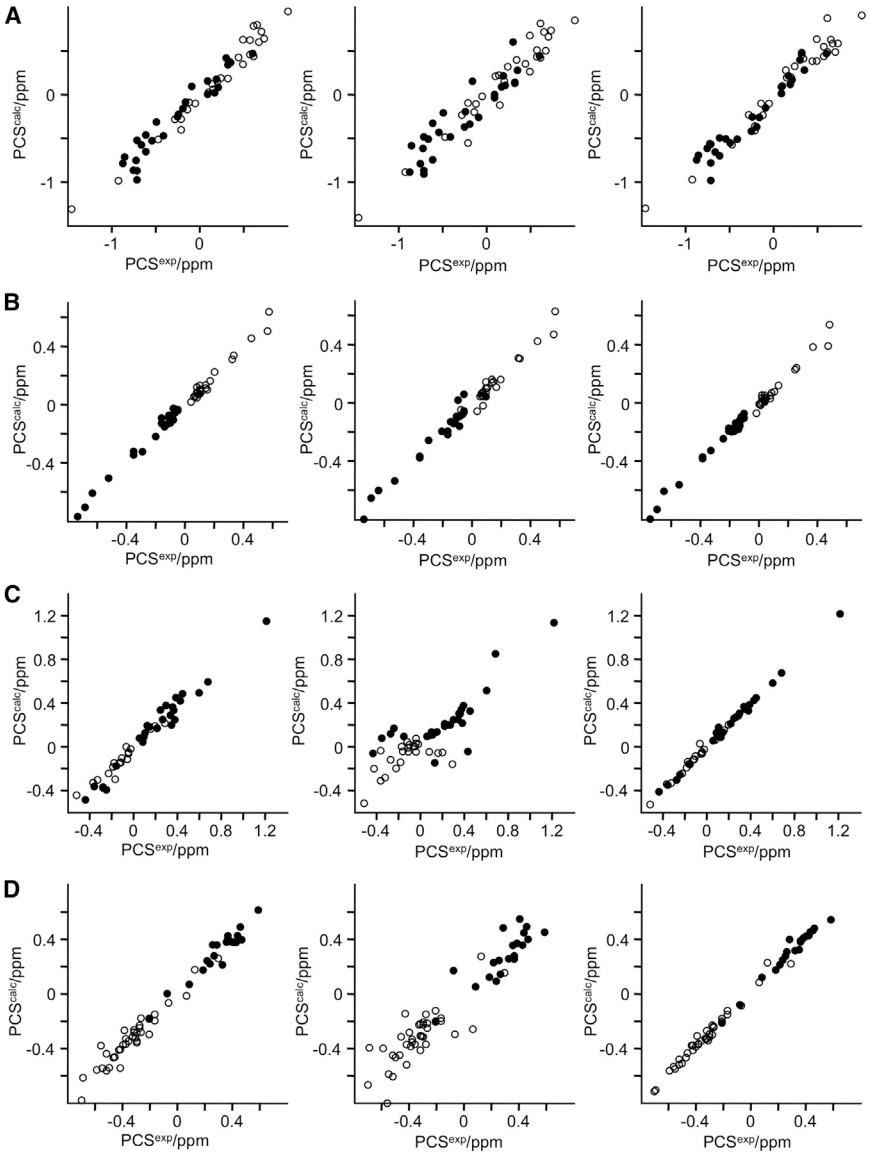

Figure 4. Correlations between Back-Calculated and Experimental PCSs of Rat ERp29-C Confirm Structural Similarity with Human ERp29-C

(A–D) The correlation plots were obtained by fitting the $\Delta\chi$ tensors to the crystal structure of human ERp29-C (left column), the NOE structure of rat ERp29-C (middle column), and the GPS-Rosetta structure of rat ERp29-C (right column). PCSs generated by Tm^{3+} and Tb^{3+} are shown as open and closed circles, respectively. Separate plots are shown for the correlations obtained for the mutants with IDA-SH tag and the wild-type protein with C1-tag. C200 mutant with IDA-SH (A), C218 mutant with IDA-SH (B), C241 mutant with IDA-SH (C), and wild-type protein with C1 tag (D). The same relative quality of the fits was observed by cross-validating with additional data obtained for wild-type ERp29-C with different lanthanides in the C1 tag.

See also Tables S2–S4 and Figure S3.

than 13 Å (Figure S2). The PCSs were thus critical for the identification of the correct fold.

DISCUSSION

The present study highlights the value of PCSs delivered by multiple site-specific lanthanide tags and the pitfalls that can be associated with the misinterpretation of NOEs, especially in the case of helical proteins where contacts between the helices are made by side chain rather than backbone atoms, requiring careful interpretation of crowded spectral regions. PCSs from multiple rather than single lanthanide sites offer particularly valuable structural restraints. First, paramagnetic relaxation enhancements (PRE) in the vicinity of a lanthanide broaden the NMR signals of many residues beyond detection. PCSs from different tagging sites thus significantly improve the coverage of the amino acid sequence with experimental PCSs (Figure 2E). Second, a single PCS value measured of a nuclear spin provides the information that the spin must be located on a PCS isosurface defined by the $\Delta\chi$ tensor. This presents a relatively weak structural restraint because the isosurface extends all around the metal ion. In contrast, if the nuclear spin must fulfill the PCSs generated by lanthanides at two different sites, it must be positioned on the line defined by the intersect of the isosurfaces of the corresponding two different $\Delta\chi$ tensors. PCSs from lanthanides at three different sites restrict the possible locations of the spin to the two points, where the line described by the intersect defined by the first two isosurfaces intersects the isosurface from the third $\Delta\chi$ tensor. The PCS from a fourth lanthanide site unambiguously locates the spin at a single point. In this way, PCSs from multiple tags allow positioning of a nuclear spin in 3D space in a manner analogous to the global positioning systems that calculate coordinates on earth by distance measurements from satellites. Similar to the situation with RDCs, which convey detailed structural information if measured with multiple alignment media (Meiler et al., 2001; Ruan et al., 2008), the availability of PCSs from four different sites presents a fundamental advance over data from a single paramagnetic center, which are compatible with multiple structure solutions as shown in previous attempts of using PCSs for protein structure determination (Bertini et al., 2002b). In the case of ERp29-C, one amide (Gln212) displayed PCSs from four sites, and PCSs from two or more sites were observed for about one-third of the amide resonances (Figure 2E).

The present study shows that, if PCS data are available only from a single lanthanide tagging site, large structural differences may not be picked up even if different lanthanides are used. This

is apparent in the correlation plots obtained for the C218 mutant, which were of similar quality for the crystal, NOE, and GPS-Rosetta structures (Figure 4B). In part, this can be attributed to the similarity in $\Delta\chi$ tensor orientations for different lanthanides, which is a consequence of the conserved ligand field presented by the compound that chelates the metal.

Nonetheless, using two different paramagnetic metal ions at each site is useful. In the present work, we tagged each of the four sites with either Tb^{3+}, Tm^{3+}, or Y^{3+} to generate two sets of PCSs from each site. Tb^{3+} and Tm^{3+} tend to shift the cross-peaks in opposite directions, supporting the resonance assignments of the paramagnetic states. In addition, Tb^{3+} is more strongly paramagnetic than Tm^{3+}, yielding PCSs at different distances from the metal ion.

The IDA-SH tag is particularly powerful for the present application because it defines the position of the lanthanide if it is located in an α-helix with an aspartate in position i+4 (Swarbrick et al., 2011; Yagi et al., 2013). Therefore, while a $\Delta\chi$ tensor determination generally requires the fit of eight parameters ($\Delta\chi_{ax}$, $\Delta\chi_{rh}$, the x, y, z coordinates of the metal ion, and three angles to relate the orientation of the tensor to the protein structure), prior knowledge of the metal coordinates reduces the number of fitting parameters to five. Notably, however, the GPS-Rosetta algorithm can also fit the metal coordinates, and this was used to locate the C1 tag for which no prior information about its location was available.

The GPS-Rosetta structure of rat ERp29-C supersedes the earlier NOE structure of the same protein. The magnitude of the structural differences suggests wrong NOE assignments. The NOE structure relied on NOE assignments made in 2D NOESY and 3D NOESY-^{15}N-HSQC spectra recorded on a 600 MHz NMR spectrometer (Liepinsh et al., 2001). In particular, the 2D spectrum allowed the observation of over a dozen NOEs with the side chain hydroxyl resonance of a buried tyrosine residue. The current GPS-Rosetta structure indicates that this hydroxyl resonance, which could only be assigned by NOEs, should have been assigned to Tyr 161 but was in fact misassigned to Tyr 202. Structure calculations using this misassignment would necessarily have resulted in further misassignments of NOEs which, in the case of poorly resolved resonances of aromatic and methyl side chains, could easily have gone undetected due to the absence of a 3D NOESY-^{13}C-HSQC spectrum of a ^{13}C-labeled sample.

The results obtained in the present work carry great promise for 3D structure determinations of perdeuterated proteins for which no side chain resonances can be observed. Selectively ^{15}N-labeled samples can be used to achieve the required spectral resolution in ^{15}N-^{1}H correlation spectra for PCS measurements of backbone amides in larger proteins. An extension of the present work could also exploit the fact that any lanthanide that generates PCSs also causes partial alignment of the protein with the magnetic field axis, giving rise to RDCs (Bertini et al., 2002a). It is well known that alignment tensors with different orientations, as obtained from lanthanide tags at different sites, greatly enhance the structural information content of RDCs (Bax et al., 2001; Tolman and Ruan, 2006; Bouvignies et al., 2007; Lange et al., 2008). Multiple orientations will augment the experimental data set particularly if only one-bond RDCs between the ^{15}N and ^{1}H spins of backbone amides can be measured.

In conclusion, paramagnetic lanthanides attached at multiple different sites offer a potent strategy to pinpoint the global fold of a protein by the use of PCSs, even in the absence of additional structural restraints. Most importantly, only PCSs of backbone amide protons are required, opening the door to solution structure determinations of proteins of limited solubility, for which other NMR parameters are difficult to measure.

EXPERIMENTAL PROCEDURES

Sample Preparation

Wild-type and mutant samples (S200C/K204D/C157S, A218C/A222D/C157S, and Q241C/N245D/C157S) of the C-terminal domain of rat ERp29-C (Liepinsh et al., 2001) were expressed with a His_6-tag at the C terminus, using the plasmid pETMCSI (Neylon et al., 2000) under control of the T7-promoter. We refer to the three different triple mutants as C200, C218, and C241, respectively. Uniformly ^{15}N- or ^{13}C/^{15}N-labeled samples were expressed in *Escherichia coli* Rosetta (λDE3)/pRARE cells. For protein production, cells were grown overnight in 10 ml of LB medium. This preculture was used to inoculate 1 l of minimal medium containing $^{15}NH_4Cl$ as the sole nitrogen source, or [U-^{13}C] glucose and $^{15}NH_4Cl$ as the sole carbon and nitrogen sources. Cells were grown at 37°C. Protein expression was induced by adding 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at OD_{600} = 0.6 and the cells were harvested 6 hr after induction. All proteins were purified using a HisTrap HP column (GE Healthcare, 5 ml). The mutants C200, C218, and C241 were ligated with the IDA-SH tag following a previously published protocol (Swarbrick et al., 2011), except that the reaction buffer contained 20 mM Tris·HCl, pH 7.6. As a fourth sample, the wild-type protein, which contains a single cysteine residue at position 157, was ligated with the C1 tag (Graham et al., 2011) as described previously (Yagi et al., 2011).

NMR Spectroscopy

All NMR spectra were recorded of about 0.3–0.6 mM solutions of the wild-type and mutant samples of rat ERp29-C in 20 mM MES (pH 4.9) at 31°C, using a Bruker Avance 800 MHz NMR spectrometer with a TCI cryoprobe. PCSs were measured as the ^{1}H chemical shifts observed in ^{15}N-HSQC spectra of ERp29-C tagged with paramagnetic lanthanides (Tm^{3+}, Tb^{3+}) minus the corresponding chemical shifts observed for the diamagnetic reference (Y^{3+}). For the wild-type protein and the C200 mutant, the paramagnetic shifts were verified for all samples by 3D HNCO spectra of the ^{13}C/^{15}N-labeled protein. To minimize residual anisotropic chemical shifts that can be associated with the chemical shifts of heteronuclei with large chemical shift anisotropies (John et al., 2005), we measured PCSs only for ^{1}H spins.

GPS-Rosetta Calculations

GPS-Rosetta is an extended version of PCS-Rosetta (Schmitz et al., 2012), which uses experimental sets of PCSs from multiple metal binding sites for the simultaneous optimization of all $\Delta\chi$ tensors during the low-resolution, backbone-only folding simulation of Rosetta. The PCS data from each of the metal binding centers were scored using

$$s_k = \sum_{q=1}^{m} \sqrt{\sum_{p=1}^{n_{pcs}} \left(PCS_{calc}^{pq} - PCS_{exp}^{pq} \right)^2}, \quad (2)$$

where m is the number of PCS data sets (one data set per metal ion) per binding site k and n_{pcs} is the number of PCSs in the data set. A total score S_{total} was then calculated as the weighted sum of the individual scores s_k and added to the low-resolution energy function of Rosetta:

$$S_{total} = \sum_{k=1}^{n} s_k . w_k, \quad (3)$$

where n is the total number of metal binding centers and w denotes the weighting factor relative to the Rosetta *ab initio* scoring function.

Rosetta 9- and 3-residue fragment libraries were generated using the experimental backbone chemical shifts of the NOE structure of rat ERp29-C (BMRB accession code 4920; Liepinsh et al., 2001) to aid in selecting fragments of

higher quality, as in the standard CS-Rosetta protocol (Shen et al., 2008). In addition, to help resolve the differences between the NOE structure (Protein Data Bank [PDB] ID 1G7D; Liepinsh et al., 2001) and the crystal structure (PDB ID 2QC7; Barak et al., 2009) of rat and human ERp29-C, respectively, the backbone dihedral angles of both structures were explicitly added to the otherwise nonhomologous fragment library.

The PCS scoring was simultaneously optimized during the backbone folding simulation for the lanthanide tags attached at Cys157, C200, C218, and C241, and the weighting factor w for each of the n centers was calculated independently by

$$w = \left(\frac{a_{high} - a_{low}}{c_{high} - c_{low}}\right) \Big/ n, \quad (4)$$

where a_{high} and a_{low} are the averages of the highest and lowest 10% of the values of the Rosetta *ab initio* score, and c_{high} and c_{low} are the averages of the highest and lowest 10% of the PCS score value obtained by scoring 1000 decoys with a unity weighting factor.

In α-helical secondary structure, lanthanide ions bound to an IDA-SH tag in position i also coordinate to the side chain carboxyl group of an aspartate in position i+4. Therefore, the metal ion was positioned 5.9 Å from the C^{α} atom of residue i+4 on a line that joins the backbone amide nitrogen of residue i+6 with the C^{α} atom of residue i+4. In contrast, the position of the lanthanide in the C1 tag, which was ligated to Cys157, was dynamically optimized during the folding simulation, using the strategy implemented in PCS-Rosetta (Schmitz et al., 2012).

ACCESSION NUMBERS

The coordinates of the GPS-Rosetta structure of rat ERp29-C were deposited in the Protein Data Bank with the accession code 2M66.

SUPPLEMENTAL INFORMATION

Supplemental Information includes four tables, Supplemental Experimental Procedures, and three figures and can be found with this article online at http://dx.doi.org/10.1016/j.str.2013.04.001.

ACKNOWLEDGMENTS

We thank Dr. Souren Mkrtchian for the plasmid encoding ERp29 and the supercomputing facility at the King Abdulla University of Science and Technology (KAUST, Saudi Arabia) for providing access to the Blue Gene/P (Shaheen) supercomputer. Financial support by the Australian Research Council, including a Future Fellowship to T.H., is gratefully acknowledged.

Received: February 18, 2013
Revised: March 28, 2013
Accepted: April 1, 2013
Published: May 2, 2013

REFERENCES

Allegrozzi, M., Bertini, I., Janik, M.B.L., 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.

Arnesano, F., Banci, L., and Piccioli, M. (2005). NMR structures of paramagnetic metalloproteins. Q. Rev. Biophys. *38*, 167–219.

Bambang, I.F., Lee, Y.K., Richardson, R., and Zhang, D. (2013). Endoplasmic reticulum protein 29 regulates epithelial cell integrity during the mesenchymal-epithelial transition in breast cancer cells. Oncogene *32*, 1240–1251.

Barak, N.N., Neumann, P., Sevvana, M., Schutkowski, M., Naumann, K., Malesević, 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.

Barthelmes, K., Reynolds, A.M., Peisach, E., Jonker, H.R., DeNunzio, N.J., Allen, K.N., Imperiali, B., and Schwalbe, H. (2011). Engineering encodable lanthanide-binding tags into loop regions of proteins. J. Am. Chem. Soc. *133*, 808–819.

Bax, A., Kontaxis, G., and Tjandra, N. (2001). Dipolar couplings in macromolecular structure determination. Methods Enzymol. *339*, 127–174.

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

Bertini, I., Luchinat, C., and Parigi, G. (2001). Solution NMR of Paramagnetic Molecules: Applications to Metallobiomolecules and Models (Amsterdam: Elsevier).

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

Bertini, I., Longinetti, M., Luchinat, C., Parigi, G., and Sgheri, L. (2002b). Efficiency of paramagnetism-based constraints to determine the spatial arrangement of α-helical secondary structure elements. J. Biomol. NMR *22*, 123–136.

Bertini, I., Luchinat, C., Parigi, G., and Pierattelli, R. (2008). NMR structures of paramagnetic metalloproteins. Dalton Trans. *2008*, 3782–3790.

Bouvignies, G., Markwick, P.R., and Blackledge, M. (2007). Simultaneous definition of high resolution protein structure and backbone conformational dynamics using NMR residual dipolar couplings. ChemPhysChem *8*, 1901–1909.

Das, S., Smith, T.D., Sarma, J.D., Ritzenthaler, J.D., Maza, J., Kaplan, B.E., Cunningham, L.A., Suaud, L., Hubbard, M.J., Rubenstein, R.C., and Koval, M. (2009). ERp29 restricts Connexin43 oligomerization in the endoplasmic reticulum. Mol. Biol. Cell *20*, 2593–2604.

Farmaki, E., Mkrtchian, S., Papazian, I., Papavassiliou, A.G., and Kiaris, H. (2011). ERp29 regulates response to doxorubicin by a PERK-mediated mechanism. Biochim. Biophys. Acta *1813*, 1165–1171.

Graham, B., Loh, C.T., Swarbrick, J.D., Ung, P., Shin, J., Yagi, H., Jia, X., Chhabra, S., Barlow, N., Pintacuda, G., et al. (2011). DOTA-amide lanthanide tag for reliable generation of pseudocontact shifts in protein NMR spectra. Bioconjug. Chem. *22*, 2118–2125.

Hus, J.C., Marion, D., and Blackledge, M. (2001). Determination of protein backbone structure using only residual dipolar couplings. J. Am. Chem. Soc. *123*, 1541–1542.

John, M., and Otting, G. (2007). Strategies for measurements of pseudocontact shifts in protein NMR spectroscopy. ChemPhysChem *8*, 2309–2313.

John, M., Park, A.Y., Pintacuda, G., Dixon, N.E., and Otting, G. (2005). Weak alignment of paramagnetic proteins warrants correction for residual CSA effects in measurements of pseudocontact shifts. J. Am. Chem. Soc. *127*, 17190–17191.

Keizers, P.H.J., and Ubbink, M. (2011). Paramagnetic tagging for protein structure and dynamics analysis. Prog. Nucl. Magn. Reson. Spectrosc. *58*, 88–96.

Koehler, J., and Meiler, J. (2011). Expanding the utility of NMR restraints with paramagnetic compounds: background and practical aspects. Prog. Nucl. Magn. Reson. Spectrosc. *59*, 360–389.

Kontaxis, G. (2012). An improved algorithm for MFR fragment assembly. J. Biomol. NMR *53*, 149–159.

Lange, O.F., Lakomek, N.A., Farès, C., Schröder, G.F., Walter, K.F., Becker, S., Meiler, J., Grubmüller, H., Griesinger, C., and de Groot, B.L. (2008). Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. Science *320*, 1471–1475.

Liepinsh, E., Baryshev, M., Sharipo, A., Ingelman-Sundberg, M., Otting, G., and Mkrtchian, S. (2001). Thioredoxin fold as homodimerization module in the putative chaperone ERp29: NMR structures of the domains and experimental model of the 51 kDa dimer. Structure *9*, 457–471.

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. Bioconjug. Chem. *24*, 260–268.

Meiler, J., Prompers, J.J., Peti, W., Griesinger, C., and Brüschweiler, R. (2001). Model-free approach to the dynamic interpretation of residual dipolar couplings in globular proteins. J. Am. Chem. Soc. *123*, 6098–6107.

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.

Otting, G. (2008). Prospects for lanthanides in structural biology by NMR. J. Biomol. NMR *42*, 1–9.

Otting, G. (2010). Protein NMR using paramagnetic ions. Annu. Rev. Biophys. *39*, 387–405.

Pintacuda, G., John, M., Su, X.C., and Otting, G. (2007). NMR structure determination of protein-ligand complexes by lanthanide labeling. Acc. Chem. Res. *40*, 206–212.

Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A.J., Read, R.J., and Baker, D. (2007). High-resolution structure prediction and the crystallographic phase problem. Nature *450*, 259–264.

Rainey-Barger, E.K., Mkrtchian, S., and Tsai, B. (2009). The C-terminal domain of ERp29 mediates polyomavirus binding, unfolding, and infection. J. Virol. *83*, 1483–1491.

Raman, S., Lange, O.F., Rossi, P., Tyka, M., Wang, X., Aramini, J., Liu, G., Ramelot, T.A., Eletsky, A., Szyperski, T., et al. (2010). NMR structure determination for larger proteins using backbone-only data. Science *327*, 1014–1018.

Rasia, R.M., Lescop, E., Palatnik, J.F., Boisbouvier, J., and Brutscher, B. (2011). Rapid measurement of residual dipolar couplings for fast fold elucidation of proteins. J. Biomol. NMR *51*, 369–378.

Rodriguez-Castañeda, F., Haberz, P., Leonov, A., and Griesinger, C. (2006). Paramagnetic tagging of diamagnetic proteins for solution NMR. Magn. Reson. Chem. *44*(Spec No), S10–S16.

Ruan, K., Briggman, K.B., and Tolman, J.R. (2008). De novo determination of internuclear vector orientations from residual dipolar couplings measured in three independent alignment media. J. Biomol. NMR *41*, 61–76.

Shen, Y., Lange, O., Delaglio, F., Rossi, P., Aramini, J.M., Liu, G.H., 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. USA *105*, 4685–4690.

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.

Su, X.C., and Otting, G. (2010). Paramagnetic labelling of proteins and oligonucleotides for NMR. J. Biomol. NMR *46*, 101–112.

Su, X.C., and Otting, G. (2011). Paramagnetic labelling of proteins and oligonucleotides for NMR – Erratum. J. Biomol. NMR *50*, 99–100.

Su, X.C., Liang, H., Loscha, K.V., and Otting, G. (2009). $[Ln(DPA)_3]^{3-}$ is a convenient paramagnetic shift reagent for protein NMR studies. J. Am. Chem. Soc. *131*, 10352–10353.

Suaud, L., Miller, K., Alvey, L., Yan, W., Robay, A., Kebler, C., Kreindler, J.L., Guttentag, S., Hubbard, M.J., and Rubenstein, R.C. (2011). ERp29 regulates DeltaF508 and wild-type cystic fibrosis transmembrane conductance regulator (CFTR) trafficking to the plasma membrane in cystic fibrosis (CF) and non-CF epithelial cells. J. Biol. Chem. *286*, 21239–21253.

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. Engl. *50*, 4403–4406.

Tian, F., Valafar, H., and Prestegard, J.H. (2001). A dipolar coupling based strategy for simultaneous resonance assignment and structure determination of protein backbones. J. Am. Chem. Soc. *123*, 11791–11796.

Tolman, J.R., and Ruan, K. (2006). NMR residual dipolar couplings as probes of biomolecular dynamics. Chem. Rev. *106*, 1720–1736.

Wang, J., Walsh, J.D., Kuszewski, J., and Wang, Y.X. (2007). Periodicity, planarity, and pixel (3P): a program using the intrinsic residual dipolar coupling periodicity-to-peptide plane correlation and ϕ/ψ angles to derive protein backbone structures. J. Magn. Reson. *189*, 90–103.

Wöhnert, J., Franz, K.J., Nitz, M., Imperiali, B., and Schwalbe, H. (2003). Protein alignment by a coexpressed lanthanide-binding tag for the measurement of residual dipolar couplings. J. Am. Chem. Soc. *125*, 13338–13339.

Wu, P., Zhang, H., Qi, L., Tang, Q., Tang, Y., Xie, Z., Lv, Y., Zhao, S., and Jiang, W. (2012). Identification of ERp29 as a biomarker for predicting nasopharyngeal carcinoma response to radiotherapy. Oncol. Rep. *27*, 987–994.

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

Yagi, H., Banerjee, D., Graham, B., Huber, T., Goldfarb, D., and Otting, G. (2011). Gadolinium tagging for high-precision measurements of 6 nm distances in protein assemblies by EPR. J. Am. Chem. Soc. *133*, 10418–10421.

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.

Zhang, D., and Richardson, D.R. (2011). Endoplasmic reticulum protein 29 (ERp29): An emerging role in cancer. Int. J. Biochem. Cell Biol. *43*, 33–36.

