

Structure restraints from heteronuclear pseudocontact shifts generated by lanthanide tags at two different sites

Benjamin J. G. Pearce¹, Shereen Jabar¹, Choy-Theng Loh¹, Monika Szabo², Bim Graham²,
Gottfried Otting^{1*}

¹ Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

² Monash Institute of Pharmaceutical Sciences, Monash University, Parkville VIC 3052, Australia

* Corresponding author

phone: +61 2 61256507

fax: +61 2 61250750

e-mail: gottfried.otting@anu.edu.au

Keywords: human ubiquitin; lanthanide tag; pseudocontact shift; residual anisotropic chemical shifts; structure determination

Abstract

Pseudocontact shifts (PCS) encode long-range information on 3D structures of protein backbones and side-chains. The level of structural detail that can be obtained increases with the number of different sites tagged with a paramagnetic metal ion to generate PCSs. Here we show that PCSs from two different sites can suffice to determine the structure of polypeptide chains and their location and orientation relative to the magnetic susceptibility tensor χ , provided that PCSs are available for ¹H as well as heteronuclear spins. In addition, PCSs from two different sites are shown to provide detailed structural information on the conformation of methyl group-bearing amino-acid side-chains. A previously published ensemble structure of ubiquitin is shown to explain the magnetic susceptibility and alignment tensors slightly better than structures that try to explain the experimental data by a single conformation, illustrating the potential of PCSs as a tool to investigate small conformational changes.

Introduction

Whilst the first NMR structures relied on the observation and assignment of many nuclear Overhauser effects (NOE) (Wüthrich 1986), other phenomena such as residual dipolar couplings (RDC), paramagnetic relaxation enhancements (PRE) and pseudocontact shifts (PCS) offer complementary data that give structural detail over different distance ranges and are particularly useful for refining and verifying protein structures (for reviews, see Bertini and Luchinat 1999; Bax et al. 2005; Nitsche and Otting 2017). The present work investigates the structural information on polypeptide conformation that can be extracted from PCSs generated by lanthanide ions at two different sites in a protein.

PCSs are manifested as changes in chemical shifts, $\Delta\delta^{\text{PCS}}$, observed in a paramagnetic versus corresponding diamagnetic protein sample. They follow the equation (Bertini et al. 2002)

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

where $\Delta\chi_{\text{ax}}$ and $\Delta\chi_{\text{rh}}$ are the axial and rhombic components of the magnetic susceptibility tensor χ respectively, and r , θ and ϕ are the polar coordinates of the nuclear spin relative to the principal axes of the $\Delta\chi$ tensor, which represents the anisotropy of the χ tensor and has the same principal axes.

PCSs present excellent structural information and, before the advent of NOEs for 3D structure determination, early efforts were directed towards obtaining the requisite restraints from PCSs generated by lanthanide ions (Barry et al. 1971; Bleaney et al. 1972; Marinetti et al. 1975). This approach proved to be compromised by the difficulty in determining the magnetic susceptibility anisotropy ($\Delta\chi$) tensor associated with a paramagnetic metal ion. In practice, the same paramagnetic metal ion attached at a different site can result in a different $\Delta\chi$ tensor, depending on the ligand field of the metal complex. Furthermore, any mobility of the metal relative to the protein modifies the PCSs in the protein. The $\Delta\chi$ -tensor parameters must thus be determined anew for each system and reflect effective $\Delta\chi$ tensors (Shishmarev and Otting 2013).

It has recently been shown that this does not exclude 3D structure determinations of proteins from PCSs only. In fact, algorithms have been developed that determine protein

structures using PCSs of backbone amide protons as the only experimental restraints, provided lanthanide ions can be attached at multiple different sites (Yagi et al. 2013; Pilla et al. 2016; 2017). This approach is computationally expensive, as a large number of 3D models of the protein need to be built and evaluated with regard to consistent $\Delta\chi$ tensor fits. Often, however, part of the protein structure is already known (e.g. from a crystal structure) and can be used for fitting $\Delta\chi$ tensors. In this situation, PCSs provide excellent restraints for structure determination of previously uncharacterised parts of the protein (de la Cruz et al. 2011; Pilla et al. 2015; Skinner et al. 2015; Chen et al. 2016).

In the past, applications of PCSs generated by lanthanide tags focused on PCSs of amide protons only, attempted to generate PCSs from multiple different tagging sites and furnished each tagging site with different paramagnetic lanthanides. In all cases, the final structure assembly was assisted by atomic force fields. The present work investigates the use of PCSs as the sole experimental restraints, which were recorded for only two samples of human ubiquitin with lanthanide tags at different sites. Apart from this methodological investigation, the PCS data present new structural restraints to validate previously published structures of ubiquitin.

We assume that $\Delta\chi$ tensors are available right from the beginning, requiring that at least part of the protein structure is known and can be used to determine the $\Delta\chi$ tensors. The $\Delta\chi$ tensors are subsequently used to determine the structure of polypeptide segments. In contrast to previous studies, we include heteronuclear PCSs, disregard force fields and omit any other experimental NMR restraints. By limiting the number of tagging sites to two and using PCSs from only a single lanthanide ion (Tb^{3+}), we probe the level of 3D structure information that can be obtained with two lanthanide-tag sites. In addition, we explore the potential of two $\Delta\chi$ tensors for determining the conformations of amino-acid side-chains.

Materials and Methods

Sample preparation

Single cysteine residues were introduced into human ubiquitin, fused with a C-terminal SerHis₆ peptide for easier purification, by making the mutants A28C and S57C. Uniformly ¹⁵N/¹³C-labelled protein samples were expressed from pETMCSI vectors (Neylon et al. 2000)

in *E. coli* BL21 (DE3). Cells were incubated in 1 L LB medium until the OD₆₀₀ reached 0.6-0.8 and were then transferred to 300 mL of M9 medium supplemented with ¹⁵NH₄Cl and D-glucose-¹³C₆, inducted with IPTG and incubated at room temperature for about 16 h. Following cell lysis by a French Press, the protein was purified by a Co-NTA column and the C-terminal SerHis₆ peptide was removed with ubiquitinase (Catanzariti et al. 2004). Typical final yields were 15 mg of purified protein per litre of culture medium. Immediately prior to tagging, DTT was removed by passing through a PD-10 desalting column equilibrated with 50 mM HEPES buffer at pH 7.5. The samples were labelled with the C1 tag (Graham et al. 2011) loaded with either Y³⁺, Tb³⁺ or Tm³⁺. The tagging reaction followed a previously published protocol (Graham et al. 2011), except that only a 5-fold excess of tag was used and the reaction was performed overnight at room temperature. The final NMR samples contained 0.1 mM solutions of tagged ubiquitin in NMR buffer (20 mM phosphate, pH 6.5).

NMR measurements

All NMR data were acquired at 25 °C on a Bruker 800 MHz NMR spectrometer equipped with a TCI cryoprobe. PCSs of ¹H-NMR signals of amide protons were measured in ¹⁵N-HSQC spectra recorded with acquisition times $t_{1\max} = 90$ ms and $t_{2\max} = 122$ ms. PCSs of ¹³C-NMR signals of carbonyl carbons were measured in 3D HNCO spectra recorded with $t_{\max}(\text{¹³C}) = 45$ ms. PCSs of ¹³C-NMR signals of α -carbons were measured in 3D HN(CO)CA spectra recorded with $t_{\max}(\text{¹³C}) = 11$ ms. PCSs of α -protons were measured in ¹³C-HSQC spectra recorded with $t_{1\max} = 16$ ms and $t_{2\max} = 122$ ms. The ¹³C-HSQC spectra were also used to extract the ¹H and ¹³C PCSs of methyl groups.

Residual dipolar couplings between ¹H and ¹⁵N spins of backbone amides, ¹D_{HN}, were measured by the IPAP experiment (Ottiger et al. 1998), using $t_{1\max} = 45$ ms and $t_{2\max} = 121$ ms. The RDCs were measured as the doublet splittings observed in the samples with C1-Tb³⁺ tag minus those observed with the C1-Y³⁺ tag.

$\Delta\chi$ -tensor determinations

PCSs were measured as the chemical shifts of the paramagnetic samples minus the chemical shifts of the diamagnetic references. To estimate the uncertainty ranges of the PCSs, the uncertainty in cross-peak position, σ , was estimated from its signal-to-noise ratio, S/N , and the acquisition time, T_{ac} , using the equation (Kontaxis et al. 2000)

$$\sigma = 0.66 N/(S \times T_{ac}) \quad (2)$$

The PCSs of backbone atoms were used to fit $\Delta\chi$ tensors to published coordinates of ubiquitin, using an in-house program based on the Levenberg-Marquardt algorithm (Marquardt 1963; More 1977) that minimizes the RMSD between experimental and back-calculated PCSs. The quality of fit was assessed by a Q factor calculated as the ratio of the root-mean-square deviation between experimental and back-calculated PCSs and the root-mean-square of the experimental PCSs.

Alignment-tensor determinations

Paramagnetic lanthanides that generate PCSs also lead to partial molecular alignment in the magnet field and, hence, to RDCs. The $^1D_{\text{HN}}$ RDCs were used to fit alignment tensors A to published coordinates of ubiquitin, using a suitably adapted version of the algorithm used for fitting $\Delta\chi$ tensors. For comparison, fitted alignment tensors A were converted into $\Delta\chi$ tensors using

$$\Delta\chi_{\text{ax,rh}} = (15\mu_0 k_B T / B_0^2) A_{\text{ax,rh}} \quad (3)$$

where μ_0 is the magnetic susceptibility of vacuum, k_B the Boltzmann constant, T the temperature and B_0 the magnetic field strength.

RACS calculation

The following equation was used to predict the change in chemical shift, $\Delta\delta^{\text{RACS}}$, for backbone nitrogens and carbonyl carbons due to the residual anisotropic chemical shift (RACS) effect (John et al. 2005)

$$\Delta\delta^{\text{RACS}} = B_0^2 / (15\mu_0 k_B T) \sum_{kl} \sigma_{kk}^{\text{CSA}} \cos^2\theta_{kl} \Delta\chi_{ll} \quad (4)$$

where σ_{kk}^{CSA} and $\Delta\chi_{ll}$ are the components of the CSA and $\Delta\chi$ tensors along the k and l principal axes, respectively, and θ_{kl} is the angle between them.

Results

Mutant samples of ubiquitin were prepared with single cysteine residues at positions 4, 14, 18, 28, 42, 44, 57 and 66, and labelled with the C1 tag loaded with Tb^{3+} . Large PCSs were observable only in ^{15}N -HSQC spectra of the mutants A28C and S57C (Fig. 1). The tagging reaction failed at positions 14, 18 and 44 and the PCSs were very small after tagging at positions 4, 42 and 66, suggesting averaging by reorientational motions of the tag. Consequently, all subsequent experiments were performed with the A28C and S57C mutants only. Lanthanide tags at positions 28 and 57 have previously been shown to generate significant PCSs and RDCs (Graham et al. 2011; Peters et al. 2011).

As observed previously for other proteins labelled with C1- Ln^{3+} tags (Graham et al. 2011; de la Cruz et al. 2011; Shishmarev et al. 2013; Yagi et al. 2013), the cross-peaks of samples tagged with C1- Tb^{3+} and C1- Tm^{3+} shifted in opposite directions relative to the diamagnetic reference produced with C1- Y^{3+} tag. In addition, the three cross-peaks obtained with the Tm^{3+} , Y^{3+} and Tb^{3+} tags for any particular spin pair were found to lie on diagonal lines, enabling resonance assignments of most of the paramagnetic cross-peaks by visual inspection. The remaining paramagnetic cross-peaks were assigned by predicting their PCSs from the $\Delta\chi$ tensors established from the PCSs measured from resolved cross-peaks.

NMR measurements

PCSs were measured for the polypeptide backbone and for side-chain atoms, using ^{15}N -HSQC spectra for backbone amides, ^{13}C -HSQC spectra for C^{α} and H^{α} atoms as well as for side-chain methyl groups, HN(CO)CA spectra for C^{α} resonances that could not be resolved in the ^{13}C -HSQC spectra, and HNCOSY spectra for carbonyl carbons. As the PCSs obtained with C1- Tm^{3+} were smaller than those obtained with C1- Tb^{3+} tags, we subsequently used only paramagnetic data obtained with Tb^{3+} . Figure 2 shows an overview of the PCSs measured for the backbone atoms of different residues.

A total of 267 and 268 PCSs of backbone atoms were measured for the A28C and S57C mutants, respectively. For about half the residues, eight or more PCSs could be measured. Cross-peaks were lost mostly due to excessive line broadening caused by PREs in the vicinity of the paramagnetic centres. In addition, $^1D_{\text{NH}}$ RDCs were measured for 52 backbone amides of the A28C mutant and 48 backbone amides of the S57C mutant.

An ensemble structure fits the PCS data best

The PCS data of backbone atoms were used to fit $\Delta\chi$ tensors to atom coordinates from different published structures of ubiquitin. Tensors were fitted by minimizing the RMSD between experimental PCSs and those back-calculated using Eq. 1. Similarly, alignment tensors were fitted using the $^1D_{\text{NH}}$ RDC data. In fitting the $\Delta\chi$ tensors to structural ensembles, we assumed the same weight for each conformation, using a single tensor to back-calculate the PCS data for each conformation and averaging the PCSs for each spin to arrive at the back-calculated PCS value for comparison with the experimental value. Individual RDC values were calculated and averaged in the same way in the alignment-tensor fits. The fits treated all structures with more than a single conformer as ensembles, even though each conformer in the structures 1D3Z and 2MJB aims to fulfil all restraints individually, whereas the structures 2K39 and 2KOX were explicitly determined to satisfy experimental RDCs by back-calculating averages of RDCs from multiple conformations.

The quality factors of the tensor fittings are reported in Table 1. The table shows that the structural ensemble 2KOX overall delivered the best fits, whereas the crystal structure 1UBQ mostly delivered relatively poor fits. The result did not change, when only residues 1-70 were included in the fit (Table S1). It also did not change, when only residues with low backbone rmsd values were included in the fit ($\text{rmsd} \leq 0.4 \text{ \AA}$ as proposed by Maltsev et al. 2014) or only residues with large backbone rmsd ($> 0.4 \text{ \AA}$), in that the 2KOX structure yielded either the smallest Q factors or Q factors that were within the uncertainty range of the smallest Q factor (Table S2).

An ensemble representation can also be generated by a collection of ubiquitin structures observed in different crystal and co-crystal structures and has previously been used successfully to interpret RDC, residual chemical shift anisotropy and $^3J_{\text{HNH}\alpha}$ data (Maltsev et al. 2014). Using the data of the present work with the weighted ensemble of ubiquitin crystal structures proposed by Maltsev et al. (2014), denoted <x-ray> in Table 1, our tensor fits produced mostly larger Q factors than the 2KOX ensemble. This result did not change when the crystal structures were used with equal weighting factor (data not shown). Similarly, the Q factors remained within 0.02 of the values in Tables 1 and S1, when average structures were used to perform the $\Delta\chi$ tensor fits, regardless of whether data for the flexible C-terminus (residues 71-76) were included or not. As expected for ensemble representations versus single-structure representations, when only the first conformer of each ensemble was

considered, the Q factors barely changed for the structures 1D3Z and 2MJB but were significantly worse for the ensemble structures 2K39 and 2KOX. In the following, we used the 2KOX ensemble for further analysis.

For both ubiquitin mutants, the long axis of the fitted $\Delta\chi$ and alignment tensors were similar in orientation as shown by the similarity in α and β angles, but the alignment tensors were smaller in magnitude than the $\Delta\chi$ tensors (Table 2). Even allowing for variable bond orientations of the backbone amides by assuming an order parameter of 0.9, the alignment tensors were smaller than expected, indicating incomplete immobilization of the metal ions relative to the protein. Nonetheless, the correlations between experimental and back-calculated PCSs were excellent (Fig. 3).

The PCS correlations could be improved slightly by taking the RACS effect into account, which cause changes in chemical shifts in paramagnetic samples due to weak molecular alignment in the magnetic field (John et al. 2005). Due to small CSA tensors, no significant RACS effects were expected for $^1\text{H}^\alpha$ and ^{13}C nuclei. Using average CSA tensors as implemented in the program Numbat (Schmitz et al. 2008), maximum RACS corrections were estimated to be 2 ppb for $^1\text{H}^\text{N}$, 34 ppb for ^{15}N and 31 ppb for ^{13}C ' spins. Average RACS corrections were predicted to be 60% smaller. Accounting for these RACS effects reduced the quality factors to 0.058 and 0.054 for the two mutants, respectively (Table S3), versus 0.061 and 0.057 without RACS correction (Table 1). Omitting residues 71-76 from the fit to account for peptide flexibility produced a minor improvement in the quality factors to 0.057 and 0.051 for the A28C and S57C mutants, respectively, when RACS corrections were included (Table S3). As the RACS corrections were small and computationally costly, we did not use them in the subsequent structure computations.

Calculations of backbone coordinates from PCSs

To probe the structural information content associated with $\Delta\chi$ -tensor qualities as obtained for the ubiquitin mutants A28C and S57C, an algorithm was developed to back-calculate the structures of segments of the ubiquitin polypeptide chain from the heteronuclear PCSs of the backbone atoms and the $\Delta\chi$ tensors reported in Table 2. The algorithm determined the backbone structures of the polypeptide segments by minimizing the RMSD between experimental and back-calculated PCSs. For each segment, the dihedral angles ϕ and ψ dihedral angles were optimised as well as six parameters describing the position and

orientation of the fragment relative to the two $\Delta\chi$ tensors. Bond lengths and angles were constrained according to previously reported values reported (Tables 2 and 3 by Engh and Huber 1991 and Table 14 by Cornell et al. 1995) and the ω angle was assumed to be 180° . The algorithm started with a pentapeptide for which ten backbone PCSs were available for at least four residues (Fig. 2). The dihedral angles and the position and orientation of the pentapeptide in space constituted a total of fourteen parameters. 50 initial sets of these parameters were generated randomly and optimized for minimal PCS RMSD using the Levenberg-Marquardt algorithm (Marquardt 1963). The initial parameters placed the peptides in a box surrounding the coordinates of the 2KOX structure. The result with the lowest RMSD was used as starting point for extending the peptide beyond five residues. Each time an amino-acid residue was added, a coarse grid search was performed for the dihedral angles of the terminal four residues and the conformation with the lowest PCS RMSD was used as the starting point for subsequent refinement by the Levenberg-Marquardt algorithm, adjusting the parameters of the whole polypeptide segment. The robustness of the structure computations was tested by repeating the calculations 20 times after the addition of random Gaussian noise to the PCSs.

For convergence of the algorithm, it was critically important that complete PCS data were available for all residues of the starting pentapeptide. Furthermore, the good structural definition of the polypeptide segments was lost as soon the segments were extended into regions with residues for which PCSs could be measured in only one of the mutants, showing the importance of PCSs generated from two different sites.

Most of the peptide conformations obtained were in remarkably good agreement with the 2KOX structure, both in position and orientation (Fig. 4). The agreement with the dihedral angles ϕ and ψ observed in the structure 2KOX was mostly good, but not always satisfactory (Fig. S1), highlighting the strength of PCSs to determine the location and direction of a polypeptide chain better than local structural detail. With regard to location and orientation, the obvious outlier observed for the fragment comprising residues 1 to 19 had a significantly higher PCS RMSD than the other conformations for that fragment. A similarly obvious outlier observed for the fragment comprising residues 38 to 44, however, had a rather low RMSD and could be rejected only on the basis of being located unacceptably far from the other polypeptide segments calculated. The fragment comprising residues 47 to 53 was most sensitive to the added noise but, on average, still matched the 2KOX structure well.

The C-terminal fragment included flexible residues from the C-terminus of ubiquitin. Due to this flexibility, the C-terminal part of this segment is an equally poor representation of the real situation as the average coordinates of the 2KOX structure. It was therefore excluded from the RMSD calculation shown in Table 3.

Determination of side-chain conformations from PCSs

In the same way in which PCSs can be used to determine backbone conformations using independently determined $\Delta\chi$ tensors, PCSs can also be used to determine the conformations of amino-acid side-chains. In particular, PCSs of methyl groups are readily measured in ^{13}C -HSQC spectra. Therefore, we investigated the possibility of determining the side-chain conformations of valine, threonine, leucine and isoleucine residues using the two $\Delta\chi$ tensors of Table 2. The average coordinates of the 2KOX ensemble were used for the backbone coordinates together with published stereospecific resonance assignments of the methyl groups of isopropyl groups of valine and leucine (BMRB 17769; Cornilescu et al. 1998). The dihedral angles of the methyl-group bearing amino-acid side-chains were established by minimizing the PCS RMSDs of the methyl groups. The back-calculated ^1H -PCS value of each methyl group was determined as the average of the PCSs calculated for each of the three protons. The χ_1 angle or χ_1/χ_2 angle combination with the lowest PCS RMSD was compared with the corresponding angles reported by the 2KOX ensemble. The analysis was performed only for the subset of residues, for which PCSs could be measured in both ubiquitin mutants.

For most of the methyl-containing side chains for which PCS data could be measured, the dihedral angles determined in this way matched the published structural ensemble with remarkable accuracy. In two cases (Thr7 and Thr9), the published dihedral angles matched RMSD minima that were higher than the lowest RMSD minimum. Notably, however, the absolute RMSD minima were only slightly lower and involved eclipsed conformations, which are unlikely to be correct (Fig. 5). Large discrepancies were observed for residues 70, 71 and 73 from the C-terminal segment, because the average coordinates of 2KOX do not adequately reflect the backbone PCSs in this flexible segment. The side chain of Ile44 is quite solvent-exposed and the PCSs seem to confirm an equilibrium between different χ_2 angles as indicated by the 2KOX ensemble (Fig. 6).

To assess the value of ^{13}C PCSs for determining the coordinates of methyl groups, we repeated the calculations by using the ^1H PCSs of the methyl groups only. For most residues,

local minima of PCS RMSDs were found at very similar positions as in Figures 5 and 6 but the minima tended to be less pronounced, reducing the confidence of dihedral angle determination (Figs S2 and S3).

Finally, the PCSs allow a comparison of the dihedral angles reported by the ensemble 2K0X with the side-chain χ_1 angles and populations determined by Chou et al. (2004) from scalar and dipolar couplings for isoleucine, threonine and valine residues. Fig. 7 shows that the side-chain conformations reported by Chou et al. (2004) and the ensemble 2K0X fulfill the experimental PCSs similarly well. For the data of Fig. 7, the RMSD between the experimental and back-calculated PCSs was 0.034 ppm for the conformations of Chou et al. (2004) and 0.038 ppm for the 2K0X conformations.

Discussion

The case for heteronuclear PCSs from two sites

PCSs from a single lanthanide-tag site present highly degenerate restraints, as can readily be appreciated by representing a $\Delta\chi$ tensor by PCS isosurfaces, which identify the possible coordinates for a nuclear spin with a certain PCS value (Fig. 8a). In principle, $\Delta\chi$ tensors from metal ions at four different positions are required for unambiguous determination of the position of a nuclear spin relative to those tensors from its PCSs. In practice, however, it can be laborious and difficult to produce four protein samples with tags and obtain sizeable PCSs from each tagging site.

Availability of PCSs from two different $\Delta\chi$ tensors restricts the coordinates of the nuclear spin to the line presented by the intersection of the respective PCS isosurfaces (Fig. 8b). It is very likely that the coordinates are restrained further, if the PCSs are available for two neighbouring spins separated by a chemical bond, as bond lengths are known with high accuracy. The spatial localization restraint becomes even stronger, if PCSs are available for three neighbouring spins linked by two chemical bonds (e.g., as in a CH_2 group or for the backbone atoms $\text{N}-\text{C}^\alpha-\text{C}'$ of an amino acid) as the bond angle between two chemical bonds is also well known.

Our results show that, indeed, the backbone structures of polypeptide chains can be determined from PCSs generated by just two $\Delta\chi$ tensors with a single paramagnetic metal ion. Besides determining the structure of the peptide, the data also yield its position and orientation relative to the $\Delta\chi$ tensors. Such experiments are eminently useful for determining

the 3D structure of a peptide in complex with a protein of known 3D structure that is labelled with paramagnetic tags. The same protocol allows re-determination of the structure of a protein segment after a change in structure, e.g. in response to ligand binding. The present study did not rely on any other information beyond bond lengths and bond angles to assist the structure determinations. Obviously, the approach can be enhanced by the addition of force fields or other experimental restraints such as RDCs, PREs, NOEs or chemical shifts indicative of specific secondary structure, but it is remarkable that PCSs alone already contain the requisite high-level structural information.

We did not use RDCs for structure calculations, although RDCs automatically result from the molecular alignment with the external magnetic field as a consequence of paramagnetic tags with anisotropic χ tensors. In general, however, RDCs are harder to measure than PCSs because RDCs are reflected in much smaller changes in peak positions. This held true also for the present mutants of ubiquitin, where $^1D_{\text{HN}}$ couplings proved to be the RDCs measured most easily. Yet, the quality factors for the alignment tensor determined from back-calculated and experimental $^1D_{\text{HN}}$ RDCs were three to four times greater than for the PCSs (Table 1), i.e. the correlation between back-calculated and experimental data was much poorer, making it correspondingly harder to derive structural information from the RDC data. Nonetheless, while not investigated in the present work, the $^1D_{\text{HN}}$ RDC data may help to reconcile any deviations remaining between the backbone dihedral angles determined here and in previous structure calculations (Fig S1).

PREs tend to deliver even worse correlations between back-calculated and experimental data than RDCs, require more experiments for accurate measurements, are affected by DSA-CSA cross-correlation effects for samples with anisotropic χ tensors (Pintacuda et al. 2004; Orton et al. 2016) and can easily be compromised by intermolecular effects. Therefore we did not attempt to measure PREs.

The quality of the correlation plots of Figure 3 is outstanding, which is all the more remarkable in view of the apparent mobility of the tags as indicated by the relatively small size of the alignment tensors compared with the $\Delta\chi$ tensors, and the relatively small magnitudes of the $\Delta\chi$ tensors compared with $\Delta\chi$ tensors observed for other proteins with C1-Ln³⁺ tags previously (Graham et al. 2011; de la Cruz et al. 2011; Yagi et al. 2013). Although the $\Delta\chi$ tensors determined by the tensor fits are thus effective rather than true $\Delta\chi$ tensors (Shishmarev and Otting 2013), the high quality of the correlations between back-calculated

and experimental PCSs indicates that the effective $\Delta\chi$ tensors are of very high predictive value.

The structure calculations of the polypeptide segments from PCSs generated by only two $\Delta\chi$ tensors critically depended on the availability of heteronuclear PCSs. If only ^1H PCSs were available, the system would be underdetermined, as there are at most two backbone hydrogens per residue and at least two backbone dihedral angles need to be determined. The heteronuclear PCSs increased the number of restraints and allowed determination of the position and orientation of most of the polypeptides in space. For samples tagged at a single site, PCS data sets can be increased by measuring samples with different paramagnetic metal ions that produce different $\Delta\chi$ tensors. In this case, however, the symmetry of $\Delta\chi$ tensors makes the position of peptide segments ambiguous (Kobashigawa et al. 2012; Guan et al. 2013).

RACS corrections

When PCSs are measured for two or more bonded nuclei, at least one of the PCSs is from a heteronucleus (^{13}C or ^{15}N). Unfortunately, backbone amide nitrogens and carbonyl carbons feature large chemical shift anisotropies, which cause changes in chemical shifts once the molecule is partially aligned in the magnetic field as a consequence of anisotropic magnetic susceptibility tensors. For nuclear spins located far from the paramagnetic centre, such residual anisotropic chemical shift (RACS) effects can seriously compromise the information content of PCSs. They appear only in the presence of the paramagnetic metal ion and thus are difficult to disentangle from PCSs. As RACS effects arise from the molecular alignment, they do not depend on the distance from the paramagnetic centre and therefore have a disproportionate impact on small PCS values of nuclear spins far from the metal ion. In the present work, this proved to be less critical, as the alignment tensors were too small to generate large molecular alignment even in the magnet of an 800 MHz NMR spectrometer. At the same time, ubiquitin is sufficiently small so that significant PCSs can be measured also for the nuclear spins furthest from the paramagnetic centre. In a larger protein and for tags creating larger $\Delta\chi$ tensors, RACS effects would have to be re-evaluated and may complicate the interpretation of paramagnetic shifts detected for nuclear spins far from the paramagnetic centre.

Ensemble versus single conformations

It is interesting that, overall, the 2KOX ensemble produced the best quality factors in the $\Delta\chi$ tensor fits (Tables 1, S1 and S2). The 2KOX ensemble was created by interpreting a large set of NMR restraints, including $^1D_{\text{HN}}$ RDCs, as averages over the members of the ensemble rather than trying to explain the data by a single conformation (Fenwick et al. 2011). While, in the present work, the 2KOX ensemble produced only a slight improvement in quality factor over other ubiquitin models, the improved fit was reflected by the quality factors of not only the $\Delta\chi$ tensors but also the alignment tensors (Tables 1, S1 and S2). In principle, paramagnetism offers a versatile alternative way over steric and electrostatic alignment media to achieve weak molecular alignment in the magnetic field, improving the chances of measuring information-rich RDCs by generating orthogonal alignments. In view of the high accuracy with which PCSs can be measured, however, PCS data alone may be as attractive as RDCs to detect small conformational changes arising from protein flexibility in solution.

Side-chain conformations

Determination of the side-chain conformations of amino acids from PCSs poses a simpler problem than the structure calculations of polypeptide backbones, if the position of the α -carbon is known. The amino-acid side-chain conformations of threonine, valine and leucine are determined by no more than two dihedral angles, which can thus be determined by two PCSs. The side-chain conformation of isoleucine is determined by three dihedral angles, but with two PCSs available for each methyl group, the number of measurements exceeds the number of parameters to be determined even if only ^1H PCSs are measured. This is one reason why the availability of ^{13}C PCSs is not as important for determining the dihedral angles of the amino-acid side-chains. Another reason are the larger errors associated with ^{13}C PCSs, if they are measured without ^{13}C - ^{13}C decoupling as in the present work. We did not choose constant-time ^{13}C -HSQC experiments in order to maintain best sensitivity in the face of PREs in the paramagnetic samples. Compared with carbonyl carbons, side-chain ^{13}C -nuclei have small CSAs, making RACS corrections unnecessary. For methyl groups far from the paramagnetic centre, it may thus be worthwhile to record more accurate ^{13}C PCSs by constant-time experiments that deliver higher resolution in the ^{13}C dimension.

Conclusion

The present work shows that outstanding correlations between back-calculated and experimental PCSs can be obtained even with single-arm tags that require fitting of effective $\Delta\chi$ tensors. PCSs generated by tags at only two different sites provide sufficient information to determine the structures of polypeptide backbones, even if PCSs are measured with only a single paramagnetic metal ion, provided that the data include PCSs of heteronuclei. We expect that the efficiency of this approach will make it popular for structure determinations of peptides bound to proteins and to assess conformational changes that may arise either from intrinsic mobility or ligand binding. In addition, the PCS data obtained in this work provide independent validation in favour of an ensemble representation of ubiquitin over representations by a single average conformation.

Acknowledgments

Financial support by the Australian Research Council is gratefully acknowledged.

References

- Barry CD, North ACT, Glasel JA, Williams RJP, Xavier AV (1971) Quantitative determination of mononucleotide conformations in solution using lanthanide ion shift and broadening NMR probes. *Nature* 232:236–245
- Bax A, Kontaxis A, Tjandra N (2005) Dipolar couplings in macromolecular structure determination. *Methods Enzymol* 339:127–174
- Bertini I, Luchinat C (1999) New applications of paramagnetic NMR in chemical biology, *Curr Opin Chem Biol* 3:145–151
- Bertini I, Luchinat C, Parigi G (2002) Magnetic susceptibility in paramagnetic NMR. *Prog NMR Spectrosc* 40:249–273
- Bleaney B, Dobson CM, Levine BA, Martin RB, Williams RJP, Xavier AV (1972) Origin of lanthanide nuclear magnetic-resonance shifts and their uses. *J Chem Soc Chem Commun* 13:791b–793
- Catanzariti AM, Soboleva TA, Jans DA, Board PG Baker RT (2004) An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Sci.* 13:1331–1339
- Chen JL, Wang X, Yang F, Cao C, Otting G, Su XC (2016) 3D structure determination of an

- unstable transient enzyme intermediate by paramagnetic NMR spectroscopy. *Angew Chem Int Ed* 55:13744–13748
- Chou JJ, Case DA, Bax A (2003) Insights into the mobility of methyl-bearing side chains in proteins from $^3J_{CC}$ and $^3J_{CN}$ coupling. *J Am Chem Soc* 125:8959–8966
- Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 117:5179–5197
- Cornilescu G, Marquardt JL, Ottiger M, Bax A (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J Am Chem Soc* 120:6836–6837
- de la Cruz L, Nguyen THD, Ozawa K, Shin J, Graham B, Huber T, Otting G (2011) Binding of low-molecular weight inhibitors promotes large conformational changes in the dengue virus NS2B-NS3 protease: fold analysis by pseudocontact shifts. *J Am Chem Soc* 133:19205–19215
- Engh RA, Huber R (1991) Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst A* 47:392–400
- Fenwick RB, Esteban-Martín S, Richter B, Lee D, Walter KF, Milovanovic D, Becker S, Lakomek NA, Griesinger C, Salvatella X (2011) Weak long-range correlated motions in a surface patch of ubiquitin involved in molecular recognition. *J Am Chem Soc* 133:10336–10339
- Graham B, Loh CT, Swarbrick JD, Ung P, Shin J, Yagi H, Jia X, Chhabra S, Barlow N, Pintacuda G, Huber T, Otting G (2011) A DOTA-amide lanthanide tag for reliable generation of pseudocontact shifts in protein NMR spectra. *Bioconjugate Chem* 22:2118–2125
- Guan JY, Keizers PHJ, Liu WM, Löhr F, Skinner SP, Heeneman EA, Schwalbe H, Ubbink M, Siegal G (2013) Small-molecule binding sites on proteins established by paramagnetic NMR spectroscopy. *J Am Chem Soc* 135:5859–5868
- John M, Park AY, Pintacuda G, Dixon NE, 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
- Kobashigawa Y, Saio T, Ushio M, Sekiguchi M, Yokochi M, Ogura K, Inagaki F (2012) Convenient method for resolving degeneracies due to symmetry of the magnetic

- susceptibility tensor and its application to pseudo contact shift-based protein-protein complex structure determination. *J Biomol NMR* 53:53–63
- Kontaxis G, Clore GM, Bax A (2000) Evaluation of cross-correlation effects and measurement of one-bond couplings in proteins with short transverse relaxation times. *J Magn Reson* 143:184–196
- Lange OF, Lakomek NA, Farès C, Schröder GF, Walter KF, Becker S, Meiler J, Grubmüller H, Griesinger C, de Groot BL (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. *Science* 320:1471–1475
- Maltsev AS, Grishaev A, Roche J, Zasloff M, Bax A (2014) Improved cross validation of a static ubiquitin structure derived from high precision residual dipolar couplings measured in a drug-based liquid crystalline phase. *J Am Chem Soc* 136:3752–3755
- Marinetti TD, Snyder GH, Sykes BD (1975) Lanthanide interactions with nitrotyrosine – specific binding-site for nuclear magnetic-resonance shift probes in proteins. *J Am Chem Soc* 97:6562–6570
- Marquardt DW (1963) An algorithm for least-squares estimation of nonlinear parameters. *J Soc Ind Appl Math* 11:431–441
- More, JJ (1978) The Levenberg-Marquardt Algorithm: Implementation and Theory. In *Numerical Analysis*. Springer Berlin, pp. 105-116
- Neylon C, Brown SE, Kralicek AV, Miles CS, Love CA, Dixon NE (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
- Nitsche C, Otting G (2017) Pseudocontact shifts in biomolecular NMR using paramagnetic metal tags. *Prog NMR Spectrosc* 98-99:20–49
- Orton HW, Kuprov I, Loh CT, Otting G (2016) Using paramagnetism to slow down nuclear relaxation in protein NMR. *J Phys Chem Lett* 7:4815–4818
- Ottiger M, Delaglio F, Bax A (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson* 131:373–378
- Peters F, Maestre-Martinez M, Leonov A, Kovačič L, Becker S, Boelens R, Griesinger C (2011) Cys-Ph-TAHA: a lanthanide binding tag for RDC and PCS enhanced protein NMR. *J Biomol NMR* 51:329–337
- Pilla KB, Koehler Leman J, Otting G, Huber T (2015) Capturing conformational states in

- proteins using sparse paramagnetic NMR data. *PLoS ONE* 10:e0127053
- Pilla KB, Otting G, Huber T (2016) Pseudocontact shift-driven iterative resampling for 3D structure determinations of large proteins. *J Mol Biol* 428:522–532
- Pilla KB, Otting G, Huber T (2017) Protein structure determination by assembling super-secondary structure motifs using pseudocontact shifts. *Structure*, in press
- Pintacuda G, Kaikkonen A, Otting G (2004) Modulation of the distance dependence of paramagnetic relaxation enhancements by CSAxDSA cross-correlation. *J Magn Reson* 171:233–243
- Schmitz C, Stanton-Cook MJ, Su XC, Otting G, 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
- Shishmarev D, Otting G (2013) How reliable are pseudocontact shifts induced in proteins and ligands by mobile paramagnetic tags? A modelling study. *J Biomol NMR* 56:203–216
- Shishmarev D, Wang Y, Mason CE, Su XC, Oakley AJ, Graham B, Huber T, Dixon NE, Otting G (2014) Intramolecular binding mode of the C-terminus of Escherichia coli single-stranded DNA binding protein determined by nuclear magnetic resonance spectroscopy. *Nucleic Acids Res* 42:2750–2757
- Skinner SP, Liu WM, Hiruma Y, Timmer M, Blok A, Hass MAS, Ubbink M (2015) Delicate conformational balance of the redox enzyme cytochrome P450cam. *Proc Natl Acad Sci* 112:9022–9027
- Vijay-Kumar S, Bugg CE, Cook WJ (1987) Structure of ubiquitin refined at 1.8 Å resolution. *J Mol Biol* 194:531–544
- Wüthrich K (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York
- Yagi H, Pilla KB, Maleckis A, Graham B, Huber T, Otting G (2013) 3D protein fold determination from backbone amide pseudocontact shifts generated by lanthanide tags at multiple sites. *Structure* 21:883–890

Tables

Table 1 Quality factors from fitting $\Delta\chi$ and alignment tensors to different published ubiquitin structures^a

	Reference	Size of Ensemble	RDC ^b		PCS ^c	
			A28C	S57C	A28C	S57C
1UBQ	Vijay-Kumar et al. 1987	1	<u>0.33(3)</u>	<u>0.33(2)</u>	<u>0.071(1)</u>	0.058(1)
1D3Z	Cornilescu et al. 1998	10	0.27(3)	0.27(2)	0.064(2)	<u>0.062(2)</u>
2K39	Lange et al. 2008	116	0.27(3)	0.23(3)	0.067(1)	0.062(1)
2K0X	Fenwick et al. 2011	640	0.27(3)	0.18(2)	0.061(2)	0.057(1)
2MJB	Maltsev et al. 2014	20	0.26(3)	0.24(3)	0.068(1)	0.061(1)
<x-ray> ^d	Maltsev et al. 2014	15	0.29(3)	0.27(2)	0.068(1)	0.054(2)

^a For residues 1-76. Table S1 shows the corresponding data for residues 1-70 and Table S2 for residues selected by rmsd. The tensors were fitted using the RDC and PCS data for the ubiquitin mutants A28C and S57C tagged with the C1-Tb³⁺ tag. Values in brackets indicate the uncertainty in the last digit.

^b Quality factors of the alignment tensors fitted to ¹D_{NH} RDCs. Underlined and bold values identify, respectively, the worst and best fits for each mutant.

^c Quality factors of the $\Delta\chi$ tensors fitted to backbone PCSs. The worst and best fits are identified as in the previous columns.

^d The coordinates of <x-ray> were calculated as the average coordinates between different crystal structures using the weighting factors given in Table S9 by Maltsev et al. (2014). Not all crystal structures report coordinates for the C-terminal residues of ubiquitin. These missing coordinates were given a weighting of zero without changing the weighting of the other structures.

Table 2 $\Delta\chi$ -tensor parameters of ubiquitin mutants A28C and S57C tagged with C1-Tb³⁺ fitted to the structure 2KOX ^a

		$\Delta\chi_{ax}/$	$\Delta\chi_{rh}/$	Metal Coordinates			Euler Angles		
		10^{-32} m^3	10^{-32} m^3	$x/\text{\AA}$	$y/\text{\AA}$	$z/\text{\AA}$	$\alpha/^\circ$	$\beta/^\circ$	$\gamma/^\circ$
A28C	PCS	-8.11(8)	-2.25(5)	-2.76(4)	1.94(3)	-15.0(5)	19.5(2)	56.8(2)	6.5(2)
	RDC	-4.8(1)	-1.4(2)				14(2)	53(1)	83(4)
S57C	PCS	-9.06(3)	-2.5(1)	17.22(4)	8.37(4)	-4(3)	11.5(4)	105(1)	151(3)
	RDC	-6.0(2)	-2.5(2)				10.2(6)	98.2(8)	40(2)

^a Fitted alignment tensors were translated into $\Delta\chi$ tensors using Eq. 3. Values in brackets indicate the uncertainties in the last digit. The uncertainties were determined by adding random noise to the experimental PCS (or RDC) values. The noise amplitude followed a Gaussian distribution with a standard deviation determined by Eq. 2 for each individual peak. $\Delta\chi$ tensors were fitted in 100 trials after the addition of random Gaussian noise to obtain the uncertainties reported for the tensor parameters. Average peak uncertainties were 0.3, 4, 10, 0.2 and 2 ppb for the chemical shifts of ¹H^N, ¹⁵N, ¹³C ^{α} , ¹H ^{α} and ¹³C ^{β} nuclei, respectively. x , y and z are the coordinates of the paramagnetic centre in the coordinate frame of the 2KOX ensemble. The Euler angles describe the orientation of the tensor relative to the protein coordinates.

Table 3 Average pairwise RMSD values of backbone atom coordinates determined for four polypeptide fragments of ubiquitin from the experimental backbone PCSs of the ubiquitin mutants A28C and S57C ^a

Residues	Average pairwise RMSD		RMSD to the 2K0X structure	
	with outliers	without outliers	with outliers	without outliers
1-19	1.5	1.2	1.5	1.2
38-44	1.5	1.3	1.9	1.7
47-53	1.8	1.8	1.6	1.6
62-76 ^b	0.5	0.5	0.9	0.9

^a The RMSD values are reported in Å. RMSD values were calculated for the individual fragments without superimposition for minimum RMSD. The $\Delta\chi$ tensors of Table 2 were used for the structure calculations as described in the main text.

^b Residues 71-76 were excluded from the RMSD calculation.

Figure legends

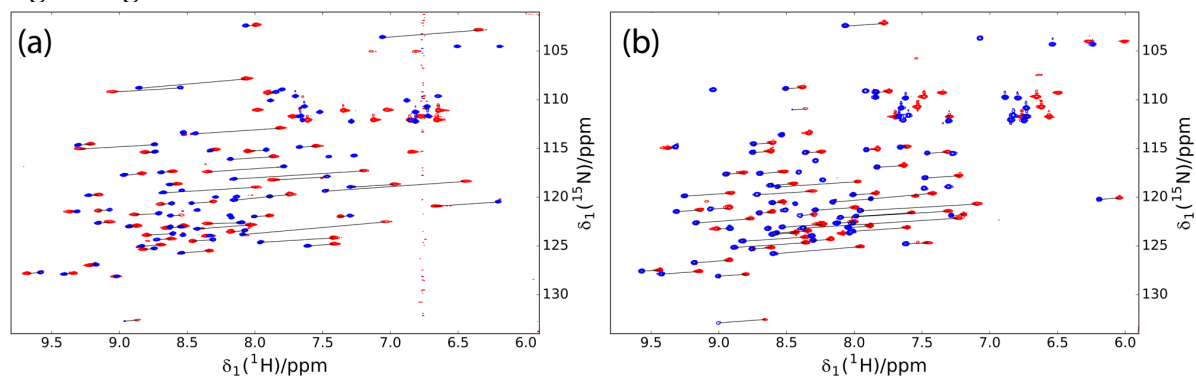


Fig. 1 Superimposition of ^{15}N -HSQC spectra of single-cysteine mutants of human ubiquitin tagged with C1- Y^{3+} (blue) or C1- Tb^{3+} (red) tags. PCSs of selected amides are identified by black lines connecting the corresponding cross-peaks. (a) Spectra of the A28C mutant. (b) Spectra of the S57C mutant

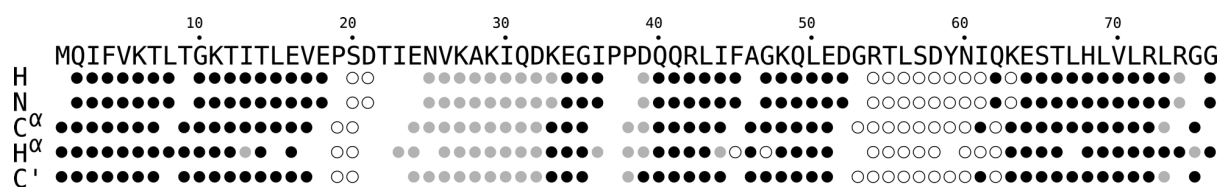


Fig. 2 Summary of the PCSs induced by Tb^{3+} in the backbone atoms of human ubiquitin. Open circles identify PCSs measured only for the A28C mutant, grey circles identify PCSs observed only for the S57C mutant and black circles represent the PCSs available for both mutants

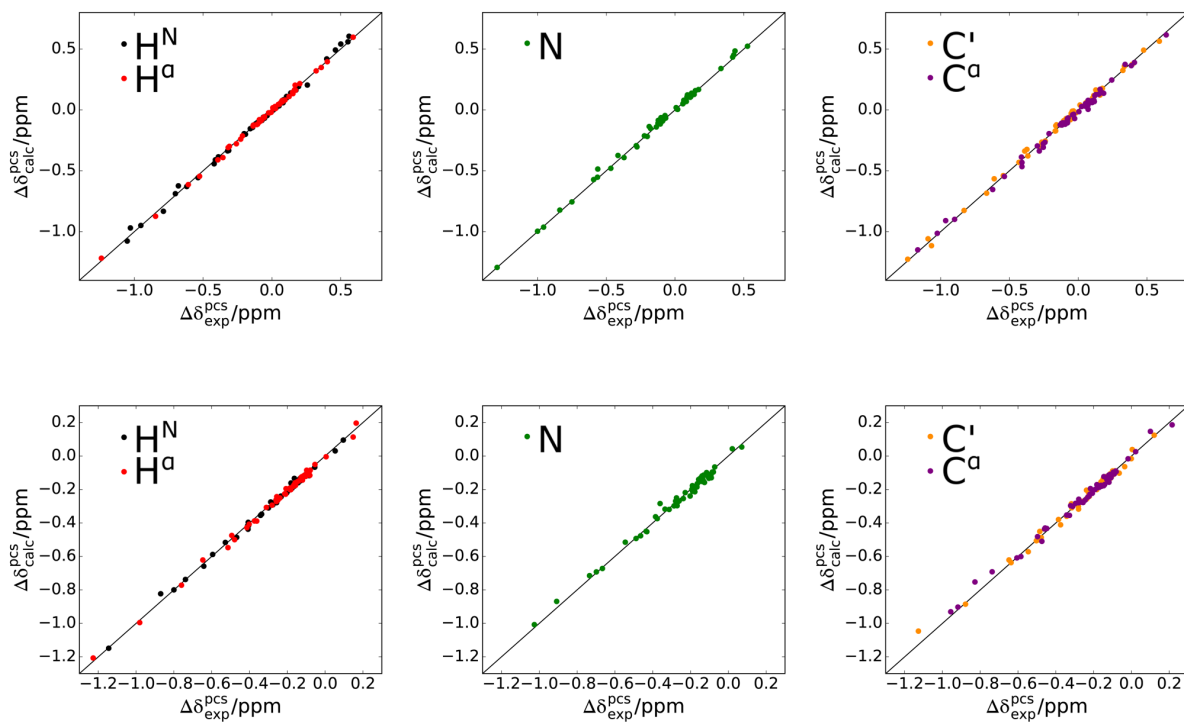


Fig. 3 Correlation plots of back-calculated versus experimental PCSs for backbone protons, nitrogens and carbons of the ubiquitin mutants A28C and S57C tagged with C1-Tb³⁺. The PCSs were back-calculated without taking RACS effects into account. Top panel: correlation plots for the mutant A28C. Bottom panel: correlation plots for the mutant S57C

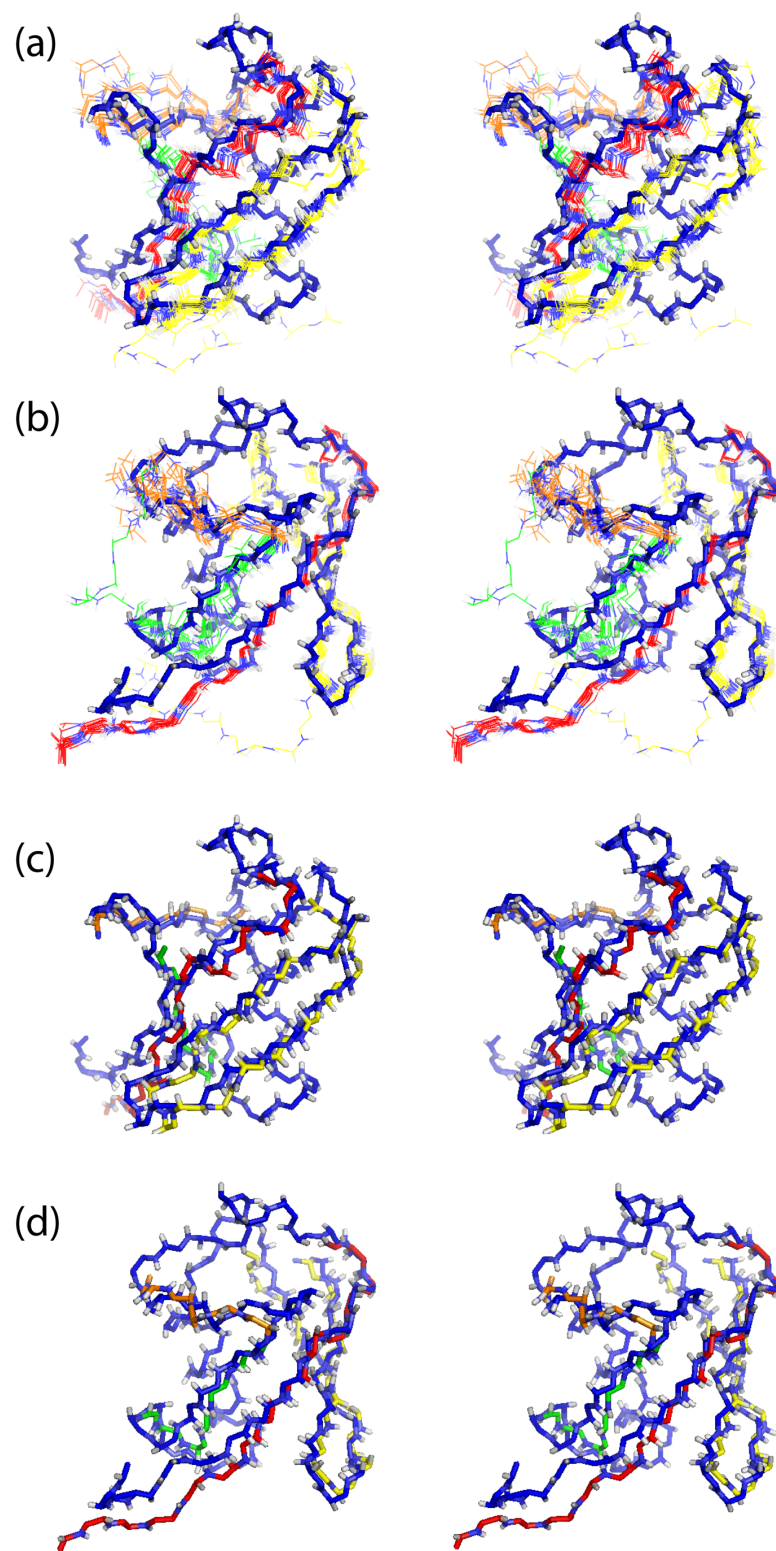


Fig. 4 Stereoviews of four fragments of the ubiquitin backbone structure determined by PCSs only, using the $\Delta\chi$ tensors of the A28C and S57C mutants with C1-Tb³⁺ tag. For comparison,

average coordinates of the 2K0X structure of ubiquitin are shown in blue. The back-calculated fragments determined from PCSs are shown in yellow (residues 1-19), green (residues 38-44), orange (residues 47-53) and red (residues 62-76). (a) and (b) Fragments are shown as ensembles of 20 structures calculated after adding Gaussian noise to the PCS data. The structures are shown from two different viewing angles. (c) and (d) Same data as in (a) and (b), except that the coordinates of the 20 fragment conformations were averaged for improved visual presentation, excluding obvious outliers among the yellow, green and orange segments

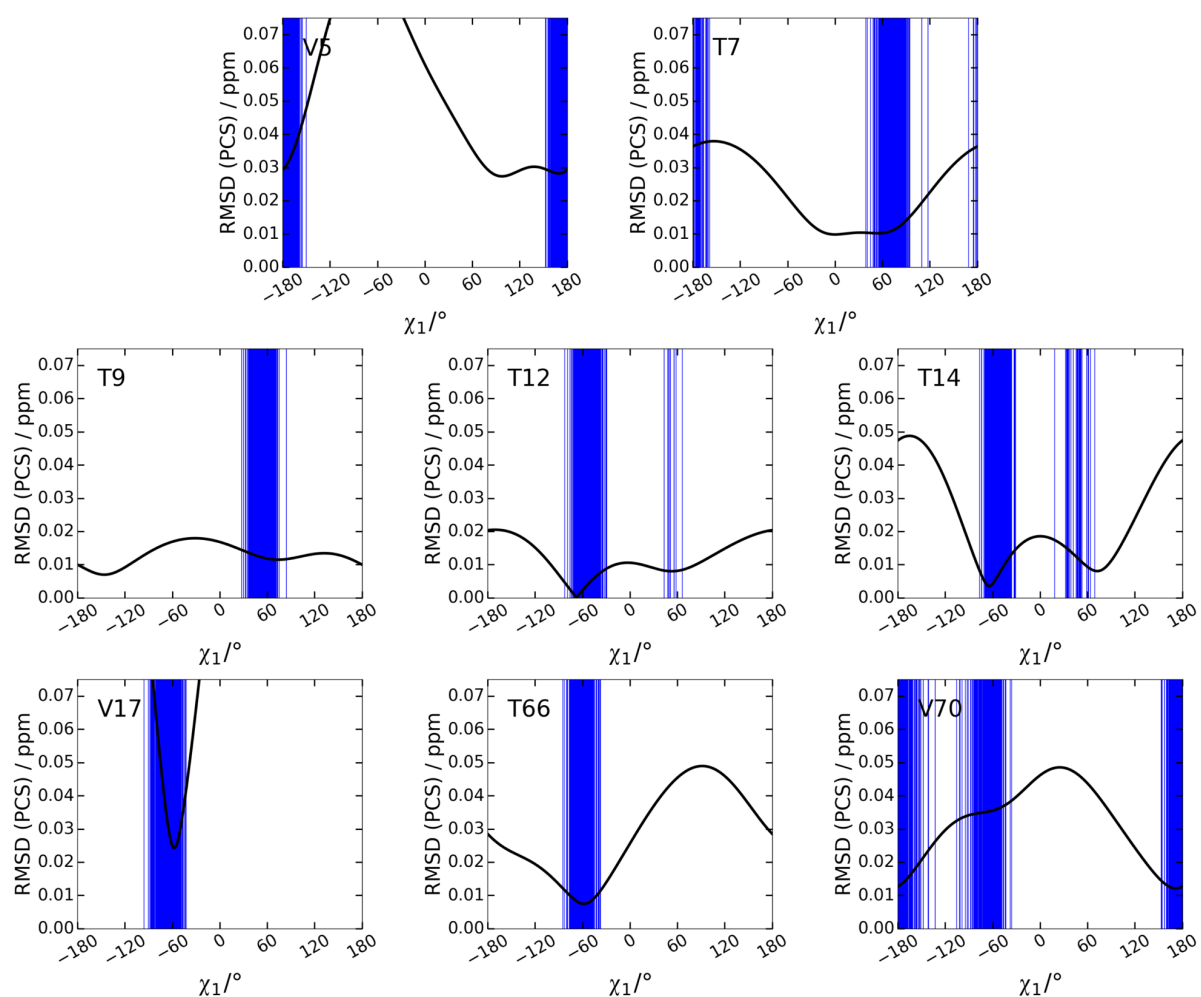


Fig. 5 RMSD between experimental and back-calculated PCSs of methyl groups as a function of χ_1 angle of threonine and valine residues. PCSs comprised the experimentally accessible ^1H and ^{13}C PCSs of the methyl groups. Blue lines identify the χ_1 angles reported by the 2KOX ensemble

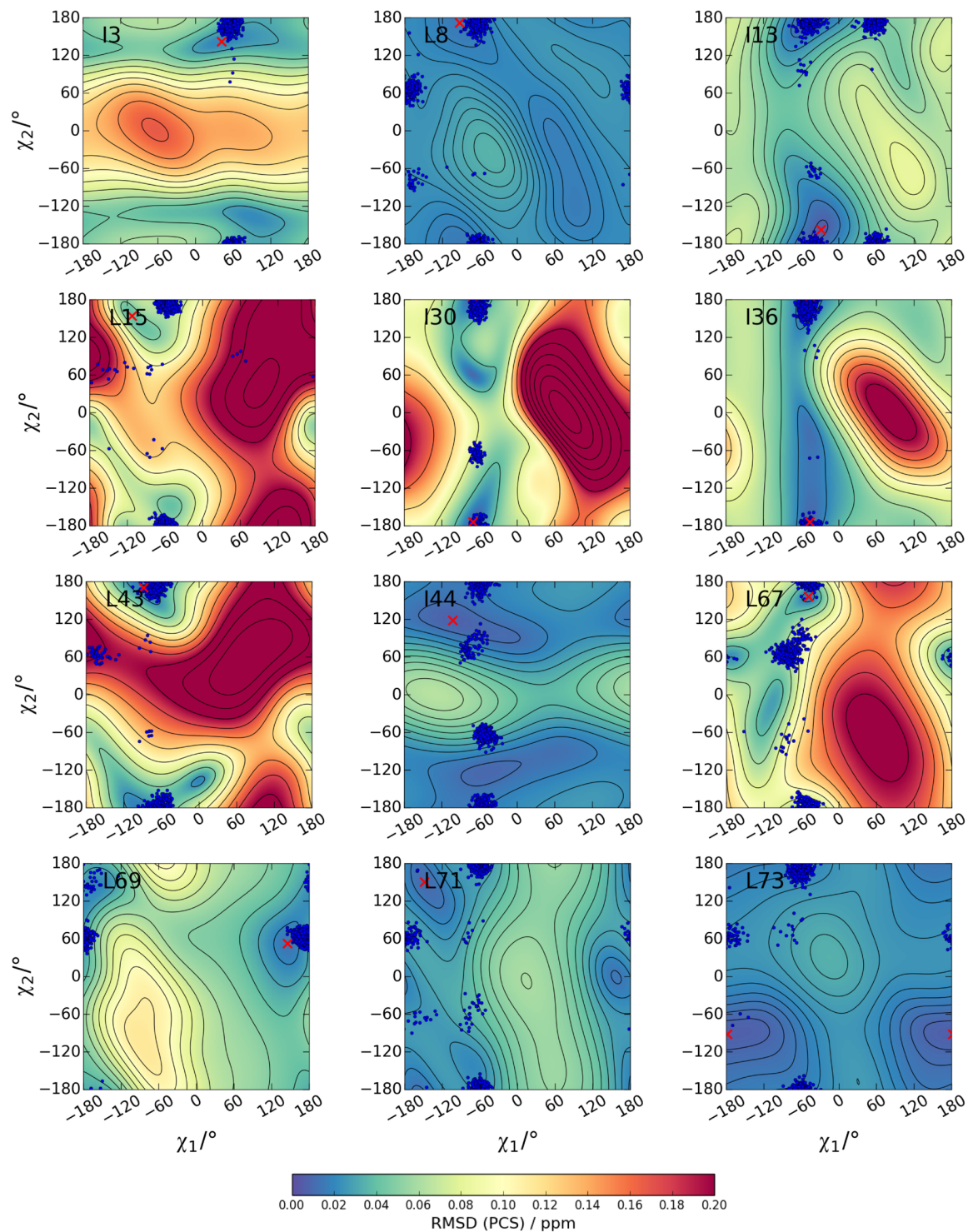


Fig. 6 RMSD between experimental and back-calculated PCSs of methyl groups as a

function of χ_1 and χ_2 angles of leucine and isoleucine residues. PCSs comprised the experimentally accessible ^1H and ^{13}C PCSs of the methyl groups. Red crosses denote the absolute minimum of the PCS RMSD, whereas blue dots indicate the χ_1 and χ_2 angle combinations reported in the 2K0X ensemble

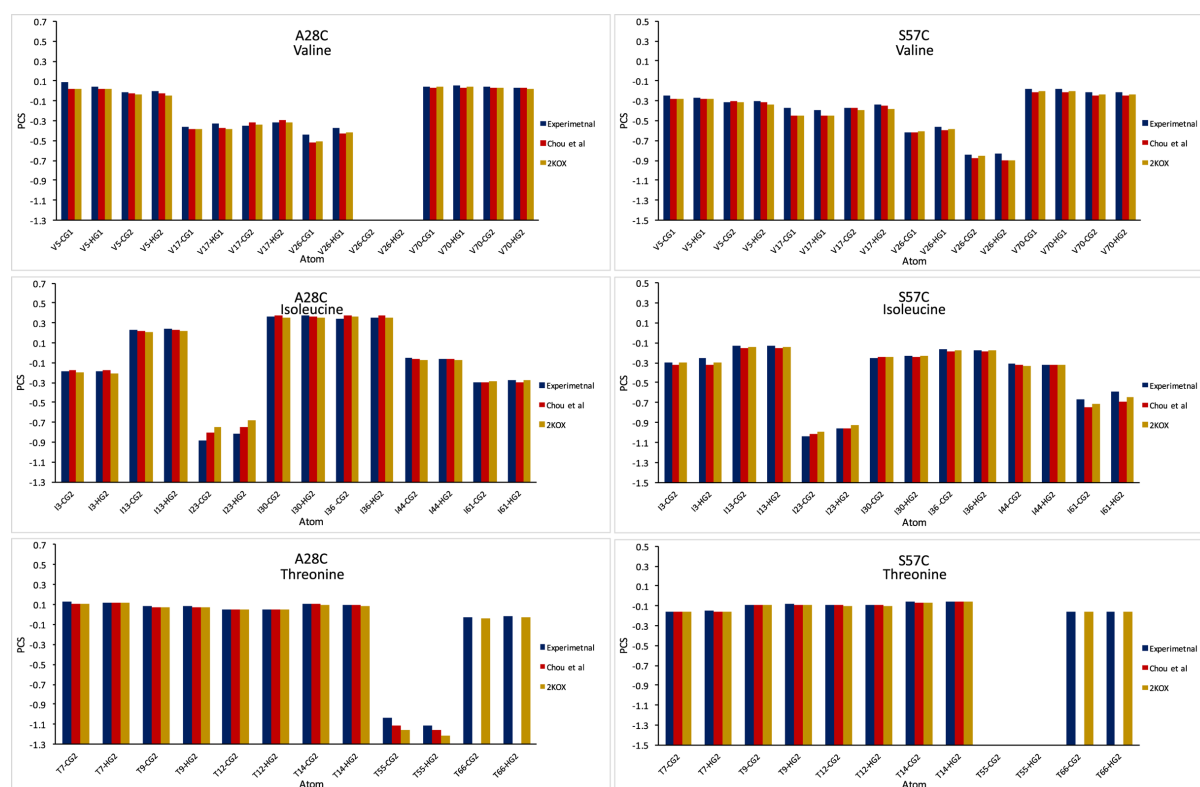


Fig. 7 Comparison of experimental and back-calculated PCSs of the side-chain γ -methyl groups of valine, isoleucine and threonine residues in ubiquitin. The experimental PCSs (blue bars) are from Table S4. $\Delta\chi$ tensors were fitted to the backbone of the 2K0X ensemble and used to calculate the PCSs expected for the respective side-chains in this ensemble (brown bars). Alternatively, the side-chains were crafted onto the average backbone structure of the 2K0X ensemble with the χ_1 dihedral angles and populations published by Chou et al. (2004) and the resulting conformations used to back-calculate the PCSs (red bars). Where no values are reported, no PCSs could be measured

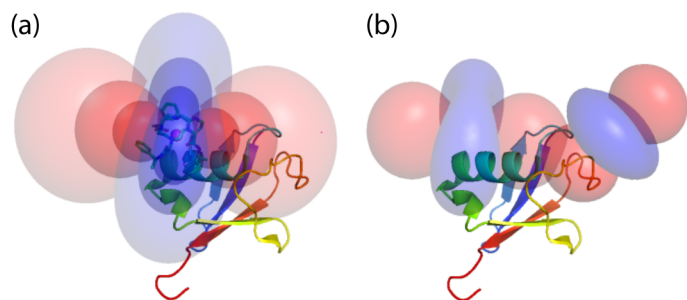


Fig. 8 PCS isosurfaces in ubiquitin tagged with a paramagnetic metal ion. Blue and red surfaces correspond to positive and negative PCS values respectively. The protein is shown in a ribbon representation. (a) Isosurfaces are shown for PCSs of 0.25, 1 and 4 ppm. Fainter isosurfaces correspond to smaller PCSs. (b) Two intersecting PCS isosurfaces from two different tagging sites. The isosurface on the left corresponds to a PCS of 1 ppm, the one on the right to a PCS of 2 ppm. A nuclear spin with PCSs corresponding to the isosurfaces shown must be positioned at the intersection between the two isosurfaces