NMR studies of protein-small molecule and protein-peptide interactions

Guan, J.

4 Structure determination of a protein-ligand complex by paramagnetic NMR spectroscopy

Abstract

Determining the three dimensional structure of a small molecule-protein complex with weak affinity can be a significant challenge. We present a paramagnetic NMR method to determine intermolecular structure restraints based on pseudocontact shifts (PCSs). Since the ligand must be in fast exchange between free and bound state and the fraction bound can be as low as a few percent, the method is ideal for ligands with high micromolar to millimolar dissociation constants. Paramagnetic tags are attached, one at a time, in a well-defined way *via* two arms at several sites on the protein surface. The ligand PCSs were measured from simple 1D ^{1}H spectra and used as docking restraints. An independent confirmation of the complex structure was carried out using intermolecular NOEs (Chapter 3). The results show that structures derived from these two approaches are similar. The best results are obtained if the magnetic susceptibility tensors of the tags are known, but it is demonstrated that with two-armed probes, magnetic susceptibility tensor can be predicted with sufficient accuracy to provide a low-resolution model of the ligand orientation and the location of the binding site in the absence of isotope labeled protein. This approach can facilitate fragment-based drug discovery in obtaining structural information on the initial fragment hits.

This work has been published as part of

J.-Y. Guan, P. H. J. Keizers, W.-M. Liu, F. Löhr, E. Heeneman, S. P. Skinner, H. Schwalbe, M. Ubbink and G. Siegal. **Small molecule binding sites on proteins established by paramagnetic NMR spectroscopy.** *J. Am. Chem. Soc.*, **2013**, *135* (15), pp 5859–5868

4.1 Introduction

The availability of three dimensional (3D) structures of protein-ligand complexes significantly improves the efficiency of refining hits towards leads in the early stages of drug discovery.[174] Typically, structure-driven hit optimization programs rely on crystallographic data. In fragment-based drug discovery (FBDD), however, weakly binding ligands (fragments) are often not observed in crystals, for a variety of reasons. NMR spectroscopy is a powerful alternative for deriving structural information, particular for weakly interacting complexes. The classical NMR approach is based on the observation of intermolecular NOEs and is suitable for most proteins smaller than 40 kDa. Although robust, the method can be time consuming and requires uniformly isotopically labeled protein. The method is strictly limited to proteins that can be functionally expressed under these conditions. Selective isotope labeling schemes have been employed in combination with deuteration to enable NMR analysis of much larger proteins. Complexation-induced chemical shift perturbation (CSP) data have been used as ambiguous interaction constraints to calculate structures of protein–ligand complexes.[175] However, this method monitors both direct and remote effects, and therefore the .binding site is not always well defined. We sought a method that could in principle be applied to proteins where no, or only limited, isotope labeling can be performed.

Paramagnetic NMR is known to be a powerful tool to study biological systems due to its versatile effects, including pseudocontact shifts (PCSs), paramagnetic relaxation enhancement (PRE) and residual dipolar coupling (RDC).[89–91] Paramagnetic NMR has been applied extensively to characterize protein-protein interactions, but very little in protein-small molecule interactions. Pioneer studies used the PRE caused by a paramagnetic metal ion[95] or spin label[92,93] to facilitate ligand screening and potentially, to obtain information on the ligand pose.[95]

A combination of PCS and RDC was applied to assist structure determination of a carbohydrate-protein complex where the paramagnetic center was introduced by creating a fusion protein with a C-terminal lanthanide-binding peptide tag (LBT).[96] PRE assisted ligand docking with a spin-labeled peptide bound specifically to a protein was also reported.[98] Recently, paramagnetic effects stemming from a two-point anchored N-terminal LBT were used to determine the structure of protein-peptide complexes.[99] PCSs have also been applied to determine the structure of a small molecule ligand in rapid exchange with a protein in which a lanthanide was bound in a natural metal binding site.[97] In this study the ligand bound very closely to the lanthanide, enabling the use of large ligand-to-protein ratios. This

latter example is the only case where PCSs have been used to elucidate the structure of a protein-ligand complexes containing a small fragment (M_w <300 Da).

We demonstrate here an alternative way to determine the location and orientation of a weakly bound fragment with PCS restraints from a ligand in rapid exchange with a protein. A rigid, double-armed lanthanide-binding tag, CLaNP-5,[144,176] was attached at three different sites, one at a time, on the protein surface *via* disulfide bond linkage. Using this tag, the magnitude and orientation of the $\Delta\chi$-tensors can be predicted. Further, the paramagnetic effects can be tuned by using different lanthanides in the tag. We selected a 12 kDa FK-506 binding protein (FKBP12) as a model protein to investigate the potential of the methodology. FKBP12 is a peptidyl-prolyl isomerase which belongs to the family of immunophilins and is a drug target for the immuno-suppressants rapamycin and FK506. There have been many structural studies on this protein.[9,10,15,20,162,163] The ligand in this study is a fragment that was identified as a hit against FKBP12 from a screen of a fragment library using Target-Immobilized NMR Screening (TINS).[84] In this work, we compare the PCS-based docking result with the structure determined by intermolecular NOEs (Chapter 3). The structure calculations were performed using parameters from both predicted and experimentally determined paramagnetic $\Delta\chi$-tensors. The result shows that, even without resonance assignments of the protein, it is possible to determine the ligand binding site and approximate orientation. The method can assist lead optimization in fragment-based drug discovery when high resolution structural information is not available.

4.2 Materials and Methods

Ligand preparation

Assignments and preparation of the ligand **1** (Figure 3.1A) was carried out as described in Chapter 3.

Protein expression and purification

Uniformly ^{15}N-labeled FKBP12 double cysteine mutants (K34C/K35C, K44C/K47C, C22V/E61C/Q65C), all of which contain an additional LEHHHHHH tag at the C-terminus were purified from *Escherichia coli* strain BL21(DE3) containing the overexpression plasmid pET20b with FKBP12 insert. The expression and purification are described in Chapters 2 and 3. The yield was in general 15-23 mgL^{-1} for the double cysteine mutants.

CLaNP-5 attachment

The details of CLaNP-5 attachment to FKBP12 variants are described in Chapter 2.

NMR measurements

All protein NMR samples contained 15 mM Tris-HCl, 25 mM NaCl, pH 7.7, 6% D_2O for ^{15}N-labeled protein and >95% D_2O for non-isotope-labeled protein. The concentration of Ln^{3+}-CLaNP-5 attached FKBP12 was 17 to 39 μM. The ligand to protein ratios were 1.3:1 for all Ln^{3+}-CLaNP-5 attached FKBP12 mutants. 1D-^{1}H and [^{1}H,^{15}N]-HSQC spectra were recorded at 290 K on a Bruker Avance DMX-600 spectrometer equipped with a TCI-Z-GRAD cryoprobe. 1D-^{1}H spectra of the complex CLaNP5-FKBP12 and ligand **1** were recorded with a spectral width of 16 ppm and 6 k complex points, resulting in a digital resolution of 1.56 Hz before zero filling. Carr–Purcell–Meiboom–Gill (CPMG) pulse sequence was used with a total echo time of 60 ms comprising 60 pulses, for suppression of macromolecules resonances. Data were processed in TopSpin (Bruker) and then spectra were analyzed in Sparky.[164]

Magnetic susceptibility tensor optimization and PCS-based structure calculations

PCSs are defined as the difference in ppm between the corresponding resonance in the paramagnetic sample and the diamagnetic sample. The PCS gives information on the distance and angle between a nucleus and the paramagnetic center according to Equation 1.3.[113] Errors were calculated by randomly excluding 10% of the data with Monte Carlo analysis implemented in Numbat.[177]

The fits of observed versus back-calculated protein PCSs and docking were performed in the XPLOR-NIH[178] program containing the PARArestraints module.[179] The predictions of the initial $\Delta\chi$-tensor positions and orientations were carried out as previously described.[144] The structure model, parameter and topology files of the ligand were generated from PRODRG server.[153] Bound ligand PCS values were used as the experimental restraints for PCS docking. PCSs of ligand methyl protons were used for methyl carbon positions instead of methyl proton positions. The protein backbone atoms and the pseudo-residues defining the metal coordinates were fixed in the docking process. The three datasets of the different tag positions on FKBP12 were used simultaneously. Ligand PCSs from all three tagging sites were used in the docking procedure with equal weighting. Each docking calculation comprised 100 steps, started with restrained rigid body docking for 20 steps (0.01 pico-seconds increments, 300 increment evaluations/step). The lowest energy structure was

subsequently subjected to restrained Langevin dynamics for 80 steps (0.001 pico-seconds increments, 2000 increment evaluations/step), which allowed the ligand and the residues within 8 Å from the ligand to be flexible. 200 independent docking calculations were performed using the solution structure of FKBP12 (PDB entry 1FKR, model 14)[9] and the predicted Δχ-tensor parameters and position of the lanthanide ion. The same calculations with experimentally determined Δχ-tensor parameters and positions were performed for comparison. The agreement between the experimental PCSs and back-calculated PCSs was evaluated using the Q factor defined in Equation 4.1 :[180]

$$Q = \sqrt{\frac{\sum_i \left(PCS_i^{\exp} - PCS_i^{calc}\right)^2}{\sum_i \left(\left|PCS_i^{\exp}\right| + \left|PCS_i^{calc}\right|\right)^2}} \qquad (4.1)$$

where $PCS_i^{\exp}$ and PCS_i^{calc} are the observed and calculated PCSs. The XPLOR-NIH script is available in Appendix B.

PCSdock prediction

The PDB structure containing all three Δχ-tensors (calculated or predicted) was imported into PCSdock using Scientific Python.[181] Ligand PCS data were imported, along with the axial and rhombic magnitudes of the Δχ-tensors. A cubic grid of user-defined size (30 Å was used) and 1 Å spacing was placed around the protein using the centre of mass as its origin. Predicted PCS values for the points on the grid were calculated for each paramagnetic Δχ-tensor. The experimental ligand PCS values for an atom *j* were then compared to the predicted PCS values for each grid point *i* using a Q score defined by Equation 4.2:

$$Q_{i,j} = \frac{1}{N}\sum_{k=1}^{N}\sqrt{\frac{(PCS_{i,k}^{pred} - PCS_{j,k}^{exp})^2}{(\left|PCS_{i,k}^{pred}\right| + \left|PCS_{,j,k}^{exp}\right|)^2}} \qquad (4.2)$$

where N is the number of paramagnetic Δχ-tensors, $PCS_{i,k}^{pred}$ is the PCS calculated for the grid point i relative to the paramagnetic Δχ-tensors k, $PCS_{j,k}^{exp}$ is the experimental PCS used for the atom j relative to the paramagnetic Δχ-tensors k. If the $Q_{i,j}$ for any atom of the ligand was lower than a user-defined level, the grid point i was accepted. If the $Q_{i,j}$ was larger for all atoms of the ligand, that grid point was discarded. The PCSdock script is kindly provided by Simon P. Skinner (Leiden University).

4.3 Results

Selection of CLaNP-5 tagging sites

In order to determine the protein-ligand complex, we also utilized paramagnetic NMR and generated three FKBP12 mutant proteins tagged with the synthetic lanthanide tag CLaNP-5. The lanthanide tag, CLaNP-5, is designed to covalently link to the protein surface *via* two disulfide bridges.[144,176] Therefore, the presence of surface accessible cysteine pairs is essential. Several criteria need to be considered for selecting the mutation sites: (1) The cysteines should be far enough from the putative ligand binding site to avoid interference with binding, yet close enough to yield appreciable paramagnetic effects; 15-30 Å is a reasonable estimation based on the location of the binding site and the total size of FKBP12. 25 Å is an estimation of the effective range for Yb^{3+}-chelated CLaNP-5. In this study, the distances between the putative ligand binding site and the three lanthanide tags are in the range of 15-25 Å. It is possible to use other lanthanides to adjust the effective range for different protein sizes.[117] (2) The two Cα atoms from the cysteines should be 6-10 Å apart, with their side chains pointing away from the protein surface and roughly in the same direction. Cysteines buried inside the protein cannot react with CLaNP-5. The locations of cysteine mutation sites (K34C/K35C, K44C/K47C and E61C/Q65C) were selected to satisfy the above requirements. As a consequence of the anisotropy of the magnetic susceptibility, the ligand PCSs can be close to zero regardless of the distance if the ligand is located close to the region in which the PCS changes its sign. Nevertheless, measuring PCSs in this situation for a different tag position will provide information. This is one of the advantages of synthetic tags compared with the methods using lanthanide-binding peptides[182,183] or metal displacement, in which usually only one tag position is available.

Ligand PCS measurement

The resonance assignments of the free ligand **1** were obtained by analysis of 1D-^{1}H, ^{13}C-APT, [^{1}H,^{13}C]-HSQC and [^{1}H,^{13}C]-HMBC spectra (Chapter 3). The binding affinity of the ligand **1** to FKBP12 was determined using [^{1}H,^{15}N]-HSQC titrations, and was found to be 1.80 (±0.05) mM. (Chapter 3). To eliminate interference from the protein resonances, 1D-^{1}H NMR spectra of the ligand in the presence of Ln^{3+}-CLaNP-5-FKBP12 were recorded using a T_2 relaxation delay of 60 ms. Ligand PCSs were measured from singly tagged protein with Yb^{3+}-CLaNP-5 in the paramagnetic sample and Lu^{3+}-CLaNP-5 in the diamagnetic sample. In total, 3 different pairs of protein samples, each containing the CLaNP-5 attached at a

different site, were used. Initially **1** was titrated into Yb^{3+}-CLaNP-5-FKBP12 to determine the optimal molar ratio of protein to ligand. The ligand and protein concentrations were 36 and 28 μM (34C/35C), 24 and 17 μM (44C/47C), and 47 and 39 μM (61C/65C). Under these conditions 0.9-1.9 % of the ligand is bound. A protein to ligand ratio 1:1.3 was used for all measurements with both Yb^{3+}-CLaNP-5 and Lu^{3+}-CLaNP-5, because this ratio represented the best compromise between size of the PCSs, the line broadening and spectral resolution. Figure 4.1 shows the spectra from which the ligand PCSs were determined for the 34C/35C tagging site.

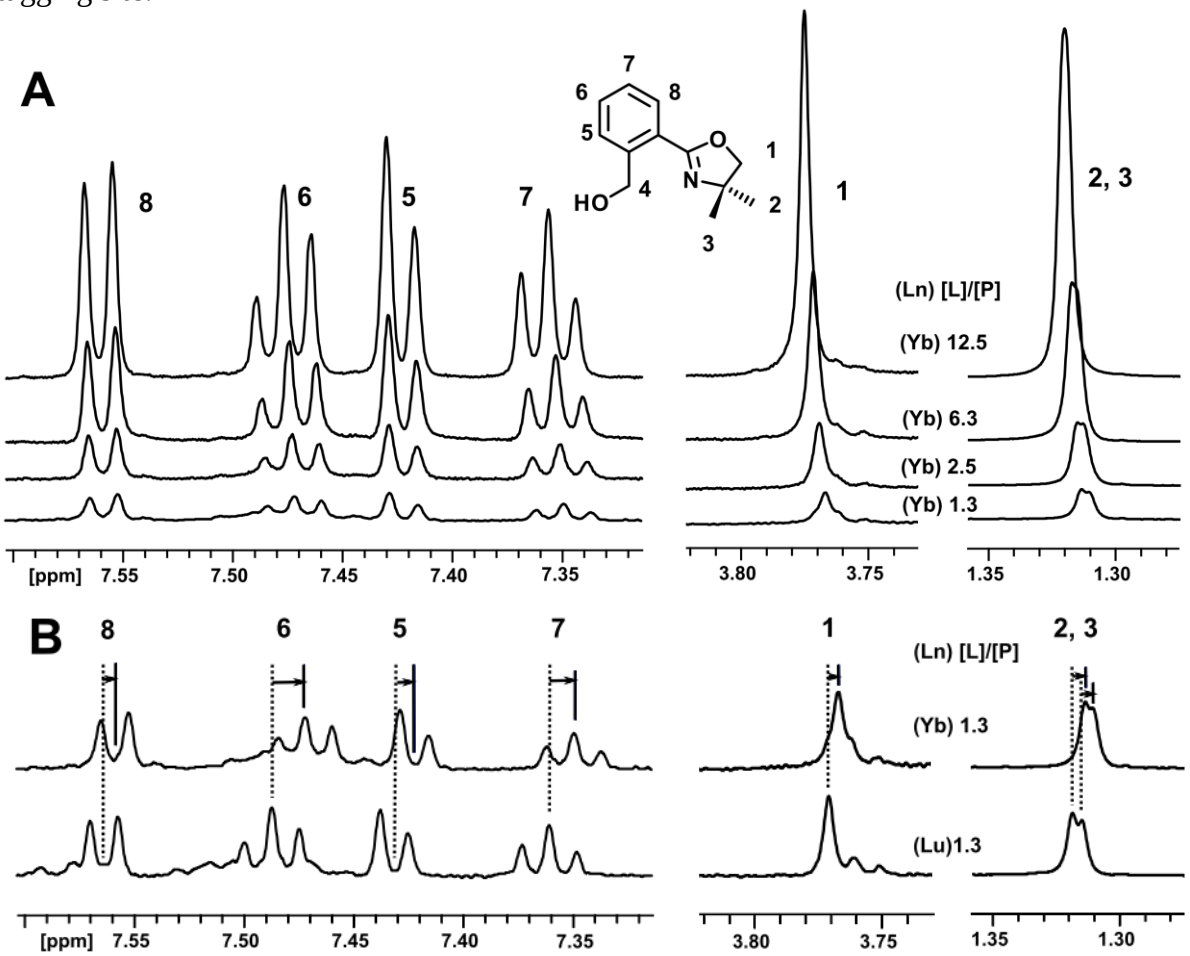

Figure 4.1: (A) Overlay of 1D-^{1}H NMR spectra of ligand **1** in the presence of 28 μM FKBP12 (34C/35C) attached to paramagnetic Yb^{3+}-CLaNP-5 with increasing ligand/protein molar ratios, which are indicated in the spectra. Proton assignments of **1** are indicated by corresponding numbers on the structure. The chemical shift of proton 4 overlapped with the water resonance and is not shown here. The resonances of the methyl groups numbered 2 and 3 are degenerate in the free form but are resolvable in the bound form. (B) The observed ligand PCS is the difference between the resonance positions for the paramagnetic (Yb^{3+}) and diamagnetic (Lu^{3+}) samples. The dashed and solid lines indicate the positions of the diamagnetic and paramagnetic ligand resonances, respectively. Spectra were recorded at 600 MHz.

The observed ligand PCS values are small (less than 10 Hz) due to the small fraction of bound ligand (1.5%). However, they can be measured precisely because of the sharp signals of the ligand. The PCSs are summarized in Figure 4.2. A control measurement of the ligand in the presence of the free probe in the absence of protein showed no paramagnetic effects (data not shown). This indicates that the observed ligand PCSs derive exclusively from the bound state of the ligand. A total of 21 PCS values were obtained from the three pairs of para- and diamagnetic spectra.

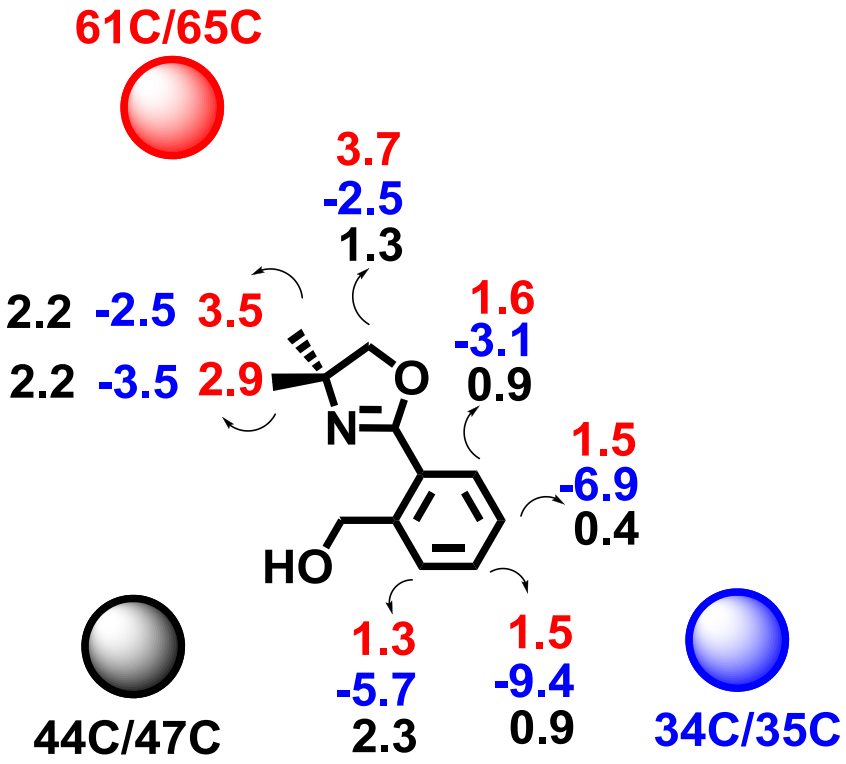

Figure 4.2: Observed ^{1}H PCS values (in Hz, for spectra acquired at 600 MHz) of ligand **1** from three different locations of the paramagnetic center. The relative positions of the paramagnetic centers are indicated and the associated PCS values are color coded. Blue, PCSs from 34C/35C; black, PCSs from 44C/47C; red, PCSs from 61C/65C.

PCS-based structure calculation with predicted Δχ-tensors

For ligands in fast-exchange on the NMR time-scale (i.e. $\Delta\Omega < k_{off}$), the observed PCSs are weighted averages of the free and bound states. Therefore, the PCS values for the bound state can be derived if the fraction of ligand bound is known. Using the known concentration of ligand and protein, as well as the experimentally determined K_D, the observed PCSs were converted to the PCSs in the bound state (Table 4.1).

Table 4.1: Ligand pseudocontact shifts converted to 100% bound (ppm) according to calculated bound fractions (indicated at bottom of the table). The observed values are given in Figure 4.2. Uncertainties (parentheses) are estimated as ± 2× standard deviation of the positions of the para- and diamagnetic resonances based on the following formulae: $\sigma = \sqrt{\sigma_{para}^2 + \sigma_{dia}^2}$

Spin	K34C/K35C	K44C/K47C	C22V/E61C/Q65C
H-1	-0.28 (0.06)	0.23 (0.08)	0.32 (0.04)
H-2	-0.29 (0.02)	0.40 (0.02)	0.31 (0.06)
H-3	-0.33 (0.02)	0.40 (0.02)	0.26 (0.04)
H-5	-0.63 (0.12)	0.42 (0.18)	0.11 (0.10)
H-6	-1.04 (0.38)	0.16 (0.34)	0.13 (0.18)
H-7	-0.77 (0.24)	0.07 (0.32)	0.14 (0.18)
H-8	-0.34 (0.12)	0.17 (0.13)	0.14 (0.01)
[Protein] (μM)	28	17	39
[Ligand] (μM)	36	24	47
Bound fraction (%)	1.5	0.9	1.9

Previous work in which CLaNP-5 was bound to a rigid protein suggested that lanthanide position and the Δχ-tensor orientation can be predicted using a simple set of rules.[144] This characteristic is important for proteins that cannot be isotopically labeled, so we first determined the structure of the complex of **1** bound to FKBP12 using the predicted Δχ-tensor orientations. As with the NOE-based structure determination, several structures of FKBP12 from the PDB were used as input. PDB file 1FKR model 14 is shown as an example. In the docking procedure, the ligand was allowed to move, and the protein was fixed, except for side chains within 8 Å of the ligand, which were allowed to rotate. The complex formed by FKBP12 and ligand **1** was energy minimized guided by the energy terms for the pseudocontact shifts and Lennard-Jones potential. The five lowest energy structures were selected, based on the total energy and overlaid with the structure derived from NOE restraints (Figure 4.3). Two orientations are found with low PCS energy, which differ by a rotation of ~90 degrees (Figure 4.3, A and B), and the lowest energy cluster (Figure 4.3A) is closest to the orientation observed in the NOE structure. The average RMSD relative to the NOE structure is 4.7 ± 0.9 Å. In the other orientation, which has on average 10% higher PCS energy, the aromatic ring of **1** points outwards from the binding site (RMSD relative to NOE 4.4 ± 0.1 Å).

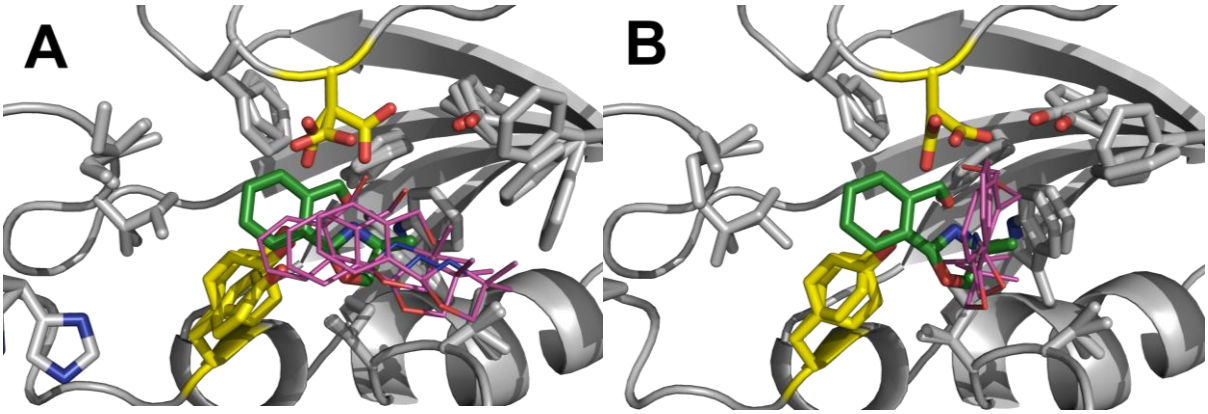

Figure 4.3: Best five PCS structures calculated using the predicted Δχ-tensors (ligands in magenta) superimposed on the averaged NOE structure (ligands in green). Two clusters with similar PCS energy are present with the ligand rotated by ~90 degrees. (A) The lowest energy cluster has an orientation parallel to the NOE structure. (B) The second cluster has the aromatic ring pointing out from the binding site.

By applying the predicted Δχ-tensor parameters for PCS-based structure calculations, the approximate location of the ligand binding site could be established. Both clusters of calculated orientations showed good agreement between the predicted and experimental ligand PCSs (Figure 4.4, A and B), as indicated by the quality factor (equation 4.1). This suggests that, using CLaNP-5 with predicted Δχ-tensor parameters, it is possible to detect the ligand binding site and obtain a low resolution structure for further ligand optimization. For this approach, prediction of the Δχ-tensor position and orientation does not require any PCSs or other NMR data on the protein, although the 3D structure of the protein is required.

Δχ-Tensor calculations

PCS values depend on the position of the nuclear spin relative to the magnetic susceptibility anisotropy tensor of the lanthanide ion, as described by equation 1.3. The Δχ-tensors are defined by eight parameters (the axial and rhombic components, the *x*, *y*, *z* coordinates of the metal, and the orientation of the Δχ-tensor relative to the molecular frame, defined as three Euler angles). Therefore, they can be determined from a minimum of eight PCSs measured from the nuclear spins with known resonance assignment. The PCS values of protein amide protons were measured as the difference in the chemical shift between the Yb^{3+}- and Lu^{3+}-CLaNP-5 attached FKBP12. Most of the [^{1}H,^{15}N]-HSQC assignments of the wild type FKBP12 could be transferred to the double cysteine mutants attached to the Lu^{3+}-CLaNP-5.

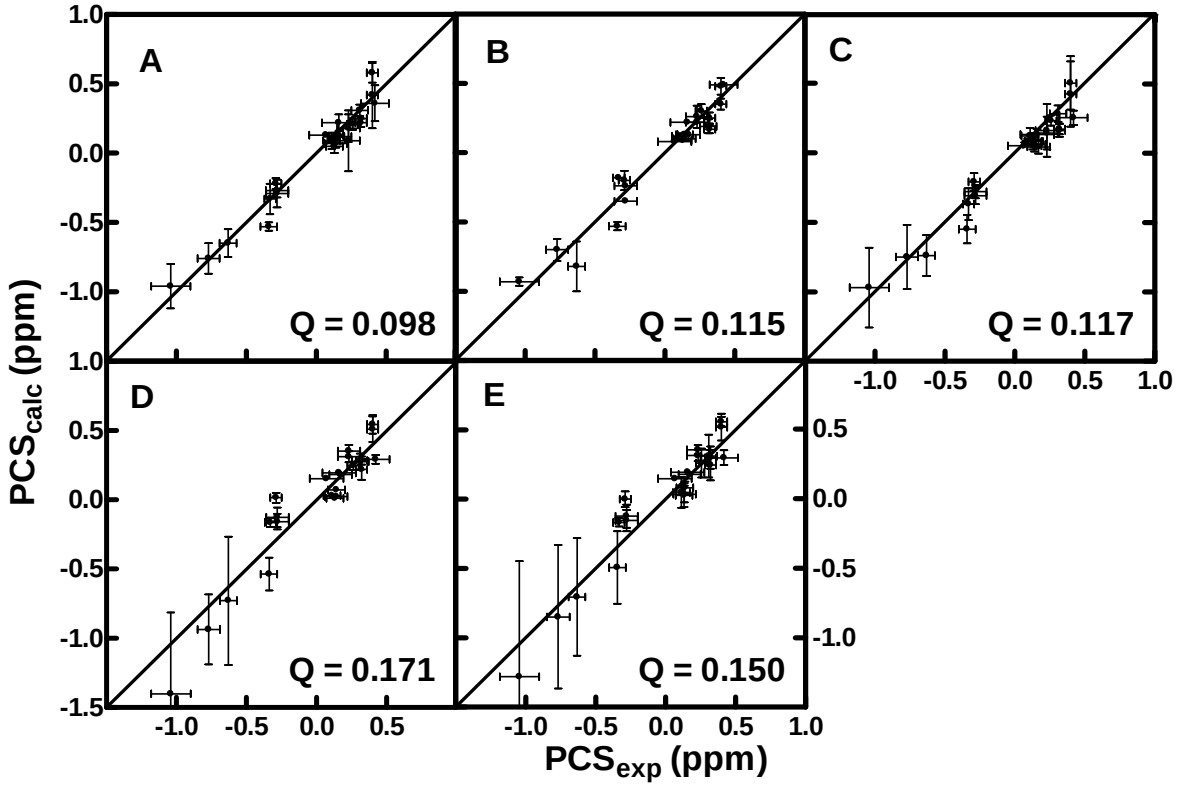

Figure 4.4: Correlation between experimental and back-calculated ligand PCSs for the top five structures of **1** bound to FKBP12 as determined from predicted and experimentally determined Δχ-tensor positions. (A) Correlation of the best structures obtained with predicted Δχ-tensors, lowest energy cluster, as shown in Figure 5A and (B) second lowest energy cluster, as shown in Figure 5B. (C) Best five structures with experimentally determined Δχ-tensors. (D) Back-calculated ligand PCS values for NOE-based structures using predicted Δχ-tensor positions. (E) Back-calculated ligand PCS values for NOE-based structures using experimentally determined Δχ-tensor positions. Vertical error bars represent 2× standard deviation of the variation in the sets of structures for the back-calculated PCSs. Horizontal error bars represent the estimated experimental error.

Although there are ambiguous assignments in the spectra of the mutants, more than 70% of the residues could readily be assigned and therefore the Δχ-tensor magnitude and orientation could be determined for each of the three mutants. The calculated parameters are summarized in Table 4.2. Overlays of the paramagnetic and diamagnetic [^{1}H,^{15}N]-HSQC spectra of the three mutants and the correlations between the experimental and back-calculated protein PCSs are shown in Figure 4.5. Compared with FKBP12 variants K34C/K35C and K44C/K47C, C22V/E61C/Q65C exhibits somewhat smaller tensor magnitudes (Table 4.2) and a slightly poorer fit (Figure 4.5), which is an indication that the tag at this position is a little more mobile than at the other two positions. Nevertheless, all

the values of the axial and rhombic components from three tagging sites are comparable to previously reported values,[144,184–188] and therefore the values for C22V/E61C/Q65C were considered reasonable.

Table 4.2: $\Delta\chi$-Tensor parameters of Yb^{3+}-CLaNP-5 attached to FKBP12.[a)]

Mutant	K34C/K35C	K44C/K47C	C22V/E61C/Q65C
$\Delta\chi_{ax}$ ($10^{-32}m^3$)	8.7 ± 0.4	8.9 ± 0.2	7.8 ± 0.1
$\Delta\chi_{rh}$ ($10^{-32}m^3$)	3.4 ± 0.4	3.3 ± 0.3	2.7 ± 0.2
Q (eqn. 4.1)	0.023	0.021	0.035
Restraints	75	73	72

[a)] Errors were estimated by randomly excluding 10% of the data with a Monte Carlo approach.[177]

PCS-based structure calculation with experimentally determined $\Delta\chi$-tensors

The docking of ligand **1** to FKBP12 was repeated using the experimentally determined tensor orientations. The orientations of the best five structures form a single cluster that is close to the NOE-derived structure (Figure 4.6). The correlation of experimental and back-calculated ligand PCSs is presented in Figure 4.4C. The average RMSD between the cluster members is 1.6 ± 0.5 Å and to the averaged NOE structure 2.8 ± 0.4 Å. With the experimental $\Delta\chi$-tensor parameters, the quality factor has slightly increased to 0.117. Given the error estimates shown in Figure 4.4, we conclude that the calculated structures fit the data and that the differences of the Q values between the clusters derived from the predicted and experimental $\Delta\chi$-tensors are not significant. The position of the ligand found with the optimized $\Delta\chi$-tensors is closer to the NOE-based position, suggesting that the optimization increases the accuracy of the solution. The PCSs were also back-calculated from the NOE-based structures, using both the predicted $\Delta\chi$-tensors (Figure 4.4D) and the experimentally determined $\Delta\chi$-tensors (Figure 4.4E). It is clear that the NOE-derived ligand position fits the PCS data worse than the calculated position, indicating that the two positions differ significantly from the experimental point of view. The NOE-based structure is used here as the standard for validation of the PCS based structure, but, as mentioned in Chapter 3, the NOE-based solution is also an approximation due to the dynamics in the binding site.

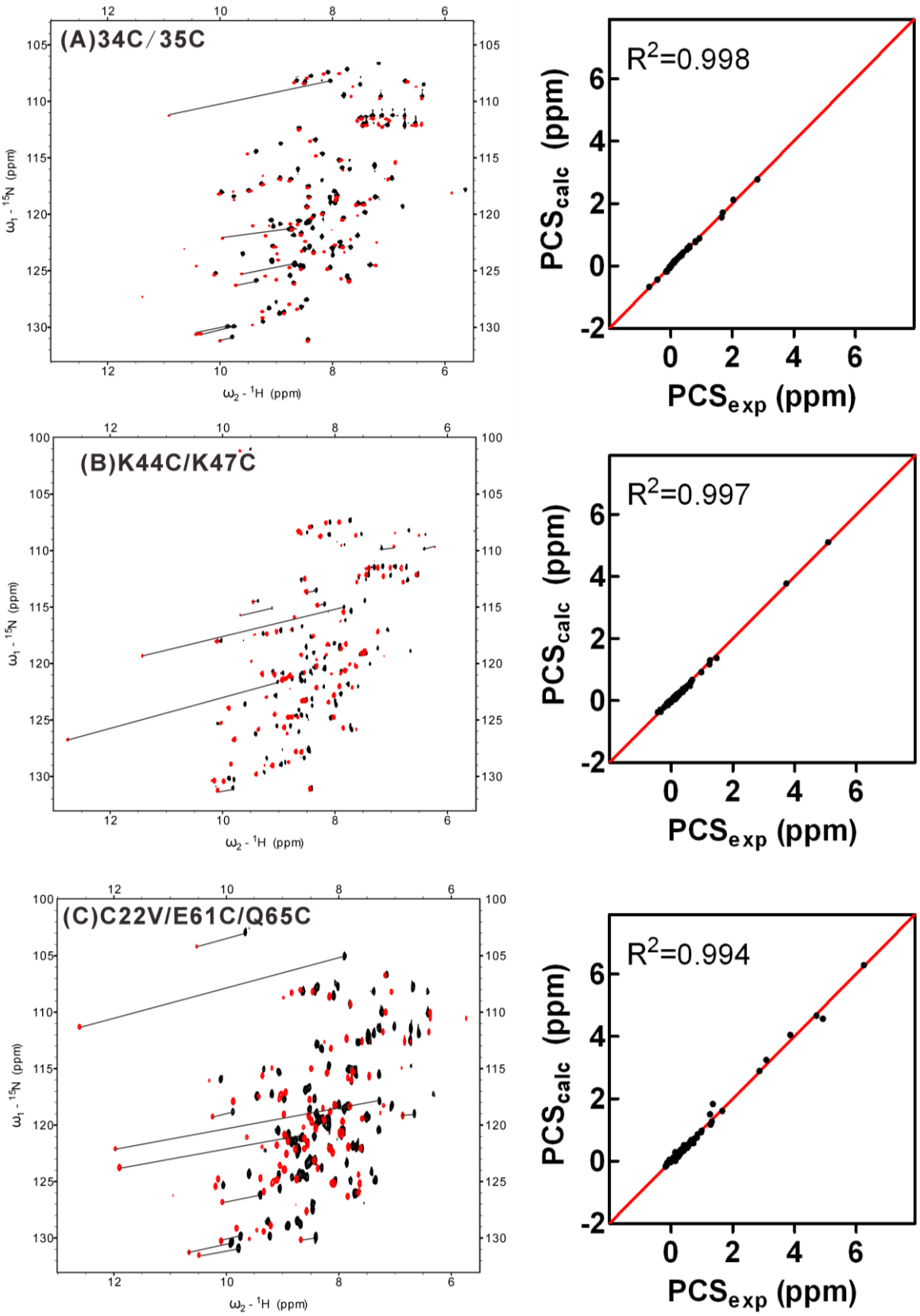

Figure 4.5: (Left) [^{1}H, ^{15}N]-HSQC overlay of CLaNP5 attached FKBP-12 mutants with Yb (red) and Lu (black). Solid lines show examples of PCS from selected corresponding residues. (Right) Correlation between experimental PCS and back-calculated PCS from FKBP-12 mutants attached to CLaNP-5(Yb). (A) K34C/K35C, (B) K44C/K47C, (C) C22V/E61C/Q65C.

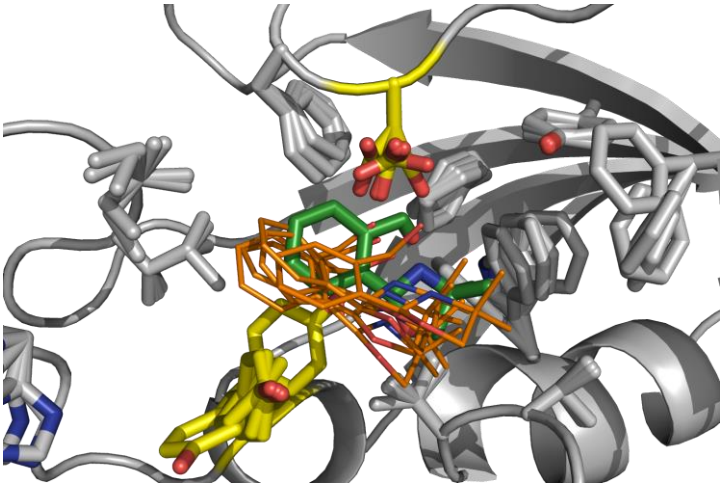

Figure 4.6: Superposition of the averaged NOE structure (in green) of the FKBP12-1 complex and the best five structures (in orange) calculated using ligand PCSs and the three experimentally determined Δχ-tensors. The protein backbone is represented as a grey ribbon except for the residues D37 and Y82 (in yellow). The average RMSD of the ligand from PCS calculations relative to the NOE calculation is 2.8 ± 0.4 Å.

A 'ghost' site found by PCS due to degeneracy of Δχ-tensors

A commonly encountered problem in paramagnetic NMR is the observation of multiple structure solutions due to the degeneracy of Δχ-tensor frames.[189 - 191] Here, we encountered a similar situation. In the calculations, another cluster of ligands was found at a site entirely different from the one shown in Figure 4.6 and which was neither identified by any intermolecular NOE, nor reported in the literature. Using the predicted Δχ-tensor parameters, this position was not identified. The position is located near a loop consisting of amino acids S39 to F46, opposite to the binding site mentioned above (Figure 4.7). At this 'ghost' site, the ligand has no van der Waal's contacts with the protein and is exposed on the protein surface and therefore it does not appear to be physically realistic. This non-physical position is apparently an artifact which originates from the degenerate nature of the Δχ-tensors. By removing solutions that have zero van der Waal's energy, it is possible to eliminate this artifact.

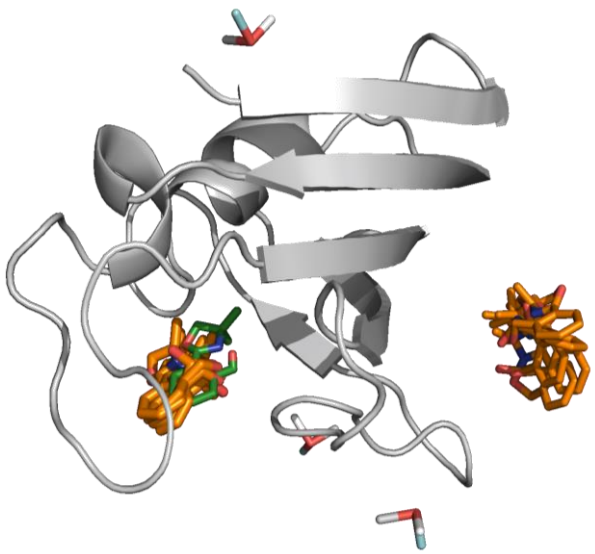

Figure 4.7: Second ligand position is defined by experimentally determined tensors. The two clusters calculated using the experimentally determined tensors are shown in orange and the averaged NOE position is shown in green. The metal positions and tensor orientations are shown in sticks, with the z-axis in cyan.

In order to visualize the degeneracy in a set of three tags, a script named PCSdock was written to produce a grid around the protein and calculate the PCS values of the pre-defined tags for each grid point. The algorithm then compares the experimental PCSs with the three calculated ones and calculates a Q value (equation 4.2). If the Q value for any atom of the ligand is below a threshold, that grid point is selected, otherwise it is discarded. Thus, PCSdock gives a fast, though crude, representation of where the ligand could be located on the basis of the PCSs. While the procedure is likely to overestimate the number of possible locations, it can be a useful tool for establishing degeneracy. Applied to FKBP12 with the three tags and the experimentally observed PCSs, the calculation clearly shows two possible, spatially distinct areas (Figure 4.8A). Each of the two low energy clusters from the calculation using experimentally determined $\Delta\chi$-tensors fits into one of the areas defined by the PCSdock calculation. Interestingly, this 'ghost' site was not found when calculated using the predicted $\Delta\chi$-tensors (Figure 4.8B).

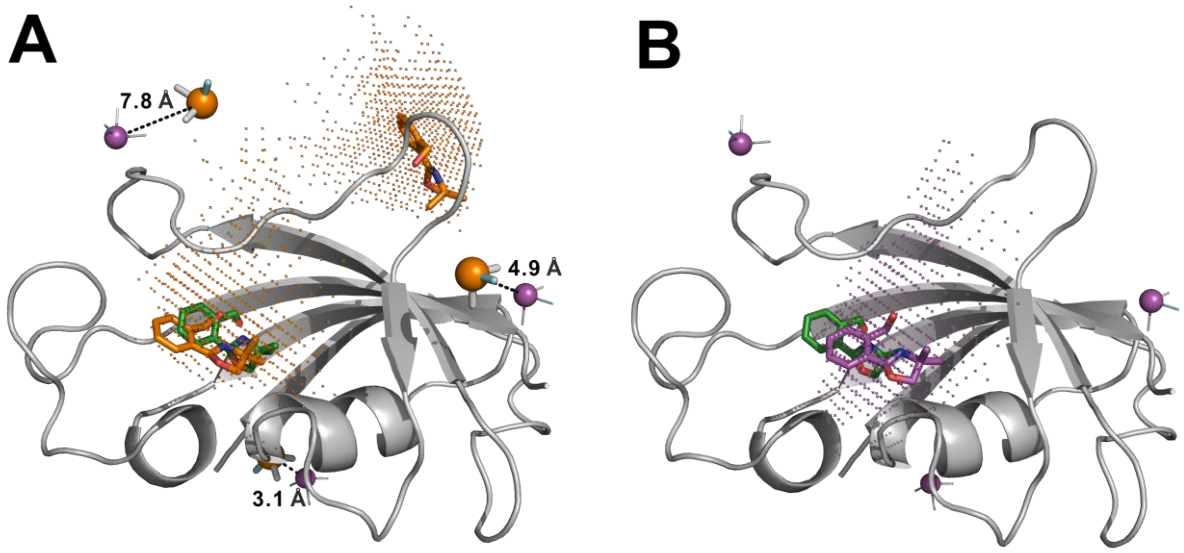

Figure 4.8: Grid points produced by PCSdock (Q = 0.15). The metal positions are shown in spheres (magenta, predicted positions; orange, experimentally determined positions), and $\Delta\chi$-tensor orientations are shown in sticks, with the z-axis in cyan. Ligands in green are the NOE-based position. (A) The two ligands in orange show the two locations determined by the experimental $\Delta\chi$-tensors. Two clusters were determined by PCSdock (orange dots): one close to the actual binding site, the other at the 'ghost' position. Distances between the metal positions determined by pure prediction and by experimental PCS are indicated. (B) Ligand in magenta shows the position determined using the predicted tensors. Only one cluster was found by PCSdock (magenta dots).

4.4 Discussion

We have demonstrated a general method to obtain structural information on complexes of weakly binding, small molecule ligands and proteins. The method seems most applicable to proteins that cannot be isotopically labeled, for which X-ray crystallography fails and for which many ligands need characterization. Although powerful, the method has certain limitations. On the target side, proteins with many surface-exposed cysteine residues are not appropriate as it would be quite challenging to attach the CLaNP-5 tag at one unique position. There are also restrictions on the ligands such that those with symmetric structures, with many scalar couplings or that exhibit intermediate exchange on the NMR time scale can cause difficulties in analyzing the data. Ligands that have symmetric structures tend to show degeneracy in NMR spectra, resulting in a reduced number of restraints that are themselves degenerate. Ligands with many scalar couplings, such as a saturated hydrocarbon ring, have ^{1}H spectra that can be difficult to resolve and the PCSs may be difficult or impossible to measure. Similarly, ligands in intermediate-exchange are difficult to study in NMR due to NMR line broadening leading to resonance overlap or less visible resonances. It should be noted that such ligands will present a challenge to any paramagnetic NMR approach, not just those based on the use of a lanthanide tag.

If the goal is to precisely define the binding site of the ligand on the protein, it is essential to obtain a sufficient number of meaningful restraints. Due to their small volume, fragments can only generate restraints covering a limited part of $\Delta\chi$-tensor space, easily leading to degenerate solutions. We addressed this issue by attaching CLaNP-5 at several sites throughout the protein surface. Three tag positions were used to generate a total of 21 PCS restraints for structure calculations. It is also possible to use CLaNP-5 with different lanthanides at a single tagging site, an approach used in previous studies for which multiple $\Delta\chi$-tensor positions were not available.[97,99] However, the tensor orientations of different lanthanides are similar in the same tag, so the PCSs merely scale with the magnitude of the anisotropy, reducing the information content of the extra restraints. In this case, a combined application of CLaNP-5 and CLaNP-7[188] would be possible using the same tag position, because the $\Delta\chi$-tensor orientations differ between these two CLaNP molecules. We expected that the three tagging sites would lead to a unique solution for the binding site. Unfortunately, degeneracy remained, leading to a physically unrealistic binding site in addition to the correct site. While the relevant site could be readily distinguished from the irrelevant one, some caution must nevertheless be exercised when interpreting the results. The presence of

this 'ghost' site suggests that, depending on the actual situation, more than three mutants might be required to fully break all degeneracy. Additionally, it is important to consider the (potential) dynamic behavior of sites within the protein when selecting positions to generate the dual cysteine mutations. Motion of the attachment site within the protein can influence both the position of the paramagnetic center and magnitude of the $\Delta\chi$-tensor.[176,185] For example, the large difference between the predicted and experimental position of the lanthanide for FKBP12 (K34C/K35C) and the orientation of its tensor (Figure 4.4A and Figure 4.5) is likely related to its location on an ill-defined protein loop, which makes the prediction less reliable and bears the risk that the tag slightly affects the average structure of the loop.

A number of NMR approaches have been proposed to determine the structure of protein-small molecule complexes. The approach based on intermolecular NOEs, which is robust and provides structural detail, is most frequently used. The method emphasizes short-range restraints (<5 Å or <8 Å depending on the isotopic labeling scheme) and requires resonance assignments of the protein. With the assistance of computational modeling, the NOE method can be relatively fast, as demonstrated in the system of matrix metalloproteinases (MMPs) and its tight-binding ligands.[192] A particular limitation of the method is the requirement for isotope-labeled protein, typically necessitating expression in *E. coli*. In contrast, methods that exclusively observe the NMR spectrum of the ligand can, in principle, be applied to protein derived from nearly any source. Epitope mapping, a method that quantitates the amplitude of ligand resonances in a saturation transfer difference spectrum (STD),[76,77] can provide information on how the small molecule binds to the protein. In theory, this information could be used for constraining molecular docking efforts. However, significant artefacts can be introduced by the inherent differences in transverse relaxation rates of ligand resonances. Therefore, the most significant advantage of the paramagnetic methods is that structural information relative to a fixed point on the protein can be reliably obtained from the ligand spectra. Previous paramagnetic NMR methodologies for determining protein-ligand structures have been limited to metalloproteins with an intrinsic metal binding site[97] or have used a lanthanide tag that can only be placed at the N-terminus of a protein.[99] In contrast, the CLaNP-5 tag can be used on non-metalloproteins and can be placed at a variety of sites on the protein, provided they are sufficiently rigid. It should be noted that in our hands, the effect of the tag on the stability of the protein is variable. In some cases, but certainly not all, the protein more easily precipitates, perhaps due to the partly hydrophobic nature of the cyclen ring system. In general we design several extra double Cys mutants and

use the most stable ones. The tag is attached via disulphide bridges, which means that the probe can be lost. The rate of dissociation is variable (days-weeks), so it is advised to prepare the samples freshly or store them at -80ºC.

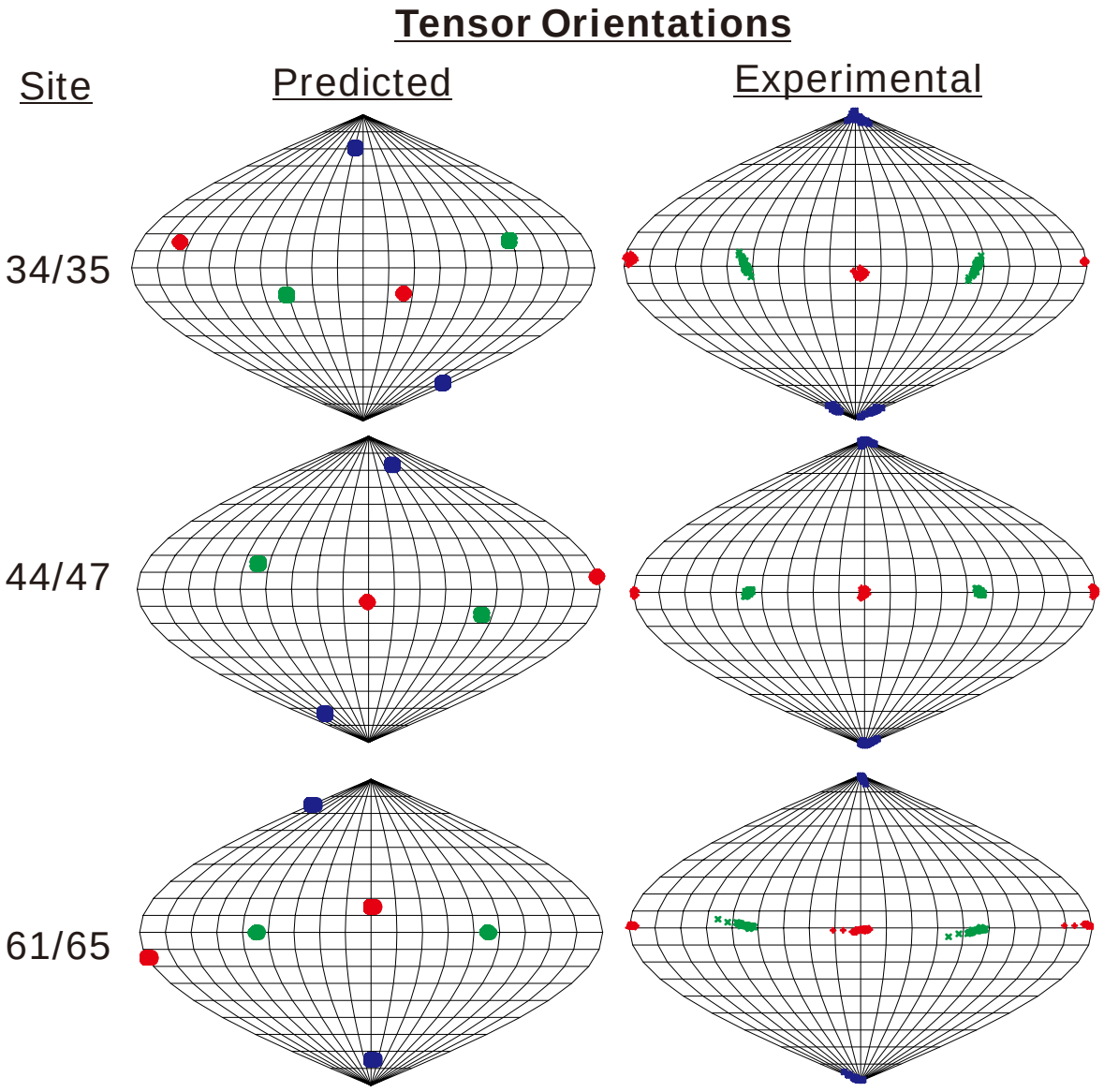

Figure 4.9: Sanson-Flamsteed projections of the predicted and experimental tensors. The orientations of the experimental tensors were translated to the axes of the Cartesian coordinates for an easier comparison with the predicted tensors. Blue, green and red colors represent z, y and x axes. The errors in the experimental tensors were calculated by randomly excluding 10% of the data. The orientations of the predicted tensors were modeled based on the protein structure. The intervals between the grid lines are 20 degrees.

The principle of the proposed technique is based on determination of the ligand position with respect to the paramagnetic Δχ-tensor. In order to predict the Δχ-tensor positions, the protein structure must be available. The results presented in this study have demonstrated that, with the predicted Δχ-tensor parameters, it is possible to identify the ligand binding site. It is not necessary to have protein resonance assignments in order to predict the Δχ-tensor position and therefore the approach can be applied to proteins that cannot be readily isotopically labeled. In principle, the approach can also be applied to proteins for which experimental structure information is not available, providing reliable structure prediction methods (such as homology modeling[193,194]) are applicable. For drug discovery, the potential binding site can be identified, which can accelerate optimization of hits to achieve higher affinity and greater biological activity even when the structure of the target is not available.

4.5 Conclusions

We have successfully determined the site and orientation of a small molecule ligand binding to a protein using ligand PCS and validated the results with an NOE structure. The use of the CLaNP-5 tag to induce the paramagnetic effects makes this approach suitable for non-metalloproteins. The results show that this strategy can identify the ligand binding site better than chemical shift perturbations. Comparison of the PCS structures from predicted and experimentally determined tensors demonstrates that the predicted tensor positions are sufficient for coarse definition of the binding site and it is therefore not necessary to experimentally optimize the tensor position. This PCS-based approach can be useful in early stages of fragment-based drug discovery to identify binding sites for proteins that are difficult to be enriched with isotopes, and in this way support optimization of early fragment hits.