Applications of paramagnetic NMR spectroscopy for protein research
Lescanne, M.

Citation
Lescanne, M. (2018, May 29). *Applications of paramagnetic NMR spectroscopy for protein research*. Retrieved from https://hdl.handle.net/1887/62455

Cover Page

Universiteit Leiden

Author: Lescanne, Mathilde
Title: Applications of paramagnetic NMR spectroscopy for protein research
Date: 2018-05-29

Chapter 2

Methyl group assignment using pseudocontact shifts with PARAssign

Abstract

A new version of the program PARAssign has been evaluated for assignment of NMR resonances of the 76 methyl groups in leucines, isoleucines and valines in a 25 kDa protein, using only the structure of the protein and pseudocontact shifts (PCS) generated with a lanthanoid tag at up to three attachment sites. The number of reliable assignments depends strongly on two factors. The principle axes of the magnetic susceptibility tensors of the paramagnetic centers should not be parallel so as to avoid correlated PCS. Second, the fraction of resonances in the spectrum of a paramagnetic sample that can be paired with the diamagnetic counterparts is critical for the assignment. With the data from two tag positions a reliable assignment could be obtained for 60% of the methyl groups and for many of the remaining resonances the number of possible assignments is limited to two or three. With a single tag, reliable assignments can be obtained for methyl groups with large PCS near the tag. It is concluded that assignment of methyl group resonances by paramagnetic tagging can be particularly useful in combination with some additional data, such as from mutagenesis or NOE-based experiments. Approaches to yield the best assignment results with PCS generating tags are discussed.

Keywords

Abbreviations

PCS, pseudocontact shift; RDC, residual dipolar coupling; PRE, paramagnetic relaxation enhancement; HSQC, heteronuclear single quantum coherence; TROSY, transverse relaxation optimized spectroscopy; CLaNP-5, caged lanthanoid NMR probe 5; ntd-HSP90, N-terminal domain of heat shock protein 90

Introduction

Liquid state NMR is widely used to study protein structure, dynamics and interactions with atomic detail. Resonance assignment is a prerequisite for all these studies. Methyl groups have become important in biomolecular NMR to probe biomolecular interactions[23] as well as dynamics of proteins over a large range of timescales, from picoseconds to seconds [28]. Furthermore, the hydrophobic nature of methyl groups makes them suitable to probe the core of a protein [26] and study specific interactions it may have with other proteins, ligands or substrates [27]. From a spectroscopic point of view, $^{13}CH_3$ methyl groups are valuable probes in very large systems because their inherent fast rotation and geometry give them favorable relaxation properties that make methyl groups suitable for TROSY experiments[24,25]. Specific isotope labeling on isoleucine, leucine and valine (ILV) methyl groups is an established approach [29] to reduce spectral crowding and to provide good spatial coverage due to dispersion of these residues through the protein. Assignment of ILV methyl groups is, however, non-trivial. To take advantage of standard *J*-coupling based assignments procedures, uniformly labeled samples are needed, requiring the production of additional samples and time-consuming 3D NMR experiments. For large systems, methyl assignment is further complicated by the need for fractional or complete deuteration of the other hydrogen positions to reduce relaxation by dipolar couplings. Various approaches have been published to assign methyl groups in large proteins. Systematic site-directed mutagenesis, combined with stereospecific labeling on methyl groups has been applied to TET2 [30,31]. To circumvent the size limitation, the so-called "divide and conquer" approach also proved to be effective. Here, individual domains or subunits are first assigned, followed by a transfer of assignments to the spectrum of the larger multidomain protein or multisubunit protein complex [31,35,36]. Another approach is to use multidimensional NOESY spectra in combination with a crystal structure [37]. This approach has been implemented in the MAP-XSII software [38] that automatically assigns methyl groups based on a known structure combined with NOEs measured on 4D NOESY experiments. Similarly, the program FLAMEnGO uses a combination of 3D NOESY, methine-methyl TOCSY and paramagnetic relaxation enhancements (PREs) and a protein structure to assign methyl groups [39].

More generally, paramagnetic NMR has become attractive for assignment [101], and software packages have been developed such as Platypus [102], Echidna [100], Possum [41] and PARAssign[42,43]. Platypus and Echidna have been used to assign NH resonances on the basis of a crystal-structure and pseudo-contact shifts (PCS). Possum assigns methyl groups with PCS measured with paramagnetic centers for which the magnetic susceptibility tensor is known. This requires assignment of at least five amides or methyl groups to be available. PARAssign does not require any assignment or prior knowledge of the susceptibility tensors. It automatically assigns ^1H, ^{13}C, or ^{15}N resonances based on PCS from several paramagnetic centers obtained from 2D HSQC

or TROSY spectra and a 3D structure of the protein and it also provides an assessment of the reliability of the assignment. It has been shown that PARAssign software could assign the amides of pseudoazurin (PAZ) (125 residues) using PCS from 3 three different paramagnetic centers. PARAssign has been tested also on P450cam (414 residues) for methyl groups assignment using synthetic data[42].

Here, we evaluate the ability of PARAssign to assign the 76 ILV methyl groups of a 25 kDa protein, the N-terminal domain of HSP90 (ntd-HSP90), using experimentally observed PCS. Ntd-HSP90 [103-105] is a well-studied chaperone protein, for which methyl group assignments are available[57]. PCS were induced by the paramagnetic tag CLaNP-5, attached to the protein via disulfide linkages to engineered cysteine residues. We find that, using two judiciously placed paramagnetic tags, PARAssign can assign up to 60% of the 76 methyl groups of ntd-HSP90 with high reliability and give many correct suggestions for the remaining 40%. The quality of the input data is an important determinant for the number of reliable assignments. With a single tag, reliable assignments can be obtained for nearby nuclei with large PCS.

Material and methods
Protein preparation

Four double cysteine mutants of the ntd-HSP90 were designed on the surface of the protein, S50C/D54C, A101C/N105C, T149C/I187C, and M130C/Q133C. Ntd-HSP90 does not have any native cysteines. The protein was produced by overexpression of the gene for ntd-HSP90 (residues 8-233, UNIPROT entry P07900) with a cleavable His$_6$ tag in the PQTEV vector in *Escherichia coli* BL21 DE3. The plasmid was kindly provided by Dr Dipen Shah[57]. Cultures of 500 mL M9 medium in 2 L Erlenmeyer flaks were incubated at 37 °C at 150 rpm. At an OD$_{600}$ of 0.6, IPTG (1 mM) was added to induce gene expression, the temperature was reduced to 18°C and the incubation was continued overnight. For ^{15}N labeled protein, ^{15}N ammonium chloride (0.3 g/L) was added as sole nitrogen source. For ^{13}CH$_3$ methyl labeling, two samples were produced labeled either at Leu-δ1-δ2/Val-γ1-γ2-[^{13}CH$_3$] or Ile-δ1-[^{13}CH$_3$]. Consequently, it was possible to identify unambiguously the residue type of the Ile methyl groups. Either 50 mg/L 2-keto-3,3-1,2,3,4-^{13}C-butyrate (Ile ^{13}CH$_3$ labeling) or 100 mg/L of 2-keto-3-methyl-3-1,2,3,4-^{13}C butyrate (Leu and Val ^{13}CH$_3$ labeling) were added to the M9 medium with 4 g/L ^{12}C glucose, one hour before induction at OD$_{600}$ ~0.5. The cells were harvested by centrifugation at 6000 rpm for 20 min at 4°C. Cell lysis was performed using lysozyme. PMSF and DNAse were added for protease inhibition and DNA breakdown. Protein purification was performed with a His-trap NTA-column, eluting with an imidazole gradient (5-500 mM imidazole in 100 mL). The His-tag was then cleaved with TEV protease during overnight incubation at 4°C. The His-tag and TEV protease were removed by a second elution from the column. CLaNP-5 was synthesized as described before[106,107]. For CLaNP-5 attachment, the protein was reduced

for one hour on ice in the presence of 5 mM DTT. Then, DTT was removed with a Sephadex G-25 PD10 desalting column (GE Healthcare) and the protein solution was mixed immediately with four molar equivalents of either Yb^{3+}-CLaNP-5 or Lu^{3+}-CLaNP5 and incubated at room temperature for one hour. The sample was loaded on an analytical Superdex 200 gel filtration column (GE Healthcare) to remove excess CLaNP-5 and protein dimers.

NMR spectroscopy

NMR samples contained 50 – 100 μM protein in 50 mM Tris-HCl, 100 mM NaCl, pH 7.5 and, 6% D_2O. All the spectra were recorded at a temperature of 298 K. ^{15}N-HSQC spectra were recorded on a Bruker Ascend Avance III 850 HD MHz spectrometer, equipped with a TCI cryoprobe, with spectral widths of 20.00 ppm (17.00 kHz) and 35.00 ppm (3.015 kHz) in the ^1H and ^{15}N dimensions, respectively. Leu-δ1-δ2/Val-γ1-γ2-[$^{13}CH_3$] and Ile-δ1-[$^{13}CH_3$] protein spectra were recorded on Bruker 800 MHz and 850 MHz spectrometers equipped with TXI cryoprobes, with a spectral width of 20.00 ppm in the ^{13}C dimension. Experiments were performed with 16 or 32 scans. Data were processed with Topspin 3.2 and NMRpipe 8.2 [108] and spectra were analyzed with CCPNMR Analysis 2.4 [109].

Determination of the PRE-derived cutoff distance

The cutoff distance, below which a signal would be too broad to be detected because of PRE, was calculated using **equation 2.1**. The line broadening due to PRE (Γ_2) was estimated to render nuclei within 9 Å from the Yb^{3+} ion invisible, based on the expected Curie relaxation, **equation 2.1** [110], which is the dominant transverse relaxation mechanism for ^1H by lanthanoids with an anisotropic magnetic susceptibility.

$$r_{\text{cutoff}} = \sqrt[6]{\frac{1}{5\Gamma_2}\left(\frac{\mu_0}{4\pi}\right)^2 \frac{\omega_H^2(g_e\mu_B)^4 J^2(J+1)^2}{(3k_BT)^2}\left(4\tau_r + \frac{3\tau_r}{1+\omega_H^2\tau_r^2}\right)} \quad \textbf{(2.1)}$$

Here, μ_0 is the permeability of a vacuum, ω_H is the Larmor frequency of the hydrogen nucleus, g_e is the Landé g-factor, μ_B is the Bohr magneton, J is the total spin quantum number of the paramagnetic lanthanoid, k_B is the Boltzmann constant, T is the absolute temperature, and τ_r is the rotational correlation time of the molecule.

Methyl group assignments

Methyl group assignments from PARAssign were compared to those obtained with traditional 3D NMR procedures. The published methyl group assignments[57] were checked and completed using CccoNH/HccoNH/HSQC spectra, kindly provided by Dr. Dipen Shah and Dr. Eiso Ab.

Assignment procedure

The PARAssign code has been refactored for the sake of readability and easier maintenance. A new reliability criterion has been implemented (see Results section). The core procedure remains unchanged[42]. The PARAssign user must supply the PCS file containing all the measured proton PCS, a PDB file and the JSON configuration

file with information about the assignment to be performed (atom names, stereospecificity of the input, double Cys mutation sites, starting point for the tensor parameters, input file directory, output file directory). ^{13}C PCS were not considered as input, because the resolution in this dimension is lower and the residual chemical shift anisotropy is significant for carbon but can be neglected for protons. The PARAssign output includes the assignment plot, tensor parameters, tensor PDB file(s) and predicted PCS for all the assignable groups in the structure. The PARAssign input consisted of the ntd-HSP90 structure, PDB entry 3T0Z [111], and the PCS datasets acquired for the S50C/D54C, A101C/N105C and T149C/I187C mutants tagged with Lu^{3+}-CLaNP-5 as diamagnetic control and Yb^{3+}-CLaNP-5 as the paramagnetic center. Although we used two different samples for isoleucines and leucines + valines, PARAssign can select the predicted type(s) of each assignable peak by entering the residue type code in the input file ('ILE' or 'VAL LEU' in our case but all possible combinations are acceptable). A run of PARAssign comprises an iteration of two consecutive procedures, the assignment procedure and the tensor optimization procedure using PCS calculated during the previous assignment procedure. In the assignment procedure of the PARAssign software, the Hungarian method for minimal cost assignment [112] is used to perform the assignment of the input PCS, using a scoring factor cost function to populate the assignment matrix:

$$Scoring\ factor = \frac{1}{Pc} * \sum_{i=1}^{Pc} \sqrt{\frac{\left(\delta_{PCS,i}^{pred} - \delta_{PCS,i}^{exp}\right)^2}{\left(\left|\delta_{PCS,i}^{pred}\right| + \left|\delta_{PCS,i}^{exp}\right|\right)^2}} \quad (\mathbf{2.2})$$

where Pc is the number of paramagnetic centers. PARAssign attributes a scoring function of 1 to assignments for which experimental and predicted PCS have opposite signs, thus disfavoring the assignment. A scoring factor threshold is user-defined (default 0.2) and is used to filter out assignments for the optimization procedure for which the scoring factor is too high. A high threshold yields many assignments with poor accuracy and a low threshold yields few assignments. The optimal threshold is the lowest possible value that yields the maximum number of assignments. For instance, when starting with a scoring factor of 0.7 and decreasing it progressively, the optimum threshold is passed when the number of assigned groups starts to drop. The tensor parameters optimization procedure consists of the minimization of the square difference between calculated and experimental PCS for all paramagnetic centers. This sequential least squares programming procedure [113] is performed by the python function scipy.optimize.fmin_slsqp (http://www.scipy.org). The output of PARAssign gives the susceptibility tensors of the paramagnetic centers and the assignment of the diamagnetic protein spectrum with a reliability scoring for each assignment (discussed below). The output figures included in this article may not reflect the true output quality. In the original .png image, the user can zoom in to read all the details. For this chapter, it is relevant to be able to judge the quality of the assignment at a glance.

Results and discussion

Selection of tagging sites

Assignment based on PCS requires accurate PCS that are not affected by averaging effects due to tag mobility. For this reason, the two-armed caged lanthanoid NMR probe (CLaNP) #5 was used [107,114], which requires two Cys residues located close together on the surface of the protein for attachment. Ntd-HSP90 has no Cys residues, so pairs of cysteines were engineered. The criteria for selection were (I) good distribution of the sites over the protein surface, (II) limited mobility of the attachment site by using secondary structure elements, (III) sufficiently exposed sites without steric hindrance of large side-chains. Four double cysteine mutations were designed, three of which are on α-helices (S50C/D54C, Cα-Cα = 6.0 Å; A101C/N105C, Cα-Cα = 5.9 Å; M130C/Q133C, Cα-Cα = 5.7 Å) and one was across neighboring strands of a beta-sheet (T149C/I187C, with Cα-Cα = 5.0 Å). Out of the four double cysteine variants, the three with the highest yields were selected, which were S50C/D54C, A101C/N105C and T149C/I187C, with yields of 30 mg/L, 30 mg/L and 20 mg/L in minimal medium, respectively. Variant M130C/Q133C yielded only 1 mg/L and was not considered further.

Reference $\Delta\chi$-tensor determination using backbone amides

For reference, the lanthanoid location and the orientation of the $\Delta\chi$-tensor were first determined on the basis of amide proton PCS and using Numbat software [99] with a crystal-structure of ntd-HSP90 (PDB entry 3T0Z [111]), **Table 2.1**.

Table 2.1 Susceptibility $\Delta\chi$-tensors obtained by fitting amide PCS of ntd-HSP90 (PDB entry 3T0Z)

Mutant tagged with Yb^{3+}-CLaNP-5	50C/54C	101C/105C	149C/187C
$\Delta\chi_{ax}$ (10^{-32} m^3)	8.4 (0.1)	8.0 (0.2)	8.3(0.3)
$\Delta\chi_{rh}$ (10^{-32} m^3)	2.0 (0.3)	1.6 (0.5)	2.2(0.3)
α (°)	173(5)	180(2)	176(3)
β (°)	155(1)	88(2)	148(1)
γ (°)	150(3)	170(3)	15(4)
x (Å)	12.4(0.5)	25.3(0.2)	-1.8(0.2)
y (Å)	1.3(0.5)	-9.5(0.2)	-5.8(0.5)
z (Å)	25.2(0.3)	6.8(0.2)	-1.2(0.2)
Restraints	80	63	69
Q value*	0.056	0.068	0.26
Q^a value*	0.028	0.034	0.12
Average pcs deviation (ppm)	0.012	0.009	0.035

* The Q-values were calculated using:

$$Q = \sqrt{\frac{\Sigma(\delta_{PCS,i}^{pred} - \delta_{PCS,i}^{exp})^2}{\Sigma(\delta_{PCS,i}^{exp})^2}}$$

$$Q^a = \sqrt{\frac{\Sigma(\delta_{PCS,i}^{pred} - \delta_{PCS,i}^{exp})^2}{\Sigma(|\delta_{PCS,i}^{pred}| + |\delta_{PCS,i}^{exp}|)^2}}$$

Diamagnetic assignments for the amide groups were available [115], BMRB entry 7003. The three mutants were ^{15}N-labeled and tagged with the diamagnetic control tag Lu^{3+}-CLaNP-5 or the paramagnetic tag Yb^{3+}-CLaNP-5 and PCS were obtained from ^{15}N-^{1}H HSQC spectra. To ameliorate the pairing problem, spectra of the ^{15}N labelled ntd-HSP90 tagged with CLaNP-5 loaded with Eu^{3+} were acquired for mutants 50C/54C and 101C/105C. The two lanthanoid $\Delta\chi$ tensors are oriented in a nearly identical orientation and only the sizes of $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ differ. Thus, the paramagnetic peaks are on the one correlation line with the diamagnetic peak and the size of the CS scales with the strength of the $\Delta\chi$ tensor. The alignment of the three peaks presents an easy way to pair paramagnetic peaks to diamagnetic ones, **Figure 2.1** For mutants S50C/D54C and A101C/N105C, the agreement between observed and calculated PCS was excellent, **Figure 2.2**. For mutant T149C/I187C the fit was of somewhat lower quality, for reasons discussed later. To check whether significant PCS could be expected for most methyl groups, PCS were predicted for the three $\Delta\chi$-tensors derived from the amide PCS. **Figure 2.3** shows that with Yb^{3+}-CLaNP-5, PCS ≥0.04 ppm are expected for most methyl groups for each of the three tag locations. These amide-based $\Delta\chi$-tensors were used solely to generate the simulated data and for comparison with the ones refined by PARAssign for the experimental methyl group data, see the sections below.
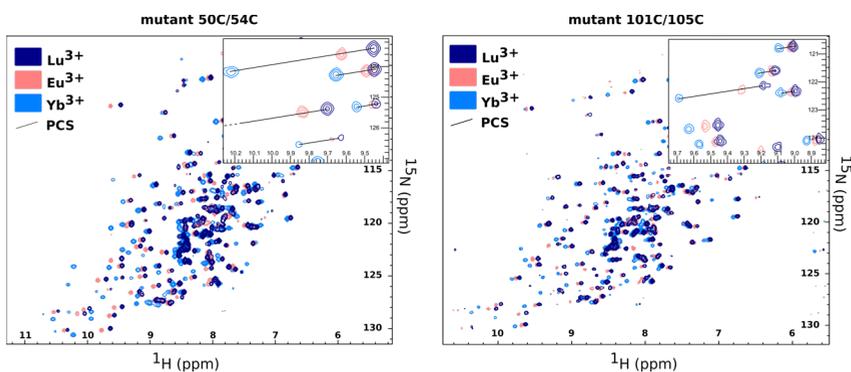
Figure 2.1. Overlay of HSQC spectra of the two ntd-HSP90 mutants 50C/54C and 101C/105C tagged with Lu^{3+}-CLaNP-5, Eu^{3+}-CLaNP-5, Yb^{3+}-CLaNP-5, in dark blue, pink and light blue, respectively. Zoom of the alignment of the three peaks are shown enclosed.
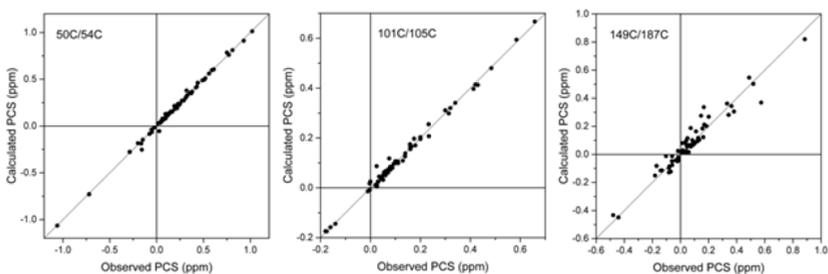
Figure. 2.2 Quality of $\Delta\chi$ fits. The calculated PCS are plotted against the observed PCS of amide groups of ntd-HSP90 tagged with Yb^{3+}-CLaNP-5 at the indicated positions. The solid lines represent a perfect correlation (x=y).

mutant 50C/54C   mutant 101C/105C   mutant 149C/187C
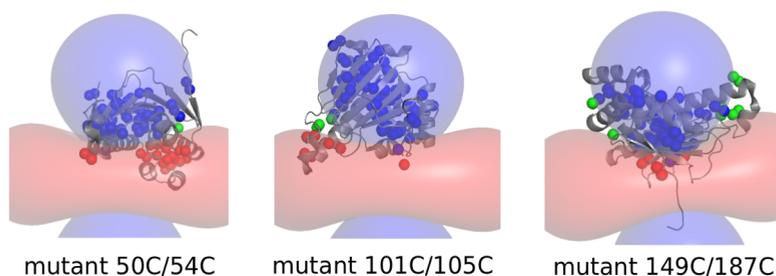
Figure. 2.3 Predicted methyl group PCS. Methyl groups that are predicted to experience positive PCS, negative PCS and PCS smaller than 0.04 ppm are represented as blue, red and green spheres, respectively. The -0.04 ppm and 0.04 ppm isosurfaces of the $\Delta\chi$-tensors based on the amide PCS are shown as red and blue surfaces.

Tagging effects on methyl group spectra

To establish the effects of a double cysteine mutation and the attachment of CLaNP-5 on methyl resonances, ^{13}C-^{1}H-HSQC spectra were acquired for the WT and Lu^{3+}-CLaNP-5 tagged mutant ntd-HSP90 with ILV methyl-labeling. The peaks from the mutant protein were matched to those of the wild-type. Most peaks could be readily matched and showed only very small chemical shift changes, **Figure 2.4a**. Two peaks, one for mutant 50C/54C and one for mutant 101C/105C, could not be matched with certainty to equivalent peaks in the spectrum of wild-type ntd-HSP90. These methyl groups were located close to the tag, **Figure 2.4b**. In mutant 149C/187C one Ile was replaced by a Cys for attachment of the tag. The limited chemical shift perturbations between WT and mutant ntd-HSP90 indicate that no major structural changes occur because of the mutations and the tagging.
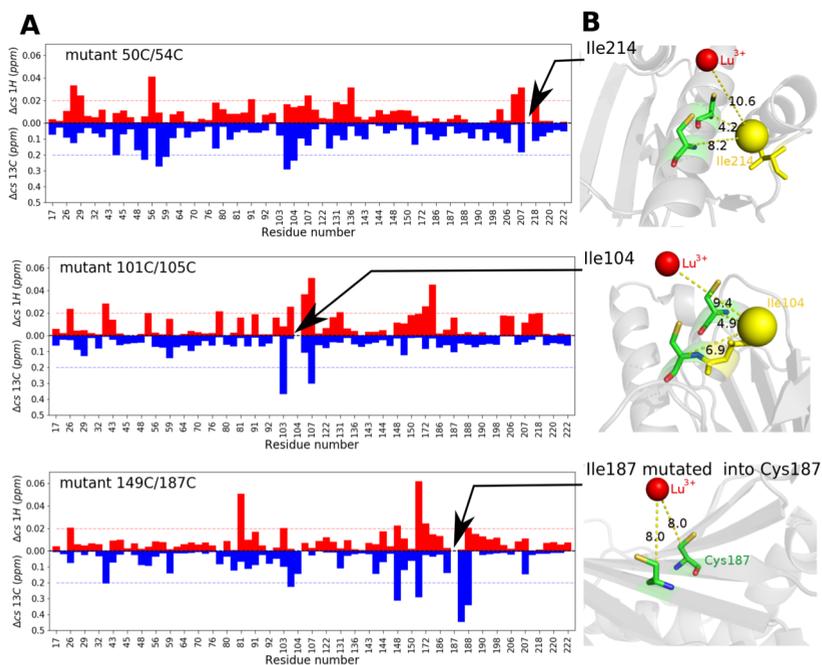
Figure. 2.4 Mutation and tagging effects. **a** Absolute chemical shift differences between WT and Lu³⁺-tagged mutant ntd-HSP90 resonances for ¹H (red) and ¹³C (blue) are plotted for each methyl group. The dotted lines mark significant changes, with 0.2 ppm for ¹³C and 0.02 for ¹H. **b** Location of methyl groups with unmatched resonances relative to the tag site. Distances are expressed in Ångström.

The pairing problem

PCS are measured as differences in chemical shifts of resonances observed in a paramagnetic and the diamagnetic spectrum. The problem to be solved is thus to pair the diamagnetic and paramagnetic resonances belonging to the same methyl group. PCS expressed in ppm units do not depend on the gyromagnetic ratio of the observed nucleus. Consequently, when the distance between the nucleus and the paramagnetic center, as well as the orientation of the vector connecting them in the $\Delta\chi$-tensor frame are similar for ¹H and ¹³C, which is almost always true for a ¹³C¹H₃ methyl group, the ¹³C and ¹H nuclei of a methyl group experience nearly the same PCS. In **Figure 2.5**, the predicted PCS for all methyl ¹³C and ¹H nuclei are plotted for ntd-HSP90 tagged at the three positions (224 points in total). The ¹H PCS are averaged over the three PCS for the individual protons in the methyl group to account for the fast methyl rotation. The ¹H PCS differ by less than 5% from the ¹³C PCS for PCS < 0.5 ppm and up to 10% for PCS larger than 0.5 ppm. Note that we neglected the effects of residual anisotropic chemical shifts and residual dipolar couplings due to partial alignment. In the plot, two exceptions can be seen. The angles of the metal-methyl vector of methyl groups L89

Cδ1 and V92 Cγ1 are very close to the angle where the PCS changes sign in mutant 149C/187C, resulting in a very high PCS gradient, **Figure 2.5**.
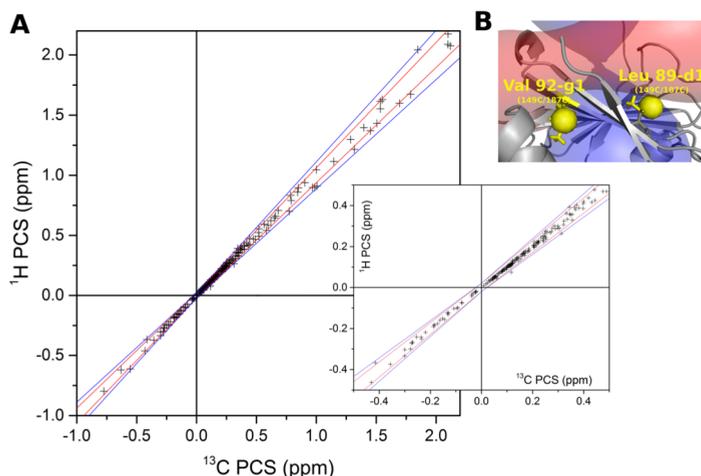
Figure. 2.5 Correlation between the ^{1}H PCS and ^{13}C PCS of methyl groups. **a** The ^{1}H and ^{13}C methyl PCS predicted for the three tag positions on ntd-HSP90 are plotted. The PCS differing less than 5% and 10% are contained between the red and blue lines, respectively. The inset shows a detail of the plot for small PCS. The lines intercept the axes at |0.02| ppm, taken as error margin. **b** Location of the two methyl groups, for the mutant 149C/187C, for which the deviation between ^{1}H and ^{13}C PCS is bigger than 10%. Isosurfaces are shown for PCS = 0.25 and −0.25 ppm. Both methyl groups are very close to the region where the PCS changes sign.

Thus, in general, the chemical shifts for the ^{1}H and ^{13}C are expected to be on a line with slope 0.95 – 1.05 for small PCS and 0.9 – 1.1 for large PCS, the correlation line. This property is used to identify which diamagnetic and paramagnetic resonances derive from the same nucleus and, thus, yield an experimental PCS (the pairing procedure). For well-dispersed parts of the spectrum pairing is trivial but in more crowded areas it can be ambiguous when more than one paramagnetic peak is on the correlation line. Pairing can either be done fully automatically with a script implemented in PARAssign based on these criteria, or in combination with manually paired peaks. When pairing manually, differences in peak intensity and shape can sometimes help, **Figure 2.6**.
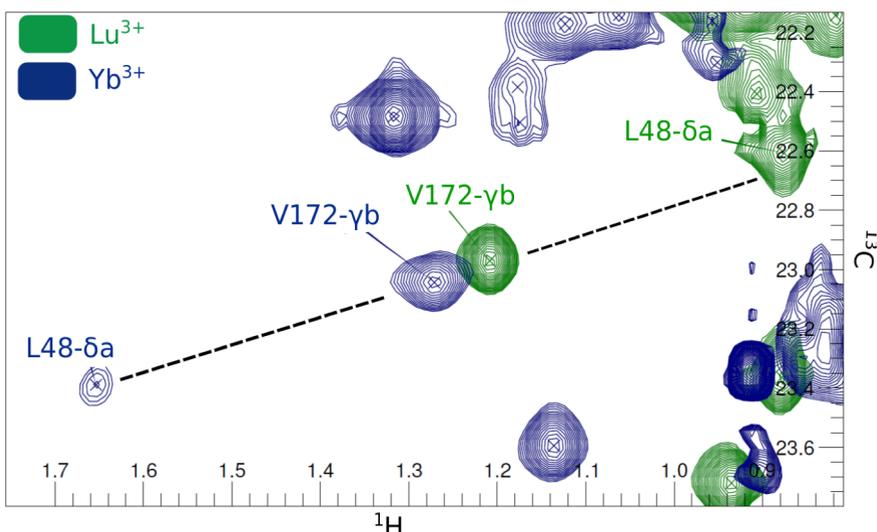
Figure. 2.6 Illustration of the peak pairing problem. The peaks for V172 and L48 are on the same correlation line in the spectrum of Yb[3+] tagged ntd-HSP90, making the pairing to the peaks in the spectrum of Lu[3+] tagged protein problematic. In this case, the observation that the V172 in the

Pairing is done iteratively, by pairing additional peaks using the predicted PCS after a run of PARAssign. In this way, 97% of ILV peaks were matched for both the mutants 50C/54C and 101C/105C and 58% of the mutant 149C/187C. The lower pairing level for the latter mutant as well as the relatively poor correlation between measured and calculated amide PCS suggest that the data for this mutant are of lower quality. It is unlikely that the tag causes a major structural modification because only small chemical shift differences between the spectra of the WT and Lu[3+]-CLaNP tagged mutant are observed, **Figure 2.4a**. The tag is also connected via both arms, because the size of the $\Delta\chi$-tensor is similar to that of the other mutants, **Table 2.1**. Single armed CLaNP yields strongly reduced $\Delta\chi$-tensor values [114]. However, some unexpected broadening and the presence of double peaks in the paramagnetic spectrum suggest that the tag may assume more than one orientation relative to the protein. The tag in mutant 149C/187C CLaNP-5 crosslinks two β-strands, a type of attachment that has not been reported before, to the best of our knowledge. Perhaps this type of attachment results in some degree of conformational freedom of the tag. The Cα-Cα distance is about 1 Å shorter for this mutant than for the two other considered mutants. In the two other mutants, the CLaNP tag connects two consecutive turns on alpha-helices, **Figure 2.4b**.

A new criterion for reliability

A consequence of the pairing problem is that the input data will, in general, not be complete, contrary to the synthetic data used in the simulation described in a previous paper[42]. To reduce possible artefacts in the assignment due to local minima in the search space, a jackknife procedure was implemented. Assignment is performed 100 times while leaving out 5% of the input PCS, randomly selected for each of the runs.

During each run, PARAssign performs the assignment automatically using an iterative procedure. The back-calculated PCS are matched with the input PCS, followed by the optimization of the Δχ-tensor parameters and lanthanoid location. PARAssign assesses the reliability of the assignments with a starring system. If all (and at least two) PCS for a given methyl group are >0.02 ppm and the combined fit of the PCS has a scoring factor, **equation 2.2**, lower than the average, the assignment is 2^*. If only one of the criteria is met, the assignment gets 1^*, otherwise 0^*. After 100 runs, the fraction of occurrence (out of the 100 runs) of each assignment is combined with the average starring of the assignment to get an overall reliability, indicated with a color code, green, yellow and white for highly reliable, suggestion and not reliable, respectively, illustrated in **Figure 2.7**.
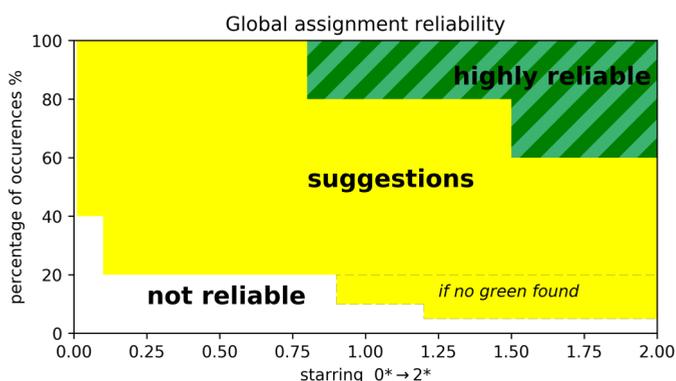
Figure. 2.7 Assignment reliability. The average quality score from the 100 assignment runs is indicated on the horizontal axis and the occurrence percentage of the runs in which the assignment is found is plotted vertically. Dark and light green refer to stereospecific and non-stereospecific assignment of Leu and Val methyl groups.

To be classified as highly reliable, colored in green, a residue assignment must occur in more than 60% of the runs with an average starring higher than 1.5 or occur more than 80% with an average starring higher than 0.8. Methyl groups that are assigned to only one of the two methyl groups in Val or Leu, i.e. methyls that are assigned stereo-specifically, are shown in dark green. If the residue is reliably assigned but both methyl groups are frequently identified as the correct assignment in the 100 consecutive assignment runs, the bar is shown in light green, indicating that the resonance has been assigned reliably but non-stereospecifically to a residue. It should be noted that the correctness of stereospecificity could not be assessed because no stereospecific assignments were available for ntd-HSP90. Assignments occurring in less than 5% of the runs are deemed unreliable (white). For all the other assignments, colored in yellow, PARAssign finds one or more assignments with reasonable occurrence frequencies (higher than 40% but not already green, higher than 20% with a starring >0.1, and, if no highly reliable assignment has been found, extra yellow assignments are proposed, see **Figure 2.7**). Such a set of suggestions can still be very useful when

combined with other strategies. Data from other sources (mutagenesis, NOE) could be used to decide which of the assignments is the correct one.

Methyl group assignment with simulated data

Ntd-HSP90 as represented in PDB entry 3T0Z contains 20 isoleucines labeled on the carbon δ1, 18 leucines labeled on both carbon δ1 and carbon δ2 and 10 valines labeled on both carbon γ1 and carbon γ2. Thus, in total 76 methyl groups need to be assigned. To establish whether PARAssign can solve the assignment problem, first the PCS for the 76 methyl groups were calculated using the $\Delta\chi$-tensors for the three tag positions obtained from the amide PCS. PARAssign was run using the WT protein structure 3T0Z and the 3 simulated PCS datasets for input. The starting parameters of the three $\Delta\chi$-tensors were $\Delta\chi_{ax}$ = 8.5 and $\Delta\chi_{rh}$ = 2.0. The starting angles and metal position were predicted by PARAssign purely based on reported positions relative to the Cys Cα and Cγ atoms and $\Delta\chi$-tensor orientations for CLaNP-5, which has been shown to be a good approximation[42,107]. Therefore, no prior experimental knowledge of the tensors is necessary. Using the simulated data for three paramagnetic centers as input, the vast majority of assignments was classified as highly reliable and all these 'green' assignments were correct. In addition, for the three assignments classified as 'yellow' the correct assignment was proposed, **Table 2.2.**

Table 2.2 Assignment of the methyl groups. The terms m1, m2 and m3 refer to the mutants 50C/54C, 101C/105C and 149C/187C, respectively. For the green category, the numbers represent the correct/total number of assignments. For the yellow category, the first number indicates for how many residues the correct residue was among the assignment suggestions, the second represents the total number of assignments in this category.

Number of Paramagnetic centers		3	2			1		
Mutant(s):		m1, m2, m3	m1, m2	m1, m3	m2, m3	m1	m2	m3
Complete simulated dataset	input	76	76	76	76	76	76	76
	Highly reliable	73/73	61/61	6/6	73/73	0/0	5/5	7/10
	Suggestion	3/3	15/15	69/69	3/3	22/52	69/70	16/53
	Not reliable	0	0	1	0	24	1	13
Experimental dataset	input	76	76	76	76	72	74	44
	Highly reliable	27/27	46/46	0/0	9/9	1/1	6/6	1/3
	Suggestion	45/47	27/28	22/55	58/66	25/57	40/56	13/33
	Not reliable	2	2	21	1	14	12	8

The PARAssign output is displayed in **Figure 2.8**. When using the data from only two tags, two of the three pairs of tags (50C/54C with 101C/105C and 101C/105C with 149C/187C) yield similar results as with three tags, **Table 2.2**, **Figure 2.9**. Combining the data of mutants 50C/54C and 149C/187C yields mostly assignments with more than one possibility ('yellow'), on the whole. The reason for this striking difference will be discussed below. We can conclude from the assignment runs with synthetic datasets that the assignment problem can be solved by PARAssign with complete and error-free data input from two or three tag positions.

Figure. 2.8 PARAssign output for a complete, simulated dataset. In green the highly reliable assignments and in yellow the suggested assignments are indicated. In each bar, the proposed residue and methyl group as well as the average starring value are shown. In this example, all the assignments are correct. See also the enlargement in **Figure 2.10**.

Figure. 2.9 PARAssign output for the three combinations of two of the complete, simulated datasets.

Assignment of methyl groups with experimental datasets

First, the methyl resonances of ntd-HSP90 were assigned using experimental datasets from the three paramagnetic centers. The PCS pairing procedure yielded 72, 74 and 44 of the expected 76 PCS for the mutants 50C/54C, 101C/105 and 149C/187C, respectively. As we had collected the spectra of Ile labeled protein samples separately from those labeled with Leu and Val (see Materials and Methods), it was possible to identify Ile methyl groups but we could not distinguish Leu and Val methyl groups. Therefore, in the input file the two possible residue types were entered as "LEU VAL". In case of other methyl group labeling schemes, the residue type can be entered with its amino acid code, for example, "MET" for the methionine methyl group labeling. PARAssign runs were performed as described above, by using either the data from three, two, or one paramagnetic center(s). The results are listed in **Table 2.2** and PARAssign output using experimental datasets of three mutants is displayed **Figure 2.10**.
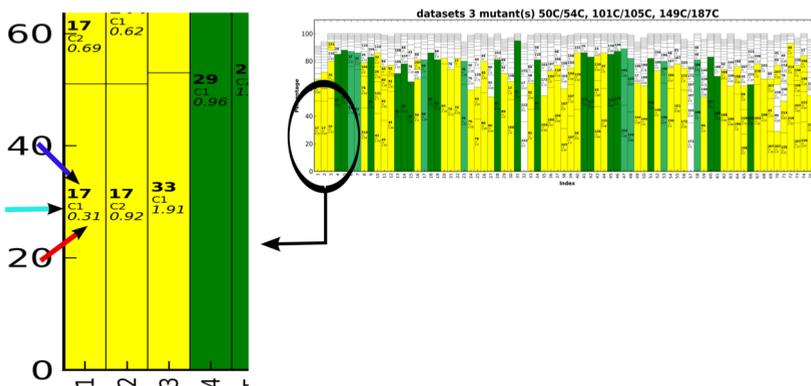
Figure. 2.10 PARAssign output for the complete experimental dataset. The highly reliable assignments are shown in green and the proposed reliable assignments are in yellow. In this example, all the green assignments are correct and 95% of the yellow assignments contain the correct assignment (among 1 to 4 possibilities). In the close-up the blue, cyan and red arrows indicate the residue number, the stereospecificity and the scoring factor, respectively.

The tensor parameters are given in **Table 2.3** and **Figures 2.11** and **2.12**.

Table 2.3 Tensor parameters refined for the different combinations of datasets. In parenthesis are the standard deviations. The Euler angles α, β, γ are defined according to the z-y-z convention. A * indicates that there are other conformations. See also **Figures 2.11** and **2.12**.

	3 paramagnetic centers			2 paramagnetic centers						1 paramagnetic center		
				50C/54C – 101C/105C		50C/54C – 149C/187C		101C/105C – 149C/187C				
	50C/54C	101C/105C	149C/187C	50C/54C	101C/105C	50C/54C	149C/187C	101C/105C	149C/187C	50C/54C	101C/105C	149C/187C
$\Delta\chi_{Ax}$ (10^{-32} m^3)	8.1 (0.7)	7.5 (0.6)	8.2 (0.9)	7.8 (0.5)	7.7 (0.7)	7.6 (0.8)	7.2 (0.5)	7.6 (0.6)	8.5 (0.8)	7.3 (0.7)	7.3 (0.5)	8.3 (0.6)
$\Delta\chi_{Rh}$ (10^{-32} m^3)	2.0 (0.4)	2.4 (0.3)	2.4 (0.3)	2.1 (0.4)	2.4 (0.2)	2.2 (0.4)	2.4 (0.3)	2.0 (0.3)	2.2 (0.3)	2.4 (0.2)	2.0 (0.4)	2.2 (0.3)
α (°)	-144 (9)*	-177 (12)	15 (81)	-38 (5)	-169 (15)	-113 (28)	-2 (38)	-174 (16)*	-47 (15)	118 (24)	-168 (6)*	44 (12)
β (°)	146 (9)	103 (8)	142 (11)	31 (5)	96 (8)	64 (18)	97 (38)	104 (6)	151 (7)	30 (4)	98 (4)	153 (6)
γ (°)	21 (6)	176 (6)	19 (16)	-160 (5)	176 (8)	-136 (23)	159 (10)*	176 (10)	6 (7)	-142 (14)	160 (2)*	-1 (14)
x (Å)	14.0 (2.3)	22.1 (5.4)	-3.8 (3.6)	13.0 (1.3)	23.8 (3.1)	14.9 (2.7)	-2.3 (6.2)	24.9 (0.4)	-1.6 (2.2)	11.2 (1.7)	25 (0.4)	-4.9 (1.3)
y (Å)	-1.6 (2.6)	-9.0 (4.3)	-4.4 (2.7)	-1.0 (1.4)	-8.8 (3.1)	-2.4 (2.6)	-4.5 (6.5)	-7.4 (1.3)	-5.8 (1.4)	-1.6 (1.1)	-6.6 (1.7)	-8.4 (1.7)
z (Å)	24.4 (1.9)	4.2 (2.9)	0.1 (1.6)	24.5 (1.9)	6.3 (2.7)	22.3 (3.3)	-0.2 (5.1)	4.8 (1.3)	-1.4 (1.0)	25.6 (0.7)	5.6 (1.1)	-0.2 (1.0)
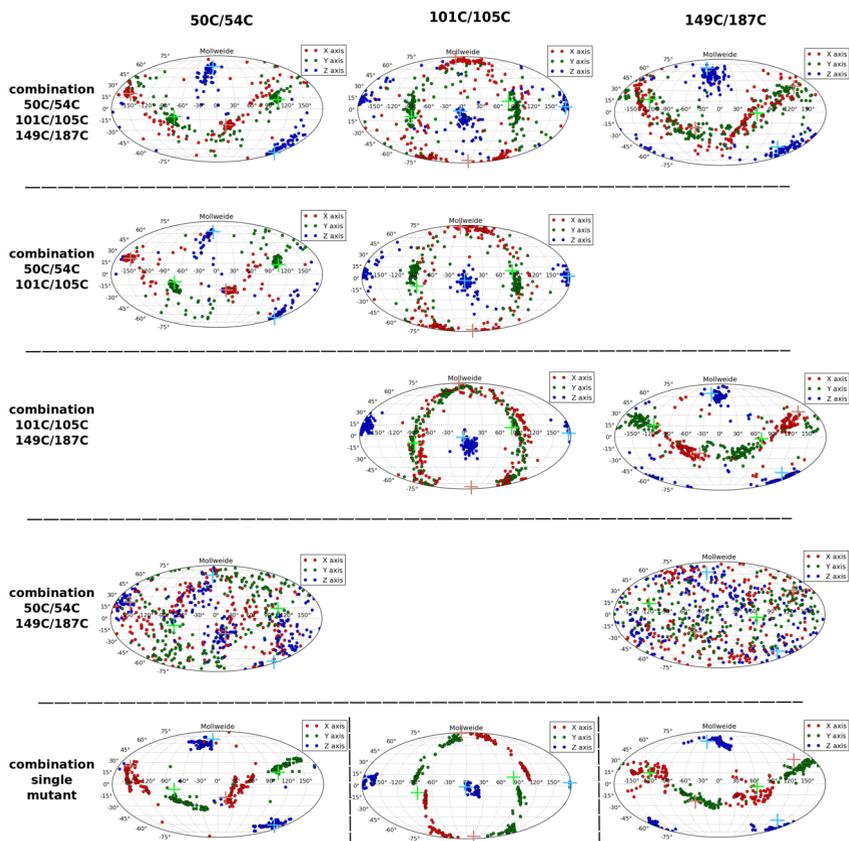
Figure 2.11. Mollweide sinusoidal representation of the tensor axes refined with PARAssign for all 7 combinations. The light crosses are the axis coordinates of the tensors refined with amide PCS, Table S1. A clear correlation can be observed between the number of highly reliable assignments for a given correlation and the quality of the tensor refinement.
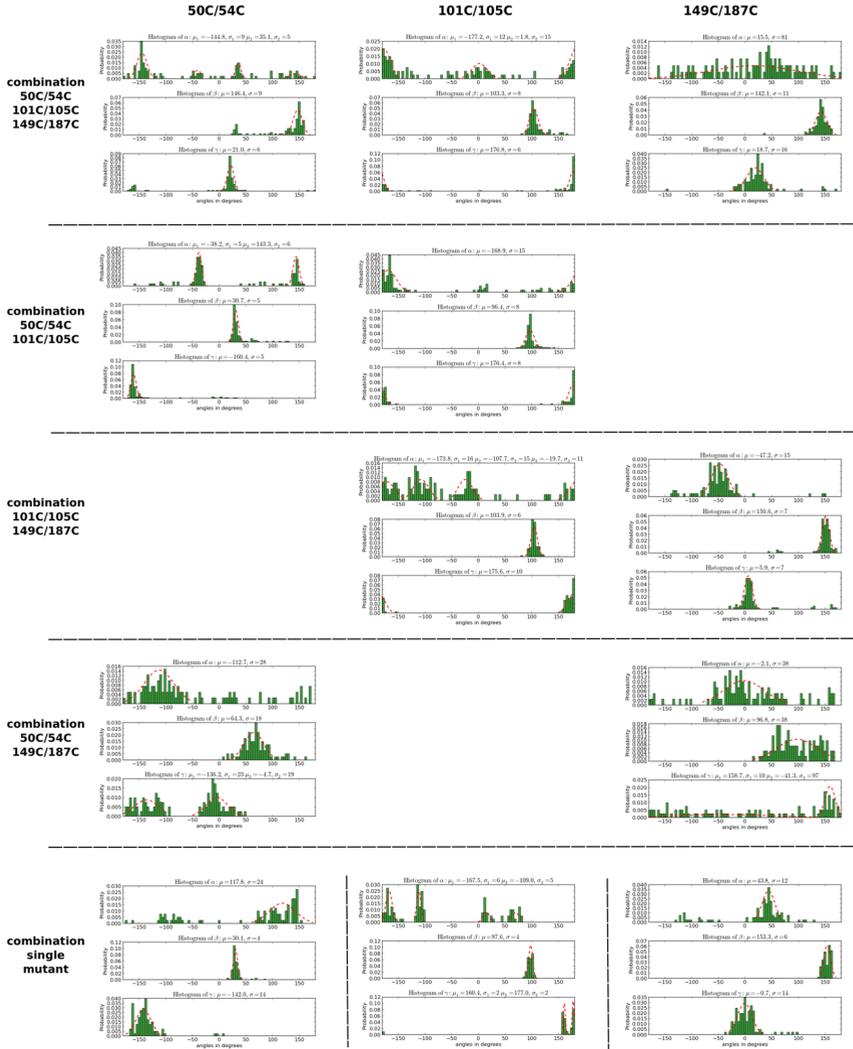
Figure 2.12. Distribution of Δχ Euler angles (convention z-y-z) for the 7 combinations. For all angle sets, the top, middle and bottom subplots are the histograms of α, β and γ, respectively. It should be noted that the zyz Euler representation $(\alpha, \beta, \gamma) \setminus \beta \in [0, 180[$ is equivalent to $(\alpha+180, -\beta, -\gamma) \setminus \beta \in [-180, 0[$ zyz Euler representation. Additionally, the xz and yz planar symmetries of Δχ lead to equivalence by 180° rotation.

In the results based on three datasets, all of the 27 highly reliable assignments were correct and 9.5% of the yellow assignments contained the correct assignment (1 to 4 possibilities). The assignment procedure was also performed with the input data of only two of the three paramagnetic centers, **Figure 2.13**. In each case, all the highly reliable (green) assignments were correct. Interestingly, however, the number of highly reliable assignments was very different for the three different combinations of paramagnetic

centers. With the combination (50C/54C-101C/105C), 46 methyl groups were assigned with high reliability, whereas only 9 assignments were highly reliable for the combination (101C/105C-149C/187C) and none for the combination (50C/54C-149C/187C). Furthermore, this combination yielded very few reliable (yellow) assignments. The much smaller input data set for mutant 149C/187C partly explains why fewer reliable assignments were obtained for combinations with this mutant. However, the large difference in assignment quality between the two combinations that contain the 149C/187C data must have another reason. We ascribe this to the relative orientations of the $\Delta\chi$-tensors, **Figure 2.13**.
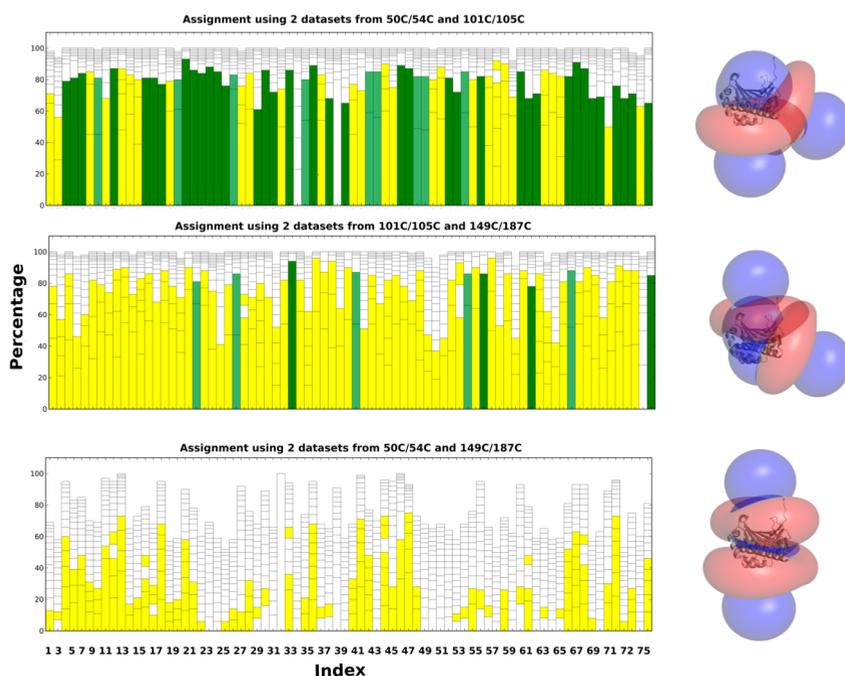
Figure. 2.13 PARAssign output for the three combinations of two of the experimental data sets. Labels were removed for clarity. On the right the corresponding relative orientations of the two CLaNP-5 tensors are shown.

For parallel $\Delta\chi$-tensors the assignment quality is poorer than for orthogonal $\Delta\chi$-tensors, which was confirmed by a run of PARAssign with complete, simulated datasets for which the combination 50C/54C-149C/187C yielded many fewer highly reliable assignments than the two other combinations, **Table 2.2**. The assignment runs were also performed using data from only one paramagnetic center. As expected, the number of correct assignments drops dramatically. Nevertheless, PARAssign identified 1 to 6 highly reliable assignments for the two mutants 50C/54C and 101C/105C, which were all correct. These assignment calculations were performed using a PRE cutoff of 9 Å, **equation 2.1**, to prevent methyl groups for which the PRE leads to extensive broadening, being considered for assignment. The 149C/187C mutant did not yield

correct assignments because of the limited input set due to low quality of the spectra. Therefore, when using **PARAssign** with only one dataset, the quality of the input becomes essential to avoid false positive assignments. From this study, we conclude that with a dataset containing less than 60% of the expected PCS, PARAssign fails to give highly reliable assignments that are all correct. The highly reliably assigned nuclei from the two mutants with good input data correspond to methyl groups that are localized close to the lanthanoid, displaying sizeable PCS, as illustrated in **Figure 2.14**. Thus, the use of a single tag could be considered for localized assignment, for instance of methyl groups in an active site.
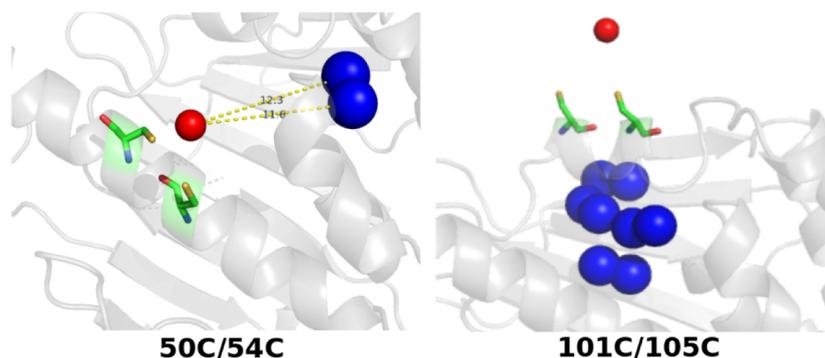
Figure. 2.14 Highly reliably assigned methyl groups, represented in blue spheres, using a single paramagnetic center for the assignment. The numbers in the 50C/54C panel indicate the distances from the lanthanoid to the methyl groups in Å.

The quality of the assignment is reflected in the accuracy and precision of the $\Delta\chi$ tensors parameters refined by PARAssign. The sinusoidal projection and histogram representation of the Euler angles of refined tensors are shown **Figures 2.11** and **2.12**, respectively. The combination (mutant **50C/54C – 101C/105C**) leading to the best assignments is also the one for which tensor orientations are the best refined. This is to be expected, because high reliability is obtained only if the same assignments are found consistently over the 100 runs. The combinations of data sets that lead to poor assignment reliability also yield a large spread in the Euler angles, indicating a poor reproducibility of tensor determination over the 100 runs. A comparison of the precision of the 50C/54C tensor calculated when using the single data set as compared to the combination with 149C/C187C shows that the former is higher. This analysis strongly suggests that the addition of a relatively poor dataset of a parallel $\Delta\chi$-tensor can worsen the assignment process.

Conclusions

The results presented in this study demonstrate possibilities and limitations for use of PCS in the assignment of protein methyl group resonances. The simulations show that, with complete input, the search problem can be solved with data from three and even from only two paramagnetic tags. The quality of the input data is vitally important. One

of the three tags yielded broader peaks and poorer spectra, resulting in a significant decrease in the number of paramagnetic peaks that could be paired with their diamagnetic counterparts. Such an incomplete PCS data set strongly affects the quality of the assignment. In our hands, CLaNP attached to two consecutive turns of an α-helix gives the best results with minimal disturbance of the protein structure. It was also found that the Δχ tensors of different tags should not be oriented along the same main axis, because this results in correlated PCS and reduced assignment quality. These factors can easily be taken into account when designing the tag positions. With two-armed tags such as CLaNP-5, the lanthanoid location and approximate directions of the axes of the Δχ tensor can be estimated well. It is then possible to predict the PCS for different tags and use these in PARAssign to predict whether complete assignment will be achieved if good spectra can be acquired.

PARAssign version 2 runs multiple times with a small part of the input data removed to yield more robust assignments. It was demonstrated that, with data from two or three independent tags, assignments classified as highly reliable are correct and that in those classified as 'yellow', the correct assignment is nearly always among the few suggested possibilities. Such information could be combined with data from NOE, PRE or mutagenesis experiments to find a unique assignment.

The critical point in obtaining complete PCS input sets is the pairing of peaks in the paramagnetic and diamagnetic spectra. Clearly, the larger the number of peaks in a spectrum, the more complicated this process is, in particular for methyl resonances, which show little dispersion. For larger systems, preparing several samples with only one type of amino acid labeled will reduce the overlap [116-118]. Labeling a single methyl group per amino acid also helps [29]. All the PCS can then be combined in PARAssign, with the amino acid types indicated. Another approach to assist the pairing of peaks is to prepare a sample with the same tag containing a weakly paramagnetic lanthanoid, such as Eu^{3+}. The Δχ tensors of different lanthanoids in the same tag are oriented to good approximation in the same direction, so that a weaker lanthanoid results in a resonance on the correlation line between the diamagnetic and paramagnetic peak, at a position determined by the ratio of the sizes of the Δχ tensors of the stronger and weaker lanthanoids. With Eu^{3+} the PCS are small and more readily attributable to the corresponding diamagnetic peak.

If some assignments are already known, these can be introduced in PARAssign to improve the assignment results. To confirm the correctness of the available assignments of the methyl resonances in ntd-HSP90, we mutated L45, L80 and L89 one by one to Ile. When these assignments were fixed in the PARAssign calculations, the number of highly reliable assignments improved by up to 15% for the best combination.

A single tag cannot be used to obtain an overall assignment, because of the symmetric nature of the paramagnetic effect, in particular for axial tensors. However, in

combination with the protein structure, resonances of nearby methyl groups can still be assigned. This approach can be helpful to highlight a certain region of a protein, for example the active site in an enzyme. The assigned nuclei can then be used to report on dynamics or structural changes that occur in its proximity.