Ligand-induced structural transitions combined with paramagnetic ions facilitate unambiguous NMR assignments of methyl groups in large proteins

Lars Mühlberg1 · Tuncay Alarcin1 · Thorben Maass1 · Robert Creutznacher1 · Richard Küchler1 · Alvaro Mallagaray1

Received: 6 December 2021 / Accepted: 24 March 2022 / Published online: 10 April 2022
© The Author(s) 2022

Abstract
NMR spectroscopy allows the study of biomolecules in close-to-native conditions. Structural information can be inferred from the NMR spectra when an assignment is available. Protein assignment is usually a time-consuming task, being specially challenging in the case of large, supramolecular systems. Here, we present an extension of existing state-of-the-art strategies for methyl group assignment that partially overcomes signal overlapping and other difficulties associated to isolated methyl groups. Our approach exploits the ability of proteins to populate two or more conformational states, allowing for unique NOE restraints in each protein conformer. The method is compatible with automated assignment algorithms, granting assignments beyond the limits of a single protein state. The approach also benefits from long-range structural restraints obtained from metal-induced pseudocontact shifts (PCS) and paramagnetic relaxation enhancements (PREs). We illustrate the method with the complete assignment of the 199 methyl groups of a MILproSVproSAT methyl-labeled sample of the UDP-glucose pyrophosphorylase enzyme from Leishmania major (LmUGP). Protozoan parasites of the genus Leishmania causes Leishmaniasis, a neglected disease affecting over 12 million people worldwide. LmUGP is responsible for the de novo biosynthesis of uridine diphosphate-glucose, a precursor in the biosynthesis of the dense surface glycocalyx involved in parasite survival and infectivity. NMR experiments with LmUGP and related enzymes have the potential to unravel new insights in the host resistance mechanisms used by Leishmania major. Our efforts will help in the development of selective and efficient drugs against Leishmania.

Keywords MILVAT labelling · 4D HMQC-NOESY-HMQC · Paramagnetic NMR · UDP-glucose pyrophosphorylase from Leishmania major · Methyl-TROSY · Complete assignment

Introduction
NMR spectroscopy enables the study of the structure, dynamics and interactions of biomolecules in close-to-native conditions (Williamson 2013; Baldwin and Kay 2009; Bax and Clore 2019; Barrett et al. 2013). Solution NMR has classically made used of 15N isotopic labeling of amido groups in the protein backbone. However, the assignment of protein backbone resonances becomes challenging in large, supramolecular systems. The advent of 1H–13C methyl-transverse relaxation optimized spectroscopy (methyl TROSY) techniques (Tugarinov et al. 2003) offered a new avenue to overcome previous limitations, allowing solution NMR with protein ensembles as large as 1 MDa (Ruschak and Kay 2012; Rosenzweig and Kay 2014; Mas et al. 2018; Tugarinov et al. 2007; Kay 2011; Sprangers and Kay 2007; Gauto et al. 2019; Rosenzweig et al. 2015; Pederson et al. 2017; Shiraishi et al. 2018). Such an approach requires selective [1H,13C]-labeling of amino acid methyl groups in an otherwise highly deuterated background. Several protocols for efficient methyl labeling from E. coli, yeast, insect and mammalian cells have been developed. Broadly speaking, labeling is achieved by the addition of selectively labeled amino acid precursors (Ile, Leu, Val, Met) or by supplementation with selectively labeled amino acids (Ala, Met, Thr) (Schütz and Sprangers 2019).
Interpretation of methyl TROSY spectra benefits from a confident assignment of methyl group signals. When backbone assignments are available from triple-resonance experiments, pulse sequences transferring magnetization through scalar (Kay et al. 1990) or dipolar (Tugarinov and Kay 2003) couplings from backbone amides or carbonyls to methyl groups can be applied. Systematic mutagenesis provides an alternative path when a backbone signal assignment is not available (Amero et al. 2011). However, mutagenesis can be costly and time consuming, and mutations can perturb the chemical environment of the methyl groups in unforeseen ways. Therefore, this approach is currently limited to large, multi-domain enzymes, or to methyl groups which cannot be assigned by other means (Sprangers and Kay 2007; Velyvis et al. 2009). Recently, structure-based assignment strategies have emerged as a powerful alternative in the cases where high-resolution crystal structures are available. In general, such methods correlate NOE-derived distance restraints from 3D or 4D HMBC-NOESY experiments (Tugarinov et al. 2012) with structural data to produce consistent assignments of the methyl resonances. Complementary, paramagnetic relaxation enhancements (PREs) and pseudocontact shifts (PCS) can be used as a source of long-range information, validating or even expanding NOE-based assignments (Velyvis et al. 2009). Following this line of thoughts several automated algorithms have been developed for methyl group assignment (Pritisanac et al. 2020). They can be classified in three categories: Those that exclusively rely on inter-methyl NOEs [(MAGIC (Monneau et al. 2017), MAGMA (Pritisanac et al. 2017), MAUS (Nerli et al. 2021) and MethylFLYA (Pritisanac et al. 2019)], those which solely depend on PCS [Possum (John et al. 2007) and PARAssign (Lescanne et al. 2017; Skinner et al. 2013)] and mixed approaches [MAP-XSII (Xu and Matthews 2013), PRE-ASSIGN (Venditti et al. 2011) and FLAMEnGO2.0 (Chao et al. 2014)].

It has been recently shown that high probe density labeling schemes facilitate unambiguous NMR assignments of methyl groups (Proudfoot et al. 2016). Simultaneous labeling of all methyl-containing amino acids offers the highest probe density, but at the price of peak overlapping in large protein complexes. This problem can be partially overcome by comparing spectra of two different protein states (i.e., apo and holo in enzymes). The presence of a ligand bound to the protein alters the chemical environment of the surrounding atoms, inducing chemical shift perturbations (CSPs) (Williamson 2013) CSPs are likely more pronounced if ligand binding stabilizes an alternative protein conformational state. In this scenario, average distances between certain protein methyl groups differ between conformers. For instance, spatially isolated methyl groups in the one state may be part of densely populated protein parts in the other state, allowing the assignment of otherwise isolated methyl groups.

Here we demonstrate how ligand-induced changes of protein states can be exploited to expand currently available strategies for structural-based assignment of methyl groups. Our approach has three prerequisites: (1) Apo and bound states exist in fast or close-to-intermediate exchange in the NMR time scale, allowing transfer of the assignment between conformations. This can often be observed for protein–ligand interactions with dissociation constants $K_{D}$ in the µM to mM range. (2) High-resolution crystal structures are available for both protein states. (3) Short- and/or long-range NMR structural information is available or can be extracted for both protein states (i.e. 3D or 4D NOESY-HMQC spectra, PCS, PREs,...).

In this work we have followed a structure-based assignment exploiting multiple ligand-induced protein states to obtain a complete assignment of Ala-β, Ile-γ1, Leu-δ2, Val-γ2, Met-ε and Thr-γ methyl [13C,1H3]-methyl-labeled enzyme UDP-glucose pyrophosphorylase from Leishmania major (MILproSYproXSAT LmUGP). L. major causes severe diseases in humans and animals, with symptoms ranging from self-healing cutaneous lesions to fatal visceral forms (WHO 1984). The enzyme plays a central role in the life cycle of L. major, as it catalyzes the reversible conversion of uridine-5′-triphosphate (UTP) and glucose 1-phosphate (Glc-1-P) into uridine diphosphate-glucose (UDP-Glc) and inorganic pyrophosphate (PPi) in the presence of Mg$^{2+}$ (Fig. 1) (Lamerz et al. 2006). UDP-Glc can be converted into UDP-galactose, which is used in the biosynthesis of a dense layer of glycoconjugates covering the parasite. Depletion of UDP-galactose pools is associated with parasite growth arrest and cell death in vitro, rendering LmUGP as an attractive target for drug development (Damerow 2015).

Methyl-labeled LmUGP exists as a soluble monomeric protein with a molecular weight of 62 kDa. The enzyme follows a sequential bifacial catalytic mechanism, the binding of UTP or UDP-Glc being the first step in the forward and backward direction. Fig. 1 Catalytic cycle postulated for LmUGP. Protein states used for the assignment are apo, UTP:Mg$^{2+}$ and UDP-Glc bound states (grey, green and blue, respectively).
catalytic reactions, respectively. Crystallographic studies and molecular dynamic (MD) simulations have shown that LmUGP undergoes significant conformational changes along the catalytic cycle (Steiner et al. 2007; Führing et al. 2013). Therefore, LmUGP constitutes an excellent model for the use of multiple protein states in the assignment of systems containing a high probe density.

Material and methods

Synthesis and purification of methyl labeled LmUGP

The gene encoding for UDP-glycopyrophosphorylase from L. major (EC 2.7.7.9) was subcloned into the pET-22b (Novagen) expression vector as described previously (Lamerz et al. 2006) and transformed into Escherichia coli BL21(DE3). Single-point mutants T96S, A145G, T172S, A183G, T226S, A291G, A345G, V413I, A419G, A454G, A470G, T492S and M495I were generated by site-directed mutagenesis (Eurofins Genomics), and confirmed by DNA sequencing (Table S1). Primers used for mutagenesis are listed in Table S1. Mutants were expressed and isotopically labeled as described below according to the substituted amino acid type. Prior purification, pellets were combined as follows: batch 1: A419G, V413I and M495I; batch 2: A291G and T172S; batch 3: A454G and T96S; batch 4: A470G and T226S; batch 5: A345G and T492S. A145G and A183G were purified separately. This approach delivers a single labeling pattern for each mutant in each sample, minimizing the number of protein purifications required.

\[
[U^{15}N, ^2H, {^{13}C, ^1H_3}]-methyl labeled UDP-glycopyrophosphorylase from LmUGP was expressed following an adapted version from previously reported protocols (Proudfoot et al. 2016; Muller-Hermes et al. 2020). Briefly, E. coli BL21(DE3) containing the gene for LmUGP was grown in 20 ml of LB Lennox medium (Roth) until an optical density of 600 nm (OD600) > 2 was reached. Ampicillin (100 µg/mL) was used as selecting agent through the expression. Unless otherwise stated, bacteria were grown at 37 °C under shaking (220 rpm). Cells for inoculation of 10 mL M9+/D2O minimal medium with a starting OD600 of 0.1 were harvested by centrifugation, and excess of TB medium was removed. This starter culture allows E. coli to acclimate to D2O, and was grown overnight. In all M9+/D2O minimal medium, 3 g/L of 15 N-amonium chloride (Deuto) and 3 g/L of deuterated 13C-glucose (1,2,3,4,5,6-d6, Deuto) were used as the principal nitrogen and carbon sources, respectively. Detailed recipes for culture media used in this study can be found in the supplementary material. The next morning 2 ml of the starter culture were spin-down, supernatant was removed and cells were transferred into 20 mL of freshly prepared M9+/D2O minimal medium. When an OD600 of 0.4 was reached, the culture volume was increased to 90 ml and cells were grown until an OD600 of 0.6–0.8 was reached. At this point, the temperature of the incubator was reduced to 16 °C and 10 mL of M9+/D2O minimal medium containing the desired labeled precursors and amino acids were added (Table S2). Protein expression was induced after 1 h using 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Cells were harvested by centrifugation when the maximal cell density was reached (OD600 3.8–4.8) and stored at -20 °C.

LmUGP was purified as described previously (Lamerz et al. 2006) with the following modifications: Bacterial pellet containing the overexpressed LmUGP was resuspended in purification buffer (20 mM Tris pH 7.8, 300 mM NaCl and 5 mM 2-mercaptoethanol, Sigma-Aldrich) containing 5 mM imidazole. Enzyme inhibitors aprotinin and leupeptin (4 µg/g wet pellet each, Roth), lysozyme (0.25 mg/g wet pellet, Merck) and benzozyme (2.5 U/g wet pellet, Novagen) were added to the suspension and cells were lysed using a Microfluidizer® (12,000 psi). Cell lysate was centrifuged at 5,000 g for 1 h, and the soluble lysate was passed through a 5 mL Ni–NTA agarose column (GE Healthcare). The column was then washed with 50 mL purification buffer containing 40 mM imidazole concentration, and LmUGP was eluted using 20 mL of purification buffer at 300 mM imidazole concentration. Fractions showing UV absorption at 280 nm were pooled together and loaded into a HiLoad 16/600 Superdex 200 pg size exclusion column (GE Healthcare). Purified LmUGP was eluted in purification buffer, the elution fractions were combined and LmUGP was concentrated with Amicon Ultra-4 Centrifugal Filter Units (Millipore, MWCO 10 kDa). Protein samples were stored at 4 °C in the presence of 2 mM tris(2-carboxyethyl)phosphine (TECP).

NMR sample preparation

Storage buffer was changed into the desired NMR buffer using 2 mL Zeba™ Spin Desalting Columns (Thermo Fisher Scientific). Buffer A was used for routine 1H-13C HMBC, methyl-methyl NOESY, TRACT experiments, titrations of uridine-5'-diphosphate-glucose (UDP-Glc, Sigma-Aldrich) and uridine-5'-triphosphate (UTP, Sigma-Aldrich) over MILproSYproSAT LmUGP and contained: 20 mM Tris-d11 (Eurisotop) pH* 7.20, 75 mM NaCl, 2 mM TCEP-d16 (CIL), 0.1 mM 2,2-Dimethyl-2-silapentane-5-sulfonate-d6 (DSS-d6, Sigma-Aldrich) and 0.2 mM imidazole in D2O (Eurisotop, 99.96%). Measurements of metal-induced PRE and PCS, and titrations of divalent and trivalent metals over UTP were conducted in buffer B, which was composed of 20 mM Bis-Tris-d19 (Sigma-Aldrich) pH* 7.06, 75 mM NaCl, 2 mM TCEP-d16 and 0.1 mM DSS-d6 in D2O. Protein concentrations were determined after buffer exchange by UV absorbance at 280 nm with ε = 42,860 M⁻¹ cm⁻¹. The
extinction coefficient ε was obtained via absolute concentration measurements of a LmUGP sample using amino acid analysis. A complete list of samples prepared can be found in Table S3.

Dissociation constants $K_D$ of UTP in the absence and in the presence of saturating concentrations of metals (10 mM MgCl$_2$ or 5 mM Lanthanoids) were inferred from 1H,13C HMQC spectra using samples containing 80–110 μM concentration of MIL$_{\text{proSVproSAT}}$ LmUGP. $K_D$s for the coordination of divalent and trivalent metals to UTP were obtained from series of 1H NMR spectra at metal concentrations ranging from 0 to 1400 μM. UTP concentration was 239.2 μM for the titration of MgCl$_2$ and LaCl$_3$, and 250 μM for the titration of LuCl$_3$, EuCl$_3$ and TbCl$_3$.

Samples for the measurement of PREs contained 199 μM MIL$_{\text{proSVproSAT}}$ LmUGP, 1.5 mM UTP and 55 μM of either MgCl$_2$ or MnCl$_2$. PCS were measured at metal concentrations ranging from 0 to 1400 μM. UTP concentration was 239.2 μM for the titration of MgCl$_2$ and LaCl$_3$, and 250 μM for the titration of LuCl$_3$, EuCl$_3$ and CeCl$_3$.

NMR experiments

All NMR samples were prepared in 3 mm NMR tubes at a final volume of 0.16–0.17 mL. NMR experiments were conducted at 293 K. Spectra were processed with Topspin 4.0.6 (Bruker) and analyzed using CCPNMR Analysis 2.4.2 software suit (Vranken et al. 2005). 1H chemical shifts were referenced to the DSS-d$_6$ peak, and 13C signals were referenced as diamagnetic references, and TbCl$_3$, TmCl$_3$, EuCl$_3$, and CeCl$_3$ as paramagnetic samples. In all the cases, ligands and metals were dissolved at high concentrations in the aforementioned buffers and the pH$_6$ was carefully adjusted to minimize dilution effects and pH-artifacts. LuCl$_3$, LaCl$_3$, TbCl$_3$, TmCl$_3$, EuCl$_3$, and CeCl$_3$ were obtained from Sigma-Aldrich. CeCl$_3$ was purchased from Alfa Aesar, whilst MgCl$_2$ and MnCl$_2$ were acquired from Merck.

4D HMQC-NOESY-HMQC experiments (Tugarinov et al. 2005) of the apo and UDP-Glc bound conformation of LmUGP were acquired with a sample containing 450 μM MIL$_{\text{proSVproSAT}}$ methyl-labeled LmUGP and 5 mM MgCl$_2$ before and after the addition of 12 mM UDP-Glc, respectively. Both experiments were measured at the Utrecht NMR SONNMRLSF facility on a 900 MHz or on a 950 MHz Bruker NMR machine equipped with a cryoprobe, respectively. For both acquisitions the mixing time was set to 180 ms. Apo conformation was measured using 30% non-uniform sampling (Robson et al. 2019) (NUS) according to a Poisson Gap sampling schedule (Hyberts et al. 2010) with 11,466 complex NUS data points in a grid of 70(13C) × 84(1H) × 52(13C) points in the indirect dimensions. The UDP-Glc bound conformation was acquired with 31.36% NUS Poisson Gap sampling schedule with 10,764 complex NUS data points in a grid of 66(13C) × 80(1H) × 52(13C) points in the indirect dimensions. For both experiments 512 points were acquired in the direct dimension, with 4 transients and a recovery delay of 1 s. Spectra were processed on a Mac-BookPro running Yosemite 10.10.5 using recursive Multi-Dimensional Decomposition (MMD, Bruker).

Transverse 1H$_M$-Γ$_2$ PRE relaxation rates were measured at 600 MHz using a 2D 1H,13C HMQC-based pulse scheme described elsewhere (Venditti et al. 2011). Seven relaxation delays (t = 0, 6, 15, 25, 36, 50 and 100 ms) were acquired in an interleaved manner with 102 ms acquisition time and spectral window of 4.2 ppm in the direct dimension. In the indirect dimension, the spectral window was set to 19 ppm with 512 increments. The recovery delay was 1.5 s, and 4 scans were acquired per experiment. Measurements were repeated twice. Data was apodized with a QSINE window function, and FIDs were zero-filled prior to Fourier-transformation to give a 2048 × 2048 data matrix. Spectra were manually phased, and decays in cross-peak intensity were fitted to an exponential decay model (Iwahara et al. 2007) using an in-house Matlab R2019b script. Only peaks showing an intensity ≥ 3 σ of spectral noise-floor at t = 6 ms were selected for fitting. Experimental random errors in $R_T$ rates were estimated as one standard deviation from a Monte Carlo simulation (Kamath and Shriver 1989) with 1,000 iterations of single exponential fits with the spectral noise-floor taken as an estimate of random uncertainties in peak intensities. Experimental $Γ_2$ were calculated as the difference in transverse relaxation rates between the sample containing...
MnCl₂ (R_{2,\text{para}}) and the sample loaded with MgCl₂ (R_{2,\text{dia}}) according to Eq. (1):

\[ \Gamma_2 = R_{2,\text{para}} - R_{2,\text{dia}} \]  

(1)

PCS were measured at 500 MHz from 2D \(^1\)H,\(^{13}\)C HMQC spectra as explained above using a spectral window of 3.7 \times 19 ppm and 1024 \times 256 increments in the direct and indirect dimensions, respectively. The relaxation delay was set to 1.5 s and 4 transients were acquired. Experimental PCS were calculated in Hz as the difference in chemical shifts in the proton dimension between the diamagnetic and the paramagnetic sample according to Eq. (2):

\[ \delta_{\text{PCS}} = \delta(1\text{H})_{\text{para}} - \delta(1\text{H})_{\text{dia}} \]  

(2)

Rotational correlation times \( \tau_c \) of \([U-^{15}\text{N},^{2}\text{H}] \) labeled LmUGP at 50 to 335 \( \mu \)M protein concentrations have been estimated using TRACT experiments (Lee et al. 2006) at 500 and 600 MHz. The relaxation delay was set to 2 s. Experiments were measured for 16 transients with 25 increasing delays of up to 0.4 s. Data were integrated from 8–10 ppm, normalized and fitted to an exponential decay model for determination of average \(^{15}\text{N} R_\alpha \) and \( R_\beta \). Samples contained 5 mM MgCl₂.

### Calculation of \( \Gamma_2 \) PREs

The contribution to PRE from Curie spin relaxation of Mn⁴⁺ at 600 MHz can be neglected, since it accounts only for 0.14\% on the total relaxation enhancement in \(^1\)H. Therefore, in this case PRE rate arising from the dipole–dipole interaction between a nucleus and an unpaired electrons (\( \Gamma_2 \)) is described by the Solomon-Bloembergen-Morgan (SBM) Eq. (3) (Bloembergen and Morgan 1961).

\[
\Gamma_2 = \frac{1}{15} \left( \frac{\mu_0}{4\pi} \right)^2 \gamma_r^2 g \mu_B^2 S(S+1) \left( \frac{4 \tau_c + \frac{13 \tau_c}{1+\omega g_2^2 \tau_c^2} + \frac{3 \tau_c}{1+\omega g_3^2 \tau_c^2} }{ \frac{S(S+1)}{ \mu_B^2 g_2^2 \mu_0^2 } } \right)
\]  

(3)

where \( r \) is the distance between the paramagnetic centre and the observed nucleus; \( \mu_0 \) is the permeability of vacuum; \( \gamma_r \) the nuclear gyromagnetic ratio; \( g \) the electron g-factor; \( \mu_B \) the electron Bohr magneton; \( S \) the electron spin quantum number; \( \omega_I \) and \( \omega_{S_e} \) the proton and electron Larmor frequencies, respectively; \( \tau_c \), the PRE correlation time defined as \( \tau_c^{-1} = \tau_{r}^{-1} + \tau_{s}^{-1} \); with \( \tau_{r} \), the electron relaxation time, which has been reported to be 9.6 ns for Mn⁴⁺ at 600 MHz (Iwahara et al. 2007). Calculated \( \Gamma_2 \) were multiplied by 0.0241 to be comparable with the measured values, as explained in Results and Discussion. No crystal structure is available for LmUGP in complex with UTP and a metal ion. Therefore, we used the following crystal structures of LmUGP as stand alone or as an ensemble: Apo (pdb 2OEF), UDP-Glc (pdb 4M2A), and in complex with UTP analog dUppCp (pdb 4M28). Coordinates of the metal position were fitted via a non-linear gradient descent using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm (Fletcher 1988) for non-linear least-square minimisation of the cost function described in Eq. (4) (Orton et al. 2020).

\[
\text{cost}_{\text{ensemble}} = \sum_i \left[ \frac{(\Gamma_{\text{obs}} - \Gamma_{\text{calc}})}{S_{\text{Γ}_2,i}} \right]^2
\]  

(4)

where \( \Gamma_{\text{obs}} \) and \( \Gamma_{\text{calc}} \) are the observed and calculated PREs, respectively. Index \( m \) is for atoms that are common between models, index \( i \) runs over every atom in the structure and \( S_{\text{Γ}_2,i} \) corresponds to the experimental uncertainty in the \( \Gamma_{m,i} \) of spin \( i \). Error in the fitting of the metal position contributes to one standard deviation.

### Fitting of alignment tensors from PCS data

Anisotropic magnetic susceptibility tensors \( \Delta \chi \) induced by lanthanoid ions were obtained by fitting of experimental PCS to Eq. (5) as described in Paramagpy software package (Orton et al. 2020).

\[
\delta_{\text{PCS}} = \frac{1}{4\pi} \left[ x^2 - y^2, \ y^2 - z^2, \ 2xy, \ 2xz, \ 2yz \right] \cdot \left[ \begin{array}{c} \Delta \chi_{xx} \\ \Delta \chi_{yy} \\ \Delta \chi_{zz} \\ \Delta \chi_{xy} \\ \Delta \chi_{xz} \\ \Delta \chi_{yz} \end{array} \right]
\]  

(5)

Being \( x, y, z \) the coordinates of the metal centre and \( \Delta \chi_{xx}, \Delta \chi_{yy}, \Delta \chi_{zz}, \Delta \chi_{xy}, \Delta \chi_{xz}, \Delta \chi_{yz} \) five explicit parameters that characterise the \( \Delta \chi \) tensor. Same combinations of structural models as described in the calculation of \( \Gamma_2 \) were used here. Fittings included corrections for residual dipolar anisotropic shifts (RADS) and for residual anisotropic chemical shifts (RACS), as indicated in Paramagpy. The latter was achieved using standard chemical shift anisotropy (CSA) tensors for \(^{13}\)C spins (Cornilescu and Bax 2000). Quality of fitted \( \Delta \chi \) tensors was evaluated via a bootstrapping approach with 1,000 iterations and 80\% randomly sampled data at each specific iteration. Errors are reported as one standard deviation.

### Calculation of Q factors

Goodness of fitting between observed and calculated PREs and PCS was evaluated by calculating Q factors according to Eq. (6)
$$Q_{\text{ensemble}} = \sqrt{\frac{\sum_i \left( \sum_m \left[ a_{i,m}^{\text{obs}} - a_{i,m}^{\text{calc}} \right] \right)^2}{\sum_i \left( \sum_m \left| a_{i,m}^{\text{obs}} \right| \right)^2}}$$  

(6)

where index $m$ is for atoms that are common between models, index $i$ runs over every atom in the structure and $a$ represents either $1/2$ or PCS.

**Use of chemical shift data in MAP-XSII**

MAP-XSII requires chemical shift data to compute assignments, although it can easily lead to assignment errors (Pritisanac et al. 2017). To overcome this problem, chemical shifts were predicted using CH3Shift (Sahakyan et al. 2011) and SHIFTX2 (Han et al. 2011) and weighted the by a factor of $10^{-12}$, de facto excluding them from the calculations. The content of the remaining input files is explained under Results and Discussion.

**Calculation of dissociation constants**

Titrations were used to derive Euclidean chemical shift perturbances (CSPs) $\Delta v_{\text{eucl}}$ according to Eq. (7).

$$\Delta v_{\text{eucl}} = \sqrt{\Delta v_{H}^2 + \Delta v_{C}^2}$$  

(7)

$\Delta v_{H}$ and $\Delta v_C$ are CSPs in the respective dimensions in units of Hz. In a simple two states model (Eq. 8), observed $\Delta v_{\text{obs}}$ at a given total ligand concentration $L_t$ are linked to the dissociation constant $K_D$ via the law of mass action (Williamson 2013) (Eq. 9).

$$P + L = PL$$  

(8)

$$\Delta v_{\text{obs}} = \frac{(P_t + L_t + K_D) - \sqrt{(P_t + L_t + K_D)^2 - 4P_tL_t}}{2P_t} \Delta v_{\text{max}}$$  

(9)

where $P_t$ is the total protein concentration, and $\Delta v_{\text{max}}$ is the maximum CSP at ligand saturation for each signal. Titration curves for nonlinear least-squares global fitting were selected according to the magnitude of the CSPs at the highest ligand concentration: CSPs larger than the mean of all CSPs + two standard deviations were used to derive a $K_D$ value. Global fittings were performed using in-house Matlab scripts. Errors were determined from a Monte Carlo approach with 1,000 iterations as previously described (Arai et al. 2012), and are given as one standard deviation.

**Results and discussion**

The overall workflow can be divided into three sequential steps, as seen in Fig. 2. First, methyl-methyl NOEs were measured from the “apo” (pdb 2OEF) and “UDP-Glc” bound (pdb 4M2A) LmUGP states. These two states exhibit the largest RMSD (3.4 Å) among all available crystal structures, and were therefore selected to maximize differences between the two sets of NOE restraints. Methyl walking delivered a 85% preliminary assignment of methyl groups, which was used as starting point for refinement of the assignment based on paramagnetic NMR. Apo LmUGP does not specifically interact with di- and trivalent ions in solution. However, metal ions can occupy the enzyme binding pocket when in complex with UTP, as it is shown for Mg$^{2+}$ as a representative example in Figure S1. PCS and PREs were measured using samples at specific UTP and metal ratios. Experiments with paramagnetic metals expanded the assignment to a total 94%, confirming the assignments allocated from methyl-methyl NOE experiments. Finally, unassigned methyl groups were mutated to complete the assignment.

**Identification of residue types**

The first step in any structure-based assignment is the correlation of the methyl group resonances with their respective amino acid types. Here, Ala-β, Ile-γ1, Leu-62, Val-γ2, Met-ε and Thr-γ methyl groups were [13C,1H3]-methyl labeled, yielding a total of 199 [13C,1H3]-methyl groups to be assigned. The labeling scheme was chosen because it combines maximal methyl probe density with minimal signal crowding. As can be seen in Fig. 3a, b, labeled methyl groups are evenly distributed over the whole enzyme. Analysis of peak intensities from a single 1H-13C HMQC spectrum suggested a uniform incorporation of isotopes into all six amino acid types (Fig. S2).

To identify the amino acid types we prepared six methyl labeled samples, according to the following labeling schemes: M-, I-, LproS,VproS-, VproS-, A- and MILproS,proS,AT-methyl labeled LmUGP. Assignment of Ala-, Ile-, Met- and Thr-methyl groups was straightforward from a simple comparison between spectra. Discrimination between Leu- and Val-methyl resonances was achieved based on leucine selective unlabeling (Mas et al. 2013). For the apo state we found 196 out of 199 expected resonances distributed as follows: 43/43 alanine, 16/16 isoleucine, 48/50 leucine, 19/19 methionine, 30/30 threonine and 40/41 valine resonances (Fig. 3c). Addition of saturating concentrations of UDP-Glc resolved signal overlapping, allowing the observation of 199 out of 199 expected resonances.
Fig. 2 Overall strategy for the structure-based assignment of MILproSVproSAT LmUGP using multiple protein states. Central to this strategy was the acquisition of 4D methyl-methyl NOESY experiments, which delivered information about neighbouring methyl groups in the apo and UDP-Glc bound protein states (grey and blue, respectively). In a first step, data from each protein state was independently loaded into MAP-XSII. Comparison of independent assignments from each state revealed a 62% overlapping. As expected, the apo state delivered NOE cross-peaks not observed in the UDP-Glc bound state, and vice versa. Analysis of such extra NOE peaks allowed ca. 8% extra assignments exclusively for each state, as shown in the upper right insert. Inclusion of the extra assignments from the complementary state as fixed in MAP-XSII calculations enlarged the assignment to a total 85%. PCSs of the 85% assigned residues were then used to derive initial, approximate Δγ tensors for each lanthamide ion with Paramagpy. Addition of experimental and calculated PCSs from all observed residues into MAP-XSII expanded the assignment to 93% resonances. PREs allowed the assignment of only 2 methyl resonances (94%), and were used as a validation tool. Finally, directed mutagenesis of the remaining 12 unassigned amino acids delivered the assignment of all MILproSVproSAT methyl-group resonances.

Preliminary assignment of methyl groups based on 4D HMQC-NOESY-HMQC experiments

Crystallographic studies have shown that LmUGP undergoes a large conformational transition upon binding of UDP-Glc (Steiner et al. 2007) According to available crystal structures (pdbs 2OEF and 4M2A), conformational changes involve rearrangement of the C-terminal domain together with functional loops of the catalytic domain, forcing the enzyme to adopt a more “compacted” shape (Fig. 4a). The structural rearrangement is also apparent in $^1$H,$^13$C HMQC spectra, where a titration of MILproSVproSAT LmUGP with UDP-Glc produced observable CSPs for virtually every methyl group (Fig. 4b). Based on these structural differences, it can be postulated that 4D HMQC-NOESY-HMQC experiments acquired from the apo and UDP-Glc bound protein states may deliver a subset of unique methyl-methyl NOEs. Such unique structural restraints can be exploited to expand the completeness of the methyl assignments beyond the limits of a single protein conformation. Assignments can be easily transferred between apo and UDP-Glc bound states from simple ligand titrations, since both protein states interconvert in fast or close-to-intermedium exchange in the NMR time scale.

4D HMQC-NOESY-HMQC experiments provided short-range structural information through NOE connectivities between neighbouring methyl groups. Methyl-methyl NOEs were analyzed from F1($^13$C)/F2($^1$H) planes as previously described (Flugge and Peters 2018). NOE patterns showing mutual connectivities were grouped into clusters allowing univocal assignment of methyl groups based on distance information from the crystal structure, in a process usually described as “methyl walk”. A comparative example of such “methyl walk” for the same region of the apo and UDP-Glc states of LmUGP can be found in Fig. 5a. As expected, each protein state shows a different pattern of NOE correlations in accordance to the distances observed in the crystal structures. Note that some amino acids could only be assigned in a specific protein state (e.g. Leu91 after saturation with UDP-Glc).

4D NOESY experiments afforded a total of 486 and 531 methyl-methyl NOEs in the absence and presence of UDP-Glc, respectively. From these NOE cross peaks, 189 NOE connections were only found in the apo state, whereas 234 NOE connections were unique to the UDP-Glc state. The higher number of methyl-methyl NOEs in the bound state as compared to the apo correlates with a more compacted, closed enzyme structure, as observed in the crystal structures. It is worth noting that differences in relaxation rates between protein states associated to local protein dynamics or the existence of invisible states could also contribute to the observed differences. Two main mechanisms contributed to the unique NOE cross-peaks observed in each protein state. Firstly, structural reorientations of specific protein
distance spatial restraints (NOEs and paramagnetic NMR, because it allows the simultaneous use of short- and long-assignments of each enzyme state. We selected between methyl signals (Fig. 5c).

The ligand in its binding pocket allowed the discrimination changes in the chemical environment due to the presence of does not alter the distance between methyl groups. However, A14 superimpose in the apo state. Addition of UDP-Glc overlapping signals. For example, resonances from A10 and L85, whose methyl moiety approaches T87 upon UDP-Glc the observation of new NOE cross-peaks. A good example is motifs alter the distance between methyl groups, allowing the observation of new NOE cross-peaks. A good example is L85, whose methyl moiety approaches T87 upon UDP-Glc addition permitting its unequivocal assignment (Fig. 5b).

Secondly, changes in chemical environment in methyl requires short computation times (Pritišanac et al. 2020). It is also used, although were not systematically explored in this study. MAP-XSII uses multiple parallel repetitions based on a Metropolis Monte Carlo (MMC) swapping routine to perform automated assignments according to the experimental restraints supplied. In a first round, methyl resonances from each enzymatic state were assigned independently. Therefore, NOE connectivities and structural models (pdb entries 2OEF for apo and 4M2A from UDP-Glc states) corresponding to each protein state together with amino acid residue types were used as inputs for the assignment in two separated runs. It should be mentioned that all available LmUGP crystal structures show two regions with poor

Fig. 3 Distribution of methyl groups in LmUGP and residue type identification. a Crystal structure of LmUGP in the apo state (pdb 2OEF). The carbons of the methyl groups are highlighted as spheres. Only the pro-(-S) methyl groups of Leu and Val residues are shown. A summary of amino acids labeled and the colour scheme used is provided in the insert. b Amino acid sequence of LmUGP showing the distribution of methyl-labeled amino acids. Colour code like in (a). c Superposition of 1H,13C HMQC spectra of apo LmUGP. Each spectrum is coloured individually (M orange, I black, LproSVproS red, VproS blue, A green and MILproSVproSAT violet). The order of the spectra in the superimposition was selected to match the colour scheme in a, b. Modelled N-terminus is indicated in red in a and b, and C-terminus with low electron density is highlighted in black in b. All samples were measured at 293 K and 500 MHz with varying protein concentrations ranging between 200 and 450 µM.

respectively). It also has a good assignment reliability and requires short computation times (Pritišanac et al. 2020). It is worth mentioning that other algorithms allowing simultaneous use of multiple restraints like PRE-ASSIGN (Venditti et al. 2011) and FLAMEnGO2.0 (Chao et al. 2014) could also be used, although were not systematically explored in this study. MAP-XSII uses multiple parallel repetitions based on a Metropolis Monte Carlo (MMC) swapping routine to perform automated assignments according to the experimental restraints supplied. In a first round, methyl resonances from each enzymatic state were assigned independently. Therefore, NOE connectivities and structural models (pdb entries 2OEF for apo and 4M2A from UDP-Glc states) corresponding to each protein state together with amino acid residue types were used as inputs for the assignment in two separated runs. It should be mentioned that all available LmUGP crystal structures show two regions with poor
electron density. These regions comprise the first seven N-terminal residues including M1 and M5, and the last 17 amino acids including T492, M495 and L498 (Fig. 3c). The web server ModLoop (Fiser and Sali 2003) was used to model structural gaps, delivering a consistent orientation for the N-terminus in ten iterative runs. The C-terminus was excluded from the final models because ModLoop produced ambiguous orientations, probably due to weak amino acid interactions.

For the initial assignment of apo and UDP-Glc bound states, 20 MMC trials were performed for each state with cut-off distances spanning between 5 and 11 Å. Each structural model together with its corresponding spatial NOE-based restraints were separately computed. Cut-offs were selected to maximize the number of signals consistently assigned to the same residue in 20 MMC runs, and corresponded to 10 and 7 Å for the apo and UDP-Glc bound states, respectively (see Fig. S3). From a total of 199 methyl groups to be assigned, 141 and 139 were consistently assigned to the same residue in the apo and UDP-Glc states, correspondingly. Notably, only 123 assignments (62% total assignment) were shared between both protein states.

As previously explained, useful spatial information is also encoded in the NOE signals producing unique assignments for each protein state. Such NOE cross-peaks can be considered as key signals, which act as bridges between otherwise isolated NOE clusters or methyl groups in each protein state. To exploit this information, we repeated the automated assignment using MAP-XSII although this time with an important caveat: assignments solely obtained in one state were included as fixed assignments during the calculations of the complementary state. Runs for each protein state produced almost identical results, expanding the assignment to a total 85% of the methyl groups (169 assignments). Subsequent manual methyl-walks of both NOE datasets confirmed the assignments obtained by MAP-XSII.

Lanthanide-induced PCS allow expansion of the assignment

No crystal structure of LmUGP in complex with UTP and Mg²⁺ is currently available. However, ab initio quantum mechanics/molecular mechanics (QM/MM) calculations have indicated that LmUGP coordinates the complex UTP:Mg²⁺ in the first step of the catalytic cycle (Führing et al. 2013). After the enzymatic reaction, Mg²⁺ is weakly hold in the binding pocket by coordination to UDP-Glc and PPI, being eventually released as a PPI:metal complex (Fig. 6a). Here, we took advantage of this feature and substituted Mg²⁺ with lanthanide ions in the presence of UTP. La³⁺ was used as diamagnetic reference for Ce³⁺ and Eu³⁺, whilst Lu³⁺ served as diamagnetic control for Tb³⁺ and Tm³⁺, in accordance with their ionic radii (D’Angelo et al. 2011). During binding, the metal ion coordinates to UTP. Due to its central location, paramagnetic relaxation enhancements (PREs) impeded the observation of a significant number of residues around the metal ion. This disadvantage could be partially compensated with the use...
of Ce$^{3+}$ and Eu$^{3+}$. Such lanthanides induce small paramagnetic effects, allowing sampling of methyl residues located close to the lanthanide ions. The use of Tb$^{3+}$ and Tm$^{3+}$ delivered significant PCS as far as 40 Å from the metal position. Magnetic anisotropy $\Delta \chi$ tensors were calculated from independent or shared metal centres for all lanthanide ions, affording in all the cases almost identical metal centre positions. PCS in the $^{13}$C dimension are prone to errors due to their lower resolution and to their larger residual chemical shift anisotropy when compared to protons. Consequently, only shifts in the $^1$H dimension were used for $\Delta \chi$ tensor refinement.
Mg\(^{2+}\) coordinates oxygens from α- and β-phosphates from UTP, establishing no further interactions with neighbouring amino acids (Führing et al. 2013). In other words, Mg\(^{2+}\) only occupies the protein metal centre as a UTP:metal complex. It was therefore crucial to ensure protein saturation with UTP. With protein concentrations of about 100 µM, this was achieved with a 35-fold ligand excess. For LmUGP, the equilibrium between free, UTP- and UTP:metal-bound protein species is in fast exchange in the NMR timescale, as it is shown for Mg\(^{2+}\) as a representative example in Fig S4. In such a case, it is possible to extract PCS and PREs when the fraction of UTP:metal:LmUGP complex matches between paramagnetic metals and their diamagnetic references. In addition, undesired paramagnetic effects could be observed in solvent-exposed methyl groups at free metal ion concentrations over ~80 µM. As a consequence, one cannot simply saturate the protein with UTP:metal complex, because a large fraction of free metal will always be present inducing unspecific paramagnetic effects. In a nutshell, accurate PCS and PREs can only be obtained when the exact fraction of UTP:metal:LmUGP (UMP) complex is known, and when the free lanthanide concentration is kept well below ~80 µM.

To obtain a precise control over the experimental set-up we developed a four-states binding model describing the interaction between UTP, a single metal ion and LmUGP (Fig. 7a). Dissociation constants \(K_{D1-3}\) were experimentally inferred or approximated from simple titrations as explained in supporting information. Binding isotherms for the calculation of \(K_{D1}\) can be inspected in Fig. S5. Fig S6 shows the binding isotherms corresponding to \(K_{D2}\) and \(K_{D3}\) for the binding of UTP to MIL\[^{proSVproS}\]AT LmUGP in the presence and absence of MgCl\(_2\) as a representative example. As a quality control, we compared the \(K_{D1}\) for the complexation of Mg\(^{2+}\) to UTP obtained from NMR titrations with a previously reported value from isothermal titration calorimetry (Zea et al. 2008). The outcome from both methods was almost identical, reassuring the accuracy of our results. With this data in hand, only four parameters needed to be calculated: concentration of UTP:metal (UM), UTP:LmUGP (UP), UTP:metal:LmUGP (UMP) complexes and \(K_{D4}\). They were approximated via nonlinear least-squares minimization of the system of Eq. (3) described in supporting information. Results are summarized in Table 1. This knowledge allowed us to prepare samples containing virtually identical fractions of UTP:metal bound to LmUGP for every lanthanoid (Fig. 8b). This set-up produced relatively small PCS owing to the low metal binding site occupation selected (50.5% ± 5%). However, it ensured low free metal concentrations (< 60 µM), preventing unspecific paramagnetic effects and enabling the extraction of accurate PCS from \[^{13}C,^1H\]-HMQC spectra.

The lack of a crystal structure of LmUGP in complex with UTP required some special considerations concerning the choice of a proper structural model. When bound to UTP, QM/MM calculations predict a protein arrangement somewhere between the apo (open) and UDP-Glc bound state (post-reactive) (Führing et al. 2013). Therefore, an “ensemble structure” was generated combining the X-ray structures of available protein states populated by the enzyme during the catalytic cycle (see Fig. S7). The ensemble contained
CeCl₃ 13 ± 4 147 ± 3 17 ± 5 
LaCl₃ 11 ± 5 152 ± 6 15 ± 7 
TbCl₃ 
EuCl₃ 28 ± 9 112 ± 2 152 ± 8 38 ± 12 
LuCl₃ 38 ± 17 138 ± 8 46 ± 21 

80 PCS for Tb³⁺ (Table S4). The "ensemble structure" as strong PREs.

K and trivalent metals (Table 1 and in the presence of saturation metal concentrations (K₁₁₁). 

the following structures: apo (pdb 2OEF), UDP-Glc bound (pdb 4M2A) and dUpCp bound states, an analog of UTP (pdb 4M2B).

The 169 consistently assigned resonances were used as a starting point for the fitting of magnetic susceptibility tensors with the help of Paramagpy (Orton et al. 2020). Signal overlapping and PREs precluded the determination of PCS for some methyl resonances, allowing the extraction of 169 PCS for Ce³⁺, 157 PCS for Eu³⁺, 133 PCS for Tm³⁺ and 80 PCS for Tb³⁺ (Table S4). The "ensemble structure" as well as the individual conformations were subjected to Δχ tensor fitting. The initial paramagnetic centre was predefined to the metal position in the crystal structure with UDP-Glc bound (Fig. 6a, pdb 4M2A), but allowed to move during the fitting. Nevertheless, the unique tensor representations (UTR) showed that in all cases coordinates of the fitted paramagnetic centre converged in a very similar position (< 1.5 Å difference). Fitted metal coordinates in the presence of UDP deviate by almost 4 Å from those observed for the UDP-Glc state. (Fig. 6a and Table S5). This observation is in good agreement with the proposed molecular mechanism for LmUGP, were Mg²⁺ leaves the binding pocket upon synthesis of UDP-Glc in the form of PPi:metal complex (Führing et al. 2013).

Q factors were used for the evaluation of the quality of protein structures determined from PCS, and are listed in Table 2. The use of the "ensemble structure" yielded the lowest Q factors for every lanthanide ion. Our results nicely reflect that LmUGP adopts several conformational states in solution along the catalytic cycle, as defined by crystal structure analysis. Overall, tensors were very robust in the error analysis with the exception of the γ angle of the principal axis of the Ce³⁺ tensor, as summarized in Table S5. PCS from Ce³⁺ were therefore considered as imprecise and excluded from further analysis.

Paramagpy provided lists of expected PCSs for all methyl protons, which were used to expand the available assignment. In the next step, these lists together with the experimental PCSs were used as additional input files in a new round of MAP-XSII. All PCSs were multiplied by a weighting factor of 10 to match the order of magnitude of PREs normally used in these calculations, as previously described (Flugge and Peters 2018). The NOE list and the crystal structure of the UDP-Glc bound state (pdb 4M2A) were introduced as additional input files. We selected this...
crystal structure because it showed the closest Q factors to the “ensemble structure” (Table 2). 20 MMC trials were performed for each PCS list with cut-off distances spanning between 7 and 13 Å (Fig. S8). Although the cut-off distances were varied, the results were very similar for every lanthanide ion. Overall, 165 out of 169 assignments previously obtained could be confirmed. The four assignments that could not be reproduced corresponded to methyl groups assigned through NOE correlations which were present only in the apo conformation. MAP-XSII generated 22 to 29 new assignments solely based on PCSs, depending on the paramagnetic metal and cut-off distance preferred. However, not every peak was assigned to the same amino acid in every PCS list. As inclusion criteria we selected only amino acids consistently assigned to the same signal in at least two $\Delta \chi$ tensors with opposite orientations (i.e., $\text{Tb}^{3+}$ and $\text{Tm}^{3+}$ or in $\text{Ce}^{3+}$ and $\text{Eu}^{3+}$). Following this approach 16 new assignments were obtained, rising the number of assigned methyl groups to 185 (93%).

Table 2  Q factors derived from measured and calculated PCSs and PREs from combined fits to different structural models

| Structural model | Q factor | PCSs | PREs |
|------------------|----------|------|------|
|                  |          | $\text{Tb}^{3+}$ | $\text{Tm}^{3+}$ | $\text{Ce}^{3+}$ | $\text{Eu}^{3+}$ | $\text{Mn}^{2+}$ |
| Apo state (pdb 2OEF) |          | 0.158 | 0.226 | 0.155 | 0.246 | 0.473 |
| UDP-Glc state (pdb 4M2A) |          | 0.159 | 0.195 | 0.171 | 0.226 | 0.441 |
| In complex with UTP analog dUpCpp (pdb 4M28) |          | 0.181 | 0.196 | 0.182 | 0.265 | 0.374 |
| “Ensemble structure” (pdb 2OEF, 4M2A, 4M28) |          | 0.133 | 0.194 | 0.128 | 0.227 | 0.172 |

$\Delta \chi$ tensors and $\Gamma_2$ were fitted to the apo, UDP-Glc, dUpCpp analog or to an “ensemble structure”
PCSs of all assigned residues were finally used to derive $\Delta \chi$ tensors depicted in Fig. 8.

**Completeness of the assignment using PREs and directed mutagenesis**

Free electrons from paramagnetic metals induce fast relaxation of signals from nuclei sited close to the metal centre. The effect is equally pronounced in all directions around the metal ion, allowing for straightforward calculations of metal coordinates. Such calculations are most robust when the paramagnetic metal is uniformly surrounded by methyl groups, as it is the case of LmUGP when bound to the UTP:metal complex. In addition to this, PREs can be used for the verification of the assignments obtained by NOEs and PCS. We selected Mn$^{2+}$ because this metal induces stronger PREs than the lanthanide ions. Furthermore, Mn$^{2+}$ is a bivalent cation having an identical coordination sphere and a similar ionic radius as the natural Mg$^{2+}$ cofactor. MnCl$_2$ was added at low concentrations (55 µM) in the presence of UTP, which allowed the measurement of site-specific PREs for methyl groups as close as 10 Å from the metal centre. Reliable PREs could be extracted from 192 signals.

Fitting of the isotropic $g$-tensor requires preliminary knowledge of the rotational correlation times $\tau_c$ of the protein in solution, which was estimated from TRACT experiments (Fig. S10) (Lee et al. 2006). Similar to $\Delta \chi$ tensor fittings, the starting metal centre was extracted from the crystal structure with pdb 4M2A, but allowed to variate during the calculations. The best Q-factor (0.172) was observed for the “ensemble structure”, in line with the results obtained from PCS (see Fig. 9 and Table 2). Metal coordinates deviated <0.5 Å from those obtained for lanthanide ions, indicating very similar binding mode to UTP during the interaction with LmUGP (Fig. 6a and Table S5). A complete list of observed and expected PREs can be found in Table S4. In addition, $^1$H,$^13$C-HMQC spectra of MIL$^{\text{proSproS}}$AT methyl-labeled samples of LmUGP in the presence and absence of UDP-Glc using a synergic combination of methyl-methyl NOESY experiments, paramagnetic NMR and directed mutagenesis. Noteworthy, neither NOEs nor paramagnetic NMR alone delivered enough information for a complete methyl side chain assignment. The acquisition of 4D HMQC-NOESY-HMQC spectra from two structurally different protein states has been instrumental to obtain the largest portion of the assignment. Protein structural rearrangements induced by ligand binding delivered unique sets of NOEs signals in each protein state. Signal overlapping was resolved owing to the different chemical shifts observed along the different enzyme states. Assignments could be transferred among protein states with simple ligand titrations. PCS were fundamental to expand the assignment beyond the limits imposed by 4D HMQC-NOESY-HMQC data alone, PREs allowed for independent validation of the assignments and directed mutagenesis permitted the assignment of isolated methyl groups. In short, this work provides new avenues in the supplementary material (Table S6). $^1$H,$^13$C HMQC spectra showing the complete assignment can be found in Fig S12 and S13 in the supplementary material, respectively.

**Conclusions**

We have achieved a complete assignment of MIL$^{\text{proSproS}}$AT methyl-labeled samples of LmUGP in the presence and absence of UDP-Glc using a synergic combination of methyl-methyl NOESY experiments, paramagnetic NMR and directed mutagenesis. Noteworthy, neither NOEs nor paramagnetic NMR alone delivered enough information for a complete methyl side chain assignment. The acquisition of 4D HMQC-NOESY-HMQC spectra from two structurally different protein states has been instrumental to obtain the largest portion of the assignment. Protein structural rearrangements induced by ligand binding delivered unique sets of NOEs signals in each protein state. Signal overlapping was resolved owing to the different chemical shifts observed along the different enzyme states. Assignments could be transferred among protein states with simple ligand titrations. PCS were fundamental to expand the assignment beyond the limits imposed by 4D HMQC-NOESY-HMQC data alone, PREs allowed for independent validation of the assignments and directed mutagenesis permitted the assignment of isolated methyl groups. In short, this work provides new avenues
for the assignment of large, methyl-labeled proteins. The complete assignment of LmUGP will also serve as basis for novel studies into the biological function of this important class of enzymes.

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1007/s10858-022-00394-0.

**Acknowledgements** We would like to thank Prof. Thomas Peters for providing access to the NMR facilities and laboratories in Lübeck. We would also like to thank Dr. Thorsten Biet for support with the NMR experiments. In particular, we would like to thank Dr. Hugo van Ingen from the Bijvoet Centre for assisting us in any respect. We also thank Prof. Dr. Francois Routier, PD Dr. Roman Fedorov and Dr. Jana Führing (Hanover Medical School, Germany) for their advice and for providing the pET-22b plasmid containing the sequence encoding for LmUGP, and Professor Marcellus Ubbink (University of Leiden, The Netherlands) for fruitful discussions.

**Author contributions** LM, TA and AM designed experiments. LM, TA and AM expressed and purified samples, and acquired the NMR experiments. RC validated experiments. LM, TA, TM and AM analyzed and interpreted NMR data. AM wrote Matlab scripts for data analysis. TM contributed on the assignment by developing software algorithms. RK produced single-point mutants. AM wrote the paper. RC and TM edited and reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding** Open Access funding enabled and organized by Projekt DEAL. The state of Schleswig-Holstein is thanked for supplying NMR infrastructure to the Institute of Chemistry and Metabolomics of the University of Lübeck (European Fonds for Regional Development, LPW-E/1.1.2/857). 4D HMQC-NOESY-HMQC experiments were measured in the high-field NMR facility at the Bijvoet Centre in Utrecht at 900 MHz and 950 MHz magnetic fields within the frame of the InEXT program (Project PID 7367).

**Data availability** Chemical shift assignments for MIL_proSVproSAT methyl-labeled UDP-glucose pyrophosphorylase from *Leishmania major* have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under the Accession Number 50749.

**Code availability** Matlab scripts used for data analysis are available from the corresponding author upon request.

**Declarations**

**Conflicts of interest** The authors have no conflicts of interest in relation to the research presented here.

**Consent for publication** All authors have read and approved the submitted manuscript.

**Ethical approval** Not applicable.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

**References**

Amero C et al (2011) A systematic mutagenesis-driven strategy for site-resolved NMR studies of supramolecular assemblies. J Biomol NMR 50:229–236

Arai M, Ferreton JC, Wright PE (2012) Quantitative analysis of multi-site protein-ligand interactions by NMR: binding of intrinsically disordered p53 transactivation subdomains with the TAZ2 domain of CBP. J Am Chem Soc 134:3792–3803

Baldwin AJ, Kay LE (2009) NMR spectroscopy brings invisible protein states into focus. Nat Chem Biol 5:808–814

Barrett PJ et al (2013) The quiet renaissance of protein nuclear magnetic resonance. Biochemistry 52:1303–1320

Bax A, Clore GM (2019) Protein NMR: Boundless opportunities. J Magn Reson 306:187–191

Bloembergen N, Morgan LO (1961) Proton relaxation times in paramagnetic solutions effects of electron spin relaxation. J Chem Phys 34:842–850

Chao FA et al (2014) FLAMEEnGO 2.0: an enhanced fuzzy logic algorithm for structure-based assignment of methyl group resonances. J Magn Reson 245:17–23

Cornilescu G, Bax A (2000) Measurement of proton, nitrogen, and carbonyl chemical shielding anisotropies in a protein dissolved in a dilute liquid crystalline phase. J Am Chem Soc 122:10143–10154

Damerow S et al (2015) Depletion of UDP-glucose and UDP-galactose using a degron system leads to growth cessation of Leishmania major. PLoS Negl Trop Dis 9:e0004205

D’Angelo P et al (2011) Revised ionic radii of lanthanoid(III) ions in aqueous solution. Inorg Chem 50:4572–4579

Fiser A, Sali A (2003) ModLoop: automated modeling of loops in protein structures. Bioinformatics 19:2500–2501

Fletcher R (1988) Practical methods of optimization, 2nd edn. Wiley, New York

Flugge F, Peters T (2018) Complete assignment of Ala, Ile, Leu, Met and Val methyl groups of human blood group A and B glycosyltransferases using lanthanide-induced pseudocontact shifts and methyl-methyl NOESY. J Biomol NMR 70:245–259

Führing J et al (2013) Catalytic Mechanism and Allostery Regulation of UDP-Glucose Pyrophosphorylase from Leishmania major. ACS Catal 3:2976–2985

Gauto DF et al (2019) Integrated NMR and cryo-EM atomic-resolution structure determination of a half-megadalton enzyme complex. Nat Commun 10:2697

Han B, Liu Y, Ginzinger SW, Wishart DS (2011) SHIFTX2: significantly improved protein chemical shift prediction. J Biomol NMR 50:43–57

Hyberts SG, Takeuchi K, Wagner G (2010) Poisson-Gap Sampling and Forward Maximum Entropy Reconstruction for Enhancing the Resolution and Sensitivity of Protein NMR Data. J Am Chem Soc 132:2145–2147

Iwahara J, Tang C, Marius Clore G (2007) Practical aspects of (1)H transverse paramagnetic relaxation enhancement measurements on macromolecules. J Magn Reson 184:185–95

John M et al (2007) Sequence-specific and stereospecific assignment of methyl groups using paramagnetic lanthanides. J Am Chem Soc 129:13749–13757
Kamath U, Shriver JW (1989) Characterization of thermotropic state changes in myosin subfragment-1 and heavy meromyosin by UV difference spectroscopy. J Biol Chem 264:5586–5592

Kay LE (2011) Solution NMR spectroscopy of supra-molecular systems, why bother? A methyl-TROSY view. J Magn Reson 210:159–170

Kay LE, Ikura M, Tschudin R, Bax A (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. J Magn Reson 89:496–514

Lamerz AC, Haselhorst T, Bergfeld AK, von Itzstein M, Gerardy-Schahn R (2006) Molecular cloning of the Leishmania major UDP-glucose pyrophosphorylase, functional characterization, and ligand binding analyses using NMR spectroscopy. J Biol Chem 281:16314–16322

Lee D, Hilty C, Wuthrich K (2006) Effective rotational correlation times of proteins from NMR relaxation interference. J Magn Reson 178:72–76

Lescanne M et al (2017) Methyl group assignment using pseudospectral shifts with PARAssign. J Biomol NMR 69:183–195

Mas G, Crublet E, Hamelin O, Gans P, Boisbouvier J (2013) Specific labeling and assignment strategies of valine methyl groups for NMR studies of high molecular weight proteins. J Biomol NMR 57:251–262

Mas G et al (2018) Structural investigation of a chaperonin in action reveals how nucleotide binding regulates the functional cycle. Sci Adv 4:eaau4196

Monneau YR et al (2017) Automatic methyl assignment in large proteins by the MAGIC algorithm. J Biomol NMR 69:215–227

Muller-Hermes C, Creutznacher R, Mallagaray A (2020) Complete assignment of Ala, Ile, Leu(ProS), Met and Val(ProS) methyl groups of the protruding domain from human norovirus GII.4 Saga. Biomol NMR Assign 14:123–130

Neri S, De Paula VS, McShan AC, Sgourakis NG (2021) Backbone-independent NMR resonance assignments of methyl probes in large proteins. Nat Commun 12:691

Orton HW, Huber T, Otting G (2020) Paramagpy: software for fitting magnetic susceptibility tensors using paramagnetic effects measured in NMR spectra. Magn Reson 1:1–12

Pederson K et al (2017) NMR characterization of HtpG, the E. coli Hsp90, using sparse labeling with (13)C-methyl alanine. J Biomol NMR 68:225–236

Pritisanac I et al (2017) Automatic assignment of methyl-NMR spectra of supramolecular machines using graph theory. J Am Chem Soc 139:9523–9533

Pritisanac I, Wurz JM, Alderson TR, Guntert P (2019) Automatic structure-based NMR methyl resonance assignment in large proteins. Nat Commun 10:4922

Pritišanac I, Reid Alderson T, Guntert P (2020) Automated assignment of methyl NMR spectra from large proteins. Prog Nucl Magn Reson Spectrosc 118–119:57–73

Proudfoot A, Frank AO, Ruggiu F, Mamo M, Lingel A (2016) Facilitating unambiguous NMR assignments and enabling higher probe density through selective labeling of all methyl containing amino acids. J Biomol NMR 65:15–27

Robson S, Arthanari H, Hyberts SG, Wagner G (2019) Nonuniform sampling for NMR spectroscopy. Enzymol 614:263–291

Rosenzweig R, Kay LE (2014) Bringing dynamic molecular machines into focus by methyl-TROSY NMR. Annu Rev Biochem 83:291–315

Rosenzweig R et al (2015) ClpB N-terminal domain plays a regulatory role in protein disaggregation. Proc Natl Acad Sci USA 112:E6872–E6881

Ruschak AM, Kay LE (2012) Proteasome allostery as a population shift between interchanging conformers. Proc Natl Acad Sci USA 109:E3454–E3462

Sahakyan AB, Vranken WF, Cavalli A, Vendruscolo M (2011) Structure-based prediction of methyl chemical shifts in proteins. J Biomol NMR 50:331–346

Schütz S, Sprangers R (2019) Methyl TROSY spectroscopy: a versatile NMR approach to study challenging biological systems. Prog Nucl Magn Reson Spectrosc 116:56–84

Shiraishi Y et al (2018) Phosphorylation-induced conformation of beta2-adrenoceptor related to arrestin recruitment revealed by NMR. Nat Commun 9:194

Skinner SP, Moshev M, Hass MA, Keizers PH, Ubbink M (2013) PARAssign—paramagnetic NMR assignments of protein nuclei on the basis of pseudocontact shifts. J Biomol NMR 55:379–389

Sprangers R, Kay LE (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature 445:618–622

Steiner T et al (2007) Open and closed structures of the UDP-glucose pyrophosphorylase from Leishmania major. J Biol Chem 282:13003–13010

Tugarinov V, Kay LE (2003) Ile, Leu, and Val Methyl Assignments of the 723-Residue Malate Synthase G Using a New Labeling Strategy and Novel NMR Methods. J Am Chem Soc 125:13868–13878

Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE (2003) Cross-correlated relaxation enhanced 1H–13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. J Am Chem Soc 125:10240–10248

Tugarinov V, Kay LE, Ibraghimov I, Orekhov VY (2005) High-resolution four-dimensional 1H–13C NOE spectroscopy using methyl-TROSY, sparse data acquisition, and multidimensional decomposition. J Am Chem Soc 127:2767–2775

Tugarinov V, Sprangers R, Kay LE (2007) Probing side-chain dynamics in the proteasome by relaxation violated coherence transfer NMR spectroscopy. J Am Chem Soc 129:1743–1750

Velyvis A, Schachman HK, Kay LE (2009) Assignment of Ile, Leu, and Val methyl correlations in supra-molecular systems: an application to aspartate transcarbamoylase. J Am Chem Soc 131:16534–16543

Venditti V, Fawzi NL, Clore GM (2011) Automated sequence- and stereo-specific assignment of methyl-labeled proteins by paramagnetic relaxation and methyl-methyl nuclear Overhauser enhancement spectroscopy. J Biomol NMR 51:319–328

Vranken WF et al (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59:687–696

Wen J, Zhou P, Wu J (2012) Efficient acquisition of high-resolution 4-D diagonal-suppressed methyl–methyl NOE/SY for large proteins. J Magn Reson 218:128–132

WHO (1984) Expert & committee. The Leishmaniasis. World Health Organization, Geneva

Williamson MP (2013) Using chemical shift perturbation to characterise ligand binding. Prog Nucl Mag Reson Spectrosc 73:1–16

Williamson MP (2017) Chemical Shift Perturbation. In: Webb GA (ed) Modern Magnetic Resonance. Springer, Cham, pp 1–19

Xu Y, Matthews S (2013) MAP-XSII: an improved program for the automatic unambiguous NMR assignments of methyl probes. J Biomol NMR 55:379–389

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.