Progress in Nuclear Magnetic Resonance Spectroscopy 118–119 (2020) 54–73

Contents lists available at ScienceDirect

Progress in Nuclear Magnetic Resonance Spectroscopy

journal homepage: www.elsevier.com/locate/pnmrs

Automated assignment of methyl NMR spectra from large proteins

Iva Pritišanac [a], T. Reid Alderson [b], Peter Güntert [a,c,d,*]

[a] Institute of Biophysical Chemistry, Center for Biomolecular Magnetic Resonance, Goethe University Frankfurt am Main, 60438 Frankfurt am Main, Germany
[b] Laboratory of Chemical Physics, NIDDK, National Institutes of Health, Bethesda, MD 20892, USA
[c] Laboratory of Physical Chemistry, ETH Zürich, 8093 Zürich, Switzerland
[d] Department of Chemistry, Tokyo Metropolitan University, Hachioji, Tokyo 192-0397, Japan

Edited by David Neuhaus and Gareth Morris

ARTICLE INFO

Article history:
Received 25 February 2020
Accepted 17 April 2020
Available online 24 April 2020

Keywords:
methyl-TROSY
Automatic resonance assignment
Large proteins
Molecular machines

ABSTRACT

As structural biology trends towards larger and more complex biomolecular targets, a detailed understanding of their interactions and underlying structures and dynamics is required. The development of methyl-TROSY has enabled NMR spectroscopy to provide atomic-resolution insight into the mechanisms of large molecular assemblies in solution. However, the applicability of methyl-TROSY has been hindered by the laborious and time-consuming resonance assignment process, typically performed with domain fragmentation, site-directed mutagenesis, and analysis of NOE data in the context of a crystal structure. In response, several structure-based automatic methyl assignment strategies have been developed over the past decade. Here, we present a comprehensive analysis of all available methods and compare their input data requirements, algorithmic strategies, and reported performance. In general, the methods fall into two categories: those that primarily rely on inter-methyl NOEs, and those that utilize methyl PRE- and PCS-based restraints. We discuss their advantages and limitations, and highlight the potential benefits from standardizing and combining different methods.

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Nuclear magnetic resonance (NMR) spectroscopy can simultaneously probe the structures and dynamics of biomolecules at atomic resolution. Before obtaining atomic-level information from NMR spectra, however, the observed NMR signals must be assigned to their corresponding nuclei of origin, henceforth referred to as the resonance assignment. Typically, resonance assignments are obtained from multi-dimensional, triple resonance NMR spectra that utilize scalar couplings to correlate ^{1}H, ^{13}C, and ^{15}N chemical shifts of the protein backbone and side chains [1–4]. For soluble, globular proteins smaller than ~30 kDa that do not suffer from severe resonance overlap or sequence degeneracy [5,6], resonance assignments can in general be determined by this method, for which several successful automated approaches are available [7,8]. Most eukaryotic proteins, however, are either larger than 30 kDa or assemble into oligomers that exceed this molecular mass threshold. In large proteins, rapid signal decay increases the linewidths of resonances and decreases the spectral resolution, rendering such proteins challenging to study using 'traditional' NMR methods.

On the other hand, the reintroduction of selectively protonated, ^{13}C-labeled methyl groups in an otherwise highly deuterated background enables the acquisition of high-resolution, [^{1}H,^{13}C]-heteronuclear multiple quantum coherence (HMQC) spectra of proteins that exceed 1 MDa [9–15]. The methyl-transverse relaxation optimized spectroscopy (TROSY) effect manifests itself through the destructive interference of dipolar interactions in the methyl $^{13}C^{1}H_3$ spin system, yielding significantly attenuated relaxation and high-quality methyl NMR spectra. However, the sparse methyl labeling approach creates isolated spin systems, which precludes resonance assignment by traditional scalar coupling-based methods.

Two current approaches for methyl resonance assignment are outlined in Fig. 1 [16,17]. Because of their isolated nature, NMR signals from methyl groups in large proteins are usually assigned either by transferring assignments obtained from smaller protein fragments to the spectra of the intact protein or an oligomer (Fig. 1a) [18,19], or, alternatively or additionally, by monitoring spectral changes after targeted site-directed mutagenesis (Fig. 1b) [20]. Both approaches are laborious, time consuming,

* Corresponding author at: Institute of Biophysical Chemistry, Center for Biomolecular Magnetic Resonance, Goethe University Frankfurt am Main, 60438 Frankfurt am Main, Germany.
E-mail address: peter.guentert@phys.chem.ethz.ch (P. Güntert).

https://doi.org/10.1016/j.pnmrs.2020.04.001
0079-6565/© 2020 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

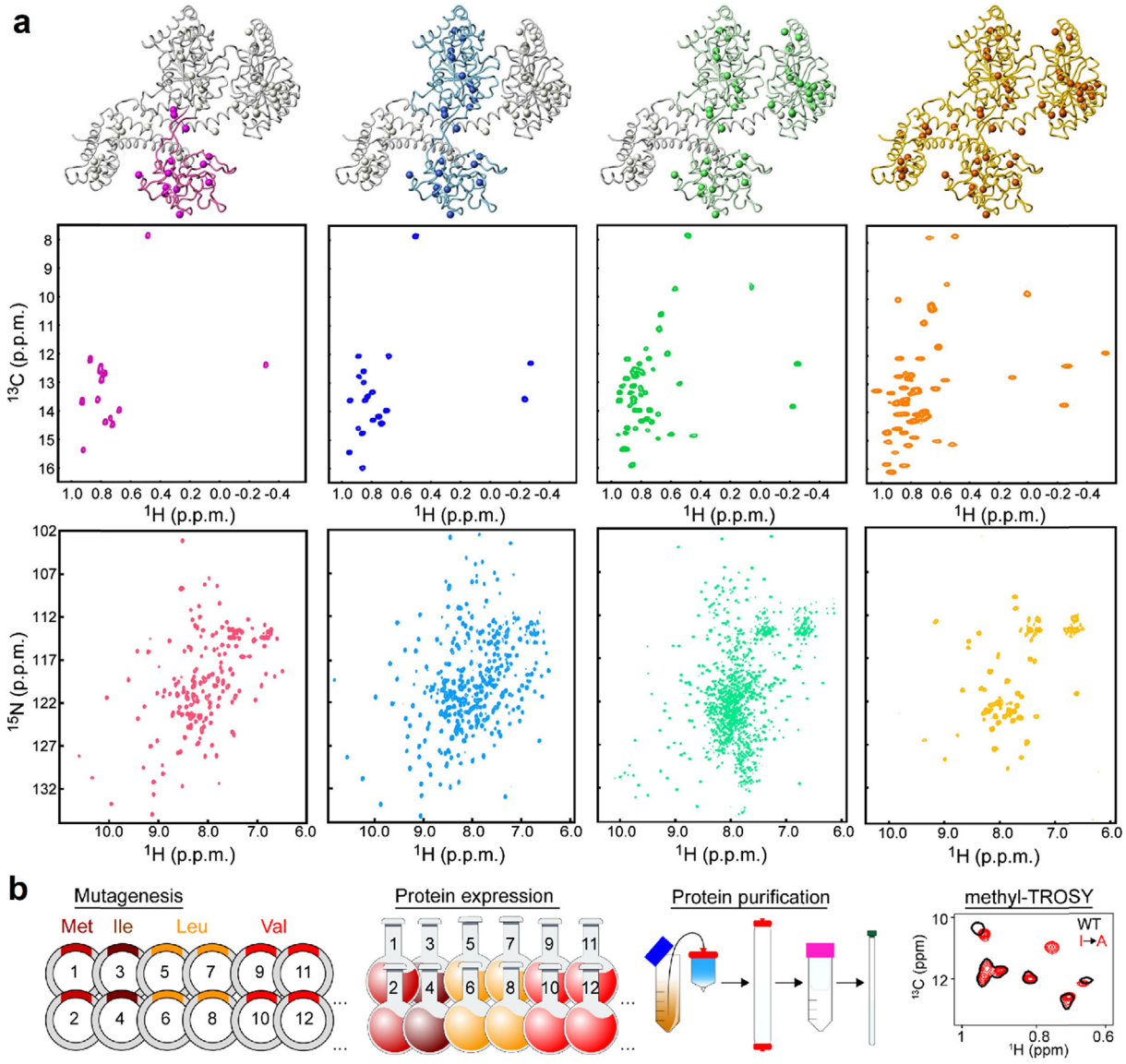

Fig. 1. Illustration of two commonly employed strategies for methyl resonance assignments. (a) Domain parsing strategy shown on the example of SecA. Figure adapted from Fig. 2 of Gelis et al. [19] with permission. Domain fragmentation dissects a large protein into individual folded domains that yield high-quality 2D [^{1}H,^{15}N]-HSQC or TROSY-HSQC spectra. These smaller domains and fragments can be assigned using the standard backbone and side-chain resonance assignment experiments. The methyl resonance assignments obtained for the individual (purple) or tandem domains (blue and green) are transferred to the methyl-TROSY spectrum of the full-length protein (orange). (b) Site-directed-mutagenesis strategy. Each methyl-bearing residue is individually mutated (mutagenesis), and the plasmid containing the mutated gene is transformed into a bacterial strain suitable for protein overexpression. Each mutant protein is expressed in a small-scale bacterial culture grown in minimal medium with D_2O, deuterated glucose, and methyl-labeling precursors (protein expression), and individually purified (protein purification). A 2D [^{1}H,^{13}C]-HMQC spectrum is recorded for each mutant (methyl-TROSY). Overlaying the spectra of the wild-type (black) and a mutant protein (red) allows the missing resonance to be identified, which yields the assignment of the mutated methyl residue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and expensive. These obstacles hinder routine application and widespread usage of methyl-TROSY methods.

In response to such difficulties, several automatic methyl resonance assignment strategies have been proposed in the past decade. Here, we review the available automatic methyl resonance assignment programs, their input requirements (Section 3.3), their algorithmic approaches (Section 4), and their reported performance (Section 5).

Alongside these software developments, advances in the selective isotopic labeling of all available methyl groups have significantly increased the number of NMR-active probes and facilitated detailed methyl NMR-based studies of large proteins. The choice of labeling directly impacts the complexity of the methyl resonance assignment search. We therefore begin with a short survey of recent developments in the isotope labeling of methyl groups (Section 2.1) and point the interested reader to comprehensive reviews focused on methyl-TROSY applications (Section 2.3). The choice of methyl labeling is further discussed in the context of data preparation (Section 3), data treatment by the different automatic resonance assignment approaches (Section 4), and in the context of algorithm performance (Section 5). The synergy between experimental design and the automation of resonance assignment is expected to advance automatic methyl resonance assignment in the future (Section 6).

2. Methyl groups as probes of macromolecular structure and dynamics

Methyl groups are particularly attractive probes in high-resolution biomolecular NMR studies of large proteins for several

reasons. First, methyl-bearing residues are well represented in the amino acid sequences of proteins and reside both in protein hydrophobic cores (e.g. Ile, Leu, Val) and at protein surfaces (e.g. Ala, Met, Thr) [21]. Such dispersed coverage yields structural and dynamical probes throughout the protein (Fig. 2). Second, the peripheral localization of methyl groups in the side chains of amino acids often results in increased dynamics, leading to slower relaxation, which is favorable for investigating large systems. In general, the methyl group dynamics increase with increasing distance from the backbone: for example, the Cβ of Ala is typically more rigid than the $C\delta_1$ of Ile, with the Cε of Met being most flexible [22]. Finally, rapid rotation around the methyl symmetry axis renders the three ^{1}H spins degenerate, leading to enhanced sensitivity, while the relatively large dispersion of ^{13}C resonances (*ca.* 20 ppm) allows for high spectral resolution. Collectively, the benefits of methyl-TROSY combined with selective methyl labeling

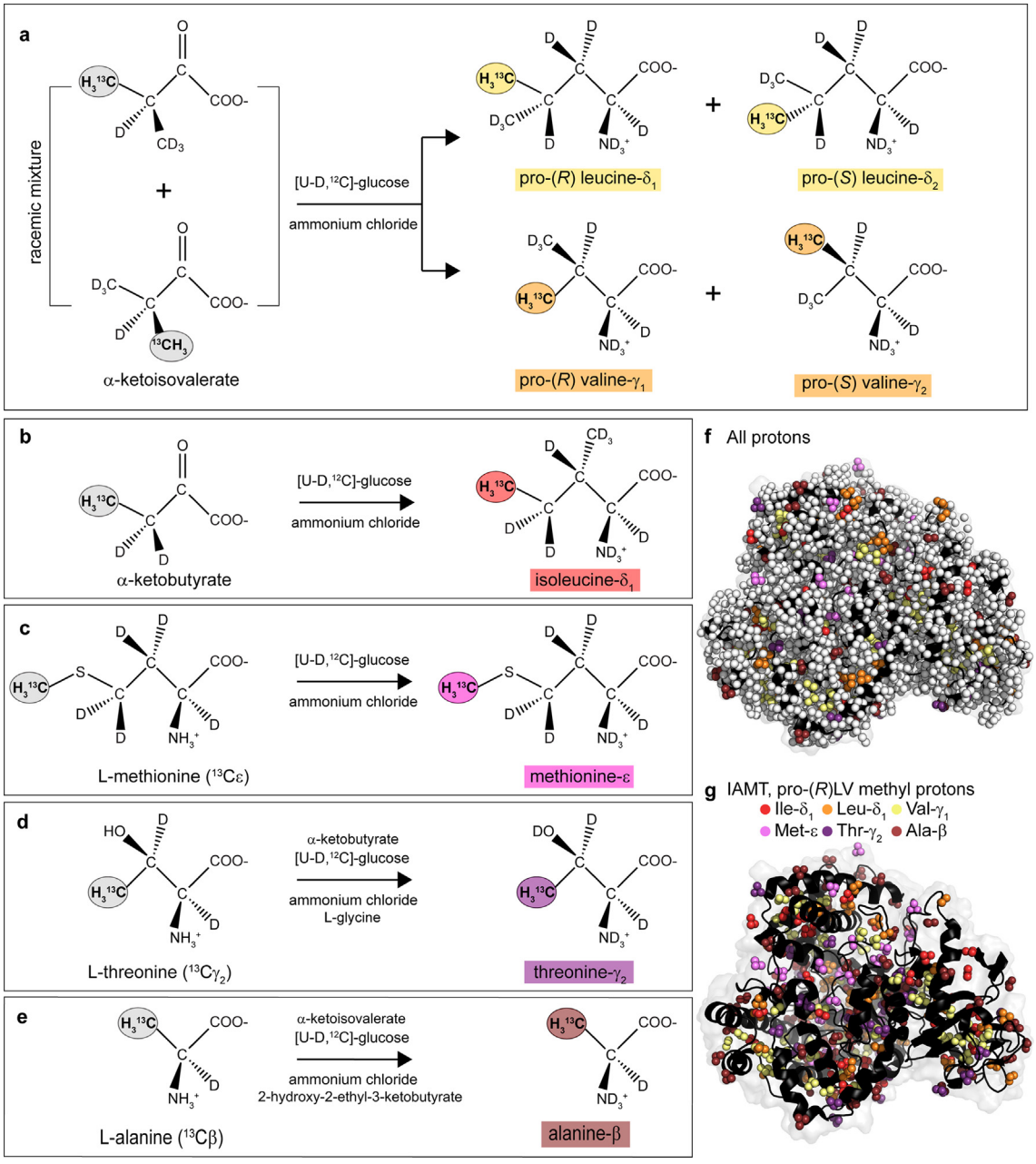

Fig. 2. Biosynthetic labeling strategies to prepare selectively $^{13}CH_3$ methyl-labeled protein in a highly deuterated background. These approaches apply for proteins produced in *E. coli* using M9 minimal media prepared in D_2O. The chemicals over each arrow indicate either metabolic conversion of the particular reagent (ammonium chloride and deuterated glucose, [U-D, ^{12}C]-glucose) into the final amino acid (right) or additional reagents that are used to suppress isotope scrambling (panels d, e: α-ketobutyrate, L-glycine, α-ketoisovalerate, and 2-hydroxy-2-ethyl-3-ketobutryate). Common precursors are listed for (a) Leu-δ_1/δ_2 and Val-γ_1/γ_2 (racemic mixture), (b) Ile-δ_1, (c) Met-ε, (d) Thr-γ_2, and (e) Ala-β labeling. The ^{13}C-labeled methyl group is indicated in a grey ellipse in the precursor and in a colored ellipse in the resultant amino acid. In panel a, the precursor used to label Leu and Val exists as a racemic mixture and therefore introduces the $^{13}CH_3$ label at either the δ_1 or δ_2 (γ_1 or γ_2) position in Leu (Val). Note that solvent-exchangeable protons are exchanged to deuterons when the precursors, and later the purified protein, are dissolved in D_2O. When preparing an AILMTV-labeled protein, the individual precursors can simply be added in combination and in the presence of L-glycine to prevent scrambling. See [72] for a more detailed discussion on labeling approaches. Selective methyl labeling greatly reduces the overall proton density, as demonstrated in (f) and (g) for malate synthase G (PDB: 1y8b) that contains a total of 5623 protons, shown as white and colored spheres. Selective protonation of the methyl groups of Ala-β, Ile-δ_1, Leu-δ_1, Met-ε, Thr-γ_2, and Val-γ_1 (IAMT, pro-*R* LV) introduces 864 protons, corresponding to *ca.* 15% of the total.

allow detailed structural and dynamical studies of large proteins and protein complexes.

2.1. Isotope labeling of methyl groups

To use methyl groups as NMR probes in studies of large systems, selectively protonated, ^{13}C-labeled methyl groups need to be introduced in an otherwise deuterated, ^{12}C protein background. While methyl labeling strategies exist for non-bacterial hosts, such as yeast [23–26] and insect cells [27–31], we here focus on methods used to obtain highly deuterated, methyl-labeled protein from *E. coli* (Fig. 2), which is the most commonly employed host organism. In the initial methyl labeling strategies, isotopically labeled amino acids (isoleucine, leucine, and valine) or pyruvate were added to the minimal medium that was used for *E. coli*-based protein production in a highly deuterated background [32,33]. Since these initial efforts, cost-effective and robust methyl labeling protocols have been established for all methyl-bearing residues (Fig. 2a–e), with significant improvements in labeling efficiency [34,35] and protein yields [36,37].

Modern approaches to selective methyl labeling exploit specific isotope-labeled biosynthetic precursors, for example, α-ketobutyrate (IUPAC name [3,3-^{2}H$_2$, 4-^{13}C$_1$]2-oxobutanoic acid) for Ile-δ_1 (Fig. 2b) [38] or α-aceto-α-hydroxybutyrate (IUPAC name 2-[^{2}H$_5$]ethyl-2-hydroxy-3-[4-^{13}C$_1$]oxobutanoic acid) for Ile-γ_2 [39]. The labeling of Ile methyl groups is advantageous because of their characteristic upfield ^{13}C chemical shifts of 13.4 ± 1.6 ppm (mean ± s.d. according to filtered BMRB statistics) for Ile-δ_1 and 17.5 ± 1.3 ppm for Ile-γ_2, and distinct metabolic pathways. This enables rapid identification of Ile resonances and the preparation of Ile-only labeled samples. For leucine and valine methyl groups, a similar strategy can be used with the shared precursor α-ketoisovalerate (IUPAC name 3-[^{13}C$_1$]methyl-[3,4,4,4-^{2}H$_4$]2-oxobutanoic acid) to label Leu-δ_1, δ_2 and Val-γ_1, γ_2 (Fig. 2a) [40]. However, significant overlap in the ^{13}C chemical shift ranges of Leu-δ_1, δ_2 (24.4 ± 2.1 ppm) and Val-γ_1, γ_2 groups (21.4 ± 2.6 ppm) motivated the establishment of separate labeling strategies for either Leu or Val. Lichtenecker et al. showed that selective Leu labeling can be achieved by using labeled α-ketoisocaproate [41], whereas the combined addition of unlabeled α-ketoisocaproate and labeled α-ketoisovalerate led to exclusive labeling of Val [42]. The addition of unlabeled precursors adds additional protons into the system, and can therefore degrade the quality of the spectrum for proteins that contain many Leu or Val residues. However, the ability to discriminate between Leu and Val methyl resonances proves highly beneficial to some of the algorithms discussed below (Section 5). The problem of additional protons can be overcome by using fully deuterated precursors when "unlabeling" either Leu or Val [42]. To avoid the racemic mixture of pro-*R* and pro-*S* labeled Leu-δ_1, δ_2 and Val-γ_1, γ_2 groups that results from the use of racemic α-ketoisovalerate (Fig. 2a), 2-hydroxy-2-[^{13}C]methyl-3-oxo-4-[^{2}H$_3$]butanoic acid can be used for selective labeling of the pro-*S* groups [43] and 2-hydroxy-2-[^{2}H$_3$]methyl-3-oxo-4-[^{13}C]butanoic acid for the pro-*R* groups [44]. Both precursors can be obtained through hydrolysis of the commercially available esters ethyl-2-hydroxy-2-[^{13}C]methyl-3-oxobutyrate (for the pro-*S* case) and ethyl-2-hydroxy-2-[^{2}H$_3$]methyl-3-oxobutyrate-4[^{13}C] (for the pro-*R* case) [43,44], with the former requiring an additional step for deuteration of the protonated 4-methyl group. In addition, selective labeling of pro-*R* or pro-*S* groups of either Val or Leu residues can be achieved by cell-free synthesis with the SAIL method [45,46] or by using *E. coli* auxotrophs [47,48]. Such stereospecific labeling significantly reduces spectral complexity, as only one of the two methyl groups of each Leu and Val residue is labeled, and enhances signal-to-noise ratios, both of which are particularly useful for studies of large systems.

Finally, biosynthetic strategies for selective labeling of the Met-ε (Fig. 2c) [49,50], Thr-γ_2 (Fig. 2d) [51,52], and Ala-β [53,54] (Fig. 2e) groups have been established, and a method for simultaneous labeling of all methyl-bearing residues (AILMTV) by combining all biosynthetic precursors in a single protein production medium has been reported [35]. This enables higher probe density with selective labeling of all methyl-containing amino acids (Fig. 2f, g) [55]. In addition, the commercial availability of most precursors has spurred the development of ^{13}CDH$_2$ and ^{13}CHD$_2$ methyl labeling for obtaining isolated ^{13}C-D and ^{13}C–^{1}H bonds in a given methyl group, which is advantageous for NMR relaxation analysis [56–63]. ^{13}CHD$_2$ labeling reduces sensitivity due to the depletion of methyl protons, which increases ^{1}H longitudinal relaxation times relative to the same protein labeled with ^{13}CH$_3$ [55]. For example, ^{13}CHD$_2$ labeling of the 360 kDa half-proteasome led to a 1.5–2-fold decrease in sensitivity and an approximately two-fold increase in ^{1}H longitudinal relaxation times [55]. The modest loss in signal is tolerable when measurement of certain methyl relaxation rates is sought. Finally, the combined usage of isotope labeling strategies and Carr-Purcell-Meiboom-Gill (CPMG) and $R_{1\rho}$ relaxation dispersion methods has brought new insight into the dynamics of methyl groups on the micro/millisecond timescale [56,64–71].

2.2. Alternative sparse labeling

While perdeuteration works well in bacterial hosts, deuterated, minimal media are ill-suited for mammalian cell growth, preventing production of highly deuterated, methyl-protonated large proteins in those systems. Producing proteins in mammalian cells is of particular interest for studies of natively glycosylated human proteins [73,74]. As an alternative, sparse labeling can be employed in which a single ^{15}N- or ^{13}C-enriched amino acid, or or a subset of such amino acids, is supplied directly to the mammalian cells [74].

The applicability of sparse methyl labeling to large glycoproteins produced in mammalian cells was recently investigated [75]. In the presence of an otherwise protonated background, deuterated leucine selectively ^{13}CH$_3$-methyl labeled at the δ_2 position was introduced into the protein, which yielded significant improvements in spectral quality compared to data obtained with non-deuterated amino acids. Selective amino acid deuteration allowed for the detection of all leucine methyl signals of an intact IgG2b antibody (150 kDa) [75]. Recently, similar advancements have been proposed using aromatic residues [76].

2.3. Methyl-TROSY: A window into the functional mechanisms of supramolecular complexes

Methyl-TROSY delivers atomic-level insight into the functional mechanisms and dynamics of proteins and protein complexes. The pioneering methyl-TROSY studies from the Kay group predominantly focused on the 20S core particle proteasome complex, which is composed of 14 subunits that total 670 kDa and reaches up to 1.1 MDa in the presence of the 11S activator [18,77]; and malate synthase G, a single polypeptide chain of 82 kDa [2,9,78,79]. Since this initial phase in the early 2000s, there have been numerous applications of methyl-TROSY, including studies on substrate recognition in the unfolded protein response [80], the self-assembly of large systems such as the *ca.* 500 kDa TET2 proteolytic complex [81] and 2–80 MDa amyloid-β protofibrils [82], interactions between the ribosome and nascent chains [83–86], and the mechanisms of enzymes [87–90], chaperones [91–94], kinases [95–97], and proteases [98], among others. Methyl groups can also be effective reporters on transient protein (mis)folding intermediates [99]. The continuing development of

experiments that probe protein conformational dynamics keeps expanding the utility of methyl groups for NMR [71,100–102].

For more insight into theoretical and practical aspects of methyl-TROSY, as well as numerous successful applications, the interested reader is directed to comprehensive reviews by Schütz et al. [72], Rosenzweig et al. [103], Ruschak et al. [104], and Tugarinov et al. [79].

3. Methyl assignment strategies

3.1. Conventional approaches to methyl resonance assignments are labor-intensive

The selective methyl labeling schemes outlined in Section 2 have significantly increased the coverage of methyl probes. While all methyl-bearing amino acids can be readily labeled, the major bottleneck for methyl-TROSY is resonance assignment: the relation of observed $^{1}H–^{13}C$ signals in a heteronuclear multiple-quantum coherence (HMQC) spectrum to specific methyl groups in the molecule [16,17]. For Leu and Val residues, this includes stereospecific assignment of the two methyl groups.

The standard approach to methyl resonance assignment in most laboratories remains large-scale site-directed mutagenesis [20], which can also be used to extend or validate the assignments generated through 'divide and conquer' or domain fragmentation strategies. Assignment by site-directed mutagenesis is especially time-consuming and laborious, as each methyl-bearing residue must be individually mutated and then each mutant protein expressed and purified. The assignment itself consists of overlaying the 2D [^{1}H,^{13}C]-HMQC spectrum of a mutant protein on that of the wild-type protein and identifying which resonance(s) are missing [20]. Since each mutant protein must be prepared with appropriate isotope labels, the large number of cultures can consume significant amounts of D_2O, deuterated glucose, and methyl-labeling precursors, even when accounting for the small volumes of cultures that are typically used for assignment purposes. The volume of the culture and consumption of isotopically labeled reagents depend on the yield of the overexpressed protein, but cultures that require 50 ml of M9 medium per mutant are not uncommon.

The 'divide-and-conquer' strategy is based on dissecting an oligomeric complex into individual monomers whereas domain fragmentation refers to dividing a large protein into its individually folded domains [18,19,105]. If the monomers or protein domains thus obtained yield high-quality HSQC- or TROSY-based spectra, they are amenable to the traditional backbone and side-chain resonance assignment strategies [1–4]. Such efforts are occasionally supplemented by inspection of methyl-methyl NOEs in the context of an available crystal structure [18,98,106].

The described strategies require that the isolated domains, monomers, or introduced mutations do not lead to structural rearrangements, such that assignments from the spectra acquired with domain fragmentation or site-directed mutagenesis can be faithfully transferred to the intact protein or oligomer. This is often not straightforward in practice, and may require additional assignment validation experiments [92,105]. Overall, the challenges and costs associated with methyl resonance assignment currently hinder routine application and widespread usage of powerful methyl-TROSY methods. Therefore, reliable automatic assignment strategies represent important advances to the field.

3.2. Automatic methyl resonance assignment approaches are structure-based

NMR studies of high molecular mass proteins often require partial or uniform deuteration, where methyl groups remain the sole protonated and isotopically (^{13}C) labeled parts of the molecule. Consequently, automatic side-chain resonance assignment strategies that rely on through-bond magnetization transfer cannot be used. Over the last decade, structure-based automatic methyl resonance assignment strategies have been developed that rely almost exclusively on sparse spatial restraints for methyl groups. These restraints can be computed from the protein structure, provided one is available, and related to those derived from NMR data. Based on their primary source of NMR data, the assignment strategies can be divided into two categories (Fig. 3): nuclear Overhauser enhancement (NOE)-based algorithms, which predominantly exploit a network of measured methyl-methyl NOEs; and paramagnetism-based approaches, which make use of paramagnetic relaxation enhancements (PREs) or pseudocontact shifts (PCS). The latter are measured on methyl spins after a paramagnetic center has been introduced into a protein in a site-specific manner.

3.2.1. NOE-based assignment approaches

Connectivity between methyl peaks from 2D [^{1}H,^{13}C]-HSQC/HMQC spectra can be established through 3D or 4D methyl-methyl NOE measurements. To calculate the most likely methyl assignments, NOE-based approaches relate NOE-established inter-methyl connectivities to those extracted from the high-resolution protein structure. The underlying idea mimics manual approaches that are typically undertaken in methyl-TROSY studies of large systems to confirm or extend assignments obtained with site-directed mutagenesis or through-bond NMR strategies [18]. The first method to automate methyl resonance assignment primarily on the basis of methyl-methyl NOEs, Methyl Assignment Prediction from X-ray Structures (MAP-XSII), was reported in 2009 by Matthews and colleagues [107,108]. Since then, four other NOE-based automatic approaches, FLAMEnGO2.0 [109,110], MAGMA [111], MAGIC [112], and MethylFLYA [113] have been developed (Fig. 3). In addition, a hierarchical NOE- and structure-based assignment protocol was developed by Xiao et al. [114], though presently without a dedicated software implementation.

The NOE-based approaches do not require any mutations or modifications of the studied protein. The short-distance nature of NOEs, stemming from their r^{-6} dependence, where r is the distance between two nuclei, and spin diffusion, may cause difficulties in the interpretation of NOEs. However, these can be partially overcome by uniform deuteration and restricted protonation of selected methyl-bearing residues [115]. 3D and 4D HMQC-NOESY-HMQC experiments [116–119] have been successfully applied to large, exclusively methyl protonated proteins in the past, generating high quality NOE data that could be used for manual or automatic assignment strategies [18,105,108,111,113] (Fig. 4). For example, Kerfah et al. applied this approach to U-[H^2], Ala-[C^{β}], Ile-[$C^{\delta1}$] and (Leu/Val)$^{pro\text{-}S}$ protonated malate synthase G (MSG, 82 kDa). The authors reported [120] that 96% of the NOEs expected based on a crystal structure in the distance range

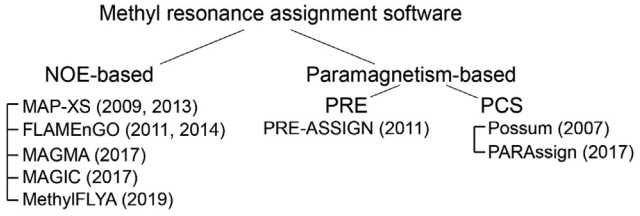

Fig. 3. Summary of automatic methyl resonance assignment strategies and their software implementations. The year of publication is given in parentheses.

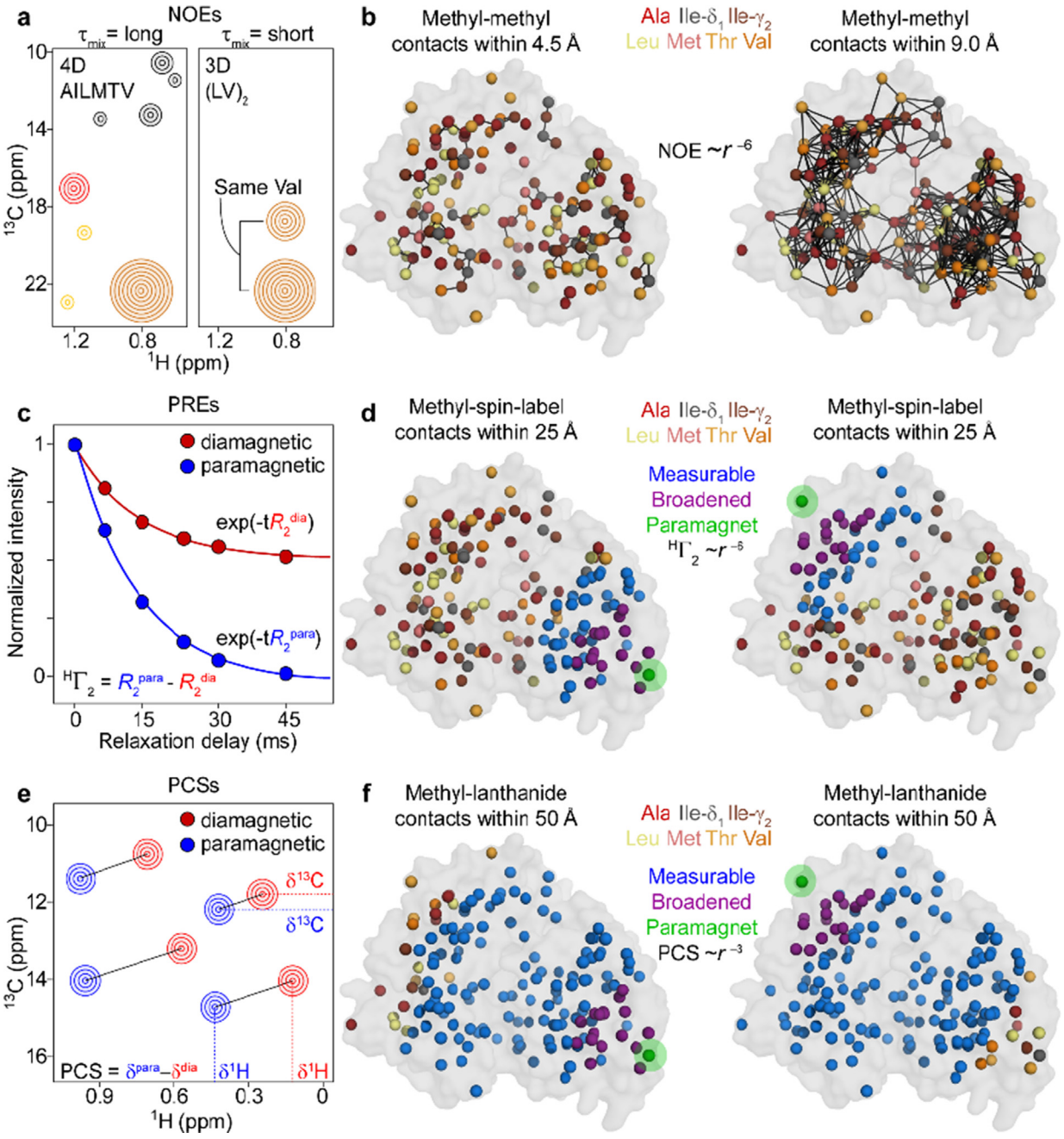

Fig. 4. NMR data used for automatic methyl resonance assignment. (a) *Left*: Example measurement of inter-residue methyl-methyl NOEs. Schematic 2D plane from a 4D HCCH methyl-methyl NOESY spectrum acquired with a long mixing time (τ_{mix}) on an AILMTV-labeled protein. The diagonal peak at 0.8/22 ppm originates from a Val methyl. All other peaks are NOE cross-peaks to the Val methyl diagonal peak. The labeling pattern produces a racemic mixture of pro-*R* and pro-*S* labeled Leu/Val residues. Thus, in each molecule either the pro-*R* or pro-*S* methyl group of a given residue is labeled and intra-residual methyl-methyl NOEs cannot be observed. *Right*: Example measurement of intra-residual methyl-methyl NOEs to identify geminal methyl pairs in Leu and Val residues. Schematic 2D plane from a 3D CCH methyl-methyl NOESY spectrum acquired with a short τ_{mix} on a (Leu-δ_1/δ_2, Val-γ_1/γ_2)-labeled protein. Simultaneous labeling of both methyl groups in Leu and Val residues enables detection of intra-residual NOEs between the methyl groups of the same residue (in this example, Val). The short τ_{mix} yields NOEs only for short distances, which is why weak NOEs with distant residues (yellow signals in the left panel) are not observed here. This facilitates linkage of the two methyl groups of a residue and simplifies automatic assignment (see text). The resonances are color coded according to panel b. (b) Crystal structure of MBP (PDB 1ANF) with methyl groups for AILMTV residues depicted as spheres. Only the δ_1 and γ_1 methyl groups are shown for Leu and Val residues. The inter-methyl connectivity network calculated from the crystal structure using a maximum inter-carbon distance cut-off of 4.5 Å (left) or 9.0 Å (right) is shown with black lines. Note that some peripheral methyl groups are disconnected from the networks. NOEs scale with inverse sixth power of the distance between the two nuclei. (c) Example PRE data with ^{1}H methyl transverse relaxation rates (R_2) measured in the paramagnetic (blue) and diamagnetic (red) state of a protein containing a single nitroxide spin label. The normalized peak intensity (I/I_0) as a function of the relaxation delay (ms) follows an exponential decay that can be fitted to obtain the R_2 value. The difference in R_2 between the paramagnetic and diamagnetic samples ($^H\Gamma_2 = R_{2,para} - R_{2,dia}$) yields the methyl ^{1}H PRE. (d) Same structure and methyl groups as in b, except that all methyl groups within a 25-Å inter-carbon distance of the spin label position (green sphere) are indicated as measurable PREs. The left and right panels depict two different cysteine mutants used to introduce the spin label (green sphere) to record PRE values. PREs scale with inverse sixth power of the distance between the paramagnetic center and the nucleus of interest; however, the larger PRE effect enables measurement over longer distances than NOEs (see text 3.2.2). Nitroxides enable measurement of PREs up to *ca.* 25 Å, while other PRE probes, such as Mn^{2+}, can reach 35 Å. (e) Example [^{1}H,^{13}C]-HMQC spectrum used to measure PCS values, defined as the difference in chemical shift (δ) for a given ^{1}H or ^{13}C nucleus between the paramagnetic and diamagnetic samples (PCS = $\delta^{para} - \delta^{dia}$). (f) Same as d, except that methyl groups within 50 Å of the lanthanide spin labels are marked in blue as probes for which PCS can be measured. PCS values scale with inverse third power of the distance between the paramagnetic center and the nucleus of interest. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

from 2.5 to 8.5 Å could be observed for this system. In addition, ~50% of the expected NOEs in the distance range from 9.5 to 10.5 Å could be detected. However, given that statistically most methyl-methyl NOEs will be short-range, the lack of long-range restraints often leads to fragmentation of the NOE network into multiple parts of high mutual similarity (Fig. 4b). Consequently, significant ambiguity in methyl resonance assignments derived exclusively from NOE restraints is expected [111] (Fig. 6). Furthermore, isolated methyl groups on the protein surface are likely to be excluded from the dominant methyl-methyl NOE networks, due to the short distance range of NOEs. The NOE-based automated approaches can be combined with site-directed mutagenesis to complete the assignment or reduce its ambiguity [111].

3.2.2. Paramagnetism-based assignment approaches

The use of paramagnetic NMR probes to assign methyl resonances in large proteins has been attractive for some time [121,122]. The availability of a high-resolution protein structure combined with knowledge of the location of the paramagnetic probe and an estimate of its magnetic susceptibility tensor allows for the theoretical paramagnetic observables to be calculated and compared to experiment (Fig. 4). Due to the high magnetic moment of electrons, paramagnetic effects can be measured on methyl groups far away from the paramagnetic center (e.g. ~25 Å for the nitroxide spin label). In contrast to the NOE, the long range of the paramagnetic effects can also enable the assignment of isolated methyl groups that reside outside of the high-density clusters of methyl-bearing residues in protein cores [122]. Two different paramagnetic mechanisms have been explored in the context of automatic methyl resonance assignment: paramagnetic relaxation enhancement (PRE) and pseudo-contact shift (PCS) (Fig. 4). The high anisotropy of the magnetic susceptibility tensor χ of most lanthanides leads to rapid relaxation of their unpaired electron spin and causes chemical shifts changes in NMR spectra. In contrast, the nearly isotropic χ of some organic radicals, such as nitroxide, results in slower relaxation of their unpaired electrons, which enables measurements of PREs from nuclear transverse relaxation rates [123].

For methyl resonance assignment, nitroxide spin labels (for PRE) and paramagnetic lanthanide tags (for PCS) must be introduced into the protein in a site-specific manner, taking into consideration the surface accessibility of the labeled sites, minimization of steric hindrance caused by the nearby side-chains, restricted mobility of the tag, and distribution of the tagged sites over the surface of the protein [122,124]. Restricting conformational freedom of the paramagnetic centers is essential for interpretability of both PREs and PCSs. To this end, both PRE- and PCS-based methyl assignment methods make use of specifically designed tags [125–127] and exploit the protein's secondary structure elements [122,124]. For instance, the PCS approach by Lescanne et al. [124] employs "the two-armed caged" lanthanide probe (CLaNP) [125] that covalently links two spatially proximal Cys residues.

To achieve a significant methyl assignment coverage, both PRE and PCS methods require data from more than one spin label (Fig. 4). This is achieved by preparing multiple protein samples, each with a different Xxx to Cys mutation (or by introducing a pair of proximal Cys residues for each CLaNP lanthanide probe [124]), and simultaneous removal (mutation) of any naturally occurring Cys residues [122]. In addition, a diamagnetic reference sample is needed. The paramagnetism-based approaches are thus considerably more demanding in terms of sample production than NOE-based approaches. On the other hand, once well-behaved samples are available, PRE/PCS measurements likely require less NMR measurement time than a 4D NOESY data set.

3.3. Input data for automatic methyl resonance assignment protocols

The input requirements of different automatic methyl resonance assignment protocols are summarized in Tables 1 and 2. In what follows, we discuss the importance of some of the input data for the quality of the assignment result and highlight the requirements characteristic for certain protocols.

3.3.1. Methyl peak lists and methyl residue types

Most of the protocols require as obligatory input $^{1}H/^{13}C$ chemical shifts recorded in [^{1}H,^{13}C]-HMQC or HSQC spectra, which represent the methyl peaks to be assigned by the algorithms (Fig. 6). Exceptions to this rule are MAGMA and PARAssign, which rely on manual analysis of the NMR data to produce an assigned NOE contact list (MAGMA) or manually determined PCS values (PARAssign), which are then used by the assignment algorithms. When provided with "raw", unassigned, peak lists, both protocols can automatically generate the required input by using their auxiliary scripts.

In addition to their chemical shift values, methyl peaks are typically given an arbitrary ID (index or random assignment) that is used both for analysis and output. For best performance, all protocols require information about the methyl residue type (Ile, Leu, Val, Ala, Met, or Thr) for each peak, which is usually specified in the input peak list. To account for overlap in $^{1}H/^{13}C$ chemical shift ranges for Leu and Val methyl groups, most protocols support ambiguity between these methyl types (Table 1). In addition, full ambiguity in methyl residue types or absence of the type specification is supported by MethylFLYA, MAGIC, and PARAssign. Except for MAGIC [112], all protocols report a considerable decline in performance when the residue type information is absent or ambiguous, which typically manifests itself either in decreased accuracy (e.g. MAP-XSII [108]) or increased assignment ambiguity (e.g. MAGMA [111]). MethylFLYA shows complete tolerance to ambiguity between Leu and Val methyl group types, providing nearly identical results with or without discrimination between these two residue types [113].

To provide methyl residue types, one can inspect [^{1}H,^{13}C]-HMQC or HSQC spectra to identify characteristic chemical shifts (e.g. $^{13}C^{\delta1}$ of Ile), or employ residue-specific methyl labeling schemes [41,42]. For PCS-based approaches that critically depend on the ability to identify corresponding methyl peaks in diamagnetic and paramagnetic spectra, the separate labeling of each methyl amino acid type [34] has been recommended [124]. In addition to unambiguously establishing residue types, such labeling significantly reduces spectral crowding.

3.3.2. Methyl-methyl NOEs

The NOE-based methods require manually or automatically picked and expert-filtered NOESY peak lists as input (Fig. 5). Columns of the NOE lists contain two carbon and one proton (3D), or two carbon and two proton (4D), frequencies of spatially proximal methyls. With the exception of MAGMA (see above), all other methods support unassigned peak lists as input. MAGIC additionally requires peak heights, as proxies for NOE intensities, to be provided in the input peak list, as it uses the intensity information in its assignment protocol (see section 4.3). Each dimension of the peak list is associated with a user-defined tolerance, which is expressed in ppm units and used in the automatic assignment of NOEs to the methyl peaks from the 2D [^{1}H,^{13}C]-HMQC or HSQC peak list. The tolerances should account for small shifts of peak maxima between different spectra, and ambiguity inherent to the NOE assignment (typical in the range of 0.2–0.4 ppm for ^{13}C and 0.025–0.04 ppm for ^{1}H). In the MAGIC study more stringent tolerances were proposed (0.1 ppm for ^{13}C and 0.01 ppm for ^{1}H) [112].

Compared to 3D NOESY data, the additional ^{1}H (or ^{13}C) dimension provided by a 4D data set can considerably reduce NOE

Table 1
Input data requirements for the NOE-based automatic methyl resonance assignment protocols.

Input data type	MAP-XSII	FLAME nGO2.0	MAGMA	MAGIC	Methyl FLYA
Peak picked, unassigned methyl-methyl NOEs (3D or 4D peak list)	Yes	Yes (Sparky)		Yes[a]	Yes
methyl-methyl NOE contact list			Yes[b]		
Measured [^{1}H,^{13}C] chemical shifts (2D peak list with residue types indicated)	Yes	Yes		Yes[c]	Yes[c]
Predicted [^{1}H,^{13}C] chemical shifts (obtained from prediction software)	Yes	Yes			
Random initial assignment list		Yes			
Protein structure (PDB format with added protons)	Yes	Yes	Yes	Yes	Yes
Protein sequence file				Yes	Yes
Known unambiguous methyl assignments (assigned peak list)	Optional		Optional		Optional
LV-geminal methyl pairing information[d]		Optional	Optional	Optional	Optional
LV methyl residue-type ambiguity (flag in the input file)			Optional		Optional
measured [^{1}H,^{15}N]-chemical shifts (peak list)		Optional			Optional
peak picked amide-methyl NOEs (3D or 4D peak list)		Optional			Optional
experimental PREs	Optional	Optional[e]			
predicted PREs	Optional				
protein regions with conformational exchange (input file with residue numbers)				Optional	

For each program, 'yes' indicates obligatory input.
[a] 3D CCH-NOESY peak list with intensity information.
[b] Two-column list with arbitrary methyl group IDs that annotate each NOE cross peak with two methyl resonances from the 2D HMQC or HSQC spectrum; IDs contain residue type labels.
[c] Residue type specification is optional.
[d] Given as described in [110] for FLAMEnGO2.0, as combined labels in the NOE contact list for MAGMA, as an addition to the 2D peak list for MAGIC, or in the form of a theoretical HCcCH TOCSY peak list for MethylFLYA.
[e] In semi-quantitative format, see Kim et al. [129].

Table 2
Input data requirements for the paramagnetism-based automatic methyl resonance assignment protocols.

Input data type	PRE-ASSIGN	PARAssign	Possum
Methyl ^{1}H-PREs[a]	Yes	Yes	
Measured diamagnetic and paramagnetic [^{1}H,^{13}C] chemical shifts[b]		Optional	Yes
List of predicted PREs based on the structure (listed with residue numbers and types)	Yes		
Initial rotational correlation time τ_c of the PRE-interaction vector	Yes		
Independently determined $\Delta\chi$-tensor parameters		Yes[c,d]	Yes
Specified atom names, stereospecificity, and double Cys mutation sites		Yes[d]	
Protein structure (PDB format)	Yes	Yes	Yes
Initial guess for the nitroxide spin label coordinates	Optional		
LV-geminal methyl pairing and pairing of Ile-δ_1,γ_2 methyls (from HCCH-TOCSY experiments)			Optional
Discrimination between Ile-δ_1/γ_2, Leu-δ_1/δ_2, and Val-γ_1/γ_2			Optional

For each program, 'yes' indicates obligatory input.
[a] Listed with the corresponding ^{1}H,^{13}C-chemical shifts in a separate list for each methyl residue-type for PRE-ASSIGN; a list of experimental PCS values with ambiguous or unambiguous residue-types indicated for PARAssign.
[b] Separate for each double Cys mutant for PARAssign in order to generate the input PCS list by an auxiliary script; including methyl residue types for Possum.
[c] Initial values.
[d] Listed in a configuration file in JSON format.

assignment ambiguity. As such, 4D peak lists are supported by MAP-XSII and FLAMEnGO2.0 [108,110], and recommended by MAGMA [111] and PRE-ASSIGN [122]. The MAGIC algorithm currently supports only 3D CCH-NOESY data, which are reportedly preferred due to faster collection and larger acquired NOE data sets [112].

3.3.3. PREs and PCSs

Methyl ^{1}H-PREs constitute the major restraints for the MC-based assignment algorithm PRE-ASSIGN [122]. The experimental data are supplied as lists of a series of methyl ^{1}H-PREs (in s^{-1}) measured with increasing relaxation time delays (Table 2). The first three columns of the lists contain a random methyl group identifier and its measured ^{1}H, ^{13}C chemical shifts, with the measured PRE-value for one of the relaxation time delays in the fourth column. Separate lists should be supplied for each methyl residue type to reduce ambiguity in the assignment [122]. To generate the predicted PREs, the protein structure file needs to be modified separately for each of the protein mutants to indicate the position of the mutated Cys residue. Taking such modified structure files as input, Xplor-NIH [130] scripts are used to generate the predicted PRE files. The program supports simultaneous optimization of the nitroxide position and the rotational correlation time (τ_c) values based on the initially predicted PRE data. The predicted PREs can then be recalculated with the optimized values of the parameters and the new PRE values supplied to the assignment routine.

Methyl PREs are supported as additional restraints by MAP-XSII [108] and FLAMEnGO2.0 [110]. FLAMEnGO2.0 employs a semi-quantitative interpretation of the measured PREs by dividing methyl peaks into three categories depending on the percentage of the signal lost due to the PRE (i.e. strongly, moderately, and weakly affected) [129,131]. An additional category is introduced for unobserved peaks. On the benchmark of the FLAMEnGO2.0 study, PRE restraints alone were reported to be insufficient to generate highly reliable methyl assignments [110]. However, in conjunction with NOEs, they had a significant positive impact on the calculation speed and accuracy, and helped resolve assignment ambiguities due to overlapping peaks [110,129].

The PCS-based method PARAssign [124] and the earlier Possum approach [121] use exclusively methyl-PCSs as input methyl NMR data. To acquire input PCSs, differences in methyl chemical shifts recorded in the [^{1}H,^{13}C]-HMQC/HSQC spectra of the paramagnetic protein and the diamagnetic reference need to be measured. Manual comparison of the spectra to pair up methyl peaks can be challenging due to the low dispersion of methyl resonances, overlapping peaks and signal broadening in the paramagnetic spectrum caused by PRE effects [121,123]. These issues can be addressed by preparation of protein samples with single and stereospecific amino acid labeling schemes, and by acquiring an additional PCS data set using a second, weaker, paramagnetic lanthanoid that gives rise to resonances situated between the dia-

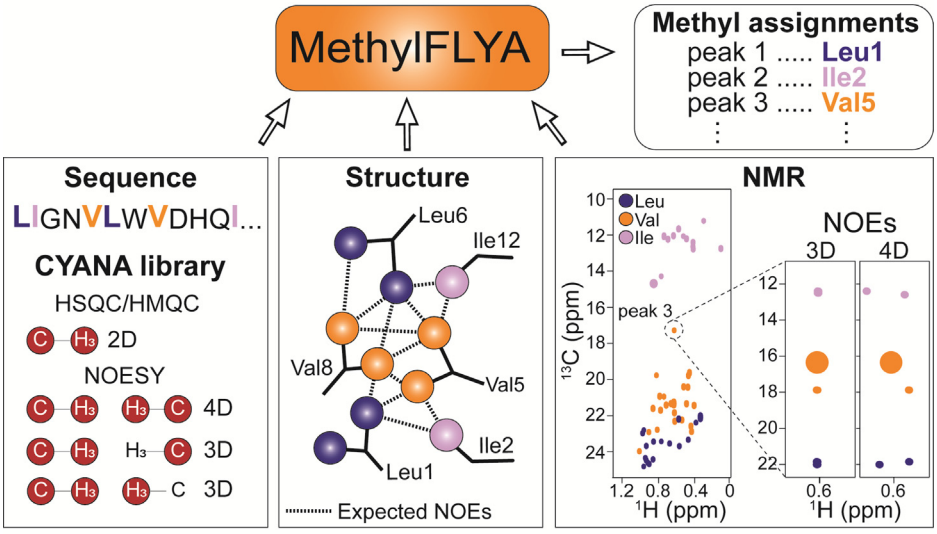

Fig. 5. Illustration of the input requirements for the MethylFLYA automatic assignment protocol. The user supplies the protein sequence (left), a 3D structure (center), and experimental methyl NMR data: a 2D [^{1}H,^{13}C]-HMQC/HSQC peak list and a 3D or 4D methyl-methyl NOESY peak list. Optionally, peak lists can be obtained by automatic peak picking with CYPICK [113,128].

magnetic and the original paramagnetic peaks in the spectrum [124]. The similarity between PCSs for methyl ^{1}H and ^{13}C nuclei can be exploited to resolve ambiguities in peak pairing [124]. Effort put into peak pairing determines the completeness of the input data, and is critical for assignment success [124]. Finally, relating the assigned diamagnetic resonances to the wild-type protein spectrum should be straightforward, provided that no major structural changes occur due to the introduced mutations and paramagnetic labeling.

Prediction of methyl ^{1}H-PCSs from a protein structure requires knowledge of the magnetic susceptibility tensor and the position of the paramagnetic lanthanide. In the Possum protocol [121], knowledge of the $\Delta\chi$-tensor parameters is assumed, which in turn assumes known assignments of amide resonances of the diamagnetic protein [132], or known assignments of a subset of methyl resonances. In contrast, the PARAssign approach proposes usage of a 'two-armed' paramagnetic probe that attaches to two Cys residues and for which the lanthanide location and the approximate orientation of the $\Delta\chi$-tensor axes can be well estimated [124,125]. The initial user-provided estimates are optimized by the algorithm during the assignment procedure (see Section 4.5). Calculating theoretical PCSs from different theoretical attachment sites is encouraged to avoid unfavorable correlation of PCS measurements from different sites, which reduces information content and negatively impacts assignment outcome [124].

3.3.4. Geminal methyl linkage

The NOE-based approaches FLAMEnGO2.0, MAGMA, MAGIC, and MethylFLYA can make use of an additional restraint that links the two geminal methyl groups of a Leu or Val residue. In the computation-time sensitive exhaustive search algorithm of MAGMA, the geminal linkage information is used to join methyl-methyl NOEs from the two methyl groups to a pseudo Leu/Val residue vertex, which reduces the effective total number of methyl peaks to be assigned and significantly reduces the calculation time [111]. The benefits of the restraint were also noted with MethylFLYA [113] and in the MAGIC study where the geminal information led to a four-fold decrease in computation time, and a small improvement in assignment accuracy [112]. In their present versions, MAP-XSII and PARAssign do not make use of this source of information.

To obtain geminal linkage information, the MAGMA study [111] recommends preparation of an additional protein sample in which both methyl groups of Leu and Val residues are protonated and ^{13}C-labeled. A short-mixing time 3D or 4D NOESY spectrum can then be used to identify the two methyl groups that originate from the same Leu or Val residue [18]. Alternatively, through-bond TOCSY-based approaches have been proposed to generate intra-residual methyl-methyl correlations [121,133], although this requires ^{13}C labeling that links the backbone to terminal methyl groups, and therefore is only applicable to relatively small proteins. Use of a precursor for Leu and Val methyl labeling that results in a protonated γ or β-methine position has been proposed in combination with a 3D TOCSY-HMQC experiment to link the two geminal methyl groups to the shared methine proton [110]. Recently, a 3D-HMBC-HMQC experiment has been devised to link the geminal methyl groups of Leu and Val [134]. Similarly to the short-mixing time NOESY strategy, this approach requires a double, pro-*R* and pro-*S*, ^{1}H,^{13}C-methyl labeled sample in an otherwise deuterated background. In the HMBC element of the pulse sequence a three-bond $^{3}J_{CH}$ coupling transfer is achieved between the protons of one of the two geminal methyl groups and the carbon of the other methyl group. The one-bond couplings are suppressed during this transfer. The following HMQC element is used for the standard one-bond $^{1}J_{CH}$ coupling. Thus, both methyl carbon frequencies and the starting proton frequency are recorded by the sequence, which allows pairing of the geminal methyl pairs. For well-resolved [^{1}H,^{13}C]-HMQC spectra, all geminal methyl pairs of Leu and Val are expected to be linked. If there are overlaps in the region of Leu/Val resonances, the approach can be combined with a long-mixing time 4D HMQC-NOESY-HMQC spectrum recorded on the same sample to obtain intra-residue NOEs as well as inter-residue NOEs. The intra-residue NOEs can then be identified based on the 3D HMBC-HMQC spectrum. As an alternative, only one of the Leu/Val-methyl groups can be labeled using stereospecific labeling [43], which reduces spectral crowding and is particularly useful for the PCS-based assignment approaches [121,124].

3.3.5. Predicted methyl chemical shifts

Structure-based prediction of methyl chemical shifts is used as an additional source of assignment restraints in the MAP-XSII and FLAMEnGO2.0 protocols. Accurate prediction of chemical shifts

from static structures is a difficult task due to their complex dependence on geometric factors and electronic environment, and the fact that measured shifts represent Boltzmann-weighted averages [135,136]. Most available structure-based chemical shift predictors are restricted to backbone resonances, but a few, e.g. PPM_One [136], ShiftX2 [137], and CH3Shift [138], provide side-chain resonance predictions, including methyl groups.

Predicted chemical shifts are used directly by the MAP-XSII and FLAMEnGO2.0 protocols, that include a chemical shift term in their scoring functions to favor assignment solutions with minimized discrepancy between the measured and the predicted methyl shifts (see section 4.1). In MAGMA, a conservative use of chemical shift predictions was tried in order to discriminate between ambiguous assignment options of individual methyl groups. To that end, assignments with predicted shifts closer to the measured ones and within the prediction tolerance were ranked higher. When used in this way and context, however, the predictions were not reliable enough to provide additional unambiguous methyl assignments [111].

3.3.6. Input protein structure

As structure-based assignment strategies, all the discussed methods assume a high degree of compatibility between the available crystal structure of the protein and structural restraints derived from NMR observables (NOEs, PREs, PCSs) (Figs. 4 and 5). Thus, availability of high-quality structures and their compatibility with the input NMR restraints are important determinants for the success of structure-based methyl assignment methods. The compatibility is not a small issue, as the NMR observables report on averages sampled by different molecular states, and are affected by nuclear spin interactions (e.g. spin diffusion) and relaxation [139]. The NMR restraints best fit an ensemble of structures whereas, for assignment purposes, typically only one static crystal structure is provided. The accuracy of the crystal structure, on the other hand, can be affected by crystal packing interactions and artificial modifications of the protein introduced to promote its crystallization. In addition, poor X-ray crystallographic resolution can limit the accuracy of atom positions, especially in protein side chains.

The impact of different input protein structures was addressed in the MAP-XSII study, where a significant decrease in assignment coverage and lower accuracy was noted for alternative crystal structures [108]. On their benchmark, the program's score could be used to identify the better-quality structures, i.e. those that were more compatible with the NOE data and yielded better assignment outcomes. The NOE score alone could not be used straightforwardly to discriminate between different structures in the MAGMA study, in which differences in assignment accuracy for different input structures were also noted, and checking assignment consistency across different structures was advised [111]. The MethylFLYA study also noted differences in assignments from different protein structures. On the presented benchmark, however, the structures that gave rise to more 'strong' (i.e. classified as reliable by the program), assignments in MethylFLYA also featured higher assignment accuracy [113]. In the MAGIC study, discrepancy in the side-chain rotamer angles between the crystal and solution-state structures was identified as a source of differences in the measured and expected NOE patterns that can affect the automatic assignment [112].

Also, the PCS-based Possum method noted instances of assignment swaps due to differences between the protein structure in solution and that captured in the only available crystal structure [121]. The authors concluded that reliable comparison of predicted and measured PCS values critically depends on the accuracy of the available protein structure, and noted a likely sub-optimal performance of the program in the presence of flexible protein segments [121].

Despite such reports, the sensitivity of the methods to the quality of the input structures has not yet been systematically evaluated. In addition, all methods except MethylFLYA presently allow input of only a single protein structure file. The potential usefulness of homology models for automatic assignment of methyl resonances in the absence of a high-resolution structure of the protein target remains to be addressed. However, given the aforementioned reports, it is unlikely that such approaches would lead to high-quality assignment results.

4. Automated methyl assignment algorithms

The high dimensionality of the search space represents a significant challenge to automatic methyl resonance assignment approaches. Given the factorial increase in the number of assignment possibilities with the number of assignable methyl peaks, different algorithmic strategies have been explored to avoid a "combinatorial explosion" while still ensuring adequate sampling and scoring of assignment solutions. In what follows, we summarize the algorithmic approaches and their performance against their respective benchmarks.

4.1. Monte-Carlo sampling – MAP-XSII and FLAMEnGO2.0

Sampling and evaluating plausible methyl assignment solutions using a Monte Carlo (MC) swapping routine is a feature of the NOE-based algorithms MAP-XSII [108] and FLAMEnGO2.0 [110], and the PRE-based method PRE-ASSIGN [122]. In these protocols, an initially random assignment given to each methyl peak is iteratively swapped and every random swap (MC move) then evaluated by a global scoring function. The function to be maximized is typically a linear combination of different terms that evaluate agreement between measured and predicted NMR data, with the range of terms and their exact form differing between the protocols. Additional differences are found in the criteria for accepting or rejecting swaps, and the details of the probability functions employed in the step.

The central component of the scoring functions of MAP-XSII and FLAMEnGO2.0 is the NOE term that quantifies the agreement between the measured methyl-methyl NOEs and those simulated based on the input structure. To simulate the NOEs, MAP-XSII uses a fixed distance cut-off that is provided by the user, whereas FLAMEnGO2.0 treats the cut-off as a variable that is optimized within a user-determined range during the assignment search. The NOE term in FLAMEnGO2.0 is defined by a Gaussian function that is maximized by reducing the differences in positions of the measured NOE and their plausible matches in the set of simulated NOEs. The simulated peak positions are generated for every tentative assignment combination from known ^{1}H,^{13}C positions in the methyl-TROSY spectrum and the current distance cut-off [110]. The NOE term is evaluated 'fuzzily' to provide a 'percentage of confidence' of an assignment after each swap and to account for NOE ambiguity inherent in the assignment process [109,110].

The NOE term of the MAP-XSII scoring function is also a multidimensional Gaussian that evaluates the scaled difference in the positions of the NOEs and their corresponding methyl peaks from the methyl-TROSY spectrum. The term is scaled in the context of every tentative methyl assignment solution by higher scoring of symmetric or reciprocal NOEs that are observed in the NOESY planes of the two methyls that give rise to an NOE [108]. The protocol considers ambiguity in automatic attribution of measured NOEs to the methyl peaks of their origin. To this end, all possible assignment options for ambiguous NOEs are considered in the con-

text of the current tentative methyl assignment, before the best one is selected and scored. Both protocols use unassigned NOESY peak lists and, in their present implementations, do not make use of NOE intensities.

In addition to the NOE score, an obligatory chemical shift score is added to the scoring functions of both protocols to evaluate discrepancies between the measured and predicted methyl chemical shifts, scaled to the prediction error. Additional scoring terms include optional restraints, such as methyl-PREs (both protocols) or methine to methyl TOCSY correlations (FLAMEnGO2.0). To obtain an estimate of assignment confidence at the level of the individual methyl peaks, both protocols rely on multiple, parallel iterations with optimized settings of the parameters. The highly reliable assignments are consistent across all or most of the parallel runs [108,110].

The target function of the PRE-ASSIGN method by Venditti et al. [122] comprises a PRE-energy term that evaluates the discrepancy between the observed and calculated methyl ^{1}H-Γ_2 values, defined as the difference in methyl proton transverse relaxation (R_2) rates between the paramagnetic and diamagnetic samples. An MC sampling protocol iteratively swaps the originally random methyl assignments. Swaps that lower the PRE-energy term are always accepted, while the rest are accepted with a Boltzmann-factor probability [140]. A manual comparison of the predicted and observed methyl-methyl NOE connectivity patterns can be performed at the end of the protocol to refine the PRE-based assignment. The careful design of the MC protocols addresses the ambiguity in the input NMR restraints. Nevertheless, local minima remain an issue inherent to the approximate optimization that becomes more severe with increasing number of methyl groups and consequent increase in the search space.

4.2. Exact graph comparison – MAGMA

In contrast to the MC-based methods, Methyl Assignment by Graph MAtching (MAGMA) relies on graph algorithms to generate methyl assignments [111]. The MAGMA protocol combines exact graph comparison algorithms with a heuristic that sorts vertex matching priorities to ensure maximum pruning of the search space. At the beginning of the protocol, MAGMA defines a simple data graph from a NOE contact list that corresponds to the measured methyl-methyl NOE cross-peaks (Fig. 6, Table 1). The graph is vertex-labeled, with labels on vertices reflecting methyl residue-types that are provided by the user directly in the NOE contact list. From a high-resolution crystal structure of the protein, a simple structure graph is calculated by computing Euclidean distances between side-chain methyl carbons, restricted to the methyl labeling scheme defined in the input file for the calculation. If the standard ILV labeling [40,141] is employed with intra-residue connectivity between two geminal Leu-δ_1,δ_2 and Val-γ_1, γ_2 methyl groups established by a separate experiment (see Section 3.3.4), the carbon coordinates for each geminal pair are averaged and the inter-carbon distances are computed from the average coordinates. To determine the optimal distance cut-off that defines edges of the structure graph, short iterations of the protocol are run with increasing distance cut-off values, while monitoring the percentage of explained inter-methyl NOEs as a function of the cut-off. On the MAGMA benchmark, it was established that the optimal distance cut-off typically corresponds to 10 Å inter-carbon distance between two methyl groups, or the distance at which all methyl-methyl NOEs are explained, if that distance is below 10 Å [111]. The derived optimal distance cut-off is introduced into the MAGMA input file for the complete methyl assignment calculation.

The assignment calculation maps the vertices from the data graph to the vertices of the structure graph in order to find the mapping(s) that maximize the number of explained data graph edges, i.e. inter-methyl NOEs. The two graphs to be compared are typically large, and using exact graph comparison algorithms can lead to unacceptably (exponentially) long calculations [143]. For this reason, it is important to find a good starting point for the graph-matching such that high-quality solutions are found early in the search. MAGMA attempts to optimize the order of the data graph traversal, i.e. walking the data graph such that the most densely connected vertices are visited first. In addition, for each of the data graph vertices, i.e. methyl peaks, the optimal order of structure graph vertices to be tested as plausible assignment solutions needs to be found, such that high-scoring mappings are tried first. MAGMA's heuristic simultaneously optimizes both the data graph traversal and the order of vertex mapping, using a combination of

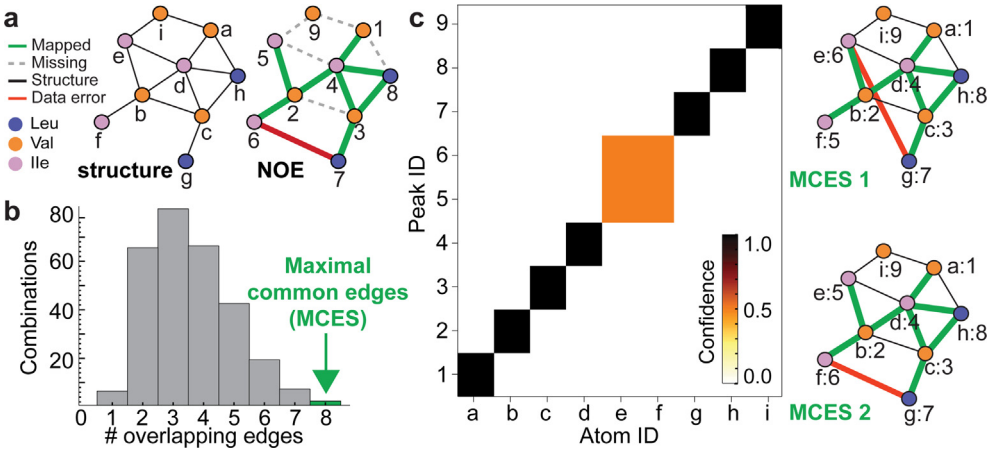

Fig. 6. MAGMA assignment strategy [111]. (a) Two simple, vertex-labeled graphs are defined based on the proximity of methyl groups in an available protein structure (structure graph), and the inter-methyl connectivities measured with methyl-methyl NOESY (NOE graph). The edges of the structure and the NOE graph are the short distances between methyls in the structure and the measured inter-methyl NOEs, respectively. The vertex labels indicate different types of methyl-bearing amino acid residues. This information is known for the structure graph. For the NOE graph, it is provided to the algorithm by the user. (b) The algorithm searches for a set of all maximal common edge subgraphs (MCESs) between the structure and the NOE graph that maximize the mapping of methyl-methyl NOEs to the structure derived inter-methyl connections and therefore represent the desired assignment solutions for the vertices of the two graphs. Many possible assignment combinations between the vertices of the structure graph and the vertices of the NOE graph lead to suboptimal solutions with a low number of overlapping edges. The algorithm heuristically attempts to minimize the time spent evaluating the low scoring solutions, and instead quickly reach the sampling of the MCES solutions. (c) This is achieved through vertex-sorting heuristics coupled to an exact algorithm that solves for MCESs [111,142].

breadth-first graph search and the Hungarian method (see [144,145] and Section 4.5) [111].

Once the optimal orders have been established, the graphs are tested for subgraph-isomorphism, which evaluates if a subgraph exists in the structure graph that is structurally equivalent (isomorphic) to the data graph. If true, all vertex mappings (i.e. assignment solutions) for which subgraph-isomorphisms hold are retrieved using the VF2 algorithm [146]. If false, the graphs are matched using the maximal common edge subgraph method as described by McGregor [142]. MAGMA's implementation of the McGregor algorithm is adapted to evaluate all maximal common subgraph solutions by looking for the structurally equivalent subgraphs of the data and the structure graph that all maximize the number of the data graph edges (Fig. 6). Here it should be noted that, due to their matching heuristics, the graph comparisons focus on all possible maximal assignment solutions, rather than all possible solutions, which would inevitably lead to exponential sampling time. For large data graphs that consist of multiple connected subgraphs, each of the subgraphs is evaluated separately, starting from the largest one. At the end of the protocol, the solutions from different subgraphs are pooled together. These solutions can be used to re-run the protocol with all the subgraphs evaluated simultaneously but restricted to the matching option identified in the individual runs.

As an output, MAGMA provides a list of all possible assignments for every methyl resonance, and thus fully describes any existing assignment ambiguity (Fig. 6). The resonances with only one assignment option are considered reliable, while for the others a site-directed mutagenesis study can be designed, guided by the revealed assignment ambiguity. The MAGMA approach offers high assignment reliability due to the completeness of the assignment search. However, it can suffer from incorrectly assigned NOEs, especially if multiple incorrect NOE assignments are attributed to a single methyl peak. It is therefore crucial to include only reliable NOEs, which can be achieved by requiring NOE reciprocity (i.e. presence of transposed peaks) and a signal-to-noise ratio threshold [111].

In the context of graph comparison, it is worth mentioning another NOE graph comparison based approach [114], which at the moment does not have a dedicated software implementation. The residue types of methyl peaks need to be established first through inspection of ^{13}C chemical shifts and through-bond correlations of methyl resonances with C^{α} and C^{β} chemical shifts. This is followed by a manual inspection of the best match between the theoretical (structure-based) and observed Ile-Ile, Ile-Val, Ile-Leu, Val-Val, Val-Leu, and Leu-Leu NOE clusters. Mapping between the clusters is hierarchical and starts from the largest observed NOE cluster, subsequently progressing towards the smallest cluster, favoring Ile-only or Ile-rich NOE clusters in the ordering. The sequential mapping of clusters from the largest to the smallest reduces the number of possibilities for the matching of the smaller clusters. Once the theoretically observed NOE cluster correspondences have been established, i.e. the NOE data have been mapped to regions of the X-ray structure, a qualitative inspection of the NOE connectivity within each cluster is performed to assign individual methyl resonances from an observed NOE cluster to the methyl-bearing residues in the associated 'theoretical' NOE cluster. Ambiguity in the initial NOE cluster mapping is resolved through an inspection of the theoretical and predicted NOE correspondences of different clusters. A qualitative inspection of NOESY cross-peak intensities in the later stages of the analysis can be performed to reduce assignment ambiguities. At present, the protocol relies on human assessment of the data matching ambiguity and an empirical determination of the optimal distance cut-offs for the definition of the theoretical NOE clusters.

4.3. Network density approach – MAGIC

To avoid the combinatorial explosion of the assignment problem, the network density approach of the MAGIC algorithm performs an exhaustive assignment search locally and hierarchically. Early assignment of the dense portions of the NOE network is prioritized, with only highly scored assignment solutions, generated in the iterative local search, allowed to advance forward. The protocol effectively combines NOE and methyl resonance assignment, providing as its output the assigned peak lists for the 3D NOESY and the 2D HMQC/HSQC spectra [112].

Like the earlier MAP-XSII and FLAMEnGO2.0 protocols, MAGIC relies on unassigned NOESY peak lists. Manual NOE assignment is replaced by an NOE-assignment procedure that precedes the methyl assignment algorithm. NOE cross-peaks from the unassigned 3D NOE peak list are first matched to the 2D HMQC methyl peaks solely based on ^{1}H, ^{13}C tolerances, allowing for multiple ambiguous assignments of methyl NOE contacts to one 2D methyl peak. To remove the many false positives, the NOE assignments are first filtered for the presence of symmetric NOEs in the NOESY planes of the plausible methyl NOE assignments, similarly to the reciprocity principle utilized by the MAGMA protocol [111]. Each remaining connection is scored to give higher confidence to connections assigned to the well-resolved 2D HMQC peaks that are less ambiguous (i.e. for which fewer NOE assignments are possible) and connect methyls with shared NOE connections (i.e. those with higher similarity of NOESY planes).

Confidence scores are stored in the peak adjacency matrix P, while an equivalent model network adjacency matrix M stores inter-methyl carbon–carbon connectivity extracted from a high-resolution protein structure. All carbon–carbon contacts shorter than a short distance cut-off (e.g. 7 Å) correspond to unity in M, with the rest of the values decreasing linearly from 1 to 0 for contacts in the distance range between the short and long cut-offs (e.g. 7–10 Å). The two distance cut-offs can be user-specified. On the MAGIC benchmark, the long distance cut-off of 10 Å was shown to achieve optimum coverage, accuracy, and calculation speed, which is equivalent to the finding reported on the MAGMA benchmark [111] (see Section 4.2). The peak network density matrix P^2 in MAGIC is used to reveal methyl peaks with high interconnectivity in the network, or the peaks that form dense clusters. The score matrix S, on which the algorithm operates, is defined as the Hadamard (element-wise) product of P and I, where I is a matrix of relative intensities of the NOE cross-peaks, indexed in the same way as P, i.e. $S_{ij} = P_{ij}\, I_{ij}$ for all index pairs ij.

The matrices P^2, S, and M are defined at the beginning of the MAGIC protocol, and the clustering threshold T_C is initialized to the maximal non-diagonal value in P^2. This threshold determines which connections form the initial peak clusters, and it is iteratively decreased to redefine the clusters (peak subnetworks). At each iteration, i.e. each re-definition of the peak subnetworks, the assignment of the subnetworks to M is exhaustively sampled such that all possible assignment permutations are evaluated. With the iterative decrease in T_C, the clusters are expanded up to a defined threshold size, at which point the best scored assignments are transferred to the global assignment process, the final stage of the algorithm in which all permutations of the earlier output cluster assignments are evaluated. The algorithm thus samples exhaustively locally, evaluating each assignment solution by summing all elements of the Hadamard product of S and M, which results in the addition of only those connection scores for which a peak-peak connection in P has a match in a methyl-methyl connection in M. Such scoring aims at finding solutions for which high-confidence, high-intensity NOE contacts are assigned to shorter methyl-methyl distances in the protein structure.

In contrast to MAGMA, the local focus of the MAGIC algorithm aims at decreasing the computational time relative to an exact search. Given that it does not consider the entire methyl connectivity network at once, it risks omitting the correct solution due to potentially erroneous connections in the local clusters, which can drive the assignment of the clusters to incorrect places in the search space, or due to connections that are critical for assignment accuracy but do not become a part of the clusters that have reached the output threshold size.

4.4. Evolutionary algorithm – MethylFLYA

MethylFLYA [113] defines a network of inter-methyl connectivity as measured by NOESY experiments and compares it to a network of expected methyl connectivity extracted from a high-resolution protein structure (Fig. 5), which makes it conceptually similar to MAGMA [111] and MAGIC [112]. However, mapping between the two networks is based on evolutionary optimization that works on a population of plausible assignment solutions in combination with a local optimization routine [8]. To assess the reliability of individual methyl group assignments, FLYA performs multiple independent parallel runs of the evolutionary protocol and consolidates their results into a final consensus assignment solution to define as reliable, or 'strong', those methyl assignments that have a high extent of consistency among the parallel runs [8], an approach equivalent to that of MAP-XSII [108] and FLAMEnGO2.0 [110].

4.4.1. MethylFLYA input data requirements

MethylFLYA takes as input peak lists from any combination of multidimensional through-bond or through-space NMR experiments [8,147]. Comparable to MAP-XSII, FLAMEnGO2.0 and MAGIC, the MethylFLYA protocol does not require the input NOESY peak lists to be annotated with the methyl resonances from the 2D HMQC or HSQC spectrum. However, knowledge of the amino acid types of methyl signals, if available, can be used to split the NOESY lists using CYANA commands. The information about the methyl residue types is given by providing a separate 2D HMQC or HSQC peak list for each type (e.g I, L,V, or joint LV). Next to the input peak lists, a protein sequence file, a 3D structure file, and a CYANA library file must be supplied.

MethylFLYA showed robust performance on NOESY peak lists containing ambiguous NOEs, i.e. NOEs that cannot be assigned to a single pair of methyl resonances, and was tolerant to differences in input crystal structures. The protocol also proved robust in applications featuring highly ambiguous and partially incorrect input information, derived from manually or automatically picked methyl-methyl NOESY and 2D [^{1}H,^{13}C]-HMQC spectra, and a 'blind' allocation of methyl resonance residue types based on the BMRB-derived chemical shift statistics [113]. Furthermore, MethylFLYA was shown to perform equally well with or without discrimination between Val and Leu resonance types. However, a significant improvement in performance was found when the two geminal methyl groups of Leu and Val residues were linked, as in the previous reports on MAGMA and MAGIC [113]. This information should be introduced to MethylFLYA in the form of a TOCSY peak list, regardless of which experiment was used to link the geminal methyl groups.

4.4.2. MethylFLYA optimal parameter values

The known protein structure and NOESY magnetization pathways specified in the CYANA library are used by MethylFLYA to construct the network of expected inter-methyl connectivity. The expected NOEs are attributed distance-dependent probabilities, which reflect a reduced likelihood of observing long-distance NOEs. The probability for every distance can be set by the user, who can consider the guidelines from an optimization performed on a benchmark of proteins with known methyl assignments [113]. To determine the optimal distance cut-off for a MethylFLYA calculation, the user can monitor the fraction of the measured peaks that are explained over a range of distance cut-offs. Cut-offs that account for a mapping of 80% or more of the measured peaks to the expected peaks are considered optimal. The optimal distance cut-offs for MethylFLYA and MAGMA were found to be generally comparable [113]. To prevent assignment errors due to differences in the expected methyl connectivity at cut-offs slightly different from the optimum, MethylFLYA is also run with distance cutoffs ±0.5 Å from the optimum. The protocol using three slightly different distance cut-offs resulted in a favorable performance of MethylFLYA on the aforementioned benchmark [113], yielding 459 correct assignments out of the total of 465 reliable methyl assignments, compared to MAGMA (333 out of 335), MAP-XSII (216 out of 259), and FLAMEnGO2.0 (113 out of 135).

4.4.3. MethylFLYA algorithm

MethylFLYA employs the general FLYA automated assignment algorithm [8] that has previously been applied to a variety of different chemical shift assignment problems [148–154], including automated assignment based exclusively on data from NOESY spectra [155,156] and methyl assignment in membrane proteins [157]. To map the expected network of methyl connectivity to the measured one, MethylFLYA employs an evolutionary algorithm, which starts from a population of random assignment solutions (i.e. individuals) that is optimized over successive generations (200 by default). A scoring function attributes an overall, global score to every individual in a population, which allows their ranking. Throughout the protocol, the top ranked individuals of a previous generation are selected for mutations and recombinations to form the next generation. The rates of mutations are optimized using a simulated annealing protocol, which starts with high rates in the early generations and subsequently reduces them following a temperature schedule [8,158]. In addition to the global score, a local score is attributed to every atom of an assignment solution, based on the assignment quality of its immediate neighbors in the network of the expected peaks. To maximize the local scores, a local optimization step is employed that identifies poorly scoring parts of individuals in a population and improves their local scores through iterative reassignments of the "problematic" parts [8,158]. The entire protocol is repeated 100 times from different random starts at three distance cut-offs, i.e. the optimum, optimum +0.5 Å, and optimum −0.5 Å [113]. Finally, a consolidation step evaluates the assignment of every methyl group over the 300 independent runs to identify 'strong' (i.e. reliable) methyl assignments. The strong assignments are those consistent across at least 80% of the runs.

Like MAP-XSII and MAGMA, MethylFLYA also supports input of any known methyl assignments, which can help increase performance. If additional experimental data, e.g. NOEs involving the backbone amide groups or data from through-bond experiments, are available, these can be included readily in the input for MethylFLYA and will be used simultaneously with the methyl NOESY data for the assignment.

Since MethylFLYA does not perform exhaustive combinatorial searches, its computation time increases only moderately (about linearly) with system size. Including additional data does not potentially lead to a "combinatorial explosion". Even though the optimization algorithm does not exhaustively explore the search space, the convergence of MethylFLYA is such that different individual runs of the optimization algorithm in general yield comparable final global score values [8], indicating that local minima are not a severe problem of the approach.

4.5. Hungarian method – PARAssign

The PCS-based protocol of PARAssign [124] aims at finding assignments for methyl resonances that minimize the difference between the measured proton PCS values and those expected from methyl positions in a known protein structure. An iterative algorithm combines methyl assignment with optimization of the $\Delta\chi$-tensor anisotropy parameters and lanthanide location for each paramagnetic center. A similar combined optimization strategy has also been proposed in the PRE-based approach by Venditti et al. [122].

The assignment part of the protocol uses the Hungarian method (Kuhn's algorithm) [144,145], which operates on a matrix of normalized differences between the predicted and measured PCS for each methyl peak/methyl atom assignment combination. The normalized differences in the matrix are sums over all paramagnetic centers for which PCSs were measured divided by the total number of paramagnetic centers. A scoring factor threshold is employed to discard assignment possibilities for which the value of the normalized difference would be inappropriately high. The threshold is optimized to maximize the number of assignments and retain the quality (i.e. not compromise the accuracy). For each tentative assignment calculated with the Hungarian method, the initially given $\Delta\chi$-tensor parameters are subjected to least-squares optimization that minimizes the squared difference between the experimental PCSs and those calculated based on the current assignment. The predicted PCSs, recalculated based on the fitting parameters, are used to populate a new assignment matrix for the next iteration of the Hungarian method. This iterative assignment and tensor parameter-fitting procedure is repeated until convergence or until a maximal number of iterations has been reached. The entire protocol is repeated 100 times using a jackknife procedure, wherein 5% of the measured PCSs are left out at random for each repetition. Reliable assignments are characterised by a higher percentage of occurrences in different parallel runs (>60%) and better than average score for the combined PCS fit. The polynomial time complexity of the Hungarian algorithm makes the approach computationally efficient. However, the Hungarian method typically gives only one optimal solution, and is not suitable for finding all similarly optimal solutions, which would reveal any ambiguity in the methyl assignment.

The previously proposed PCS-based Possum approach [121] required the knowledge of the $\Delta\chi$-tensor parameters, which could be obtained automatically if the backbone resonance assignments for the diamagnetic protein were known [132], or deduced from a minimum of five amide or methyl assignments [124]. PARAssign effectively removes this requirement.

4.6. Sparse labeling-based assignment – ASSIGN_SLP

For sparsely labeled systems, in which a single ^{15}N- or ^{13}C-enriched amino acid or a subset of such amino acids is supplied, for instance, directly to mammalian cells, structure-based assignment methods have been explored using measured proton and heteronuclear chemical shifts (^{13}C, ^{15}N) in combination with sparse NOEs and one-bond RDCs [159,160]. These data types can be predicted on the basis of the three-dimensional protein structure [136–138,161–163].

After outlining an assignment strategy that compared each of the data types to their predictions in a sequential manner, increasingly excluding assignment possibilities upon each comparison [160], Gao et al. introduced a genetic algorithm to search for an optimal set of assignments across all data types simultaneously [159]. On a benchmark of four proteins with known assignments, it could be shown that the most informative data type was NOE, with RDC and chemical shift data proving valuable for resolving assignment ambiguities and improving accuracy [159]. The approach uses genetic principles of mutation and cross-over, the rates of which are optimized during the assignment search to generate potential assignment solutions. These are evaluated with an objective function that is a combination of scores for each data type with adjustable weighting factors [159]. A software implementation for the method, entitled ASSIGNments for Sparsely Labeled Proteins (ASSIGN_SLP), is freely available. The program outputs multiple possible assignment solutions and advocates for the selection of highly reliable and consistent assignment results rather than relying solely on the top-ranked solutions, which may provide a complete assignment yet result in errors due to the inherent ambiguity of assignments derived from sparse data [159].

Although originally designed for assignment of sparsely labeled glycosylated proteins, the ASSIGN_SLP protocol has recently been extended for the complete assignment of the methyl-TROSY spectrum of the full-length bacterial Hsp90 homologue (HtpG), a 145 kDa homo-dimer [164]. Given that the protein was perdeuterated and exclusively alanine ^{13}C–^{1}H methyl-labeled, inter-methyl NOE data were very scarce. As a result, only chemical shifts and ^{13}C–^{1}H methyl RDC data could be used for the automatic resonance assignment, which was supplemented with partial methyl assignments obtained for the individual protein domains by standard triple resonance methods. The program was able to assign 60% of the observed methyl resonances with high confidence, allowing the interpretation of structural changes in HtpG upon the addition of an ATP analog [164].

Finally, it is worth mentioning an early resonance assignment strategy for sparsely labeled proteins that proposed correlating the amide proton-deuterium exchange rates measured by NMR and mass spectrometry (MS) [165], which has similarities with a recent combined NMR-MS method for methionine methyl resonance assignment, which focuses on comparison of methionine oxidation levels [166].

5. Performance on test benchmarks

The automated methyl assignment methods were tested by their authors on different benchmarks of synthetic and real NMR data. The author-reported performance on real NMR data is summarized in Table 3. The performance was evaluated against independently obtained reference methyl resonance assignments for the proteins in the benchmarks. The reference assignments were generally available either from the authors of the original methyl-TROSY studies of these proteins, who typically acquired the assignments through a combination of site-directed mutagenesis and divide-and-conquer or domain fragmentation strategies, or from the BMRB database. Occasionally, the authors of the automatic assignment methods resorted to divide-and-conquer approaches themselves to determine the independent reference methyl assignments for evaluation of their automatic method. The reference assignments typically covered a large fraction of all methyls groups of the proteins in the benchmark (>90%).

The NOE-based methods were generally tested on larger sets of proteins, which is likely due to the fact that NOE data can be collected from a single protein sample, whereas multiple samples are required for the collection of paramagnetic data. Relative to the paramagnetism-based methods, the NOE-based approaches have also been tested on larger proteins, although most of the methods focused on mid-size proteins (25–40 kDa). The largest protein in these benchmarks is the 14-mer $\alpha_7\alpha_7$ "half-proteasome" (360 kDa), which was assigned by MAP-XSII, MAGMA, and MethylFLYA.

Table 3
Summary of reported performance of automated methyl assignment protocols on benchmarks using experimental NMR data.

Method	Data sets	Mass (kDa)	Labeling	Labeled methyl groups	Assigned as reliable (%)[a]	Correctly assigned (%)[b]	Calculation time (min)
NOE-based							
MAP-XSII [108]	3	27–360	ILV, AILV	93–139	71–90	93–98	<1–20
FLAMEnGO2.0 [110]	1	41	ILV	121	98	84	N.R.
MAGMA [111]	9	9–360	LV, ILV, AILV	26–268	32–92[c]	100	3×10^{-4}–3×10^{3}[d]
MAGIC [112]	8	25–42	AILMTV, AILMV, AILV, ILMV[e]	76–209	32–97	>95	1–5700
MethylFLYA [113]	5	28–360	AILV, ILV	62–268	63–84	94–100	23–74
PRE/PCS-based							
PRE-ASSIGN [122]	1	27	AILMV	140	94[f]	100	<1
Possum [121]	1	30	AILMTV	125	95	94	<120
PARAssign [124]	1	25	ILV	76	60	100	N.R.

Ranges correspond to the lowest and highest value across the benchmarks. N.R. stands for not reported.
[a] Percentage relative to the number of labeled methyl groups.
[b] Percentage relative to the number of reliably assigned methyl groups with reference assignment available.
[c] Values recalculated to indicate the percentage of assigned methyl groups, rather than methyl residues, as reported in the original work [111].
[d] Time required for the calculation to converge.
[e] With dimethyl, monomethyl, or stereospecific labeling of Leu/Val methyls for different proteins [112].
[f] Using PRE-data and cross-validation by methyl-methyl NOE data.

Overall, for most of the methods the authors reported a high coverage of reliable (as classified by the respective algorithm) methyl assignments, with six out of eight algorithms reaching over 90% of the methyl groups with reference assignment for at least one protein in their respective benchmarks. With one exception, the realiably assigned methyl groups are reported to be more than 90% correct. Interestingly, all three methods that were applied to the 360 kDa $\alpha_7\alpha_7$ "half-proteasome" achieved their best coverage on this data set, yielding reliable assignments for 98, 92, and 84% of the methyl groups, respectively. This may be due to the fact that, with two symmetric heptameric rings, the $\alpha_7\alpha_7$ assembly gives rise to a single set of methyl resonances for its ~26 kDa α-subunit monomer, which greatly simplifies the assignment task. The next largest protein is the single polypeptide chain of malate synthase G (MSG), with a mass of 82 kDa and 268 ILV-labeled methyl groups that were automatically assigned by MAGMA and MethylFLYA. Reliable assignments were achieved for 37% (MAGMA) and 65% (MethylFLYA) of the methyl groups, as compared to the reference assignment of MSG, in both cases with complete accuracy [111,113]. A partial, yet reliable, assignment from purely automatic procedures is very valuable for methyl-NMR studies of large proteins. Furthermore, it can significantly aid manual efforts for the assignment of additional methyl groups, since both programs also report the ambiguous assignment possibilities for the remaining, unassigned methyls.

The three paramagnetism-based methods were each applied to a single protein (Table 3). Two of the algorithms report about 95% reliable assignments, probably reflecting the longer distance range of the paramagnetic effects in comparison with NOEs. For the AMILV methyl-labeled protein EIN, assignments could be determined with PRE-ASSIGN for 94% of the methyl resonances using PRE-data from five labeled sites and methyl NOEs in conjunction, while PRE and NOE data were reported to be in disagreement for 20 out of 140 methyl groups [122]. On the other hand, using exclusively methyl-methyl NOE data collected for AILV-methyl labeled EIN from the same study [122], MAGMA, MAP-XSII, and MethylFLYA were able to assign 42%, 48% and 61% of the methyl resonances, respectively. For the N-domain of HSP90, using PCS data from two paramagnetic centers, another paramagnetism-based algorithm, PARAssign [124], was able to reliably assign 60% of the ILV-methyl resonances, whereas MAGMA and MethylFLYA assigned 49% and 60% of the resonances, respectively, using inter-methyl NOEs [111,113].

From these examples it is apparent that, next to the quality of the NMR data, the shape of the protein can also be a determinant for assignment success. Even though EIN and the N-domain of HSP90 have similar molecular masses (27 and 22 kDa), EIN has a more elongated shape that leads to a significant fragmentation of the network of inter-methyl NOEs [111]. In this case, the long-range methyl PRE-restraints are particularly well-suited for resonance assignment. On the other hand, globular proteins with highly inter-connected methyl networks are best suited for the NOE-based methods. Combining inter-methyl NOE restraints with paramagnetic restraints provides valuable complementary data [108,110,122], which can also be used for assignment cross-validation.

A comparison of the performance of the different algorithms on the basis of the results in Table 3 is hampered by the fact that the programs were applied to different proteins and/or different input data. To remedy this situation, at least for the NOE-based methods,

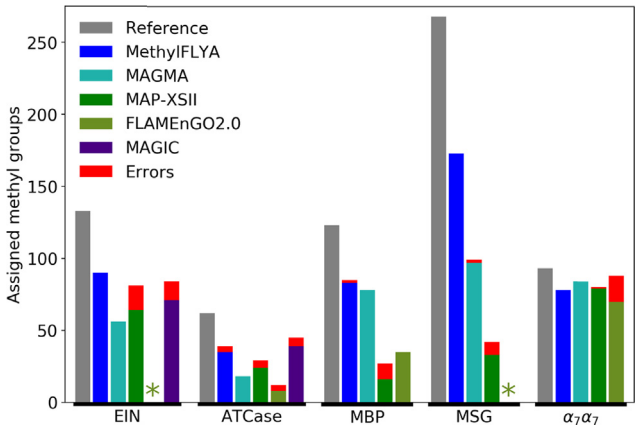

Fig. 7. Summary of the performance of the available NOE-based automatic methyl assignment protocols on the MAGMA benchmark [111]. Data from Table 1 of [113]. For each of the five proteins, the same input data were used for all programs, except for MAGIC that requires also NOE signal intensity information, which was available only for EIN and ATCase. Reference assignments for the labeled methyl groups (grey), which were independently determined and made available by the authors of the original methyl-TROSY studies of these proteins, are compared to the reliable assignments as determined by the algorithms. Erroneous assignments are shown in red. Asterisks indicate that no reliable assignments were found by FLAMEnGO2.0 for EIN and MSG [113].

a benchmark of methyl-NOESY data was compiled for five proteins of 28–360 kDa size [111], and the NOE-based automated assignment algorithms were tested with identical input data [113]. This enabled a direct comparison of their performance, which is summarized in Fig. 7.

It is evident that the NOE-based programs deliver in most cases a significant number of reliable methyl assignments. However, methyl groups always remain that could be assigned manually (using additional information, e.g. from specific labeling and/or site-directed mutagenesis) but not by the automated algorithms that rely exclusively on methyl-methyl NOEs. For instance, the MethylFLYA algorithm, which delivers overall the largest number of reliable assignments, was able to reliably assign between 56% and 84% of the available reference assignments. At present, the most accurate NOE-based assignment method is MAGMA, which owes its high accuracy to exhaustive sampling of the methyl assignment possibilities using exact graph comparison approaches [111], with 333 correct assignments and only two errors (Fig. 7). MethylFLYA is the next most accurate method [113], accurately assigning 459 methyl groups while making only 6 errors. Interestingly, all applicable algorithms performed similarly well for the largest target, the 360 kDa symmetric 14-mer $\alpha_7\alpha_7$ half proteasome, where they yielded reliable assignments for the vast majority (84–95%) of methyl groups with known reference assignments. For the other proteins, there is more variation among the different algorithms, both in terms of coverage and accuracy. The MAGIC algorithm [112] yields, for the two applicable proteins in the benchmark, a similar number of reliable assignments to MethylFLYA, albeit with a slightly higher error rate of 10%.

The mutual agreement among the sets of reliable assignments found by the different methods is relatively low (Fig. 8a). This indicates a complementarity of the results, which can be exploited to increase the completeness and accuracy of the methyl assignments. All methods agreed on only 11% of the reference assignments (Fig. 8a), which was in part due to the absence of reliable assignments from FLAMEnGO2.0 for two proteins (Fig. 7). MAGMA and MethylFLYA shared the largest intersection, which contained only correct methyl assignments (Fig. 8b). In fact, all the assignments in the intersections of any two methods were completely accurate (100%), except for the intersection of FLAMEnGO2.0 and MAP-XSII. Also applying MAGIC to the subset of proteins with available intensity information (Fig. 8c) gave complete agreement between all methods albeit for very few methyls (< 2%). MethylFLYA, MAGIC, MAGMA, and MAP-XSII agreed on about 18% of the

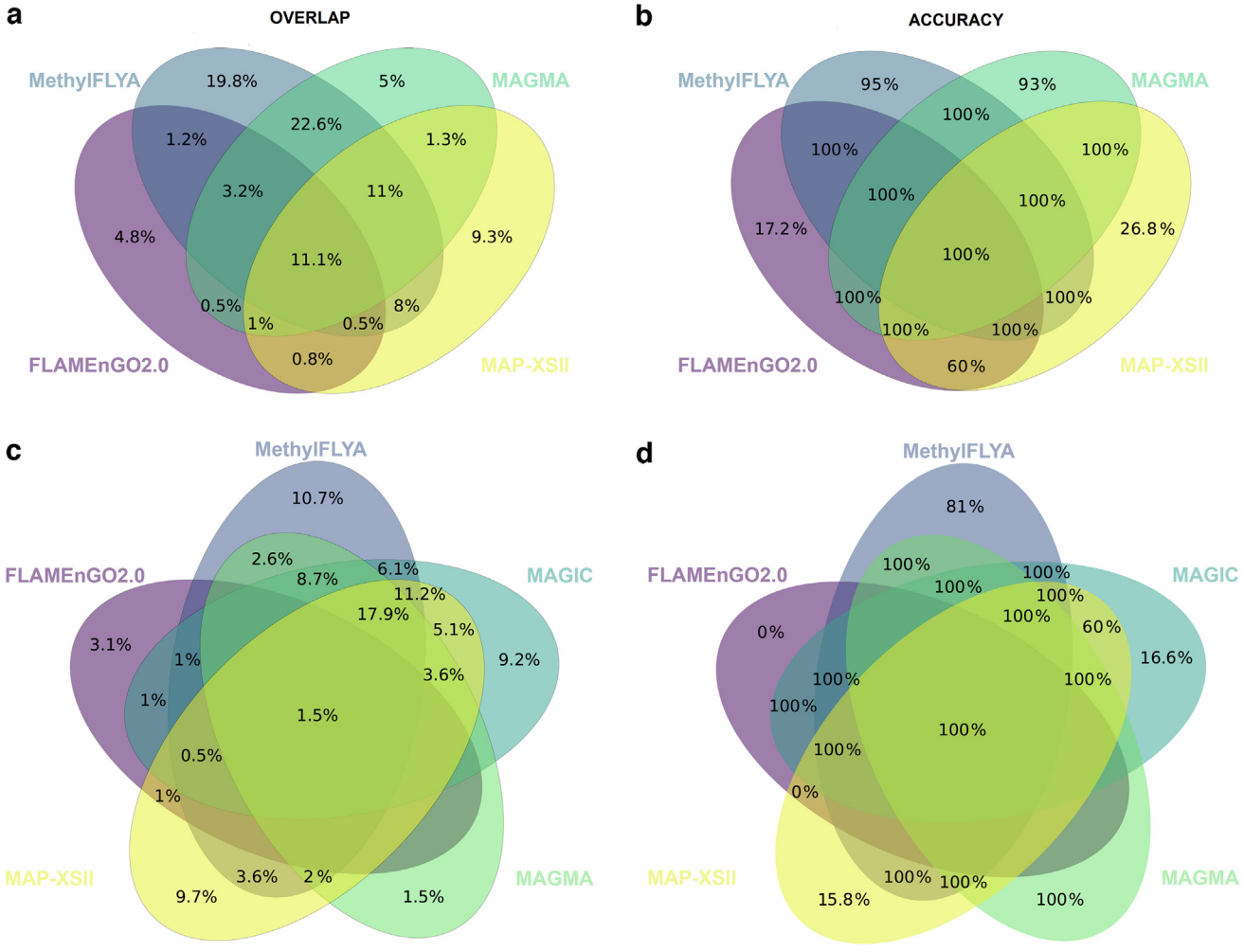

Fig. 8. Agreement among the reliable methyl assignments generated by the NOE-based methods on the MAGMA benchmark [111,113]. (a Upper left) Overlap among the assignments generated by MethylFLYA, MAGMA, FLAMEnGO2.0 and MAP-XSII compared over the five proteins in the MAGMA benchmark. The percentages of automatically generated assignments that agree between the different methods are given in the intersections. 100% corresponds to the number of known reference assignments available. Percentages outside of any intersection refer to the assignments generated by only one of the methods, and intersections without a percentage are empty. (b Upper right) Accuracy of the reliable methyl assignments generated by the different methods. (c) and (d) Same as above, but including application of MAGIC to the applicable subset (EIN and ATCase) of the MAGMA benchmark.

reference assignments. Again, the assignments in the intersections of any combination of two methods were completely accurate, apart from the small intersection of MAGIC and MAP-XSII (Fig. 8d).

The performance of MethylFLYA on the benchmark was further investigated with reduced experimental information [113]. In the best-case scenario, which was used with all programs for the comparison in Fig. 7, both knowledge of the amino acid types of methyl resonances and linkage of the two geminal methyl groups of Leu and Val were provided in the input. Without discrimination between Leu and Val resonances, MethylFLYA performed very similarly to the best-case scenario. Leu/Val residue discrimination is therefore not crucial, at least for MethylFLYA. In contrast, removing the geminal linkage for the Leu and Val methyl groups had a significant negative impact, reducing the percentage of reliably assigned methyls by about 20% for EIN, ATCase, MBP, and MSG, and up to 30% for $\alpha_7\alpha_7$ [113]. Nevertheless, the overall accuracy of the reliable assignments remained high. The critical importance of the geminal linkage information for automatic methyl assignment was reported previously for MAGMA [111]. In the MAGIC study, a four-fold decrease in computation time and a somewhat improved assignment accuracy were noted as benefits of the geminal linkage information [112]. Alternatively, the Leu/Val geminal linkage information can be substituted with stereospecific labeling schemes that restrict isotopic labeling to only pro-*R* or pro-*S* methyl groups, and thus reduce the number of methyl resonances to be assigned [43].

6. Conclusions

Methyl-TROSY is a powerful method for elucidating the mechanisms of action of macromolecular machines, unveiling their dynamical interactions and ligand binding sites. Wider applications of the method are hampered by the laborious, time-consuming, and expensive nature of methyl resonance assignment. Throughout the last decade, various structure-based automatic methyl resonance assignment approaches have been developed, which primarily rely on measured methyl chemical shifts, inter-methyl NOEs, paramagnetic restraints, or combinations thereof. The measured NMR data can be related to that predicted from the existing three-dimensional structure of a protein, or protein complex, using a variety of automated approaches, which range from Monte-Carlo and evolutionary algorithms to exact graph comparisons.

The sparsity of the NMR data for large proteins underlies assignment ambiguity. To reveal ambiguous assignment options, some protocols rely on exhaustive sampling (MAGMA, MAGIC), while others employ multiple parallel repetitions of the assignment protocol (MethylFLYA, FLAMEnGO2.0, MAP-XSII). Paramagnetism-based methods, which exploit the longer distance range of paramagnetic effects, are an alternative for proteins and complexes of elongated shapes, and can also be supplemented with NOE restraints to extend and cross-validate the generated methyl assignments.

All approaches benefit, to various degrees, from knowledge of the amino acid types of methyl resonances, as this significantly reduces the search space. Knowledge of Leu/Val geminal methyl pairing strongly assists the NOE-based approaches by limiting assignment ambiguity. Some advances can be expected from stereospecific labeling of Leu/Val methyl groups, which reduces the size of the assignment problem, but also the number of probes and NMR restraints. All approaches can profit from known assignments for some of the resonances, which decrease the number of remaining assignment possibilities and can reduce calculation times. The automatic methods also face common challenges, such as overlap of methyl peaks in the 2D [^{1}H,^{13}C]-HMQC spectra, which can limit both unambiguous NOE assignment and accurate pairing of peaks between the diamagnetic and paramagnetic spectra for the PCS-based approaches.

In the future, the NOE-based methods might benefit from considering NOESY cross-peak intensities, for which treatments introduced by ASSIGN_SLP and MAGIC could provide useful guidelines. The complementarity between different approaches can be exploited for better coverage and reliability of assignments. Combining paramagnetism- and NOE-based approaches is expected to be particularly powerful, as it would provide both assignment cross-validation, and complementary extension.

The existing automatic methyl resonance assignment methods demonstrate a diversity of strategies devised to aid the assignment problem. The inherent ambiguity of methyl resonance assignment is due to the sparsity of NMR data, which remains a limitation for automatic assignment approaches that should be addressed in the future by a synergy of experiments and computation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Sattler, J. Schleucher, C. Griesinger, Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, Prog. Nucl. Magn. Reson. Spectrosc. 34 (1999) 93–158.
[2] V. Tugarinov, L.E. Kay, 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 (2003) 13868–13878.
[3] D. Sheppard, C.Y. Guo, V. Tugarinov, 4D ^{1}H–^{13}C NMR spectroscopy for assignments of alanine methyls in large and complex protein structures, J. Am. Chem. Soc. 131 (2009) 1364–1365.
[4] S. Grzesiek, A. Bax, Amino acid type determination in the sequential assignment procedure of uniformly ^{13}C/^{15}N-enriched proteins, J. Biomol. NMR 3 (1993) 185–204.
[5] D.P. Frueh, A.C. Goodrich, S.H. Mishra, S.R. Nichols, NMR methods for structural studies of large monomeric and multimeric proteins, Curr. Opin. Struct. Biol. 23 (2013) 734–739.
[6] R. Konrat, NMR contributions to structural dynamics studies of intrinsically disordered proteins, J. Magn. Reson. 241 (2014) 74–85.
[7] M.C. Baran, Y.J. Huang, H.N.B. Moseley, G.T. Montelione, Automated analysis of protein NMR assignments and structures, Chem. Rev. 104 (2004) 3541–3555.
[8] E. Schmidt, P. Güntert, A new algorithm for reliable and general NMR resonance assignment, J. Am. Chem. Soc. 134 (2012) 12817–12829.
[9] V. Tugarinov, P.M. Hwang, J.E. Ollerenshaw, L.E. Kay, Cross-correlated relaxation enhanced ^{1}H–^{13}C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes, J. Am. Chem. Soc. 125 (2003) 10420–10428.
[10] R. Sprangers, A. Velyvis, L.E. Kay, Solution NMR of supramolecular complexes: providing new insights into function, Nat. Methods 4 (2007) 697–703.
[11] R. Kerfah, M.J. Plevin, R. Sounier, P. Gans, J. Boisbouvier, Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins, Curr. Opin. Struct. Biol. 32 (2015) 113–122.
[12] S. Wiesner, R. Sprangers, Methyl groups as NMR probes for biomolecular interactions, Curr. Opin. Struct. Biol. 35 (2015) 60–67.
[13] Y.J. Jiang, C.G. Kalodimos, NMR studies of large proteins, J. Mol. Biol. 429 (2017) 2667–2676.
[14] C.D. Huang, C.G. Kalodimos, Structures of large protein complexes determined by nuclear magnetic resonance spectroscopy, Ann. Rev. Biophys. 46 (2017) 317–336.
[15] Z.K. Boswell, M.P. Latham, Methyl-based NMR spectroscopy methods for uncovering structural dynamics in large proteins and protein complexes, Biochemistry 58 (2019) 144–155.
[16] S. Alphonse, R. Ghose, Methyl NMR spectroscopy: Measurement of dynamics in viral RNA-directed RNA polymerases, Methods 148 (2018) 100–114.
[17] S.D. Gorman, D. Sahu, K.F. O'Rourke, D.D. Boehr, Assigning methyl resonances for protein solution-state NMR studies, Methods 148 (2018) 88–99.
[18] R. Sprangers, L.E. Kay, Quantitative dynamics and binding studies of the 20S proteasome by NMR, Nature 445 (2007) 618–622.
[19] I. Gelis, A.M.J.J. Bonvin, D. Keramisanou, M. Koukaki, G. Gouridis, S. Karamanou, A. Economou, C.G. Kalodimos, Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR, Cell 131 (2007) 756–769.

[20] C. Amero, M.A. Durá, M. Noirclerc-Savoye, A. Perollier, B. Gallet, M.J. Plevin, T. Vernet, B. Franzetti, J. Boisbouvier, A systematic mutagenesis-driven strategy for site-resolved NMR studies of supramolecular assemblies, J. Biomol. NMR 50 (2011) 229–236.

[21] J. Janin, S. Miller, C. Chothia, Surface, subunit interfaces and interior of oligomeric proteins, J. Mol. Biol. 204 (1988) 155–164.

[22] A. Mittermaier, L.E. Kay, J.D. Forman-Kay, Analysis of deuterium relaxation-derived methyl axis order parameters and correlation with local structure, J. Biomol. NMR 13 (1999) 181–185.

[23] R. Suzuki, M. Sakakura, M. Mori, M. Fujii, S. Akashi, H. Takahashi, Methyl-selective isotope labeling using α-ketoisovalerate for the yeast *Pichia pastoris* recombinant protein expression system, J. Biomol. NMR 71 (2018) 213–223.

[24] L. Clark, J.A. Zahm, R. Ali, M. Kukula, L.Q. Bian, S.M. Patrie, K.H. Gardner, M.K. Rosen, D.M. Rosenbaum, Methyl labeling and TROSY NMR spectroscopy of proteins expressed in the eukaryote *Pichia pastoris*, J. Biomol. NMR 62 (2015) 239–245.

[25] M. Miyazawa-Onami, K. Takeuchi, T. Takano, T. Sugiki, I. Shimada, H. Takahashi, Perdeuteration and methyl-selective ^{1}H, ^{13}C-labeling by using a *Kluyveromyces lactis* expression system, J. Biomol. NMR 57 (2013) 297–304.

[26] L. Clark, I. Dikiy, D.M. Rosenbaum, K.H. Gardner, On the use of *Pichia pastoris* for isotopic labeling of human GPCRs for NMR studies, J. Biomol. NMR 71 (2018) 203–211.

[27] A.S. Solt, M.J. Bostock, B. Shrestha, P. Kumar, T. Warne, C.G. Tate, D. Nietlispach, Insight into partial agonism by observing multiple equilibria for ligand-bound and G_s-mimetic nanobody-bound β_1-adrenergic receptor, Nat. Commun. 8 (2017) 12.

[28] Y. Kofuku, T. Ueda, J. Okude, Y. Shiraishi, K. Kondo, T. Mizumura, S. Suzuki, I. Shimada, Functional dynamics of deuterated β_2-adrenergic receptor in lipid bilayers revealed by NMR spectroscopy, Angew. Chem. Int. Ed. 53 (2014) 13376–13379.

[29] J. Okude, T. Ueda, Y. Kofuku, M. Sato, N. Nobuyama, K. Kondo, Y. Shiraishi, T. Mizumura, K. Onishi, M. Natsume, M. Maeda, H. Tsujishita, T. Kuranaga, M. Inoue, I. Shimada, Identification of a conformational equilibrium that determines the efficacy and functional selectivity of the μ-opioid receptor, Angew. Chem. Int. Ed. 54 (2015) 15771–15776.

[30] B. Franke, C. Opitz, S. Isogai, A. Grahl, L. Delgado, A.D. Gossert, S. Grzesiek, Production of isotope-labeled proteins in insect cells for NMR, J. Biomol. NMR 71 (2018) 173–184.

[31] Y. Kofuku, T. Yokomizo, S. Imai, Y. Shiraishi, M. Natsume, H. Itoh, M. Inoue, K. Nakata, S. Igarashi, H. Yamaguchi, T. Mizukoshi, E. Suzuki, T. Ueda, I. Shimada, Deuteration and selective labeling of alanine methyl groups of β_2-adrenergic receptor expressed in a baculovirus-insect cell expression system, J. Biomol. NMR 71 (2018) 185–192.

[32] W.J. Metzler, M. Wittekind, V. Goldfarb, L. Mueller, B.T. Farmer II, Incorporation of ^{1}H/^{13}C/^{15}N-{Ile, Leu, Val} into a perdeuterated, ^{15}N-labeled protein: Potential in structure determination of large proteins by NMR, J. Am. Chem. Soc. 118 (1996) 6800–6801.

[33] M.K. Rosen, K.H. Gardner, R.C. Willis, W.E. Parris, T. Pawson, L.E. Kay, Selective methyl group protonation of perdeuterated proteins, J. Mol. Biol. 263 (1996) 627–636.

[34] V. Tugarinov, V. Kanelis, L.E. Kay, Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy, Nat. Protoc. 1 (2006) 749–754.

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

[36] M.L. Cai, Y. Huang, R.B. Yang, R. Craigie, G.M. Clore, A simple and robust protocol for high-yield expression of perdeuterated proteins in *Escherichia coli* grown in shaker flasks, J. Biomol. NMR 66 (2016) 85–91.

[37] E.S. O'Brien, D.W. Lin, B. Fuglestad, M.A. Stetz, T. Gosse, C. Tommos, A.J. Wand, Improving yields of deuterated, methyl labeled protein by growing in H2O, J. Biomol. NMR 71 (2018) 263–273.

[38] K.H. Gardner, L.E. Kay, Production and incorporation of ^{15}N, ^{13}C, ^{2}H (^{1}H–d1 Methyl) isoleucine into proteins for multidimensional NMR studies, J. Am. Chem. Soc. 119 (1997) 7599–7600.

[39] A.M. Ruschak, A. Velyvis, L.E. Kay, A simple strategy for ^{13}C,^{1}H labeling at the Ile-γ2 methyl position in highly deuterated proteins, J. Biomol. NMR 48 (2010) 129–135.

[40] V. Tugarinov, L.E. Kay, An isotope labeling strategy for methyl TROSY spectroscopy, J. Biomol. NMR 28 (2004) 165–172.

[41] R.J. Lichtenecker, N. Coudevylle, R. Konrat, W. Schmid, Selective isotope labelling of leucine residues by using α-ketoacid precursor compounds, ChemBioChem 14 (2013) 818–821.

[42] R.J. Lichtenecker, K. Weinhäupl, L. Reuther, J. Schörghuber, W. Schmid, R. Konrat, Independent valine and leucine isotope labeling in *Escherichia coli* protein overexpression systems, J. Biomol. NMR 57 (2013) 205–209.

[43] P. Gans, O. Hamelin, R. Sounier, I. Ayala, M.A. Durá, C.D. Amero, M. Noirclerc-Savoye, B. Franzetti, M.J. Plevin, J. Boisbouvier, Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins, Angew. Chem. Int. Ed. 49 (2010) 1958–1962.

[44] L.C. Shi, L.E. Kay, Tracing an allosteric pathway regulating the activity of the HslV protease, Proc. Natl. Acad. Sci. USA 111 (2014) 2140–2145.

[45] M. Kainosho, P. Güntert, SAIL – Stereo-array isotope labeling, Q. Rev. Biophys. 42 (2009) 247–300.

[46] M. Kainosho, T. Torizawa, Y. Iwashita, T. Terauchi, A.M. Ono, P. Güntert, Optimal isotope labelling for NMR protein structure determinations, Nature 440 (2006) 52–57.

[47] Y.R. Monneau, Y. Ishida, P. Rossi, T. Saio, S.R. Tzeng, M. Inouye, C.G. Kalodimos, Exploiting *E. coli* auxotrophs for leucine, valine, and threonine specific methyl labeling of large proteins for NMR applications, J. Biomol. NMR 65 (2016) 99–108.

[48] Y. Miyanoiri, Y. Ishida, M. Takeda, T. Terauchi, M. Inouye, M. Kainosho, Highly efficient residue-selective labeling with isotope-labeled Ile, Leu, and Val using a new auxotrophic *E. coli* strain, J. Biomol. NMR 65 (2016) 109–119.

[49] Z. Serber, W. Straub, L. Corsini, A.M. Nomura, N. Shimba, C.S. Craik, P.O. de Montellano, V. Dötsch, Methyl groups as probes for proteins and complexes in in-cell NMR experiments, J. Am. Chem. Soc. 126 (2004) 7119–7125.

[50] M. Fischer, K. Kloiber, J. Häusler, K. Ledolter, R. Konrat, W. Schmid, Synthesis of a ^{13}C-methyl-group-labeled methionine precursor as a useful tool for simplifying protein structural analysis by NMR spectroscopy, ChemBioChem 8 (2007) 610–612.

[51] A. Velyvis, A.M. Ruschak, L.E. Kay, An economical method for production of ^{2}H, $^{13}CH_3$-threonine for solution NMR studies of large protein complexes: application to the 670 kDa proteasome, PLoS One 7 (2012) 8.

[52] K. Sinha, L. Jen-Jacobson, G.S. Rule, Specific labeling of threonine methyl groups for NMR studies of protein-nucleic acid complexes, Biochemistry 50 (2011) 10189–10191.

[53] I. Ayala, R. Sounier, N. Use, P. Gans, J. Boisbouvier, An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein, J. Biomol. NMR 43 (2009) 111–119.

[54] R.L. Isaacson, P.J. Simpson, M. Liu, E. Cota, X. Zhang, P. Freemont, S. Matthews, A new labeling method for methyl transverse relaxation-optimized spectroscopy NMR spectra of alanine residues, J. Am. Chem. Soc. 129 (2007) 15428–15429.

[55] T.L. Religa, L.E. Kay, Optimal methyl labeling for studies of supra-molecular systems, J. Biomol. NMR 47 (2010) 163–169.

[56] A.J. Baldwin, T.L. Religa, D.F. Hansen, G. Bouvignies, L.E. Kay, $^{13}CHD_2$ methyl group probes of millisecond time scale exchange in proteins by ^{1}H relaxation dispersion: an application to proteasome gating residue dynamics, J. Am. Chem. Soc. 132 (2010) 10992–10995.

[57] O. Millet, D.R. Muhandiram, N.R. Skrynnikov, L.E. Kay, Deuterium spin probes of side-chain dynamics in proteins. 1. Measurement of five relaxation rates per deuteron in ^{13}C-labeled and fractionally ^{2}H-enriched proteins in solution, J. Am. Chem. Soc. 124 (2002) 6439–6448.

[58] N.R. Skrynnikov, O. Millet, L.E. Kay, Deuterium spin probes of side-chain dynamics in proteins. 2. Spectral density mapping and identification of nanosecond time-scale side-chain motions, J. Am. Chem. Soc. 124 (2002) 6449–6460.

[59] T.I. Igumenova, K.K. Frederick, A.J. Wand, Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution, Chem. Rev. 106 (2006) 1672–1699.

[60] X.L. Liao, V. Tugarinov, Selective detection of $^{13}CHD_2$ signals from a mixture of $^{13}CH_3$/$^{13}CH_2D$/$^{13}CHD_2$ methyl isotopomers in proteins, J. Magn. Reson. 209 (2011) 101–107.

[61] X.L. Liao, D. Long, D.W. Li, R. Brüschweiler, V. Tugarinov, Probing side-chain dynamics in proteins by the measurement of nine deuterium relaxation rates per methyl group, J. Phys. Chem. B 116 (2012) 606–620.

[62] A. Hsu, P.A. O'Brien, S. Bhattacharya, M. Rance, A.G. Palmer, Enhanced spectral density mapping through combined multiple-field deuterium $^{13}CH_2D$ methyl spin relaxation NMR spectroscopy, Methods 138 (2018) 76–84.

[63] R. Ishima, A.P. Petkova, J.M. Louis, D.A. Torchia, Comparison of methyl rotation axis order parameters derived from model-free analyses of ^{2}H and ^{13}C longitudinal and transverse relaxation rates measured in the same protein sample, J. Am. Chem. Soc. 123 (2001) 6164–6171.

[64] U. Brath, M. Akke, D.W. Yang, L.E. Kay, F.A.A. Mulder, Functional dynamics of human FKBP12 revealed by methyl ^{13}C rotating frame relaxation dispersion NMR spectroscopy, J. Am. Chem. Soc. 128 (2006) 5718–5727.

[65] R. Otten, J. Villali, D. Kern, F.A.A. Mulder, Probing microsecond time scale dynamics in proteins by methyl ^{1}H Carr-Purcell-Meiboom-Gill relaxation dispersion NMR measurements. Application to activation of the signaling protein NtrC^{r}, J. Am. Chem. Soc. 132 (2010) 17004–17014.

[66] U. Weininger, A.T. Blissing, J. Hennig, A. Ahlner, Z.H. Liu, H.J. Vogel, M. Akke, P. Lundström, Protein conformational exchange measured by ^{1}H $R_{1\rho}$ relaxation dispersion of methyl groups, J. Biomol. NMR 57 (2013) 47–55.

[67] U. Weininger, Z.H. Liu, D.D. McIntyre, H.J. Vogel, M. Akke, Specific $^{12}C^{\beta}D_2$ $^{12}C^{\gamma}D_2$ $^{S1}3^{C}\varepsilon H_D2$ isotopomer labeling of methionine to characterize protein dynamics by 1H and ^{13}C NMR relaxation dispersion, J. Am. Chem. Soc. 134 (2012) 18562–18565.

[68] E. Rennella, A.K. Schuetz, L.E. Kay, Quantitative measurement of exchange dynamics in proteins via ^{13}C relaxation dispersion of $^{13}CHD_2$-labeled samples, J. Biomol. NMR 65 (2016) 59–64.

[69] A.B. Gopalan, T. Yuwen, L.E. Kay, P. Vallurupalli, A methyl ^{1}H double quantum CPMG experiment to study protein conformational exchange, J. Biomol. NMR 72 (2018) 79–91.

[70] T. Yuwen, P. Vallurupalli, L.E. Kay, Enhancing the sensitivity of CPMG relaxation dispersion to conformational exchange processes by multiple-quantum spectroscopy, Angew. Chem. Int. Ed. 55 (2016) 11490–11494.

[71] A.B. Gopalan, P. Vallurupalli, Measuring the signs of the methyl ^{1}H chemical shift differences between major and 'invisible' minor protein conformational

states using methyl ^{1}H multi-quantum spectroscopy, J. Biomol. NMR 70 (2018) 187–202.
[72] S. Schütz, R. Sprangers, Methyl TROSY spectroscopy: A versatile NMR approach to study challenging biological systems, Prog. Nucl. Magn. Reson. Spectrosc. 116 (2020) 56–84.
[73] A.W. Barb, J.H. Prestegard, NMR analysis demonstrates immunoglobulin G N-glycans are accessible and dynamic, Nat. Chem. Biol. 7 (2011) 147–153.
[74] J.H. Prestegard, D.A. Agard, K.W. Moremen, L.A. Lavery, L.C. Morris, K. Pederson, Sparse labeling of proteins: Structural characterization from long range constraints, J. Magn. Reson. 241 (2014) 32–40.
[75] S. Yanaka, H. Yagi, R. Yogo, M. Yagi-Utsumi, K. Kato, Stable isotope labeling approaches for NMR characterization of glycoproteins using eukaryotic expression systems, J. Biomol. NMR 71 (2018) 193–202.
[76] Y. Miyanoiri, M. Takeda, T. Terauchi, M. Kainosho, Recent developments in isotope-aided NMR methods for supramolecular protein complexes –SAIL aromatic TROSY, BBA-General Subjects 1864 (2020) 129439.
[77] T.L. Religa, R. Sprangers, L.E. Kay, Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR, Science 328 (2010) 98–102.
[78] V. Tugarinov, L.E. Kay, Quantitative ^{13}C and ^{2}H NMR relaxation studies of the 723-residue enzyme malate synthase G reveal a dynamic binding interface, Biochemistry 44 (2005) 15970–15977.
[79] V. Tugarinov, L.E. Kay, Methyl groups as probes of structure and dynamics in NMR studies of high-molecular-weight proteins, ChemBioChem 6 (2005) 1567–1577.
[80] G.E. Karagöz, D. Acosta-Alvear, H.T. Nguyen, C.P. Lee, F.X. Chu, P. Walter, An unfolded protein-induced conformational switch activates mammalian IRE1, eLife 6 (2017) e30700.
[81] P. Macek, R. Kerfah, E.B. Erba, E. Crublet, C. Moriscot, G. Schoehn, C. Amero, J. Boisbouvier, Unraveling self-assembly pathways of the 468-kDa proteolytic machine TET2, Sci. Adv. 3 (2017) 9.
[82] N.L. Fawzi, D.S. Libich, J.F. Ying, V. Tugarinov, G.M. Clore, Characterizing methyl-bearing side chain contacts and dynamics mediating amyloid β protofibril interactions using $^{13}C_{methyl}$-DEST and lifetime line broadening, Angew. Chem. Int. Ed. 53 (2014) 10345–10349.
[83] L.D. Cabrita, S.T.D. Hsu, H. Launay, C.M. Dobson, J. Christodoulou, Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy, Proc. Natl. Acad. Sci. USA 106 (2009) 22239–22244.
[84] C.A. Waudby, H. Launay, L.D. Cabrita, J. Christodoulou, Protein folding on the ribosome studied using NMR spectroscopy, Prog. Nucl. Magn. Reson. Spectrosc. 74 (2013) 57–75.
[85] A. Deckert, C.A. Waudby, T. Wlodarski, A.S. Wentink, X.L. Wang, J.P. Kirkpatrick, J.F.S. Paton, C. Camilloni, P. Kukic, C.M. Dobson, M. Vendruscolo, L.D. Cabrita, J. Christodoulou, Structural characterization of the interaction of α-synuclein nascent chains with the ribosomal surface and trigger factor, Proc. Natl. Acad. Sci. USA 113 (2016) 5012–5017.
[86] A.M.E. Cassaignau, H.M.M. Launay, M.E. Karyadi, X.L. Wang, C.A. Waudby, A. Deckert, A.L. Robertson, J. Christodoulou, L.D. Cabrita, A strategy for co-translational folding studies of ribosome-bound nascent chain complexes using NMR spectroscopy, Nat. Protoc. 11 (2016) 1492–1507.
[87] E.F. DeRose, T.W. Kirby, G.A. Mueller, W.A. Beard, S.H. Wilson, R.E. London, Transitions in DNA polymerase β μs-ms dynamics related to substrate binding and catalysis, Nucleic Acids Res. 46 (2018) 7309–7322.
[88] D. Oyen, R.B. Fenwick, P.C. Aoto, R.L. Stanfield, I.A. Wilson, H.J. Dyson, P.E. Wright, Defining the structural basis for allosteric product release from *E. coli* dihydrofolate reductase using NMR relaxation dispersion, J. Am. Chem. Soc. 139 (2017) 11233–11240.
[89] Y. Peng, A.L. Hansen, L. Brüschweiler-Li, O. Davulcu, J.J. Skalicky, M.S. Chapman, R. Brüschweiler, The Michaelis complex of arginine kinase samples the transition state at a frequency that matches the catalytic rate, J. Am. Chem. Soc. 139 (2017) 4846–4853.
[90] A.M. Whitaker, M.T. Naik, R.E. Mosser, G.D. Reinhart, Propagation of the allosteric signal in phosphofructokinase from Bacillus stearothermophilus examined by methyl-transverse relaxation-optimized spectroscopy nuclear magnetic resonance, Biochemistry 58 (2019) 5294–5304.
[91] W.L. Meng, E.M. Clerico, N. McArthur, L.M. Gierasch, Allosteric landscapes of eukaryotic cytoplasmic Hsp70s are shaped by evolutionary tuning of key interfaces, Proc. Natl. Acad. Sci. USA 115 (2018) 11970–11975.
[92] R. Rosenzweig, A. Sekhar, J. Nagesh, L.E. Kay, Promiscuous binding by Hsp70 results in conformational heterogeneity and fuzzy chaperone-substrate ensembles, eLife 6 (2017) 22.
[93] A. Sekhar, A. Velyvis, G. Zoltsman, R. Rosenzweig, G. Bouvignies, L.E. Kay, Conserved conformational selection mechanism of Hsp70 chaperone-substrate interactions, eLife 7 (2018) 29.
[94] K. Weinhäupl, C. Lindau, A. Hessel, Y. Wang, C. Schütze, T. Jores, L. Melchionda, B. Schönfisch, H. Kalbacher, B. Bersch, D. Rapaport, M. Brennich, K. Lindorff-Larsen, N. Wiedemann, P. Schanda, Structural basis of membrane protein chaperoning through the mitochondrial intermembrane space, Cell 175 (2018).
[95] M. Larion, A.L. Hansen, F.L. Zhang, L. Brüschweiler-Li, V. Tugarinov, B.G. Miller, R. Brüschweiler, Kinetic cooperativity in human pancreatic glucokinase originates from millisecond dynamics of the small domain, Angew. Chem. Int. Ed. 54 (2015) 8129–8132.
[96] G.P. Lisi, J.P. Loria, Solution NMR spectroscopy for the study of enzyme allostery, Chem. Rev. 116 (2016) 6323–6369.
[97] T. Saleh, P. Rossi, C.G. Kalodimos, Atomic view of the energy landscape in the allosteric regulation of Abl kinase, Nat. Struct. Mol. Biol. 24 (2017) 893–901.
[98] S. Vahidi, Z.A. Ripstein, M. Bonomi, T. Yuwen, M.F. Mabanglo, J.B. Juravsky, K. Rizzolo, A. Velyvis, W.A. Houry, M. Vendruscolo, J.L. Rubinstein, L.E. Kay, Reversible inhibition of the ClpP protease via an N-terminal conformational switch, Proc. Natl. Acad. Sci. USA 115 (2018) E6447–E6456.
[99] T.R. Alderson, L.E. Kay, Unveiling invisible protein states with NMR spectroscopy, Curr. Opin. Struct. Biol. 60 (2019) 39–49.
[100] F.A. Chao, Y.F. Li, Y. Zhang, R.A. Byrd, Probing the broad time scale and heterogeneous conformational dynamics in the catalytic core of the Arf-GAP ASAP1 via methyl adiabatic relaxation dispersion, J. Am. Chem. Soc. 141 (2019) 11881–11891.
[101] T.R. Yuwen, R. Huang, P. Vallurupalli, L.E. Kay, A methyl-TROSY-based ^{1}H relaxation dispersion experiment for studies of conformational exchange in high molecular weight proteins, Angew. Chem. Int. Ed. 58 (2019) 6250–6254.
[102] Y. Toyama, M. Osawa, M. Yokogawa, I. Shimada, NMR method for characterizing microsecond-to-millisecond chemical exchanges utilizing differential multiple-quantum relaxation in high molecular weight proteins, J. Am. Chem. Soc. 138 (2016) 2302–2311.
[103] R. Rosenzweig, L.E. Kay, Bringing dynamic molecular machines into focus by methyl-TROSY NMR, Annu. Rev. Biochem. 83 (2014) 291–315.
[104] A.M. Ruschak, L.E. Kay, Methyl groups as probes of supra-molecular structure, dynamics and function, J. Biomol. NMR 46 (2010) 75–87.
[105] A. Velyvis, H.K. Schachman, L.E. Kay, Assignment of Ile, Leu, and Val methyl correlations in supra-molecular systems: An application to aspartate transcarbamoylase, J. Am. Chem. Soc. 131 (2009) 16534–16543.
[106] A.K. Schuetz, L.E. Kay, A dynamic molecular basis for malfunction in disease mutants of p97/VCP, eLife 5 (2016) 25.
[107] Y.Q. Xu, M.H. Liu, P.J. Simpson, R. Isaacson, E. Cota, J. Marchant, D.W. Yang, X. D. Zhang, P. Freemont, S. Matthews, Automated assignment in selectively methyl-labeled proteins, J. Am. Chem. Soc. 131 (2009) 9480–9481.
[108] Y.Q. Xu, S. Matthews, MAP-XSII: an improved program for the automatic assignment of methyl resonances in large proteins, J. Biomol. NMR 55 (2013) 179–187.
[109] F.-A. Chao, L. Shi, L.R. Masterson, G. Veglia, FLAMEnGO: A fuzzy logic approach for methyl group assignment using NOESY and paramagnetic relaxation enhancement data, J. Magn. Reson. 214 (2012) 103–110.
[110] F.A. Chao, J.G. Kim, Y.L. Xia, M. Milligan, N. Rowe, G. Veglia, FLAMEnGO 2.0: An enhanced fuzzy logic algorithm for structure-based assignment of methyl group resonances, J. Magn. Reson. 245 (2014) 17–23.
[111] I. Pritišanac, M.T. Degiacomi, T.R. Alderson, M.G. Carneiro, A.B. Eiso, G. Siegal, A.J. Baldwin, Automatic assignment of methyl-NMR spectra of supramolecular machines using graph theory, J. Am. Chem. Soc. 139 (2017) 9523–9533.
[112] Y.R. Monneau, P. Rossi, A. Bhaumik, C.D. Huang, Y.J. Jiang, T. Saleh, T. Xie, Q. Xing, C.G. Kalodimos, Automatic methyl assignment in large proteins by the MAGIC algorithm, J. Biomol. NMR 69 (2017) 215–227.
[113] I. Pritišanac, J.M. Würz, T.R. Alderson, P. Güntert, Automatic structure-based NMR methyl resonance assignment in large proteins, Nat. Commun. 10 (2019) 12.
[114] Y. Xiao, L.R. Warner, M.P. Latham, N.G. Ahn, A. Pardi, Structure-based assignment of Ile, Leu, and Val methyl groups in the active and inactive forms of the mitogen-activated protein kinase extracellular signal-regulated kinase 2, Biochemistry 54 (2015) 4307–4319.
[115] R. Sounier, L. Blanchard, Z.R. Wu, J. Boisbouvier, High-accuracy distance measurement between remote methyls in specifically protonated proteins, J. Am. Chem. Soc. 129 (2007) 472–473.
[116] J. Cavanagh, W.J. Fairbrother, A.G. Palmer III, N.J. Skelton, M. Rance, Protein NMR Spectroscopy, Principles and Practice, second ed., Academic Press, San Diego, CA, 2007.
[117] C. Zwahlen, K.H. Gardner, S.P. Sarma, D.A. Horita, R.A. Byrd, L.E. Kay, An NMR experiment for measuring methyl-methyl NOEs in ^{13}C-labeled proteins with high resolution, J. Am. Chem. Soc. 120 (1998) 7617–7625.
[118] G.W. Vuister, G.M. Clore, A.M. Gronenborn, R. Powers, D.S. Garrett, R. Tschudin, A. Bax, Increased resolution and improved spectral quality in 4-dimensional ^{13}C/^{13}C-separated HMQC-NOESY-HMQC spectra using pulsed-field gradients, J. Magn. Reson. B 101 (1993) 210–213.
[119] V. Tugarinov, L.E. Kay, I. Ibraghimov, V.Y. Orekhov, High-resolution four-dimensional ^{1}H–^{13}C NOE spectroscopy using methyl-TROSY, sparse data acquisition, and multidimensional decomposition, J. Am. Chem. Soc. 127 (2005) 2767–2775.
[120] R. Kerfah, M.J. Plevin, O. Pessey, O. Hamelin, P. Gans, J. Boisbouvier, Scrambling free combinatorial labeling of lanine-β, isoleucine-δ1, leucine-proS and valine-proS methyl groups for the detection of long range NOEs, J. Biomol. NMR 61 (2015) 73–82.
[121] M. John, C. Schmitz, A.Y. Park, N.E. Dixon, T. Huber, G. Otting, Sequence-specific and stereospecific assignment of methyl groups using paramagnetic lanthanides, J. Am. Chem. Soc. 129 (2007) 13749–13757.
[122] V. Venditti, N.L. Fawzi, G.M. Clore, Automated sequence- and stereo-specific assignment of methyl-labeled proteins by paramagnetic relaxation and methyl-methyl nuclear overhauser enhancement spectroscopy, J. Biomol. NMR 51 (2011) 319–328.
[123] G. Otting, Protein NMR using paramagnetic ions, Ann. Rev. Biophys. 39 (2010) 387–405.
[124] M. Lescanne, S.P. Skinner, A. Blok, M. Timmer, L. Cerofolini, M. Fragai, C. Luchinat, M. Ubbink, Methyl group assignment using pseudocontact shifts with PARAssign, J. Biomol. NMR 69 (2017) 183–195.

[125] P.H.J. Keizers, J.F. Desreux, M. Overhand, M. Ubbink, Increased paramagnetic effect of a lanthanide protein probe by two-point attachment, J. Am. Chem. Soc. 129 (2007) 9292–9293.
[126] D. Häussinger, J.R. Huang, S. Grzesiek, DOTA-M8: an extremely rigid, high-affinity lanthanide chelating tag for PCS NMR spectroscopy, J. Am. Chem. Soc. 131 (2009) 14761–14767.
[127] N.L. Fawzi, M.R. Fleissner, N.J. Anthis, T. Kalai, K. Hideg, W.L. Hubbell, G.M. Clore, A rigid disulfide-linked nitroxide side chain simplifies the quantitative analysis of PRE data, J. Biomol. NMR 51 (2011) 105–114.
[128] I. Pritisanac, J.M. Würz, P. Güntert, Fully automated assignment of methyl resonances of a 36 kDa protein dimer from sparse NOESY data, J. Phys. Conf. Ser. 1036 (2018) 012008.
[129] J. Kim, Y.J. Wang, G. Li, G. Veglia, A semiautomated assignment protocol for methyl group side chains in large proteins, Methods Enzymol. 566 (2016) 35–57.
[130] C.D. Schwieters, J.J. Kuszewski, N. Tjandra, G.M. Clore, The Xplor-NIH NMR molecular structure determination package, J. Magn. Reson. 160 (2003) 65–73.
[131] J.L. Battiste, G. Wagner, Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear Overhauser effect data, Biochemistry 39 (2000) 5355–5365.
[132] C. Schmitz, M. John, A.Y. Park, N.E. Dixon, G. Otting, G. Pintacuda, T. Huber, Efficient c-tensor determination and NH assignment of paramagnetic proteins, J. Biomol. NMR 35 (2006) 79–87.
[133] H.L. Eaton, S.W. Fesik, S.J. Glaser, G.P. Drobny, Time-dependence of ^{13}C–^{13}C magnetization transfer in isotropic mixing experiments involving amino-acid spin systems, J. Magn. Reson. 90 (1990) 452–463.
[134] L. Siemons, H.W. Mackenzie, V.K. Shukla, D.F. Hansen, Intra-residue methyl-methyl correlations for valine and leucine residues in large proteins from a 3D-HMBC-HMQC experiment, J. Biomol. NMR 73 (2019) 749–757.
[135] D.S. Wishart, Interpreting protein chemical shift data, Prog. Nucl. Magn. Reson. Spectrosc. 58 (2011) 62–87.
[136] D.W. Li, R. Brüschweiler, PPM_One: a static protein structure based chemical shift predictor, J. Biomol. NMR 62 (2015) 403–409.
[137] B. Han, Y.F. Liu, S.W. Ginzinger, D.S. Wishart, SHIFTX2: significantly improved protein chemical shift prediction, J. Biomol. NMR 50 (2011) 43–57.
[138] A.B. Sahakyan, W.F. Vranken, A. Cavalli, M. Vendruscolo, Structure-based prediction of methyl chemical shifts in proteins, J. Biomol. NMR 50 (2011) 331–346.
[139] G. Wagner, S.G. Hyberts, T.F. Havel, NMR structure determination in solution: A critique and comparison with X-ray crystallography, Annu. Rev. Biophys. Biomol. Struct. 21 (1992) 167–198.
[140] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (1953) 1087–1092.
[141] R. Lichtenecker, M.L. Ludwiczek, W. Schmid, R. Konrat, Simplification of protein NOESY spectra using bioorganic precursor synthesis and NMR spectral editing, J. Am. Chem. Soc. 126 (2004) 5348–5349.
[142] J.J. McGregor, Backtrack search algorithms and the maximal common subgraph problem, Softw.-Pract. Exp. 12 (1982) 23–34.
[143] H. Bunke, Graph matching: Theoretical foundations, algorithms, and applications, in: Proceedings of Vision Interface 2000, Montreal, Algorithmica, 2000.
[144] H.W. Kuhn, The Hungarian method for the assignment problem, Nav. Res. Logist. 52 (2005) 7–21.
[145] J. Munkres, Algorithms for the assignment and transportation problems, J. Soc. Indust. Appl. Math. 5 (1957) 32–38.
[146] L.P. Cordella, P. Foggia, C. Sansone, M. Vento, A (sub)graph isomorphism algorithm for matching large graphs, IEEE Trans. Pattern Anal. Mach. Intell. 26 (2004) 1367–1372.
[147] E. Schmidt, P. Güntert, Automated structure determination from NMR spectra, Meth. Mol. Biol. 1261 (2015) 303–329.
[148] T. Aeschbacher, E. Schmidt, M. Blatter, C. Maris, O. Duss, F.H.-T. Allain, P. Güntert, M. Schubert, Automated and assisted RNA resonance assignment using NMR chemical shift statistics, Nucleic Acids Res. 41 (2013) e172.
[149] D.F. Gauto, L.F. Estrozi, C.D. Schwieters, G. Effantin, P. Macek, R. Sounier, A.C. Sivertsen, E. Schmidt, R. Kerfah, G. Mas, J.P. Colletier, P. Güntert, A. Favier, G. Schoehn, P. Schanda, J. Boisbouvier, Integrated NMR and cryo-EM atomic-resolution structure determination of a half-megadalton enzyme complex, Nat. Commun. 10 (2019) 2697.
[150] B. Krähenbühl, I. El Bakkali, E. Schmidt, P. Güntert, G. Wider, Automated NMR resonance assignment strategy for RNA via the phosphodiester backbone based on high-dimensional through-bond APSY experiments, J. Biomol. NMR 59 (2014) 87–93.
[151] E. Schmidt, J. Gath, B. Habenstein, F. Ravotti, K. Székely, M. Huber, L. Buchner, A. Böckmann, B.H. Meier, P. Güntert, Automated solid-state NMR resonance assignment of protein microcrystals and amyloids, J. Biomol. NMR 56 (2013) 243–254.
[152] E. Schmidt, T. Ikeya, M. Takeda, F. Löhr, L. Buchner, Y. Ito, M. Kainosho, P. Güntert, Automated resonance assignment of the 21 kDa stereo-array isotope labeled thioldisulfide oxidoreductase DsbA, J. Magn. Reson. 249 (2014) 88–93.
[153] J.M. Würz, S. Kazemi, E. Schmidt, A. Bagaria, P. Güntert, NMR-based automated protein structure determination, Arch. Biochem. Biophys. 628 (2017) 24–32.
[154] J. Stanek, T. Schubeis, P. Paluch, P. Güntert, L.B. Andreas, G. Pintacuda, Automated backbone NMR resonance assignment of large proteins using redundant linking from a single simultaneous acquisition, J. Am. Chem Soc. 142 (2020) 5793–5799.
[155] E. Schmidt, P. Güntert, Reliability of exclusively NOESY-based automated resonance assignment and structure determination of proteins, J. Biomol. NMR 57 (2013) 193–204.
[156] T. Ikeya, J.-G. Jee, Y. Shigemitsu, J. Hamatsu, M. Mishima, Y. Ito, M. Kainosho, P. Güntert, Exclusively NOESY-based automated NMR assignment and structure determination of proteins, J. Biomol. NMR 50 (2011) 137–146.
[157] L. Kooijman, P. Ansorge, M. Schuster, C. Baumann, F. Löhr, S. Jurt, P. Güntert, O. Zerbe, Backbone and methyl assignment of bacteriorhodopsin incorporated into nanodiscs, J. Biomol. NMR 74 (2020) 45–60.
[158] C. Bartels, P. Güntert, M. Billeter, K. Wüthrich, GARANT - A general algorithm for resonance assignment of multidimensional nuclear magnetic resonance spectra, J. Comput. Chem. 18 (1997) 139–149.
[159] Q. Gao, G.R. Chalmers, K.W. Moremen, J.H. Prestegard, NMR assignments of sparsely labeled proteins using a genetic algorithm, J. Biomol. NMR 67 (2017) 283–294.
[160] Q. Gao, C.Y. Chen, C. Zong, S. Wang, A. Ramiah, P. Prabhakar, L.C. Morris, G.J. Boons, K.W. Moremen, J.H. Prestegard, Structural aspects of heparan sulfate binding to Robo1-Ig1-2, ACS Chem. Biol. 11 (2016) 3106–3113.
[161] Y. Shen, A. Bax, SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network, J. Biomol. NMR 48 (2010) 13–22.
[162] M. Zweckstetter, NMR: prediction of molecular alignment from structure using the PALES software, Nat. Protoc. 3 (2008) 679–690.
[163] H. Valafar, J.H. Prestegard, REDCAT: a residual dipolar coupling analysis tool, J. Magn. Reson. 167 (2004) 228–241.
[164] K. Pederson, G.R. Chalmers, Q. Gao, D. Elnatan, T.A. Ramelot, L.C. Ma, G.T. Montelione, M.A. Kennedy, D.A. Agard, J.H. Prestegard, NMR characterization of HtpG, the *E. coli* Hsp90, using sparse labeling with ^{13}C-methyl alanine, J. Biomol. NMR 68 (2017) 225–236.
[165] W.K. Nkari, J.H. Prestegard, NMR resonance assignments of sparsely labeled proteins: Amide proton exchange correlations in native and denatured states, J. Am. Chem. Soc. 131 (2009) 5344–5349.
[166] F. Yu, J. Qiao, J. Robblee, D. Tsao, J. Anderson, I. Capila, An integrated approach to unique NMR assignment of methionine methyl resonances in proteins, Anal. Chem. 89 (2017) 1610–1616.

Glossary

$\alpha_7\alpha_7$: "half-proteasome" 20S core particle
ATCase: dimer of regulatory chains of aspartate transcarbamoylase from *E. coli*
BMRB: Biological Magnetic Resonance Data Bank
EIN: N-terminal domain of *E. coli* Enzyme I
HMQC: heteronuclear multiple quantum coherence
HSP90: N-terminal domain of heat shock protein 90
HSQC: heteronuclear single quantum coherence
MC: Monte Carlo
MBP: maltose binding protein
MSG: malate synthase G
NOE: nuclear Overhauser effect
NOESY: NOE spectroscopy
PCS: pseudocontact shift
PDB: Protein Data Bank
PRE: paramagnetic relaxation enhancement
SAIL: stereo-array isotope labeling
s.d.: standard deviation
TOCSY: total correlation spectroscopy
TROSY: transverse relaxation optimized spectroscopy