Published in final edited form as:

Methods Enzymol. 2016 ; 566: 35–57. doi:10.1016/bs.mie.2015.08.033.

A Semi-Automated Assignment Protocol for Methyl Group Side-Chains in Large Proteins

Jonggul Kim¹, Yingjie Wang¹, Geoffrey Li¹, and Gianluigi Veglia^{1,2,*}

¹Department of Chemistry- University of Minnesota, Minneapolis, MN 55455

²Department of Biochemistry, Molecular Biology, and Biophysics— University of Minnesota, Minneapolis, MN 55455.

Abstract

The developments of biosynthetic specific labeling strategies for side-chain methyl groups have allowed structural and dynamic characterization of very large proteins and protein complexes. However, the assignment of the methyl-group resonances remains an Achilles' heel for NMR, as the experiments designed to correlate side chains to the protein backbone become rather insensitive with the increase of the transverse relaxation rates. In this chapter, we outline a semi-empirical approach to assign the resonances of methyl group side chains in large proteins. This method requires a crystal structure or an NMR ensemble of conformers as an input, together with NMR data sets such as NOEs and PREs, to be implemented in a computational protocol that provides a probabilistic assignment of methyl group resonances. As an example, we report the protocol used in our laboratory to assign the side chains of the 42-kDa catalytic subunit of the cAMP-dependent protein kinase A. Although we emphasize the labeling of isoleucine, leucine, and valine residues, this method is applicable to other methyl group side chains such as those of alanine, methionine, and threonine, as well as reductively-methylated cysteine side chains.

INTRODUCTION

Isotopic labeling is at the heart of NMR spectroscopy. After the pioneering work in the early '60s by several research groups (H. L. Crespi & Katz, 1969; H. L. Crespi, Rosenberg, R.M. Katzz, J.J., 1968; Markley, Putter, & Jardetzky, 1968; Putter, Barreto, Markley, & Jardetzky, 1969), more complex labeling schemes have been used to attenuate transverse relaxation and push the boundaries of NMR analysis of large macromolecular complexes (Kainosho et al., 2006; Kasinath, Valentine, & Wand, 2013; Meissner & Sorensen, 1999). A significant step forward was made when non-labile protons in proteins were substituted by deuterons, ameliorating the dipolar relaxation due to ^1H - ^{13}C , ^1H - ^{15}N , and ^1H - ^1H interactions, increasing sensitivity and resolution, thereby rendering proteins larger than 20 kDa amenable to structural NMR studies (Nietlispach et al., 1996; Venters et al., 1995). Although complete deuteration has been used for backbone assignments in conjunction with transverse relaxation optimized spectroscopy (TROSY) (Pervushin, Riek, Wider, & Wüthrich, 1997),

* To whom correspondence should be addressed: Gianluigi Veglia, Department of Biochemistry, Biophysics, and Molecular Biology, University of Minnesota, 6-155 Jackson Hall, MN 55455. Telephone: (612) 625-0758. Fax: (612) 625-2163. vegli001@umn.edu.

NOESY-based experiments benefit from incomplete (or fractional) deuteration, which preserves some protons for distance measurements (Nietlispach et al., 1996). Another quantum leap in the NMR structure determination of large complexes was made by Kay and co-workers, who developed a biosynthetic strategy for ^1H , ^{13}C methyl group labeling in a highly deuterated background (Gardner & Kay, 1997; Goto, Gardner, Mueller, Willis, & Kay, 1999). Since then, this labeling strategy has expanded to include selective methyl group labeling such as isoleucine, valine, leucine, methionine (Gelís et al., 2007), threonine (Velyvis, Ruschak, & Kay, 2012), and alanine (Ayala, Sounier, Use, Gans, & Boisbouvier, 2009). Although selective labeling schemes for aromatic side chains will certainly have an impact both in structure determination and dynamic characterization of large proteins (Kasinath et al., 2013), here we will focus on the use of methyl groups for the spectroscopy of large systems (Tugarinov, Hwang, Ollerenshaw, & Kay, 2003) akin to the TROSY method developed for backbone amides (Pervushin et al., 1997). Following these methyl labeling schemes, quantitative studies on binding, structure, and conformational dynamics of proteins that are several hundred kDa are emerging, such as large molecular machinery (Religa, Sprangers, & Kay, 2010; Ruschak, Religa, Breuer, Witt, & Kay, 2010; Sprangers & Kay, 2007), allosteric enzymes (Shi & Kay, 2014), chaperones (Saio, Guan, Rossi, Economou, & Kalodimos, 2014), and protein thermodynamics (Tzeng & Kalodimos, 2012).

Labeling of side chain methyl groups for large proteins

The first methyl labeling scheme was published in 1997 by Gardner and Kay (Gardner & Kay, 1997), who obtained a selective protonation of amino acids in which the direct biosynthetic precursor to isoleucine, 2-ketobutyrate, was enzymatically and chemically prepared with selective $^{13}\text{CH}_3$ labeling in a highly deuterated background. Exploiting the *Escherichia coli* (*E. coli*) biosynthetic pathway, this protocol enables the specific $^{13}\text{CH}_3$ labeling of the C δ of Ile, while the remaining non-labile protons are replaced by deuterons. The incorporation of 2-ketoisovalerate, in a similar fashion, leads to selective labeling on the C δ and C γ of leucine and valine, respectively (Goto et al., 1999), resulting in the so-called “ILV labeling” scheme (**Figure 1**). Nowadays, these biosynthetic precursors are commercially available in a variety of labeling schemes for assignment, structure determination, and dynamics studies. Selective labeling of the C γ methyl of isoleucine has also been devised (Ruschak, Velyvis, & Kay, 2010), but is uncommon due to the superior spectral qualities of the C γ methyl group. Usually, selective labeling of other amino acids is achieved through direct addition of the amino acid to the growth medium during cell growth. For alanine and methionine, this is most commonly accomplished by directly incorporating the amino acid prior to induction (30 minutes to 1 hour) at final concentrations of 100-250 mg/L (Saio et al., 2014; Tzeng & Kalodimos, 2012) for Met and from 100-800 mg/L for alanine (Ayala et al., 2009; Tzeng & Kalodimos, 2012). An alternate approach is to engineer cysteine and react with ^{13}C -methyl-methanethiosulfonate (^{13}C -MMTS) to produce a methionine mimic (Religa, Ruschak, Rosenzweig, & Kay, 2011). Kay and coworkers have also developed an *in vitro* biosynthetic method to produce U-[^2H], Thr- γ 2[$^{13}\text{CH}_3$] ($^{13}\text{CH}_3$ labeled threonine in a deuterated background) (Velyvis et al., 2012). Threonine is a direct biosynthetic precursor to 2-ketobutyrate, leading to isoleucine methyl (C δ) labeling. Since threonine is converted to glycine either from threonine aldolase or threonine dehydrogenase/2-amino-3-ketobutyrate ligase, the authors used 100 mg/L of d_5 -glycine

along with 50 mg/L of U-[²H] Thr-γ2[¹³CH₃] and 50 mg/L of 2-ketobutyrate for specific threonine and isoleucine labeling (**Figure 1**).

Methyl labeling protocol for the cAMP-dependent Protein Kinase A (PKA-C)

In this outline, we highlight a method to label the ¹³C isoleucine (C8), leucine, and valine methyl groups for the catalytic subunit of cAMP-dependent Protein Kinase A (PKA-C) (**Figure 2**). Specific labeling of side-chain methyl groups is achieved using the M9 minimal medium typically used for ¹⁵N labeling of proteins with the addition of appropriate amounts of methyl labeling precursors. As the protocols for overexpression and purification of PKA-C have been described extensively (Hemmer, McGlone, & Taylor, 1997; Narayana, Cox, Shaltiel, Taylor, & Xuong, 1997; Yonemoto, McGlone, Slice, & Taylor, 1991), here we report only the main steps of the protocol (**Figure 3**). Note that for binding titrations studies we used fully protonated kinase and 2D NMR experiments; however, for methyl group assignment experiments (i.e., 3D [¹³C,¹³C]-methyl NOESY experiments) we utilized fractionally deuterated kinase expressed in 80% D₂O. In our hands, the utilization of fully perdeuterated media resulted in a scarce expression of the kinase. Nonetheless, for larger proteins with molecular weights greater than 100 kDa, high levels of deuteration are often necessary. In the latter case, over-expression of proteins in 100% D₂O is highly strain-, vector-, and protein-dependent and requires optimization for each of these variables. Several tips for the optimization of perdeuterated growths have been provided by Gardner *et al.* (Gardner, Zhang, Gehring, & Kay, 1998).

Materials

Note: All stocks should be sterile before usage.

LB Media (1L)—10.0 g Tryptone

5.0 g Bacto-Yeast

10.0 g NaCl

M9 Media (1L)—3.0 g KH₂PO₄

12.8 g Na₂HPO₄·7H₂O

1.0 g NH₄Cl (for ¹⁵N labeling use ¹⁵NH₄Cl)

0.5 g NaCl

Vitamins (50mL)—5.0 mg Ca D(+)-panthothenate

5.0 mg Biotin

5.0 mg Folic Acid

5.0 mg Niacinamide USP

5.0 mg Pyridoxal 5-phosphate monohydrate

1000 mg Thiamin

Trace Metals (1x, 100mL)—50 mM FeCl₃

20 mM CaCl₂

10 mM MnCl₂

10 mM ZnSO₄

2 mM CoCl₂

2 mM CuCl₂

2 mM NiCl₂

2 mM Na₂MoO₄

2 mM Na₂SeO₃

2 mM H₃BO₃

Note: More details in the preparation of Trace Metal solution is provided at Studier et. al. (Studier, 2005)

Keto-acid Stock (10 mL, water)—2-ketobutyric acid-4-¹³C sodium salt (70mg/L of media)

2-keto-3-(methyl-¹³C)-butyric acid-4-¹³C sodium salt (90mg/L of media)

Calcium Chloride

Magnesium Sulfate

D-glucose

Ampicillin

Isopropyl β-D-1-thiogalactopyranoside (IPTG)

1. Place 500 mL of M9 media into baffled flasks and autoclave.
2. Grow BL21 (DE3) cells containing the pRSET plasmid with PKA-C overnight in LB media (with Ampicillin).
3. In the morning the culture should be in log phase growth. Then pellet the BL21 (DE3) cells under 2200xg for 10 minutes. During this time prepare add the following per 500 mL flask containing M9 media:
 - a. 10 ml of 20% w/v glucose
 - b. 5 mL of Vitamin stock

- c. 1 mL of 1 M MgSO₄.
 - d. 1 mL of 50 mM CaCl₂
 - e. 500 µL of 1000x Trace Metals
 - f. 500 µL of 10% w/v Ampicillin stock
4. Resuspend the cell pellet in M9 media and allow to grow for at least an hour to allow for the culture to adapt to the adjusted metabolism. We usually have an initial OD₆₀₀ of ~0.8.
 5. Once the cells reach log phase growth, lower the temperature to 24°C and add in the appropriate concentration of keto-acids into each flask (70 mg/L of 2-ketobutyric acid-4- ¹³C sodium salt and 90 mg/L of 2-keto-3-(methyl-¹³C)-butyric acid-4-¹³C sodium salt).
 6. After one hour, induce with 0.4 mM IPTG for at least 5 hours. Harvest under 8000xg for 25 minutes at 4°C.

Purification of PKA-C

Buffer A (1L)

- 30 mM MOPS (3-(N-morpholino)propanesulfonic acid)	(628g)
- 15 mM MgCl ₂	(305g)
- 5 mM β-mercaptoethanol	(352µl/L of buffer)
- pH 8.0	

Buffer B (1L)

- 30 mM MOPS	(6.28g)
- 15 mM MgCl ₂	(3.05g)
- 25 mM KCl	(1.86g)
- 5 mM β-mercaptoethanol	(352µl/L of buffer)
- pH 8.0	

Elution Buffer (100mL)

- 30 mM MOPS	(628 mg)
- 15 mM MgCl ₂	(305 mg)
- 25 mM KCl	(186 mg)
- 1 mM cAMP	(33 mg)
- 5 mM β-mercaptoethanol	(35.2µl/100 mL of buffer)
- pH 8.0	

Cell Lysis

- 1) Re-suspend PKA and the His₆-RII β (R213K) overexpressed cells in Buffer A. Make sure the pellet containing RII α was grown in at least 1/3 of the volume of your PKA.
- 2) Grind the cells in ice for approximately 10 minutes then lyse using a French press using 1000 psi of pressure.
- 3) Spin down the solution at 20,000 rpm for 40 minutes at 4°C. Mix with 10 mL of NiNTA slurry and batch bind at 4°C for a minimum of 3 hours.

Elution

- 1) Run the slurry through the column (make sure the resin does not dry).
- 2) Wash with Buffer B (~200 mL). If preferred the wash buffer can also include 20 mM Imidazole to remove some non-specific binding proteins. This will not affect the PKA:RII α interaction.
- 3) Elute PKA-C using 150 mL of elution buffer.
- 4) To remove RII α wash with ~50 mL of Buffer B with 250 mM imidazole.

Although mass spectrometry is a viable method to assess the labeling of proteins with stable isotopes, the most straightforward way to determine the correct isotopic incorporation is to acquire the [¹H, ¹³C]-HMQC spectra of the system of interest. The characteristic chemical shift patterns allow one to easily identify isoleucine (C δ), leucine, and valine methyl groups as they resonate in distinct regions of the ¹H/¹³C correlated spectra. Due to the high sensitivity of methyl groups, 2D spectra of large proteins (~40 kDa) with reasonable sensitivity (S/N ratio 200-400) can be measured with modern spectrometers (700-900 MHz) with relatively modest concentrations (~200 μ M). A [¹H, ¹³C]-HMQC spectrum of apo PKA-C shows the labeling and spectral quality of ILV-methyl groups labeling (**Figure 4**). Below, we outline the NMR sample preparation that we use for *apo* form of PKA-C:

Materials

Potassium Phosphate Monobasic

Potassium Chloride

Dithiothreitol

Magnesium Chloride

Sodium Azide

10kDa Centrifugal Concentrator

99+% Deuterium Oxide

Long-tip Pipet, 13-1/4"

5 mm Shigemi Tube

1. Dialyze purified PKA-C under NMR buffer (20 mM KH_2PO_4 , 90 mM KCl, 10 mM DTT, 10 mM MgCl_2 , 1mM NaN_3 at pH 6.5) overnight at 4°C. Initial concentration of PKA-C should < 5 μM .
2. Concentrate PKA-C at 2400xg using a 10kDa cutoff centrifugal concentrator at 4°C until the enzyme is concentrated to ~200 μM . The protein concentration can reliably be measured by absorbance measurement at 280 nm using an extinction coefficient of 52060 $\text{M}^{-1}\text{cm}^{-1}$.
3. Pipette 250 μL (for Agilent spectrometers) or 300 μL (for Bruker spectrometers) and add 12.5 or 15 μL of D_2O (~5% v/v), respectively. Transfer the protein solution into a shigemi tube using a long-tip pipet. Place the plunger inside the tube until the plunger is evenly covered with no visible bubbles in the solution.
4. After spectrometer set up (spectrometer details are dependent on model and spectrometer) acquire an $[\text{H},^{13}\text{C}]$ -HMQC spectrum.

Semi-Automated Methyl Group Resonance Assignment Strategies

For backbone studies, the assignment is carried out by walking through the peptide backbone through a series of triple resonance NMR experiments (Kay, Ikura, Tschudin, & Bax, 1990; Salzmänn, Pervushin, Wider, Senn, & Wuthrich, 1998; Salzmänn, Wider, Pervushin, Senn, & Wuthrich, 1999; Yang & Kay, 1999). For the assignment of methyl groups resonances, the classical approach requires TOCSY-based spectroscopy with the magnetization starting on the methyl groups and transferred *via* ^{13}C - ^{13}C couplings to the protein backbone (Sattler, Schleucher, & Griesinger, 1999). This approach, however, is not applicable to methyl group side-chains of large proteins, such as the 82 kDa malate synthase G with a correlation time of ~46 ns (Tugarinov & Kay, 2003b). In the latter case, the branched amino acids present a bifurcation of the magnetization pathways with a significant loss of transfer. Although several experiments have been tailored to correlate backbone $\text{C}\alpha$, $\text{C}\beta$ and C' chemical shifts to methyl groups (Sheppard, Guo, & Tugarinov, 2009; Tugarinov & Kay, 2003a; Tugarinov, Venditti, & Marius Clore, 2014), the relatively fast transverse relaxation rates of these nuclei for larger proteins or conformational exchange phenomena (see the case of PKA-C (Srivastava et al., 2014)) do not allow to carry out sequential backbone assignments. A possible solution is to use a *divide and conquer* approach in which proteins are dissected into several smaller domains and the assignments are pieced together (Saio et al., 2014; Velyvis, Schachman, & Kay, 2009). However, there are many large monomeric proteins that are unamenable to this approach. An alternative strategy is to use site-specific mutations to exchange amino acids with methylated side chains for glycine or other amino acids. The latter method, however, is very cumbersome and time-consuming, particularly for very large systems. In response to this need, Matthews and coworkers proposed an automatic method that uses both NMR and X-ray structural data to back calculate methyl side chain assignments (Xu et al., 2009). This automated procedure was designed to rapidly assign methyl groups using full-length proteins and without mutagenesis (Xu et al., 2009). This approach requires a crystal structure as a starting point to calculate

methyl group chemical shifts and synthetic NOEs. An algorithm is then used to rank the scores of the initial (seed) assignment against the experimental NMR data. The latest version of this method features an improved algorithm with no manual intervention in NOESY spectra peak picking.

Inspired by this work, we developed an approach that uses a similar philosophy, but utilizes a probabilistic method to tackle assignment problem. Specifically, we improved the efficiency of the phase space search using *fuzzy logic* coupled with Monte Carlo sampling. This approach, named FLAMEnGO (Fuzzy Logic Assignment of Methyl Groups), utilizes NOESY data as the primary input in concert with other data sets, such as methine-methyl TOCSY data, as well as paramagnetic relaxation enhancements (PREs). In our newer version of this software, FLAMEnGO 2.0 (Chao et al., 2014), we introduced a graphic interface that enables the interactive assignment of the methyl group resonances with the option of including secondary constraints, such as mutagenesis or supplemental NMR data.

In the following synopsis, we describe the construction and usage of FLAMEnGO 2.0, as well as its application to the side-chain methyl groups of PKA-C.

Semi-Automated Assignment Protocol using FLAMEnGO 2.0

All of the available methyl auto-assignment algorithms share the same philosophy: searching for the best assignments by comparing experimental NOE connectivities and chemical shifts to back-calculated data from the X-ray coordinates (Chao et al., 2014; Chao, Shi, Masterson, & Veglia, 2012; Venditti, Fawzi, & Clore, 2011; Xu et al., 2009; Xu & Matthews, 2013). Starting from the available crystal structure or NMR structural bundle, FLAMEnGO 2.0 back-calculates a 3D/4D NOESY spectrum from the internuclear methyl-methyl distances and compares synthetic data to an experimental NOESY spectrum (**Figure 5**). The algorithm iteratively swaps the resonance assignments until the scoring function is maximized. This process is repeated iteratively for an array of different simulated NOE distances until no further improvement is found. Given the probabilistic nature of this algorithm, multiple runs are performed in parallel to provide a statistical weight to the resonance assignments. Although FLAMEnGO is designed to handle sparse NOE data sets, sufficient coverage of the NOE network and other restraints (i.e., methine-methyl TOCSY or methyl-HN COSY, PREs or site-directed mutagenesis) are required for a confident assignment. The outline of the procedure for using FLAMEnGO GUI is shown in **Figure 6**.

Below, we detail the experimental data and data formatting, as well as brief instructions on usage of FLAMEnGO 2.0.

Materials

- 3D C,C,H ^{13}C -HMQC NOESY ^{13}C -HMQC data set (These pulse sequences are part of the TOPSPIN release for Bruker spectrometers, and are downloadable from the NMR-FAM website http://pine.nmrfam.wisc.edu/download_pulseprogs.html for Agilent spectrometers).
- 2D [^1H , ^{13}C]-HMQC data set (required)

- Predicted chemical shift table (required, sources listed below)
- Paramagnetic Relaxation enhancement data (optional)
- Methine-methyl TOCSY data (optional)
- Assigned amide chemical shifts (optional, if amide-methyl NOESY data are incorporated)
- PDB file for protein. Note that hydrogen atoms need to be added to the PDB. Molprobability (<http://molprobability.biochem.duke.edu/>) is a resource to perform this function.

System Requirements and Specifications

The python-based FLAMEnGO 2.0 program includes two main components: 1) the frontend GUI program `flame.py`, and 2) the backend engine `FLAMEnGO.py`. The former provides a convenient interface for configuring the setup and displaying the results, while the latter is the driver for a Monte-Carlo search. The software is ready to run under Python 2.7, and requires wxPython and matplotlib modules for GUI and plotting. An integrated scientific python package, such as Enthought Canopy (<https://www.enthought.com/products/canopy/>) or Anaconda Python (<https://store.continuum.io/cshop/anaconda/>) are needed for full functionality.

Input File Formats

FLAMEnGO 2.0 uses a Monte Carlo search algorithm to swap resonance assignments from given input data and computes the *fitness* of the resonance assignments based on a score function. The details of the scoring function are discussed in Chao *et al.* (Chao et al., 2014; Chao et al., 2012).

Initial Random Assignment—The algorithm needs an *arbitrary* initial assignment to start the swapping process associated with the Monte Carlo algorithm. This *seed* assignment must be consistent with PDB numbering and included in all input data. The data are formatted in a series of columns by <assignment #> <amino acid type> <resonance atom ID> -> <assignment #> <amino acid type> <resonance atom ID>. Note that there is no change in residue, rather this indicates to the program which residues to assign. To constrain a specific resonance assignment (*e.g.*, found using site-directed mutagenesis), the user should place an asterisk at the beginning of the line:

296	VAL	HG2	->	296	VAL	HG2
*301	ILE	HD1	->	301	ILE	HD1
321	ILE	HD1	->	321	ILE	HD1
323	VAL	HG1	->	323	VAL	HG1
323	VAL	HG2	->	323	VAL	HG2
325	ILE	HD1	->	325	ILE	HD1

3D C,C,H¹³C-HMQC-NOESY-¹³C-HMQC—The 3D C,C,H ¹³C-HMQC-NOESY-¹³C-HMQC data need to be formatted in a series of columns in the following order: <¹³C (ω2) chemical shift> <¹H (ω3) chemical shift> <¹³C (ω1) chemical shift> <¹³C (ω2) linewidth> <¹H (ω3) linewidth> <¹³C (ω1) linewidth>

18.32	1.28	27.41	0.4	0.025	0.4
22.76	0.88	19.53	0.4	0.025	0.4
22.76	0.88	15.69	0.4	0.025	0.4
22.76	0.88	21.55	0.4	0.025	0.4
22.76	0.88	18.66	0.4	0.025	0.4
20.97	0.78	17.77	0.4	0.025	0.4
25.05	0.68	23.33	0.4	0.025	0.4

[¹H,¹³C]-HMQC/¹⁵N Amide—The data is formatted in a series of columns by <assignment #> <amino acid type> <resonance atom ID> <chemical shift value>. Note that the assignment reported in this file at the beginning of the calculation is arbitrary for the [¹H,¹³C]-HMQC. The amide chemical shifts are optional and only required if one is using amide to methyl NOE data. The amide chemical shifts must be assigned based on previous experiments.

7	LEU	HD1	0.34
7	LEU	HD2	0.56
7	LEU	CD1	25.39
7	LEU	CD2	24.56
20	LEU	HD1	0.88
20	LEU	HD2	0.64
20	LEU	CD1	25.10
20	LEU	CD2	27.22

Paramagnetic Relaxation Enhancement Data (PREs)—Although the PRE data are listed as an option, we demonstrated that the inclusion of these constraints in the computation dramatically increases both the speed and accuracy of the resonance assignments (Chao et al., 2014; Chao et al., 2012). In particular, PRE data are crucial to solve the ambiguities caused by partial peak overlap (**Figure 7**). PRE data are included as distances using the semi-quantitative approach described by Battiste and Wagner (Battiste & Wagner, 2000), in which the ratios of the peak intensities in the presence of the oxidized versus the reduced spin label are converted into three different categories: weakly affected (< 20% of signal is lost through the PRE effect), moderately affected (20-80% of the signal is lost) or strongly effected (<20% of the signal is lost). These categories are flagged as W, M, and S in the input file. If a resonance is unobserved regardless of the PRE effect, then a flag of 999 is used in the 4th column of the input file. The data are formatted in four columns, including <assignment #> <amino acid type> <resonance atom ID> <Flag (W, M, S, 999)>:

89	HD1	LEU	999
92	HD2	LEU	999
102	HD1	LEU	M
102	HD2	LEU	M
71	HD1	ILE	S
92	HD1	LEU	S
138	HD1	LEU	W
13	HD1	LEU	W

Prediction of Methyl group chemical shifts from the PDB file—Various algorithms are used to predict the side chain chemical shift values. Output from any commonly used program, such as CH3SHIFT(Sahakyan, Vranken, Cavalli, & Vendruscolo, 2011) (<http://www.sidechain.ch.cam.ac.uk/CH3Shift/>), SHIFTX2 (Han, Liu, Ginzinger, & Wishart, 2011), (<http://www.shiftx2.ca/>) and PPM (Li & Bruschweiler, 2012) (<http://spin.ccic.ohio-state.edu/index.php/ppm>) can be implemented in FLAMEnGO 2.0. Since the output is not uniform between programs, data must be reformatted as follows: <residue #> <amino acid type> resonance atom ID> <predicted chemical shift value> <error of prediction>

1	VAL	CG2	19.992	1.128
1	VAL	HG2	0.828	0.173
5	LEU	CD1	25.747	1.412
5	LEU	HD1	0.969	0.176

Residue Type—Amino acid-type assignments can be included as constraints in the search algorithm. The data format is identical to that of the initial assignment file.

Instructions to run FLAMEnGO 2.0 GUI

1. Open FLAMEnGO
“python flame.py”
2. Specify the location of each input file on the main window..
3. To set up FLAMEnGO go under **File > SetUp (Figure 8A)**. Set the NOE range between 5.0-10.0 with step size of 0.5. The sampling number is the number of Monte Carlo calculations run at each point and should be over 100k (**Figure 8B**).
4. Select the **Run** button. Please note that calculations take a considerable amount of time.
5. Once the runs are finished a plot of the score with respect to the NOE distance will appear (**Figure 8C**). To minimize the uncertainty of the

assignment, click on the point where the NOE value with the maximum score is first reached (**Figure 8D**)

6. Select the number of calculations you wish to run (10). The final output will give the assignment and the statistical weight of each assignment.

Output Files

FLAMEnGO provides two types of output files. The first is the assignment file, which provides the calculated assignment at each given calculation at a specified NOE cutoff distance. Once the aggregate calculations are performed at a cutoff distance, the probability of each assignment is calculated. The probability of each assignment from the calculations is what is used to determine the *fitness* of the assignment.

Score and Assignment—The output file provides the user with the percent of the residues assigned, the assignment score and the assignment from the single calculation. The assignment is provided as the following <initial assignment> -> <calculated assignment. On the first line the first two columns provide the <final scoring value> <% of residues assigned>.

(205.2936878260806, 93.079554272662747, 193, 17.844744345905379)

#####

158	LEU	HD1	->	148	LEU	HD1
237	VAL	HG1	->	90	VAL	HG2
321	ILE	HD1	->	196	ILE	HD1
259	LEU	HD2	->	191	LEU	HD2
197	LEU	HD1	->	75	LEU	HD1

Summary—The summary file will provide the assignment, as well as the number of times the particular assignment was made across all the calculations. A confident assignment is defined in this case as one that is consistent through all calculations at least 80% of the time; a reliable statistic for this assessment requires 10 calculations. For each assignment the output is as follows: <initial assignment> -> <calculated assignment> <number of times the assignment was calculated>.

105	VAL	HG1	->	137	VAL	HG1	10
254	LEU	HD1	->	68	LEU	HD1	10
75	LEU	HD2	->	263	LEU	HD2	6
143	LEU	HD2	->	45	LEU	HD2	7
13	LEU	HD1	->	259	LEU	HD1	10
241	VAL	HG2	->	137	VAL	HG2	10

Conclusions and Perspectives

Selective labeling of the methyl group side-chains has extended the application of NMR methods to the analysis of both structure and dynamics of proteins up to 1 MDa. However, as the systems become larger, classical methods for resonance assignments fail. In this chapter, we reported a semi-empirical approach that enables the assignment of methyl group resonances in large proteins. This structure-based approach requires the X-ray or the NMR ensemble of structures to generate a probability-based assignment. Inclusion of NOEs, PRE, site-specific mutagenesis data, methine-methyl TOCSY, and partial assignment from methyl-HN COSY data dramatically improves both the accuracy and the speed of the assignment procedure. This method promises to extend biomolecular NMR studies beyond the MDa limit.

Acknowledgements

This work was supported in part by the NIH (GM100310 and GM72701 to GV and T32 AR007612 to JK). NMR experiments were carried out at the Minnesota NMR Center and FLAMEnGO 2.0 calculations at the Minnesota Supercomputing Institute.

References

- Ayala I, Sounier R, Use N, Gans P, Boisbouvier J. An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. *J Biomol NMR*. 2009; 43(2):111–119. doi: 10.1007/s10858-008-9294-7. [PubMed: 19115043]
- Battiste JL, Wagner G. 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*. 2000; 39(18):5355–5365. [PubMed: 10820006]
- Chao FA, Kim J, Xia Y, Milligan M, Rowe N, Veglia G. FLAMEnGO 2.0: An enhanced fuzzy logic algorithm for structure-based assignment of methyl group resonances. *J Magn Reson*. 2014; 245:17–23. doi: 10.1016/j.jmr.2014.04.012. [PubMed: 24915505]
- Chao FA, Shi L, Masterson LR, Veglia G. FLAMEnGO: a fuzzy logic approach for methyl group assignment using NOESY and paramagnetic relaxation enhancement data. *J Magn Reson*. 2012; 214(1):103–110. doi: 10.1016/j.jmr.2011.10.008. [PubMed: 22134225]
- Crespi HL, Katz JJ. High resolution proton magnetic resonance studies of fully deuterated and isotope hybrid proteins. *Nature*. 1969; 224(5219):560–562. [PubMed: 5346595]
- Crespi HL, Rosenberg RM, Katzz JJ. proton magnetic resonance of proteins fully deuterated except for 1H-leucine side chains. *Science*. 1968; 161:795–796. [PubMed: 5663808]
- Gardner KH, Kay LE. Production and Incorporation of 15N, 13C, 2H (1H-δ1 Methyl) Isoleucine into Proteins for Multidimensional NMR Studies. *Journal of the American Chemical Society*. 1997; 119(32):7599–7600. doi: 10.1021/ja9706514.
- Gardner KH, Zhang X, Gehring K, Kay LE. Solution NMR Studies of a 42 KDa Escherichia Coli Maltose Binding Protein/β-Cyclodextrin Complex: Chemical Shift Assignments and Analysis. *Journal of the American Chemical Society*. 1998; 120(45):11738–11748. doi: 10.1021/ja982019w.
- Gelis I, Bonvin AM, Keramisanou D, Koukaki M, Gouridis G, Karamanou S, Kalodimos CG. Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. *Cell*. 2007; 131(4):756–769. doi: 10.1016/j.cell.2007.09.039. [PubMed: 18022369]
- Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE. A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated 15N-, 13C-, 2H-labeled proteins. *J Biomol NMR*. 1999; 13(4):369–374. doi: 10.1023/A:1008393201236. [PubMed: 10383198]
- Han B, Liu Y, Ginzinger SW, Wishart DS. SHIFTX2: significantly improved protein chemical shift prediction. *J Biomol NMR*. 2011; 50(1):43–57. doi: 10.1007/s10858-011-9478-4. [PubMed: 21448735]

- Hemmer W, McGlone M, Taylor SS. Recombinant strategies for rapid purification of catalytic subunits of cAMP-dependent protein kinase. *Anal Biochem.* 1997; 245(2):115–122. doi: 10.1006/abio.1996.9952. [PubMed: 9056191]
- Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Mei Ono A, Guntert P. Optimal isotope labelling for NMR protein structure determinations. *Nature.* 2006; 440(7080):52–57. doi: 10.1038/nature04525. [PubMed: 16511487]
- Kasinath V, Valentine KG, Wand AJ. A ¹³C labeling strategy reveals a range of aromatic side chain motion in calmodulin. *J Am Chem Soc.* 2013; 135(26):9560–9563. doi: 10.1021/ja4001129. [PubMed: 23767407]
- Kay LE, Ikura M, Tschudin R, Bax A. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *Journal of Magnetic Resonance (1969).* 1990; 89(3):496–514. doi: 10.1016/0022-2364(90)90333-5.
- Li DW, Bruschweiler R. PPM: a side-chain and backbone chemical shift predictor for the assessment of protein conformational ensembles. *J Biomol NMR.* 2012; 54(3):257–265. doi: 10.1007/s10858-012-9668-8. [PubMed: 22972619]
- Markley JL, Putter I, Jardetzky O. High-resolution nuclear magnetic resonance spectra of selectively deuterated staphylococcal nuclease. *Science.* 1968; 161(3847):1249–1251. doi: 10.1126/science.161.3847.1249. [PubMed: 5673435]
- Meissner A, Sorensen OW. Optimization of three-dimensional TROSY-type HCCH NMR correlation of aromatic (1)H-(13)C groups in proteins. *J Magn Reson.* 1999; 139(2):447–450. doi: 10.1006/jmre.1999.1796. [PubMed: 10423385]
- Narayana N, Cox S, Shaltiel S, Taylor SS, Xuong N. Crystal structure of a polyhistidine-tagged recombinant catalytic subunit of cAMP-dependent protein kinase complexed with the peptide inhibitor PKI(5-24) and adenosine. *Biochemistry.* 1997; 36(15):4438–4448. doi: 10.1021/bi961947+ [PubMed: 9109651]
- Nietlispach D, Clowes RT, Broadhurst RW, Ito Y, Keeler J, Kelly M, Laue ED. An Approach to the Structure Determination of Larger Proteins Using Triple Resonance NMR Experiments in Conjunction with Random Fractional Deuteration. *Journal of the American Chemical Society.* 1996; 118(2):407–415. doi: 10.1021/ja952207b.
- Pervushin K, Riek R, Wider G, Wüthrich K. Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences.* 1997; 94(23):12366–12371.
- Putter I, Barreto A, Markley JL, Jardetzky O. Nuclear magnetic resonance studies of the structure and binding sites of enzymes. X. Preparation of selectively deuterated analogs of staphylococcal nuclease. *Proc Natl Acad Sci U S A.* 1969; 64(4):1396–1403. [PubMed: 5271760]
- Religa TL, Ruschak AM, Rosenzweig R, Kay LE. Site-directed methyl group labeling as an NMR probe of structure and dynamics in supramolecular protein systems: applications to the proteasome and to the ClpP protease. *J Am Chem Soc.* 2011; 133(23):9063–9068. doi: 10.1021/ja202259a. [PubMed: 21557628]
- Religa TL, Sprangers R, Kay LE. Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. *Science.* 2010; 328(5974):98–102. doi: 10.1126/science.1184991. [PubMed: 20360109]
- Ruschak AM, Religa TL, Breuer S, Witt S, Kay LE. The proteasome antechamber maintains substrates in an unfolded state. *Nature.* 2010; 467(7317):868–871. doi: 10.1038/nature09444. [PubMed: 20944750]
- Ruschak AM, Velyvis A, Kay LE. A simple strategy for (1)(3)C, (1)H labeling at the Ileγ2 methyl position in highly deuterated proteins. *J Biomol NMR.* 2010; 48(3):129–135. doi: 10.1007/s10858-010-9449-1. [PubMed: 20949307]
- Sahakyan AB, Vranken WF, Cavalli A, Vendruscolo M. Structure-based prediction of methyl chemical shifts in proteins. *J Biomol NMR.* 2011; 50(4):331–346. doi: 10.1007/s10858-011-9524-2. [PubMed: 21748266]

- Saio T, Guan X, Rossi P, Economou A, Kalodimos CG. Structural basis for protein antiaggregation activity of the trigger factor chaperone. *Science*. 2014; 344(6184):1250494. doi: 10.1126/science.1250494. [PubMed: 24812405]
- Salzmann M, Pervushin K, Wider G, Senn H, Wuthrich K. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc Natl Acad Sci U S A*. 1998; 95(23):13585–13590. doi: 10.1073/pnas.95.23.13585. [PubMed: 9811843]
- Salzmann M, Wider G, Pervushin K, Senn H, Wuthrich K. TROSY-type triple-resonance experiments for sequential NMR assignments of large proteins. *Journal of the American Chemical Society*. 1999; 121(4):844–848. doi: 10.1021/Ja9834226.
- Sattler M, Schleucher J, Griesinger C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solutions employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* 1999; 34:93–158.
- Sheppard D, Guo C, Tugarinov V. Methyl-detected ‘out-and-back’ NMR experiments for simultaneous assignments of Alabeta and Ilegamma2 methyl groups in large proteins. *J Biomol NMR*. 2009; 43(4):229–238. doi: 10.1007/s10858-009-9305-3. [PubMed: 19274445]
- Shi L, Kay LE. Tracing an allosteric pathway regulating the activity of the HslV protease. *Proc Natl Acad Sci U S A*. 2014; 111(6):2140–2145. doi: 10.1073/pnas.1318476111. [PubMed: 24469799]
- Sprangers R, Kay LE. Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature*. 2007; 445(7128):618–622. doi: 10.1038/nature05512. [PubMed: 17237764]
- Srivastava AK, McDonald LR, Cembran A, Kim J, Masterson LR, McClendon CL, Veglia G. Synchronous Opening and Closing Motions Are Essential for cAMP-Dependent Protein Kinase A Signaling. *Structure*. 2014; 22(12):1735–1743. doi: 10.1016/j.str.2014.09.010. [PubMed: 25458836]
- Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif*. 2005; 41(1):207–234. [PubMed: 15915565]
- Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE. Cross-correlated relaxation enhanced ¹H[¹bond]¹³C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J Am Chem Soc*. 2003; 125(34):10420–10428. doi: 10.1021/ja030153x. [PubMed: 12926967]
- Tugarinov V, Kay LE. 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*. 2003a; 125(45):13868–13878. doi: 10.1021/ja030345s. [PubMed: 14599227]
- Tugarinov V, Kay LE. Side chain assignments of Ile delta 1 methyl groups in high molecular weight proteins: an application to a 46 ns tumbling molecule. *J Am Chem Soc*. 2003b; 125(19):5701–5706. doi: 10.1021/ja021452+ [PubMed: 12733908]
- Tugarinov V, Venditti V, Marius Clore G. A NMR experiment for simultaneous correlations of valine and leucine/isoleucine methyls with carbonyl chemical shifts in proteins. *J Biomol NMR*. 2014; 58(1):1–8. doi: 10.1007/s10858-013-9803-1. [PubMed: 24346684]
- Tzeng SR, Kalodimos CG. Protein activity regulation by conformational entropy. *Nature*. 2012; 488(7410):236–240. doi: 10.1038/nature11271. [PubMed: 22801505]
- Velyvis A, Ruschak AM, Kay LE. An economical method for production of (2)H, (13)CH3-threonine for solution NMR studies of large protein complexes: application to the 670 kDa proteasome. *PLoS One*. 2012; 7(9):e43725. doi: 10.1371/journal.pone.0043725. [PubMed: 22984438]
- Velyvis A, Schachman HK, Kay LE. Assignment of Ile, Leu, and Val methyl correlations in supra-molecular systems: an application to aspartate transcarbamoylase. *J Am Chem Soc*. 2009; 131(45):16534–16543. doi: 10.1021/ja906978r. [PubMed: 19860411]
- Venditti V, Fawzi NL, Clore GM. Automated sequence- and stereo-specific assignment of methyl-labeled proteins by paramagnetic relaxation and methyl-methyl nuclear Overhauser enhancement spectroscopy. *J Biomol NMR*. 2011; 51(3):319–328. doi: 10.1007/s10858-011-9559-4. [PubMed: 21935714]
- Venters RA, Huang CC, Farmer BT 2nd, Trolard R, Spicer LD, Fierke CA. High-level 2H/13C/15N labeling of proteins for NMR studies. *J Biomol NMR*. 1995; 5(4):339–344. doi: 10.1007/BF00182275. [PubMed: 7647552]

- Xu Y, Liu M, Simpson PJ, Isaacson R, Cota E, Marchant J, Matthews S. Automated assignment in selectively methyl-labeled proteins. *J Am Chem Soc.* 2009; 131(27):9480–9481. doi: 10.1021/ja9020233. [PubMed: 19534551]
- Xu Y, Matthews S. MAP-XSII: an improved program for the automatic assignment of methyl resonances in large proteins. *J Biomol NMR.* 2013; 55(2):179–187. doi: 10.1007/s10858-012-9700-z. [PubMed: 23292498]
- Yang DW, Kay LE. TROSY triple-resonance four-dimensional NMR spectroscopy of a 46 ns tumbling protein. *Journal of the American Chemical Society.* 1999; 121(11):2571–2575. doi: 10.1021/Ja984056t.
- Yonemoto WM, McGlone ML, Slice LW, Taylor SS. Prokaryotic expression of catalytic subunit of adenosine cyclic monophosphate-dependent protein kinase. *Methods Enzymol.* 1991; 200:581–596. [PubMed: 1956337]

Figure 1. Biosynthetic pathways for the specific labeling of A) isoleucine, B) leucine and valine, C) alanine, D) methionine and E) threonine. Note the off-pathway dilution of alanine and threonine amino acids.

Figure 2.

General growth scheme used to express ^{15}N ^{13}C -ILV labeled PKA-C. All cultures are grown in 32°C with induction occurring at 24°C . This is to prevent the formation of inclusion bodies during expression (Yonemoto et al., 1991). In M9 media, $^{15}\text{NH}_4\text{Cl}$ (1 g/L) and D-glucose (4 g/L) are used as the only nitrogen and carbon sources respectively.

Figure 3.
SDS-PAGE of the purification of PKA-C using a His₆-RII α immobilized construct.

Figure 4. $[\text{}^1\text{H}, \text{}^{13}\text{C}]$ -HMQC of Apo PKA-C acquired with a 220 μM sample on an Avance III 900 MHz Bruker spectrometer. The spectrum was acquired with 2048×200 complex points and 16 transients at 27°C for a total acquisition time of approximately 1 ½ hours.

Figure 5. Overview of the FLAMEnGO algorithm. An input structure and the $[^1\text{H}, ^{13}\text{C}]$ -HMQC spectrum are used to simulate a NOE spectrum. This spectrum is compared with the experimental NOE data, the assignments swapped and the process is repeated until a best match is found. Other experimental restraints, such as PRE data, are used. Figure adapted from Chao, et al. (Chao et al., 2012)

Figure 6.
Outline for running FLAMEnGO GUI.

Figure 7. Spectra of oxidized (green) and reduced (red) PKA-C with a spin label on residue 244. Note the spectrum should be nearly identical before and after the reduction of the spin label.

Figure 8. Graphical Interface for FLAMEnGO. A) Main menu of FLAMEnGO. Input files are directly placed inside the selection. B) Parameter set up for the NOE distances, amino acid types you wish to assign, and the number of Monte Carlo steps. C) Output from the run. Note that the program performs the calculation at each NOE distance. D) Once the score has plateaued from increasing the NOE distance, select the point and repeat the calculation at this distance to provide a statistical assignment of each residue. Figure adapted from (Chao et al., 2014)