Developement of strategies for the isotopic labeling of methyl groups for the NMR study of large protein assemblies

Rime Kerfah

UNIVERSITÉ DE GRENOBLE

THÈSE
Pour obtenir le grade de
DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE
Spécialité : Biologie structurale et nanobiologie

Présentée par
Rime Kerfah

Thèse dirigée par **Jérôme Boisbouvier**

Préparée au sein du **Groupe de Spéctrosopie RMN Biomoléculaire**

École Doctorale Chimie et Sciences du Vivant

Développement de stratégies de marquage isotopique des groupements méthyles pour l'étude d'assemblages protéiques de grande taille par RMN

Date de soutenance le 26 Septembre 2014,
devant le jury composé de :

Bruno Kieffer	Président
Carine Tisné	Rapporteur
Ewen Lescop	Rapporteur
Olivier Hamelin	Membre
Michael Plevin	Membre
Jérôme Boisbouvier	Directeur de thèse

ACKNOWLEDGMENTS

I wish to thank my supervisor, Dr. Boisbouvier, for the guidance and generosity that he has provided me. I consider it a privilege to have had the opportunity in working within his research group.

I would like to acknowledge with sincere gratitude to Elodie and Pavel for their assistance in my early life in science and also Guillaume for his ingenious IT advises.

I am also really grateful for the scientific support of Pierre and Dominique and for their cultural discussions.

I would like to thank the whole NMR, group, especially Adrien, Karine, Catherine, Lionel, Isabel, Paul, Hugo and Jia-ying for the collaboration and the excellent work environment that have proportioned to me.

My thesis would not be possible without the presence of my parents, brothers and my best friend, Amal. For all of them I would like to express my gratitude for being always by my side.

To conclude, I would like to thank Victor, who has been my energy source to pursue this step of my life and who has always managed to put a smile on my face in difficult circumstances.

Content

Abstract

Solution NMR spectroscopy has been limited to small biological objects for a long time. Nowadays, it is unequivocally recognized that the strategy of specific isotope labeling of methyl groups in a perdeuterated protein has significantly extended the frontier of this technique. Indeed, proteins as large as 1 MDa could be investigated by NMR. Conversely, this strategy presents an important drawback consisting of the drastically reduced number of protonated probes. The project of this thesis falls within the framework of developing new methodologies to cope with this scarce structural information, relying on the simultaneous labeling of several methyl groups to increase the number of probes.

For optimized combinatorial labeling, the choice of the amino acids to label simultaneously and the precursors as well as the protocol for their incorporation has to be carefully studied. In this work, a new protocol was introduced for the scrambling-free and optimized isotopic labeling of $A^{\beta}I^{\delta1}(LV)^{proS}$ methyl groups. In comparison to the "standard $A^{\beta}I^{\delta1}LV$" labeling scheme, the proposed pattern induces a 2-fold decrease of number of Leu and Val NMR signals and enhances the intensity of detectable long-range nOes by a factor 4. The described protocol also permits the suppression of spurious correlations, especially harmful for structural studies based on detection/analysis of nOes.

To make an efficient use of the obtained high quality NMR spectra using this protocol, assignment of the methyl groups signals is mandatory. Two strategies were then proposed. The first is suitable for systems whose molecular weight does not exceed 100 kDa. It relies on the use of isotopically linearized precursors (with different isotope topologies to discriminate each methyl group) to assign in a regio- and stereo-specific manner the isoleucine, leucine and valine methyl groups in a single step, employing an optimized "out and back" ^{13}C-TOCSY pulse sequence. While the second, adapted to supra-molecular proteins (> 100 kDa), consists of optimizing the previously reported SeSAM approach (Sequence-Specific Assignment of Methyl groups by Mutagenesis). Indeed, thanks to the developed enriched culture medium for the specific labeling of Ala, the minimal required culture volume was significantly decreased, enabling the proteins expression in 24 well plates and their parallel purification in 96 well plates. This improved SeSAM version was estimated to allow the assignment of *ca.* 100 methyl cross-peaks in 2 weeks, including 4 days of NMR time and less than 2 k€ of isotopic materials.

To illustrate the pertinence of using selectively protonated methyl groups, either in a single or combined fashion, several applications were presented, namely the real-time NMR study of self-assembly process of a ~0.5 MDa supra-molecular protein (PhTET-2). The use of combinatorial labeling for the detection of long-range nOes to up to 10 Å (8 Å) in proteins of 82 kDa (respectively 0.5 MDa) was also investigated. This same approach was exploited for the filtering of inter-monomeric long-range nOes in the same symmetrical and homo-oligomeric PhTET-2 protein.

Résumé

Pendant longtemps la spectroscopie RMN en solution a été limitée à de petits objets biologiques. Aujourd'hui, il est clairement reconnu que la stratégie du marquage isotopique spécifique de groupements méthyle dans une protéine perdeutérée a considérablement repoussé la frontière de cette technique. En effet, des protéines aussi grande que 1 MDa ont pu être récemment étudiées par RMN. Cependant, cette stratégie présente un inconvénient important lié au nombre réduit de sondes protonées. Le projet de cette thèse s'inscrit dans le cadre du développement de nouvelles méthodes pour faire face à cette rareté d'information structurale, en s'appuyant sur le marquage simultané (combinatoire) de plusieurs groupements méthyle afin d'augmenter le nombre de sondes.

Pour un marquage combinatoire optimal, le choix des acides aminés à marquer et les précurseurs à utiliser ainsi que le protocole pour leur incorporation doivent être judicieusement étudié. Dans ce travail, un nouveau protocole a été mis en place pour un marquage $A^{\beta}I^{\delta 1}(LV)^{proS}$ optimisé, exempt de toutes fuites isotopiques. En comparaison avec le marquage "$A^{\beta}I^{\delta 1}LV$ standard", le modèle proposé mène à la diminution d'un facteur de 2 du nombre de signaux RMN des Leu et Val et améliore par un facteur de 4 l'intensité des nOes à long portée qui sont expérimentalement détéctable. Par ailleurs, ce protocole permet également la suppression des corrélations parasites (i.e., Ile-γ_2), particulièrement nocives pour les études structurales basées sur la détection / analyse de nOes.

Afin d'exploiter les spectres RMN obtenus en utilisant le protocole ci-dessus mentioné, l'attribution des signaux des méthyles est obligatoire. Deux stratégies ont été donc proposées. La première est applicable aux systèmes dont le poids moléculaire ne dépasse pas 100 kDa (e.g. MSG). Elle se repose sur la linéarisation du marquage isotopique des acides aminés pemettant ainsi l'utilisation de l'expérience ^{13}C-TOCSY pour attribuer d'une manière régio et stéréo-spécifique les methyles de l'isoleucine, leucine et la valine en une seule étape. Quant à la seconde approche, adaptée aux proteines supra-moléculaire (> 100 kDa), consiste à une optimisation de l'approche SeSAM (Sequence-Specific Assignment of Methyl groups by Mutagenesis) précédemment rapportée dans la litérature. En effet, grâce au milieu de culture enrichi qui a été mis au point pour le marquage spécifique de l'Ala, le volume minimal de culture requis a été considérablement diminué. Ceci a permis par conséquence l'expression des protéines dans des plaques à 24 puits et leur purification parallèle dans des plaques à 96 puits, raccourcissant ainsi le temps global de préparation des échantillons. Il a été estimé que l'utilisation de cette version améliorée de SeSAM offre la possibilité d'attribuer environ 100 méthyles en 2 semaines, dont 4 jours de temps de la RMN, en consommant moins de 2 k € de matériaux isotopiques.

Pour illustrer la pertinence du marquage isotopique et protonation selectifs des méthyles, d'une façon combinée ou pas, de nombreuses applications ont été présentés, à savoir l'étude en temps réel du processus d'auto-assemblage d'une protéine supramoléculaire de ~ 0,5 MDa (PhTET-2) par RMN. Le marquage combinatoire pour la détection de nOes longue portée jusqu'à 10 Å (8 Å) en protéines de 82 kDa (respectivement 0,5 MDa) a également été étudié. Cette même approche a aussi été exploitée pour le filtrage de nOes inter-monomères à long portée, qui sont particulièrement importants pour le calcul de structure, dans des systèmes symétrique et homo-oligomèriques (PhTET-2).

CHAPTER
1

INTRODUCTION

1.1 Overview of isotope labeling in solution NMR spectroscopy:

Solution state nuclear magnetic resonance (NMR) spectroscopy is now established as a powerful tool to study structure, dynamics and interactions of biomolecules in atomic details. Indeed, the applicability of this technique has evolved significantly over the years.

The standard multidimensional ^{1}H homonuclear NMR, commonly used in the 1980s, was limited to objects up to 10 kDa (Qian, Billeter et al. 1989). Larger molecules were challenging due to the inherent spectral overlap and rapid transverse relaxation (R_2) of the nuclei. The presence of many protons in these systems results in overcrowding of NMR signals, hindering single peak identification (assignment) and, therefore, NMR data exploitation. Additionally, the rapid magnetization losses originated by the increase in correlation time (τ_c) and ^{1}H-^{1}H dipolar interactions, leads to broadening of the NMR signal and even to its complete disappearance.

To overcome the abovementioned limitations, several solutions have been subsequently developed in terms of hardware, e.g., cryogenic probes, high-field spectrometers (up to 1GHz) and NMR pulse sequences. Nevertheless, the relevant step that has considerably increased the range of biochemical systems that can be addressed by solution NMR spectroscopy was the introduction of protein isotopic labeling techniques.

Initially, multidimensional heteronuclear NMR, where ^{15}N, ^{13}C and ^{1}H are correlated, has permitted a better signal dispersion and, thus, studies of proteins with slightly larger sizes (~20 kDa) (Torchia, Sparks et al. 1988, Kay, Torchia et al. 1989, Ikura, Marion et al. 1990). Partial protein deuteration (e.g., 70% deuteration) (Yamazaki, Lee et al. 1994, Muchmore, Sattler et al. 1996) was used to improve relaxation properties of the manipulated spins, allowing the recovery of a certain sensitivity in systems as large as ~40 kDa. In fact, due to its low gyromagnetic ratio ($\gamma[^{2}H]/\gamma[^{1}H]=0,15$), when deuterium ($^{2}H$) substitutes the protons, ^{1}H-^{1}H dipolar interactions are reduced leading to slower ^{1}H and ^{13}C transverse relaxation. For even larger systems, a more drastic proton dilution is required, resulting in a low proton occupancy per single site. This protonation pattern

hampers the acquisition of a suitable signal-to-noise ratio for the NMR spectra analysis. At this point, the ambition to visualize all the ^{1}H sites, which were poorly populated, had to be sacrificed. Thus, a new concept arises whose principle relies on silencing most of the visible sites in favor of fully protonating only a fraction, offering the possibility to acquire restricted but high quality NMR spectra. Reintroducing exchangeable protons, such as backbone amide protons in perdeuterated proteins, is the most commonly used approach (Venters, Huang et al. 1995, Garrett, Seok et al. 1997). However, representing an average of 20% of the total proton density in proteins, this type of protonation reaches rapidly its limit when the NMR study targets supra-molecular systems (> 100 kDa).

To push the size limit of protein NMR, the strategy of selective protonation of methyl groups was proposed (Rosen, Gardner et al. 1996, Gardner and Kay 1997). In this approach, the favorable relaxation properties and the proton multiplicity of the methyl groups have been widely exploited to obtain high quality NMR spectra of large proteins up to 1 MDa (see fig. 1.1 and §1.3).

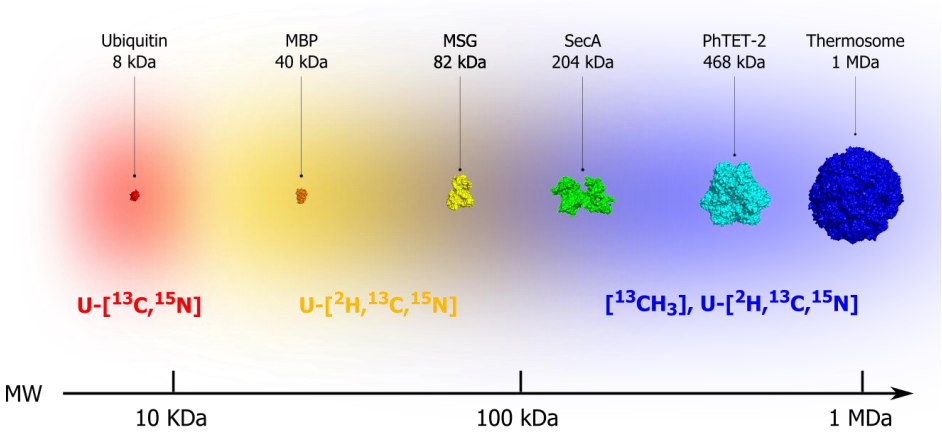


Figure 1.1: Evolution of the size of biological systems studied by solution NMR *vs* isotopic labeling. Ubiquitin (1ubq); maltose binding protein (MBP) (1dmb); malate synthase G (MSG) (1p7t); ATPase motor of Sec translocase (SecA) (2vda); tetrahedral aminopeptidase (PhTET-2) (1y0r);Thermosome (1a6d). Molecular weight ranges to which different isotope labeling strategies are suited.

The sparse structural information obtained from the selective protonation strategy, however, impedes 3D structure determination. Nevertheless, a broad range of NMR analyses remain realizable, providing valuable insights into the structure (Sprangers and

Kay 2007, Velyvis, Yang et al. 2007), dynamics (Sprangers and Kay 2007, Tugarinov, Sprangers et al. 2007), interactions and functions of supra-molecular (Sprangers, Velyvis et al. 2007). Furthermore, the combined selective protonation of both amide and methyl groups has helped to extend, although not in an infinite manner, the size limit of structure determination by NMR. This has been demonstrated in the calculation of the global fold of a 82 kDa monomeric protein (Tugarinov, Choy et al. 2005).

The work presented in this thesis focuses on the development of novel methodologies for studying large proteins by solution NMR spectroscopy, relying on the selective protonation and labeling of methyl groups technology, a prerequisite to examine such systems.

In this manuscript, I will first provide an overview on the state-of-the-art strategies for the selective protonation in large systems (> 50 kDa), with a focus on labeling of methyl group (Chapter 1). Then, I will introduce a new protocol for scrambling-free and optimized combinatorial labeling of alanine, isoleucine, leucine and valine residues (chapter 2). To assign the labeled methyl probes, I will describe two methods (chapter 3): The first, suitable for systems whose molecular weight does not exceed 100 kDa, relies on the use of differently labeled precursors to assign in a regio- and stereo-specific manner the isoleucine, leucine and valine methyl groups in a single step. While the second, adapted to supra-molecular proteins (> 100 kDa), consists of optimizing the previously reported SeSAM approach (Sequence-Specific Assignment of Methyl groups by Mutagenesis). To illustrate the pertinence of using selectively protonated methyl groups, either in a single or combined fashion, I will present three applications (Chapter 4): The real-time NMR study of self-assembly process of a supra-molecular protein (~0.5 MDa), the use of combinatorial labeling for the detection of long-range nOes and the filtering of inter-monomeric long-range nOes in symmetrical and homo-oligomeric proteins. Finally, I will conclude by summarizing the presented work and perspectives towards future improvements for isotope labeling methodologies for NMR studies (Chapter 5).

1.2 Selective protonation and isotopic labeling in a perdeuterated biological system

The diversity of NMR experiments offers the possibility of monitoring a broad range of biological questions at atomic resolution. However, their application is limited when the particle size increases, requiring the perdeuteration labeling where few protons are reintroduced at given sites (≥ 50 kDa). In this chapter, I will present the most commonly used strategies for producing selectively protonated proteins and will describe the strategic atomic sites for selective protonation.

1.2.1 Production of labeled proteins

Generally, proteins are expressed using *in vivo* systems, such as prokaryotic or eukaryotic cells. In recent years, the *in vitro* approach has evolved and has become a complementary useful approach. However, the use of this approach remains limited because it requires particular expertise, e.g., knowledge of cell extract preparation, expensive RNase-free consumables and specific laboratory spaces.
In this section, I will discuss the primary advantages and drawbacks of the most used expression systems for NMR sample preparation.

1.2.1.1 *In vivo protein production*

- Bacteria

Among the available expression systems, the use of *Escherichia coli* (*E. coli*), particularly the BL21 (DE3) strain, is by far the most popular and cost-effective system. The accumulated knowledge concerning its genetic and metabolism, together with its short generation time and ability to adapt to adverse conditions, make this microorganism the most appropriate candidate for $^{2}H_2O$-labeled protein expression.
The use of heavy water affects negatively its growth and overall kinetics; however, compared with other cells, *E. coli* still presents a relatively high $^{2}H_2O$ tolerance after adaptation steps.
Briefly, over-expressing perdeuterated proteins is performed in a completely deuterated culture medium, where the only proton source is the specifically protonated precursor (or amino acid) of the targeted residue for labeling.

Thus far, the use of labeled precursors represents the best approach. However, the feasibility of this strategy remains dependent on the availability of the labeled precursor and the existence of a metabolic pathway that allows its efficient incorporation without isotope scrambling. This topic will be extensively reviewed in §1.3.2.

The *E. coli* system has many positive aspects but still presents some limitations, namely, reduced protein yields (associated with decreased growth caused by $^{2}H_2O$ toxicity) and difficulty in expressing many eukaryotic protein targets.

In an attempt to cope with the deficient growth of this bacterium in $^{2}H_2O$, enriched culture media are alternatively used. Such media can be readily purchased (e.g., BioExpress®1000) or prepared by supplementing minimum media with algal (e.g., Isogro®, Celtone®) (Crublet, Kerfah et al. 2014) or bacterial hydrolysates (e.g., Silantes® *E.coli*-OD2). These products, available with different patterns of [^{13}C,^{15}N,^{2}H] uniform labeling, have been proven to be extremely efficient in boosting cell growth and, consequently, improving the protein yield (Crublet, Kerfah et al. 2014). However, when selective amino acid labeling is intended, their use is more challenging. In fact, their high content in unlabeled amino acids creates a "strong" competition against the supplemented labeled version, leading to the low incorporation of such expensive material (see §1.2.2.11).

In conclusion, although $^{2}H_2O$ toxicity can be managed in certain cases, the incapability of *E. coli* to produce many eukaryotic proteins cannot be solved. Therefore, the use of a eukaryotic system is mandatory.

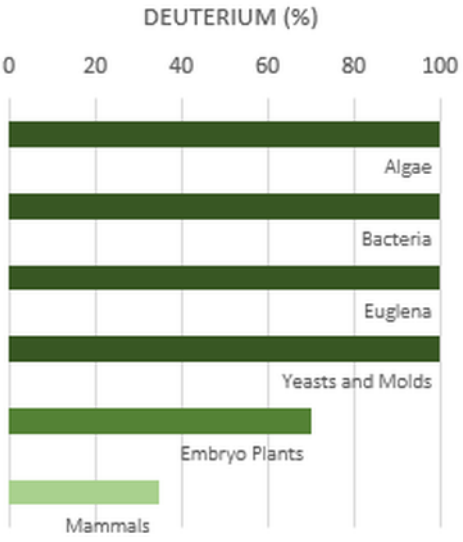


Figure 1.2: Deuterium tolerance of representative members of diverse organism classes. (Katz and Crespi 1966)

- Eukaryotic Systems

For their correct folding and stability, several eukaryotic proteins may present the requirement of post-translational modifications, and/or disulfide bond formation, and/or presence of some chaperones. Unable to provide the cited requisites, the *E. coli* utilization for their expression ceases to be appropriate.

In the literature, various investigations have reported the use of mammalian and insect cells, such as the mouse myeloma Sp2/0 cell line (Hansen, Petros et al. 1992), Chinese hamster ovary cells (CHO)(Archer, Bax et al. 1993, Wyss, Dayie et al. 1997), the mouse hybridoma cell line (Shindo, Masuda et al. 2000), the human embryonic kidney cells (HEK293)(Werner, Richter et al. 2008) and *Spodoptera frugiperda* insect ovary cells (Strauss, Bitsch et al. 2003, Strauss, Bitsch et al. 2005, Gossert, Hinniger et al. 2011) to produce isotopically (^{13}C and ^{15}N) labeled proteins. However, the expensiveness together with the limited $^{2}H_2O$ tolerance of these systems (see fig. 1.2) imposes the use of yeast cells as an alternative for expressing deuterated eukaryotic proteins.

Yeast cells

During the last fifteen years, the use of yeast for the NMR protein preparation has attracted considerable attention for the following reasons: (*i*) ability to perform many mammalian post-translational modifications; (*ii*) tolerance to $^{2}H_2O$ (Katz and Crespi 1966); (*iii*) high expression yields with simple and affordable isotopic labeling; (*iv*) ease of genetic manipulation; and finally (*iv*) ability to secrete the overexpressed protein, coping with its eventual toxicity. Several yeast strains, such as *Saccharomyces cerevisiae* (Lemercinier, Muskett et al. 2001, Uhrinova, Uhrin et al. 2001, Chang, Keller et al. 2002, Landon, Barbault et al. 2008, Jaremko, Jaremko et al. 2014), *Pichia angusta* (Massou, Puech et al. 1999) , *Pichia pastoris* (Wood and Komives 1999, Morgan, Kragt et al. 2000, Rodriguez and Krishna 2001, Pickford and O'Leary 2004) and *Kluyveromyces lactis* (Sugiki, Shimada et al. 2008) have been explored to express isotopically-labeled proteins. Thus far, *P. pastoris* is the most "popular" for generating higher expression yields (up to 500 mg/L in fermenters) and faithful native patterns of glycosylation. Indeed, various studies reported its successful use for the *U*-[^{2}H, ^{13}C or ^{15}N] protein production in 95-99% $^{2}H_2O$ medium culture (Massou, Puech et al. 1999, Morgan, Kragt et al. 2000, Ichikawa, Osawa et al. 2007). Depending on the use or not of deuterated carbon sources along the whole culture (in growth and induction phases); the protein deuteration levels were reported to vary from 85% to more than 95%.

Recently, the production of perdeuterated maltose binding protein (MBP) (42 kDa), selectively labeled and protonated at the isoleucine (Ile or I), leucine (Leu or L) and valine (Val or V) methyl groups in *K. lactis* has been described (Miyazawa-Onami,

Takeuchi et al. 2013). This labeling was first tested using α-ketoacid precursors (see §1.3.3), i.e., labeled α-ketoisovalerate and α-ketobutyrate in the wild-type microorganism. The resulted enrichment was estimated to be 67 ± 6%, 2.2 ± 0.3% and 2.4 ± 0.4% for Ile-δ_1, Leu$^{pro\mathit{R}/pro\mathit{S}}$ and Val $^{pro\mathit{R}/pro\mathit{S}}$, respectively. Subsequently, the MBP was co-expressed with the bacterial branched chain amino acid aminotransferase (BCAT)[a] to compensate for the low uptake of α-ketoisovalerate through the mitochondria, where yeast BCAT is thought to be exclusively localized. The obtained enrichment using the same precursors was quantified to 88 ± 7%, 10 ± 2% and 15 ± 3% for Ile, Leu and Val, respectively. The result was generally improved but remained unsatisfactory in the cases of Leu and Val. Finally, these latter residues could be more efficiently labeled, i.e., 85% and 95% respectively, through [2,3-$^{2}H_2$, *U*-^{13}C] Val addition. The high isotopic labeling of the Ile residue using its cost-effective precursor alone represents an interesting result. Indeed, this finding indicates that using affordable precursors for amino acid labeling is feasible in yeast. Nevertheless, more optimizations, e.g., genetic lesions, and better understanding of metabolism are certainly required for the efficient manipulation of this system.

In conclusion, the yeast cell is clearly an extremely powerful expression system that will have a broader application in NMR. However, its ability to accomplish some post-translational modifications, such as prolyl hydroxylation as well as a certain type of phosphorylation and high mannose glycosylation, remains limited (Takahashi and Shimada 2010).

1.2.1.2 *In vitro protein production*

The cell-free (CF) approach consists of artificially using the protein transcriptional/translational machinery of eukaryotic or prokaryotic cells to synthesize the protein of interest (see table 1.1 and fig. 1.3). Cell extracts are used after depleting the DNA and RNA of the original organism, allowing the exclusive expression of the supplied DNA/RNA of the protein target, facilitating in turn its subsequent purification. Generally, this type of approach is requested when the particle of interest presents one of the following characteristics: (*i*) is toxic for the cell expression system, (*ii*) is

[a] Branched chain aminotransferases catalyzes transamination of the branched chain amino acids: leucine, isoleucine, and valine.

subjected to *in vivo* proteolysis degradation, or (*iii*) is aggregated into inclusion bodies. The CF is an open system, offering the possibility to add important substances, e.g., cofactors, ligands and protein stabilizers, at any time to improve protein synthesis. For instance, membrane proteins could be correctly produced because their expression was feasible in the presence of a wide variety of detergents/lipids (Cappuccio, Hinz et al. 2009, Goren, Nozawa et al. 2009, Chalmeau, Monina et al. 2011, May, Andreasson-Ochsner et al. 2013, Linser, Gelev et al. 2014).

Over the years, CF approach knew a great technical enhancement. The initially confronted obstacles, such as rapidly depleting essential substrates (due to e.g., the presence nucleases, proteases and ATPases); producing heterogeneous N-termini of proteins (Torizawa, Shimizu et al. 2004); difficulty to produce stable and reproducible cell extracts and to obtain suitable protein yields, have been or still being investigated (Carlson, Gan et al. 2012).

In the present context, i.e., NMR protein sample preparation, CF is particularly beneficial because of the possibility to reduce both isotope scrambling and dilution. In fact, during the cell extract preparation process, a certain inactivation of aminotransferases occurs (Kigawa, Muto et al. 1995, Torizawa, Shimizu et al. 2004) and the residual transaminase activities, particularly of those transaminases involved in Gly, Ala, Asx[b], Glx[c] residue metabolism, can be suppressed either by inhibitors or reducing agents (e.g. $NaBH_4$) (Morita, Shimizu et al. 2004, Tonelli, Singarapu et al. 2011). However, an excess of these molecules was reported to negatively affect the protein yields (Takeda and Kainosho 2012). Furthermore, isotope dilution caused by endogenous unlabeled amino acids contained in the extract was resolved by its dialysis before utilization (Torizawa, Shimizu et al. 2004). A combined gel filtration step can subsequently be added to optimize the removal of unlabeled amino acids (Takeda and Kainosho 2012). As a result, optimized isotope enrichment can be achieved using lower quantities of expensive RNase-free labeled amino acids.
Notably, CF protein synthesis can also be performed in ^2H_2O, enabling a higher deuteration level (Etezady-Esfarjani, Hiller et al. 2007). The buffer of the cell extracts is

[b] Asp and Asn
[c] Gln and Glu

exchanged for $^{2}H_2O$ by several filtration runs (MWCO 10 kDa) or by dialysis before its utilization for protein synthesis. This CF variant was investigated in our laboratory and in contrary to the reported results (Etezady-Esfarjani, Hiller et al. 2007), a significant reduction (~ 50%) in the protein yield was obtained, constituting a real drawback of this application.

The *E. coli* extract is the most widely used. However, because of the increasing necessity of enlarging NMR structural studies to eukaryotic proteins, cell-free extracts from eukaryotic cells, such as wheat germ (Madin, Sawasaki et al. 2000, Morita, Sawasaki et al. 2003, Vinarov, Lytle et al. 2004, Makino, Beebe et al. 2014), *Leischmanii* (Kovtun, Mureev et al. 2011), *Thermus thermophilus* (Uzawa, Hamasaki et al. 1993, Zhou, Asahara et al. 2012) and even human *HeLa* cells (Mikami, Kobayashi et al. 2008), are progressively being introduced (see table 1.1). The best-known and, consequently, the most optimized system is most likely the extract from the wheat germ system.

Both *E. coli* and wheat germ cell-free extracts are commercially available for isotopic labeling applications. Alternatively, homemade cell extracts can be used since protocols describing their laboratory preparation are also available in the literature (Kigawa, Yabuki et al. 2004, Torizawa, Shimizu et al. 2004). However, the difficulty in achieving a good reproducibility makes this option risky and expensive considering the time that may be spent for their optimization.

Another cell-free technology, which is called the PURE[d] system (Shimizu, Inoue et al. 2001), is gradually emerging as a powerful alternative. In this approach, the cell extract is substituted by a reaction medium composed exclusively of recombinant proteins. The potential of this technology lies in the complete enzyme control that can be exploited to improve the final synthesis in a protein-dependent manner. For instance, the non-addition of transaminases would make specific isotope labeling more efficient, whereas the use of given recombinants chaperones would assist in the correct folding of certain proteins (Shimizu, Kanamori et al. 2005, Whittaker 2013). This technology has great

[d] PURE: Protein synthesis Using Recombinant Elements expression systems

potential; however, it still requires a significant optimization to be affordable for large scale production of labeled proteins.

In comparision to the *in vivo* systems, CF approach offers the possibility to overcome several complications for the expression of labeled proteins, such as protein toxicity, control of isotope scrambling, addition of detergents for membrane proteins …etc.; but remains in most of cases more expensive.

Table 1.1: Characteristics of the most commonly used cell extracts for cell-free protein synthesis. (Carlson, Gan et al. 2012)

Type	Advantages	Disadvantages
E. coli extract	1. Simple and cost-effective preparation 2. High protein synthesis yield 3. High rate of protein synthesis 4. Clearly elucidated biochemical knowledge and well established tools for genetic modifications 5. Low-cost energy sources 6. Able to fold complex proteins	1. Limited post-translational modifications
Wheat-germ extract	1. Wide spectrum expression of eukaryotic proteins has been achieved repeatedly 2. Medium yield of complex proteins 3. Sophisticated high-throughput method for proteomics	1. Low yield of extract from cells 2. Extract preparation is lengthy and complex 3. Poor genetic modification tools 4. Commercially available but expensive
Rabbit reticulocyte lysate	1. Easy cell breakage and quick preparation 2. Eukaryotic-specific post-translational modifications	1. Complex manipulation of animal tissue required 2. Narrow spectrum of proteins expressed to date 3. High background of endogenous globin mRNAs and abundance of RNase M 4. Poor genetic modification tools 5. Low protein synthesis yields
Insect cell extract	1. Easy cell breakage and quick preparation 2. Eukaryotic-specific post-translational modifications 3. Signal sequence processing	1. Cell cultivation is expensive and time-consuming 2. Poor genetic modification tools

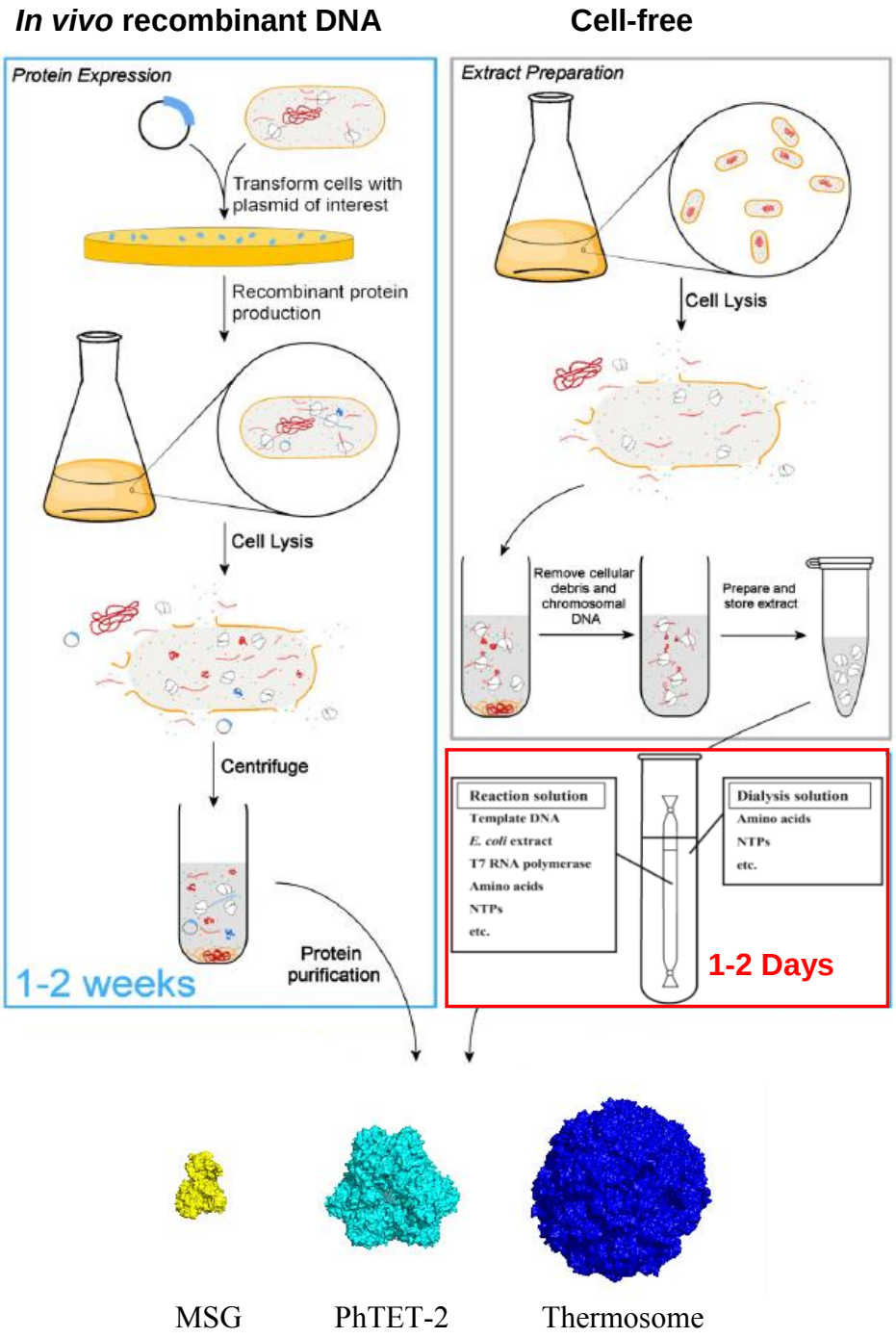


Figure 1.3: Comparison of in vivo recombinant DNA protein expression with cell-free protein synthesis. Cartoon representation of malate synthase G (MSG) (1p7t); tetrahedral aminopeptidase (PhTET-2) (1y0r) and Thermosome (1a6d). Adapted from (Carlson, Gan et al. 2012)

1.2.2 Optimal sites for selective protonation and isotopic labeling

Selective protonation in a deuterated background represents a unique approach for tackling large proteins by NMR. Here, an instinctive question arises regarding the choice and possibilities of atomic sites that can be protonated. In fact, depending on the required NMR analysis, the proton probe can be placed on either of the following location: (*i*) The backbone, such as the widely used NH position. This labeling is well adapted for monitoring systems whose molecular weight does not exceed c.a. 50-100 kDa limits. (*ii*) The side-chain, such as methyl group positions. This labeling can be applied to any size system but becomes indispensable in supra-molecular particles (> 100 kDa). (*iii*) Both backbone and side-chain. This labeling is powerful within a certain size range (< 100 kDa). For instance the selective protonation of both amide and methyl groups allowed the global fold calculation of a 82 kDa monomeric protein (Tugarinov, Choy et al. 2005).

In this section, I will discuss the most important strategies reported in the literature for selective protonation and labeling at specific site in proteins. The focus will be on the approaches using *E. coli* as the expression system.

1.2.2.1 Selective protonation and isotopic labeling on the backbone:

1.2.2.1.1 Selective protonation and labeling of amide groups

This labeling scheme is the most commonly used for large proteins (≤ 100 kDa) because of its broad NMR application and relative cost-effectiveness. According to the required NMR experiment or analysis, NH groups can be uniformly or specifically labeled. Consequently, the labeling protocol must be adapted for each case.

Uniform protonation and ^{15}N-labeling of amide groups:

Generally, uniform isotopic labeling of amide groups is achieved by adding a ^{15}N source (e.g., $^{15}NH_4Cl$) in the deuterated medium culture. However, their selective protonation is much more demanding, as discussed below:

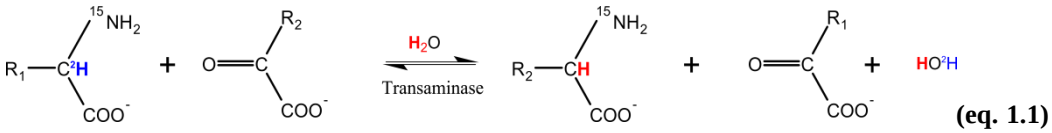
- *Uniform protonation of amide groups using in vivo system*

The re-protonation of amide groups can be easily achieved by placing the deuterated particle in water (protonated buffer). Its incubation in a basic buffer is favorable for an effective and fast amide $^{2}H/^{1}H$ exchange. This situation is true as long as amide groups are accessible to the solvent. Because this contact can be heavily restricted in large protein cores, other alternatives had to be developed.

Depending on the studied particle, two distinct approaches can be used:
The first, consists of unfolding the deuterated particle and subsequently refolding it in H_2O (Venters, Huang et al. 1995, Tugarinov, Choy et al. 2005). This strategy offers extremely selective protonation at the exchangeable protein sites (e.g., NH_2, NH, OH, SH). However, this treatment is not universal because it cannot be applied to all proteins. Even when refolding is possible, crucial protein loss is usually obtained during this process, making this approach not optimal, particularly when protein expression yield is originally limited.

The second alternative, relies on expressing the protein target using *U*-[^{2}H, ^{15}N]-algae hydrolysates (with or without *U*-^{13}C) dissolved in H_2O. Using this protocol, an overall deuteration extent of 80% was reported (Markus, Dayie et al. 1994). The simultaneous addition of a limited amount of protonated glucose (1 g/L), to ensure the energy production, seems to not affect the final deuterium incorporation (~85% using 3g/L CELTONE-dN-supplemented medium) (Fiaux, Bertelsen et al. 2004).

In contrast to the important deuteration of most side-chain residues, high protonation at C_α positions (and also moderately in certain C_β positions) was observed in a residue-type-dependent manner (see table 1.2)(Lohr, Katsemi et al. 2003). This extra-protonation is particularly harmful in the specific case of NMR triple resonances experiment, where the $^{13}C_\alpha$ nucleus relaxes rapidly due to its ^{1}H proximity (^{13}C-^{1}H dipole-dipole interaction) (Yamazaki, Weontae et al. 1994, Venters, Huang et al. 1995). This low C_α deuteration level of certain amino acids can be partially explained by the activity of transaminases that principally exchange $^{2}H_\alpha$ for ^{1}H from the medium according to the following equation:



(eq. 1.1)

Amino acid (1) **α-ketonic acid (2)** **Amino acid (2)** **α-ketonic acid (1)**

For some others, usually Gly, Cys, Trp, Glx[e] and Asx[f], this is primarily due to their initial lack or low amounts in the used hydrolysates (Lohr, Katsemi et al. 2003).

Amino acid type	Pro (14)	Arg (5)	His (4)	Lys (14)	Ile (11)	Thr (12)	Leu (13)	Tyr (7)	Val (17)
Deuteration level[a] (%)	98.2± 4.2	97.1± 2.4	96.4 ± 2.1	94.2 ± 3.9	88.3± 5.3	83.5 ±7.1	83.4 ±6.6	79.3± 7.0	73.6 ± 6.5
Amino acid type	Ala (12)	Ser (5)	Phe (13)	Met (4)	Glx[c] (30)	Cys (6)	Gly[d] (17)	Asx[b] (26)	Trp (4)
Deuteration level[a] (%)	73.1 ± 4.3	71.3±7.4	58.9 ±7.1	54.3 ±9.1	49.9 ±9.9	37.0 ±8.3	27.1±5.7	21.6±6.1	18.9 ±3.8

Table 1.2: Fractional Cα deuteration levels of DFPase from *Loligo vulgaris* expressed in *E. coli* using ^{2}H labeled algal lysate (Bio-Express 100 ADC) dissolved in $^{1}H_2O$. The number of residues contained in the protein is indicated between brackets. [a] Average percentage deuteration determined from intensity ratios of cross peaks in a pair of 2D HN(CO)(CA) spectra with and without $^{1}JC_{\alpha}H_{\alpha}$ dephasing. Values are not corrected for differential $^{13}C_{\alpha}$ T_2 times of protonated and deuterated carbons. [b] no systematic differences were found between Glu and Gln and between Asp and Asn. [c] for glycine the experiment measures the fraction of fully deuterated α-CH_2 groups. Table obtained from (Lohr, Katsemi et al. 2003).

Indeed, the currently used protocols (based on low pH treatment) for the algae or bacteria hydrolysis are drastic and lead to the complete degradation of the aforementioned amino acids. Therefore, these residues are rapidly depleted during bacterial culture or protein expression. Thus, these amino cids are endogenously synthesized within the protonated water from the protonated glucose and/or deuterated amino acids. Consequently, high and non-uniform proton incorporation (at C_{α} and at other aliphatic positions when protonated glucose is used) is obtained. The use of *U*-[^{2}H]-glucose, together with the supplementation of the missed amino acids (e.g., Gly, Cys, Trp, Glx and Asx) can be helpful but expensive. Nevertheless, this method is insufficient because aminotransferase activity remains present.

In *E. coli*, transaminases belong to the large pyridoxal-phosphate (PLP)-dependent enzymes family. Four major aminotransferases have been biochemically and genetically

[e] Gln and Glu

[f] Asn and Asp

identified: (1) aspartate aminotransferase (AspC), which is able to catalyze Asp, Phe, Trp and Tyr biosynthesis (Hayashi, Inoue et al. 1993); (2) aromatic amino acid aminotransferase (TyrB), which acts on the same substrates as AspC with the difference to be retro-inhibited by Tyr (Hayashi, Inoue et al. 1993); (3) branched-chain amino acid aminotransferase (IlvE), which is involved in Val, Leu and Ile production; and finally (4) alanine-valine transaminase (AvtA), which is primarily responsible for the Ala synthesis and also participates in the Leu and Ile formation (Yoneyama, Hori et al. 2011). Because of their substrate specificity overlaps, bacteria deficient of an ensemble of transaminases were proposed to reduce either the isotope scrambling or the protonation at undesired sites (Griffey, Redfield et al. 1985, Waugh 1996, Weigelt, van Dongen et al. 2002, Fiaux, Bertelsen et al. 2004). For instance, up to 90% of the deuteration level with complete amide protonation could be achieved using 1 g/L of protonated glucose and deuterated algae hydrolysate (CELTONE®-dN) in H_2O. This protocol was applied to overexpress *U*-[^{2}H],[^{2}H,^{15}N-Leu]-systems in the DL39 (DE3) or CT19 strains which are deficient in a certain set of transaminases (Fiaux, Bertelsen et al. 2004). However, similar deuteration level obtained from prototrophic (85%) *E. coli*, renders the use of auxotrophic strains less attractive. In fact, the use of bacteria with such extensive genetic lesions is not advisable because of their heavy effect on microorganism growth that is exacerbated by $^{2}H_2O$ toxicity.

Alternatively, the *in vivo* inhibition of PLP-dependent enzymes using chemical compounds, e.g, (aminooxy)-acetic acid, was also described (Lopukhov, Ponomareva et al. 2002). However, PLP-dependent enzymes ensure a broad spectrum of reactions in the cell (not only transamination) and the non-toxicity of this type of inhibition appears unlikely, especially when coupled to the $^{2}H_2O$ *E. coli* growth difficulty. Indeed, from our experience in the laboratory, such approaches seem to not be viable as no growth was observed, when inhibitors of PLP-dependent enzymes were applied to the *E. coli* culture at the used concentrations in cell-free protein synthesis.

Hence, undesired extra-protonation can most likely be decreased in an efficient manner if *U*-[^{2}H,^{15}N]-algae or bacterial hydrolysates are supplemented with their missing residues and applied together at high and inhibitory aminotransferases concentrations (see table 1.3). Assuming hypothetical efficiency (which must be verified and

quantified), the proposed solution is expensive, particularly because of the high cost of the individually added [^{2}H,^{15}N] amino acids. In conclusion, the investment in optimizing protocols for amino acids preparation (hydrolysis or other) from algae or bacterial cultures seems to represent the quickest way to obtain more satisfactory results.

- *Uniform protonation of amide groups using an in vitro system*

Despite not being a topic of this manuscript, it is worth mentioning that the recently proposed protocols for cell-free protein synthesis may cope with the undesired C_α protonation, reported above. These protocols rely on the previously discussed alternative, where inhibition of PLP-dependent enzymes is targeted. This inhibition is reached using *e.g.,* sodium borohydride (Su, Loh et al. 2011) or aminooxyacetic acid (Morita, Shimizu et al. 2004, Tonelli, Singarapu et al. 2011), together or not with glutamate synthase inhibitors (L-methionine sulfoximine) (Tonelli, Singarapu et al. 2011). Using these protocols, the protein of interest can be produced in H_2O, allowing protonation of the desired amide group with reduced extra protonation. Nonetheless, the use of these inhibitors at high concentrations to achieve effective inhibition has been associated with lowered protein yields (Takeda and Kainosho 2012).

Optimizing the PURE approach (see §1.2.1.2) is expected to be helpful for synthesizing perdeuterated proteins with a specific protonation. The absence of transaminases would allow specific amide group protonation without any ^{2}H/^{1}H exchange at C_α sites.

<u>Selective protonation and ^{15}N-labeling of a specific amide group</u>:

The specific protonation and ^{15}N labeling of amide groups is one of the most commonly used approach in biomolecular NMR. However, in large proteins the 2D [^{15}N,^{1}H] spectra are overcrowded. Therefore, the selective labeling of a given residue was proposed for their simplification.

Thus far, residue-specific protonation of amide groups is not possible, the above-discussed strategies, i.e., unfolding/folding or protein expression in H_2O, must be

applied even when only a given residue type is targeted. However, residue specific isotope labeling is realizable, allowing its possible filtering by NMR spectroscopy.

- *^{15}N -Labeling of amide group of a specific residue using in vivo system*s:

a) *By inhibition*

Residue-specific ^{15}N labeling, can be achieved by [^{15}N]-residue addition (Lee, Androphy et al. 1995). Unfortunately, due to the transaminase activity two major concerns are usually confronted (Muchmore, McIntosh et al. 1989). The first is isotope dilution, which translates into exchanging ^{15}N of the supplied residue against ^{14}N from other nitrogen-containing molecules produced by the bacteria. The second is isotope scrambling observed in certain residues (e.g., Glu, Gln, Asp, Ala, Gly, Ser, Thr…etc.) (Muchmore, McIntosh et al. 1989).

Amino acid	Inhibition (mg/L)[a]	Amino acid	Inhibition (mg/L)[a]
Ala	400 (800)[b]	Leu	100 (200)[b]
Arg	400	Lys	100
Asn	250	Met	250 (500)[b]
Asp	50	Phe	50
Cys	400 (800)[b]	Pro	*100*
Glu	400 (800)[b]	Ser	1600
Gln	400	Thr	100
Gly	250-375	Trp	50
His	100	Tyr	100
Ile	100 (200)[b]	Val	100

Table 1.3: Amino acid concentrations for selective labeling by inhibition.[a] from (Griffey, Redfield et al. 1985).[b] from (Ramesh, Frederick et al. 1994)

Amino acid biosynthesis in bacteria is regulated at the enzymatic activity and gene expression levels. Supplying the host bacteria with the desired [*U*-^{2}H, ^{15}N]-amino acid at transaminases inhibitory concentrations (Griffey, Redfield et al. 1985, Ramesh, Frederick et al. 1994) leads to the reduction of its endogenous synthesis, resolving the isotopic dilution problem. Additionally, the concomitant addition of the remaining [*U*-^{2}H]-amino acids, was reported to limit the isotope leak (no enrichment quantifications are available). As formerly discussed in §1.2.1.1, algae lysates are the more affordable source of [*U*-^{2}H]-amino acids. In the present case of residue-specific labeling, the lysate should be depleted of the intended residue to label. Otherwise, the ^{15}N-amino acid enrichment cannot be optimal due to competition with its unlabeled version contained in the used lysate. Thus, labeling based on the inhibition strategy can not be an affordable approach because deuterated lysates which are depleted of a given residue are not yet commercialized. Consequently, the deuterated amino acids must be individually mixed, rendering the total cost significant. Alternatively, higher quantities

of [^{15}N] amino acid can be applied to dilute the unlabeled content. Depending on its amount in the used lysate, the required over-supplementation of the [^{15}N] amino acid can easily be doubled leading to an unreasonable total cost too.
In conclusion, the development of new methodologies, such as chromatographic methods, that enable the removal of selective amino acids in these media will be useful for the development of this field.

b) *Using auxotrophic E.coli strains*

Auxotrophic bacteria strains can also be used as an alternative to the inhibition approach for coping with isotope dilution (Griffey, Redfield et al. 1985, Muchmore, McIntosh et al. 1989, Waugh 1996). Such genetic lesions impose the metabolization of the exogenously added amino acid because the cell is not designed to produce it. Nevertheless, the requirement for [*U*-^{2}H]-amino acids supplementation to mitigate the ^{15}N scrambling, together with the associated low growth and genetic design complexity, make this approach less attractive.

1.2.2.1.2 Selective protonation and isotopic labeling at the C_α position

Protocols aiming the obtaining of C_α protonation, through chemical (Yamazaki, Tochio et al. 1997, Upson and Hruby 1977) or enzymatic (Homer, Kim et al. 1993) treatments of amino acids mixtures (with or without ^{13}C enrichment) have been proposed.
Such labeling was applied to assign the backbone a small protein (~98 amino acids) at 10ºC, simulating a system of 30 kDa (Yamazaki, Tochio et al. 1997). The novelty of this idea relies on taking advantage of the long life $^{1}H_\alpha$-$^{13}C_\alpha$ multiple quantum coherences (double/zero quantum) in large proteins, whose relaxation properties become more favorable when $^{1}H_\alpha$ surrounding protons are exchanged against deuterium. For that approach, a certain set of triple resonance NMR experiments i.e., HACAN, HACACB, HACACO and HACA(N)CO, where the HSQC scheme in these pulses were replaced by HMQC, were optimized to provide high NMR signal sensitivity. This labeling scheme becomes particularly interesting when high pH solutions or temperatures are required. In such conditions, the inherently fast exchange of amide protons renders this method more efficient for backbone resonances detection than the $^{1}H_N$ based experiments. In this work, the protein sample was prepared in H_2O

using a *U*-[^{2}H, ^{13}C, ^{15}N], $^{1}H_{\alpha}$- amino acids mixture and then dissolved in $^{2}H_2O$ buffer for the amide proton exchange. Highly selective C_α protonation (80%) was achieved in only 11 amino acids. This strategy is extremely expensive because a high concentration of amino acids mixture (5 g/L) should be applied to compensate for the generated racemic fraction from the chemical treatment. Moreover, the lack of a significant part of residues, which is due to either the commercial hydrolysate mixture composition or degradation upon chemical treatment, makes this strategy less attractive.
Extrapolating this idea to larger proteins (≥ 100 kDa), this method is clearly not beneficial. In fact, for optimal C_α protonation these amino acids should be added in a H_2O culture medium, which is unfavorable for large proteins, otherwise transaminases will exchange the C_α proton against deuterium taken from $^{2}H_2O$ culture medium. Based on table 1.2, the C_α $^{1}H/^{2}H$ exchange process in a deuterated culture is expected to behave similarly to the reported C_α protonation in a H_2O medium. Thus far, no investigation concerning the extent of the remained C_α protonation in this context was reported. These data would help to evaluate the applicability of this approach, at least for monitoring some residues whose C_α atoms maintain a high protonation level in deuterated cultures (e.g., as a complementary tool for assignment).

In contrast, cell-free synthesis performed in $^{2}H_2O$ (Etezady-Esfarjani, Hiller et al. 2007) would permit the discussed protonation to be obtained. This open system (CF) offers the possibility to reduce transaminase activity via inhibitors addition and, consequently, the H_α exchange. Neglecting the high cost of the *U*-[^{2}H, $^{1}H_{\alpha}$] amino acid mixture and the inherently lowered protein yield (see §1.2.1.2), this approach could represent an alternative to the selective amide groups protonation.

1.2.2.2 Selective protonation and isotopic labeling on the side chain

Side-chain protonation and labeling of several residues (e.g., Pro, Arg, Met, Lys, Ile, Leu) in a specific manner is feasible by simply adding them in a *U*- [^{1}H, ^{13}C]-residue pattern to the medium culture. In fact, this labeling is much more controllable than amide groups because their lateral moieties are less frequently used for synthesizing other residues. Using the above-described approaches in 1.2.2.1.1. (*e.g.* inhibition, auxotrophic bacterial strains etc.), selective side-chain labeling of the remaining

"recyclable" amino acids can also be achieved (Brodin, Drakenberg et al. 1989). However, the juxtaposed presence of several protons in the side-chain leads to a rapid transverse relaxation rate (T_2), which translates into low sensitivity of the NMR signal in general. This effect is even more pronounced in large proteins (large τ_c). Additionally, this protonation leads to intra-residue spin diffusion in nOes-based experiments, complicating the quantification of inter-nuclear distances (Gardner and Kay 1998). Therefore, labeling of a specific atomic position in the side chain has subsequently been explored.

1.2.2.2.1 Stereo-array Isotope Labeling (SAIL)

This innovative labeling was proposed to enlarge the size range of proteins whose 3D structure can be determined by solution NMR spectroscopy. Indeed, although proton density is significantly decreased by more than 50% (see table 1.4) (Kainosho, Torizawa et al. 2006, Ikeya, Takeda et al. 2009), the site specific information required for the structure calculation (for proteins of ~50 KDa) is maintained measurable. Briefly, this labeling consists of uniformly replacing in a stereo-selective fashion one ^{1}H by ^{2}H in methylene groups and two ^{1}H by ^{2}H in each methyl group. For prochiral methyl groups of Leu and Val, one specific methyl is kept "silent" in NMR as it is labeled as $^{12}C(^{2}H)_3$. And finally, aromatic rings are labeled with alternative ^{12}C–^{2}H and ^{13}C–^{1}H moieties (Kainosho, Torizawa et al. 2006) (see fig. 1.4). The obtained overall pattern renders this approach well adapted to NMR applications.

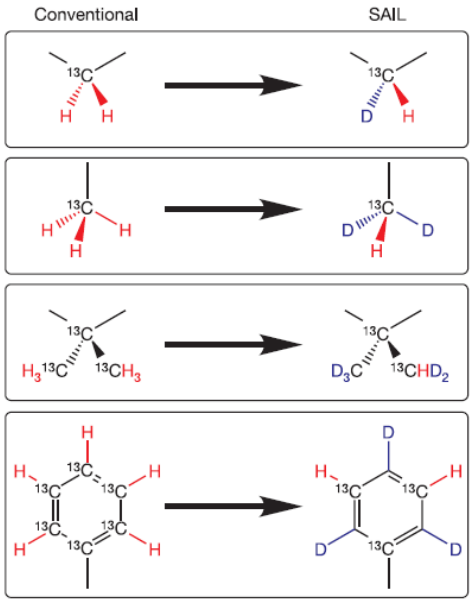


Figure 1.4: SAIL amino acids labeling. From (Kainosho, Torizawa et al. 2006)

Through this labeling more than "two birds are killed with one stone": (*i*) Backbone and side-chain assignment, in a stereospecific manner, is possible because through-bond connectivity is conserved and protons are stereo-specifically labeled. (*ii*) Distances obtained through nOes are more accurate because spin diffusion is significantly reduced. (*iii*) The $^{2}J_{H\text{-}H}$ couplings are removed and $^{3}J_{H\text{-}H}$ couplings are significantly

reduced. (*iv*) Both the sensitivity and resolution of NMR spectra are enhanced because the proton number is diminished by more than half (see Table 1.4), giving a high quality spectra.(*v*) The automation of assignment and structure calculations is finally realizable because NMR signal overlap is alleviated and stereo-specific assignment of methylene and methyl protons is readily available (Ikeya, Terauchi et al. 2006, Ikeya, Takeda et al. 2009).

Such amino acids are extremely useful but expensive. Therefore, cell-free expression system was adopted as part of the SAIL approach. Through this expression method, isotope dilution (Torizawa, Shimizu et al. 2004) and scrambling can be concomitantly reduced (see §1.2.1.2.), allowing the decrease in the required isotope quantity. Nevertheless, the currently practiced prices (10 K€/sample) are not accessible for the majority of laboratories.

Quantity	CaM		MBP	
	Uniform Labeling	**SAIL**	**Uniform labelling**	**SAIL**
^{1}H atoms per molecule in $^{1}H_2O$	1,095	697 (64%)	2,860	1,802 (63%)
^{1}H atoms per molecule in $^{2}H_2O^{*}$	851	453 (53%)	2,249	1,191 (53%)
Non-exchangeable side-chain ^{1}H atoms	692	305 (44%)	1,850	821 (44%)
NOE cross-peaks expected Ⱨ	9,812	5,642 (58%)	17,076	9,382 (55%)
NOE cross-peaks used	-	4,576 (47%)	-	7,485 (44%)
NOE distance restraints expected ⫲	2,883	2,720 (94%)	4,293	4,347 (101%)
NOE distance restraints used	-	2,422 (84%)	-	3,818 (89%)

* The number was calculated by assuming that all exchangeable ^{1}H nuclei are replaced by ^{2}H.
Ⱨ NOE cross-peaks are assumed to be observable between all aliphatic, aromatic, backbone amide and Asn/Gln side-chain amide ^{1}H nuclei that are less than 4.75 Â apart in the 20 conformers of the SAIL-CaM solution structure, or less than 4.1 Â apart in the 20 conformers of the SAIL-MBP solution structure.
⫲ Excluding duplicate restraints from symmetry-related peak pairs and NOEs for fixed distances. For uniform labeling, pseud-atoms were assumed.

Table 4: Expected and observed features of CaM and BMP samples prepared by conventional uniform labeling and SAIL methods. From (Kainosho, Torizawa et al. 2006)

In contrast, comparing the associated costs of the more affordable but time-consuming approaches, e.g., salaries, consumables for long periods of wet laboratory work and data analysis ranging from several months to few years, leads to conclude that SAIL approach effectiveness may compensate for its high price. Indeed, the significant enhancement of the obtained SAIL NMR spectra allows their faster analysis.

From the information that has been provided so far, this approach is certainly innovative and appropriate for proteins ~up to 40 kDa but remains inapplicable to the targeted objects in this manuscript. In fact, the obtained proton reduction does not fall into the protonation limit ranges of supra-molecular proteins. Nonetheless, SAIL amino acids can be used individually. Thus, instead of labeling all amino acids, a few of them can be exploited as sole probes in the deuterated protein.

In fact, this utilization has already been reported in the literature where the labeled proteins were produced in either cell-free (Takeda, Ono et al. 2010) or bacterial expression systems (Miyanoiri, Takeda et al. 2011, Takeda, Terauchi et al. 2012). In these publications, several derivative of SAIL amino acid labeling patterns were efficiently exploited for diverse useful applications in small proteins (10-12 kDa), such as the assignment of the ring signals of aromatic residues, the determination of their χ_2 conformations and the acquisition of high quality nOe restraints (Takeda, Ono et al. 2010, Miyanoiri, Takeda et al. 2011). Recently, an interesting conformational study using nOes measurements of the β-proton pairs across individual Cys disulfide bonds was reported (Takeda, Terauchi et al. 2012). Thus far, no application has been reported in large proteins, most likely due to the already mentioned cost limitation. Of note, derivative of SAIL Leu and Val amino acids were newly reported for MSG (82 kDa) labeling (Miyanoiri, Takeda et al. 2013). This application will be extensively discussed in §1.3.2.1.2.

1.2.2.2.2 Specific protonation position in aromatic amino acids

Aromatic residues occur frequently in the hydrophobic cores of globular proteins, which represent up to 15% of their amino acid compositions (Miller, Janin et al. 1987). Therefore, nOes related to aromatic rings are extremely important distance constraints to determine high-resolution protein structures. These aromatic residues play a decisive role in protein-protein or protein-ligand interactions. Additionally, due to its largest indole ring, Trp is considered the most influent residue that significantly affects the spatial packing and dynamics of the surrounding amino acids in the protein hydrophobic core (Miyanoiri, Takeda et al. 2011).

Protocols using different carbon sources for the atomic protonation and labeling of aromatic residues are available in the literature (Venters, Calderone et al. 1991, Teilum, Brath et al. 2006). These protocols are not here discussed because they result in an extensive ^{13}C labeling (affect other residues) and partial enrichment.

Thus initially, the protonated ε-^{13}C-Phe was first used for its assignment in a deuterated system of 21 kDa (Wang, Janowick et al. 1999). Further publications suggested protonated shikimic acid (Rajesh, Nietlispach et al. 2003), phenylpyruvate and 4-hydroxyphenylpyruvate (Raap J, Niewenhuis S et al. 1999) as candidates for selectively protonating and labeling the side-chains of Phenylalanine (Phe), Tyrosine (Tyr) and Tryptophan (Trp) residues in perdeuterated contexts (see fig 1.6). However, these approaches are not appropriate for large proteins because all ring carbons are protonated (low signal dispersion and strong dipolar interactions).

Interesting alternatives that offer similar SAIL patterns with more affordable prices have recently been proposed (Kasinath, Valentine et al. 2013, Lichtenecker, Weinhaupl et al. 2013). In the first study, the biosynthesis pathway of aromatic residues (Phe, Tyr, Trp) where erythrose 4-phosphate is condensated with phosphoenolpyruvate was exploited (Kasinath, Valentine et al. 2013). The [4-^{13}C]-erythrose was used as a sole carbon source in conjunction with deuterated ^{12}C-pyruvate (to decrease ^{13}C scrambling) in the $^{2}H_2O$ *E. coli* culture (see fig. 1.6). As a result, a single delta carbon position was ^{13}C enriched at 67% in both Phe and Tyr. All protons of the added erythrose were deuterated in the culture except for Phe-$H_{\delta,\zeta}$, Tyr-H_{δ} and Trp-H_{δ}. Thanks to this labeling (isolated ^{1}H-^{13}C pairs in the aromatic ring) analysis of ^{13}C-relaxation aromatic groups of deuterated calcium-saturated calmodulin was performed. In the second protocol, Phe-specific protonation and labeling at ε_1 and ε_2 positions was proposed (Lichtenecker, Weinhaupl et al. 2013). This protocol relies on sodium [3,3-$^{2}H_2$] ([3,5-$^{13}C_2$; 2,4,6-$^{2}H_3$]phenyl) pyruvate which is prepared from the cost-effective isotope sources [1,3-$^{13}C_2$] acetone and $^{2}H_2O$. The detailed protocol is not available yet; however, the authors announced a future article containing more precise synthesis data with a protein application.

The selective protonation at aromatic groups is helpful but its applicability remains restricted by protein size.

1.2.2.2.3 Methyl groups labeling

Methyl groups represent the probe of choice for the NMR study of large proteins. This atomic position is the "core" of my thesis. Therefore, a special subchapter will be dedicated for describing this selective labeling technology (see §1.3).

1.3 Selective protonation and isotopic labeling of methyl groups in perdeuterated biological systems

Despite their small number, methyl-containing proteinogenic[g] amino acids have been reported to generally cover up to 30% of the protein systems (from statistical study performed on 37 monomeric proteins ranging from 8 to 35 kDa), and to represent up to 50% of their hydrophobic cores (Miller, Janin et al. 1987). A further study demonstrated that these amino acids occupy up to 24% of the protein-protein interface area (Lo Conte, Chothia et al. 1999). In larger and oligomeric proteins, the content of methyl-containing amino acids is expected to be higher because the ratio of solvent inaccessible-to-accessible areas is more important. Indeed, the simple analysis of a few large proteins studied in our laboratory, e.g., Thermosome (hexadecamer of 1 MDa), PhTET-2 (dodecamer of 470 kDa), ClpP (tetradecamer of 308 kDa) and MSG (monomer 82 kDa), demonstrates that methyl-containing residues represent more than 40% of the total amino acids composition in these systems (see fig. 1.1).

Certainly, such protein coverage is not ideal in the sense that an important part of the analyzed object is not monitored. Nevertheless, the literature has proven that this specific labeling strategy can be sufficient to address critical problems. In §1.3, a minimal biochemical background considered required for understanding the currently used labeling strategies of each methyl-containing residue (six in total) is described.

1.3.1 Why methyl groups?

Methyl groups are the best candidates for selective protonation (see fig. 1.5). This argument can be justified as follows:

[g] amino acids that are incorporated into proteins without undergoing any posttranslational modifications.

1. Methyl groups enrich protein hydrophobic cores and protein-protein interfaces. NOEs between their protons provide valuable long-range distance restraints that are indispensable for the protein folding calculation. An analysis of over 290 non-homologous proteins was reported (Gardner and Kay 1997), indicating that unlike NH-NH and $^{13}CH_3$-NH nOes that involve residues with a respective median separation of two and three amino acids in the primary sequence, $^{13}CH_3$-$^{13}CH_3$ nOes involve residues whose median separation is approximately 30 amino acids (Gardner and Kay 1997).
2. Methyl groups allow the acquisition of highly sensitive and resolved NMR signals. The proton multiplicity of methyl groups results in a resonance three times more intense than a single proton. Furthermore, due to rapid rotation about their three fold symmetric axis and due to their flexibility (when located at the extremities or amino acid long side-chain peripheries), methyl groups show a more favorable transverse and longitudinal relaxation than a single backbone or side-chain proton. The quantification of the signal-to-noise ratios (S/N) of ^{15}NH and $^{13}CH_3$ HSQC spectra acquired on the same [^2H,^{15}N,$^{13}CH_3$]-protein sample (two systems of 42 and 110 kDa were analyzed) expressed in 100% ^2H_2O with [*U*-^2H]-glucose, demonstrated a gain of a factor of 4-7 in the $^{13}CH_3$ 2D NMR spectrum acquired on a 500 MHz spectrometer (Hajduk, Augeri et al. 2000). This enhancement can be even more significant using the HMQC experiment. Indeed, a sensitivity gain of 2.6-fold was obtained at 5 C^o for HMQC data set relative to its HSQC correspondent in a *U*-[^2H,^{15}N]- MSG sample (τ_c=118 ns), where Ile-δ1, Leu and Val methyl groups are labeled (Tugarinov and Kay 2004). Considering the mentioned results, the S/N in a $^{13}CH_3$ HMQC spectrum is expected to be 10-18 higher than ^{15}NH. In a further study, the [^1H,^{15}N]-HSQC correlation spectra of a protein of 204 kDa (SecA) was reported to contain only ~10% of the expected signals whereas the [^1H,^{13}C]-HMQC spectra showed the complete set of the expected methyl groups (220 probes)(Gelis, Bonvin et al. 2007).
3. Methyl groups resonate in an uncrowded region of the 2D [^1H,^{13}C] spectrum. Therefore, a better spectral resolution is expected compared with the remaining atom positions.

Isotope scrambling in methyl group labeling strategy

The selective protonation and isotopic labeling of the methyl groups in a perdeuterated protein relies on adding the labeled precursor/amino acid to the *E. coli* culture before inducing protein expression. As repeatedly noted, the metabolic pathway through which this added compound can be processed is usually not isolated. Indeed, *in vivo* biochemical reactions cross-talk via their catalytic enzymes (see fig. 1.6). Therefore, the risk that undesired labeling appears in the NMR spectrum is highly probable. This undesired

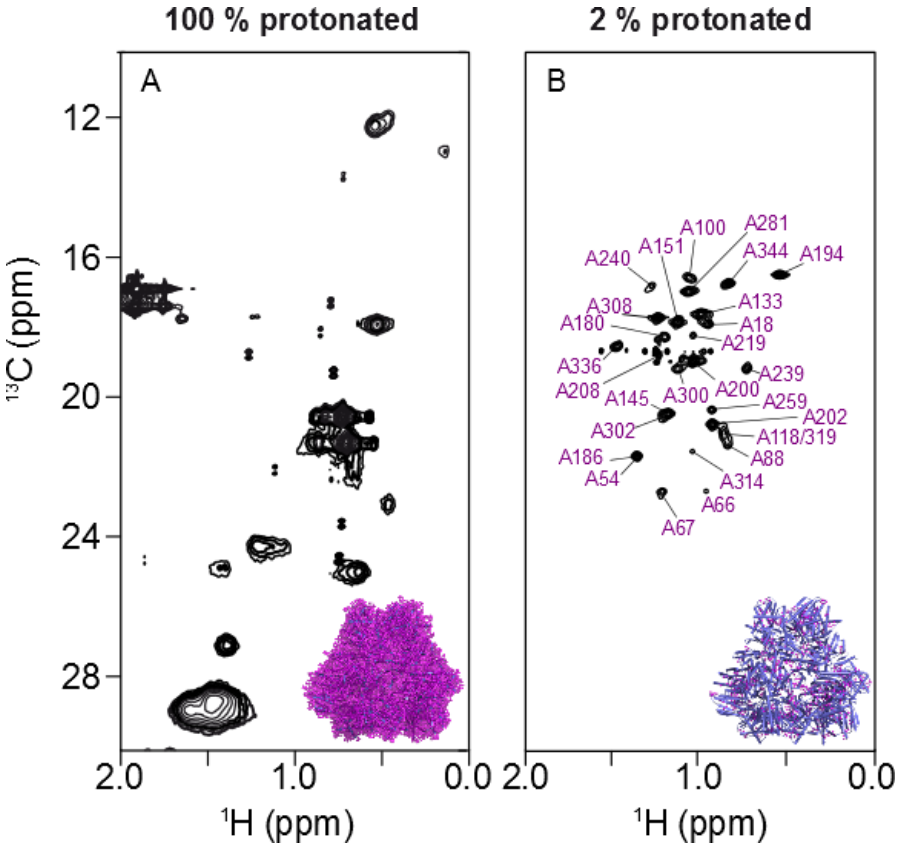


Figure 1.5: NMR signal enhancement upon selective protonation of Alanine methyl group positions in a perdeuterated PhTET-2 protein (468kDa). Protein was produced with A) [U-^{1}H,^{13}C,^{15}N] and B) [U-^{2}H,^{15}N],[$^{13}CH_3$]-Ala-β labeling. 2D [^{1}H,^{13}C] NMR spectra were recorded using 0.5 mM samples (monomer concentration) and a 800 MHz spectrometer equipped with a cryogenically cooled probehead. The 3D structures of PhTET (PDB 1Y0R) specifically or uniformly protonated are represented in their respective spectra. Balls depicted in purple represent ^{1}H atoms.

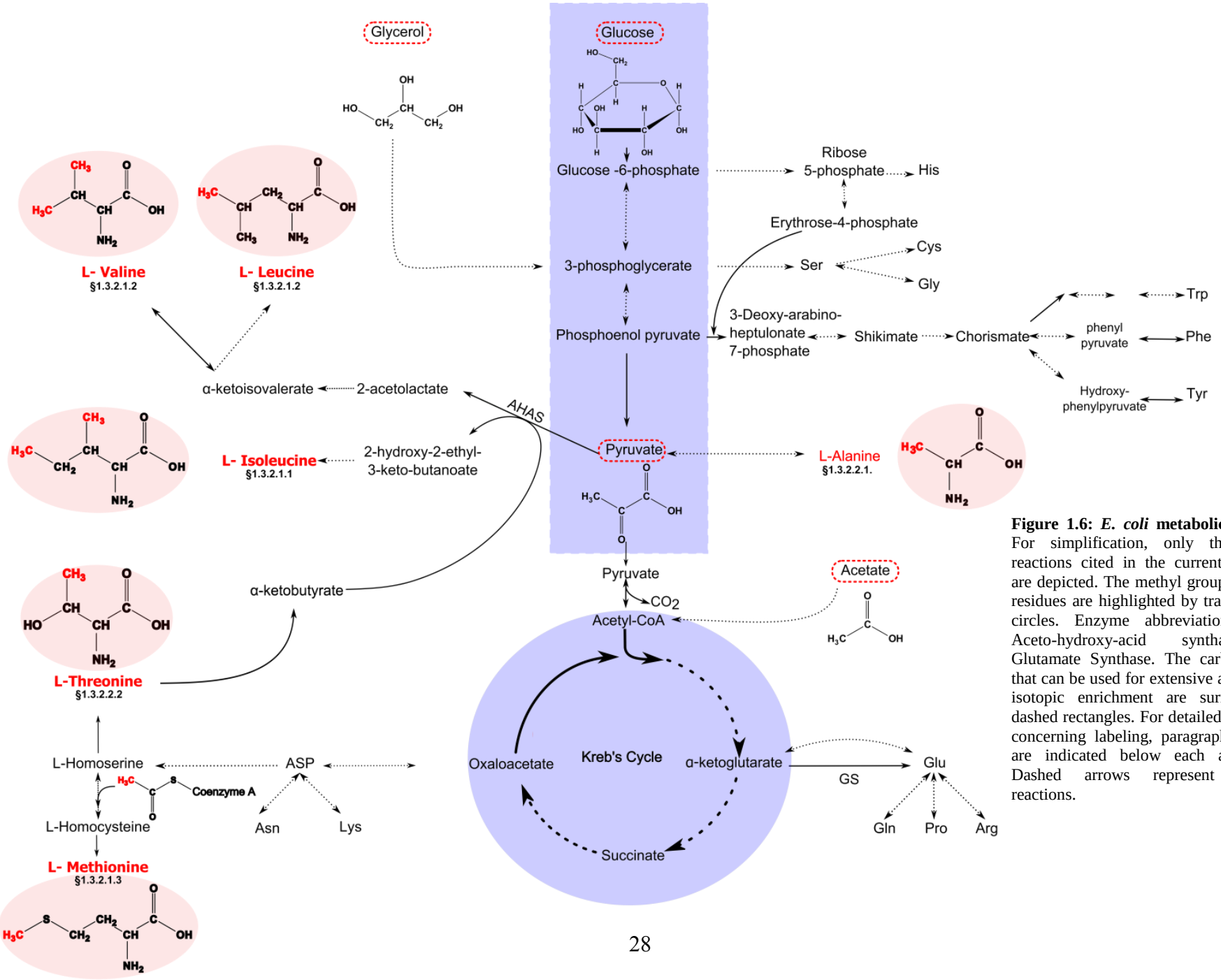


Figure 1.6: *E. coli* metabolic pathways. For simplification, only the principal reactions cited in the current manuscript are depicted. The methyl group–containing residues are highlighted by transparent red circles. Enzyme abbreviations: AHAS: Aceto-hydroxy-acid synthase. GS: Glutamate Synthase. The carbon sources that can be used for extensive and selective isotopic enrichment are surrounded by dashed rectangles. For detailed information concerning labeling, paragraph references are indicated below each amino acid. Dashed arrows represent multi-step reactions.

scrambling can affect the final data by obtaining misleading NMR signals or low isotope enrichment of the desired residue. Therefore, thorough knowledge of *E .coli* metabolism is required to be able to efficiently "manipulate" this system. Although the discussed ^{15}N scrambling is difficult to control, the isotope leak from side chain is luckily more restricted and well defined. The primary enzymes involved in isotope scrambling for each methyl group labeling strategy will be detailed in their respective categories.

1.3.2 Labeling strategy for each methyl-containing amino acid

According to their labeling strategies, the methyl-containing amino acids can be classified into two principle categories. In the first, which is the easiest to handle, residues are located at the end of metabolic pathways and result from irreversible reactions. When added to the culture, [$^{13}CH_3$]-labeled residues are readily incorporated into the expressed protein without any scrambling, as observed for methionine. However, the high cost of certain labeled amino acids and even their non-availability in some cases, led to search for a cost-effective approach. The best alternative in this situation consists of adding the labeled precursor (Gardner and Kay 1997) instead of the amino acid itself, whose synthesis is more affordable. Ile, Leu and Val amino acids are concerned by this strategy. Notably, despite this apparent simplicity other types of complexity might appear when these amino acids are intended to be simultaneously labeled. This complexity is observed particularly when precursors share the same enzymatic machinery (e.g., preferential enzymatic processing of a precursor at the cost of another) as will be shown in §2.2.

When amino acids are involved in reversible or intermediate metabolic pathways, then the conditions are even more complex. Indeed, whatever the used compound, amino acid or its precursor, the introduced labeling, will definitively appear elsewhere. For such residues, typically Ala and Thr, adding the labeled amino acid is used. Synthesis protocols were developed to make these molecules as affordable as possible, and solutions were set up to limit or to suppress the undesired isotope scrambling.

1.3.2.1 Amino acids involved in irreversible metabolic pathways

1.3.2.1.1 Isoleucine residue methyl groups labeling

Isoleucine occurs naturally in proteins at an abundance of 4.9% (Miller, Janin et al. 1987) and contains two methyl groups at delta 1 (δ_1) and gamma 2 (γ_2) positions. In *E.coli*, this residue is synthesized by condensating one molecule of pyruvate with another molecule of α-ketobutyrate via aceto-hydroxy-acid synthase (AHAS) activity. Figure 1.8 illustrates the molecular origin of each carbon of the Ile methyl group. The biosynthetic pathway of Ile was one of the first to be exploited by methyl-specific labeling protocols (Gardner and Kay 1997).

Labeling of the delta 1 methyl group position of isoleucine

Ile-δ_1 methyl group represents a probe of choice for NMR analysis due to its flexible side-chain that shows favorable longitudinal and transverse relaxation properties. Moreover, Ile-δ_1 correlations are well dispersed and usually present a high quality NMR spectrum. From the previously referred analysis (see §1.3.1), this probe presents an interesting number of nOe contacts since each Ile-δ_1 methyl group was determined to be, in average, within 6 Å distance from ~5.1 ± 1.9 backbone NH and 2.8 ± 1.5 Ile-δ_1, Val-γ, and Leu-δ and Met-ε methyl groups (Gardner and Kay 1997).

In the late 1990's, [3,3-$^2H_2$, *U*-^{13}C] α-ketobutyrate (60 mg/L) was shown to be an interesting metabolic precursor for the specific labeling of Ile δ_1-methyl group without measurable isotopic scrambling. Initially, this compound was prepared from [*U*-^{13}C] threonine (Thr or T) using the threonine deaminase (TD) enzyme (see §1.3.3.2.2). Thus, threonine was converted into the [3-$^2H_2$, *U*-^{13}C] α-ketobutyrate, whose protons were subsequently exchanged for deuterium by simple incubation in ^2H_2O at basic pH (Gardner and Kay 1997). Later, several chemical synthesis protocols were proposed as alternatives to the expensive commercially labeled α-ketobutyrate at that time (Hajduk, Augeri et al. 2000, Gross, Gelev et al. 2003, Lichtenecker, Ludwiczek et al. 2004). Currently, the relatively accessible prices practiced by the isotope suppliers do not justify the application of such chemical synthesis whose realization would be complicated for a routine biochemical laboratory.
Nowadays, α-ketobutyrate exists in diverse labeling patterns, such as [3,3-$^2H_2$, *U*-^{13}C],

[3,3-$^{2}H_2$, 4-^{13}C], [4,4,3,3-$^{2}H_4$, 4-^{13}C] or even [4,3,3-$^{2}H_3$, 4-^{13}C], allowing an optimized Ile labeling scheme for different types of NMR experiments. Nevertheless, the employment of NMR pulses that correlate δ1-methyl group resonance with the nuclei backbone requires the use of *U*-[^{13}C,^{2}H] glucose. *In vivo,* Ile-C_β is provided by pyruvate which results from glucose metabolism (see fig. 1.8). In parallel, the use of *U*-[^{13}C,^{2}H] glucose will necessarily label simultaneously Ile-γ_2 methyl group, preventing the possibility of solely linking the δ1 correlation to the backbone with a linear ^{13}C spin system. Such bifurcation in ^{13}C labeled side-chain has been shown to decrease the efficiency of ^{13}C-^{13}C magnetization transfer in a ^{13}C-TOCSY experiment, which is widely used for the residue side-chain assignments (see §3.1) (Tugarinov and Kay 2003).

The obtained or purchased α-ketobutyrate is directly added to the minimal *E. coli* expression medium. The compound is taken up by the bacteria and undergoes a sequence of biochemical reactions to be converted into Ile. In the next point, I will present the most important enzymes involved in its *in vivo* processing.

- *In vivo* α-ketobutyrate enzymatic processing

In vivo, five steps are required to convert the α-ketobutyrate into Ile. Both the first and third reactions deserve the most details (see fig. 1.8) for involving enzymes that show an interesting behavior toward Ile precursors, namely aceto-hydroxy-acid synthase (AHAS) and Ketol-acid reductoisomerase (KARI).

a) *Aceto-hydroxy-acid synthase*

Three AHASs are present in *E. coli* (Umbarger 1996, Bar-Ilan, Balan et al. 2001) (EC 2.2.1.6: also called acetolactate synthase). These synthases catalyze the first common biosynthetic step of the branched-chain amino acids Ile, Leu and Val (see figures 1.7 and 1.8). Those enzymes ensure the condensation of pyruvate (first substrate) with either α-ketobutyrate or another molecule of pyruvate, producing 2-aceto-2-hydroxybutanoate (Ile precursor) and acetolactate (Leu and Val precursor), respectively. Except for the AHAS II, the other enzymes are mainly retro-inhibited by Val (Chipman, Barak et al. 1998). The activity of AHAS II will be further exploited for the *in vitro*

synthesis of Ile precursor (see §2.2); therefore, some mechanistic details are here provided.

AHAS enzymes require thiamin diphosphate (ThDp), divalent metal ions and flavin adenine dinucleotide (FAD) as cofactors for their activity. For a long time, the role of FAD in the AHAS mechanism was unknown because no oxidation or reduction occurs during the catalysis. Later, this characteristic was found to represent an evolutionary relic of the ancestry of its sub-family (Duggleby 2006).

Based on colorimetric assays, the high affinities (50-60 fold) of AHAS II and III to α-ketobutyrate than pyruvate as a second substrate (acceptor) have been rapidly demonstrated (Chipman, Barak et al. 1998). This preference is perfectly justifiable by the fact that the α-ketobutyrate cell content is 2 orders of magnitude lower than that of pyruvate (Duggleby, McCourt et al. 2008, Steinmetz, Vyazmensky et al. 2010). Thus, biosynthesis reaction involving α-ketobutyrate are more "protected". The existence of three isoenzymes was explained by the fact that AHAS II and III satisfy the requirements for Ile, Leu and Val synthesis during growth on glucose. In contrast, when the pyruvate concentration in cell is low (e.g., growth on acetate), then this function is ensured by AHAS I, which is not specific to either of its natural substrates (Barak, Chipman et al. 1987).

No AHAS catalytic subunit (CSU) structure of *E. coli* is available. Nevertheless, structures of homologous CSU (yeast, *Arabidopsis thaliana*) helped to have an architectural view of the AHAS CSU from *E. coli*. Reported investigations have shown that Trp 464, which is located at the acceptor substrate pocket, is responsible for AHAS II preferential binding to α-ketobutyrate as an acceptor (Chipman, Barak et al. 1998). Indeed, the hydrophobic interaction of its ethyl moiety with the indol ring of this Trp is stronger than that of the methyl moiety of pyruvate. Furthermore, the high affinity of the enzyme to pyruvate as a donor was explained by the structural roles of both Phe109 and Val375 residues, which are situated at the donor substrate pocket. Indeed, Val 375 seems to cause a steric hindrance that favor pyruvate binding and the mutation of either

Val 375 or Phe 109 leads to a disruption in the proper orientation of the donor substrate and, consequently, the entire catalysis reactions (Steinmetz, Vyazmensky et al. 2010).

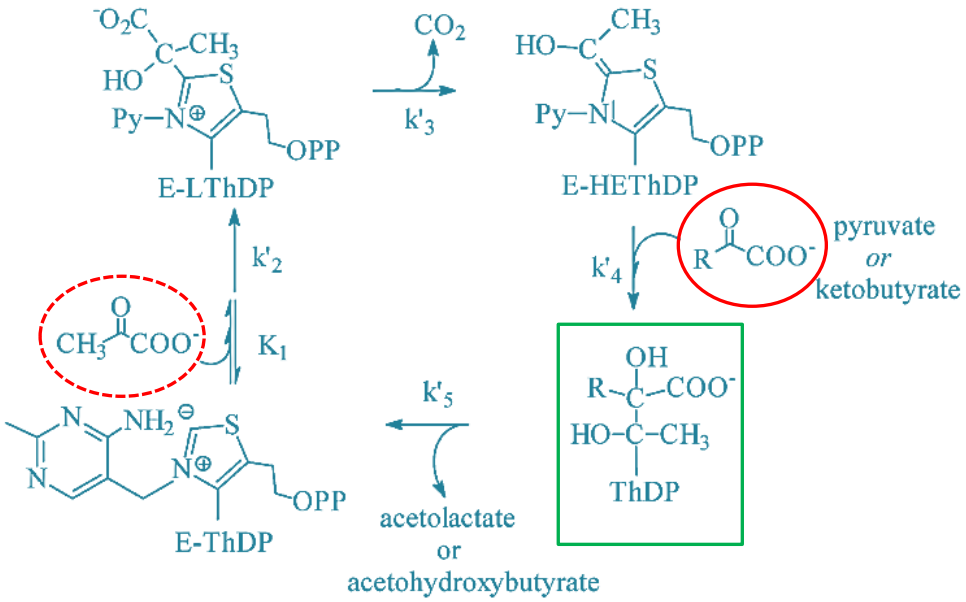


Figure 1.7: : Scheme of the aceto-hydroxy acid synthase mechanism. Intermediate and rate constants are identified. The complex Enzyme-Thiamine diphosphate is indicated as E-ThDP. E-LThDP represents 2-hydroxyethyl-ThDP intermediate and E-HEThDP represents the 2-hydroxyethyl-ThDP intermediate. First substrate is surrounded by red dashed circle. Second substrate is surrounded by red circle. The covalent product-ThDP adduct (in green rectangle) can be either Actetolactate-Thiamine diphosphate (ALThDP) when pyruvate is the acceptor or Aceto-hydroxy butanoate-Thiamine diphosphate (AHBThDP) when 2-ketobutyrate is the acceptor. Adapted from (Steinmetz, Vyazmensky et al. 2010)

b) *Ketol-acid reductoisomerase*

Ketol-acid reductoisomerase (EC. 1.1.1.86), which is known as KARI, is a tetramer involved in branched-chain amino acids biosynthesis. The KARI enzyme catalyzes an unusual two-step reaction where the first step consists of a stereospecific alkyl migration in either 2-acetolactate or 2-aceto-2-hydroxybutanoate followed by a reduction as depicted in figure 1.9.

Despite being unable to isolate intermediates, the entire process is assumed to be performed in two different stages. This enzyme presents another singular characteristic of an absolute requirement for Mg^{+2}. Several studies reported that KARI processes more

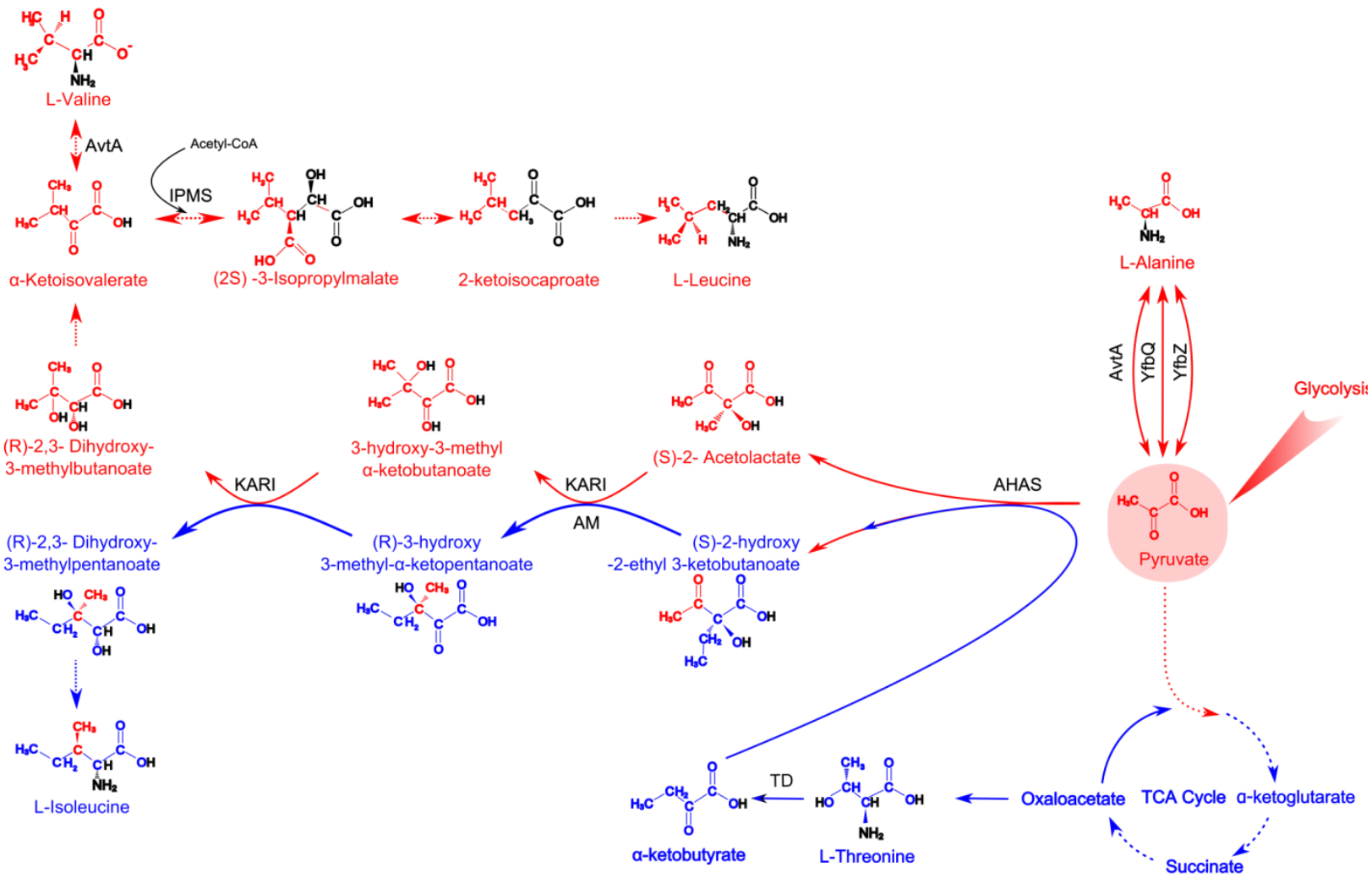
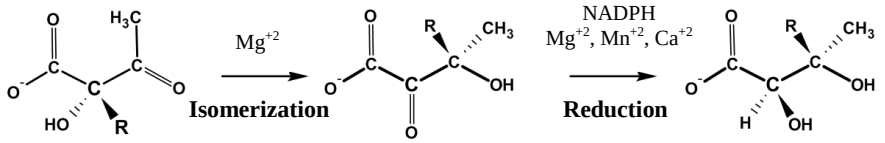


Figure 1.8: Biosynthesis of Ala, Ile, Leu and Val in *E. coli*. The amino acids highlighted in red are directly generated from pyruvate. The amino acids in blue are generated from the TCA intermediate. Moieties depicted in black are from acetyl-CoA. Amine groups and protons from the solvent are also depicted in black. Enzyme abbreviation: AHAS: aceto-hydroxy-acid synthase; AM: 2-acetolactate mutase; AvtA: alanine-valine transaminase; IPMS: 2-isopropylmalate synthase; KARI: ketol-acid reductoisomerase; TD: threonine deaminase; YfbQ and YfbZ: alanine transaminases**.** Dashed arrows represent multi-step reactions.



(1): R= CH_3, (2S)-acetolactate or R =CH_2-CH_3, (2S)-2-aceto-2-hydroxybutyrate
(2): R= CH_3, 2,3-Dihydroxy-3-methylbutanoate or R= CH_2-CH_3, 2,3-Dihydroxy-3-methylpentanoate

Figure 1.9: Scheme of Ketol-acid reductoisomerase reaction catalysis.

efficiently the Ile precursors than the acetolactate compound. In fact, despite the similar K_m values for both these molecules, KARI presents a 5- to 8-fold higher activity with 2-aceto 2-hydroxybutanoate (Dumas, Biou et al. 2001). This "natural behavior" helps balance the metabolites biosynthesis equilibrium, where the low α-ketobutyrate cell content is compensated by enzyme kinetics. Recently, the *E. coli* KARI structure was determined by X-ray crystallography (PDB 3E9Y) (Wang, Lee et al. 2009) but no structural mechanistic information that can explain its higher affinity for the Ile precursors has been published.

Labeling the gamma 2 methyl group position of isoleucine

Isoleucine γ_2 methyl groups are interesting probes from the NMR signal quality point of view; however, depending on the required NMR analysis their use should be well pondered. For instance, these methyl groups present a relatively good dispersion in 2D [$^{13}C^1H_3$] NMR spectrum but at the meantime they appear at a crowded region overlapping with Ala and Val residues. Therefore, their combined labeling with other methyl-groups (e.g., Ala and Val) is not advisable (see §2.1). Furthermore, the reported comparison of the proton longitudinal relaxation time (T_1) of δ_1 and γ_2 methyl groups in perdeuterated proteasome samples (α subunits ring of 360 kDa) revealed a 2-fold difference (^1H T_1 Ile-δ_1 > Ile-γ_2) (Ruschak, Velyvis et al. 2010). Taking advantage of this data, the signal-to-noise ratio of Ile-γ_2 methyl groups from an optimized SOFAST-HMQC experiment, where recycle times were adjusted for each sample, was estimated to be higher than the Ile-δ_1 by 30%. However, due to the same property (i.e., rapid R_1), the use Ile-γ_2 as probes for long-range nOes detection is not optimal.

Furthermore, the Ile-γ_2 methyl group is derived from pyruvate (see fig. 1.8). Initially,

[*U*-^{13}C] pyruvate was used as the sole carbon source in perdeuterated minimal expression medium for its labeling (Rosen, Gardner et al. 1996). However, in addition to the inevitable simultaneous labeling of Val, Leu and some methylene/methine sites, the final methyl protonation pattern of those residues is not homogeneous. For instance, the reported protonation of Ile-γ_2 methyl groups in the expressed protein in 99.9% ^2H_2O where pyruvate was incrementally added, was estimated to be 56%, with CH_3,$CH_2{}^2H$, CH^2H_2 and C^2H_3 isotopomer distribution of 32.3%, 19.7%, 31.5% and 16.5%, respectively (Rosen, Gardner et al. 1996). The appearance of these chemical species deteriorates NMR signal sensitivity and increases spectra complexity.

2-hydroxy-2-ethyl- -3-ketobutanoic acid is only one step further than α-ketobutyrate in *E. coli* Ile metabolism (see fig. 1.8). Its use in the 2-hydroxy-2-ethyl-$^2H_5$-3-ketobutanoate-4-^{13}C pattern was initially reported for the selective labeling of Ile-γ_2 methyl groups (Ruschak, Velyvis et al. 2010, Gans, Boisbouvier et al. 2010b). Missing the Ile ^{13}C enrichment at the C_β and C_α positions, the assignment of the obtained Ile-γ_2 through t bound magnetization transfer- based NMR experiments (e.g., relayed COSY or TOCSY) was not possible. Therefore, one year later, a new labeling scheme of the same molecule, i.e., 2-hydroxy-2-[$^2H_5$]ethyl-[1,2,3,4-$^{13}C_4$]-3-ketobutanoate, which links the methyl group to the backbone, was developed (Gans, Boisbouvier et al. 2010b, Ayala, Hamelin et al. 2012). In both of these studies, an interesting and unexplained isotope scrambling (~5%) to Leu-δ_1 and Val-γ_1 methyl group positions (pro-*R*) was observed (this scrambling will be discussed in greater detail later in this subsection). Nonetheless, the simultaneous addition of deuterated α-ketoisovalerate (at 200 mg/L) could dilute this undesired isotopic labeling.

Thus far, the synthesis of the used labeled 2-aceto-2-hydroxybutanoic acid is only feasible by chemical process (Ayala, Hamelin et al. 2012). In this protocol, the compound is synthesized in both R and S stereoisomers. Knowing that the R form is not metabolized by bacteria, half of the isotope enriched material is not used, which represents a drawback of this chemical production. A concentration of 100 mg/l should be added in the culture to obtain a high incorporation (>90%). The use of this protocol for obtaining other labeling patterns, e.g. 2-hydroxy-2-[$^2H_5$]ethyl-[4-^{13}C]-3-ketobutanoate or 2-hydroxy-2-[2'-^{13}C-1'-$^2H_2$]ethyl -3-keto-4-[$^2H_3$] butanoic acid is

conceptually possible. Unfortunately, the high cost or non-availability of the differently labeled primary reactants (i.e., [^{2}H, $^{13}CH_3$] iodo-ethane and ethyl-protected [*U*-^{13}C]-3-ketobutyrate) limits the applicability of this protocol.

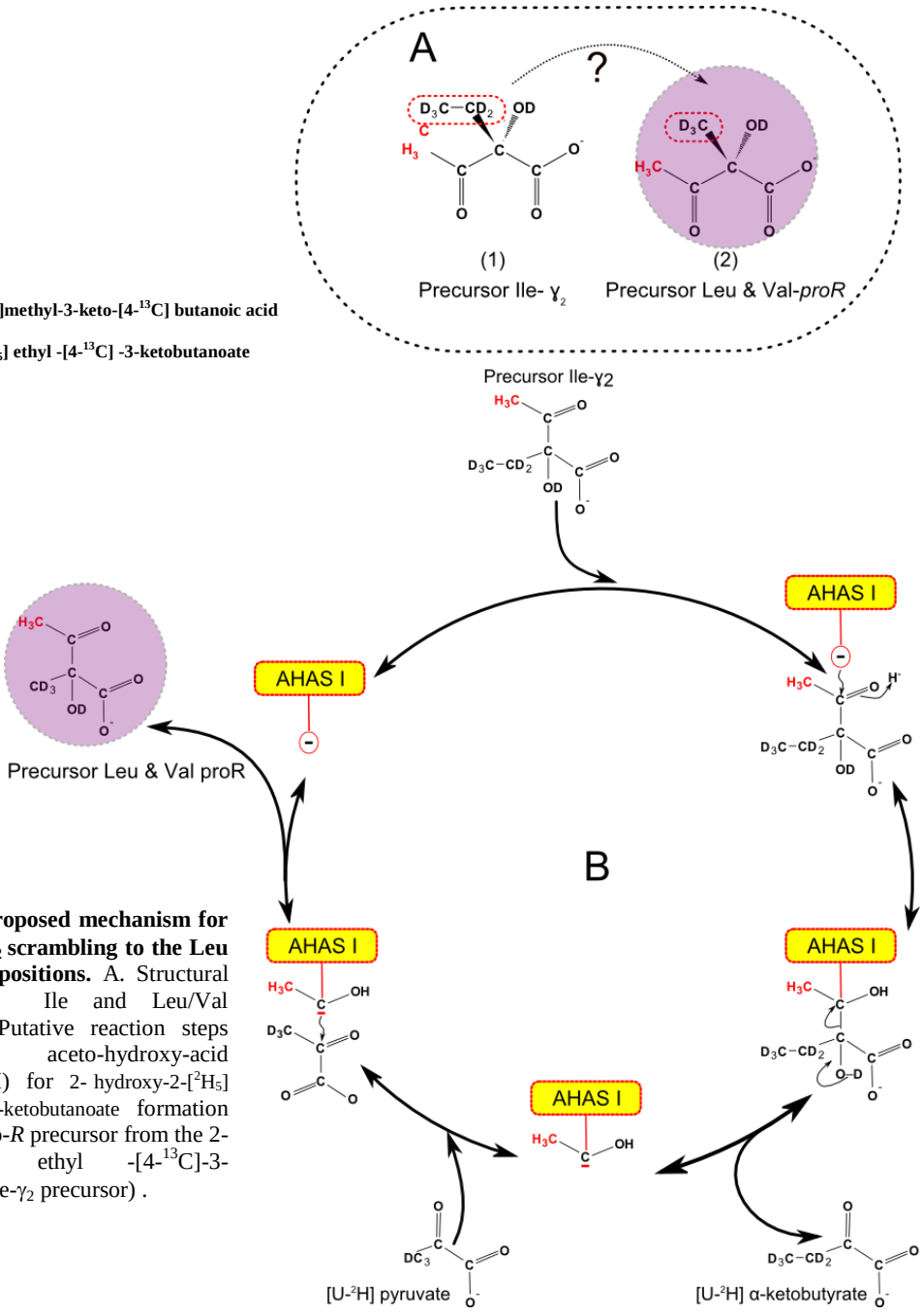
- *In vivo* 2-hydroxy-2-ethyl butanoate enzymatic processing and the origin of the observed Ile-γ_2 scrambling

The 2-hydroxy-2-ethyl butanoic acid is a product of the condensation of the previously discussed α-ketobutyrate with pyruvate for which the enzyme machinery was already presented. At this point, the novel aspect that can be discussed is the eventual enzymatic origin of the reported isotope leak to Leu-δ_1 and Val-γ_1. Indeed, the appearance of Leu-pro-*R* and Val-pro-*R* enriched with ^{13}C, indicates that their precursor is generated from 2-hydroxy-2-[$^{2}H_5$] ethyl-[4-^{13}C]-3-ketobutanoate, the unique ^{13}C source in the *E. coli* culture.

The structural comparison of this precursor with 2-acetolactate (Leu/Val proR precursor) indicates that the only difference between these compounds is the existence of an ethyl group, instead of a methyl group, at 2´ position of the Ile precursor (see fig. 1.10). Recently, an interesting work regarding the AHAS I characterization (Belenky, Steinmetz et al. 2012) reported that, unlike AHAS II and III, two steps (4 and 5 in fig. 1.7) of their well-known catalysis reactions can be reversibly performed by this enzyme. Considering this information, one can easily speculate that AHAS I deactetylates 2- hydroxy-2-[$^{2}H_5$]ethyl-[4-^{13}C]-3-ketobutanoate and releases the deuterated α-ketobutyrate. Subsequently, the labeled moiety is condensated with another deuterated pyruvate, as α-ketobutyrate is less abundant, forming [4-^{13}C]-2-acetolactate (2-hydroxy-2-[$^{2}H_3$]methyl-3-keto-[4-^{13}C] butanoic acid), as illustrated in figure 1.10. The same process should occur with 2-hydroxy-2-[$^{2}H_3$]methyl-3-keto-[4-^{13}C] butanoic acid; however, because the labeled carbon will always be at the same position (pro-*R*) after re-condensation, this reaction is not noticed.

1.3.2.1.2 Leucine and valine residues methyl groups labeling

The methyl groups of Leu and Val represent more than 50% of the available CH_3 probes



(1): 2-hydroxy-2-[$^2H_3$]methyl-3-keto-[4-^{13}C] butanoic acid

(2): 2- hydroxy-2-[$^2H_5$] ethyl -[4-^{13}C] -3-ketobutanoate

Figure 1.10: Proposed mechanism for the *in vivo* Ile-γ_2 scrambling to the Leu and Val proR positions. A. Structural comparison of Ile and Leu/Val precursors. B. Putative reaction steps catalyzed by aceto-hydroxy-acid synthase (AHA I) for 2- hydroxy-2-[$^2H_5$] ethyl -[4-^{13}C] -3-ketobutanoate formation (Leu and Val-pro-*R* precursor from the 2-hydroxy-2-[$^2H_5$] ethyl -[4-^{13}C]-3-ketobutanoate (Ile-γ_2 precursor) .

in proteins (McCaldon and Argos 1988). As illustrated in figure 1.8, both of these residues are formed from acetolactate, which is generated from the condensation of two molecules of pyruvate. These residues share several biosynthesis steps; therefore, the concomitant labeling of Leu and Val can be simply achieved using one common precursor.

Initially, [2,3-$^{2}H_2$]-^{13}C- Val was proposed for the selective protonation of both methyl groups of Leu and Val residues (Oda, Nakamura et al. 1992). To correct the observed low Leu-δ isotope enrichment, a new compound, [$^{13}CH_3$]-α-ketoisovalerate, was subsequently developed (Goto, Gardner et al. 1999). Both strategies were inefficient for the NMR study of large proteins because of the intense intra-residue ^{1}H-^{1}H dipolar interactions between the two prochiral methyl groups. Later, the introduction of a single labeled and protonated methyl group into the same precursor (i.e., [$^{13}CH_3$,$^{12}CD_3$]- α-ketoisovalerate) permitted the enhancement of the NMR spectra sensitivity, generating a better resolution (Gross, Gelev et al. 2003, Tugarinov and Kay 2004). The [$^{13}CH_3$,$^{12}CD_3$]- α-ketoisovalerate molecule is commercialized only in a racemic mixture and its use leads to the labeling of 50% of each prochiral Leu and Val methyl groups, maintaining all of them detectable by NMR. Because CH_3 groups appear in a narrow window range, the typical spectral congestion problem of large proteins renders this labeling scheme less appropriate compared to Ile-δ_1. Furthermore, despite the beneficial 50% reduction of ^{1}H-^{1}H dipolar interactions within the same molecule, the fact that each prochiral methyl group is only labeled at 50% represents another disadvantage. Indeed, a 4-fold intensity reduction is expected for the detectable long-range nOes when compared with the obtained intensity from the 100% stereospecific labeled prochiral methyl groups.

To improve even more resolution and sensitivity, stereospecific labeling of Leu and Val was finally proposed using 2-acetolactate (at 300 mg/L), a prior precursor relatively to α-ketoisovalerate in the metabolic pathway of these residues (see fig.1.8)(Gans, Hamelin et al. 2010). *In vivo*, this chemical compound undergoes an enzymatic and stereospecific rearrangement of its methyl groups via KARI enzyme catalysis.

Depending on its labeling pattern, 2-acetolactate allows the exclusive and scrambling-free labeling of either pro-*R* or pro-*S* in Leu and Val residues. Hence, the total number of the signals is reduced by a factor of two, alleviating the spectral overlap.
The reported isotope enrichment is high (>95%) but can be drastically decreased if high [*U*-^{2}H] glucose concentrations (>2 g/L) are used. This is mainly due to the higher amount of [*U*-^{2}H]-acetoacetate, originated from glucose-catabolism derived pyruvate, which competes with the exogenous labeled 2-acetolacte, leading to the low isotope enrichment. Because of their co-incorporation incompatibility, a similar effect is also observed when Ile precursors are simultaneously used with acetolactate (see §2.2).

Similarly to the α-ketobutyrate, the use of *U*-[^{13}C, ^{2}H] glucose is a prerequisite for Leu-C_{α} and carbonyls labeling in both of the above-explained situations (use of α-ketoisovalerate and 2-acetolacate as precursors). This pattern is usually required to correlate the resonance of the methyl group with nuclei in the polypeptide backbone for the side-chain assignment purposes.

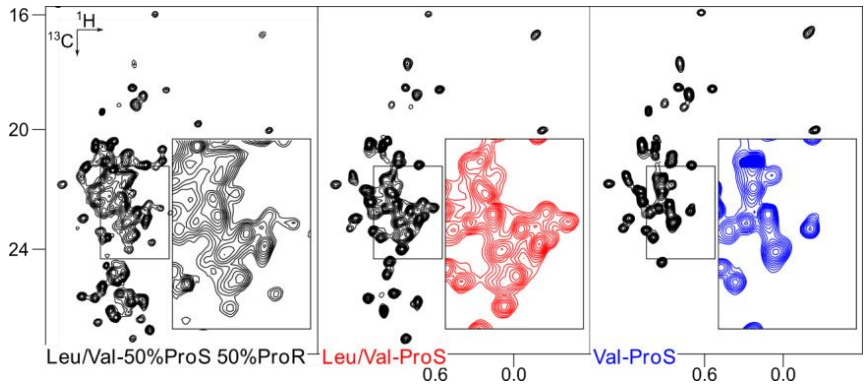


Figure 1.11: Improvement of NMR spectra resolution of supramolcular proteins. 2D [^{1}H,^{13}C] HMQC spectra of PhTET-2 (468 kDa). The protein was produced using [$^{13}CH_3$,$^{12}CD_3$]-α-ketoisovalerate (black), [$^{13}CH_3$]- 2-acetolactate (red) , [$^{13}CH_3$]-2-acetolactate and deuterated Leucine (blue). From (Mas, Crublet et al. 2013)

- Residue-specific stereo-labeling

Motivated by the same goal, i.e., reducing NMR spectral congestion, further protocols were reported aiming the labeling of only one of the two Leu and Val residues (see fig. 1.11). This objective can be primarily reached by the following approaches:

a) *By addition of specific precursor*

To reach the selective and exclusive labeling of a given residue, one can simply add its specific labeled precursor, which is generally more cost-effective than the amino acid itself. For Leu-δ methyl groups, labeled 2-ketoisocaproate, an irreversible metabolic intermediate, was suggested as a candidate (see fig 1.8) (Lichtenecker, Coudevylle et al. 2013). However, the utilization of 2-ketoisocaproate does not represent an ideal alternative to the described 2-acetolactate, despite of its affordability and ability to remove, *de facto*, the Leu/Val overlap. Indeed, this precursor is only available in a racemic mixture, resulting in 50% labeling of both Leu prochiral methyl groups (pro-*R* and pro-*S*). In comparison to the stereospecific labeling, the suppressed Val signals are compensated by doubling Leu NMR correlations, maintaining a similar number of peaks and leading to the 2-fold decrease in Leu NMR signal sensitivity. Obviously, this estimation is true as long as the numbers of Leu and Val contained in the studied particle are similar (which is generally the case). This approach is certainly useful but only up to a certain protein molecular weight, where signal overlap remains manageable. Therefore, a greater effort should be made to create a new protocol where stereospecific labeled 2-ketoisocaproate can be synthesized.

In contrast, the application of this approach to selective Val labeling is not possible. The analysis of the Val metabolic pathways clearly demonstrates that no irreversible Val precursor exists. Hence, other approaches had to be investigated (see b and c).

b) By a retro-inhibition approach

Recently, two published studies described the selective labeling of Val methyl groups using simple and robust protocols. Both strategies take advantage of the ability of 2-isopropylmalate synthase (IPMS) (EC 2.3.3.13) to be retro-inhibited by its end-product, Leu.

In the first protocol, the addition of stereo-specifically labeled Val (at 100 mg/L) together with deuterated Leu (20 mg/L) was suggested (Miyanoiri, Takeda et al. 2013). However, the obtained low Val incorporation (83%), in comparison to other protocol discussed below, in addition to the high price of this compound does not justify its utilization. Nevertheless, when the simultaneous labeling of Leu and Val at distinct prochiral methyl groups positions is required, then the use of both stereo-specifically

labeled amino acids cannot be avoided. This combination has been successfully applied to MSG (82 kDa) (Miyanoiri, Takeda et al. 2013). In the second publication, the affordable, labeled 2-acetolactate (300 mg/L) supplemented concomitantly with

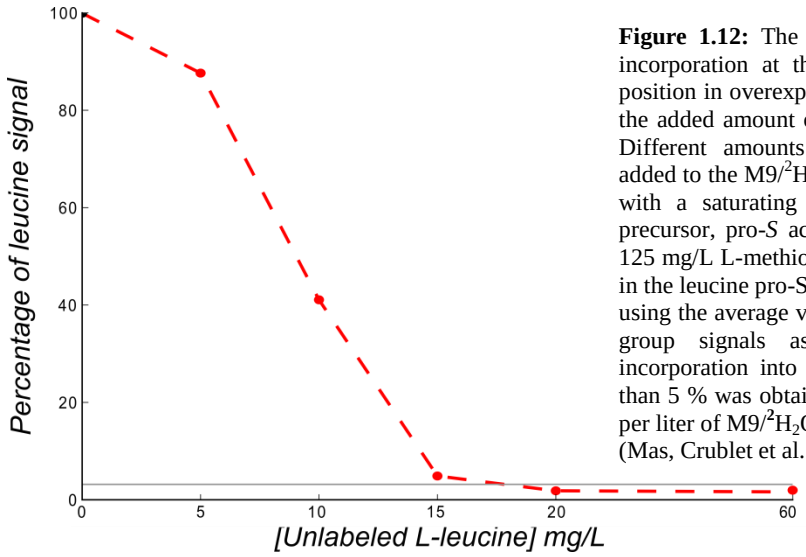


Figure 1.12: The decrease in $^{13}C^{1}H_3$ isotopomer incorporation at the leucine pro-*S* methyl group position in overexpressed ubiquitin as a function of the added amount of exogenous unlabeled leucine. Different amounts of non-labeled leucine were added to the M9/$^{2}H_2O$ medium 1 h prior to together with a saturating quantity of the leucine/valine precursor, pro-*S* acetolactate- ^{13}C (300 mg/L) and 125 mg/L L-methionine (methyl-^{13}C). The decrease in the leucine pro-S methyl group signal is analyzed using the average volume of the methionine methyl group signals as a reference. A level of incorporation into the leucine side chains of less than 5 % was obtained by adding 20 mg of leucine per liter of M9/$^{2}H_2O$ culture medium. Adapted from (Mas, Crublet et al. 2013)

deuterated Leu (20 mg/L) was used for the Val methyl groups assignment of PhTET-2 (468 kDa)(Mas, Crublet et al. 2013). Val-pro-*S* was incorporated at 90% with a residual labeling of 2% at Leu-pro-*S* methyl groups (see fig. 1.12).

c) *By addition of a stereo-specifically labeled amino acid*

The previously mentioned Leu spectral congestion obtained using 2-ketoisocaproate in large proteins could be easily avoided using stereo-specifically labeled Leu (pro-*R* or pro-*S* at 20 mg/L). In fact, this approach is currently possible for both Leu and Val (see b) as their chemical synthesis is already described (Miyanoiri, Takeda et al. 2013). Nonetheless, despite their usefulness, particularly stereo-specifically labeled Leu, the employment of these compounds is seriously compromised by the high associated costs of their chemical synthesis.

- *In vivo* Leu and Val precursor enzymatic processing

The primarily enzymes that modulate the efficiency of the proposed protocols for Leu and Val labeling are KARI and IPMS.

a) *Ketol-acid reductoisomerase*

This enzyme was already described in §1.3.3.1.1. The primary importance of KARI in this context remains in its role in stereo-specifically transferring the methyl group (pro-*S*) from the 2-acetolactate to the subsequent intermediate (2-keto-3-hydroxyisovalerate) ensuring the desired labeling.

b) 2-isopropylmalate synthase

The 2-isopropylmalate synthase enzyme (IPMS) (E:C: 2.3.3.13) catalyzes Claisen-condensation between acetyl coenzyme A (AcCoA) and an α-ketoisovalerate acid. This enzyme requires a divalent-metal for its activity (Casey, Baugh et al. 2012). This reaction represents the first step for L-Leu biosynthesis. IPMS was classified into the LeuA dimer superfamily, which is characterized by two conserved domains: an N-terminal hydroxymethylglutaryl-CoA lyase-like catalytic domain and a C-terminal LeuA dimer regulatory domain (Frantom 2012) (see fig. 1.13). IPMS was isolated from various species, and invariably, this enzyme is subject to feedback inhibition by L-leucine. Because of being a potential drug target, IPMS from *Mycobacterium tuberculosis* (MtIPMS) is the one that has been extensively characterized.

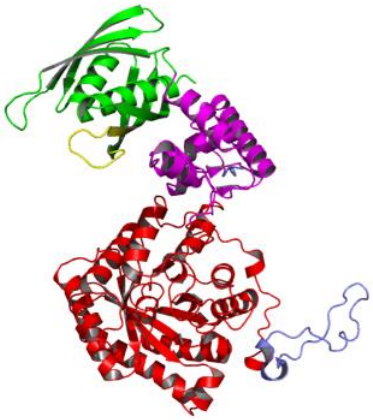


Figure 1.13: Representation of the monomer 2-isopropylmalate synthase. Green: regulatory domain; Magenta: linker; Red: catalytic domain.; Yellow: the flexible loop involved in the onset-step inhibition (PDB 3FIG).

Kinetic studies of MtIPMS revealed that L-leucine acts as a non-competitive inhibitor versus α-ketoisovalerate, with comparable K_{ii} and K_{is} values of 22 μM and 8 μM, respectively. L-Leu exerts a reversible and slow-onset inhibition mechanism on MtIPMS, which is particular because all reported inhibitors that have shown the same mechanism were characterized as competitive. Thus, L-Leu binds rapidly to the enzyme forming an E–I complex that subsequently undergoes a slow equilibration with a more inhibited form of the enzyme, E*–I. Structural studies revealed that L-Leu binding does not provoke any changes in the quaternary organization and in the rigid domain motion of MtIPMS. The Leu allosteric regulation of the enzyme was found to be ensured by a discrete inhibitory signal transmitted from the regulatory domain to the active site (Carvalho, Frantom et al. 2009). In a recent work, the flexible loop located in the regulatory

domain was described to mediate the slow-onset step of allosteric inhibition (Casey, Baugh et al. 2012) (see fig. 1.13).

1.3.2.1.3 Methionine residue methyl group labeling

Methionine (Met or M) possesses the most dynamic side-chain of the methyl-containing residues in proteins. A statistical study reported that the order parameters $(S^2)^h$ of Met methyl axes tend to range between 0.1 and 0.2 (Yang, Mittermaier et al. 1998).

This Met property translates into favorable T_2 properties, which are reflected by a high signal-to-noise ratio of NMR signal. For instance, although attempts using ^{15}NH have failed, the labeling of [$^{13}CH_3$]-Met allowed the acquisition of sensitive NMR spectra of the dynamic N-terminus in the archaeal 20S core particle proteasome (670 kDa) (Religa, Sprangers et al. 2010). Consequently, dynamic insights into the gating regulation mechanism of proteasome were obtained (Religa, Sprangers et al. 2010).

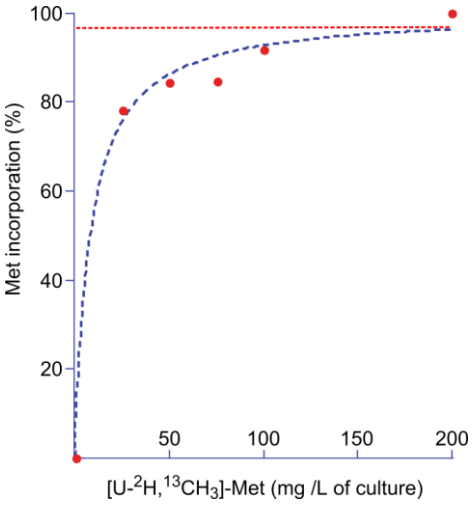


Figure 1.14: The [U-^{2}H, $^{13}CH_3$]-methionine incorporation curve. The Ubiquitin was produced in deuterated culture with increasing Met concentrations. NMR signal integration were performed using NMRDraw (Mas et. al. *in preparation*).

From another point of view but still related to the NMR signal quality, this residue is also characterized by resonating in an uncrowded region of the [^{1}H,^{13}C] correlation spectrum, offering by that high signal resolution (Stoffregen, Schwer et al. 2012).

The selective protonation and isotope labeling of Met-ε methyl groups can be achieved without noticeable scrambling, either by adding the labeled residue (Gelis, Bonvin et al. 2007, Religa, Sprangers et al. 2010, Gifford, Ishida et al. 2011, Weininger, Liu et al. 2012) or its labeled precursor (Fischer, Kloiber et al. 2007). To reach high L-labeled Met incorporation (>90%), this residue must be added at 100 mg/L. The Met incorporation curve is depicted in figure 1.14. Obviously, in the case where L,D-Met is

[h] S^2 is the spatial restriction of internal motion (order parameter) varying from 0 to 1. If the internal motion is not present, then S^2 approaches 1.

used, then the supplemented quantity must be doubled once the D form is not metabolized by the bacteria.

Similar to most of the precursors described in this introduction, Met, in different isotopic variations (i.e, *U*-[^{1}H,^{13}C]-L Met, *U*-[^{2}H,^{13}CH$_3$]-L Met) can be purchased from isotope suppliers. Alternatively, this residue can also be chemically synthesized by condensing ^{13}C-methyl iodide with D,L-homocysteine (deuterated or not) in liquid ammonia (Melville, Rachele et al. 1947, Gifford, Ishida et al. 2011, Weininger, Liu et al. 2012). The use of α-keto-methionine, which is a metabolic precursor of methionine, has also been proposed (Fischer, Kloiber et al. 2007) but no incorporation studies are available to evaluate the extent of the generated Met isotope enrichment. In case that α-keto-methionine presents similar incorporation to that of Met molecule, its utilization is expected to be more affordable. Indeed, its amination occurs within the cell and directly in the correct stereoisomer, avoiding the waste of isotope material as is the case of L,D-Met.

Furthermore, the methyl group of methionine is isolated from the lateral carbon skeleton through its bonding to a sulfur atom. Their protons are distant from the other hydrogen atoms of the side-chain therefore ^{1}H-^{1}H-dipolar interactions are reduced. Thus, the use of [*U*-^{2}H, ^{13}CH$_3$]-Met may be judged useless. Indeed, large machineries (1 MDa) were studied using protonated Met, a fact that can argue this point of view (Religa, Sprangers et al. 2010). In our laboratory, the putative effect of the deuterated side-chain [^{13}CH$_3$]-Met on NMR signal quality was investigated. A preliminary analysis of the NMR signal intensity of an ensemble of separated peaks in two samples of 1 MDa protein (prepared with protonated and deuterated Met side-chain) revealed an average gain of approximately 15% when using [*U*-^{2}H, ^{13}CH$_3$]-Met.

1.3.2.2 Amino acids involved in reversible metabolic pathways

1.3.2.2.1 Alanine residue methyl group labeling

Alanine (Ala or A) is one of the most abundant residues in proteins. Statistically, this amino acid represents approximately 8-9% of the total number of amino acids, with a particularly large distribution in both protein core and surface (Miller, Janin et al. 1987,

McCaldon and Argos 1988). This residue is used as a backbone secondary structure (Wishart and Sykes 1994) and dynamic (Godoy-Ruiz, Guo et al. 2010) reporter.

Specific β-methyl group labeling can be achieved using several protocols; however, all of these protocols rely on commercially available [3-^{13}C] L-Alanine addition (Isaacson, Simpson et al. 2007, Ayala, Sounier et al. 2009). Notably, unlike the above-mentioned amino acids, which are synthesized by irreversible metabolic pathways in *E. coli*, Ala results from a single reversible reaction from pyruvate. Figure 1.8 clearly illustrates the primary isotope scrambling that pyruvate metabolism generates. Therefore, in an initial protocol, [3-^{13}C] L-Alanine, which was previously incubated in $^{2}H_2O$ with Tryptophan Synthase (other transaminases can be alternatively used *e.g* BCAT) to exchange $^{1}H_{\alpha}$ for ^{2}H, was added simultaneously with a high concentration of perdeuterated amino acids (deuterated algae lysates)(Nibeilliu and Malthouse 2004, Isaacson, Simpson et al. 2007). This important deuterated amino acid reservoir effectively decreases the undesired isotope leak and may even enhance the *E. coli* growth. However, the level of Ala incorporation into the targeted protein was negatively affected due to labeled Ala dilution (incorporation estimated to be at 30%)(Ayala, Sounier et al. 2009). To solve this drawback, an alternative solution was published a few years later, consisting of adding labeled Ala and concomitantly saturating the primary metabolic pathways that involves pyruvate. This approach allows to reach a high Ala incorporation (>95%) where the isotopic scrambling is reduced from 25% to less than 0.1% by supplementing deuterated (60 mg/l) L-isoleucine, α-ketoisovalerate (200 mg/L) and succinate (2.5 g/L).

- *In vivo* Alanine enzyme processing

Although at the end of a metabolic pathway, Ala involvement in a reversible reaction causes isotope scrambling of its labeled methyl group to pyruvate. As a central block of amino acids biosynthesis, the pyruvate labeled fraction must be diluted by adding deuterated precursors. The enzymes implicated in the conversion of Ala into pyruvate are described below.

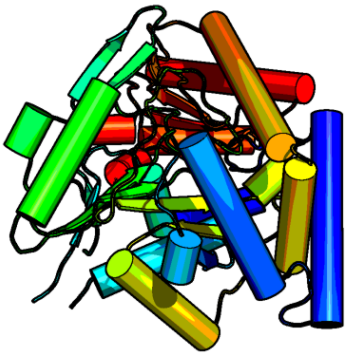


Figure 1.15: Asymmetric unit of the crystal structure of valine-pyruvate aminotransferase (AvtA) from Salmonella typhimurium LT2 (PDB: 3G7Q).

In the present manuscript, the major part of transaminases that use Ala as substrate with high specificity, either as donor or receptor of amine groups, are assumed to be those transaminases involved in its biosynthesis. Surprisingly, the literature search regarding theses enzymes revealed an important lack of biochemical and structural data. Based on the available literature, three major transaminases ensure Ala biosynthesis. Together with the well-known Alanine-valine transaminase (AvtA), two additional enzymes (YfbQ and YfdZ) were recently identified (Kim, Schneider et al. 2010, Yoneyama, Hori et al. 2011). Indeed, although single or double deletions of their respective genes do not affect the growth of *E. coli*, the triple mutant (*avtA yfbQ yfdZ)* displayed drastically decreased growth (Kim, Schneider et al. 2010)˙(Yoneyama, Hori et al. 2011). In both studies, normal growth was restored upon Ala addition. Moreover, in addition to 3-phosphoserine amino-transferase other transaminases can cover the same biosynthetic activity but in a minor way.

a) *Alanine-Valine Transaminase*

Alanine-Valine Transaminase (AvtA), also known as transaminase C, catalyzes pyruvate amination to form L-Ala, using Val as an amine donor. This enzyme is also involved in the final step of Val synthesis through α-ketoisovalerate amination with an amine group taken from 2-aminobutyrate. Both Leu and Ala seem to repress the *avt*A gene; however, contradictory results have been reported regarding the Val effect on *avt*A gene (McGilvray and Umbarger 1974, Falkinham 1979, Whalen and Berg 1982, Whalen and Berg 1984).

The addition of 1 mM Ala to *E. coli* culture seems to repress *avt*A gene expression by 2-fold (Whalen and Berg 1984, Kim, Schneider et al. 2010). In the reported Ala labeling protocol (Ayala, Sounier et al. 2009), an important quantity of this residue is usually added to the culture (>500 mg/L) to optimize its incorporation (> 90%). This concentration represents 5.5- to 9-fold of the tested concentration (1mM) in the above-mentioned work. One would expect that AvtA activity is significantly decreased during the labeled protein expression, at least as long as the Ala cell concentration is high. Thus, the observed isotopic scrambling can be speculated to be either caused by other transaminases or simply starts to appear when the Ala cell content decreases (due to its consumption for protein synthesis). The NMR analysis of the expressed Ala-[$^{13}CH_3$]-

proteins resulting from different induction times could somehow elucidate this question. As previously highlighted, no AvtA structural analysis was reported in *E .coli*. Nevertheless, an X-ray crystallographic structure with 1.8 Å resolution of valine-pyruvate Amino-transferase (AvtA) from *Salmonella typhimurium* LT2 was deposited in the Protein Data Bank in 2009 and modified in 2011 (see fig. 1.15). Unfortunately, no publication is associated with this structure.

b) *YfbQ and YfdZ enzymes*

As previously discussed, the identification of YfbQ and YfdZ enzymes is relatively recent, and no structures and mechanistic models have yet been proposed (Kim, Schneider et al. 2010). An initial biochemical characterization of both enzymes reported that they are 87 kDa homodimers. YfbQ seems to possess a higher affinity to pyruvate than YfdZ because their respective K_m values were estimated to be 0.55 ±0.06 mM and 0.94 ±0.19 mM. These values are within the range of the in-cell pyruvate concentrations. The K_m of YfbQ for alanine was approximatively 4.9 ±1.2 mM, indicating that the deamination of Ala is most likely more favored by this enzyme than its amination.

c) *Transaminases inhibition*

Instead of inhibiting the metabolic reactions for isotopic scrambling suppression, one would prefer to "silence" the involved transaminases. The use of microorganisms deficient in a specific aminotransferase has been proposed (see §1.2.2.1.1)(Griffey, Redfield et al. 1985, Waugh 1996, Weigelt, van Dongen et al. 2002, Fiaux, Bertelsen et al. 2004). However, because of the overlapping specificity of transaminases, the simultaneous deletion of several of them is often required. For a compensation purpose, the concomitant supplementation of amino acids is therefore needed. In the special case of Ala, because the mechanism of its biosynthetic pathway remains unclear, the use of a microorganism that defect all major *E. coli* transaminases was suggested (*ilv*E, *tyr*B, *asp*C and *avt*A)(Waugh 1996). Submitting the bacterium to such drastic modification, together with ^2H_2O toxicity, is expected to be detrimental for cell growth and protein overexpression yield. Nonetheless, no study using this approach was reported. Considering the previously discussed results regarding the transaminases involved in

Ala metabolism, the *avtA yfbQ yfdZ serC E. coli* mutant could be a good choice to test the overexpression of Ala labeled proteins in $^{2}H_2O$.

Alternatively, the use of chemicals to inhibit transaminases in *E.coli* is conceptually an interesting approach since their addition is easier, universal and should be cost-effective than deuterated precursors. In fact, the previously cited example (see §1.2.2.1.1) of a PLP-dependent enzyme inhibitor, i.e., aminooxy acetic acid (120 mg/L), has been reported to suppress efficiently the scrambling in the case of the [Val-^{15}N]-protein *E. coli* expression (Lopukhov, Ponomareva et al. 2002). However, due to the combined toxic effects of $^{2}H_2O$ and these chemical compounds, this approach is expected to not be viable.

In a cell-free protein synthesis, adding β-chloro L-alanine was reported to successfully inhibit the Alanine aminotransferases because no Ala scrambling has been observed (Morita, Shimizu et al. 2004). However, thus far, the addition of this compound in an *E. coli* culture has not been reported, most likely due to its cell toxicity.

1.3.2.2.2 Threonine residue methyl group labeling

Unlike the remaining methyl group-containing residues, Thr is the only amino acid that has a higher propensity to be located in protein surfaces rather than the interior (7.1% against 4.6%, respectively) (Miller, Janin et al. 1987). Due to its hydroxyl and methyl groups, Thr is the only residue that can participate in both hydrogen bonds and non-polar interactions. In contrast to the aforementioned residues, Thr represents an intermediate metabolite for another amino acid (i.e., Ile). It was the last resistant residue to be tamed for the isotopic labeling.

Initially, a protocol using [2-^{13}C] glycerol as a unique carbon source was proposed to specifically label Thr methyl group (Sinha, Jen-Jacobson et al. 2011) (see figures 1.8 and 1.17). The use of [2-^{13}C] glucose was avoided because of the 50% ^{13}C dilution that results from its conversion into 2 molecules of pyruvate (one labeled and the other unlabled). The [2,5-^{13}C]-glucose pattern could be a solution; however, its high cost makes this idea impracticable. In this protocol, supplementation with $NaH^{13}CO_3$ (10 g/L), deuterated Ala (800 mg/L), α-ketoisovalerate (200 mg/L), Ile (80 mg/L), Met (150 mg/L) and Lys (200 mg/L) was applied to prevent the decrease in Thr ^{13}C enrichment

(see figures 1.8 and 1.17). Unfortunately, the application of this protocol on large proteins is severely compromised by the fact that $^{1}H/^{2}H$ exchange occurs within the cell resulting in the predominant appearance of $^{13}CH^{2}H_2$ isotopomer (90%) representing only ~25% of the methyl groups in $^{2}H_2O$-overexpressed protein. The high and non-uniform proton density, which is inevitably achieved because of the used protonated glycerol, represents another important drawback. The obtained labeling pattern was described to be optimal for ^{13}C relaxation study, as it generates only a $^{13}CH^{2}H_2$-Thr population, and was applied to the RNA binding domain of the transcriptional termination factor rho (130 residues) (Sinha, Jen-Jacobson et al. 2011).

For a long time the development of labeled Thr precursors was not possible because of the following (see fig. 1.17): (*i*) The first candidate, i.e., phospho L-homoserine would necessarily take at least one deuterium from the deuterated medium culture, generating a $CH_2{}^{2}H$ isotopomer specie that is not the ideal labeling specie for the TROSY NMR studies of high molecular weight complexes (Religa and Kay 2010). (*ii*) The second candidate, acetyl-CoA, which forms Thr through its condensation with glycine, would result in an uncontrollable isotope scrambling; (*iii*) The third candidate, acetaldehyde, which undergoes the same reaction as acetyl-CoA, is commercially available but expensive (Velyvis, Ruschak et al. 2012). Therefore, the direct addition of Thr represented the more appropriate approach and the synthesis of [α,β-^{2}H, γ_2-$^{13}C^{1}H_3$]-L-Thr had to be investigated.

Threonine residue synthesis and protocols for its incorporation

The chemical synthesis of [α,β-^{2}H, γ_2-$^{13}C^{1}H_3$]-L-Thr is challenging because this molecule possesses two asymmetric centers. After this long delay in setting up a protocol for its labeling, an *in vitro* enzymatic synthesis was finally established two years ago (Velyvis, Ruschak et al. 2012). This protocol consists of reproducing the major enzymatic reactions that occur *in vivo* for the Thr biosynthesis. Starting from the condensation of deuterated pyruvate with the affordable ^{13}C-formaldehyde, [α,β-^{2}H, γ_2-$^{13}C^{1}H_3$]-L-Thr was produced in four steps with a yield of 60%. All reactions were enzymatically performed (see fig. 1.16).

The use of the synthesized [α,β-^{2}H, γ_2-^{13}C^{1}H$_3$]- L-Thr to label proteasome (β subunits) led to the expected isotope scrambling into the Ile–δ_1 methyl groups that results from the threonine deaminase (TD) activity (Velyvis, Ruschak et al. 2012).

Additionally, due to the threonine dehydrogenase (TDH), another "invisible" scrambling occurs. Indeed, as demonstrated in figure 1.17, [α,β-^{2}H, γ_2-^{13}C^{1}H$_3$]-L-Thr is degraded into fully deuterated Gly and AcetylCoA to which the selectively protonated and isotopically labeled methyl group of Thr is transferred. This latter molecule enters the citric acid cycle (TCA) and forms Glu (the precursor for other amino acids) that will incorporate ^{1}H and ^{13}C arising from [^{13}CH$_3$]-Acetyl-CoA, at the carbon γ position.

Notably, when *U*-^{13}C Thr is used, additional ^{13}C scrambling is widespread through Gly that subsequently integrates Ala and Met biosynthesis pathways, via Ser and C_1 metabolisms, respectively (see fig. 1.17).

Thus, in addition to the non-desired proton incorporation, which appears to be low, this pathway decreases the *in vivo* availability of labeled Thr (Velyvis, Ruschak et al. 2012).

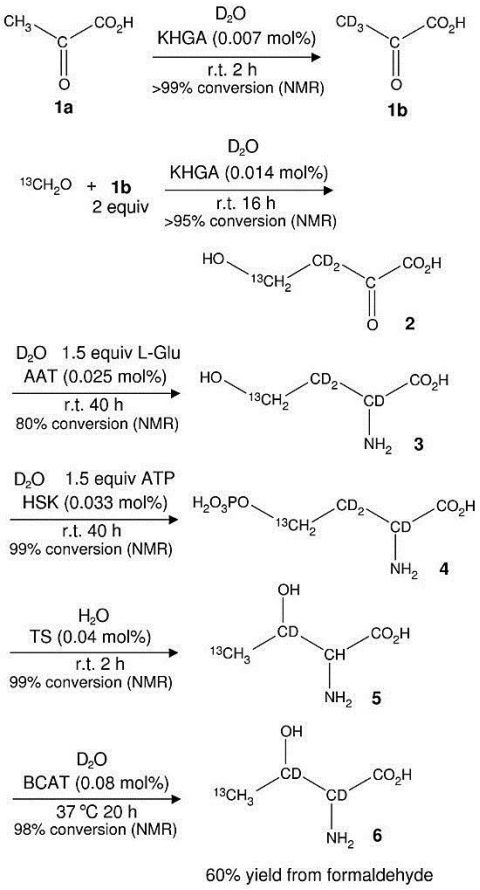


Figure 1.16: Scheme representing the enzymatic biosynthesis of U-[^{2}H],Thr-γ_2[^{13}CH$_3$].1a. pyruvate. 1b. [U-^{2}H]-pyruvate.2. 2-keto-4-hydroxybutyrate.3.Homoserine.4.O-phospho-L-homoserine. 5-6 Threonine. Enzyme name abbreviations: 2-keto-4-hydroxyglutarate aldolase(KHGA), aspartate aminotransferase (AAT), homoserine Kinase (HSK), threonine synthase (TS), branched-chain-amino-acid Aminotransferase (BCAT). From (Velyvis, Ruschak et al. 2012)

The consequence of this "Thr consumption" implies either low labeled-Thr incorporation or the requirement of a higher quantity to achieve the highest incorporation. Thus, aiming also the Ile signal labeling, addition of 50 mg/L [α,β-^{2}H, γ_2-^{13}C^{1}H$_3$]-L-Thr together with 50 mg/L [4-^{13}C, 3-^{2}H$_2$]-α-ketobutyrate and 100 mg/L deuterated Gly was suggested as a good combination to reach ~90% and >95% of Thr-γ_2 and Ile-δ_1 isotope enrichment, respectively.

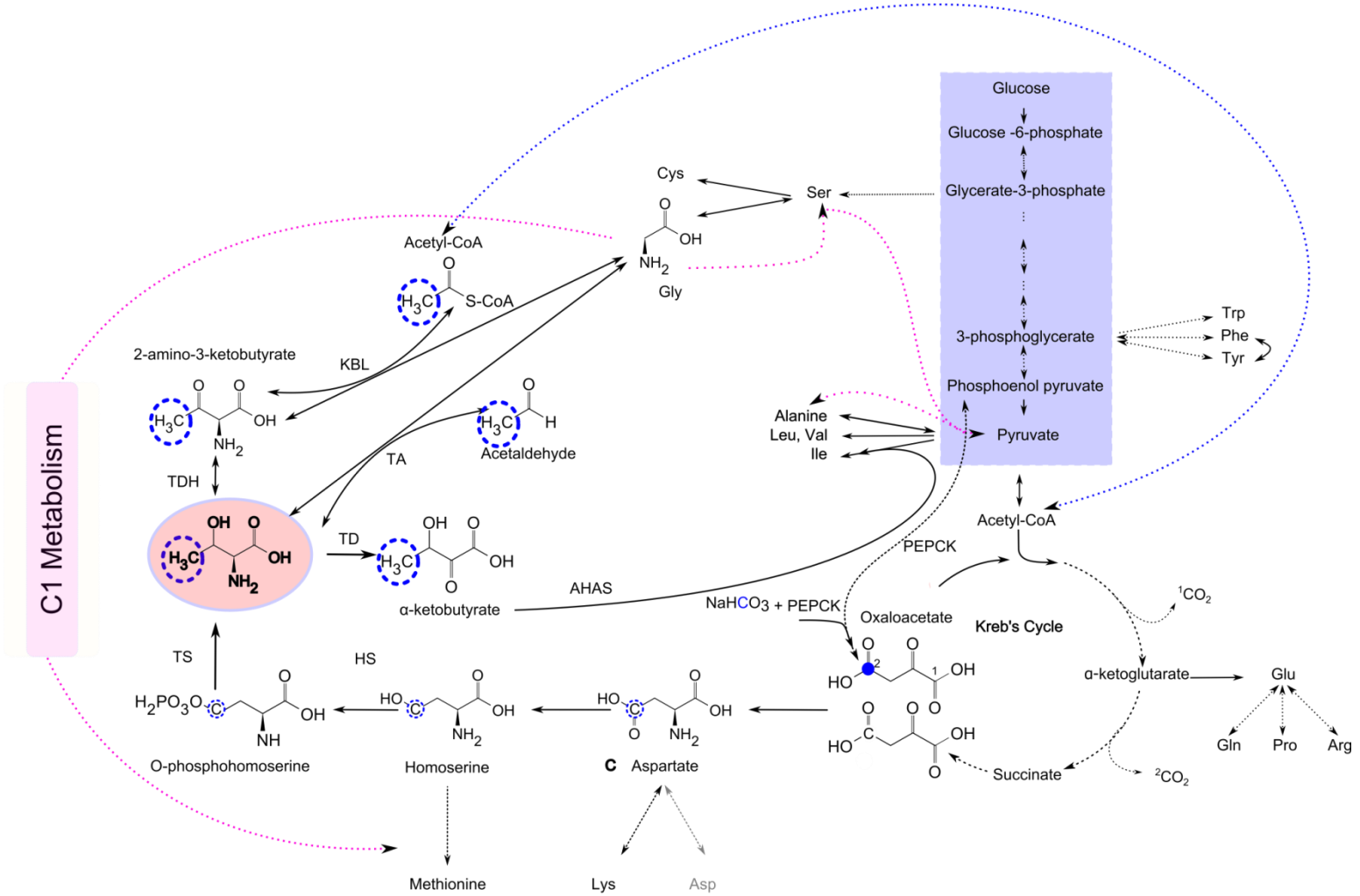


Figure 1.17: ***E. coli* metabolic pathways for Threonine biosynthesis and degradation.** Arrows depicted in pink represent reactions through which ^{13}C scrambling, via glycine, is obtained when U-^{13}C- Thr is used. Carbons highlighted in blue represent a parallel pathway of the Thr methyl group origin (Sinha et. al, 2011). The carbons surrounded with dashed blue circles represent atoms that can be either from the TCA oxaloacetate or from phosphoenolpyruvate. The 1, 2 numbers in oxaloacetate molecule indicate the carbon atoms that are lost in the TCA. Enzyme abbreviation: TD:Threonine Deaminase; CP:Pyruvate Carboxylase, HS: Homoserine Kinase; TS: Threonine Synthase; TD: Threonine Deaminase; AHAS: Aceto-hydroxy-acid synthase; TDH: Threonine Deshydrogenase; TA: Threonine Aldolase; KBL: 2-amino-3-ketobutyrate ligase. Dashed arrows represent multi-steps reactions

This labeling scheme was applied to the 20S proteasome core particle (670 kDa)(Velyvis, Ruschak et al. 2012).

Although, this synthesis protocol allows the use of [$^{13}CH_3$]-Thr as an NMR probe, its application is laborious and time-consuming because 5 recombinant enzymes must be previously produced and purified.

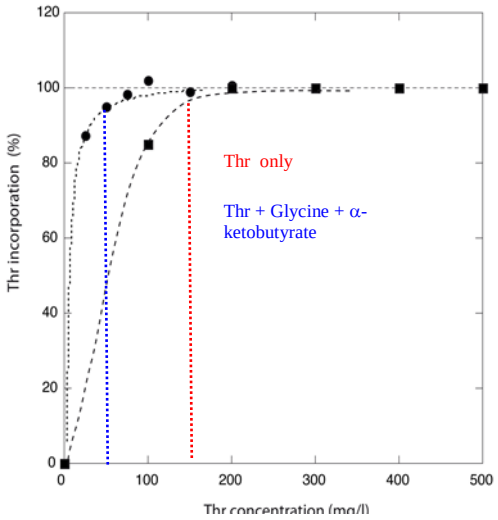


Figure 1.19: Optimization of the incorporation of Thr [α,β-^{2}H, γ_2-^{13}C^{1}H$_3$]-L-Thr. The Threonine curve incorporation was performed using [U-^{2}H] ubiquitin. Met was used as internal reference for the quantification of Thr incorporation.

Recently, an alternate chemical synthesis protocol has been set up by our collaborators (Hamelin et al. *in preparation*) to generate the same Thr labeling. Using this protocol, labeled Thr can be produced with a yield of ~35 % within 2 weeks using ^{13}C acetyl chloride as starting labeled chemicals.

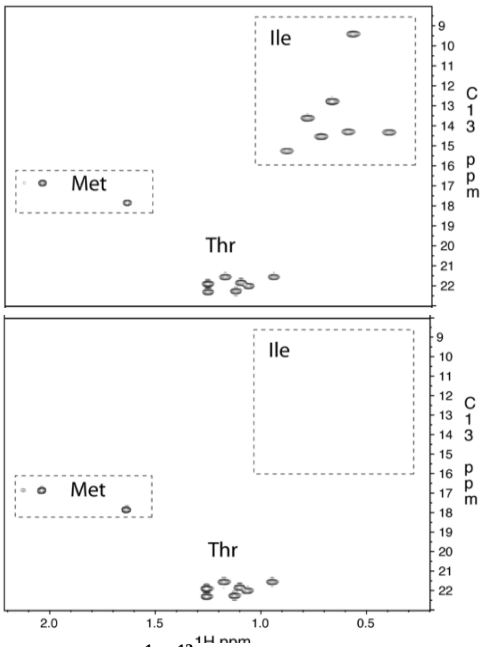


Figure 1.19: : [^{1}H,^{13}C] HSQC spectra of [U-^{2}H, Thr-Cγ_2] human ubiquitin. Ubiquitin was produced using (A) [α,β-^{2}H, γ_2-^{13}C^{1}H$_3$]-L-Thr (50 mg/L) + Met (100mg/L) and (B) [α,β-^{2}H,γ_2-^{13}C^{1}H$_3$]-L-Thr (50 mg/L) + [U-^{2}H]-α–ketobutyrate (100 mg/L) + [U-^{2}H]-glycine (500mg/L) + Met (100mg/L). Met was used as an internal reference for the quantification of Thr incorporation.

According to our results, the simple addition of labeled Thr would require a quantity of ~150 mg/L to ensure its high incorporation (> 90%) into $^{2}H_2O$-expressed protein (see fig. 1.18). Similar to the formerly presented results, reducing the [α,β-^{2}H, γ_2-^{13}C^{1}H$_3$]-L-Thr concentration by 3-fold is feasible when added together with [*U*-^{2}H]-α-ketobutyrate (60 mg/L) and [*U*-^{2}H]-Gly (500 mg/L), respectively (see figures 1.18 and 1.17). Through this work, α-ketobutyrate was clearly demonstrated to "turn-off" Ile synthesis from Thr efficiently, confirming previous hypothesis in the published work (see fig. 1.18 & 1.19) (Velyvis, Ruschak et al. 2012) .

- *In vivo* threonine enzymatic processing

Threonine differs from the remaining amino acids by being an intermediate metabolite. To produce Ile endogenously, Thr residue is degraded by TD in *E. coli*. Thr degradation via TDH leads to glycine and acetyl-CoA synthesis, which integrate Ala/Met biosynthesis pathways and TCA, respectively. These enzymes are discussed below (see fig. 1.17).

a) Threonine deaminase or dehydratase

E. coli TD (EC 4.2.1.16) is a homo-tetramer (MW 220 kDa) belonging to the PLP-dependent enzyme family (Gallagher, Gilliland et al. 1998) and exists in two types: The biosynthetic (BTD), involved in an aerobic process, and the catabolic (CTD) that acts in an anaerobic conditions (Yu, Li et al. 2013).

Both of these types catalyze the conversion of L-Threonine to α-ketobutyrate and ammonia. Additionally, BTD is also involved in the deamination of L-Ser to pyruvate. Whereas the BTD unit possesses the catalytic (N-terminal) and regulatory (C-terminal) domains, the CTD only has the former one. BTD, encoded by the gene *ilv*A, was one of the first allosteric enzymes to be discovered (Umbarger 1992). BTD participates in the carbon flow regulation of the branched amino acids through an allosteric control where Ile is the inhibitor whereas Thr and Val are the activators (Gallagher, Gilliland et al. 1998, Yu, Li et al. 2013). The apparent affinity of TD for Ile was estimated to be 100 times its affinity for Thr (Umbarger and Brown 1957). The previously discussed Thr labeling protocol (Hamelin et al., *in preparation*) takes advantage of this Ile retro-control of the TD enzyme. A sufficient quantity of [*U*-^{2}H]- α-ketobutyrate, which generates [*U*-^{2}H]-Ile production, is applied to "inhibit" Ile synthesis from labeled Thr. This inhibition will help to avoid the decrease in its quantity and suppress the non-desired NMR signals that otherwise would be present simultaneously with [$^{13}CH_3$]-Thr. Furthermore, [3,3-$^{2}H_2$]- α-ketobutyrate (50 mg/L) incorporation is reportedly slightly decreased when [3-^{2}H]-^{13}C-α-ketoisovalerate (100 mg/L) is concomitantly added (Goto, Gardner et al. 1999). The authors explained that

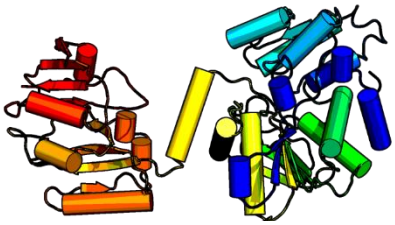


Figure 1.20: TD crystallographic structure (PDB 1TDJ)

the Val generated from the high amount of the exogenously supplemented α-ketoisovalerate will cope with the Ile retro-inhibition on TD. The lifting of this feedback inhibition leads to the production of unlabeled α-ketobutyrate from the available Thr cell content, diluting, consequently, the isotope enrichment.

b) *Threonine dehydrogenase*

Threonine dehydrogenase (TDH) (1.1.1.103) catalyzes Thr oxidation to 2-amino-3-ketobutanoate (see fig. 1.17), using NAD^+ as a cofactor. This product is subsequently either decomposed to aminoacetone and CO_2 or converted into glycine by 2-amino 3- ketobutanoate CoA ligase (Nakano, Okazaki et al. 2014). The Thr catabolism pathway initiated by TDH represents the major route (87% in rat hepatocytes) (Bird and Nunn 1983) for its degradation in *E. coli* where Leu was demonstrated to be its inducer.

Figure 1.21: Crystallographic asymmetric unit of Threonine Dehydrogenase (PDB 2DFV). Three monomers of *P. horikoshii* TDHs were located in the crystallographic asymmetric unit, however, the crystal structure exhibits a homo-tetramer structure

TDH is a homo-tetramer (148 kDa) that belongs to the long chain alcohol dehydrogenases/reductases family (Epperly and Dekker 1991). Each subunit contains one Zinc ion that is indispensable for the enzyme activity (Epperly and Dekker 1991). The divalent metal ion is thought to be in giving TDH a higher specificity toward threonine and reduced activity with unnatural substrates (Johnson, Chen et al. 1998). The K_m values of *E. coli* TDH for L-threonine and NAD^+ were reported to be 1.43 mM and 0.19 mM, respectively, with a kcat value of 8100 min^{-1} and a specific activity of 57 U/mg-protein at 37ºC. The first TDH X-ray crystallographic structure was from *Pyrococcus horikoshii* (Ishikawa, Higashi et al. 2007)(see fig. 1.21), which shares 41.3% identity with the E. coli TDH. Its kinetic characterization revealed that the K_m values of PhTDH at 65ºC for L-threonine and NAD^+ were 0.013 mM and 0.01 mM, respectively, with a kcat value of 33100 min^{-1} and specific activity of 1750 U/mg-protein (Higashi, Fukada et al. 2005).

c) *2-amino 3-ketobutyrate CoA ligase*

KBL (EC 2.3.1.29) is a PLP-dependent enzyme that catalyzes the previously referred reaction step in the primary main metabolic degradation pathway for threonine. Subsequent to TDH catalysis, KBL converts the obtained 2-amino-3-ketobutyrate, with the participation of cofactor CoA, to glycine and acetyl-CoA (see §1.17). These two enzymes act as a complex, most likely containing one tetramer of TDH and two dimers of KBL (84 kDa each dimer), which is perfectly consistent with the highly reactive nature of 2-amino-3-ketobutyrate that can undergo spontaneous decarboxylation. Its structure was determined by X-ray crystallography (PDB 1FC4) (see fig. 1.22) (Schmidt, Sivaraman et al. 2001).

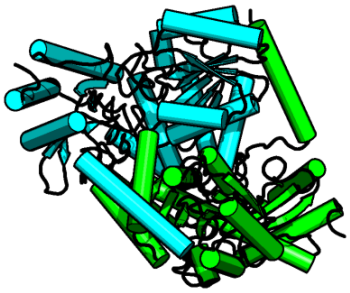


Figure 1.22: Structure of 2-amino 3-ketobutyrate CoA ligase enzyme (PDB 1FC4)

1.3.2.3 Grafted Methyl group

The lack of methyl group-containing residues in a targeted region is detrimental for the NMR study of high molecular weight proteins using selective protonation and isotope labeling of methyl groups. In contrast, the NMR signals overlap resulting from a high number of available probes can also hamper this study. In some cases, the introduction of only few probes can be sufficient to address the biological question being examined. The formerly mentioned work performed on the *Thermoplasma acidophilum* proteasome is a good example (Religa, Sprangers et al. 2010). Indeed, the providential existence of two dynamic Met probes (the third was artificially introduced by mutagenesis) in the region of interest, allowed the accomplishment of an NMR study that led to a better understanding of proteasome gate regulation. The possibility of grafting a methyl group at a desired site in the studied particle is discussed below.

The concept of introducing an artificial probe at a given site to perform a targeted investigation is not new. This probe insertion was concretized mostly by mutation and sparsely by Cys/Lys chemical modification (Jentoft and Dearborn 1983, Bokoch, Zou et al. 2010, Religa, Ruschak et al. 2011, Liu, Horst et al. 2012, Hattori, Furuita et al. 2013). The latter strategy relies on the natural availability of Cys and Lys residues in proteins. Therefore, its utilization could be easier and more practical than the mutagenesis-based approach. However, the success of this approach is tightly related to

Cys/Ly solvent accessibility, which definitively modulates the occurrence of the chemical modification reaction. The combination of both techniques, mutation and chemical modification, can also be considered, as presented below.

1.3.2.3.1 Grafted methyl group on cysteine

Generally, Cys is not abundant in proteins (2%). However, this residue can particularly enrich some interfaces, such as those interfaces between protease and an inhibitor (Lo Conte, Chothia et al. 1999). Furthermore, Cys residue has been extensively used for the introduction of spin labels. Through the oxidation of its SH group, several "labeling moieties" could be "stuck", e.g., nitroxide or metal ion paramagnetic labels. Similarly, $^{13}CH_3$ can be attached to Cys by chemical treatment using [^{13}C]-methylmethanethiosulfonate (MMTS; IUPAC name, methylsulfonylsulfanylmethane)(Kenyon and Bruice 1977, Religa, Ruschak et al. 2011). The MMTS molecule reacts with free cysteine residues in proteins to produce (S)-methylthiocysteine (MTC). This latter molecule, which is structurally similar to MET (see fig.1.23), occupies the same volume, suggesting a high probability of obtaining intense $^{13}CH_3$ signals that resonate in an uncrowded region of [^{1}H,^{13}C] correlation spectra. Recently, ^{13}C-MMTS was used to investigate the gating mechanism of both *Thermoplasma acidophilum* proteasome and the *E. coli* ClpP protease (Religa, Ruschak et al. 2011).

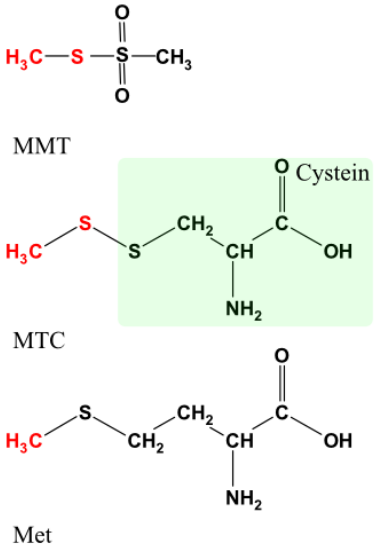
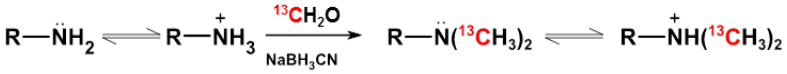


Figure 1.23: Introduction of an artificially labeled methyl groups through the MMTS molecule. MMT: methylsulfonylsulfanylmethane, MTC: (S)-methylthiocysteine, Met:Methionine The labeled methyl group is highlighted in red.

1.3.2.3.2 Grafted methyl group on lysine:

Lys occurs more frequently than Cys in an overall protein and in the protein–protein interface (Chakrabarti and Janin 2002). Therefore, its utilization may be more pertinent in some cases. Lys methylation, through labeled formaldehyde, was initially described in the early 1980s and then applied in various studies (Jentoft and Dearborn 1983, Rayment 1997). This modification is performed according to the following reaction:



(eq. 1.2)

This approach was particularly useful for investigating the conformational changes of a specific ligand on the extracellular surface of the β2 adrenergic receptor , which belongs to the G-protein-coupled receptor family (Bokoch, Zou et al. 2010). In this study, from the existing 16 lysine residues, 7 had to be mutated to Arg to decrease the NMR signal superposition. Obviously these mutated Lys residues have been chosen carefully, saving the region of interest, i.e., around the Lys305-Asp192 salt bridge.

CHAPTER
II

COMBINATORIAL LABELING OF METHYL GROUPS

2.1 Context

Selective protonation and labeling at an atomic position is useful for the study of large proteins by NMR. The ability to obtain this selectivity for several amino acid types, within a certain proton density range, would be even more helpful as the increase in the number of probes permits better coverage of the studied particle.

Currently, protocols for the selective labeling of methyl-containing residues are available for each of them (nine of methyl groups). Having all the methyl groups simultaneously labeled would allow monitoring of 30-40% of the amino acids in targeted proteins.

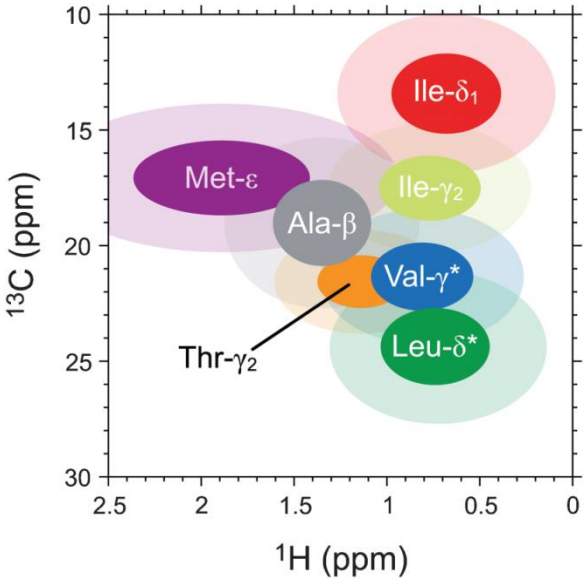


Figure 2.1: Expected chemical shifts of residue-specific methyl groups. From (Plevin and Boisbouvier 2012)

In the literature, various works have reported the labeling of more than one single amino acid to tackle the studied question (Gelis, Bonvin et al. 2007, Amero, Schanda et al. 2009, Velyvis, Schachman et al. 2009, Esposito, Sankar et al. 2010, Godoy-Ruiz, Guo et al. 2010, Religa, Sprangers et al. 2010, Yang, Welch et al. 2010, Karagoz, Duarte et al. 2011, Kato, van Ingen et al. 2011). Most of these investigations rely on diverse combinations of Ile-δ_1, non-stereo-specifically labeled Leu/Val, Ala and Met.

Despite the apparent knowledge of these combinatorial schemes, certain of them are not optimized (see §2.2). Furthermore, the evolution of the methyl group labeling field is ongoing and new molecules are constantly introduced, e.g., 2-acetolactate for the stereo-specific labeling of Leu and Val, 2-hydroxy ethyl butanoic acid for the Ile-γ_2 labeling, L-Thr molecule for the Thr-γ labeling. Therefore, before choosing the residues to combine, several parameters should be taken into account, as discussed below:

- Spectral overlap

Methyl groups resonate in a narrow 2D [^{1}H,^{13}C] NMR spectrum region, within a 17 ppm and 2 ppm range for ^{13}C and ^{1}H, respectively. Severe superimposition of their signals is expected in high molecular weight systems. Therefore, the simultaneous labeling of certain methyl probes is obviously not achievable without signal overlap, requiring a compromise to extract the maximum of information. Generally, the resonance dispersion is residue type-dependent. Figure 2.1 clearly illustrates the expected chemical shift for each type of methyl group. For example, both Met and Ile-δ_1 methyl groups appear in a highly characteristic, separated spectral window, making them easy to combine with any other residues. Nevertheless, depending on the intended NMR analysis, their monitoring along less flexible residues, such as Ala, is not advisable because of their significantly different relaxation properties.

While the Ala, Met and Thr are simple cases for which the decision to employ them or not is easy, the Ile, Leu and Val present a higher complexity due to their double methyl groups containing.

Combinatorial labeling involving residues containing two methyl groups:

The regio- and stereo-specific labeling of residues that contain two methyl groups, typically Ile, Leu, and Val, is obviously more beneficial (see §1.3.3.1.1 and §1.3.3.1.2.). Indeed, the strong ^{1}H-^{1}H dipolar interactions normally present between the two proximal protonated methyl groups ($\leq$ 3 Å) deteriorate the spectral quality due to inherently rapid relaxation. Using one probe per amino acid, permits a decrease in signal overlap and optimizes sensitivity. Consequently, a dilemma arises regarding which specific methyl group to label.

Provided that the combined amino acids appear in minimally separated regions of the 2D NMR spectrum, any Ile, Leu and Val methyl group can be used. However, the Leu and Val pro-*S* methyl groups (δ_2 and γ_2) present naturally broader chemical shift dispersion than pro-*R* groups (London, Wingad et al. 2008). Thus, their analysis is "easier" because there are fewer expected overlaps. Figure 2.2 illustrates the difference in the discussed signal dispersion for the Val residues. For the Ile-γ_2 methyl groups, the

most critical point is the fact that their resonances overlap with those of Ala, Val, Met and Thr, making Ile-δ_1 more pertinent for combinatorial labeling.

In conclusion, Ile-δ_1, Leu-pro-*S* and Val- pro-*S* probes are in general the most appropriate residues for simultaneous labeling with other amino acids.

- Isotope scrambling:

The undesired isotope scrambling previously discussed in §1.3 can be cleverly manipulated. For instance, in the special case of Ile-γ_2 labeling, the isotope leak to Leu and Val at the pro-*R* methyl group (~5%) was reported (see §1.3.3.1.1). Therefore, when a label is intended for Leu and Val along with Ile, it is advisable to combine the pro-*R* prochiral with Ile-γ_2 groups. This scheme would avoid using a greater amount of acetolactate or α-ketoisovalerate, which is necessary for the dilution of the intruder pro-*R* signals in case where the (Leu/Val)- pro-*S* pattern is chosen.

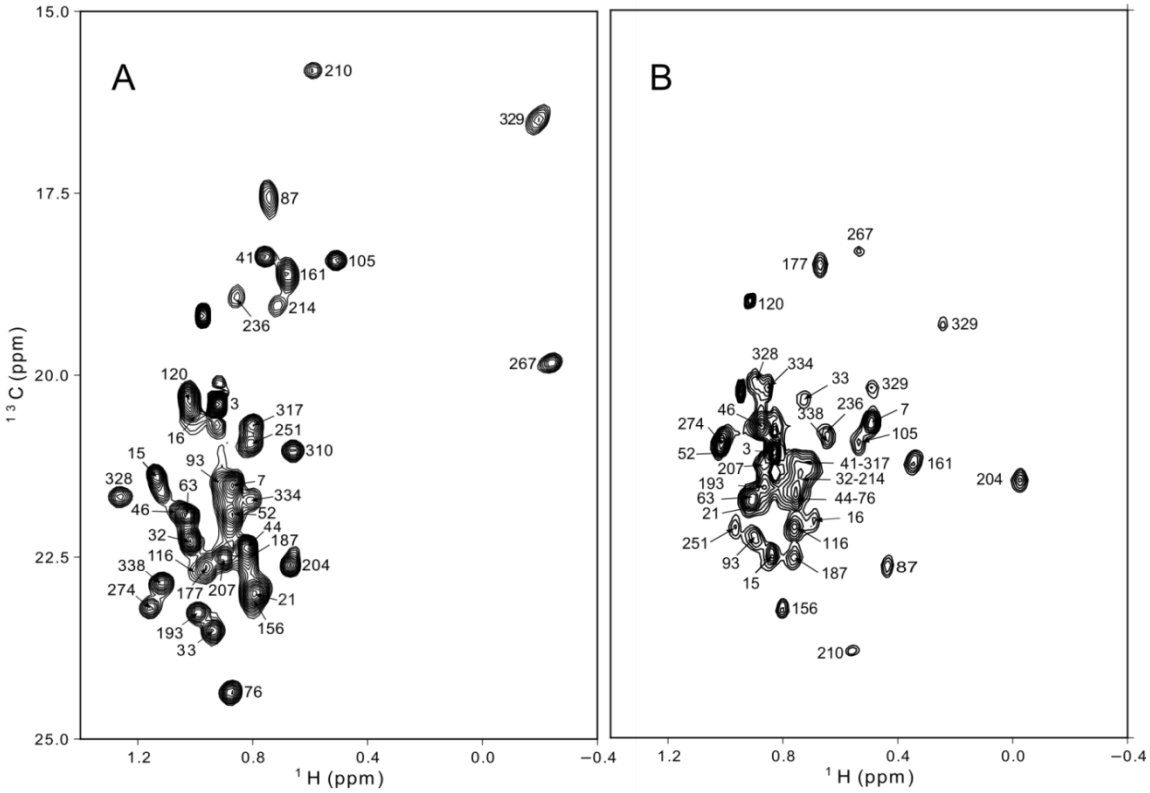


Figure 2.2: Comparison of 2D [U-^{2}H, $^{13}CH_3$-Val^{proS}]-PhTET-2 and [U-^{2}H, $^{13}CH_3$-Val^{proR}]-PhTET-2 NMR spectra. From Mas, Crublet et al., 2013.

Another case, previously described in the literature (Velyvis, Ruschak, et. al., 2012), is the isotope scrambling from Thr to Ile-δ_1 (see §1.3.3.2.2). In *E. coli*, the Thr residue is

endogenously degraded into α-ketobutyrate by TD, which serves as a precursor for the Ile residue. Consequently, the addition of only labeled Thr will also label Ile-δ_1. This choice is extremely practical and may be more convenient than combining the labeling of Thr with Ile-γ_2.

Finally, the "adverse" scrambling, that results from the combination of Ala and Ile labeling is also expected. *In vivo*, [$^{13}CH_3$]-Ala is converted into [$^{13}CH_3$]-pyruvate by several transaminases (see §1.3.3.2.1 and figure 1.8). The simultaneous addition of [$^{13}CH_3$]-α-ketobutyrate generates, through its condensation with the labeled pyruvate, a fraction of Ile residues that are doubly labeled at both δ_1 and γ_2 positions. Because these groups are very close (~3 Å), artifactual NMR signals are expected, which are especially harmful in NOESY experiments (see §2.2).

- Co-incorporation incompatibilities of the precursors used for labeling

Nature preserves the equilibrium of cellular metabolism by tuning either enzyme expression or its kinetic properties. Therefore, enzymes can show different affinities towards their natural substrates. Careful attention is needed when the chosen residues are processed by the same enzymatic machinery, as in the case of Ile, Leu and Val (see fig. 1.8). All precursors of these residues undergo a series of biochemical reactions catalyzed by common enzymes, such as aceto-hydroxy-acid synthase and ketol-acid reductoisomerase (see §1.3.2.1.1). Therefore, the *in vivo* processing of certain of their precursors can be more favored than others, affecting the final isotope enrichment of the targeted residues (Ile, Leu and Val) (see §2.2).

Based on the information depicted in figure 2.1 and considering the aforementioned points, two combinatorial labeling schemes can be proposed: [Ala, Ile-δ_1, Leu-pro-*S*, Val-pro-*S*] and [Met-ε, Ile-δ_1, Thr-γ]. As demonstrated in figure 2.3, optimal peak dispersion was obtained for labeled MSG, where overlaps between the methyl groups were reduced. The superimposition of the Leu and Val signals can be alleviated through the specific labeling of one of the two residues (see §1.3.3.1.2 and see fig. 1.11), which is especially helpful for supra-molecular systems.

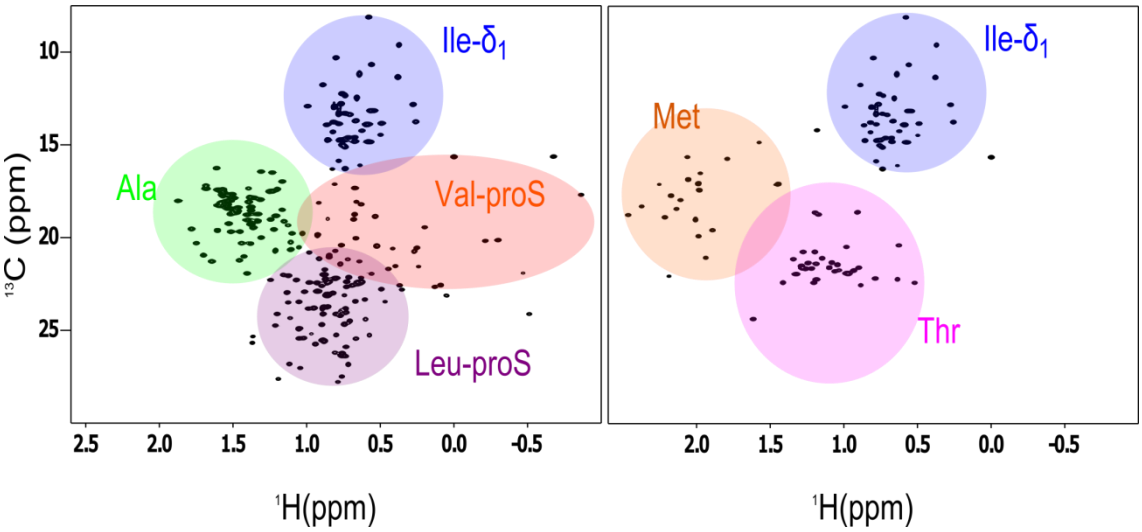


Figure 2.3: 2D [^{1}H,^{13}C] spectra of [Ala-β, Ile-δ_1, Leu-pro-*S*, Val-pro-*S*] and [Ile-δ_1, Met-ε, Thr-γ_2]-MSG (82 kDa) illustrating a scrambling free combinatorial labeling where both sensitivity and resolution are optimized. Spectra were acquired at 37ºC on 800 MHz spectrometer equipped with a cryoprobe. The analyzed samples were 1 mM in a 20 mM MES, pH 7.1 and 25 mM $MgCl_2$ $^{2}H_2O$ buffer.

2.2 Article I: scrambling-free combinatorial labeling of alanine-β, isoleucine-δ_1, leucine-pro*S* and valine-pro*S* groups for the detection of long-range NOEs

The points discussed above in §2.1 are well illustrated in the work presented herein as a a submitted scientific publication. Accordingly, I will describe a novel, cost-effective protocol for the enzymatic synthesis of an Ile precursor, 2-hydroxy-2-[2'-^{13}C-1'-$^{2}H_2$]ethyl-3-keto-4-[$^{2}H_3$]butanoic acid. The utilization of this precursor, instead of α-ketobutyrate, prevents the aforementioned scrambling to Ile-γ_2 when Ala is simultaneously labeled. A solution for the mentioned co-incorporation incompatibility with the Ile and Val/Leu precursors (i.e., acetolactate) will be also presented. The optimized $A^{\beta}I^{\delta 1}(LV)^{proS}$ labeling scheme will be used for the detection of long-range nOes.

Scrambling Free Combinatorial Labeling ofnAlanine-β, Isoleucine-δ1 Leucine-pro*S* and Valine-pro*S* methyl groups for the detection of Long Range NOEs

Authors: Rime Kerfah[a,b,c], Michael J Plevin[d], Ombeline Pessey[a,b,c], Olivier Hamelin[b,c,e], Pierre Gans [a,b,c], Jérôme Boisbouvier[a,b,c,*]

Affiliations:
[a] Univ. Grenoble Alpes, Institut de Biologie Structurale (IBS), F-38027 Grenoble, France
[b] CNRS, F-38027 Grenoble, France
[c] CEA, DSV, F-38027 Grenoble, France
[d] Department of Biology, University of York, York,YO10 5DD, UK.
[e] Univ. Grenoble Alpes, Chemistry and Biology of Metals Laboratory, F-38027 Grenoble, France

* Corresponding author:
Jérôme Boisbouvier
Institut de Biologie Structurale
71 Avenue des Martyrs – CS 10090
38044 GRENOBLE CEDEX 9 – FRANCE
Email: jerome.boisbouvier@ibs.fr

Running head title: Specific Labeling of AILV Methyl Groups

Key words: NMR, Distance Restraints, Structure Determination, Methyl Groups, Specific Protonation, Stereospecific Labeling, Valine, Acetolactate, Large Proteins, NO

Abstract

Specific isotopic labeling of methyl groups in proteins has greatly extended the applicability of solution NMR spectroscopy. Simultaneous labeling of the methyl groups of several different amino acid types can offer a larger number of useful probes that can be used for structural characterisations of challenging proteins. Herein, we propose an improved AILV methyl-labeling protocol in which L and V are stereo-specifically labeled. We show that 2-ketobutyrate cannot be combined with A and 2-acetolactate (for the stereo-specific labeling of L and V) as this results in co-incorporation incompatibility and isotopic scrambling. Thus, we developed a robust and cost-effective enzymatic synthesis of the isoleucine precursor, 2-hydroxy-2-[2'-^{13}C-1'-$^{2}H_2$] ethyl-3-keto-4-[$^{2}H_3$] butanoic acid, as well as an incorporation protocol that eliminates metabolic leakage. We show that application of this labeling scheme to a large 82 kDa protein permits the detection of long-range ^{1}H-^{1}H NOEs between methyl probes separated by up to 10 Å.

1. Introduction

It is unequivocally recognized that strategies for the specific isotopic labeling of methyl groups in proteins have substantially extended the applicability of solution NMR spectroscopy. Indeed, these advances have permitted solution NMR studies of supra-molecular complexes (> 100 kDa), which were previously inaccessible to this technique (Ruschak, Religa et al. 2010, Plevin and Boisbouvier 2012, Rosenzweig and Kay 2014). While such protein systems remain challenging for NMR structure determination, there are a growing number of elegant NMR studies of dynamics (Sprangers, Velyvis et al. 2007, Religa, Sprangers et al. 2010), interactions (Sprangers and Kay 2007, Amero, Asuncion Dura et al. 2011) and function (Ruschak, Religa et al. 2010) of such systems; all of which have benefited from specific methyl group labeling technology.

The first report of methyl selective labeling concerned the δ_1-methyl group of isoleucine (I) (Gardner and Kay 1997). In this protocol, 2-keto, 3,3-[$^{2}H_2$],4-[^{13}C]-butyrate was the sole source of protons added to perdeuterated culture media, to generate a [*U*-^{2}H], I-[$^{13}CH_3$] labeled overexpressed protein. Subsequently, numerous protocols and an array of precursors have been developed to label the remaining methyl-containing amino acids, including: leucine (L) and valine (V) labeling with either 2-keto-isovalerate (Goto, Gardner et al. 1999, Gross, Gelev et

al. 2003, Lichtenecker, Ludwiczek et al. 2004, Tugarinov and Kay 2004); 2-ketoisocaproate (Lichtenecker, Weinhaupl et al. 2013) or acetolactate (Gans, Hamelin et al. 2010, Mas, Crublet et al. 2013) for either stereospecific or non-stereospecific labeling of L and V; alanine (A) (Isaacson, Simpson et al. 2007, Ayala, Sounier et al. 2009); the γ_2 methyl group of isoleucine using 2-aceto-hydroxy butanoic acid (Ruschak, Velyvis et al. 2010, Ayala, Hamelin et al. 2012); methionine (M) (Fischer, Kloiber et al. 2007, Gelis, Bonvin et al. 2007); and recently threonine (T) (Velyvis, Ruschak et al. 2012).

Working with [$^{13}CH_3$]-labeled methyl groups in a perdeuterated background is a prerequisite for detecting high-quality NMR spectra of large proteins. A side effect of this strategy is the considerable loss of structural information due to the low number of remaining protonated probes. Methyl-containing residues represent 30-40% of the amino acids in proteins and they are particularly abundant in the hydrophobic cores of folded proteins. Therefore, simultaneous labeling of the methyl groups of several amino acids represents an obvious route for increasing the number of NMR-visible sites from which useful structural restraints can be obtained. Various combinatorial methyl labeling strategies have been reported, including ILV(Gross, Gelev et al. 2003, Lichtenecker, Ludwiczek et al. 2004, Tugarinov, Choy et al. 2005) and MILV(Gelis, Bonvin et al. 2007). For the ILV scheme, most studies reported the use of 2-ketoisovalerate and 2-ketobutyrate to label LV and the $I^{\delta 1}$ methyl groups, respectively. Because of the high abundance of alanine in proteins, this residue was later added to the ILV ensemble to further increase the measureable number of distance constraints (Godoy-Ruiz, Guo et al. 2010). However, 2-ketoisovalerate leads to racemic labeling of the LV methyl groups such that each prochiral methyl group is protonated at 50%. Consequently, the intensity of inter-methyl group NOEs will be reduced by a factor of 4. Furthermore, the labeling of all methyl groups of leucine and valine generates two NMR-visible sites per residue and thus renders the analysis of the NMR 3D and 4D matrices more complex and time consuming.

In this work, we propose an alternative AILV labeling scheme, in which L and V are stereo-specifically labeled. Additionally, we demonstrate that combining diverse precursors for the simultaneous labeling of AILV can lead to "cross-talk" in the metabolic pathways. We verify that adding 2-ketobutyrate in combination with alanine and 2-

acetolactate results in co-incorporation incompatibility and isotopic scrambling, which affects the quality of the prepared samples. Indeed, *E. coli* enzymatic machinery preferentially processes I precursors over 2-acetolactate. Therefore, these metabolites cannot be simultaneously added to the bacterial culture. Furthermore, 2-ketobutyrate cannot be used in combination with alanine as this leads to isotopic scrambling at $I^{\delta 1}$positions (Ayala, Sounier et al. 2009).

Herein, we suggest a modified protocol to enhance 2-acetolactate incorporation in the presence of isoleucine precursors. We also present a robust and cost-effective enzymatic synthesis of an alternative candidate for the $I^{\delta 1}$ labeling: 2-hydroxy-2-[2'-^{13}C-1'-$^{2}H_2$] ethyl-3-keto-4-[$^{2}H_3$]butanoic acid. This precursor can be successfully combined with [$^{13}CH_3$] alanine to obtain $I^{\delta 1}$ and A^{β} methyl probes with no detectable isotopic scrambling. Finally, we demonstrate that our proposed AILV labeling scheme is a useful tool for the detection of long-range ^{1}H-^{1}H NOEs between methyl probes separated by up to 10 Å in large proteins.

2. Materials and Methods

2.1 Preparation of aceto-hydroxy-acid synthase II (AHAS II):

The overexpression and purification of AHAS II (ALS II) followed the protocol of D. Chipman (Vyazmensky, Sella et al. 1996). *E. coli* BL21(DE3) cells carrying the plasmid containing the AHAS II gene were grown at 37°C in Luria Broth (LB) media. When the O.D. (600 nm) reached 0.5-0.7, AHAS II expression was induced by the addition of IPTG to a final concentration of 0.4 mM. Expression was performed for 12 hours at 20°C. The cells were harvested by centrifugation at 5000 g for 15 min at 4°C, ressuspended in 10 ml of 0.1 M TRIS-HCl, pH 7.5, and centrifuged at 4000 g for 15 min at 4°C. The cells were ressuspended in 10 ml of buffer A (50 mM TRIS, pH 8, 0.5 M KCl, 10 mM imidazole and 20 µM FAD). The cells were disrupted by sonication for 2 min and the insoluble materials were removed by centrifugation at 45000 g for 45 min at 4°C. The supernatant was passed over a Ni-NTA column pre-equilibrated with buffer A. After washing with 5 equivalent volumes of buffer A, the protein was eluted using buffer B (50 mM TRIS, pH 8, 0.5 M KCl, 400 mM imidazole and 20 µM FAD). The fractions containing AHAS II were pooled, concentrated (55 mg/L), dialyzed against pure water and lyophilized. The activity of AHAS II was determined by measuring the

decrease in absorbance at 333 nm of pyruvate.

2.2 Synthesis of 2-hydroxy-2-[2'-^{13}C-1'-D_2]ethyl-3-keto-4-[D_3] butanoic acid:

The synthesis of 2-hydroxy-2-[2' ^{13}C-1'-$^{2}H_2$]ethyl-3-keto-4-[$^{2}H_3$] butanoic acid was performed following the protocol previously described by D. Chipman (Engel, Vyazmensky et al. 2004). The reaction was initiated by adding an aliquot of AHAS II at 6 μM (420 ng/ml) to an equimolar (33 mM) mixture of deuterated pyruvate (perdeuterated by incubating unlabeled pyruvate in $^{2}H_2O$ at pH 10.7 for 72 hours) and 3,3 [^{2}H],4-^{13}C-2-ketobutyrate in 3 ml of D_2O buffer containing 50 mM potassium phosphate, pH 7.8, 10 mM $MgCl_2$, 1 mM thiamine diphosphate, and 20 μM FAD. This reaction was followed by 1D NMR.

2.3 *Protein expression and purification:*

E. coli BL21(DE3) cells carrying the plasmids for ubiquitin or MSG were progressively adapted to M9/D_2O media containing 1 g/L $^{15}ND_4Cl$ and 2 g/L D-glucose-d_7 (Isotec) in three stages over 24 h. In the final culture, the bacteria were grown at 37°C in M9 media prepared with 99.85% $^{2}H_2O$ (Eurisotop). When the O.D. at 600 nm reached 0.7, a solution containing the labeled precursors was added.

- **For the production of [*U*-^{2}H], I-[$^{13}CH_3$]$^{\delta 1}$, L-[$^{13}CH_3$]^{proS}, V-[$^{13}CH_3$]^{proS} MSG sample**:

2-[$^{13}CH_3$], 4-[$^{2}H_3$] acetolactate at 300 mg/L was added 1 hour prior to induction, 40 minutes later (i.e. 20 minutes prior to induction), 3,3-[$^{2}H_2$],4-[^{13}C]-2-ketobutyrate was added to a final concentration of 60 mg/L. The incorporation of $^{13}CH_3$ isotopomers was 50 % lower when both precursors are added simultaneously.

- **For the production of a [*U*-^{2}H], A-[$^{13}CH_3$]$^{\beta}$, I-[$^{13}CH_3$]$^{\delta 1}$ ubiquitin sample:**

A mixture of 60 mg/L 3,3-[$^{2}H_2$],4-[^{13}C]-2-ketobutyrate, 700 mg/L 3-[^{13}C], 2-[^{2}H] alanine and 200 mg/L *U*-[^{2}H] 2-ketoisovalerate was added 1 h prior to induction, according to Godoy-Ruiz et al. (Godoy-Ruiz, Guo et al. 2010). $^{13}CH_3$ scrambling in the $I^{\gamma 2}$ position was observed using this protocol.

- **For the production of a scrambling-free [*U*-^{2}H], A-[$^{13}CH_3$]$^{\beta}$, I-[$^{13}CH_3$]$^{\delta 1}$ ubiquitin sample:**

A mixture of 60 mg/L 2-hydroxy-2-[2'-^{13}C-1'-$^{2}H_2$]ethyl-3-keto-4-[$^{2}H_3$] butanoic acid, 700 mg/L 3-[^{13}C], 2-[^{2}H]-alanine and 200 mg/L *U*-[^{2}H] 2-ketoisovalerate was added 1 h prior to induction

- **For the production of a scrambling-free [U-^{2}H], A-[$^{13}CH_3$]$^{\beta}$, I-[$^{13}CH_3$]$^{\delta 1}$, L-[$^{13}CH_3$]^{proS}, V-[$^{13}CH_3$]^{proS} MSG sample:**

2-[$^{13}CH_3$], 4-[$^{2}H_3$] acetolactate at 300 mg/L was added 1 h prior to induction. Forty minutes later (20 minutes prior to induction), 60 mg/L 2-hydroxy-2-[2'-^{13}C-1'-$^{2}H_2$]ethyl-3-keto-4-[$^{2}H_3$] butanoic acid and 700 mg/L 3-[^{13}C] and 2-[^{2}H]-alanine were added. Protein expression was induced by a final IPTG concentration of 1 mM. The expression was performed overnight at 20°C for MSG and 3 hours at 37°C for ubiquitin before harvesting. Ubiquitin was purified by Ni-NTA (Qiagen) chromatography in a single step. Malate Synthase G (MSG) was purified initially by Chelating Sepharose chromatography (GE Healthcare) followed by gel filtration chromatography (Superdex 200 pg; GE Healthcare). The typical final yields after purification were 40 and 80 mg/L of methyl-specific protonated ubiquitin and MSG, respectively.

2.3 NMR Spectroscopy

The typical concentrations of ubiquitin and MSG in the NMR samples were 2 and 1 mM, respectively, in a 100% $^{2}H_2O$ buffer containing either 20 mM Tris and 20 mM NaCl at pH 7.4 (ubiquitin) or 25 mM MES, 20 mM $MgCl_2$ and 5 mM DTT at pH 7.1 (MSG).

The 2D (^{1}H,^{13}C) NMR spectra of ubiquitin and MSG were recorded at 37°C on a Varian (Agilent) DirectDrive spectrometer operating at a proton frequency of 800 MHz equipped with a cryogenic triple resonance probehead. The 3D HMQC-NOESY experiment was recorded for 82 h with a 1 mM [*U*-^{2}H], A-[$^{13}CH_3$]$^{\beta}$, I-[$^{13}CH_3$]$^{\delta 1}$, L-[$^{13}CH_3$]^{proS}, V-[$^{13}CH_3$]^{proS} MSG sample and a NOE mixing time of 500 ms. The experiment was recorded with 4 scans per increment and maximum acquisition times of 20 ms in both the ^{13}C and ^{1}H indirect dimensions. All data were processed and analyzed using nmrPipe/nmrDraw (Delaglio *et al.* 1995) and CCPN software (Vranken et al. 2005).

3 Results and Discussion

3.1 Co-incorporation of precursors for isoleucine and stereospecific leucine and valine labeling:

ILV combinatorial labeling has been used as a tool for the study of several biological systems (Gross, Gelev et al. 2003, Lichtenecker, Ludwiczek et al. 2004). Most of the previous reports have used 2-ketobutyrate and 2-ketoisovalerate for the labeling of I$^{\delta 1}$ and the non-stereospecific labeling of the prochiral methyl groups of

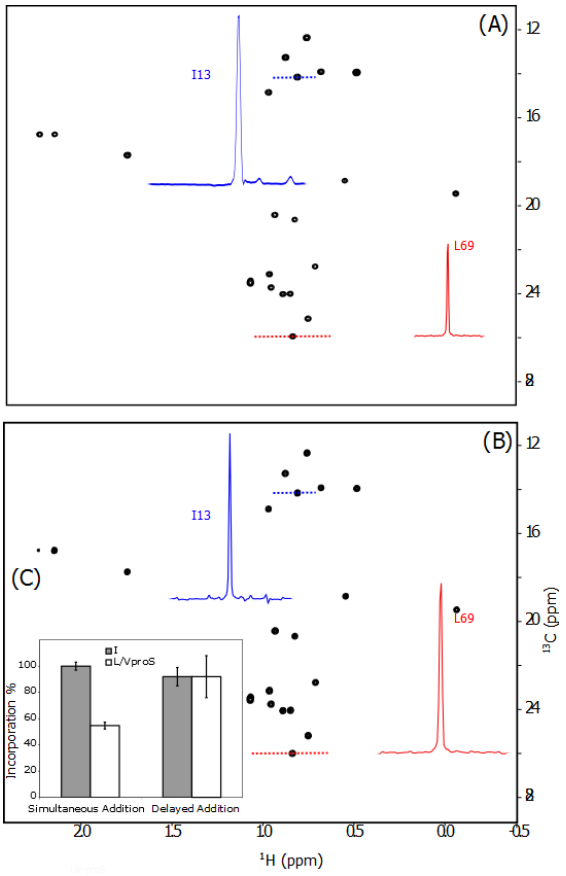


Figure 1: The co-incorporation of $I^{\delta 1}$ and $(LV)^{proS}$ precursors in a ubiquitin sample: 2D HSQC NMR spectra were recorded at 37°C in $^{2}H_2O$ buffer (20 mM Tris, pH 7.4, and 20 mM NaCl) on a 800 MHz NMR spectrometer equipped with a cryogenic probe. The [*U*-^{2}H], A-[$^{13}CH_3$]$^{\beta}$, I-[$^{13}CH_3$]$^{\delta 1}$-ubiquitin sample was prepared using 2-[$^{13}CH_3$], 4-[$^{2}H_3$] acetolactate and 3,3-[$^{2}H_2$], 4-[^{13}C]-2-ketobutyrate. A) Both precursors were added simultaneously to the culture 1 hour before induction (3 h). B) Labeled acetolactate was added to the culture 1 hour before induction (3 h), while the labeled 2-ketobutyrate was added 20 minutes before induction. C) Quantification of the I-δ1 and (LV)-pro-*S* signals in both cases using the Met-ε methyl groups were used as an internal reference. The experimental I-δ1/M-ε and LV-pro-*S*/Met-ε ratios were compared to those obtained for conditions with complete incorporation of each precursor.

L and V, respectively. Depending on the size of the system studied, reducing the number of resonances and resonance overlap can be crucial for interpretation of spectra. The use of 2-acetolactate offers a robust solution for crowded spectra through stereospecific labeling of only one of the prochiral methyl groups of leucine and valine.

In addition to spectral resolution enhancement, stereospecific labeling can increase the intensity of long-range NOEs by a factor of 4 compared to the labeling pattern obtained with 2-ketoisovalerate precursor (Gans, Hamelin et al. 2010).

To test the $I^{\delta 1}(LV)^{proS}$ combination, a culture of ubiquitin was grown as described in 2.3. A 2D (^{1}H,^{13}C) spectrum of this sample of ubiquitin is shown in figure 1A.

Interestingly, signals corresponding to $(LV)^{proS}$ methyl groups were significantly less intense than those of $I^{\delta 1}$ methyl groups. Integration of these signals shows a substantial reduction in isotopic incorporation at $(LV)^{proS}$ sites (Figure 1C). Under our culture conditions, isotopic incorporation at $(LV)^{proS}$ sites was estimated to be 50-60% of that at $I^{\delta 1}$ sites.

Low incorporation rate severely affects the detection of NMR data. Indeed, the intensity of the detectable NOE correlations between the $I^{\delta 1}$ and $(LV)^{proS}$ residues is estimated to be diminished by a factor of 1.66. For NOEs involving L^{proS} and V^{proS} methyl groups, the intensity reduction is expected to be 2.8-fold.

Notably, this co-incorporation incompatibility is not observed with the 2-ketoisovalerate precursor. Thus, the co-incorporation of 2-ketobutyrate and 2-acetolactate had to be investigated.

3.2 Enhancement of the co-incorporation level of 2-acetolactate and the isoleucine precursor:

Thorough analysis of the enzymatic machinery involved in the processing of both 2-ketobutyrate and 2-acetolactate indicates that the ketol-acid reductoisomerase (EC.1.1.1.86), known as KARI, more efficiently processes the isoleucine precursors than those of leucine and valine. KARI presents a 5- to 8-fold higher activity with 2-hydroxy-2-ethyl-3-keto-butanoic acid (isoleucine precursor) than with 2-acetolactate (Dumas, Biou et al. 2001). This information could explain the more efficient incorporation of 2-ketobutyrate compared to 2-acetolactate (Panel 1A). It may also explain why the incorporation of 2-ketoisovalerate, whose enzymatic processing occurs after KARI step, is not affected in presence of 2-ketobutyrate.

Considering this information, we hypothesized that adding 2-ketobutyrate to the culture after 2-acetolactate would promote enzymatic processing of the LV precursor. To test this hypothesis, a ubiquitin culture was prepared in which the two precursors were added at separate times: 2-acetolactate was added 1 hour before induction, while 2-ketobutyrate was added 40 minutes later (i.e. 20 minutes before induction). Quantification of the resulting isotopic-labeling patterns clearly shows that this 2-step approach enhances the incorporation of the 2-acetolactate precursors to more than 90%, without significantly affecting the level of $I^{\delta 1}$ labeling (Figure 1C).

3.3 Isotopic scrambling at the $I^{\gamma 2}$ position in the standard alanine and isoleucine ($A^{\beta}I^{\delta 1}$) labeling scheme:

To achieve a $A^{\beta}I^{\delta 1}(LV)^{proS}$ labeling pattern, 3-[^{13}C], 2-[^{2}H]-alanine must be used in conjunction with 2-[$^{13}CH_3$], 4-[$^{2}H_3$]-acetolactate and 3,3-[$^{2}H_2$], 4-[^{13}C]-2-ketobutyrate. However, the simultaneous availability of 2-ketobutyrate and labeled alanine in the cell is expected to generate an isotopic leak to the $I^{\gamma 2}$ methyl group position (Ayala, Sounier et al. 2009). In *E. coli,* the first common step in the biosynthesis of the branched-chain amino acids (I, L and V) is catalyzed by three aceto-hydroxy-acid synthases (AHAS)(Umbarger 1996, Bar-Ilan, Balan et al. 2001).

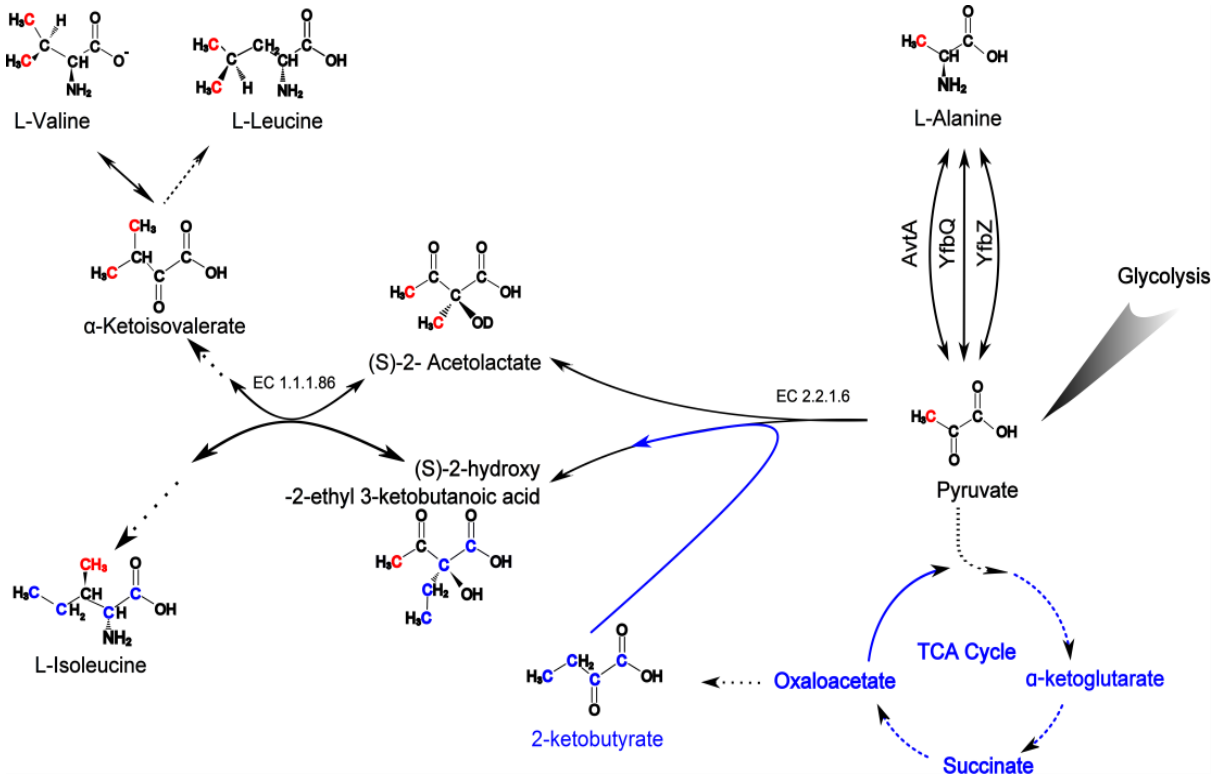


Figure 2: Biosynthetic pathway of the AILV residues. The chemistry of $^{13}CH_3$ group incorporation (in red) from alanine into isoleucine, leucine and valine. The carbons from 2-ketobutyrate are depicted in blue. Each biosynthetic intermediate has been named according to the Kyoto Encyclopedia of Genes and Genomes (KEGG). The enzymes responsible for catalyzing each reaction are indicated by EC number. EC 2.2.1.6: aceto-hydroxy-acid synthase; EC 1.1.1.86: ketol-acid reductoisomerase (for simplification, we indicated only the enzymes referenced in the text). Dashed arrows indicate multistep reactions. Further information on the Ile metabolic pathway can be found online: http://www.genome.jp/kegg/.

These enzymes catalyse two condensation reactions involving pyruvate: (1) reaction of pyruvate with 2-ketobutyrate to produce the I precursor 2-hydroxy-2-ethyl butanoic acid; or (2) reaction of two molecules of pyruvate to producing the L/V precursor 2-acetolactate (fig. 2). The origin of the $I^{\gamma 2}$ methyl group is the pyruvate. The addition of an excess of 3-[^{13}C], 2-[^{2}H]-alanine to the culture medium causes the cellular pyruvate pool to become partially labeled due to the transamination activity of various enzymes (Ayala, Sounier et al. 2009), such as alanine-valine transaminase (AvtA), YfbQ and YfbZ. Deamination of 3-[^{13}C], 2-[^{2}H]-alanine produces 3-[^{13}C]-pyruvate, which is subsequently combined with the added 3,3-[$^{2}H_2$], 4-[^{13}C]-2-ketobutyrate, leading to the appearance of a fraction of isoleucine that is labeled at both the δ_1 and γ_2 methyl groups. To demonstrate this phenomenon, “standard”

$A^{\beta}I^{\delta 1}$ labeling was applied to ubiquitin (see fig. 3). As expected, signals corresponding to the $I^{\gamma 2}$ methyl groups were observed, demonstrating the predicted scrambling pathway. Integration of the crosspeaks revealed that these signals represent approximately 2-5% of the $I^{\delta 1}$ methyl group content. For this calculation, we considered solely the signals of the CH_3 isotopomer because the $^{13}CH_2{}^2H$ and $^{13}CH^2H_2$ isotopomers represent less than 5% of the $I^{\gamma 2}$ $^{13}CH_3$ signals. Similar artifacts were also detected when MSG (82 kDa) was overexpressed using the same protocol (data not shown).

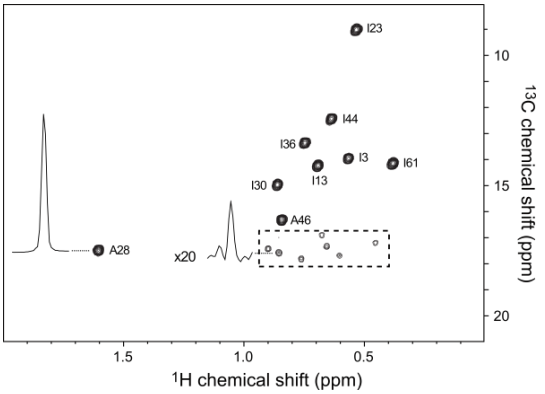


Figure 3: Isotope scrambling to I-γ_2 methyl groups in the standard combinatorial labeling of $A^{\beta}I^{\delta 1}$ in ubiquitin sample. 2D HSQC NMR spectra were recorded at 37°C in ^2H_2O buffer (20 mM Tris, 20 mM NaCl pH 7.4) on a NMR spectrometer operating at a proton frequency of 600 MHz. [*U*-^{2}H], A-[$^{13}CH_3$]$^{\beta}$, I-[$^{13}CH_3$]$^{\delta 1}$ubiquitin was prepared using 3,3-[$^2H_2$],4-[^{13}C]-2-ketobutyrate, 3-[^{13}C], 2-[^{2}H] alanine and *U*-[^{2}H] 2-ketoisovalerate. Spectrum was shown at 10% of the maximal intensity. Boxed part is represented at 1% of the maximum intensity. Signals for β alanine and δ1 methyl carbons of isoleucine were observed in ^{13}C spectra, but also peaks corresponding to the resonance γ2 methyl carbons of isoleucine. The level of isotopic scrambling in γ2 methyl carbons of isoleucine was estimated to be c.a. 2%.

3.4 Suppression of isotopic scrambling at the $I^{\gamma 2}$ position in $A^{\beta}I^{\delta 1}$ labeling:

Low level scrambling to $I^{\gamma 2}$ methyl groups can be neglected for many types of NMR applications (e.g., interaction, dynamics and triple experiments for assignments). However, for structural studies based on the detection and analysis of NOEs, these spurious correlations could cause erroneous assignments and consequently incorrect distance restraints. The intensity of an NOE between two sites is inversely proportional to sixth power of the inter-site distance. Therefore, the intensity of an NOE between a site enriched at 2 % and another at 100 % that are separated by a given distance would be equivalent to the intensity of the NOE from those same sites fully occupied multiplied by $\sqrt[6]{50}$ (i.e., approximately 2). In the case of isoleucine, where the distance between the δ_1 and γ_2 methyl groups is on the order of 3 Å, the expected NOE intensity between a fully-labeled δ_1 site and a fractionally-labeled γ_2 site would correspond to a distance of ~6 Å between two fully-labelled sites. Given that it is possible to detect inter-methyl NOEs for distances greater than 6 Å (Sounier, Blanchard et al. 2007), fractional labeling of γ_2 methyl sites due to isotopic scrambling is unacceptable.

3.4.1 Synthesis of 2-hydroxy-2-[2'-^{13}C-1'-$^{2}H_2$] ethyl-3-keto-4-[$^{2}H_3$] butanoic acid:

To avoid scrambling to $I^{\gamma 2}$ methyl groups it would be necessary to use an alternative I precursor to 2-ketobutyrate. Analysis of the relevant metabolic pathways shows that 2-hydroxy-2-ethyl-3-keto-butanoic acid (see figure 2) may be an alternative option. This molecule has already been used for $I^{\gamma 2}$ labeling (Ruschak, Velyvis et al. 2010, Ayala, Hamelin et al. 2012). In these studies, this precursor was produced chemically and the protocol for its synthesis is well described. However, utilization of the same compound for $I^{\delta 1}$ labeling would require an alternative synthesis strategy that uses [2-^{13}C-1-$^{2}H_2$] ethyl iodide and deuterated acetoacetate. However, neither compound is commercially available. Therefore, we sought to develop an alternative synthesis that could be easily performed in any biochemistry laboratory.

We evaluated whether ALS II could be used to synthesize the required precursor by catalyzing the condensation of pyruvate with 2-ketobutyrate. Using this approach, 2-hydroxy-2-[2'-^{13}C-1'-$^{2}H_2$]ethyl-3-keto-4-[$^{2}H_3$] butanoic acid was produced using ALS II from an equimolar mixture of deuterated pyruvate and 3,3-[$^{2}H_2$],4-[^{13}C]-2-ketobutyrate, according to the protocol described in 2.2. The ^{1}H NMR spectra of the initial and final compounds from the synthetic reaction are presented in figure 4. The reaction reached completion in 2 hours, as shown by the disappearance of the signals from 2-ketobutyrate. The 2-hydroxy-2-[2'-^{13}C-1'$^{2}H_2$] ethyl-3-keto-4-[$^{2}H_3$] butanoic acid appears as a doublet at 0.8 ppm. A fully deuterated 2-acetolactate (not visible by NMR) is also synthesized as a side-product during the reaction. This compound is produced at approximately 10 % of the 2-hydroxy-2-ethyl-3-ketobutanoic acid, as determined in comparable reactions performed with protonated and non-labeled reactants (data not shown).

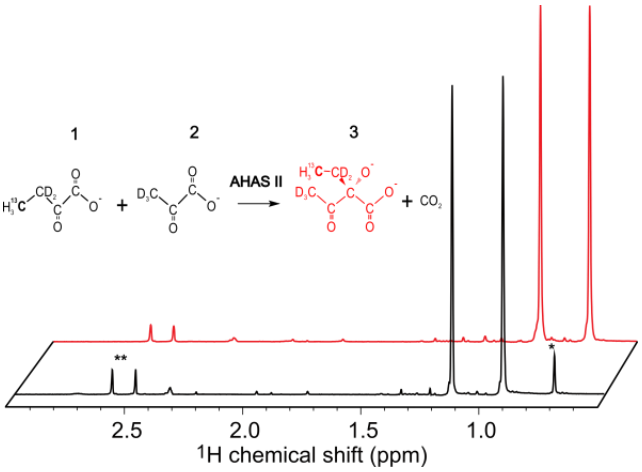


Figure 4: 1D NMR spectra representing the synthesis of the isoleucine precursor. 2-hydroxy-2-[2' ^{13}C-1'-$^{2}H_2$]ethyl-3-keto-4-[$^{2}H_3$] butanoic acid (3) was synthesized by condensation of 3,3-[$^{2}H_2$],4-[^{13}C]-2-ketobutyrate (1) with [*U*-^{2}H]-pyruvate (2) as described in 2.2. In black: 1D spectrum of the reaction solution before condensation (deuterated pyruvate is not visible). In red: 1D spectrum after condensation. * the minor form of 2-ketobutyrate where the 2nd carbon is hydroxylated.** buffer signal.

3.4.2 Incorporation of 2-hydroxy-2-[2'-^{13}C- 1'-$^{2}H_2$] ethyl-3-keto-4-[$^{2}H_3$] butanoic acid:

The ability of the synthesized precursor to suppress undesired enrichment of $I^{\gamma 2}$ methyl groups was tested. Ubiquitin was expressed using the produced compound without any purification (100 mg/L) in conjunction with an excess of labeled alanine (700 mg/L) and [*U*-^{2}H]-2-ketoisovalerate (200 mg/L).

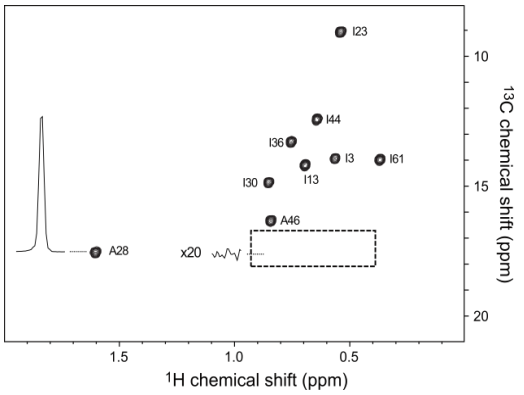


Figure 5: Scrambling-Free Combinatorial labeling of Alanine and Isoleucine-δ_1 in the ubiquitin sample. The 2D HSQC NMR spectrum were recorded at 37°C in a $^{2}H_2O$ buffer (20 mM Tris, pH 7.4, and 20 mM NaCl) on a NMR spectrometer operating at a proton frequency of 600 MHz. The [*U*-^{2}H], [*U*-^{12}C], I-[$^{13}C^{1}H_3$]$^{\delta 1}$ ubiquitin was prepared using 2-hydroxy-2-[2'-^{13}C-1'-$^{2}H_2$] ethyl-3-keto-4-[$^{2}H_3$]-butanoate and 3-[$^{13}CH_3$]-2-[^{2}H]-alanine. The spectrum is shown and was quantified as described in fig.3 The incorporation level of the $^{13}CH_3$ groups in the δ1 position of isoleucine and the β position of Ala was estimated to be higher than 95%, based on the integration of the NMR signals observed in a two-dimensional 2D (^{1}H-^{13}C) spectrum of the labeled proteins. Only the signals for the δ1 isoleucine methyl carbons were observed in the ^{13}C spectra, indicating that the $^{13}C^{1}H_3$ groups of (S)-2-hydroxy-2-[2'-^{13}C,1'-$^{2}H_2$]ethyl-3-keto-[4-$^{2}H_3$]-butanoic acid were not incorporated into the metabolic pathways of the other amino acid precursors, respectively.

NMR analysis of the resulting ubiquitin sample shows that the $I^{\gamma 2}$ methyl group signals were efficiently removed without affecting those of the $I^{\delta 1}$ methyl groups (fig. 5). We then sought to determine the efficiency of precursor incorporation.

A previous protocol for synthesizing 2-hydroxy-ethyl-3-keto-butanoic acid produced a racemic mixture of R and S forms, with only the S form being utilized by the cell (Ayala, Hamelin et al. 2012). The enzymatic synthesis scheme outlined here produces only the S form. Consequently, only half the amount of 2-hydroxy-2-[2'-^{13}C-1'-$^{2}H_2$] ethyl-3-keto-4-[$^{2}H_3$] butanoic acid should be required to achieved maximal $I^{\delta 1}$ incorporation compared to the chemically synthesized precursor. To test this hypothesis, increasing amounts of both compounds were added to different ubiquitin cultures. The resulting obtained incorporation curves are shown in fig. 6.

95% incorporation at $I^{\delta 1}$ sites was achieved using only 60 mg/mL of the enzymatically synthesized compound was 60 mg/L compared to 100 mg/L of the chemically synthesized compound (fig. 6). In addition, the enzymatic synthesis described offers a considerable amount of flexibility in the final labeling pattern. For example, simply

by modifying the initial reactants, this approach can be used to produce

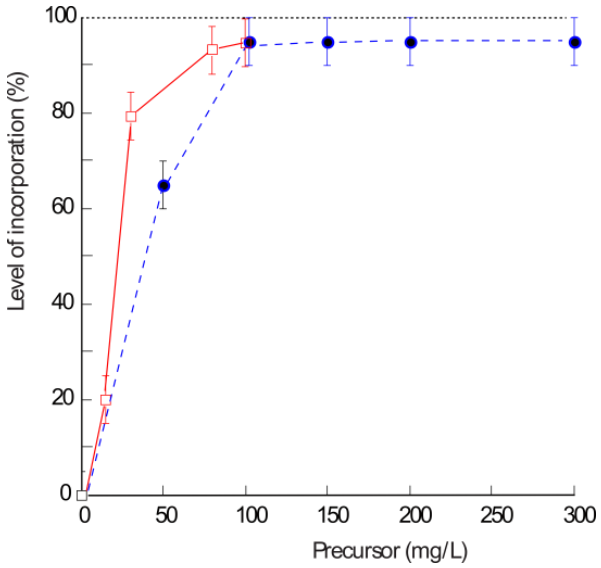


Figure 6. Incorporation level of the chemically and enzymatically synthesized precursors in overexpressed ubiquitin as a function of the amount of exogenous labeled precursor added. The ubiquitin samples were prepared in *E. coli* cultures in M9 minimal media in the presence of increasing amounts of either racemic chemically synthesized precursor (full blue circles) or enzymatically synthesized precursor (open red squares) along with 200 mg/L *U*-[^{13}C]-methionine. An isoleucine side-chain incorporation level of up to 95 % was achieved by adding 100 mg and 60 mg of leucine per liter of M9/$^{2}H_2O$ culture medium for the chemically and enzymatically synthesized compounds, respectively.

precursors for both $I^{\delta 1}$ and $I^{\gamma 2}$ labeling. The same synthetic protocol can also be used for the production of labeled 2-acetolactate, which would allow the labeling of both prochiral methyl groups in leucine and valine.

3.5 $A^{\beta}I^{\delta 1}(LV)^{proS}$ application on MSG for the detection of long-range NOEs:

Protein structure determination by NMR relies on the extraction of a large set of meaningful structural restraints, such as NOEs, dihedral angle restraints and residual dipolar couplings (RDCs). However, while accurate local structural information can be easily achieved, information on overall folding is seriously compromised, particularly for elongated or modular biological systems. Therefore, long-range ^{1}H-^{1}H NOEs are valuable for insight into the global shape of the protein. NOE correlations for ^{1}H-^{1}H distances of 8-12 Å have been previously reported in a perdeuterated protein where a restricted number of methyl groups were selectively protonated (Sounier, Blanchard et al. 2007).

In such an example, the reduction in spin diffusion afforded by the reduced numbers of protons permits accurate distance calibration of long-range NOEs. While advantageous, reducing proton concentration also decreases the number of detectable NOEs and therefore the overall numbers of distance restraints obtainable.

In this study, we were interested in confirming whether these long-range NOEs remain detectable in large perdeuterated systems where a higher number of protonated methyl probes has been introduced. Taking advantage of the optimised $A^{\beta}I^{\delta 1}(LV)^{proS}$ labeling protocol

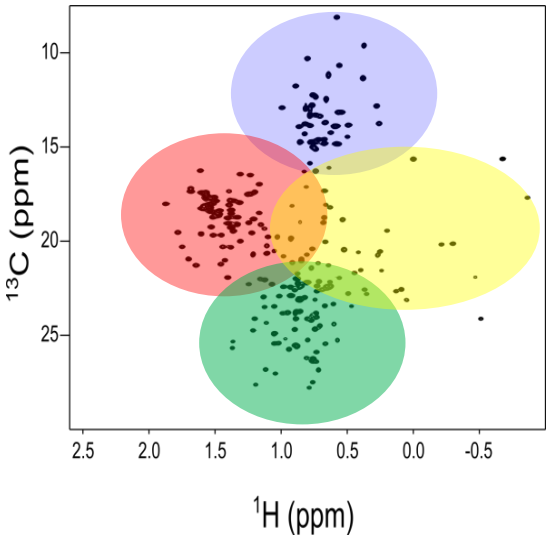


Figure 7: 2D [^{1}H,^{13}C] spectra of [*U*-^{2}H], A-[^{13}CH$_3$]$^{\beta}$, I-[^{13}CH$_3$]$^{\delta1}$, (LV)-[^{13}CH$_3$]^{proS} MSG (82 kDa), illustrating scrambling-free combinatorial labeling where both sensitivity and resolution are optimized. Spectra were acquired at 37ºC on a 800 MHz spectrometer equipped with a cryoprobe. The samples concentrations were 1mM in a 20 mM MES, pH 7.1 and 25 mM $MgCl_2$ $^{2}H_2O$ buffer. In blue circle: $I^{\delta1}$ signals, in red circle: A^{β} signals, in yellow circle: V^{proS} signals and green circle: L^{proS}

described above, a [*U*-^{2}H], A-[^{13}CH$_3$]$^{\beta}$, I-[^{13}CH$_3$]$^{\delta1}$, (LV)-[^{13}CH$_3$]^{proS} labeled sample of MSG was prepared. The 2D (^{1}H,^{13}C) HSQC spectrum recorded of this sample is shown in figure 7, confirming the high quality of the sample prepared and the excellent fidelity of the labeling strategy.

Analysis of a 3D (^{1}H,^{1}H,^{13}C) HMQC-NOESY spectrum indicated that 96% of the expected NOEs between the methyl pairs separated by 2.5-8.5 Å were detected (theoretical distances were predicted from the 1Y0R PDB co-ordinates). Moreover, ~50% of the NOEs arising from methyl groups separated by 9.5-10.5 Å could also be detected. Examples of 2D (^{1}H,^{1}H) strips showing long-range NOEs are presented in Figure 8. Thus, despite higher proton density (14%) than the sample used in previous analyses, a large number of long-range NOEs are still detectable. Likewise a higher number of distance restraints could be determined: out of 2294 expected NOEs within a distance range of 2.5-10.5 Å, 1714 were experimentally observed. These long-range distances will be particularely useful for structure determination of large, elongated proteins and complexes.

3. Conclusion

We have outlined an optimized protocol for scrambling-free $A^{\beta}I^{\delta1}(LV)^{proS}$ labeling that will be widely applicable. Simultaneous labeling of the methyl groups of 4 different amino acids serves to increase the number of useful probes for measuring long-range NOEs in large or challenging proteins. Thanks to this labeling scheme, which optimises both resolution and sensitivity, we could detect long-range ^{1}H-^{1}H NOEs between methyl probes separated by up to 10 Å. This labeling scheme will be useful for structural characterization of large proteins and protein complexes.

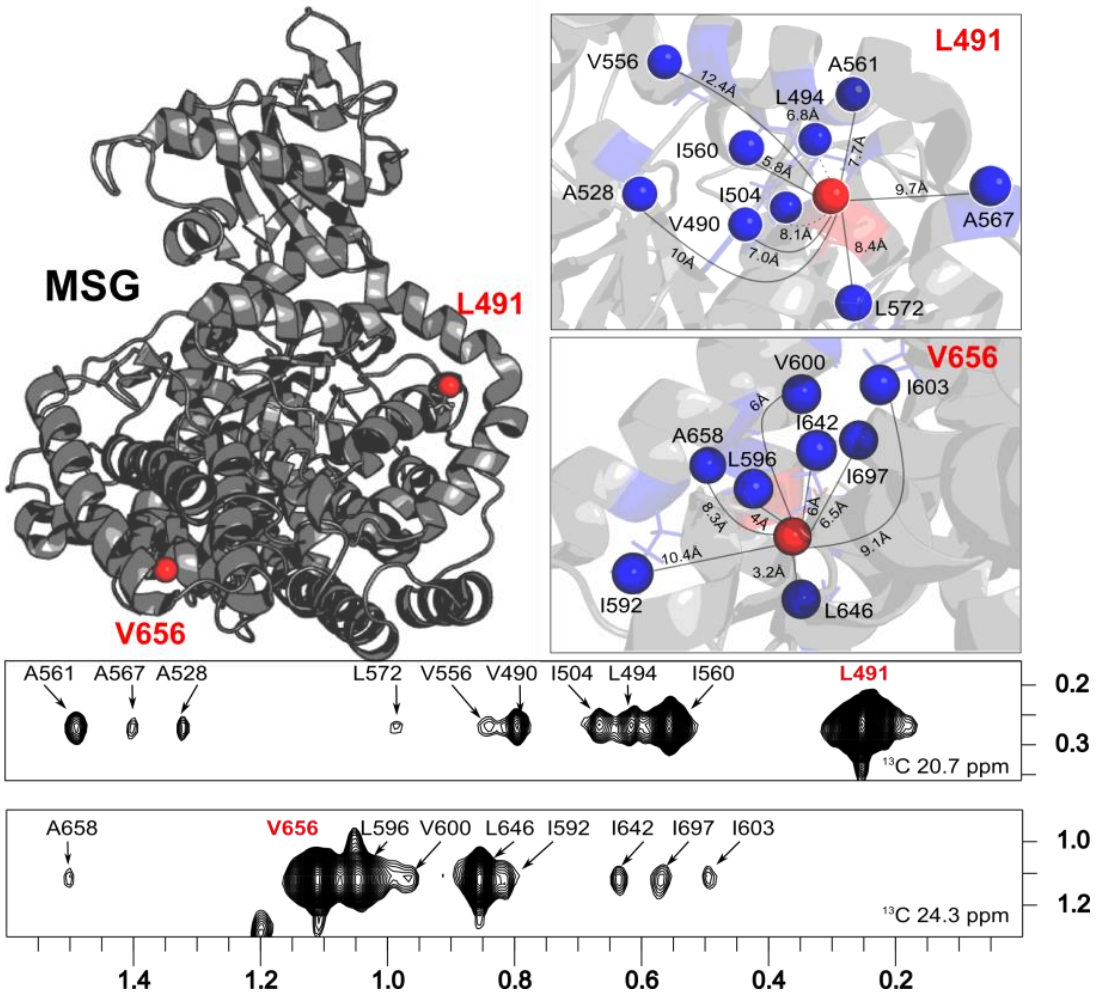


Figure 8: Detection of long-range nOes in the [*U*-D], A-[$^{13}CH_3$]$^{\beta}$, I-[$^{13}CH_3$]$^{\delta 1}$, (LV)-[$^{13}CH_3$]$^{pro\mathit{S}}$-MSG sample (82 kDa). 2D extracts of the 3D ^{13}C-HMQC–NOESY spectrum are presented along with the extracted distances between each pair of methyl groups from the 3D structure of MSG (1D8C). The experiment was acquired on a 800 MHz spectrometer equipped with a cryoprobe at 37°C for 3.5 days with recycling delay d_1= 1.3 s and NOE mixing time τ_m=500ms. The [*U*-^{2}H, $A^{\beta}I^{\delta 1}(LV)^{pro\mathit{S}}$]-MSG sample was at 1 mM in a $^{2}H_2O$ buffer containing 20 mM MES, pH 7.1, and 25 mM $MgCl_2$. Diagonal peaks are highlighted in red.

Acknowledgement

We would like to thank Dr. S.J. Remington for providing MSG plasmid, Dr. D. Chipman for providing ALSII plasmid and Mrs I. Ayala as well as Dr. R. Sounier for stimulating discussions. This work used the high-field NMR and the isotopic labeling facilities at the Grenoble Instruct Centre (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Program FP7/2007-2013 Grant Agreement no. 260887.

References

Amero C, Asuncion Dura M, Noirclerc-Savoye M, Perollier A, Gallet B, Plevin MJ, Vernet T, Franzetti B, Boisbouvier J. 2011. A systematic mutagenesis-driven strategy for site-resolved NMR studies of supramolecular assemblies. J Biomol NMR 50: 229-236.

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

Ayala I, Hamelin O, Amero C, Pessey O, Plevin MJ, Gans P, Boisbouvier J. 2012. An optimized isotopic labelling strategy of isoleucine-gamma2 methyl groups for solution NMR studies of high molecular weight proteins. Chem Commun (Camb) 48: 1434-1436.

Bar-Ilan A, Balan V, Tittmann K, Golbik R, Vyazmensky M, Hubner G, Barak Z, Chipman DM. 2001. Binding and activation of thiamin diphosphate in acetohydroxyacid synthase. Biochemistry 40: 11946-11954.

Clore GM, et al.Dumas R, Biou V, Halgand F, Douce R, Duggleby RG. 2001. Enzymology, structure, and dynamics of acetohydroxy acid isomeroreductase. Acc Chem Res 34: 399-408.

Engel S, Vyazmensky M, Berkovich D, Barak Z, Chipman DM. 2004. Substrate range of acetohydroxy acid synthase I from Escherichia coli in the stereoselective synthesis of alpha-hydroxy ketones. Biotechnol Bioeng 88: 825-831.

Fischer M, Kloiber K, Hausler J, Ledolter K, Konrat R, Schmid W. 2007. Synthesis of a 13C-methyl-group-labeled methionine precursor as a useful tool for simplifying protein structural analysis by NMR spectroscopy. Chembiochem 8: 610-612.

Gans P, Hamelin O, Sounier R, Ayala I, Dura MA, Amero CD, Noirclerc-Savoye M, Franzetti B, Plevin MJ, Boisbouvier J. 2010. Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. Angew Chem Int Ed Engl 49: 1958-1962.

Gardner KH, Kay LE. 1997. Production and Incorporation of 15N, 13C, 2H (1H-d1 Methyl) Isoleucine into Proteins for Multidimensional NMR Studies. Journal American Chemistry Society 119: 7599-7600.

Gelis I, Bonvin AM, Keramisanou D, Koukaki M, Gouridis G, Karamanou S, Economou A, Kalodimos CG. 2007. Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. Cell 131: 756-769.

Godoy-Ruiz R, Guo C, Tugarinov V. 2010. Alanine methyl groups as NMR probes of molecular structure and dynamics in high-molecular-weight proteins. J Am Chem Soc 132: 18340-18350.

Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE. 1999. 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 13: 369-374.

Gross JD, Gelev VM, Wagner G. 2003. A sensitive and robust method for obtaining intermolecular NOEs between side chains in large protein complexes. J Biomol NMR 25: 235-242.

Isaacson RL, Simpson PJ, Liu M, Cota E, Zhang X, Freemont P, Matthews S. 2007. A new labeling method for methyl transverse relaxation-optimized spectroscopy NMR spectra of alanine residues. J Am Chem Soc 129: 15428-15429.

Lichtenecker R, Ludwiczek ML, Schmid W, Konrat R. 2004. Simplification of protein NOESY spectra using bioorganic precursor synthesis and NMR spectral editing. J Am Chem Soc 126: 5348-5349.

Lichtenecker RJ, Weinhaupl K, Reuther L, Schorghuber J, Schmid W, Konrat R. 2013. Independent valine and leucine isotope labeling in Escherichia coli protein overexpression systems. J Biomol NMR 57: 205-209.

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

Plevin M, Boisbouvier J. 2012. Isotope-Labelling of Methyl Groups for NMR Studies of Large Proteins. in J CMaP, ed. Recent Developments in Biomolecular NMR, Royal Society of Chemistry.

Religa TL, Sprangers R, Kay LE. 2010. Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. Science 328: 98-102.

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

Ruschak AM, Velyvis A, Kay LE. 2010a. A simple strategy for (1)(3)C, (1)H labeling at the Ile-gamma2 methyl position in highly deuterated proteins. J Biomol NMR 48: 129-135.

Ruschak AM, Religa TL, Breuer S, Witt S, Kay LE. 2010b. The proteasome antechamber maintains substrates in an unfolded state. Nature 467: 868-871.

Sounier R, Blanchard L, Wu Z, Boisbouvier J. 2007. High-accuracy distance measurement between remote methyls in specifically protonated proteins. J Am Chem Soc 129: 472-473.

Sprangers R, Kay LE. 2007. Probing supramolecular structure from measurement of methyl (1)H-(13)C residual dipolar couplings. J Am Chem Soc 129: 12668-12669.

Sprangers R, Velyvis A, Kay LE. 2007. Solution NMR of supramolecular complexes: providing new insights into function. Nat Methods 4: 697-703.

Tugarinov V, Kay LE. 2004. An isotope labeling strategy for methyl TROSY spectroscopy. J Biomol NMR 28: 165-172.

Tugarinov V, Choy WY, Orekhov VY, Kay LE. 2005. Solution NMR-derived global fold of a monomeric 82-kDa enzyme. Proc Natl Acad Sci U S A 102: 622-627.

Umbarger HE. 1996. Biosynthesis of the branched-chain amino acids. In Escherichia Coli and Salmonella Typhimurium: Cellular and Molecular Biology. American Society for Microbiology Press: pp. 442–457.

Velyvis A, Ruschak AM, Kay LE. 2012. 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 7: e43725.

Vyazmensky M, Sella C, Barak Z, Chipman DM. 1996. Isolation and characterization of subunits of acetohydroxy acid synthase isozyme III and reconstitution of the holoenzyme. Biochemistry 35: 10339-10346.

CHAPTER III

ASSIGNMENT OF METHYL GROUPS

3.1 Context

High-quality NMR spectra in which both sensitivity and resolution are optimized, either by using adapted labeling schemes or NMR pulse sequences, are only meaningful when they are assigned. Without the identification of each NMR signal, structures cannot be calculated and atomic information cannot be extracted.

In the case of small proteins (≤ 30 kDa), a conventional set of multi-dimensional heteronuclear NMR experiments is routinely used for the backbone and side-chain assignments. Generally, the backbone atoms are assigned first and then the lateral moieties are subsequently identified by correlating side-chain signals to the backbone assignments.

Due to the associated relaxation problem, these same pulses cease to be effective for large proteins (≥ 30 kDa). Therefore, the assignment of NMR signals in large proteins has been a challenge for a long time. The advent of the triple labeling schemes (^{2}H, ^{13}C and ^{15}N) as well as TROSY experiment and its variants has helped significantly to alleviate this bottleneck. For example, the backbone assignments of several large proteins, such as the human tumor suppressor protein p53 (67 kDa) (Mulder, Ayed et al. 2000), Malate Synthase G (82 kDa) (Tugarinov, Muhandiram et al. 2002) and 7,8-dihydroneopterin aldolase (110 kDa) (Salzmann M , Pervushin K et al. 2000), have already been reported. Later, after backbone assignment obstacle was overcome for such size range (~100 kDa); the focus was redirected to the side-chains. Moreover, for the study of supramolecular systems (> 100 kDa) by solution NMR, selective methyl group labeling and protonation technology was proven to be a unique available tool. With no possibility of assigning the backbone, alternative strategies had to be developed for methyl group assignments.

3.2 Methyl groups assignment in large proteins

The assignment of methyl groups consist of identifying the residue to which the NMR signal belongs. In the specific case of Ile, Leu and Val, an additional level of information, the regio- and stereospecificity, is also required. To obtain the indicated information in large proteins (≤ 100 kDa), several approaches can be explored:

3.2.1 Methyl groups assignment based on bond-transfer NMR experiments:

Classically, the assignment of the side-chain carbon skeleton of the methyl group-containing amino acids is based on TOCSY pulses, where magnetization is transferred from the methyl groups to the previously assigned backbone (Montelione, Lyons et al. 1992). Nonetheless, due to their non-linearity, the side-chains assignment of Ile, Leu and Val is particularly compromised in large systems (≤ 100 kDa). In fact, the relatively fast relaxation of the NMR signal of these systems, even when they are deuterated, together with the inefficient magnetization transfer at the branch point bifurcations are detrimental for signal sensitivity. As a solution, the linearization of Leu and Val labeling through the use of the [$^{13}CH_3$,$^{12}CD_3$]-α-ketoisovalerate precursor was proposed (Tugarinov and Kay, 2003). A better magnetization transfer is achievable because only one of the two prochiral methyl groups is connected to the backbone through a ^{13}C linear spin system. However, this alternative is still limited because of the obtained racemic scheme where both prochiral methyl groups (pro-*R* and pro-*S*) are simultaneously labeled at 50%, resulting in a loss of sensitivity, increased in signal overlap, and no information regarding stereo-specificity can be extracted (§1.3.2.1.2).

Moreover, the linear labeling of the Ile-δ_1 residue remains unfeasible thus far. As previously explained in §1.3.2.1.1, to link the Ile-δ_1 methyl group to the backbone, the use of [^{13}C]-α-ketobutyrate (the Ile precursor) is not sufficient. The reason is that the C_β atom of Ile arises from pyruvate that derives in turn from the glucose metabolism. Therefore, [*U*-^{13}C] glucose should be added, which leads to simultaneous labeling of Ile-γ_2 as well (see fig. 1.8).

Using the available tools for labeling at the time, a series of sophisticated "out-and-back" COSY-based transfer pulses had to be optimized for each type of residue to assign the ILV methyl groups in MSG (82 kDa) (Tugarinov and Kay 2003).

3.2.2 Methyl group assignment using structure-based methods

Alternatively, automated approaches can also be considered when the 3D structure of the protein target is known. Despite of not being routinely used, an overview of the available softwares is provided. The methyl groups assignment using these softeware can relie on separate or combined NMR data, such as pseudocontact shifts (PCSs), nOes and paramagnetic relaxation enhancements (PREs), as described below.

- Assignment based on pseudocontact shifts data:

Initially, an algorithm called POSSUM[i] was described, which is based on the comparison of the experimental PCSs, induced by a site-specifically bound paramagnetic lanthanide ion, with those back-calculated from the crystal structure (John, Schmitz et al. 2007). This program was used for the methyl groups assignment of a 30 kDa complex where the divalent ions of its active site were replaced by a single lanthanide ion (La^{+3} or Dy^{+3}/Yb^{+3}). The use of several samples labeled with different paramagnetic lanthanides (Dy^{+3} /Yb^{+3}) with different paramagnetic ranges led to the assignment of more than 80% of the Met, Ala, Ile, Leu, Thr and Val methyl groups. These results were enhanced when supplemental stereospecific information was added. To my knowledge, no further work reported the use of this program despite its applicability to deuterated proteins where only the methyl groups are protonated. This can be justified by numerous reasons: (*i*) The differences between the crystal and solution structures generates divergences between the measured and predicted PCSs. (*ii*) Flexible segments and side-chain residues that are not determined in the crystallographic structure cannot be addressed. However, *i* and *ii* are common disadvantages to all 3D structure-based approaches. (*iii*) The application of this approach in a larger protein would give rise to heavier signal overlap, requiring additional input of 3 or 4D NMR spectra from both diamagnetic and paramagnetic samples. (*iv*) Despite the feasible attachment of site-specific lanthanide tags, the metal-binding proteins are still better adapted for this approach. (*iv*) The necessity of $\Delta\chi$ tensors, which are obtained either from known assignments of a subset of cross-peaks or from the [^{15}N,^{1}H] correlation spectra of backbone amides, makes this strategy very laborious and, in the latter case, not suitable for proteins larger than 100 kDa.

- Assignment based on Nuclear Overhauser effect data:

Later, a more practical program with a completely different principle was published, called MAP-XS[j] (Xu, Liu et al. 2009, Xu and Matthews 2013). In this work, the required input consists of the protein crystal structure together with various NMR spectra, which are: a [^{13}C,^{1}H] HMQC (TROSY), an HmCmC experiment correlating the methyl resonances with the directly bonded ^{13}C atoms and a 3D CCH-NOESY, giving the nOes network among

[i] Paramagnetically Orchestrated Spectral Solver of Unassigned Methyl
[j] Methyl Assignment Prediction from X-ray structures

methyl groups. The principle of this strategy relies on predicting the chemical shifts and nOe correlations from the crystal structure and comparing them with the experimental NMR data. Based on the resulting scores, which reflect the match between the predicted and experimental nOes and chemical shifts, an initial assignment is obtained that is subsequently "refined" through an automated assignment-swapping protocol.

In the initial version of MAP-XS, the assignment of up to 90% of Ile, Leu and Val methyl groups in the proteasome using NMR spectra of its α7α7 ring (300 kDa) was achieved when the predicted nOes corresponding to distance cut-off of ~9 Å were used. The use of such a high distance cut-off can be over-fitting and may introduce uncertainties, as long-range nOes in supra-molecular proteins are not expected. This method was reported to be efficient when the experimental nOe data represented at least 50% of the back-calculated data from the crystal structure (Chao, Shi et al. 2012). This condition is very stringent in the case of large proteins where fewer nOes are generally observed. MAP-XS is also limited to ordered regions where a high density of methyl groups is observed, but it fails in regions where methyl groups are sparse and "isolated". To alliviate this latter limitation, combinatorial labeling where the maximum number of methyl-containing residues are targeted, could be useful to reduce this residues isolation. The algorithm also relies on the prediction of chemical shifts, which are not that accurate because the differences between calculated and experimental data can be significantly large. Finally, the softeware seems to depend strongly on the accuracy of the mapping of nOes, which may require 4D experiments that are not very suitable for large molecules (Chao, Shi et al. 2012). Notably, this algorithm was applied to the perdeuterated N-terminal domain of *E. coli* Enzyme I (EIN) (27 kDa), which was labeled at the Ala, Ile-δ_1, Leu, Val and Met methyl groups, afforded the assignment of only 32 out of 140 methyl cross-peaks (Venditti, Fawzi et al. 2011).

Some of the exposed weak points of the initial version of this algorithms seem to be solved in the recently published MAP-XS version (Xu and Matthews 2013). The previous requirement of accurate and manual peak-picking of the nOe correlations was removed. Moreover, a platform to integrate all types of available data (e.g., RDC, PRE and PCs) was also added, offering higher applicability to large systems.

Alternatively, other nOe-based softwares, such as NOEnet (Stratmann, van Heijenoort et al. 2009, Stratmann, Guittet et al. 2010) and FLYA (Lopez-Mendez and Guntert 2006, Ikeya,

Takeda et al. 2009, Schmidt and Guntert 2012) designed for the overall protein assignment (not only methyl groups), can be considered.

- Assignment based on combined NMR data:

After the two previous approaches, a methodology combining different types of NMR data arose (Venditti, Fawzi et al. 2011). It involves producing the target protein with a nitroxide spin label introduced at a single position. The distances (up to ~25 Å) between the spin label and NMR active nuclei can be determined from the increased R_2 relaxation rates. Thus, the measured ^1H_M-Γ2 PRE rates are matched to those predicted from the structure. The resulting resonance assignments with the best score are then refined by comparison of to the experimentally observed [$^{13}CH_3$] nOe patterns from the protein structure. This combined PRE/nOe approach led to the complete assignment of Ala, Ile (δ_1) and Met methyl groups while 87% of Leu and Val were stereo-specifically assigned in perdeuterated EIN. Despite its apparent efficiency and predicted utility with high molecular weight proteins, this approach remains laborious because of the special sample preparation (*i.e.*, mutations, spin labeling) and the complex data analysis.

Recently, another algorithm called FLAMEnGO[k] was published using Monte Carlo sampling (Chao, Shi et al. 2012). This program seems to manage efficiently the ambiguous and scarce nOes information, indeed, a lack of up to 70% of the predicted nOes correlations was described to be tolerated for the correct assignment of 70% of MBP methyl groups. It can also incorporate several types of NMR data, including 3D and 4D NOESY (even from amide-methyl contacts) and PREs.

3.2.3 Stereospecific assignment of methyl groups

The stereospecific assignment of Leu and Val prochiral methyl groups is importante for increasing the precision of the calculated 3D structure. This type of assignment is particularly critical in the case of large molecules (~100 kDa), especially when an important part of nOe information is not available due to the prerequisite partial deuteration (Hilty, Wider et al. 2003). In this case, several strategies have been described:

- Stereospecific assignment using fractional labeling:

[k] Fuzzy Logic Assignment of MEthyl GrOups

The first efficient method that has been proposed, relies on the fractional labeling of 10% *U*-[^{13}C] glucose (Neri, Szyperski et al. 1989). This labeling scheme generates the formation of isolated $^{13}CH_3$-pro-*S* groups and coupled ^{13}C-$^{13}CH_3$-pro-*R* groups in Leu and Val. Thus, stereospecific discrimination is achieved through a CT-HMQC experiment where the signals of the prochiral group resonances are inverted due to the existence or not of ^{13}C-^{13}C coupling (Tugarinov and Kay 2004) (see fig. 3.1). Despite its pertinence, this elegant strategy highlights an important drawback related to the low sensitivity, particularly in large proteins. In fact, low isotope enrichment level (5-10%) and high protonation are unavoidably obtained in the overexpressed protein. Indeed, cultures should be carried out in water; otherwise, various isotopomers of the Val and Leu methyl groups are inevitably obtained, exacerbating the deterioration of both sensitivity and resolution of the NMR spectra.

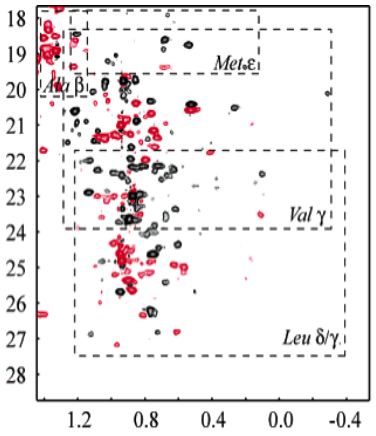


Figure 3.1 ^{1}H-^{13}C CT HSQC correlation spectrum (800 MHz) recorded on protonated and fractionally labeled MSG (82 kDa). ProR (red) and proS (black) prochiral methyl groups. From Tugarinov et al. (2004)

Furthermore, this approach is considered expensive since the fractionally labeled sample, which has to be highly concentrated to compensate for "weak labeling", cannot be exploited for any other NMR experiments.

Later, this low isotope enrichment in both pro-*R* and pro-*S* positions could be enhanced through the 2-[^{13}C]methyl-acetolactate addition to [*U*-^{13}C]-glucose-based cultures (Plevin, Hamelin et al. 2011). An isotope enrichment of 33% instead of 5-10% was reported. However, even with this possible improvement, such strategy is obviously not "useful" for the large proteins (> 100 kDa) due to their protonation condition.

- Stereospecific assignment using stereospecifically labeled 2-acetolactate:

When the incorporation of stereo-specifically labeled Leu and Val in $^{2}H_2O$-overexpressed protein was finally achieved (Kainosho, Torizawa et al. 2006, Gans, Hamelin et al. 2010, Miyanoiri, Takeda et al. 2013), "easy" and non-ambiguous assignment became possible even

in large proteins. The use of commercially available, stereo-specifically labeled Leu/Val is the most straightforward way to achieve their assignment. However, the high cost of these amino acid does not promote their utilization. Therefore, interest was refocused on more affordable stereo-specifically labeled precursors, which currently represent the best alternative. Recently, the Val-pro-*S* and Val-pro-*R* stereospecific assignment was reported using different acetoacetate labeling patterns in a 0.5 MDa protein (Mas, Crublet et al. 2013).

- Methods based on the J^3

In parallel, methods based on the J^3 measurements (Sattler, Schwalbe et al. 1992, Vuister, Yamazaki et al. 1993, Karimi-Nejad, Schmidt et al. 1994, Tugarinov and Kay 2004, Tang, Iwahara et al. 2005) and computational approaches (Pristovsek and Franzoni 2006, Hansen and Kay 2011) have also been proposed. However, they have proven very complex to set-up, entailing a heavy, laborious data analysis. Generally, they are useful for small- and medium-sized proteins, and their application to larger proteins is challenging (Tugarinov and Kay 2004).

Considering the information discussed in this introduction, it is easy to see that the regio- and stereo-specific assignment of ILV is not a straightforward and simple process. In §3.2.4, I describe a practical and cost-effective method for regio- and stereospecific assignment, which can be characterized by simple spectroscopic analysis.

3.2.4 Article II: A single step regio- and stereo-specific NMR assignment of isoleucine, leucine and valine residues in large proteins

In this chapter, I describe a novel method (adapted to systems <100 kDa) that allows simple, practical regio- and stereo-specific ILV assignment using only one sample and utilizing an optimized ^{13}C-^{13}C TOCSY experiments. In this interdisciplinary work, my contribution consisted of developing a protocol that enables the connection of Ile-δ_1 methyl groups to the backbone through a linearized ^{13}C-^{13}C spin system (Patent # 12305984.2, August 2012). This labeling scheme allows an efficient TOCSY magnetization transfer in all three residues (Ile, Leu and Val) simultaneously. The spectra that I analyzed were acquired using an NMR pulse sequence developed and processed by Dr. D. Marion.

A Single Step Regio- and Stereospecific NMR Assignment of Isoleucine, Leucine and Valine Residues in Large Proteins

Rime Kerfah[a],[b],[c], Jérome Boisbouvier [a],[b],[c] and Dominique Marion* [a],[b],[c]

[a] Univ. Grenoble Alpes, IBS, F-38044 Grenoble, France
[b] CNRS, IBS, F-38044 Grenoble, France
[c] CEA, IBS, F-38044 Grenoble, France
* *Biomolecular NMR Spectroscopy Group*
Institut de Biologie Structurale
71 Avenue des Martyrs
CS 10090
38044 GRENOBLE CEDEX 9 - FRANCE

Abstract

A new strategy for the NMR assignment of aliphatic side-chains in large perdeuterated proteins is proposed. It involves an alternative isotopic labelling protocol, the use of an out-and-back ^{13}C-^{13}C TOCSY experiment and an optimized non-uniform sampling protocol. It is long known that the non-linearity of aliphatic spin-system (Ile, Val, Leu) substantially compromises the efficiency of the TOCSY transfers. To permit the use of this pulse scheme, a series of linear precursors were designed and produced to be added to the culture medium and yield linear ^{13}C perdeuterated side-chains with a single protonated CH_3 group. For Val and Leu residue, the topologically different spin-systems introduced for the pro-R and pro-S methyl groups provide stereospecific assignment. In contrast to relayed COSY experiments, no fine tuning of the pulse sequence is required except for the isotropic mixing time. Enhanced resolution and sensitivity have been afforded by the non-uniform sampling and remaining J_{CC} couplings removed by deconvolution prior to the processing by iterative soft thresholding (IST). This strategy has been used on malate synthase G where a large amount of the CH_3 (60 to 90% depending on the residue type) could be correlated directly up to the backbone Cα. It is anticipated that this robust combined strategy could be in the near future routinely applied to large proteins.

Introduction

Resonance assignment is a prerequisite for most NMR studies of biomacromolecules [1]. For proteins, the assignment of the backbone resonance is generally the easy step while the side-chains assignment is more tedious and not always completed. To provide the link to the backbone, J-correlated experiments are preferred to the ambiguous through-space connectivities via nOe. From an historical perspective, it is of interest to note that the J_{HH}-based NMR experiments designed for side-chains assignment have evolved over the years: COSY spectra were first replaced by relayed-COSY[2] and ultimately by TOCSY (Hartman-Hahn polarization transfer) experiments [3]. Despite of the fact that the last two experiments yield similar correlation between non-adjacent spins, the TOCSY pulse scheme, though proposed only a few months later, has nowadays superseded

the relayed-COSY as for being more efficient (the coherence transfer is faster) and easier to implement (no delays need to be adjusted). Obviously, relaxation remains a limit that cannot be lifted by purely spectroscopic means.

When ^{13}C labeled proteins became available, J_{HH} experiments were substituted by similar ones involving J_{CC}. The efficiency of the magnetization transfer is improved due to the occurrence of larger (35 Hz) coupling that are moreover not conformational dependent. Similarly to the proton approach, HCCH-COSY or HCCH-TOCSY [4] experiments were proposed. In a 3D (H)CCH-TOCSY or HC(C)H-TOCSY an entire side-chain can be mapped. Eaton *et al* (1990) [5] have analyzed the time-dependence of the ^{13}C-^{13}C TOCSY transfer in protein side-chains. For uniform coupling constant, the transfer function only depends upon the coupling topologies. For branched aliphatic side-chains such as Leu, Val or Ile, a sharp decrease of the efficiency was reported. Note that these simulations have not included spin-relaxation and mixing times as long as 40 ms are not compatible with fast relaxing spin-systems found in large proteins.

In larger proteins (typically above 50 kDa), sensitivity and resolution become critical and alternate labelling strategies has been designed where the complete protein is deuterated except some CH_3 groups[6,7]. As the non-linearity of branched residues such as Leu, Val and Ile compromises the efficiency of the TOCSY based approaches, Tugarinov *et al*[8,9] have proposed a set of COSY-type experiments to correlate the CH_3 resonances to the backbone. The coherences are step-wise transferred from one carbon to its neighbor using suitably designed building blocks and selective pulses are applied to avoid that the magnetization can go in two separate directions at the branch points.

As proteins contain diastereotopic groups ($C_\beta H_2$, Val- and Leu-CH_3), the standard assignment should be finalized with the stereospecific discrimination[10]. It has been shown that the quality of the NMR based structures dramatically improves with stereospecific assignment. This would be also true for any other dynamic information such as relaxation or exchange rates.

Stereospecific assignment can be obtained either by NMR triangulation[11-14] computational prediction[15] or stereospecific labelling[13,16-18].

For Val, triangulation methods based on the quantitative measurement of $^3J_{HH}$, $^3J_{CH}$, $^3J_{NH}$ and $^3J_{CC}$ have been proposed [19]. This time-demanding method provides unambiguous stereospecific assignment only if a single canonical ($\chi_1 = \pm 60^o, 180^o$) rotameric state is

present. Furthermore, the precise measurement of ^{3}J as small as 1 or 2 Hz becomes unrealistic for high molecular weight proteins due to broader signals.

The prediction-based stereospecific assignment[15] relies on the availability of the 3D structure and its efficiently is tightly dependent on the large number of carefully picked nOes. Assuming the availability of the protein target 3D structure, the abundance and accuracy of the detectable nOes is compromised in large systems.

Among the currently available isotopic labelling strategies, the fractional labelling with 10% of *U*-[^{13}C] glucose[16] was the first reported solution for the stereospecific assignment. This labelling scheme yields an isolated $^{13}CH_3$-*pro-S* groups and a coupled $^{13}CH_3$-*pro-R* groups to the $^{13}C_\beta$ (Val) and $^{13}C_\gamma$ (Leu). Thus, the prochiral groups are easily discriminated in a CT-HMQC experiment by the presence or absence of a ^{13}C-^{13}C J-coupling. This strategy, though efficient, suffers from a low NMR sensitivity owing to the fractional labelling and to the inevitable protonation. This low ^{13}C enrichment could be significantly enhanced using the previously described methyl labelled-acetoacetate (pro-S) [20] together with ^{13}C-glucose[17], whereas the need to perform the protein expression in water cannot be avoided. If the protein over-expression were carried out in $^{2}H_2O$, then some protons of the pyruvate arising from the added glucose would be exchanged[21] . Moreover, several isotopomers of the Val and Leu methyl groups (i.e. $^{13}CH_3$, $^{13}CH_2{}^{2}H$, $^{13}CH^{2}H_2$, $^{13}C^{2}H_3$) would coexist and by detrimental to experimental sensitivity and resolution. Thus, this strategy is not fully compatible with the perdeuteration imperative for high molecular weight proteins. The fractional labelling has been applied to malate synthase G (82 kDa)[13] and due to its poor sensitivity this strategy has to be supplemented by quantitative $^{3}J_{C\gamma N}$ and $^{3}J_{C\gamma C'}$ coupling measurements. In perdeuterated proteins, protonated Leu and Val methyl groups can be obtained using α–ketoisovalerate as precursor. From the obtainable patterns (both $^{13}CH_3$ or a racemic mixture of (LV)-*pro-R* and (LV)-*pro-S*) no stereospecific information is retained. However, the use of the methyl-labeled acetolactate[20] leads to the straightforward prochiral differenciation.

Although this strategy exhibits a higher sensitivity than the fractional labelling approach proposed earlier, some practical limitations remain. Both *pro-R*- and *pro-S*-moieties have the same spin topology and thus, if mixed, the stereoselective information get lost. Thus, duplicate experiments on two samples are necessary.

An alternate approach for assigning stereospecifically branched aliphatic residues has been recently proposed by Mas *et al* [18]: it aims at producing two labelling patterns that can be distinguished spectroscopically. A first sample is labelled with Val-[2,3-$^{2}H_2$; 1,2,3-$^{13}C_3$; [$^{13}C^{1}H_3$]pro-R/ [$^{12}C^{2}H_3$]pro-S] (identified as Val-*pro-R*) and another Val-[2,3-$^{2}H_2$; 3-^{13}C; [$^{13}C^{2}H_3$] pro-R / [$^{13}C^{1}H_3$] pro-S (Val-*pro-Y*). The *pro-R* methyl group is linked to the backbone resonances using the first isotopomer (resulting from the *linear labelling*) and subsequently the *pro-S* CH_3 is connected to *pro-R* CH_3 using the other one (the *Y-shaped labelling*).

In this communication we explore a new strategy to achieve the stereo-specific assignment on a single sample with the minimal number of experiments. The ^{13}C, ^{2}H labelling strategy for Val, Leu and Ile side-chains should meet the following requirements: (a) The entire side-chains are deuterated with the exception of a single protonated CH_3 group; (b) A linear ^{13}C labelling arrangement is required to permit an efficient ^{13}C-^{13}C TOCSY transfer, (c) The spin-systems containing each of the two CH_3 have to be topologically or spectroscopically different. For both Ile-γ1 and -δ2 CH_3, a linear chain to the backbone can be selected. In the case of Val and Leu, the discrimination of the γ1 from γ2 CH_3 and δ1 from δ2 CH_3 is achieved by connecting the *pro-R* prochiral methyl group to the backbone (linear labelling) while the *pro-S* is directly connected to the first one (Y-shaped labelling) [18]. The proposed methodology has been applied on the Malate Synthase G (MSG) (81KDa)[22], a 723-residue monomeric enzyme.

Results and Discussion:

The objective of the current work is to achieve the regio- and stereospecific assignment of all ILV side-chains with a minimal number of samples as well as the shorter experimental time. For this purpose, an original labelling strategy will be combined with TOCSY-based experiments.

As branched ^{13}C spin systems compromise the efficient of isotropic transfers in TOCSY experiment[5], linearized ^{13}C-labelled ILV side chains are required and thus linearized precursors have to be produced (cf. fig 1). While such compounds have been already described for Ile-γ2[23] Val and Leu[9,18,20], analogs for Ile-δ1 are not yet accessible. Here we use the previously developed 2 hydroxy-2-ethyl-3-oxo-butanoic acid and acetolactate molecule to regio- and stereospecifically link the Ile-γ2[23] and the Leu-pro-R and Val-pro-R methyl groups to the backbone[18,20]. Additionally, we benefit from an enzymatic synthesis protocol described elsewhere (Kerfah *et al,* to be published), that yields the Ile-δ1 linearization.

Synthesis of a linearized Ile-δ1 precursor:

A precursor for Ile, the 2-hydroxy-2-ethyl-3-oxo-butanoic acid, can be obtained enzymatically from generic small molecules. As reported elsewhere (Kerfah *et al*, to be published), the enzyme aceto-hydroxy-acid synthase II from *E. coli* was used to condensate a molecule of 3,3 [^{2}H], 1,2,3,4-$^{13}C_4$-2-ketobutyrate with 3,3,3[^{2}H], 2-^{13}C pyruvate (see Experimental section for details). The obtained product of this reaction is labelled as follows: 2-hydroxy-2-[1',2'-$^{13}C_2$-1'-$^{2}H_2$]ethyl -3-oxo-[1,2,3-$^{13}C_3$]-4-[$^{2}H_3$] butanoic acid (compound (3) in fig. 1). By this means, the δ1-CH_3 in Ile is linked to the backbone through a linear ^{13}C spin system while the γ2-CH_3 remains unlabelled.

The availability of linearized labelled ILV side-chains will allow us to substitute the residue-specific COSY-based experiments [8,9] with a TOCSY-based pulse scheme. This experiment is expected to perform well for all residues without much optimization.

Stereospecific labeling schemes for the Leu and Val assignment:

In the previous paragraphs, we have addressed the easy case of Ile where Ile-δ1 and Ile-γ2 can be discriminated with linearized precursors. However, for Val and Leu, the distinction of the two prochiral groups is a more challenging task. Due to the smaller number of measurable ^{1}H-^{1}H nOes in (partially) deuterated proteins, the stereospecific assignment of Leu and Val prochiral CH_3 groups[24] is pivotal.

In the present work, we rather opted for labelling stereospecifically Leu and Val in the presence of a deuterated background. In fact, the stereoassignment approach based on such labelling has been recently reported [18] : this is the most straightforward strategy that remains compatible with the favorable relaxation features associated with deuteration.

Thus, for the key precursor of both Val and Leu acetolactate (also known as 2-hydroxy-2-methyl-3-oxobutanoic acid), two different labelling patterns will be considered: (i) using 2-hydroxy -2-[$^{2}H_3$]methyl- 3-oxo-[1,2,3,4-$^{13}C_4$] butanoic acid, a linear labelling is obtained from the ^{13}CH-*pro-R* down to the backbone ^{13}Cα. (ii) with 2-hydroxy -2-[^{13}C]methyl -3-oxo-[3,3,4,4,4-$^{2}H_5$],[3,4-$^{13}C_2$] butanoic acid, a linear linkage between the ^{13}CH-*pro-S* and the previously assigned ^{13}CH-*pro-R* is generated via the ^{13}Cβ (for Val) and ^{13}Cγ (for Leu). This later labelling pattern (Y-shaped labelling) lead to Val-*pro-Y* and Leu-*pro-Y* respectively (see caption to fig. 1). These two molecules that are topologically and thus spectroscopically discernible can be mixed into a single sample and provide first the assignment of the ^{13}CH-*pro-R* and then that of the ^{13}CH-*pro-S*.

Optimized MSG labeled sample for both regio- and stereospecific assignment of ILV: Using precursors (1) to (4) depicted in fig.1 we can produce a single MSG sample that will contains ILV side-chains with the 6 linearized labelling. To ensure that the all labelling patterns are present in the final protein with the same yield, two conditions should be fulfilled: the precursors should be added at the equivalent concentration in the culture medium and their incorporation should occur with nearly identical kinetics. The in-cell conversion of the precursors into ILV residue involves a common enzyme, the ketol-acid reductoisomerase (EC 1.1.1.86): it was reported[25] to have a 5- to 8-fold higher activity with the Ile precursors than with the Val/Leu precursor. This differential activity, that would lead to a lower incorporation of labelled Val/Leu should be compensated for: the precursors for Ile are added to the medium only 20 mn before induction while

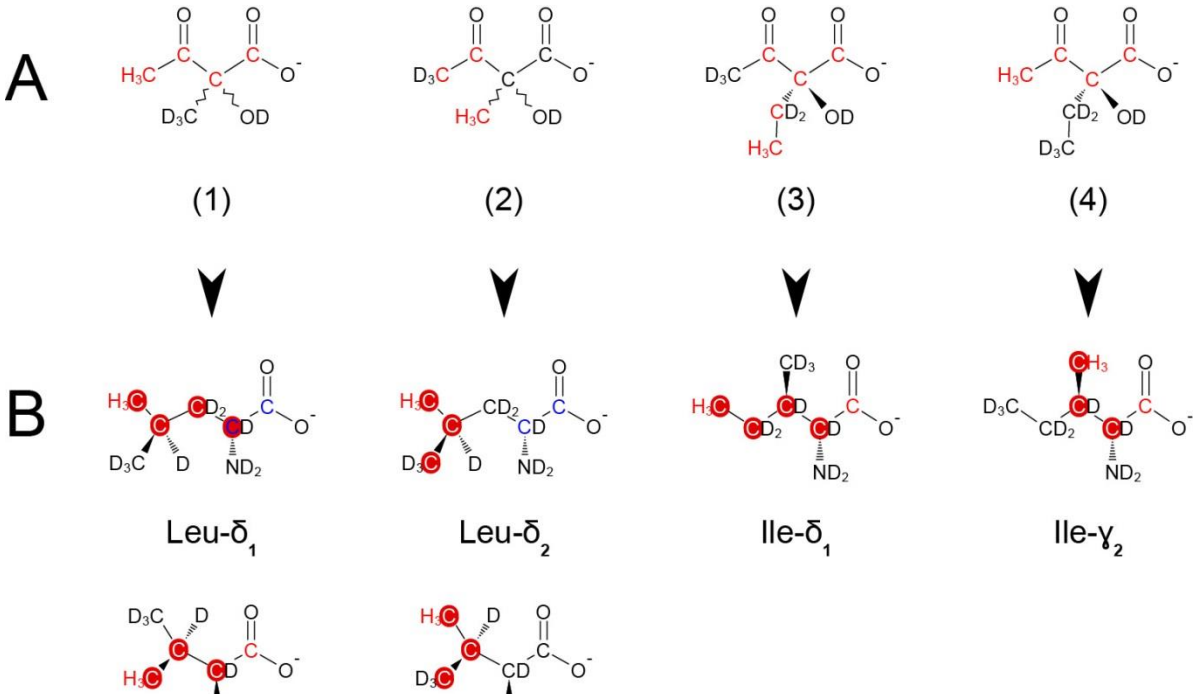


Fig 1: Precursors and resulting residues for the labelling of Leu, Val and Ile side-chains of MSG. In **(A)** are depicted the 2D structure of the used precursors; **(1)** 2-hydroxy-2-[$^{2}H_3$]methyl-3-oxo-[1,2,3,4-$^{13}C_4$] butanoic acid, **(2)** 2-hydroxy-2-[^{13}C]methyl-3-oxo-[4,4,4-$^{2}H_3$],[3,4-$^{13}C_2$] butanoic acid, **(3)** 2-hydroxy-2-[1',2'-$^{13}C_2$- 1'-$^{2}H_2$]ethyl -3-oxo-[1,2,3-$^{13}C_3$]-4-[$^{2}H_3$] butanoic acid and **(4)** 2-hydroxy-2-[$^{2}H_5$]ethyl -3-oxo-[1,2,3,4-$^{13}C_4$] butanoic acid. Protocols for the individual incorporation of compounds (1) and (2) have been reported by Gans *et al* [20] and for compound (3) and (4) by Ayala *et al*[23]. In **(B),** the isotopic labeling patterns obtained for ILV residues. In the figure, these patterns are named according to the $^{13}CH_3$ moiety present in the residue. The following alternate nomenclature has been used in the litterature: Val–γ1 = Val–*pro-R*, Val–γ2 = Val–*pro-S* = Val–*pro-Y*, Leu–δ1 = Leu–*pro-R* and Leu–δ2 = Leu–*pro-S* = Leu–*pro-Y*. ^{12}C carbons are in black, ^{13}C carbons are in red when the labelling stems from the precursors and in blue when provided by the [*U*-^{2}H,^{13}C] glucose added to the culture. The filled red circles identify the 3- or 4-spin systems along which the magnetization is transferred in the out-and-back" HCCH TOCSY pulse sequence (cf fig 2). Linear 4-spin chains are obtained for Leu-δ1 and Ile-δ1 and 3-spin chains for the 4 other residues. Note that except for Val-γ2, the carbonyl carbons are also ^{13}C labelled and their decoupling may improve the spectral resolution of the adjacent Cα.

labelling for each ILV residue as shown in fig. 1.

«Out-and-back» -TOCSY based pulse sequence: The pulse sequence used for assigning and correlating the $^{13}CH_3$ group with the backbone is depicted in fig. 2. A TOCSY start on the protons and end on the same nuclei. An "out-and-back" design is thus preferred although the return transfer has a cost in terms of sensitivity in large macromolecules. In the precursors (1), (3) and (4) (cf fig.1) that will lead to the linear labelling (Ile–δ1, Ile–γ2, Val–γ1 and Leu–δ1), the nucleus that will lead to the carbonyl nucleus is ^{13}C labelled. However, the considerable difference in chemical shift with respect to other ^{13}C would require a rf field strength that is not compatible with the current probe design. Though not part of the TOCSY transfer, the ^{13}CO have been decoupled during the t_2 evolution period to improve the resolution of the Cα.

For a given experimental time, an enhanced digital resolution can be obtained using non-uniform sampling in the indirect dimensions. As discussed in the caption to fig. 2, the sampling schedule has been optimized for the ^{13}C of the methyl groups ($^1J_{CC}$=35Hz).

Comparison with relayed COSY approaches:

pulse train is used to transfer magnetization along the side-chain from the CH_3 to the Cα carbon. As the two labelling patterns proposed earlier contains only ^1H on the methyl groups the magnetization will

As discussed above, TOCSY-based correlations experiments are nowadays exclusively used for resonance assignment of side-chains in non-uniformly labelled proteins. This statement is not completely accurate for uniformly ^{13}C labelled proteins for two reasons: the ^{13}C chemical shift range is larger than the ^1H one and some residues such as Val and Leu are branched. The design of the most recent NMR (cryo-) probes permit the use of strong rf field that cover easily 8 to 10 kHz for isotropic mixing scheme. Thus, the larger chemical shift range is no longer a rationale for using relayed-COSY rather than TOCSY experiment.

Let us now compare the transfer efficiency in these two experiments. Coherences can be transferred between J-coupled spins in two different ways: in the absence of radio-frequency by free J-coupling evolution (relayed COSY) or in the presence of a strong r.f. field by isotropic mixing (TOCSY).

The Hamiltonian for the scalar coupling is given by

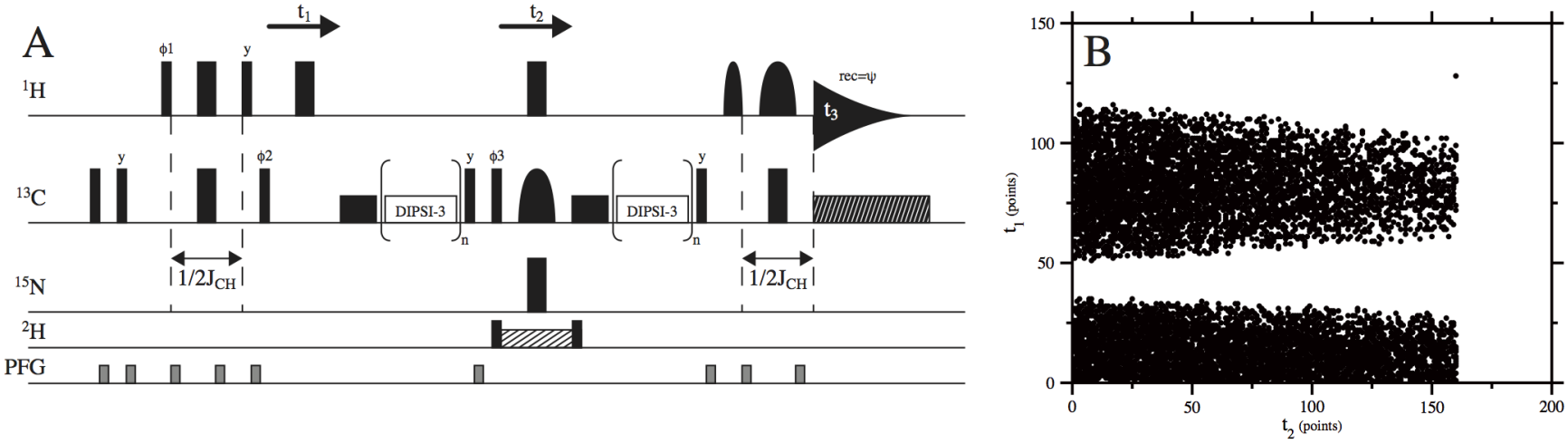


Figure 2: (A) "Out-and-back" HCCH TOCSY pulse sequence for correlating the 13C of the aliphatic side-chains Leu, Ile and Val. For sensitivity reasons, the magnetization starts from ^{1}H and is finally detected on ^{1}H during t_3. Via an INEPT transfer, the proton magnetization on the CH_3 is transferred to ^{13}C and edited during t_1. During a first isotropic mixing step (DIPSI-3), it migrates to other side-chain carbons. The ^{13}C frequency is identified during the t_2 period, while the ^{1}H, the ^{13}CO, ^{15}N and ^{2}H are decoupled. The return path brings the magnetization back to the methyl carbon and then back to protons. For the final INEPT transfer, selective pulses (EBURP2 and REBURP [35]) have been used to minimize the residual water signal present in the sample. All 90° (180°) pulses are indicated by narrow (wide) rectangles and are applied along the x-axis, unless indicated otherwise. The duration of the DIPSI-3 [36] mixing scheme was either 15.5 ms (2-loop experiment) or 23.3 ms (3-loop experiment). The purging of isotropic mixing was achieved by a 1.5ms trim pulse prior mixing and by 90º pulse after the mixing. The respective spectral widths were SW_1 (^{13}C CH_3) = 22 ppm, SW_2 (^{13}C aliph) = 66 ppm and SW_3 (^{1}H)= 12 ppm and the ^{1}H and ^{13}C carriers were set at 4.66 and 37.4 ppm. The ^{13}CO decoupling during t_2 is achieved using a ISNOB5 shaped pulse [37] at the ^{13}CO frequency (175 ppm) covering a bandwidth of 20 ppm. The following phase cycling aiming at the suppression of axial peaks has been used: $\phi1$ = x, –x, $\phi2$ = x, x, –x, –x, $\phi3$= 4(x), 4(–x), rec(ψ) = x, –x, –x, x, –x, x, x, –x. Quadrature detection along t_1 and t_2 is achieved via States-TPPI of phases $\phi2$ and $\phi3$ respectively. (B) Sampling schedule used for the non-uniform sampling of the "out-and-back" HCCH TOCSY. The largest increment used along t_1 (CH_3) and along t_2 (aliphatic carbons) is respectively 128 (= 26 ms) and 160 (= 16 ms). As all methyl ^{13}C are coupled with a single neighbor ($J_{CC} \approx 35$Hz), a cosine-modulated schedule along t_1 has been selected and no data have been collected in the vicinity of increment 43 (14.3 ms = $1/2J_{CC}$). No similar optimization was performed along t_2 as the aliphatic ^{13}C have an unequal number of neighbors (the ^{13}C$\square$ is decoupled from the ^{13}CO). To maximize the experimental S/N ratio, the schedule accounts for the transverse relaxation along t1 and t2 with more points sampled for shorter increment value. The first increments (t_1 or t_2=0) have been omitted as they cannot practically be reached due to the finite duration of both ^{1}H and ^{13}C pulses.

$$\mathcal{H} = 2\pi J_{IS}\ I \cdot S \qquad [1]$$

During free precession periods (such as in relayed COSY), the weak coupling regime applies generally

$|\pi J_{IS}| \ll |\omega_I - \omega_S|$

and only one component of the Hamiltonian remains active

$$\mathcal{H}_f = 2\pi J_{IS}\ IzSz \qquad [2]$$

In contrast, during isotropic mixing (TOCSY), all chemical shifts are eliminated and all terms survive

$$\mathcal{H}_i = 2\pi J_{IS}\,(IxSx + IySy + IzSz) \qquad [3]$$

Inspection of the two expressions evidences that the transfer of coherence is faster under isotropic mixing than under the weak coupling regime ("relayed COSY"). In a TOCSY experiment, an in-phase to in-phase coherence transfer ($Ix \rightarrow Sx$) requires $1/2J_{IS}$; in contrast, in the absence of r.f., an in-phase to anti-phase coherence transfer ($Ix \rightarrow 2IySz$) requires the same time and should be followed by an identical delay for refocusing ($2IzSy \rightarrow Sx$) (see caption to fig 3 for details). As relaxation is a major burden for large macromolecular systems, it can be anticipated that isotropic mixing is more efficient for coherence transfer than free J-evolution.

The linearized side-chains contain several types of linear 3- and 4 spin systems with the following pattern:

$C^1H_3 - C^2 - C^3$ and $C^1H_3 - C^2 - C^3 - C^4$

To transfer the coherence from C^1 to C^3 or C^4, two options are open: a single TOCSY mixing of suitable length or a series of free J-evolution periods separated by 90º pulses. This latter approach has been used by Tugarinov *et al*[8,9] because it still performs well on branched amino-acids. In fig. 3, we compare the efficiency of the TOCSY strategy with that of the relayed COSY *in the absence of relaxation*. For 3-spin and linear 4-spin systems, 76% and 55% of the magnetization reaches the end of the spin system within 18.3 ms and 23.4 ms respectively. In contrast, 100% of the magnetization is transferred with 2 and 3 relays respectively, i.e. in 28 ms and 42 ms. Note that both approaches are of "out-and-back" nature and that the refocusing for the relay COSY occurs in the second half of the sequence. In summary, the TOCSY strategy leads to a maximal transfer (76% and 55%) earlier than the relayed COSY solution. Taking into account that these methods are applied on large proteins that exhibit fast transverse relaxation, it may be advantageous to opt for a slightly less efficient but faster technique. However, if the side-chains would not have been linearized, the relayed COSY should be preferred.

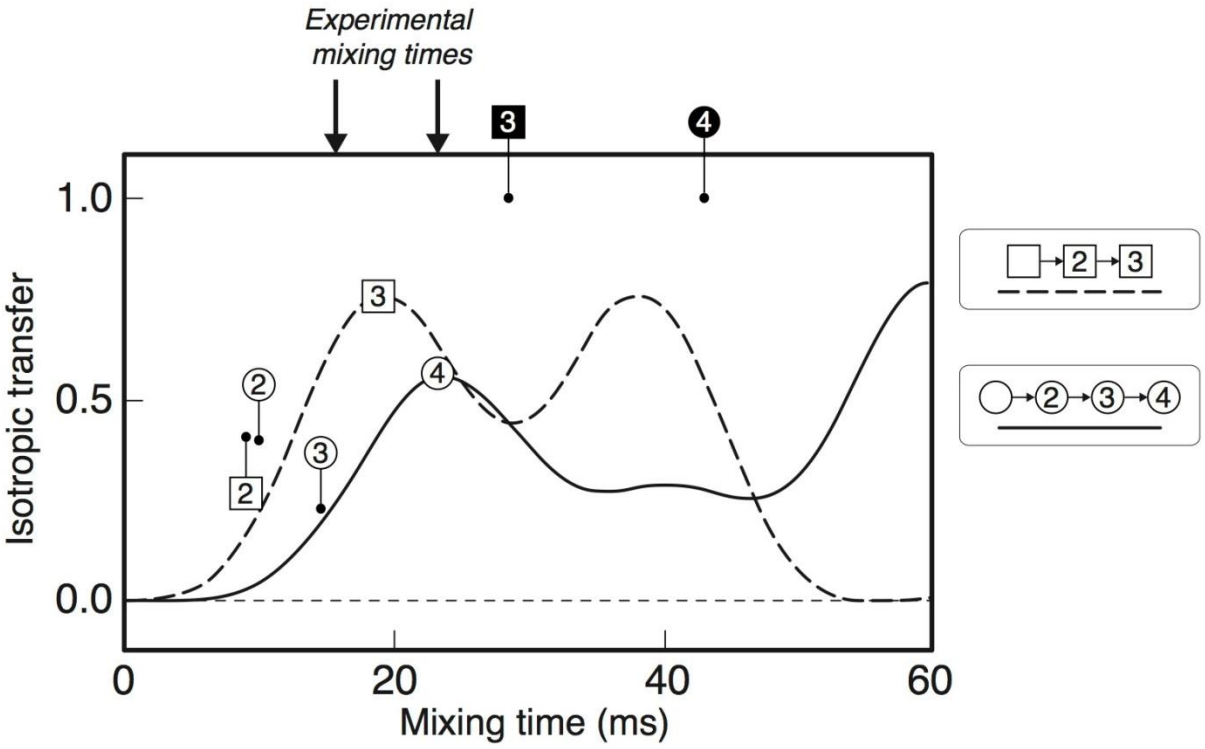


Figure 3: Time dependence of the coherence transfer intensity in isotropic (TOCSY) mixing for various spin systems with uniform J-couplings in the absence of relaxation. Our ILV-labelled protein contains linear 3-spin systems (Ile-γ2, Val-γ1, Val-γ2 and Leu- δ2) identified by squares (□) and linear 4-spin systems (Ile-δ1, Leu-δ1) by circles (○). These plots, adapted from Eaton *et al* (1990) [5], show the full transfer for in-phase coherences (Ix →Sx) from one end (spin 1) to the opposite end of the spin-system (spin 3 or 4). Their first local maxima occur respectively at t= 18.3ms (symbol **[3]** in a square) for 3-spin systems and at t= 23.4ms for 4-spin systems (symbol **(4)** in a circle). The two arrows above the plot indicate the experimental mixing times for the 2-loop experiment (15.5 ms) and for the 3-loop experiment (23.3 ms) (see caption to fig 2). On the same plot, the efficiency of other intermediate transfers have been indicated: the open symbols **[2]**, **(2)** and **(3)** correspond to the first maxima of the [1]→[2] transfer in 3-spin systems (occuring at 9.5 ms) and of the (1)→(2) and (1)→(3) transfer in 4-spin systems (at 10 and 14.3 ms). For the chosen experimental times (15.5 and 23.3 ms), these intermediate transfers are thus smaller than the full transfer (< 15% for 3-spins and < 30% for 4-spins). Alternatively, the coherences in 3- and 4 spins systems can be transferred stepwise in a relayed COSY experiment: for each step, the transfer from an in-phase coherence to an anti-phase coherence (Ix → 2 IxSz) evolves as sin (πJt) leading to a full transfer at t= 1/(2J) = 14ms for $^{1}J_{CC}$. Thus, in the absence of relaxation, a 100% transfer is obtained for 28 ms and 42 ms using relayed COSY (black symbols **[3]** and **[4]**).

To get convinced of the detrimental effect of branched spin-system, the reader should get back to the original work of Eaton et al [5]. In summary,

one can benefit from the simplicity of the isotropic mixing only on suitably linearized side-chains.

ILV methyl groups assignment:

The regio- and stereoassignment of the branched amino acids is the aim of our work. able to label, in a linearized manner, all of Ile–δ_1, Ile–γ_2, Leu–δ_1, Leu–δ_2, Val–γ_1 and Val–γ_2 in MSG. Thanks to this labelling, the TOCSY experiment could be used instead of the relayed COSY approach.

Our approach relies on the principle that Ile–δ_1, Ile–γ_2, Leu–δ_1 and Val–γ_2 methyl groups can be assigned, since they are linked to the backbone in a linear ^{13}C spin system.

Through the matching of its respective Cα and Cβ resonances with the ones previously obtained from the backbone assignment, one should be able to identify each methyl group. Once this step realized and from each identified $^{13}CH_3$-*pro-R* group, the $^{13}CH_3$-pro-S group can be assigned since they are linked through a $^{13}C\gamma$ or $^{13}C\beta$ in Leu and Val, respectively.

Two TOCSY experiments have been recorded on the prepared MSG sample. According to the previously estimated evolution of the coherences[5], two mixing times have been selected. The first one at 18.3ms (2 DIPSI-3 blocks) has been chosen for being optimal for the Cβ detection while the second was at 23.4 ms (3 DIPSI-3 blocks) is more favourable to the Cα detection, i.e. magnetization evolves for a longer time reaching more distant nuclei.

In fig. 4, examples of ^{13}C correlation strips of Ile, Leu and Val side-chains to the backbone are depicted for a mixing time of 23.4 ms. Taken at the frequency of the ^{13}C of the methyl groups in the F_1 dimension, they provide complementary information for the two labelling patterns, the linear one and the Y-branched. For nuclei in linear labelling isotopomers (Ile–δ_1, Ile–γ_1, Leu–δ_1, Val–γ_1), cross-peaks are visible to nearly all nuclei up to the Cα carbon. The choice of this longer mixing time gives rise to stronger Cα cross-peaks but to weaker Cβ peaks. This trend is bolstered up by the CO decoupling which suppresses one of the two ^{1}JCC coupling for the Cα only. On the other hand, in Y-shaped labelling isotopomers (Leu–δ_2, Val–γ_2), the strip at the *pro-S* $^{13}CH_3$ frequency exhibits a cross-peak at the *pro-R* $^{13}C^2H_3$ frequency as well as with the spins in-between. Matching this *pro-R* $^{13}C^2H_3$ frequency with the diagonal $^{13}CH_3$ peak in the linear labelling isotopomers is made difficult due to isotopic shifts. In the Y-shaped labelling isotopomers, the *pro-R* ^{13}C bears 3 deuterium nuclei while it is protonated in the linear labelling isotopomers. This 0.7-0.9 ppm isotopic shift[21] is indicated by blue arrows in fig. 4 for Val and Leu residues.

The two indirect dimensions (F1 for the methyl ^{13}C, F2 for all aliphatic ^{13}C) are not sampled in a constant-time manner and thus the $^1J_{CC}$ (≈35 Hz) coupling with the adjacent carbon(s) remains active throughout the evolution period. Consequently a splitting (doublet for $^{13}CH_3$ and a triplet for other nuclei with two neighbors is anticipated. When the digital resolution increases, such multiplets may become visible and may hamper the peak picking. This issue is less tractable in the F2 dimension (aliphatic carbons may have 1 or 2 neighbors) but also less detrimental due to the larger spectral width (66 ppm). In the F1 dimension (methyl ^{13}C), the suppression of the $^1J_{CC}$

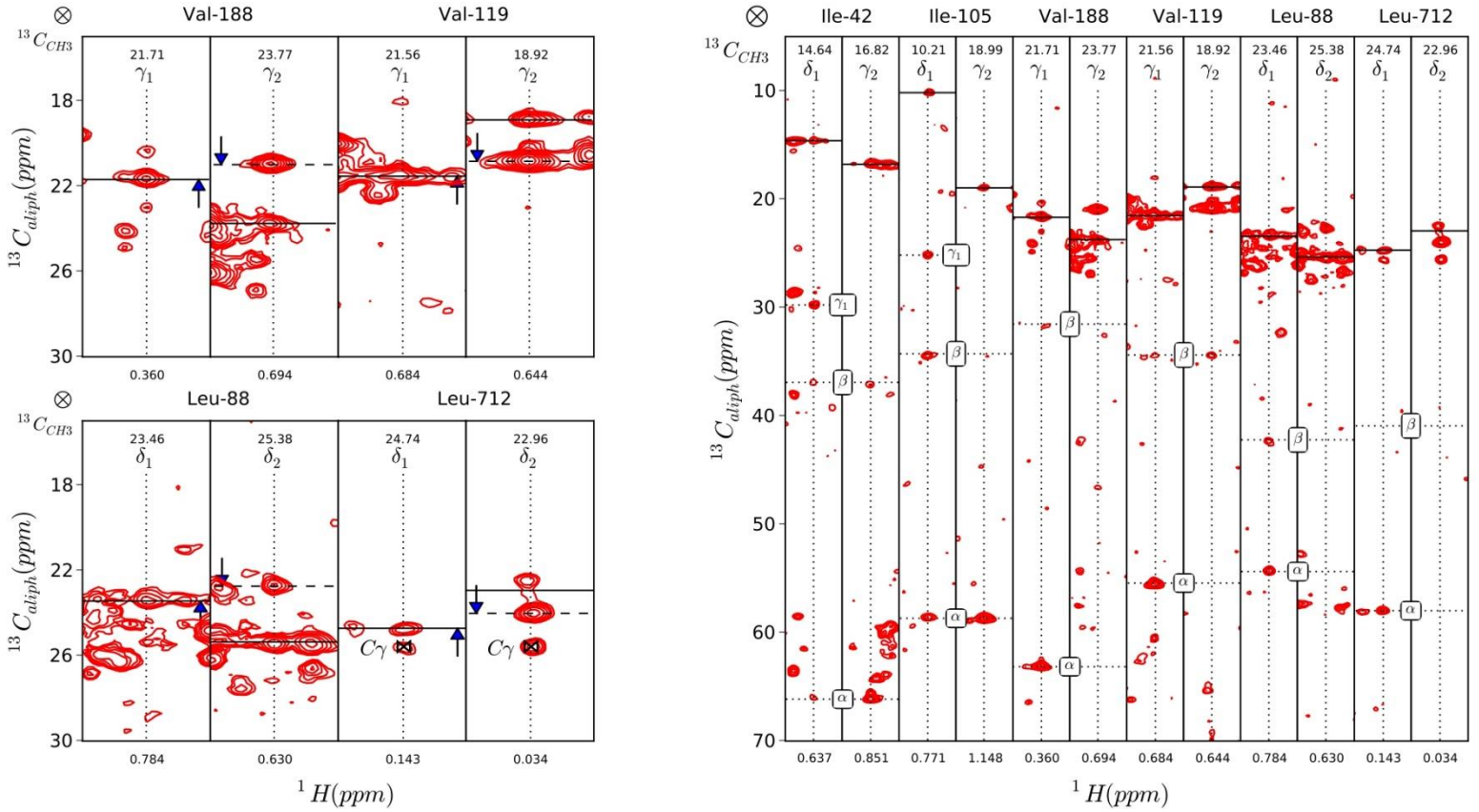


Figure 4: Selected strips from the out-and-back HCCH TOCSY recorded on MSG with a mixing time of 23.4ms. The panel on the left is an expansion of the full plots on the right only for Val and Leu residues. The axis of the 3D data sets are F_1(^{13}C CH3), F_2(^{13}C aliph) and F_3 (^{1}H) as defined in the pulse sequence shown in fig 2. For each residue, two strips taken at the ^{13}C chemical shift (F1) of the methyl carbons (respectively Ile-γ2 and Ile-δ1, Val-γ1 and Val-γ2 , Leu-δ1 and Leu-δ2) are shown. The corresponding frequency is indicated at the top of the strip and the diagonal peak (F1=F2) marked with a continuous line. As a result of labelling scheme shown in fig 1, only one methyl group in each Val and Leu correlates with the backbone (Cα) (Val-γ1 and Leu- δ1) but both for Ile (Ile-γ2 and Ile-δ1). The correlation between the two CH_3 in Val and Leu is illustrated in the left panel: in the strips taken at the Val-γ2 and Leu-δ2 frequency, the correlation with the other CH_3 (Val-γ1 and Leu-δ1) is visible but due to the deuterium isotopic shift ($CH_3 \rightarrow CD_3$), it is shifted by 0.7 ppm (this difference is indicated by blue arrows).

would be feasible at the acquisition stage by means of selective 180º pulse on the ^{13}C adjacent to the methyl group. However on a sample containing 3 different types of labelled residues (Ile, Val and Leu), it is by no means an easy task to design a pulse able to invert all adjacent carbons without perturbing the CH_3.

For lack of generally applicable decoupling method by selective pulses, we have chosen to remove this coupling by deconvolution prior to iterative soft thresholding processing. Fig. 5 compares two cross-peaks (Val–Cγ_1 to Cα) without and with J-removal: when the $^1J_{CC}$ between the Cγ_1 and Cβ is present, the occurence of two signals in the F_1 dimension make any assigment more ambiguous.

The complete analysis of the two 3D TOCSY matrices demonstrated that the Cα correlations have been detected in a proportion of 60%, 86%, 85% and 91% for the Ile-δ_1, Ile-γ_2, Leu-δ_1 and Val-γ_1, respectively. The average signal-to-noise ratio was around 30 for all the residues except for shorted Val residue where it reached 60. Note that signals whose intensity is lower than 5 times fold of noise intensity have been discarded. In agreement with the simulation curves, the amount of detected Cβ resonance is lower than for Cα in the 23.3 ms experiment: however, including the second experiment with shorter mixing time (15.5 ms) up to 80 % of the correlation between the $^{13}CH_3$ groups and the Cβ have been detected. Due to severe overlaps in the NMR spectra of large proteins, it might turn to be necessary to record another experiment with a shorter mixing time to complement the Cα correlations with potentially less overlapping Cβ cross-peaks.

Conclusion

In this paper, a new strategy combining advanced isotopic labelling and NMR spectroscopic tools has been proposed to achieve the regio- and stereospecific assignment of ILV side-chains in larger proteins. An unique ^{13}C-^{13}C TOCSY experiment with enhanced resolution afforded by optimized non-uniform sampling has been used. This experiment, which performs well only on linear spin topologies, requires a protein sample where all side-chain spin systems have been suitably linearized. In addition, we have shown that different labelling patterns could be designed to discriminate all CH_3 ILV groups. First, in the easy instance of Ile, where Ile-δ1 and Ile-γ2 are not topologically equivalent. But also for Val and Leu, where the *pro-R* and *pro-S* methyl groups have being anchored to non-

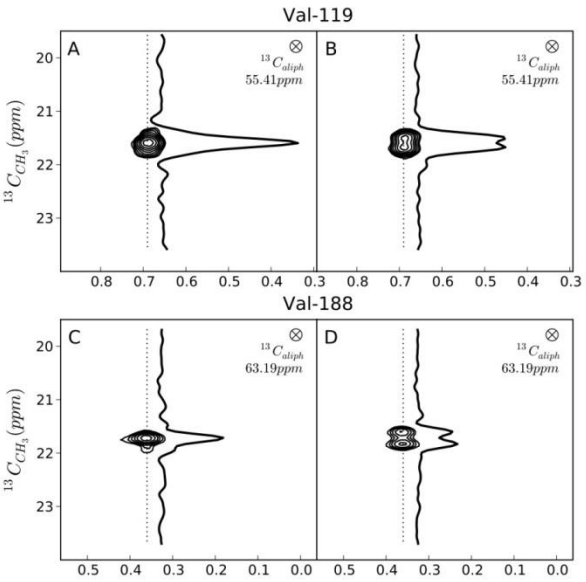


Fig. 5. J-coupling removal in the $F_1(^{13}C\ CH_3)$ dimension. During the t_1 evolution period, the $^1J_{CC}$ (35 Hz) coupling between the methyl ^{13}C and the adjacent carbon is active. While its suppression during acquisition by selective pulses is not feasible for all residue types, it can be eliminated by suitable deconvolution during processing. The (F1-F3) plane are taken orthogonally to the strips shown in figure 4 at the frequency of the $^{13}C\alpha$ cross-peak. The couplings visible in panels B and D are eliminated in panels A and C. The contour levels for the 2D plots and the vertical scale for the 1D slices are identical in all panels.

equivalent spin-systems that could be discriminated spectroscopically. It is anticipated that this combined strategy will be important for the study of large proteins and that the TOCSY approach will offer simplicity advantages that will convince an increasing number of spectroscopists.

Experimental Section

Precursors Preparation:

Preparation of aceto-hydroxy-acid synthase II enzyme (AHAS II):

Overexpression and purification of AHAS II have been carried out according to the previously published protocol [26,27].

Precursor for the linearized Ile-δ_1 [$^{13}CH_3$] labeling:

The AHAS II from *E. coli* was used in order to condensate a molecule of 3,3-[^{2}H], 1,2,3,4-$^{13}C_4$-2-oxo-butyrate with another of 3,3,3-[^{2}H], 2-^{13}C pyruvate. The product of this reaction is the 2-hydroxy-2-[1',2'-$^{13}C_2$-1'-$^2H_2$]ethyl-3-oxo-[1,2,3-$^{13}C_3$]-4-[$^2H_3$]butanoic acid (compound (3) in fig.1), a precursor of the Ile. Through this labeled precursor, the $^{13}CH_3$ in the Ile-δ1 position can be correlated to the backbone via a linear ^{13}C spin system since the Ile-γ2 methyl group remains unlabelled. The reaction was carried out by adding the AHAS II at 6 μM (420 ng/ml) in an equimolar mixture at 33 mM of deuterated pyruvate (perdeuteration has been achieved by treatment of unlabeled pyruvate in ^2H_2O at pH 10.7 during 72 hours) and ^{13}C, ^{2}H labeled 2-oxo-butyrate in a ^2H_2O buffer of potassium phosphate 50 mM at pH 7.8 (uncorrected), $MgCl_2$ 10 mM, Thiamine diphosphate 1 mM, FAD 20 μM. The complete condensation of the referred molecules has been monitored by 1D NMR. The initial labeled chemicals have been purchased from Sigma-Aldrich.

Precursors for the linearized Ile-γ_2, Leu-δ_1/Val-γ_1 and Y-shaped Leu-δ_2/Val-γ_2 labelings:

These compounds have been purchased from NMR-BIO (www.nmr-bio.com). Of note, the

chemical synthesis of the 2-hydroxy-2-[D_5] ethyl -3-oxo-[1,2,3,4-$^{13}C_4$] butanoic acid (linearized Ile-γ_2 precursor, compound (4) in fig. 1) has been described by Ayala *et al* (2012) [23] while the synthesis of both of 2-hydroxy-2-[D_3] methyl -3-oxo-[1,2,3,4-$^{13}C_4$] butanoic acid (linearized Leu-δ_1/Val-γ_1 precursor) and 2-hydroxy-2-[^{13}C]methyl-3-oxo-[4,4,4-$^{2}H_3$],[3,4-$^{13}C_2$] butanoic acid (Y-shaped Leu-δ_2/Val-γ_2) have been reported by Gans *et al* [20].

MSG production and purification:

The MSG expression and purification has been carried out as described in Godoy-Ruiz *et al* [28]. However, to achieve the desired combinatorial ILV labeling of MSG, a first solution containing a equimolar concentration of compounds (1) and (2) (Leu and Val precursor, cf fig.1A) was added to the *U*-[^{13}C,^{15}N, ^{2}H] culture 1 hour before the induction of the MSG expression. In order to optimize the co-incorporation of the ensemble of the used compounds, a second equimolar mixture of compounds (3) and (4) (Ile precursors, cf fig.1A) was added only 20 minutes before induction. The final labeling pattern is described in fig. 1B.

The analyzed MSG sample was at 1mM in $^{2}H_2O$ buffer containing 25 mM MES (pH 7.0 uncorrected), 20mM $MgCl_2$, 5mM DTT.

Nuclear magnetic resonance:

The NMR data were recorded on a Varian-Agilent spectrometer operating at ^{1}H=600 MHz equipped with a triple-resonance cryoprobe. The pulse sequence was adapted from the HCCH-TOCSY [29] as implemented in the Biopack library ("hcch_tocsy.c"). This experiment, designed for fully protonated proteins, achieves the following magnetization flow:

$H(t_1) \rightarrow C(t_2) \rightarrow TOCSY \rightarrow C \rightarrow H(t_3)$.

The detected protons during t_1 or t_3 can be any proton of the side-chain, from the terminal CH_3 to the backbone H□. On deuterated samples with protonated CH_3, an "out-and-back: design is required for sensitivity reasons: starting from the CH_3, the magnetization migrates along the side-chain (possibly up to the C□), is then labelled during t_1 and migrates back to the CH_3 according to the following scheme:

$H \rightarrow C(t_2) \rightarrow TOCSY \rightarrow C(t_1) \rightarrow TOCSY \rightarrow C \rightarrow H(t_3)$

The "out-and-back" HCCH-TOCSY pulse sequence is depicted in fig. 2. The 3D data set has been acquired using non-uniform sampling along the two indirect dimensions: the largest increments in t_1 and t_2 are respectively $128\times DW_1$ and $160\times DW_2$ but only 6400 (t_1,t_2) pairs have been sampled which corresponds to a 31% undersampling. A relaxation delay of 1.1 s was used along with 8 scans/ FID, giving rise to a net

acquisition time of 70 h. The sampling schedule was generated with *ScheduleTool*, a java program written by Jeff Hoch and colleagues at the University of Connecticut (USA). Details on the pulse sequence and the sampling schedule are given in the caption to fig. 2.

Data processing

The spectra were transformed using MddNMR 2.2, a reconstruction software using multi-dimensional decomposition (MDD) and compressed sensing (CS) written by Vladislav Orekhov and colleagues at the University of Gothenburg (Sweden). The processing scripts were generated using the qMDD graphical user interface. Compressed sensing algorithms use a l_1-norm regularization [30] to minimize the artefacts due to the non-uniform acquisition: with the iterative soft thresholding (IST) algorithm [31], convergence could be achieved in less that 500 iterations. nmrPipe was then used to phase along all dimensions after the CS reconstruction. The processing was carried out on a MacBook Pro running under MacOS 10.8.

During the t_1 evolution period, the ^{13}C chemical shift of the CH_3 groups is labelled but the homonuclear ^{13}C-^{13}C scalar coupling remains active. An alternative to experimental decoupling involves the J-deconvolution during the processing stage[32-34]. In a ^{13}C methyl group coupled with a single neighbor, the signal is multiplied by $\cos(\pi \times J_{CC}t_1)$ due to the $J_{CC} \approx 35$ Hz. By multiplying the time-domain signal by $1/\cos(\pi \times J_{CC}t_1)$ prior to processing, the corresponding J-splitting can be eliminated. Note that no data points have been acquired for t_1 values close to $t_1=1/(2\times J_{CC})$ and that the scaling factor never reaches large values (typically less than 4).

References

[1] K. Wüthrich, G. Wider, G. Wagner, W. Braun, *J Mol Biol* 1982, *155*, 311–319.

[2] G. Eich, G. Bodenhausen, R. R. Ernst, *J Am Chem Soc* 1982, *104*, 3731–3732.

[3] L. Braunschweiler, R. R. Ernst, *J. Magn. Reson.* 1983, *53*, 521–528.

[4] A. Bax, G. M. Clore, A. M. Gronenborn, *J. Magn. Reson.* 1990, *88*, 425–431.

[5] H. L. Eaton, S. W. Fesik, S. J. Glaser, G. P. Drobny, *J. Magn. Reson.* 1990, *90*, 452–463.

[6] K. H. Gardner, L. E. Kay, *J Am Chem Soc* 1997, *119*, 7599–7600.

[7] N. K. Goto, K. H. Gardner, G. A. Mueller, R. C. Willis, L. E. Kay, *J. Biomol. NMR* 1999, *13*, 369–374.

[8] V. Tugarinov, L. E. Kay, *J Am Chem Soc* 2003, *125*, 5701–5706.

[9] V. Tugarinov, L. E. Kay, *J Am Chem Soc* 2003, *125*, 13868–13878.

[10] P. Guentert, W. Braun, M. Billeter, K. Wüthrich, *J Am Chem Soc* 2001, *111*, 3997–4004.

[11] S. G. Hyberts, W. Märki, G. Wagner, *Eur. J. Biochem.* 1987, *164*, 625–635.

[12] G. W. Vuister, T. Yamazaki, D. A. Torchia, A. Bax, *J. Biomol. NMR* 1993, *3*, 297–306.

[13] V. Tugarinov, L. E. Kay, *J Am Chem Soc* 2004, *126*, 9827–9836.

[14] C. Tang, J. Iwahara, G. M. Clore, *J. Biomol. NMR* 2005, *33*, 105–121.

[15] P. Pristovšek, L. Franzoni, *J. Comput. Chem.* 2006, *27*, 791–797.

[16] D. Neri, T. Szyperski, G. Otting, H. Senn, K. Wüthrich, *Biochemistry* 1989, *28*, 7510–7516.

[17] M. J. Plevin, O. Hamelin, J. Boisbouvier, P. Gans, *J. Biomol. NMR* 2011, *49*, 61–67.

[18] G. Mas, E. Crublet, O. Hamelin, P. Gans, J. Boisbouvier, *J. Biomol. NMR* 2013, *57*, 251–262.

[19] Y. Karimi-Nejad, J. M. Schmidt, H. Ruterjans, H. Schwalbe, C. Greisinger, *Biochemistry* 1994, *33*, 5481–5492.

[20] P. Gans, O. Hamelin, R. Sounier, I. Ayala, M. A. Durá, C. D. Amero, M. Noirclerc-Savoye, B. Franzetti, M. J. Plevin, J. Boisbouvier, *Angew. Chem. Int. Ed.* 2010, *49*, 1958–1962.

[21] M. K. Rosen, K. H. Gardner, R. C. Willis, W. E. Parris, T. Pawson, L. E. Kay, *J Mol Biol* 1996, *263*, 627–636.

[22] V. Tugarinov, W.-Y. Choy, V. Y. Orekhov, L. E. Kay, *Proc. Natl. Acad. Sci. U.S.A.* 2005, *102*, 622–627.

[23] I. Ayala, O. Hamelin, C. Amero, O. Pessey, M. J. Plevin, P. Gans, J. Boisbouvier, *Chem. Commun.* 2012, *48*, 1434.

[24] C. Hilty, G. Wider, C. Fernández, K. Wüthrich, *J. Biomol. NMR* 2003, *27*, 377–382.

[25] R. Dumas, V. Biou, F. Halgand, R. Douce, R. G. Duggleby, *Acc. Chem. Res.* 2001, *34*, 399–408.

[26] C. M. Hill, S. S. Pang, R. G. Duggleby, *Biochem J* 1997, *327 (Pt 3)*, 891–898.

[27] D. Chipman, Z. Barak, J. V. Schloss, *Biochim Biophys Acta* 1998, *1385*, 401–419.

[28] R. Godoy-Ruiz, C. Guo, V. Tugarinov, *J Am Chem Soc* 2010, *132*, 18340–18350.

[29] A. Bax, G. M. Clore, A. M. Gronenborn, *J. Magn. Reson.* 1990, *88*, 425–431.

[30] A. S. Stern, D. L. Donoho, J. C. Hoch, *J. Magn. Reson.* 2007, *188*, 295–300.

[31] K. Kazimierczuk, V. Y. Orekhov, *Angew Chem Int Ed Engl* 2011, *50*, 5556–5559.

[32] N. Shimba, A. S. Stern, C. S. Craik, J. C. Hoch, V. Dötsch, *J Am Chem Soc* 2003, *125*, 2382–2383.

[33] I. Scholz, S. Jehle, P. Schmieder, M. Hiller, F. Eisenmenger, H. Oschkinat, B.-J. Van Rossum, *J Am Chem Soc* 2007, *129*, 6682–6683.

[34] D. Marion, *J. Magn. Reson.* 2010, *206*, 81–87.

[35] H. Geen, R. Freeman, *J. Magn. Reson.* 1991, *93*, 93–141.

[36] A. J. Shaka, C. J. Lee, A. Pines, *J. Magn. Reson.* 1988, *77*, 274–293.

[37] E. Kupce, J. Boyd, I. D. Campbell, *J Magn Reson Ser B* 1995, *106*, 300–303.

3.3 Methyl groups assignment in supramolecular assemblies

3.3.1 Historical overview

Apart from the structure-based approaches already described in §3.2.2, and the possible employment of solid-state NMR spectroscopy (not discussed in this manuscript), the main strategies used for the methyl group assignments of supramolecular systems (>100 kDa) are presented below.

3.3.2 Methyl groups assignment using the "divide and conquer" strategy

The idea relies on disassembling the studied system into small and stable fragments, and subsequently transferring their resonances assignment to the full-size native particle. Initially, the methyl groups assignment of a 670 kDa oligomeric protein (the 20S proteasome) has been reported (Sprangers and Kay 2007). This supra-molecular system was decomposed by mutation into its individual monomer (α_7) (21 kDa), which was easily assigned by NMR. Given the similarity of the 2D [^{13}C, ^{1}H] NMR spectra of the generated oligomeric intermediates to their native protein, the methyl groups assignment of the smallest unit was used as a basis to assign the larger assembly ($\alpha_7\alpha_7$) (360 kDa) that spontaneously forms from native monomers and finally to assign the entire particle ($\alpha_7\beta_7\beta_7\alpha_7$) (670 kDa). By combining this strategy with NOESY, COSY experiments and additional mutagenesis, 89% of Ile, Leu and Val residues from the native proteasome could be assigned.

Using the same "philosophy", the assignment of the Ile-δ_1, Leu, Val and Met-ε methyl groups in the multi-domain protein SecA (204 kDa) was also described (Gelis, Bonvin et al. 2007). In this work, the protein was expressed in constructs of different lengths; starting from an individual domain, the remaining parts were gradually inserted to finally reconstitute the native protein. Again, confirming that the NMR profile of each protein segment was "superimposable" to the full-length particle, the assignment was successfully transferred from the smaller "piece" to the larger. To complement and corroborate the obtained data, NOESY experiments and mutagenesis were also used. In this study, the complete methyl group assignment of Ile and Met residues is reported, while those of Leu and Val were restricted to the region of interest only.

In another investigation, this strategy was also applied to the 300 kDa aspartate transcarbamoylase (ATCase) (Velyvis A, Schachman H.K. et al. 2009) but revealed less effective. Nonetheless, the additional application of complementary approaches (*e.g.*, PCSs and mutagenesis) led to the assignment of 86% of the Ile, Leu and Val methyl groups.

The referenced works are clearly impressive, but the used approaches present several disadvantages, such as the following: (*i*) the multi-domain and oligomeric features are not universal; (*ii*) even when the particle of interest responds to these conditions, the possibility of individually expressing each domain in a stable form is not guaranteed, neither is the dismantlement of the oligomeric protein. In the latter context, considerable work is required to identify the conditions that efficiently disturb the quaternary structure without affecting the tertiary folding of the small subunits. Neglecting all the aforementioned biochemical bottlenecks, (*iii*) the lack of similarity between NMR spectral profiles of the native particle and its subunits remains the main risk of this approach. Indeed, this specific situation was observed with the tetrahedral amino peptidase (PhTET-2) studied in our laboratory, which shows similar signal dispersion but completely different chemical shifts between its dodecameric and monomeric states (see fig. 3.2). Therefore, a mutagenesis-based approach was chosen for the assignment of its methyl groups (see §3.1.1.3).

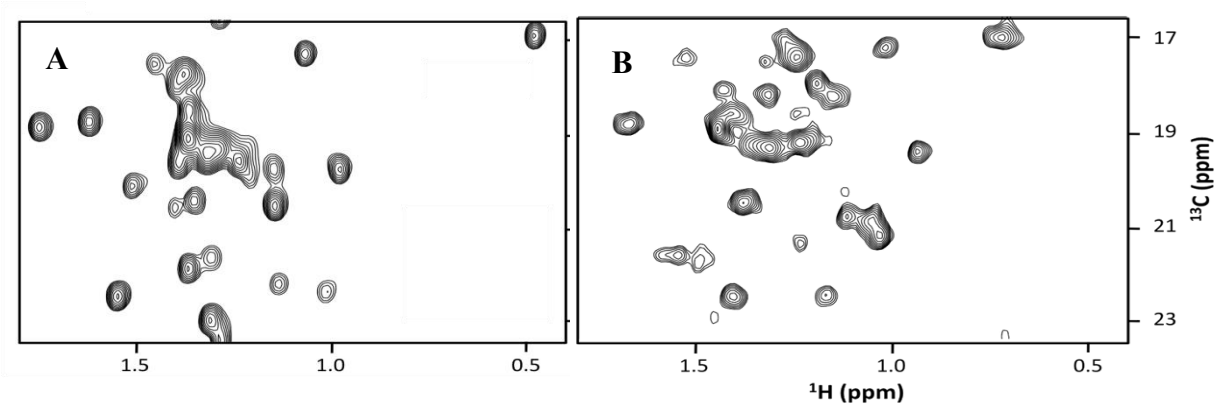


Figure 3.2: [^{13}C,^{1}H] HMQC spectra of [U-^{2}H, A$^{\beta}$]-PhTET-2: Monomeric (A) and Native (B) states (see §4.3)

3.3.3 Methyl groups assignments using mutagenesis

The practice of mutagenesis for the assignment of "punctual NMR signals" has been used since the beginning of protein NMR spectroscopy. By mutating an isotopically labeled residue (e.g., Ile) to a structurally similar residue that is not labeled (e.g., Leu), its respective methyl resonance would, in the simplest case, disappear from the NMR spectrum. Through

subsequent comparison of 2D [^{1}H,^{13}C] correlation spectra of the native and mutant proteins, the mutated residue is readily assigned. Several examples can be cited where methyl group assignments have been achieved by this site-by-site mutation strategy, such as those of the Met residues in the poliovirus RNA polymerase (52 kDa) (Yang, Welch et al. 2010) and proteasome (678 kDa)(Religa, Sprangers et al. 2010); the Ile-δ_1 amino acids in ClpP protease (300 kDa)(Sprangers, Gribun et al. 2005); and the Ile, Leu and Val residues in the chaperone, ClpB (92 kDa)(Rosenzweig, Moradi et al. 2013).

Two significant drawbacks have to be considered in this approach. The first is related to the structural alteration that may be caused by the residue used to replace the mutated amino acid. This can be "counteracted" somewhat by choosing a so-called conservative mutation as explained in 3.2.2. While the second is the laborious work required for the NMR sample preparation. Great efforts are required for the generation of the mutants plasmids and then for their production and purification.

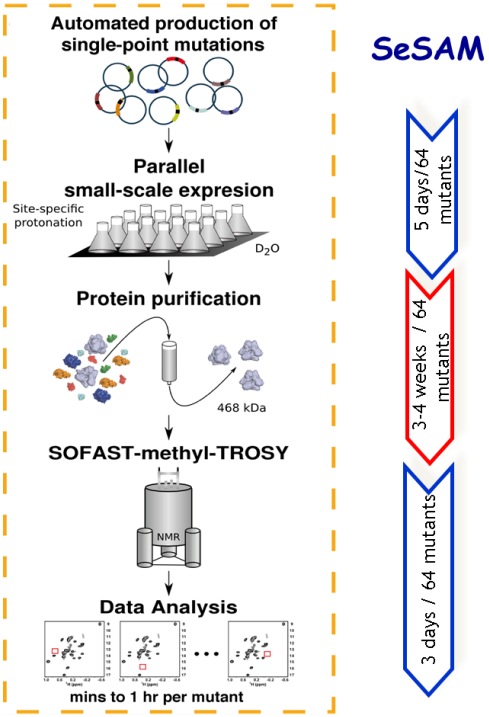


Figure 3.3: Principles of Sequence-Specific Assignment of methyl groups by Mutagenesis (SeSAM). From Amero et al. 2011.

Taking these bottlenecks into account, a strategy called SeSAM (Sequence-Specific Assignment of methyl groups by Mutagenesis) was proposed for the extensive assignment of all methyl groups (Amero, Asuncion Dura et al. 2011). The SeSAM strategy takes advantage of semi-automated protein production and preparation of its library of "single-site" mutants as well as sensitivity-optimized NMR experiments. Thanks to this approach, the complete resonance assignment of 34 Ile-δ_1 and 30 Ala-β methyl groups of a 468 kDa supramolecular protein oligomer (PhTET-2) was realized in less than 2 months (see fig. 3.3). This was feasible because of: (*i*) the involved automated mutagenesis robots, (*ii*) the parallel protein expression; and (*iii*) the SOFAST NMR experiment that allows the rapid acquisition of a high-quality 2D spectrum from samples containing in average 1 mg of proteins. Despite its effectiveness, SeSAM still encounters bottlenecks, such as the following: (*i*) difficulty in

managing the experimental procedure due to the numerous samples (*ii*) the sequential protein purification, which is a time consuming step (1 month/64 mutants), (*iii*) the need to acquire NMR spectra of entire mutant libraries (64 Ile and Ala residues in PhTET-2), which forces the spectrometer user to be present to change the samples.

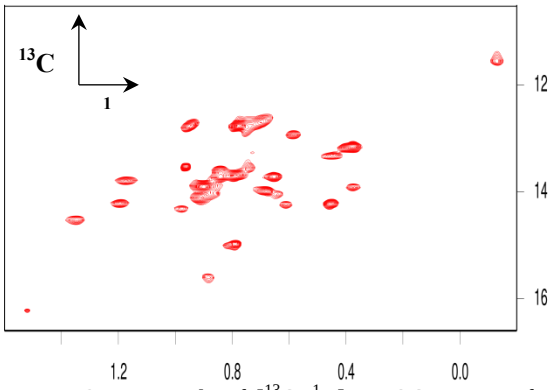


Figure 3.4: Example of [^{13}C, ^{1}H]-HMQC spectra of [U-^{2}H, I$^{\delta1}$]-PhTET-2 acquired at 50ºC on a 800 MHz spectrometer equipped with a 1.7 mm adapted cryogenic probe on 0.1 mM sample with an experimental time of 2 h.

The improvement of the two first points was successfully realized, as discussed in the work presented in §3.3.5. The reduction of culture volumes was made possible by the use of an optimized, enriched culture medium. Consequently, this allowed both the production and purification steps to be parallelized. However, optimization of the spectral acquisition time was later realized within our laboratory. An automated sample changer with a 24 sample capacity was purchased along with a 1.7 mm diameter probe adapted for small volumes. Using this equipment, the sensitivity improved by 4-fold per unit of mass, compared with the same amount of sample diluted in a standard 5 mm Shigemi tube. The use of this ensemble allows both comfortable spectral acquisitions by exploiting even the non-working times (e.g., nights and weekends) and reduced sample volumes, which decreases the total preparation cost (see fig. 2.5).

3.3.4 Article III: A cost-effective protocol for the parallel production of libraries of $^{13}CH_3$-specifically labeled mutants for NMR studies of high molecular weight proteins

In this subchapter, I describe the aforementioned improvements for the sequence-specific assignment of methyl groups by mutagenesis (SeSAM). My specific contribution was the development of an enriched culture medium that enabled the decreased culture volume, which allowed the parallelization of both production and purification.

This improved version of SeSAm has been published as protocol in Methods in Molecular Biology.

A cost-effective protocol for the parallel production of libraries of $^{13}CH_3$-specifically labeled mutants for NMR studies of high molecular weight proteins

Names: Elodie Crublet[a,b,c,*], Rime Kerfah[a,b,c], Guillaume Mas[a,b,c], Marjolaine Noirclerc-Savoye[a,b,c], Violaine Lantez[a,b,c], Thierry Vernet[a,b,c], Jerome Boisbouvier[a,b,c,*]

Affiliations :
[a] CEA, Institut de Biologie Structurale Jean-Pierre Ebel, Grenoble, France
[b] CNRS, Institut de Biologie Structurale Jean-Pierre Ebel, Grenoble, France
[c] Université Joseph Fourier – Grenoble 1, Institut de Biologie Structurale Jean-Pierre Ebel, Grenoble, France

e-mail address of the corresponding authors: elodie.crublet@ibs.fr; jerome.boisbouvier@ibs.fr

Running head: NMR assignment of large protein assemblies using mutagenesis

Summary

There is increasing interest in applying NMR spectroscopy to the study of large protein assemblies. Development of methyl-specific labeling protocols combined with improved NMR spectroscopy enable nowadays studies of proteins complexes up to 1 MDa. For such large complexes, the major interest lies in obtaining structural, dynamic and interaction information in solution, which requires sequence-specific resonance assignment of NMR signals. While such analysis is quite standard for small proteins, it remains one of the major bottlenecks when the size of the protein increases.

Here, we describe implementation and latest improvements of SeSAM, a fast and user-friendly approach for assignment of methyl resonances in large proteins using mutagenesis. We have improved culture medium to boost the production of methyl specifically labeled proteins, allowing us to perform small-scale parallel production and purification of a library of $^{13}CH_3$-specifically labeled mutants. This optimized protocol is illustrated by assignment of Alanine, Isoleucine and Valine methyl groups of the homododecameric aminopeptidase PhTET2. We estimated that this improved method allows assignment of *ca.* 100 methyl cross-peaks in 2 weeks, including 4 days of NMR time and less than 2 k€ of isotopic materials.

Key words: methyl-group, isotopic labeling, high molecular weight proteins, NMR spectroscopy, SeSAM, assignment, site-directed mutagenesis

Introduction

Supramolecular systems are involved in many of the key processes that occur in cells. Therefore, understanding their local structure and dynamics is critical. For such investigations, NMR spectroscopy is a technique of choice and now allows studies of assemblies up to 1 MDa (Sprangers and Kay 2007). This was made possible by the development of protocols for the selective protonation of methyl groups in perdeuterated proteins (Gardner and Kay 1997, Tugarinov, Kanelis et al. 2006, Ayala, Sounier et al. 2009, Ruschak and Kay 2010, Plevin and Boisbouvier 2012). This strategy is based on some very favorable relaxation properties of methyl groups in proteins, that show increased sensitivity compared to backbone amide proton. Moreover, methyl group containing residues are usually common and well dispersed within the polypeptide sequences, covering homogeneously the

protein space. Thus, methyl groups are excellent probes of protein structure, dynamics and interactions, particularly for very large proteins. In addition to studying naturally occurring methyl groups, methyl-containing amino acids can also be used to replace solvent-exposed residues as NMR reporters of protein interaction in order to, for instance, validate an expected binding site (Religa, Ruschak et al. 2011, Stoffregen, Schwer et al. 2012).

In all cases, analysis of structural and dynamic information yielded by methyl groups requires sequence-specific assignment of methyl resonances. Conventional through bonds assignment approaches (Bax 2011) to assign backbone and side chain methyl groups resonances work efficiently for small proteins but cease to be applicable to proteins over 100 kDa. In this case, alternative approaches are required. Several examples of different methyl group assignment procedures that have been successfully applied to large proteins are provided below. To date, many of the supramolecular systems studied by NMR spectroscopy are multimeric; an option to assign such assemblies is thus to try to split the quaternary complex into smaller fragments. This "divide-and-conquer" technique (Gelis, Bonvin et al. 2007, Sprangers and Kay 2007) relies on disassembling the oligomeric system and transferring the resonances assignment to the full-size complex. This method however requires considerable optimization to find conditions that destabilize oligomeric interfaces without significantly disrupting the structure of the monomer or domain. Another alternative to overcome size limitation is solid-state NMR spectroscopy (Turano, Lalli et al. 2010), in which the linewidth is independent of the molecular weight (Marassi, Ramamoorthy et al. 1997). Yet, this approach requires crystal preparation giving high quality spectra similar to solution state NMR, and time-consuming analysis of complex ^{13}C-^{13}C correlation spectra acquired using solid-state NMR. Methyl-methyl Nuclear Overhauser Enhancement (NOE) experiments can also be used in combination with chemical shift prediction programs to assign methyl groups in proteins (Xu, Liu et al. 2009). This method is so far limited to small systems due to the complexity of detecting long range NOE at more than 7 Å in very large protein assemblies. Another approach to assign large proteins is to analyze paramagnetic relaxation enhancements (PRE) induced by nitroxide spin-labels in combination with an available 3D structure (Venditti, Fawzi et al. 2011).

Alternatively, several groups have reported a mutagenesis-based approach for assignment of some methyl resonances in large proteins. Using Leu or Val mutations along with stereospecific $^{13}CH_3$ labeling of Leu/Val residues, A. Seven and J. Rizo (Seven and Rizo

2012) were able to assign methyl resonances to a 73 kDa protein domain. Similarly, 15 out of 17 methyl resonance frequencies of methionine methyl groups of an RNA polymerase were assigned by site-by-site mutation strategy (Yang, Welch et al. 2010). A similar approach was used to obtain some Ile, Leu or Val methyl group assignments in the protease ClpB (Rosenzweig, Moradi et al. 2013). In parallel, the assignment-by-mutagenesis strategy has also been applied to a 468 kDa supramolecular protein oligomer, for the first time in a highly systematic way (Amero, Asuncion Dura et al. 2011). The method called SeSAM (Sequence-Specific Assignment of methyl groups by Mutagenesis) is based on automated molecular biology techniques, small-scale parallel preparation of residue-specific isotope-labeled samples and sensitivity-optimized NMR experiments. Each mutant construct is expressed on a small-scale using fully perdeuderated expression media supplemented with isotope labeled metabolic precursors designed for the specific protonation of a single class of methyl group (Gardner and Kay 1997, Goto, Gardner et al. 1999, Tugarinov, Kanelis et al. 2006, Fischer, Kloiber et al. 2007, Gelis, Bonvin et al. 2007, Ayala, Sounier et al. 2009, Gans, Hamelin et al. 2010, Ruschak and Kay 2010). A conservative mutation of one methyl-containing residue to another non-labeled one causes the disappearance of its NMR correlation from NMR spectrum of a specifically methyl-labeled sample. This systematic strategy led to complete resonance assignment of the 34 isoleucine-δ1 and 30 alanine-β methyl groups in less than 2 months.

Although very effective, this method remains difficult to implement. The production step is achieved from 50 mL culture medium per mutant, *i.e.* one culture flask/mutant. It therefore requires a lot of manipulation from the user and may be the source of handling mistakes. Moreover, the purification step is time consuming because each mutant is purified sequentially. Therefore, we attempted to optimize and simplify this strategy (Fig 1). First we improved the culture medium to enhance cell density, allowing us to decrease the culture volume and perform all the cultures in parallel, in 24 deep-well plates. Then, all the mutants were purified in parallel, on a 96-well plate format, therefore enabling preparation of the samples in a few hours instead of weeks. Using this improved approach, we were able to reduce experiment time by a factor of 4 and isotope cost by a factor of 2 compared to previously published implementation (Amero, Asuncion Dura et al. 2011).

Materials

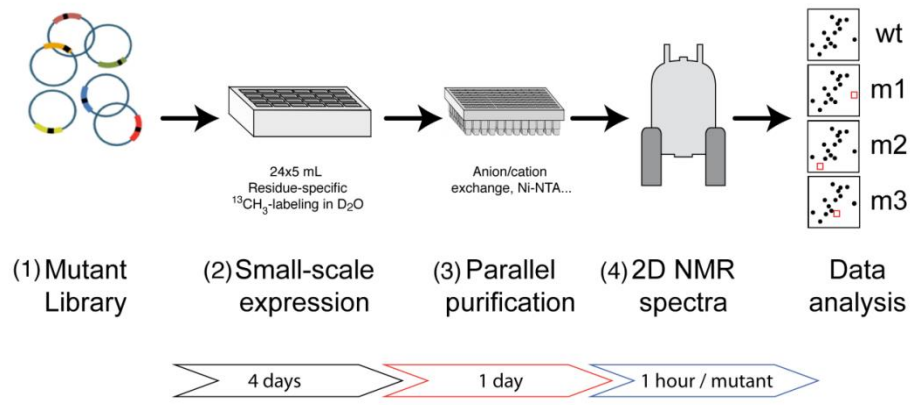
Expression of methyl-specifically labeled proteins

1. Freshly transformed *E. coli* cells (BL21(DE3), BL21(DE3)RIL…) to overexpress protein of interest
2. LB broth
3. 2X M9 medium prepared in H_2O (for 1L: 20 g of Na_2HPO_4, 7 H_2O; 6 g of KH_2PO_4; 1 g of NaCl; 2 g of NH_4Cl). Autoclave to sterilize.
4. M9 prepared in D_2O (for 1L: 5.3 g of anhydrous Na_2HPO_4; 3 g of anhydrous KH_2PO_4; 0.5 g of NaCl; 1 g of NH_4Cl). Use sterile D_2O.
5. Oligo-elements (for 1L of M9 medium: 1 mL of 1 M $MgSO_4$, 1 mL of 0.1 M $CaCl_2$, 1 mL of 0.1 M $MnCl_2$, 1 mL of 50 mM $ZnSO_4$, 0.5 mL of 100 mM $FeCl_3$). Sterilize on 0.22 μm filter. Stocks solutions should be prepared in H_2O when used in M9 100% H_2O or 50% H_2O/D_2O (for the precultures) and in D_2O when used in M9 100% D_2O. In this case, all powders should be dissolved and lyophilized twice in D_2O to remove residual water before preparing stock solutions.
6. Vitamin cocktail (for 50 mL: 25 mg of pyridoxine, 25 mg of biotin, 25 mg of panthothenate hemi-calcium, 25 mg of folic acid, 25 mg of choline chloride, 25 mg of niacineamide, 2.5 mg of riboflavin, 125 mg of thiamine). Solubilize by adjusting the pH around 7, sterilize on 0.22 μm filter and decrease the pH around 5 for long term storage. Use 2 mL for 1L of M9 medium. Vitamins should be prepared in H_2O for precultures and in D_2O when used in M9 100% D_2O (see above).
7. Isotopes: D_2O ($^2H \geq 99.8\%$), D-(^2H, ^{12}C)-glucose ($^2H \geq 98\%$), deuterated rich cell extract. Several sources of cell extract are commercially available (Spectra 9 (CIL), Celtone® Complete Medium (CIL), BioExpress® 1000 (CIL), Silantes® E.Coli-OD2 (Silantes), Isogro® (Isotec)…) In this study, we chose Isogro® (noted as ^2H-cell extract in the following text), but other cell extracts are likely to give the same results.
8. IPTG (1M in D_2O)
9. $^{13}CH_3$-methyl specifically labeled precursors were purchased on a deuterated form ready for direct introduction into the culture medium (NMR-Bio): ^2H-$^{13}CH_3$-Alanine ($^{13}C \geq 99\%$; $^2H \geq 98\%$), ^2H-$^{13}CH_3$-2-ketobutyric acid ($^{13}C \geq 99\%$; $^2H \geq 98\%$), ^2H-$^{13}CH_3$-2-hydroxy-2-methyl-3-oxo-4-butanoic acid ($^{13}C \geq 99\%$; $^2H \geq 95\%$), ^2H-L-Isoleucine ($^2H \geq 98\%$), ^2H-L-Leucine ($^2H \geq 98\%$), ^2H-α-ketoisovalerate ($^2H \geq 98\%$).
10. 10 mL 24-well DeepWell plates
11. Gas permeable adhesive seals

Purification

1. 10X BugBuster® buffer (Merck-Millipore)
2. DNAse, RNAse, lysozyme (Euromedex)
3. 96-well filter plates (Macherey-Nagel)
4. 2,2 mL-96 deep-well plates
5. Aluminum seals
6. QIAvac 96 Vacuum manifold (Qiagen) or any vacuum manifold for processing 96-well plates
7. Appropriate resin (for non-tagged proteins: any ion exchange resin (Q sepharose, SP sepharose…) or affinity resin (Protein A, Protein G, Heparin sepharose…); for tagged proteins: Ni-NTA, Talon resin, Strep-tactin sepharose…). In this study, we used Q sepharose resin (GE Healthcare Life Sciences).
8. For anion exchange chromatography: equilibration/washing buffer (20 mM Tris pH 7.5, 160 mM NaCl); elution buffer (20 mM Tris pH 7.5, 350 mM NaCl)
9. Dialysis membrane (Gebaflex, Dialysis system for small-volume samples)

Figure 1: The Principle of improved SeSAM. Schematic illustration of the parallel mutation-based NMR assignment Strategy.(1) Each methyl-containing residue in the target sequence is mutated, on a site-by-site basis, to another similar methyl containing amino acid (*e.g.* Val-to-Ala) (see Table 1). (**2**) Mutant constructs are expressed on a small-scale in 24 deep-well plates, using M9 medium supplemented with 2 g/L ^{2}H-cell extract. One hour before induction fully perdeuderated expression medium is complemented with isotope labeled metabolic precursors designed for the specific incorporation of $^{13}CH_3$ isotopes of a single class of methyl groups. The volume of culture is adjusted to ensure a minimal yield of 0.3- 0.5 mg of purified protein. (**3**) Cell pellets are then lysed in parallel using chemical lysis buffer and proteins are purified in 96-well plates filled with anion exchange (or any other suitable) resin. Each mutant sample is dialyzed against H_2O, lyophilized and dissolved in NMR suitable buffer. (**4**) NMR spectra can be acquired using the SOFAST-methyl TROSY pulse sequence and a NMR spectrometer operating at high magnetic field. Sequence-specific assignment of each NMR signal is inferred by comparing the 2D ^{1}H-^{13}C correlation spectrum of each member of the mutant library with a spectrum of the native protein. A conservative mutation of one methyl-containing residue to another non-labeled one causes the disappearance of the methyl group signal from NMR spectra recorded for a specifically methyl-labeled sample.



NMR spectroscopy

NMR spectra were recorded on a Agilent Direct Drive spectrometer operating at a proton frequency of 800 MHz equipped with a 5 mm cryogenically-cooled triple resonance pulsed field gradient probe head. Samples were loaded in a 2.5 mm shigemi tube inserted coaxially into a 5 mm tube.

Methods

The goal of the approach is both to improve yields and simplify the procedure previously published (Amero, Asuncion Dura et al. 2011). For that purpose, all the small-scale parallel production, lysis and purification steps should first be setup on uniformly deuterated native protein before applying the protocol to the whole isotopically labeled library of mutants. The hypothesis is made that all mutants behave identically to the native protein and that methyl labeling does not change the purification profile (same expression level, same purification conditions…).

Mutated CH_3-containing residue	Suggested substitution	Acceptable substitution
Ala	Ser	Val (Amero, Asuncion Dura et al. 2011), Gly, Cys, Thr
Ile	Val (Sprangers, Gribun et al. 2005, Chan, Weissbach et al. 2012)	Leu (Sprangers, Gribun et al. 2005, Amero, Asuncion Dura et al. 2011, Rosenzweig, Moradi et al. 2013), Met
Leu	Ile (Rosenzweig, Moradi et al. 2013), Met	Val, Phe
Val	Ile (Rosenzweig, Moradi et al. 2013)	Met, Leu, Ala (Mas, Crublet et al. 2013) , Thr
Thr	Ser	Ala, Asn, Val
Met	Leu	Ile, Val

Table 1: Suggested substitutions for methyl containing residues
The BLOSUM (BLOcks SUbstitution Matrix) matrix is a substitution matrix used for sequence alignment of proteins (Henikoff and Henikoff 1992). BLOSUM matrices are used to score alignments between evolutionarily divergent protein sequences. They are based on local alignments. In this table, suggested/acceptable substitutions for each of the 6 methyl-containing amino acids, for the purpose of assignment using SeSAM strategy (Amero, Asuncion Dura et al. 2011), are listed along with an application reference when it exists.

Generation of mutant plasmids libraries

Constructs carrying single point mutations can be purchased commercially or prepared using an automated molecular biology platform. Here, they were generated by in-house automated molecular biology Platform (RoBioMol - Institut de Biologie Structurale, Jean-Pierre Ebel) using an automated PCR-based protocol adapted from the QuikChange site-directed

mutagenesis method. PCR amplification was performed with Phusion Hot Start enzyme (Finnzymes) using the expression plasmid pET41c-PhTET2 as template and the specific mutagenic primers. Products were purified and digested by DpnI. Final mutations were selected by transformation and verified by sequencing. Valine residues were mutated into Alanine residues, Alanine into Valine, Isoleucine into Leucine. Amino acids must be substituted by an isosteric one to avoid considerable changes in the structure and minimize secondary shift effects (see Note 1 and Table 1).

Protein labeling for methyl detection

Selective methyl protonation provides excellent probes for monitoring interactions and dynamics, and high quality spectra can be recorded in very large systems. For assignment, a prerequisite is to restrict labeling to only one type of methyl containing residue at once, in order to minimize secondary chemical shift perturbations due to the mutation. Currently, methods are available to label Isoleucine (δ1(Gardner and Kay 1997) or γ2 (Ayala, Hamelin et al. 2012)), Alanine (Isaacson, Simpson et al. 2007, Ayala, Sounier et al. 2009) Threonine (Sinha, Jen-Jacobson et al. 2011) or Methionine (Fischer, Kloiber et al. 2007, Gelis, Bonvin et al. 2007) methyl groups efficiently. For Leucine and Valine, however, the labeling scheme is more difficult because these amino acids share the same biogenesis pathway and each residue has two prochiral methyl groups. Consequently, compared to other residues, for a similar spectral window, the number of methyl group resonances is 4 times higher, potentially resulting in peak overlap and loss of resolution. Thus, until recently, the assignment-by-mutagenesis strategy was not attractive for Valine residues. Since then, a stereospecific isotopic labeling method was developed for Leucine/Valine labeling (Gans, Hamelin et al. 2010) . More recently, our group has also developed the stereoselective labeling of Valine without Leucine using co-incorporation of pro-*S* Acetolactate and ^{2}H-Leucine (Mas, Crublet et al. 2013). This new isotopic labeling strategy increased the resolution and reduced overlaps and potential secondary chemical shifts by a factor of 4 on average. Thus, it enabled easy assignment-by-mutagenesis of Valine residues in large protein assemblies, compared to standard labeling using 2 keto-acids (Tugarinov and Kay 2003).
In this section we describe the protocol to specifically label large proteins for methyl detection and produced them on a small-scale using enriched cell extract.

1. Bacteria must first be trained to adapt progressively in 100% ^{2}H medium. All these adaptation steps will be performed in 24-well plates covered with gas permeable adhesive seal. Pick a fresh colony of transformed cells from each mutant of a LB plate (see Note 2) and start a 2-mL bacterial culture of LB medium prepared in H_20. Once OD_{600} reaches 3-4 (6-8 hours) at 37°C, transfer 50 µL of the cell culture into 2 mL of M9 medium prepared in H_2O. The starting OD_{600} of this new culture should be about 0.1. Incubate the culture in a shaking incubator (220-250 rpm) at 37°C overnight (the OD_{600} should be around 2-2.5) and transfer 100-200 µL of the cell culture into 2 mL of M9 medium prepared in 50% D_2O (see Note 3) (starting OD_{600}=0.2). Let the culture grow at 37°C until the OD_{600} reaches 2-2.5 and transfer 300 µL of the cell culture into 3 mL of M9 medium in 100% D_2O (starting OD_{600}=0.25). Incubate the culture at 37°C overnight until OD_{600} is 1.5-2 and use this cell culture as your starting culture.
2. Prepare the volume of M9 needed for the whole culture (2 x 5 mL/mutant) in a sterile flask (M9 in D_2O + oligo-elements in D_2O + vitamins in D_2O + 2 g/L of ^{12}C, ^{2}H glucose + antibiotics). Add 2g/L of ^{2}H-cell extract (see Notes 4,5,6).
3. Fill a 24 deep-well block (see Note 7) with 3.5 mL of this medium. Inoculate each well with the overnight culture at a starting OD_{600} of 0.3. Cover with gas permeable adhesive seal and grow at 37°C until OD_{600} reaches 1.5.
4. Add the precursors (see Note 8) diluted in 1.5 mL of M9/D_2O, 1 h prior to IPTG induction. The amount of $^{13}CH_3$-precursors to add was optimized to ensure complete labeling of proteins overexpressed in rich medium (Fig 2a). The yield of the protein was improved by a factor of 1.6 when ^{2}H-cell extract was added to the culture medium, while the frequency of labeling was still almost 100%, proving there is no isotopic dilution by ^{2}H-cell extract (Fig 2b).
5. Let the culture grow for 1 h. The OD_{600} should reach a value of 1.5.
6. Add IPTG (in D_2O) to 0.5 mM to induce protein expression. Continue incubation at 37°C for 4 hours (depending on your protein).
7. Centrifuge the entire plate in a swing-out rotor for 20 min at 4000 rpm. Discard the supernatant and store the pellet at –80°C or process directly.

Lysis

Protein purification and lysis strategy may vary according to the protein of interest (presence and type of fusion tag…). Here we described the lysis strategy optimized

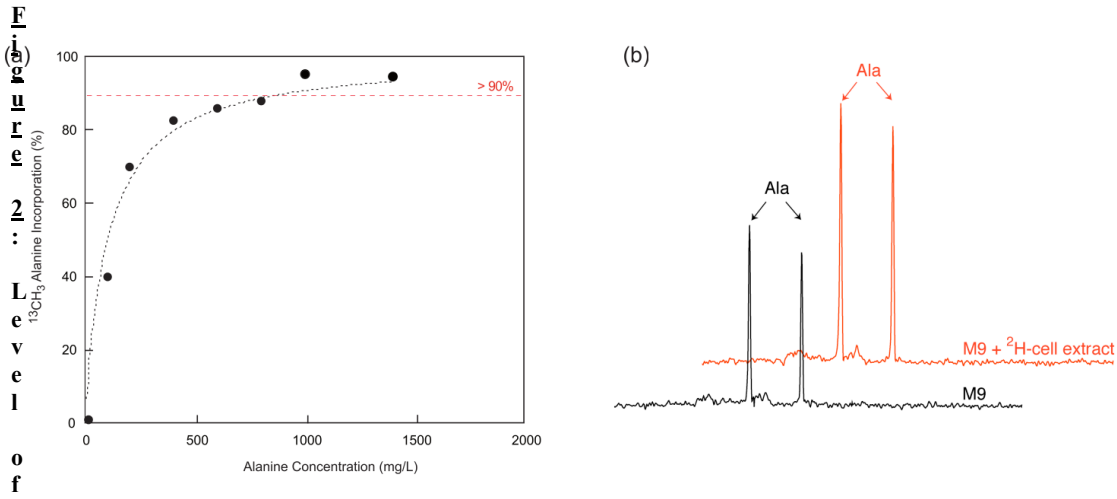


Fig.2: incorporation of $^{13}CH_3$-alanine in proteins expressed in M9 medium supplemented with ^{2}H-cell extract. Level of incorporation of ^{13}C at the Cβ-alanine position in overexpressed ubiquitin as a function of the amount of the exogenous alanine added in culture medium 1 h prior to induction. Ubiquitin was expressed in *E. coli* in M9/D_20 culture medium supplemented with 2g/L ^{2}H, ^{12}C glucose and 2g/L ^{2}H, ^{12}C Isogro®. Different amounts of $^{13}CH_3$ L-alanine (0-1.4 g/L) were added one hour before induction. ^{2}H-Isovalerate (400 mg/L) and ^{2}H-isoleucine (120 mg/L) were added to prevent scrambling from alanine. Fixed amount of $^{13}CH_3$ L-methionine (0.5 g/L) was used as an internal reference. The level of incorporation was determined by analyzing the intensities of one alanine methyl resonance with respect to signal of methyl group of one methionine residue. A level of incorporation in Alanine side chains ≥ 90% is obtained by adding 1 g of alanine per liter of culture medium. **(b)** Comparison of 1D ^{13}C-filtered NMR spectra of U-[^{12}C, ^{2}H], Ala-[$^{13}CH_3$]$^{\beta}$ ubiquitin expressed in equal volume of M9 medium (solid line) or M9 medium supplemented with ^{2}H-cell extract (dashed line). Data were recorded at 37°C using a 2.5 mm Shigemi tube, on an 800 MHz NMR spectrometer equipped with a cryogenic probe head. Comparison of 1D spectra shows an increase of the *ratio* Signal/Noise by a factor of 1.65. These results are in agreement with an increase of both the cell density (OD_{600} x 1.7 with ^{2}H-cell extract) and the culture yields (yields x 1.75 with ^{2}H-cell extract), indicating that ^{13}C-Alanine is fully incorporated in these conditions.

for PhTET2 expressed in BL21(DE3)RIL.

1. Lyse the cells (see Note 9). Add 250 µL of lysis buffer/well in the 24-well culture plate (1X BugBuster®, 20 mM Tris pH 7.5, 100 mM NaCl (see Note 10) 5 mM $MgCl_2$, 2 µg/mL DNAse, 10 µg/mL RNAse, 0.3 mg/mL Lysozyme) and resuspend the cells by pipetting up and down. Pool the 2 pellets of the same mutant (see Note 7) and incubate 40 min at room temperature with (occasional) shaking.
2. Heat the crude extract at 85°C (see Note 11) for 15 min by making the plate directly float in a water bath.
3. Centrifuge the plate at 4000 rpm for 20 min. Both soluble- and whole-cell pellets should be analyzed (see 3.4.6).

Small-scale parallel purification

In this section, we describe a typical protein purification procedure using an anion exchange resin, but alternative resins can be used (see Note 12).

1. Prepare the purification 96-well filter plate. Resuspend the Q sepharose resin thoroughly. Pipet 800 µL of resin suspension (bed volume of 400 µL) into each well of the plate. Wash each well twice with 1 mL of water and 3 times with 1 mL of equilibration buffer using a vacuum manifold (or alternatively a centrifuge with a swing-out rotor).
2. Transfer the clear supernatants (section 3.3.3) into the 96-well filter plate containing the resin. Seal the block with aluminum seal and caps to avoid leaks and place at room temperature for 1 h with gentle shaking.
3. Place the 96-well filter plate over a 96 deep-well plate in the vacuum manifold. Remove the aluminum seal and caps (only over the used wells) and let the samples flow through the resin first by gravity, then by applying vacuum until the samples have been completely drawn through the plate.
4. Place a drain deep-well plate and wash the resin by adding 400 µL of washing buffer to each well and then apply vacuum as above. Repeat the wash 4 times.
5. Place the filter plate on top of a new deep-well plate and add 400 µL of elution buffer in each well. Incubate 2 min and proceed as above. Repeat the elution 4 times and store the eluate.
6. Analyze by SDS-PAGE the total, soluble and eluted fractions for some of the mutants, according to standard procedures.

NMR spectroscopy

To assign a large protein such as PhTET2 (468 kDa, 353 residues per subunit), the SeSAM strategy (Amero, Asuncion Dura et al. 2011) is followed.

The type of probe and the tube configuration can affect the required amount of material. For small sample amounts, users should always choose the most sensitive probe available (best S/N per mg of protein). Using a standard 5-mm cryogenically-cooled probe, the protein should be concentrated to optimize the amount of spins present in the most sensitive area of the active volume (*i.e.* near the axial symmetry axis). This can be achieved simply by placing the sample in a small diameter tube (1 to 3-mm) centered inside a standard 5-mm tube. As a result, compared to a 5-mm NMR tube, the sensitivity gain is *ca.* a factor of 2 per mg of protein. Alternatively, using cryoprobes optimized for small volumes (3 mm coldprobe (Agilent); 1.7 mm microcryoprobe (Bruker)) can further increase senstivity two fold allowing a reduction of the culture volume (1* 5 mL-well/mutant) or a division of the acquisition time

by a factor of 4. In this study, the samples were loaded in a 2.5 mm-shigemi tube placed coaxially to a regular 5-mm NMR tube as a sample holder and NMR spectra were recorded on a spectrometer equipped with a 5-mm cryprobe.

1. Prepare methyl labeled mutants as described above. Transfer the protein in a NMR suitable buffer. In our case, The protein was extensively dialyzed in H_2O, lyophilized and resuspended in 60 µL (± 0.15 mM) of 20 mM Tris-DCl 20 mM NaCl pH 7.4 in 100% D_2O. Alternatively (if the protein is not stable in pure H_2O), buffer can be exchanged by a series of dilution/concentration steps in D_2O buffer using a concentration unit with an appropriate cut-off.
2. Record ^{1}H-^{13}C SOFAST-methyl-TROSY spectra (more sensitive experiment by unit of time (Amero, Schanda et al. 2009)) for each mutant sample and native protein. Here, taking advantage of the thermostability of PhTET2, NMR data are recorded at 50°C. The angle of proton excitation pulse is set to 30° and the recycling delay is optimized to 0.4 s. The length of each NMR experiment is adjusted depending on the concentration of the sample typically 1h for 9 nmol of sample.
3. All data are processed and analyzed with NMRPipe (Delaglio, Grzesiek et al. 1995). Sequence-specific assignments of each NMR signal are inferred by comparing the 2D ^{1}H-^{13}C correlation spectrum recorded for each member of the mutant library with a spectrum recorded for the wild-type protein (Fig 3). Conceptually, assignment-by-mutagenesis is straightforward. In practice, however, the overlap of resonances and the occurrence of secondary chemical shifts changes can confuse the analysis. Perturbations are likely to occur but can be minimized (using conservative mutations). In the first round, all spectra with only one missing peak are considered for a straightforward assignment of a first set of resonances. Then, more complex spectra are studied with consideration of structure, first set of unambiguous assignment and the whole set of spectra taken into account, in order to analyze secondary chemical shift perturbations. That is why considering the full library of single-site mutations greatly simplifies the process of resonance assignment by cross-validating the results several times (see Note 13).

Conclusion

Thanks to the development of protocols and molecules that allow residue-specific

protonation of methyl groups in highly perdeuterated proteins, it is now feasible to apply solution NMR techniques to protein systems as large as 1 MDa. Structural and dynamic information yielded by methyl groups is most useful when a sequence-specific assignment for the probe is known. Easily obtaining these assignments remains the major hurdle in many studies of large proteins. Here we describe a fast, efficient and user-friendly protocol for resonance assignment that has allowed us to assign up to 100 methyl cross-peaks in about 2 weeks with 4 days of NMR time and a isotopic cost of less than 2 k€. We demonstrated the feasibility of this protocol on samples labeled on Alanine, Isoleucine or Valine. This method can also be extended to Methionine- or Threonine-labeled proteins. As for Valine labeling, the methyl-specific labeling of Leucine residues will require the development of a stereoselectively labeled amino acid (or precursor) to label Leucine residues independently from Valines.

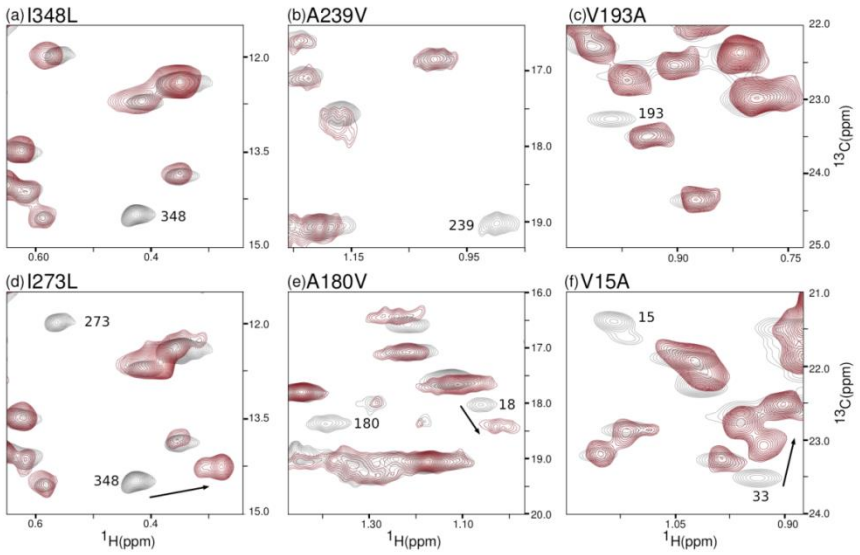


Figure 3: Examples of spectra of mutants with specific Ile-δ1, Ala-β or Val pro-*S* labeled methyl probes
Spectra of mutants displaying modest secondary chemical shift perturbations (a,b,c) were initially chosen for analysis and, when possible, sequence-specific assignment of methyl groups were made. This initial set of unambiguous assignments assisted the analysis of spectra displaying larger chemical shift perturbations (d, e, f). Spectra of this figure are extracted from the work of Amero *et al.* (Amero, Asuncion Dura et al. 2011) and (Mas, Crublet et al. 2013).SOFAST Methyl-TROSY spectra were recorded at 50°C, on an 800 MHz NMR spectrometer equipped with a cryogenic probe head, using samples of *U*-[^{12}C, ^{2}H], Ile-[$^{13}CH_3$]$^{\delta 1}$ (a and d) (Amero, Asuncion Dura et al. 2011), *U*-[^{12}C, ^{2}H], Ala-[$^{13}CH_3$]$^{\beta}$ (b and e) (Amero, Asuncion Dura et al. 2011) or *U*-[^{12}C, ^{2}H], Val-[$^{13}CH_3$]$^{pro\text{-}S}$ (c and f)(Mas, Crublet et al. 2013) labeled mutant PhTET2 protein (red). Each spectrum was overlaid with the reference spectrum of the native particle (black). Experimental acquisition times were adjusted to the sample protein concentration (from 0,26 to 0,42 mg/sample, *i.e.* 7-11 nmol of PhTET2 monomer) with a maximum experimental time set to 1h. The assignment inferred for the missing resonance in the mutant spectrum is indicated and secondary chemical shift perturbations are annotated with arrows.

1. Alanine according to Amero *et al* (Amero, Asuncion Dura et al. 2011), who mutated Alanine into Valine. However, according to BLOSUM matrices for amino acid substitutions (Henikoff and Henikoff 1992), the Valine to Isoleucine mutations (score 3) would have been wiser and most likely to have minimal effect on the structure and NMR spectra. Indeed, Valine differs from Isoleucine by the loss of only one CH_2 group, whereas it differs from Alanine by the loss of $CH\text{-}CH_3$. According to these matrices, Chan *et al* (Chan, Weissbach et al. 2012) generated a set of mutants in which each Isoleucine was substituted by a Valine (score 3). The resulting chemical shift perturbations were smaller than those from the corresponding Ile to mutations Ala (score -1), facilitating the assignment (Table 1).
2. Antibiotics are not specified. Put appropriate antibiotics considering the plasmid and the cells used. However, ampicillin is not suitable because it can be inactivated by the β-lactamases produced by the cell, resulting in plasmid loss and drop of protein expression. This effect seems to be worse in D_2O medium because of the successive adaptation steps that promote loss of selective pressure. Here, we routinely used the pET-41 plasmid, which contains the kanamycin resistance gene.
3. M9 50 % D_2O is prepared with 2X M9 H_2O and same volume of D_2O. The oligo-elements, vitamins, antibiotics and glucose used at this step are still prepared in water.
4. To increase cell density, we supplemented M9 medium with ^{2}H-cell extract. We optimized the protocol with Isogro® but many similar rich bacterial cell growth media are commercially available (Spectra 9, Celtone® Complete Medium, BioExpress® 1000, Silantes® E.Coli-OD2…). We did not test them but they can most likely be used to replace Isogro®, after suitable optimization.
5. Rich medium can be the source of isotopic dilution as it contains $^{12}CD_3$-labeled amino acids. To minimize this, we tested different concentrations of ^{2}H-cell extract (2, 3 or 10 g/L). Growth is significantly increased from 0 to 2 g/L, however higher concentrations of ^{2}H-cell extract did not show further significant improvement. Moreover, ^{2}H-cell extract is added at the start of the culture, whereas methyl specifically labeled precursors are added 1h before induction. This early addition reduces isotopic dilution because most of deuterated amino acids are consumed for the cell growth, while $^{13}CH_3$-precursors are incorporated for the production of the labeled protein. Optimization was done using Isogro®, but similar results are expected for other types of commercial ^{2}H-cell extracts.

6. For ubiquitin, yields with 2g/L of ^{2}H-cell extract were increased by a factor of 1.6 (Fig 2b).
7. Considering both yields, parallelization of the method and simplification of the purification steps, we were able to reduce the volume of production to 10 mL/mutant. One 24 deep-well plate is thus suitable for 12 mutants (2 wells of 5 mL/mutant). The final concentration in a 2.5 mm Shigemi tube is around 0.15 mM (9 nmol or 0,35 mg of PhTET2 monomer) in 60 µL.
8. Incorporation level of $^{13}CH_3$-alanine was determined using 2 g/L of ^{12}C, ^{2}H-glucose and 2 g/L of ^{2}H-cell extract in culture medium. Using higher concentrations of glucose may cause isotopic dilution of the $^{13}CH_3$-precursor (Plevin, Hamelin et al. 2011) and the incorporation curves must then be modified accordingly.
9. One crucial step to optimize in the protein production process is the bacterial cell lysis. Different lysis methods should be tested. However, when tens or hundreds of mutants are produced, only a few lysis methods can be reasonably used in parallel. Sonication proved problematic unless using High Throughput sonicators (which are expensive) or ultrasonic bath (but the ultrasonic energy is not always equally distributed throughout the plate and the results may not be reproducible from well to well). For these reasons, we chose to optimize lysis conditions using chemical treatment (all the preliminary tests were performed on the native protein). This allowed us to achieve lysis directly on the culture plates, in parallel. We tested different lysis buffers and selected the BugBuster®, as it showed the best solubility yields for our protein of interest.
10. The NaCl concentration of the lysis buffer has to be settled according to the protein of interest. Decrease it if the protein does not bind to the ion exchange resin.
11. The protein studied is an aminopeptidase from a thermophilic organism, adapted to high temperatures. The protein is thus stable at 85°C for 15 min, whereas most of other *E. coli* proteins will precipitate. Bacterial contaminants are then removed by centrifugation.
12. Fill your filter plate with any suitable resin (anion or cation exchange resin, Ni-NTA, StrepTactin…). Adapt bead volume to the capacity of the resin. Typically, we prepare wells containing 400 µL of anion exchange resin for a 10-mL cell culture. This can be scaled up or down to suit your needs based on the expected protein yield.
13. In almost 50 % of cases (using conservative mutations (see Note 1)), the only

difference between mutant and reference spectra is a single missing cross-peak in the spectrum of the mutant (Fig 3 a,b,c). In such instances, the missing peak can be unambiguously assigned to the methyl group of the mutated residue. In the remaining spectra, the disappearance of the signal is accompanied by small changes in the chemical shift of a few additional correlations (Fig 3 d,e,f). This effect is expected and has previously been observed (Sprangers, Gribun et al. 2005, Balayssac, Bertini et al. 2008, Amero, Asuncion Dura et al. 2011). Peak movements that do not directly concern the mutated resonance can complicate the process of obtaining a sequence-specific assignment from a single experiment, especially in an overcrowded region of the spectrum. Conservative mutations enable minimization of secondary chemical shift perturbations. In the same way, stereospecific labeling of a single methyl group (Valine pro-*S*) *vs* labeling of 4 methyl groups (Val/Leu) using α-ketoisovalerate (Tugarinov and Kay 2003) reduces peak overlapping as well as secondary chemical shift up to a factor 4.

Nonetheless, secondary chemical shift perturbations reflect modifications in the local electronic environment and can therefore provide complementary information that can be used to confirm the proposed assignment. The key point is that the information provided by secondary chemical shift changes only becomes interpretable when data from a full library of methyl group mutants is considered (Amero, Asuncion Dura et al. 2011). Any ambiguous assignment can therefore be readily cross-validated using structurally-close, straightforward-assigned resonances. Using an incomplete library of mutants would not permit the same level of confidence in the final assignments.

Acknowledgements

We would like to thank Dr P. Macek, M. Plevin, O. Hamelin, P. Gans, I. Ayala, C. Amero and A. Favier for stimulating discussions and assistance in sample preparation or analysis. This work used the RobioMol, High-Field NMR, Isotopic Labeling and Seq3A platforms of the Grenoble Instruct centre (ISBG; UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB). The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme FP7/2007-2013 Grant Agreement no. 260887.

References

1. Sprangers R., Kay L.E. (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* **445 (7128)**,618-622.

2. Gardner K., Kay L.E. (1997) Production and incorporation of 15N, 13C, 2H (1H-delta 1 methyl) isoleucine into proteins for multidimensional NMR studies. *J Am Chem Soc* **119**,7599-7600.

3. Ayala I., Sounier R., Use N. et al. (2009) An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. *J Biomol NMR* **43 (2)**,111-119.

4. Tugarinov V., Kanelis V., Kay L.E. (2006) Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat Protoc* **1 (2)**,749-754.

5. Ruschak A.M., Kay L.E. (2010) Methyl groups as probes of supra-molecular structure, dynamics and function. *J Biomol NMR* **46 (1)**,75-87.

6. Plevin M.J., Boisbouvier J. (2012) Isotope-Labelling of Methyl Groups for NMR Studies of Large Proteins. In: Recent Developments in Biomolecular NMR. Royal Society of Chemistry. doi:10.1039/9781849735391

7. Stoffregen M.C., Schwer M.M., Renschler F.A. et al. (2012) Methionine scanning as an NMR tool for detecting and analyzing biomolecular interaction surfaces. *Structure* **20 (4)**,573-581.

8. Religa T.L., Ruschak A.M., Rosenzweig R. et al. (2011) 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* **133 (23)**,9063-9068.

9. Bax A. (2011) Triple resonance three-dimensional protein NMR: before it became a black box. *J Magn Reson* **213 (2)**,442-445.

10. Gelis I., Bonvin A.M., Keramisanou D. et al. (2007) Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. *Cell* **131 (4)**,756-769.

11. Turano P., Lalli D., Felli I.C. et al. (2010) NMR reveals pathway for ferric mineral precursors to the central cavity of ferritin. *Proc Natl Acad Sci U S A* **107 (2)**,545-550.

12. Marassi F.M., Ramamoorthy A., Opella S.J. (1997) Complete resolution of the solid-state NMR spectrum of a uniformly 15N-labeled membrane protein in phospholipid bilayers. *Proc Natl Acad Sci U S A* **94 (16)**,8551-8556.

13. Xu Y., Liu M., Simpson P.J. et al. (2009) Automated assignment in selectively methyl-labelled proteins. *J Am Chem Soc* **131 (27)**,9480-9481.

14. Venditti V., Fawzi N.L., Clore G.M. (2011) Automated sequence- and stereo-specific assignment of methyl-labeled proteins by paramagnetic relaxation and methyl-methyl nuclear overhauser enhancement spectroscopy. *J Biomol NMR* **51 (3)**,319-328.

15. Seven A., Rizo J. (2012) Assigning the methyl resonances of the 73 kDa Munc13-1 MUN domain by mutagenesis. 25 th ICMRBS-poster n° P302 TU, Lyon.

16. Yang X., Welch J.L., Arnold J.J. et al. (2010) Long-range interaction networks in the function and fidelity of poliovirus RNA-dependent RNA polymerase studied by nuclear magnetic resonance. *Biochemistry* **49 (43)**,9361-9371.

17. Rosenzweig R., Moradi S., Zarrine-Afsar A. et al. (2013) Unraveling the mechanism of protein disaggregation through a ClpB-DnaK interaction. *Science* **339 (6123)**,1080-1083.

18. Amero C., Asuncion Dura M., Noirclerc-Savoye M. et al. (2011) A systematic mutagenesis-driven strategy for site-resolved NMR studies of supramolecular assemblies. *J Biomol NMR* **50 (3)**,229-236.

19. Goto N.K., Gardner K.H., Mueller G.A. et al. (1999) 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* **13 (4)**,369-374.

20. Fischer M., Kloiber K., Hausler J. et al. (2007) Synthesis of a 13C-methyl-group-labeled methionine precursor as a useful tool for simplifying protein structural analysis by NMR spectroscopy. *Chembiochem* **8 (6)**,610-612.

21. Gans P., Hamelin O., Sounier R. et al. (2010) Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. *Angew Chem Int Ed Engl* **49 (11)**,1958-1962.

22. Ayala I., Hamelin O., Amero C. et al. (2012) An optimized isotopic labelling strategy of isoleucine-gamma2 methyl groups for solution NMR studies of high molecular weight proteins. *Chem Commun (Camb)* **48 (10)**,1434-1436.

23. Isaacson R.L., Simpson P.J., Liu M. et al. (2007) A new labeling method for methyl transverse relaxation-optimized spectroscopy NMR spectra of alanine residues. *J Am Chem Soc* **129 (50)**,15428-15429.

24. Sinha K., Jen-Jacobson L., Rule G.S. (2011) Specific labeling of threonine methyl groups for NMR studies of protein-nucleic acid complexes. *Biochemistry* **50 (47)**,10189-10191.

25. Mas G., Crublet E., Hamelin O. et al. (2013) An efficient protocol for selective labeling of Valine methyl groups in high molecular weight protein assemblies. In preparation.

26. Tugarinov V., Kay L.E. (2003) Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J Am Chem Soc* **125 (45)**,13868-13878.

27. Amero C., Schanda P., Dura M.A. et al. (2009) Fast two-dimensional NMR spectroscopy of high molecular weight protein assemblies. *J Am Chem Soc* **131 (10)**,3448-3449.

28. Delaglio F., Grzesiek S., Vuister G.W. et al. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6 (3)**,277-293.

29. Henikoff S., Henikoff J.G. (1992) Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A* **89 (22)**,10915-10919.

30. Chan P.H., Weissbach S., Okon M. et al. (2012) Nuclear Magnetic Resonance Spectral Assignments of alpha-1,4-Galactosyltransferase LgtC from Neisseria meningitidis: Substrate Binding and Multiple Conformational States. *Biochemistry* **51 (41)**,8278-8292.

31. Plevin M.J., Hamelin O., Boisbouvier J. et al. (2011) A simple biosynthetic method for stereospecific resonance assignment of prochiral methyl groups in proteins. *J Biomol NMR* **49 (2)**,61-67.

32. Sprangers R., Gribun A., Hwang P.M. et al. (2005) Quantitative NMR spectroscopy of supramolecular complexes: dynamic side pores in ClpP are important for product release. *Proc Natl Acad Sci U S A* **102 (46)**,16678-16683.

33. Balayssac S., Bertini I., Bhaumik A. et al. (2008) Paramagnetic shifts in solid-state NMR of proteins to elicit structural information. *Proc Natl Acad Sci U S A* **105 (45)**,17284-17289.

CHAPTER IV

APPLICATIONS OF METHYL GROUPS LABELING STRATEGY

(SINGLE AND COMBINATORIAL LABELING)

4.1 Labeling of methyl groups for real-time NMR studies of protein oligomerization process

4.1.1 Context

In nature, oligomeric proteins are highly abundant. In fact, it has been estimated that more than one-third of the total proteins *in vivo* are oligomeric (Ali and Imperiali 2005). Several reasons were suggested to explain the clear protein evolution toward assemblies: (*i)* the introduction of new active sites. It seems that one-sixth of oligomeric enzymes have an active site at the inter-subunit interface; (*ii*) it is easier to synthesize, faithfully, proteins consisting of multiple short subunits than a single chain protein of comparable size; (*iii*) in case of thermophilic organisms, oligomerization confers higher stability of proteins against denaturation; and (*iv*) enzymatic activity can be controlled by regulation of the oligomeric state (Goodsell and Olson 2000).

Nearly 50 years ago, it was demonstrated that a denatured ribonuclease protein can spontaneously recover its native folding (Anfinsen and Haber 1961). Therefore, the information contained in the protein sequence was suggested to be enough to guide refolding toward the correct conformation. Today, the governing principles of folding kinetics have been extensively studied. Indeed, the prediction of protein structures is becoming more reliable (Salvatella, Dobson et al. 2005). However, significant information gaps still subsist regarding the assembly process of large complexes. Studies to obtain mechanistic details about oligomerization are scarce. The complexity of following all of the assembly intermediates, which are generated almost simultaneously and often with a very short lifetime, can perhaps explain this deficit. Indeed, any of the currently available techniques used for structural studies present a limitation for such purposes. Crystallography provides only static analysis, electron microscopy (EM) has low resolution, native mass spectroscopy (NMS) informs about the mass and stoichiometry of the formed intermediates but does not supply direct structural data and the real-time NMR was, for a long time, limited to small biological objects (~20 kDa). Nevertheless, some interesting studies, such as the capsid maturation of the herpes virus and bacteriophage HK96 (Heymann, Cheng et al. 2003, Gan, Speir et al. 2006), using mainly time-lapse cryo-EM monitoring, have been reported. Additionally, more recently, an *in vitro* ribosome-assembly mechanism model was proposed (Mulder, Yoshioka et al. 2010).

Furthermore, the band-Selective Optimized-Flip-Angle Short Transient (SOFAST) (Schanda, Kupce et al. 2005) technique, established in our laboratory, offers the ability to monitor in real-time kinetics of proteins at a site-resolved level. Its use, in combination with the described labeling strategies and complementary structural tools (i.e., EM and isotopic hybridization MS analyses (NMS), represents a potential tool for the study of the oligomerization mechanisms of high molecular weight complexes (Amero, Schanda et al. 2009).

4.1.2 Article IV: Probing protein self-assembly by real-time NMR spectroscopy

In this subchapter, I will describe the study of the self-assembly mechanism of a dodecameric amino-peptidase (PhTET-2; ~0.5 MDa). This study will be presented as a draft for a scientific publication where my contribution was first in setting up the protocols for deoligomerization and reoligomerization processes and then in preparing and characterizing the oligomeric state of the isotopically labeled sample used for the subsequent NMR analysis, which was performed by Dr. P. Macek[l].

[l] Post-Doctor at Dr. Boisbouvier team from 2011 to 2014.

Title: Probing Protein Self-Assembly by Real-Time NMR spectroscopy

Authors: P. Macek[1], R. Kerfah[1], E. Boeri Erba[1], C. Moriscot[1], E. Crublet[1], G. Schoehn[1], C. Amero[2], J. Boisbouvier[1]*

Affiliations:

[1]Institut de Biologie Structurale-J.P. Ebel, 41 Rue Jules Horowitz, 38027 Grenoble, France.

[2]Centro Investigaciones Quimicas, Av. Universidad 1001, CP 62210 Cuernacava, Mexico.

*correspondence to: jerome.boisbouvier@ibs.fr and carlosamero@uaem.mx

Abstract: The spontaneous formation of higher order structures from building blocks -self-assembly, is a fundamental attribute of life. Although self-assembly is a time-dependent process that occurs at the molecular level, its current understanding originates from static structures, low-resolution techniques and modelling. Nuclear magnetic resonance (NMR) spectroscopy offers the unique ability to monitor structural changes at the atomic level in real-time; however, its size and time resolution constraints remain a practical challenge in studies of self-assembly. Here, we report the unique application of methyl specific labeling, in an otherwise deuterated protein, combined with relaxation optimized, fast acquisition real-time NMR to overcome both size and time scale limits, respectively. For the first time the self-assembly of a nearly half mega Dalton protein complex was monitored at the structural level, including the intermediate states.

One Sentence Summary: Monitoring the self-assembly process at the structural level by real-time NMR, EM and native MS revealed oligomeric intermediates and parallel pathways in the assembly of a 468 kDa nanomachine.

Main Text: Self-assembly, the essential process in nature is also a desired fabrication strategy in technology. Molecular self-assembly reflects information encoded in the protein structure as a combination of weak non-covalent interactions (*1*). Spontaneous organization of protein oligomers from a small number of subunits has been studied and explored experimentally and theoretically for complexes with dihedral and cyclic symmetries (*2*, *3*). On the contrary, protein assemblies with cubic group symmetries comprise a high number of subunits and are therefore difficult to study experimentally due to their inherently large size.

In biology, cubic symmetries, uniquely suited for the creation of large hollow shells optimal for the encapsulation of substrates, drew attention for their appearance in spherical viruses (*4*). Both thermodynamic and kinetic models of self-assembly into particles with cubic symmetries have been presented (*5*, *6*), in particular for proteins with the icosahedral symmetry common in spherical viruses. The physical properties of cubic symmetries lead to their prevalence in large spherical protein complexes, which are optimal for the design of protein nanocages and nanovaults. The paradigm of self-assembly was recently transferred from nature to nanotechnology as a bottom-up approach. In addition, developments in protein and DNA technologies led to *de novo* design of self-assembled particles and materials, thus demonstrating significant progress in bionanotechnology (*7*, *8*).

The function of molecular machines is linked to conformational rearrangements (*9*). Similarly, as accepted for protein-protein associations, which is the repetitive process that occurs during self-assembly, the protein subunits must undergo conformational rearrangements in order to assemble into a supramolecular structure (*10*). Tracking the conformational changes in off-equilibrium conditions is challenging due to the added time dimension and low populations of the intermediate states. Several experimental tools were developed for time-resolved structural studies. Nuclear magnetic resonance spectroscopy (NMR) can probe the conformational changes in real-time, but its use in studies of self-assembly was limited by the size of the particles (*11*). Electron microscopy (EM) images large biomolecules, but is restricted to fixed samples and provides low structural and time resolution (*12*), while native mass spectrometry (MS) can determine the mass of intact protein complexes and their precise stoichiometry (*13*), but provides no direct structural information.

In this work, we overcome the main limitations of NMR in order to elucidate the self-assembly mechanism of a half mega Dalton complex. In particular, we implement methyl specific labeling in a perdeuterated background and relaxation optimized fast acquisition in real-time NMR. NMR data were complemented by, and corroborated with, time-resolved data from both EM and isotopic hybridization MS analyses.

This remarkable combination of structural techniques allowed us to dissect the self-assembly pathway, including the kinetic intermediates, of a protein nanomachine with tetrahedral symmetry. Tetrahedral aminopeptidase 2 (TET2) from *Pyrococcus horikoshii*, an assembly with a molecular weight of 468 kDa, comprises twelve subunits. Back-to-back dimers form the edges of a hollow truncated tetrahedron with cavity accesses located at the

apexes and facets (*14*).

Previously, we had reported the fast acquisition of high quality ^{1}H-^{13}C methyl transverse relaxation optimized spectroscopy (SOFAST-methyl-TROSY) data sets of alanine$^{\beta}$ (Ala$^{\beta}$) at 1.2 s (*15*) in steady-state conditions. In this work we benefit from our previous assignment of methyl resonance frequencies of TET2 Ala$^{\beta}$ (*16*) and utilize them as site-specific structural probes. As the C$^{\beta}$ chemical shift is a well-known reporter of secondary structures (*17*), we chose the Ala$^{\beta}$-[$^{13}CH_3$] specific labeling in an otherwise uniformly deuterated background *U*-[^{15}N,^{12}C,^{2}H] (*18*) to follow the structural changes during the self-assembly process. To observe the self-assembly process in real-time, the system was homogeneously brought into off-equilibrium using the stopped flow device located inside the NMR magnet (*19*), and an Ala$^{\beta}$-[$^{13}CH_3$] spectra series was collected with a 15 s acquisition time per spectrum (Fig. 1A).

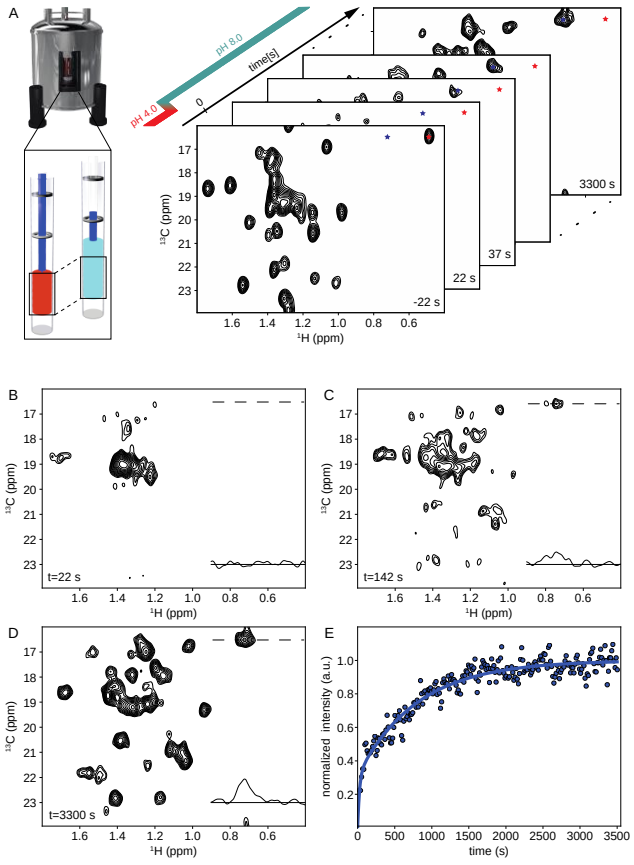


Fig. 1. Real-time kinetic NMR and progress of TET2 self-assembly. (A) *In situ* stopped flow pH jump and real-time NMR spectra acquisition (see supplementary information). (B) Ala$^{\beta}$-[$^{13}CH_3$] TET2 spectrum after assembly initiation upon basic pH jump. (C) Ala$^{\beta}$-[$^{13}CH_3$] TET2 intermediate spectrum. (D) Spectrum of the reassembled TET2. Traces in NMR spectra show 1D projection of A192. (E) Build of the non-overlapping peak intensity evolution (blue dots) with fit of double exponential (solid line).

In addition to NMR experiments, we used EM, which visualizes individual molecules as two-dimensional projections, to resolve the heterogeneous oligomeric states during self-assembly. In particular, we applied negative-stain sample preparation to rapidly trap off-equilibrium assembly intermediates at various time-points (*20*). Moreover, we monitored the initial events during the self-assembly process by native MS, in which the experimental settings preserve intact complexes in the gas phase. The combination of NMR with EM and MS allowed us to characterize the oligomerization states and topologies during the self-assembly process. Using these methods, we sought to elucidate the mechanism of TET2 self-assembly.

Low pH induces the disassembly of a TET2 dodecamer into monomers. The methyl Ala (Ala^{β}) spectrum of TET2 in acidic conditions (Fig. 1A) comprises narrow peaks with a ^{1}H and ^{13}C chemical shift dispersion similar to the native dodecamer, but with a different spectral fingerprint (fig. S1). In such conditions, there were no evident particle projections observed in EM. Further biophysical characterization of TET2 in acidic conditions by gel filtration/MALLS and analytical ultracentrifugation revealed a higher than theoretical (39 kDa) molecular mass of 45 kDa (fig. S2). Narrow peaks and chemical shift dispersion in the NMR spectrum indicate the presence of a compact structured monomeric protein, which is corroborated by the lack of distinct particle projections in EM and the determined molecular mass. However, the slight overestimation of molecular mass suggests the presence of a monomer in equilibrium with a small population of the higher molecular weight form.

The self-assembly of TET2 monomers into a dodecamer was initiated by a rapid jump in pH from acidic conditions to a neutral pH. The Ala^{β} NMR spectrum and EM projections of the self-assembled particle are characteristic of the native TET2 dodecamer (Fig. 1C, fig. S3). During the self-assembly process, the intensities of non-overlapping peaks evolved with double exponential growth (Fig. 1E, fig. S4), reflecting the complex formation of (i) native dodecamers and (ii) assembly intermediates with dodecamer-like interfaces.

A flexible monomer is the initial intermediate in the TET2 self-assembly. The first NMR spectrum acquired after initiation of the assembly process contains a significant major peak of broad line width located in the central region (Fig. 2A), which decays during the self-assembly course. This spectrum shows no connection to the spectrum of the monomer, and differs significantly from that of the dodecamer. The broad peak in the center

of the spectrum is typical for an unstructured protein (*21*) and indicates an increase of flexibility in the molecule. The broad peak and dodecamer peaks overlap and influence one another as they simultaneously decay and build, respectively. To extract the decay of the flexible intermediate, the scaled spectrum of the final dodecamer was subtracted from each spectrum of the series (fig. S5), with the scaling factor obtained from the double exponential fit of the dodecamer evolution (Fig. 1E, fig. S4).

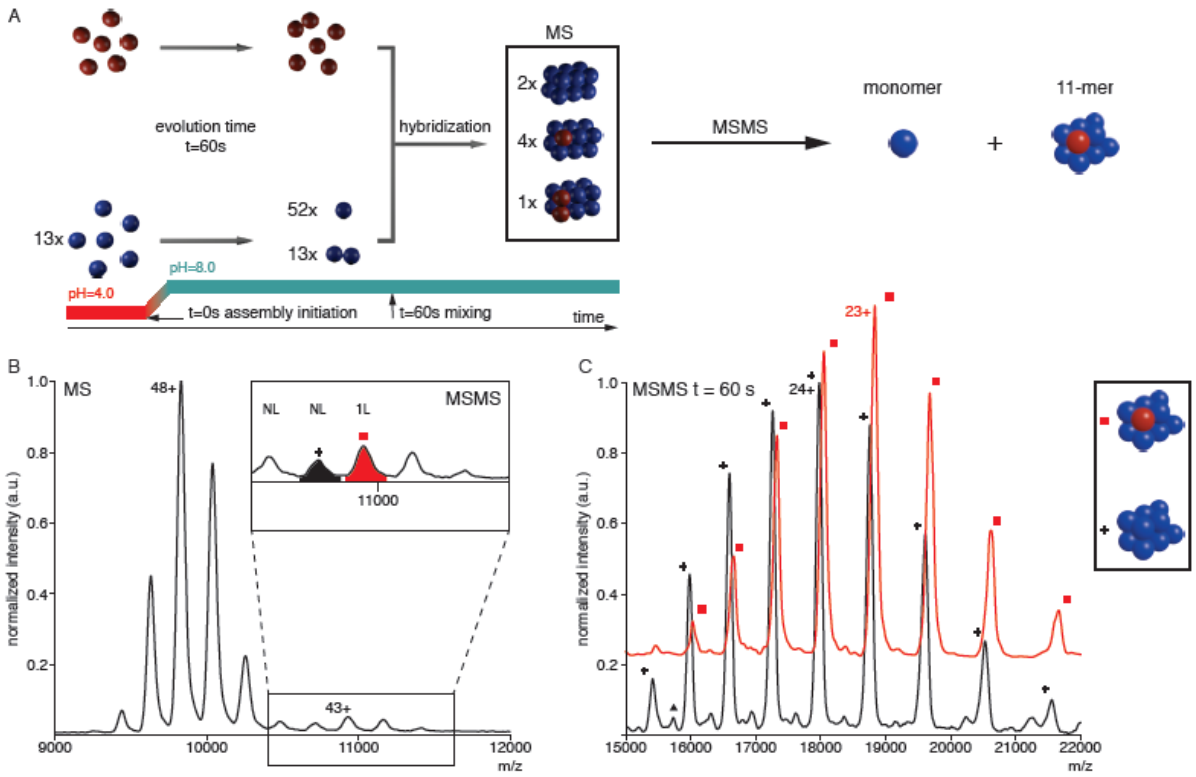


Fig. 2. Characterization of the flexible intermediate by NMR and MS. (A) Schematics of the isotopic chase experiment. The red and blue color represent U-[^{15}N,^{13}C,^{2}H] labeled and non-labeled proteins, respectively. (B) MS spectrum of a self-assembly reaction quenched at 60 s. Insert shows peaks subjected to further MS/MS analysis and their determined isotopic composition (NL – fully non-labelled TET2, 1 L 1 labelled monomer in TET2 complex). (C) Collision induced dissociation of monomer during MS/MS. Tandem MS spectra of the isotopically labeled intermediates chased at 60 s showing parent and 11-mer ions. Black and red correspond to the MS peaks at 10700 and 10900 m/z, respectively. Note the m/z difference between fully non-labeled (black +) and single unit labeled (red square) 11-mer ions.

The peak intensity time-course after the subtraction is typical for the intermediate state, with zero intensity in the starting spectrum, a maximum in the first spectrum after self-assembly initiation, and a decay to zero over time (Fig. 4B).

To uncover the nature of this flexible intermediate, we used EM and MS. EM projections of the sample were stained immediately after the initiation of assembly, and revealed heterogeneous particles of a small size and without well-defined shapes (fig. S6). To

dissect the oligomerization state of the flexible intermediate, we coupled an isotopic chase experiment with native MS analysis. In this method, the oligomeric intermediate is encoded ("chased down") within the native structure by isotopic quenching at different evolution times and hybridization into the final complex (Fig. 2A). The *U*-[^{15}N,^{13}C,^{2}H] isotopic labelling of TET2 significantly shifts the mass of dodecamer to higher m/z values (Fig. 2C and S4). Based on the mass difference after the isotopic quench by the excess of non-labelled reaction at 60 s, we tracked the oligomerization state of the flexible intermediate. Peaks corresponding to the chased intermediates were further analyzed by tandem MS, which allowed us to determine their exact mass and composition (*22*). The chased intermediates contain majority of the single unit labelled and traces of the double unit labelled complexes alongside the non-labelled complex revealing the monomeric state of the initial flexible intermediate (Fig. 2C).

TET2 self-assembly proceeds through the ensemble of oligomeric intermediates. NMR spectra series, after the subtraction of dodecamer evolution, contain an additional set of peaks superimposed on those of the flexible intermediate (Fig. 1C, fig. S5). These additional peaks are located next to, or flanking, the peak positions of the dodecamer, reflecting an increase in peak linewidths in the spectra series before dodecamer subtraction, in comparison to the peaks of the mature dodecamer. The broader peaks represent increased structural heterogeneity of the dodecamer-like structure fold that is characteristic of different oligomerization states and exchanges between different conformations. The time evolution of these additional peaks, characterized by an initial build followed by decay, indicates the presence of a folded intermediate (Fig. 4B). The EM snapshots obtained at a time where the concentration of the folded intermediate was at its maximum show well-structured particles with high heterogeneity in both size and shape (Fig. 3 and fig. S5). Distinct particle projections, such as triangle, V-shaped, horseshoe and square like forms, can be recognized. These projections were strikingly similar to the back projections of various oligomeric states and topologies generated from the TET2 structure (Fig. 3B). Using these projections, we assigned the oligomeric states of the EM particles as tetramer (V-shape), hexamer (triangle, horseshoe), and octamer (square). This outcome confirms that the increased linewidth in the NMR spectra originates from a superposition of different assembly intermediates with various oligomeric states. The initial build rate of this intermediate comprising an ensemble

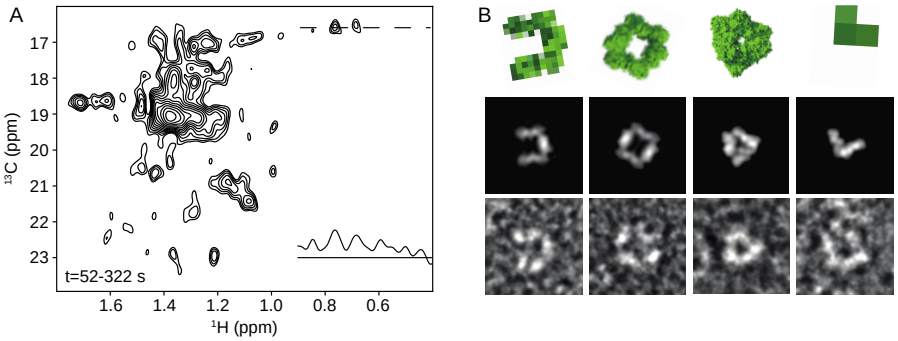


Fig. 3. TET2 self-assembly intermediate states. (A) Sum of the dodecamer subtracted spectra between 52 and 322 s, including the 1D projection of native TET2 A192. (B) EM snapshots of oligomeric intermediates captured at 2 min. (bottom row) and back projections (middle row) of corresponding oligomeric structures (top row). From right to left, tetramer, triangular and horseshoe like hexamers, and octamer.

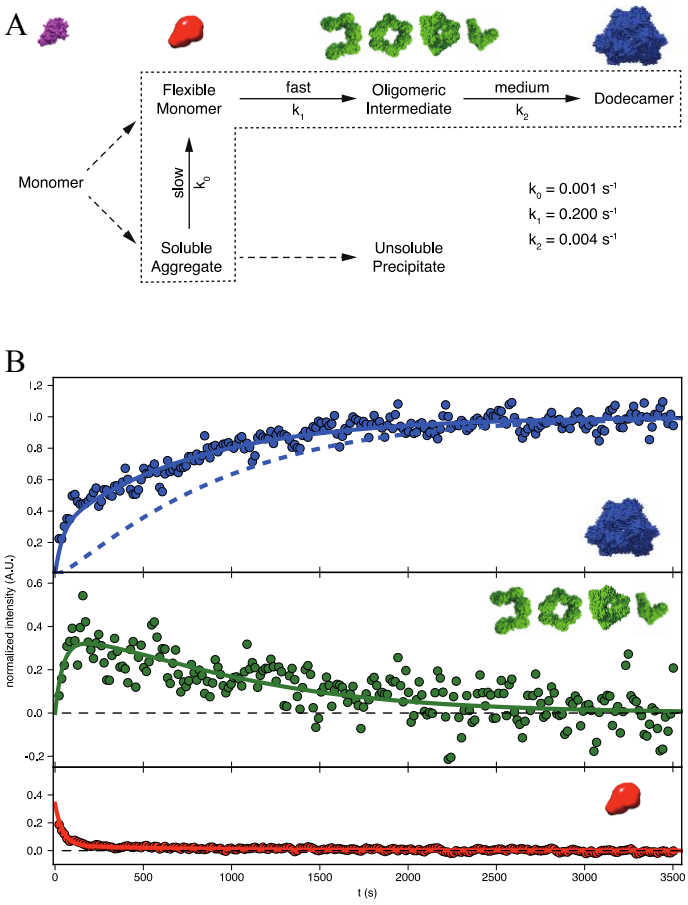


Fig. 4. (A) Fit of burst-phase four state model into NMR acquired progress curves. Solid blue line represents the sum of dodecamer evolution (blue dashed line) with ensemble of oligomeric intermediates (green). Flexible intermediate evolution is shown in red. (B) Full TET2 self-assembly model. Dotted area surrounds the states included in the burst-phase four state fit. Dashed arrows indicate unobservable fast transition of monomer to pre-equilibrium between the flexible intermediate and soluble aggregate and accumulation of solution NMR invisible precipitate.

which implies that the flexible monomer is transformed into an ensemble of assembly of oligomeric states corresponds to the initial decay rate of the flexible monomer (Fig. 4A),intermediates of various oligomerization states and topologies.

The flexible monomer is in pre-equilibrium with the soluble aggregate. To determine the kinetic model of TET2 assembly, the NMR derived progress curves of flexible monomer, oligomeric intermediate and dodecamer were fit simultaneously (Fig. 4A). The minimal kinetic model that fit all datasets (Fig. 4B) comprises (i) pre-equilibrium between the flexible intermediate and the NMR invisible soluble aggregate, (ii) the ensemble of oligomeric intermediates and (iii) the final dodecamer. The model includes burst formation of an ensemble of oligomeric intermediates from the equilibrium between the flexible monomer and soluble aggregate. The existence of the soluble aggregate was inferred from the presence of a macroscopic precipitate in the reassembled sample. Direct experimental evidence of a soluble aggregate was obtained from EM and analytical ultracentrifugation, where the aggregation appears in the acid disassembled TET2.

The overall kinetic model (Fig. 4B) includes an acid stabilized monomer, which is transformed into the flexible monomeric intermediate in equilibrium with the soluble aggregate at a rate beyond the time resolution of fast acquisition NMR. The flexible monomer is subsequently stabilized in an oligomeric intermediate that assembles into the final dodecamer. The transition of the acid stabilized monomer into the flexible intermediate, which subsequently collapses into the natively folded structure bound in the oligomeric intermediate, implies the off-pathway nature of the acid stabilized monomer (*23*). Self-assembly of the TET2 dodecamer from the pre-equilibrium between the flexible monomer and soluble aggregate follows a sigmoidal evolution (fig. 4A) and comprises events with fast, medium and slow kinetics events. The fast event is the transition of the flexible monomer into the ensemble of properly folded assembly intermediates of various oligomeric states, the medium event represents the evolution of oligomeric intermediates into the final dodecamer and the slow step describes the partial replenishment of the flexible monomer pool from the soluble aggregate. The soluble aggregate belongs to the misassembled pathway, which leads to the formation of a precipitate.

The dimer is the building block of TET2 self-assembly. The EM determined intermediates contain an even number of subunits with the topology of a broken tetrahedron with preserved edges, which indicate that the dimer is a building block of TET2 self-assembly (fig. 3B). The self-assembly can be viewed as an evolution over long time scales

with a sequence of steps characterized by the interface size (*3*). In this model, the hierarchy of interface size and strength defines the order during the self-assembly. Therefore, the comparison of TET2 dimerization and trimerization interfaces, which comprise 38 and 20 aminoacids (*26*), suggest an initial formation of the dimer, followed by the random association of dimers into the final dodecamer. In addition, *in vivo* mutation studies (*27*) support the conclusion that the dimer is formed first and serves as the building block in TET2 self-assembly.

Multiple assembly pathways lead to the TET2 dodecamer. The ensemble of assembly intermediates comprises different oligomerization states and topologies. The presence of intermediates with the same oligomerization state, but different topologies, implies that they belong to different assembly pathways (e.g., hexamers with triangle and horseshoe form) (fig. 3B). The presence of multiple assembly pathways indicates that the assembly proceeds through parallel sequential addition or completely stochastic assembly (*24*, *25*). The lack of a single kinetic pathway in the stochastic mechanism is distinct from the self-organization into dihedral oligomers, which progresses by serial second order associations of two identical well defined intermediates (*2*, *3*). During stochastic assembly, the building block is formed first through the association of larger interfaces, and is then followed by the random association of the building block with equally sized smaller interfaces to form the final structure. In contrast, the hierarchy of interfaces in the dihedral complexes dictates the order of sequential association of two identical intermediates. Therefore, we can conclude, that in the final topology, stoichiometry as well as the mechanism of assembly of large oligomeric complexes is structurally encoded in complementary interactions at the building block interfaces (*25*).

Our study demonstrated that the key processes of self-assembly into high order structures can be described and understood through direct observation at the structural level. These results highlight the potential of integrating real-time NMR, EM and MS for increasing our understanding of protein self-assembly and its use in the development of novel therapeutic approaches and the rational design of artificial nanosystems.

References and Notes:

1. G. M. Whitesides, B. Grzybowski, Self-assembly at all scales, *Science* **295**, 2418–2421 (2002).

2. R. Jaenicke, H. Lilie, Folding and association of oligomeric and multimeric proteins, *Adv. Protein Chem.* **53**, 329–401 (2000).

3. E. D. Levy, E. Boeri Erba, C. V. Robinson, S. A. Teichmann, Assembly Reflects Evolution of Protein Complexes, *Nature* **453**, 1262–1265 (2008).

4. F. H. Crick, J. D. Watson, Structure of small viruses, *Nature* **177**, 473–475 (1956).

5. A. Zlotnick, To Build a Virus Capsid: An Equilibrium Model of the Self Assembly of Polyhedral Protein Complexes, *Journal of Molecular Biology* **241**, 59–67 (1994).

6. D. Endres, A. Zlotnick, Model-Based Analysis of Assembly Kinetics for Virus Capsids or Other Spherical Polymers, *Biophysical Journal* **83**, 1217–1230 (2002).

7. D. Han *et al.*, DNA origami with complex curvatures in three-dimensional space, *Science* **332**, 342–346 (2011).

8. N. P. King *et al.*, Computational Design of Self-Assembling Protein Nanomaterials with Atomic Level Accuracy, *Science* **336**, 1171–1174 (2012).

9. G. Bhabha *et al.*, A Dynamic Knockout Reveals That Conformational Fluctuations Influence the Chemical Step of Enzyme Catalysis, *Science* **332**, 234–238 (2011).

10. G. Schreiber, Kinetic studies of protein-protein interactions, *Curr. Opin. Struct. Biol.* **12**, 41–47 (2002).

11. E. Rennella *et al.*, Real-Time NMR Characterization of Structure and Dynamics in a Transiently Populated Protein Folding Intermediate, *J. Am. Chem. Soc.* **134**, 8066–8069 (2012).

12. N. Fischer, A. L. Konevega, W. Wintermeyer, M. V. Rodnina, H. Stark, Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy, *Nature* **466**, 329–333 (2010).

13. N. Morgner, F. Montenegro, N. P. Barrera, C. V. Robinson, Mass Spectrometry—From Peripheral Proteins to Membrane Motors, *Journal of Molecular Biology* **423**, 1–13 (2012).

14. L. Borissenko, M. Groll, Crystal Structure of TET Protease Reveals Complementary Protein Degradation Pathways in Prokaryotes, *Journal of Molecular Biology* **346**, 1207–1219 (2005).

15. C. Amero *et al.*, Fast Two-Dimensional NMR Spectroscopy of High Molecular Weight Protein Assemblies, *J. Am. Chem. Soc.* **131**, 3448–3449 (2009).

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

17. G. Cornilescu, F. Delaglio, A. Bax, Protein backbone angle restraints from searching a database for chemical shift and sequence homology, *J Biomol NMR* **13**, 289–302 (1999).

18. I. Ayala, R. Sounier, N. Usé, P. Gans, J. Boisbouvier, An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein, *J Biomol NMR* **43**, 111–119 (2008).

19. P. Schanda, V. Forge, B. Brutscher, Protein folding and unfolding studied at atomic resolution by fast two-dimensional NMR spectroscopy, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11257–11262 (2007).

20. A. M. Mulder *et al.*, Visualizing ribosome biogenesis: parallel assembly pathways for the 30S subunit, *Science* **330**, 673–677 (2010).

21. A. B. Sahakyan, W. F. Vranken, A. Cavalli, M. Vendruscolo, Structure-based prediction of methyl chemical shifts in proteins, *J Biomol NMR* **50**, 331–346 (2011).

22. M. Sharon, Structural MS Pulls Its Weight, *Science* **340**, 1059–1060 (2013).

23. Y. Bai, Kinetic evidence for an on-pathway intermediate in the folding of cytochrome c, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 477–480 (1999).

24. M. Hemberg, S. N. Yaliraki, M. Barahona, Stochastic Kinetics of Viral Capsid Assembly Based on Detailed Protein Structures, *Biophysical Journal* **90**, 3029–3042 (2006).

25. A. J. Olson, Y. H. E. Hu, E. Keinan, Chemical mimicry of viral capsid self-assembly, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 20731–20736 (2007).

26. E. D. Levy, J. B. Pereira-Leal, C. Chothia, S. A. Teichmann, 3D Complex: a Structural Classification of Protein Complexes, *PLoS Comp Biol* **2**, e155 (2005).

27. A. Appolaire *et al.*, Pyrococcus horikoshii TET2 peptidase assembling process and associated functional regulation, *Journal of Biological Chemistry* (2013), doi:10.1074/jbc.M113.450189.

Title: Probing Protein Self-Assembly by Real-Time NMR

Authors: P. Macek[1], R. Kerfah[1], E. Boeri Erba[1], C. Moriscot[1], E. Crublet[1], G. Schoehn[1], C. Amero[2], J. Boisbouvier[1]*

Affiliations:
[1]Institut de Biologie Structurale-J.P. Ebel, 41 Rue Jules Horowitz, 38027 Grenoble, France.
[2]Centro Investigaciones Quimicas, Av. Universidad 1001, CP 62210 Cuernacava, Mexico.
*Correspondence to: jerome.boisbouvier@ibs.fr

Supplementary Materials:

Materials and Methods

Protein expression and purification:

Preparation of 2-[^{2}H], 3-[^{13}C]L-alanine: The labeled alanine was synthesized using a procedure described in (*1*). The α-proton of the ^{13}C-labeled L-alanine, commercially available, was exchanged into deuterium by Tryptophane synthase-rich cell lysate prepared from LB culture of E.coli transformed with pSTB7. A quantity of 1.0g of ^{13}C-labeled L-alanine was dissolved in 60 ml of a 50mM Tris solution in D2O at pH 7.6, in presence of 1.0mg of pyridoxal phosphate (sigma). Lyophilized lysate, corresponding to 1.0L of E.coli culture was added and the mixture incubated at 35ºC for 24h. After lyophilization, the resulted lysate was resuspended in 60ml D2O and incubated for 24h at 35ºC for complete exchange. The reaction was stopped and Tryptophane synthase precipitated by heating to 80ºC for 30min. Finally, the 2-[^{2}H], 3-[^{13}C] L-alanine was recovered in the supernatant after centrifugation. The completion of the H/D exchange was confirmed by 1H- NMR spectroscopy (1D). Data were acquired at 25ºC in D_2O.

PhTET2 production and purification: Expression of *U*-[^{2}H], *U*-[^{15}N], *U*-[^{12}C], *U*-[$^{13}C^{1}H_3$]-Ala-β PhTET2 was performed in the E.coli BL21(DE3) RIL strain. The detailed protocol is described in (*2*). E. coli BL21(DE3) RIL carrying the pET41c plasmid with the PhTET2 were grown in three stages during 24 h. to progressively adapt to M9/$D_2$0 media. Constituents of the basic M9 medium (Sigma) were anhydrous. All of the [^{15}N] NH_4Cl, [1,2,3,4,5,6,6-$^{2}H_7$, *U*-^{12}C] glucose, kanamycine, IPTG and M9 complements were resuspended in D2O (99.85%) and lyophilized, twice. The final culture was grown at 37°C in M9 media prepared with 99.85% D_2O. When the O.D. (600 nm) reached 0.8, a D2O solution containing 2-[^{2}H], 3-[^{13}C] L-alanine and other perdeuterated precursors was added. After addition the concentrations in culture medium were: 800 mg/L of 2-[2H], 3-[13C] L-alanine; 2.5 g/L of succinate-d4 (Cambridge Isotope Laboratories, Inc.); 200 mg/L for 2-ketoisovalerate-d7 (CDN Isotopes, Inc.); and 60 mg/L for isoleucine-d10 (Cambridge Isotope Laboratories, Inc.). After 1 hour, protein expression was induced by the addition of IPTG to a final concentration of 0.5 mM, then cells were grown at 37°C during 4 hours before harvesting. Finally, cells were harvested by centrifugation.

Cell pellet was resuspended and lysed in buffer containing 50 mM Tris, 150 mM NaCl, 0.1% Triton X-100, 0.25 mg/ml Lysozyme, 0.05 mg/ml of DNAse, 20 mM MgSO4 and 0.2 mg/ml of RNAse, at pH 8. Cells were disrupted by decompression carried out by Microfluidizer using three passes at 15000psi. The crude extract was heated at 85ºC for 15min. and then centrifuged at 17500rcf for 1h. at 4ºC. The supernatant was dialyzed overnight at room temperature against 20mM Tris, 100 mM NaCl at pH 7.5. The dialyzed extract was centrifuged at 17500rcf for 10 minutes at 4ºC and supernatant was loaded on Resource Q column (GE Healthcare). The PhTET2 was eluted with linear gradient (0 – 1 M NaCl in 20mM Tris, pH 8 over 10CV). The fractions containing protein with similar mass (39 kDa) according to SDS- PAGE (12.5% polyacrylamide) were pooled and concentrated using an Amicon cell (Millipore) with a molecular mass cutoff of 30kDa. The protein solution was then loaded onto a Superdex 200 HiLoad 16/60 column (GE Healthcare) equilibrated by 20mM Tris, pH 8, 100mM NaCl.

Diassembly and Reassembly:

Dodecameric PhTET2 was disassembled into monomer by Zn precipitation - resolubilization procedure. The 0.2mM PhTET2 in 20mM Tris, 20mM NaCl, pH 7.8 in D_2O, was precipitated by 3M ZnSO4, pH 4,2 in D2O, in a volume ratio of 1:7 ($ZnCl_2$: protein). Then, the resolubilization of PhTET2 solution was carried out by the chelation of Zn^{+2} by 70mM EDTA in 20mM Tris, 20mM NaCl, pH 6,0 in D_2O, in a volume ratio of 1:15 (protein:EDTA).
PhTET2 monomer after diasassembly was exchanged to 20mM Tris-D_{11}, 20mM NaCl, pH 4.0 in D_2O. The final NMR sample comprise 350μL of 250μM PhTET2 in was supplemented with $ZnCl_2$ in D_2O, pH 4.0 to final concentration 500μM. The self-assembly was initiated by stopped flow pH jump within magnet (*3*) by rapid injection of 75μL of 283mM Tris-D_{11}, pH 8.0 in D_2O.

NMR Data Acquisition and processing:

All NMR experiments were performed on an INOVA spectrometer (Varian, Palo Alto, CA) operating at 800MHz ^1H frequency and equipped with a cryogenic probe. All experiments weree acquired at 50°C. SOFAST-methyl-TROSY (*2*) was acquired with PC9 shape pulse of flip-angle 30° and length δ = 4.26ms. The d1 delay set to 289.7ms. Each spectrum in kinetic dimension was acquired with spectral width of 1600Hz and 20 complex points in 13C dimension. The total acquisition time of one 2D spectrum was 14.4s. Each experiment was performed with only one scan per increment in t1, ensuring highest repetition rates of experiments, and the sign of the ^{13}C excitation pulse phase and receiver phase in SOFAST-methyl-HMQC was alternated between subsequent experiments.
Spectra processing and analysis were performed usig nmrPipe (*4*) and nmrglue (*5*) module of python programming language (*6*). Two succeeding spectra (1 and 2, 2 and 3, 3 and 4, . . ., n-1 and n) along the kinetic time were added prior processing of 2D kinetic planes. Such data treatment removes spectral artifacts such as axial peaks or t1 noise at the water frequency and improves the base line, allowing a more accurate and precise measurement of spectral parameters, such as peak positions and intensities (*3*).

Fitting:

Numerical modeling and fitting were performed using numpy and scipy (*7*) python modules.

The normalized build-up of dodecamer and dodecamer-like species was fitted with double exponential (fig. S4) and used for substraction of dodecamer/dodecamer-like contribution from the kinetic spectra serie.

The reaction steps of the TET2 self-assembly process were defined as ν0 (release of monomer from the soluble aggregate), ν1 (sum of steps leading to ensemble of oligomeric intermediates from monomer) and ν2 (formation of mature dodecamer). Due to the significant differences in monomer, flexible intermediate and dodecamer spectra and measured timescales much smaller than diffusion limit ($k_a << 10^5$) , all steps were considered as conformation controlled and therefore modeled as a first order reactions.

$$\nu 0 = k_0 \cdot \mathrm{Ag} - k_{0r} \cdot \mathrm{Mo} \quad (\mathrm{Ag} \leftrightarrow \mathrm{Mo})$$

$$\nu 1 = k_1 \cdot \mathrm{Mo} - k_{1r} \cdot \mathrm{Ol} \quad (\mathrm{Mo} \leftrightarrow \mathrm{Ol})$$

$$\nu 2 = k_2 \cdot \mathrm{Ol} - k_{2r} \cdot \mathrm{Do} \quad (\mathrm{Ol} \leftrightarrow \mathrm{Do})$$

The changes in concentrations are given by

$$\mathrm{Ag} = -\nu 0$$

$$\mathrm{Mo} = \nu 0 - \nu 1$$

$$\mathrm{Ol} = \nu 1 - \nu 2$$

$$\mathrm{Do} = \nu 2$$

Where Ag, Mo, Ol, and Do represent the concentrations of soluble aggregate, flexible monomer, ensemble of oligomeric states, and dodecamer. The dissociation rate constants were assumed negligibly small (k_{nr} = 0) under consideration of significant dodecamer stability (*8*, *9*) and low probability of oligomer dissociation coupled to unfolding into flexible monomer. The initial aggregate to flexible intermediate equilibrium was included as the fit parameter. As the peak intensities depend on the protein correlation time and internal dynamics and their significant differences between self-assembly intermediates and dodecamer, the fitted intermediate progression curves were linearly scaled to the corresponding datasets of normalized peak intensities within the fitting procedure.

Figures:

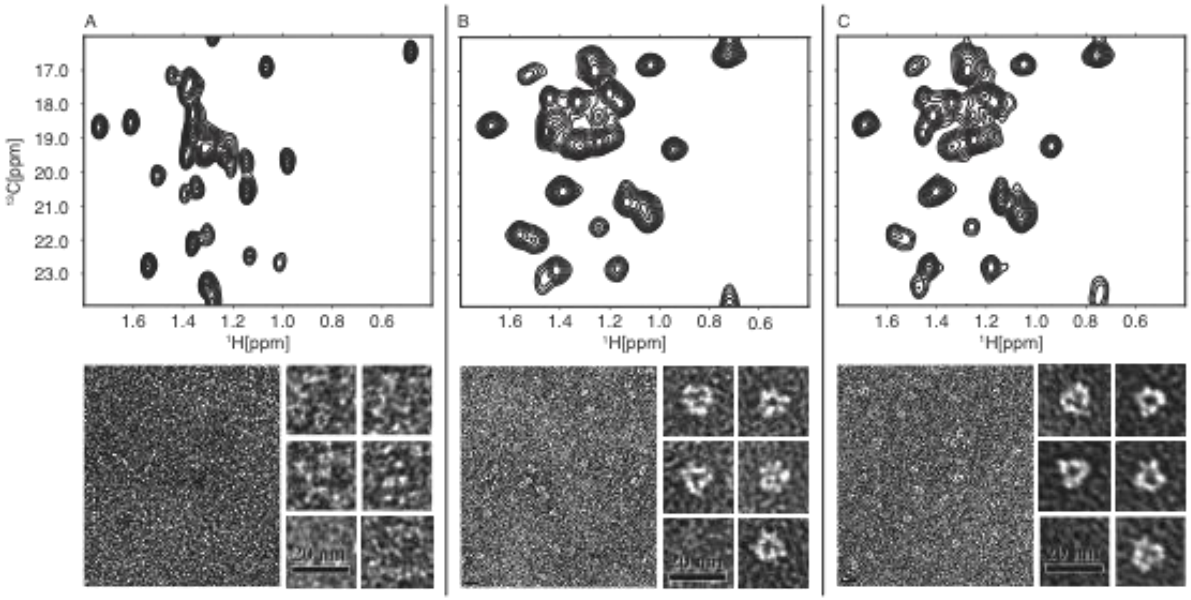


fig. S1. TET2 Ala$^{\beta}$-[$^{13}CH_3$] HMQC spectrum and EM projection of (A) diassembled monomer, (B)resassembled dodecamer, and (C) native dodecamer.

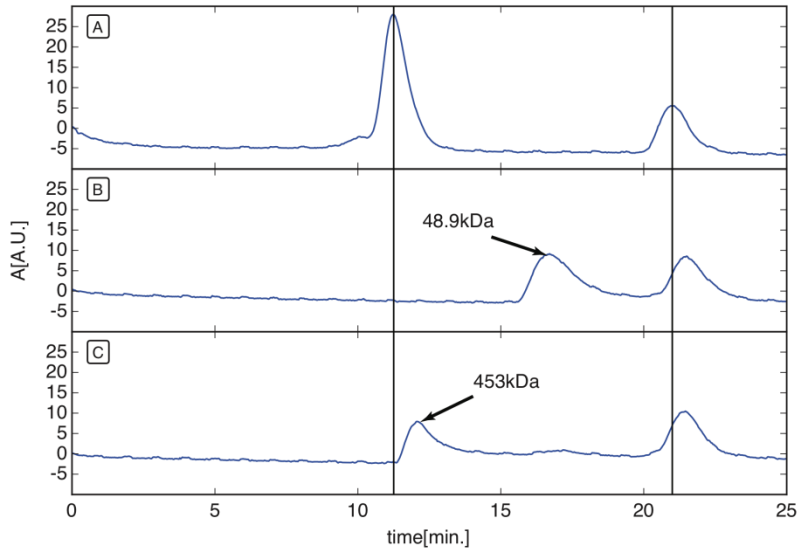


fig. S2. Gel filtration/MALLS characterization of TET2 disassembly/reassembly. (A) TET2 before disassembly T20N20 pH7.4, (B) disassembled TET2 AcAc20 N100 pH4.0, and (C) reassembled TET2 T20N20 pH7.4.

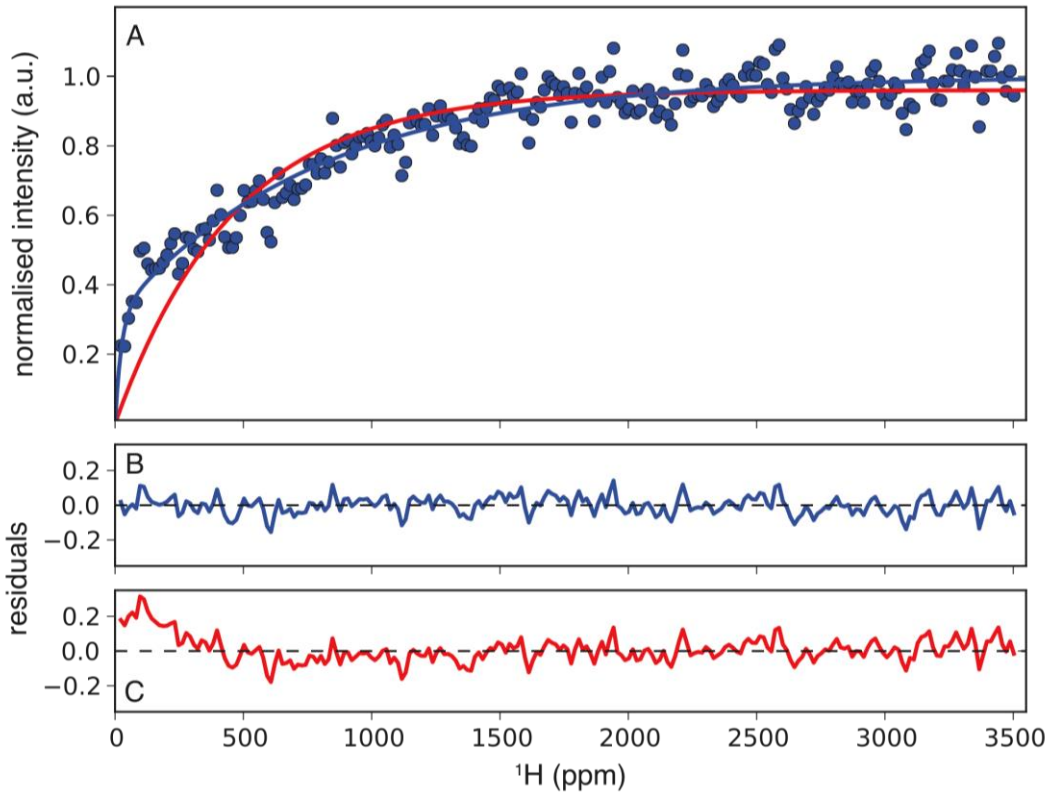


fig. S3. (A) Double and single exponential fit of dodecamer/dodecamer-like specie evolution used for the substraction of the dodecamer build from kinetic spectra series. (B, C) Residuals of double and single exponential fit.

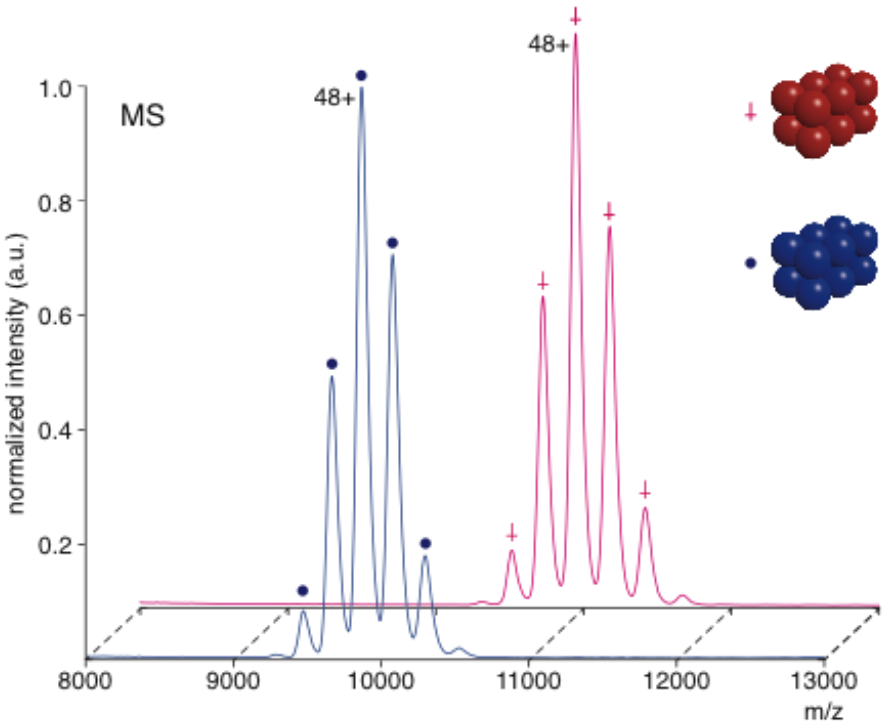


fig. S4. MS spectra of TET2 at natural abundance (blue) and fully HCN labeled (red) TET2

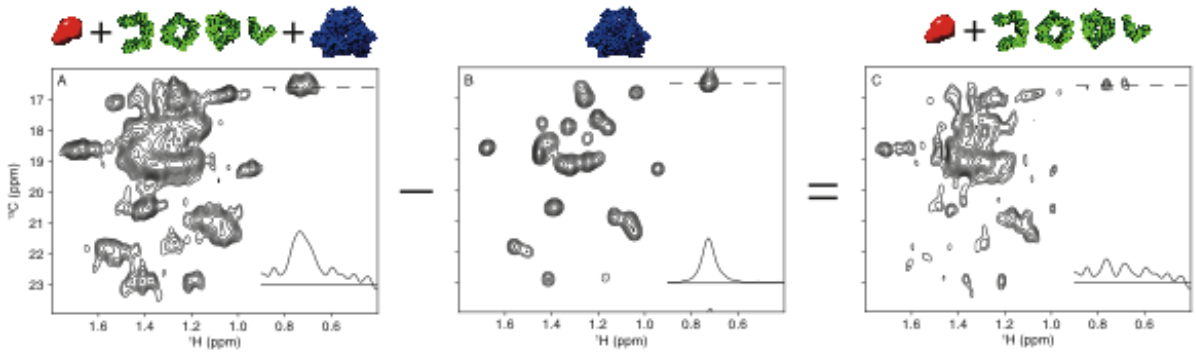


fig. S5. Subtraction of the dodecamer evolution from the kinetic spectra series for deconvolution of the intermediate evolution from the dodecamer build-up. Traces show the position of the native TET2 A192. Note the increased linewidth in the NMR spectrum during the self-assembly time-course (off-equilibrium) (A) in comparison to NMR spectrum of reassembled TET2 in steady-state (B).

-

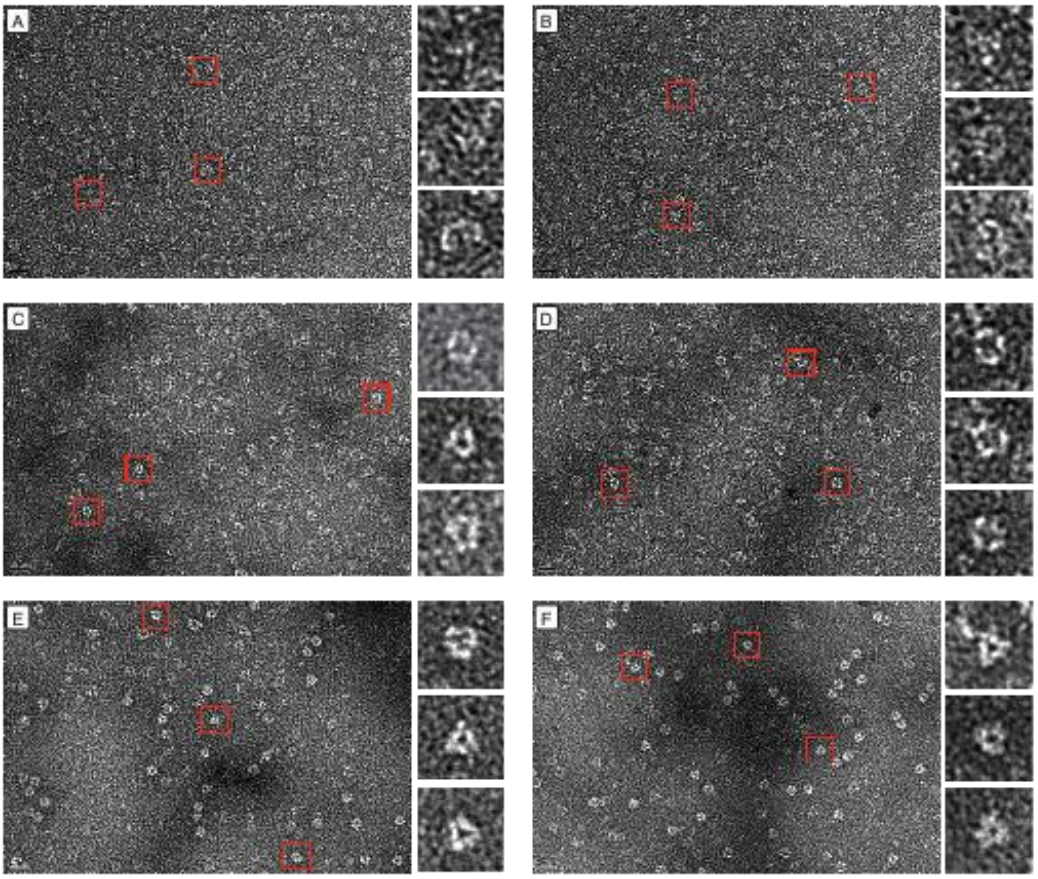


fig. S6. EM images of the order increase during the course of TET2 self-assembly. (A, B, C) flexible intermediate, (D, E, F) oligomeric intermediate, and (G, H, I) reassembled dodecamer.

References and Notes:
1. I. Ayala, R. Sounier, N. Usé, P. Gans, J. Boisbouvier, An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein, *Journal of Biomolecular NMR* 43, 111–119 (2009).

2. C. Amero *et al.*, Fast Two-Dimensional NMR Spectroscopy of High Molecular Weight Protein Assemblies, *Journal of the American Chemical Society* 131, 3448–3449 (2009).

3. P. Schanda, V. Forge, B. Brutscher, Protein folding and unfolding studied at atomic resolution by fast two-dimensional NMR spectroscopy, *Proceedings of the National Academy of Sciences* 104, 11257–11262 (2007).

4. F. Delaglio *et al.*, NMRPipe: a multidimensional spectral processing system based on UNIX pipes., *Journal of biomolecular NMR* 6, 277–93 (1995).

5. J. J. Helmus, C. P. Jaroniec, Nmrglue: an open source Python package for the analysis of multidimensional NMR data, *Journal of Biomolecular NMR* , 355–367 (2013).

6. A. B. Downey, *Think Python How to Think Like a Computer Scientist* (O'Reilly Media, 2012).

7. E. Jones, T. Oliphant, P. Peterson, and Others, SciPy: Open Source Scientific Tools for Python.

8. E. Rosenbaum, M. Ferruit, M. A. Durá, B. Franzetti, Studies on the parameters controlling the stability of the TET peptidase superstructure from Pyrococcus horikoshii revealed a crucial role of pH and catalytic metals in the oligomerization process., *Biochimica et biophysica acta* 1814, 1289–94 (2011).

9. E. Rosenbaum *et al.*, Effects of hydrostatic pressure on the quaternary structure and enzymatic activity of a large peptidase complex from Pyrococcus horikoshii., *Archives of biochemistry and biophysics* 517, 104–10 (2012).

4.2 Combinatorial methyl groups labeling for long-range nOes detection

4.2.1 Context

The Nuclear Overhauser Effect (nOe) is one of the most important observables used for the 3D structure determination of proteins by solution NMR, providing the major source structural restraints. The Overhauser effect occurs via a dipole-dipole interaction where two proximal protons mutually transfer magnetization through space. In an isolated pair of protons, its intensity is inversely proportional to the sixth power of their distance from each other.

In a protonated protein, the longest ^{1}H-^{1}H distance from which nOe is still detectable is approximately 6 Å. The detection of longer-range nOes is primarily hampered by the fact that the dipole-dipole interaction becomes weaker (due to the r^{-6} dependence) and also hindered by the existence of high proton density, which leads to spin diffusion. Indeed, efficient transfers that occur by relay truncate the magnetization arising from weak long-range ^{1}H-^{1}H nOe interactions. Thus, this phenomenon prevents the use of long mixing times, a factor that favors the detection of the desired weak interactions. Due to spin diffusion, the interpretation of nOes is only semi-quantitative in protonated proteins.

The easily detectable short-range nOes at 3-6 Å (separated by ~3 residues, on average) provide important local information, which is sufficient for the high-resolution 3D structure determination of small and globular proteins. This situation ceases to be true when the protein target is non-globular (e.g., elongated, modular or complexes). In fact, additional information about the relative position of secondary structure elements with respect to each other is a fundamental necessity. For instance, distances between two amides (d_{NN}) of less than 6 Å are required to properly pack α-helices against β-sheets or β strands together, while longer distances of approximately 7-8 Å are needed to pack two α-helices with respect to each other (Koharudin, Bonvin et al. 2003).

To detect long-range nOes, spin diffusion must be reduced. Therefore, the use of perdeuterated proteins that are selectively protonated at the amide groups was proposed as a solution (Mal, Matthews et al. 1998, Koharudin, Bonvin et al. 2003). Experimentally, this

approach permits the detection of up to 80% of the expected nOes between protons separated by up to 7-8 Å in small systems (Mal, Matthews et al. 1998, Koharudin, Bonvin et al. 2003). The exploitation of the detected long-range nOes from the amide protons supplemented with the dihedral angle restraints derived from ^{13}C and ^{15}N experiments, together with the information from the ^{1}HN secondary chemical shifts, was sufficient to obtain a reliable fold model of chymotrypsin inhibitor type 2 with a backbone RMSD of 1.3 ±0.3 Å (compact, α/β protein that consists of 64 amino acids)(Koharudin, Bonvin et al. 2003).

Later, in an attempt to detect nOes between protons separated by more than 8 Å, a heavier dilution of the total proton density in the protein was proposed (Sounier, Blanchard et al. 2007). For this purpose, the selective protonation at methyl groups of a specific residue type was demonstrated as the best choice for the following reasons: (*i*) due to proton multiplicity, the transferred magnetization between two remote methyl groups is nine-fold more intense than other proton pairs that are equally separated. (*ii*) The relaxation properties (R_1 and R_2) of the methyl groups are highly favorable in perdeuterated systems, giving rise to high resolution and signal intensity. (*iii*) The spin diffusion is drastically reduced. In fact, the total proton density decreases from ~20%, the average fraction of the exchangeable protons (e.g., OH, NH_2, NH, etc.) in perdeuterated proteins, to ~2%, where only one type of methyl group (i.e., Ile-δ_1) is protonated in perdeuterated protein. Thanks to this labeling, accurate nOe-derived distance restraints arising from protons that are separated by up to 12 Å were reported on two small systems (9 and 19 kDa). Given that the total proton density is kept at 2-3%, spin diffusion was estimated to contribute to ~10% of the nOe signal intensity in small proteins (Sounier, Blanchard et al. 2007).

Long-range nOes are easily obtained when a single type of methyl group is protonated (2% proton density). However, the number of extracted distance restraints in this case is very restricted, rendering the applicability of this approach very limited. For instance, for Ile-δ_1-MSG, which contains 44 Ile residues, the theoretical number of nOes derived from distances between 8.5 to 10.5 Å is only 17 (nOes pairs). Notably, ~50% were experimentally detected, making the number of available nOes down to 8 (Sounier, PhD thesis, 2008). Therefore, combinatorial labeling of more than one amino acid appears to be appropriate for increasing the total number of detectable long-range nOes.

In the literature, previous studies reported this possibility using either $I^{\delta 1}LV$ (Tugarinov, Choy et al. 2005) or $(LV)^{proS}$ labeling schemes in MSG (Gans, Hamelin et al. 2010) . Analysis of the corresponding 4D HMQC–NOESY–HMQC data shows that up to ~27 % of the 110 expected long-range nOe pairs arising from the protons separated by 8.5-10.5 Å in *U*-[^{2}H], (LV) [$^{13}C^{1}H_3$]^{proS}-MSG are still observable (Gans, Hamelin et al. 2010). This decrease (from ~50 % in [*U*-^{2}H, $I^{\delta 1}$]-MSG to 27% in [*U*-^{2}H, $(LV)^{proS}$]-MSG) in the experimentally detectable long-range nOes is primarily due to the different longitudinal relaxation properties of the monitored amino acids (Ile and Leu/Val). However, by analyzing these results we can conclude that increasing the number of methyl group probes (from 44 Ile to 104 Leu and Val) by 2.5-fold in MSG, resulted in the augmentation of the number of the obtained nOes by a factor of ~3.75. Therefore, combinatorial labeling is an interesting path to be explored and optimized for maximizing the distance information that can be experimentally acquired.

In this chapter, [*U*-^{2}H, $AI^{\delta 1}(LV)^{proS}$]-MSG is used as a model to demonstrate the strenght of the combinatorial labeling. Despite the high proton density, the labelling of more than two amino acids is still useful for obtaining long-range distance restraints. Application of this approach to supra-molecular proteins is shown using PhTET-2 (468 kDa) as an exemple.

4.2.2 Combinatorial labeling of methyl groups for the detection of long-range nOes in perdeuterated large systems

Principles of the Nuclear Overhauser Effect

The rate of cross-relaxation (origin of the Overhauser effect) between two methyl groups can be described as follows:

$$\sigma_{AB} = \frac{1}{10}\frac{1}{\langle r^3\rangle^2}\left(\frac{\mu_0}{4\pi}\gamma^2\hbar\right)^2 [6J(2\omega_H) - J(0)] \quad \textbf{(4.1)}$$

A and B: two methyl groups

$\langle r^3\rangle$: average third power of the ^{1}H-^{1}H distance between the A and B methyl groups

$\mu_{0:}$ magnetic field constant

γ: proton gyromagnetic ratio

ω_H: proton Larmor frequency

$\hbar$: Planck constant divided by 2π

J(ω): spectral density that informs about the motion at frequency ω

For the simple case of wobbling in a cone, the spectral density of a ^{1}H-^{1}H vector within a protein with isotropic movement is formulated as follows:

$$J(\omega) = S^2 \left(\frac{\tau_c}{1+(\omega\tau_c)^2}\right) \qquad \textbf{(4.2)}$$

where S^2 is the parameter order and τ_c is correlation time

For long-range nOes (≥ 6 Å), we can approximate that the order parameter (S^2) tends to be 1. Furthermore, at high magnetic fields and for large proteins, the $J(2\omega_H)$ contribution is negligible compared to J(0). Taking these two approximations into account, 4.2 can be written as follows:

$$\sigma_{AB} = \frac{-\tau_c}{\langle r^3\rangle^2} \left(\frac{\mu_0}{4\pi}\gamma^2\hbar\right)^2 \qquad \textbf{(4.3)}$$

Finally, we can conclude that:

$$\sigma_{SI} \propto \frac{\tau_c}{\langle r^3\rangle^2} \qquad \textbf{(4.4)}$$

In conclusion, the magnetization transfer between two isolated methyl groups via a cross-relaxation mechanism is theoretically only dependent on the correlation time of the protein (τ_c) and the sixth power of their distance from each other $\langle r^3\rangle^{-2}$. Therefore, nOe magnetization transfer is as efficient as the protein size increase. Nevertheless, this is more complex because spin diffusion, longitudinal relaxation R_1 and transverse relaxation R_2 limit the detection of long-range nOes (see fig. 4.2.2). To favor nOe observation in large systems, certain parameters of the NOESY experiment, such as mixing time (τ_m), need to be optimized to take longitudinal relaxation properties into account.

Optimization of the NOESY experiment:

For the large proteins NMR study, the high signal overlap is unavoidable. Therefore, the use of 4D HMQC–NOESY–HMQC was previously proposed to cope with this poor resolution (see fig. 4.2.1)(Tugarinov, Kay et al. 2005). However, the edition of a supplemental dimension in the pulse sequence leads to the loss of sensitivity by a factor of $\sqrt{2}$. Thus, the acquired NMR signal by a 4D experiment is expected to be diminished by a factor of 1.4 compared to its intensity in a 3D experiment.

Moreover, to optimize the resolution of a 4D experiment, an extremely long experimental time is required. Therefore, a compromise should be instituted, e.g., reduction in d_1 and editing time, where the signal sensitivity is necessarily reduced. After weighing the advantages and disadvantages of this 4D experiment, its 3D version was used because it is easier to analyze, and optimized resolution and sensitivity can be obtained. In my work, I used 2 variants of the 3D experiment, HCH-edited NOESY for the MSG (82 kDa) protein and to enhance resolution, HCC-edited NOESY was applied for the PhTET-2 system (468 kDa).

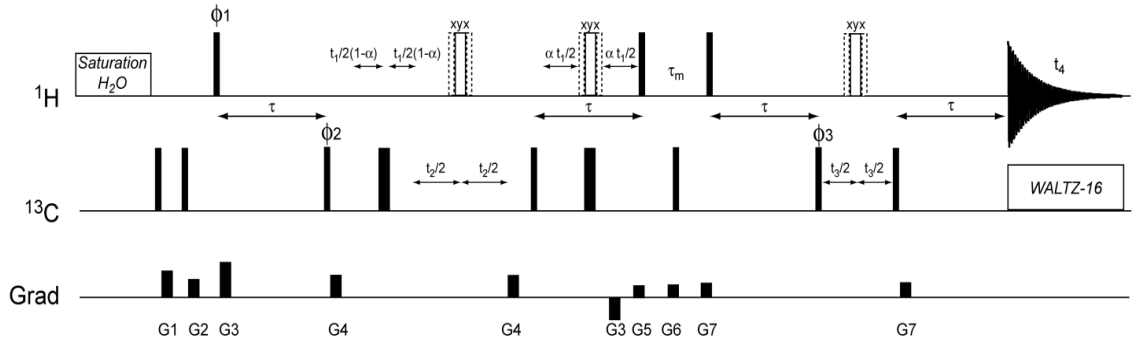


Figure 4.2.1: 4D HCCH NOESY pulse scheme. Narrow and wide rectangular pulses are applied with flip angles of 90° and 180°. Unless specified, pulses are applied along the x-axis. The ^{1}H and ^{13}C frequencies are positioned at the center of the monitored methyl regions. The shaped ^{1}H pulse is of the 180° REBURP (Geen et al., 1991) type (3 ms duration at 800 MHz for a 1600 Hz inversion bandwidth centered in the methyl region).The durations and strengths of the z-gradients in units of (ms; G/cm) are: G1 = (0.2; 26), G2 = (1; 6), G3 = (1; 22), G4 = (0.2; 58), and G5 = (1; 14). Phase cycle is: ϕ1 = 4(x), 4(-x); ϕ2 = x, -x, x, x, -x, x, -x, -x, x, -x, x, x, -x, x, -x, -x; ϕ3 = 8(x), 8(-x); ϕ4 = 2(x), 2(y); ϕ5 = x, -x; and rec = x, -x, -x, x, -x, x, x, -x, -x, x, x, -x, x, -x, -x, x. Quadrature detection in t_1, t_2 and t_3 is achieved by incrementing ϕ_1, ϕ_2, and ϕ_3, respectively, in the usual States-TPPI manner. τ_m and d_1 are optimized for each sample.

Mixing time optimization

In biological macromolecules, the intensities of the nOe correlation signals (I_{SI}) depend on the cross-relaxation rate (σ_{SI}) between the magnetization of each pair of protons, A and B, on the mixing time (τ_m) during which the dipolar interaction is active, and on the T_1 relaxation of the methyl protons. Therefore, the optimization of τ_m is very important to maximize the nOe signal intensity. The optimal τ_m value can be formulated as follows:

$$\tau^{\circ}_m = \frac{1}{(R_{1S}+\sum_I \sigma_{SI})} < T_1 \quad \textbf{(4.5)}, \qquad \text{(Macura and Ernst 1980)}$$

where

R_1 is the longitudinal relaxation rate of proton methyl group

τ°_m: optimal mixing time

According to equation (4.5), two distinct situations may arise:

a) Optimal mixing time in small proteins:

Considering the equation (**4.4)**, the $\sum_I \sigma_{SI}$ term is negligible for small proteins (or sytems with weak proton density). Thus, the $\tau°_m$ in such systems can be approximated to $\frac{1}{R1}$. The mixing time in small molecules can be estimated by measuring T_1 (using an inversion-recovery NMR experiment).

b) Optimal mixing time in large proteins:

This case is compatible with the present study because the protein targets are 82 and 468 kDa and contain a proton density of 14% and 8%, respectively. Therefore, $\sum_I \sigma_{SI}$ cannot be neglected; however, $\sum_I \sigma_{SI}$ is difficult to measure and to estimate. Thus, the $\tau°_m$ parameter must be experimentally optimized by monitoring the variation of nOe signal intensities as a function of different mixing times in a 2D NOESY experiment (nOe build-up curves). For the proteins that are simultaneously labeled at different methyl groups, such as [*U*-^{2}H, AI$^{\delta1}$(LV)^{proS}]-MSG, the determination of $\tau°_m$ encounters two critical points that should be taken into account. Because it is a large protein, the signal overlap in certain regions of the spectrum is inevitable, requiring the choice of a restricted set of "separated peaks". In the given exemple of MSG, the four types of labeled methyl groups present different relaxation properties. The optimal mixing time for each $^{13}CH_3$ is expected to be quite different. Therefore, careful attention should be paid to represent the existing types of methyl groups in the selected set of signals.

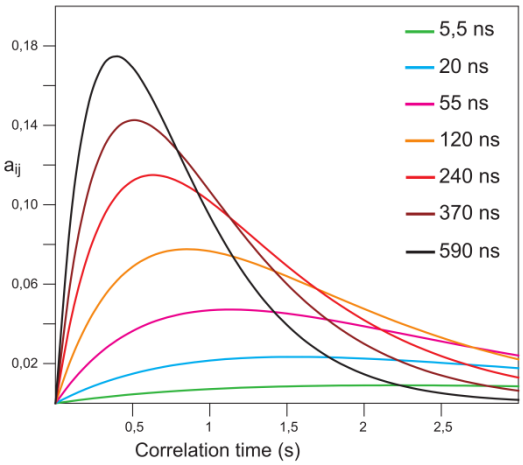


Figure 4.2.2: Simulation of the magnetization transfer evolution between two methyl groups separated by 5.5 Å a_{AB} in a function of the mixing time (τ_m) for different correlation times. (Sounier, PhD thesis, 2008).

In this work, different mixing times were applied for the used labeled samples (respective τ_m values will be indicated for each case in the presented results).

4.2.3 Detection of long-range nOes in MSG labeled simultaneously at $AI^{\delta1}(LV)^{proS}$

NOEs between protons that are separated by up to 10 Å remained detectable in [*U*-^{2}H]-Ile$^{\delta1}$-MSG when analyzed by ^{13}C-edited NOESY (Sounier, PhD thesis, 2008). The aim of my work is to study the possibility of generalizing this approach for the combinatorial labeling of several methyl groups (see §2.2).

MSG is a globular protein that contains 44 Ile, 72 Ala, 69 Leu and 42 Val residues. A perdeuterated sample specifically protonated and isotopically labeled at the methyl groups of Ala-β, Ile-δ_1, Leu-pro-*S* and Val-pro-*S* was prepared in a scrambling-free manner, as described in §2.2. Figure 2.3 presents the obtained 2D HMQC spectrum. After the optimization of several parameters (τ_m=500 ms and d_1=1.3 s), a 3D HMQC-NOESY-HMQC experiment was acquired at 37ºC for 3.5 days on a 800 MHz spectrometer equipped with a cryoprobe. In figure 4.2.3, two 2D extracts from the acquired spectrum are represented, where examples of long-range nOe contacts are clearly illustrated (additional 2D strips are represented in §2.2).

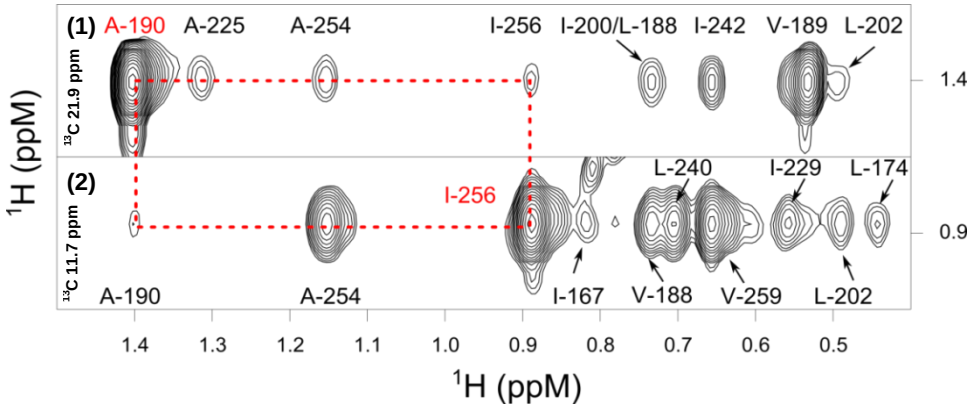


Figure 4.2.3: Detection of long-range nOes in a [U-^{2}H, AI$^{\delta1}$(LV)^{proS}]-MSG sample (82 kDa). 2D strip plot of 3D ^{13}C-HMQC–NOESY spectrum. The extracted distances from the MSG crystallographic structure (1D8C) between each pair of methyl groups are: (1) **190-225:** 7.6 Å; **190-254:** 6.8 Å; **190-256:** 9.8 Å; **190-200:** 10 Å; **190-188:** 8.9 Å; **190-189:** 3.6 Å; **190-202:** 8.8 Å. (2) **256-174:** 8 Å; **256-229:** 9.6 Å; **256-259:** 3.5 Å; **256-240:** 5 Å; **256-188:** 6.3 Å; **256-167:** 8 Å; **256-254:** 3.4. The experiment was acquired at 37ºC on a 800 MHz spectrometer equipped with a cryoprobe for 3.5 days with d_1= 1.3 s and τ_m=500ms. The [U-^{2}H, AI$^{\delta1}$(LV)^{proS}]-MSG sample was at 1mM concentration in 20mM MES, 25mM $MgCl_2$ pH 7.1 $^{2}H_2O$ buffer. Diagonal peaks are highlighted in red.

To statistically estimate the distances from which nOe remain detectable in [*U*-^{2}H, $AI^{\delta1}(LV)^{proS}$]-MSG, I manually assigned all observed nOe signals for each type of residue (except the Leu residues; their analysis is ongoing). Then, I extracted all methyl-methyl distances from the crystallographic structure (PDB 1D8C) of MSG with a cut-off of 15 Å. Finally, for each distance range, I verified whether the expected nOes were experimentally

detected. The obtained results, depicted in figure 4.2.4, indicate that more than 90% of the expected nOes within a distance of less than 8.5 Å were observed.

Analysis of the histograms also reveals that up to 65% (295 correlations out of 450) of the nOes occurring between two methyl groups separated by 8.5 to 10.5 Å are experimentally observed with a minimum signal-to-noise ration of 2.5:1.

In comparison to MSG selectively protonated at the Ile-δ_1 methyl groups, the proton density was increased by a factor of 7 in the [*U*-^{2}H, A$^{\beta}$I$^{\delta1}$(LV)^{proS}]-MSG, where the protons of the AI$^{\delta1}$(LV)^{proS} methyl groups represent 14% of the total number of ^{1}H in MSG. However, augmenting the number of probes by a factor of 5 allowed an increase in the number of detected long-range nOes by approximately 18.5-fold (solely considering the 8.5-10.5 Å range). This improvement is clearly underestimated because data from the Leu residues are not included yet.

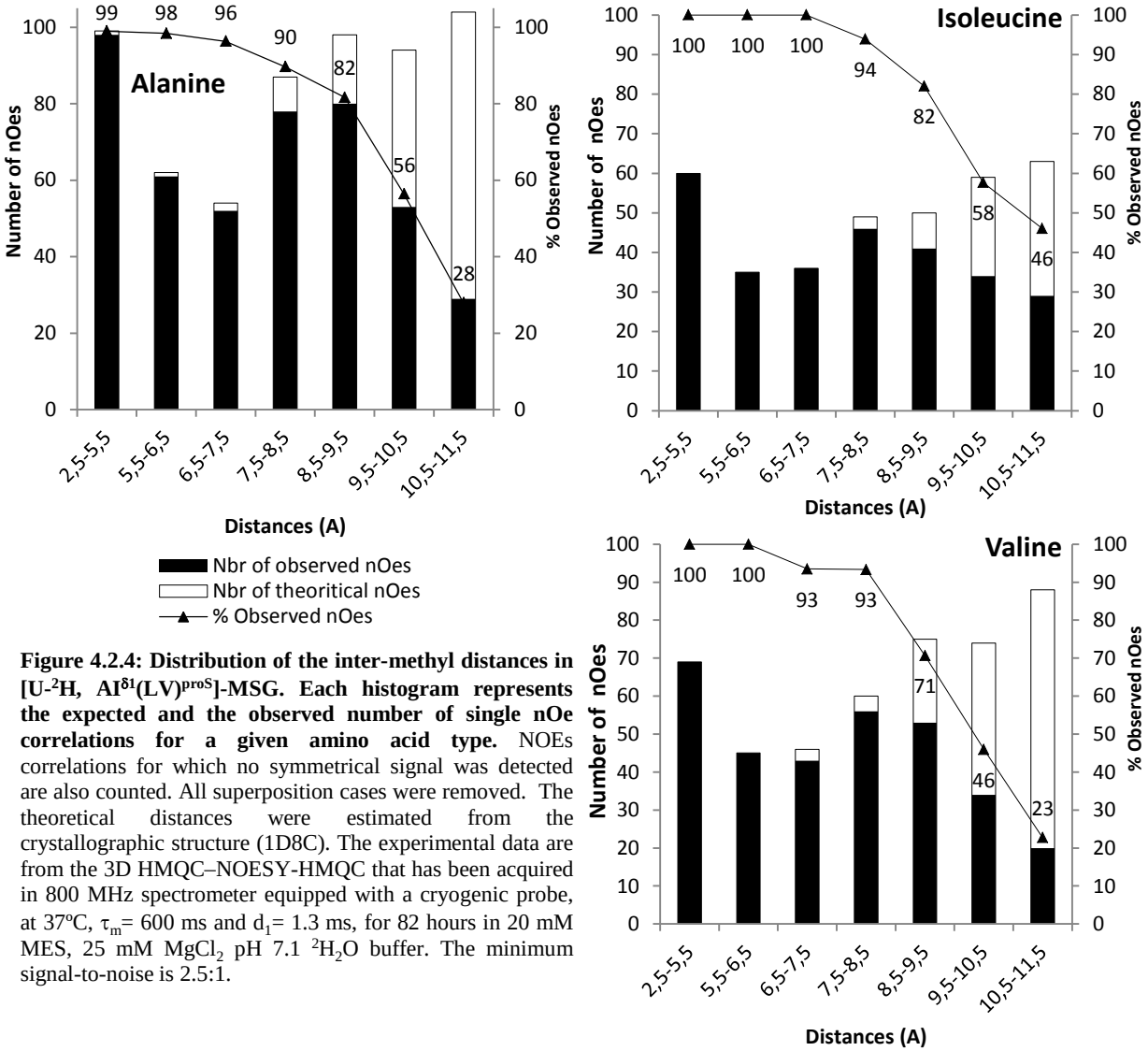


Figure 4.2.4: Distribution of the inter-methyl distances in [U-^{2}H, AI$^{\delta1}$(LV)^{proS}]-MSG. Each histogram represents the expected and the observed number of single nOe correlations for a given amino acid type. NOEs correlations for which no symmetrical signal was detected are also counted. All superposition cases were removed. The theoretical distances were estimated from the crystallographic structure (1D8C). The experimental data are from the 3D HMQC–NOESY-HMQC that has been acquired in 800 MHz spectrometer equipped with a cryogenic probe, at 37ºC, τ_m= 600 ms and d_1= 1.3 ms, for 82 hours in 20 mM MES, 25 mM $MgCl_2$ pH 7.1 ^2H_2O buffer. The minimum signal-to-noise is 2.5:1.

However, from the obtained Ile and Val statistics (amino acids that present the most similar relaxation properties to Leu), the number of detectable nOes involving Leu residues can be theoretically predicted (174 nOes out of 291). Thus, its inclusion to the previously presented result leads to an increase by 30-fold of the observable long-range nOes (within the same distance range). Considering the overall number of detectable nOes, this can be estimated to increase by a 15-fold within 2.5-10.5Å in the discussed combinatorial labeling.

My initial results show that, despite the higher proton density, the total number of long-range distance restraints that can be extracted from the detected long-range nOes is higher in comparison with the reported results from the [*U*-^{2}H, Ile-δ_1]-MSG sample (Sounier, PhD thesis, 2008). The contribution of spin diffusion is expected to be important. However, their use in a semi-quantitative manner is valuable, particularly for large proteins, where the number of nOes is scarce due the prerequisite of high deuteration.

4.2.4 Generalization to a supramolecular protein (PhTET-2)

The efficiency of nOe magnetization transfer is proportional to the mixing time, as discussed in §4.2.2.1. However, as the molecular weight of the protein target increases, the line of the detected signal becomes broader. Thus, the τ_c parameter limits the sensitivity threshold of the NMR signal, consequently affecting the maximal distance that can be observed. In a previous simulation, the nOes arising from protons separated by ~7-8 Å were estimated to remain detectable in particles with a correlation time of 240 ns, such as PhTET-2 at 50ºC (see fig. 4.2.4) (Sounier, PhD thesis, 2008).

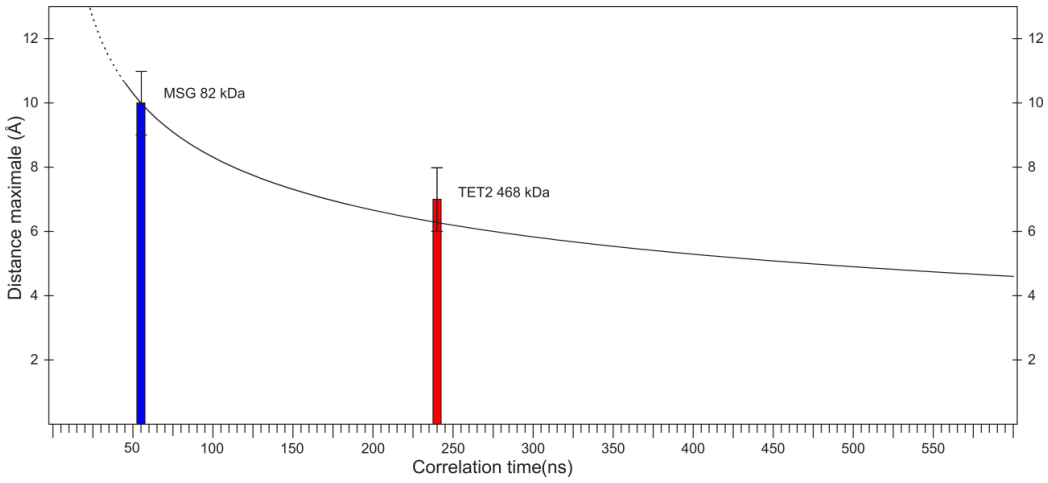


Figure 4.2.5 Maximal distance (r_{max}) to observe nOes between two methyl groups as a function of the correlation time of the protein. This curve was simulated using the equation: $\boldsymbol{r_{max} = r_{max}^{ref}\left(\frac{\tau_c^{ref}}{\tau_c}\right)^{\frac{1}{4}}\left(\frac{(R_{1S}+\sum\sigma_{SI})^{ref}}{(R_{1S}+\sum\sigma_{SI})}\right)^{\frac{1}{6}}}$, where the reference is MSG: T_1 = 1. 3 s (at 37ºC), τ_C= 55 ns and r_{ref} = 10 Å. From Sounier, PhD thesis, 2008.

This simulation was experimentally confirmed using the [*U*-^{2}H, I$^{\delta 1}$TV^{proS}]-PhTET-2 sample were long-range nOes of up to 7.5 Å remained detectable in our conditions. Figure 4.2.5 represents a 2D extract of a 3D ^{13}C-HMQC–NOESY-HMQC spectrum, highlighting examples of the detected long-range nOes.

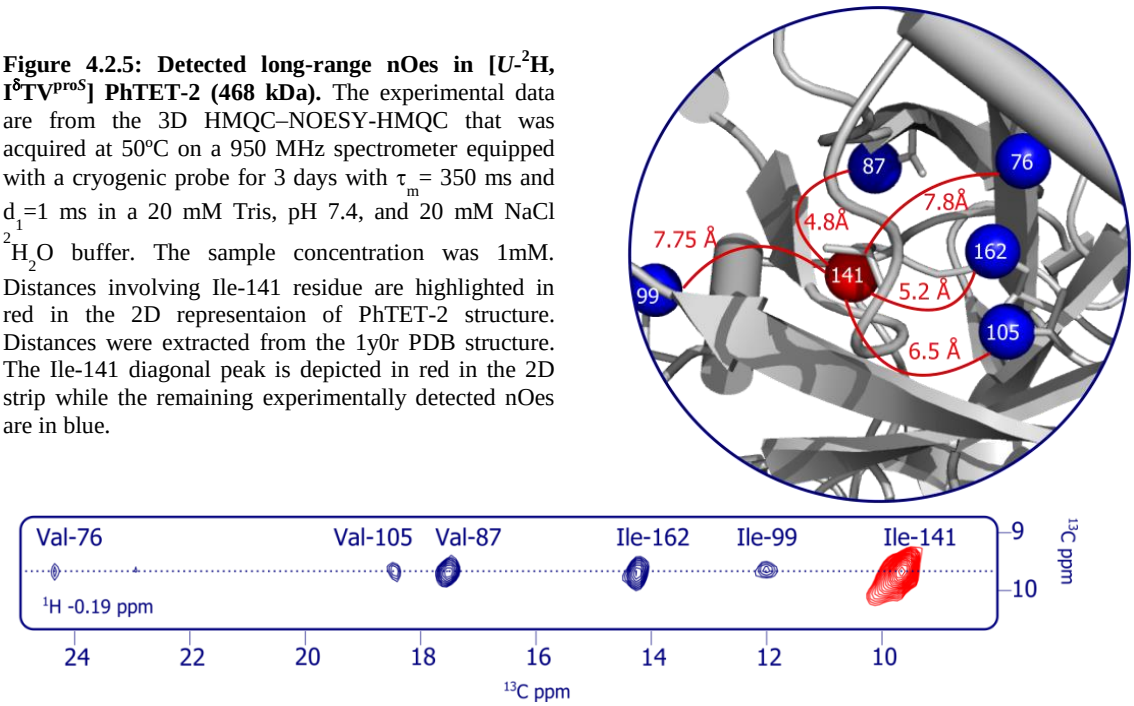


Figure 4.2.5: Detected long-range nOes in [*U*-^{2}H, I$^{\delta}$TV^{proS}] PhTET-2 (468 kDa). The experimental data are from the 3D HMQC–NOESY-HMQC that was acquired at 50ºC on a 950 MHz spectrometer equipped with a cryogenic probe for 3 days with τ_m = 350 ms and d_1=1 ms in a 20 mM Tris, pH 7.4, and 20 mM NaCl $^{2}H_2O$ buffer. The sample concentration was 1mM. Distances involving Ile-141 residue are highlighted in red in the 2D representaion of PhTET-2 structure. Distances were extracted from the 1y0r PDB structure. The Ile-141 diagonal peak is depicted in red in the 2D strip while the remaining experimentally detected nOes are in blue.

Based on this same approach, the cross-validation of the Val residue assignments for PhTET-2 was successfully performed on a [I$^{\delta}$V^{proS}]-PhTET-2 sample (Mas, Crublet et al. 2013). Analysis of the data obtained from this sample demonstrated that ~40% of the expected nOes corresponding to distances ranging from 6-8 Å are detectable.

In summary, I have shown that long-range nOes can be observed in a large oligomeric protein such as PhTET2. Subsequently, I was interested in applying this method to detect and filter long-range inter-monomeric nOes (§4.3).

4.3 Combinatorial methyl groups labeling for the filetering of inter-monomeric nOes in homo-oligomeric proteins

4.3.1 Context

Considerable attention has been devoted to develop methods for the measurement of intermolecular nOes in protein complexes. This type of data is required for the solution

structure determination of biological complexes, such as complexes of protein-protein, protein-ligand or protein-nucleic acid and oligomeric proteins. Initially, the measurement of intermolecular nOes was challenging because of the inability to distinguish the individual components from one another. As a solution, strategies exploiting combinations of ^{15}N, ^{13}C and ^{2}H enrichment were adopted to differentially label the protein complexes. Thus, the resulting intermolecular nOes in these samples could be selected using isotope-edited NMR experiments (Ikura and Bax 1992, Omichinski, Clore et al. 1993, Battiste, Mao et al. 1996, Walters, Ferentz et al. 2001). These approaches are efficient in small proteins but are poorly sensitive in large systems because of efficient transverse relaxation and increased spin diffusion.

Later, a robust and sensitive strategy for the detection of intramolecular nOes, involving methyl and aromatic groups, was proposed (Gross, Gelev et al. 2003). This strategy involves the selective protonation of one partner of the protein complex at the Ile, Leu and Val methyl groups, while the second component remains unlabeled (protonated). This approach was applied to a 35 kDa system but can be easily transposed to larger systems where every component of the protein complex is deuterated. Thus, selective protonation at specific sites in the aromatic residues (§1.2.2.2.2) in combination with methyl groups could give rise to valuable constraints.

In the case of large oligomeric proteins, selective protonation of methyl groups in a perdeuterated context is a prerequisite for their study by NMR. The discrimination between intra- and inter-monomeric methyl-methyl nOes is possible when the protein target is a hetero-oligomer, as chemical shift of each subunit is distinct. However, this advantageous situation ceases to be true when the protein target is a symmetrical homo-oligomer. In these types of systems, each residue shows a single chemical shift due to the symmetrical chemical environment in all subunits of the protein assembly. Consequently, it is difficult to distinguish whether the methyl groups involved in a given nOe correlation belong to the same monomer. Therefore, the use of different methyl groups to asymmetrically label oligomeric particles was proposed. This approach was successfully applied to refine a small pentameric, helical membrane protein structure (Traaseth, Verardi et al. 2008). In this study, the monomers of a perdeuterated protein were either selectively protonated at Ile-δ_1 or the Leu/Val$^{proS/proR}$ methyl groups. Thus, in addition to the unambiguous obtained data, long-range inter-monomeric nOes were also detected (up to 12 Å).

Inspired by the aforementioned strategy, in this subchapter I will demonstrate that the combinatorial labeling of a larger number of methyl groups is a potential tool to obtain unambiguous, long-range inter-monomeric nOes in supramolecular systems (468 kDa). Thus, I will use the dodecameric PhTET-2 sample whose monomeric units are asymmetrically labeled.

4.3.2 Filtering inter-monomeric nOes in the homo-dodecameric PhTET-2

The choice of PhTET-2 for this study is pertinent because is a symmetrical homo-dodecameric assembly with a molecular weight of 468kDa. The back-to-back dimers form the edges of this hollow truncated providing such an architecture that allows the monitoring of two different interface types (see fig. 4.3.1).

4.3.2.1 Strategy

This study relies on the preparation of two PhTET-2 samples that are distinctively labeled which were deoligomerized into individual monomers. The native particle is reconstituted by triggering a random reoligomerization process in solution, where the two dissociated PhTET-2 samples are mixed at a ratio of 1:1. As a result, a dodecameric PhTET-2 particle whose monomers are differentially labeled is obtained (see fig. 4.3.2). Thus, taking advantage of this labeling asymmetry, the inter-monomers nOes can be easily and unambiguously identified.

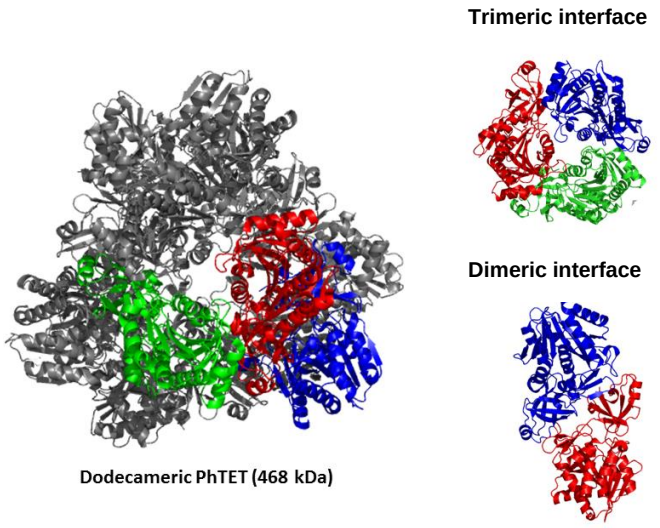


Figure 4.3.1: Scheme representing the PhTET-2 interfaces types. Each color represents an equivalent but different monomer

As previously discussed, long-range methyl-methyl nOes at up to 7-8 Å remain detectable in systems as large as PhTET-2. However, in the present reoligomerized PhTET-2 sample, the isotope enrichment of each methyl group was diluted by a factor of 2. Therefore, the detection of nOes involving these distance ranges is compromised. For a single pair of monomers, two cases occur during the reoligomerization process: in 50% of the dimers, both

units are similarly labeled and in the remaining half, the monomers are distinctively labeled. Thus, while the intra-monomer nOe signals were decreased by a factor of 2, the intensity of the obtained inter-monomer nOes is reduced by a factor of 4. Neglecting the spin diffusion effect, the expected maximal distance from which the inter- monomeric nOes remain

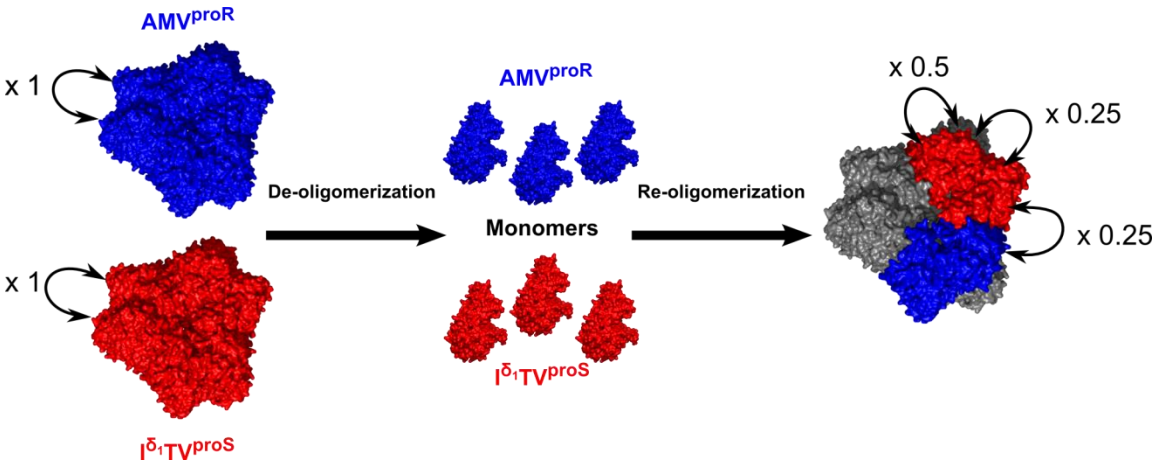


Figure 4.3.2: Scheme representing the sample preparation for the filtering of inter-monomeric nOes. In red, PhTET-2 is [U-^{2}H, I$^{\delta1}$V^{proS} T] labeled and in blue, the protein is [U-^{2}H, MAV^{proR}] labeled. The probability of the occurrence of each type of inter-monomeric nOe (from equivalently or differentially labeled monomers) in the native and reoligomerized particles is indicated.

observable in the system represents in the system represents $\sqrt[6]{\frac{1}{4}}$ (~80%) of the nOe detection threshold. The limit of 7-8 Å was previously simulated and subsequently confirmed experimentally with a [*U*-^{2}H, I$^{\delta1}$TV^{proS}]-PhTET-2 sample containing entirely labeled methyl groups (§4.2.4). Thus, the maximal distance that gives rise to detectable nOe correlations is estimated to be ~5-6 Å. Taking into account this dilution factor, figure 4.2.5 can be updated to represent the cut-off distance for which inter-monomeric nOes remain detectable (see fig. 4.3.3).

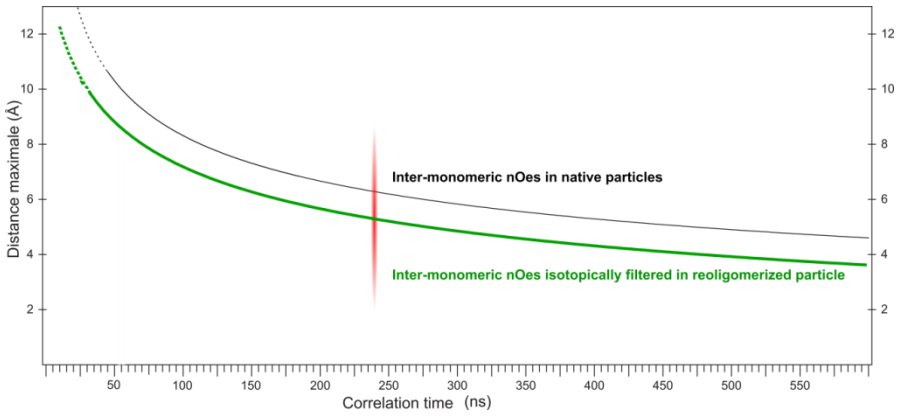


Figure 4.3.3 Maximal distance (r_{max}) to observe nOes between two methyl groups as a function of the correlation time of the protein. The black curve represents the same simulation as 4.2.4, where the methyl groups are labeled and selectively protonated at 100%. The green curve takes into account the 4-fold nOe intensity reduction, which results from 50% PhTET-2 labeling at [U-^{2}H, I$^{\delta1}$V^{proS} T] and [MAV^{proR}]. The correlation time (τ_c) of PhTET-2 at 50ºC is highlighted by a red line.

4.3.2.2 Sample preparation

Applying the strategy discussed above requires the ability to de- and re-oligomerize the protein target. The set-up of a reproductible and reliable protocol enabling the performance of this reversible process is critical. In the present study, samples were prepared according to the described protocol in §4.1.

Choice of labeling scheme:

The choice of the labeling scheme for each of the two native particle samples plays an important role in the quality of the obtained data. To maximize the number of detectable inter-monomeric nOes, both samples should contain an equivalent number of probes, as numerous as possible. As previously discussed in §2.1, careful attention should be paid regarding the combination of methyl groups to avoid undesired isotope scrambling, which may introduce artifacts. For instance, Ile-γ_2 labeling leads to an isotope leak into the Leu and Val pro-*R* methyl group positions. Therefore, their concomitant use is advisable. Thus, the best choice would be to use Ile-δ_1, (Leu,Val)^{proS}, Thr and Met in one component of the protein complexe and Ile-δ_2, (Leu,Val)^{proR} and Ala in another. Statistically, Ala residues are more abundant than the individual Thr and Met residues in proteins. Therefore, Ala is able to balance the total number of both residues.

PhTET-2 contains 30 Ala, 32 Ile, 7 Met, 18 Thr and 35 Val. Based on the available crystallographic structure (PDB1y0r), analysis of the expected nOes revealed that inter-monomeric nOes involving Ile-γ_2 methyl groups are rare. Therefore, I decided to exclude this probe. Moreover, the Leu and Val present a severe overlap (see §2.1), thus, I choose to retain Val residue for its affordable labeling protocol and the availability of its assignment. Considering these modifications, I redistributed my labeling combination using [*U*-^{2}H, I$^{\delta1}$V^{proS} T] and [*U*-^{2}H, MAV^{proR}] schemes (85 and72 methyl probes, respectively).

Precursors for labeling:

Subsequently, I overexpressed [*U*-^{2}H, I$^{\delta1}$V^{proS} T] and [*U*-^{2}H, MAV^{proR}] samples of PhTET-2 according to the protocol described in §2.1. To obtain the specific, combinatorial labeling of the Ile-δ1, Val-pro-*S* and Thr residues, I first added 2- hydroxy-2-[^{13}C]methyl-3-keto-4-[$^2H_3$]

butanoic acid (labeled 2-acetolactate) with [*U*-^{2}H]-Leu to the *E.coli* culture for the selective labeling of Val (see §1.3.2.1). After 40 minutes, [α,β-^{2}H, γ_2-^{13}C^{1}H$_3$]-L-Thr, [^{2}H]-Gly and 2-keto-3[^{2}H$_2$]-4[^{13}C] butyrate were simultaneously added to label Thr and Ile (see §1.3.2.1.1 and §1.3.2.2.2). This shifted addition of precursors is required to avoid the co-incorporation incompatibility between the Val and Ile precursors.

For the second, [MAV^{proR}] labeling scheme, the protocol was easier because all required precursors could be concomitantly added without any complications (see §1.3.2.2.2, §1.3.2.1.3 and §1.3.2.2.1). The 2D HMQC NMR spectra obtained for both samples are presented in figure 4.3.4.

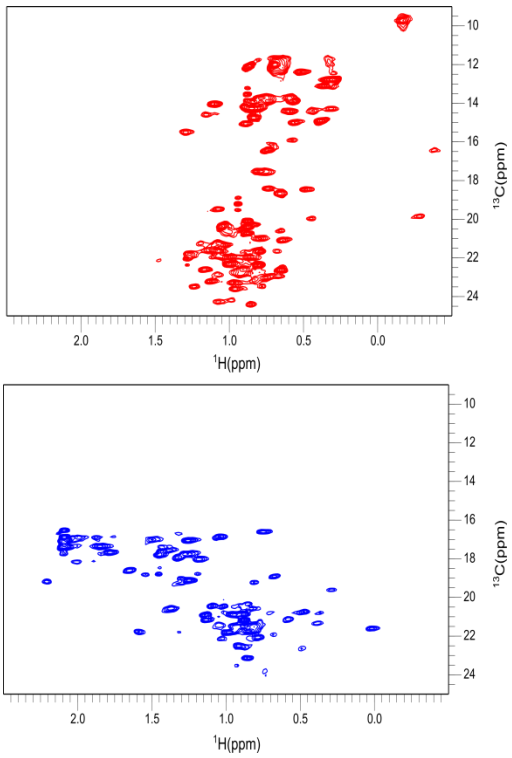


Figure 4.3.4: 2D HMQC spectra of the PhTET-2 samples. In red, PhTET-2 is [U-^{2}H, I$^{\delta1}$V^{proS} T] labeled and in blue, PhTET-2 is [U-^{2}H, MAV^{proR}] labeled. The spectra were acquired in at 50ºC on a 950 MHz spectrometer equipped with a cryoprobe. The samples are at a concentration of 1mM in ^{2}H$_2$O buffer with 20 mM Tris, pH 7.4, and 20 mM NaCl.

Part of these two samples were deoligomerized and subsequently reoligomerized according to the protocol described in §4.1.2 (see fig. 4.3.2) to ultimately obtain a 0.9 mM sample where the required mixed labeling was introduced into the dodecamer; henceforth, this sample will be called Mix-PhTET-2. Figure 4.3.6 shows the 2D HMQC spectrum obtained for Mix-PhTET-2 superimposed on the [*U*-^{2}H, I$^{\delta1}$V^{proS} T]- and [*U*-^{2}H, MAV^{proR}]-PhTET-2 samples.

4.3.2.3 Filtering of long-range inter-monomer nOes

Thanks to the adopted asymmetric labeling, the assignment of part of the inter-monomeric nOes is unambiguously achieved. However, indistinguishable inter-monomeric-nOe correlations, originating from similarly labeled monomers are also present (see fig. 4.3.2). To easily and efficiently filter the targeted long-range nOes from the numerous other correlations, superimposition of the acquired native and reoligomerized spectra ([*U*-^{2}H, AI$^{\delta1}$(LV)^{proS}]-PhTET-2, [*U*-^{2}H, MAV^{proR}]-PhTET-2 and Mix-PhTET-2) is required. The subsequent interpretation is as follows:

- The intra-monomeric nOes are detected in the three spectra. However, differences in signal intensities and maximal detectable distances are expected. Indeed, this effect is due to the labeling dilution from the native to the reoligomerized sample.
- The inter-monomeric nOes originating from equivalently labeled monomers are also observed in all spectra.
- The inter-monomeric nOes arising from asymmetrically labeled monomers are detectable only in the Mix-PhTET-2 spectrum. Hence, the simultaneous analysis of these 3 matrices (superimposed planes) allows straightforward identification of the desired nOes.

After the optimization of different parameters such as τ_m= 350 ms and d_1=1 ms (see §4.1.2.1) for the three samples ([*U*-^{2}H, AI$^{\delta1}$(LV)^{proS}]-PhTET-2, [*U*-^{2}H, MAV^{proR}]-PhTET-2 and Mix-PhTET-2), the 3D HMQC-NOESY-HMQC (HCC variant) was run at 50ºC on a 950 MHz spectrometer equipped with a cryoprobe for 3 days.

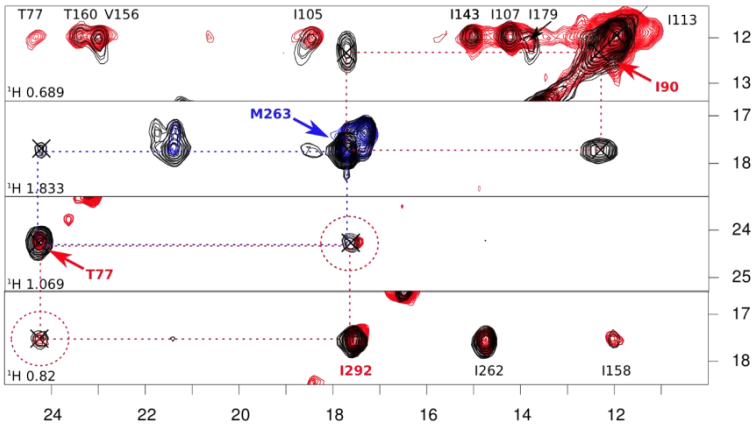


Figure 4.3.5: Examples of 2D strips from the 3D HMQC-NOESY-HMQC experiment acquired on the [U-^{2}H, I$^{\delta1}$TV^{proS}] (red), [U-^{2}H, MAV^{proR}] (blue) and Mix-PhTET-2 (black) samples. The spectra were acquired in a 950 MHz spectrometer equipped with a cryoprobe. The samples range from 0.9 to 1 mM in ^2H_2O buffer with 20 mM Tris (pH 7.4) and 20 mM NaCl. The experiment was run for 3 days at 50ºC with τ_m= 350 ms and d_1=1 ms. The cross-correlation peaks surrounded by dashed red circles highlight inter-nOe signals arising from the equivalently labeled monomers.

For analysis of the acquired data, I utilized the previously assigned Ile, Ala and Val residues in PhTET-2 (Amero, Asuncion Dura et al. 2011, Mas, Crublet et al. 2013). Thanks to the combinatorial labeling, which enabled the availability of numerous nOe contacts, additional methyl groups could be assigned, including M263, T77 and T245.

In Mix-PhTET-2, 172 intra-monomeric nOe pairs are expected based on a cut-off of 6 Å, whereas a total of 19 inter-monomeric nOes pairs are predicted from the X-ray crystallographic structure (PDB 1y0r). From the 12 nOe pairs involved in differentially labeled monomers, 40% (5 of 12) were observed with a minimum signal-to-noise ratio of

(5:1). However, two unexpected results were observed and should be discussed. According to the crystallographic structure (PDB 1y0r), 4 of these detected nOes arise from distances ranging from 5.2 to 5.6 Å, whereas 4 of the expected nOes from distances between 3.5 and 4.5 Å were not observed. Moreover, most of the inter-monomer nOes (5 out of 6) ranging from 2.9 to 4.8 Å that result from equivalently labeled monomers were not detected as well. This absence may be explained by the low quality of the prepared sample, as the reoligomerized NMR spectrum was not completely faithful to the native spectrum (see fig. 4.3.6). Because the methyl groups in the reoligomerized particle did not recover the expected native position, atomic distances are certainly affected and consequently the observed nOes. However, for those that were identical to the native dodecameric particle (see fig. 4.3.6), it was straightforward to determine inter-monomeric nOes from methyl groups separated by up to 5 Å, as previously predicted.

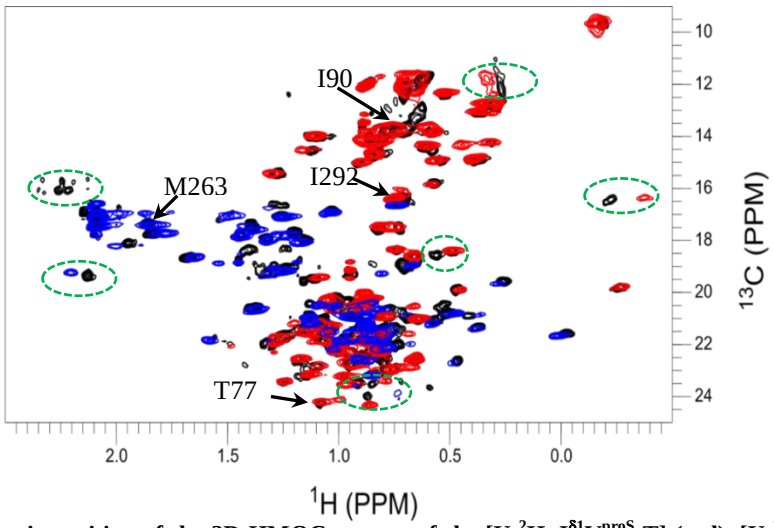


Figure 4.3.6: Superimposition of the 2D HMQC spectra of the [U-^{2}H, I$^{\delta1}$V^{proS} T] (red), [U-^{2}H, MAV^{proR}] (blue) and Mix-PhTET-2 (in black) samples. The spectra were acquired in a 950 MHz spectrometer equipped with a cryoprobe. The samples are ranged from 0.9 to 1 mM in $^{2}H_2O$ buffer containing 20 mM Tris (pH 7.4) and 20 mM NaCl. Examples of the chemical shift differences between the native and reoligomerized samples are highlighted by green, dashed circles. Examples of the residues for which inter-monomer nOes were observed are represented by an indication of their number in the primary protein sequence.

The presented data could be compatible with the context where native PhTET-2 particles were not efficiently monomerized due to some "unexpected" experimental reason. As the re-assembly of the generated oligomeric state was not optimal the predicted inter-monomer contacts could be detectable in a given interface but not in another. For instance, if the applied chemical treatment led to the PhTET-2 dimerization instead of its monomerization (which is more easily achievable), one can speculate that the methyl groups that originated the experimentally detected inter-monomeric nOes are primarily localized at the

monomer/monomer interfaces (see fig. 4.3.1). The others, predicted at trimerization contacts could not be observed since the overall dodecamer was not correctly assembled (see fig. 4.3.1). Despite being plausible, this theory is not verified. Indeed, the plotting of the respective methyl groups of the expected nOes (detected or not) in the structure indicates that these probes are spreaded in all interfaces and localized in different secondary structures (i.e., loops, β-strand and flexible α helix).

Although a sample with a better oligomerization state is required for more accurate analysis, the preliminary results are already promising. The results clearly show that 40% of the expected long-range inter-monomer nOes within a cutt-off of 6 Å were detected in a supramolecular 0.5 MDa protein. In addition to obtaining both long-range and unambiguous data, the filtering of these nOes from the numerous intra-monomeric correlations was particularly easy. This approach was applied to the PhTET-2 system, as a proof of concept. Its use for other challenging systems, such as membrane proteins, would be of great interest.

CHAPTER V

CONCLUSION AND PERSPECTIVES

Solution NMR spectroscopy has been limited to small biological objects for a long time. Nowadays, it is unequivocally recognized that the strategy of specific isotope labeling of methyl groups in a perdeuterated protein has significantly extended the frontier of this technique. Expression of [^{2}H, $^{13}CH_3$]-proteins is generally performed by adding selectively protonated and isotopically labeled metabolic precursors to the *E. coli* deuterated culture.

In the last years, protocols for the synthesis and incorporation of new isotope-labeled precursors have been made available to extend the library of methyl labeling methods. Indeed, the use of these compounds has proven efficient in improving significantly the sensitivity and resolution of large proteins NMR spectra (up to 1 MDa). Conversely, this strategy also leads to a significant loss of structural information due to the low number of protonated probes, which represents its main drawback.

The project of my thesis falls within the framework of developing new methodologies to cope with this scarce structural information, relying on the simultaneous labeling of several methyl groups to increase the number of probes.

To obtain optimized combinatorial labeling, the choice of the labeled precursors as well as the protocol for their incorporation have to be carefully studied. For the $A^{\beta}I^{\delta 1}(LV)^{proS}$ methyl groups combination, I demonstrated that due to the "cross-talk" in the metabolic pathways of these residues, the co-incorporation of α-ketobutyrate and 2-acetolactate is not compatible leading to a low isotope enrichment of Leu and Val. Moreover, the simultaneous addition of labeled α-ketobutyrate and alanine to *E. coli* culture, generates an Ile faction (2-5%) labeled at both δ_1 and γ_2 methyl groups positions. To cope with these undesired side-effects, I first set up a new enzymatic procedure to synthesize an alternate Ile precursor, 2-hydroxy-2ethyl-3-keto butanoic acid, preventing Ile-γ_2 scrambling. Furthermore, I have proposed a new strategy where the supplementation for Ile is delayed with respect to the Leu/Val precursor (2-acetolactate) to promote their incorporation.

In comparison to the "standard $A^{\beta}I^{\delta 1}LV$" labeling scheme, the proposed $A^{\beta}I^{\delta 1}(LV)^{proS}$ pattern induces a 2-fold decrease of number of Leu and Val NMR signals and enhances the intensity of detectable long-range nOes by a factor 4. The described protocol also permits the

suppression of spurious correlations, especially harmful for structural studies based on detection/analysis of nOes. Protein preparation following such protocol gives rise to NMR spectra, with higher sensitivity and resolution.

To make an efficient use of these optimized NMR spectra, assignment of the methyl groups signals is mandatory. The TOCSY-based assignment of methyl groups remains feasible for systems with less than 100 kDa molecular weight. However, for branched aliphatic side-chains such as Leu, Val or Ile, a sharp decrease of its efficiency was reported. Furthermore, the stereospecific assignment is particularly pivotal to obtain accurate 3D high-resolution of large protein structures where nOes information is by default scarce due the high deuteration prerequisite. Thus, I took advantage of the setup enzymatic protocol to produce a linearized Ile-δ_1 precursor that was so far unavailable in contrast with the case of Ile-γ_2, Leu and Val residues. To be able to regio- and stereo-specifically assign these residues in a single step, I used a mixture of precursors with different isotope topologies to discriminate each methyl group. Having all Ile, Leu and Val isotopically linearized, an optimized "out-and-back" ^{13}C-TOCSY could be designed. Applying this approach on MSG (82 kDa), I demonstrated that the Cα correlations have been detected in a proportion of 60%, 86%, 85% and 91% for the Ile-δ_1, Ile-γ_2, Leu-γ_1 and Val-γ_1, respectively with a minimal signal-to-noise ratio of 5:1 while 70% and 89% of Leu-γ_2 and Val-γ_2 (pro-*S*), respectively, were clearly linked to their respective pro-R prochiral groups providing the stereospecific assignment.

The above described approach ceases to be applicable to supra-molecular proteins. Thus, to assign 64 (Ile and Ala) methyl groups in a 468 kDa system, the SeSAM approach has been described to be practical and efficient. However, due to the protein production and purification processes, the published method was laborious and time-consuming. To simplify these two steps, I developed an enriched culture medium for the specific labeling of Ala. Indeed, by supplying the minimal culture medium with deuterated algae hydrolysate, where the [^{2}H, ^{13}CH$_3$]-Ala is over-supplemented, an increase in protein yield by a factor of 1.75 was obtained with an optimized labeled Ala enrichment. Such an improvement permitted the decrease in the minimal required culture volume, enabling its performance in 24 well plates. The small volume of culture (5 ml/well) ain turn the parallel purification in 96 well plates, offering a more comfortable and a faster manipulation. This improved SeSAM version was

estimated to allow the assignment of *ca.* 100 methyl cross-peaks in 2 weeks, including 4 days of NMR time and less than 2 k€ of isotopic materials.

The use of such culture media for the specific labeling of methyl group is transposable to the large scale *E. coli* culture, fact that copes with the $^{2}H_2O$ toxicity and in some cases with the protein solubility issues. Nevertheless, due to the required over-supplementation of the labeled amino acid (or precursor), the cost per sample would be significantly elevated. Furthermore, the addition of high quantities of these molecules can also lead to the inhibition of certain enzyme, which may interfere with their incorporation. Therefore, great efforts should be made to make available such enriched medium depleted with the intended residue to label. It is anticipated that, one or several amino-acids could be removed selectively by means of chromatography from a hydrolyzed mixture. This would extend even more the applicability of solution NMR spectroscopy in conjunction with selective protonation as the $^{2}H_2O$ toxicity will be significantly reduced.

After optimization of the obtainable NMR spectra, where several methyl groups are simultaneously labeled, and developing alternatives for their assignment, I become interested in exploring novel strategies for their exploitation. As previously mentioned, the combinatorial labeling of methyl groups leads to an increased number of NMR probes in large proteins. In the developed work, I estimated that the label of 4 types of methyl groups, $AI^{\delta1}(LV)^{proS}$-MSG, increases the number of detectable nOes within 2.5-10.5 Å distance range by a factor of 15 compared with the [^{2}H, Ile-δ1]-MSG. I also showed that 60% of the expected long-range nOes arising from methyl separated by up to 8.5- 10.5 remain detectable, making this approach a tool of choice for the 3D structure calculation of equivalently sized proteins.

The same study was performed on $I^{\delta1}V^{proS}$- PhTET-2 (468 kDa) indicating that 40% of nOes from methyl groups of 6-8Å apart are still detectable. Using such advantage I devised the following approach where long-range inter-monomeric nOes are easily and unambiguously filtered in this symmetrical and homo-oligomeric system. Generally, such key information is valuable for the correct packing of subunits against each others.

Finally, the pertinence of the selective protonation of methyl group strategy was also illustrated by trucking in a real-time by SOFAST NMR the evolution of diverse intermediates

towards the final dodecameric re-oligomerization of PhTET-2. Combining these technologies with other structural techniques, EM and MS, a kinetic model could be proposed for such process.

The selective protonation and labeling of methyl groups is a unique technology to tackle large proteins. Despite its usefulness, the applicability of this approach remains somehow limited by its associated costs, which are still not affordable for many laboratories. The decreased protein yield due to the $^{2}H_2O$ toxicity makes it even more expensive. Therefore, the development of adapted rich medium seems to be the most urgent necessity to push upward the limits of selective methyl groups labeling.

CHAPTER VI

BIBLIOGRAPHY

Ali, M. H. and B. Imperiali (2005). "Protein oligomerization: how and why." Bioorg Med Chem **13**(17): 5013-5020.

Amero, C., M. Asuncion Dura, M. Noirclerc-Savoye, A. Perollier, B. Gallet, M. J. Plevin, T. Vernet, B. Franzetti and J. Boisbouvier (2011). "A systematic mutagenesis-driven strategy for site-resolved NMR studies of supramolecular assemblies." J Biomol NMR **50**(3): 229-236.

Amero, C., P. Schanda, M. A. Dura, I. Ayala, D. Marion, B. Franzetti, B. Brutscher and J. Boisbouvier (2009). "Fast two-dimensional NMR spectroscopy of high molecular weight protein assemblies." J Am Chem Soc **131**(10): 3448-3449.

Anfinsen, C. B. and E. Haber (1961). "Studies on the reduction and re-formation of protein disulfide bonds." J Biol Chem **236**: 1361-1363.

Archer, S. J., A. Bax, A. B. Roberts, M. B. Sporn, Y. Ogawa, K. A. Piez, J. A. Weatherbee, M. L. Tsang, R. Lucas, B. L. Zheng and et al. (1993). "Transforming growth factor beta 1: NMR signal assignments of the recombinant protein expressed and isotopically enriched using Chinese hamster ovary cells." Biochemistry **32**(4): 1152-1163.

Ayala, I., O. Hamelin, C. Amero, O. Pessey, M. J. Plevin, P. Gans and J. Boisbouvier (2012). "An optimized isotopic labelling strategy of isoleucine-gamma2 methyl groups for solution NMR studies of high molecular weight proteins." Chem Commun (Camb) **48**(10): 1434-1436.

Ayala, I., R. Sounier, N. Use, P. Gans and J. Boisbouvier (2009). "An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein." J Biomol NMR **43**(2): 111-119.

Balayssac, S., I. Bertini, A. Bhaumik, M. Lelli and C. Luchinat (2008). "Paramagnetic shifts in solid-state NMR of proteins to elicit structural information." Proc Natl Acad Sci U S A **105**(45): 17284-17289.

Bar-Ilan, A., V. Balan, K. Tittmann, R. Golbik, M. Vyazmensky, G. Hubner, Z. Barak and D. M. Chipman (2001). "Binding and activation of thiamin diphosphate in acetohydroxyacid synthase." Biochemistry **40**(39): 11946-11954.

Barak, Z., D. M. Chipman and N. Gollop (1987). "Physiological implications of the specificity of acetohydroxy acid synthase isozymes of enteric bacteria." J Bacteriol **169**(8): 3750-3756.

Battiste, J. L., H. Mao, N. S. Rao, R. Tan, D. R. Muhandiram, L. E. Kay, A. D. Frankel and J. R. Williamson (1996). "Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex." Science **273**(5281): 1547-1551.

Bax, A. (2011). "Triple resonance three-dimensional protein NMR: before it became a black box." J Magn Reson **213**(2): 442-445.

Belenky, I., A. Steinmetz, M. Vyazmensky, Z. Barak, K. Tittmann and D. M. Chipman (2012). "Many of the functional differences between acetohydroxyacid synthase (AHAS) isozyme I and other AHASs are a result of the rapid formation and breakdown of the covalent acetolactate-thiamin diphosphate adduct in AHAS I." FEBS J **279**(11): 1967-1979.

Bird, M. I. and P. B. Nunn (1983). "Metabolic homoeostasis of L-threonine in the normally-fed rat. Importance of liver threonine dehydrogenase activity." Biochem J **214**(3): 687-694.

Bokoch, M. P., Y. Zou, S. G. Rasmussen, C. W. Liu, R. Nygaard, D. M. Rosenbaum, J. J. Fung, H. J. Choi, F. S. Thian, T. S. Kobilka, J. D. Puglisi, W. I. Weis, L. Pardo, R. S. Prosser, L. Mueller and B. K. Kobilka (2010). "Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor." Nature **463**(7277): 108-112.

Brodin, P., T. Drakenberg, E. Thulin, S. Forsen and T. Grundstrom (1989). "Selective proton labelling of amino acids in deuterated bovine calbindin D9K. A way to simplify 1H-NMR spectra." Protein Eng **2**(5): 353-357.

Cappuccio, J. A., A. K. Hinz, E. A. Kuhn, J. E. Fletcher, E. S. Arroyo, P. T. Henderson, C. D. Blanchette, V. L. Walsworth, M. H. Corzett, R. J. Law, J. B. Pesavento, B. W. Segelke, T. A. Sulchek, B. A. Chromy, F. Katzen, T. Peterson, G. Bench, W. Kudlicki, P. D. Hoeprich, Jr. and M. A. Coleman (2009). "Cell-free expression for nanolipoprotein particles: building a high-throughput membrane protein solubility platform." Methods Mol Biol **498**: 273-296.

Carlson, E. D., R. Gan, C. E. Hodgman and M. C. Jewett (2012). "Cell-free protein synthesis: applications come of age." Biotechnol Adv **30**(5): 1185-1194.

Carvalho, L. P. S., P. A. Frantom, A. Argyrous and J. S. Blanchard (2009). "Kinetic Evidence for Inter-Domain Communication in the Allosteric Regulation of á-Isopropylmalate Synthase from M.tuberculosis." Biochemistry **48**(9): 1996–2004.

Casey, A. K., J. Baugh and P. A. Frantom (2012). "The slow-onset nature of allosteric inhibition in alpha-isopropylmalate synthase from Mycobacterium tuberculosis is mediated by a flexible loop." Biochemistry **51**(24): 4773-4775.

Chakrabarti, P. and J. Janin (2002). "Dissecting protein-protein recognition sites." Proteins **47**(3): 334-343.

Chalmeau, J., N. Monina, J. Shin, C. Vieu and V. Noireaux (2011). "alpha-Hemolysin pore formation into a supported phospholipid bilayer using cell-free expression." Biochim Biophys Acta **1808**(1): 271-278.

Chan, P. H., S. Weissbach, M. Okon, S. G. Withers and L. P. McIntosh (2012). "Nuclear Magnetic Resonance Spectral Assignments of alpha-1,4-Galactosyltransferase LgtC from Neisseria meningitidis: Substrate Binding and Multiple Conformational States." Biochemistry **51**(41): 8278-8292.

Chang, X., D. Keller, S. I. O'Donoghue and J. J. Led (2002). "NMR studies of the aggregation of glucagon-like peptide-1: formation of a symmetric helical dimer." FEBS Lett **515**(1-3): 165-170.

Chao, F. A., L. Shi, L. R. Masterson and G. Veglia (2012). "FLAMEnGO: a fuzzy logic approach for methyl group assignment using NOESY and paramagnetic relaxation enhancement data." J Magn Reson **214**(1): 103-110.

Chipman, D., Z. Barak and J. V. Schloss (1998). "Biosynthesis of 2-aceto-2-hydroxy acids: acetolactate synthases and acetohydroxyacid synthases." Biochim Biophys Acta **1385**(2): 401-419.

Crublet, E., R. Kerfah, G. Mas, M. Noirclerc-Savoye, V. Lantez, T. Vernet and J. Boisbouvier (2014). "A cost-effective protocol for the parallel production of libraries of 13CH3-specifically labeled mutants for NMR studies of high molecular weight proteins." Methods Mol Biol **1091**: 229-244.

Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Bax (1995). "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." J Biomol NMR **6**(3): 277-293.

Duggleby, R. G. (2006). "Domain relationships in thiamine diphosphate-dependent enzymes." Acc Chem Res **39**(8): 550-557.

Duggleby, R. G., J. A. McCourt and L. W. Guddat (2008). "Structure and mechanism of inhibition of plant acetohydroxyacid synthase." Plant Physiol Biochem **46**(3): 309-324.

Dumas, R., V. Biou, F. Halgand, R. Douce and R. G. Duggleby (2001). "Enzymology, structure, and dynamics of acetohydroxy acid isomeroreductase." Acc Chem Res **34**(5): 399-408.

Engel, S., M. Vyazmensky, D. Berkovich, Z. Barak and D. M. Chipman (2004). "Substrate range of acetohydroxy acid synthase I from Escherichia coli in the stereoselective synthesis of alpha-hydroxy ketones." Biotechnol Bioeng **88**(7): 825-831.

Epperly, B. R. and E. E. Dekker (1991). "L-threonine dehydrogenase from Escherichia coli. Identification of an active site cysteine residue and metal ion studies." J Biol Chem **266**(10): 6086-6092.

Esposito, D., A. Sankar, N. Morgner, C. V. Robinson, K. Rittinger and P. C. Driscoll (2010). "Solution NMR investigation of the CD95/FADD homotypic death domain complex suggests lack of engagement of the CD95 C terminus." Structure **18**(10): 1378-1390.

Etezady-Esfarjani, T., S. Hiller, C. Villalba and K. Wuthrich (2007). "Cell-free protein synthesis of perdeuterated proteins for NMR studies." J Biomol NMR **39**(3): 229-238.

Falkinham, J. O., 3rd (1979). "Identification of a mutation affecting an alanine-alpha-ketoisovalerate transaminase activity in Escherichia coli K-12." Mol Gen Genet **176**(1): 147-149.

Fiaux, J., E. B. Bertelsen, A. L. Horwich and K. Wuthrich (2004). "Uniform and residue-specific 15N-labeling of proteins on a highly deuterated background." J Biomol NMR **29**(3): 289-297.

Fischer, M., K. Kloiber, J. Hausler, K. Ledolter, R. Konrat and W. Schmid (2007). "Synthesis of a 13C-methyl-group-labeled methionine precursor as a useful tool for simplifying protein structural analysis by NMR spectroscopy." Chembiochem **8**(6): 610-612.

Frantom, P. A. (2012). "Structural and functional characterization of alpha-isopropylmalate synthase and citramalate synthase, members of the LeuA dimer superfamily." Arch Biochem Biophys **519**(2): 202-209.

Gallagher, D. T., G. L. Gilliland, G. Xiao, J. Zondlo, K. E. Fisher, D. Chinchilla and E. Eisenstein (1998). "Structure and control of pyridoxal phosphate dependent allosteric threonine deaminase." Structure **6**(4): 465-475.

Gan, L., J. A. Speir, J. F. Conway, G. Lander, N. Cheng, B. A. Firek, R. W. Hendrix, R. L. Duda, L. Liljas and J. E. Johnson (2006). "Capsid conformational sampling in HK97 maturation visualized by X-ray crystallography and cryo-EM." Structure **14**(11): 1655-1665.

Gans, P., J. Boisbouvier, I. Ayala and O. Hamelin (2010b). Brevet Procédé de marquage isotopique spécifique de groupements méthyle de val, leu et ile. Commissariat A L'energie Atomique Et Aux Energies Alternatives, Centre National De La Recherche Scientifique and U. J. Fourier. France. **Patent WO2011083356 A1**.

Gans, P., O. Hamelin, R. Sounier, I. Ayala, M. A. Dura, C. D. Amero, M. Noirclerc-Savoye, B. Franzetti, M. J. Plevin and J. Boisbouvier (2010). "Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins." Angew Chem Int Ed Engl **49**(11): 1958-1962.

Gardner, K. H. and L. E. Kay (1997). "Production and Incorporation of 15N, 13C, 2H (1H-d1 Methyl) Isoleucine into Proteins for Multidimensional NMR Studies." Journal American Chemistry Society **119**: 7599-7600.

Gardner, K. H. and L. E. Kay (1997). "Production and Incorporation of 15N, 13C, 2H (1H-δ1 Methyl) Isoleucine into Proteins for Multidimensional NMR Studies." Journal of the American Chemical Society **119**(32): 7599–7600.

Gardner, K. H. and L. E. Kay (1998). "The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins." Annu Rev Biophys Biomol Struct **27**: 357-406.

Garrett, D. S., Y. J. Seok, A. Peterkofsky, G. M. Clore and A. M. Gronenborn (1997). "Identification by NMR of the binding surface for the histidine-containing phosphocarrier protein HPr on the N-terminal domain of enzyme I of the Escherichia coli phosphotransferase system." Biochemistry **36**(15): 4393-4398.

Gelis, I., A. M. Bonvin, D. Keramisanou, M. Koukaki, G. Gouridis, S. Karamanou, A. Economou and C. G. Kalodimos (2007). "Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR." Cell **131**(4): 756-769.

Gifford, J. L., H. Ishida and H. J. Vogel (2011). "Fast methionine-based solution structure determination of calcium-calmodulin complexes." J Biomol NMR **50**(1): 71-81.

Godoy-Ruiz, R., C. Guo and V. Tugarinov (2010). "Alanine methyl groups as NMR probes of molecular structure and dynamics in high-molecular-weight proteins." J Am Chem Soc **132**(51): 18340-18350.

Goodsell, D. S. and A. J. Olson (2000). "Structural symmetry and protein function." Annu Rev Biophys Biomol Struct **29**: 105-153.

Goren, M. A., A. Nozawa, S. Makino, R. L. Wrobel and B. G. Fox (2009). "Cell-free translation of integral membrane proteins into unilamelar liposomes." Methods Enzymol **463**: 647-673.

Gossert, A. D., A. Hinniger, S. Gutmann, W. Jahnke, A. Strauss and C. Fernandez (2011). "A simple protocol for amino acid type selective isotope labeling in insect cells with improved yields and high reproducibility." J Biomol NMR **51**(4): 449-456.

Goto, N. K., K. H. Gardner, G. A. Mueller, R. C. Willis and L. E. Kay (1999). "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 **13**(4): 369-374.

Griffey, R. H., A. G. Redfield, R. E. Loomis and F. W. Dahlquist (1985). "Nuclear magnetic resonance observation and dynamics of specific amide protons in T4 lysozyme." Biochemistry **24**(4): 817-822.

Gross, J. D., V. M. Gelev and G. Wagner (2003). "A sensitive and robust method for obtaining intermolecular NOEs between side chains in large protein complexes." J Biomol NMR **25**(3): 235-242.

Hajduk, P. J., D. J. Augeri, J. Mack, R. Mendoza, J. Yang, S. F. Betz and S. W. Fesik (2000). "NMR-Based Screening of Proteins Containing 13C-Labeled Methyl
Groups." J. Am. Chem. Soc. **122**(22): 7898-7904.

Hansen, A. P., A. M. Petros, A. P. Mazar, T. M. Pederson, A. Rueter and S. W. Fesik (1992). "A practical method for uniform isotopic labeling of recombinant proteins in mammalian cells." Biochemistry **31**(51): 12713-12718.

Hansen, D. F. and L. E. Kay (2011). "Determining valine side-chain rotamer conformations in proteins from methyl 13C chemical shifts: application to the 360 kDa half-proteasome." J Am Chem Soc **133**(21): 8272-8281.

Hattori, Y., K. Furuita, I. Ohki, T. Ikegami, H. Fukada, M. Shirakawa, T. Fujiwara and C. Kojima (2013). "Utilization of lysine (1)(3)C-methylation NMR for protein-protein interaction studies." J Biomol NMR **55**(1): 19-31.

Hayashi, H., K. Inoue, T. Nagata, S. Kuramitsu and H. Kagamiyama (1993). "Escherichia coli aromatic amino acid aminotransferase: characterization and comparison with aspartate aminotransferase." Biochemistry **32**(45): 12229-12239.

Henikoff, S. and J. G. Henikoff (1992). "Amino acid substitution matrices from protein blocks." Proc Natl Acad Sci U S A **89**(22): 10915-10919.

Heymann, J. B., N. Cheng, W. W. Newcomb, B. L. Trus, J. C. Brown and A. C. Steven (2003). "Dynamics of herpes simplex virus capsid maturation visualized by time-lapse cryo-electron microscopy." Nat Struct Biol **10**(5): 334-341.

Higashi, N., H. Fukada and K. Ishikawa (2005). "Kinetic study of thermostable L-threonine dehydrogenase from an archaeon Pyrococcus horikoshii." J Biosci Bioeng **99**(2): 175-180.

Hilty, C., G. Wider, C. Fernandez and K. Wuthrich (2003). "Stereospecific assignments of the isopropyl methyl groups of the membrane protein OmpX in DHPC micelles." J Biomol NMR **27**(4): 377-382.

Homer, R. J., M. S. Kim and D. M. LeMaster (1993). "The use of cystathionine gamma-synthase in the production of alpha and chiral beta deuterated amino acids." Anal Biochem **215**(2): 211-215.

Ichikawa, O., M. Osawa, N. Nishida, N. Goshima, N. Nomura and I. Shimada (2007). "Structural basis of the collagen-binding mode of discoidin domain receptor 2." EMBO J **26**(18): 4168-4176.

Ikeya, T., M. Takeda, H. Yoshida, T. Terauchi, J. G. Jee, M. Kainosho and P. Guntert (2009). "Automated NMR structure determination of stereo-array isotope labeled ubiquitin from minimal sets of spectra using the SAIL-FLYA system." J Biomol NMR **44**(4): 261-272.

Ikeya, T., T. Terauchi, P. Guntert and M. Kainosho (2006). "Evaluation of stereo-array isotope labeling (SAIL) patterns for automated structural analysis of proteins with CYANA." Magn Reson Chem **44 Spec No**: S152-157.

Ikura, M. and A. Bax (1992). "Isotope-Filtered 2D NMR of a Protein-Peptide Complex:
Study of a Skeletal Muscle Myosin Light Chain Kinase
Fragment Bound to Calmodulin." Journal American Chemistry Society **114**: 2433-2440.

Ikura, M., D. Marion, L. E. Kay, H. Shih, M. Krinks, C. B. Klee and A. Bax (1990). "Heteronuclear 3D NMR and isotopic labeling of calmodulin. Towards the complete assignment of the 1H NMR spectrum." Biochem Pharmacol **40**(1): 153-160.

Isaacson, R. L., P. J. Simpson, M. Liu, E. Cota, X. Zhang, P. Freemont and S. Matthews (2007). "A new labeling method for methyl transverse relaxation-optimized spectroscopy NMR spectra of alanine residues." J Am Chem Soc **129**(50): 15428-15429.

Ishikawa, K., N. Higashi, T. Nakamura, T. Matsuura and A. Nakagawa (2007). "The first crystal structure of L-threonine dehydrogenase." J Mol Biol **366**(3): 857-867.

Jaremko, M., L. Jaremko, M. Nowakowski, M. Wojciechowski, R. H. Szczepanowski, R. Panecka, I. Zhukov, M. Bochtler and A. Ejchart (2014). "NMR structural studies of the first catalytic half-domain of ubiquitin activating enzyme." J Struct Biol **185**(1): 69-78.

Jentoft, N. and D. G. Dearborn (1983). "Protein labeling by reductive alkylation." Methods Enzymol **91**: 570-579.

John, M., C. Schmitz, A. Y. Park, N. E. Dixon, T. Huber and G. Otting (2007). "Sequence-specific and stereospecific assignment of methyl groups using paramagnetic lanthanides." J Am Chem Soc **129**(44): 13749-13757.

Johnson, A. R., Y. W. Chen and E. E. Dekker (1998). "Investigation of a catalytic zinc binding site in Escherichia coli L-threonine dehydrogenase by site-directed mutagenesis of cysteine-38." Arch Biochem Biophys **358**(2): 211-221.

Kainosho, M., T. Torizawa, Y. Iwashita, T. Terauchi, A. Mei Ono and P. Guntert (2006). "Optimal isotope labelling for NMR protein structure determinations." Nature **440**(7080): 52-57.

Karagoz, G. E., A. M. Duarte, H. Ippel, C. Uetrecht, T. Sinnige, M. van Rosmalen, J. Hausmann, A. J. Heck, R. Boelens and S. G. Rudiger (2011). "N-terminal domain of human Hsp90 triggers binding to the cochaperone p23." Proc Natl Acad Sci U S A **108**(2): 580-585.

Karimi-Nejad, Y., J. M. Schmidt, H. Ruterjans, H. Schwalbe and C. Greisinger (1994). "Conformation of valine side chains in ribonuclease T1 determined by NMR studies of homonuclear and heteronuclear 3J coupling constants." Biochemistry **33**(18): 5481-5492.

Kasinath, V., K. Valentine and A. Wand (2013). " A 13C labeling strategy reveals a range of aromatic side chain motion in calmodulin." Journal American Chemistry Society **135**(26): 9560-9563.

Kato, H., H. van Ingen, B. R. Zhou, H. Feng, M. Bustin, L. E. Kay and Y. Bai (2011). "Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR." Proc Natl Acad Sci U S A **108**(30): 12283-12288.

Katz, J. J. and H. L. Crespi (1966). "Deuterated organisms: cultivation and uses." Science **151**(3715): 1187-1194.

Kay, L. E., D. A. Torchia and A. Bax (1989). "Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease." Biochemistry **28**(23): 8972-8979.

Kenyon, G. L. and T. W. Bruice (1977). "Novel sulfhydryl reagents." Methods Enzymol **47**: 407-430.

Kigawa, T., Y. Muto and S. Yokoyama (1995). "Cell-free synthesis and amino acid-selective stable isotope labeling of proteins for NMR analysis." J Biomol NMR **6**(2): 129-134.

Kigawa, T., T. Yabuki, N. Matsuda, T. Matsuda, R. Nakajima, A. Tanaka and S. Yokoyama (2004). "Preparation of Escherichia coli cell extract for highly productive cell-free protein expression." J Struct Funct Genomics **5**(1-2): 63-68.

Kim, S. H., B. L. Schneider and L. Reitzer (2010). "Genetics and regulation of the major enzymes of alanine synthesis in Escherichia coli." J Bacteriol **192**(20): 5304-5311.

Koharudin, L. M., A. M. Bonvin, R. Kaptein and R. Boelens (2003). "Use of very long-distance NOEs in a fully deuterated protein: an approach for rapid protein fold determination." J Magn Reson **163**(2): 228-235.

Kovtun, O., S. Mureev, W. Jung, M. H. Kubala, W. Johnston and K. Alexandrov (2011). "Leishmania cell-free protein expression system." Methods **55**(1): 58-64.

Landon, C., F. Barbault, M. Legrain, M. Guenneugues and F. Vovelle (2008). "Rational design of peptides active against the gram positive bacteria Staphylococcus aureus." Proteins **72**(1): 229-239.

Lee, K. M., E. J. Androphy and J. D. Baleja (1995). "A novel method for selective isotope labeling of bacterially expressed proteins." J Biomol NMR **5**(1): 93-96.

Lemercinier, X., F. W. Muskett, B. Cheeseman, P. B. McIntosh, L. Thim and M. D. Carr (2001). "High-resolution solution structure of human intestinal trefoil factor and functional insights from detailed structural comparisons with the other members of the trefoil family of mammalian cell motility factors." Biochemistry **40**(32): 9552-9559.

Lichtenecker, R., M. L. Ludwiczek, W. Schmid and R. Konrat (2004). "Simplification of protein NOESY spectra using bioorganic precursor synthesis and NMR spectral editing." J Am Chem Soc **126**(17): 5348-5349.

Lichtenecker, R. J., N. Coudevylle, R. Konrat and W. Schmid (2013). "Selective isotope labelling of leucine residues by using alpha-ketoacid precursor compounds." Chembiochem **14**(7): 818-821.

Lichtenecker, R. J., K. Weinhaupl, L. Reuther, J. Schorghuber, W. Schmid and R. Konrat (2013). "Independent valine and leucine isotope labeling in Escherichia coli protein overexpression systems." J Biomol NMR **57**(3): 205-209.

Lichtenecker, R. J., K. Weinhaupl, W. Schmid and R. Konrat (2013). "alpha-Ketoacids as precursors for phenylalanine and tyrosine labelling in cell-based protein overexpression." J Biomol NMR **57**(4): 327-331.

Linser, R., V. Gelev, F. Hagn, H. Arthanari, S. G. Hyberts and G. Wagner (2014). "Selective Methyl Labeling of Eukaryotic Membrane Proteins Using Cell-Free Expression." J Am Chem Soc.

Liu, J. J., R. Horst, V. Katritch, R. C. Stevens and K. Wuthrich (2012). "Biased signaling pathways in beta2-adrenergic receptor characterized by 19F-NMR." Science **335**(6072): 1106-1110.

Lo Conte, L., C. Chothia and J. Janin (1999). "The atomic structure of protein-protein recognition sites." J Mol Biol **285**(5): 2177-2198.

Lohr, F., V. Katsemi, J. Hartleib, U. Gunther and H. Ruterjans (2003). "A strategy to obtain backbone resonance assignments of deuterated proteins in the presence of incomplete amide 2H/1H back-exchange." J Biomol NMR **25**(4): 291-311.

London, R. E., B. D. Wingad and G. A. Mueller (2008). "Dependence of amino acid side chain 13C shifts on dihedral angle: application to conformational analysis." J Am Chem Soc **130**(33): 11097-11105.

Lopez-Mendez, B. and P. Guntert (2006). "Automated protein structure determination from NMR spectra." J Am Chem Soc **128**(40): 13112-13122.

Lopukhov, L. V., A. A. Ponomareva and L. O. Yagodina (2002). "Inhibition of bacterial pyridoxal-depending enzymes by (aminooxy)-acetic acid improves selective 15N isotope labeling of bacterially expressed protein." Biotechniques **32**(6): 1248-1250.

Madin, K., T. Sawasaki, T. Ogasawara and Y. Endo (2000). "A highly efficient and robust cell-free protein synthesis system prepared from wheat embryos: plants apparently contain a suicide system directed at ribosomes." Proc Natl Acad Sci U S A **97**(2): 559-564.

Makino, S., E. T. Beebe, J. L. Markley and B. G. Fox (2014). "Cell-free protein synthesis for functional and structural studies." Methods Mol Biol **1091**: 161-178.

Mal, T. K., S. J. Matthews, H. Kovacs, I. D. Campbell and J. Boyd (1998). "Some NMR experiments and a structure determination employing a [15N,2H] enriched protein." J Biomol NMR **12**(2): 259-276.

Marassi, F. M., A. Ramamoorthy and S. J. Opella (1997). "Complete resolution of the solid-state NMR spectrum of a uniformly 15N-labeled membrane protein in phospholipid bilayers." Proc Natl Acad Sci U S A **94**(16): 8551-8556.

Markus, M. A., K. T. Dayie, P. Matsudaira and G. Wagner (1994). "Effect of deuteration on the amide proton relaxation rates in proteins. Heteronuclear NMR experiments on villin 14T." J Magn Reson B **105**(2): 192-195.

Mas, G., E. Crublet, O. Hamelin, P. Gans and J. Boisbouvier (2013). "An efficient protocol for selective labeling of Valine methyl groups in high molecular weight protein assemblies.": In preparation.

Mas, G., E. Crublet, O. Hamelin, P. Gans and J. Boisbouvier (2013). "Specific labeling and assignment strategies of valine methyl groups for NMR studies of high molecular weight proteins." J Biomol NMR **57**(3): 251-262.

Massou, S., V. Puech, F. Talmont, P. Demange, N. D. Lindley, M. Tropis and A. Milon (1999). "Heterologous expression of a deuterated membrane-integrated receptor and partial deuteration in methylotrophic yeasts." J Biomol NMR **14**(3): 231-239.

May, S., M. Andreasson-Ochsner, Z. Fu, Y. X. Low, D. Tan, H. P. de Hoog, S. Ritz, M. Nallani and E. K. Sinner (2013). "In vitro expressed GPCR inserted in polymersome membranes for ligand-binding studies." Angew Chem Int Ed Engl **52**(2): 749-753.

McCaldon, P. and P. Argos (1988). "Oligopeptide biases in protein sequences and their use in predicting protein coding regions in nucleotide sequences." Proteins **4**(2): 99-122.

McGilvray, D. and H. E. Umbarger (1974). "Regulation of transaminase C synthesis in Escherichia coli: conditional leucine auxotrophy." J Bacteriol **120**(2): 715-723.

Melville, D. B., J. R. Rachele and E. B. Keller (1947). "A synthesis of methionine containing radiocarbon in the methyl group." J Biol Chem **169**(2): 419-426.

Mikami, S., T. Kobayashi, M. Masutani, S. Yokoyama and H. Imataka (2008). "A human cell-derived in vitro coupled transcription/translation system optimized for production of recombinant proteins." Protein Expr Purif **62**(2): 190-198.

Miller, S., J. Janin, A. M. Lesk and C. Chothia (1987). "Interior and surface of monomeric proteins." J Mol Biol **196**(3): 641-656.

Miyanoiri, Y., M. Takeda, J. Jee, A. M. Ono, K. Okuma, T. Terauchi and M. Kainosho (2011). "Alternative SAIL-Trp for robust aromatic signal assignment and determination of the chi(2) conformation by intra-residue NOEs." J Biomol NMR **51**(4): 425-435.

Miyanoiri, Y., M. Takeda, K. Okuma, A. M. Ono, T. Terauchi and M. Kainosho (2013). "Differential isotope-labeling for Leu and Val residues in a protein by E. coli cellular expression using stereo-specifically methyl labeled amino acids." J Biomol NMR **57**(3): 237-249.

Miyazawa-Onami, M., K. Takeuchi, T. Takano, T. Sugiki, I. Shimada and H. Takahashi (2013). "Perdeuteration and methyl-selective (1)H, (13)C-labeling by using a Kluyveromyces lactis expression system." J Biomol NMR **57**(3): 297-304.

Montelione, G. T., A. B. Lyons, S. D. Emerson and M. Tashiro (1992). "An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins." Journal American Chemistry Society **114**(27): 10974–10975.

Morgan, W. D., A. Kragt and J. Feeney (2000). "Expression of deuterium-isotope-labelled protein in the yeast pichia pastoris for NMR studies." J Biomol NMR **17**(4): 337-347.

Morita, E. H., T. Sawasaki, R. Tanaka, Y. Endo and T. Kohno (2003). "A wheat germ cell-free system is a novel way to screen protein folding and function." Protein Sci **12**(6): 1216-1221.

Morita, E. H., M. Shimizu, T. Ogasawara, Y. Endo, R. Tanaka and T. Kohno (2004). "A novel way of amino acid-specific assignment in (1)H-(15)N HSQC spectra with a wheat germ cell-free protein synthesis system." J Biomol NMR **30**(1): 37-45.

Muchmore, D. C., L. P. McIntosh, C. B. Russell, D. E. Anderson and F. W. Dahlquist (1989). "Expression and nitrogen-15 labeling of proteins for proton and nitrogen-15 nuclear magnetic resonance." Methods Enzymol **177**: 44-73.

Muchmore, S. W., M. Sattler, H. Liang, R. P. Meadows, J. E. Harlan, H. S. Yoon, D. Nettesheim, B. S. Chang, C. B. Thompson, S. L. Wong, S. L. Ng and S. W. Fesik (1996). "X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death." Nature **381**(6580): 335-341.

Mulder, A. M., C. Yoshioka, A. H. Beck, A. E. Bunner, R. A. Milligan, C. S. Potter, B. Carragher and J. R. Williamson (2010). "Visualizing ribosome biogenesis: parallel assembly pathways for the 30S subunit." Science **330**(6004): 673-677.

Mulder, F. A., A. Ayed, D. Yang, C. H. Arrowsmith and L. E. Kay (2000). "Assignment of 1H(N), 15N, 13C(alpha), 13CO and 13C(beta) resonances in a 67 kDa p53 dimer using 4D-TROSY NMR spectroscopy." J Biomol NMR **18**(2): 173-176.

Nakano, S., S. Okazaki, H. Tokiwa and Y. Asano (2014). "Binding of NAD+ and L-Threonine Induces Stepwise Structural and Flexibility Changes in Cupriavidus necator L-Threonine Dehydrogenase." J Biol Chem **289**(15): 10445-10454.

Neri, D., T. Szyperski, G. Otting, H. Senn and K. Wuthrich (1989). "Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional 13C labeling." Biochemistry **28**(19): 7510-7516.

Nibeilliu, M. E. and J. P. Malthouse (2004). "The stereospecificity and catalytic efficiency of the tryptophan synthase-catalysed exchange of the alpha-protons of amino acids." Biochem J **381**(Pt 3): 847-852.

Oda, Y., H. Nakamura, T. Yamazaki, K. Nagayama, M. Yoshida, S. Kanaya and M. Ikehara (1992). "1H NMR studies of deuterated ribonuclease HI selectively labeled with protonated amino acids." J Biomol NMR **2**(2): 137-147.

Omichinski, J. G., G. M. Clore, O. Schaad, G. Felsenfeld, C. Trainor, E. Appella, S. J. Stahl and A. M. Gronenborn (1993). "NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1." Science **261**(5120): 438-446.

Pickford, A. R. and J. M. O'Leary (2004). "Isotopic labeling of recombinant proteins from the methylotrophic yeast Pichia pastoris." Methods Mol Biol **278**: 17-33.

Plevin, M. and J. Boisbouvier (2012). Isotope-Labelling of Methyl Groups for NMR Studies of Large Proteins. Recent Developments in Biomolecular NMR. C. M. a. P. J, Royal Society of Chemistry.

Plevin, M. J. and J. Boisbouvier (2012). Isotope-Labelling of Methyl Groups for NMR Studies of Large Proteins. Recent Developments in Biomolecular NMR, Royal Society of Chemistry.

Plevin, M. J., O. Hamelin, J. Boisbouvier and P. Gans (2011). "A simple biosynthetic method for stereospecific resonance assignment of prochiral methyl groups in proteins." J Biomol NMR **49**(2): 61-67.

Pristovsek, P. and L. Franzoni (2006). "Stereospecific assignments of protein NMR resonances based on the tertiary structure and 2D/3D NOE data." J Comput Chem **27**(6): 791-797.

Qian, Y. Q., M. Billeter, G. Otting, M. Muller, W. J. Gehring and K. Wuthrich (1989). "The structure of the Antennapedia homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors." Cell **59**(3): 573-580.

Raap J, Niewenhuis S, Creemers A, Hexspoor S, Kragl U and L. J (1999). "Synthesis of Isotopically Labelled L-Phenylalanine and L-Tyrosine." Eur J Org Chem: 2609–2621.

Rajesh, S., D. Nietlispach, H. Nakayama, K. Takio, E. D. Laue, T. Shibata and Y. Ito (2003). "A novel method for the biosynthesis of deuterated proteins with selective protonation at the aromatic rings of Phe, Tyr and Trp." J Biomol NMR **27**(1): 81-86.

Ramesh, V., R. O. Frederick, S. E. Syed, C. F. Gibson, J. C. Yang and G. C. Roberts (1994). "The interactions of Escherichia coli trp repressor with tryptophan and with an operator oligonucleotide. NMR studies using selectively 15N-labelled protein." Eur J Biochem **225**(2): 601-608.

Rayment, I. (1997). "Reductive alkylation of lysine residues to alter crystallization properties of proteins." Methods Enzymol **276**: 171-179.

Religa, T. L. and L. E. Kay (2010). "Optimal methyl labeling for studies of supra-molecular systems." J Biomol NMR **47**(3): 163-169.

Religa, T. L., A. M. Ruschak, R. Rosenzweig and L. E. Kay (2011). "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 **133**(23): 9063-9068.

Religa, T. L., R. Sprangers and L. E. Kay (2010). "Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR." Science **328**(5974): 98-102.

Rodriguez, E. and N. R. Krishna (2001). "An economical method for (15)N/(13)C isotopic labeling of proteins expressed in Pichia pastoris." J Biochem **130**(1): 19-22.

Rosen, M. K., K. H. Gardner, R. C. Willis, W. E. Parris, T. Pawson and L. E. Kay (1996). "Selective methyl group protonation of perdeuterated proteins." J Mol Biol **263**(5): 627-636.

Rosenzweig, R. and L. E. Kay (2014). "Bringing dynamic molecular machines into focus by methyl-TROSY NMR." Annu Rev Biochem **83**: 291-315.

Rosenzweig, R., S. Moradi, A. Zarrine-Afsar, J. R. Glover and L. E. Kay (2013). "Unraveling the mechanism of protein disaggregation through a ClpB-DnaK interaction." Science **339**(6123): 1080-1083.

Ruschak, A. M. and L. E. Kay (2010). "Methyl groups as probes of supra-molecular structure, dynamics and function." J Biomol NMR **46**(1): 75-87.

Ruschak, A. M., T. L. Religa, S. Breuer, S. Witt and L. E. Kay (2010). "The proteasome antechamber maintains substrates in an unfolded state." Nature **467**(7317): 868-871.

Ruschak, A. M., A. Velyvis and L. E. Kay (2010). "A simple strategy for (1)(3)C, (1)H labeling at the Ile-gamma2 methyl position in highly deuterated proteins." J Biomol NMR **48**(3): 129-135.

Salvatella, X., C. M. Dobson, A. R. Fersht and M. Vendruscolo (2005). "Determination of the folding transition states of barnase by using PhiI-value-restrained simulations validated by double mutant PhiIJ-values." Proc Natl Acad Sci U S A **102**(35): 12389-12394.

Salzmann M , Pervushin K , Wider G , Hans S and W. K (2000). "NMR Assignment and Secondary Structure Determination of an Octameric 110 kDa Protein Using TROSY in Triple Resonance Experiments." J. Am. Chem. Soc. **122**(31): 7543–7548.

Sattler, M., H. Schwalbe and C. Griesinger (1992). "Stereospecific assignment of leucine methyl groups with carbon-13 in natural abundance or with random 13C labeling." Journal of the American Chemical Society **114**(3).

Schanda, P., E. Kupce and B. Brutscher (2005). "SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds." J Biomol NMR **33**(4): 199-211.

Schmidt, A., J. Sivaraman, Y. Li, R. Larocque, J. A. Barbosa, C. Smith, A. Matte, J. D. Schrag and M. Cygler (2001). "Three-dimensional structure of 2-amino-3-ketobutyrate CoA ligase from Escherichia coli complexed with a PLP-substrate intermediate: inferred reaction mechanism." Biochemistry **40**(17): 5151-5160.

Schmidt, E. and P. Guntert (2012). "A new algorithm for reliable and general NMR resonance assignment." J Am Chem Soc **134**(30): 12817-12829.

Seven, A. and J. Rizo (2012). Assigning the methyl resonances of the 73 kDa Munc13-1 MUN domain by mutagenesis. t. I.-p. n. P. TU. Lyon

Shimizu, Y., A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa and T. Ueda (2001). "Cell-free translation reconstituted with purified components." Nat Biotechnol **19**(8): 751-755.

Shimizu, Y., T. Kanamori and T. Ueda (2005). "Protein synthesis by pure translation systems." Methods **36**(3): 299-304.

Shindo, K., K. Masuda, H. Takahashi, Y. Arata and I. Shimada (2000). "Backbone 1H, 13C, and 15N resonance assignments of the anti-dansyl antibody Fv fragment." J Biomol NMR **17**(4): 357-358.

Sinha, K., L. Jen-Jacobson and G. S. Rule (2011). "Specific labeling of threonine methyl groups for NMR studies of protein-nucleic acid complexes." Biochemistry **50**(47): 10189-10191.

Sounier, R., L. Blanchard, Z. Wu and J. Boisbouvier (2007). "High-accuracy distance measurement between remote methyls in specifically protonated proteins." J Am Chem Soc **129**(3): 472-473.

Sprangers, R., A. Gribun, P. M. Hwang, W. A. Houry and L. E. Kay (2005). "Quantitative NMR spectroscopy of supramolecular complexes: dynamic side pores in ClpP are important for product release." Proc Natl Acad Sci U S A **102**(46): 16678-16683.

Sprangers, R. and L. E. Kay (2007). "Probing supramolecular structure from measurement of methyl (1)H-(13)C residual dipolar couplings." J Am Chem Soc **129**(42): 12668-12669.

Sprangers, R. and L. E. Kay (2007). "Quantitative dynamics and binding studies of the 20S proteasome by NMR." Nature **445**(7128): 618-622.

Sprangers, R., A. Velyvis and L. E. Kay (2007). "Solution NMR of supramolecular complexes: providing new insights into function." Nat Methods **4**(9): 697-703.

Steinmetz, A., M. Vyazmensky, D. Meyer, Z. E. Barak, R. Golbik, D. M. Chipman and K. Tittmann (2010). "Valine 375 and phenylalanine 109 confer affinity and specificity for pyruvate as donor substrate in acetohydroxy acid synthase isozyme II from Escherichia coli." Biochemistry **49**(25): 5188-5199.

Stoffregen, M. C., M. M. Schwer, F. A. Renschler and S. Wiesner (2012). "Methionine scanning as an NMR tool for detecting and analyzing biomolecular interaction surfaces." Structure **20**(4): 573-581.

Stratmann, D., E. Guittet and C. van Heijenoort (2010). "Robust structure-based resonance assignment for functional protein studies by NMR." J Biomol NMR **46**(2): 157-173.

Stratmann, D., C. van Heijenoort and E. Guittet (2009). "NOEnet--use of NOE networks for NMR resonance assignment of proteins with known 3D structure." Bioinformatics **25**(4): 474-481.

Strauss, A., F. Bitsch, B. Cutting, G. Fendrich, P. Graff, J. Liebetanz, M. Zurini and W. Jahnke (2003). "Amino-acid-type selective isotope labeling of proteins expressed in Baculovirus-infected insect cells useful for NMR studies." J Biomol NMR **26**(4): 367-372.

Strauss, A., F. Bitsch, G. Fendrich, P. Graff, R. Knecht, B. Meyhack and W. Jahnke (2005). "Efficient uniform isotope labeling of Abl kinase expressed in Baculovirus-infected insect cells." J Biomol NMR **31**(4): 343-349.

Su, X. C., C. T. Loh, R. Qi and G. Otting (2011). "Suppression of isotope scrambling in cell-free protein synthesis by broadband inhibition of PLP enymes for selective 15N-labelling and production of perdeuterated proteins in H2O." J Biomol NMR **50**(1): 35-42.

Sugiki, T., I. Shimada and H. Takahashi (2008). "Stable isotope labeling of protein by Kluyveromyces lactis for NMR study." J Biomol NMR **42**(3): 159-162.

Takahashi, H. and I. Shimada (2010). "Production of isotopically labeled heterologous proteins in non-E. coli prokaryotic and eukaryotic cells." J Biomol NMR **46**(1): 3-10.

Takeda, M. and M. Kainosho (2012). "Cell-free protein synthesis using E. coli cell extract for NMR studies." Adv Exp Med Biol **992**: 167-177.

Takeda, M., A. M. Ono, T. Terauchi and M. Kainosho (2010). "Application of SAIL phenylalanine and tyrosine with alternative isotope-labeling patterns for protein structure determination." J Biomol NMR **46**(1): 45-49.

Takeda, M., T. Terauchi and M. Kainosho (2012). "Conformational analysis by quantitative NOE measurements of the beta-proton pairs across individual disulfide bonds in proteins." J Biomol NMR **52**(2): 127-139.

Tang, C., J. Iwahara and G. M. Clore (2005). "Accurate determination of leucine and valine side-chain conformations using U-[15N/13C/2H]/[1H-(methine/methyl)-Leu/Val] isotope labeling, NOE pattern recognition, and methine Cgamma-Hgamma/Cbeta-Hbeta residual dipolar couplings: application to the 34-kDa enzyme IIA(chitobiose)." J Biomol NMR **33**(2): 105-121.

Teilum, K., U. Brath, P. Lundstrom and M. Akke (2006). "Biosynthetic 13C labeling of aromatic side chains in proteins for NMR relaxation measurements." J Am Chem Soc **128**(8): 2506-2507.

Tonelli, M., K. K. Singarapu, S. Makino, S. C. Sahu, Y. Matsubara, Y. Endo, M. Kainosho and J. L. Markley (2011). "Hydrogen exchange during cell-free incorporation of deuterated amino acids and an approach to its inhibition." J Biomol NMR **51**(4): 467-476.

Torchia, D. A., S. W. Sparks and A. Bax (1988). "NMR signal assignments of amide protons in the alpha-helical domains of staphylococcal nuclease." Biochemistry **27**(14): 5135-5141.

Torizawa, T., M. Shimizu, M. Taoka, H. Miyano and M. Kainosho (2004). "Efficient production of isotopically labeled proteins by cell-free synthesis: a practical protocol." J Biomol NMR **30**(3): 311-325.

Traaseth, N. J., R. Verardi and G. Veglia (2008). "Asymmetric methyl group labeling as a probe of membrane protein homo-oligomers by NMR spectroscopy." J Am Chem Soc **130**(8): 2400-2401.

Tugarinov, V., W. Y. Choy, V. Y. Orekhov and L. E. Kay (2005). "Solution NMR-derived global fold of a monomeric 82-kDa enzyme." Proc Natl Acad Sci U S A **102**(3): 622-627.

Tugarinov, V., V. Kanelis and L. E. Kay (2006). "Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy." Nat Protoc **1**(2): 749-754.

Tugarinov, V. and L. E. Kay (2003). "Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods." J Am Chem Soc **125**(45): 13868-13878.

Tugarinov, V. and L. E. Kay (2003). "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 **125**(19): 5701-5706.

Tugarinov, V. and L. E. Kay (2004). "An isotope labeling strategy for methyl TROSY spectroscopy." J Biomol NMR **28**(2): 165-172.

Tugarinov, V. and L. E. Kay (2004). "Stereospecific NMR assignments of prochiral methyls, rotameric states and dynamics of valine residues in malate synthase G." J Am Chem Soc **126**(31): 9827-9836.

Tugarinov, V., L. E. Kay, I. Ibraghimov and V. Y. Orekhov (2005). "High-resolution four-dimensional 1H-13C NOE spectroscopy using methyl-TROSY, sparse data acquisition, and multidimensional decomposition." J Am Chem Soc **127**(8): 2767-2775.

Tugarinov, V., R. Muhandiram, A. Ayed and L. E. Kay (2002). "Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase g." J Am Chem Soc **124**(34): 10025-10035.

Tugarinov, V., R. Sprangers and L. E. Kay (2007). "Probing side-chain dynamics in the proteasome by relaxation violated coherence transfer NMR spectroscopy." J Am Chem Soc **129**(6): 1743-1750.

Turano, P., D. Lalli, I. C. Felli, E. C. Theil and I. Bertini (2010). "NMR reveals pathway for ferric mineral precursors to the central cavity of ferritin." Proc Natl Acad Sci U S A **107**(2): 545-550.

Uhrinova, S., D. Uhrin, J. Nairn, N. C. Price, L. A. Fothergill-Gilmore and P. N. Barlow (2001). "Solution structure and dynamics of an open beta-sheet, glycolytic enzyme, monomeric 23.7 kDa phosphoglycerate mutase from Schizosaccharomyces pombe." J Mol Biol **306**(2): 275-290.

Umbarger, H. E. (1992). "The origin of a useful concept--feedback inhibition." Protein Sci **1**(10): 1392-1395.

Umbarger, H. E. (1996). "Biosynthesis of the branched-chain amino acids. In Escherichia Coli and Salmonella Typhimurium: Cellular and Molecular Biology." American Society for Microbiology Press: pp. 442–457.

Umbarger, H. E. and B. Brown (1957). "Threonine deamination in Escherichia coli. II. Evidence for two L-threonine deaminases." J Bacteriol **73**(1): 105-112.

Uzawa, T., N. Hamasaki and T. Oshima (1993). "Effects of novel polyamines on cell-free polypeptide synthesis catalyzed by Thermus thermophilus HB8 extract." J Biochem **114**(4): 478-486.

Velyvis A, Schachman H.K. and K. L.E (2009). "Assignment of Ile, Leu, and Val Methyl Correlations in Supra-Molecular Systems: An Application to Aspartate Transcarbamoylase." Journal American Chemistry Society **131**(45): 16534–16543.

Velyvis, A., A. M. Ruschak and L. E. Kay (2012). "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 **7**(9): e43725.

Velyvis, A., H. K. Schachman and L. E. Kay (2009). "Assignment of Ile, Leu, and Val methyl correlations in supra-molecular systems: an application to aspartate transcarbamoylase." J Am Chem Soc **131**(45): 16534-16543.

Velyvis, A., Y. R. Yang, H. K. Schachman and L. E. Kay (2007). "A solution NMR study showing that active site ligands and nucleotides directly perturb the allosteric equilibrium in aspartate transcarbamoylase." Proc Natl Acad Sci U S A **104**(21): 8815-8820.

Venditti, V., N. L. Fawzi and G. M. Clore (2011). "Automated sequence- and stereo-specific assignment of methyl-labeled proteins by paramagnetic relaxation and methyl-methyl nuclear Overhauser enhancement spectroscopy." J Biomol NMR **51**(3): 319-328.

Venters, R. A., T. L. Calderone, L. D. Spicer and C. A. Fierke (1991). "Uniform 13C isotope labeling of proteins with sodium acetate for NMR studies: application to human carbonic anhydrase II." Biochemistry **30**(18): 4491-4494.

Venters, R. A., C. C. Huang, B. T. Farmer, 2nd, R. Trolard, L. D. Spicer and C. A. Fierke (1995). "High-level 2H/13C/15N labeling of proteins for NMR studies." J Biomol NMR **5**(4): 339-344.

Vinarov, D. A., B. L. Lytle, F. C. Peterson, E. M. Tyler, B. F. Volkman and J. L. Markley (2004). "Cell-free protein production and labeling protocol for NMR-based structural proteomics." Nat Methods **1**(2): 149-153.

Vuister, G. W., T. Yamazaki, D. A. Torchia and A. Bax (1993). "Measurement of two- and three-bond 13C-1H J couplings to the C delta carbons of leucine residues in staphylococcal nuclease." J Biomol NMR **3**(3): 297-306.

Vyazmensky, M., C. Sella, Z. Barak and D. M. Chipman (1996). "Isolation and characterization of subunits of acetohydroxy acid synthase isozyme III and reconstitution of the holoenzyme." Biochemistry **35**(32): 10339-10346.

Walters, K. J., A. E. Ferentz, B. J. Hare, P. Hidalgo, A. Jasanoff, H. Matsuo and G. Wagner (2001). "Characterizing protein-protein complexes and oligomers by nuclear magnetic resonance spectroscopy." Methods Enzymol **339**: 238-258.

Wang, H., D. A. Janowick, J. M. Schkeryantz, X. Liu and S. W. Fesik (1999). "A Method for Assigning Phenylalanines in Proteins." Journal American Chemistry Society **121**(7): 1611–1612.

Wang, J. G., P. K. Lee, Y. H. Dong, S. S. Pang, R. G. Duggleby, Z. M. Li and L. W. Guddat (2009). "Crystal structures of two novel sulfonylurea herbicides in complex with Arabidopsis thaliana acetohydroxyacid synthase." FEBS J **276**(5): 1282-1290.

Waugh, D. S. (1996). "Genetic tools for selective labeling of proteins with alpha-15N-amino acids." J Biomol NMR **8**(2): 184-192.

Weigelt, J., M. van Dongen, J. Uppenberg, J. Schultz and M. Wikstrom (2002). "Site-selective screening by NMR spectroscopy with labeled amino acid pairs." J Am Chem Soc **124**(11): 2446-2447.

Weininger, U., Z. Liu, D. D. McIntyre, H. J. Vogel and M. Akke (2012). "Specific 12CbetaD(2)12CgammaD(2)S13CepsilonHD(2) isotopomer labeling of methionine to characterize protein dynamics by 1H and 13C NMR relaxation dispersion." J Am Chem Soc **134**(45): 18562-18565.

Werner, K., C. Richter, J. Klein-Seetharaman and H. Schwalbe (2008). "Isotope labeling of mammalian GPCRs in HEK293 cells and characterization of the C-terminus of bovine rhodopsin by high resolution liquid NMR spectroscopy." J Biomol NMR **40**(1): 49-53.

Whalen, W. A. and C. M. Berg (1982). "Analysis of an avtA::Mu d1(Ap lac) mutant: metabolic role of transaminase C." J Bacteriol **150**(2): 739-746.

Whalen, W. A. and C. M. Berg (1984). "Gratuitous repression of avtA in Escherichia coli and Salmonella typhimurium." J Bacteriol **158**(2): 571-574.

Whittaker, J. W. (2013). "Cell-free protein synthesis: the state of the art." Biotechnol Lett **35**(2): 143-152.

Wishart, D. S. and B. D. Sykes (1994). "The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data." J Biomol NMR **4**(2): 171-180.

Wood, M. J. and E. A. Komives (1999). "Production of large quantities of isotopically labeled protein in Pichia pastoris by fermentation." J Biomol NMR **13**(2): 149-159.

Wyss, D. F., K. T. Dayie and G. Wagner (1997). "The counterreceptor binding site of human CD2 exhibits an extended surface patch with multiple conformations fluctuating with millisecond to microsecond motions." Protein Sci **6**(3): 534-542.

Xu, Y., M. Liu, P. J. Simpson, R. Isaacson, E. Cota, J. Marchant, D. Yang, X. Zhang, P. Freemont and S. Matthews (2009). "Automated assignment in selectively methyl-labeled proteins." J Am Chem Soc **131**(27): 9480-9481.

Xu, Y., M. Liu, P. J. Simpson, R. Isaacson, E. Cota, J. Marchant, D. Yang, X. Zhang, P. Freemont and S. Matthews (2009). "Automated assignment in selectively methyl-labelled proteins." J Am Chem Soc **131**(27): 9480-9481.

Xu, Y. and S. Matthews (2013). "MAP-XSII: an improved program for the automatic assignment of methyl resonances in large proteins." J Biomol NMR **55**(2): 179-187.

Yamazaki, T., L. Lee, Arrowsmith C.H, D. R. Muhandiram and L. E. Kay (1994). "A Suite of Triple Resonance NMR Experiments for the Backbone Assignment of 15N, 13C, 2H Labeled Proteins with High Sensitivity." Journal American Chemistry Society **116**(26): 11655–11666.

Yamazaki, T., L. Weontae, R. Muhandiram and L. E. Kay (1994). "NMR Experiments for the Measurement of Carbon Relaxation Properties in Highly Enriched, Uniformly 13C,15N-Labeled Proteins: Application to 13C.alpha. Carbons." Journal American Chemistry Society **116**(18): 8266–8278.

Yang, D., A. Mittermaier, Y. K. Mok and L. E. Kay (1998). "A study of protein side-chain dynamics from new 2H auto-correlation and 13C cross-correlation NMR experiments: application to the N-terminal SH3 domain from drk." J Mol Biol **276**(5): 939-954.

Yang, X., J. L. Welch, J. J. Arnold and D. D. Boehr (2010). "Long-range interaction networks in the function and fidelity of poliovirus RNA-dependent RNA polymerase studied by nuclear magnetic resonance." Biochemistry **49**(43): 9361-9371.

Yoneyama, H., H. Hori, S. J. Lim, T. Murata, T. Ando, E. Isogai and R. Katsumata (2011). "Isolation of a mutant auxotrophic for L-alanine and identification of three major aminotransferases that synthesize L-alanine in Escherichia coli." Biosci Biotechnol Biochem **75**(5): 930-938.

Yu, X., Y. Li and X. Wang (2013). "Molecular evolution of threonine dehydratase in bacteria." PLoS One **8**(12): e80750.

Zhou, Y., H. Asahara, E. A. Gaucher and S. Chong (2012). "Reconstitution of translation from Thermus thermophilus reveals a minimal set of components sufficient for protein synthesis at high temperatures and functional conservation of modern and ancient translation components." Nucleic Acids Res **40**(16): 7932-7945.