

An Economical Method for Production of ^2H , $^{13}\text{CH}_3$ -Threonine for Solution NMR Studies of Large Protein Complexes: Application to the 670 kDa Proteasome

Algirdas Velyvis^{1*}, Amy M. Ruschak¹, Lewis E. Kay^{1,2*}

1 Departments of Molecular Genetics, Biochemistry, and Chemistry, University of Toronto, Toronto, Ontario, Canada, **2** Program in Molecular Structure and Function, Hospital for Sick Children, Toronto, Ontario, Canada

Abstract

NMR studies of very high molecular weight protein complexes have been greatly facilitated through the development of labeling strategies whereby $^{13}\text{CH}_3$ methyl groups are introduced into highly deuterated proteins. Robust and cost-effective labeling methods are well established for all methyl containing amino acids with the exception of Thr. Here we describe an inexpensive biosynthetic strategy for the production of L- $[\alpha\text{-}^2\text{H}; \beta\text{-}^2\text{H}; \gamma\text{-}^{13}\text{C}]$ -Thr that can then be directly added during protein expression to produce highly deuterated proteins with Thr methyl group probes of structure and dynamics. These reporters are particularly valuable, because unlike other methyl containing amino acids, Thr residues are localized predominantly to the surfaces of proteins, have unique hydrogen bonding capabilities, have a higher propensity to be found at protein nucleic acid interfaces and can play important roles in signaling pathways through phosphorylation. The utility of the labeling methodology is demonstrated with an application to the 670 kDa proteasome core particle, where high quality Thr ^{13}C , ^1H correlation spectra are obtained that could not be generated from samples prepared with commercially available U- $[\text{}^{13}\text{C}, \text{}^1\text{H}]$ -Thr.

Citation: Velyvis A, Ruschak AM, Kay LE (2012) An Economical Method for Production of ^2H , $^{13}\text{CH}_3$ -Threonine for Solution NMR Studies of Large Protein Complexes: Application to the 670 kDa Proteasome. PLoS ONE 7(9): e43725. doi:10.1371/journal.pone.0043725

Editor: Ramón Campos-Olivas, Spanish National Cancer Center, Spain

Received: June 13, 2012; **Accepted:** July 24, 2012; **Published:** September 11, 2012

Copyright: © 2012 Velyvis et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Salary support for A.M.R. was obtained in the form of a Canadian Institutes of Health Research (CIHR) post-doctoral fellowship. This work was supported by grants from the CIHR and the Natural Sciences and Engineering Research Council of Canada (L.E.K.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: kay@pound.med.utoronto.ca (LEK); algirdas.velyvis@utoronto.ca (AV)

Introduction

Many of the advances in protein NMR spectroscopy can be directly traced to the development of isotope labeling strategies that have substantially increased the range of biomolecular systems that can be explored [1–6]. For example, applications to very high molecular weight complexes have benefited significantly from the preparation of highly deuterated molecules where the relaxation times of the remaining NMR probes, typically backbone amide moieties or side-chain methyl groups, are significantly increased [7,8]. Concomitant with the emergence of these important labeling approaches has been the advancement of new NMR experiments that exploit the labeling in ways that permit the recording of spectra of both increased sensitivity and resolution [9,10].

Over the past 15 years our laboratory has developed a strategy for studying high molecular weight protein complexes that involves $^{13}\text{CH}_3$ labeling of Ile ($\delta 1$ or $\gamma 2$), Leu and Val methyl positions in an otherwise highly deuterated ^{12}C background [11–13]. Spectra are recorded that make use of a methyl-TROSY effect that results in significant line-narrowing [14]. Applications of this methodology to a large number of systems have now been reported [15–27], along with schemes for extending the labeling to Ala [21,28] and Met [16,29,30] methyl positions or for stereospecific incorporation of methyl labels at either proR or proS positions of Leu and Val side-chains [31]. More recently an

approach for placement of methyl groups at positions of interest has been introduced involving substitution of the native residue with Cys and subsequently reacting with ^{13}C -methyl-methanethio-sulfonate (^{13}C -MMTS) [32].

Ile, Leu, Val comprise approximately 20% of the amino acids in a 'typical' protein, and Ala, Met approximately 10% and 2%, respectively [33]. It is thus expected that in many cases these residues, in various combinations, will provide 'excellent coverage' of the protein in the sense that they will be found in regions that contribute in important ways to the structure or dynamics of the molecule studied. However, as pointed out by Rule and coworkers [34] these residues are under-represented at protein-nucleic acid interfaces. Moreover, Ile, Leu, Val and Met are predominantly partitioned inside proteins, while Ala has a small preference for the interior as well [35]. Thus, these residues are not effective probes of protein surfaces. The one remaining methyl containing residue, Thr, has both a much higher relative propensity for placement at protein-nucleic acid interfaces [34] and, not surprisingly, also a higher composition on protein surfaces relative to the interior [35]. In addition, of all methyl-containing residues the hydrogen bonding functionality of the Thr side-chain is unique. Finally, like other amino acids, Thr residues can play critically important roles in protein function, such as is the case for the proteasome, a Thr protease [36,37], that forms the basis of a large research effort in our laboratory (see below).

Rule and coworkers have recently introduced a method to label methyl groups of Thr in recombinant proteins with the $^{13}\text{CHD}_2$ isotopomer by using $2\text{-}^{13}\text{C}$ -glycerol and bicarbonate during bacterial growth in D_2O [34]. This methodology is most certainly applicable to small-intermediate sized proteins, as illustrated by this group. However, applications to very large complexes with aggregate molecular weights in the hundreds of kDa will be compromised by an extent of labeling of only approximately 25%, by the fact that fully protonated glycerol is used as the carbon source so that protons will be introduced in positions other than methyl groups and because it is not possible to generate the $^{13}\text{CH}_3$ methyl isomer while still retaining a high level of protein deuteration. Kainosho and coworkers have developed an elegant approach for protein labeling (termed the SAIL method) in which amino-acids with the desired labeling pattern are prepared via organic chemistry and then added, typically to a cell-free protein synthesis system, to produce suitably labeled proteins [4]. Recently, Thr labeled with ^2H and ^{13}C at the $\text{C}\beta$ (C_3) position was prepared for studies of side-chain hydrogen exchange [38]. Unfortunately, the cost of such labeled amino acids can be high. Moreover, Thr with the labeling pattern that we seek here for studies of supra-molecular protein complexes (see below) is not commercially available presently from SAIL Technologies Inc., suppliers of SAIL amino acids.

With this in mind we describe here the biosynthesis of L- $[\alpha\text{-}^2\text{H};\beta\text{-}^2\text{H};\gamma\text{-}^{13}\text{C}]$ Thr ($\text{U-}[^2\text{H}],\text{Thr-}\gamma 2[^{13}\text{CH}_3]$) starting from the relatively inexpensive precursor, ^{13}C formaldehyde (60% yield). All of the enzymes necessary for the synthesis have been expressed and purified and are available upon request. The utility of the methodology is illustrated with an application to the 20S proteasome core particle from *T. acidophilum*, 670 kDa, that plays an integral role in cellular homeostasis [37]. We show that all 15 of the expected correlations from Thr methyl groups of the labeled β -subunit of the enzyme are present in $^{13}\text{C},^1\text{H}$ NMR correlation maps; notably only a modest fraction of the expected peaks could be observed when commercially available, uniformly ^{13}C and fully protonated Thr was used as a precursor. Assignments of the majority of the Thr resonances could be made from NOESY data sets correlating previously assigned Ile, Leu, Val methyls with Thr methyl groups. With the development of a cost effective strategy for the biosynthesis of $\text{U-}[^2\text{H}],\text{Thr-}\gamma 2[^{13}\text{CH}_3]$ it is now possible to generate highly deuterated samples of very high molecular weight protein complexes with Thr probes of structure and dynamics.

Results and Discussion

Biosynthesis of $\text{U-}[^2\text{H}],\text{Thr-}\gamma 2[^{13}\text{CH}_3]$

Prior to the development of a strategy for the production of $\text{U-}[^2\text{H}],\text{Thr-}\gamma 2[^{13}\text{CH}_3]$ labeled proteins we first examined the biosynthetic pathway of Thr to evaluate whether an appropriately labeled precursor of this amino acid could be generated that would then be added during protein expression. Such an approach has been used successfully in the development of 2-ketobutyrate and 2-ketovalarate, precursors for Ile($\delta 1$) and Leu/Val respectively [11,12], that are used to produce $\text{U-}[^2\text{H}],\text{Ile-}\delta 1[^{13}\text{CH}_3]$ - or $\text{U-}[^2\text{H}],\text{Leu,Val-}[^{13}\text{CH}_3,^{12}\text{CD}_3]$ -labeled proteins [13] and more recently for the generation of precursors that result in proteins with Ile $\gamma 2$ CH_3 labeling [39] or stereospecific CH_3 methyl group labeling of Leu and Val residues [31]. Figure 1 illustrates the biosynthesis of Thr (boxed compound) starting from Asp. Of particular interest is that the final step, catalyzed by threonine synthase (TS), involves addition of a solvent hydrogen to the 4 position of O-phosphohomoserine. Because protein expression is carried out in D_2O , a prerequisite for the generation of highly

deuterated proteins that is required for NMR studies of very high molecular weight complexes, the methyl group of the Thr so produced will contain at least one deuterium. It has been shown previously that the optimum labeling strategy for methyl groups is $^{13}\text{CH}_3$ so that a methyl-TROSY effect can be exploited [40,41]. Addition of a precursor in the Thr synthesis pathway is thus not a viable option for the production of proteins with $\text{U-}[^2\text{H}],\text{Thr-}\gamma 2[^{13}\text{CH}_3]$.

An alternative strategy is one where a precursor is introduced to a cell line where the main, TS-dependent, biosynthetic pathway is inactivated. In this manner Thr could be assembled from glycine and a $^{13}\text{CH}_3$ -labeled acetyl fragment (either acetaldehyde or acetyl-CoA) via threonine aldolase or 2-amino-3-ketobutyrate ligase, Figure 1. Neglecting questions relating to the efficiency of such a scheme, the central role of acetyl-CoA in cellular metabolism will inevitably result in any acetyl fragment labeling other amino acids with ^1H and ^{13}C .

We therefore turned to an approach whereby $\text{U-}[^2\text{H}],\text{Thr-}\gamma 2[^{13}\text{CH}_3]$ would be added to growth media during protein expression. Several schemes for the ‘in-vitro’, enzymatic synthesis of Thr are possible, as indicated in Figure 1. First, condensation of $1\text{-}^2\text{H},2\text{-}^{13}\text{C}$ acetaldehyde with deuterated glycine via threonine aldolase [42,43] generates Thr with the desired labeling pattern. However, acetaldehyde labeled in this manner is expensive (\$5500/g, minimum order of 3 g) and both L-Thr (2S,3R) and L-allo-Thr (2S,3S) are produced, necessitating separation. Second, Thr can be produced via a pathway involving steps catalyzed by 2-amino-3-ketobutyrate ligase and threonine dehydrogenase [44], with ^{13}C -labeled sodium acetate serving as a source of acetyl-CoA (Figure 1). However, the production of large quantities of Thr is likely to be compromised because (i) the intermediate 2-amino-3-ketobutyrate has a half-life estimated to be less than 1 minute [45] or 10 minutes [46] at pH 7, with spontaneous decarboxylation diverting a large fraction of the isotopically labeled material into aminoacetone [44] and (ii) 2-amino-3-ketobutyrate ligase also possesses some threonine aldolase activity [47] which would destroy the desired product (Thr) in a coupled enzymatic reaction. A third approach is also suggested in Figure 1, starting from 2-ketobutyrate (2KB, $^{13}\text{CH}_3^{12}\text{CD}_2\text{COCOONa}$), a well-established and inexpensive commercially available precursor for labeling of Ile $\delta 1$ positions with ^1H and ^{13}C [11]. 2KB can be transaminated into L-2-aminobutyrate either by branched chain aminotransferase or an engineered dehydrogenase [48]. If an enzyme could be found that oxidizes 2-aminobutyrate to Thr it would enable conversion of 2KB into the desired labeled product. While it is known that monooxygenases of the cytochrome P450 family can hydroxylate aliphatic carbon positions [49,50] and indeed P450BM3 which is specific for fatty acids can be engineered to hydroxylate butyrate [51], to the best of our knowledge a P450 that produces Thr has not been reported.

The strategy that we have adopted, therefore, is to synthesize Thr by ‘reproducing’ the major Thr biosynthesis pathway in the ‘test-tube’, Figure 1 top, whereby Asp is converted into homoserine in 3 steps and then to Thr by homoserine kinase (HSK) and threonine synthase. A simplification to the path taken by nature involves generating homoserine **3** via transamination of the keto acid, 2-keto-4-hydroxybutyrate (KHB, **2**), as indicated in Figure 2, which summarizes the Thr biosynthesis route used here. The first critical step is condensation of ^2H , pyruvate **1b** with ^{13}C -formaldehyde, catalyzed via the enzyme 2-keto-4-hydroxyglutamate aldolase [52,53] (KHGA, also known as 2-keto-3-deoxy-6-phosphogluconate aldolase), a member of the pyruvate aldolase family that has been used in biocatalysis previously [54]. The product of this reaction (referred to as reaction 1 in what follows) is

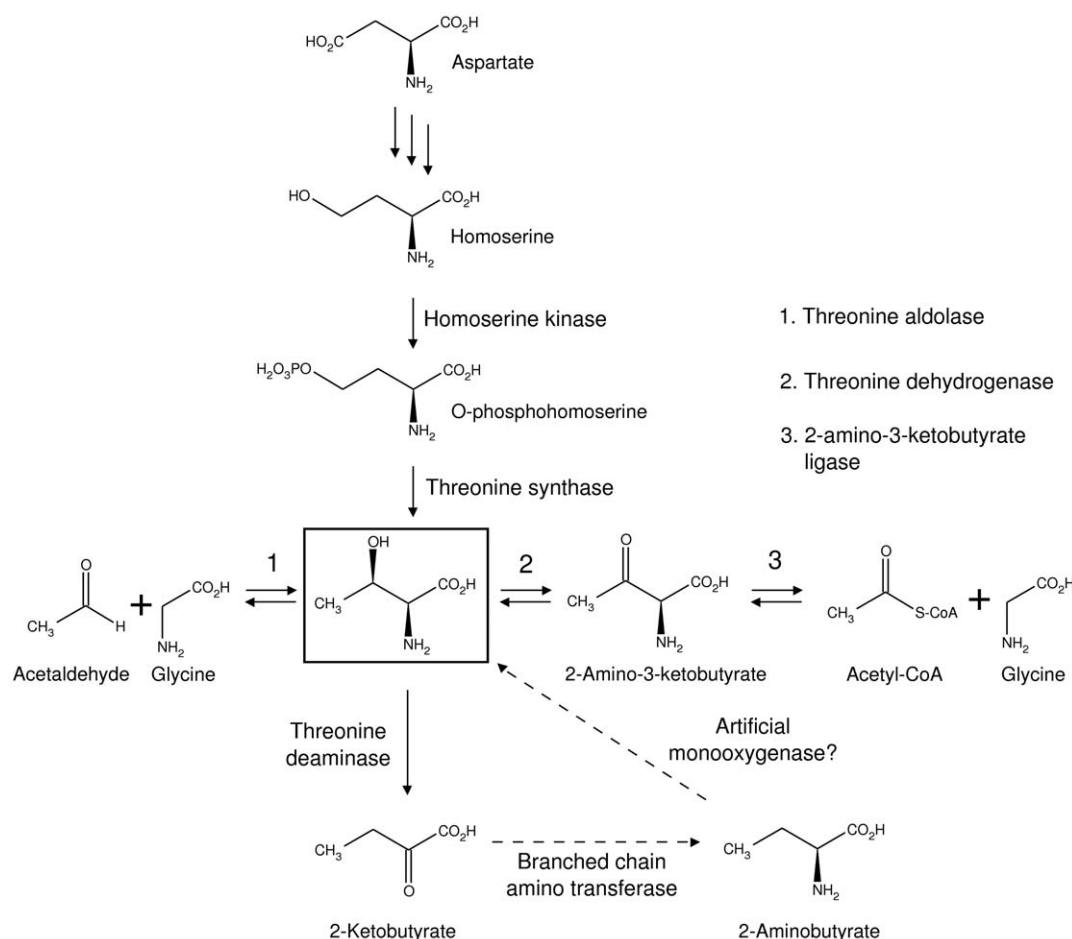


Figure 1. Metabolism of Thr starting from Asp. The structure of Thr is highlighted in the box and the Thr to Asp pathway is along the vertical. Also shown are pathways that potentially can be exploited to synthesize Thr and the enzymes that are involved (horizontal). Dashed lines (bottom) indicate the non-physiological conversion of 2-ketobutyrate into L-2-aminobutyrate and then into Thr. doi:10.1371/journal.pone.0043725.g001

KHB **2**, in which the entire carbon chain for Thr is assembled with the correct deuteration and ^{13}C labeling pattern at C3 and C4. Importantly, the ^{13}C -formaldehyde precursor is available commercially and is relatively inexpensive, with a list price for ^{13}C -paraformaldehyde of \$720/gram that compares very favorably to the cost of a custom synthesis of 1- ^2H ,2- ^{13}C acetaldehyde (\$16,500 for a minimum order of 3 grams). Additionally, ^2H -pyruvate, **1b**, is readily generated from pyruvate **1a** via incubation with catalytic amounts of KHGA in D_2O based buffer for 1 hour at pH 7.5. The equilibrium constant for the production of **2** from **1b** is 250 M^{-1} [52] so that by adding excess pyruvate the reaction can be made to go to completion; in a mixture starting as 0.1 M formaldehyde and 0.2 M pyruvate 96% of the starting formaldehyde is predicted to be converted to KHB. Indeed, when depolymerized ^{13}C -paraformaldehyde (0.11 M) is mixed with a two-fold excess of pyruvate in the presence of KHGA the aldehyde signals in ^1H NMR spectra disappear, while those from KHB appear, with greater than 95% conversion (see Supporting Information S1). Moreover, when the reaction is carried out in 99.9% D_2O only the C4 attached hydrogens in **2** are observed in spectra, as expected (Supporting Information S1). Upon attainment of equilibrium (within several hours when 0.014 mole percent of KHGA is used at room temperature) the reaction solution is filtered to remove the biocatalyst and used in the next step without further purification.

The transamination of KHB **2** to homoserine **3** can reportedly be catalyzed by aspartate and alanine aminotransferases [52]. Aspartate aminotransferase (AAT, see Supporting Information S1) was the most efficient of several enzymes tested with 60% conversion to homoserine in 3 hours using 1.5 equivalents of Glu as the amino donor. A point of concern regarding transamination reactions is their reversibility since ‘similar’ types of compounds are interconverted (amino acid+keto acid \rightarrow keto acid+amino acid). A large excess of Glu may be required to drive the reaction of interest to completion, unless it can be coupled to an essentially irreversible downstream reaction, such as phosphorylation by ATP. For this reason it is beneficial to combine transamination of **2** and phosphorylation of **3** into a single procedure (we designate this “reaction 2–3”). NMR analysis indicates that reaction 2–3 stops after about 40 hours with approximately 20% of the original 0.1 M KHB **2** remaining and only 1% homoserine **3** (Supporting Information S1). At intermediate times we observed much higher levels of homoserine as illustrated in Supporting Information S1, where the reaction on fully protonated precursors in H_2O solvent is shown. The simplest explanation for these levels of conversion is that transamination of **2** proceeds rapidly but that AAT has a high K_M for KHB (so that the reaction effectively terminates at a low concentration of KHB, 20 mM). In contrast, because homoserine **3** is the canonical substrate for HSK it has a much higher affinity, $K_M < 1 \text{ mM}$ [55], but the kinetics of the HSK reaction are slower

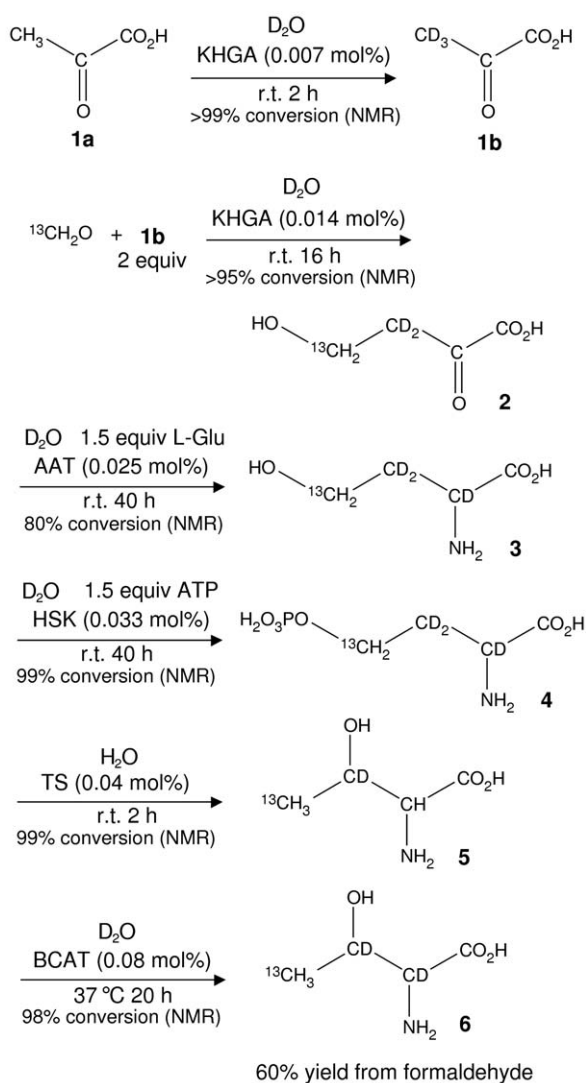


Figure 2. Scheme used for the biosynthesis of U-[²H], Thr- γ 2[¹³CH₃], 6. Enzyme name abbreviations: KHGA = 2-keto-4-hydroxyglutarate aldolase, AAT = aspartate aminotransferase, HSK = homoserine kinase, TS = threonine synthase, BCAT = branched-chain-amino-acid aminotransferase.

doi:10.1371/journal.pone.0043725.g002

than for the AAT catalyzed step (hence the observation of the homoserine intermediate).

O-phospho-L-homoserine, OPHS **4**, was subsequently purified from the 2–3 reaction mixture using ion exchange chromatography [56] (Supporting Information S1), removing any protonaceous organic contaminants such as 2-ketoglutarate, glutamate and possibly ADP/ATP from the transamination step (Supporting Information S1). Purified OPHS (~0.1 M, also containing up to 1 M NaCl from the ion exchange eluent) is efficiently converted by TS in H₂O to Thr **5** and one equivalent of inorganic phosphate. However, an inevitable consequence of the TS-catalyzed reaction in H₂O is protonation of **5** at C2. This necessitates a final H/D exchange step. Water is evaporated, the resultant Thr **5**, phosphate and NaCl mixture dissolved in 99.9% D₂O and the enzyme, BCAT (branched-chain-amino-acid aminotransferase) added to catalyze H/D exchange, producing Thr with the desired labeling pattern, **6**.

NMR spectra of the final product **6** are shown in Figure 3, establishing excellent purity, a high degree of perdeuteration at C2 and C3 and low amounts of contaminants. After filtering out the last biocatalyst the D₂O solution containing 78 mM Thr **6**, ~80 mM sodium phosphate, ~1 M NaCl, and ~0.5 mM pyridoxal phosphate can be used directly for recombinant protein labeling or stored in frozen form. Two batches of Thr were synthesized in this manner and in each case 2.5 mmol amino acid was produced (300 mg) from a starting amount of 4.1 mmol (127 mg) paraformaldehyde, corresponding to a yield of 60%.

Testing labeling strategies

In order to establish the efficiency of Thr labeling in proteins and to determine both the amount of precursor to add and whether other compounds should be included simultaneously in the growth medium we have prepared several samples of U-[¹³C,¹H] labeled Abp1p SH3 domain where ¹²C labeled precursors such as Thr, 2KB or 2-ketoisovalerate (2KIV) have been added approximately one hour prior to the induction of protein overexpression (see Supporting Information S1) in standard BL21(DE3) *E. coli* cells. Because of the small size of the domain (~60 amino acids) it is straightforward to quantify peak intensities accurately in a series of constant time ¹³C,¹H HSQC data sets. Decreases in peak intensities reflect incorporation of (unlabeled) precursor, providing a gauge of the extent of both desired and undesired labeling.

Figure 4 displays the fraction of ¹³C intensity lost at Thr C γ (A), Ile C δ 1 (B) and Glu C γ (C) positions as a function of added precursor. The nature of the precursor(s) and the amounts added are indicated along the x-axis. Addition of 50 mg/L Thr leads to ~50% labeling at both Thr γ 2 and Ile δ 1 positions (sample 1). Labeling of Ile is not unexpected since Thr is a precursor of Ile biosynthesis [57]. In order to avoid it Rule and coworkers add perdeuterated Ile to their growth media [34], similar to Ayala *et al* in their strategy for labeling Ala methyl groups [28]. Here we have taken an opposite approach. Because cross peaks for Ile residues fall in an isolated region of the ¹³C,¹H correlation map and are among the most well resolved of all methyl types we prefer to include Ile labeling in all of our Thr samples (and often also Leu, Val). The advantages are two-fold. First, the increased proton density in the protein, relative to Thr only ¹³CH₃ methyl samples, reduces ¹H T₁ values leading to sensitivity improvements. Second, addition of the inexpensive 2KB (Ile) precursor ensures that the added U-[²H], Thr- γ 2[¹³CH₃] is not diverted to Ile, leading to higher incorporation of the desired Thr label. Thus, by adding both 50 mg/mL Thr and 50 mg/mL 2KB the fractional labeling increases to 75% and >95% for Thr γ 2 and Ile δ 1, respectively. Alternatively, addition of larger quantities of Thr (100 mg, sample 6) ensures that, despite the diversion of this compound to Ile, there is still a sufficient amount to achieve high labeling of Thr C γ 2 (>90%). Increasing the quantities of added Thr does lead to some isotope scrambling, however. ¹³C,¹H HSQC spectra of the SH3 domain indicate that when 100 mg/L of ¹H,¹²C Thr is added to the growth medium there is an approximately 40% reduction in peaks from Gly (Supporting Information S1). Note that Thr can be converted to Acetyl-CoA and Gly via the threonine dehydrogenase pathway [44], as shown in Figure 1. In a ‘real’ sample where U-[²H], Thr- γ 2[¹³CH₃] is added to a D₂O based growth the Gly produced is fully deuterated so that additional protons are not added to the protein. However, this pathway does dilute the Thr precursor. Moreover, the ¹³CH₃ acetyl-CoA so produced is a precursor in the synthesis of Glu, with the H γ and C γ positions derived from the methyl group. The level of incorporation at these sites does not appear to be high, however, Figure 4C. Addition of

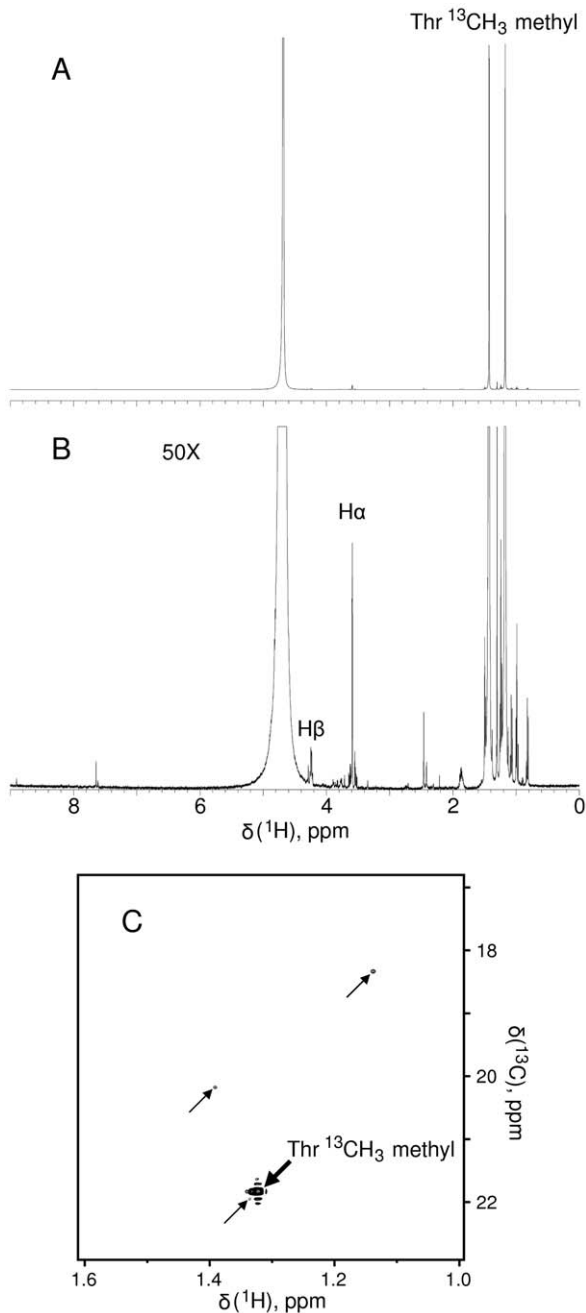


Figure 3. NMR spectra of L-[α - ^2H ; β - ^2H ; γ - ^{13}C]-Thr, **6.** (A) ^1H NMR spectrum with vertical scale adjusted to fit entire methyl doublet. (B) Same spectrum as in A with a 50-fold enhanced vertical scale. (C) ^{13}C , ^1H HSQC spectrum of the same sample, **6**. Contour level is set to 0.5% of the Thr methyl signal. The only 3 other peaks (all under 1% of the intensity of the Thr methyl) derive from impurities and are indicated with arrowheads. Approximately 2% residual protonation remains at the α position and much less at β . doi:10.1371/journal.pone.0043725.g003

^{13}C -Gly during the production of SH3 domain samples generated from unlabeled Thr and 2KB does reduce the extent of scrambling from Thr to Gly and slightly to Glu. Importantly, the extent of Thr labeling increases to close to 90% from 75% (compare samples 2 and 4), indicating that diversion of Thr into undesirable pathways is at least partially inhibited. Given that perdeuterated glycine is inexpensive we suggest 50 mg/L labeled U- ^2H , Thr- γ 2[$^{13}\text{CH}_3$],

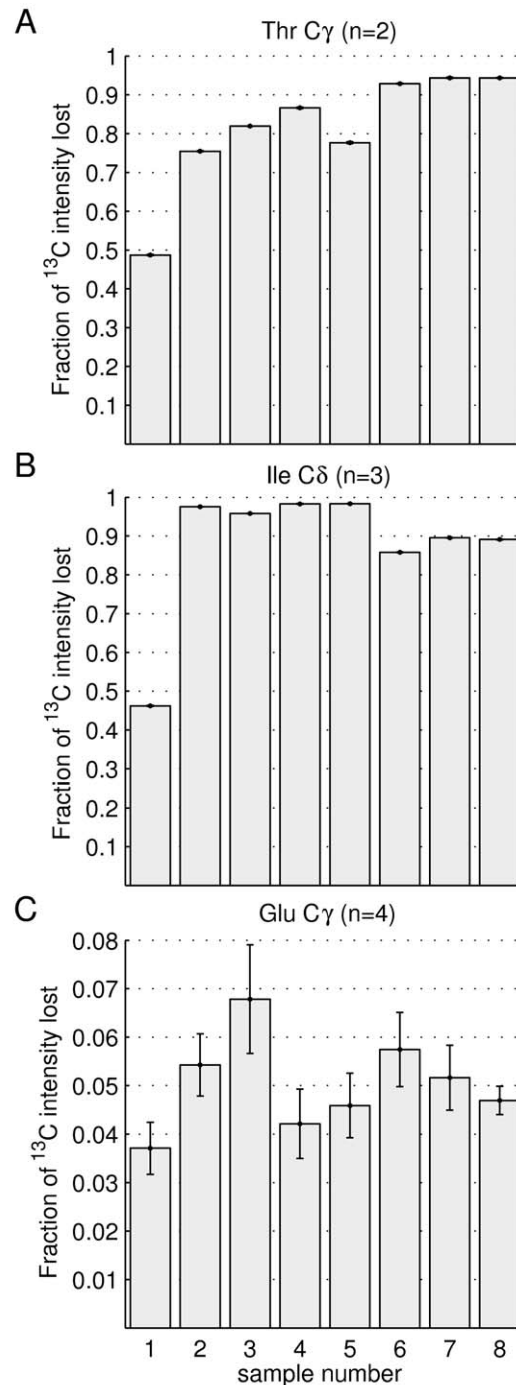


Figure 4. Efficacy of Thr labeling in proteins. Fraction of ^{13}C intensity lost at Thr $\text{C}\gamma$ (A), Ile $\text{C}\delta$ (B) and Glu $\text{C}\gamma$ (C) positions of a U- ^{13}C , ^1H labeled Abp1p SH3 domain sample where ^{12}C labeled precursors such as Thr, 2-ketobutyrate or 2-ketoisovalerate (2KIV) are added in the amounts indicated at the bottom of the figure prior to induction of protein expression. The value of 'n' corresponds to the number of residues of a particular type that are averaged over in each measurement, with the error bars denoting the complete range of values obtained. Details are in Supporting Information S1. doi:10.1371/journal.pone.0043725.g004

50 mg/L labeled 2KB ($^{13}\text{CH}_3\text{CD}_2\text{COCOONa}$) and 100 mg/L d_5 -glycine as an optimal combination for production of highly deuterated proteins labeled with $^{13}\text{CH}_3$ at Thr and Ile ($\delta 1$) methyl positions.

An application to the proteasome core particle

The 20S proteasome core particle (CP) is a 670 kDa barrel-like structure that catalyzes the majority of proteolysis in the cell [58]. It is made up of four axially stacked heptameric rings and in the case of the *T. acidophilum* proteasome that we study each of the rings consists of 7 equivalent subunits of either α or β polypeptides, arranged as $\alpha_7\beta_7\beta_7\alpha_7$. Initial studies by our laboratory focused on the α components of the proteasome for which we had obtained very nearly complete Ile, Leu, Val methyl assignments [24]. More recently we have focused on the β rings as these contain the catalytic residues for proteolysis, including Thr 1 whose nucleophilic hydroxyl moiety attacks a carbonyl carbon at the site of substrate cleavage [36]. Prior to this work we had prepared CP proteasome samples using commercially available U- ^{13}C , ^1H -Thr in a highly deuterated background. However, ^{13}C , ^1H HMQC spectra, recorded with $^{13}\text{C}\beta$ selective decoupling, were of poor quality and only a modest fraction of the expected cross peaks could be observed, Figure 5A. No doubt, the high level of protonation at the Thr β position contributes significantly to the poor quality of the data by increasing both transverse spin relaxation rates of methyl protons, leading to significant line broadening [10], and effective linewidths through the introduction of homonuclear scalar couplings. The added U- ^{13}C , ^1H -Thr results in labeling at the Ile $\delta 1$ position, as expected. The Ile correlations are also poor, reflecting both unresolved $^{13}\text{C}\delta$ - $^{13}\text{C}\gamma$ couplings and a single ^1H at the C γ position that derives from the protonated Thr. While improved Ile $\delta 1$ spectra could be obtained by addition of 2-ketobutyrate, it is not clear whether this precursor could completely turn off production of Ile biosynthesis from Thr.

With the development of a robust and economical biosynthesis scheme for the production of U- ^{13}C , ^1H -Thr- $\gamma 2$ [$^{13}\text{CH}_3$], optimally labeled for NMR studies of high molecular weight complexes, we have now revisited the NMR spectrum of the $^{13}\text{CH}_3$ -Thr labeled CP proteasome. To this end we have prepared a sample of U- ^{13}C , ^1H -Ile- $\delta 1$ [$^{13}\text{CH}_3$], Thr- $\gamma 2$ [$^{13}\text{CH}_3$] $\alpha_7\beta_7\beta_7\alpha_7$, with labeling confined to the β rings. The HMQC data set is illustrated in Figure 5B and it is clear that the quality of the Thr region of the spectrum is much improved relative to when fully protonated precursor is used (compare Figures 5A and B). All 15 of the expected Thr correlations are now observed, although the intensities of a number of the cross peaks are low, indicating significant conformational heterogeneity that results from dynamics that are intrinsic to this allosteric protease.

Assignments for the majority of the Thr methyl groups could be obtained from NOE correlations (Supporting Information S1) connecting Thr peaks with those derived from the previously assigned Ile, Leu, Val methyls. Initially, a 2D ^{13}C , ^1H edited NOESY data set was recorded (mixing time of 250 ms) on the U- ^{13}C , ^1H -Ile- $\delta 1$ [$^{13}\text{CH}_3$], Thr- $\gamma 2$ [$^{13}\text{CH}_3$] $\alpha_7\beta_7\beta_7\alpha_7$ sample; analysis of the spectrum with the assistance of the X-ray structure of the *T. acidophilum* proteasome [59] lead to the assignment of 7 of the Thr methyls. Subsequently, analysis of 3D NOESY data recorded on a U- ^{13}C , ^1H -Leu,Val-[proR- $^{13}\text{CH}_3$, $^{12}\text{CD}_3$], Thr- $\gamma 2$ [$^{13}\text{CH}_3$] $\alpha_7\beta_7\beta_7\alpha_7$ sample confirmed the assignments, lead to the assignment of T81 $\gamma 2$ and T39 $\gamma 2$ and narrowed down the identities of a pair of peaks as belonging to either T106 $\gamma 2$ or T147 $\gamma 2$. Interestingly, assignments of 4 peaks derived from a cluster that includes Thr 1, 3, 16 and 44 and that are weak have yet to be fully confirmed.

In summary, we have presented an enzymatic method for the synthesis of U- ^{13}C , ^1H -Thr- $\gamma 2$ [$^{13}\text{CH}_3$], optimized for NMR studies of high molecular weight proteins, starting from ^{13}C -paraformaldehyde, natural abundance pyruvate and D_2O . Metabolic enzymes from *E. coli* serve as catalysts. In total, 3 weeks are required to recombinantly produce and purify all five enzymes and to synthesize 1–2 grams of Thr. High quality spectra could be recorded on a 0.5 mM (subunit concentration) U- ^{13}C , ^1H -Ile- $\delta 1$ [$^{13}\text{CH}_3$], Thr- $\gamma 2$ [$^{13}\text{CH}_3$] $\alpha_7\beta_7\beta_7\alpha_7$ sample and the quality of the data was far superior than in previous data sets obtained on samples generated with fully protonated Thr precursor. With the availability of U- ^{13}C , ^1H -Thr- $\gamma 2$ [$^{13}\text{CH}_3$] and previously described precursors it is now possible to label the methyl groups of any methyl containing amino acid. There are most certainly advantages for labeling each. The utility of Thr is made clear when one considers that, unlike other methyl containing residues, it has a high propensity for surface exposure [35] and it is more abundant at protein nucleic acid interfaces [34]. Moreover, Thr plays a critical role in the mechanisms of a number of important enzymes [36,37] and in several eukaryotic signaling complexes where biological activity is regulated through phosphorylation [60–62]. The demonstration that high quality NMR spectra can be obtained of Thr methyl groups even for a complex of molecular mass approaching 700 kDa suggests that there will be a substantial number of systems that will benefit from the present labeling scheme.

Materials and Methods

NMR spectra of compounds **2** and **4** are shown in Supporting Information S1 (**6** is shown in Figure 3) along with a detailed description of the biosynthesis of U- ^{13}C , ^1H -Thr- $\gamma 2$ [$^{13}\text{CH}_3$] and the expression and purification of the 5 enzymes (see Supporting Information S1) that are required for Thr production.

Production of the Abp1p SH3 domain and NMR

Samples of U- ^{13}C , ^1H -Abp1p SH3 domain were produced in *E. coli* BL21(DE3) cells, 0.3 L M9 minimal media per sample with ^{13}C -glucose as the sole carbon source, using a protocol described in detail in Supporting Information S1. Samples ranged from 1.65–2 mM in protein, dissolved in 99.9% D_2O , 50 mM $\text{NaH}_2\text{PO}_4\cdot\text{H}_2\text{O}$, 0.1 M NaCl, 1 mM NaN_3 , pH* = 7.0 buffer. ^{12}C labeled precursors such as Thr, 2KB, Gly or 2-ketoisovalerate (2KIV) were added (depending on the sample) as indicated in Figure 4 and Supporting Information S1. The degree of precursor incorporation into protein has been assessed by quantifying cross peak intensities in constant-time ^{13}C , ^1H HSQC data sets [63,64] that were recorded of each sample and compared to a reference sample where precursors were not added. In order to account for slightly different protein concentrations in each sample the peaks of interest were normalized by the relative intensities of Ala methyl correlations. Ala methyl groups were chosen as an internal reference because they derive from pyruvate (glucose) and thus should not be affected by the addition of precursor.

Proteasome production and analysis by NMR

Proteasome samples were prepared as described in detail in Supporting Information S1. A U- ^{13}C , ^1H -Ile- $\delta 1$ [$^{13}\text{CH}_3$], Thr- $\gamma 2$ [$^{13}\text{CH}_3$] $\alpha_7\beta_7\beta_7\alpha_7$ sample (β subunit labeled) was generated by expression of the β subunit with 50 mg/L U- ^{13}C , ^1H -Thr- $\gamma 2$ [$^{13}\text{CH}_3$], 50 mg/L of labeled sodium 2-ketobutyrate (4 - ^{13}C -3,3- d_2) and 100 mg/L of d_5 -glycine, all added 1 hour before induction. A second sample, U- ^{13}C , ^1H -Ile- $\delta 1$ [$^{13}\text{CH}_3$], Leu,Val-[proR- $^{13}\text{CH}_3$, $^{12}\text{CD}_3$], Thr- $\gamma 2$ [$^{13}\text{CH}_3$] $\alpha_7\beta_7\beta_7\alpha_7$, was generated

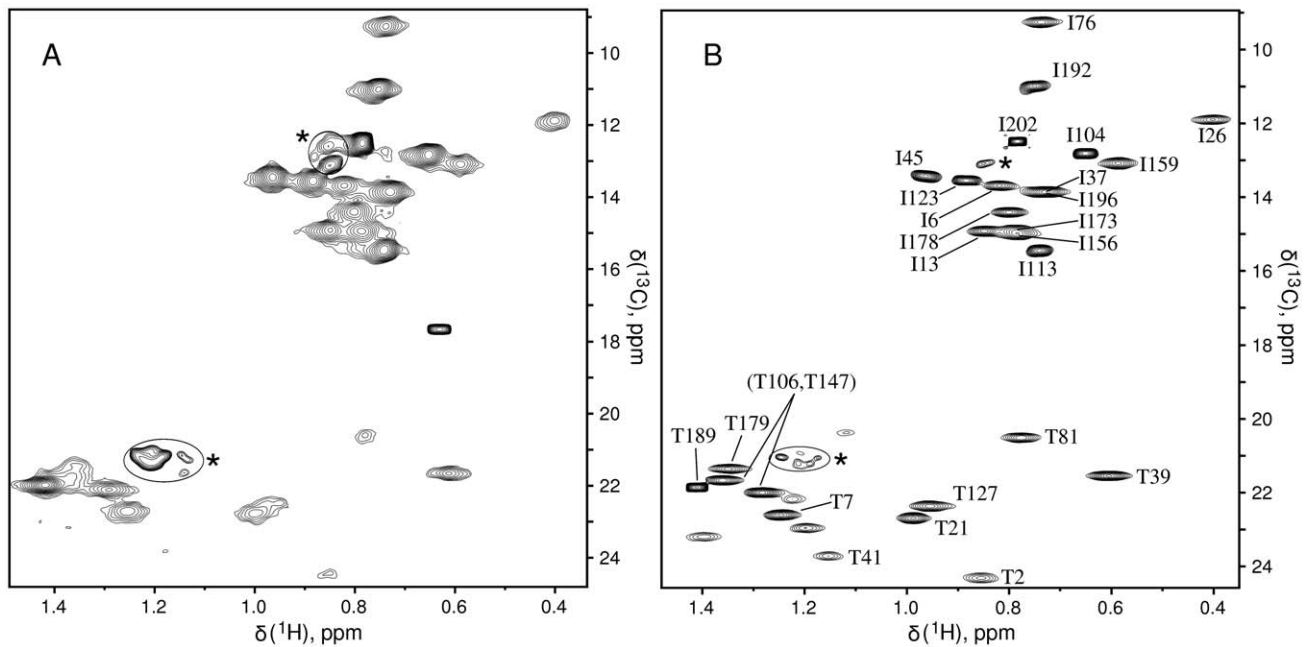


Figure 5. NMR spectroscopy of Thr- γ 2[$^{13}\text{C}_3$]-labeled proteins. (A) $^{13}\text{C},^1\text{H}$ HMQC spectrum of U- $[\text{}^2\text{H}]$ $\alpha_7\beta_7\beta_7\alpha_7$ prepared with U- $[\text{}^{13}\text{C},^1\text{H}]$ -Thr (labeled in the β subunit, 200 μM subunit concentration) recorded in 18 hours, 70°C, 18.8T. $^{13}\text{C}\beta$ Thr band selective adiabatic decoupling [65] was applied to narrow the Thr $^{13}\text{C}\gamma$ resonance lines. (B) $^{13}\text{C},^1\text{H}$ HMQC spectrum of 0.54 mM (subunit concentration) U- $[\text{}^2\text{H}]$, Ile- δ 1[$^{13}\text{CH}_3$], Thr- γ 2[$^{13}\text{CH}_3$] $\alpha_7\beta_7\beta_7\alpha_7$ (labeled in the β subunit) recorded in 2 hours at 70°C, 18.8T. Partial assignments are indicated; stars denote degradation product peaks. The difference in acquisition times is intended to compensate for the concentration differences. It is clear, however, that much higher quality spectra are derived from samples prepared with the current labeling scheme.
doi:10.1371/journal.pone.0043725.g005

with 232 mg/L of an acetolactate based precursor added (purchased from Isotec, cat# 729558) in addition to 2-ketobutyrate. Samples were 0.54 (Ile) and 1.3 (Ile, Leu, Val) mM in protein concentration (subunit concentration), dissolved in 25 mM phosphate, 50 mM NaCl, 99.9% D_2O buffer, pD = 6.8.

All NMR spectra were recorded at a static magnetic field of 18.8T, 70°C on a spectrometer equipped with a room temperature probe. Each $^{13}\text{C},^1\text{H}$ HMQC data set was obtained in 2 hours. A 2D $^{13}\text{C},^1\text{H}$ NOESY-HMQC data set was measured on the U- $[\text{}^2\text{H}]$, Ile- δ 1[$^{13}\text{CH}_3$], Thr- γ 2[$^{13}\text{CH}_3$] $\alpha_7\beta_7\beta_7\alpha_7$ sample with a 16.5 ppm spectral window in F_1 , 80 complex data points in the indirect dimension, 250 ms NOE mixing time and 1.5 s repetition delay for a total of 14.5 hours of accumulation. A 3D $^{13}\text{C},^{13}\text{C}$ -separated NOESY experiment was recorded on the U- $[\text{}^2\text{H}]$, Ile- δ 1[$^{13}\text{CH}_3$], Leu,Val-[proR- $^{13}\text{CH}_3$, $^{12}\text{CD}_3$], Thr- γ 2[$^{13}\text{CH}_3$] $\alpha_7\beta_7\beta_7\alpha_7$ sample, 18.5 ppm spectral window in both carbon dimensions, 8 scans per increment, 72 (t_1) and 64 (t_2) complex points in the two indirect dimensions, 200 ms NOE mixing time and 1.5 s interscan delay (75 hour experiment time).

References

- Bax A, Grzesiek S (1993) Methodological advances in protein NMR. *Acc Chem Res* 26: 131–138.
- Clore GM, Gronenborn AM (1991) Structures of larger proteins in solution: three- and four-dimensional heteronuclear NMR spectroscopy. *Science* 252: 1390–1399.
- Grzesiek S, Anglister J, Ren H, Bax A (1993) Carbon-13 line narrowing by deuterium decoupling in deuterium/carbon-13/nitrogen-15 enriched proteins. Application to triple resonance 4D J connectivity of sequential amides. *J Am Chem Soc* 115: 4369–4370.
- Kainosho M, Torizawa T, Iwashita Y, Terauchi T, Mei Ono A, et al. (2006) Optimal isotope labelling for NMR protein structure determinations. *Nature* 440: 52–57.
- Goto NK, Kay LE (2000) New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Curr Opin Struct Biol* 10: 585–592.
- LeMaster DM, Kushlan DM (1996) Dynamical Mapping of E. coli Thioredoxin via ^{13}C NMR Relaxation Analysis. *J Am Chem Soc* 118: 9255–9264.
- Fiaux J, Bertelsen EB, Horwich AL, Wüthrich K (2002) NMR analysis of a 900K GroEL GroES complex. *Nature* 418: 207–211.
- Tugarinov V, Hwang PM, Kay LE (2004) Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins. *Annu Rev Biochem* 73: 107–146.
- Pervushin K, Riek R, Wider G, Wüthrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A* 94: 12366–12371.
- Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE (2003) Cross-correlated relaxation enhanced ^1H - ^{13}C NMR spectroscopy of methyl groups in very high

- molecular weight proteins and protein complexes. *J Am Chem Soc* 125: 10420–10428.
11. Gardner KH, Kay LE (1997) Production and Incorporation of ¹⁵N, ¹³C, ²H (1H- δ 1 Methyl) Isoleucine into Proteins for Multidimensional NMR Studies. *J Am Chem Soc* 119: 7599–7600.
 12. Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE (1999) A robust and cost-effective method for the production of Val, Leu, Ile (δ 1) methyl-protonated ¹⁵N-, ¹³C-, ²H-labeled proteins. *J Biomol NMR* 13: 369–374.
 13. Tugarinov V, Kay LE (2004) An isotope labeling strategy for methyl TROSY spectroscopy. *J Biomol NMR* 28: 165–172.
 14. Ollerenshaw JE, Tugarinov V, Kay LE (2003) Methyl TROSY: explanation and experimental verification. *Magn Reson Chem* 41: 843–852.
 15. Hamel DJ, Dahlquist FW (2005) The contact interface of a 120 kD CheA-CheW complex by methyl TROSY interaction spectroscopy. *J Am Chem Soc* 127: 9676–9677.
 16. Gelis I, Bonvin AMJJ, Keramisanou D, Koukaki M, Gouridis G, et al. (2007) Structural basis for signal-sequence recognition by the translocase motor SecE as determined by NMR. *Cell* 131: 756–769.
 17. Kreishman-Deitrick M, Goley ED, Burdine L, Denison C, Egile C, et al. (2005) NMR analyses of the activation of the Arp2/3 complex by neuronal Wiskott-Aldrich syndrome protein. *Biochemistry* 44: 15247–15256.
 18. Kato H, van Ingen H, Zhou B-R, Feng H, Bustin M, et al. (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: 12283–12288.
 19. Baldwin AJ, Hilton GR, Lioe H, Bagn ers C, Benesch JLP, et al. (2011) Quaternary dynamics of α -crystallin as a direct consequence of localised tertiary fluctuations in the C-terminus. *J Mol Biol* 413: 310–320.
 20. Velyvis A, Yang YR, Schachman HK, Kay LE (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: 8815–8820.
 21. Isaacson RL, Simpson PJ, Liu M, Cota E, Zhang X, et al. (2007) A new labeling method for methyl transverse relaxation-optimized spectroscopy NMR spectra of alanine residues. *J Am Chem Soc* 129: 15428–15429.
 22. Sprangers R, Gribun A, Hwang PM, Houry WA, Kay LE (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: 16678–16683.
 23. Amero C, Schanda P, Dur  MA, Ayala I, Marion D, et al. (2009) Fast two-dimensional NMR spectroscopy of high molecular weight protein assemblies. *J Am Chem Soc* 131: 3448–3449.
 24. Sprangers R, Kay LE (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. *Nature* 445: 618–622.
 25. Ruschak AM, Religa TL, Breuer S, Witt S, Kay LE (2010) The proteasome antechamber maintains substrates in an unfolded state. *Nature* 467: 868–871.
 26. Religa TL, Sprangers R, Kay LE (2010) Dynamic regulation of archaeal proteasome gate opening as studied by TROSY NMR. *Science* 328: 98–102.
 27. Karag z GE, Duarte AMS, Ippel H, Uetrecht C, Sinnige T, et al. (2011) N-terminal domain of human Hsp90 triggers binding to the cochaperone p23. *Proc Natl Acad Sci U S A* 108: 580–585.
 28. Ayala I, Sounier R, Us  N, Gans P, Boissbouvier J (2009) An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. *J Biomol NMR* 43: 111–119.
 29. Stoffregen MC, Schwer MM, Renschler FA, Wiesner S (2012) Methionine Scanning as an NMR Tool for Detecting and Analyzing Biomolecular Interaction Surfaces. *Structure* 20: 573–581.
 30. Fischer M, Kloiber K, H usler J, Ledolter K, Konrat R, 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: 610–612.
 31. Gans P, Hamelin O, Sounier R, Ayala I, Dur  MA, et al. (2010) Stereospecific isotopic labeling of methyl groups for NMR spectroscopic studies of high-molecular-weight proteins. *Angew Chem Int Ed Engl* 49: 1958–1962.
 32. Religa TL, Ruschak AM, Rosenzweig R, Kay LE (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: 9063–9068.
 33. Gerstein M (1997) A structural census of genomes: comparing bacterial, eukaryotic, and archaeal genomes in terms of protein structure. *J Mol Biol* 274: 562–576.
 34. Sinha K, Jen-Jacobson L, Rule GS (2011) Specific labeling of threonine methyl groups for NMR studies of protein-nucleic acid complexes. *Biochemistry* 50: 10189–10191.
 35. Miller S, Janin J, Lesk AM, Chothia C (1987) Interior and surface of monomeric proteins. *J Mol Biol* 196: 641–656.
 36. Seem ller E, Lupas A, Stock D, L we J, Huber R, et al. (1995) Proteasome from *Thermoplasma acidophilum*: a threonine protease. *Science* 268: 579–582.
 37. Goldberg AL (2003) Protein degradation and protection against misfolded or damaged proteins. *Nature* 426: 895–899.
 38. Takeda M, Jee J, Ono AM, Terauchi T, Kainosho M (2011) Hydrogen exchange study on the hydroxyl groups of serine and threonine residues in proteins and structure refinement using NOE restraints with polar side-chain groups. *J Am Chem Soc* 133: 17420–17427.
 39. Ruschak AM, Velyvis A, Kay LE (2010) A simple strategy for ¹³C, ¹H labeling at the Ile- γ 2 methyl position in highly deuterated proteins. *J Biomol NMR* 48: 129–135.
 40. Ollerenshaw JE, Tugarinov V, Skrynnikov NR, Kay LE (2005) Comparison of ¹³CH₃, ¹³CH₂D, and ¹³CHD₂ methyl labeling strategies in proteins. *J Biomol NMR* 33: 25–41.
 41. Religa TL, Kay LE (2010) Optimal methyl labeling for studies of supramolecular systems. *J Biomol NMR* 47: 163–169.
 42. Steinreiber J, Fesko K, Reisinger C, Sch rmmann M, van Assema F, et al. (2007) Threonine aldolases - an emerging tool for organic synthesis. *Tetrahedron* 63: 918–926.
 43. Fesko K, Reisinger C, Steinreiber J, Weber H, Sch rmmann M, et al. (2008) Four types of threonine aldolases: Similarities and differences in kinetics/thermodynamics. *J Mol Catal B Enzym* 52–53: 19–26.
 44. Marcus JP, Dekker EE (1993) Threonine formation via the coupled activity of 2-amino-3-ketobutyrate coenzyme A lyase and threonine dehydrogenase. *J Bacteriol* 175: 6505–6511.
 45. Laver WG, Neuberger A, Scott JJ (1959) 287. α -Amino- β -keto-acids. Part II. Rates of decarboxylation of the free acids and the behaviour of derivatives on titration. *J Chem Soc* 1483.
 46. Marcus JP, Dekker EE (1993) pH-dependent decarboxylation of 2-amino-3-ketobutyrate, the unstable intermediate in the threonine dehydrogenase-initiated pathway for threonine utilization. *Biochem Biophys Res Commun* 190: 1066–1072.
 47. Marcus JP, Dekker EE (1993) Identity and some properties of the L-threonine aldolase activity manifested by pure 2-amino-3-ketobutyrate ligase of *Escherichia coli*. *Biochim Biophys Acta* 1164: 299–304.
 48. Zhang K, Li H, Cho KM, Liao JC (2010) Expanding metabolism for total biosynthesis of the nonnatural amino acid L-homoalanine. *Proc Natl Acad Sci U S A* 107: 6234–6239.
 49. Kim D, Ortiz de Montellano PR (2009) Tricistronic overexpression of cytochrome P450cam, putidaredoxin, and putidaredoxin reductase provides a useful cell-based catalytic system. *Biotechnol Lett* 31: 1427–1431.
 50. Haines DC, Tomchick DR, Machius M, Peterson J (2001) Pivotal role of water in the mechanism of P450BM-3. *Biochemistry* 40: 13456–13465.
 51. Ost TW, Miles CS, Murdoch J, Cheung Y, Reid G, et al. (2000) Rational redesign of the substrate binding site of flavocytochrome P450 BM3. *FEBS Lett* 486: 173–177.
 52. Lane RS, Shapley A, Dekker EE (1971) 2-keto-4-hydroxybutyrate aldolase. Identification as 2-keto-4-hydroxyglutarate aldolase, catalytic properties, and role in the mammalian metabolism of L-homoserine. *Biochemistry* 10: 1353–1364.
 53. Patil RV, Dekker EE (1992) Cloning, nucleotide sequence, overexpression, and inactivation of the *Escherichia coli* 2-keto-4-hydroxyglutarate aldolase gene. *J Bacteriol* 174: 102–107.
 54. Clap s P, Fessner W-D, Sprenger GA, Samland AK (2010) Recent progress in stereoselective synthesis with aldolases. *Curr Opin Chem Biol* 14: 154–167.
 55. Huo X, Viola RE (1996) Substrate specificity and identification of functional groups of homoserine kinase from *Escherichia coli*. *Biochemistry* 35: 16180–16185.
 56. Laber B, Gerbling KP, Harde C, Neff KH, Nordhoff E, et al. (1994) Mechanisms of interaction of *Escherichia coli* threonine synthase with substrates and inhibitors. *Biochemistry* 33: 3413–3423.
 57. McCourt JA, Duggleby RG (2006) Acetoacetyl synthase and its role in the biosynthetic pathway for branched-chain amino acids. *Amino Acids* 31: 173–210.
 58. L we J, Stock D, Jap B, Zwickl P, Baumeister W, et al. (1995) Crystal structure of the 20S proteasome from the archaeon *T. acidophilum* at 3.4 Å resolution. *Science* 268: 533–539.
 59. F rster A, Masters EI, Whitby FG, Robinson H, Hill CP (2005) The 1.9 Å structure of a proteasome-11S activator complex and implications for proteasome-PAN/PA700 interactions. *Mol Cell* 18: 589–599.
 60. Romeo Y, Zhang X, Roux PP (2012) Regulation and function of the RSK family of protein kinases. *Biochem J* 441: 553–569.
 61. Freeman AK, Morrison DK (2011) 14-3-3 Proteins: diverse functions in cell proliferation and cancer progression. *Semin Cell Dev Biol* 22: 681–687.
 62. Gardino AK, Yaffe MB (2011) 14-3-3 proteins as signaling integration points for cell cycle control and apoptosis. *Semin Cell Dev Biol* 22: 688–695.
 63. Santoro J, King GC (1992) A constant-time 2D overbroadening experiment for inverse correlation of isotopically enriched species. *J Magn Reson* 97: 202–207.
 64. Vuister GW, Bax A (1992) Resolution enhancement and spectral editing of uniformly ¹³C-enriched proteins by homonuclear broadband ¹³C decoupling. *J Magn Reson* 98: 428–435.
 65. Kupce E, Wagner G (1996) Multisite Band-Selective Decoupling in Proteins. *J Magn Reson B* 110: 309–312.