

Stereo-selective deuteration in aspartate, asparagine, lysine, and methionine amino acid residues using fumarate as a carbon source for *E. coli* in D₂O

Gaddafi I. Danmaliki[§], Philip B. Liu[‡], and Peter M. Hwang^{*§}

[§]Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada T6G 2H7

[‡]Department of Medicine, University of Alberta, Edmonton, Alberta, Canada T6G 2R3

ABSTRACT: We propose an isotope labeling strategy for expressing proteins in *E. coli* using protonated natural abundance ¹³C carbon sources in D₂O. The strategy eliminates dominant ¹H-¹³C and ¹H-¹H dipolar relaxation mechanisms to produce long-lived magnetization for solution NMR spectroscopy. Isolated ¹H magnetization can be transferred via through-space NOEs to slowly relaxing ¹H-¹⁵N or ¹H-¹³C methyl groups, providing the resolution needed for accurate structure determination of large protein systems by solution NMR.

To fully characterize ¹H/²H-isotope incorporation by NMR, we applied our labelling strategy to a model system, the WW domain from human Pin1 protein. Using glucose as sole carbon source in D₂O, a very high level of protonation was observed in the aromatic side chains, as well as at the H^{β2} and H^{β3} positions of serine and tryptophan, consistent with previous studies. With our specialized FROMP media (fumarate, rhamnose, oxalate, malonate, pyruvate), more

aromatic ring positions were deuterated, and stereo-selective protonation of $H^{\beta 2}$ with deuteration at H^{α} and $H^{\beta 3}$ was achieved in Asp, Asn, Lys, and Met residues.

In solution NMR, stereospecific chemical shift assignments for H^{β} are typically obtained in conjunction with χ_i dihedral angle determinations using 3-bond J-coupling (${}^3J_{N-H\beta}$, ${}^3J_{CO-H\beta}$, ${}^3J_{H\alpha-H\beta}$) experiments. However, due to motional averaging, the assumption of a pure rotameric state can yield incorrect χ_i dihedral angles with incorrect stereospecific assignments. In three residues, Lys28, Met30, and Asn44, the use of stereo-selective isotope labeling reversed the stereospecific chemical shift assignments and χ_i dihedral angles obtained in the traditional manner, demonstrating how stereo-selective isotope labeling can improve the accuracy of χ_i dihedral angles obtained by solution NMR.

INTRODUCTION

Nuclear magnetic resonance (NMR) spectroscopy is a powerful technique for the elucidation of protein structure and dynamics at atomic resolution, contributing about 10% of all the structures deposited in the Protein Data Bank.¹ The development of multinuclear multidimensional solution NMR combined with uniform isotopic enrichment (¹³C, ¹⁵N) supports the determination of protein structures up to 25 kDa.^{2,3} However, for systems in excess of 25 kDa, NMR studies are hampered by rapid signal decay, determined by the transverse relaxation rate, which scales roughly with molecular weight. One strategy for minimizing the transverse relaxation rate is to replace the ¹H nuclei in a protein with ²H to minimize ¹H-¹H and ¹H-¹³C dipolar relaxation.⁴ However, this approach reduces the number of available NOE-based ¹H-¹H distance restraints required for high resolution structure determination.⁵ Reincorporating ¹H at specific sites in a highly deuterated background is therefore necessary. Several approaches have been developed over the past decades to achieve this goal.

Random fractional deuteration has been utilized by several groups,^{4,9} and depending on the size of the protein, a deuteration level of 50-90% provides a good compromise between spectral quality and available distance information necessary for high resolution structure determination. The disadvantage of this technique is that it produces numerous isotopomers that disperse signals into multiple peaks due to deuterium isotope shift, compromising sensitivity and resolution.^{4,9}

Methyl groups have proven to be ideal molecular probes for solution NMR spectroscopy studies of large proteins. Protocols for selective protonation of methyl groups of Ala, Thr, Ile, Leu, Val, and Met in perdeuterated background have been developed.¹⁰⁻¹⁷ Rapid rotation about the methyl symmetry axis attenuates ¹H-¹³C dipolar relaxation. An inability to obtain structural information for

all non-methyl-containing amino acid side chains is the major limitation of this approach. For instance, the side chains of the aromatic amino acids are very important for hydrophobic packing in proteins and ligand binding interfaces. ^{13}C reverse labeling of the aromatic side chains of phenylalanine and tyrosine simplifies spectral crowding and overlap, improving sensitivity while providing long range NOEs.^{18,19} However, one drawback of this approach is that strong ^1H - ^1H dipolar relaxation still occurs between vicinal protons in the aromatic rings.

The stereo-array isotope labeling (SAIL) technique developed by Kainosho and co-workers is a creative approach that provides a unique solution to rapid transverse relaxation in NMR.²⁰ In this technique, all twenty amino acids are chemically and enzymatically synthesized and then incorporated into protein with a cell-free expression system. The amino acids are synthesized with a stereospecific and regiospecific pattern fully ^{13}C labelled, with only one ^1H attached to every ^{13}C atom and all other sites completely deuterated. The technique greatly enhances spectral sensitivity and resolution, but the major drawbacks are that it is extremely labor-intensive and costly, and fast dipolar relaxation still dominates any non-methyl ^1H - ^{13}C group even in a highly deuterated background.

Despite advances to judiciously incorporate ^1H into proteins in an otherwise highly deuterated background, new methods are needed to obtain structural information from all 20 amino acid residues in a protein. Here, we present a protocol for producing proteins in *Escherichia coli* (*E. coli*) using inexpensive non-isotopically enriched carbon sources, taking advantage of the inherent amino acid biosynthetic pathways in *E. coli* to stereo-specifically incorporate ^1H into proteins produced in D_2O -based media. This approach eliminates the dominant sources of dipolar relaxation (^1H - ^1H and ^1H - ^{13}C dipoles), producing isolated ^1H - ^{13}C groups in a largely deuterated background. This provides additional ^1H magnetization that can be transferred via through-space

NOEs to nearby ^1H - ^{15}N and ^1H - ^{13}C methyl groups, which can then be resolved via multinuclear NMR to provide the necessary structural restraints for high resolution structure determination of large protein systems.

EXPERIMENTAL PROCEDURES

Construct design

We have applied our labeling technique to the WW domain from the human Pin1 protein, the smallest folded protein domain we could identify containing all 20 amino acids except cysteine (about 30 residues). It has been characterized by extensive biophysical studies and mutagenesis. The rate limiting step for the folding of this protein is the formation of a “loop 1” structure (between the first two β -strands) consisting of an unusual four-residue type-II turn inserted within a larger six residue loop that is implicated in protein-protein interactions.^{21,22} Replacing the wild-type loop 1 with a shorter sequence has been shown to improve both folding and thermal stability by an order of magnitude.²¹ The wild-type protein has a low yield when expressed in *E. coli*, typically 2 mg/L, explaining why most studies utilized peptide synthesis over recombinant expression.^{21,23,24} Due to these reasons, we designed, optimized, and expressed a mutant human Pin1 WW domain construct in *E. coli* possessing a shortened loop 1,²¹ and a His-tag connected by a long N-terminal linker (Figure 1). An expression vector for this was synthesized by ATUM (formerly DNA2.0), with a high copy number origin of replication, ampicillin selection marker, strong ribosome binding site, lacI repressor gene, and a T5 promoter controlled by two flanking lacO sites that allow induction by the addition of IPTG (isopropyl β -D-1-thiogalactopyranoside) in *E. coli*. The T5 promoter is recognized by *E. coli* RNA polymerase, so the construct can be expressed in

any strain. Codons were optimized by ATUM for expression in *E. coli*, and the 53-amino acid construct yielded ~75 mg/L when expressed in 1 L minimal M9 media (10 g glucose).

10 20 30 40 50
MGHHHHHSS GGSSGMADEE KLPPG**WEKRM** S-ADGRVYYFN HITNASQWER PSG

Figure 1. Mutant human Pin1 WW domain sequence. Red lettering denotes modifications made to the native amino acid sequence: an N-terminal His-tag followed by a flexible linker, as well as an Ala-Asp substitution/deletion for Ser-Gly-Arg in the middle of the protein in loop 1.²¹ The three anti-parallel β -strands are shown in bold.

Production of (¹⁵N, ¹³C)-mutant human Pin1 WW domain in H₂O minimal M9 media

To facilitate complete stereospecific resonance assignments, fully protonated uniformly d (¹⁵N, ¹³C)-labeled mutant human Pin1 WW domain was expressed in *E. coli* BL21(DE3) with slight modifications to methods described earlier.^{3,25,26} The protein was expressed in 1 L M9 minimal media, containing 9 g Na₂HPO₄ and 2.5 g KH₂PO₄ dissolved in 950 mL H₂O, yielding a pH 7.3-7.4. A 50 mL solution containing 1 g ¹⁵NH₄SO₄, 3g ¹³C-glucose, 1 mL 1 M MgSO₄, 1 mL 0.1M CaCl₂, 1 mL 5 % ampicillin solution, 1 mg biotin, and 100 mg thiamine was filtered and added to the autoclaved 950 mL M9 salts.

Four to six transformed colonies were inoculated into 10 mL LB media containing 100 mg/L ampicillin and incubated at 37 °C. After reaching A₆₀₀ ~1.0, the cells were diluted into 1 L minimal M9 media and allowed to grow further until A₆₀₀ ~0.8. The cells were induced with 1 mM IPTG and allowed to grow for six hours post-induction. The cells were centrifuged at 5000 rpm (4420 x g) for 20 min and harvested. The cell pellet was re-suspended in 20 mL lysis buffer: 50 mM Tris (pH 8.0), 10 mM MgSO₄, 10 ug/mL DNase I. 200 mg sodium deoxycholate and 20 mg lysozyme, each pre-dissolved in 1 mL distilled water, was added to the lysis buffer to disrupt cell membranes and cell walls. The lysate was centrifuged at 15000 rpm (27,000 x g) for 15 minutes,

and the supernatant was syringe-filtered with a 0.45 μm cut-off filter. The supernatant was applied to a Qiagen Ni-NTA column equilibrated with binding buffer (20mM Tris-HCl, 300mM NaCl, 10 mM imidazole), and then washed with the same buffer containing 80 mM imidazole. Protein was eluted with the same buffer containing 250 mM imidazole. Fractions were assessed by SDS-PAGE, and fractions containing pure protein were dialyzed against 6.5 mM ammonium bicarbonate for three days and then lyophilized. Mass spectrometry was used to confirm the identity of the protein and showed that the initial methionine in the sequence had been removed. To facilitate the stereospecific assignments of methyl groups, a mutant human Pin1 WW domain sample was produced using 10% uniformly (^1H , ^{13}C)-enriched glucose and 90% unenriched glucose as described previously.²⁷

Production of (^{15}N , ^{13}C)-mutant human Pin1 WW domain in D_2O -M9 minimal media

To assess the specific $^1\text{H}/^2\text{H}$ incorporation into individual amino acids, mutant human-Pin1 WW domain was expressed in D_2O , using protonated natural abundance ^{13}C -glucose rather than ^{13}C -glucose as described in the previous section. The expression and purification process was basically the same as the previous section, except that cells were diluted into D_2O M9 media prior to induction. The bacterial cells were grown in D_2O media for at least one doubling time, which was three hours. Cells were induced at $A_{600\text{nm}} \sim 0.8$ using 1 mM IPTG and allowed to grow for six hours post-induction. The purification was the same as the previous section, yielding isotopically enriched protein that we will denote as Pin1(glucose). The level of $^1\text{H}/^2\text{H}$ incorporation in every amino acid was determined by comparing the peak intensities from a ^1H - ^{13}C HSQC spectrum of Pin1(glucose) with that of a fully protonated sample, Pin1(unlabeled). The HSQC was performed with long ^{13}C indirect acquisition times (20 ms) without constant time, which was made possible because only natural abundance (1%) ^{13}C isotope was used.

Rationale for isotope labeling using D₂O-“FROMP” minimal media

The major concern in producing perdeuterated proteins is the phenomenon of isotopic “scrambling”, whereby metabolic precursors are processed variably through multiple pathways before they are incorporated into amino acids, resulting in multiple isotopomers, like CH₂, CHD, and CD₂.^{11,28} The presence of multiple isotopomers spreads signals into multiple peaks, compromising both sensitivity and resolution. One potential cause of scrambling is the interconversion between phosphoenolpyruvate, pyruvate, and oxaloacetate at the juncture of the key metabolic pathways: glycolysis, tricarboxylic acid (TCA) cycle, and gluconeogenesis (Figure 2).²⁹ Oxalate has been shown to inhibit many of the enzymes at this key metabolic juncture including the gluconeogenic enzymes, phosphoenolpyruvate carboxykinase (PEPCK), malic enzyme, and phosphoenolpyruvate synthetase (Figure 2).³⁰⁻³² We reasoned that oxalate could thus be used to limit scrambling, so long as precursors are supplied on both sides of the blockade: pyruvate on the side of the TCA cycle and rhamnose on the glycolysis/gluconeogenesis side (Figure 2). Rhamnose was chosen instead of glucose because it is metabolized into L-lactaldehyde (which can be converted to pyruvate) and dihydroxyacetone phosphate. Gluconeogenic synthesis of glucose from dihydroxyacetone phosphate and subsequent metabolism to erythrose-4-phosphate via the pentose phosphate pathway would result in a higher degree of deuteration at the aromatic amino acids than if glucose were supplied in the media. (see Figure 2).

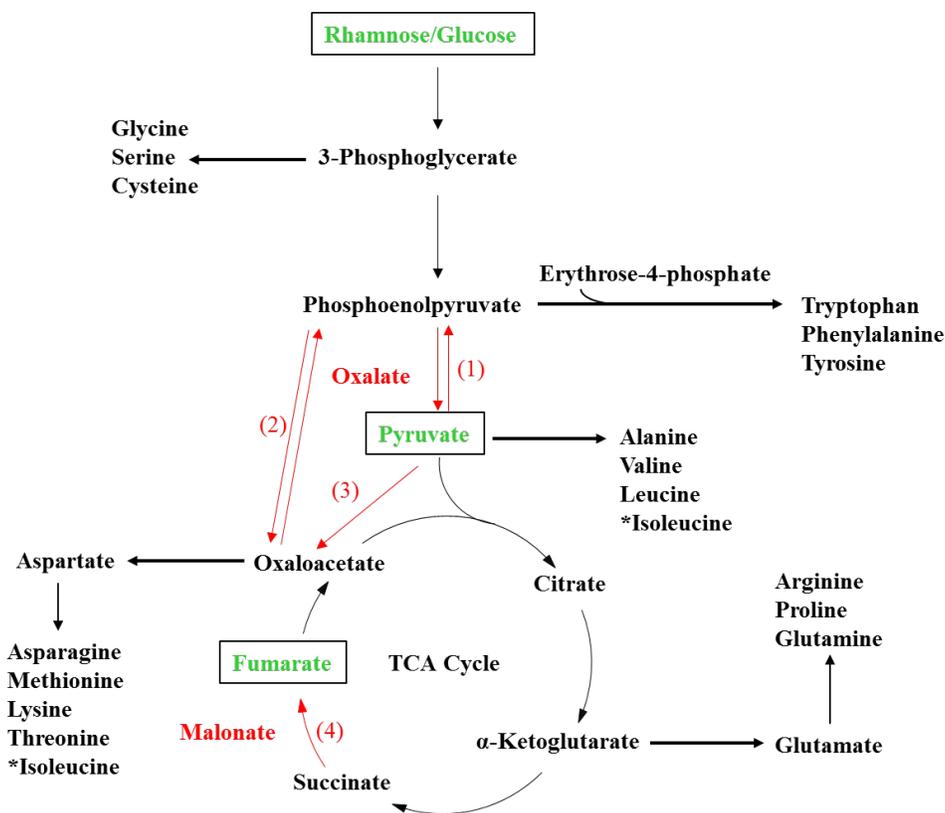


Figure 2. Overview of amino acid biosynthetic pathways. Carbon sources added to the growth media: rhamnose/glucose, pyruvate, and fumarate (green). Enzyme inhibitors: oxalate and malonate (red). Oxalate inhibits multiple enzymes involved in the interconversion of phosphoenolpyruvate, pyruvate, and oxaloacetate/malate: (1) phosphoenolpyruvate synthetase and pyruvate kinase, (2) phosphoenolpyruvate carboxykinase, and (3) pyruvate carboxylase. Malonate inhibits succinate dehydrogenase (4), preventing oxidation of succinate to fumarate.

We also postulated that the amino acids belonging to the oxaloacetate family in the TCA cycle (aspartate, asparagine, lysine, threonine, and methionine) could be stereospecifically labeled with ^1H at the $\text{H}^{\text{O}2}$ position and ^3H at the $\text{H}^{\text{O}3}$ position by supplying fumarate as an additional carbon source. Fumarate is converted into malate in a reaction catalyzed by fumarase, with one ^3H stereospecifically incorporated from D_2O (Figure 3).³³ Malate is then metabolized within the TCA cycle to oxaloacetate, which is converted into aspartate by transamination. Endogenous fumarate production by the TCA cycle can be suppressed by the addition of malonate to the growth media (Figure 2).^{34,35}

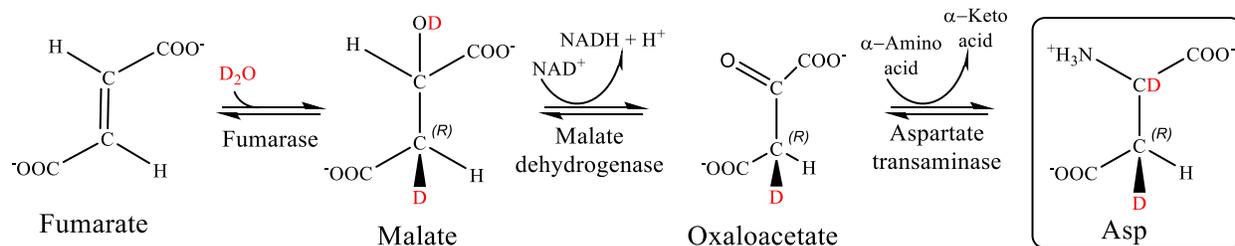


Figure 3. Stereo-selective incorporation of ^3H from D_2O into the H^{β} position in aspartate. Fumarate supplied in the growth media is converted into malate in a reaction catalyzed by fumarase, with one ^3H stereospecifically incorporated from D_2O . Malate dehydrogenase catalyzes the oxidation of malate to oxaloacetate which is converted to aspartate via a transamination reaction.

Production of mutant human Pin1 WW domain in D_2O -“FROMP” (fumarate, rhamnose, oxalate, malonate, and pyruvate) media

Cells were grown initially in 10 mL LB till an $A_{600\text{nm}} \sim 1$ was reached. The cells were diluted in 1 L M9/ H_2O medium containing 1g NH_4Cl and 3g ^{13}C -rhamnose. Doubling time was two hours, and the cells were allowed to grow till $A_{600\text{nm}} \sim 0.7$. The cells were then harvested and re-suspended in 0.9 L M9/ D_2O , containing 8 g Na_2HPO_4 , 2.2 g K_2HPO_4 , 1g NH_4Cl and 3g ^{13}C -rhamnose, 0.11 g MgSO_4 (anhydrous), 0.01 g CaCl_2 (anhydrous), 90 mg ampicillin, 100 mg thiamine, and 1 mg biotin that was sterile filtered. The bacterial cells were grown for one hour prior to the addition of 3 g sodium pyruvate, 3 g sodium fumarate, 1 g oxalic acid and 1 g malonic acid, induced at $A_{600\text{nm}} \sim 0.9$, and allowed to grow for six hours post-induction. Protein purification was the same as in the previous sections, yielding ~ 10 mg per liter of growth media. We will denote the NMR sample from this preparation as Pin1(FROMP). Peak intensities obtained from Pin1(FROMP) were compared to Pin1(unlabeled) and Pin1(glucose) using ^1H - ^{13}C HSQC spectra.

NMR Spectroscopy

All NMR samples were 500 μ L in volume. The buffer conditions were: 100 mM KCl, 10 mM imidazole, 0.5 mM 2,2-dimethyl-2-silapentane-5-sulfonate-d6 sodium salt (DSS-d6) as an NMR chemical shift internal reference, and 0.01 % NaN₃ in 90% H₂O, 10% D₂O or 100 % D₂O, pH 6.8. Concentrations ranged from 0.5 to 2 mM protein.

All NMR experiments were conducted at 30 °C on a Varian Inova 500 MHz spectrometer. The spectrometer was equipped with triple resonance probes and Z-pulsed field gradients. All one-dimensional experiments were processed using VNMRJ (Varian Associates) and all two-dimensional and three-dimensional NMR experiments were processed using NMRPipe/NMRDraw software.³⁶ The 2D and 3D spectra were analyzed further using NMRviewJ (One Moon Scientific).³⁷

Backbone ¹H, ¹⁵N, ¹³C chemical shift assignments for the mutant human Pin1 WW domain were obtained by analyzing the 3D HNCACB and 3D CBCA(CO)NH experiments. Sidechain ¹H and ¹³C chemical shift assignments were obtained by analyzing the (H)C(CO)NH-TOCSY and H(C)(CO)NH-TOCSY experiments. Aromatic side-chain resonances were assigned using an aromatic 3D ¹³C-edited NOESY-HSQC (mixing time 100 ms). The 2D constant time ¹H,¹³C-HSQC experiment on an NMR sample prepared with 10% ¹³C-glucose labeling was used to obtain stereospecific assignments of Leu and Val methyl groups in the protein (Pro-S methyl groups are in phase with methionine methyls, which are free from 1-bond ¹³C-¹³C J couplings). The chi-1 dihedral angle and the stereospecific assignment for H ^{β} protons were obtained by analyzing the following experiments: 3D HNHB, 3D HN(CO)HB, and ¹H-TOCSY-¹⁵N-HSQC (28 ms mixing time). For these experiments, the intensity of the correlations between HN and H ^{$\beta 2$} /H ^{$\beta 3$} depends on

$^3J_{N-H\beta}$, $^3J_{CO-H\beta}$, $^3J_{H\alpha-H\beta}$ respectively, giving an internally redundant dataset for determining if $H^{\beta2}/H^{\beta3}$ forms 60° (small 3J) or 180° (large 3J) dihedral angles with N, CO, and $H\alpha$, respectively, as determined by the three major χ_1 rotameric states: trans, gauche, and gauche'.

RESULTS AND DISCUSSION

Chemical shift assignment of Pin1 WW domain

The 2D 1H - ^{15}N HSQC NMR spectrum of fully protonated Pin1 WW domain showed well dispersed amide proton signals, characteristic of a folded protein domain. A near-complete backbone and sidechain assignment was achieved for the structured region of the protein including all aromatic sidechains, providing probes of isotope labeling patterns in 19 out of the 20 naturally occurring amino acids (all except cysteine).

Quantitation of $^1H/^2H$ incorporation in Pin1 WW domain grown in D_2O - ^{13}C -glucose or D_2O -FROMP media

The 2D 1H - ^{13}C HSQC spectrum of Pin1(unlabeled), shown in Figure 4A, serves as a reference against which deuterated samples can be compared, facilitating the testing of various unlabeled carbon sources for biosynthetic incorporation in *E. coli*. All HSQC spectra relied on natural abundance ^{13}C , allowing high resolution of the indirect ^{13}C dimension without the use of constant time evolution. Peak intensities from Pin1(unlabeled) were compared to the peak intensities of Pin1(glucose) (Figure 4B) or Pin1(FROMP) (Figure 4C) to determine the level of $^1H/^2H$ incorporation into every amino acid site, normalized by a scaling factor. The scaling factor corrects for differences in concentration between the samples, and was very close to 1. For aliphatic positions, the scaling factor was derived by comparing intensities of serine β -methylene

hydrogens, since these were the most highly protonated positions in all samples studied (see Table 1). The scaling factor was calculated to ensure that the sum of isotopomers at the serine β position was the same for all three protein samples. It is also important to note that in quantitating each isotopomer, individual peaks corresponding to CH_3 and $\text{CH}_3(\text{D})$ have to be scaled down by factors of 3 and 2, respectively, when compared to $\text{CH}(\text{D})$, due to the signal contribution of every proton. For aromatic sidechain positions, the most highly protonated positions were the $\text{H}\delta$ ring protons of Phe/Tyr, and these appeared to be protonated $\sim 100\%$ in all samples. The aromatic region of Pin1(unlabeled) is shown in Figure 5A and compared that of Pin1(glucose) (Figure 5B) or Pin1(FROMP) (Figure 5C).

When *E. coli* is grown in D_2O , almost all alpha hydrogens are derived biosynthetically from solvent during transamination reactions, so these positions are highly deuterated in Pin1(glucose) (Figure 4B) or Pin1(FROMP) (Figure 4C). The result agrees well with the findings of Rosen *et al*¹¹ and Otten *et al*³⁸, who observed ($\sim 95\%$) deuteration at the H^α of all amino acids. A notable exception in our study, however, is the H^α positions of methionine residues, which retain a small degree of protonation (see Figures 4B and 4C), a distinction we are at a loss to explain.

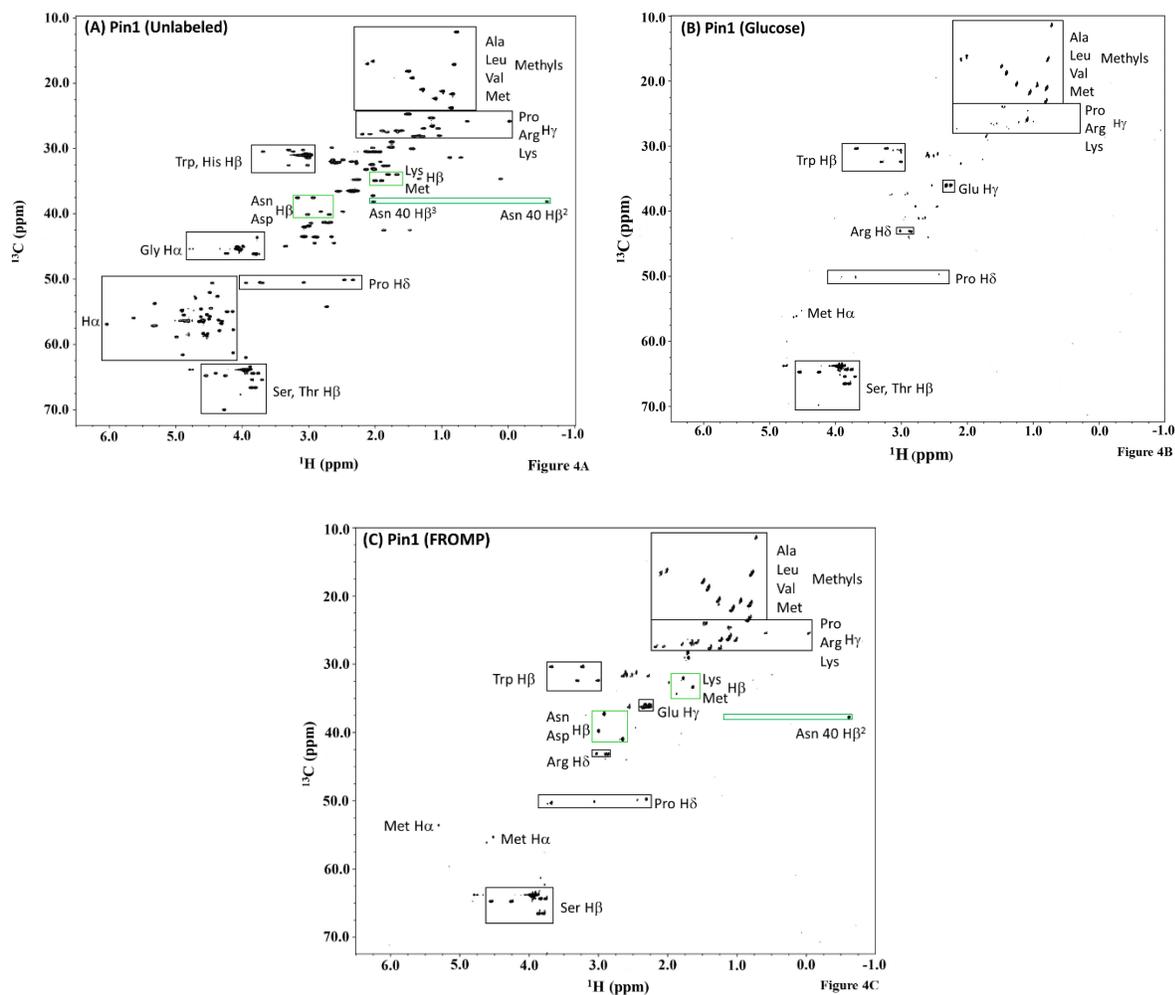


Figure 4. Natural abundance 2D ^1H - ^{13}C HSQC (aliphatic region) (A) Fully protonated Pin1 (unlabeled) (B) Deuterated Pin1 (glucose) (C) Deuterated Pin1 (FROMP). Highlighted in black are the fingerprint regions of human Pin1 for methyl-containing amino acids, H^α of all residues, H^β positions of Trp, His, Ser, Thr, H^β positions of Pro, Arg, Lys, and H^δ positions of Pro. Highlighted in green are the H^β protons of Asp, Asn, Lys, and Met. Protein concentration was 2 mM, pH ~ 6.8 in each sample. The acquisition and processing parameters for each spectrum were identical for comparison.

Table 1. Average percent isotopomers in the amino acids of mutant Pin1 WW domain grown in glucose/FROMP and D₂O

Amino acid type	C β		C γ			C δ			C ϵ			
Val	CH: 0/0	CD: 100/100										
Ile	CH: 0/0	CD: 100/100		CH ₂ : 0/0	CHD: 27/0	CD ₂ : 73/100						
Leu	CH ₂ : 0/0	CHD: 0/0	CD ₂ : 100/100	CH: 0/0	CD: 100/100							
Glu	CH ₂ : 0/0	CHD: 0/0	CD ₂ : 100/100	CH ₂ : 0/35	CHD: 30/38	CD ₂ : 70/27						
Gln	CH ₂ : 0/0	CHD: 0/0	CD ₂ : 100/100	CH ₂ : 0/35	CHD: 32/54	CD ₂ : 68/11						
Arg ^a	CH ₂ : 0/0	CHD: 0/0	CD ₂ : 100/100	CH ₂ : 0/53	CHD: 43/45	CD ₂ : 57/2	CH ₂ : 0/0	CHD: 80/50	CD ₂ : 20/50			
Pro ^a	CH ₂ : 0/0	CHD: 0/0	CD ₂ : 100/100	CH ₂ : 0/40	CHD: 30/33	CD ₂ : 70/27	CH ₂ : 0/0	CHD: 33/48	CD ₂ : 67/52			
Asp ^a	CH ₂ : 0/0	CHD: 0/156	CD ₂ : 100/0									
Asn	CH ₂ : 0/0	CHD: 20/160	CD ₂ : 80/0									
Met	CH ₂ : 0/0	CHD: 0/30	CD ₂ : 100/70	CH ₂ : 0/0	CHD: 35/45	CD ₂ : 65/55						
Lys ^a	CH ₂ : 0/0	CHD: 0/90	CD ₂ : 100/10	CH ₂ : 0/0	CHD: 38/35	CD ₂ : 62/65	CH ₂ : 0/0	CHD: 22/76	CD ₂ : 78/24	CH ₂ : 0/0	CHD: 0/0	CD ₂ : 100/100
Thr	CH: 27/0	CD: 73/100										
Ser ^b	CH ₂ : 90/75	CHD: 10/25	CD ₂ : 0/0									
Trp ^a	CH ₂ : 80/80	CHD: 20/20	CD ₂ : 0/0									
Phe	CH ₂ : 0/0	CHD: 40/35	CD ₂ : 60/65									
Tyr ^a	CH ₂ : 0/0	CHD: 30/30	CD ₂ : 70/70									
His	CH ₂ : 0/0	CHD: 0/0	CD ₂ : 100/100									

^aaverage of two residues, ^baverage of three residues.

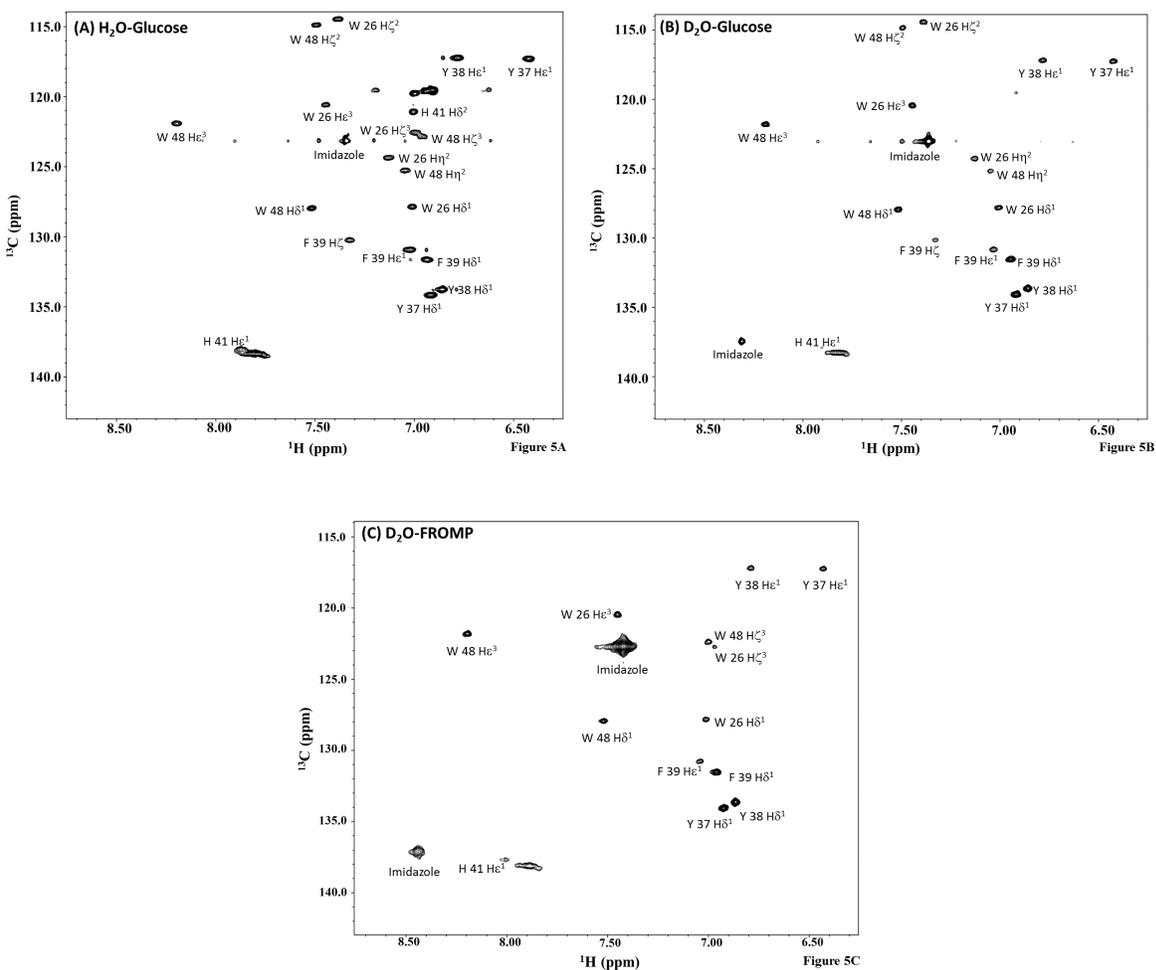


Figure 5. Natural abundance 2D ^1H - ^{13}C HSQC (aromatic region) (A) Fully protonated human Pin1 expressed in H_2O -glucose (B) Human Pin1 expressed in D_2O -glucose (C) Human Pin1 expressed in D_2O -FROMP. Protein concentration was 2 mM, pH \sim 6.8 in each sample. The acquisition and processing parameters for each spectrum were identical for comparison.

Pyruvate family (alanine, valine, leucine, isoleucine- γ 2)

Methyl groups are excellent probes of structure and dynamics in solution NMR studies because of their favourable relaxation properties and their occurrence in protein hydrophobic cores and protein-protein interfaces.^{39,40} Experimental protocols for the production of highly deuterated proteins with selective ^1H - ^{13}C labeling at methyl groups of alanine, valine, leucine and isoleucine have been established in many studies.^{3,38} In theory, these protocols can be combined with our

FROMP-media to produce ^1H - ^{13}C methyl labeling at these amino acids. Thus, we will not discuss the biosynthetic pathways for these amino acids in detail, focusing more on other residue types for which isotope labeling strategies have not been as well established.

Isotopomeric distributions for methyl groups have been studied in detail using ^1H -glucose as carbon source by Otten *et al.*³⁸ and Shekhtman *et al.*⁴¹, as well as by Rosen *et al.*¹¹ for ^1H -pyruvate. The results obtained for Pin1(FROMP) are similar to those obtained by Rosen *et al.*¹¹ for ^1H -pyruvate, as expected, given that pyruvate is the source of Ala, Val, Leu, and Ile- γ 2 methyl groups in both studies. It is noteworthy that we obtain the same isotopomeric distributions for these amino acids despite the fact that our FROMP media also contains rhamnose and fumarate, which would be metabolized to phosphoenolpyruvate and oxaloacetate, respectively. There does not appear to be substantial production of pyruvate from these metabolites supplied in the FROMP media, suggesting that oxalate in the media was successful in suppressing metabolic flux through the PEP–pyruvate–oxaloacetate node.²⁹ As pointed out by Rosen *et al.*,¹¹ the high degree of scrambling observed in pyruvate may be due to the enzymatic activity of alanine aminotransferase, which catalyzes the interconversion of pyruvate and alanine, introducing solvent deuterons to the methyl group in the process^{42,44}.

The hydrogen atoms in the Ile- δ 1 and Thr- γ 2 methyl groups are not derived from pyruvate and display a different isotopomer distribution from the methyl groups of Val and Leu. The Ile- δ 1 methyl group is derived from Thr- γ 2, and these two methyl groups show almost identical isotopomer distributions (Table 2).

When ^1H -glucose is utilized as carbon source in D_2O media, much higher degrees of deuteration are obtained compared to ^1H -pyruvate, since all pyruvate for biosynthesis would have

to be derived from phosphoenolpyruvate, oxaloacetate, or malate. The levels of methyl group deuteration shown in Table 2 are far greater than those obtained by Otten *et al*³⁸, due to a larger estimation of the population of “invisible” CD₃ isotopomers, ~50% versus ~0%. Aside from this discrepancy, both studies agree with a roughly 1:1 ratio of CH₃D:CHD₂ isotopomers.

Table 2. Average percent isotopomers in methyl-containing residues of Pin1 WW domain grown in glucose/FROMP and D₂O

	Ala (3)*-β	Val-γ2	Val-γ1	Leu-δ1	Leu-δ2	Ile-δ1	Ile-γ2	Met (2)*-ε	Thr-γ2
CH ₃	3/17	4/43	5/50	4/52	0/63	0/0	6/29	0/2	0/4
CH ₂ D	15/17	18/27	26/28	21/31	20/30	5/8	21/23	7/8	6/10
CHD ₂	21/24	18/22	26/26	24/26	23/17	26/31	18/23	7/7	31/37
CD ₃	61/42	60/8	43/0	51/0	57/0	69/61	55/25	86/83	63/49

* Average of three alanines and two methionines, all other values are derived from single residues.

Phosphoglycerate family (serine, cysteine, tryptophan, and glycine)

Serine is a major metabolic precursor in the biosynthesis of cysteine, tryptophan, and glycine. The **human** Pin1 WW domain possesses no cysteine residue. All alpha hydrogens exchange with solvent in a D₂O-based growth, so glycine is entirely deuterated. Serine and tryptophan have similarly high levels of protonation at the H^β position in Pin1(glucose), with 80-90% CH₃ isotopomer. These hydrogen atoms are derived from the 3-position of 3-phosphoglycerate in the glycolytic pathway (Figure 6 and Figure 2). It is somewhat surprising that the protonation is so markedly greater than 50%, since one might expect the 3-position of 3-phosphoglycerate to have an isotopomeric ratio CH₃:CHD of 1:1, with the two isotopomers derived from the glycolytic catabolism of a single glucose molecule. However, glycolysis yields 100% CH₃ isotopomer, because the enzyme, phosphohexose isomerase (which catalyzes the conversion of glucose-6-phosphate to fructose-6-phosphate), transfers a proton from the 2-position of glucose to the 1-position of fructose, instead of deriving a deuteron from solvent.⁴⁵ Thus, glycolysis in D₂O

yields two identical molecules of 3-³H-3-phosphoglycerate from a single molecule of ¹H-glucose. Rosen *et al.* also observed very high levels of protonation at H^β for Ser, Trp, and Cys grown with glucose as carbon source, but a much lesser degree of protonation when ¹H-pyruvate was used.¹¹ In contrast, a high proportion of Ser and Trp CH₂ isotopomer was maintained in Pin1(FROMP) Table 1, suggesting that most of the 3-phosphoglycerate used in the biosynthesis of serine was derived from catabolism of rhamnose and not from gluconeogenic precursors derived from pyruvate/oxaloacetate. This result is likely due to the availability of rhamnose as a carbon source, as well as the presence of oxalate in FROMP media as a tight inhibitor of PEP carboxykinase and PEP synthetase, limiting the production of phosphoenolpyruvate from oxaloacetate and pyruvate, respectively.

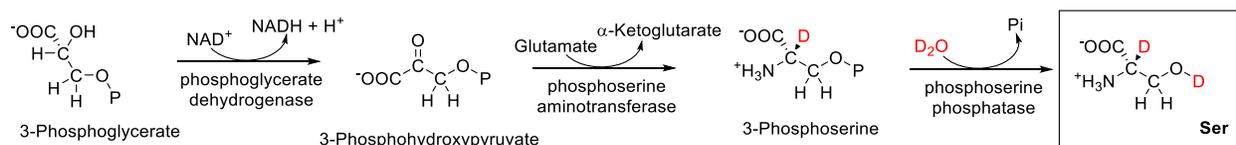


Figure 6. Biosynthesis of serine from 3-phosphoglycerate.

Oxaloacetate family (aspartate, asparagine, methionine, threonine, lysine, isoleucine)

The amino acids asparagine, methionine, lysine, threonine, and isoleucine (C^α, C^{γ1} and C^{δ1}) are derived from aspartate, formed in a transamination reaction involving oxaloacetate. Oxaloacetate is generated as a TCA cycle intermediate from the oxidation of (L)-malate, which in turn is generated by the addition of a water molecule across the double bond of fumarate. This reaction is catalyzed by the enzyme fumarase, and when the reaction takes place in D₂O, the result is (2S,3R)-3-³H-malate (Figure 3).³³ We hypothesized that using fumarate as a carbon source would result in stereospecific incorporation of ¹H at the H^{β2} position and ³H at the H^{β3} position of aspartate.

In fact, this is what was observed, with a strong ^1H signal at the $\text{H}^{\beta 2}$ position and no detectable ^1H signal at the $\text{H}^{\beta 3}$ position in Pin1(FROMP) Figure 4C, as opposed to virtually no detectable signal at either position in Pin1(glucose) Figure 4B. Based on signal intensities, one would estimate 160% ^1H incorporation into the $\text{H}^{\beta 2}$ position of Asp/Asn Pin1(FROMP) Table 1, which is clearly an overestimate. The signal is more intense than that observed in Pin1(unlabeled) because the transverse relaxation of the CHD group in Pin1(FROMP) is much slower than that of the CH_2 group in Pin1(unlabeled). In contrast, Asp16 is part of the mobile N-terminal tail of the protein, so its signals were twice as intense as those originating from other Asp/Asn residues. For Asp16, the observed signal intensity ratio was closer to 100% rather than 160%, because relaxation differences are less of a factor in determining signal intensity.

The high level of stereospecific ^1H incorporation at the $\text{H}^{\beta 2}$ position of Asp/Asn implies that the ^1H atom originating on fumarate remained attached through its enzymatic conversion to malate, oxaloacetate, aspartate, and asparagine. In contrast, when pyruvate is converted to alanine by alanine aminotransferase, methyl protons are readily replaced by solvent deuterons,⁴²⁻⁴⁴ explaining why protonation of the Ala methyl is substantially lower than that of Leu and Val methyls, despite the fact that all are derived from pyruvate. Thus, aspartate aminotransferase is able to catalyze the conversion from oxaloacetate to aspartate without perturbing hydrogens at the β position, in contrast to alanine aminotransferase, which tends to exchange them with solvent.

A prerequisite for the stereospecific labeling of $^1\text{H}^{\beta 2}$ and $^2\text{H}^{\beta 3}$ in Asp/Asn is that the oxaloacetate precursor is derived entirely from the fumarate supplied in FROMP media. Oxaloacetate is produced from succinate through the activity of succinate dehydrogenase, which is inhibited by malonate. Oxaloacetate is also produced by anaplerotic reactions, most notably the carboxylation of phosphoenolpyruvate by PEP carboxylase, a reaction inhibited both by oxalate

and TCA cycle intermediates.^{46,47} Hence, the FROMP media used in the current study was effective at preventing isotope scrambling related to the key PEP-oxaloacetate-pyruvate node³⁹.

Lys, Thr, and Met are also derived from aspartate. The biosynthesis of these amino acids requires reduction of the C γ carboxyl of Asp, consuming two reducing equivalents from NADPH. The C γ positions of Lys, Thr, and Met are partially protonated, with the fully deuterated isotopomer predominating (>50%) in both Pin1(glucose) and Pin1(FROMP) Table 1. This incorporation indicates that a small proportion of the hydride in NADPH is ^1H , rather than ^2H . In *E. coli*, NADP $^+$ is converted to NADPH through the activities of glucose-6-dehydrogenase and 6-phosphogluconate dehydrogenase in the pentose phosphate pathway, as well as isocitrate dehydrogenase in the TCA cycle.⁴⁸ NADPH is required in the biosynthesis of a number of amino acids, so the ^1H : ^2H ratio of NADPH will impact isotopic labeling at many different sites, resulting in isotopic mixtures. Improving the homogeneity of labeling will require either finding a way to control the ^1H : ^2H ratio of NADPH or bypassing NADPH-dependent synthetic steps with downstream metabolites.

In the biosynthesis of lysine (Figure 7), pyruvate is added to the skeleton of aspartate- β -semialdehyde, accounting for the high level of protonation observed at the H $^{\beta}$ position in Pin1(FROMP). One intermediate in the biosynthetic pathway, L,L- α,ϵ -diaminopimelate, is a symmetric molecule (Figure 7), so that the β and δ positions are interchangeable, as are the (deuterated) α and ϵ positions. Thus, the aspartate- β -semialdehyde component contributes equally to both β and δ positions, and the pyruvate component contributes equally to these positions as well. In Pin1(FROMP), the H $^{\beta 2}$ position has an apparent protonation of 90% (although given the relaxation effects observed in Asp/Asn, the actual percentage may be closer to 60%), while the H $^{\beta 3}$ position is entirely deuterated. Therefore, the stereospecificity of protonation introduced by the

aspartate- β -semialdehyde precursor is the same as that introduced by the pyruvate precursor (coincidentally). Given the symmetry of the L,L- α,ϵ -diaminopimelate, the corresponding H^{8s} position must be equally protonated and the H^{8r} position deuterated. As is usually the case, the Lys sidechain δ hydrogens have chemical shifts that are too overlapped to make this distinction. Thus, it is not possible to make stereospecific assignments on the basis of NMR spectra, yet it is possible to assign the stereochemistry based on symmetry considerations in the biosynthesis of lysine.

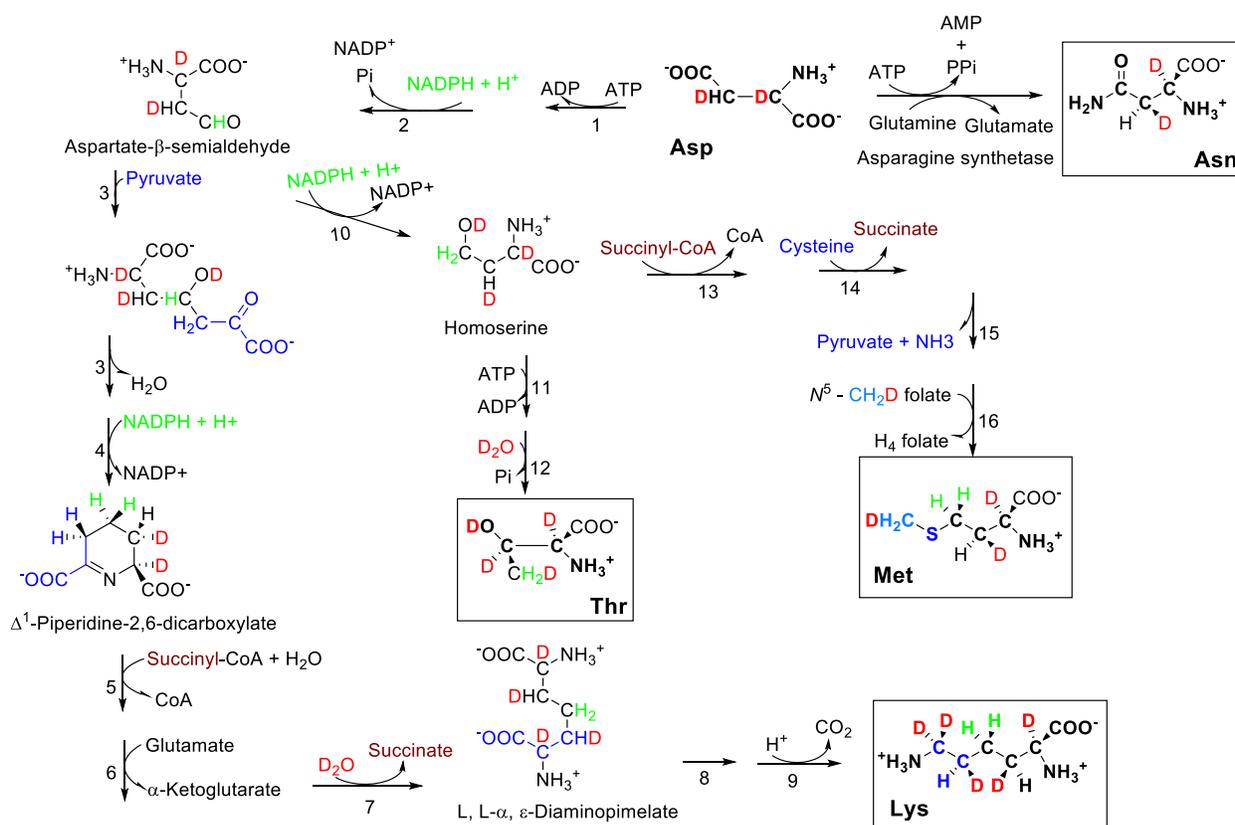


Figure 7. Pathways for the biosynthesis of asparagine, lysine, threonine, and methionine. Atoms derived from solvent (red), protons derived from NADPH (green), protons derived from pyruvate (blue), protons derived from methyltetrahydrofolate (light blue), sulfur derived from cysteine in methionine biosynthesis (blue). The pathway enzymes are: 1. Aspartokinase 2. Aspartate- β -semialdehyde dehydrogenase (entry point of a proton to H^r of all amino acids belonging to the oxaloacetate family) 3. Dihydropicolinate synthase 4. Δ^1 -piperidine-2,6-dicarboxylate dehydrogenase (entry point of a proton to the H^r of lysine) 5. *N*-succinyl-2-amino-6-ketopimelate synthase 6. Succinyl diaminopimelate aminotransferase (PLP enzyme) 7. Succinyl diaminopimelate desuccinylase 8. Diaminopimelate epimerase 9. Diaminopimelate decarboxylase

10. Homoserine dehydrogenase (entry point of proton to the H^γ of threonine and methionine) 11. Homoserine kinase 12. Threonine synthase (PLP dependent enzyme) 13. Homoserine acyl transferase 14. Cystathionine-γ-synthase 15. Cystathionine-β-lyase 16. Methionine synthase.

Aspartate-β-semialdehyde is also converted into homoserine (Figure 7), a common branch point for the synthesis of threonine and methionine. Isomerization of homoserine to threonine occurs through a phosphorylated intermediate that displaces a hydrogen atom from C^β. Thr H^β retained around 27 % protonation in Pin1(glucose) Table 1, similar to the findings of Otten *et al*³⁸. However, in Pin1(FROMP), it is the H^{β3} from homoserine that is retained (unfortunately), so that the H^β is completely deuterated at this position. Homoserine is also converted into homocysteine as a synthetic precursor to methionine. Homocysteine is then methylated at its sulfur atom by 5-methyltetrahydrofolate, producing methionine and tetrahydrofolate. 5-methyltetrahydrofolate can be regenerated either through the catabolism of serine or glycine. Given the very high level of deuteration observed at the methionine εCH₃ group of Pin1(glucose) and Pin1(FROMP) (see Table 2), it would appear that glycine is the dominant source of single carbons in the folate cycle. The H^{β2} position of Met is substantially less protonated (30%) than the corresponding position in Asp/Asn/Lys, presumably as a result of one of the many enzymatic steps leading from aspartate-β-semialdehyde to methionine.

α-ketoglutarate family (glutamate, glutamine, proline, arginine)

α-ketoglutarate is a TCA cycle intermediate that gives rise to glutamate, glutamine, proline and arginine. As shown in Figure 8, α-ketoglutarate formed in the TCA cycle is deuterated at the H^β positions because it originates from the oxaloacetate ketone carbon, which condenses with acetyl-CoA and subsequently acquires ³H from D₂O solvent. However, the C^γ of α-ketoglutarate are derived from the methyl group of pyruvate, hence Glu, Gln, Arg, and Pro are highly protonated

at H⁷ in Pin1(FROMP) and to a much lesser degree in Pin1(glucose) Table 1. The relative isotopomeric populations are consistent with those observed in the methyl groups of Val and Leu, which are also derived from pyruvate.

The H⁶ hydrogens in proline are derived from NADPH or NADH. In the biosynthesis of proline, glutamate is phosphorylated, followed by the conversion of γ -glutamyl phosphate into glutamate γ -semialdehyde, resulting in the incorporation of a hydrogen at the H⁶ of proline from NADH (Figure 8). Δ^1 -Pyrroline-5-carboxylate, the product of non-enzymatic cyclization of glutamate γ -semialdehyde, undergoes a reduction reaction with NADH or NADPH providing a second hydrogen at H⁶. There appeared to be some stereospecificity to the low protonation pattern at the proline H⁶ position, but we were unable to resolve this further.

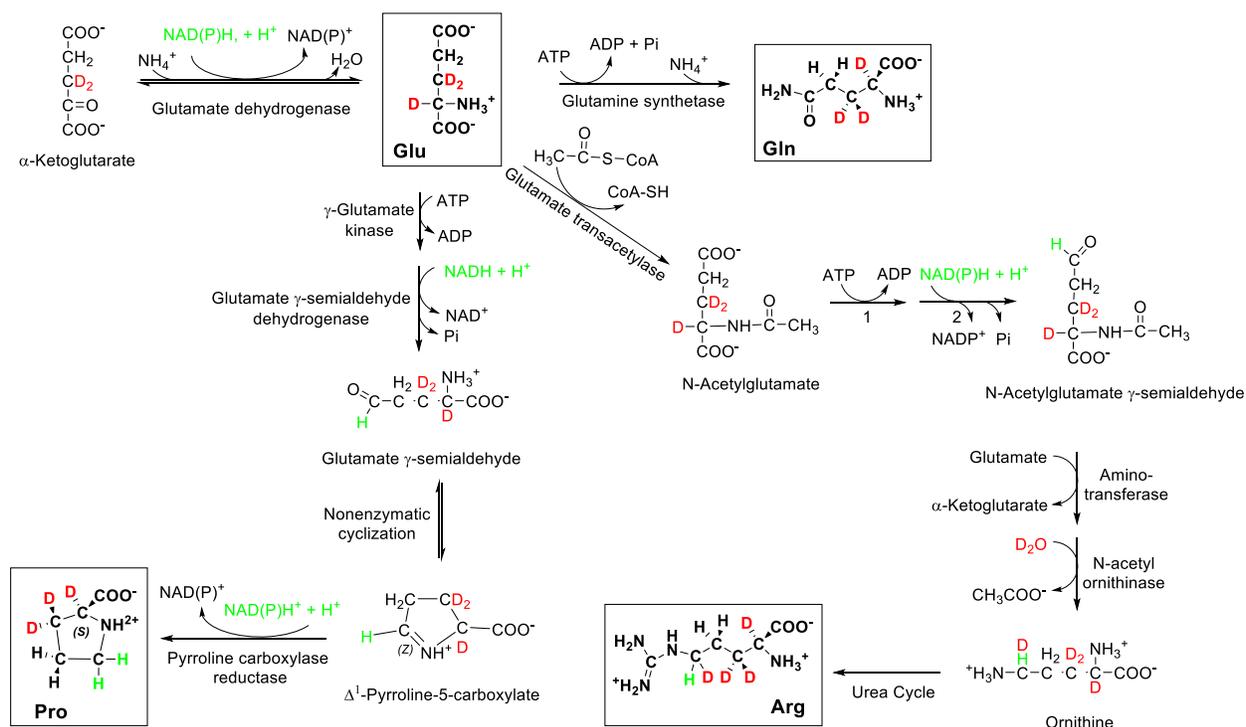


Figure 8. Biosynthesis of glutamate, glutamine, proline and arginine from α -ketoglutarate. Protons incorporated from NADPH are shown in green. NADPH supplies protons to the H⁶ positions of proline and arginine. Atoms incorporated from solvent are shown in red. Pathway

enzymes labeled 1 and 2 are *N*-acetylglutamate kinase and *N*-acetylglutamate dehydrogenase respectively.

In arginine, the reduction at the H^b position requires the conversion of N-acetyl- γ -glutamyl phosphate into N-acetylglutamate γ -semialdehyde using NADH or NADPH as cofactors (Figure 8). The semialdehyde is then converted to N-acetylornithine through a transamination reaction that converts glutamate into α -ketoglutarate. Ornithine is subsequently converted into arginine via the urea cycle. Thus, the H^b position of arginine acquires one H atom from NADH/NADPH and one solvent deuteron from a transamination reaction. We have no explanation why the protonation level observed at H^b of arginine was higher in Pin1(glucose) than Pin1(FROMP) Table 1. It is possible that there are cytosolic pools of NADPH with varying ratios of ¹H:²H depending on the local enzymes that regenerate it.

Phosphoenolpyruvate + erythrose-4-phosphate family (tryptophan, phenylalanine, tyrosine)

Erythrose-4-phosphate and two molecules of phosphoenolpyruvate (PEP) are the precursors for chorismate, the common branch point leading to the synthesis of the aromatic amino acids, phenylalanine, tyrosine, and tryptophan.

The six-member ring in the tryptophan indole group is derived from chorismate. The 1- (aldehyde), 2-, 3-, and 4-positions of erythrose-4-phosphate become the C², C², C³, and C³ positions of Trp. H² and H² are partially protonated in Pin1(glucose) but entirely deuterated in Pin1(FROMP) (Figure 9). In contrast, The H³ position in Trp is derived from the 4-position of erythrose-4-phosphate, and it is ~100% protonated in both Pin1(glucose) and Pin1(FROMP). The H³ position of Trp is derived from NADPH, so it would be expected to be partially protonated and partially deuterated in both Pin1(glucose) and Pin1(FROMP). However, the protonation level was higher in Pin1(FROMP) for unknown reasons. The five-member ring of the Trp indole group is

derived from phosphoribose (Figure 10). The Trp H^{δ1} position is derived from position 2 of phosphoribose and is highly protonated, although less so in Pin1(FROMP) than in Pin1(glucose) (Figure 9).

The labeling pattern observed at the H^δ and H^ε of both phenylalanine and tyrosine residues is similar in both Pin1(glucose) and Pin1(FROMP) Figure 9. The H^{δ1} position of phenylalanine and tyrosine is derived from the methylene group of PEP, while the 4-position of erythrose-4-phosphate is the source of protons at H^{δ2}, accounting for the high level of protonation observed at the H^δ site (the H^{δ1} and H^{δ2} positions are interchangeable) Figure 11. The H^{ε1} and H^{ε2} positions of Phe/Tyr have the same biosynthetic origin as the H^{ε3} and H^{ε2} positions of tryptophan, respectively. Although the H^{ε1} and H^{ε2} positions of Phe/Tyr cannot be distinguished from each other spectroscopically due to ring flips, one would expect levels of protonation similar to their corresponding sites in tryptophan. A higher degree of protonation at the H^{ε3} position of phenylalanine was observed for Pin1(glucose) versus Pin1(FROMP), which is consistent with the corresponding H^{ε2} position in tryptophan, both derived from the 2-position of erythrose-4-phosphate.

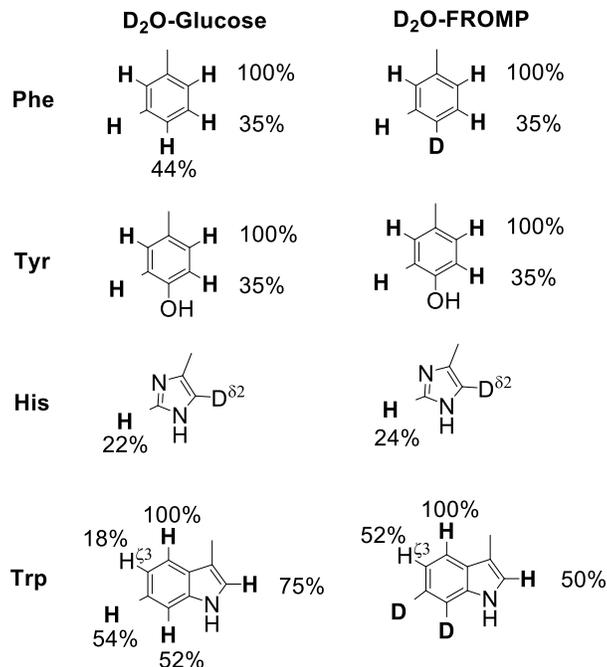


Figure 9. Average percent protonation observed at the aromatic sidechains of phenylalanine, tyrosine, tryptophan and histidine for Pin1(glucose) and Pin1(FROMP).

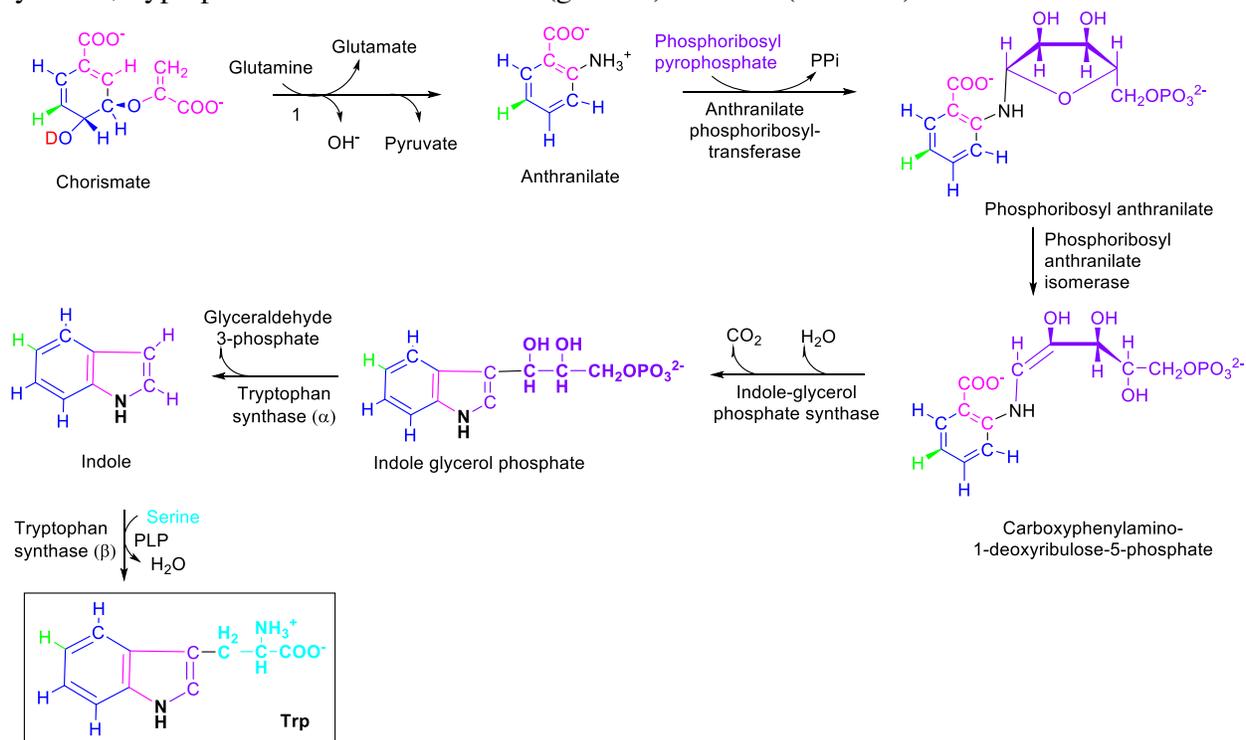


Figure 10. Biosynthesis of tryptophan from chorismate. Atoms derived from phosphoenolpyruvate, erythrose-4-phosphate, phosphoribosyl pyrophosphate, NADPH, serine and solvent (D₂O) are shown in pink, blue, purple, green, light blue, and red, respectively. The pathway enzyme 1 is anthranilate synthase.

The H^β positions of phenylalanine and tyrosine are derived from a molecule of phosphoenolpyruvate, which condenses with shikimate to eventually form chorismate (Figure 11). It is unknown why the protonation levels at H^β for Phe and Tyr are so low (30-40%, non-stereospecific) compared with Ser and Trp, despite all being derived from phosphoenolpyruvate. It is possible that exchange with solvent occurs at the transamination step as it does with alanine, but this seems unlikely, given that the aromatic amino acid aminotransferase is similar to aspartate aminotransferase (which leaves H^β protonation intact, as we have noted) in protein structure, substrate specificity, and spectroscopic properties.⁴⁹ There are a number of enzymes catalyzing reactions between the condensation of phosphoenolpyruvate with phosphoshikimate and the final transamination reaction at which solvent exchange could occur: 3-enoylpyruvateshikimate-5-phosphate synthase, chorismate synthase, and chorismate mutase (Figure 11). Of these, chorismate mutase catalyzes the formation of prephenate from chorismate via a Claisen rearrangement, establishing the C^β-C^γ bond, so it is quite possible that the solvent exchange occurs at this step.

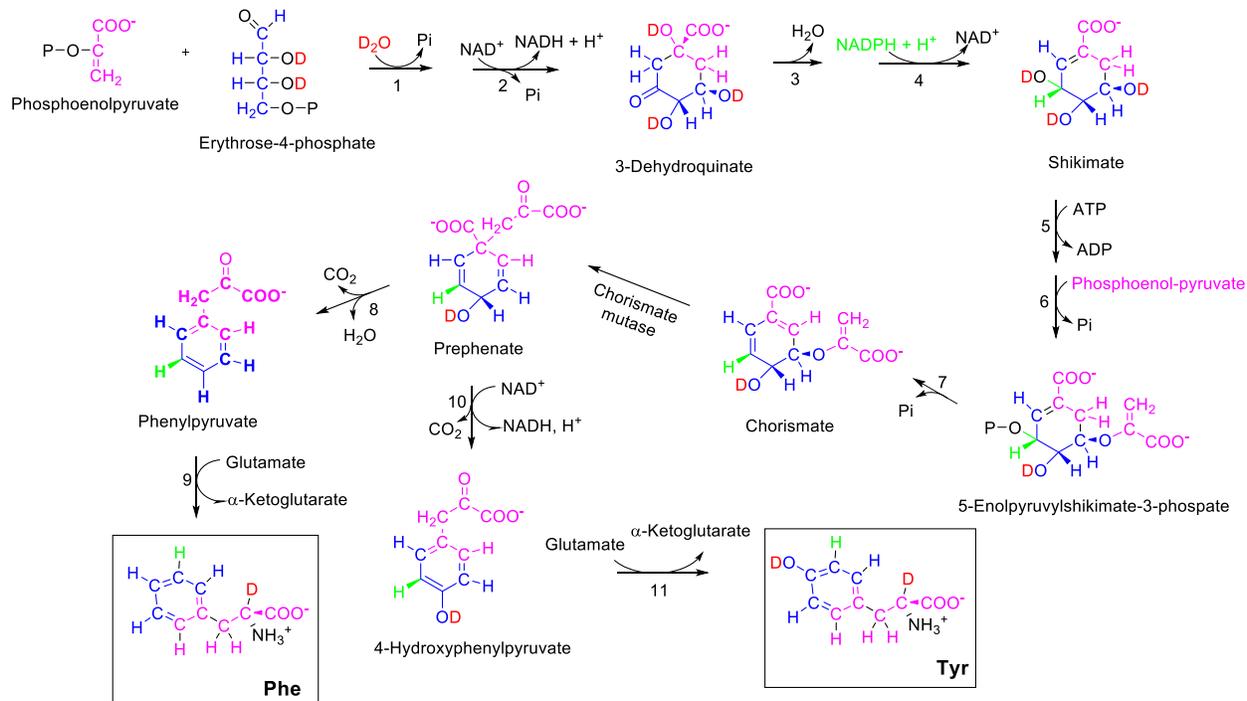


Figure 11. Biosynthesis of phenylalanine and tyrosine. Atoms derived from phosphoenolpyruvate, erythrose-4-phosphate, NADPH and solvent (D₂O) are shown in pink, blue, green and red, respectively. The pathway enzymes are: 1. 2-keto-3-deoxy-D-arabinoheptulosonate-7-phosphate synthase 2. dehydroquinate synthase 3. 3-dehydroquinate dehydratase 4. shikimate dehydrogenase 5. shikimate kinase 6. 3-enolpyruvateshikimate-5-phosphate synthase 7. chorismate synthase 8. prephenate dehydratase 9. aminotransferase 10. prephenate dehydrogenase 11. aminotransferase.

Ribose phosphate family (histidine)

Histidine is the only amino acid belonging to this family, and it is derived from three precursors. Phosphoribosyl pyrophosphate (PRPP) contributes five carbon atoms, the purine ring of ATP contributes a carbon and a nitrogen, and glutamine contributes the second nitrogen in the imidazole ring (Figure 12). The labeling pattern observed in histidine was consistent for both Pin1(glucose) and Pin1(FROMP), producing complete deuteration at H^β and H^{β2}. A small degree of protonation at H^α (~20%, Figure 9) derives from the six-membered ring of adenine, which is contributed by N¹⁰-formyl-tetrahydrofolate.

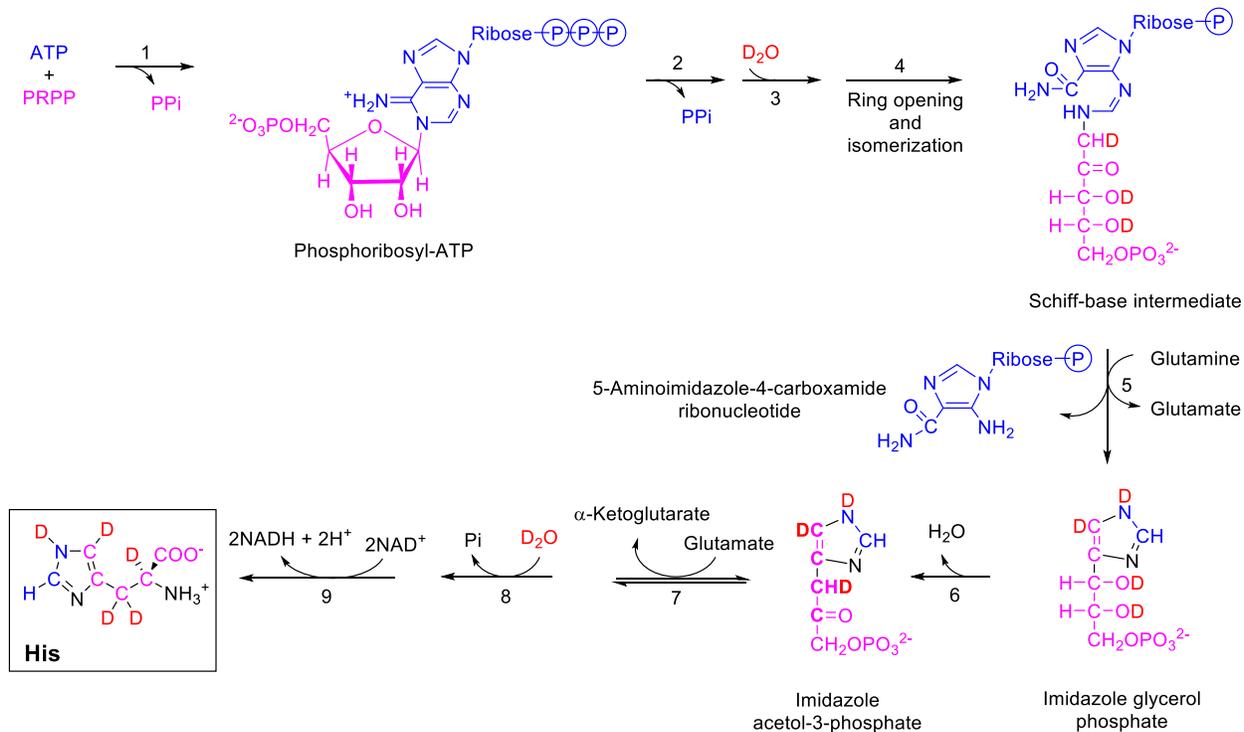


Figure 12. Outline of the biosynthesis of histidine. Atoms derived from PRPP and ATP are shaded in pink and blue, respectively. Deuterium atoms derived from solvent are shown in red. The pathway enzymes are: 1. ATP phosphoribosyl transferase 2. pyrophosphohydrolase 3. phosphoribosyl-AMP cyclohydrolase 4. phosphoribosyl-formimino-5-aminoimidazole-4-carboxamide ribonucleotide isomerase 5. glutamine amidotransferase 6. imidazole glycerol-3-phosphate dehydratase 7. L-histidinol phosphate aminotransferase 8. histidinol phosphate phosphatase 9. histidinol dehydrogenase.

Application of FROMP labeling to χ_i dihedral angle determination

The precision and accuracy of NMR-derived protein structures could be significantly improved with stereospecific assignments of β -methylene protons and χ_i dihedral angle restraints.⁵⁰ The utility of three-bond J coupling for determining backbone and side chain dihedral angles has long been established,⁵¹⁻⁵³ and 3D experiments have been developed to measure these. One such experiment is the 3D HNHB, which correlates amide proton and nitrogen resonances with intra-residue H β resonances, providing semi-quantitative information on the size of $^3J(\text{NH}^\beta)$ Figure 13b and 13c.⁵¹ An analogous experiment is the HN(CO)HB, which transfers magnetization via the 3J -

coupling between carbonyl carbon and H^β protons.⁵⁴ $^3J_{\alpha\beta}$ can also be estimated using relative peak intensities in a 3D ¹H-TOCSY-¹⁵N-HSQC.⁵⁰ Together, these three experiments provide redundant information about the χ_1 dihedral angle, along with stereospecific assignment of the β -methylene protons (Figure 13a).

Figures 13b and 13c show strip plots of the HNHB, HN(CO)HB, and ¹H-TOCSY-¹⁵N-HSQC spectra for human Pin1 Asn40 and Asn44, respectively. Asn40 is found in the structured region of the second β -strand of Pin1 (Figure 1), and the crystal structure (1ZCN) shows that it adopts a “trans” (180°) χ_1 dihedral angle. All of the experiments in Figure 13b are consistent with the “trans” conformation, also allowing the stereospecific assignment of the upfield resonance to H^{β2}. The application of FROMP labeling technique confirmed this assignment (Figure 4c).

Asn44 is found in the loop region connecting the second and third β -strands of human Pin1 WW domain (Figure 1). The χ_1 dihedral angle of this residue is gauche⁻ (-60°) from the crystal structure. However, the conformation predicted using ³J-coupling measurements was gauche⁺ (+60°) (Figure 13c), with the upfield resonance assigned as H^{β3}. However, stereospecific labeling with FROMP media unambiguously indicated the upfield resonance to be H^{β2}. These apparently incongruous results are most readily explained by internal motions, more specifically, averaging between two χ_1 dihedral angle conformers, trans and gauche⁻. This is more in keeping with the gauche⁻ conformation observed in the crystal structure. Consistent with motional averaging, the difference between strong and weak signals in the ³J-coupling experiments was less pronounced in Asn44 than in Asn40. Thus, the semi-quantitative assessment of ³J-couplings was unable to distinguish between a pure gauche⁻ conformation and rotameric averaging between gauche⁻ and trans conformations, requiring the correct stereospecific chemical shift assignments to make the distinction.

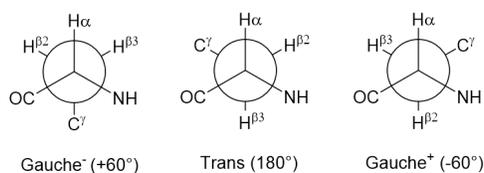


Figure 13A

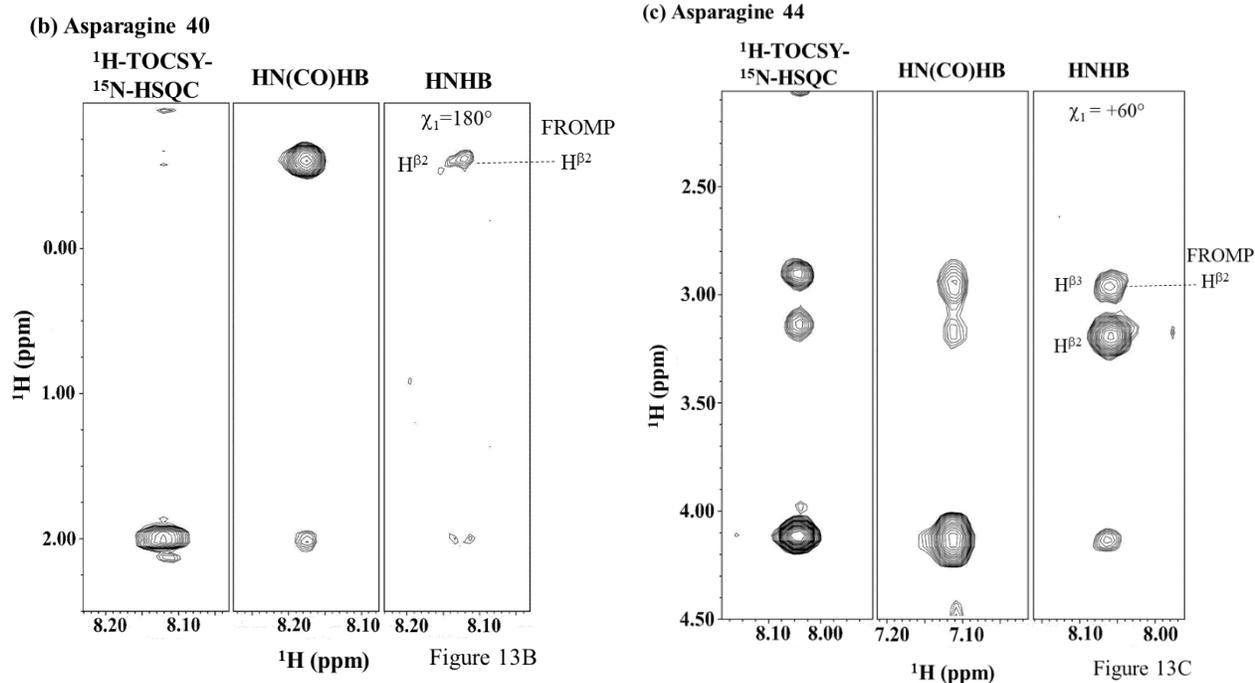


Figure 13. Application of FROMP labeling technique for stereospecific assignment of β -methylene protons. (A) Rotamers showing gauche ($+60^\circ$), trans (180°), and gauche (-60°) χ_1 dihedral angles. Strip plots from 3D HNHB, HN(CO)HB, and ^1H -TOCSY- ^{15}N -HSQC spectra for (B) Asn 40 and (C) Asn 44 of the mutant Pin1 WW domain. 3D HNHB and ^1H -TOCSY- ^{15}N -HSQC strip plots are taken at the amide frequency of Asn 40 and Asn 44, while HN(CO)HB strip plots are taken at the amide frequency of His 41 and Ala 45. The intensity of each correlation is related to the size of each 3-bondJ-coupling.

Asp18 and Lys21 are found in the flexible N-terminal region of Pin1, and both residues are absent in the crystal structure. The H^β signals from these residues were too overlapped in 3D-NMR experiments to determine a χ_1 dihedral angle. However, using FROMP labeling, we were still able to obtain stereospecific assignments for these residues.

The residue Asp33 was absent in the crystal structure because it was mutated in our Pin1 construct²¹. We could not detect Asp33 H^β signals in two out of the three χ_1 dihedral angle

experiments due to exchange broadening of Asp33 NH. However, we were still able to obtain stereospecific assignments for this residue using our isotope labeling technique.

Lys28 and Met30 are found in the first β -strand of Pin1. In the crystal structure, both of these residues adopt trans (180°) χ_1 dihedral angles. However, the conformation predicted through 3J -coupling measurement was gauche ($+60^\circ$), with the assignment of the downfield H^{β} signals to H^{α} for both residues. However, stereo-selective labeling using FROMP media assigned the upfield signals to H^{α} , meaning that the χ_1 dihedral angle assignment was incorrect, similar to Asn44. The χ_1 dihedral angles of Lys28 and Met30 are better described as a rotameric average between gauche and trans conformations, which is again more consistent with the X-ray crystal structure.

In high resolution X-ray crystal structures, a small percentage of residues can be seen to adopt multiple χ_1 dihedral angles. In NMR studies of residual dipolar couplings in multiple alignment media, up to 50% of residues adopt multiple χ_1 dihedral angles.^{55,56} Our limited data on Asp/Asn/Lys/Met residues seem to support the notion that multiple side chain conformations are common in solution. Our analysis of Lys28, Met30, and Asn44 suggests that the use of 3J -coupling measurement experiments alone to determine stereospecific chemical shift assignments and χ_1 dihedral angles is unreliable when rotameric averaging is present. Therefore, it is anticipated that stereo-selective labeling will improve the accuracy of NMR-based χ_1 dihedral angle determinations.

Application of the FROMP labeling technique to structure determination of large proteins

Structure determinations of larger proteins by solution NMR have focused on the production of highly deuterated proteins with selective 1H - ^{13}C labeling at methyl groups (present in Ala, Ile, Leu, Met, Thr, and Val). Structures are then solved on the basis of NH-NH, NH-methyl

and methyl-methyl NOEs.⁵⁸ Our labeling strategy provides a complimentary approach that provides additional stereospecific protonation at Asp, Asn, and Lys (H β) amino acid sidechains. It also provides a high degree of protonation for at least one site in an additional 9 out of the 20 amino acids (Glu, Phe, Pro, Gln, Arg, Ser, Trp, Tyr, and probably Cys). When combined with methyl-specific protonation of aliphatic amino acids, it should now be possible to obtain NOE-based distance restraints for a majority of protein residues via NOEs from these new ¹H-¹³C positions to ¹H-¹³C methyl groups.

CONCLUSIONS

We thus demonstrate a convenient and inexpensive technique for obtaining stereospecific β -methylene deuteration in Asp, Asn, Lys, and Met amino acid residues. Not only does the method produce isolated ¹H magnetization that is optimal for solution NMR studies of large (>30 kDa) proteins, but the stereospecific assignments it provides can be used to better delineate sidechain rotamers and dynamics in smaller protein systems as well. Additional techniques could be developed to obtain stereospecific labeling in all 15 amino acid residues that contain β -methylene groups, though this will likely require the use of more expensive amino acid precursors.

AUTHOR INFORMATION

Corresponding Author

Department of Medicine, University of Alberta, Edmonton, Alberta, Canada T6G 2R3 780-492-4158, 780-492-2722 (Fax) phwang1@ualberta.ca

Notes

The authors declare that they have no conflicts of interest with the content of this article.

ACKNOWLEDGMENT

This work was supported by a Discovery Grant, RGPIN-2015-06664, from NSERC (Natural Sciences and Engineering Research Council of Canada), startup funds from the Department of Medicine and Faculty of Medicine and Dentistry and the University of Alberta, and private funding from the Hwang Professional Corporation. We would like to acknowledge the anonymous NSERC reviewer who recommended WW domains as a potential model system, as well as Monica Li, for her assistance with acquiring NMR spectra.

REFERENCES

- (1) <http://www.rcsb.org/pdb/statistics/holdings.do>
- (2) Clore, G. M., Gronenborn, A. M. (1998) NMR structure determination of proteins and protein complexes larger than 20 kDa. *Curr. Opin. Chem. Biol.* 2, 564-570.
- (3) 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, 749-754.
- (4) LeMaster, D. M., Richards, F. M. (1988) NMR sequential assignment of Escherichia coli thioredoxin utilizing random fractional deuteration. *Biochemistry.* 27, 142-150.
- (5) Farmer II, B. T., Venters, R. A. (1999) NMR of Perdeuterated Large Proteins. *Mod. Techniques Protein NMR.* 16, 75-120.
- (6) Gardner, K. H., Kay, L. E. (1998) The use of ^2H , ^{13}C , ^{15}N multidimensional NMR to study the structure and dynamics of proteins. *Annu. Rev. Biophys. Biomol. Struct.* 27, 357-406.
- (7) Goto, N. K., Kay, L. E. (2000) New Developments in Isotope Labeling Strategies for Protein Solution NMR Spectroscopy. *Curr. Opin. Struct. Biol.* 10, 585-592.
- (8) Gardner, K. H., Rosen, M. K., Kay, L. E. (1997) Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR. *Biochemistry.* 36, 1389-1401.
- (9) LeMaster, D. M. (1987) Chiral β and random fractional deuteration for the determination of protein side-chain conformation by NMR. *FEBS Lett.* 223, 191-196.

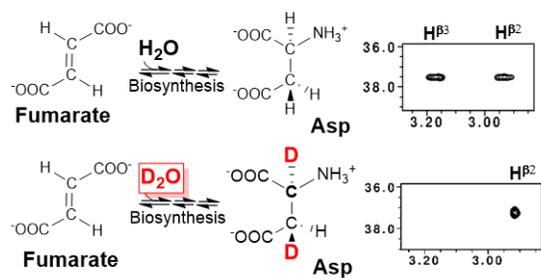
- (10) Gardner, K. H., Kay, L. E. (1997) Production and Incorporation of ^{15}N , ^{13}C , ^3H (^1H - δ^1 Methyl) Isoleucine into Proteins for Multidimensional NMR Studies. *J. Am. Chem. Soc.* 119, 7599-7600.
- (11) Rosen, M. K., Gardner, K. H., Willis, R. C., Parris, W. E., Pawson, T., Kay, L. E. (1996) Selective Methyl Group Protonation of Perdeuterated Proteins. *J. Mol. Biol.* 263, 627-636.
- (12) Goto, N. K., Gardner, K. H., Mueller, G. A., Willis, R. C., Kay, L. E. (1999) A robust and cost-effective method for the production of Val, Leu, Ile (δ^1) methyl-protonated ^{15}N -, ^{13}C -, ^3H -labeled proteins. *J. Biomol. NMR.* 13, 369-374.
- (13) Ayala, I., Sounier, R., Usé, 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.
- (14) Tugarinov, V., Kay, L. E. (2004) An isotope labeling strategy for methyl TROSY spectroscopy. *J. Biomol. NMR.* 28, 165-172.
- (15) Fischer, M., Kloiber, K., Häusler, J., Ledolter, K., Konrat, R., Schmid, W. (2007) Synthesis of a ^{13}C -Methyl-Group-Labeled Methionine Precursor as a Useful Tool for Simplifying Protein Structural Analysis by NMR Spectroscopy. *ChemBioChem.* 8, 610-612.
- (16) Velyvis, A., Ruschak, A. M., Kay, L. E. (2012) An Economical Method for Production of ^3H , $^{13}\text{CH}_3$ - Threonine for Solution NMR Studies of Large Protein Complexes: Application to the 670 kDa Proteasome. *PLoS One.* 7, e43725
- (17) Gans, P., Hamelin, O., Sounier, R., Ayala, I., Durá, M. A., Amero, C. D., Noirclerc-Savoie, M., Franzetti, B., Plevin, M. J., 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.
- (18) Vuister, G. W., Kim, S. J., Wu, C., Bax, A. (1994) 2D and 3D NMR-study of phenylalanine residues in proteins by reverse isotopic labeling. *J. Am. Chem. Soc.* 116, 9206-9210.
- (19) Aghazadeh, B., Zhu, K., Kubiseski, T. J., Liu, G. A., Pawson, T., Zheng, Y., Rosen, M. K. (1998) Structure and mutagenesis of the Dbl homology domain. *Nat. Struct. Biol.* 5, 1098-1107.
- (20) Kainosho, M., Torizawa, T., Iwashita, Y., Terauchi, T., Ono, A. K., Güntert, P. (2006) Optimal isotope labelling for NMR protein structure determinations. *Nature*, 440, 52-57.

- (21) Jäger, M., Zhang, Y., Bieschke, J., Nguyen, H., Dendle, M., Bowman, M. E., Noel, J. P., Gruebele, M., Kelly, J. W. (2006) Structure–function–folding relationship in a WW domain. *Proc. Natl. Acad. Sci. U. S. A.*, 103, 10648-10653.
- (22) Verdecia, M. A., Bowman, M. E., Lu, K. P., Hunter, T., Noel, J. P. (2000) Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nat. Struct. Biol.* 7, 639-643.
- (23) Kowalski, J. A., Liu, K., Kelly, J. W. (2002) NMR Solution Structure of the Isolated Apo Pin1 WW Domain: Comparison to the X-Ray Crystal Structures of Pin1. *Biopolymers.* 63, 111-121.
- (24) Wintjens, R., Wieruszkeski J. M., Drobecq, H., Rousselot-Pailley, P., Buée, L., Lippens, G., Landrieu, I. (2001) ¹H NMR Study on the Binding of Pin1 Trp-Trp Domain with Phosphothreonine Peptides. *J. Biol. Chem.* 276, 25150-25156.
- (25) Cai, F., Li, M. X., Pineda-Sanabria, S. E., Gelozia, S., Lindert, S., West, F., Sykes, B. D., Hwang, P. M. (2016) Structures reveal details of small molecule binding to cardiac troponin. *J. Mol. Cell. Cardiol.* 101, 134-144.
- (26) Zahran, S., Pan, J. S., Liu, P. B., Hwang, P. M. (2015) Combining a PagP fusion protein system with nickel ion-catalyzed cleavage to produce intrinsically disordered proteins in E. coli. *Protein Expression Purif.* 116, 133-138.
- (27) Neri, D., Szyperski, T., Otting, G., Senn, H., Wuethrich, K. (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 carbon-13 labeling *Biochemistry.* 28, 7510-7516.
- (28) Su, X. C., Loh, C. T., Qi, R., Otting, G. (2011) Suppression of isotope scrambling in cell-free protein synthesis by broadband inhibition of PLP enzymes for selective ¹⁵N-labelling and production of perdeuterated proteins in H₂O. *J. Biomol. NMR.* 50, 35-42.
- (29) Sauer, U., Eikmanns, B. J. (2005) The PEP–pyruvate–oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiol. Rev.* 29, 765-794.
- (30) Stiffin, R. M., Sullivan, S. M., Carlson, G. M., Holyoak, T. (2008) Differential Inhibition of Cytosolic PEPCCK by Substrate Analogues. Kinetic and Structural Characterization of Inhibitor Recognition. *Biochemistry.* 47, 2099-2109.
- (31) Yang, Z., Floyd, D. L., Loeber, G., Tong, L. (2000) Structure of a closed form of human malic enzyme and implications for catalytic mechanism. *Nat. Struct. Biol.* 7, 251-257.

- (32) Narindrasorasak, S., Bridger, W. A. (1978) Probes of the structure of phosphoenolpyruvate synthetase: Effects of a transition state analogue on enzyme conformation. *Can. J. Biochem.* 56, 816-819.
- (33) Kasperek, G. J., Pratt, R. F. (1977) The Fumarase Reaction A nuclear magnetic resonance experiment for biological chemistry students. *J. Chem. Educ.* 54, 515-516.
- (34) Thorn, M. B. (1953) Inhibition by Malonate of Succinic Dehydrogenase in Heart-muscle Preparations. *Biochem. J.* 54, 540-547.
- (35) Pardee, A. B., Potter, V. R. (1949) Malonate Inhibition of Oxidations in the Krebs Tricarboxylic Acid Cycle. *J. Biol. Chem.* 178, 241-250.
- (36) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR.* 6, 277-293.
- (37) Johnson, B. A., Blevins, R. A. (1994) NMR view: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR.* 4, 603-614.
- (38) Otten, R., Chu, B., Krewulak, K. D., Vogel, H. J., Mulder, F. A. (2010) Comprehensive and Cost-Effective NMR Spectroscopy of Methyl Groups in Large Proteins. *J. Am. Chem. Soc.* 132, 2952-2960.
- (39) Tugarinov, V., Kay, L. E. (2005) Methyl groups as probes of structure and dynamics in NMR studies of high-molecular-weight proteins. *ChemBiochem.* 6, 1567-1577.
- (40) Kerfa, R., Plevin, M. J., Sounier, R., Gans, P., Boisbouvier, J. (2015) Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. *Curr. Opin. Struct. Biol.* 32, 113-122.
- (41) Shekhtman, A., Ghose, R., Goger, M., Cowburn, D. (2002) NMR structure determination and investigation using a reduced proton (REDPRO) labeling strategy for proteins. *FEBS Lett.* 524, 177-182
- (42) Malaisse, W. J., Malaisse-Lagae, F., Liemans, V., Ottinger, R., Willem, R. (1990) Phosphoglucoisomerase-catalyzed interconversion of hexose phosphates: isotopic discrimination between hydrogen and deuterium. *Mol. Cell. Biochem.* 93, 153-165.
- (43) Golichowski, A., Harruff, R. C., Jenkins, W. T. (1977) The Effects of pH on the Rates of Isotope Exchange Catalyzed by Alanine Aminotransferase. *Arch. Biochem. Biophys.* 178, 459-467.

- (44) Cooper, A. J. L. (1976) Proton Magnetic Resonance Studies of Glutamate-Alanine Transaminase-catalyzed Deuterium Exchange. *J. Biol. Chem.* 251, 1088-1096.
- (45) Babu, U. M., Johnston, R. B. (1976) Nuclear magnetic resonance studies of D₂O-substrate exchange reactions catalyzed by glutamic pyruvic and glutamic oxaloacetic transaminases. *Biochemistry.* 15, 5671-5678.
- (46) Mildvan, A. S., Scrutton, M. C., Utter, M. F. (1966) Pyruvate carboxylase. VII. A possible role for tightly bound manganese. *J. Biol. Chem.* 241, 3488-3498.
- (47) Zeczycki, T. N., Maurice, M. S., Attwood, P. V. (2010) Inhibitors of Pyruvate Carboxylase *Open Enzym Inhib. J.* 3, 8-26.
- (48) Spaans, S. K., Weusthuis, R. A., Oost, J., Kengen, S. W. M. (2015) NADPH-generating systems in bacteria and archaea. *Front Microbiol.* 6, 742.
- (49) Hayashi, H., Inoue, K., Nagata, T., Kuramitsu, S., Kagamiyama, H. (1993) Escherichia coli Aromatic Amino Acid Aminotransferase: Characterization and Comparison with Aspartate Aminotransferase. *Biochemistry.* 32, 12229-12239.
- (50) Clore, G. M., Bax, A., Gronenborn, A. M. (1991) Stereospecific assignment of β -methylene protons in larger proteins using 3D ¹⁵N-separated Hartmann-Hahn and ¹³C-separated rotating frame Overhauser spectroscopy. *J. Biomol. NMR.* 1, 13-22.
- (51) Archer, S. J., Ikura, M., Torchia, D. A., Bax, A. (1991) An Alternative 3D NMR Technique for Correlating Backbone ¹⁵N with Side Chain H β Resonances in Larger Proteins. *J. Magn. Reson.* 95, 636-641.
- (52) Bax, A., Vuister, G. W., Grzesiek, S., Delaglio, F., Wang, A. C., Tschudin, R., Zhu, G. (1994) Measurement of Homo- and Heteronuclear *J* Couplings from Quantitative *J* Correlations *Methods Enzymol.* 239, 79-105.
- (53) Montelione, G. T., Winkler, M. E., Rauenbuehler, P., Wagner, G. (1989) Accurate Measurements of Long-Range Heteronuclear Coupling Constants from Homonuclear 2D NMR Spectra of Isotope-Enriched Proteins. *J. Magn. Reson.* 82, 198-204.
- (54) Grzesiek, S., Ikura, M., Clore, G. M., Gronenborn, A. M., Bax, A. (1992) A 3D Triple-Resonance NMR Technique for Qualitative Measurement of Carbonyl-H β *J* Couplings in Isotopically Enriched Protein. *J. Magn. Reson.* 96, 215-221.
- (55) Mittermaier, A., Kay, L. E. (2001) χ_1 Torsion Angle Dynamics in Proteins from Dipolar Couplings. *J. Am. Chem. Soc.* 123, 6892-6903.

(56) Li, F., Grishaev, A., Ying, J., Bax, A. (2015) Side Chain Conformational Distributions of a Small Protein Derived from Model-Free Analysis of a Large Set of Residual Dipolar Couplings. *J. Am. Chem. Soc.* 137, 14798-14811.



For Table of Contents Only.