

# Selective $^{13}\text{C}$ labeling of nucleotides for large RNA NMR spectroscopy using an *E. coli* strain disabled in the TCA cycle

Chandar S. Thakur · Jacob N. Sama ·  
Melantha E. Jackson · Bin Chen · T. Kwaku Dayie

Received: 20 May 2010 / Accepted: 28 July 2010 / Published online: 6 November 2010  
© The Author(s) 2010. This article is published with open access at Springerlink.com

**Abstract** *Escherichia coli* (*E. coli*) is an ideal organism to tailor-make labeled nucleotides for biophysical studies of RNA. Recently, we showed that adding labeled formate enhanced the isotopic enrichment at protonated carbon sites in nucleotides. In this paper, we show that growth of a mutant *E. coli* strain DL323 (lacking succinate and malate dehydrogenases) on  $^{13}\text{C}$ -2-glycerol and  $^{13}\text{C}$ -1,3-glycerol enables selective labeling at many useful sites for RNA NMR spectroscopy. For DL323 *E. coli* grown in  $^{13}\text{C}$ -2-glycerol without labeled formate, all the ribose carbon atoms are labeled except the C3' and C5' carbon positions. Consequently the C1', C2' and C4' positions remain singlet. In addition, only the pyrimidine base C6 atoms are substantially labeled to ~96% whereas the C2 and C8 atoms of purine are labeled to ~5%. Supplementing the growth media with  $^{13}\text{C}$ -formate increases the labeling at C8 to ~88%, but not C2. Not unexpectedly, addition of exogenous formate is unnecessary for attaining the high enrichment levels of ~88% for the C2 and C8 purine positions in a  $^{13}\text{C}$ -1,3-glycerol based growth. Furthermore, the ribose ring is labeled in all but the C4' carbon position, such that the C2' and C3' positions suffer from multiplet splitting but the C5' position remains singlet and the C1' position shows a small amount of residual C1'–C2' coupling. As expected, all the protonated base atoms, except C6, are labeled to ~90%. In addition, labeling with  $^{13}\text{C}$ -1,3-glycerol affords an isolated methylene ribose with high enrichment at the

C5' position (~90%) that makes it particularly attractive for NMR applications involving  $\text{CH}_2$ -TROSY modules without the need for decoupling the C4' carbon. To simulate the tumbling of large RNA molecules, perdeuterated glycerol was added to a mixture of the four nucleotides, and the methylene TROSY experiment recorded at various temperatures. Even under conditions of slow tumbling, all the expected carbon correlations were observed, which indicates this approach of using nucleotides obtained from DL323 *E. coli* will be applicable to high molecular weight RNA systems.

**Keywords** Alternate-site specific labeling ·  $\text{CH}_2$ -methylene TROSY · Ribose and nucleobase · RNA · Structure and dynamics

## Abbreviations

AMP	Adenosine 5'-monophosphate
CMP	Cytidine 5'-monophosphate
DHAP	Dihydroxyacetone phosphate
FBP	Fructose-6-bisphosphate
F6P	Fructose-6-phosphate
GA3P	Glyceraldehyde-3-phosphate
Gly	Glycine
GMP	Guanosine 5'-monophosphate
noPPP	Non-oxidative pentose phosphate pathway
OAA	Oxaloacetate
oPPP	Oxidative pentose phosphate pathway
R5P	Ribose-5-phosphate
rNMPs	Ribonucleoside monophosphates
rNTPs	Ribonucleoside triphosphates
Ser	Serine
TIM	Triosephosphate isomerase
UMP	Uridine 5'-monophosphate

C. S. Thakur · J. N. Sama · M. E. Jackson · B. Chen ·  
T. K. Dayie (✉)

Department of Chemistry and Biochemistry, Center for  
Biomolecular Structure & Organization, University of Maryland,  
1115 Biomolecular Sciences Bldg (#296), College Park,  
MD 20742-3360, USA  
e-mail: dayie@umd.edu

## Introduction

Uniform labeling technology ushered in an era of high resolution NMR studies of RNA structures and RNA-ligand interactions (Batey et al. 1995; Hall 1995; Pardi 1995; Fürtig et al. 2003; Latham et al. 2005; Dayie 2008; Lu et al. 2010). Milligram quantities of uniformly  $^{13}\text{C}/^{15}\text{N}$  labeled nucleotides can be extracted from bacteria grown on minimal media supplemented with  $^{15}\text{NH}_4\text{Cl}$  as the sole nitrogen source and on uniformly labeled glucose, or acetate or methanol as the carbon source (Batey et al. 1995; Hoffman and Holland 1995; Pardi 1995). RNAs transcribed using these labels inspired new multidimensional NMR experiments that enabled the assignment of chemical shifts to each nucleus in a polynucleotide chain, and made possible the routine structure determination and dynamics characterizations of RNAs 30 nucleotides or less (Latham et al. 2005; Dayie 2008; Lu et al. 2010).

Yet the uniform isotopic labeling methods introduced earlier do not overcome the inherent signal degeneracy problems of large RNAs. Instead, uniform labeling exacerbates the degeneracy problem and negates the benefits of monitoring RNA-ligand interactions, assignment of resonances and structural characterizations in the following fashion. As shown previously, for uniformly labeled nucleotides, the ribose  $\text{C}2'$  and  $\text{C}4'$  positions form a triplet that increases the spectral complexity by three-fold. Similarly, the  $\text{C}1'$  and  $\text{C}5'$  positions form doublets that increase two-fold the number of resonances in non-constant time HJSC spectra (Dayie and Thakur 2010). Uniform labeling also introduces considerable direct one-bond scalar couplings and residual dipolar couplings from adjacent carbons.

These unwanted splittings can be removed using either constant time experiments (Bax et al. 1979; Bax and Freeman 1981; Grzesiek and Bax 1992; van de Ven and Philippens 1992), adiabatic band selective decoupling schemes (Kupce and Wagner 1996; Brutscher et al. 2001; Dayie 2005), or maximum entropy reconstruction-deconstruction (Shimba et al. 2003). But each method has its attendant disadvantages. Constant time experiments limit the acquisition times to  $t_1^{\text{max}} = n/J_{\text{CC}}$  where  $n$  is an integer,  $J_{\text{CC}}$  is the homonuclear coupling constant (e.g.  $J_{\text{C}5\text{C}6} = 67$  Hz and  $J_{\text{C}1'\text{C}2'} = 43$  Hz). During the constant-time period,  $T$ , the decay of the transverse magnetization is proportional to  $\exp(-R_2T)$ , where  $R_2$  is the homogenous transverse rate constant. A compromise must therefore be found between sensitivity (small values of  $T$ ) and resolution (large values of  $T$ ): the long constant-time delays needed to improve resolution lead to significant signal attenuation for RNA molecules larger than 30 nucleotides (Dayie 2005). Additionally, band selective decoupling experiments preclude the observation of the regions to which the decoupling schemes are applied. Finally, the

maximum entropy reconstruction-deconstruction requires uniform coupling constants, but in RNA these couplings vary substantially (32–42 Hz in ribose ring).

Selective  $^{13}\text{C}$  labeling schemes can potentially remove most of these drawbacks. By selectively labeling each carbon site, the severe overlap of  $\text{C}2'$  with  $\text{C}3'$  is eliminated. In addition, the three-fold complexity created by the triplets formed by  $\text{C}2'$  and  $\text{C}4'$  carbons and the doublets formed by the  $\text{C}1'$  and  $\text{C}5'$  carbons are removed (Dayie and Thakur 2010). Similarly the direct one-bond scalar couplings and residual dipolar couplings from adjacent carbons introduced by uniform labeling is diminished. By this means the coupling topology is simplified, and signal sensitivity and spectral resolution are enhanced.

These favorable properties make selective  $^{13}\text{C}$  labeling attractive for diverse NMR applications. For example, straightforward measurements of various NMR parameters using standard multidimensional experiments can be envisioned. First, quantitative measurements of relaxation rates of  $^{13}\text{C}$ - $^1\text{H}$  spin pairs, previously complicated by the presence of  $^{13}\text{C}$ - $^{13}\text{C}$  dipolar interactions and  $^1J_{\text{CC}}$  couplings in uniformly labeled samples (Yamazaki et al. 1994; Dayie et al. 2002; Boisbouvier et al. 2003), can now be carried out (Johnson and Hoogstraten 2008) even for large RNA molecules. Second, direct carbon detection experiments suffer from reduction in signal intensity during acquisition because of the evolution of  $^{13}\text{C}$ - $^{13}\text{C}$  scalar couplings (Bermel et al. 2006; Fiala and Sklenár 2007; Farès et al. 2007). Using selectively labeled samples, these experiments can now be run without complicated pulse programs, or complicated data processing, or both. Third, a number of previously developed methods for measuring residual dipolar couplings and chemical shift anisotropy resulting from weak alignment (Boisbouvier et al. 2004; Miclet et al. 2004; Ying et al. 2006) will benefit considerably using the selective labels. Again straightforward measurements of these couplings can be undertaken without recourse to using filtering elements to remove unwanted  $^{13}\text{C}$ - $^{13}\text{C}$  J couplings. Finally transverse relaxation optimized spectroscopy (TROSY) studies promise to extend the size of macromolecules that can be studied by NMR (Pervushin et al. 1997). Here again selective labeling can provide a clean and simple alternative for recording TROSY spectra without obliterating the information content of decoupled regions. For example, we expect a two-fold improvement in resolution in the  $\text{C}5'$  region using methylene TROSY experiments without the need for a constant time evolution and band-selective decoupling, and a two-fold improvement in resolution in the  $\text{C}1'$  region without sacrificing the information content in the  $\text{C}2'$  region using non-constant time methine TROSY experiments that dispense with band-selective decoupling.

While alternate  $^{13}\text{C}$ – $^{12}\text{C}$  labeling schemes have been shown to be important for tackling large protein structural analysis (Religa and Kay 2010; Takeuchi et al. 2010), fewer such studies have been carried out for RNA molecules (Johnson et al. 2006; Hoogstraten and Johnson 2008; Dayie and Thakur 2010) despite some of the potential applications of selective labels enumerated above. For RNA analysis, specific-labeled nucleoside triphosphates (NTPs) are readily synthesized using wildtype and mutant *E. coli* bacteria as attractive molecular factories. For example, the growth of *E. coli* K10-1516 strain (glucose-6-phosphate dehydrogenase enzyme (*zwf* gene) is disabled, hereafter referred to as K10*zwf*) on  $^{13}\text{C}$ -2-glycerol provides labeling mostly at the C2' and C4' ribose positions, and at both C5 and C6 positions of pyrimidines (Johnson et al. 2006; Hoogstraten and Johnson 2008; Dayie and Thakur 2010). Similarly, growth of an *E. coli* strain deficient in tricarboxylic acid cycle enzymes malate and succinate dehydrogenases (LeMaster and Kushlan 1996) (DL323) on  $^{13}\text{C}$ -1,3-glycerol yields  $^{13}\text{C}$  label at the C5 position of pyrimidines, and growth on  $^{13}\text{C}$ -2-glycerol yields  $^{13}\text{C}$  label at the C6 position of pyrimidines (Johnson et al. 2006). We recently demonstrated that spiking the growth medium with  $^{13}\text{C}$  labeled formate increases the number of sites selectively enriched in K10*zwf* (Dayie and Thakur 2010). We observed that labels produced by K10*zwf* at C1' and C5' are  $\sim 7$  and  $\sim 15\%$ , respectively, whereas the fully enzymatic method can yield  $>90\%$  at those sites. However the fully enzymatic method is limited to labeling each ribose position using site specific-labeled glucose that also requires the coupling of the base to the labeled sugar component, a costly undertaking (Tolbert and Williamson 1996; Schultheisz et al. 2008).

Here we explored the potential to preserve a higher level of label at these two important ribose positions and the protonated base sites while retaining the advantage of having an isolated two spin system that will limit the magnetization transfers to desired pathways. The mutant *E. coli* DL323 strain has the two tricarboxylic acid cycle enzymes succinate and malate dehydrogenases disabled, making it an alternative strain that could compensate for the drawbacks of the K10*zwf* strain and the enzymatic method. We therefore extended those earlier studies to DL323 and present an analysis of the isotopic enrichment using labeled glycerol and formate in DL323. As expected, we find that it is not necessary to spike the  $^{13}\text{C}$ -1,3-glycerol growth media with  $^{13}\text{C}$ -labeled formate to label the C2 and C8 purine sites. These labeling patterns contrast with the ones obtained using  $^{13}\text{C}$ -2-glycerol wherein exogenous labeled formate is necessary to label the C2 and C8 purine positions. Moreover to explore the usefulness of these site specific labels under conditions of broadened resonances of large RNAs, we dissolved the labeled nucleotides in

perdeuterated glycerol and show that  $\text{CH}_2$ -methylene TROSY gives the expected carbon correlations, making this approach likely applicable to high molecular weight RNAs. Obtaining a high degree of labeling at ribose C1', C2', C5' and purine C2 and C8 positions increases the scope of the isotope labeling patterns that are useful for structural, functional and dynamics studies of important RNA macromolecules.

## Materials and methods

### Bacterial strains

The mutant strain DL323 (CGSC # 7538, F-, *sdh-1*,  $\lambda^-$ , *mdh-2*, *rph-1*) used in this work was obtained from the Yale Coli Genetic Stock Center.

### Isotopes

The following isotopically labeled compounds were purchased from Cambridge Isotope Laboratory (Andover, MA) and Isotec-Sigma-Aldrich (Miamisburg, OH):  $^{13}\text{C}$ -sodium formate (99%),  $^{13}\text{C}$ -1,3-glycerol (99%),  $^{15}\text{N}$ - $(\text{NH}_4)_2\text{SO}_4$  (99%), and  $^{13}\text{C}$ -2-glycerol (99%).

### Stock solutions

The stock solutions required for bacterial growth were prepared using distilled and deionized water heat sterilized. The Studier media (SPG) contains 25 mM  $(\text{NH}_4)_2\text{SO}_4$ , 50 mM  $\text{KH}_2\text{PO}_4$ , 50 mM  $\text{Na}_2\text{HPO}_4$ , 2 mM  $\text{MgSO}_4$  and trace metal solution, supplemented with either labeled glycerol or formate and glycerol (Studier 2005). The LeMaster-Richards (LMR) media contains 176 mM  $\text{KH}_2\text{PO}_4$ , 25 mM  $\text{NaOH}$ , 10  $\mu\text{l}$   $\text{H}_2\text{SO}_4$ , 12.6 mM  $(\text{NH}_4)_2\text{SO}_4$ , 2 mM  $\text{MgSO}_4$ , 10  $\mu\text{M}$   $\text{FeSO}_4$  and 0.2% trace metals (LeMaster and Richards 1982).

### Media for bacterial growth

Luria–Bertani (LB), LMR and SPG minimal media were prepared as described (Sambrook and Russell 2001; Studier 2005).

### Growth optimization procedure

The optimization of mutant DL323 *E. coli* growth was performed for high yields of biomass per input gram of carbon source (Thakur et al. 2010). Briefly, 5 ml starter cultures in unlabeled SPG or LMR medium were inoculated from a single colony of DL323 grown on LB plates without antibiotics incubated at 37°C overnight. The next

morning, cells were washed twice in  $1\times$  phosphate-buffered saline (PBS) and re-suspended in 5 ml of fresh SPG medium without carbon source; 1 ml from this re-suspension was added to 50 ml of SPG medium and incubated at 37°C. After 12–14 h of incubation the cells were washed twice in  $1\times$  PBS and resuspended in 50 ml SPG medium with no carbon source, 5 ml from this resuspension were added to 500 ml of SPG medium supplemented with labeled carbon sources and incubated at 37°C for 12 h. Growth in LMR was carried out identically as described above for SPG.

#### Nucleic acid digestion and boronate column purification

The cell pellet from the 500 ml culture was resuspended in 20 ml lysis buffer (STE buffer: 0.1 M NaCl, 10 mM Tris, 1 mM EDTA, pH 8 and 0.5% SDS). The nucleic acids were separated from the cellular proteins using a 25:24:1 phenol:chloroform:isoamyl alcohol mixture. The nucleic acids present in the upper aqueous layer were extracted multiple times with an equal volume of a 24:1 chloroform:isoamyl alcohol mixture, then the nucleic acids were precipitated overnight at  $-20^{\circ}\text{C}$  with ethanol and acetate buffer pH 5.2. The total nucleic acids were subjected to nuclease P1 digestion in 15 mM sodium acetate, 0.1 mM  $\text{ZnSO}_4$ , pH 5.2, the mixture was lyophilized and re-suspended in 10 ml of 1 M triethylammonium bicarbonate (TEABC). Later, the mixture was applied to a cis-diol boronate affinity column to separate deoxyribonucleotides from ribonucleotides. The boronate resin was hydrated in 10 mM Tris HCl, pH 8, rinsed with acidified water, pH 4.5, and equilibrated with 1 M triethylammonium bicarbonate (TEABC), pH 9.5, prior to the separation of the nucleotides. The nucleotide mixture was applied to the column, unbound deoxyribonucleotides and salts were washed with five column volumes of 1 M TEABC, and the unbound deoxyribonucleotides were collected and pooled. The bound ribonucleotides were eluted with acidified water, pH 4.5, and the purity of both dNMPs and rNMPs were checked using NMR.

#### Separation and desalting of rNMPs by reversed-phase FPLC

The lyophilized rNMPs were dissolved in a minimal amount of solution A (0.1 M monobasic potassium phosphate prepared in distilled water) and passed through a 0.22- $\mu\text{m}$  syringe membrane filter (Fisherbrand) prior to injection on a Targa C18 column (250  $\times$  20 mm, 10  $\mu\text{m}$ , 120 Å, Higgins Analytical, Inc.) using RP-FPLC on an ÄKTA Design<sup>TM</sup> system with a detector at a wavelength of

254 nm. Two segment linear gradients of solution B (0.1 M monobasic potassium phosphate, 20%  $\text{CH}_3\text{CN}$ , prepared in distilled water) were employed with a flow rate of 10 ml/min and gradients of 0.3 and 10% of buffer B per minute. The sample purity was examined with an analytical RP-FPLC on a Targa C18 column (250  $\times$  4.6 mm, 5  $\mu\text{m}$ , Higgins Analytical, Inc.) with a flow rate of 0.8 ml/min. The purified NMPs were desalted prior to phosphorylation using the same preparative column washed with solution C (0.05% aqueous trifluoroacetic acid (TFA)) and a gradient of 5% solution D (20%  $\text{CH}_3\text{CN}$ , 0.05% TFA) per minute with a flow rate of 10 ml/min.

#### NMR experiments

NMR experiments were run on a four channel Bruker Avance III 600 MHz spectrometer equipped with actively shielded z-axis gradient triple resonance probe at temperatures of 15–45°C. The NMR data sets were processed and the peak positions and intensities were analyzed with Bruker's TOPSPIN 2.1 as described previously (Dayie and Thakur 2010). One dimensional (1D)  $^{13}\text{C}$  spectra and two-dimensional non-constant-time ( $^1\text{H}$ ,  $^{13}\text{C}$ ) heteronuclear single quantum correlation (HSQC) spectra (Bodenhausen and Ruben 1980; Bax et al. 1990) were acquired to analyze the rNMP fractions from the DL323 bacterial strain. To ascertain the labeling pattern of carbon in both the ribose and the base rings, the fractional  $^{13}\text{C}$  enrichment at each carbon site was quantified directly by 1D proton methods or indirectly using 2-bond ( $^2\text{J}_{\text{HN}}$ ) HSQC as described previously (Dayie and Thakur 2010). In addition, the fractional  $^{13}\text{C}$  enrichment at each carbon site was quantified by comparing the labeling pattern and intensity of resonances in proton decoupled direct carbon spectra of uniformly  $^{13}\text{C}/^{15}\text{N}$  rNMP from commercial sources and rNMPs produced with DL323 *E. coli*. The samples were matched in concentration. However to account for slight differences in sample concentrations, the intensities of each of the spectra were normalized by the average peak intensities in the  $^2\text{J}_{\text{HN}}$   $^1\text{HN}-^{15}\text{N}$  HSQC data sets because all nucleotides are uniformly  $^{15}\text{N}$ -labeled. The fractional enrichment is defined as  $\text{Frac\_E}_i = (I_{s,i}/I_{u,i})(N_u/N_s)$ , where  $I_s$  and  $I_u$  are the peak intensities of the  $i$ th resonance in the direct 1D or 2D HSQC carbon spectra of the selectively and uniformly labeled rNMPs that were recorded identically,  $N_s$  and  $N_u$  are normalization factors derived from the average peak intensities in the 2D  $^2\text{J}_{\text{HN}}$   $^1\text{HN}-^{15}\text{N}$  HSQC data sets of the different samples (Lundström et al. 2007).  $\text{CH}_2$ -TROSY experiments (Miclet et al. 2004) were run with the following slight modifications: the WURST-4 decoupling waveform was not applied during the carbon  $t_1$  evolution period, and the selective pulse on the C5' carbon was replaced with a non-selective 180° pulse.

## Results and discussion

### Incorporation of $^{13}\text{C}$ into ribose ring of nucleotides using the pentose phosphate pathway (PPP)

To put our results in the context of the *E. coli* mutant DL323 grown in labeled glycerol with or without labeled formate, we present a brief overview of the metabolic pathways for nucleotide metabolism via glycolysis, gluconeogenesis, and the Krebs cycle. Details can be found in standard textbooks (Nelson and Cox 2008; Voet et al. 2008). The ribose moiety of the ribonucleotide is derived exclusively from ribose 5-phosphate (R5P) produced via glucose-6-phosphate (G6P) directly from gluconeogenesis using glycerol (Fig. 1). Because both the oxidative PPP (oPPP) and the non-oxidative PPP (noPPP) are operative in DL323 *E. coli*, the partitioning of the carbon flux leads to scrambled labels in the ribose ring. For instance, placement of the  $^{13}\text{C}$ -label at the central C-2 carbon of glycerol leads to the following labeling pattern. When the oPPP is operative, fructose 6-phosphate (F6P), derived from dihydroxyacetone phosphate (DHAP) and glyceraldehyde 3-phosphate (GA3P), is labeled exclusively at C2 and C5 (Fig. 1). In turn, G6P derived from F6P loses its C1 carbon as carbon dioxide to form R5P. Thus under oPPP conditions, this R5P is labeled exclusively at the C1 and C4 positions (Fig. 1). In contrast, when the noPPP is operative, three reversible reactions lead to three molecules of R5P with different combinations of carbon labels. In the first reverse reaction, transketolase (TKL) catalyzes the transfer of a two carbon fragment from the ketose donor F6P to the aldose acceptor GA3P to form xylulose 5-phosphate (X5P) and erythrose 4-phosphate (E4P). This X5P, eventually isomerized to R5P, is labeled at only the C2 and C4 positions. In the second reverse reaction, transaldolase (TAL) removes a three carbon fragment from F6P and condenses it with E4P to form sedoheptulose 7-phosphate (S7P) and GA3P. In the third reverse reaction, transketolase again transfers a two carbon fragment from S7P to GA3P to form X5P and R5P. In this case, X5P is labeled at C2 and C4 whereas R5P is labeled at only C4. The predicted net effect is a 2:1 labeling pattern of [2, 4- $^{13}\text{C}_2$ ]ribose and [4- $^{13}\text{C}$ ]ribose, and no label at the 3 or 5 positions via the noPPP. The oPPP will lead to a [1, 4- $^{13}\text{C}_2$ ] ribose label (Fig. 1) and unlabeled carbon dioxide. A similar analysis indicates that placement of the  $^{13}\text{C}$ -label at the terminal C-1,3 carbons of glycerol leads to the following labeling pattern. The reverse noPPP leads a 2:1 labeling pattern of [1, 3, 5- $^{13}\text{C}_3$ ]ribose and [1, 2, 3, 5- $^{13}\text{C}_4$ ]ribose with no label at the 4 position. The oPPP will lead to a [2, 3, 5- $^{13}\text{C}_3$ ] ribose label, but no label at the 4-position (Fig. 1). Thus one expects coupling between ribose C1' and C2', and between C2' and C1' and C3'.

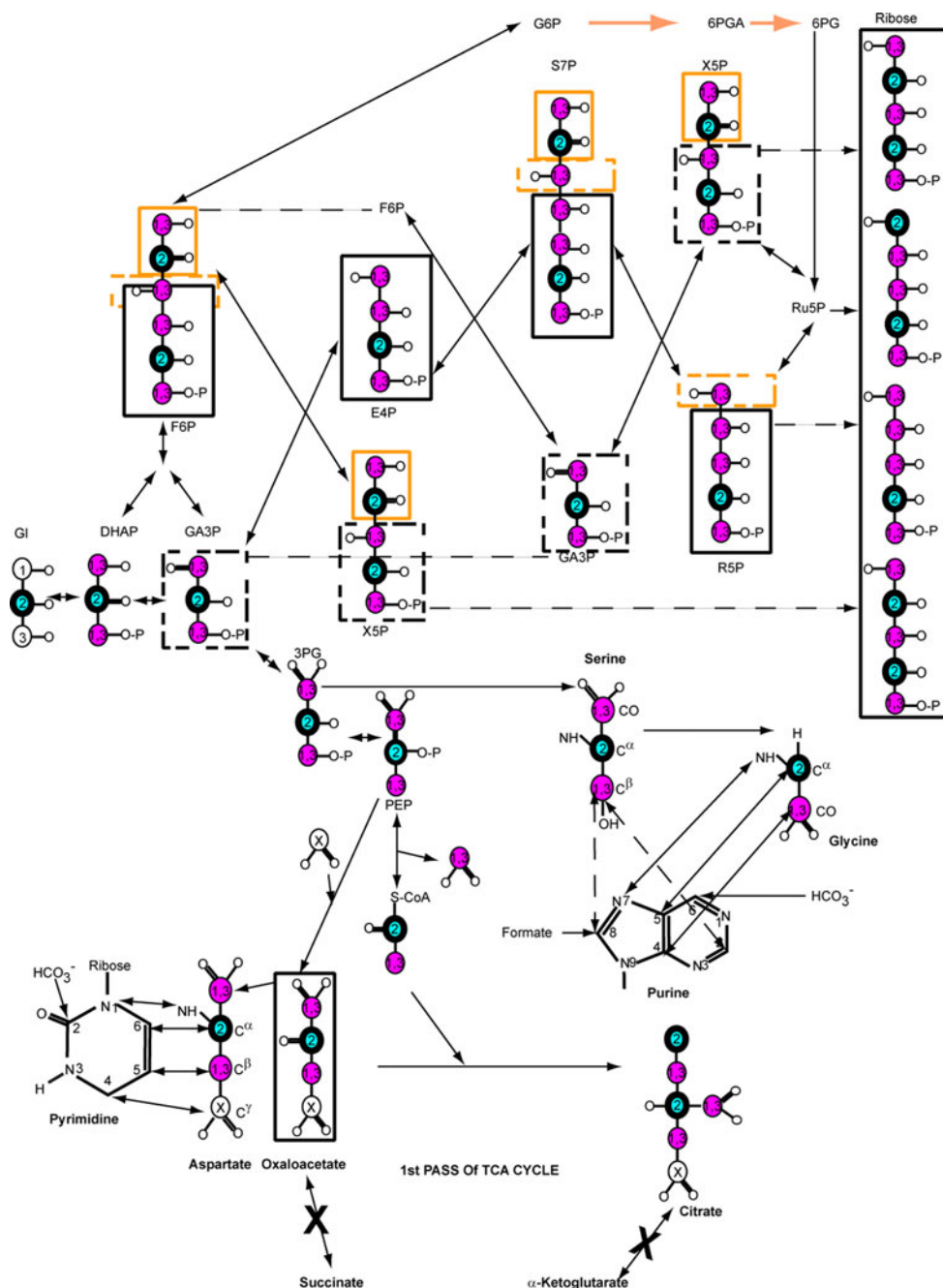
### Incorporation of $^{13}\text{C}$ into base ring of nucleotides via the glycolysis, gluconeogenesis and the Krebs cycle

The main advantage of the DL323 strain is that the flux through the TCA cycle is reduced to zero thereby preventing the dilution of the labels arising from oxaloacetate (LeMaster and Kushlan 1996). The nucleotide bases are synthesized from the precursors of 3PG that give rise to glycine (Gly) and serine (Ser) and from oxaloacetate that gives rise to aspartic acid (Asp). The amide group, the C $^{\alpha}$ -, C $^{\beta}$ - and C $^{\gamma}$ -carbon positions of Asp become the N1, C6, C5, and C4 ring atoms, respectively of pyrimidines (Fig. 1). The bicarbonate pool arises from several sources: the decarboxylation of PEP to form acetyl-CoA, carbon 1 of G6P and carbon 1 of pyruvate (derived from carbon 1 or 3 of glycerol). The formate pool also arises from the C $^{\beta}$ H $_2$ OH of Ser (derived from carbon 1 or 3 of glycerol). In addition, exogenous formate could enter the metabolic cycle by exchanging the carboxyl group of pyruvate by consuming acetyl-CoA (Thauer et al. 1972; Knappe et al. 1974) possibly by the reversible action of pyruvate formate lyase (Kirkpatrick et al. 2001) (Fig. 1). The purine ring atoms C2 and C8 also derive from the formate pool described above. Thus the purine C2 and C8 sites are expected to be labeled with  $^{13}\text{C}$  using  $^{13}\text{C}$ -1,3-glycerol. If  $^{13}\text{C}$ -1,3-glycerol is used addition of exogenous formate should make no difference to labeling at the purine C2 and C8 positions.

For DL323 *E. coli* grown in  $^{13}\text{C}$ -1,3-glycerol without labeled formate, the ribose ring is labeled in all but the C4' carbon position (Fig. 2) in agreement with metabolic analysis (Fig. 1) and previous work (Johnson et al. 2006). As a result, the C2' and C3' positions suffer from splitting by carbon-carbon coupling but the C5' position remains singlet (Fig. 3); the C1' position also remains substantially singlet with some admixture of residual C1'–C2' coupling. These latter contributions arise from the reverse noPPP (Fig. 2c); at most, 20% of the labeled C1' molecules are expected to have an adjacent C2' label. As expected, all but the C6 base atoms (C5 carbon position of pyrimidine, the C2 and C8 atoms of purine) are substantially labeled to ~90%. No increased labeling is observed in the ribose region on addition of labeled formate to the  $^{13}\text{C}$ -1, 3-glycerol media. Addition of labeled formate, however, increases the labeling of the C6 carbon positions of the Pyr ring from <5 to ~23% (Table 1).

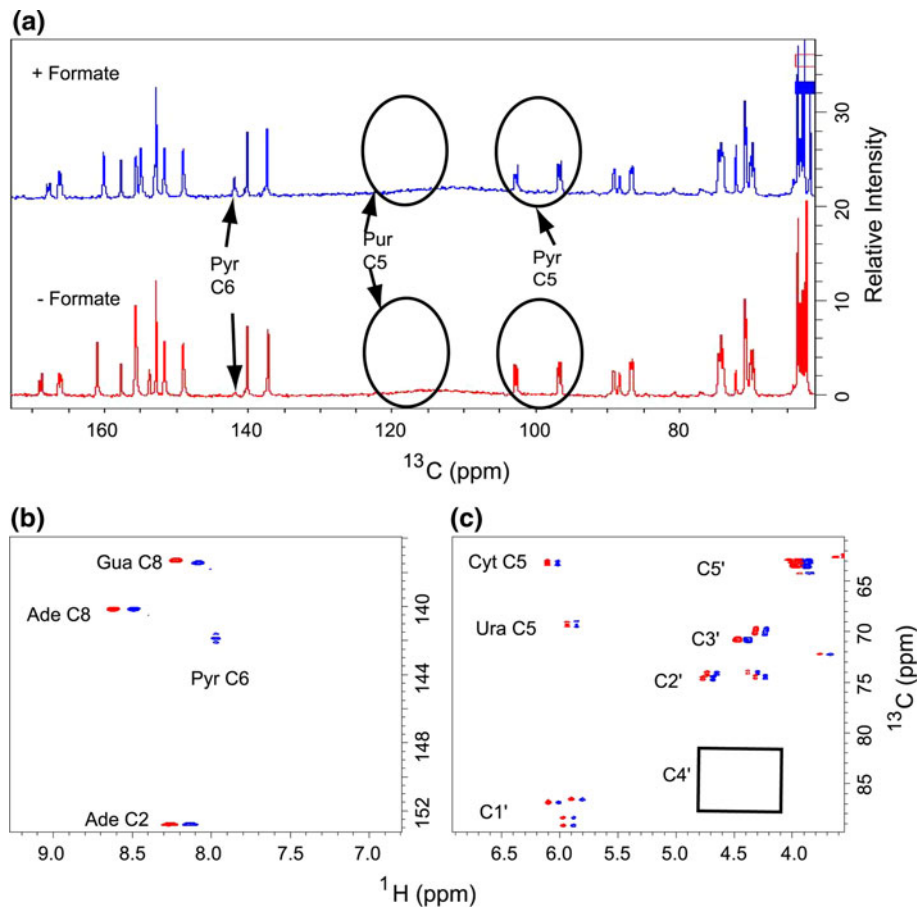
For DL323 *E. coli* grown in  $^{13}\text{C}$ -2-glycerol without labeled formate, the ribose ring is labeled in all but the C3' and C5' carbon positions (Fig. 3) in agreement with metabolic analysis (Fig. 1) and previous work (Johnson et al. 2006). As a result, the C2' and C4' positions are not split by the carbon-carbon coupling observed in strain DL323 grown in  $^{13}\text{C}$ -1,3-glycerol. The C1' positions also remain





**Fig. 1** Major metabolic pathways involved in the production of nucleic acid nucleotides from glycerol, including key steps in glycolysis, gluconeogenesis and one pass through the tricarboxylic (TCA) cycle. With the *E. coli* strain lacking succinate and malate dehydrogenase (DL323), the oxidative branch of the pentose phosphate pathway remains intact but the TCA cycle is severed in two places such that the oxaloacetate is derived exclusively from carboxylation of PEP and the resulting label is not diluted by the TCA cycle. Atom labels for the terminal (1,3) carbons (magenta and thin circle) and central (2) carbon (cyan and thick circle) of glycerol are highlighted. Positions that are enriched due to the presence of  $^{13}\text{CO}_2$  in the growth medium are shown with an X. Pyrimidine base derived from the Oxaloacetate (OAA) produced by carboxylation of

phosphoenolpyruvate (PEP) is shown via the aspartate intermediate. This OAA cannot be used as a substrate in the first and subsequent rounds of the TCA cycle because of the two mutations. Consequently OAA derived aspartate amino acid can be produced with  $^{13}\text{C}$  labeling at only the  $\text{C}^\alpha$  position if  $^{13}\text{C}$ -2-glycerol is used. If  $^{13}\text{C}$ -1,3-glycerol is used only  $\text{C}^\beta$  position is labeled. In either case carboxylation of PEP leads to labeling of the  $\text{C}^\gamma$  position. Similarly the labeling pattern of purines from glycine derived from 3PG are shown such that if  $^{13}\text{C}$ -2-glycerol is used only the  $\text{C}^\alpha$  position of Gly and therefore C5 position of the purine ring is labeled. Otherwise if  $^{13}\text{C}$ -1,3-glycerol is used the  $\text{CO}$  of Gly and therefore C4 of purine ring is labeled, and the  $\text{C}^\beta$  position of Ser leads to labeling of the purine C2 and C8 positions



**Fig. 2** Labeling pattern of a mixture of four rNMPs isolated from DL323 *E. coli* strain grown without (*red*) and with (*blue*)  $^{13}\text{C}$ -formate in a  $^{13}\text{C}$ -1,3-glycerol background. **a** Direct carbon detection 1D spectrum showing all the labeled carbon positions for nucleotides labeled with  $^{13}\text{C}$ -1,3 -glycerol and no formate (*bottom, red*) or with formate (*top, blue*). A long recycle delay of 5 s was used to allow for sufficient magnetization recovery and proton decoupling was limited to the acquisition period only. The level of enrichment at the adenine (Ade) and guanine (Gua) C8 positions remain at the same high level but that of Pyr C6 increases only on addition of  $^{13}\text{C}$ -labeled formate. Slight differences in pH and salt conditions between samples leads a

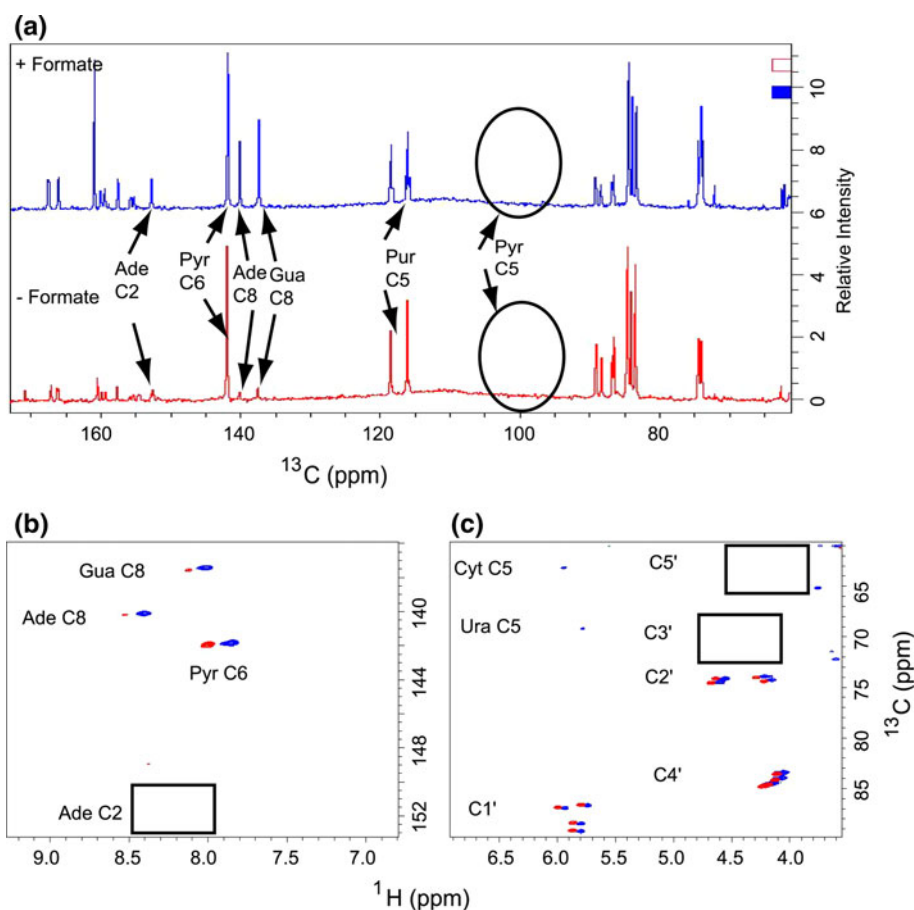
noticeable shift in the chemical positions for peaks downfield of Pyr C6. **b** 2D non-constant time HSQC spectrum of a mixture of all four labeled rNMPs showing the protonated base region. For ease of comparison the spectrum obtained without labeled formate (*red contours*) are displaced to the left of the formate labeled spectrum (*blue contours*). The level of enrichment at the Pyr C6 increases slightly by spiking with  $^{13}\text{C}$ -labeled formate. **c** 2D non-constant time HSQC spectrum of a mixture of all four labeled nucleotides showing the ribose region. The cytosine (Cyt) and Uracil (Ura) C5 resonances at 96.67 ppm and 102.69 ppm respectively are folded into the spectrum. The C4' region is *boxed* to highlight the absence of labeling

singlet (Fig. 3), presumably the labeled C2' arises in a different molecule (Fig. 1). As expected, only the pyrimidine C6 base atoms are substantially labeled to  $\sim 96\%$ . The C2 and C8 atoms of purine are labeled to  $\sim 5\%$ , comparable to that observed for K10zwf *E. coli* grown in  $^{13}\text{C}$ -2-glycerol without labeled formate (Dayie and Thakur 2010).

In contrast to K10zwf *E. coli* grown in  $^{13}\text{C}$ -2-glycerol with labeled formate (Dayie and Thakur 2010), addition of labeled formate to the  $^{13}\text{C}$ -2-glycerol media does not lead to new labels at the C3' and C5' positions (Fig. 3c). Similar to K10zwf *E. coli*, addition of labeled formate to  $^{13}\text{C}$ -2-glycerol leads to increased labeling of the C8 carbon positions of the purine ring to  $\sim 88\%$  (Fig. 3). Surprisingly C2 remains labeled at low levels unlike the case seen in the K10zwf strain (Dayie and Thakur 2010).

#### Estimating the level of carbon labeling efficiency

The level of  $^{13}\text{C}$  labeling efficiency is easily estimated using 1D  $^1\text{H}$  or natural abundance  $^{13}\text{C}$  carbon spectra. However for cases of overlap, two other methods can be used. The first method exploits long range proton-nitrogen correlations in  $^{15}\text{N}$ -HSQC spectra to estimate the labeling efficiency of the purine C2 and C8 carbon sites, the pyrimidine C5 and C6 carbon sites and the purine C1' carbon site (Fig. 4). The relaxation properties and transfer efficiencies are different for long range and one-bond magnetization transfers; nonetheless, the use of the long range  $^{15}\text{N}$ -HSQC method has been shown to be useful for estimating the level of  $^{13}\text{C}$  incorporation (Dayie and Thakur 2010). The labeling efficiency is estimated by



**Fig. 3** Labeling pattern of a mixture of four rNMPs isolated from DL323 *E. coli* grown without (*red*) and with (*blue*)  $^{13}\text{C}$ -formate in a  $^{13}\text{C}$ -2-glycerol background. **a** Direct carbon detection 1D spectrum showing all the labeled carbon positions for nucleotides labeled with  $^{13}\text{C}$ -2-glycerol and no formate (*bottom, red*) or with formate (*top, blue*). A long recycle delay of 5 s was used to allow for sufficient magnetization recovery and proton decoupling was limited to the acquisition period only. Spiking the growth media with  $^{13}\text{C}$ -labeled formate increases the level of enrichment at the adenine (Ade) and guanine (Gua) C8 positions but that of Pyr C6 remains the same. **b** 2D non-constant time HSQC spectra of a mixture of all four labeled

rNMPs showing the protonated base region. For ease of comparison the spectrum obtained without labeled formate (*red contours*) is displaced to the left of the formate labeled spectrum (*blue contours*). Again the level of enrichment at the Pyr C6 does not change with spiking with  $^{13}\text{C}$ -labeled formate whereas that of Pur C8 increases from  $\sim 17$  to 88%. **c** 2D non-constant time HSQC spectra of a mixture of all four labeled nucleotides showing the ribose region. The cytosine (Cyt) and Uracil (Ura) C5 resonances at 96.67 and 102.69 ppm, respectively are folded into the spectrum. The C3' and C5' regions are *boxed* to highlight the absence of labeling

comparing the intensity of each  $^{13}\text{C}$  satellite peak to the intensity of the center peak. For a mixture of the four rNMPs extracted from DL323 *E. coli* grown in  $^{13}\text{C}$ -1,3-glycerol with or without labeled formate, there is significant overlap in both the base and ribose regions even for an isolated rUMP (Fig. 2). Analysis of the 2D  $^2\text{J}_{\text{HN}}$  HSQC spectra (Fig. 4) allows a straightforward estimate of the labeling at purine C2 and C8 and pyrimidine C6 (Fig. 4). For site specific- $^{13}\text{C}/^{15}\text{N}$ -labeled rNMP, the 2D method is in excellent agreement with the 1D  $^1\text{H}$  method (Fig. 4). In the absence of formate, the level of enrichment was  $\sim 90\%$  for the purine C8 and C2. In the presence of formate the level of enrichment remains at  $\sim 90\%$  for purine C8 and  $\sim 90\%$  for Ade C2 (Table 1). As expected, each of the H2 and H8 proton resonance is split into a doublet with little

central peak in the acquisition dimension (Fig. 4a). In contrast, each H6 proton resonance gives rise to a singlet suggestive of almost no label at the pyrimidine C6 position (Fig. 4a); each H5 proton resonance shows a doublet suggestive of almost complete labeling (not shown). The C1' and C5' regions were sufficiently isolated to enable an estimate of the labeling efficiencies using the 1D  $^1\text{H}$  NMR spectrum: the enrichment level was  $\sim 58\%$  for the ribose C1' and 90% for the ribose C5'.

Similarly, analyses of the multiplet pattern of the four labeled nucleotides derived from DL323 *E. coli* grown in  $^{13}\text{C}$ -2-glycerol without formate show the opposite labeling pattern in the purine C2/C8 and pyrimidine C5/C6 regions. As expected, the H2 proton resonance gives rise to a singlet (at N1 and N3) suggestive of little labeling at the purine C2



**Table 1**  $^{13}\text{C}$  enrichment levels at various carbon positions within ribonucleotides harvested from DL323 *E. coli* grown on  $^{13}\text{C}$ -1,3-glycerol with and without  $^{13}\text{C}$ -formate as carbon sources

Carbon position labeled	$^{13}\text{C}$ -Carbon source: 1,3-Glycerol only	$^{13}\text{C}$ -Carbon source: 1,3-Glycerol and Formate
Purine <sup>a</sup>		
Ade C2	0.88 ± 0.04	0.82 ± 0.13
C8	0.92 ± 0.07	0.81 ± 0.13 <sup>b</sup>
Pyrimidine <sup>a</sup>		
C5	0.88 ± 0.04	0.92 ± 0.03
C6	<0.05	0.22 ± 0.02
Ribose		
C1 <sup>a</sup>	0.58 ± 0.08	0.47 ± 0.11
C2 <sup>b</sup>	0.50 ± 0.04	0.57 ± 0.09
C3 <sup>b</sup>	0.93 ± 0.01	0.77 ± 0.07
C4 <sup>b</sup>	<0.02	<0.02
C5 <sup>c</sup>	0.94 ± 0.03	0.85 ± 0.06

<sup>a</sup> The percentage label is calculated as an average of three methods: (i) the ratio of the sum of the intensities of satellite peaks to the sum of the intensities of the satellite and center peaks using the 2-bond  $^{15}\text{N}$  HSQC without  $^{13}\text{C}$  decoupling during acquisition (Dayie and Thakur 2010); (ii) the ratio of the sum of the intensities of satellite peaks to the sum of the intensities of the satellite and center peaks using the 1D  $^1\text{H}$  experiment without  $^{13}\text{C}$  decoupling during acquisition and (iii) using the fractional enrichment (Frac\_Ei) method of Lundström et al. (2007) as described in the text. We find that direct 1D and 2D methods (i and ii) give consistently slightly higher values than the fractional enrichment method (iii)

<sup>b</sup> The percentage label (Plabel) is calculated as in (a) but this time with only method (iii)

<sup>c</sup> The percentage label (Plabel) is calculated as in (a) but this time with only methods (ii) and (iii)

position (Fig. 4c, d). Each H6 proton resonance, unlike the case for growth in  $^{13}\text{C}$ -1,3-glycerol, gives rise to a doublet with little central peak in the acquisition dimension (Fig. 4a) indicating substantial labeling ( $\sim 98\%$ ) at the pyrimidine C6 position (Fig. 4c).

This method using the long range proton-nitrogen correlations in  $^{15}\text{N}$ -HSQC spectra is adequate for estimating the labeling efficiency in the aromatic ring (Tables 1, 2). The method is, however, limited to purine C2 and C8 carbon sites, the pyrimidine C5 and C6 carbon sites and the purine C1' carbon site. The second method of fractional enrichment (Lundström et al. 2007) gets around this limitation. However this fractional enrichment method cannot be carried out on a mixture of the four nucleotides obtained from boronate affinity purification. The four nucleotides must first be separated. Then each component can be analyzed accurately. Using this method we obtain enrichment values that are comparable to but consistently larger (by about 5–30%) than those obtained with the two previous methods. The values reported in Tables 1 and 2 are the averages of the three methods.

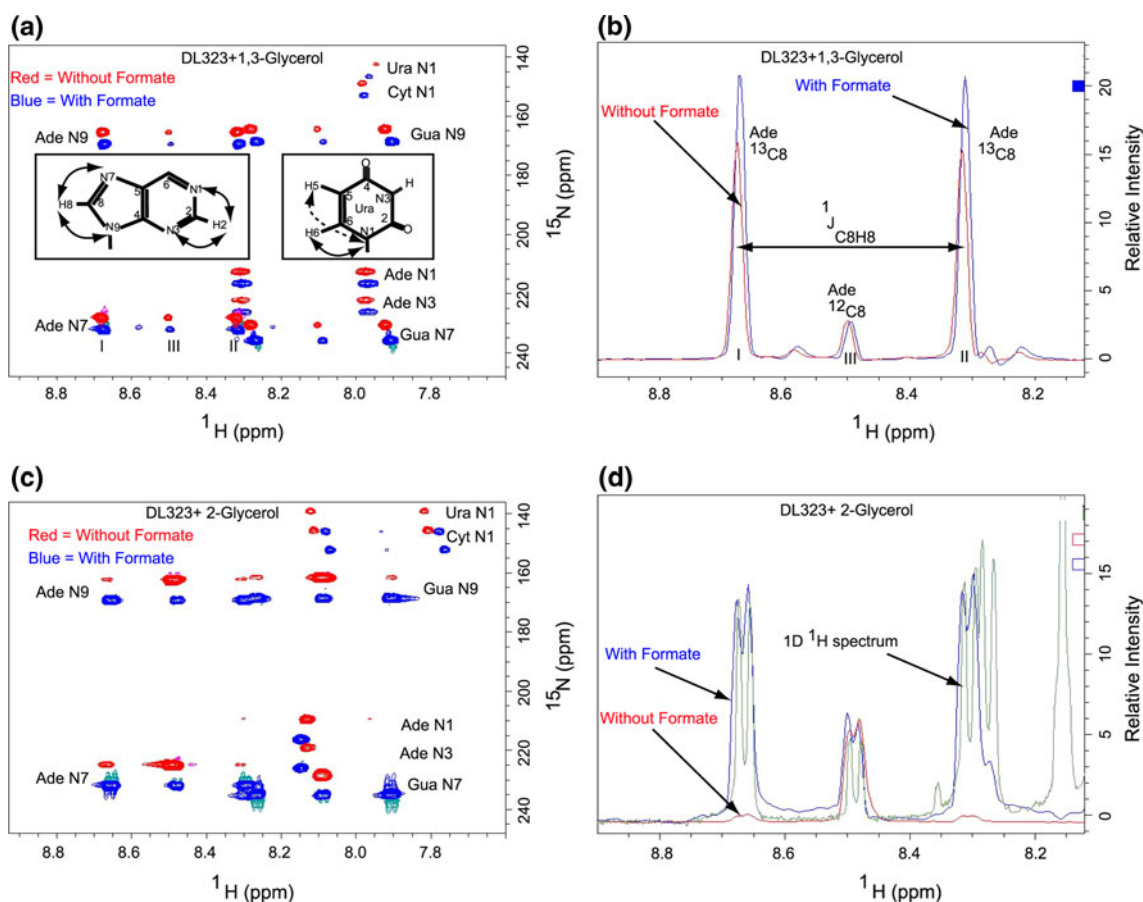
#### Applications of selective labels for $^{13}\text{C}$ NMR study of nucleic acids

An example of the high quality uncluttered spectra obtained using strain DL323 is shown in Fig. 5. The labeling pattern of rNMPs derived from DL323 grown in  $^{13}\text{C}$ -2-glycerol and formate is particularly attractive for

relaxation studies of not only the ribose C2' and C4' positions, as elegantly demonstrated recently by Hoogstraten and his colleagues (Johnson et al. 2006; Johnson and Hoogstraten 2008), but also the ribose C1' and the base C6 and C8 positions. Each of these five positions is essentially singlet. Their isolation from directly labeled adjacent neighbors means that the interference arising from strong  $^{13}\text{C}$ - $^{13}\text{C}$  magnetic interactions in the base and ribose rings is no longer an issue for extracting accurate relaxation parameters (Johnson et al. 2006; Johnson and Hoogstraten 2008).

The labeling pattern of rNMPs derived from DL323 grown in  $^{13}\text{C}$ -1,3-glycerol and formate is less attractive for relaxation studies of the ribose region because the C2' position is doublet, the C1' retains some residual doublet arising from  $^{13}\text{C}2'$ - $^{13}\text{C}1'$  isotopomers, and the C4' has no label (Fig. 5a, c). This is consistent with the metabolic analysis that predicts that at most 20% of the molecules have both C1' and C2' labeled, whereas 75% of the molecules have either C2' and C1' labeled or C2' and C3' labeled (Fig. 1). In principle, measurements on isolated  $^{13}\text{C}1'$  spins can be performed so long as suitable purging pulse sequences are employed to remove  $^{13}\text{C}2'$ - $^{13}\text{C}1'$  pairs. In the base region only the purine C2 and C8 sites are available for clean relaxation measurements; the pyrimidine C5 site has multiplet structure precluding its use for accurate relaxation measurements.

Nonetheless, an attractive feature of the labels derived from DL323 grown in  $^{13}\text{C}$ -1,3-glycerol and formate is the



**Fig. 4** Estimation of C2, C5, C6 and C8- $^{13}\text{C}$  labeling efficiency using two- and three-bond  $^{15}\text{N}$ -HSQC experiment without carbon decoupling during acquisition for a mixture of four rNMPs extracted from DL323 *E. coli* grown on either  $^{13}\text{C}$ -1,3-glycerol or  $^{13}\text{C}$ -2-glycerol. **a**  $2\text{D } ^2J_{\text{HN}}$  HSQC spectra for the mixture of four rNMPs obtained from the DL323 bacterial culture (grown in  $^{13}\text{C}$ -1,3-glycerol) are superimposed (the spectrum obtained without labeled formate, *red contours and upper*; formate labeled spectrum, *blue contours and lower*). The two insets show the observable long range  $^1\text{H}$ - $^{15}\text{N}$  correlations in the purine and uracil rings. **b** 1D section of the Ade N7 position (from Fig. 4a) is depicted for labeled rAMP without formate (*red*) and with formate (*blue*). The satellite peaks are labeled I and II,

and the center peak is labeled III. **c**  $2\text{D } ^2J_{\text{HN}}$  HSQC spectra for the mixture of the four rNMPs obtained from DL323 bacterial culture (grown in  $^{13}\text{C}$ -2-glycerol) are superimposed (the spectrum obtained without labeled formate, *red contours and upper*; formate labeled spectrum, *blue contours and lower*). **d** The  $^1\text{H}$  spectrum with no  $^{13}\text{C}$ -decoupling during acquisition (*green*) is superimposed on 1D section of the Ade N7 position. The 1D sections are depicted for labeled rAMPs without formate (*red*) and with formate (*blue*). The 1D slices from the  $2\text{D } ^2J_{\text{HN}}$  HSQC spectra overlay with the proton spectrum, suggesting the percentage label can be estimated using either the 2D or 1D experiment. The 2D experiment is preferable in case of overlap

complete isolation of the C5' labels from C4'. This label is therefore appropriate for NMR applications that exploit the methylene carbon position in RNA. As demonstrated by Bax and coworkers, considerable gain in  $^1\text{H}$  resolution is possible for methylene C5' regions in a 19-base-pair DNA fragment (Miclet et al. 2004).

Therefore, to ascertain the usefulness of site specific labels under conditions of broadened resonances that accompany RNA of increased size, we dissolved the labeled nucleotides in perdeuterated glycerol. At a temperature of  $40^\circ\text{C}$ , the ribose C5' resonances are not only broadened in a normal non-constant time  $^{13}\text{C}$  HSQC experiment (Fig. 6c) but also exhibit extensive spectral overlap. To estimate the overall tumbling time constant ( $\tau_c$ )

at this temperature, we use the high frequency approximation (Farrow et al. 1995; Ishima and Nagayama 1995; Peng and Wagner 1995; Dayie et al. 1996) and assume isotropic tumbling, such that the overall correlation time is independent of the order parameter (Kay et al. 1989; Eldho and Dayie 2007):

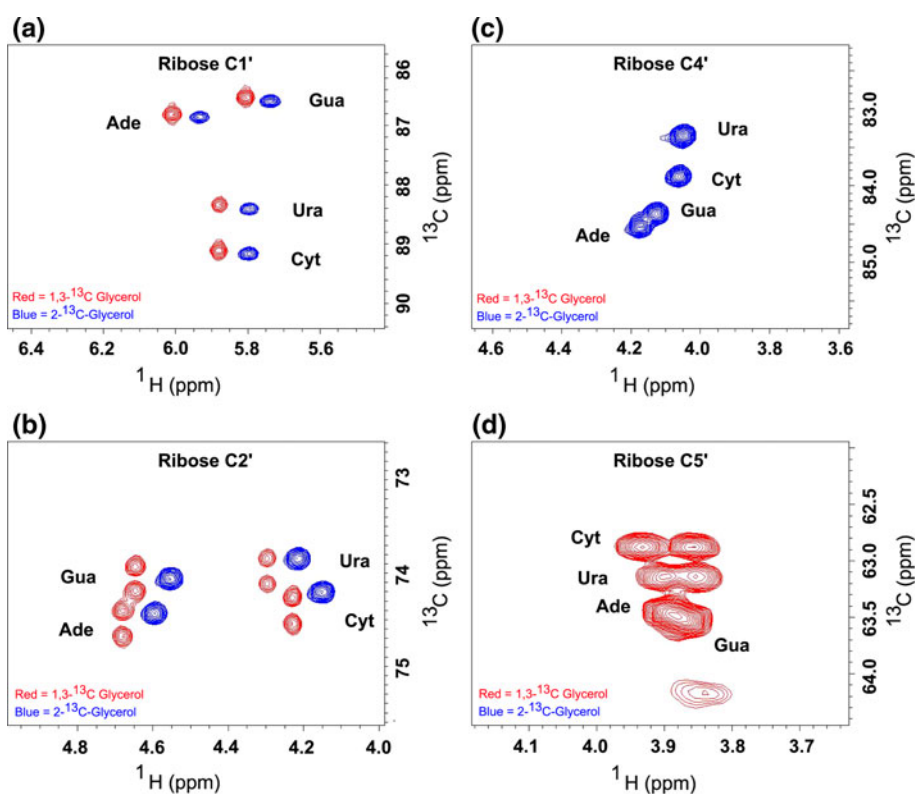
$$\tau_c \approx \left( \sqrt{6R_2/R_1 - 7} \right) / 2\omega_X \quad (1)$$

$R_1$  is the longitudinal relaxation rate constant,  $R_2$  is the transverse relaxation rate constant, and  $\omega_X$  is the Larmor frequency of the X ( $^{15}\text{N}$  or  $^{13}\text{C}$ ) spin. At  $40^\circ\text{C}$ , we estimate the rotational correlation time to be  $9.5 \pm 0.7$  ns. Given 6 ns as the correlation time for a 36 nucleotide RNA at

**Table 2**  $^{13}\text{C}$  enrichment levels at various carbon positions within ribonucleotides harvested from DL323 *E. coli* grown on  $^{13}\text{C}$ -2-glycerol with and without  $^{13}\text{C}$ -formate as carbon sources

Carbon position labeled	$^{13}\text{C}$ -Carbon source: 2-Glycerol only	$^{13}\text{C}$ -Carbon source: 2-Glycerol and Formate
Purine <sup>a</sup>		
Ade C2	$0.16 \pm 0.10$	$0.12 \pm 0.06$
C8	$0.17 \pm 0.12$	$0.88 \pm 0.03$
Pyrimidine		
C5	<0.03	<0.03
C6	$0.98 \pm 0.003$	$0.95 \pm 0.02$
Ribose		
C1'	$0.51 \pm 0.11$	$0.33 \pm 0.05$
C2'	$0.30 \pm 0.09$	$0.60 \pm 0.08$
C3'	<0.02	<0.02
C4'	$0.87 \pm 0.22$	$1.0 \pm 0.17$
C5'	<0.02	<0.02

<sup>a</sup> The percentage labels are calculated as described in the footnote to Table 1



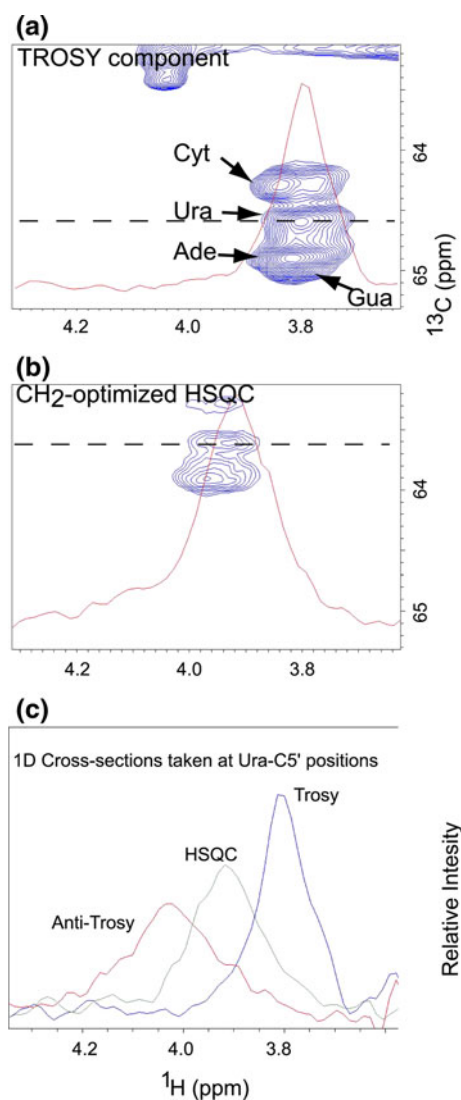
**Fig. 5** 2D non-constant time HSQC spectra of all four labeled nucleotides extracted from DL323 *E. coli* grown on either  $^{13}\text{C}$ -1,3-glycerol (*red contours, left shifted*) or  $^{13}\text{C}$ -2-glycerol (*blue contours*) with media supplemented with  $^{13}\text{C}$ -formate. Formate has no effect on the enrichment in the ribose region. Growth on  $^{13}\text{C}$ -2-glycerol results in label at C1', C2' and C4' and absence of label at C3' and C5', whereas growth on  $^{13}\text{C}$ -1,3-glycerol results in label at all but the C4'

carbon. **a** Ribose C1', **b** Ribose C2', **c** Ribose C4' and **d** Ribose C5'. The resonances from each of the four nucleotides are annotated for adenine (Ade), cytosine (Cyt), guanine (Gua), and uracil (Ura). Slight differences in pH and salt conditions between samples leads a noticeable shift in the chemical positions such that some of the singlets are not exactly centered relative to the doublets in the carbon dimension

25°C (Eldho and Dayie 2007), and assuming a linear correlation between molecular weight and correlation time, then these nucleotides are behaving roughly as a 50–60 nucleotide RNA. Experiments carried out at lower temperatures (down to 15°C) indicate useful signals can

be obtained for RNAs of up to 150 nucleotides ( $\sim 30$  ns correlation time) in agreement with previous observations (Hansen and Al-Hashimi 2007).

Use of the non-constant time version of the  $^{13}\text{CH}_2$  TROSY experiment (Miclet et al. 2004) enables the rescue



**Fig. 6** CH<sub>2</sub>-TROSY (Miclet et al. 2004) and CH<sub>2</sub>-optimized HSQC (Schleucher et al. 1994; Sattler et al. 1995) spectra of the C5' region of all four labeled nucleotides extracted from DL323 *E. coli* grown on <sup>13</sup>C-1,3-glycerol and <sup>13</sup>C-formate. **a** TROSY component. The four nucleotide peaks are labeled for cytosine (Cyt), uracil (Ura), adenine (Ade), and guanine (Gua). **b** CH<sub>2</sub>-optimized HSQC. **c** 1D traces showing linewidths. Experiments were carried at 40°C with rNMPs dissolved in 99% D<sub>8</sub>-glycerol to simulate a macromolecule of ~50–60 nucleotides. Identical acquisition and processing parameters were used: time domain matrices of 64\* × 512\* data points with acquisition times of 26.5 ms (*t*<sub>1</sub>) and 48 ms (*t*<sub>2</sub>). Spectra are plotted at identical contour levels, and the inset shows the 1D cross sections taken at Ura chemical shift position. As expected, the TROSY peak corresponding to the (C<sup>-</sup>/H<sup>+</sup>) transitions (Miclet et al. 2004) is narrow, right and down shifted from the peak associated with the decoupled HSQC spectra

of the slowest relaxing multiplet component with improved resolution and sensitivity compared to the fastest relaxing and broadened signals (Fig. 6a, b). Compared to the normal CH<sub>2</sub>-methylene optimized HSQC (Schleucher et al. 1994; Sattler et al. 1995), all four rNMPs in the mixture are

resolved at 800 MHz proton frequency. It is anticipated that these and other new experiments that incorporate the <sup>13</sup>CH<sub>2</sub> TROSY module can be designed to probe RNA-ligand interactions at very high resolution using the site specific labels described here.

As discussed previously for the K10z<sub>wf</sub> labeled rNMPs (Dayie and Thakur 2010), the DL323 labeled rNMPs will also benefit a number of important spin relaxation applications. For instance, obtaining accurate relaxation parameters from <sup>13</sup>C-CPMG-based relaxation dispersion experiments for quantifying millisecond time-scale processes is predicated on removing modulations that complicate analysis of the data (Yamazaki et al. 1994; Johnson et al. 2006; Johnson and Hoogstraten 2008; Hansen et al. 2009).

## Conclusion

*Escherichia coli* strains with lesions in specific metabolic pathways are attractive for synthesizing selectively labeled nucleotides for tackling large RNA macromolecules (Johnson et al. 2006; Johnson and Hoogstraten 2008; Dayie and Thakur 2010). Here we show that we can preserve a high level of label at the ribose C1' and C5' and base C2 and C8 without introducing labels in adjacent carbon positions, thereby ensuring an isolated two spin system that limits extraneous relaxation pathways. The use of the mutant *E. coli* DL323 strain that is deficient in succinate and malate dehydrogenases has the added advantage that the flux through the TCA cycle is reduced to zero such that maximal labeling is obtained at either the C5 or C6 pyrimidine position without compromising the label of the ribose ring. The use of <sup>13</sup>C-1,3-glycerol produces rNMPs enriched at C1' (~60%) and C5' (~90%) in the ribose ring and C2 (~90%) and C8 (~90%) in the purine ring without the deleterious one-bond <sup>13</sup>C–<sup>13</sup>C scalar couplings that could interfere with accurate spin-relaxation measurements. The use of <sup>13</sup>C-2-glycerol also produces rNMPs with enrichment of <sup>13</sup>C at C2' and C4' of the ribose groups and C6 groups in the Pyr ring without incorporation of label at either base C5 sites or ribose C3' or C5'. Thus the one-bond <sup>13</sup>C–<sup>13</sup>C scalar couplings that could interfere with accurate spin-relaxation measurements are absent. As expected, we find that the C2 and C8 purine sites are highly <sup>13</sup>C-enriched in rNMPs derived from DL323 strain grown in <sup>13</sup>C-1,3-glycerol growth media without <sup>13</sup>C-labeled formate; rNMPs derived from DL323 strain grown in <sup>13</sup>C-2-glycerol growth media require exogenous formate for high enrichment of mostly the C8 positions. The isolated but highly enriched C5' position opens up an attractive avenue for CH<sub>2</sub>-methylene TROSY applications for large RNA molecules. Overall then, these labeling patterns

provide a large number of ‘isolated’ backbone ribose and base side-chain probes that are available for the study of dynamics and structure of large RNAs.

**Acknowledgments** This work was supported in part by the University of Maryland Nano-Biotechnology Award and the National Institutes of Health grant GM077326 to T. K. Dayie, and J. N. Sama is a recipient of a Percy Julian Fellowship (NSF grant CHE0746446) and an HHMI undergraduate fellowship.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## References

- Batey RT, Battiste JL, Williamson JR (1995) Preparation of isotopically enriched RNAs for heteronuclear NMR. *Methods Enzymol* 261:300–322
- Bax A, Freeman R (1981) Investigation of complex networks of spin-spin coupling by two-dimensional NMR. *J Magn Reson* 44:542–561
- Bax A, Mehlkopf AF, Smidt J (1979) Homonuclear broad-band decoupled absorption spectra. *J Magn Reson* 35:167–169
- Bax A, Ikura M, Kay LE, Torchia DA, Tschudin R (1990) Comparison of different modes of two-dimensional reverse correlation NMR for the study of proteins. *J Magn Reson* 86:304–318
- Bermel W, Bertini I, Felli IC, Piccioli M, Pierattelli R (2006) C-13-detected protonless NMR spectroscopy of proteins in solution. *Prog Nucl Magn Reson Spectrosc* 48:25–45
- Bodenhausen G, Ruben DJ (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem Phys Lett* 69:185–189
- Boisbouvier J, Wu Z, Ono A, Kainosho M, Bax A (2003) Rotational diffusion tensor of nucleic acids from  $^{13}\text{C}$  NMR relaxation. *J Biomol NMR* 27:133–142
- Boisbouvier J, Bryce DL, O’Neil-Cabello E, Nikonowicz EP, Bax A (2004) Resolution-optimized NMR measurement of  $^1\text{D}_{\text{CH}}$ ,  $^1\text{D}_{\text{CC}}$  and  $^2\text{D}_{\text{CH}}$  residual dipolar couplings in nucleic acid bases. *J Biomol NMR* 30:287–301
- Brutscher B, Boisbouvier J, Kupce E, Tisne C, Dardel F, Marion D, Simorre JP (2001) Base-type-selective high-resolution  $^{13}\text{C}$  edited NOESY for sequential assignment of large RNAs. *J Biomol NMR* 19:141–151
- Dayie KT (2005) Resolution enhanced homonuclear carbon decoupled triple resonance experiments for unambiguous RNA structural characterization. *J Biomol NMR* 32:129–139
- Dayie KT (2008) Key labeling technologies to tackle sizeable problems in RNA structural biology. *Int J Mol Sci* 9:1214–1240
- Dayie TK, Thakur CS (2010) Site-specific labeling of nucleotides for making RNA for high resolution NMR studies using an *E. coli* strain disabled in the oxidative pentose phosphate pathway. *J Biomol NMR* 47:19–31
- Dayie KT, Wagner G, Lefevre JF (1996) Theory and practice of nuclear spin relaxation in proteins. *Annu Rev Phys Chem* 47:243–282
- Dayie KT, Brodsky AS, Williamson JR (2002) Base flexibility in HIV-2 TAR RNA mapped by solution  $^{15}\text{N}$ ,  $^{13}\text{C}$  NMR relaxation. *J Mol Biol* 317:263–278
- Eldho NV, Dayie KT (2007) Internal bulge and tetraloop of the catalytic domain 5 of a group II intron ribozyme are flexible: implications for catalysis. *J Mol Biol* 365:930–944
- Farès C, Amata I, Carlomagno T (2007)  $^{13}\text{C}$ -detection in RNA bases: revealing structure-chemical shift relationships. *J Am Chem Soc* 129:15814–15823
- Farrow NA, Zhang O, Szabo A, Torchia DA, Kay LE (1995) Spectral density function mapping using  $^{15}\text{N}$  relaxation data exclusively. *J Biomol NMR* 6:153–162
- Fiala R, Sklenár V (2007)  $^{13}\text{C}$ -detected NMR experiments for measuring chemical shifts and coupling constants in nucleic acid bases. *J Biomol NMR* 39:153–163
- Fürtig B, Richter C, Wöhnert J, Schwalbe H (2003) NMR spectroscopy of RNA. *ChemBioChem* 4:936–962
- Grzesiek S, Bax A (1992) Improved 3D triple resonance NMR techniques applied to a 31 kDa protein. *J Magn Reson* 96:432–440
- Hall KB (1995) Uses of  $^{13}\text{C}$ - and  $^{15}\text{N}$ -labeled RNA in NMR of RNA-protein complexes. *Methods Enzymol* 261:542–559
- Hansen AL, Al-Hashimi HM (2007) Dynamics of large elongated RNA by NMR carbon relaxation. *J Am Chem Soc* 129:16072–16082
- Hoffman DW, Holland JA (1995) Preparation of C-13-labeled ribonucleotides using acetate as an isotope source. *Nucleic Acids Res* 23:3361–3362
- Hoogstraten CG, Johnson JE (2008) Metabolic labeling: taking advantage of bacterial pathways to prepare spectroscopically useful isotope patterns in proteins and nucleic acids. *Concepts Magn Reson Part A* 32:34–55
- Ishima R, Nagayama K (1995) Protein backbone dynamics revealed by quasi spectral density function analysis of amide N-15 nuclei. *Biochemistry* 34:3162–3171
- Johnson JE, Hoogstraten CG (2008) Extensive backbone dynamics in the GCAA RNA tetraloop analyzed using  $^{13}\text{C}$  NMR spin relaxation and specific isotope labeling. *J Am Chem Soc* 130:16757–16769
- Johnson JE, Julien KR, Hoogstraten CG (2006) Alternate-site isotopic labeling of ribonucleotides for NMR studies of ribose conformational dynamics in RNA. *J Biomol NMR* 35:261–274
- Kay LE, Torchia DA, Bax A (1989) Backbone dynamics of proteins as studied by  $^{15}\text{N}$  inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* 28:8972–8979
- Kirkpatrick C, Maurer LM, Oyelakin NE, Yoncheva YN, Maurer R, Slonczewski JL (2001) Acetate and formate stress: opposite responses in the proteome of *Escherichia coli*. *J Bacteriol* 183:6466–6477
- Knappe J, Blaschkowski HP, Gröbner P, Schmitt T (1974) Pyruvate formate-lyase of *Escherichia coli*: the acetyl-enzyme intermediate. *Eur J Biochem* 50:253–2563
- Kupce E, Wagner G (1996) Multisite band-selective decoupling in proteins. *J Magn Reson B* 110:309–312
- Latham MR, Brown DJ, McCallum SA, Pardi A (2005) NMR methods for studying the structure and dynamics of RNA. *ChemBiochem* 6:1492–1505
- LeMaster DM, Kushlan DM (1996) Dynamical mapping of *E. coli* thioredoxin via  $^{13}\text{C}$  NMR relaxation analysis. *J Am Chem Soc* 118:9255–9264
- LeMaster DM, Richards FM (1982) Preparative-scale isolation of isotopically labeled amino acids. *Anal Biochem* 122:238–247
- Lu K, Miyazaki Y, Summers MF (2010) Isotope labeling strategies for NMR studies of RNA. *J Biomol NMR* 46:113–125
- Lundström P, Teilum K, Carstensen T, Bezsonova I, Wiesner S, Hansen DF, Religa TL, Akke M, Kay LE (2007) Fractional  $^{13}\text{C}$  enrichment of isolated carbons using [1- $^{13}\text{C}$ ]- or [2- $^{13}\text{C}$ ]-glucose facilitates the accurate measurement of dynamics at backbone C $\alpha$  and side-chain methyl positions in proteins. *J Biomol NMR* 38:199–212
- Miclet E, Williams DC Jr, Clore GM, Bryce DL, Boisbouvier J, Bax A (2004) Relaxation-optimized NMR spectroscopy of methylene



- groups in proteins and nucleic acids. *J Am Chem Soc* 126: 10560–10570
- Nelson DL, Cox MM (2008) *Lehninger: principles of biochemistry*. W. H. Freeman, New York
- Pardi A (1995) Multidimensional heteronuclear NMR experiments for structure determination of isotopically labeled RNA. *Methods Enzymol* 261:350–380
- Peng JW, Wagner G (1995) Frequency spectrum of NH bonds in eglinc from spectral density mapping at multiple fields. *Biochemistry* 34:16733–16752
- 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* 94:12366–12371
- Religa TL, Kay LE (2010) Optimal methyl labeling for studies of supra-molecular systems. *J Biomol NMR* 47:163–169
- Sambrook J, Russell D (2001) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Sattler M, Schwendinger MG, Schleucher J, Griesinger C (1995) Novel strategies for sensitivity enhancement in heteronuclear multi-dimensional NMR experiments employing pulsed field gradients. *J Biomol NMR* 5:11–22
- Schleucher J, Schwendinger M, Sattler M, Schmidt P, Schedletsky O, Glaser SJ, Sørensen OW, Griesinger C (1994) A general enhancement scheme in heteronuclear multidimensional NMR employing pulsed field gradients. *J Biomol NMR* 4:301–306
- Schultheisz HL, Szymczyna BR, Scott LG, Williamson JR (2008) Pathway engineered enzymatic de novo purine nucleotide synthesis. *ACS Chem Biol* 3:499–511
- Shimba N, Stern AS, Craik CS, Hoch JC, Dotsch V (2003) Elimination of  $^{13}\text{C}\alpha$  splitting in protein NMR spectra by deconvolution with maximum entropy reconstruction. *J Am Chem Soc* 125:2382–2383
- Studier FW (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expr Purif* 41:207–234
- Takeuchi K, Frueh DP, Hyberts SG, Sun ZY, Wagner G (2010) High-resolution 3D CANCA NMR experiments for complete main-chain assignments using  $\text{C}\alpha$  direct detection. *J Am Chem Soc* 132: 2945–2951
- Thakur CS, Brown ME, Sama JN, Jackson ME, Dayie TK (2010) Growth of wildtype and mutant *E. coli* strains in minimal media for optimal production of nucleic acids for preparing labeled nucleotides. *Appl Microbial Biotechnol* 88:771–779
- Thauer RK, Kirchniawy FH, Jungermann KA (1972) Properties and function of the pyruvate–formate–lyase reaction in Clostridia. *Eur J Biochem* 27:282–290
- Tolbert TJ, Williamson JR (1996) Preparation of specifically deuterated RNA for NMR studies using a combination of chemical and enzymatic synthesis. *J Am Chem Soc* 118:7929–7940
- van de Ven FJM, Philippens MEP (1992) Optimization of constant-time evolution in multidimensional experiments. *J Magn Reson* 97:637–644
- Voet D, Voet JG, Pratt CW (2008) *Fundamentals of biochemistry*, 3rd edn. Wiley, New York
- Yamazaki T, Muhandiram R, Kay LE (1994) NMR experiments for the measurement of carbon relaxation properties in highly enriched, uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled proteins: application to  $^{13}\text{C}\alpha$  carbons. *J Am Chem Soc* 114:8266–8278
- Ying J, Grishaev A, Bax A (2006) Carbon-13 chemical shift anisotropy in DNA bases from field dependence of solution NMR relaxation rates. *Magn Reson Chem* 44:302–310