J|A|C|S ARTICLES

Published on Web 12/17/2010

Enzymatic De Novo Pyrimidine Nucleotide Synthesis

Heather L. Schultheisz,[†] Blair R. Szymczyna,[†] Lincoln G. Scott,[‡] and James R. Williamson*,[†]

Department of Molecular Biology, The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, MB33, La Jolla, California 92037 and Cassia, LLC, 3030 Bunker Hill Street, Suite 105, San Diego, California 92109

Received July 14, 2010; E-mail: jrwill@scripps.edu

Abstract: The use of stable isotope labeling has revolutionized NMR studies of nucleic acids, and there is a need for methods of incorporation of specific isotope labels to facilitate specific NMR experiments and applications. Enzymatic synthesis offers an efficient and flexible means to synthesize nucleoside triphosphates from a variety of commercially available specifically labeled precursors, permitting isotope labeling of RNAs prepared by in vitro transcription. Here, we recapitulate de novo pyrimidine biosynthesis in vitro, using recombinantly expressed enzymes to perform efficient single-pot syntheses of UTP and CTP that bear a variety of stable isotope labeling patterns. Filtered NMR experiments on ^{13}C, ^{15}N, ^{2}H-labeled HIV-2 TAR RNA demonstrate the utility and value of this approach. This flexible enzymatic synthesis will make implementing detailed and informative RNA stable isotope labeling schemes substantially more cost-effective and efficient, providing advanced tools for the study of structure and dynamics of RNA molecules.

Introduction

Enzymatic synthesis has emerged as an important tool in the production of biochemicals. Enzyme catalyzed reactions are often more specific, efficient, cost-effective, and afford reduced environmental impact over traditional chemical methods.[1−4] Because of these advantages, many straightforward enzymatic schemes have been adopted in industrial biochemical production.[5] Recent advances in metabolic engineering have greatly broadened the targets and applications for enzymatic synthesis, which now contributes significantly to the biofuel industry[6,7] and natural products synthesis.[8] In particular, the development of enzymatic tools to synthesize nucleotides has had a tremendous impact on the study of RNA, DNA, and other nucleic acids by nuclear magnetic resonance (NMR) methods.[9] Multinuclear NMR experiments require the incorporation of one or more stable isotope labels, and therefore synthetic methods to incorporate isotope labels into nucleic acid building blocks have driven a significant advance in structural studies, including allowing the study of larger molecules.[10] In particular, there is a need for pyrimidine nucleotides that are selectively isotope labeled to complement the specifically labeled purine nucleotides that are commercially available and/or obtained efficiently using published methods.[11]

The primary method to obtain stable isotope labeled nucleotides is the harvest of uniformly ^{13}C, ^{15}N-labeled nucleotides from bacteria grown on labeled media containing $^{15}NH_4Cl$ and ^{13}C-glucose or ^{13}C-methanol.[12−14] This biomass method is suitable for economic production of uniformly labeled nucleotides in large quantity, but does not allow for efficient site-specific incorporation of stable isotopes at designated locations without specialized labeled medium, metabolically modified bacteria, or both. Selective incorporation of ^{15}N, ^{13}C into nucleic acids is extremely useful to highlight specific resonances, to facilitate selective magnetization transfer pathways, and to simplify spectra. Using ^{13}C, ^{2}H- glucose and unlabeled bases, enzymatic methods for the selective deuteration of the ribose moiety of nucleotides have allowed the study of RNA of increasing size by effectively editing spectra to allow for the assignment of intermolecular and sugar−sugar NOEs.[15] Using this method, the $H_{1'}$ and $H_{2'}$ protons can be selectively observed, providing important structural information, and a variety of

[†] Department of Molecular Biology, The Skaggs Institute for Chemical Biology, The Scripps Research Institute.

[‡] Cassia, LLC.

(1) Boumendjel, A.; Blanc, M.; Williamson, G.; Barron, D. *J. Agric. Food Chem.* **2009**, *57*, 7264–7267.
(2) Ahmad, A. L.; Oh, P. C.; Abd Shukor, S. R. *Biotechnol. Adv.* **2009**, *27*, 286–296.
(3) Nuijens, T.; Cusan, C.; Kruijtzer, J. A.; Rijkers, D. T.; Liskamp, R. M.; Quaedflieg, P. J. *J. Org. Chem.* **2009**, *74*, 5145–5150.
(4) Rowan, A. S.; Hamilton, C. J. *Nat. Prod. Rep.* **2006**, *23*, 412–443.
(5) Chang, S. W.; Shaw, J. F. *N. Biotechnol.* **2009**, *26*, 109–116.
(6) Ognjanovic, N.; Bezbradica, D.; Knezevic-Jugovic, Z. *Bioresour. Technol.* **2009**, *100*, 5146–5154.
(7) Zhao, H.; Song, Z.; Olubajo, O.; Cowins, J. V. *Appl. Biochem. Biotechnol.* **2010**, *162*, 13–23.
(8) Comini, M. A.; Dirdjaja, N.; Kaschel, M.; Krauth-Siegel, R. L. *Int. J. Parasitol.* **2009**, *39*, 1059–1062.
(9) Shajani, Z.; Varani, G. *Biopolymers* **2007**, *86*, 348–359.
(10) Davis, J. H.; Tonelli, M.; Scott, L. G.; Jaeger, L.; Williamson, J. R.; Butcher, S. E. *J. Mol. Biol.* **2005**, *351*, 371–382.
(11) Lu, K.; Miyazaki, Y.; Summers, M. F. *J. Biomol. NMR* **2010**, *46*, 113–125.
(12) Batey, R. T.; Battiste, J. L.; Williamson, J. R. *Method Enzymol.* **1995**, *261*, 300–322.
(13) Batey, R. T.; Cloutier, N.; Mao, H.; Williamson, J. R. *Nucleic Acids Res.* **1996**, *24*, 4836–4837.
(14) Nikonowicz, E. P.; Sirr, a.; Legault, P.; Jucker, F. M.; Baer, L. M.; Pardi, A. *Nucleic Acids Res.* **1992**, *20*, 4507–4513.
(15) Tolbert, T. J.; Williamson, J. R. *J. Am. Chem. Soc.* **1997**, *119*, 12100–12108.

labeling patterns are possible due to the availability of labeled glucose. Incorporation of isotope labeling into the base moiety of nucleotides is also extremely important due to their structural role in hydrogen bonding and base stacking.

We have previously developed a method to selectively incorporate ^{13}C, ^{15}N into the purine bases via a total enzymatic synthesis of ATP and GTP starting from serine, glucose, NH_4^+ and CO_2 with up to 66% isolated yield.[16] Historically, labeled UTP and CTP have been prepared from chemically synthesized bases. For example, chemical synthesis of $^{15}N_3$ uracil from $(^{15}NH_4)_2SO_4$ is very efficient giving up to 75% yield.[17] An efficient chemoenzymatic synthesis of U-^{13}C, ^{15}N-UTP has also been demonstrated starting from U-^{15}N, ^{13}C-uracil and U-^{13}C-glucose giving an isolated yield of 60%.[18] Specifically labeled $^{13}C_{2'}$, $^{15}N_{1,3}$- uridine has been synthesized from K^{13}CN and ^{15}N-urea with overall yields of 6.1% and 35.2% based on respective starting materials.[19] A chemical synthesis of $^{13}C_6$-UTP from K^{13}CN was also achieved with an overall yield of 24%.[20] Specifically ^{15}N-labeled cytidine has been chemically synthesized from uridine and $^{15}NH_4Cl$ with yields of 94% for $^{15}N_{amino}$-cytidine, 72% for $^{15}N_3$-cytidine, and 62% for $^{15}N_{amino,\ 3\text{-cytidine}}$.[21] Specifically deuterated CDP has been chemically synthesized from tert-butyldimethylsilyl cytidine derivatives with ~17% overall yield for both $^2H_{2'}$ and $^2H_{3'}$ labels.[22] Here we report a total enzymatic synthesis of UTP and CTP through the expansion of our current enzymatic methods to the de novo pyrimidine biosynthetic pathway. The result is an efficient flexible method to incorporate ^{15}N, ^{13}C, and ^2H into both the base and ribose moieties of pyrimidine nucleotides.

Results and Discussion

Design of the Enzymatic Synthesis. In contrast to de novo purine synthesis, where the base is constructed step by step onto the ribose moiety, in de novo pyrimidine synthesis, the preformed nucleobase is enzymatically coupled to the ribose moiety, as shown in Scheme 1A. While the reconstitution of entire *E. coli* pathways in vitro has proven successful in many cases,[16,23] several key substitutions proved necessary for efficient in vitro orotate synthesis using the de novo pyrimidine pathway. The carbamoyl-phosphate synthase-like carbamate kinase enzyme from thermophile *Pyrococcus furiosus* (*cpkA*)[24,25] was substituted for *E. coli* carbamoyl phosphate synthase (*carA/B*) due to three advantages. First, *cpkA* utilizes NH_4^+ as the nitrogen source, in contrast to the strict requirement for glutamine by *carA/B*, making the reaction more cost-effective. Second, the single subunit *Pyrococcus* enzyme *cpkA* is extremely stable and is easy to prepare. In contrast, *E. coli* carbamoyl phosphate *carA/B* proved unstable and required a challenging preparation of both subunits for optimal activity. Third, *Pyrococcus cpkA* worked under the conditions desired for pyrimidine nucleotide synthesis, eliminating the effects of the regulatory feedback of *E.coli carA/B* by UMP, UDP, and UTP in the one pot synthesis.[26]

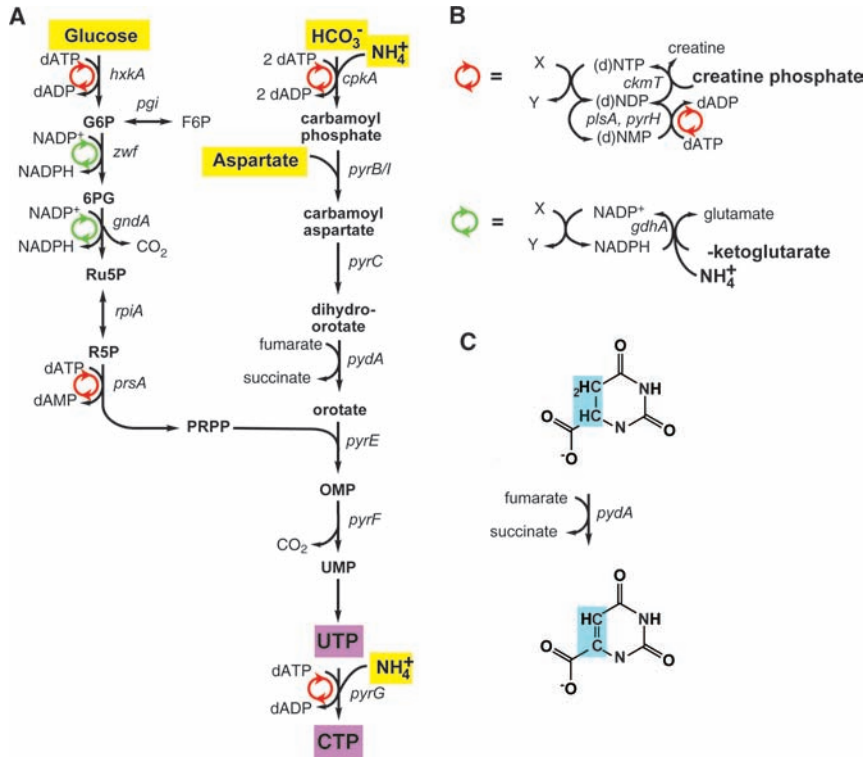
Although the *E. coli* enzymes for the de novo pyrimidine pathway downstream from *E. coli carA/B* functioned well, the requirement for ubiquinone as a hydrogen acceptor by *E. coli* dihydro-orotate dehydrogenase (*pyrD*) in the fourth step of de novo pyrimidine synthesis was very inconvenient. The possibility of ubiquinone recycling was daunting because the multisubunit membrane associated enzymes such as succinate dehydrogenase or NADH dehydrogenase would be required. However, dihydro-orotate dehydrogenase A (*pydA*) from *Lactoccocus lactis* (a class 1A member of the dihydro-orotate dehydrogenase family) uses fumarate as its hydrogen acceptor, and this enzyme-cofactor pair proved to be a suitable replacement.[27]

Regeneration of ATP and $NADP^+$, as shown in Scheme 1B, provides a driving force for the overall pyrimidine synthesis reaction.[28,29] Several equivalents of creatine phosphate drive the immediate recharging of NMPs and NDPs to NTP by nucleotide kinases through mass action. Similarly, α-ketoglutarate and NH_4^+ fuel the conversion of NADPH to $NADP^+$ providing the proper substrate for the $NADP^+$ dependent reactions to move forward. In addition dATP was substituted for ATP in all cases to prevent dilution of the NTP pool with unlabeled ATP. Although de novo pyrimidine enzymes *cpkA* and *pyrG* accept dATP with slightly reduced activity, the deoxynucleotide byproducts can be easily separated from the product NTPs during the affinity purification step.

With all of the proper components in place, the engineered de novo pyrimidine pathway was combined with pentose phosphate enzymes and cofactor regeneration of ATP and $NADP^+$ to synthesize UTP from HCO_3^-, NH_4^+, aspartate, and glucose in one pot. CTP was synthesized from UTP and NH_4^+ in a separate reaction due to inhibition from components present in the first reaction.[30] Figure 1A,B shows the overall reaction for synthesis of UTP and CTP. A relatively small number of components are necessary to build the pyrimidine ring. The metabolic origin of each atom is shown in Figure 1C,D. The major contributor is aspartate, and due to the wide availability of specifically labeled forms of aspartate, many specific pyrimidine base labeled patterns can be synthesized including a combination of ^{13}C, ^{15}N, and ^2H. The exchange of C_6 proton with solvent during the decarboxylation of OMP is convenient for specific deuteration schemes. Deuteration at the C_5 position is often desired since the H_5 and H_1 chemical shifts are very similar, and C_5 deuteration has been accomplished previously through a separate chemical synthesis of $^2H_5$- uracil.[31] Using the *de novo* enzymatic synthesis scheme, selective deuteration on both the base and ribose with simultaneous ^{15}N labeling was

(16) Schultheisz, H. L.; Szymczyna, B. R.; Scott, L. G.; Williamson, J. R. *ACS Chem. Biol.* **2008**, *3*, 499–511.
(17) Griffey, R. H.; Poulter, C. D. *Nucleic Acids Res.* **1983**, *11*, 6497–6504.
(18) Gilles, A. M.; Cristea, I.; Palibroda, N.; Hilden, I.; Jensen, K. F.; Sarfati, R. S.; Namane, A.; Ughetto-Monfrin, J.; Barzu, O. *Anal. Biochem.* **1995**, *232*, 197–203.
(19) Patching, S. G.; Middleton, D. A.; Henderson, P. J.; Herbert, R. B. *Org. Biomol. Chem.* **2003**, *1*, 2057–2062.
(20) SantaLucia, J.; Shen, L. X.; Cai, Z. P.; Lewis, H.; Tinoco, I. *Nucleic Acids Res.* **1995**, *23*, 4913–4921.
(21) Ariza, X.; Vilarrasa, J. *J. Org. Chem.* **2000**, *65*, 2827–2829.
(22) Sandbrink, J.; Stromberg, R. *Nucleosides, Nucleotides Nucleic Acids* **2003**, *22*, 1657–1659.
(23) Schultheisz, H. L.; Szymczyna, B. R.; Williamson, J. R. *J. Am. Chem. Soc.* **2009**, *131*, 14571–14578.
(24) Durbecq, V.; Legrain, C.; Roovers, M.; Pierard, A.; Glansdorff, N. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12803–12808.
(25) Alcantara, C.; Cervera, J.; Rubio, V. *FEBS Lett.* **2000**, *484*, 261–264.
(26) Anderson, P. M.; Meister, A. *Biochemistry* **1966**, *5*, 3164–3169.
(27) Fagan, R. L.; Jensen, K. F.; Bjornberg, O.; Palfey, B. A. *Biochemistry* **2007**, *46*, 4028–4036.
(28) Crans, D. C.; Kazlauskas, R. J.; Hirschbein, B. L.; Wong, C. H.; Abril, O.; Whitesides, G. M. *Methods Enzymol.* **1987**, *136*, 263–280.
(29) Chenault, H. K.; Whitesides, G. M. *Appl. Biochem. Biotechnol.* **1987**, *14*, 147–197.
(30) Tolbert, T. J.; Williamson, J. R. *J. Am. Chem. Soc.* **1996**, *118*, 7929–7940.
(31) Kiritani, R.; Asano, T.; Fujita, S.; Dohmaru, T.; Kawanishi, T. *J. Labelled Compd. Radiopharm.* **1986**, *23*, 207–214.

Scheme 1. Pyrimidine Nucleotide Synthesis with Cofactor Recycling[a]



[a] (A) *Conversion of glucose to phosphoribosyl-pyrophosphate (PRPP).* Glucose and dATP are converted to glucose-6-phosphate (G6P) by the action of hexokinase (*hxkA)* with the production of dATP. G6P is interconverted to fructose-6-phosphate (F6P) by the action of phosphoglucose isomerase (*pgi*). G6P and $NADP^+$ are converted to 6-phosphogluconic acid (6PG) by the action of G6P- dehydrogenase (*zwf*) with the production of NADPH. 6PG and $NADP^+$ are converted to ribulose-5-phosphate (Ru5P) by the action of 6PG-dehydrogenase (*gndA*) with the production of NADPH. Ru5P is interconverted to ribose-5-phosphate (R5P) by the action of phosphoribose isomerase (*rpiA*). R5P and dATP are converted to PRPP by the action of PRPP synthase (*prsA*) with the production of dAMP. *Conversion of HCO_3^- and NH_4^+ to UTP and CTP.* HCO_3^-, NH_4^+, and dATP are converted to carbamoyl phosphate by the action of carbamate kinase-like carbamoyl phosphate synthase (*cpkA*) with the production of dADP. Carbamoyl phosphate and aspartate are converted to carbamoyl aspartate by the action of aspartate carbamoyl transferase (*pyrB/I*). Carbamoyl aspartate is converted to dihydro-orotate by the action of dihydro-orotase (*pyrC*). Dihydro-orotate and fumarate are converted to orotate by the action of dihydro-orotate dehydrogenase (*pydA*) with the production of succinate. Orotate and PRPP are converted to orotate monophosphate (OMP) by the action of orotate phosphoribosyl transferase (*pyrE*). OMP is converted to uridine monophosphate (UMP) by the action of orotate decarboxylase (*pyrF*) with the production of CO_2. UMP is charged up to uridine triphosphate (UTP) as shown in (B). For cytidine triphosphate (CTP): UTP, NH_4^+, and dATP are converted to CTP by the action of CTP synthase (*pyrG*) with the production of dADP. (A) *Cofactor regeneration schemes.* NTP regeneration (red). (d)NMP and dATP are converted to (d)NDP by the action of myokinase (*plsA*) and uridylate kinase (*pyrH*), with the production of dADP. (d)NDP and creatine phosphate are converted to (d)NTP by the action of creatine phosphokinase (*ckmT*) with the production of creatine. $NADP^+$ regeneration (green). NADPH, NH_4^+ and α-ketoglutarate are converted to $NADP^+$ by the action of *gdhA* with the production of glutamate. (B) *Conversion of Dihydro-ortate to Orotate.* C_5 and C_6 are highlighted in cyan. C_6 loses its proton in this step. It is replaced by H from solvent during decarboxylation of OMP to UMP.

successful in a single pot reaction. There is one limitation placed on possible labeling patterns using this scheme due to the production of CO_2 by decarboxylation of OMP to UMP and 6-phosphogluconate to ribulose-5-phosphate during PRPP generation. These steps couple the carbon source for the C_2 of the pyrimidine ring to C_1 of aspartate and C_1 of glucose. Degassing and inert atmosphere are sufficient to prevent isotope dilution from air and solvent at the C_2 position (see Supplementary Figure 8). If C_2 labeling is desired without ribose labeling, PRPP can be generated enzymatically from unlabeled ribose.[30] In addition, $^{13}C_1$-aspartate is readily available and can be included to label C_2 without C_4, C_5, C_6 labeling.

Implementation of the Enzymatic Synthesis. There are 18 enzymes required to implement the de novo enzymatic synthesis of UTP and CTP. The majority of the pentose phosphate pathway and cofactor recycling enzymes are commercially available or have been cloned previously.[16] The pyrimidine specific enzymes were cloned from *E. coli* given our previous success with other *E. coli* pathways. In addition, carbamoyl-phosphate synthase like carbamate kinase from *Pyrococcus furiosus* was cloned and substituted in the first step of the de novo pyrimidine pathway. Similarly, dihydro-orotate dehydrogenase was cloned from *Lactococcus lactis* and used in place of the *E. coli* enzyme in the fourth step of de novo pyrimidine synthesis. Due to the lack of suitable activity assays, it was not possible to determine the specific activities for steps 2−4 of de novo pyrimidine synthesis. Instead, an assumed specific activity of 0.5 U/mg was used for aspartate carbamoyl transferase, dihydro-orotase, and dihydro-orotate dehydrogenase. The amounts for the unassayed enzymes were empirically optimized during pilot scale reactions. For best results, it is advisible to use fresh preparations all of the enzymes to prepare nucleotides on a scale sufficient for transcription reactions. To illustrate the flexibility and utility of the enzymatic synthesis, five different labeled nucleotides were synthesized, as shown in Figure 2.

Synthesis of U-^{15}N, $^2H_{5,3',4',5',5''}$- UTP. A 0.75 mmol scale UTP synthesis was performed with U-^{2}H- glucose, ^{15}N, $^2H_{3,3,2}$-aspartate, and $^{15}NH_4Cl$. The labeled glucose, aspartate and ammonium chloride were combined with stoichiometric substrates ($KHCO_3$ and sodium fumarate), fuel reagents (α-ketoglutarate, creatine phosphate), catalytic cofactors (dATP, $NADP^+$), and the 16 enzymes listed in column 1 of Table 1 in

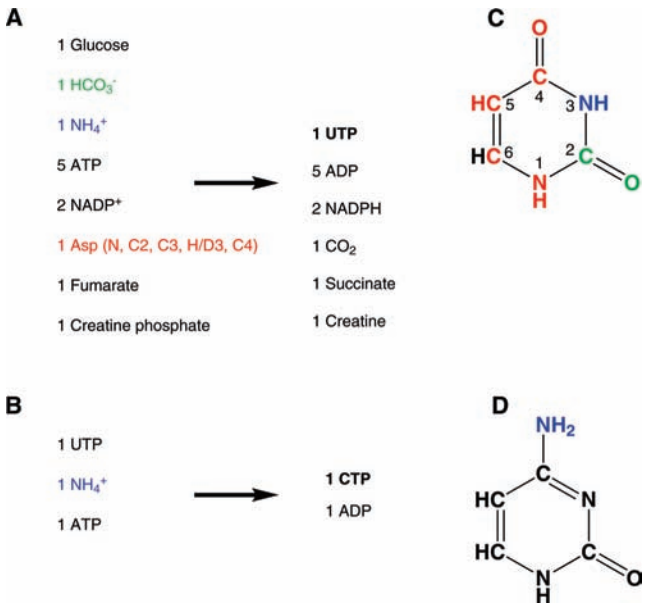


Figure 1. Substrates and products for biosynthesis of (A) UTP and (B) CTP. Metabolic origin of pyrimidine atoms for (C) uracil and (D) cytidine. The substrates and metabolic source for each pyrimidine atom are color-coded: bicarbonate (green), ammonium (blue), and aspartate (red). The H6 hydrogen is derived from water.

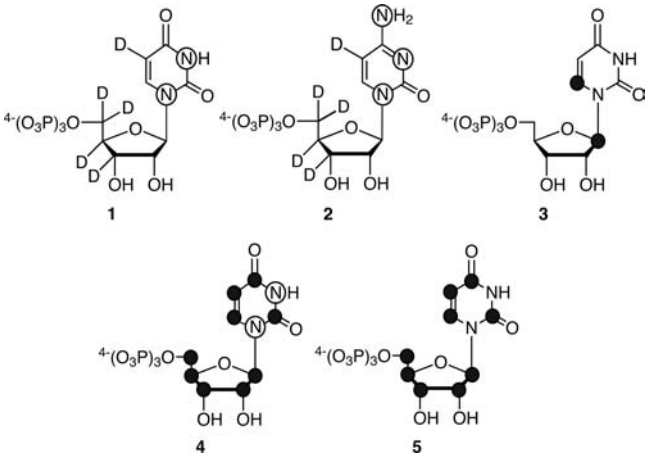


Figure 2. Molecular structures of specifically labeled pyrimidine nucleotides. (1) U-^{15}N,^{2}H$_{5,3',4',5',5''}$- UTP, (2) U-^{15}N,^{2}H$_{5,3',4',5',5''}$- CTP, (3) ^{13}C$_{6,1'}$- UTP, (4) U-^{15}N,^{13}C- UTP, (5) U-^{13}C- UTP. ^{2}H labels are indicated by D,^{13}C labels are indicated by black circles and ^{15}N labels are indicated by open circles.

a 75 mL buffered reaction incubated at 37 °C in two steps. UTP formation was monitored by HPLC over the course of 3 days, when the reaction ceased to progress. The time course for this reaction is shown in Supporting Information Figure S6A. Affinity purification afforded a 30% isolated yield of U-^{15}N, ^{2}H$_{5,3',4',5',5''}$- UTP based on input glucose. Subsequent reactions using different preparations of the unassayed de novo enzymes were more efficient.

Synthesis of U-^{15}N, ^{2}H$_{5,3',4',5',5''}$- CTP. A 0.125 mmol scale CTP synthesis was performed with U-^{15}N, ^{2}H$_{5,3',4',5',5''}$- UTP and ^{15}NH$_4$Cl. The labeled UTP and ammonium chloride were combined with fuel reagent (creatine phosphate), catalytic cofactor (dATP) and the 4 enzymes listed in column 2 of Table 1 in a 125 mL buffered reaction. CTP formation was monitored by HPLC over the course of 20 h when the reaction ceased to progress. The time course for this synthesis is shown in Supporting Information Figure S6B. Affinity purification afforded a 48% isolated yield of U-^{15}N, ^{2}H$_{5,3',4',5',5''}$- CTP based on input UTP.

Synthesis of ^{13}C$_{6,1'}$- UTP. A 0.3 mmol scale synthesis was performed using ^{13}C$_2$- glucose and ^{13}C$_2$-aspartate. The labeled glucose and aspartate were combined with stoichiometric substrates (KHCO$_3$, sodium fumarate, NH$_4$Cl), fuel reagents (α-ketoglutarate, creatine phosphate), catalytic cofactors (dATP, NADP$^+$) and the 15 enzymes listed in column 3 of Table 1 in a 30 mL buffered reaction. UTP formation was monitored by HPLC over the course of 3 days when the reaction ceased to progress. The time course of this synthesis is shown in Supporting Information Figure S6C. Affinity purification afforded a 25% yield based on input glucose.

Synthesis of U-^{13}C, ^{15}N- UTP. A 0.75 mmol scale synthesis was performed using U-^{13}C-glucose, U-^{13}C, ^{15}N- aspartate, ^{15}NH$_4$Cl, and NaH^{13}CO$_3$. The labeled substrates were combined with stoichiometric substrate (sodium fumarate), fuel reagents (α-ketoglutarate, creatine phosphate), catalytic cofactors (dATP, NADP$^+$), and the 16 enzymes listed in column 4 of Table 1 in a 75 mL buffered reaction. UTP formation was monitored by HPLC over the course of 1 day when the reaction ceased to progress. The time course of this synthesis is shown in Supporting Information Figure S6D. Affinity purification afforded a 45% yield based on input glucose. Since equimolar amounts of aspartate and glucose are included in the reaction, the yield is also 45% based on input aspartate.

Synthesis of U-^{13}C-UTP. A 0.75 mmol scale synthesis was performed using U-C-glucose, U-^{13}C- aspartate, and NaH^{13}CO$_3$. The labeled substrates were combined with stoichiometric substrates (sodium fumarate, ammonium chloride), fuel reagents (α-ketoglutarate, creatine phosphate), catalytic cofactors (dATP, NADP$^+$), and the 15 enzymes listed in column 5 of Table 1 in a 75 mL buffered reaction. UTP formation was monitored by HPLC over the course of 2 days when the reaction ceased to progress. The time course of this synthesis is shown in Figure 3. Affinity purification afforded a 40% yield based on input glucose.

Incorporation of Labeled Nucleotides into RNA. To demonstrate the utility of the labeled nucleotides, HIV-2 TAR RNA was transcribed in vitro using U-^{13}C, ^{15}N-ATP, U-^{13}C, ^{15}N-GTP, and the U-^{15}N, ^{2}H$_{5,3',4',5',5''}$- UTP and U-^{15}N, ^{2}H$_{5,3',4',5',5''}$- CTP prepared as described above. The ^{1}H,^{15}N-HSQC and ^{1}H,^{13}C-HSQC spectra of the labeled RNA are shown in Supporting Information Figures S9 and S10, respectively. The proton spectra are of particular interest since vital secondary and tertiary structural information can be gained through understanding the network of hydrogen bonding, base stacking and conformational features of base and ribose moieties. However, the informative areas of RNA proton spectra are naturally crowded, and it is extremely difficult to discern specific correlations.

Employing our specific labeling synthesis scheme, we set out to observe a particular interaction between the H$_{1'}$ and H$_{2'}$ protons of pyrimidine nucleotides with protons bonded to ^{13}C atoms on purines in HIV-2 TAR. The experiment focuses on the internucleotide distance constraints between the H$_{1'}$ and H$_{2'}$ protons of pyrimidine nucleotides and the ^{13}C bound protons of purine nucleotides. The ^{13}C, ^{15}N filtered-edited NOESY experiment, shown in Figure 4B, revealed the two expected NOE correlations in HIV-2 TAR where the protons are ~2 Å apart: H$_{2'}$ of C4 to ^{13}C−H$_8$ of A5 and H$_{2'}$ of U26 to ^{13}C−H$_8$ of G27. Relative to the ^{13}C-edited NOESY spectrum, these two correlations represent a small fraction of the total number of

Table 1. Complete List of Enzymes for De Novo Enzymatic Synthesis of UTP and CTP[a]

enzyme	EC	gene	source	1	2	3	4	5
hexokinase	2.7.1.1	*hxkA*	BY	3.6 U		1.4 U	3.6 U	3.6 U
glucose-6-phosphate isomerase	5.3.1.9	*pgi*	BY	15 U				
glucose-6-phosphate dehydrogenase	1.1.1.49	*zwf*	BY	4 U		1.6 U	4 U	4 U
phosphogluconate dehydrogenase	1.1.1.44	*gndA*	RE	4.4 U		1.7 U	4.4 U	4.4 U
ribose-5-phosphate isomerase	5.3.1.6	*rpiA*	SP	4.8 U		1.9 U	4.8 U	4.8 U
ribose-phosphate diphosphate kinase	2.7.6.1	*prsA*	RE	5.3 U		2.1 U	5.3 U	5.3 U
carbamate kinase-like carbamoyl-P synthase	6.3.4.16	*cpkA*	PF	1.8 U		0.5 U	1.4 U	1.4 U
aspartate carbamoyl transferase	2.1.3.2	*pyrl/B*	RE	42 mg		7 mg	21 mg	21 mg
dihydro-orotase	3.5.2.3	*pyrC*	RE	10 mg		4 mg	6.5 mg	10 mg
dihydro-orotate dehydrogenase	1.3.3.1	*pydA*	LL	25 mg		4.4 mg	10.8 mg	11.5 mg
orotate phosphoribosyl transferase	2.4.2.10	*pyrE*	RE	5.9 U		2.3 U	5.9 U	5.9 U
orotate monophosphate decarboxylase	4.1.1.23	*pyrF*	RE	6.4 U		2.6 U	6.4 U	6.4 U
cytidine triphosphate synthase	6.3.4.2	*pyrG*	RE		20 U			
uridine monophosphate kinase	2.7.4.22	*pyrH*	RE	7.1 U		2.8 U	4.7 U	7 U
cytidine monophosphate kinase	2.7.4.14	*cmk*	RE		1 U		3 U	
adenylate kinase	2.7.4.3	*plsA*	CM	24 U	1 U	8 U	22 U	22 U
glutamic dehydrogenase	1.4.1.3	*gdhA*	BL	18 U		7.3 U	18 U	18 U
creatine phosphokinase	2.7.3.2	*ckmT*	CM	42 U	10 U	10.5 U	26 U	26 U

[a] Amounts of each enzyme used for synthesis of compounds 1−5 in Figure 2 are listed as either Units of enzyme or mg of enzyme. RE = recombinant from *E. coli*, PF = recombinant from *Pyrococcus furiosus*, LL = recombinant from *Lactobacillus lactis*, BY = baker's yeast, SP = spinach, CM = chicken muscle, BL = bovine liver.

NOE correlations typically observed. The application of filtered-filtered and edited−edited experiments using this labeling scheme would reveal proton−proton distance constraints within the populations of pyrimidines and purines, respectively.

Accessible site-specific labeling of pyrimidine bases will also advance the collection and quantitative analysis of dynamics measurements on the ps−ns and μs−ms time scales. To date, ^{13}C relaxation studies have been conducted on uniformly labeled samples and the associated data are complicated by magnetic interactions between adjacent ^{13}C atoms. Through the use of the one-pot synthesis scheme we present here, a variety of pyrimidine nucleotides that have ^{13}C at isolated positions in the base can be rapidly generated. In addition to extending the size of RNA molecules that can be studied, this labeling strategy will also enable the use of sensitive relaxation dispersion experiments to measure dynamics on the μs−ms time scale, instead of depending on the power dependence methodologies.[9] The relative ease and efficiency with which these nucleotides are synthesized should facilitate informative structural and dynamic studies of RNA molecules and complexes of increasing size.

Conclusions

Utilizing enzymes from several different species, we have designed an efficient and flexible method for site-specific labeling of pyrimidine nucleotides in the base and ribose moiety from readily available precursors. This method has allowed the synthesis of new labels with up to 45% yield based on input glucose and aspartate. We have shown that when used in well-designed combinations, these labeled nucleotides enable extremely specific secondary structural features in RNA to be detected by NMR. In the future, we hope to expand our enzymatic synthesis to deoxyribonucleotides to create an equally valuable set of isotope editing tools for probing the secondary structures of DNA molecules.

Experimental Methods

Materials. Chemicals were purchased from Sigma. D-glucose ($^{13}C_2$, 99%), D-glucose ($^{13}C_1$, 98%), D-glucose ($^{2}H_{1,2,3,4,5,6,6}$, 98%), D-glucose ($^{13}C_{1,2,3,4,5,6}$, 99%), D/L- aspartic acid ($^{13}C_2$, 99%), L-aspartic acid ($^{13}C_1$, 99%), L-aspartic acid ($^{15}N_{amino}$, 98%, ^{2}H-$_{2,3,3}$, 98%), L-aspartic acid ($^{13}C_{1,2,3,4}$, 98%), L-aspartic acid ($^{15}N_{amino}$, 98%, $^{13}C_{1,2,3,4}$, 98%), NH_4Cl (^{15}N, 99%), $NaHCO_3$ (^{13}C, 99%) and ATP (U-^{15}N, 98%, ^{13}C, 98%) were purchased from Cambridge Isotope Laboratories. Hexokinase from baker's yeast, glucose-6-phosphate isomerase from baker's yeast, glucose-6-phosphate dehydrogenase from baker's yeast, ribose-5-phosphate isomerase from spinach, creatine phosphokinase from rabbit muscle, adenylate kinase from chicken muscle, and glutamic dehydrogenase from bovine liver were purchased from Sigma. Recombinant enzymes *prsA*, *gndA*, and U-^{15}N, ^{13}C-GTP were prepared as previously described.[16]

Enzyme Cloning. The genes for *pyrI*, *pyrB*, *pyrC*, *pyrE*, *pyrF*, *pyrG*, *pyrH*, and *cmk* were cloned from *E. coli* K12 MRE600 genome. The genes for *cpkA* from *Pyrococcus furiosus* and *pydA* from *Lactococcus lactis* were cloned from genomic DNA purchased from the ATCC (#19435 and #43587, respectively). All genes were cloned using standard procedures based on the reported sequence in Genbank. Gene specific primers with compatible sites for either *Bam*HI, *Eco*RI or *XhoI* restriction endonucleases were used to insert all genes into ampicillin resistant plasmid pET22-HT (derived from pET22b, Novagen) encoding an N-terminal hexahistidine tag. In addition, *pyrB* was inserted into the kanamycin resistant plasmid pET28b (Novagen), also encoding an N-terminal hexahistidine tag. *E. coli* strain DH5α was used for cloning and plasmid maintenance. All constructs were verified by DNA sequencing.

Enzyme Purification. All plasmids were transformed into *E. coli* BL21 (DE3) cells for protein overexpression. Additionally, *cpkA* was transformed into the *E. coli* Rosetta (DE3) strain that enhanced expression. For *pyrI/B* coexpression, the cells were transformed with both *pyrI*-pET22HT and *pyrB*-pET28b, and grown with 200 μg/mL amplicillin and 100 μg/mL kanamycin. Briefly, overnight 5 mL cultures in LB with 200 μg/mL ampicillin were diluted 100 fold and grown at 37 °C with shaking at 260 rpm until the OD_{600} was ~0.5. Protein expression was induced by addition of 1 mM IPTG, followed by growth for an additional 4−5 h. Cells were harvested by centrifugation at 10 000 $\times g$ for 30 min. All steps after cell growth were carried out at 4 °C. The cell pellets from 2 L of growth were gently resuspended in 200 mL 50 mM K-Hepes (pH 7.6), 1 M NaCl, 20 mM imidazole with 10% glycerol. Cell lysis was accomplished by sonication with 20 cycles of a 30 s pulse and 2 min rest at 70% power. The lysates were centrifuged at 31 000 $\times g$, and the supernatants were loaded directly to a Ni-NTA (Qiagen) column equilibrated with lysis buffer minus glycerol. The column was washed with 6 volumes of the same buffer and protein was eluted with 250 mM imidazole. The purified protein was dialyzed against 25 mM K-Hepes (pH 7.6), 50 mM NaCl, 5 mM β-mercaptoethanol (βME) with 50% glycerol. Dialysis for *pyrG*

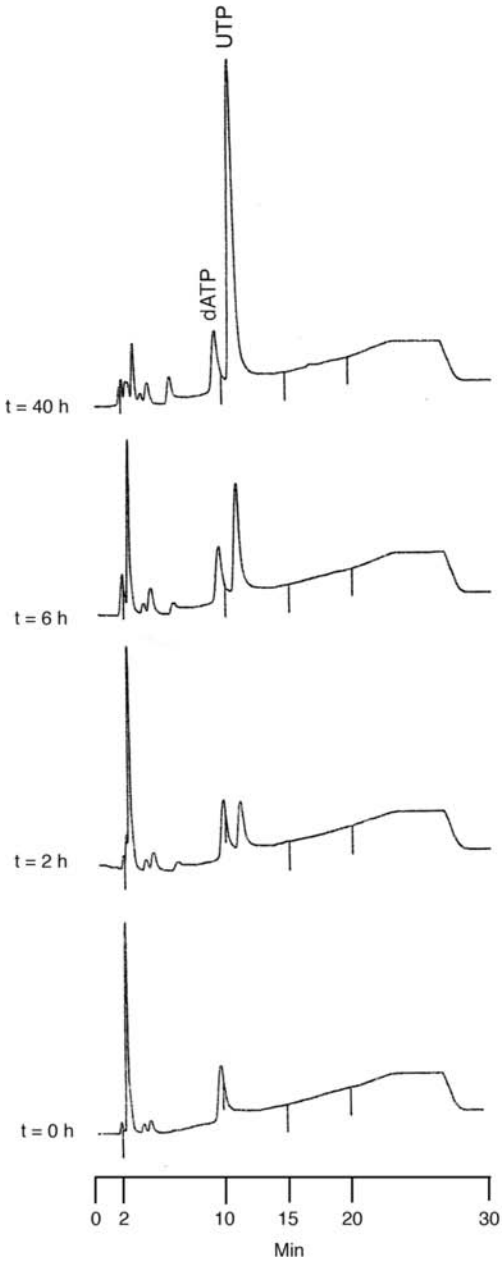


Figure 3. HPLC chromatograms showing the time course of the U-^{13}C-UTP-forming reaction. The positions of the catalytic cofactor dATP and the product UTP are indicated.

additionally contained 4 mM glutamine and 20 mM βME. Dialyzed enzymes were stored at −20 °C and were active for at least 6 months with the exception of *pyrG* and *pydA* which began to lose activity after ∼2 weeks. Yields ranged from ∼30− 100 mg of purified protein per liter of culture depending on the enzyme, but different preparations of each enzyme were generally consistent. Supporting Information Table S1 lists typical yields and specific activities for each de novo pyrimidine enzyme.

Enzymatic Assays. Enzymatic activity is reported in units (U), where 1 U is the amount of enzyme required to convert 1 μmol of substrate to product per minute. Specific activity is reported as U per mg of enzyme. Enzymatic activities were determined for all enzymes except *pyrI/B*, *pyrC*, and *pydA* by coupling the reaction to the consumption of ATP monitored by the change in A_{340} due to the action of pyruvate kinase and lactate dehydrogenase ADP and NADH.[32] For the other enzymes, the estimated specific activity was assumed to be 0.5 U/mg. Table 1 shows the amount of enzyme added in units or milligrams to each synthesis.

Nucleotide Synthesis. Synthesis of $^{13}C_{2,1',2',3',4',5'}$- UTP. Stoichiometric substrates: 18 mg (100 μmol) U-^{13}C-glucose, 20 mg (200 μmol) NaH$^{13}CO_3$, 54 mg (1 mmol) NH_4Cl, 34.2 mg (200 μmol) $^{13}C_1$-aspartate, 32 mg (200 μmol) sodium fumarate. Fuel reagents: 90.4 mg (400 μmol) α-ketoglutarate, 0.5 g (1.55 mmol) creatine phosphate. Catalytic cofactors: 0.7 mg (1 μmol) $NADP^+$, 5.5 mg (10 μmol) dATP. Substrates and cofactors were combined in 20 mM $MgCl_2$, 20 mM DTT, 50 mM KCl, 100 μg/mL ampicillin, 50 μg/mL kanamycin. The pH was adjusted to 9 with 1 M KOH and the solution was thoroughly flushed with Argon. Enzymes were added 0.17 U *cpkA*, 1.3 mg *pyrI/B*, 1.5 mg *pyrC*, 1.6 mg *pydA*, 0.78 U *pyrE*, 0.86 U *pyrF*, 0.94 U *cmk*, 3.5 U *ckmT*, 0.48 U *hxkA*, 0.53 U *zwf*, 0.58 U *gndA*, 0.64 U *rpiA*, 0.70 U *prsA*, 2.4 U *gdhA*, 2.6 U *plsA* to give a final volume of 10 mL. Argon was gently bubbled again, and the reaction was sealed in a glass flask. This label was synthesized for use in the C_2- labeling efficiency experiment, shown in Supplementary Figure 8.

Synthesis of U-^{15}N, $^2H_{5,\ 3',4',5',\ 5''}$-UTP. Stoichiometric substrates: 140 mg (0.75 mmol) $^2H_{1,2,3,4,5,6,6}$-glucose, 413 mg (7.5 mmol) $^{15}NH_4Cl$. Fuel reagent: 0.5 g (1.5 mmol) creatine phosphate. Catalytic cofactor: 41 mg (0.75 mmol) dATP. Substrates and cofactor were combined in 20 mM $MgCl_2$. The pH was adjusted to 8.5 with 1 M KOH. Enzymes were added: 3.6 U *hxkA*, 15 U *pgi*, 4 U *plsA*, 6 U *ckmT* to a final volume of 50 mL. The reaction was incubated for 90 h to allow sufficient exchange with solvent to protonate the C2 position of glucose. After ∼90 h the remaining reaction components for UTP synthesis were added. Stoichiometric substrates: 150 mg (1.5 mmol) $KHCO_3$, 263 mg (1.5 mmol) ^{15}N, $^2H_{2,3,3}$-aspartic acid, 240 mg (1.5 mmol) sodium fumarate. Fuel reagents: 4 g (12 mmol) creatine phosphate, 0.68 g (3 mmol) α-ketoglutarate. Catalytic cofactor: 5.5 mg (7.5 μmol) $NADP^+$. Substrates and cofactors were combined in the original solution. 100 μg/mL ampicillin, 50 μg/mL kanamycin, and 20 mM DTT were added and pH was readjusted to 8.5 with 1 M KOH. The remaining enzymes were added as listed in Table 1 (1) to give a final volume of 75 mL.

Synthesis of U-^{15}N, $^2H_{5,\ 3',4',5',\ 5''}$-CTP. 125 μmoles purified U-^{15}N, $^2H_{5,\ 3',4',5',\ 5''}$-UTP, 7 mg (12.5 μmoles) dATP, 380 mg (6.9 mmol) $^{15}NH_4Cl$, and 327 mg (1 mmol) creatine phosphate were combined in 8 mM $MgCl_2$, 4 mM KH_2PO_4/K_2HPO_4 (pH 7.6) with 40 μg/mL ampicillin and 20 μg/mL kanamycin. Enzymes were added as listed in Table 1 (2) give a final volume of 125 mL.

Synthesis of $^{13}C_{1',6'}$- UTP. Stoichiometric substrates: 56 mg (300 μmol) $^{13}C_2$-glucose, 60 mg (600 μmol) $KHCO_3$, 247 mg (1.4 mmol) 42% L/58% D- $^{13}C_2$- aspartic acid, 162 mg (3 mmol) NH_4Cl, 96 mg (600 μmol) sodium fumarate. Fuel reagents: 271 mg (1.2 mmol) α- ketoglutarate, 1.3 g (4.2 mmol) creatine phosphate. Catalytic cofactors: 2.2 mg (3 μmol) $NADP^+$, 16.5 mg (30 μmol) dATP. Substrates and cofactors were combined in 20 mM $MgCl_2$, 20 mM DTT, 100 μg/mL ampicillin, 50 μg/mL kanamycin and the pH was adjusted to 8.5 with 1 M KOH. Enzymes were added according Table 1(3) giving a final volume of 30 mL.

Synthesis of U-^{13}C, ^{15}N- UTP. Stoichiometric substrates: 139.5 mg (0.75 mmol) U-^{13}C-glucose, 151.5 mg (1.5 mmol) NaH$^{13}CO_3$, 262.5 mg (1.5 mmol) U-^{15}N, ^{13}C- L-aspartic acid, 412 mg (7.5

(32) Scott, L. G.; Tolbert, T. J.; Williamson, J. R. *Methods Enzymol.* **2000**, *317*, 18–38.
(33) Peterson, R. D.; Theimer, C. A.; Wu, H.; Feigon, J. *J. Biomol. NMR* **2004**, *28*, 59–67.
(34) Zwahlen, C.; Legault, P.; Vincent, S. J. F.; Greenblatt, J.; Konrat, R.; Kay, L. E. *J. Am. Chem. Soc.* **1997**, *119*, 6711–6721.
(35) Grzesiek, S.; Bax, A. *J. Am. Chem. Soc.* **1993**, *115*, 12593–12594.
(36) Piotto, M.; Saudek, V.; Sklenar, V. *J. Biomol. NMR* **1992**, *2*, 661–665.
(37) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277–293.
(38) Carlomagno, T.; Amata, I.; Williamson, J. R.; Hennig, M. *Biomol. NMR Assign.* **2008**, *2*, 167–169.

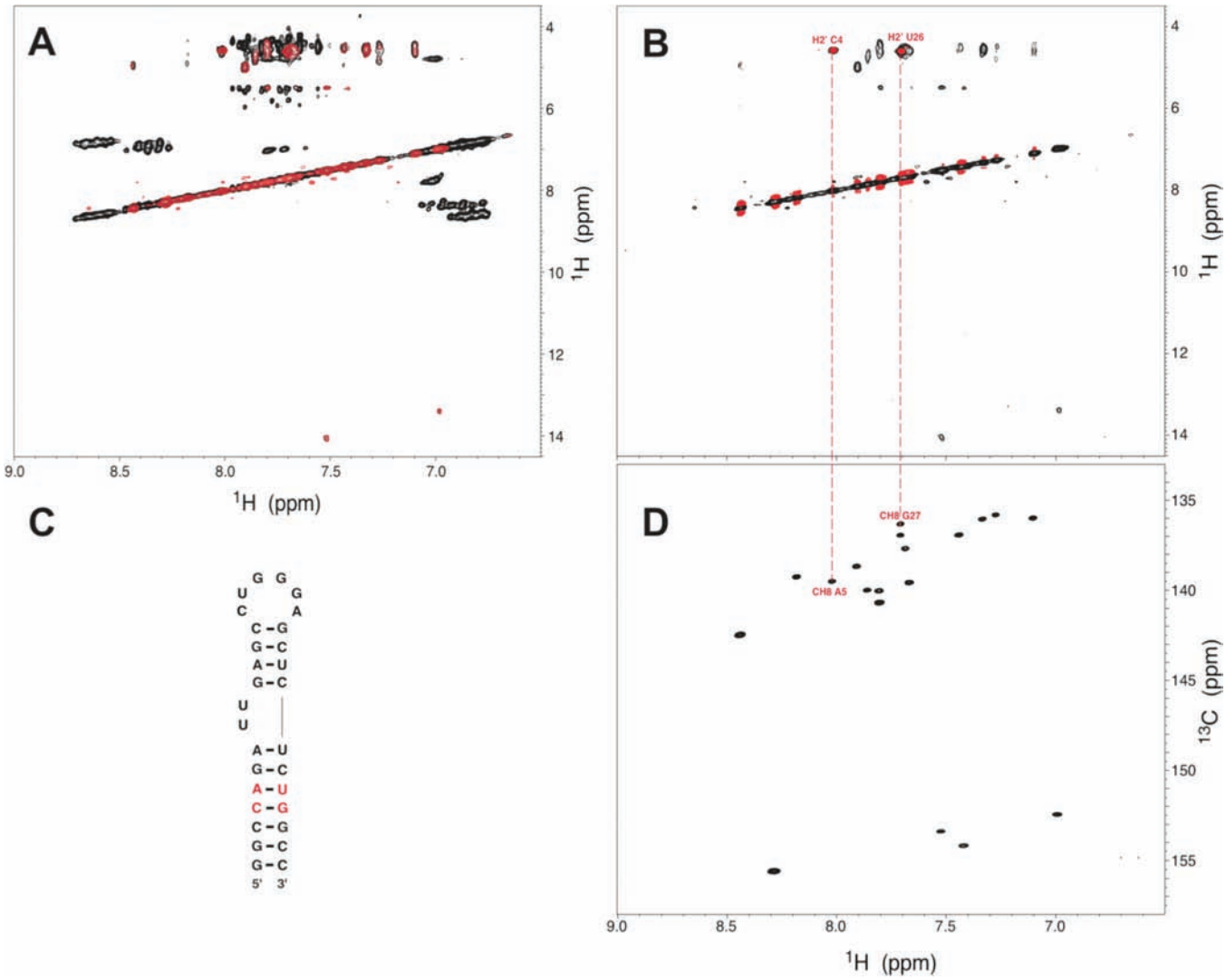


Figure 4. Filtered-edited NOESY experiments on the HIV-2 RNA synthesized from U-^{15}N, $^2H_{5,3',4',5',5''}$- pyrimidine nucleotides and U-^{13}C-purine nucleotides reveal the two pyrimidine-purine dinucleotide helical segments in HIV-2 TAR. (A) The homonuclear NOESY for HIV-2 TAR is shown in black with ^{13}C-edited spectra overlaid in red. (B) The ^{13}C,^{15}N filtered-edited (red) and ^{13}C-edited (black) NOESY spectra reveals two correlations corresponding to the close special proximity (~2 Å) of the $H_{2'}$ of a pyrimidine and the H_8 of a subsequent purine in A-form helix: $H_{2'}$ C4—H_8 A5 and $H_{2'}$ U26—H_8 G27. (C) The secondary structure of HIV-2 TAR is shown where the nucleotides highlighted in the filtered-edited experiment are also in red. (D) ^{1}H, ^{13}C -HSQC of the aromatic protons from the ^{13}C -labeled purines. Resonance assignments were transferred from published values,[38] which were determined under similar conditions.

mmol) $^{15}NH_4Cl$, 240 mg (1.5 mmol) sodium fumarate. Fuel reagents: 678 mg (3 mmol) α-ketoglutarate, 3.4 g (10.5 mmol) creatine phosphate. Catalytic cofactors: 5.5 mg (7.5 μmol) $NADP^+$, 40.1 mg (75 μmol) dATP. Substrates and cofactors were combined in 20 mM $MgCl_2$, 20 mM DTT, 100 μg/mL ampicillin, 50 μg/mL kanamycin and the pH was adjusted to 8.5 with 1 M KOH. Enzymes were added according Table 1(4) giving a final volume of 75 mL.

Synthesis of U-^{13}C- UTP. Stoichiometric substrates: 139.5 mg (0.75 mmol) U-^{13}C-glucose, 151.5 mg (1.5 mmol) $NaH^{13}CO_3$, 253 mg (1.4 mmol) U-^{13}C- L-aspartic acid, 405 mg (7.5 mmol) NH_4Cl, 240 mg (1.5 mmol) sodium fumarate. Fuel reagents: 689 mg (3 mmol) α-ketoglutarate, 3.5 g (10.7 mmol) creatine phosphate. Catalytic cofactors: 5.5 mg (7.5 μmol) $NADP^+$, 40.1 mg (75 μmol) dATP. Substrates and cofactors were combined in 20 mM $MgCl_2$, 22 mM DTT, 100 μg/mL ampicillin, 50 μg/mL kanamycin and the pH was adjusted to 8.5 with 1 M KOH. Enzymes were added according Table 1(5) giving a final volume of 75 mL.

UTP and CTP formation was monitored by HPLC using a Vydac nucleotide ion exchange column (250 mM × 4.6 mM), using a linear gradient of Buffer A (25 mM Na_2HPO_4/ NaH_2PO_4 (1:1) adjusted to pH 2.8 with acetic acid) and Buffer B (125 mM Na_2HPO_4/NaH_2PO_4 (1:1) adjusted to pH 2.9 with acetic acid) over 30 min at a flow rate of 2 mL/min. NTPs were monitored at 260 nm. HPLC chromatograms depicting the time course for each synthesis are included in Supporting Information.

Nucleotide Purification. The reaction was sterilized by filtration through a 0.2 μm filter, ammonium bicarbonate was added to a final concentration of 0.5 M, and the pH was adjusted to 10 with ammonium hydroxide. The solution was filtered again and loaded to a 20 g column of Affi-gel 601 (Biorad) boronate affinity resin equilibrated with 0.5 M ammonium bicarbonate pH 10 at 4 °C. The column was washed with the same buffer and the nucleotides were eluted with water acidified with CO_2. The products were verified by HPLC, NMR, and mass spectrometry. Final HPLC chromatograms and mass spectra are given in Supporting Information.

RNA Synthesis. HIV-2 TAR RNA (5′GGCCAGAUUGAGCCUGGGAGCUCUCUGGCC3′) was synthesized by in vitro transcription with T7 RNA polymerase using a mixture of U-^{13}C, ^{15}N-ATP purchased from Cambridge Isotope Laboratories, U-^{13}C, ^{15}N-GTP synthesized as described previously,[16] and U-U-^{15}N, $^2H_{5,\ 3',4',5',\ 5''}$-UTP and U-^{15}N, $^2H_{5,\ 3',4',5',\ 5''}$-CTP. RNA was synthesized in a 20 mL reaction under optimized conditions: 8 mM total NTPs (2 mM

each), 40 mM Tris HCl (pH 8.1), 0.1 mM spermidine, 10 mM DTT, 24 mM $MgCl_2$, 0.001% Triton X-100, 80 mg/mL poly(ethylene glycol) (8000 MW), 300 nm each DNA strand (Invitrogen), and 1 mg/mL T7 RNA polymerase, incubated at 37 °C for 5 h. The RNA was purified on denaturing 20% polyacrylamide gels, electro-eluted and desalted, lyophilized and diluted in 10 mM Na_2HPO_4/NaH_2PO_4 (pH 6.5), 150 mM NaCl, 10% D_2O for recording NMR spectra.

NMR Experiments. The spectral benefits of site-specifically labeling base residues with ^{13}C and ^{15}N isotopes were assessed with two-dimensional correlation spectroscopy. All spectra were collected at 25 °C on a 900 MHz Bruker Avance spectrometer equipped with a triple resonance 5 mm TXI-HCN probe with triple-axis gradients. A ^{13}C, ^{15}N-filtered-edited NOESY spectrum of type "F1fF2e" was optimized for aromatic C−H correlations, acquired using a 250 ms mixing time according to published pulse programs,[33,34] and compared to the analogous ^{13}C−F2-edited NOESY spectrum. ^{1}H, ^{15}N-, and ^{1}H, ^{13}C-HSQC experiments were also collected. Water suppression for most of the edited experiments was achieved through the use of flip-back pulses and the Watergate scheme.[35,36] ^{1}H, ^{13}C-HSQC spectra employed echo/antiecho gradient selection to remove the water signal. All spectra were referenced to 2,2-dimethyl-2-silapentane-5-sulfonate, processed with NMRPipe and viewed in NMRDraw.[37]

Acknowledgment. The authors thank Zahra Shajani for helpful discussions and Dee-Hua Huang for collection of the NMR spectra shown in Supporting Information Figure S8. This work was supported by a grant from the NIH (P50-GM082545).

Supporting Information Available: HPLC chromatograms and mass spectra for each nucleotide synthesis, ^{13}C NMR spectra for C_2-^{13}C labeling efficiency, ^{1}H, ^{15}N- and ^{1}H, ^{13}C-HSQC spectra of specifically labeled HIV-2 TAR, and a table of enzyme yields and specific activities. This material is available free of charge via the Internet at http://pubs.acs.org.

JA1059685