Published in final edited form as:
Chembiochem. 2014 July 21; 15(11): 1573–1577. doi:10.1002/cbic.201402130.

Regio-selective Chemical-Enzymatic Synthesis of Pyrimidine Nucleotides Facilitates RNA Structure and Dynamics Studies

Luigi J. Alvarado[a], **Regan M. LeBlanc**[a], **Andrew P. Longhini**[a], **Sarah C. Keane**[b], **Niyati Jain**[c], **Zehra F. Yildiz**[d], **Blanton S. Tolbert**[c], **Victoria M. D'Souza**[d], **Michael F. Summers**[b], **Christoph Kreutz**[e], and **T. Kwaku Dayie**[a,*]

[a] Center for Biomolecular Structure & Organization Department of Chemistry & Biochemistry University of Maryland 1115 Biomolecular Sciences Building College Park, MD 20782 (USA)

[b] Howard Hughes Medical Institute at University of Maryland Baltimore Country Baltimore, MD 21250 (USA)

[c] Case Western Reserve University Cleveland, OH 44106 (USA)

[d] Department of Molecular and Cellular Biology Harvard University Cambridge, MA 02138 (USA)

[e] Institute of Organic Chemistry and Center for Molecular Biosciences Innsbruck (CMBI) University of Insbruck 6020 Innsbruck, Austria

Abstract

Isotope labeling revolutionized NMR studies of small nucleic acids, but to extend this technology to larger RNAs requires site-specific labeling tools to expedite NMR structural and dynamics studies. Using enzymes from the pentose phosphate pathway, we couple chemically synthesized uracil nucleobase with specifically ^{13}C-labeled ribose to synthesize both UTP and CTP with nearly quantitative yields. This chemo-enzymatic method affords a cost-effective preparation of labels that are unattainable by current methods. The methodology generates versatile ^{13}C and ^{15}N labeling patterns which, when employed with relaxation-optimized NMR spectroscopy, effectively mitigates problems of rapid relaxation that result in low resolution and sensitivity. The methodology is demonstrated with RNAs of various sizes, complexity, and importance: the exon splicing silencer 3 (27 nt), iron responsive element (29 nt), Pro-tRNA (76 nt), and HIV-1 core encapsidation signal (155 nt).

Keywords

RNA; NMR spectroscopy; Labeling; Structure; Dynamics

Ribonucleic acid (RNA) plays critical roles in essentially all forms of life including signaling, gene regulation, catalysis, and retroviral infection.[1–5] RNA partakes in such a broad range of functions because it can adopt intricate and pliable three-dimensional

*Dayie@umd.edu.

Supporting information for this article is available

structures. However there is a dearth of structural dynamics information on RNA, even though RNA functions are dependent on structural dynamics.

The reasons for this paucity of structural information include conformational heterogeneity, uniform negative surface charge, and extensive chemical shift overlap and fast relaxation properties of the constituent nuclei (^{1}H, ^{13}C, ^{15}N, ^{31}P). These problems give rise to low resolution and sensitivity. The chemical shift overlap problem has been partially addressed by uniform labeling and fast relaxation by transverse relaxation optimized spectroscopy (TROSY).[6–8] However, uniform labeling introduces direct one-bond and residual dipolar ^{13}C-^{13}C couplings that exacerbate resolution, sensitivity, and linewidth problems, and hinder accurate measurement of ^{13}C relaxation parameters.[9]

To address the limitations of uniform labeling, five approaches are utilized to obtain site-specifically and isotopically labeled RNA. (i) Total chemical synthesis of nucleotides, followed by solid-phase synthesis of RNA using phosphoramidite chemistry. This method is powerful for incorporation of isotopes into RNA at any desired position, but it faces problems of regio- and stereo-selectivity during phosphoramidite synthesis, and low yields (<10 %) that drop significantly for larger oligonucleotides > 50 nt. [10–12] (ii) *De novo* biosynthesis of NTPs, followed by *in vitro* RNA transcription. This robust technique produces labeled nucleotides at significant yields (<60 %),[13] but comes at the cost of using ~18-23 enzymes, expensive precursor substrates, and inaccessibility to some labeling patterns: labeling the ribose C1′ and C5′ in the context of pyrimidine C4, C5, and C6 (readily done with our new method) requires $^{13}C_2$-2,6-glucose that is not commercially available and ^{13}C-labeled D/L-aspartic acid that is prohibitively costly. (iii) Phosphorylated biomass-produced NMPs used for *in vitro* RNA transcription. This method again provides useful labeled nucleotides; however, the overall yield is low per labeled input metabolite, and isotopic scrambling often leads to inadequate suppression of ^{13}C-^{13}C coupling.[7,14,15] (iv) Phosphorylated selective biomass-produced NMPs used for *in vitro* RNA transcription. Again this method overcomes the isotopic scrambling problem with adequate suppression of ^{13}C-^{13}C couplings; however, low overall yields remain an issue.[16–19] (v) Chemo-enzymatic synthesis of NTPs, followed by *in vitro* RNA transcription. This approach is potentially the most versatile method available. However until now, lack of commercially available selectively labeled ribose and base precursors has hampered the realization of its full potential.[20]

Here we combine an improved organic synthetic approach that selectively places labels in the pyrimidine nucleobase (either ^{15}N1, ^{15}N3, ^{13}C2,^{13}C4, ^{13}C5, or ^{13}C6 or any combination) and a very efficient enzymatic method to couple ribose with uracil to produce previously unattainable labeling patterns (Scheme 1) containing isolated two-spin systems in both the ribose and the nucleobase. We show that these labels are ideal for both structural and dynamic studies for large RNAs.

Using the method of Santa Lucia and co-workers,[21] we synthesized 6-$^{13}C_1$-1,3-^{15}-uracil, **3**, using K^{13}CN, bromoacetic acid and $^{15}N_2$-urea (Scheme 1) with yields >82 %.[10] The identity of **3** was confirmed by NMR (SI-1).

To demonstrate the versatility of this approach, we coupled 1′,5′-$^{13}C_2$-D-ribose (Omicron Biochemicals, Southbend, IN) to **3** in enzymatic condensation reactions (Scheme 1) using the following three/four enzymes from the pentose phosphate pathway: Ribokinase (E.C. 2.7.1.15), phosphoribosyl pyrophosphate synthetase (E.C. 2.7.6.1); uridine phosphoribosyl transferase (E.C. 2.4.2.9); and CTP synthetase (E.C. 6.3.4.2).[6,22–27] All four enzymes were expressed and purified as described in our previous work.[28]

The *in vitro* one-pot, two-step synthesis of UTP, **4**, was accomplished quickly, in high yields, and with extreme ease. First, UMP was synthesized from 1′,5′-$^{13}C_2$-D-ribose and **3** within 5 h (SI-2). This labeled UMP was readily converted to **4** *in situ* using nucleoside monophosphate kinase. The final yield of **4** was ~90% with respect to starting uracil (Supplementary methods). The identity of **4** was confirmed by HPLC and NMR (SI-2). From UTP, CTP, **5**, was directly synthesized and purified (Scheme 1).[29] The yield for CTP from UTP was nearly quantitative after 6 hours (SI-3) and the recovery of isolated product after purification was nearly 90% with respect to UTP.

These site-specifically labeled nucleotides were then used for *in vitro* transcriptions of four RNAs of interest (Scheme 2): (i) exon splicing silencer 3 (ESS3, 27 nt) involved in inhibition of splicing through occlusion of the spliceosome assembly;[30] (ii) iron responsive element (IRE, 29 nt), involved in cellular iron homeostasis;[31] (iii) Pro-tRNA (76 nt) as a primer for reverse transcription for murine leukemia virus; [32] and (iv) HIV-1 core encapsidation signal (155 nt) involved in genomic packaging.[33]

The site-specific labeling of these RNAs significantly reduced spectral crowding, eliminated ^{13}C-^{13}C scalar couplings, increased signal-to-noise ratios, and facilitated direct carbon detection experiments. The average linewidth improvement factor (the ratio of the average ^{1}H linewidth in a regular HSQC to a CH-/CH_2-optimized TROSY experiment) was ~1.7 (with a range from 1.1 to 2.5) for C5′-methylene region and 1.9 (with a range of 1.0 to 3.0) for C6-methine (Figure 2a,b,d insets; SI-4).[34] Compared to traditional HSQC experiments, TROSY experiments afforded linewidth and sensitivity improvement (Figure 1).

In addition to these spectral quality improvements, our labels open up new avenues for structural analysis of RNA. In a standard duplex RNA helix, H1′ to H5′ and H5′ to H6/H8 nuclear Overhauser effect (NOE) contacts are often difficult to analyse because of spectral crowding in the C5′ region due to the limited chemical shift dispersion of H2′, H3′, H4′, and H5′/H5″ resonances (Figure 2).[35] With our new labels, we demonstrate that the interaction between sequential nucleotides “i” and “i+1” involving the pairs $H1_i'$-$H5'_i$/$H6_i$-$H1'_{(i+1)}$ can be readily observed in a 3D NOESY-HSQC experiment (Figure 2b,c). The expected sequential NOEs are clearly visible and permit sequential assignments to be made readily. With this approach we unambiguously traced and assigned all the H1′, H5′, H5″, and H6 proton resonances with contiguous U-C stretches in IRE RNA.

Another set of NMR applications particularly suited for our site-specific labels are direct ^{13}C detection experiments. Advances in NMR probe technology enable acquisition of NMR experiments that bypass the fast-relaxing ^{1}H bottleneck.[36,37] Unlike

uniformly $^{13}C/^{15}N$ labeled samples, applying our selective labels is advantageous for the following reasons: implementation does not require IPAP, S^3E, selective pulse implementation, or constant time elements for carbon decoupling in the direct dimension (SI-4). Similarly, the transfer efficiency is not adversely affected by undesired coherence transfers such as C1′↔C2′, C5↔C6, and N1↔C2 without the use of selective pulses (SI-5). For instance, in a selectively U,C-labeled IRE RNA we observed all the expected C1′-N1 and C6-N1 correlations in a 2D carbon-detected CN experiment (Figure 3).

Finally, we also show that RNAs transcribed with our custom nucleotides have better derived NMR dynamics parameters (e.g. T_1, $T_{1\rho}$). For instance, both IRE and ESS3 RNAs showed a near perfect data fit for T_1 and $T_{1\rho}$ relaxation experiments, respectively, with our site-specific labels (Figure 4a,c). From our relaxation experiments, best-fit parameters for all RNAs followed a similar pattern, where uniformly labeled RNA consistently had larger standard errors in the measurements.

Next, we compared the ns-ps dynamics of both C1′-H1′ and C6-H6 bond vectors as a ratio of R_2/R_1 decay rate constants. Both IRE and ESS3 RNAs showed a few residues with different decay rate values between uniformly and selectively labeled RNA (Figure 4b,d). Our results underscore the importance of using site-specifically labeled nucleotides to study the structure and dynamics of RNA, especially for RNAs >25 nt. Experimental parameters are detailed in the supplementary information (SI-6).

In conclusion, using a chemical enzymatic method, we have designed an efficient and flexible method for site-specific labeling of gram quantities of pyrimidine nucleotides with up to 90 % yield (based on input uracil). These labeled ribonucleotides enable facile measurement of a range of NMR dynamics (T_1, $T_{1\rho}$) and structural parameters in RNA. In addition, our labels afford NMR spectra with improved linewidths and signal-to-noise ratios for RNAs ranging in size from 27- to 155-nt. Finally, the use of these labels would (i) enable accurate measurement of residual dipolar couplings,[38] (ii) improve the acquisition of RNA heteronuclear chemical shifts to facilitate rapid identification of noncanonical RNA structures within RNA-ligand/protein interaction interfaces,[39] (iii) be useful for solid-state NMR of RNA,[40,41] and (iv) be easily expanded to deuterate C5, C6 or both positions.

Experimental Section

Buffers and reagents

All buffers were prepared with chemicals bought from Sigma-Aldrich or Fisher Scientific. 1′,5′-$^{13}C_2$-D-ribose was purchased from Omicron Biochemicals.

Protein expression and purification

Ribokinase (E.C. 2.7.1.15), Phosphoribosyl pyrophosphate synthetase (E.C. 2.7.6.1), Uridine phosphoribosyl transferase (E.C. 2.4.2.9), and CTP synthethase (E.C. 6.3.4.2) were all expressed, purified and assayed for activity as previously described.[28]

UTP synthesis

The site-specifically labeled 1′,5′,6-$^{13}C_3$-1,3-$^{15}N_2$-Uridine triphosphate was synthesized by phosphorylation of uridine monophosphate (UMP); UMP was synthesized from uracil, ribose, and dATP with an overall yield of ~ 90 % with respect to input uracil (Supplementary information).

CTP synthesis

The site-specifically labeled 1′,5′,6-$^{13}C_3$-1,3-$^{15}N_2$-Cytidine triphosphate was enzymatically synthesized *in vitro* in a single-step reaction as previously described.[13] Our typical yield was 95 % (Supplementary information).

RNA synthesis and purification

All four RNAs (ESS3, IRE, Pro-tRNA, and HIV-1) were synthesized *in vitro* following standard published protocols. RNAs were then purified by denaturing polyacrylamide gel electrophoresis, followed by electroelution. Finally, RNAs were dialyzed into their buffer of choice (Supplementary information).

Nuclear magnetic resonance spectroscopy

NMR experiments were carried out on 0.2 to 1 mM RNA samples. Various 1D and 2D proton detected and direct carbon-detect CN experiments were collected at 25 °C on either a 600 and/or 800 MHz Bruker Avance III spectrometer equipped with a HCN triple resonance cryoprobe. All NMR data were processed using TopSpin 3.2, NMRpipe/NMRDraw, and NMRviewJ (Supplementary information, Tables S1 and S2).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

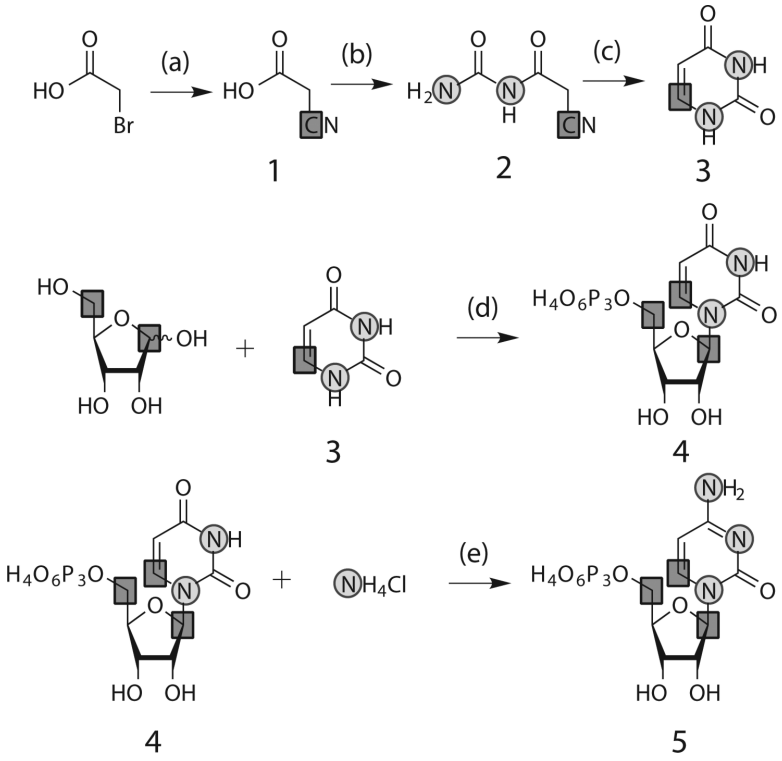
Acknowledgments

We thank the National Institute for General Medical Sciences (P50 GM103297 to BST, MFS, TKD, and VMD), the National Science Foundation (DBI1040158 to TKD), and the Austrian Sciences Fund (I844 to CK).

References

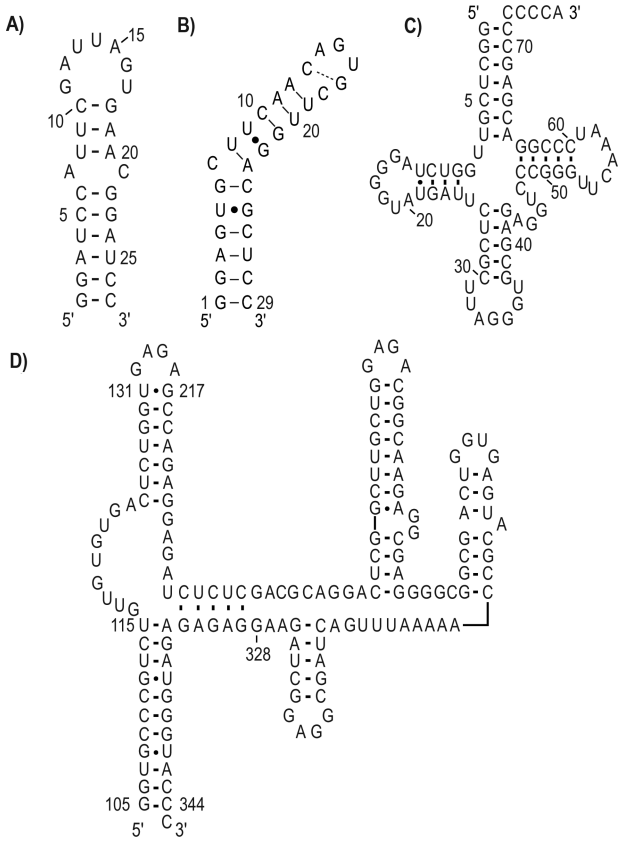
1. Steitz TA. Nat. Rev. Mol. Cell Biol. 2008; 9:242–253. [PubMed: 18292779]
2. Breaker RR. Future Microbiol. 2009; 4:771–773. [PubMed: 19722830]
3. Serganov A, Nudler E. Cell. 2013; 152:17–24. [PubMed: 23332744]
4. Mattick JS. J. Exp. Biol. 2007; 210:1526–1547. [PubMed: 17449818]
5. Lu K, Heng X, Garyu L, Monti S, Garcia EL, Kharytonchyk S, Dorjsuren B, Kulandaivel G, Jones S, Hirem ath A, et al. Science. 2011; 334:242–245. [PubMed: 21998393]
6. Scott LG, Tolbert TJ, Williamson JR. Methods Enzymol. 2000; 317:18–38. [PubMed: 10829270]
7. Nikonowicz EP, Sirr A, Legault P, Jucker FM, Baer LM, Pardi A. Nucleic Acids Res. 1992; 20:4507–4513. [PubMed: 1383927]
8. Pervushin K, Riek R, Wider G, Wüthrich K. Proc. Natl. Acad. Sci. U. S. A. 1997; 94:12366–12371. [PubMed: 9356455]
9. Dayie KT. Int. J. Mol. Sci. 2008; 9:1214–1240. [PubMed: 19325801]

10. Wunderlich CH, Spitzer R, Santner T, Fauster K, Tollinger M, Kreutz C. J. Am. Chem. Soc. 2012; 134:7558–7569. [PubMed: 22489874]
11. Milecki J. J. Label. Compd. Radiopharm. 2002; 45:307–337.
12. Quant S, Wechselberger RW, Wolter MA, Wörner K-H, Schell P, Engels JW, Griesinger C, Schwalbe H. Tetrahedron Lett. 1994; 35:6649–6651.
13. Schultheisz HL, Szymczyna BR, Scott LG, Williamson JR. J. Am. Chem. Soc. 2011; 133:297–304. [PubMed: 21166398]
14. Batey RT, Inada M, Kujawinski E, Puglisi JD, Williamson JR. Nucleic Acids Res. 1992; 20:4515–4523. [PubMed: 1383928]
15. Hoffman DW, Holland JA. Nucleic Acids Res. 1995; 23:3361–3362. [PubMed: 7667118]
16. Lemaster DM, Kushlan DM. J. Am. Chem. Soc. 2001; 118:9255–9264.
17. Johnson JE, Julien KR, Hoogstraten CG, Biomol J. NMR. 2006; 35:261–274.
18. Thakur CS, Sama JN, Jackson ME, Chen B, Dayie TK, Biomol J. NMR. 2010; 48:179–192.
19. Thakur CS, Dayie TK, Biomol J. NMR. 2012; 52:65–77.
20. Tolbert TJ, Williamson JR. J. Am. Chem. Soc. 1996; 118:7929–7940.
21. SantaLucia J, Shen LX, Cai Z, Lewis H, Tinoco I. Nucleic Acids Res. 1995; 23:4913–4921. [PubMed: 8532537]
22. Park J, van Koeverden P, Singh B, Gupta RS. FEBS Lett. 2007; 581:3211–3216. [PubMed: 17585908]
23. Li S, Lu Y, Peng B, Ding J. Biochem. J. 2007; 401:39–47. [PubMed: 16939420]
24. Sin IL, Finch LR, Bacteriol J. 1972; 112:439–444.
25. Lundegaard C, Jensen KF. Biochemistry. 1999; 38:3327–3334. [PubMed: 10079076]
26. Sinha SC, Smith JL. Curr. Opin. Struct. Biol. 2001; 11:733–739. [PubMed: 11751055]
27. Lunn FA, MacDonnell JE, Bearne SL. J. Biol. Chem. 2008; 283:2010–2020. [PubMed: 18003612]
28. Arthur PK, Alvarado LJ, Dayie TK. Protein Expr. Purif. 2011; 76:229–237. [PubMed: 21111048]
29. MacDonnell JE, Lunn FA, Bearne SL. Biochim. Biophys. Acta. 2004; 1699:213–220. [PubMed: 15158730]
30. Levengood JD, Rollins C, Mishler CHJ, Johnson CA, Miner G, Rajan P, Znosko BM, Tolbert BS. J. Mol. Biol. 2012; 415:680–698. [PubMed: 22154809]
31. Pantopoulos K, Porwal SK, Tartakoff A, Devireddy L. Biochemistry. 2012; 51:5705–5724. [PubMed: 22703180]
32. Harada F, Peters GG, Dahlberg JE. J. Biol. Chem. 1979; 254:10979–10985. [PubMed: 115865]
33. Heng X, Kharytonchyk S, Garcia EL, Lu K, Divakaruni SS, LaCotti C, Edme K, Telesnitsky A, Summers MF. J. Mol. Biol. 2012; 417:224–239. [PubMed: 22306406]
34. Miclet E, Williams DC Jr, Clore GM, Bryce DL, Boisbouvier J, Bax A. J. Am. Chem. Soc. 2004; 126:10560–10570. [PubMed: 15327312]
35. Wijmenga SS, van Buuren BNM. Prog. Nucl. Magn. Reson. Spectrosc. 1998; 32:287–387.
36. Bermel W, Felli IC, Kummerle R, Pierattelli R. Concepts Magn. Reson. Part A. 2008; 32A:183–200.
37. Bermel W, Bertini I, Felli IC, Pierattelli R. J. Am. Chem. Soc. 2009; 131:15339–15345. [PubMed: 19795864]
38. Ying J, Wang J, Grishaev A, Yu P, Wang Y-X, Bax A, Biomol J. NMR. 2011; 51:89–103.
39. Barton S, Heng X, Johnson BA, Summers MF, Biomol J. NMR. 2013; 55:33–46.
40. V Cherepanov A, Glaubitz C, Schwalbe H. Angew. Chemie. 2010; 49:4747–4750.
41. Marchanka A, Simon B, Carlomagno T. Angew. Chemie. 2013; 52:9996–10001.
42. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Swanstrom R, Burch CL, Weeks KM. Nature. 2009; 460:711–716. [PubMed: 19661910]



Scheme 1.

Chemical-enzymatic synthesis of UTP and CTP. [a]Reagents and conditions: (a) sodium carbonate, pH 9, 3 h at 80 °C, 20 h at rt; (b) urea in acetic anhydride, 30 min at 90 °C, (c) 5 % Pd/BaSO4 in 50 % aqueous acetic acid, 36 h at rt; (d) ribokinase, phosphoribosyl pyrophosphate synthetase, uridine phosphoribosyl transferase, nucleoside monophosphate kinase in sodium phosphate pH 7.5, 12 h at 37 °C; (e) cytidine triphosphate synthetase in Tris-HCl pH 8.0, 6 h at 37 °C. Squares: ^{13}C; Circles: ^{15}N.



Scheme 2.

Secondary structures for (a) ESS3 RNA (27 nt), (b) IRE RNA (29 nt), (c) Mouse Pro-tRNA (76 nt), and (d) HIV-1 core encapsidation signal RNA (155 nt, based on chemical probing of the intact genome by Watts et al.[42] and partially confirmed by NMR). Dash: Watson-Crick base pairing, Dot: Wobble base pairing.

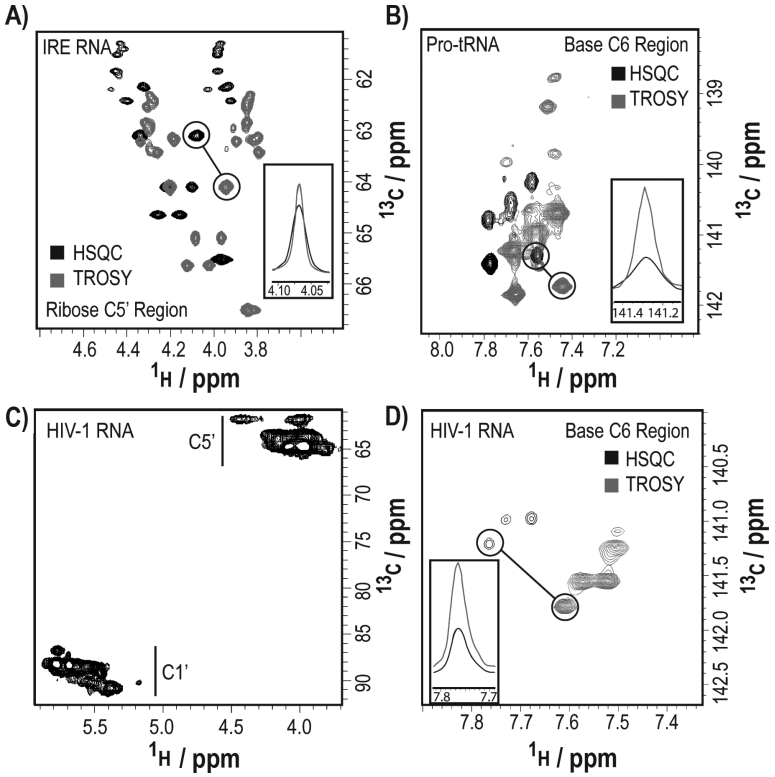


Figure 1.

Spectral resolution and improved signal-to-noise ratios for site-specifically labeled RNAs. (a) Comparison of HSQC (black) and CH2-TROSY (gray) spectra of the C5′ region of the site-specifically U,C-labeled IRE RNA (29 nt). (b) Comparison of HSQC (black) and CHTROSY (gray) spectra of the C6 region of the site-specifically U-labeled Pro-tRNA (76 nt). (c) HSQC (C1’ and C5’ regions) of the site-specifically U-labeled HIV-1 core encapsidation signal RNA (155 nt). (d) Comparison of HSQC (black) and CH-TROSY (gray) spectra of the C6 region of the site-specifically U-labeled HIV-1 core encapsidation signal RNA (155 nt).

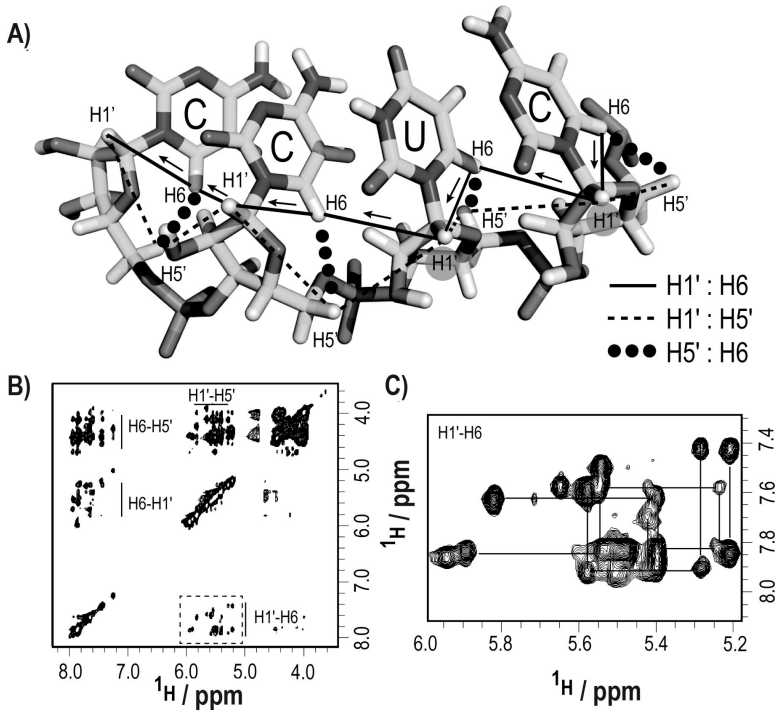


Figure 2.

Through-space NOE resonance assignments for the IRE RNA become straightforward with specifically labeled nucleotides. (a) Intra- and inter-nucleotide sequential ^{1}H walk between H1′, H5′, and H6 along a pyrimidine tract. (b) The spectral crowding in a 3D NOESYHSQC experiment is significantly reduced when using our specifically labeled nucleotides. (c) The NOE sequential walk between $H1'_{i}$-$H5'_{i}$/$H6_{i}$-$H1'_{(i+1)}$ was readily followed for all contiguous pyrimidine tracts. Experimental param eters are detailed in the supplementary information (SI-6).

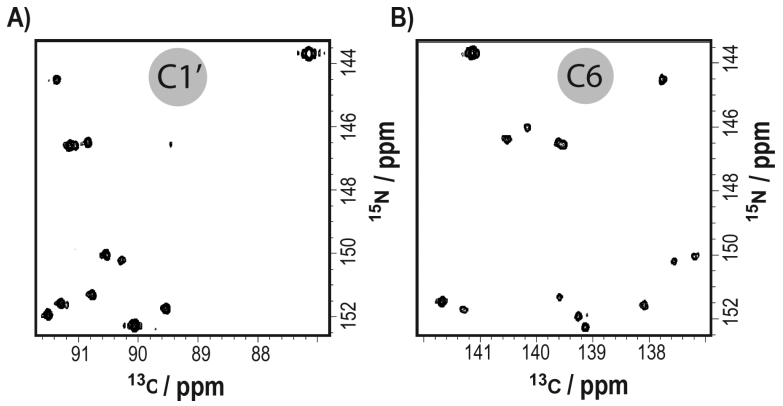


Figure 3.

Two-dimensional direct carbon-detected CN experiments of 0.7mM IRE RNA. (a) C1′-N1 region. (b) C6-N1 region.

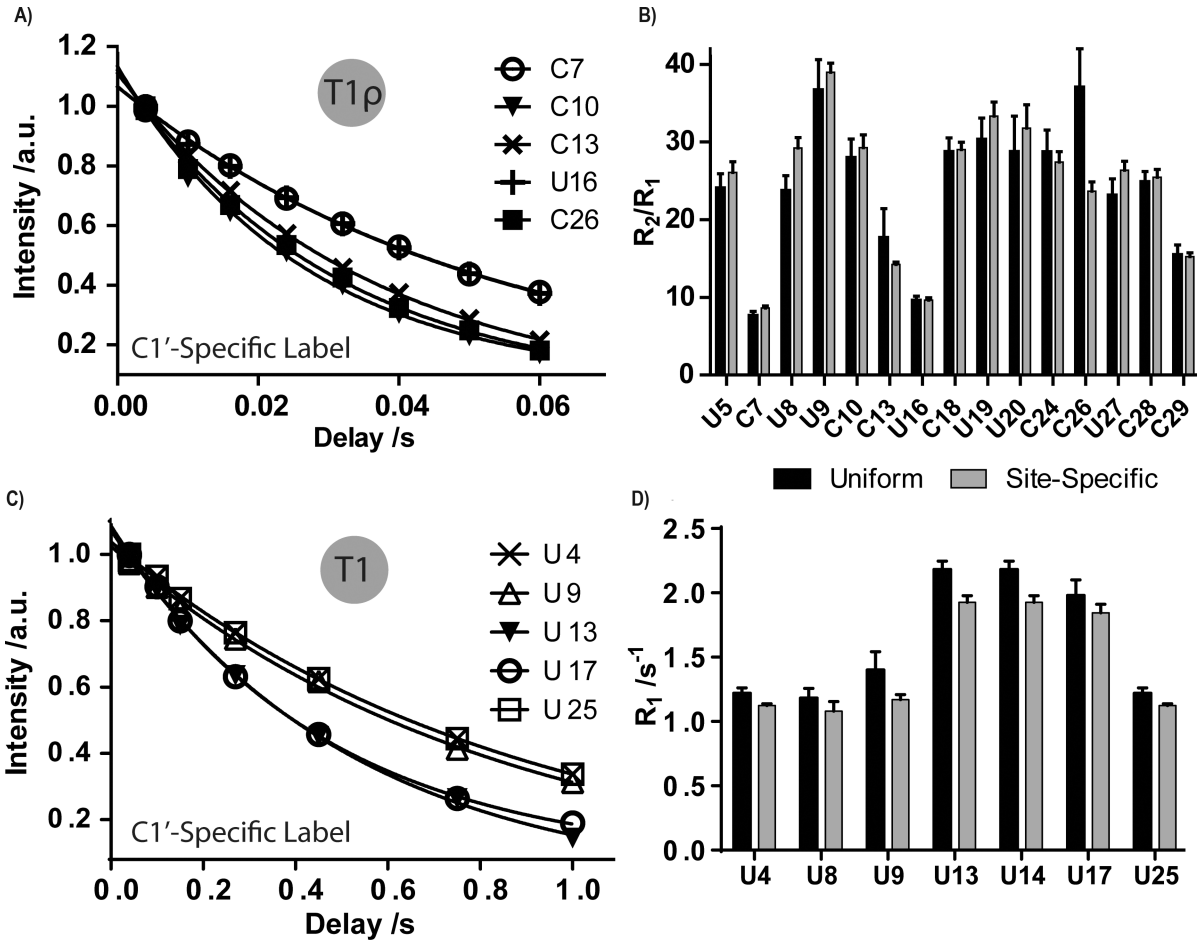


Figure 4.

Site-specific labeling improves dynamics measurements of the IRE and ESS3 RNAs. (a,c) Curve fits of $T_{1\rho}$ or T_1 experiments on five residues of the IRE RNA and ESS3 RNA, respectively. (b,d) Comparison of overall dynamics (R_2/R_1 ratio or R_1 values) of all U and/or C residues of IRE and ESS3 RNAs, respectively, when the molecule was $^{13}C/^{15}N$ uniformly or site-specifically labeled. In the IRE RNA, U8, C13, and C26 show significant differences in R2/R1 ratios. In ESS3 RNA, U9, U13, and U14 show significant differences in R_1 decay rates. Other residues remain unaffected. Errors bars are shown as standard deviation from three independent experiments.