JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

Pinpointing RNA-Protein Cross-Links with Site-Specific Stable Isotope-Labeled Oligonucleotides

Victor S. Lelyveld,[†,§,#] Anders Björkbom,[†,∥,§,○,#] Elizabeth M. Ransey,[▽] Piotr Sliz,[▽,⊥] and Jack W. Szostak*[,†,‡,§]

[†]Department of Molecular Biology and Center for Computational and Integrative Biology, Howard Hughes Medical Institute, Massachusetts General Hospital, Boston, Massachusetts 02114, United States

[‡]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, United States

[∥]Department of Biosciences, Åbo Akademi University, Åbo FI-20520, Finland

[§]Department of Genetics, [▽]Department of Biological Chemistry and Molecular Pharmacology, and [⊥]Department of Pediatrics, Harvard Medical School, Boston, Massachusetts 02115, United States

Supporting Information

ABSTRACT: High affinity RNA-protein interactions are critical to cellular function, but directly identifying the determinants of binding within these complexes is often difficult. Here, we introduce a stable isotope mass labeling technique to assign specific interacting nucleotides in an oligonucleotide-protein complex by photo-cross-linking. The method relies on generating site-specific oxygen-18-labeled phosphodiester linkages in oligonucleotides, such that covalent peptide-oligonucleotide cross-link sites arising from ultraviolet irradiation can be assigned to specific sequence positions in both RNA and protein simultaneously by mass spectrometry. Using Lin28A and a let-7 pre-element RNA, we demonstrate that mass labeling permits unambiguous identification of the cross-linked sequence positions in the RNA-protein complex.

Proteins recognize structured RNAs to form regulatory and catalytic complexes. UV and chemical cross-linking have been widely used to interrogate these biochemical interactions under native solution-phase conditions, but it is still challenging to pinpoint the specific interacting residues within the linear RNA and protein sequences.[1] Photo-cross-linking under ultraviolet illumination followed by purification and sequencing[2,3] or mass analysis[1] may yield significant structural insights, but neither method is, by itself, sufficient to unambiguously assign cross-link sites to specific side chains on both proteins and nucleic acids simultaneously. The incorporation of artificial photoreactive bases (e.g., 4-thiouridine or 6-thioguanosine) is useful to resolve ambiguities and increase the yield of covalent photoproducts.[3] It is, however, possible that the resultant cross-links do not precisely represent biologically relevant interactions since the modified nucleotides themselves may alter the interaction of interest.

Here we present an alternative approach based on site-specific stable isotope labeling that can be used for efficient identification of interacting sites by liquid chromatography-mass spectrometry (LC-MS). The technique relies on a facile, inexpensive, and fully automated method to generate individual ^{18}O-labeled phosphodiester linkages during oligonucleotide synthesis. These site-specifically mass labeled oligonucleotides can be prepared in a straightforward manner by standard solid-phase synthesis, such that no cumbersome organic synthesis is necessary to generate specific mass labeled DNA or RNA probes. Site-specifically labeled oligonucleotides can be cross-linked to interacting proteins (Figure 1A), and the resulting covalent product can be subjected to hydrolysis and analyzed by MS. Covalent nucleotide-labeled peptides prepared in this manner exhibit a unique isotopic distribution exclusively when an ^{18}O mass label is retained adjacent to the cross-link site.

To demonstrate the utility of site-specific RNA mass labeling, we examined the solution-phase interaction of a precursor fragment of the let-7 miRNA and Lin28A, a high affinity RNA-binding protein that is one of four critical regulators whose overexpression yields an induced pluripotent stem cell (iPS) state.[4] Mature let-7 plays a regulatory role in inflammation[5] and has been implicated in tumorigenesis.[6] Dicer activity on the let-7 precursor, prelet-7, releases a stem-loop fragment "pre-element," preE-let-7, and the mature let-7 miRNA. Binding of Lin28A to the preE region of prelet-7 inhibits let-7 microRNA processing.[7] Based on structural evidence, the interaction between let-7 precursor RNAs and Lin28A protein is mediated through contacts in both loop and stem regions of preE-let-7 (PDB 3TS0).[8] Consistent with crystallographic contacts between a truncated preE-let-7 RNA (dubbed preE$_M$-let-7f) and the cold shock domain of a loop-minimized Lin28A (Lin28ΔΔ), we recently observed[9] a native solution-phase cross-link between these domains by tandem MS.

To pinpoint the precise ribonucleotides in the let-7 pre-element that are photo-cross-linked to Lin28A upon UV exposure, we now report the use of stable isotope labeling of a pre-element RNA and high-resolution mass spectrometry. Our recent study implicated a uridine at the cross-linking site, most likely in the loop region of the pre-element hairpin.[9] We therefore synthesized the previously co-crystallized 25-nt stem-loop preE$_M$-let-7f RNA[8] carrying a single ^{18}O mass label on one of two possible uridine positions in the pre-element hairpin: U11 or U12 in the RNA sequence

Received: October 9, 2015
Published: November 19, 2015

DOI: 10.1021/jacs.5b10596
J. Am. Chem. Soc. 2015, 137, 15378−15381

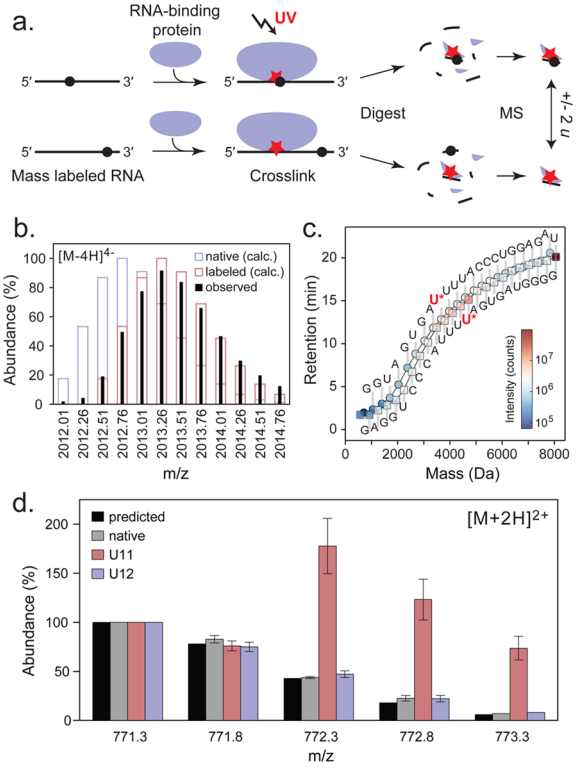
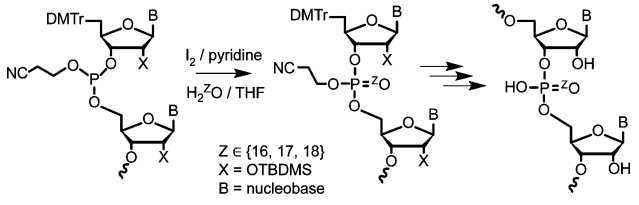


Figure 1. Identifying RNA-protein contacts with site-specific RNA mass labeling. (a) Sample preparation workflow. Oligonucleotides are prepared with selectively enriched ^{18}O phosphates at distinct positions (black circle), separately incubated with protein, UV cross-linked, hydrolyzed to peptide-nucleotide fragments, and analyzed by MS. (b) Calculated and observed isotopic distributions for $[M - 4H]^{4-}$ ion of 25-nt $preE_M$-let-7f RNA (GGGGUAGUGAU^{11}UUUACCCUGGAGAU) labeled with ^{18}O at the phosphodiester following U11. (c) Direct LC-MS sequencing of U11-labeled RNA, with the mass labeled position indicated as U* (red). (d) Isotope distribution of $[MGFGFLSMTAR + UMP + 2H]^{2+}$, a tryptic peptide ion derived from cross-linking Lin28A in the presence of $preE_M$-let-7f with natural abundance (native) or ^{18}O mass labeled at the 3′ phosphodiester following U11 or U12. All spectra are normalized to the abundance of the monoisotopic ion 771.3 *m*/*z*.

GGGGUAGUGAU^{11}U^{12}UUACCCUGGAGAU, where the mass label is 3′ to the indicated nucleoside in each case. Solid-phase oligonucleotide synthesis by the phosphoramidite method proceeds by iterative cycles of protected nucleoside phosphite addition to a deprotected terminal hydroxyl on the growing oligonucleotide chain, followed by rapid iodine-mediated oxidation to generate an O-protected phosphate triester (Scheme 1).[10,11] The source of oxygen equivalents is H_2O, typically applied to the solid support in a solution of I_2, pyridine, and tetrahydrofuran (THF). This oxidation method has been used to generate regiospecifically isotope-labeled DNA dinucleotides and short oligonucleotides for NMR studies using isotope-enriched water.[12−14] It has also recently been shown that the oxidation mix can be completely substituted with one formulated with enriched water ($H_2{}^{18}O$) to generate RNA that is uniformly labeled at every backbone position in the resulting oligonucleotide.[15] Rather than labeling all positions, we used two alternative oxidation mixes in an automated synthesizer, such that isotopically enriched phosphodiester nonbridging oxygens were incorporated in a highly site-specific manner. Here, the heavy oxidation mix was formulated with 20 mM I_2 and 97% enriched $H_2{}^{18}O$ water in the volumetric ratio 2:78:20 $H_2{}^{18}O$:THF:pyridine.

Scheme 1. Site-Specific Stable Isotope Labeling by Iodine Oxidation during Solid-Phase Oligonucleotide Synthesis



Mass labeled oligonucleotides maintained the expected +2 Da shift following a typical deprotection protocol (Figure 1b). Crude oligonucleotides were purified by the DMT-on method using C18 cartridges, followed by HPLC purification. Oligonucleotides mass labeled at a single position were typically enriched by 90% starting from 97% enriched $H_2{}^{18}O$ (Figure 1b).

We confirmed the position of the mass label in the oligonucleotide sequence by direct LC-MS sequencing (Figure 1c) by a variation of the method recently described.[16] A database of chemical formulas of all possible single-cut hydrolytic fragments with or without ^{18}O enrichment was generated, such that misincorporation of the label would be observable. For the U11-labeled RNA, the label appeared exclusively as a mass difference corresponding to [UMP + 2 Da − H_2O] between 11-nt and 12-nt hydrolytic fragments on both 5′ and 3′ sequence ladders. Fragments with higher mass carried the +2 Da label on both 5′ and 3′ ladders.

While the labeled nonbridging phosphodiester oxygens are stable under biological conditions,[15] we noted that phosphodiester hydrolysis of mass labeled RNA can result in label exchange with bulk solvent[17] when the newly generated nucleotide carries a 3′(2′)-monophosphate, which would result in label loss. Acid hydrolysis of RNA generates a terminal 3′(2′)-monophosphate product,[18] and we therefore sought to minimize the impact of exchange when generating digests by this method. We compared 50% (v/v) FA:water (~12 N) digestion at 60 and 80 °C over time (Figure S1). The ratio of ^{16}O:^{18}O UMP increased steadily over digest time at 80 °C, indicating significant phosphate oxygen exchange, such that the +2 Da species was nearly at natural abundance levels after 2 h. Under these conditions, the first-order rate constant for exchange was 1.1 h^{-1}. When digested at 60 °C, the ^{18}O mass label was well retained even after 2.5 h of digestion, with a significantly slower exchange rate of 0.28 h^{-1}. Alternatively, enzymatic RNA digestion with nuclease P1, which leaves a 5′-phosphorylated product after cleavage, resulted in negligible loss of the mass label (Figure S3). For the cross-linked RNA-peptide species analyzed in Figures 1d, 2, and S2, we chose a 2 h RNA digest in 50% (v/v) FA at 60 °C.

Recombinant Lin28ΔΔ protein was complexed with unlabeled, U11-labeled, or U12-labeled $preE_M$-let-7f RNA,[8] photo-cross-linked with 254 nm UV light, and hydrolyzed with trypsin and FA. Using this procedure, we recently identified a tryptic peptide, MGFGFLSMTAR, cross-linked to uridine monophosphate (UMP) by acid digestion and tandem mass spectrometry.[9] Using mass labeled $preE_M$-let-7f RNA oligonucleotides, we observed no significant deviation from the expected isotopic distribution for the peptide-UMP cross-linked species arising from unlabeled RNA or RNA mass labeled in the U12 position (Figure 1d). However, highly significant ^{18}O enrichment ($[M + 2H + 2\ Da]^{2+}$ = 772.3 *m*/*z*) is apparent for cross-

DOI: 10.1021/jacs.5b10596
J. Am. Chem. Soc. 2015, 137, 15378−15381

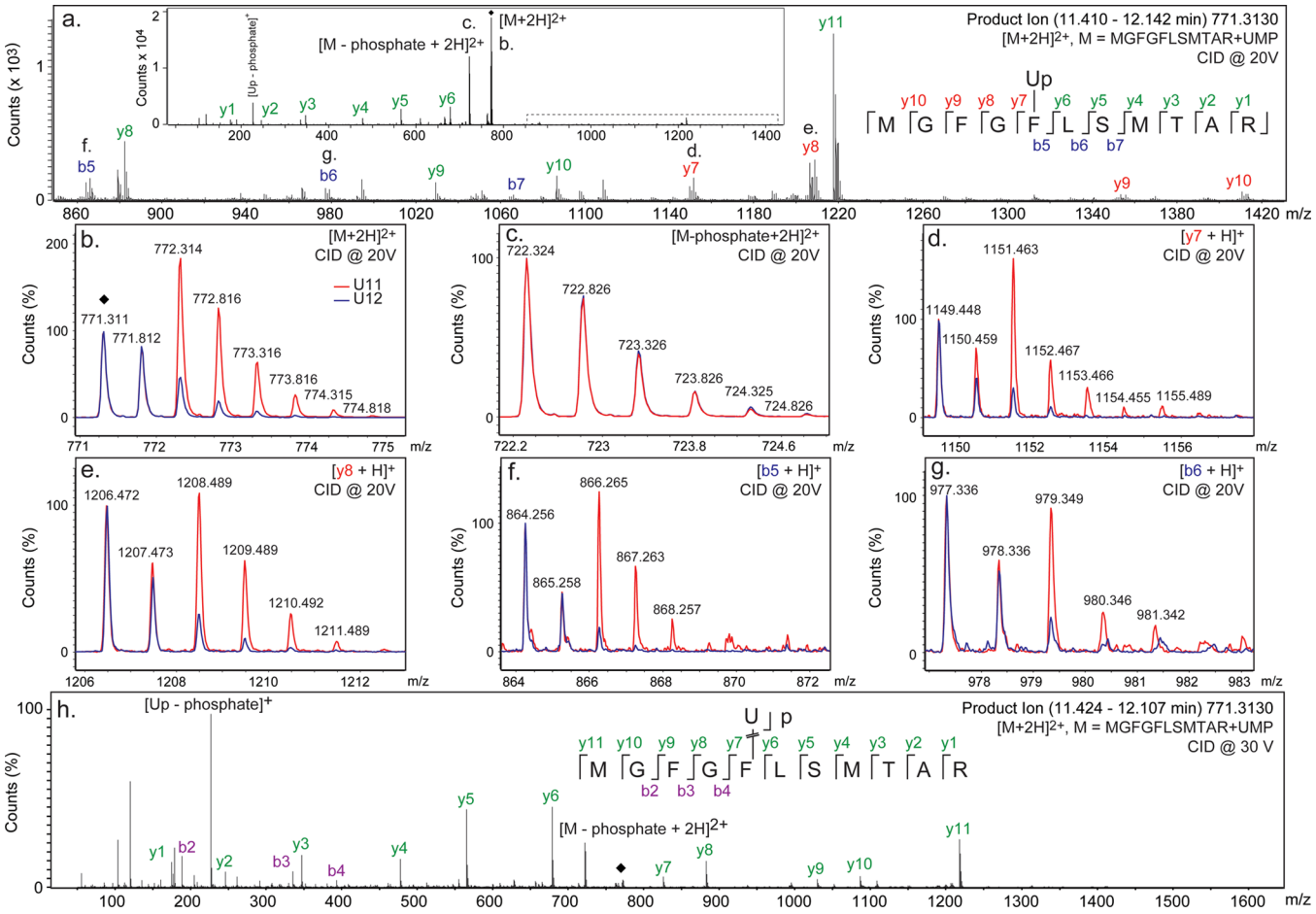


Figure 2. Simultaneous assignment of cross-linked nucleotide and amino acid sequence position by MS/MS. (a) A subset of product ions (840−1420 m/z) generated by CID at 20 V from the selected ion 771.3130 m/z (z = 2, isolation width 4 m/z), showing a fragmentation pattern consistent with uridine monophosphate (Up) attached by a mass-neutral linkage to phenylalanine in the fifth sequence position of the peptide MGFGFLSMTAR from Lin28A. Green y-ions are those derived from peptide fragmentation following loss of Up, shown fully in panel (h). Inset: Full range of product ions showing region (dotted box) magnified in the main panel. (b−g) Selected product ions arising from samples prepared with ^{18}O labeling after position U11 (red) or U12 (blue) in preE$_M$-let-7f, with apparent +2 Da enrichment derived only from U11 products. (b) Isotope distribution of remaining unfragmented selected ion (black diamond) magnified from the spectra in (a). (c−g) Magnified views of the product ion isotope distributions in panel (a). (h) Full scan of product ions from CID at 30 V with the same selected ion (black diamond), showing peptide sequence fragments for y1−11 and b2−4 arising from Up loss, with $[Up - PO_4H_2]^{1+}$ shown as a prominent product ion. Intensities in panels (b−g) are normalized to the first isotope.

linked peptide generated from U11-labeled RNA, thereby assigning the uridine in the cross-linked species to U11 in the preE$_M$-let-7f sequence.

Using tandem MS/MS with CID, we confirmed the position and identity of the observed peptide-nucleotide species (Figure 2). The cross-linked peptide ion corresponding to MGFGFLSMTAR-UMP (771.3 m/z) was isolated for fragmentation with a 4 m/z isolation width (771−775 m/z), such that the isotope distributions for both unlabeled and labeled peptides would be simultaneously monitored (Figure 2a, windowed on the high mass range for clarity). At lower CID energy (20 V), the characteristic product ions y7 (Figure 2d) and b5 (Figure 2f) carrying uridine monophosphate are observed, as well as larger ions (Figure 2e,g) consistent with UMP modification at phenylalanine-55 (F55). Higher CID energy confirms the underlying peptide sequence (Figure 2h).

Product ions carrying UMP showed a characteristic mass-shifted isotope distribution when carrying the mass label. When the RNA used for complexation was isotope labeled following U12, the resulting fragmentation spectrum for the cross-linked peptide shows a native isotopic distribution. However, when the same complexes were formed from U11 ^{18}O-labeled RNA, an enriched isotope distribution is observed for all peptide fragments carrying UMP. Loss of phosphate in the CID fragmentation spectra of U11-labeled peptide (Figure 2c) showed recovery of a natural abundance isotope distribution in the remaining uridine-peptide fragment ion.

We also observed the same peptide cross-linked to di- and trinucleotides, as a result of incomplete RNA digestion (Figure S2). From U11-labeled peptide-RNA complexes, we observe the signature mass-shifted isotope distribution for the cross-linked nucleotides U + 2 Da, AU + 2 Da, and GAU + 2 Da, AUU + 2 Da, strongly implying that U11 is the cross-link site. For the U12-labeled species, the peptide cross-linked to U, AU, and GAU all show natural abundance isotope levels, but we also observe some enrichment in the isotope distribution on the cross-linked trinucleotide AUU + 2 Da (Figure S2b). For this latter fragment, the enriched cross-linked trinucleotide must compositionally include the U12 mass label, but its absence in smaller cross-linked species strongly implies that U12 is adjacent to the cross-link site but not directly linked in the UV-catalyzed reaction.

DOI: 10.1021/jacs.5b10596
J. Am. Chem. Soc. 2015, 137, 15378−15381

Given the known sequence of preE$_M$-let-7, these data taken together unambiguously assign the UV-induced photo-cross-link position to uridine U11 in the RNA and phenylalanine F55 in Lin28A in this complex. Furthermore, mass labeling also allows for non-native isotopic enrichment to be used as an analytical signature for identification of interesting ions. Peptide preparations for tandem mass analysis are typically highly complex, and UV cross-linking further increases complexity in the absence of additional purification steps. The ability to differentiate cross-linked RNA-peptide species from a complex background could significantly improve precursor ion selection and analysis, possibly alleviating the need for rigorous sample enrichment steps. RNA with partially enriched phosphodiester positions could be used to uniquely identify cross-linked peptide-nucleotide products by examining high-resolution survey spectra for bimodal isotope distributions exhibiting +2 Da mass shifts.[17,19−21]

Crude maps of the interaction surfaces in RNA-protein complexes could in principle be elucidated by an extension of our method that makes use of an efficient search strategy based on oligonucleotides labeled at multiple phosphate positions. Uridine cross-links are predominantly observed with 254 nm UV irradiation and ESI-LC-MS.[1] Therefore, the number of mass labeled positions in a small RNA need only be, in the worst case, a function of the uridine content, and not all adjacent uridines must be labeled since trinucleotide-carrying peptides are often observed. The expected number of uridines is ca. six for single-stranded small RNA of mean base composition and length 20−25. All singly labeled oligonucleotides could be synthesized and analyzed individually in a cross-linking MS experiment, or multiple labels could be used for larger RNAs to search the space with higher efficiency, ruling out candidate positions by deduction.

Because our method relies only on ^{18}O-enriched water, introducing stable site-specific phosphodiester mass labels into synthetic oligonucleotides is significantly less expensive and more generally accessible than other conservative RNA labeling schemes that seek to minimize chemical perturbation, such as those that rely on isotopically enriched ^{13}C or ^{15}N nucleosides. Photoreactive thiolated nucleoside analogs such as 4-thiouridine or 6-thioguanosine have the potential to affect RNA structure and protein interactions, since their affinity and specificity within RNA duplexes is measurably different from their native counterparts.[22,23] As such, isotopic phosphodiester labeling is a prudent strategy for identification and confirmation of RNA-protein cross-linking sites under physiological conditions.

■ ASSOCIATED CONTENT

Supporting Information

Methods and Figures S1−S3. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b10596.

■ AUTHOR INFORMATION

Corresponding Author

*szostak@molbio.mgh.harvard.edu

Present Address

[○]Statens Serum Institut, Artillerivej 5, DK-2300 Copenhagen S, Denmark

Author Contributions

[#]These authors contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

J.W.S. is an Investigator of the Howard Hughes Medical Institute. This work was supported by grants from the Simons Foundation to J.W.S. (290363) and from the National Cancer Institute to P.S. (NIH R01CA163647). A.B. was supported by a fellowship from the Academy of Finland, and E.M.R. is supported by the National Science Foundation Graduate Research Fellowship (DGE1144152) and by the UNCF-Merck Graduate Research Science Initiative. We further thank Drs. L. Li, A. Pal, and A. Fahrenbach for helpful discussions.

■ REFERENCES

(1) Kramer, K.; Sachsenberg, T.; Beckmann, B. M.; Qamar, S.; Boon, K.-L.; Hentze, M. W.; Kohlbacher, O.; Urlaub, H. *Nat. Methods* **2014**, *11*, 1064.

(2) Ule, J.; Jensen, K. B.; Ruggiu, M.; Mele, A.; Ule, A.; Darnell, R. B. *Science* **2003**, *302*, 1212.

(3) Hafner, M.; Landthaler, M.; Burger, L.; Khorshid, M.; Hausser, J.; Berninger, P.; Rothballer, A.; Ascano, M.; Jungkamp, A.-C.; Munschauer, M.; Ulrich, A.; Wardle, G. S.; Dewell, S.; Zavolan, M.; Tuschl, T. *Cell* **2010**, *141*, 129.

(4) Yu, J.; Vodyanik, M. A.; Smuga-Otto, K.; Antosiewicz-Bourget, J.; Frane, J. L.; Tian, S.; Nie, J.; Jonsdottir, G. A.; Ruotti, V.; Stewart, R.; Slukvin, I. I.; Thomson, J. A. *Science* **2007**, *318*, 1917.

(5) Iliopoulos, D.; Hirsch, H. A.; Struhl, K. *Cell* **2009**, *139*, 693.

(6) Yu, F.; Yao, H.; Zhu, P.; Zhang, X.; Pan, Q.; Gong, C.; Huang, Y.; Hu, X.; Su, F.; Lieberman, J.; Song, E. *Cell* **2007**, *131*, 1109.

(7) Piskounova, E.; Polytarchou, C.; Thornton, J. E.; LaPierre, R. J.; Pothoulakis, C.; Hagan, J. P.; Iliopoulos, D.; Gregory, R. I. *Cell* **2011**, *147*, 1066.

(8) Nam, Y.; Chen, C.; Gregory, R. I.; Chou, J. J.; Sliz, P. *Cell* **2011**, *147*, 1080.

(9) Ransey, E. M.; Björkbom, A.; Holton, K.; Lelyveld, V. S.; Han, A.; Daley, G. Q.; Szostak, J. W.; Sliz, P. Manuscript in preparation.

(10) Wincott, F.; DiRenzo, A.; Shaffer, C.; Grimm, S.; Tracz, D.; Workman, C.; Sweedler, D.; Gonzalez, C.; Scaringe, S.; Usman, N. *Nucleic Acids Res.* **1995**, *23*, 2677.

(11) Scaringe, S. A.; Francklyn, C.; Usman, N. *Nucleic Acids Res.* **1990**, *18*, 5433.

(12) Potter, B. V. L; Eckstein, F.; Uznański, B. *Nucleic Acids Res.* **1983**, *11* (11), 7087.

(13) Shah, D. O.; Lai, K.; Gorenstein, D. G. *J. Am. Chem. Soc.* **1984**, *106*, 4302.

(14) Connolly, B. A.; Eckstein, F. *Biochemistry* **1984**, *23*, 5523.

(15) Hamasaki, T.; Matsumoto, T.; Sakamoto, N.; Shimahara, A.; Kato, S.; Yoshitake, A.; Utsunomiya, A.; Yurimoto, H.; Gabazza, E. C.; Ohgi, T. *Nucleic Acids Res.* **2013**, *41*, e126.

(16) Björkbom, A.; Lelyveld, V. S.; Zhang, S.; Zhang, W.; Tam, C. P.; Blain, J. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2015**, *137*, 14430.

(17) Castleberry, C. M.; Lilleness, K.; Baldauff, R.; Limbach, P. A. *J. Mass Spectrom.* **2009**, *44*, 1195.

(18) Oivanen, M.; Kuusela, S.; Lonnberg, H. *Chem. Rev.* **1998**, *98*, 961.

(19) Back, J. W.; Notenboom, V.; de Koning, L. J.; Muijsers, A. O.; Sixma, T. K.; de Koster, C. G.; de Jong, L. *Anal. Chem.* **2002**, *74*, 4417.

(20) Meng, Z.; Limbach, P. A. *Anal. Chem.* **2005**, *77*, 1891.

(21) Wallis, T. P.; Pitt, J. J.; Gorman, J. J. *Protein Sci.* **2001**, *10*, 2251.

(22) Sheng, J.; Larsen, A.; Heuberger, B. D.; Blain, J. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2014**, *136*, 13916.

(23) Heuberger, B. D.; Pal, A.; Del Frate, F.; Topkar, V. V.; Szostak, J. W. *J. Am. Chem. Soc.* **2015**, *137*, 2769.

DOI: 10.1021/jacs.5b10596
J. Am. Chem. Soc. 2015, 137, 15378−15381