Published in final edited form as:

Chem Commun (Camb). 2013 August 25; 49(66): 7322–7324. doi:10.1039/c3cc44454f.

Universal Labeling of 5'-Triphosphate RNAs by Artificial RNA Ligase Enzyme with Broad Substrate Specificity

John C. Haugner III and Burckhard Seelig*

Department of Biochemistry, Molecular Biology and Biophysics & BioTechnology Institute,
University of Minnesota, 1479 Gortner Ave, Saint Paul, MN 55108 USA.

Abstract

An artificial RNA ligase specific to RNA with a 5'-triphosphate (PPP-RNA) exhibits broad sequence specificity on model substrates and secondary siRNAs with direct applications in the identification of PPP-RNAs through sequencing.

The development of novel enzymes to manipulate nucleic acids has been a driving force in pioneering methods for molecular biology, genomics and transcriptomics. One such type of enzymes are ligases, which join together two strands of nucleic acids by forming a phosphodiester bond with a 3'-5' linkage. Many known natural ligases couple the phosphodiester bond formation with ATP hydrolysis through a series of group transfers.¹ Ligases have been used extensively in the analysis of large mixtures of RNA to add adapters of known sequence to the RNAs,² thus enabling reverse transcription, PCR amplification and ultimately their identification by microarray analysis or sequencing.³ As all known natural ligase enzymes require a 5'-monophosphate terminus (P-RNA) for the ligation reaction to occur, RNA mixtures are typically enzymatically modified to uniformly introduce the 5'-monophosphate in preparation for ligation and sequencing. This can prove problematic for interpreting subsequent RNA sequencing data as classes of RNA characterized by a different 5'-phosphorylation can no longer be easily distinguished in sequencing data. Identifying primary transcripts, also called immature RNA, can be particularly difficult because they have a 5'-triphosphate (PPP-RNA) and cannot be selectively purified from other classes of RNAs. While the triphosphate of eukaryotic mRNA is eventually capped, other classes of RNA retain the 5'-triphosphate as one of their only identifying features. This includes prokaryote mRNA, some viral mRNAs in eukaryotes,⁴ transcripts from mitochondria⁵ and chloroplasts,⁶ and secondary siRNAs.⁷ To identify these RNAs from cell isolates, methods exist which enrich for PPP-RNAs prior to sequencing, but they commonly identify false positives and require laborious independent verification with a complementary method.⁸ Therefore, an alternative, simplified method for identifying PPP-RNAs with improved accuracy would be highly desirable.

This journal is © The Royal Society of Chemistry [year]

Fax: +1 612 625-5780; Tel: +1 612 626-6281; seelig@umn.edu.

†Electronic Supplementary Information (ESI) available: Detailed experiments examining sequence specificity, schematic of general RNA-sequencing protocol, table of oligonucleotide substrates, calculated T_m values for PPP-RNA/splint hybrids of the application experiments, materials and methods. See DOI: 10.1039/b000000x/

The splinted, 5'-triphosphate-dependent ligation of two RNA strands has repeatedly been used as a model reaction to generate novel catalysts. Both artificial ribozymes⁹ and deoxyribozymes¹⁰ have been generated that catalyze this reaction. However, these catalytic nucleic acids often have strict sequence requirements near the ligation site, limiting the choice of PPP-RNA sequences that can be ligated. Furthermore, these nucleic acid enzymes act as both catalyst and complementary splint, which requires a new catalyst to be designed and synthesized specifically for each new substrate sequence. Previously, we reported the *de novo* selection and evolution of triphosphate dependent artificial RNA ligases.¹¹ These protein enzymes ligate a 5'-triphosphate RNA substrate to the 3'-hydroxyl of a second RNA exclusively forming a 3'-5' linkage while both substrates are aligned by a complementary oligonucleotide splint (Fig. 1). The artificial ligases are zinc-dependent metalloproteins of ~10 kDa and were derived from the two zinc finger protein hRXXRα.^{11a} We solved the structure of RNA ligase 10C, a highly active variant, revealing a novel protein fold with highly unusual structural features.¹²

To determine if this enzyme could be used for the general identification of PPP-RNAs, we investigated the sequence specificity of ligase 10C. This artificial enzyme had been generated to ligate a single substrate pair, but had never been challenged to evolve an ability to ligate alternative RNA sequences. To determine the sequence specificity of 10C, we varied the nucleotides adjacent to the ligation site. We also tested a substrate pair of entirely unrelated sequence to assess the impact of large scale changes in sequence. Finally, to extend 10C for practical applications, we investigated whether this ligase can perform selective ligation in a mixture of secondary siRNAs. These experiments demonstrate RNA ligase 10C exhibits broad sequence specificity and therefore should be a valuable tool for the whole-cell isolation and identification of PPP-RNA.

RNA ligase 10C was assayed for the ligation of chemically synthesized RNA oligonucleotides to PPP-RNA substrates synthesized and ³²P-labeled by *in vitro* transcription. Assay conditions were chosen to mimic conditions used in commercial RNA isolation and sequencing kits.¹³ Under these conditions, 10C was able to ligate the original substrates used during its evolution as well as an unrelated sequence, even though the nucleotides flanking the ligation site were completely different (Fig. 2). Ligation rates were similar for these two substrates with the k_{obs} of $0.45 \pm 0.01 \text{ h}^{-1}$ for the original substrates and $0.24 \pm 0.01 \text{ h}^{-1}$ for the substrate pair with unrelated sequence. We expanded upon these experiments by testing different nucleotides at the 3'-hydroxyl and 5'-triphosphate sides of the ligation site independently (Table S1). As we were limited to the transcription by T7 RNA polymerase, we only created PPP-RNA substrates beginning with a G or A.¹⁴ For all substrates tested, the k_{obs} were similar within a 10-fold range (Fig. S1). By comparison, the most commonly used natural enzyme T4 DNA ligase has been shown to have an approximately 3-fold difference in k_{cat} in the ligation of nicked DNA depending on the sequence surrounding the ligation site.¹⁵ If the ligation using 10C is allowed to proceed overnight as is commonly done with commercial ligases, the differences in yield for different substrates are reduced to < 3-fold. Although the ligase 10C was originally selected to ligate only one specific pair of substrate sequences, the enzyme does not require any specific nucleotide sequence near the ligation site.

The ligation rate of ligase 10C is comparable to previously reported deoxyribozymes^{10a} which catalyze the same reaction under similar reaction conditions, but is lower than the typical rate observed for natural ligase protein enzymes.¹⁶ Nevertheless, the current activity of 10C is sufficient for immediate applications, and engineering efforts to improve the catalytic efficiency of 10C are underway. In contrast to previously generated nucleic acid enzymes,^{9–10} ligase 10C can ligate all different RNA sequences without the need to tailor-make a modified catalyst for each substrate because the complementary splint is not part of, or attached to, the catalyst.

To directly demonstrate the utility of 10C, we tested the ligase with two secondary short interfering RNAs (siRNAs). siRNAs are involved in the regulation of gene expression through the RNAi pathway. Secondary siRNAs are a unique class of siRNA that were identified in *C. elegans* and have a 5'-triphosphate as they are synthesized by an RNA-dependent RNA polymerase to amplify the RNAi response.^{7, 17} We chose two previously published siRNAs (S2, S3 in Fig. 3a)^{7a} which we transcribed with T7 polymerase. Those siRNAs lack secondary structure that could interfere in base pairing to the complementary DNA splint. We also included the original PPP-RNA substrate (S1) used to evolve ligase 10C as it shares a similar 5'-sequence with the secondary siRNAs. As expected, ligase 10C was capable of ligating 5'-triphosphorylated versions of all three substrates to an adapter sequence 8 nucleotides in length. While the 8 nt adapter was chosen to facilitate separation of RNAs in the gel, longer adapters of constant sequence can be used to enable subsequent RNA sequencing. Substrates of the same sequence but with a 5'-monophosphate instead of the 5'-triphosphate were not ligated (Fig. 3b). We then combined the three 5'-triphosphorylated substrates in a single reaction and tested if individual splints could specifically promote the ligation of their complementary substrates in a mixture of similar sequences. No cross-reactivity between splints was detected, despite sequence identity of the substrates in three of the four nucleotides adjacent to the ligation site (Fig. 3b). The discrimination between substrates with a complementary splint and those with imperfect complementarity is >100-fold, using the limit of detection of the assay as a conservative estimate. Calculated melting temperatures (T_m) of the splint/substrate combinations indicate that only the matching sequence pairs will hybridize at the temperature and ionic strength of the ligation mixture (Table S2). By using splints with relatively short overhangs (e.g. 8 base pairs), the hybridization stringency increases dramatically as even single mismatches can reduce the T_m by more than 10 °C. Therefore, the ligation of sequencing adapters to biologically relevant RNAs by ligase 10C can be controlled through the use of a splint which specifically base pairs to the desired PPP-RNA substrate.

The artificial RNA ligase enzyme 10C has a potential application in the sequencing of PPP-RNAs (Fig. S2). Deep RNA sequencing (RNA-seq) is replacing traditional microarray technology for large scale RNA analysis in part because of its increased dynamic range and lower detection limit. RNA-seq has been used to characterize specific RNA subpopulations, including those which are of low cellular abundance.³ One of the primary challenges in the analysis of specific RNA classes is the minimization of false positive results due to sample contamination with other classes of RNA or cleavage products.³ For PPP-RNAs, ligase 10C offers a simple and direct route for the ligation of adapters without false positives as the

triphosphate provides the energy needed to drive bond formation. We envision that, in addition to sequencing specific sequences, total PPP-RNA could be analyzed by using oligonucleotide splints with a short degenerate sequence overhang. This would allow for ligation of any RNA sequence and has been used previously for RNA-seq in a similar fashion.^{13a}

In summary, these results highlight the value of the artificial RNA ligase 10C for the direct isolation and identification of virtually any PPP-RNAs. Efforts are underway to utilize ligase 10C in RNA sequencing efforts to evaluate the efficacy of 10C to identify a broad range of PPP-RNAs from biological samples. Finally, this exemplary application of the artificial RNA ligase highlights the potential of *in vitro* selection and evolution methods for the creation of useful artificial biocatalysts.¹⁸

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by the US National Aeronautics and Space Administration (NASA) Agreement no. NNX09AH70A through the NASA Astrobiology Institute–Ames Research Center, the US National Institutes of Health (NIH) (T32 GM08347) and by a grant from the Biocatalysis Initiative of the BioTechnology Institute at the University of Minnesota (to B.S.). We thank M. D. Lane, D. Morrone and M. V. Golynskiy for comments on the manuscript.

Notes and references

1. (a) Lehman IR. *Science*. 1974; 186:790–797. [PubMed: 4377758] (b) Shuman S. J. *Biol. Chem.* 2009; 284:17365–17369. [PubMed: 19329793]
2. (a) Moore MJ, Sharp PA. *Science*. 1992; 256:992–997. [PubMed: 1589782] (b) Moore, MJ.; Query, CC. *Methods Enzymol.* Vol. 317. Academic Press; 2000. p. 109–123.
3. Ozsolak F, Milos PM. *Nat. Rev. Genet.* 2011; 12:87–98. [PubMed: 21191423]
4. Hornung V, Ellegast J, Kim S, Brzózka K, Jung A, Kato H, Poeck H, Akira S, Conzelmann K-K, Schlee M, Endres S, Hartmann G. *Science*. 2006; 314:994–997. [PubMed: 17038590]
5. Clayton DA. *Annu. Rev. Biochem.* 1984; 53:573–594. [PubMed: 6383200]
6. Sugiura M. *Plant Mol. Biol.* 1992; 19:149–168. [PubMed: 1600166]
7. (a) Sijen T, Steiner FA, Thijssen KL, Plasterk RHA. *Science*. 2007; 315:244–247. [PubMed: 17158288] (b) Pak J, Fire A. *Science*. 2007; 315:241–244. [PubMed: 17124291]
8. Qiu Y, Cho B-K, Park YS, Lovley D, Pálsson BØ, Zengler K. *Genome Res.* 2010; 20:1304–1311. [PubMed: 20592237]
9. Bartel DP, Szostak JW. *Science*. 1993; 261:1411–1418. [PubMed: 7690155]
0. (b) Jaeger L, Wright MC, Joyce GF. *Proc. Natl. Acad. Sci. U. S. A.* 1999; 96:14712–14717. [PubMed: 10611278] (c) Ikawa Y, Tsuda K, Matsumura S, Inoue T. *Proc. Natl. Acad. Sci. U. S. A.* 2004; 101:13750–13755. [PubMed: 15365187]
10. (a) Purtha WE, Coppins RL, Smalley MK, Silverman SK. *J. Am. Chem. Soc.* 2005; 127:13124–13125. [PubMed: 16173722] (b) Flynn-Charlebois A, Wang YM, Prior TK, Rashid I, Hoadley KA, Coppins RL, Wolf AC, Silverman SK. *J. Am. Chem. Soc.* 2003; 125:2444–2454. [PubMed: 12603132]
11. (a) Seelig B, Szostak JW. *Nature*. 2007; 448:828–831. [PubMed: 17700701] (b) Seelig B. *Nat. Protocols*. 2011; 6:540–552.
12. Chao F-A, Morelli A, Haugner JC III, Churchfield L, Hagmann LN, Shi L, Masterson LR, Sarangi R, Veglia G, Seelig B. *Nat. Chem. Biol.* 2013; 9:81–83. [PubMed: 23222886]

13. Kuersten SA. US Pat. US20100279305A1. 2010 (b) An excess of enzyme was used to reduce ligation times.
14. 5'-triphosphate RNA is commonly synthesized using T7 polymerase, which strongly disfavors the incorporation of 5'-C or U.
15. Lohman GJS, Chen L, Evans TC. J. Biol. Chem. 2011; 286:44187–44196. [PubMed: 22027837]
16. While k_{obs} of ligase 10C is approximately 3 orders of magnitude lower than k_{cat} of T4 DNA ligase, ligations are completed after overnight incubation.
17. Aoki K, Moriguchi H, Yoshioka T, Okawa K, Tabara H. EMBO J. 2007; 26:5007–5019. [PubMed: 18007599]
18. (a) Golynskiy MV, Seelig B. Trends Biotechnol. 2010; 28:340–345. [PubMed: 20483496] (b) Golynskiy MV, Haugner JC III, Morelli A, Morrone D, Seelig B. Meth. Mol. Biol. 2013; 978:73–92.

Fig. 1.

Artificial RNA ligase catalyzing the formation of a 3'-5'-phosphodiester bond between a 5'-triphosphate RNA (red) and 3'-hydroxyl RNA (blue) substrate. The proposed mechanism of the reaction is shown. A complementary splint DNA oligonucleotide (black) anneals to the RNA and is necessary for ligation.

Fig. 2.

Ligation of two pairs of RNA substrates by artificial ligase 10C. The arrow represents the site of ligation. The data were fit to first-order kinetics $Y = Y_f (1 - e^{-kt})$ where Y_f = final yield and $k = k_{\text{obs}}$. The solid lines show the progress of the reaction with the substrates that were originally used to generate ligase 10C (5' CUAACGUUCGC↓GGAGACUCUUU). Dashed lines indicate the ligation of a second substrate pair of unrelated sequence (5' GCAUGUCAGCA↓AGGCCUAUCAA) that has the same GC content. Data are the mean of 3 replicates $\pm$ SD.

Fig. 3.

Application of the artificial RNA ligase enzyme to selectively ligate secondary siRNA. **a)** Alignment of the original ligation substrate (S1) and secondary siRNA substrates (S2) and (S3). Sequence similarities are highlighted in yellow. For each of the three substrates, a matching complementary splint oligodeoxynucleotide (black) was used to ligate them to a 3'-hydroxy-substrate "adapter". The adapter is identical in all three cases (blue). The first eight nucleotides of each substrate base-pair to its respective splint and are capitalized. **b)** Denaturing PAGE gel depicting ligation experiments with substrates S1 to S3 to form

products P1 to P3, demonstrating the necessity of a triphosphate (PPP) at the 5'-terminus (lanes 3, 7, 11). No ligation is observed for substrates with a 5'-monophosphate (P) (lanes 2, 6, 10). Lanes 4, 8 and 12 show the selective ligation of a single substrate in a mixture of substrates with highly similar sequences, which is dependent on the presence of the correct complementary splint. The substrate bands are marked by empty triangles and the product bands by filled triangles.