Supporting Information

Copying of mixed sequence RNA templates inside model protocells

Derek K. O'Flaherty,[†,§] Neha P. Kamat,[*,†,‡,§] Fatima N. Mirza,[†,⊥] Li Li,[†] Noam Prywes,[†,∥] and Jack W. Szostak*[,†]

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

[‡] Biomedical Engineering Department, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, United State

Table of Contents

Materials and Methods

Materials. Myristoleic acid (C14:1, MA), monomyristolein (the glycerol monoester of myristoleic acid, GMM), decanoic acid (C10:0, DA), decanol (DOH), oleic acid (OA), and monocaprin (the glycerol monoester of decanoic acid, GMD) were purchased from Nu Chek Prep. Adenosine-5'-monophosphate (^{14}C-labeled) was purchased from Moravek Biochemicals and Radiochemicals and nicotinamide adenine dinucleotide and adenosine triphosphate (^{32}P- labeled) were purchased from Perkin Elmer. The other RNA oligonucleotide dimer and trimers were synthesized on a Mermade 6 synthesizer with

phosphoramidites from Glen Research or purchased from GE Healthcare. RNA oligomers used for vesicle stability and primer extension studies were purchased from IDT. All other chemicals were purchased from Sigma-Aldrich.

Vesicle preparation. Small, unilamellar vesicles were prepared by mixing fatty acid with buffer (0.2 M Na^+-bicine, pH 8.5), tumbling overnight, extruding through 200 nm pore filters using a MiniExtruder system (Avanti Polar Lipids), and freeze/thawed 3 times. For mixed composition membranes, the neat fatty acid and glycerol monoester were mixed at the desired molar ratio before mixing with buffer. Encapsulation of nucleotides or oligomers was achieved by mixing the solute with buffer before addition to the neat amphiphiles. Vesicles were allowed to tumble for > 12 hours after extrusion and were purified from unencapsulated solutes on a Sepharose 4B size exclusion column. A running buffer containing 5 mM vesicles with matching membrane composition to the purified vesicles was used to ensure that the total fatty acid concentration remained above the critical aggregation concentration of the amphiphile, to prevent disruption of vesicles during purification. For permeability studies, a small amount of calcein (100 μM) was added to the vesicle hydration buffer during their initial formation to act as a tracer molecule to enable accurate detection of the vesicle and unencapsulated fractions in initial and subsequent fractionation steps.

Radiolabelled oligomers. Trimers and longer oligonucleotides were 5′ labeled by using ^{32}P-γ-ATP (Perkin Elmer) and T4 polynucleotide kinase (New England Biolabs), followed by Phenol/Chloroform extraction of the kinase and purification on a Sephadex G-10 column. Labeled oligomers were analyzed by PAGE to ensure the removal of unreacted ^{32}P-γ-ATP.

Permeability measurements. To the extruded and purified vesicles containing the radioactive nucleotide or oligomer was added a premixed solution of Mg^{2+}-citrate ($MgCl_2$ mixed with Na^+-citrate at a

1:4 $MgCl_2$: citrate molar ratio as described by Adamala et. al.[1]) to achieve the desired final concentration of 50 mM Mg^{2+} and 200 mM Na^{+}-citrate at pH 8.5. Aliquots from vesicle samples were taken at various time points after sample preparation (2, 4, 8, and 24 h for time trials and 2 h for heat exposure tests) and loaded onto a size exclusion column (Sepharose 4B) to separate vesicles from free solute. The ratio of encapsulated and unencapsulated nucleotide or oligomer was determined by scintillation counting with UniverSol scintillation fluid (MP Biomedicals) on a Beckman Coulter LS 6500 multipurpose scintillation counter. For certain RNA retention studies, a poly(A) 10-mer or 30-mer with a 5′-fluorescein or 5′-Cy3 dye was encapsulated in vesicles. No calcein tracer was used during studies with fluorescently-labeled RNA. To determine the extent of oligomer leakage, vesicles were separated from free oligomer after exposure to test conditions. After measuring the fluorescence intensity of eluted fractions on a SpectraMax Gemini EM (Molecular Devices), the integrated fluorescence intensity of the vesicle and free dye peaks was determined and used to quantitate leakage of oligomer. Thermal cycling experiment were composed of 2 min at 25 ^{o}C followed by 2 min at 90 ^{o}C, repeated over 20 cycles.

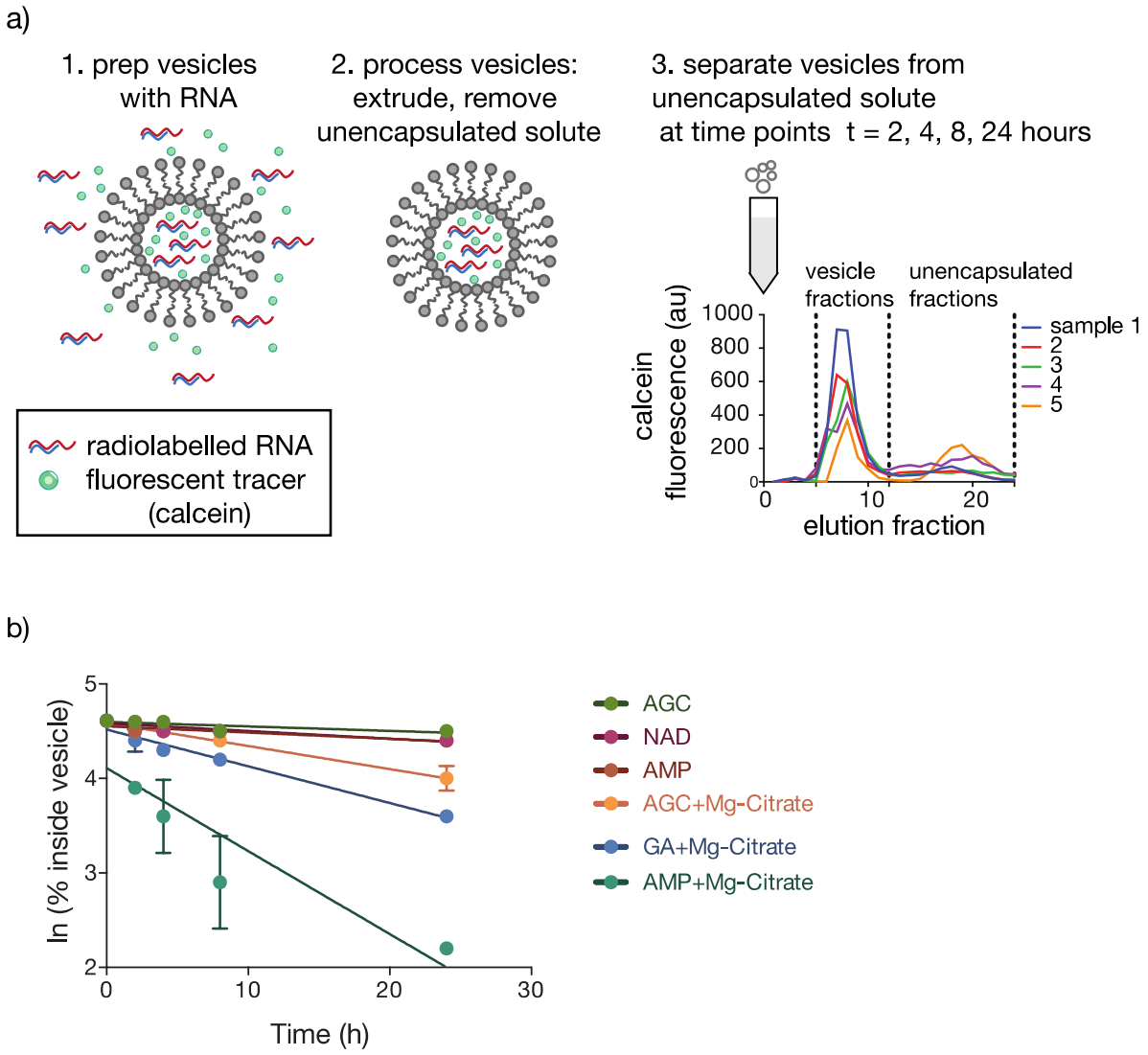
Synthesis of activated nucleotides. Cytosine 5′-phosphoro-2-aminoimidazolide, adenosine 5′-phosphoro-2- aminoimidazolide, 2-thioribothymidine 5′-phosphoro-2-aminoimidazolide and guanosine 5′-phosphoro-2- aminoimidazolide were synthesized according to a previously described procedure.[2-3] The oligomers, UGA, GAG, AGG, GGC, GCG, and CGC were similarly conjugated to 2-aminoimidazole to form activated helper oligonucleotides as described previously.[3-4] All nucleotides were purified by reverse phase chromatography using a Teledyne Isco Combiflash on a RediSepRf C18Aq column (Teledyne Isco), equilibrated with 20 mM triethylammonium bicarbonate (pH 7.5), and eluted with an acetonitrile gradient.

2-thioribothymidine 5′-phosphoro-2-aminoimidazolide: ^{1}H NMR (400 MHz, D_2O) δ: 7.81 (m, 1H), 6.81 (m, 1H), 6.76 (d, 1H, J = 4 Hz), 6.63 (m, 1H), 4.35 (m, 1H), 4.28-4.16 (m, 3H), 4.11 (m, 1H), 1.96 (s, 3H). Peaks corresponding to residual TEAB observed at 3.05 and 1.13 ppm. ^{13}C{1H} NMR (100 MHz, D2O) δ: 179.1,

171.9, 154.3 (d, J = 4 Hz), 140.0, 126.3 (d, J = 11 Hz), 119.0, 118.0 (d, J = 6 Hz), 95.7, 84.7 (d, J = 9 Hz), 77.2, 71.5, 67.1 (d, J = 5 Hz), 15.4. Peaks corresponding to residual TEAB observed at 49.2 ppm and 10.8 ppm. ^{31}P (162 MHz, D_2O): -11.05 λ_{max}: 275 nm (H_2O). LRMS (ESI-ion trap): calc for $[C_{13}H_{18}N_5O_7PS - (H^+)]^-$ 418.0592, found 418.0. LRMS (ESI- ion trap): calc for $[C_{13}H_{18}N_5O_7PS + (Na^+)]^+$ 442.0557, found 442.2. HRMS (ESI-QTOF): calc for $[C_{13}H_{18}N_5O_7PS - (H^+)]^-$ 418.0592, found 418.0607.

Primer extension studies in vesicles. Vesicles were prepared with 15 μM primer, 20 μM template, and 0.2 M Na^+-bicine pH 8.5, tumbled overnight, extruded, subjected to 5 freeze/thaw cycles, and allowed to tumble for at least 10 hours. Vesicles were purified away from unencapsulated primer and template and the reaction was initiated by adding a premixed solution of Mg^{2+}-citrate followed by 2-aminoimidazole activated nucleotides and/or trimers at 22°C. The Mg^{2+}-citrate premixed solution was prepared using a solution of $MgCl_2$ (1 M) and trisodium citrate (1.5 M) in a 1:4 ratio. All reactions were conducted under the following conditions: 0.2 M Na^+-bicine pH 8.5, 50 mM Mg^{2+}: 200 mM citrate, 10 mM activated mononucleotides, and 0.5 mM activated trimers. The primer sequence was 5′-FAM-GCG UCG ACU G -3′ and the template sequence was 5′-**G CCU CAG** CAG UCG AGC G -3′, with the template region in bold. Samples were analyzed by electrophoresis on a denaturing 20% polyacrylamide gel as described previously.[1] Reaction products were visualized using a Typhoon 9410 PhosphorImager (GE Healthcare). The second template sequence was 5'- **CG CGC CUA** CAG UCG AGC G -3′, with the template region in bold. Analysis was carried out as described above.

Supporting Figures and Tables



SI Figure 1. Determining RNA permeability through fatty acid vesicle membranes

(a) Schematic of the RNA permeability assay. Vesicles were prepared containing radiolabelled RNA and a fluorescent tracer (calcein). Vesicles were extruded and then purified on a Sepharose 4B size exclusion column to remove unencapsulated RNA and fluorescent molecules. Vesicles were then exposed to various conditions (Mg^{2+}-citrate or temperature conditions) or left to tumble on a rotator. At various time points, aliquots of the vesicle sample were re-purified on a size exclusion column. Vesicle and unencapsulated fractions were identified by monitoring calcein fluorescence and were then collected. (b) The fraction of RNA that had crossed the MA: GMM vesicle membrane was quantified via scintillation counting and the ln [fraction encapsulated] was plotted vs. time (h).

RNA name / # nucleotides)	Vesicle composition	Permeability coefficient (10^{-12} cm/s)		
		No Mg^{2+}	+50 mM Mg^{2+}-citrate	4 mM $MgCl_2$
AMP/ 1	MA:GMM	5.3±1.4	64±8.8	15±1.7
NAD or GA / 2	MA:GMM	5.9±1.6	32±0.3	-
AGC / 3	MA:GMM	1.7±1.1	22±5.2	-
AAAA / 4	MA:GMM	-	6.5±1.1	-
AAAAA / 5	MA:GMM	-	1.4±2.3	-
AMP / 1	OA	0.8±1.9	-	-
NAD / 2	OA	0.7±0.4	-	-
AGC / 3	OA	0.21±0.003	-	-

SI Table 1. Permeability coefficients of RNA through fatty acid membranes

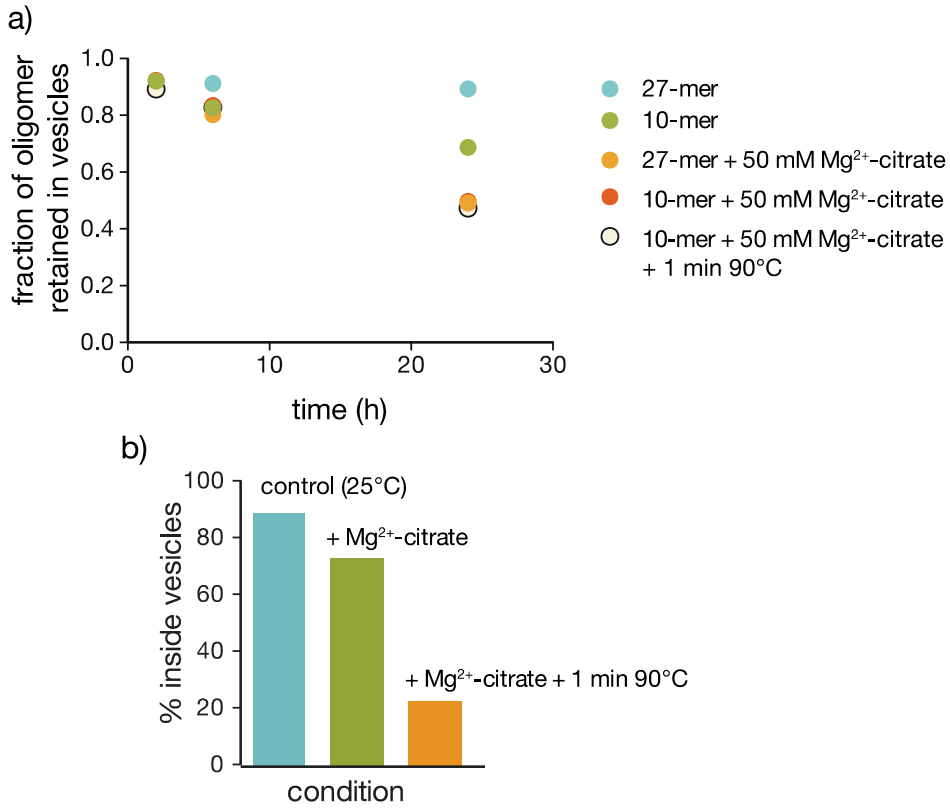
The permeability coefficient (*P*) of different RNA molecules was calculated according to a classical diffusion mechanism via Fick's law and follows the procedure used by Clary et. al.[5] Dashes (-) indicate samples for which permeability coefficients were not obtained. Briefly, the slope of the curve of ln[fraction RNA encapsulated] vs. time was determined (see SI Figure 1b). *P* is related to the slope (*K*) by,

$$P = \left(\frac{r}{3}\right) K$$

where r is the internal radius of the vesicle and relates to the external diameter of a vesicle, *d*, by

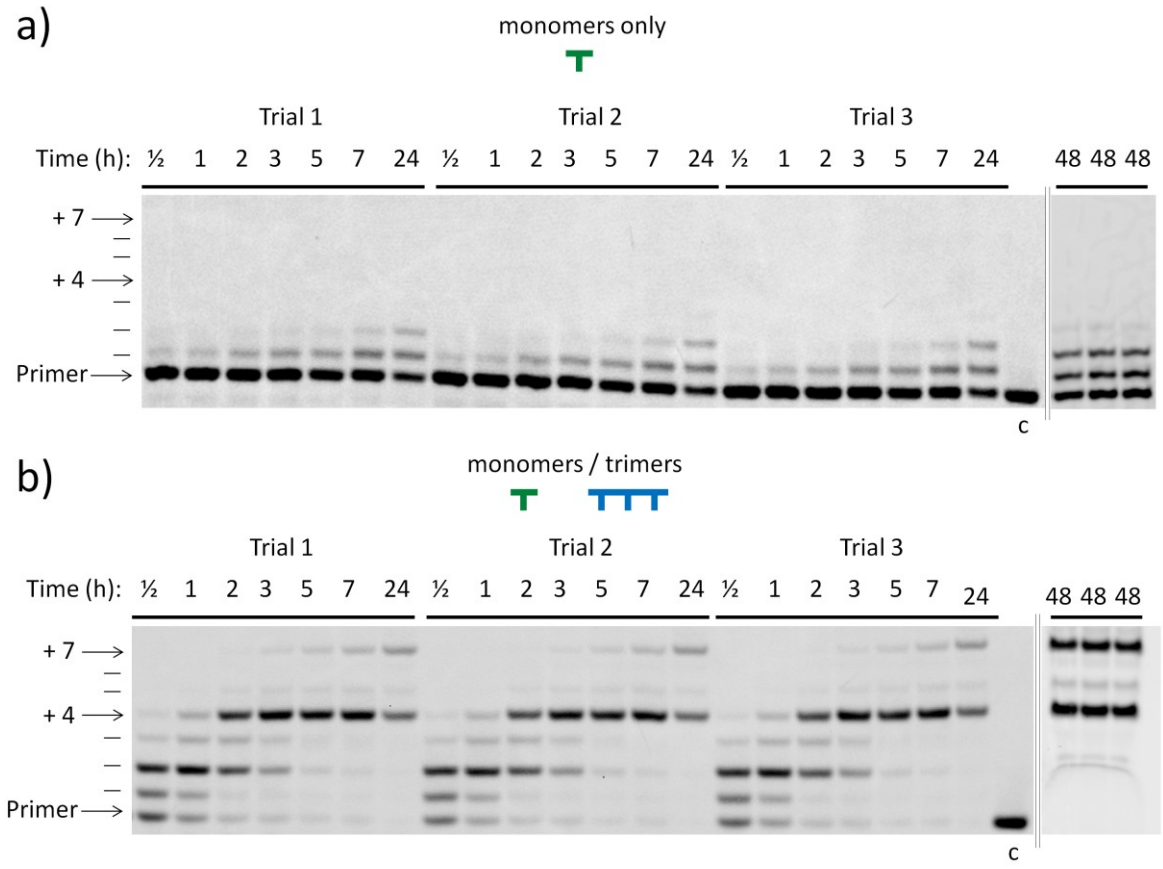
$$r = \frac{d}{2} - \Delta x$$

where Δx is the thickness of the bilayer. We used d=200nm (the diameter of vesicles used upon extrusion in our study) and Δx was estimated to be 4.6nm from Walde et al.[6] N=3.



SI Figure 2. RNA retention in decanoic acid: decanol: glycerol monodecanoate 4: 1: 1 vesicles in the presence of Mg^{2+}-citrate.

(a) The leakage of oligonucleotides (5'FAM-10-mer and a 5'Cy3-27-mer) were monitored overtime. The sequences of the RNA oligomers were 5'FAM-AAA AAA AAA A-3' and 5' Cy3-GCG UAG ACU GAC UGG AGC AGC AGC AGC-3'. In the absence of Mg^{2+}-citrate, a 10-mer leaks from vesicle membranes after 1 day. The retention of a larger 27-mer indicates the vesicles, though leaky to smaller 10-mer oligomers, are stable. The leakage of both oligomers upon addition to Mg^{2+}-citrate indicates a likely rupture/ instability of some portion of these vesicles under this condition. (b) Trimer (AGC) permeability through decanoic: decanol: glycerol monodecanoate 4: 1: 1 vesicles after 2 hours under various conditions.



c)

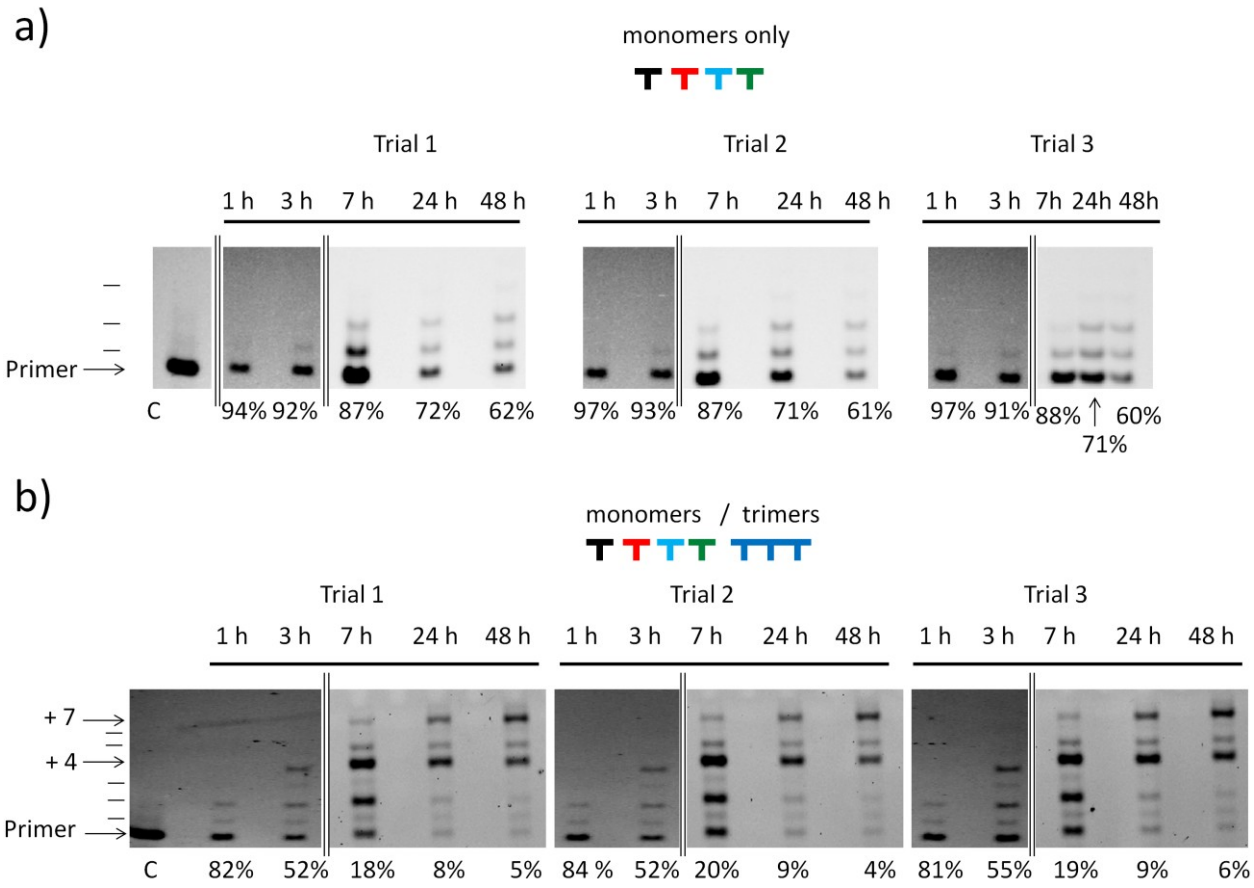
Primer (%) remaining (In solution)

	Monomers only			Monomers / trimers		
Time (h)	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
0.5	99	99	98	30	26	28
1	98	98	97	15	13	15
2	94	93	94	7	5	7
3	92	92	91	5	5	5
5	88	85	85	3	3	3
7	81	80	80	2	2	3
24	63	60	61	0	2	4
48	52	52	54	0	0	0

SI Figure 3. In-solution nonenzymatic copying of a mixed RNA template with (a) monomers only (without the use of trimers) and (b) with monomers and trimer helpers.

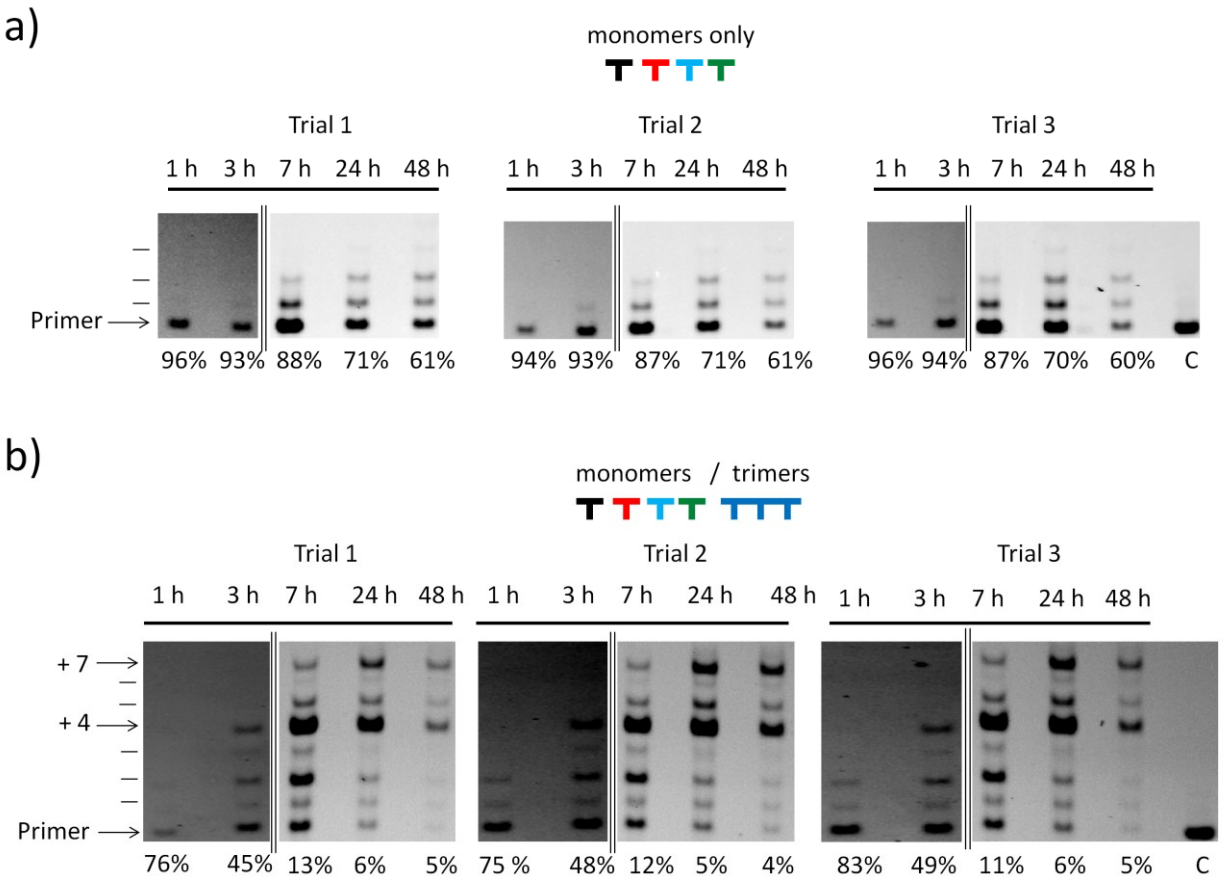
Primer extension experiments were carried out in triplicates and the fraction of primer remaining at different reaction times is shown in table (c). All primer extension reactions were carried out using 10

mM activated ribonucleotide monomers (2-NH_2Imp-rA, 2-NH_2Imp-rG, 2-NH_2Imp-rC, 2-NH_2Imp-s^2rT), 50 mM $MgCl_2$, 200 mM sodium citrate, 200 mM Na^+-bicine pH 8.5, and either without (a) or with (b) 0.5 mM activated trimers. The sequence context of the template and primer strands were as follows: 5'-**G CCU CAG** CAG UCG AGC G and 5'-(6-FAM)-CGCUCGACUG, respectively (template region in bold). c stands for the primer alone.



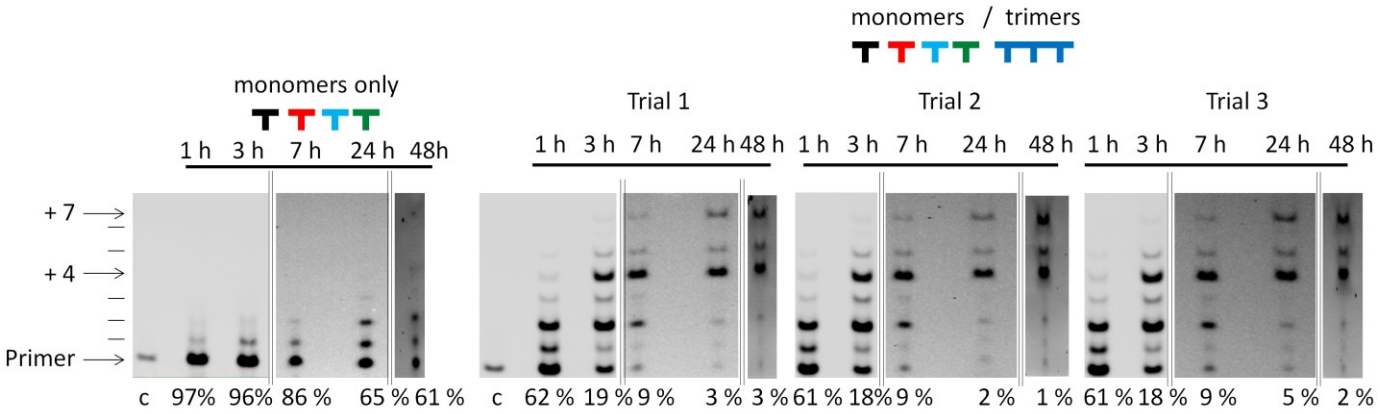
SI Figure 4. Nonenzymatic copying of a mixed RNA template within MA : GMM (2:1) vesicles using activated (a) mononucleotides only or (b) mononucleotides and trinucleotide helpers.

Primer extension experiments were carried out in triplicates. The fraction of primer remaining at different reaction times is shown below each lane. All primer extension reactions were carried out using 10 mM activated ribonucleotide monomers (2-NH_2Imp-rA, 2-NH_2Imp-rG, 2-NH_2Imp-rC, 2-NH_2Imp-s^2rT), 50 mM $MgCl_2$, 200 mM sodium citrate, 200 mM Na^+-bicine pH 8.5, and in the absence (a) or presence (b) of 0.5 mM activated trinucleotides. The sequence context of the template and primer strands were as follows: 5'-**G CCU CAG** CAG UCG AGC G and 5'-(6-FAM)-CGCUCGACUG, respectively (template region in bold). c stands for the primer alone.



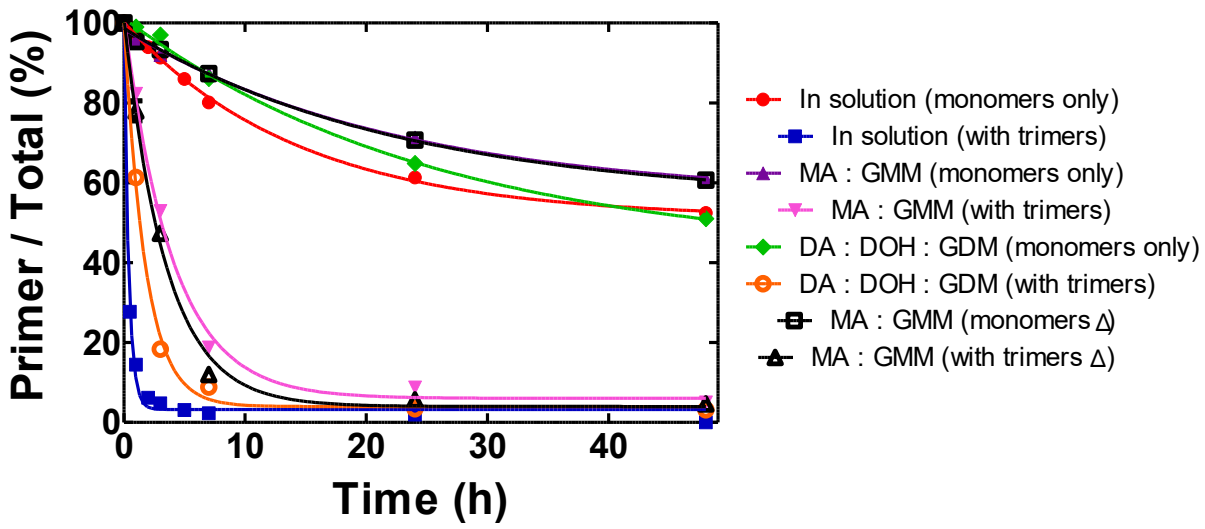
SI Figure 5. Nonenzymatic copying of a mixed RNA template within MA : GMM (2:1) vesicles, subject to an initial 1 min 90°C heat-pulse, using activated (a) mononucleotides only or (b) mononucleotides and trinucleotide helpers.

Primer extension experiments were carried out in triplicates. The initial heat pulse was carried out prior to the addition of monomers. The fraction of primer remaining at different reaction times is shown below each lane. All primer extension reactions were carried out using 10 mM activated ribonucleotide monomers (2-NH_2Imp-rA, 2-NH_2Imp-rG, 2-NH_2Imp-rC, 2-NH_2Imp-s^{2}rT), 50 mM $MgCl_2$, 200 mM sodium citrate, 200 mM Na^+-bicine pH 8.5, and in the absence (a) or presence (b) of 0.5 mM activated trinucleotides. The sequence context of the template and primer strands were as follows: 5'-**G CCU CAG** CAG UCG AGC G and 5'-(6-FAM)-CGCUCGACUG, respectively (template region in bold). c stands for the primer alone.



SI Figure 6. Nonenzymatic copying of a mixed RNA template within DA: DOH: GDM (4:1:1) vesicles using activated mononucleotides only, or mononucleotides and trinucleotide helpers.

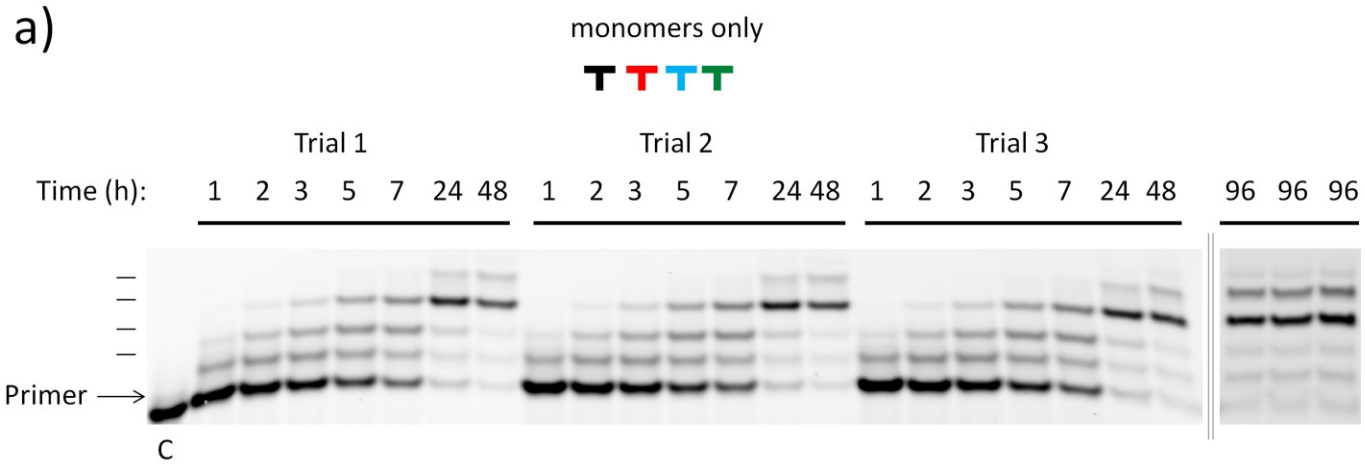
Primer extension experiments with mononucleotides and trinucleotide helpers were carried out in triplicates. The fraction of primer remaining at different reaction times is shown below each lane. All primer extension reactions were carried out using 10 mM activated ribonucleotide monomers (2-NH_2Imp-rA, 2-NH_2Imp-rG, 2-NH_2Imp-rC, 2-NH_2Imp-s^2rT), 50 mM $MgCl_2$, 200 mM sodium citrate, 200 mM Na^+-bicine pH 8.5, and in the absence (left) or presence (right) of 0.5 mM activated trinucleotides. The sequence context of the template and primer strands were as follows: 5'-**G CCU CAG** CAG UCG AGC G and 5'-(6-FAM)-CGCUCGACUG, respectively (template region in bold). c stands for the primer alone.



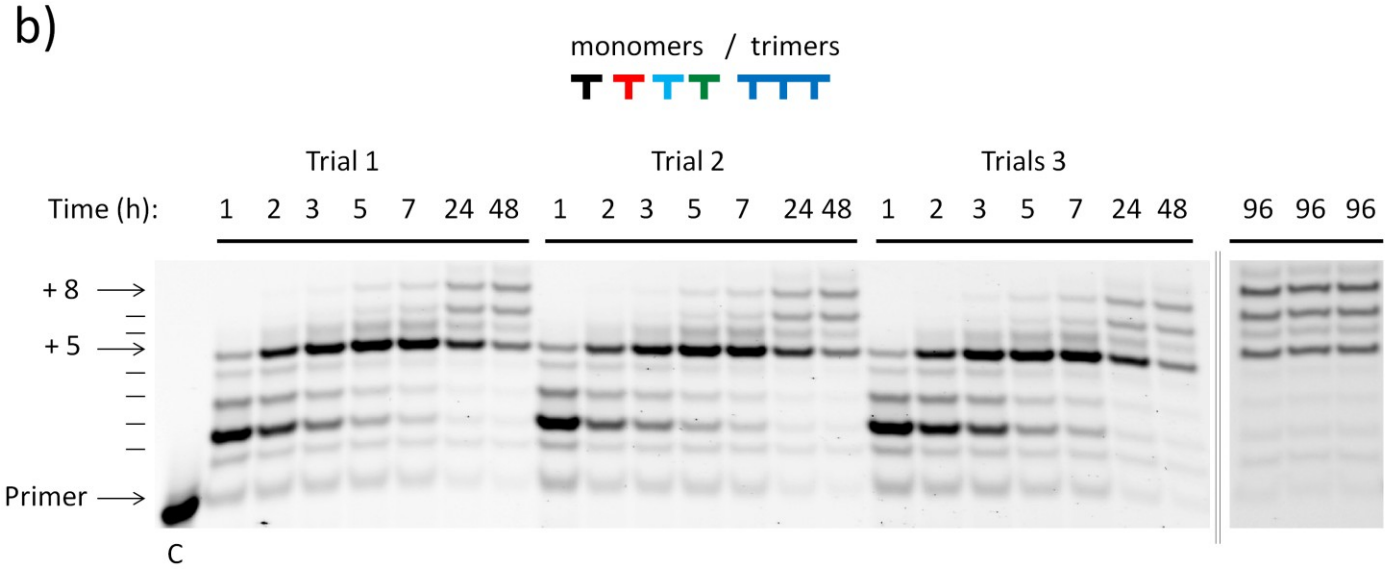
SI Figure 7. Summary of primer decay over time in solution, within MA : GMM (2:1), or within DA : DOH : GDM (4:1:1) vesicles.

A similar decay profile was observed for the primer extensions conducted with monomers only, whereas a difference was observed for primer extensions conducted with mononucleotides and trinucleotides (see Figure 5c in the main text for bar graph). The sequence context of the template and primer strands were as follows: 5'-**G CCU CAG** CAG UCG AGC G and 5'-(6-FAM)-CGCUCGACUG, respectively (template region in bold). Δ represents a 1 minute pulse at 90 °C. Data points are reported as the mean +/- s.d. from triplicate experiments.

a)



b)



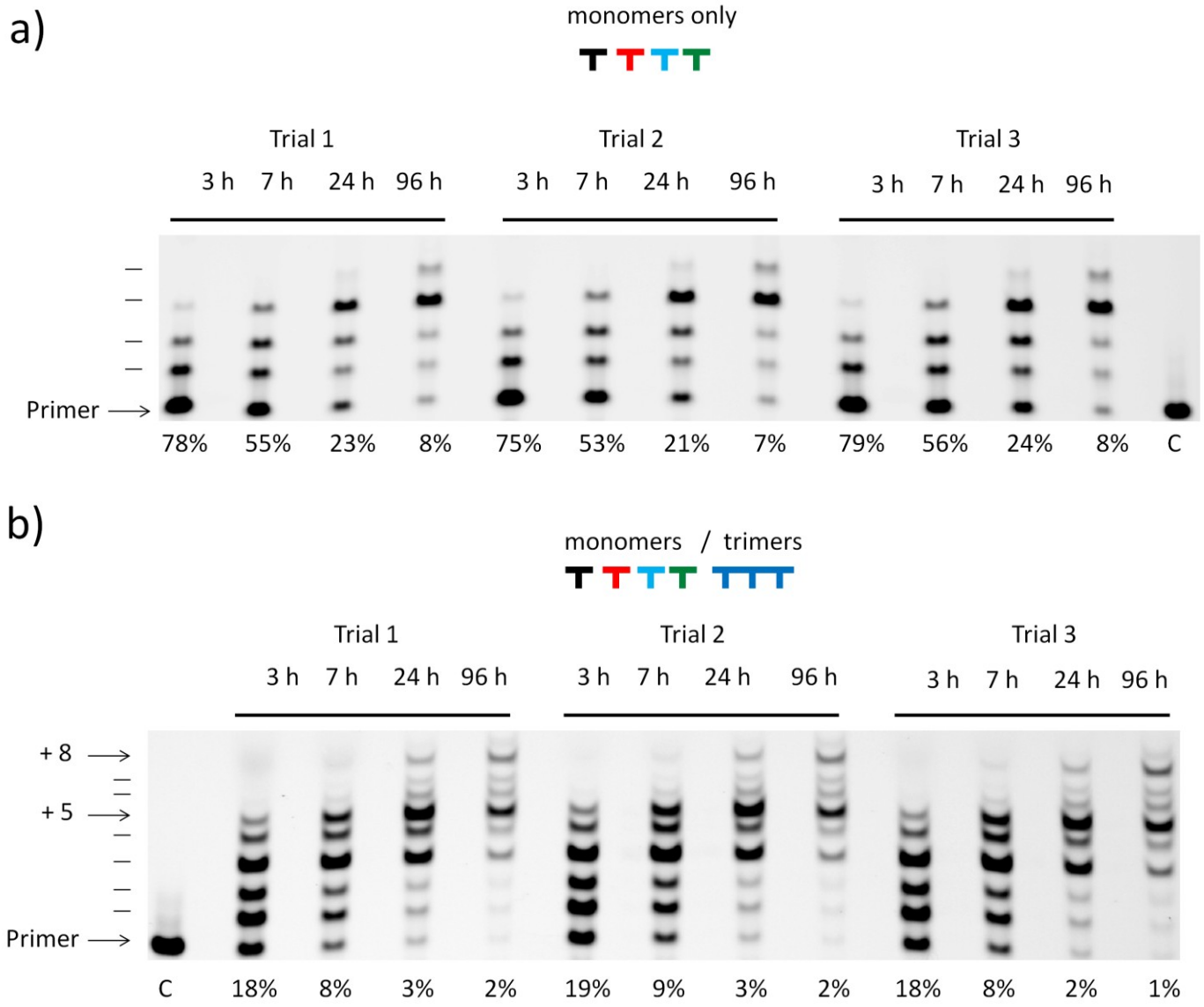
c)

Primer (%) remaining (In solution)

	Monomers only			Monomers / trimers		
Time (h)	Trial 1	Trial 2	Trial 3	Trial 1	Trial 2	Trial 3
1	90	89	88	10	10	10
2	77	77	76	8	8	9
3	66	67	67	7	7	7
5	51	52	52	6	5	6
7	41	41	42	5	5	5
24	14	14	14	4	4	3
48	10	11	9	4	4	2
96	5	4	5	3	2	2

SI Figure 8. In-solution nonenzymatic copying of a second mixed RNA template with (a) monomers only (without the use of trimers) and (b) with monomers and trimer helpers.

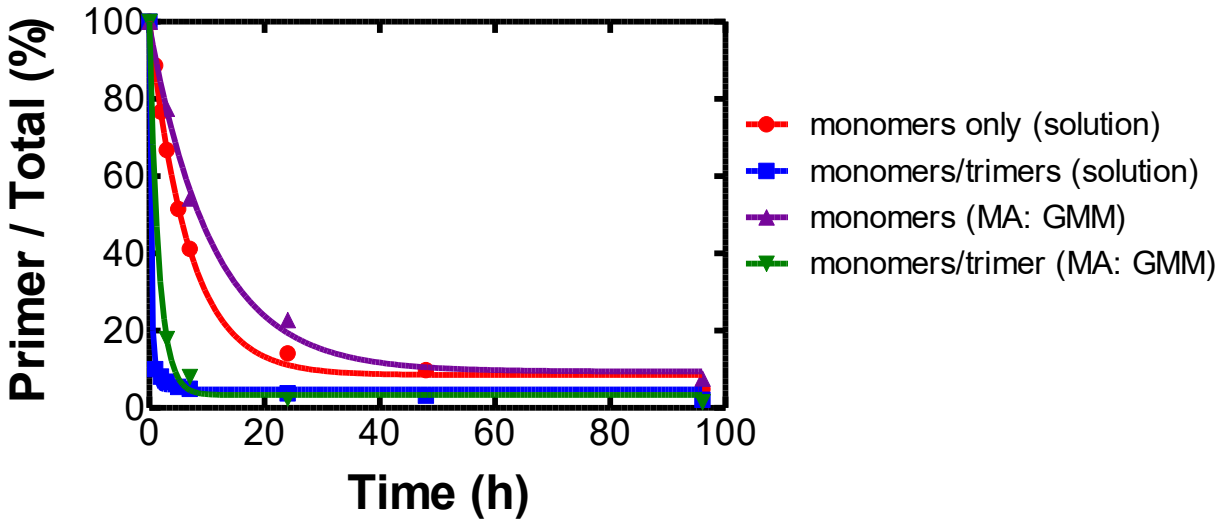
Primer extension experiments were carried out in triplicates and the fraction of primer remaining at different reaction times is shown in table (c). All primer extension reactions were carried out using 10 mM activated ribonucleotide monomers (2-NH_2Imp-rA, 2-NH_2Imp-rG, 2-NH_2Imp-rC, 2-NH_2Imp-s^2rT), 50 mM $MgCl_2$, 200 mM sodium citrate, 200 mM Na^+- bicine pH 8.5, and either without (a) or with (b) 0.5 mM activated trinucleotides (except for the 5′-GCG, which was 1 mM given that it is used for two extensions), with an initial 1 minute pulse at 90 °C. The sequence context of the template and primer strands were as follows: 5'- **CG CGC CUA** CAG UCG AGC G and 5'-(6-FAM)-CGCUCGACUG, respectively (template region in bold). c stands for the primer alone.



SI Figure 9. Nonenzymatic copying of a second mixed RNA template within MA : GMM (2:1) vesicles, subject to an initial 1 min 90°C heat-pulse, using activated (a) mononucleotides only or (b) mononucleotides and trinucleotide helpers.

Primer extension experiments were carried out in triplicates. The initial heat pulse was carried out prior to the addition of monomers. The fraction of primer remaining at different reaction times is shown below each lane. All primer extension reactions were carried out using 10 mM activated ribonucleotide monomers (2-NH_2Imp-rA, 2-NH_2Imp-rG, 2-NH_2Imp-rC, 2-NH_2Imp-s^{2}rT), 50 mM $MgCl_2$, 200 mM sodium citrate, 200 mM Na^+-bicine pH 8.5, and in the absence (a) or presence (b) of 0.5 mM activated

trinucleotides (except for the 5′-GCG, which was 1 mM given that it is used for two extensions), with an initial 1 minute pulse at 90 °C. The sequence context of the template and primer strands were as follows: 5'- **CG CGC CUA** CAG UCG AGC G and 5'-(6-FAM)-CGCUCGACUG, respectively (template region in bold). c stands for the primer alone.



SI Figure 10. Summary of primer decay over time in solution or within MA : GMM (2:1) for a second sequence.

The sequence context of the template and primer strands were as follows: 5'- **CG CGC CUA** CAG UCG AGC G and 5'-(6-FAM)-CGCUCGACUG, respectively (template region in bold). Data points are reported as the mean +/- s.d. from triplicate experiments.

References

1. Adamala, K.; Szostak, J. W. *Science* **2013,** *342* (6162), 1098-100.

2. Heuberger, B. D.; Pal, A.; Del Frate, F.; Topkar, V. V.; Szostak, J. W *. J. Am. Chem. Soc.* **2015,** *137* (7), 2769-75.

3. Li, L.; Prywes, N.; Tam, C. P.; O'Flaherty, D. K.; Lelyveld, V. S.; Izgu, E. C.; Pal, A.; Szostak, J. W. *J. Am. Chem. Soc.* **2017,** *139* (5), 1810-1813.

4. Prywes, N.; Blain, J. C.; Del Frate, F.; Szostak, J. W. *eLife* **2016,** *5*, e17756.

5. Clary, L.; Verderone, G.; Santaella, C.; Vierling, P. *Biochim. Biophys. Acta* **1997,** *1328* (1), 55-64.

6. Walde, P.; Namani, T.; Morigaki, K.; Hauser, H. *Liposome Technology* **2006,** *1*, 1-19.