Supporting Information

for manuscript entitled

Cascade of Reduced Speed and Accuracy After Errors in Enzyme-Free Copying of Nucleic Acid Sequences

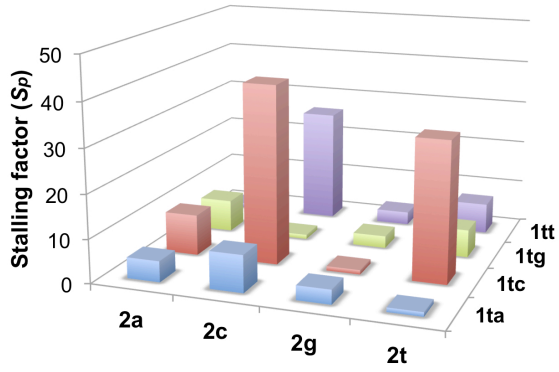
Kevin Leu, Eric Kervio, Benedikt Obermayer, Rebecca M. Turk-MacLeod, Caterina Yuan, Jesus-Mario Luevano Jr., Eric Chen, Ulrich Gerland, Clemens Richert, Irene A. Chen

Contents

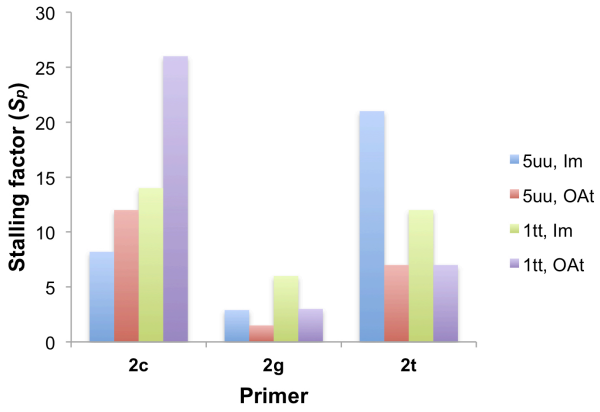
Supporting Figure S1. Stalling factors (S_p) when considering extension by any monomer.
Supporting Tables S1-S7. Rate constants, fidelities, and primer conversion.
Supporting Table S8. Comparison of background dimerization rate, extension rate after mismatches, and extension rate after matches.
Supporting Figure S2. Mass spectra of primers.
Supporting Figures S3-S9. Kinetic data from MS assay.
Supporting Figures S10-S22. Kinetic data from PAGE assay.
Supporting Figure S23: Comparison of mutation profiles for different nucleic acid systems, including LNA.
Supporting Text and Figures S24-S29: Computer simulation.
Supporting Figures S30-S41: Characterization data for activated monomers.

Supporting Figure S1. Stalling factors (S_p) when considering extension by any monomer (not just the correct one). The factor (S_p) is analogous to S_G but uses the total rate of primer extension instead of k_G alone. Compare with Figure 2A and 2B from the main text. (Legend in panel B applies to panel B only.) Corresponding data tables are given below each panel.

A



B

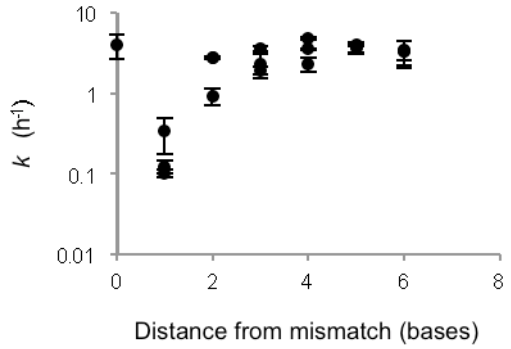


Reactants	k_G (h^{-1})	k_{total} (h^{-1})	S_G	S_p
5uu, 2a, 4a-t	5.2	5.4	-	-
5uu, 2c, 4a-t	0.22	0.66	24	8.2
5uu, 2g, 4a-t	1.3	1.9	4	2.9
5uu, 2t, 4a-t	0.09	0.26	58	21
1tt, 2a, 4a-t	1.9	1.9	-	-
1tt, 2c, 4a-t	0.04	0.14	52	14
1tt, 2g, 4a-t	0.12	0.32	15	6
1tt, 2t, 4a-t	0.06	0.16	30	12
5uu, 2a, 3a-t	0.65	0.95	-	-
5uu, 2c, 3a-t	0.02	0.08	34	12
5uu, 2g, 3a-t	0.2	0.64	3.2	1.5
5uu, 2t, 3a-t	0.03	0.14	19	7
1tt, 2a, 3a-t	2.5	2.5	-	-
1tt, 2c, 3a-t	0.03	0.09	79	26
1tt, 2g, 3a-t	0.15	0.8	16	3
1tt, 2t, 3a-t	0.15	0.4	16	7

Reactants	k_G (h^{-1})	k_{total} (h^{-1})	S_G	S_p
1tg, 2a, 3a-t	0.06	0.18	14	7.4
1tg, 2c, 3a-t	0.86	1.3	-	-
1tg, 2g, 3a-t	0.09	0.46	9.3	2.8
1tg, 2t, 3a-t	0.1	0.2	8.4	6.4
1tc, 2a, 3a-t	0.06	0.19	20	9.4
1tc, 2c, 3a-t	0.01	0.04	110	41
1tc, 2g, 3a-t	1.2	1.8	-	-
1tc, 2t, 3a-t	0.02	0.06	74	32
1ta, 2a, 3a-t	0.04	0.13	13	4.8
1ta, 2c, 3a-t	0.02	0.08	27	8.5
1ta, 2g, 3a-t	0.05	0.19	9.4	3.3
1ta, 2t, 3a-t	0.48	0.64	-	-

Supporting Table S1. Polymerization rates progressively downstream of a single mismatch in the RNA system. Apparent first-order rate constants (k) upon addition of the activated monomer were measured by PAGE assay. Data corresponds to Figure 4, Supporting Figures S13-15. Figure at bottom shows the same experimental data and from Table 2 graphically. For theoretical calculations, the non-templated rate k_n was assumed to be 1/100th of the templated rated k_t.

Reactants		**# bases from mismatch to terminus**	k_G **(h^{-1})**
6.2c, 7g, 8g	3′…GAGUG**C**CCUC… 5′…CUCAC**U**G	2	2.7 ± 0.09
6.2g, 7g, 8g	3′…GAGUG**G**CCUC… 5′…CUCAC**U**G	2	2.8 ± 0.006
6.2u, 7g, 8g	3′…GAGUG**U**CCUC… 5′…CUCAC**U**G	2	0.9 ± 0.2
6.3a, 7g, 8g	3′…GAGU**A**ACCUC… 5′…CUCA**C**UG	3	3.5 ± 0.3
6.3c, 7g, 8g	3′…GAGU**C**ACCUC… 5′…CUCA**C**UG	3	1.9 ± 0.2
6.3u, 7g, 8g	3′…GAGU**U**ACCUC… 5′…CUCA**C**UG	3	2.3 ± 0.8
6.4a, 7g, 8g	3′…GAG**A**GACCUC… 5′…CUC**A**CUG	4	4.9 ± 0.2
6.4c, 7g, 8g	3′…GAG**C**GACCUC… 5′…CUC**A**CUG	4	3.6 ± 0.08
6.4g, 7g, 8g	3′…GAG**G**GACCUC… 5′…CUC**A**CUG	4	2.3 ± 0.5
6.5a, 7g, 8g	3′…GA**A**UGACCUC… 5′…CU**C**ACUG	5	3.8 ± 0.6
6.5c, 7g, 8g	3′…GA**C**UGACCUC… 5′…CU**C**ACUG	5	4.0 ± 0.2
6.5u, 7g, 8g	3′…GA**U**UGACCUC… 5′…CU**C**ACUG	5	3.7 ± 0.6
6.6c, 7g, 8g	3′…G**C**GUGACCUC… 5′…C**U**CACUG	6	3.5 ± 0.9
6.6g, 7g, 8g	3′…G**G**GUGACCUC… 5′…C**U**CACUG	6	3.3 ± 1.2
6.6u, 7g, 8g	3′…G**U**GUGACCUC… 5′…C**U**CACUG	6	3.4 ± 1.1



Supporting Table S2. Stalling and mutation rate after two consecutive errors in the RNA system. Mismatched bases in bold. Rates determined by PAGE assay. See Figure 6, Supporting Figures S18-19.

Reactants		k (h^{-1})
6.7c, 7g, 8g	3′…GAGUG**CA**CUC… 5′…CUCAC**UG**	0.032 $\pm$ 0.004
6.7g, 7g, 8g	3′…GAGUG**GA**CUC… 5′…CUCAC**UG**	0.028 $\pm$ 0.0006
6.7u, 7g, 8g	3′…GAGUG**UA**CUC… 5′…CUCAC**UG**	0.035 $\pm$ 0.01
6.7g, 7g, 8a	3′…GAGUG**GA**CUC… 5′…CUCAC**UG**	0.00035 $\pm$ 0.00032
6.7g, 7g, 8c	3′…GAGUG**GA**CUC… 5′…CUCAC**UG**	0.00046 $\pm$ 0.00029
6.7g, 7g, 8u	3′…GAGUG**GA**CUC… 5′…CUCAC**UG**	0.0027 $\pm$ 0.0006

Supporting Table S3. Effect of upstream sequence context of the template in the RNA system. Experimental reaction rates (*k*) corresponding to Figure 7A, determined by PAGE assay. Supporting Figures S12, S15-18.

Template **Primer**		***k*** **(h^{-1})**
6.0aa, 7g, 8u	3′…GAGUGACAAC… 5′…CUCACUG	0.79 ± 0.24
6.0ca, 7g, 8u	3′…GAGUGACACC… 5′…CUCACUG	1.2 ± 0.2
6.0ga, 7g, 8u	3′…GAGUGACAGC… 5′…CUCACUG	0.8 ± 0.2
6.0ua, 7g, 8u	3′…GAGUGACAUC… 5′…CUCACUG	1.3 ± 0.5
6.0ac, 7g, 8g	3′…GAGUGACCAC… 5′…CUCACUG	0.14 ± 0.05
6.0cc, 7g, 8g	3′…GAGUGACCCC… 5′…CUCACUG	22.4 ± 0.7
6.0gc, 7g, 8g	3′…GAGUGACCGC… 5′…CUCACUG	0.137 ± 0.001
6.0uc, 7g, 8g	3′…GAGUGACCUC… 5′…CUCACUG	3.1 ± 1.1
6.0ag, 7g, 8c	3′…GAGUGACGAC… 5′…CUCACUG	3.0 ± 1.1
6.0cg, 7g, 8c	3′…GAGUGACGCC… 5′…CUCACUG	1.5 ± 0.5
6.0gg, 7g, 8c	3′…GAGUGACGGC… 5′…CUCACUG	19.4 ± 0.3
6.0ug, 7g, 8c	3′…GAGUGACGUC… 5′…CUCACUG	15 ± 4
6.0au, 7g, 8a	3′…GAGUGACUAC… 5′…CUCACUG	0.10 ± 0.05
6.0cu, 7g, 8a	3′…GAGUGACUCC… 5′…CUCACUG	0.09 ± 0.02
6.0gu, 7g, 8a	3′…GAGUGACUGC… 5′…CUCACUG	0.17 ± 0.02
6.0uu, 7g, 8a	3′…GAGUGACUUC… 5′…CUCACUG	0.62 ± 0.32

Supporting Table S4. Effect of upstream sequence context of the template on fidelity after a mismatch in the DNA system. A single time point was taken for analysis by the MS assay to determine fidelity. Incorporation frequency of correct base is underlined. See Figure 7B.

Reactants		$f_C/f_T/f_A/f_G$
1aa, 2g, 4a-t	3′-GCGTGCACAG… 5′-CGCACGG	0.03/0.18/0.22/<u>0.57</u>
1ca, 2g, 4a-t	3′-GCGTGCACCG… 5′-CGCACGG	0.03/0.03/0.14/<u>0.78</u>
1ga, 2g, 4a-t	3′-GCGTGCACGG… 5′-CGCACGG	0.34/0.06/0.22/<u>0.33</u>
1ta, 2g, 4a-t	3′-GCGTGCACTG… 5′-CGCACGG	0.08/0.23/0.42/<u>0.27</u>
1at, 2g, 4a-t	3′-GCGTGCTCAG… 5′-CGCACGG	0.02/0.15/0.14/<u>0.69</u>
1ct, 2g, 4a-t	3′-GCGTGCTCCG… 5′-CGCACGG	0 /0/0.05/<u>0.95</u>
1gt, 2g, 4a-t	3′-GCGTGCTCGG… 5′-CGCACGG	0.43/0.04/0.21/<u>0.32</u>
1tt, 2g, 4a-t	3′-GCGTGCTCTG… 5′-CGCACGG	0.07/0.02/0.68/<u>0.23</u>
1ct, 2g, 3a-t	3′-GCGTGCTCCG… 5′-CGCACGG	0.01/0/0.01/<u>0.97</u>
1tt, 2g, 3a-t	3′-GCGTGCTCTG… 5′-CGCACGG	0.01/0.01/0.86/<u>0.12</u>

Supporting Table S5. Reaction rates and incorporation frequencies with a DNA/locked nucleic acid template, measured by PAGE assay. Starred reactions used less monomer due to the high rate of correct incorporation (1 mM for **4g, 4c, 4a**; 4 mM for **4t**). All other reactions used 5 mM for **4g, 4c, 4a** and 20 mM for **4t.** See Figure 5B of main text, Supporting Figures S19-22.

Reactants		***k*** (h^{-1})	***f***
9c, 10, 4g*	3′…GAGT$_L$GA$_L$CCT$_L$C… 5′…CTCA CT G…	19 ± 5	0.99970 ± 0.00009
9c, 10, 4t*	3′…GAGT$_L$GA$_L$CCT$_L$C… 5′…CTCA CT G…	0.0026 ± 0.00005	0.00015 ± 0.00004
9c, 10, 4c*	3′…GAGT$_L$GA$_L$CCT$_L$C… 5′…CTCA CT G…	0.0017 ± 0.001	0.00008 ± 0.00005
9c, 10, 4a*	3′…GAGT$_L$GA$_L$CCT$_L$C… 5′…CTCA CT G…	0.002 ± 0.002	0.00007 ± 0.00007
9g, 10, 4c	3′…GAGT$_L$GA$_L$CGT$_L$C… 5′…CTCA CT G…	2.8 ± 0.2	0.765 ± 0.059
9g, 10, 4t	3′…GAGT$_L$GA$_L$CGT$_L$C… 5′…CTCA CT G…	0.35 ± 0.15	0.097 ± 0.041
9g, 10, 4g	3′…GAGT$_L$GA$_L$CGT$_L$C… 5′…CTCA CT G…	0.46 ± 0.07	0.127 ± 0.018
9g, 10, 4a	3′…GAGT$_L$GA$_L$CGT$_L$C… 5′…CTCA CT G…	0.041 ± 0.002	0.011 ± 0.0004
9a, 10, 4t	3′…GAGT$_L$GA$_L$CAT$_L$C… 5′…CTCA CT G…	1.8 ± 0.4	0.542 ± 0.006
9a, 10, 4c	3′…GAGT$_L$GA$_L$CAT$_L$C… 5′…CTCA CT G…	0.012 ± 0.0004	0.0038 ± 0.0008
9a, 10, 4g	3′…GAGT$_L$GA$_L$CAT$_L$C… 5′…CTCA CT G…	1.5 ± 0.3	0.446 ± 0.003
9a, 10, 4a	3′…GAGT$_L$GA$_L$CAT$_L$C… 5′…CTCA CT G…	0.028 ± 0.0003	0.009 ± 0.002
9t, 10, 4a	3′…GAGT$_L$GA$_L$CTT$_L$C… 5′…CTCA CT G…	1.8 ± 0.2	0.853 ± 0.004
9t, 10, 4c	3′…GAGT$_L$GA$_L$CTT$_L$C… 5′…CTCA CT G…	0.008 ± 0.002	0.0037 ± 0.0006
9t, 10, 4g	3′…GAGT$_L$GA$_L$CTT$_L$C… 5′…CTCA CT G…	0.27 ± 0.06	0.122 ± 0.013
9t, 10, 4t	3′…GAGT$_L$GA$_L$CTT$_L$C… 5′…CTCA CT G…	0.044 ± 0.017	0.021 ± 0.010

Supporting Table S6. Primer conversion at end of assay for extension after different matches and mismatches (refer to Table 1, MS assay). Underlined reactions represent matched base pairs.

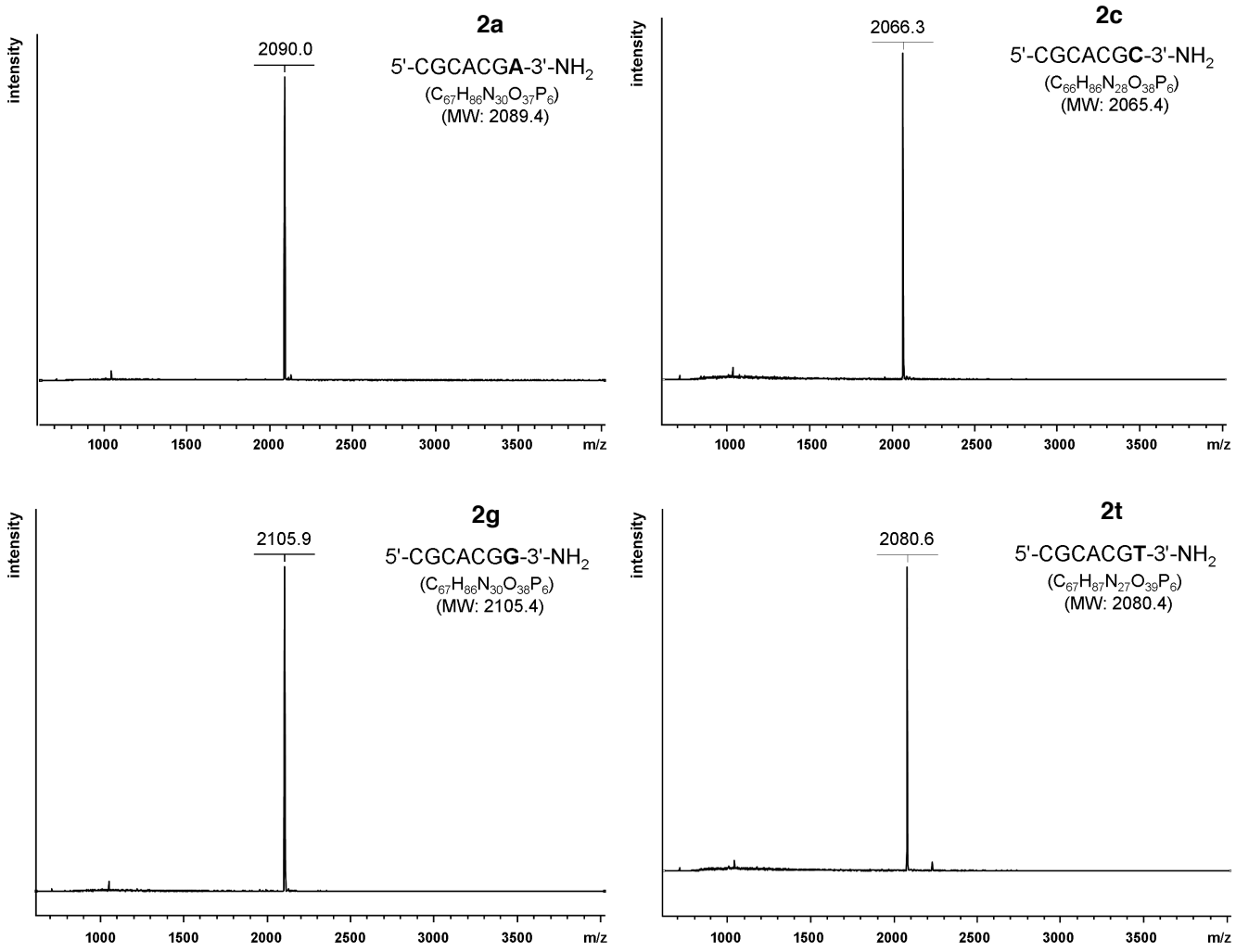
Template Primer		Primer conversion
1tg, 2a, 3a-t	3′-GCGTGC**G**CTG... 5′-CGCACG**A**	98%
<u>1tg, 2c, 3a-t</u>	3′-GCGTGC**G**CTG… 5′-CGCACG**C**	97%
1tg, 2g, 3a-t	3′-GCGTGC**G**CTG… 5′-CGCACG**G**	96%
1tg, 2t, 3a-t	3′-GCGTGC**G**CTG… 5′-CGCACG**T**	99%
1tc, 2a, 3a-t	3′-GCGTGC**C**CTG... 5′-CGCACG**A**	99%
1tc, 2c, 3a-t	3′-GCGTGC**C**CTG… 5′-CGCACG**C**	97%
<u>1tc, 2g, 3a-t</u>	3′-GCGTGC**C**CTG… 5′-CGCACG**G**	99%
1tc, 2t, 3a-t	3′-GCGTGC**C**CTG… 5′-CGCACG**T**	96%
1ta, 2a, 3a-t	3′-GCGTGC**A**CTG... 5′-CGCACG**A**	99%
1ta, 2c, 3a-t	3′-GCGTGC**A**CTG… 5′-CGCACG**C**	95%
1ta, 2g, 3a-t	3′-GCGTGC**A**CTG… 5′-CGCACG**G**	99%
<u>1ta, 2t, 3a-t</u>	3′-GCGTGC**A**CTG… 5′-CGCACG**T**	99%

Supporting Table S7. Primer conversion at end of assay for extension after a single mismatch (refer to Table 2, MS assay).

Reactants		**Primer conversion**
5uu, 2a, 4a-t	3′-GCGUGC**U**CUG… 5′-CGCACG**A**	99%
5uu, 2c, 4a-t	3′-GCGUGC**U**CUG… 5′-CGCACG**C**	13%
5uu, 2g, 4a-t	3′-GCGUGC**U**CUG… 5′-CGCACG**G**	95%
5uu, 2t, 4a-t	3′-GCGUGC**U**CUG… 5′-CGCACG**T**	28%
1tt, 2a, 4a-t	3′-GCGTGC**T**CTG... 5′-CGCACG**A**	99%
1tt, 2c, 4a-t	3′-GCGTGC**T**CTG… 5′-CGCACG**C**	20%
1tt, 2g, 4a-t	3′-GCGTGC**T**CTG… 5′-CGCACG**G**	98%
1tt, 2t, 4a-t	3′-GCGTGC**T**CTG… 5′-CGCACG**T**	48%
5uu, 2a, 3a-t	3′-GCGUGC**U**CUG… 5′-CGCACG**A**	95%
5uu, 2c, 3a-t	3′-GCGUGC**U**CUG… 5′-CGCACG**C**	19%
5uu, 2g, 3a-t	3′-GCGUGC**U**CUG… 5′-CGCACG**G**	98%
5uu, 2t, 3a-t	3′-GCGUGC**U**CUG… 5′-CGCACG**T**	52%
1tt, 2a, 3a-t	3′-GCGTGC**T**CTG... 5′-CGCACG**A**	99%
1tt, 2c, 3a-t	3′-GCGTGC**T**CTG… 5′-CGCACG**C**	99%
1tt, 2g, 3a-t	3′-GCGTGC**T**CTG… 5′-CGCACG**G**	99%
1tt, 2t, 3a-t	3′-GCGTGC**T**CTG… 5′-CGCACG**T**	99%

Supporting Figure S2. MALDI TOF mass spectra of 3′-aminoterminal primers (**2a, 2c, 2g, 2t**) (negative, linear mode). Labeled peaks are pseudomolecular ions ($[M-H]^-$) determined of maxima of unresolved isotope patterns. Molecular formulae and their masses are those of the neutral molecule. Mass accuracy ± 0.1 %.

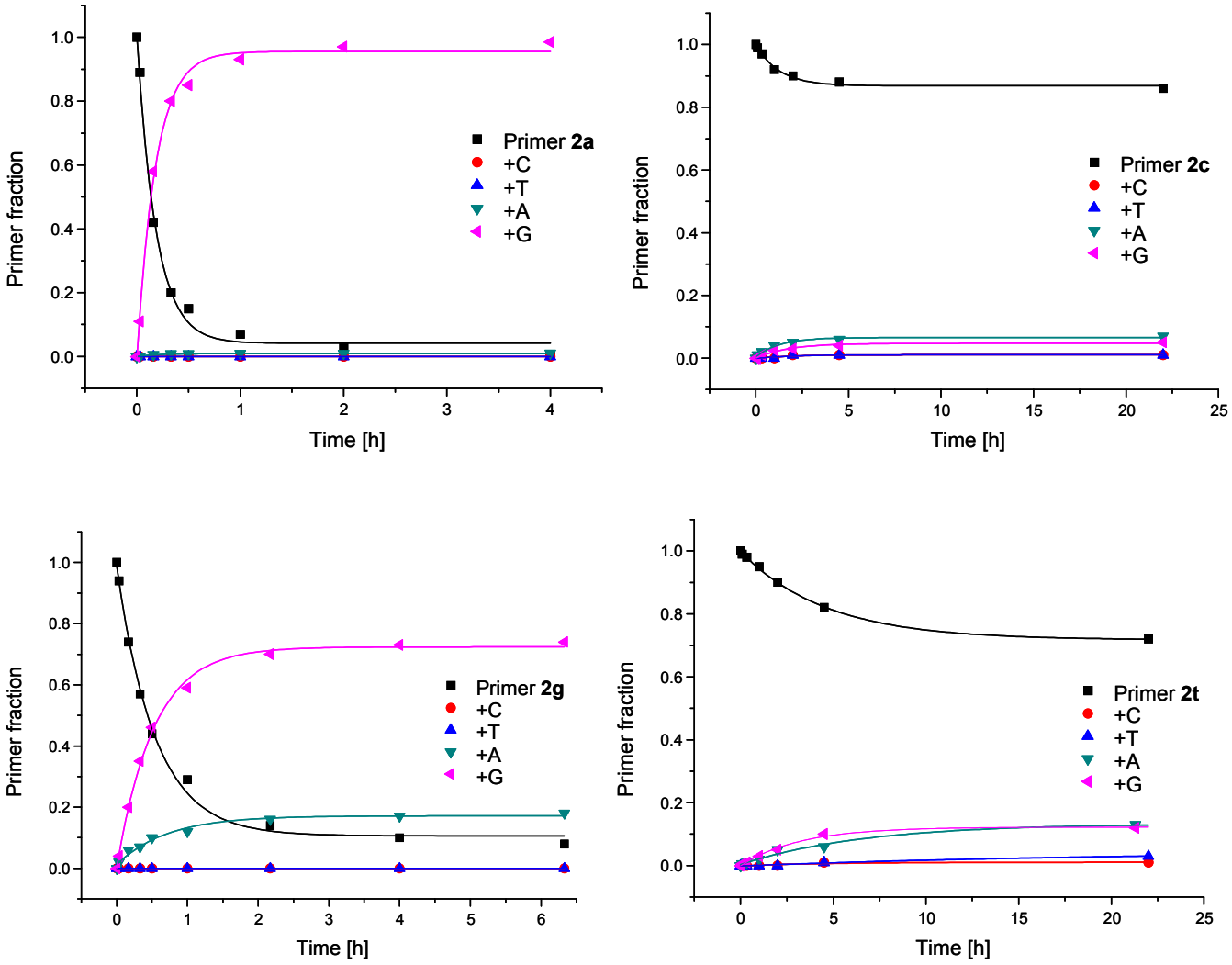
2a: 5′-CGCACGA-3′-NH_2 MALDI-TOF MS calc for $C_{67}H_{86}N_{30}O_{37}P_6$ 2089.4, found 2090.0;
2c: 5′-CGCACGC-3′-NH_2 MALDI-TOF MS calc for $C_{66}H_{86}N_{28}O_{38}P_6$, 2065.4 found 2066.3;
2g: 5′-CGCACGG-3′-NH_2 MALDI-TOF MS calc for $C_{67}H_{86}N_{30}O_{38}P_6$ 2105.4, found 2105.9;
2t: 5′-CGCACGT-3′-NH_2 MALDI-TOF MS calc for $C_{67}H_{87}N_{27}O_{39}P_6$ 2080.4, found 2080.6.



Supporting Figure S3. Kinetic data and fits used for the determination of rate constants presented in Table 2. Conditions: mixture of imidazolides of monomers (**4a-t**, 3.6 mM each), 36 μM 3′-aminoterminal primer (**2a, 2c, 2g, 2t**), 54 μM RNA template **5uu**, 20°C, 200 mM HEPES buffer (400 mM NaCl, 80 mM $MgCl_2$), pH 7.0.

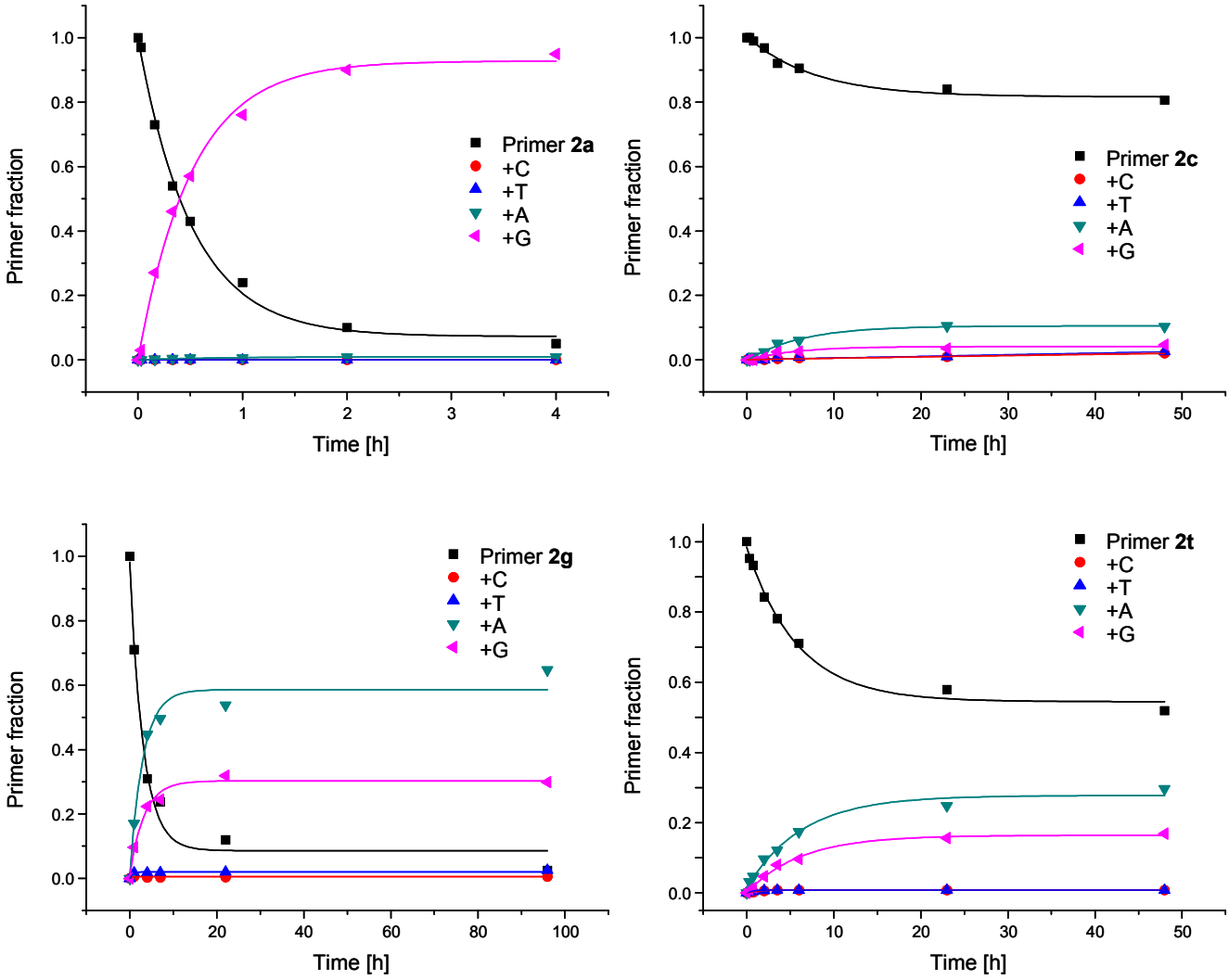
Plots of Kinetics of Primer Extension, MS assay

Monomers: **4a, 4c, 4g, 4t**
Template: **5uu**



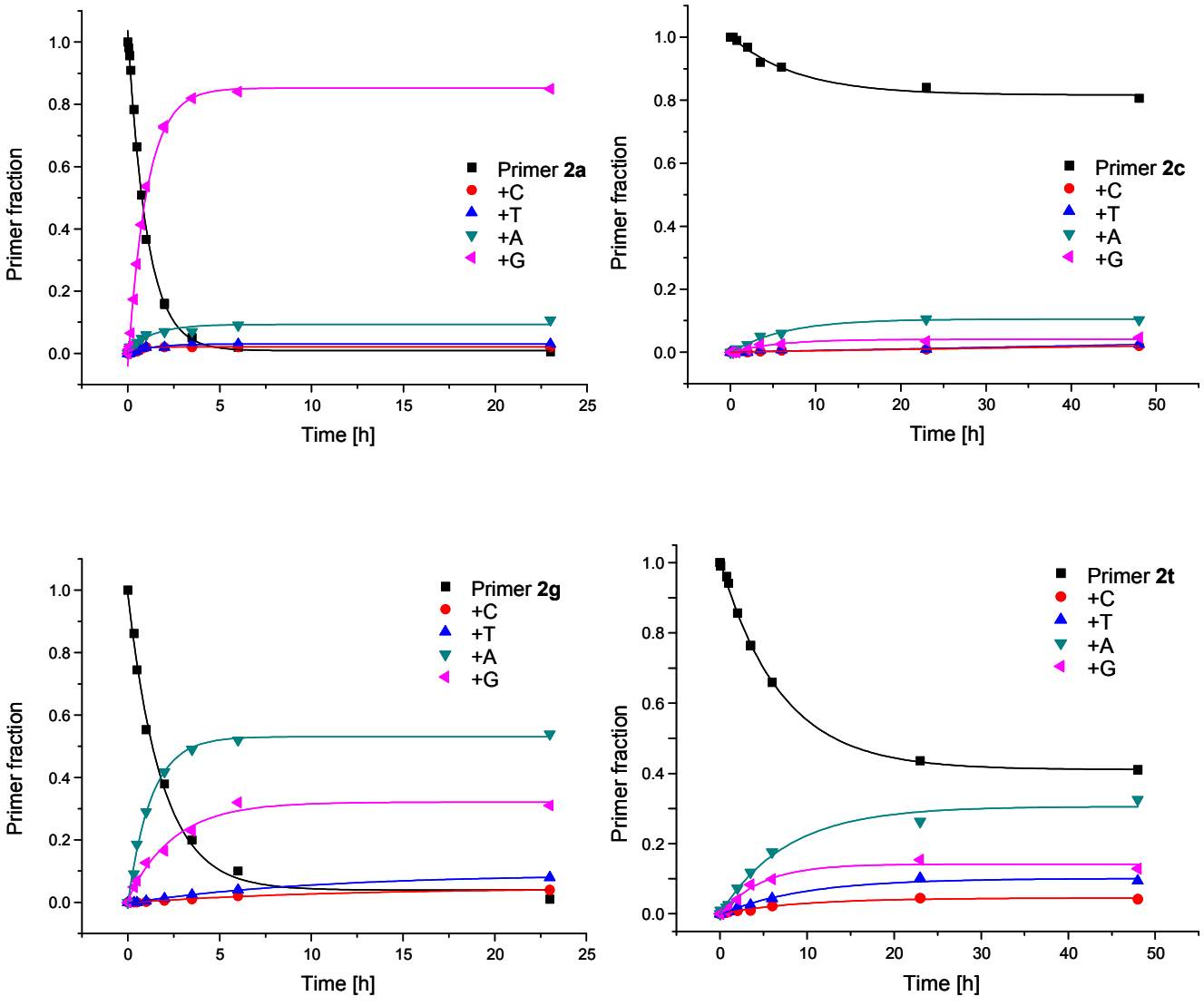
Supporting Figure S4. Kinetic data and fits used for the determination of rate constants presented in Table 2. Conditions: mixture of imidazolides of monomers (**4a-t**, 3.6 mM each), 36 μM 3′-aminoterminal primer (**2a, 2c, 2g, 2t**), 54 μM DNA template **1tt**, 20°C, 200 mM HEPES buffer (400 mM NaCl, 80 mM $MgCl_2$), pH 7.0.

Monomers: **4a, 4c, 4g, 4t**
Template: **1tt**



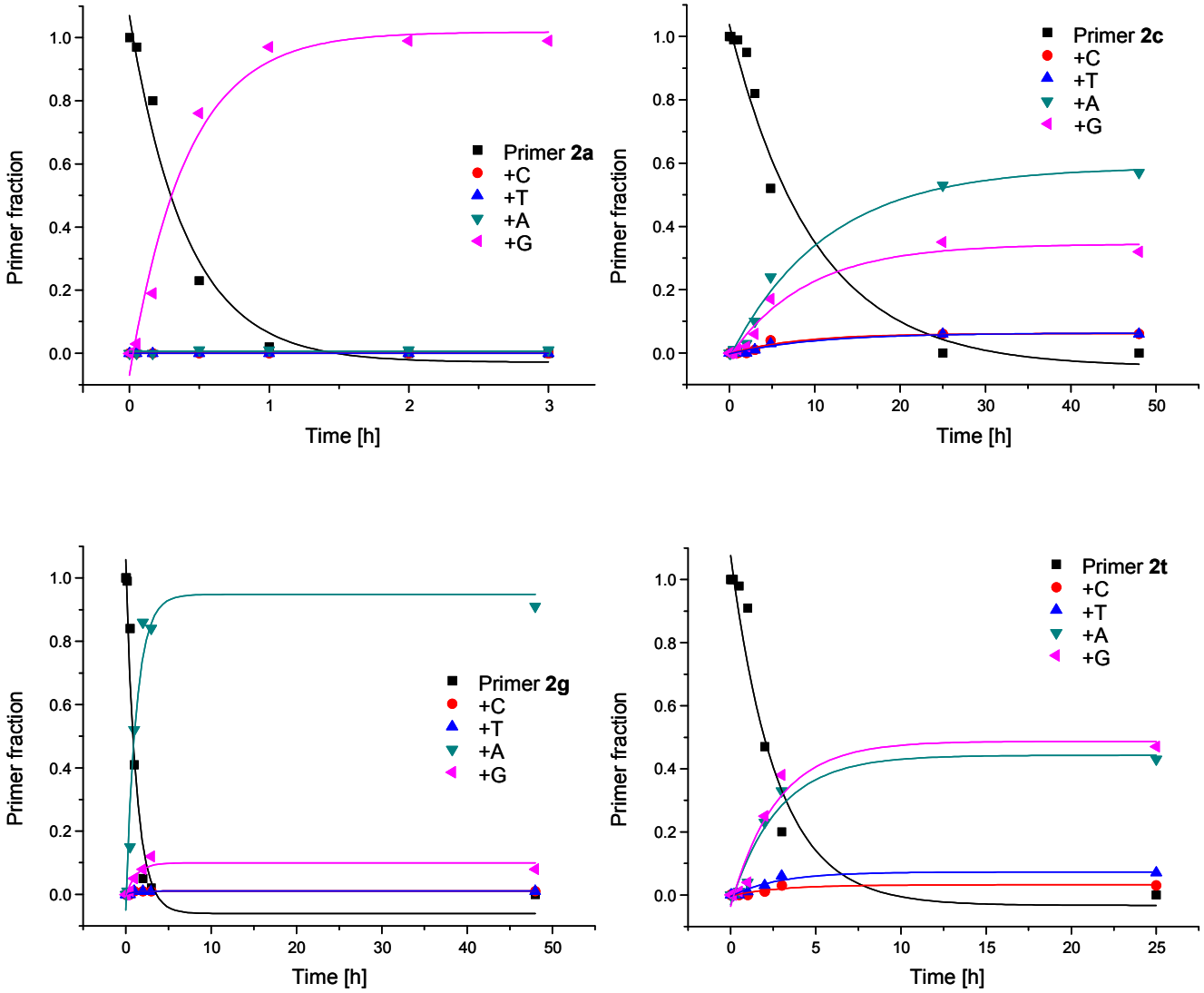
Supporting Figure S5. Kinetic data and fits used for the determination of rate constants presented in Table 2. Conditions: mixture of OAt esters of monomers (**3a-t**, 3.6 mM each), 36 μM 3′-aminoterminal primer (**2a, 2c, 2g, 2t**), 54 μM RNA template **5uu**, 20°C, 200 mM HEPES buffer (400 mM NaCl, 80 mM $MgCl_2$), pH 8.9.

Monomers: **3a, 3c, 3g, 3t**
Template: **5uu**



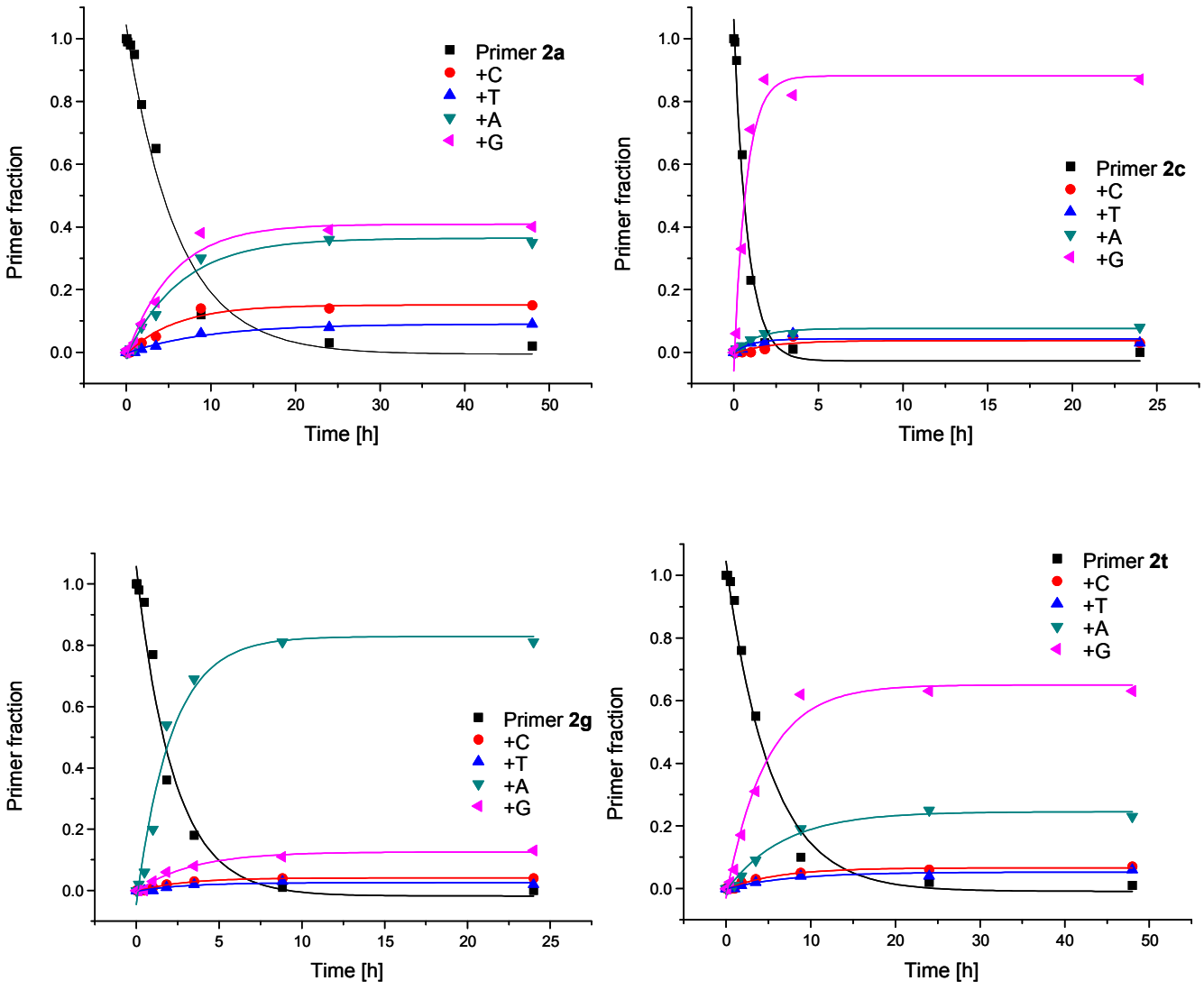
Supporting Figure S6. Kinetic data and fits used for the determination of rate constants presented in Table 2. Conditions: mixture of OAt esters of monomers (**3a-t**, 3.6 mM each), 36 μM 3′-aminoterminal primer (**2a, 2c, 2g, 2t**), 54 μM DNA template **1tt**, 20°C, 200 mM HEPES buffer (400 mM NaCl, 80 mM $MgCl_2$), pH 8.9.

Monomers: **3a, 3c, 3g, 3t**
Template: **1tt**



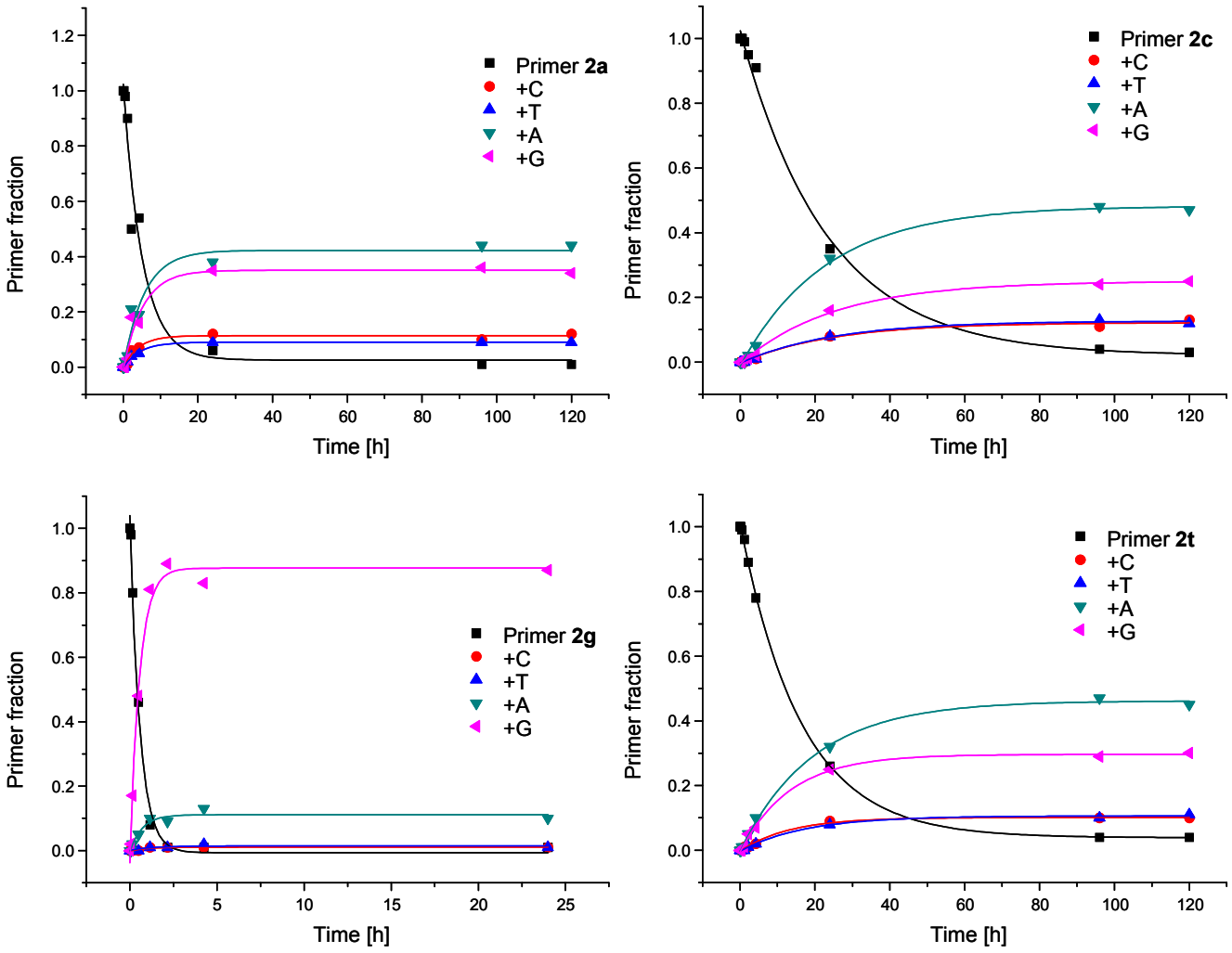
Supporting Figure S7. Kinetic data and fits used for the determination of rate constants presented in Table 1. Conditions: mixture of OAt esters of monomers (**3a-t,** 3.6 mM each), 36 μM 3′-aminoterminal primer (**2a, 2c, 2g, 2t**), 54 μM DNA template **1tg**, 20°C, 200 mM HEPES buffer (400 mM NaCl, 80 mM $MgCl_2$), pH 8.9.

Monomers: **3a, 3c, 3g, 3t**
Template: **1tg**



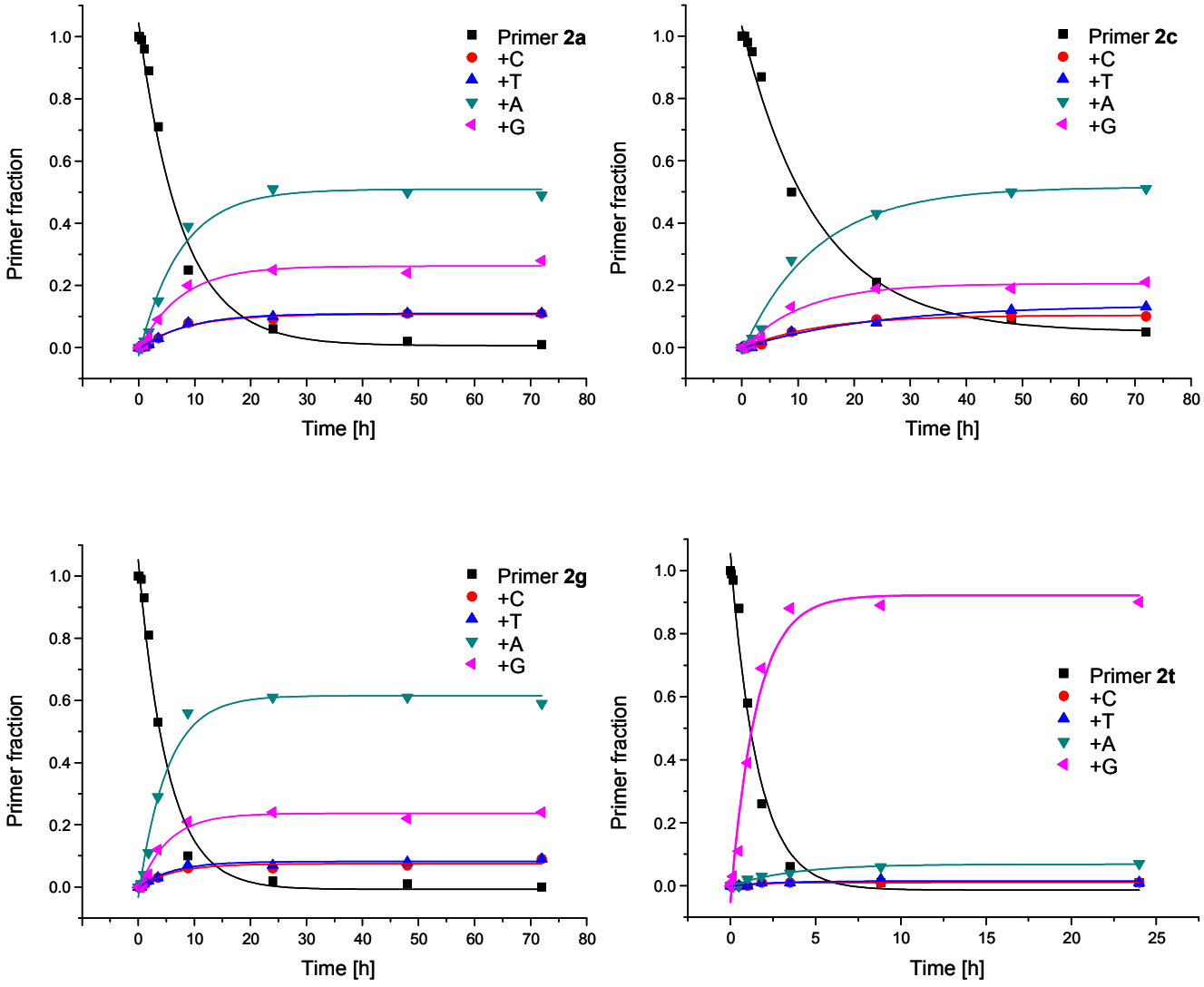
Supporting Figure S8. Kinetic data and fits used for the determination of rate constants presented in Table 1. Conditions: mixture of OAt esters of monomers (**3a-t,** 3.6 mM each), 36 μM 3′-aminoterminal primer (2a, 2c, 2g, 2t), 54 μM template **1tc**, 20°C, 200 mM HEPES buffer (400 mM NaCl, 80 mM $MgCl_2$), pH 8.9.

Monomers: **3a, 3c, 3g, 3t**
Template: **1tc**

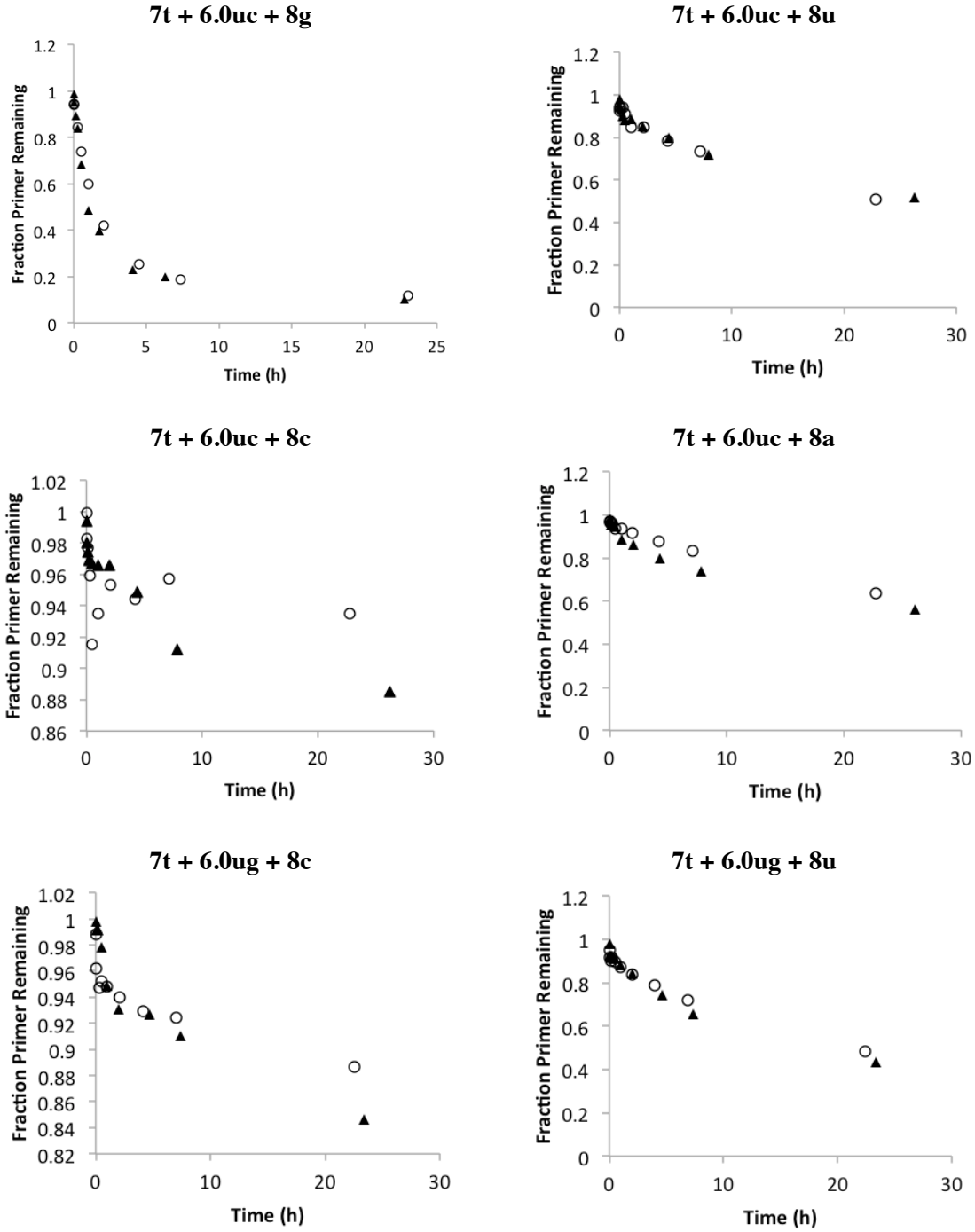


Supporting Figure S9. Kinetic data and fits used for the determination of rate constants presented in Table 1. Conditions: mixture of OAt esters of monomers (**3a-t**, 3.6 mM each), 36 μM 3′-aminoterminal primer (**2a, 2c, 2g, 2t**), 54 μM DNA template **1ta**, 20°C, 200 mM HEPES buffer (400 mM NaCl, 80 mM $MgCl_2$), pH 8.9.

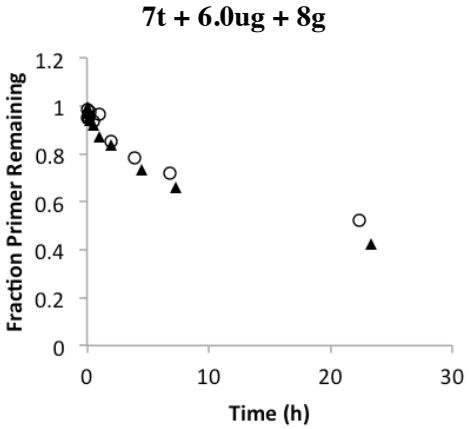
Monomers: **3a, 3c, 3g, 3t**
Template: **1ta**

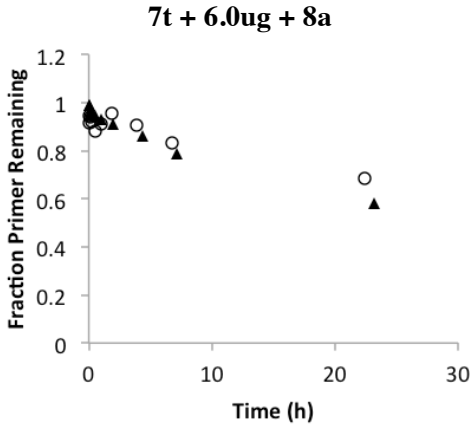


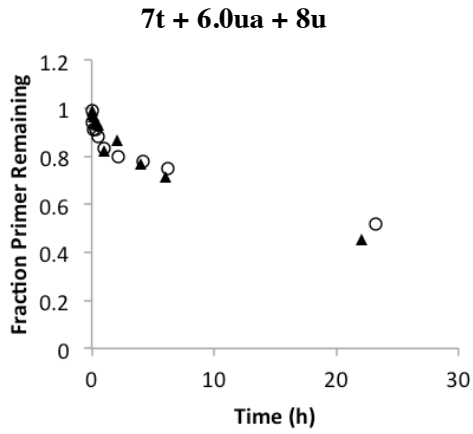
Supporting Figure S10. Plots of kinetics of primer extension: PAGE assay. Corresponding reaction is indicated above each graph. Independent experimental replicates are shown within each plot.

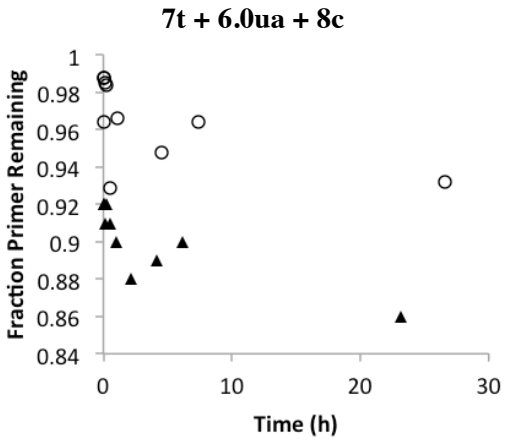


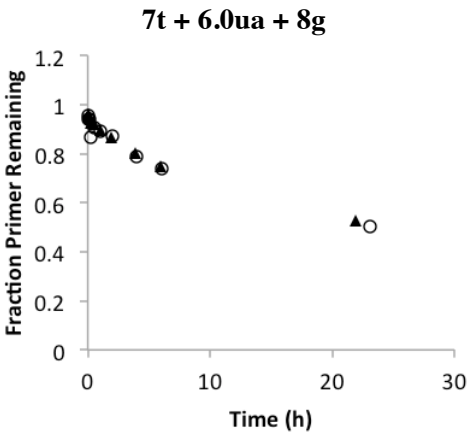
Supporting Figure S11. Plots of kinetics of primer extension: PAGE assay; cont.

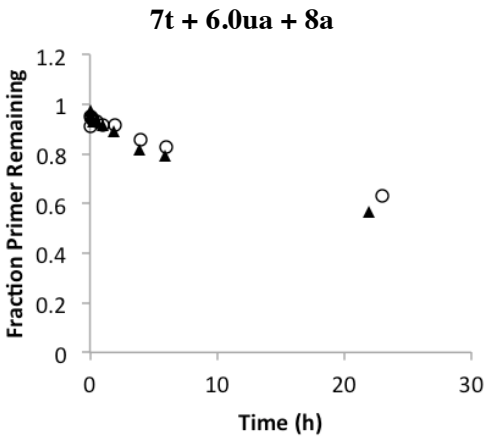




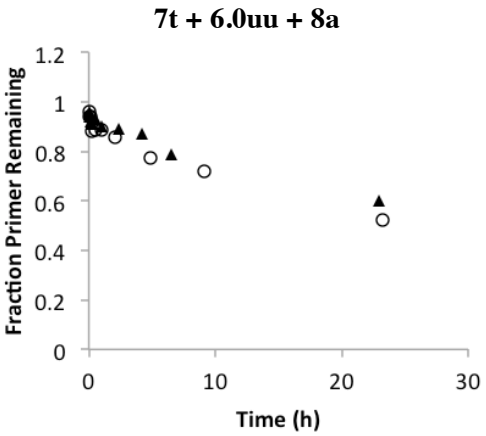


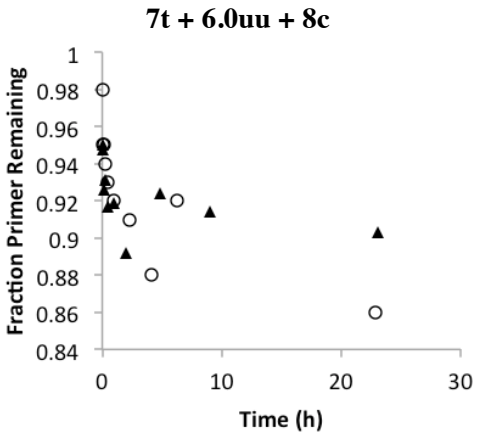


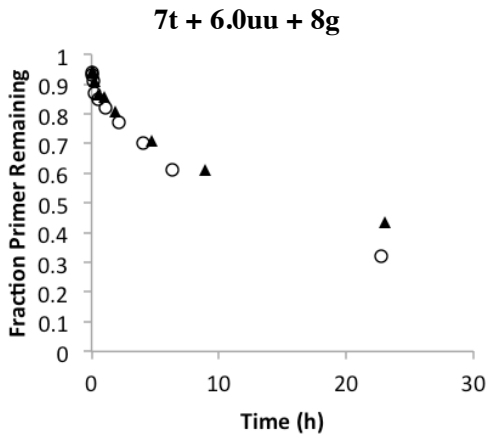


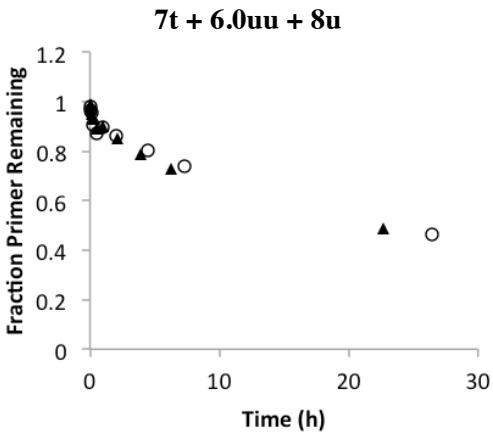


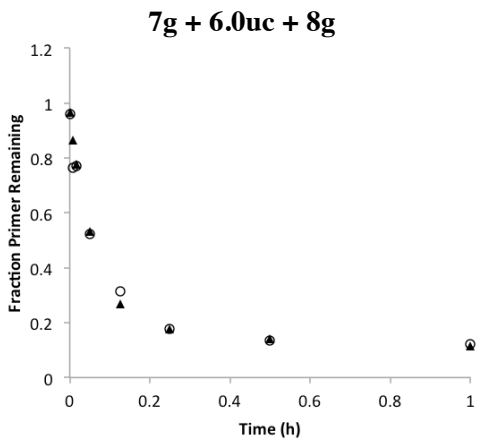
Supporting Figure S12. Plots of kinetics of primer extension: PAGE assay; cont.

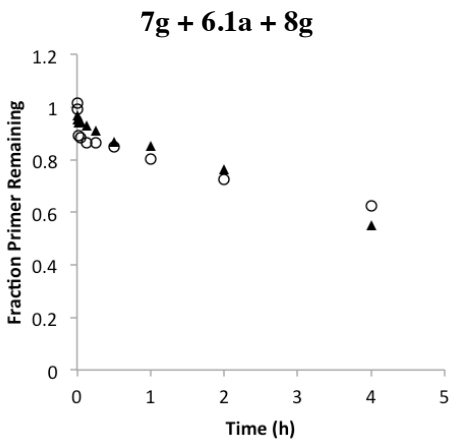




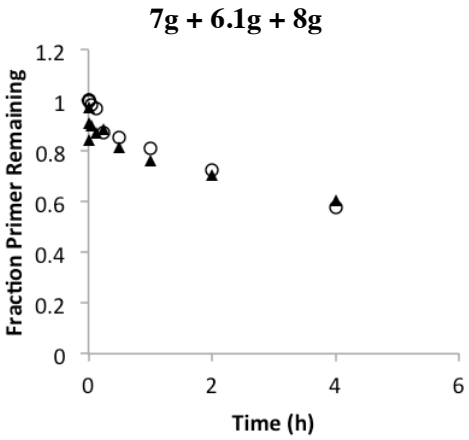


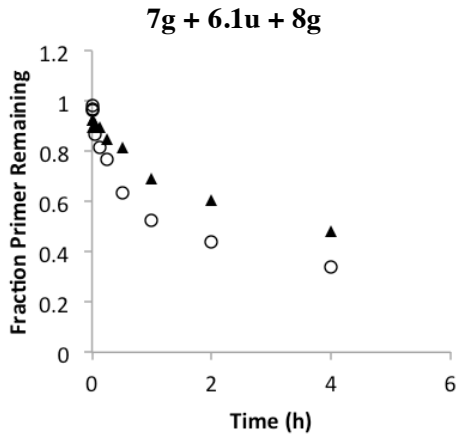


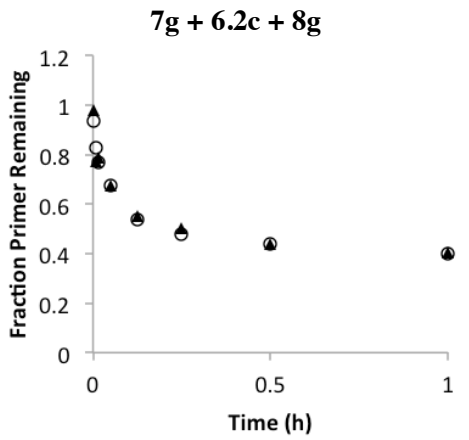


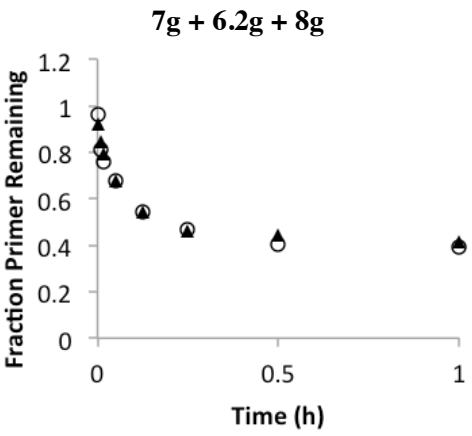


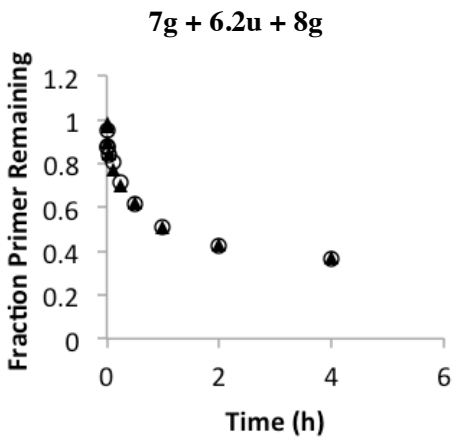
Supporting Figure S13. Plots of kinetics of primer extension: PAGE assay; cont.

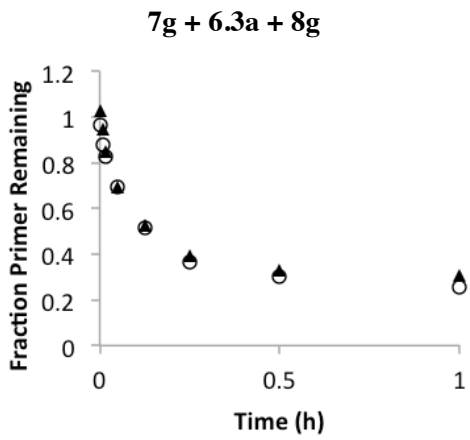




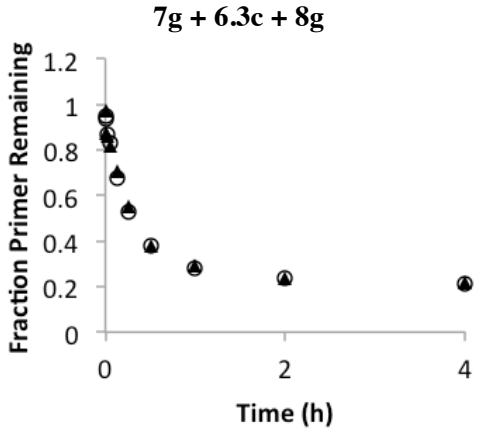


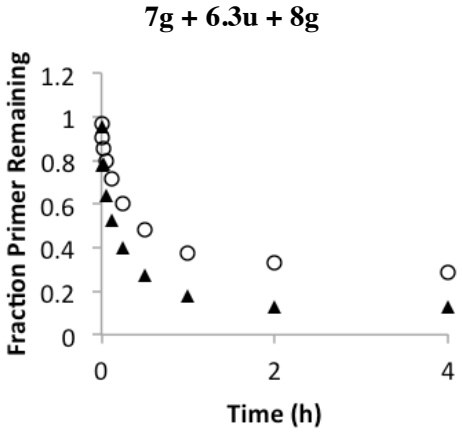


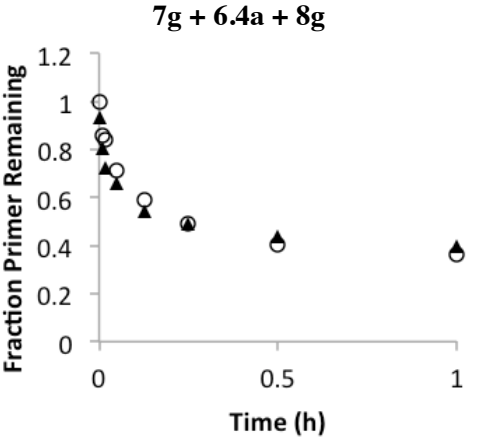


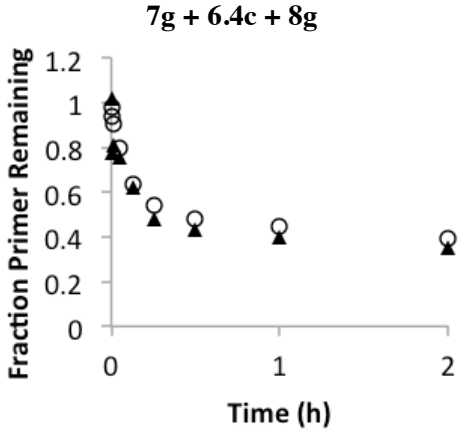


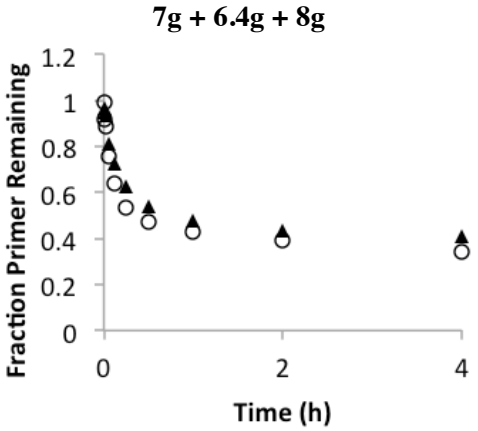
Supporting Figure S14. Plots of kinetics of primer extension: PAGE assay; cont.

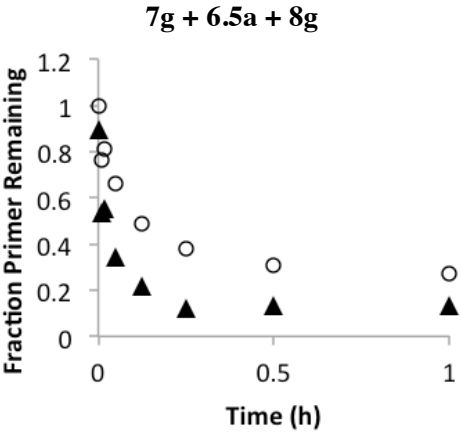




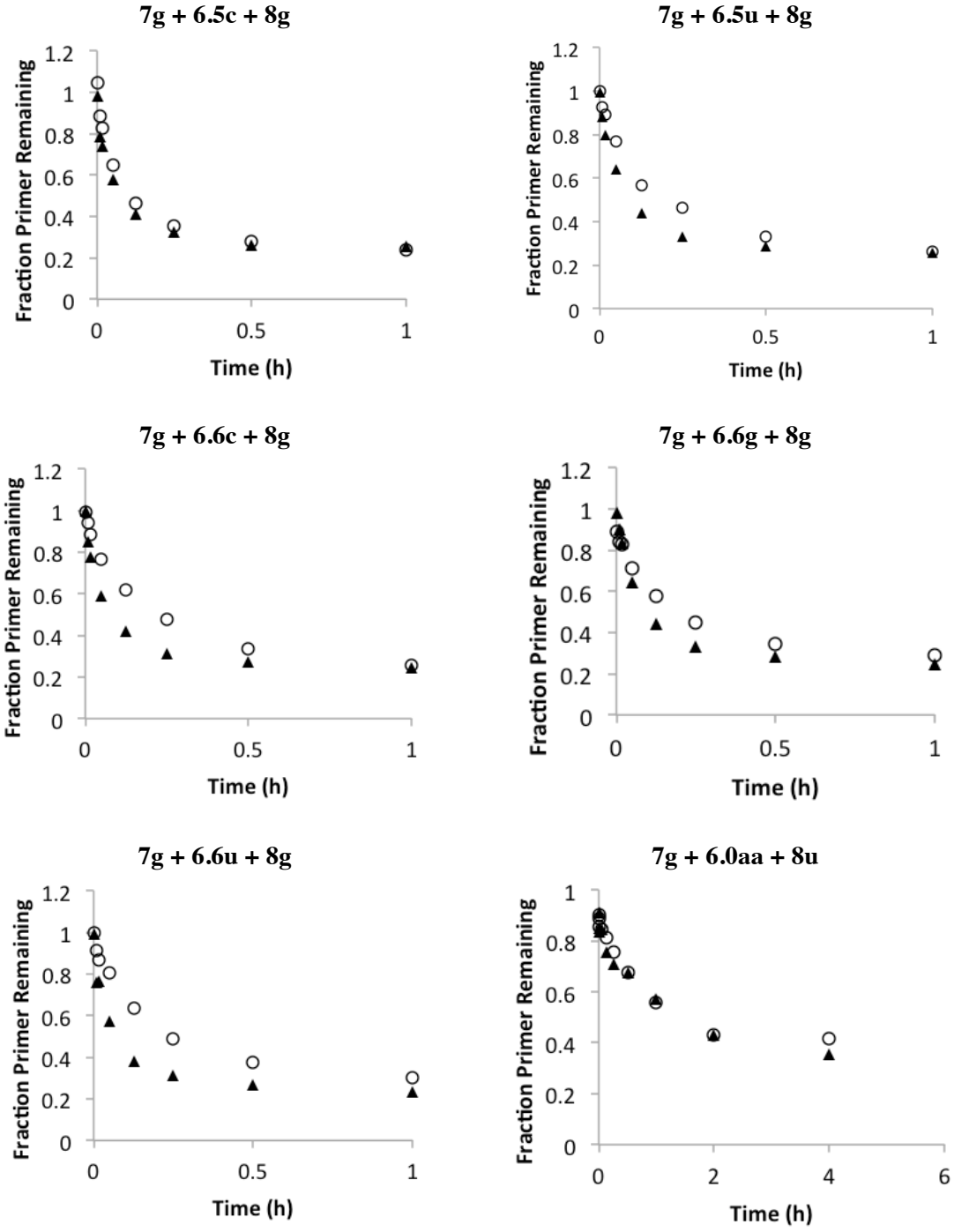




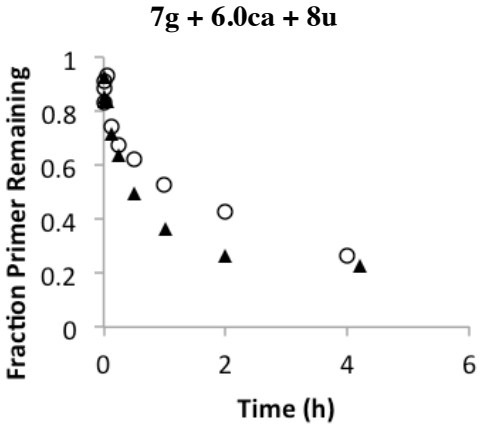


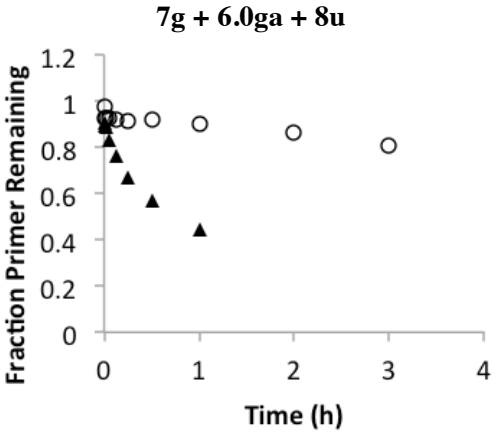


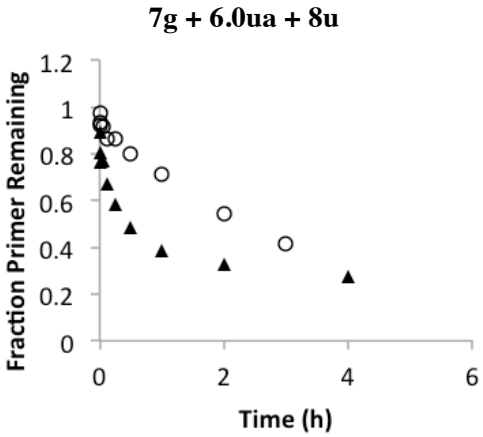
Supporting Figure S15. Plots of kinetics of primer extension: PAGE assay; cont.

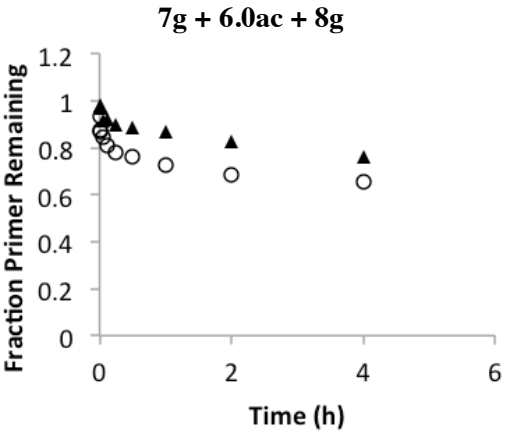


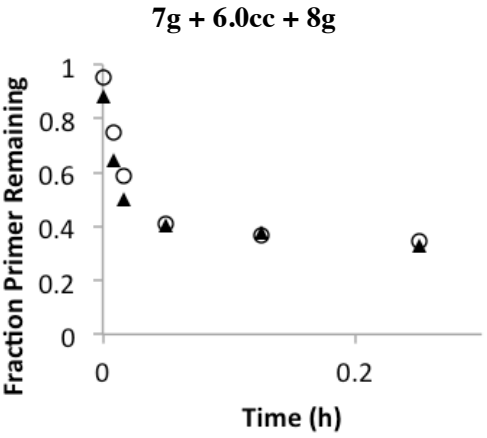
Supporting Figure S16. Plots of kinetics of primer extension: PAGE assay; cont.

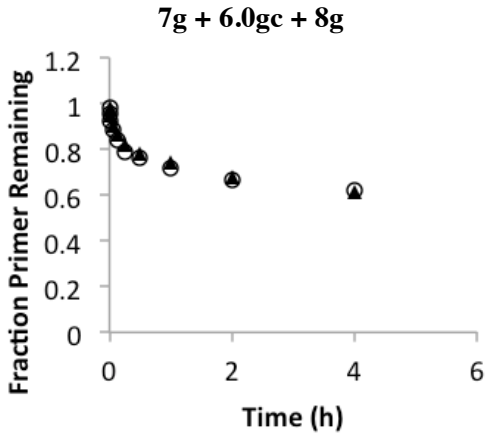




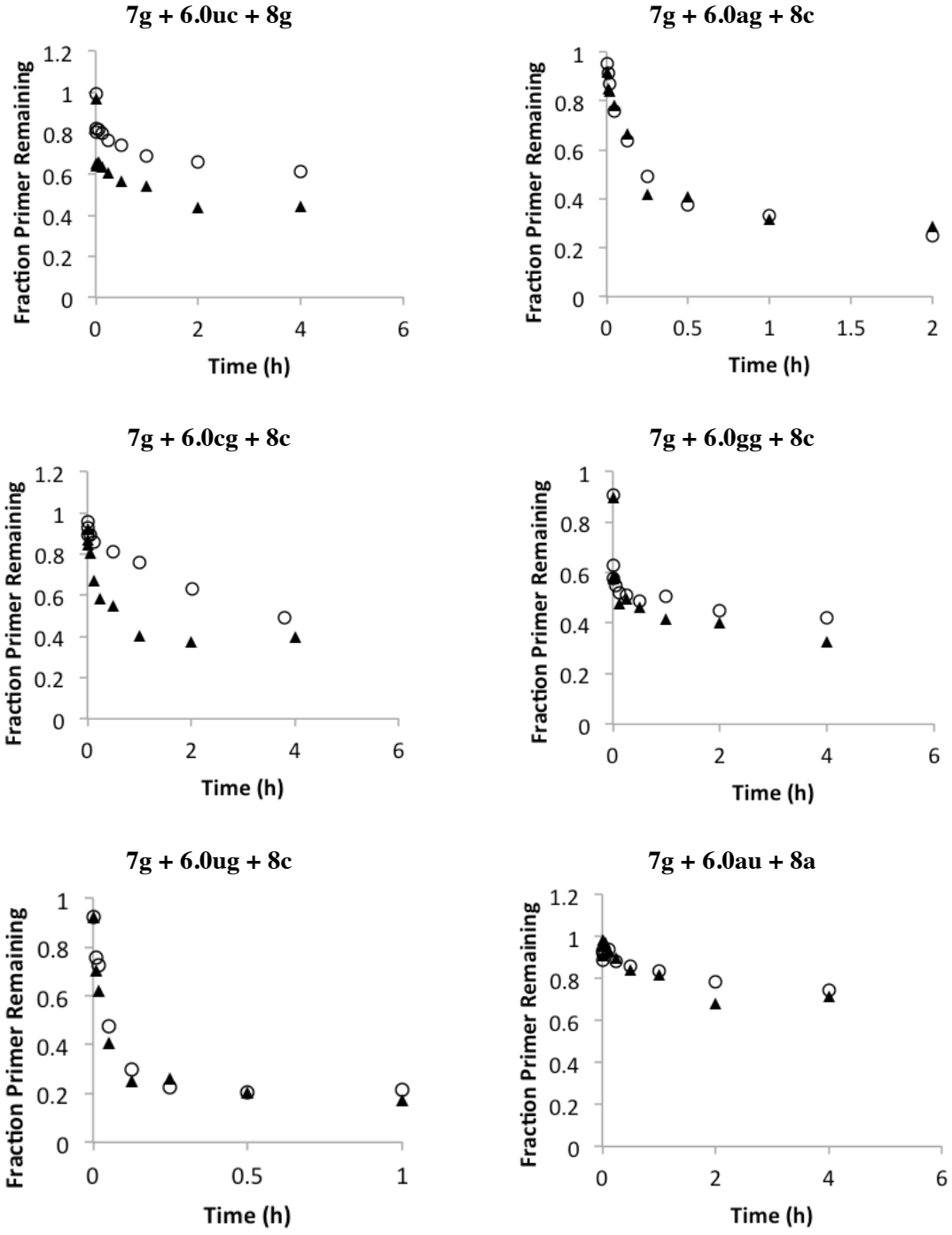




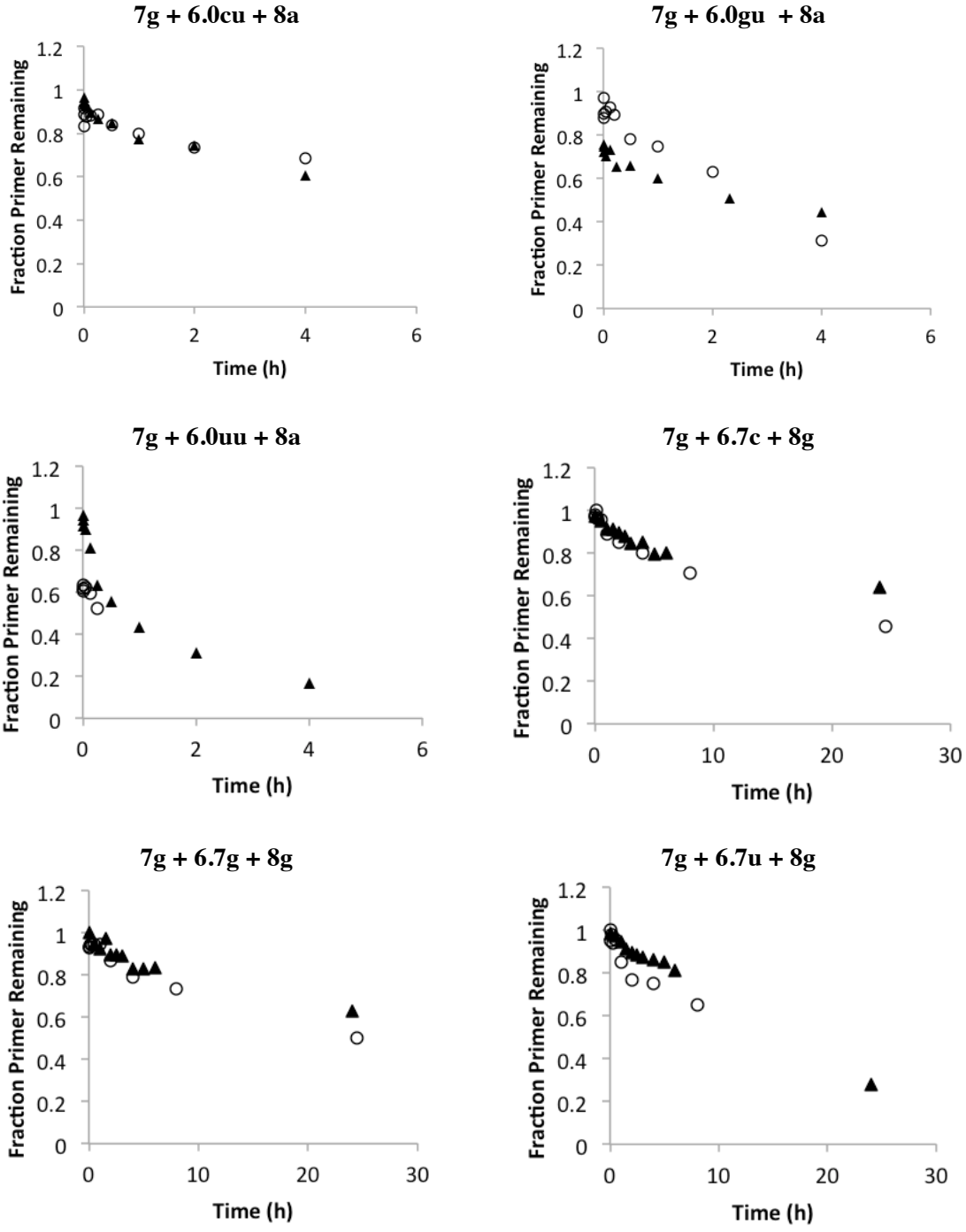




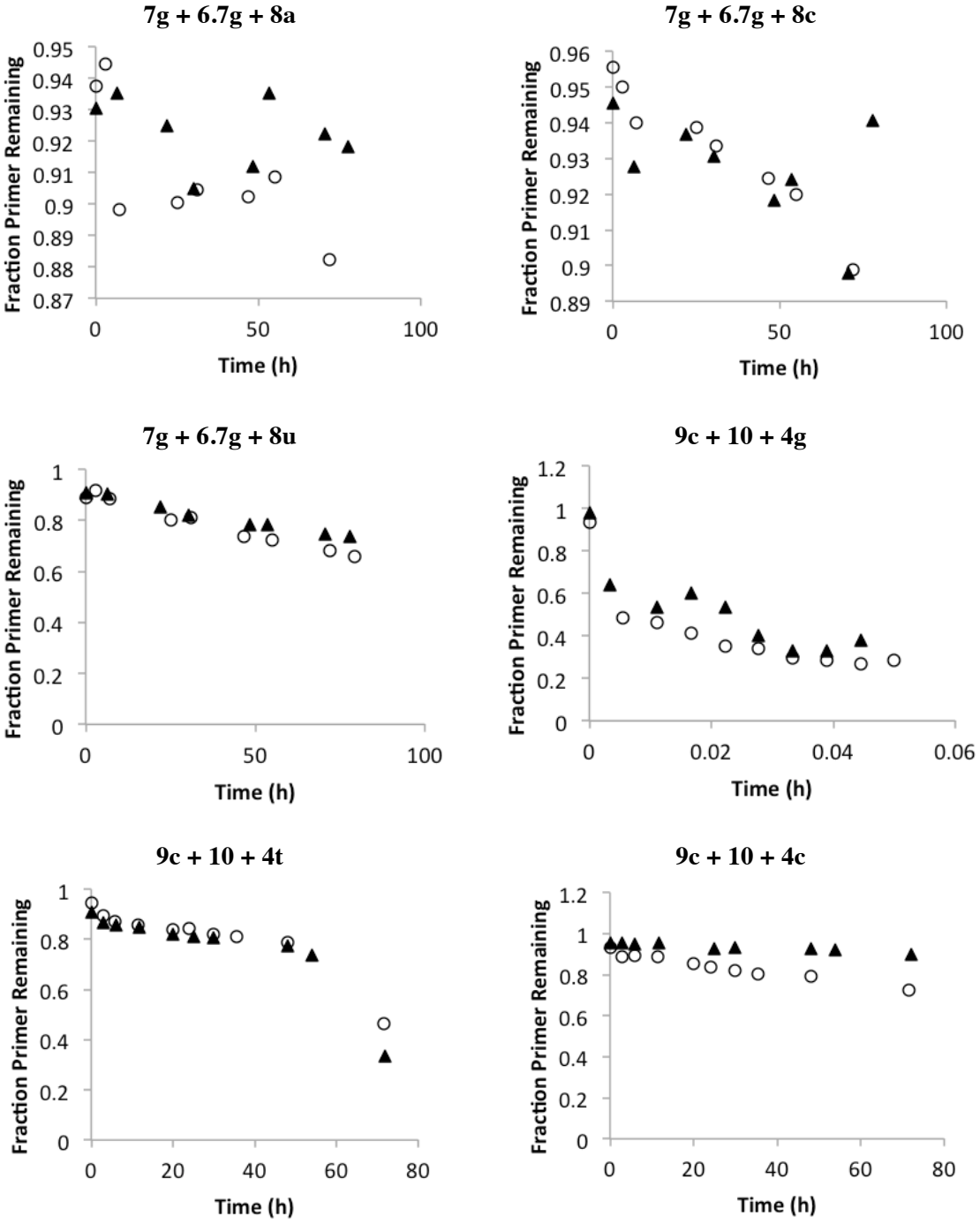
Supporting Figure S17. Plots of kinetics of primer extension: PAGE assay; cont.



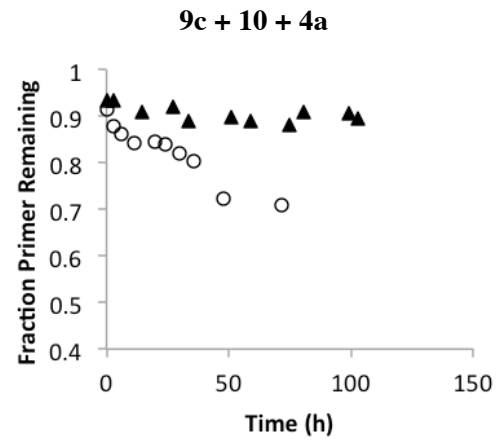
Supporting Figure S18. Plots of kinetics of primer extension: PAGE assay; cont.

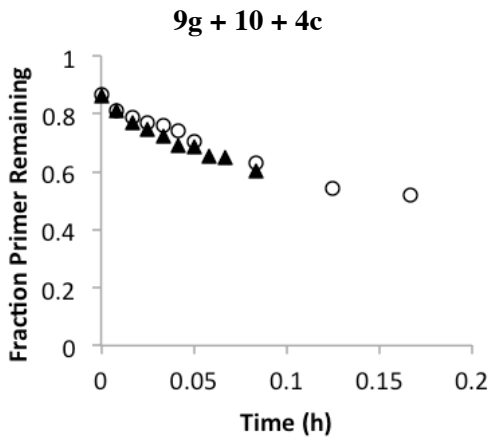


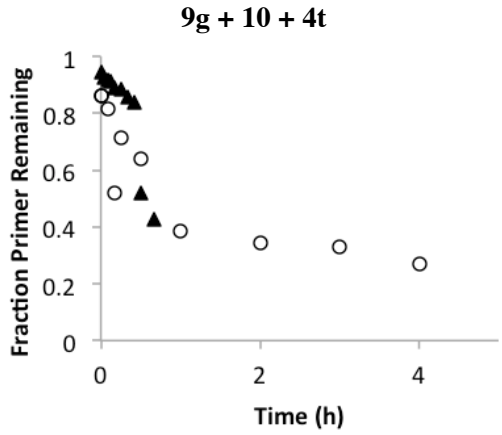
Supporting Figure S19. Plots of kinetics of primer extension: PAGE assay; cont.

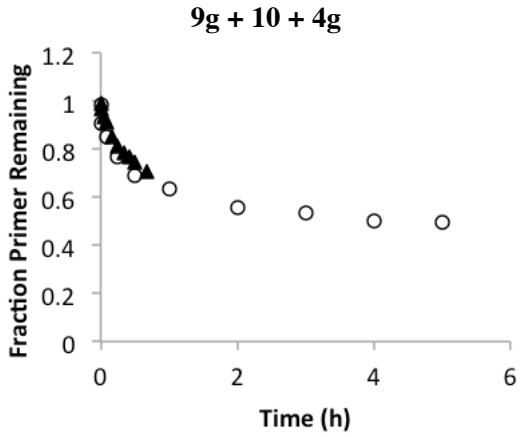


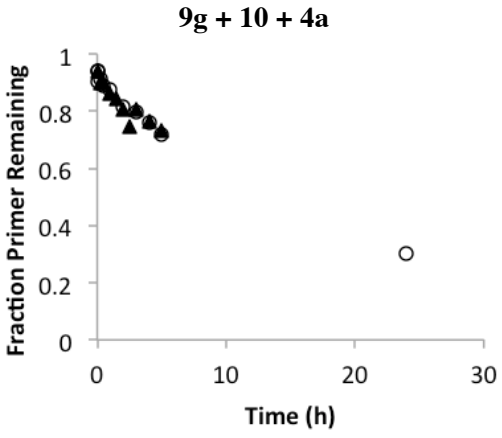
Supporting Figure S20. Plots of kinetics of primer extension: PAGE assay; cont.

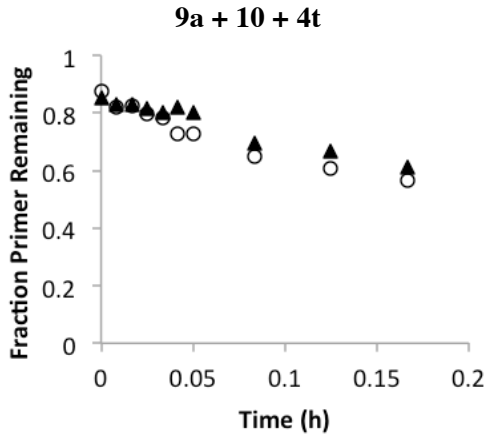




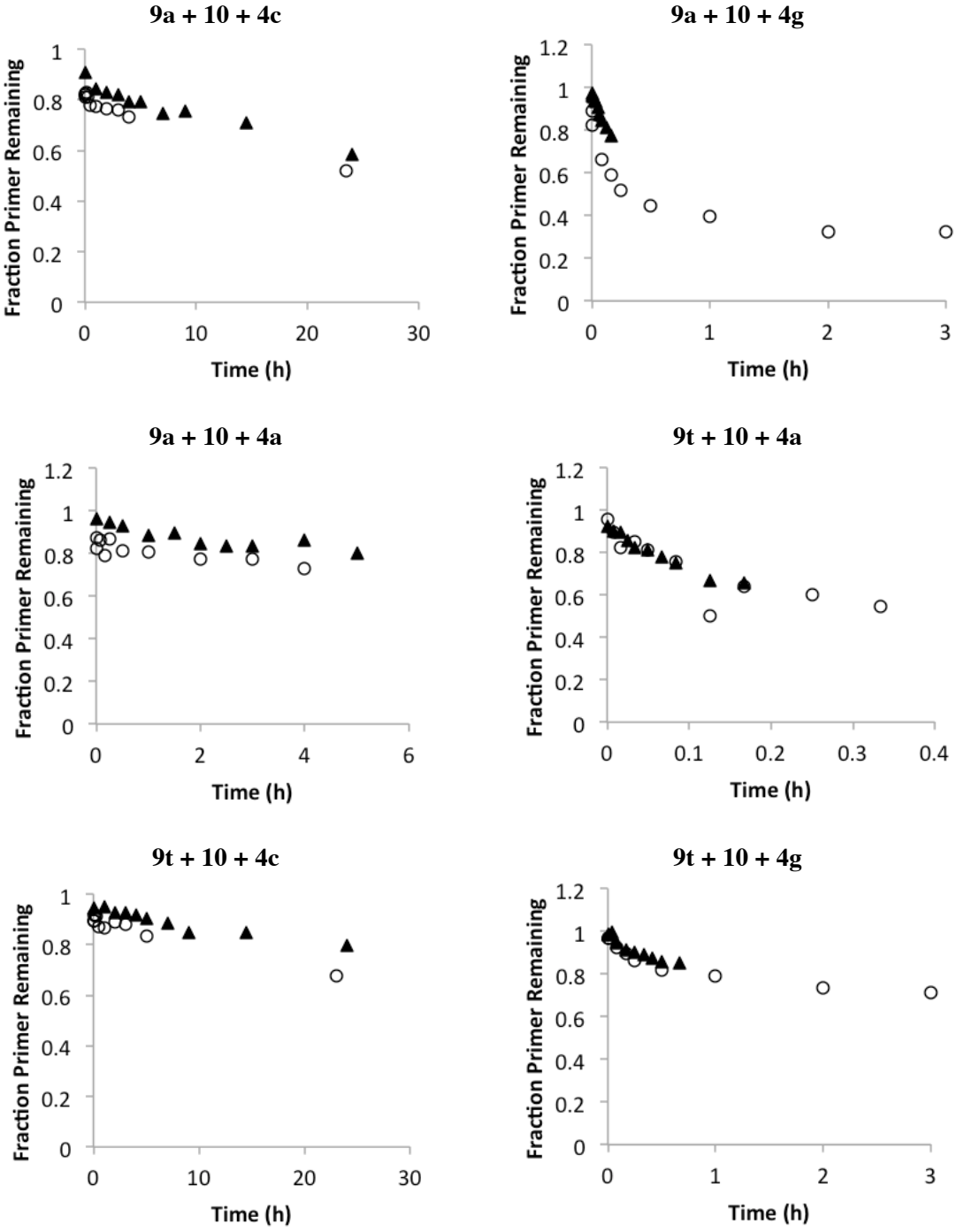




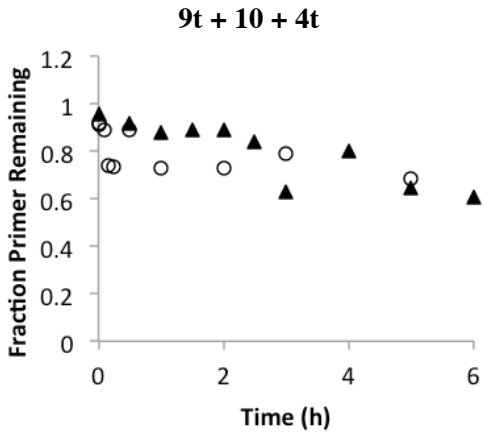




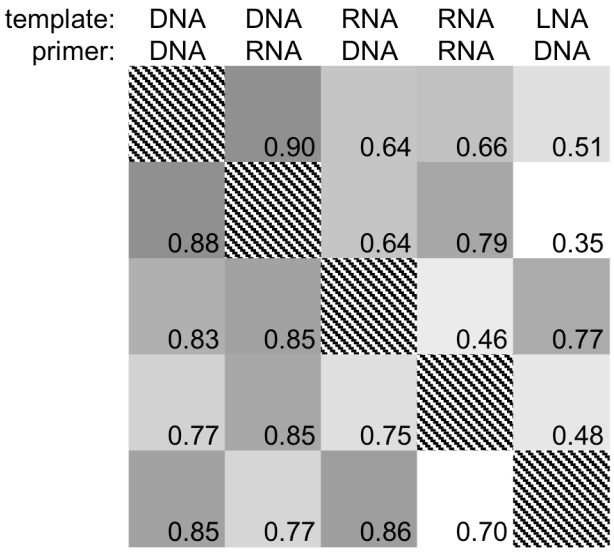
Supporting Figure S21. Plots of kinetics of primer extension: PAGE assay; cont.



Supporting Figure S22. Plots of kinetics of primer extension: PAGE assay; cont.

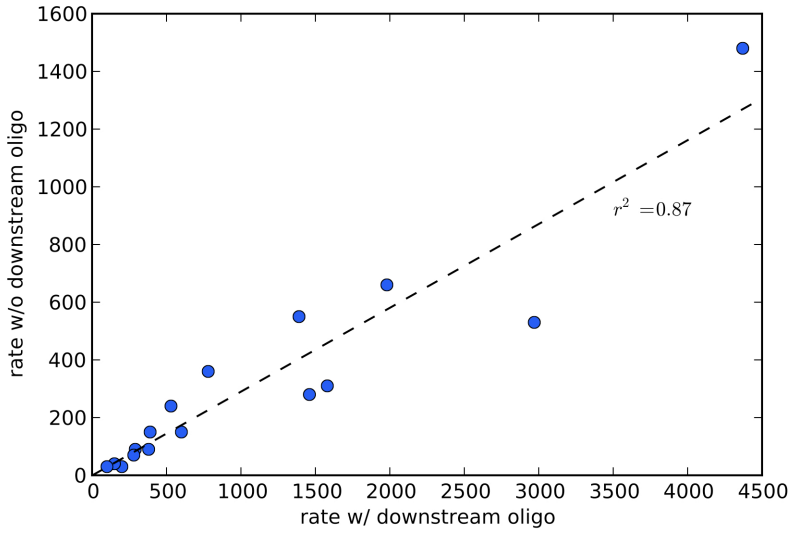


Supporting Figure S23: Comparison of mutation profiles for different nucleic acid systems. Data are from the current paper and two literature references (Rajamani et al. 2010, *JACS* **132**, 5880-5885; Leu et al. 2011, *NAR* **39**, 8135-8147). Overall similarity was estimated in two ways for comparing the mutation profiles of different systems: a Pearson correlation between the logarithm of the error rates for two systems or the dot product between these values. In both cases we used only the 12 values for misincorporations in each system. Lower triangular part is the Pearson correlation; upper triangular part is the normalized dot product. Diagonal entries (striped) have values of 1. Note that LNA/DNA is most similar to RNA/DNA.



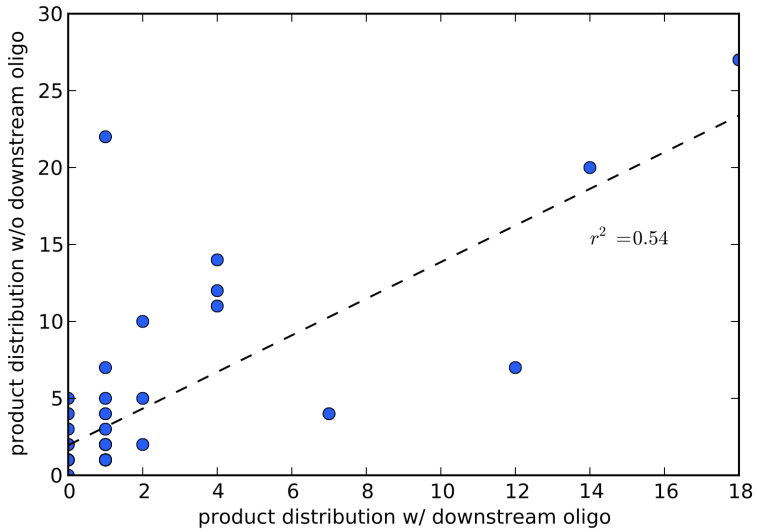
Parameter Extrapolation

1. Incorporation rates: based on Kervio et al. 2010, *PNAS* **107**:12074, Tables 1 and 2. Table 1 contains a complete set of rates (triplet context) in the presence of a downstream 'helper' oligo, which increases the rate by a factor of ~3 on average. Table 2 contains a subset of rates determined in the absence of the helper oligo. To obtain a full set of values for a scenario without helper oligo, we regressed the rates from Table 1 and Table 2 against each other (when both values were available; Supporting Figure S24). The resulting relationship was used to estimate the rate of reaction in the absence of helper oligo for the complete set of contexts.



Supporting Figure S24. Relationship between extension rate with and without helper oligo.

2. Mutation rates: based on Kervio et al. 2010, *PNAS* **107**:12074, Table 5 and Table S3. The complete set of values was estimated for the scenario in the absence of helper oligo by linear regression as described in #1 above (Supporting Figure S25). In addition, the rates were adjusted for a scenario with equimolar monomers, assuming a first-order dependence of reaction rate on the concentration of monomer.



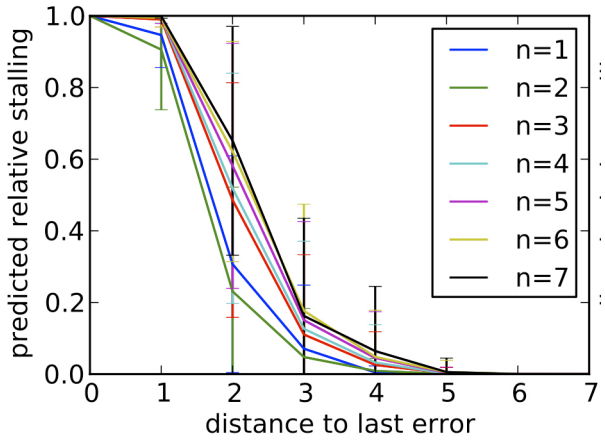
Supporting Figure S25. Relationship between mutation frequencies with and without helper oligo.

3. Stalling factor after single or multiple mismatch: for a single mismatch, stalling factors are from Supporting Figure S1. We observed that a double mismatch caused 6-fold greater stalling than a single mismatch (Figure 6 from main text). This factor might be greater for even larger stretches of mismatches. For the simulation, we assume that a duplex with 2 or more consecutive mismatches at the primer terminus has a 10-fold increased stalling factor over the corresponding single mismatch. This assumption was changed (see robustness analysis below), without substantial effect on the results.

4. Mutation rates after mismatch: based on Table 1 and 2 in the main text. Experimental observations were limited to template base C. Fidelities after a mismatch were compared to the corresponding matched reaction to determine the factor by which the presence of a mismatch lowered fidelity. This factor was then applied to other template bases (matched fidelities from #1). These values were used regardless of the number of consecutive mismatches. Incorporation of incorrect monomers after a mismatch was assumed to be unbiased. This was because the product distribution after multiple errors should essentially reflect the preference of untemplated reactions, which have not been measured in detail. However, it is only important whether or not the base complementary to the corresponding template base is incorporated. Since the template is random, any

incorporation bias should not be very important once results have been averaged over different template sequences.

5. Polymerization rate after recovery from mismatches. We observed that the polymerization rate recovers to nearly normal if a mismatch is followed by a correctly matched base pair (main text Figure 4). This observation matched a theoretical prediction based on the base pairing probability of the last pair, calculated using RNA folding simulations. Therefore, to extend this observation to more cases (i.e., different sequences and larger mutation clusters), we used a theoretical prediction for the probability of formation of the terminal base pair to calculate the extension rate (Supporting Figure S26). In general, these adjustments were relatively minor.



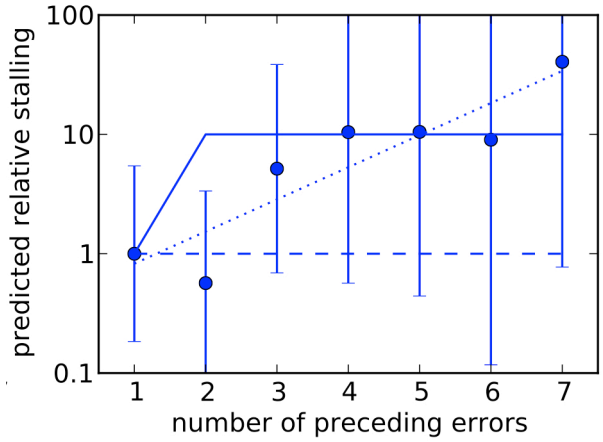
Supporting Figure S26. Predictions for recovery after stalling after *n* successive errors, obtained from RNA folding simulations using 100 random sequences (disallowing GU pairs to more closely approximate the DNA scenario). Error bars denote standard deviations over the sequence ensemble. The *y*-axis is a normalized measure of stalling (0 for no stalling, 1 for the amount of stalling directly after a mismatch).

Sensitivity Analysis

6. Sensitivity to sequence correlations. Simulations were run with randomly chosen rates from the rate tables. Effectively, instead of walking along the sequence and looking at successive triplets, we generated two independent random triplets for extension rate and mutation profile. This destroys correlations within a sequence

(because normally neighboring triplets overlap and the rates for successive polymerization steps are therefore not independent). This alteration did not give significant changes in the results.

7. Sensitivity to extrapolated rate of extension following a region of 2 or more mismatches. Our main assumption is described in #3 above. Note that the main assumption (#3) is a relatively conservative assumption since we know that stalling after a double mismatch is greater than after a single mismatch, and we then assume that double, triple, etc, mismatches behave similarly. However, we also tested two alternative scenarios: 1) No additional stalling beyond that exhibited after a single mismatch (i.e., the most conservative assumption that stalling after a mismatched region is the same regardless of the size of the mismatched region); 2) an exponentially increasing stalling factor, suggested by the RNA folding simulations described in the following paragraph. These three possible assumptions are summarized in Supporting Figure S27.

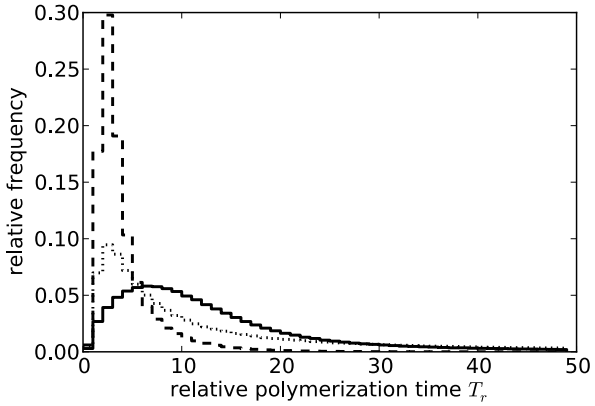


Supporting Figure S27. Predictions for stalling directly after multiple errors following 3 different assumptions. The solid line assumes that stalling factors after more than one successive errors are increased by a constant factor 10 (see #3 above); the dotted line shows an exponential fit to the folding results; the dashed line assumes no additional stalling. RNA folding simulations used 100 random sequences, disallowing GU pairs; error bars denote standard deviations over the sequence ensemble. Stalling factors after more than one mismatch might be dominated by non-templated reactions rather than base-pairing probability.

To estimate the reduction in the incorporation rate after mismatches, we assume that the reduced rate $k' = pk_0 + (1 - p)k_{nt}$, where k_0 is the regular incorporation rate, k_{nt} is the rate of the untemplated reaction and p is the

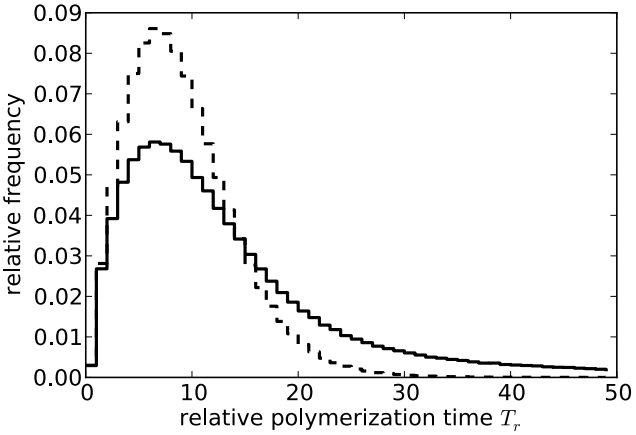
probability that the primer terminus forms a base pair with the correct base in the template. This probability can be calculated using RNA folding simulations, where we take 100 random sequences with complementary primers, and introduce *n* successive errors at a distance *x* to the primer terminus, and record *p* using Vienna RNA secondary structure folding, disallowing GU pairs to mimic the DNA situation. Generally, the results depend strongly on the sequence; we use averaged values. Also, we do not consider multiple neighboring error clusters. Finally, since we do not know all rates (and preferences) of the untemplated reactions, these results were used to increase or decrease the stalling factors measured for incorporation immediately after one single mismatch. Note that the overall values of the stalling factors, especially after multiple errors, do have a quantitative effect on the results: results from the RNA folding simulations suggest an exponential increase with the size of the preceding error cluster, but much depends on the efficiency of non-templated reactions.

The results obtained from the different assumptions are given in Supporting Figure S28. The results under different assumptions are qualitatively similar.



Supporting Figure S28. Simulation results for 1000 runs each of 1000 random sequences of length L = 50. This figure shows the relative polymerization time, with three different assumptions for the amount of stalling after more than one mismatch. The solid line assumes that stalling factors after more than one successive error are increased by a constant factor 10 (see #3 above); the dotted line shows an exponential fit to the folding results; the dashed line assumes no additional stalling. Solid line is also shown in Figure 8C in the main text.

8. Sensitivity to reduced stalling factors. Stalling factors were effectively decreased by assuming that the non-templated reaction rate was at least 10% of the total extension rate. There is a quantitative difference but the qualitative behavior is similar (Supporting Figure S29).

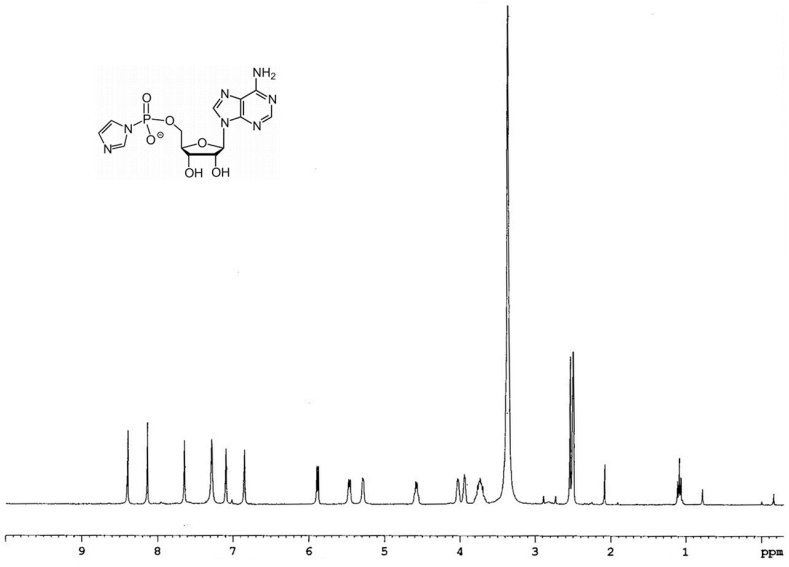


Supporting Figure S29. Histogram of polymerization times with reduced stalling. Solid line: simulations with the usual experimental parameters described in the main text (also shown in the main text, Figure 8C). Dashed line: simulations with reduced stalling.

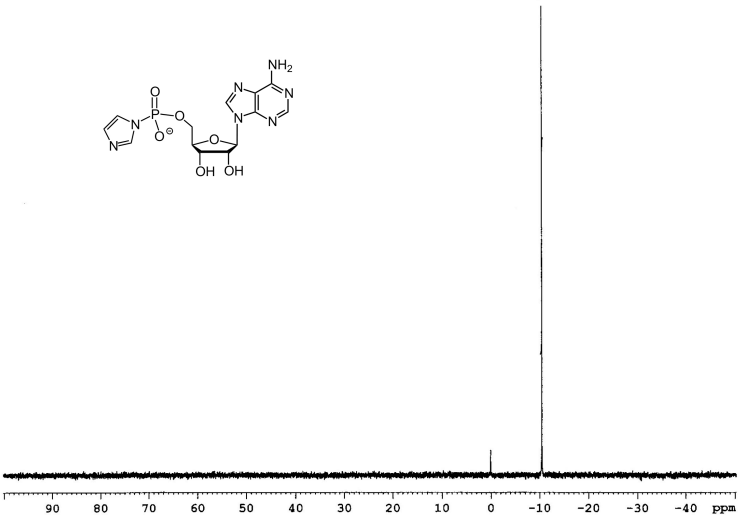
Note: NMR data for Supporting Figures S30-S33 are courtesy of GLSynthesis, Inc.; ESI-MS spectra were collected by the Small Molecule Mass Spectrometry Facility at Harvard University using an Agilent ESI-TOF 6220 high resolution mass spectrometer.

Supporting Figure S30. ^{1}H-NMR (300 MHz, **A**) and ^{31}P-NMR (121 MHz, **B**) spectra (DMSO) of AMP-Im (**8a**), with negative mode ESI mass spectrum (**C**) on following page. Molecular formula for the $(M-H)^-$ ion: $C_{13}H_{15}N_7O_6P$.

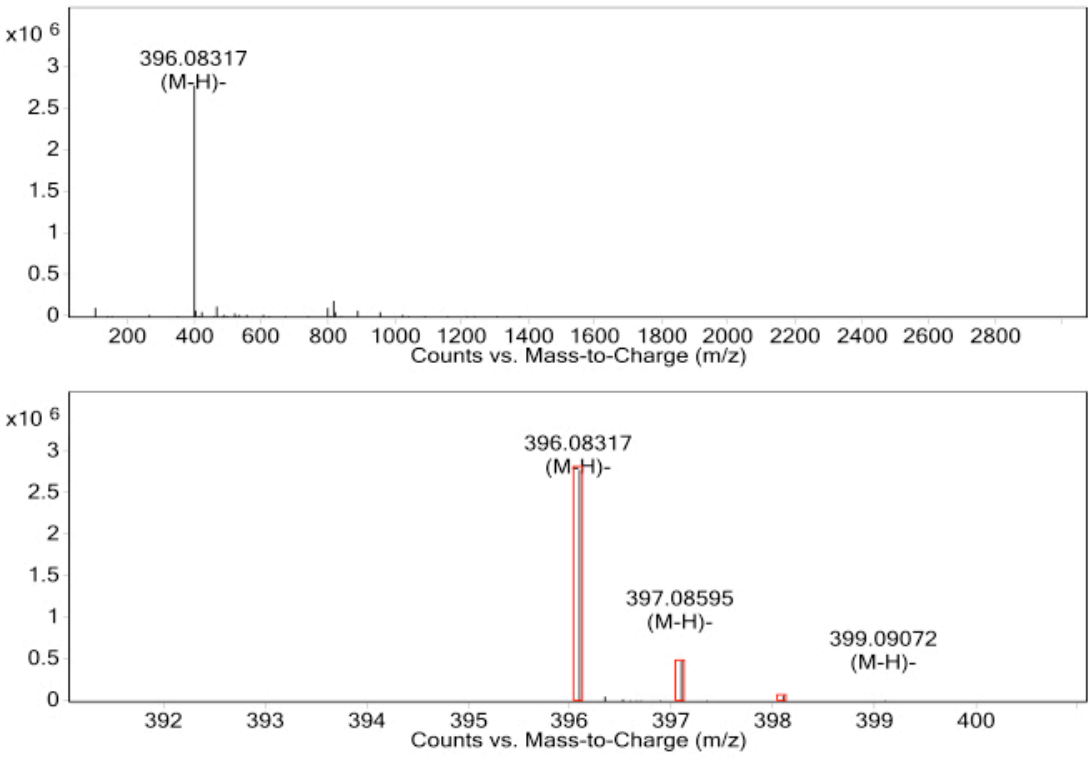
A



B



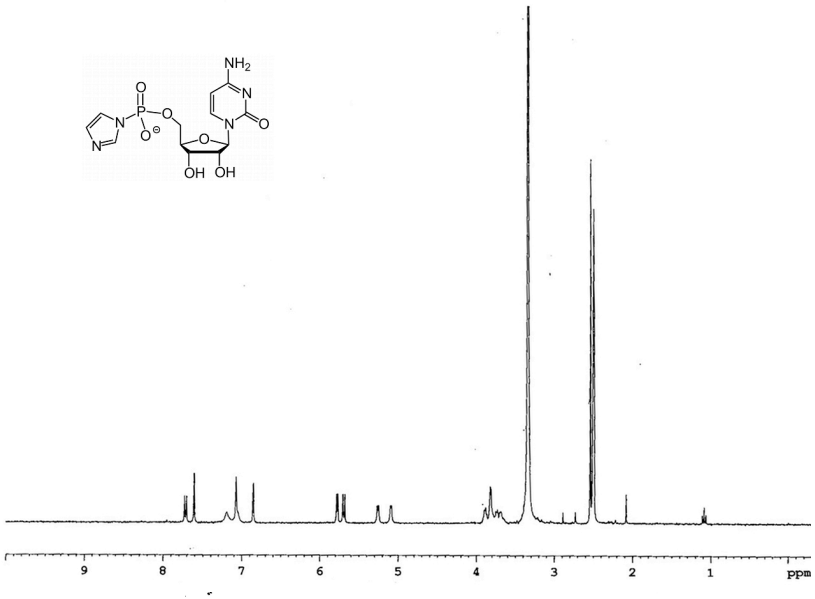
C



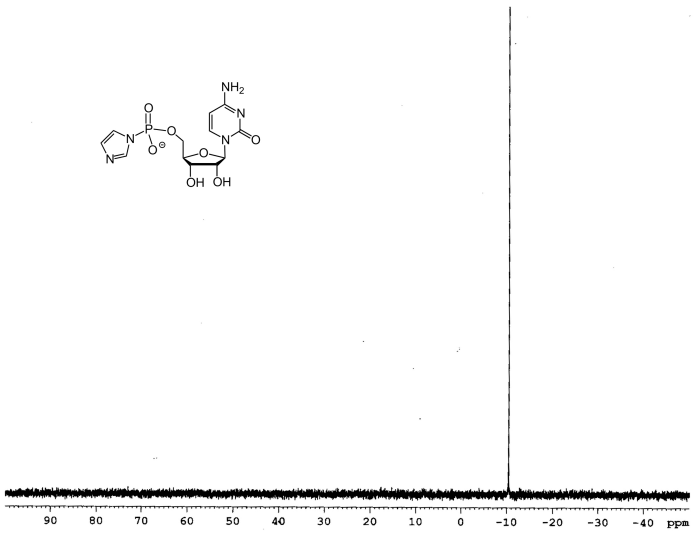
Observed m/z	**Calculated m/z**
396.08317	396.08269

Supporting Figure S31. ^{1}H-NMR (300 MHz, **A**) and ^{31}P-NMR (121 MHz, **B**) spectra (DMSO) of CMP-Im (**8c**), with negative mode ESI mass spectrum (**C**) on following page. Molecular formula for the $(M-H)^-$ ion: $C_{12}H_{15}N_5O_7P$.

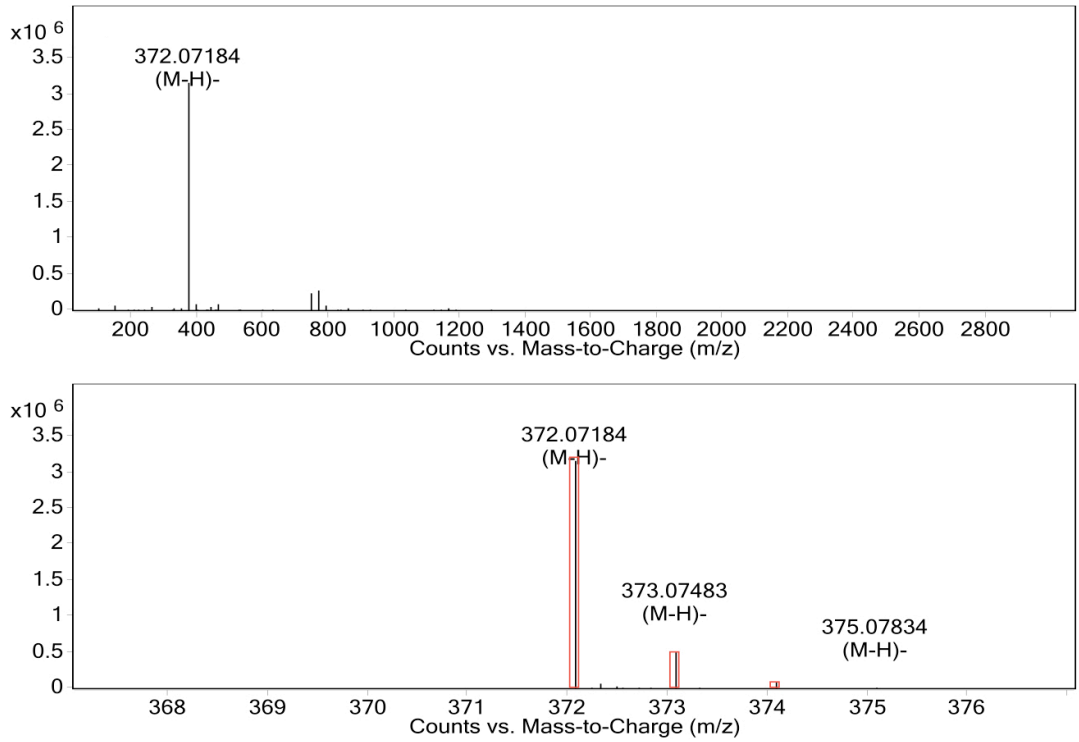
A



B



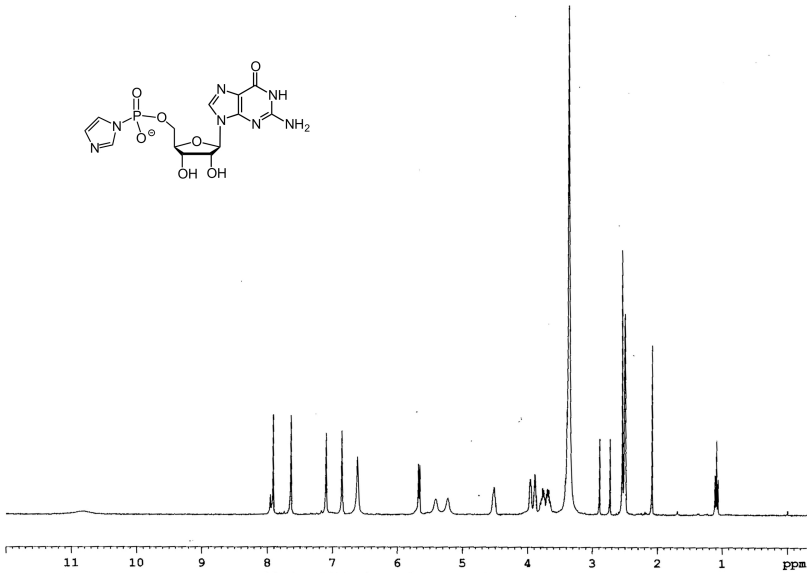
C



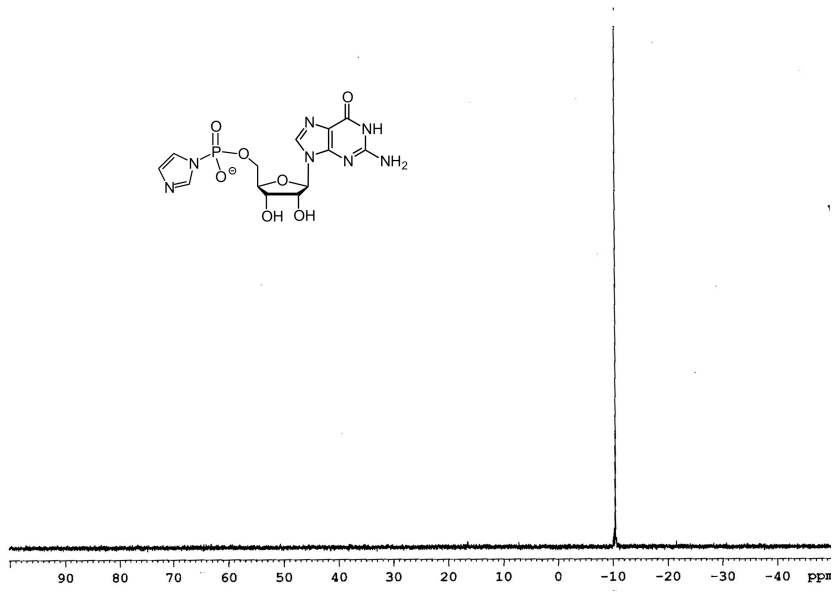
Observed m/z	**Calculated m/z**
372.07184	372.07146

Supporting Figure S32. ^{1}H-NMR (300 MHz, **A**) and ^{31}P-NMR (121 MHz, **B**) spectra (DMSO) of GMP-Im (**8g**), with negative mode ESI mass spectrum (**C**) on following page. Molecular formula for the $(M-H)^-$ ion: $C_{13}H_{15}N_7O_7P$.

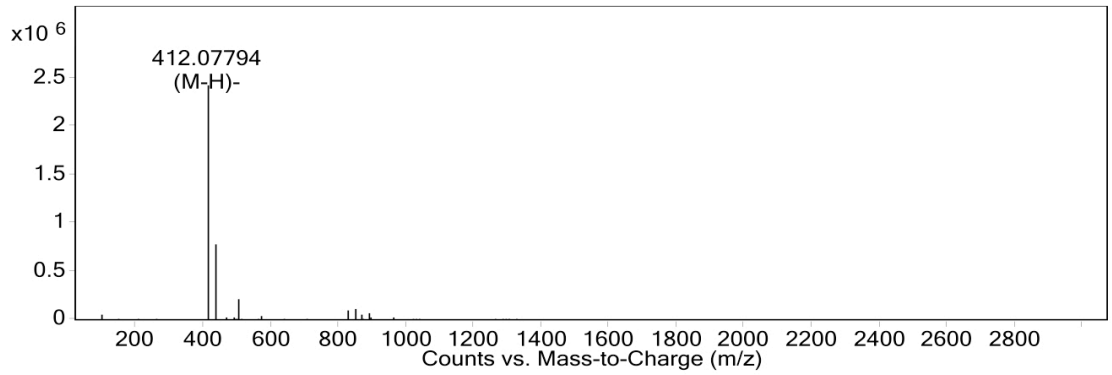
A

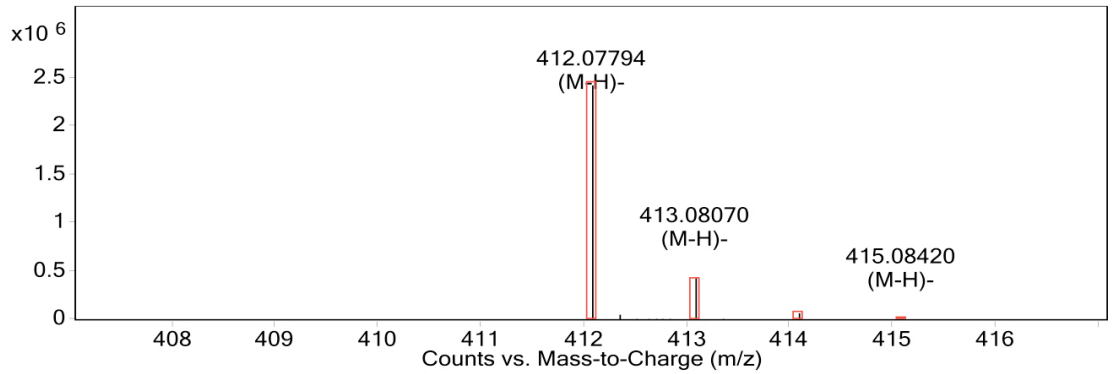


B



C

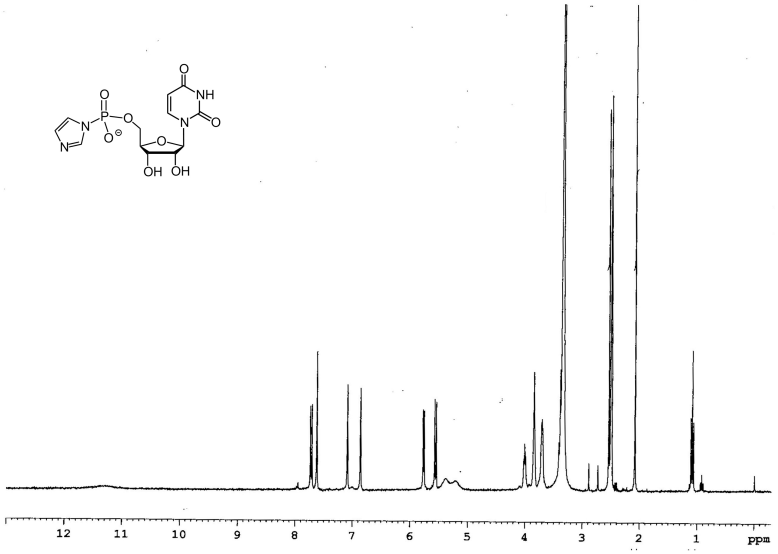




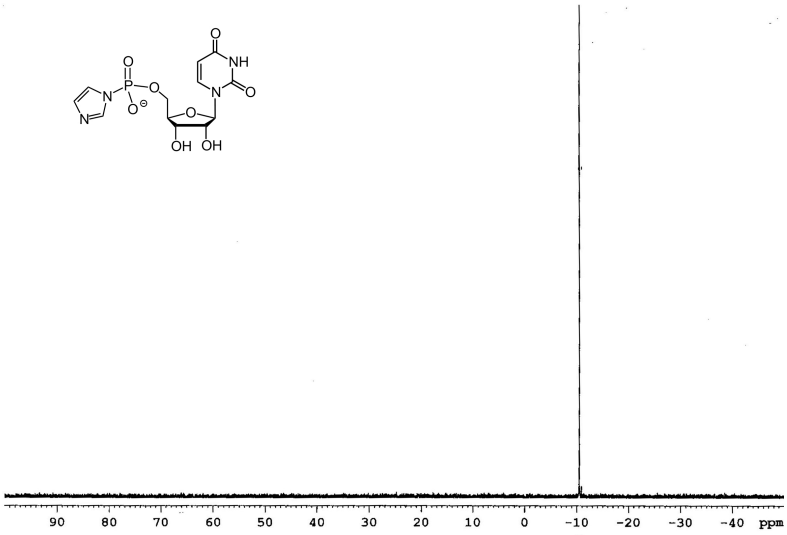
Observed m/z	**Calculated m/z**
412.07794	412.07761

Supporting Figure S33. ^{1}H-NMR (300 MHz, **A**) and ^{31}P-NMR (121 MHz, **B**) spectra (DMSO) of UMP-Im (**8u**), with negative mode ESI mass spectrum (**C**) on following page. Molecular formula for the $(M-H)^-$ ion: $C_{12}H_{14}N_4O_8P$.

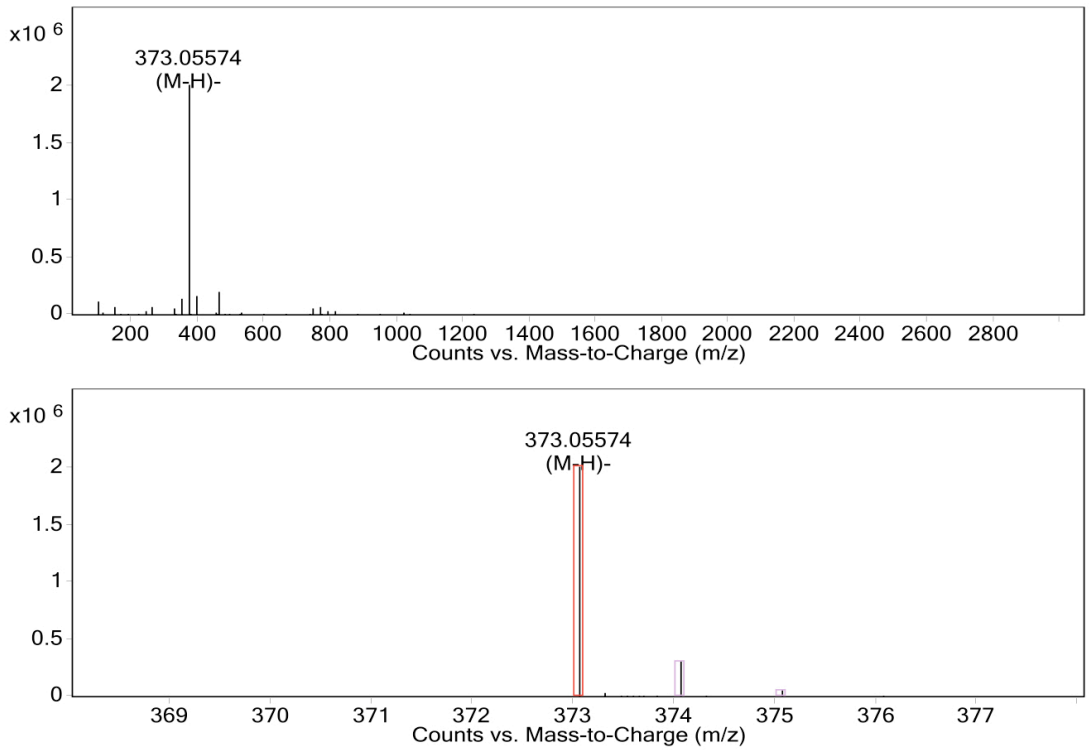
A



B



C

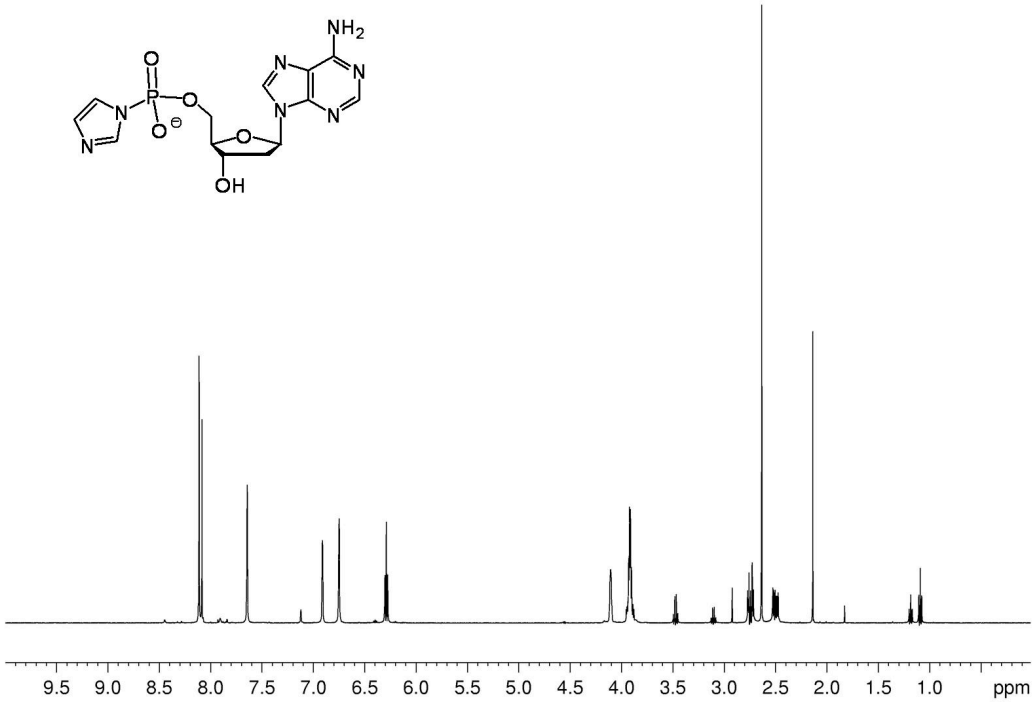


Observed m/z	**Calculated m/z**
373.05574	373.05547

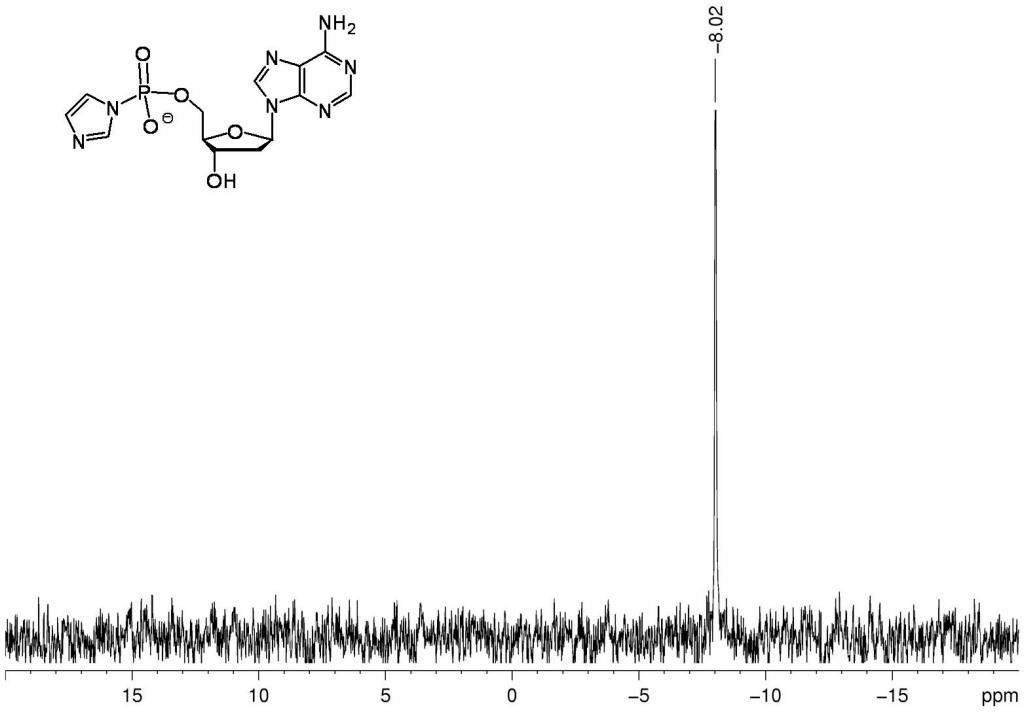
Note: Data for Supporting Figures S34-S41 were obtained using instrumentation described in the main text under ***Materials and methods for assays monitored by mass spectrometry (MS)****.*
High-resolution ESI-mass spectra were acquired on a Bruker micrOTOF-Q spectrometer, using Bruker Daltonics Data Analysis 3.4 software.

Figure S34. Spectra of the imidazolide of dAMP (dAMP-Im, **4a**). (**A**) ^{1}H-NMR spectrum (500 MHz, D_2O), and (**B**) ^{31}P-NMR spectrum (203 MHz, D_2O). The figure is continued on the next page.

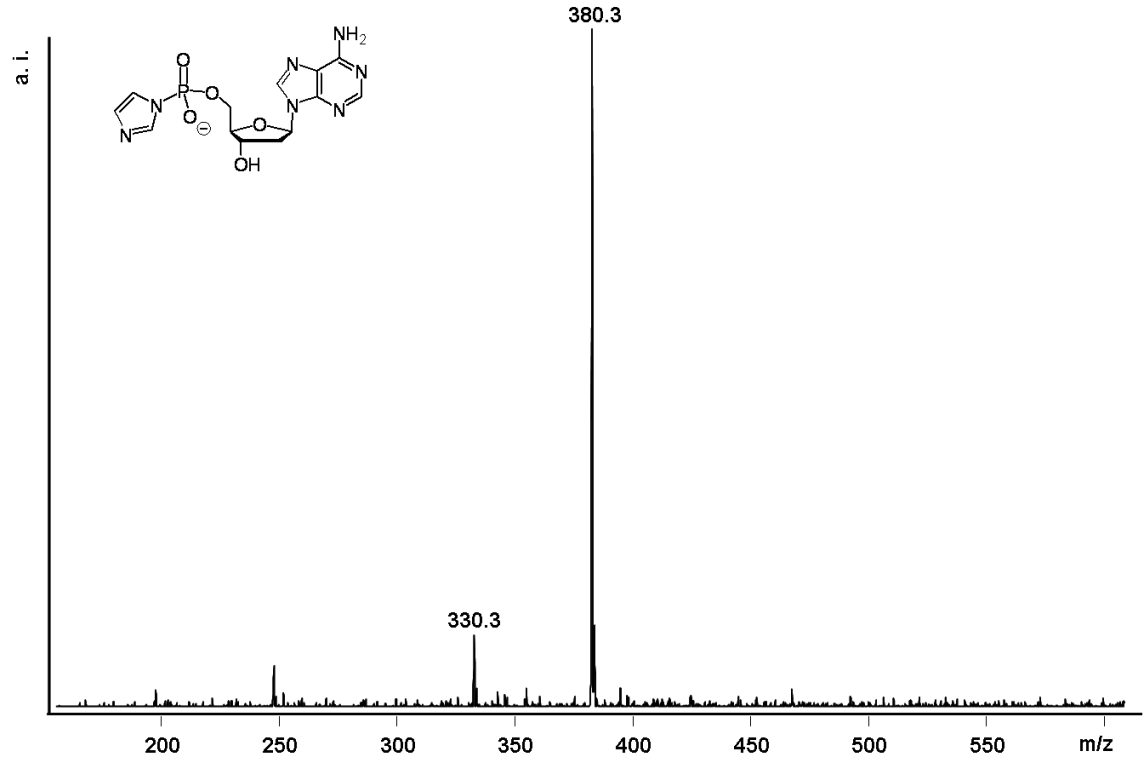
A



B



C



D

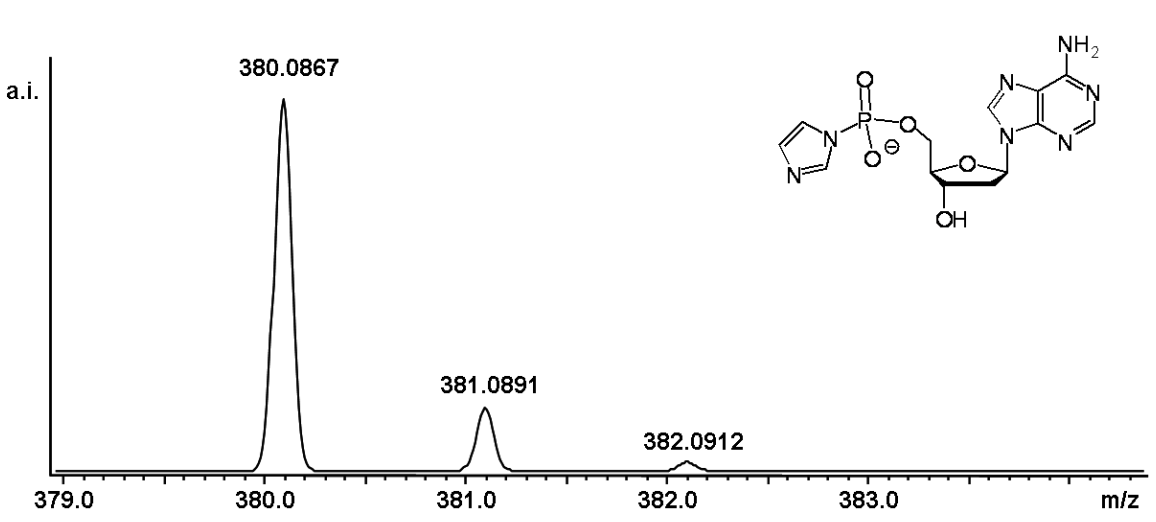
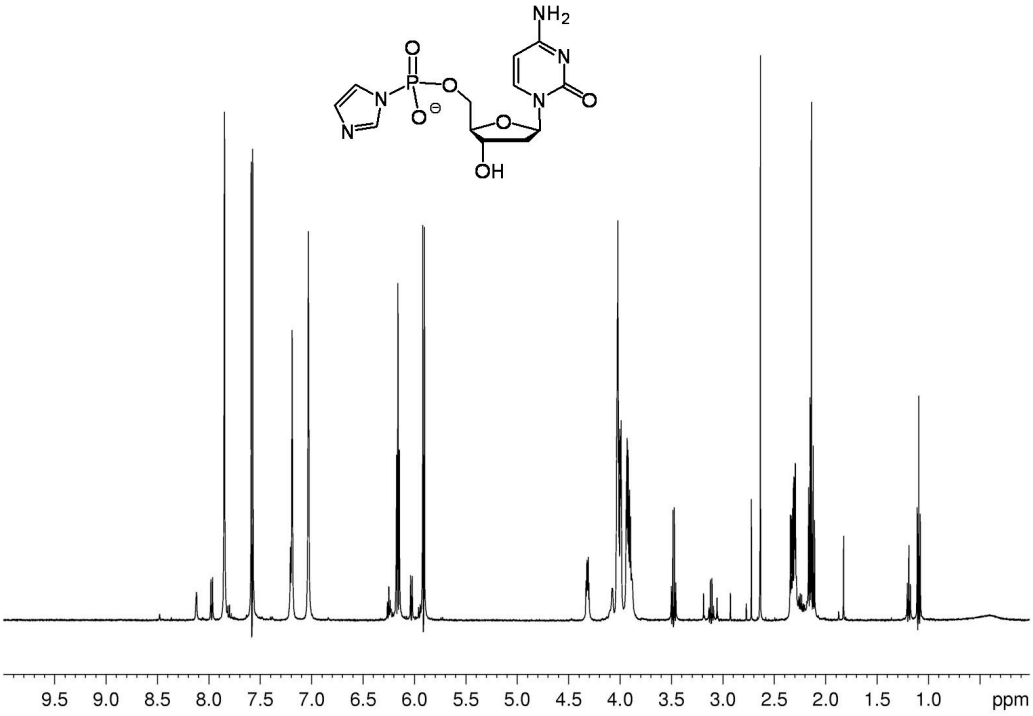


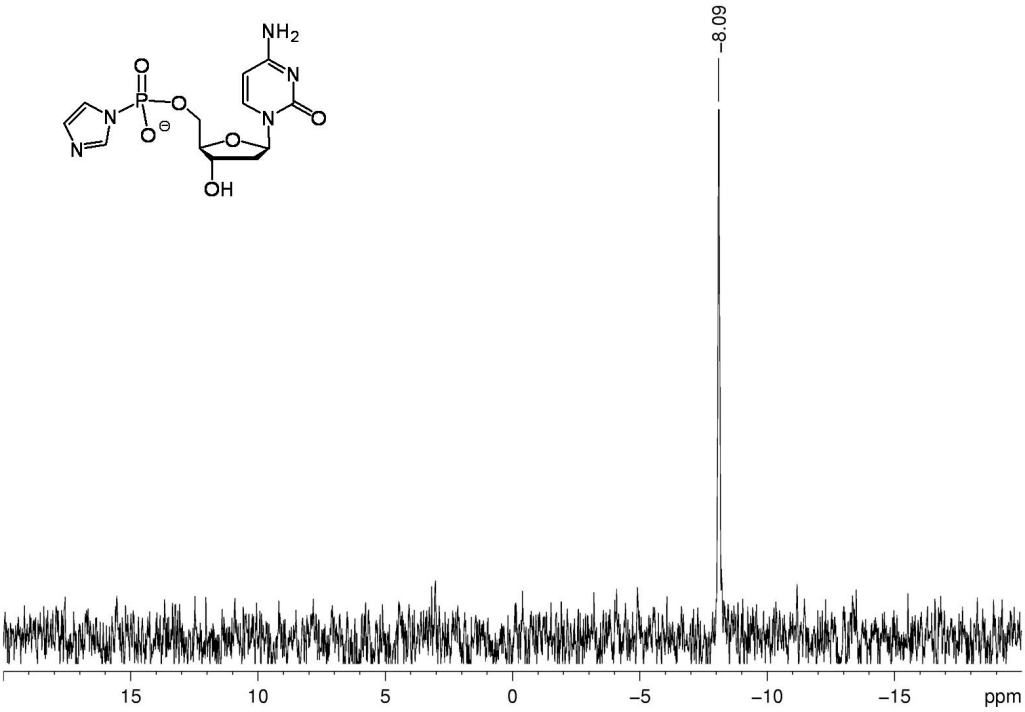
Figure S34, continued. (**C**) Low-resolution ESI-MS, and (**D**) high-resolution ESI-MS of dAMP-Im (**4a**); mass calculated for monoisotopic peak of $C_{13}H_{15}N_7O_5P^-$ (negative mode): 380.0878 Da.

Figure S35. Spectra of the imidazolide of dCMP (dCMP-Im, **4c**). (**A**) ^{1}H-NMR spectrum (500 MHz, D_2O), and (**B**) ^{31}P-NMR spectrum (203 MHz, D_2O). The figure is continued on the next page.

A



B



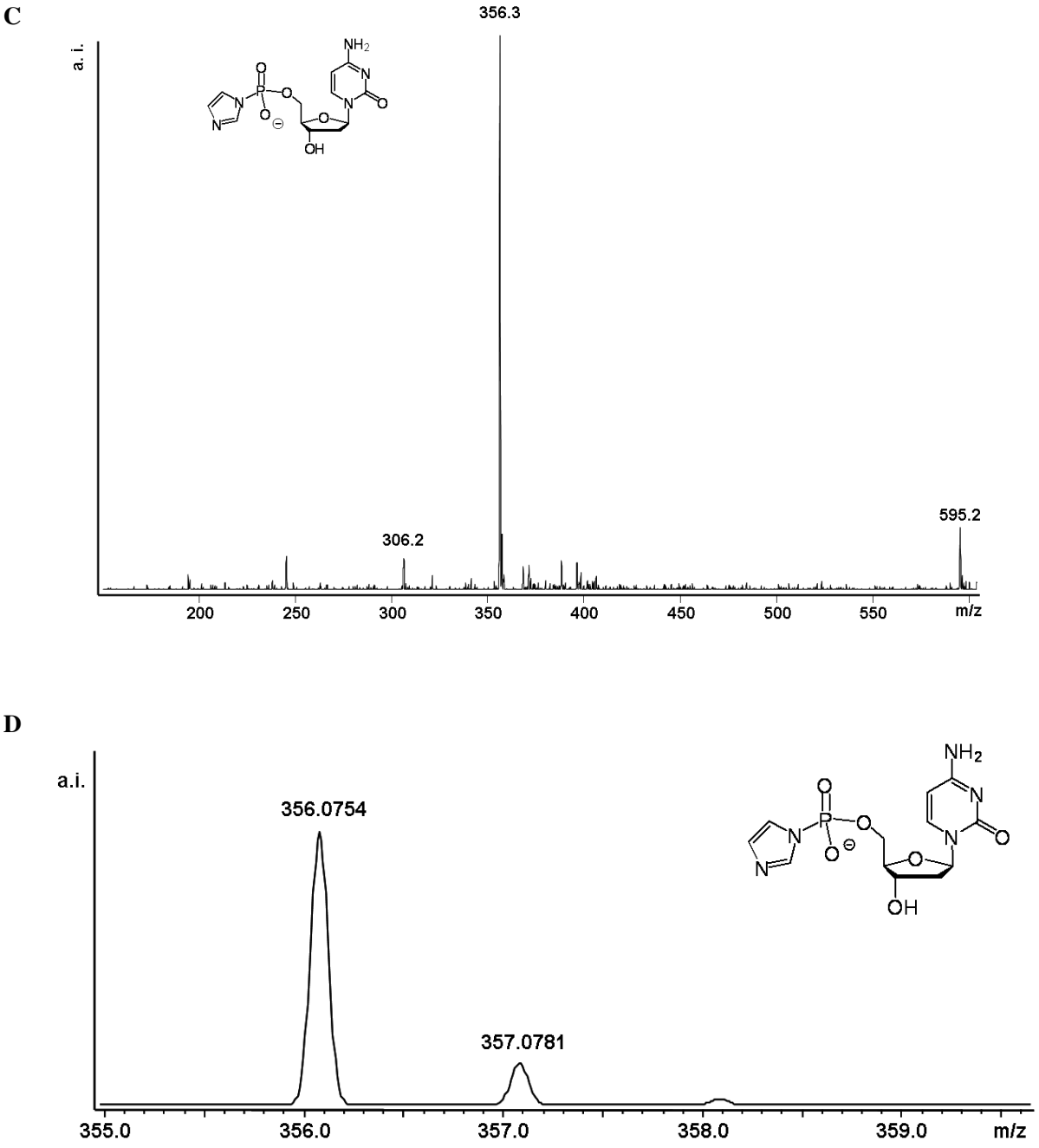
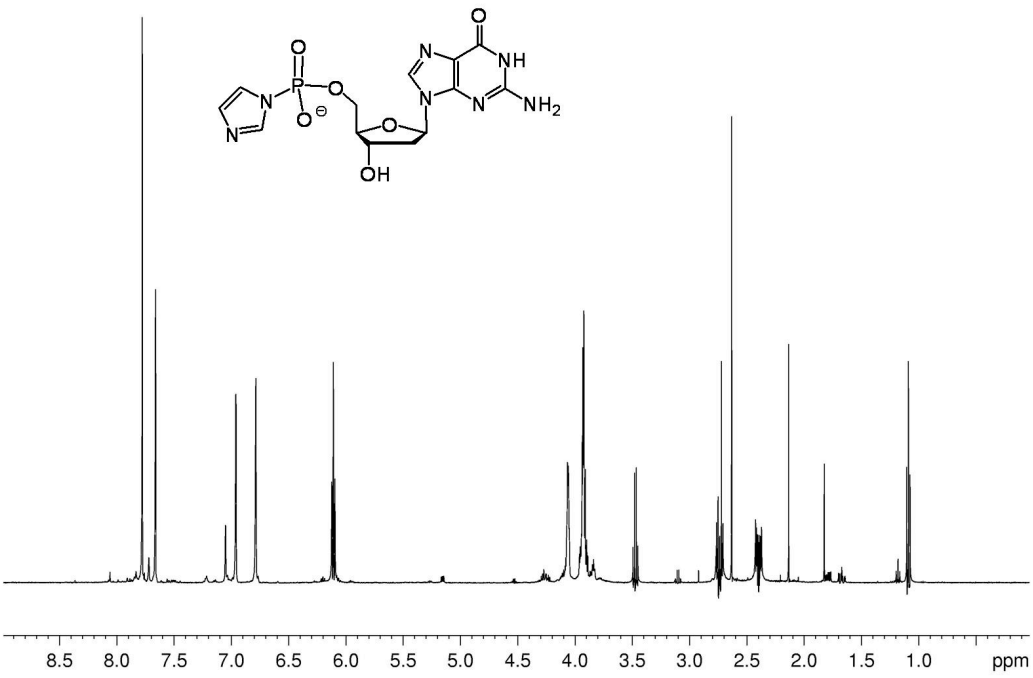


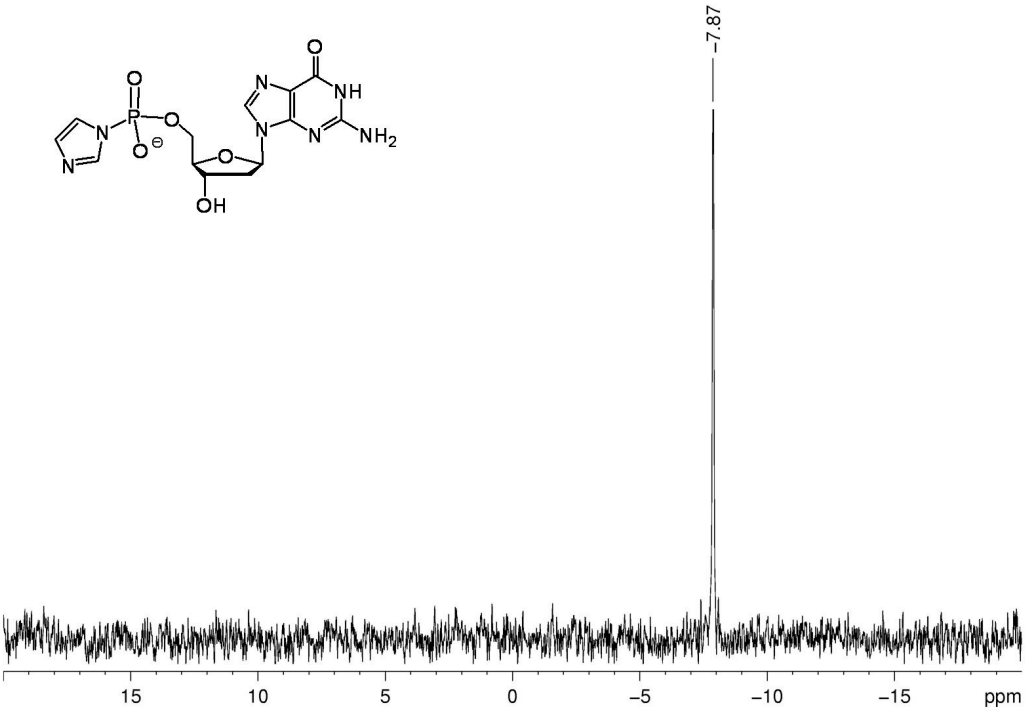
Figure S35, continued. (**C**) Low-resolution ESI-MS, and (**D**) high-resolution ESI-MS of dCMP-Im (**4c**); mass calculated for monoisotopic peak of $C_{12}H_{15}N_5O_6P^-$ (negative mode): 356.0765 Da.

Figure S36. Spectra of the imidazolide of dAMP (dGMP-Im, **4g**). (**A**) ^{1}H-NMR spectrum (500 MHz, D_2O), and (**B**) ^{31}P-NMR spectrum (203 MHz, D_2O). The figure is continued on the next page.

A



B



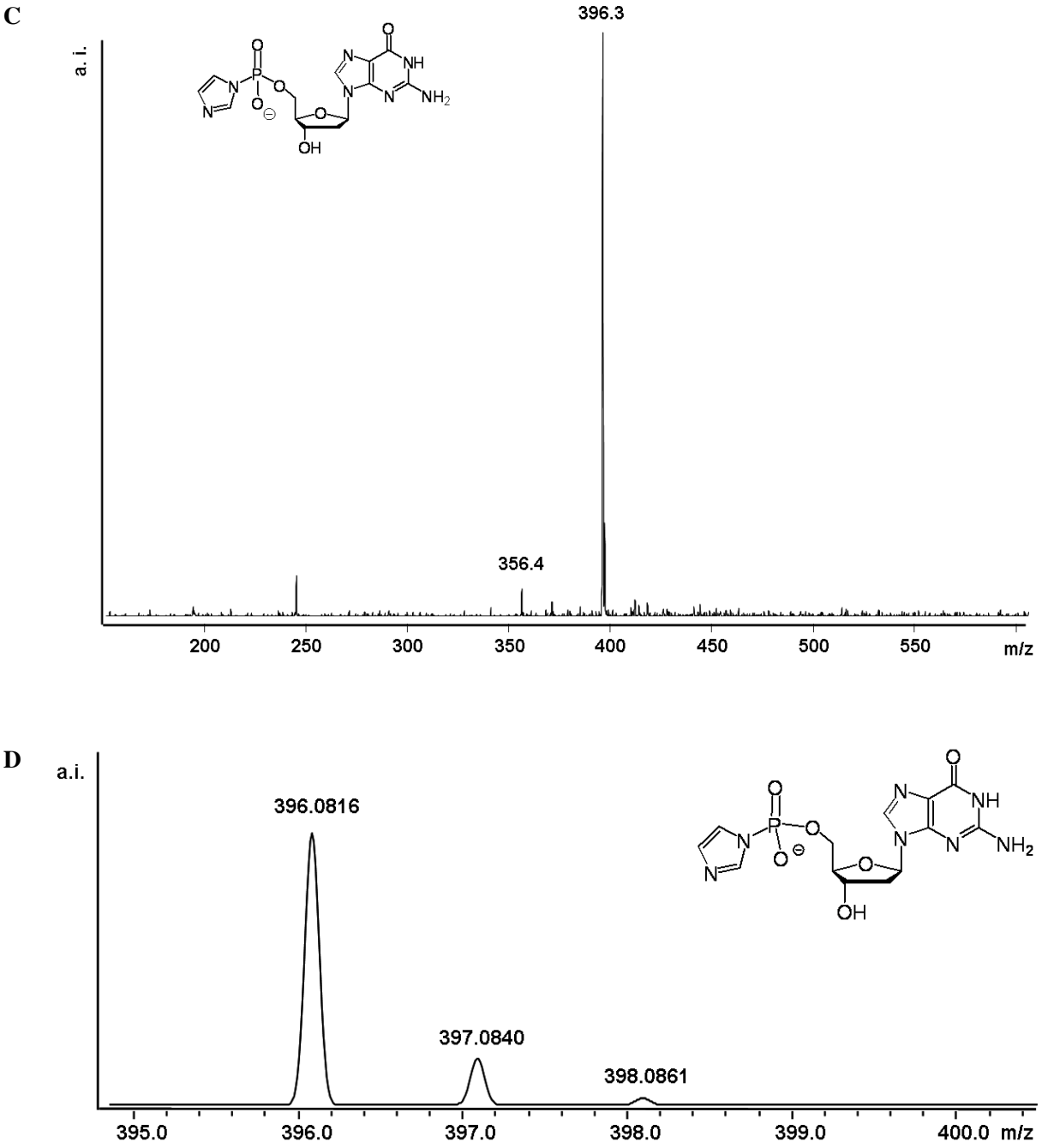
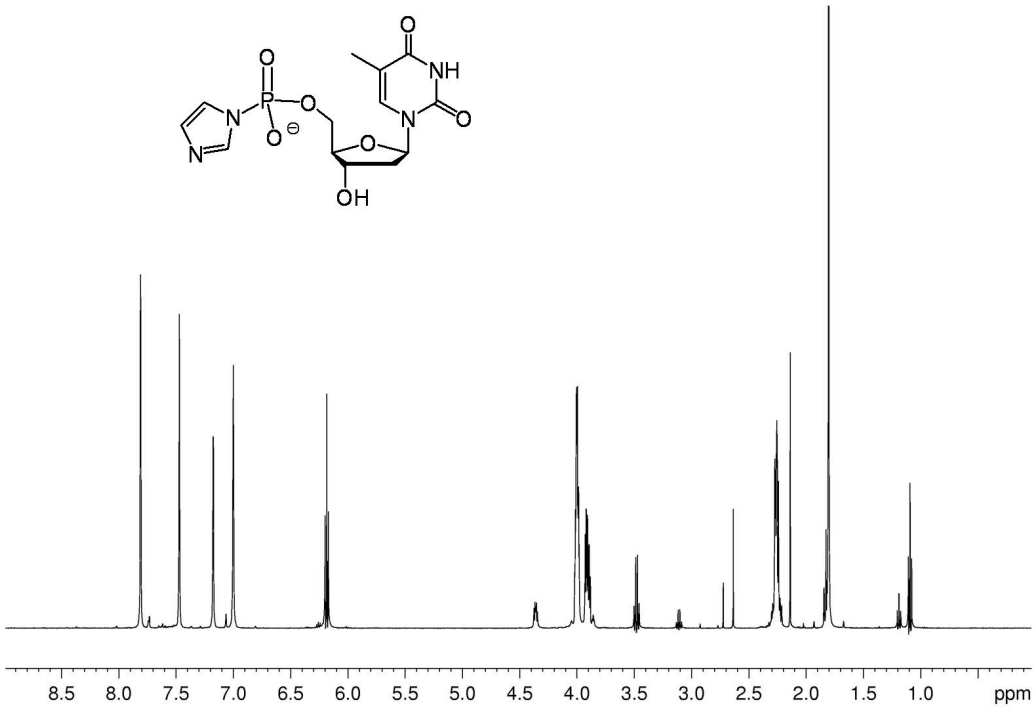


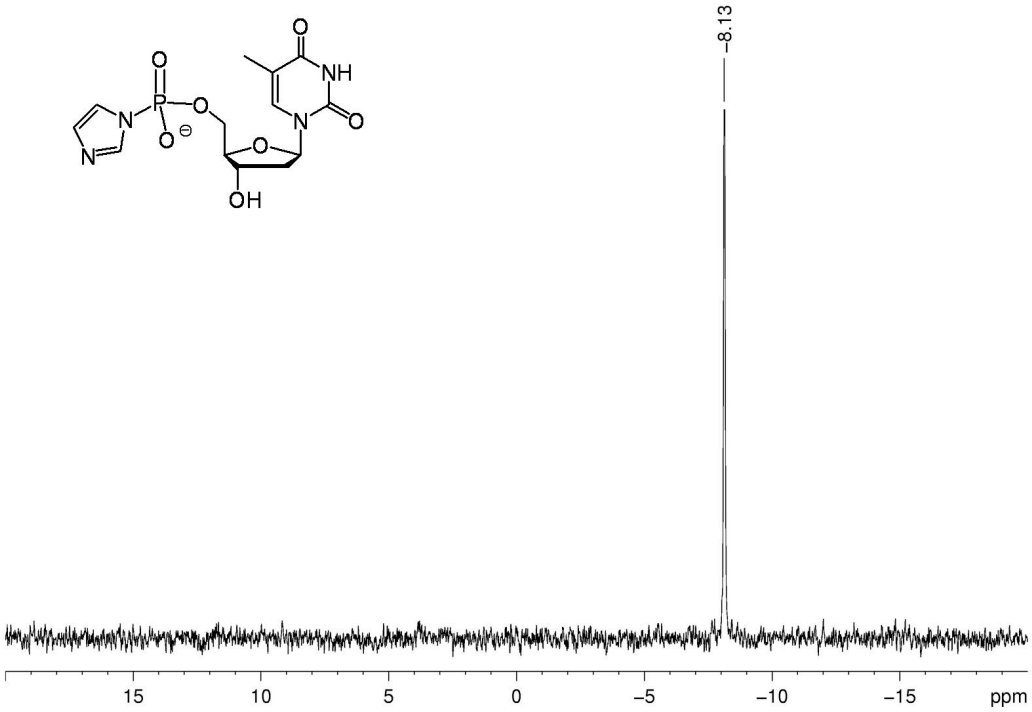
Figure S36, continued. (**C**) Low-resolution ESI-MS, and (**D**) high-resolution ESI-MS of dGMP-Im (**4g**); mass calculated for monoisotopic peak of $C_{13}H_{15}N_7O_6P^-$ (negative mode): 396.0827 Da.

Figure S37. Spectra of the imidazolide of TMP (TMP-Im, **4t**). (**A**) ^{1}H-NMR spectrum (500 MHz, D_2O), and (**B**) ^{31}P-NMR spectrum (203 MHz, D_2O). The figure is continued on the next page.

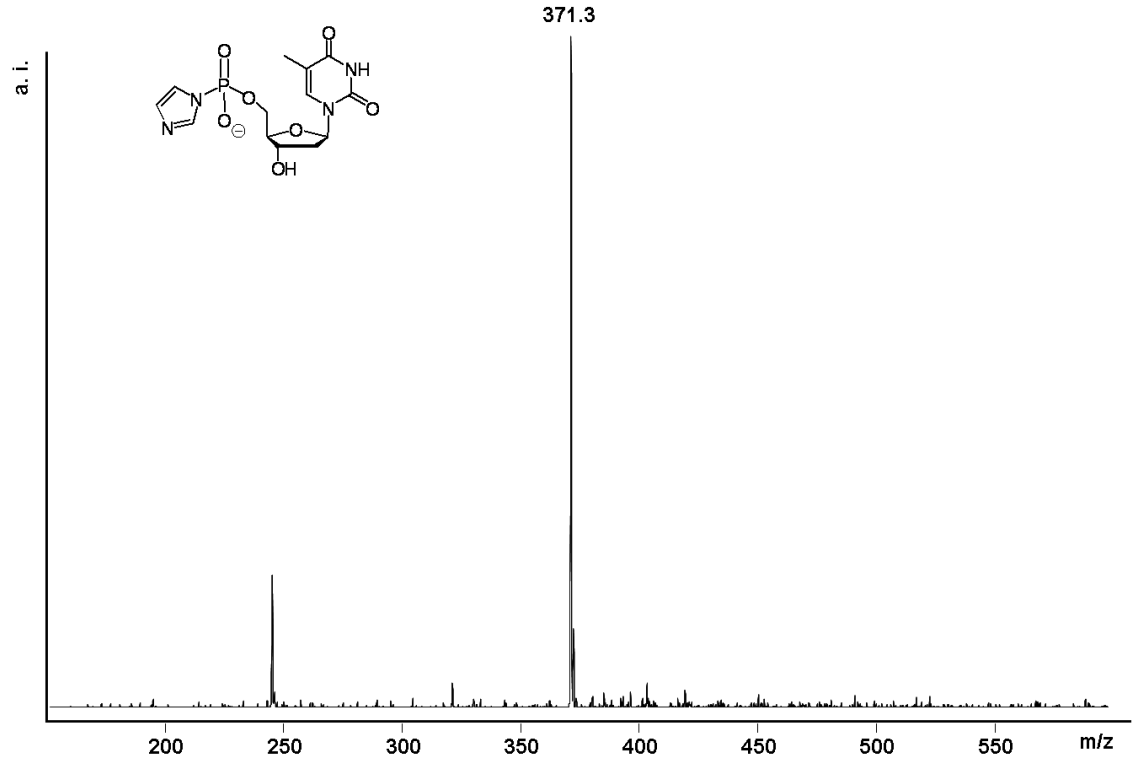
A



B



C



D

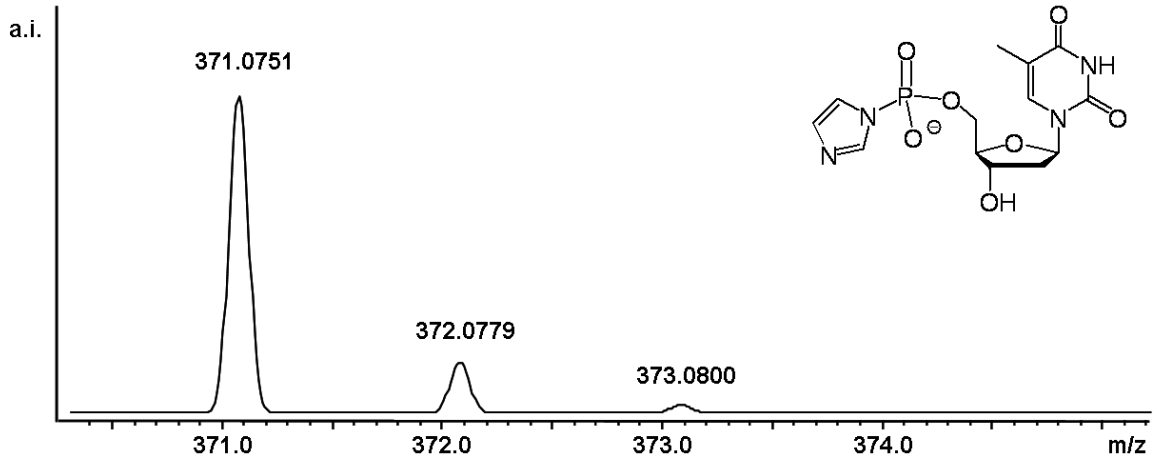


Figure S37, continued. (**C**) Low-resolution ESI-MS, and (**D**) high-resolution ESI-MS of TMP-Im (**4t**); mass calculated for monoisotopic peak of $C_{13}H_{16}N_4O_7P^-$ (negative mode): 371.0762 Da.

Figure S38. (**A**) Low-resolution ESI-MS, and (**B**) high-resolution ESI-MS of dAMP-OAt (**3a**); mass calculated for monoisotopic peak of $C_{15}H_{15}N_9O_6P^-$ (negative mode): 448.0888 Da.

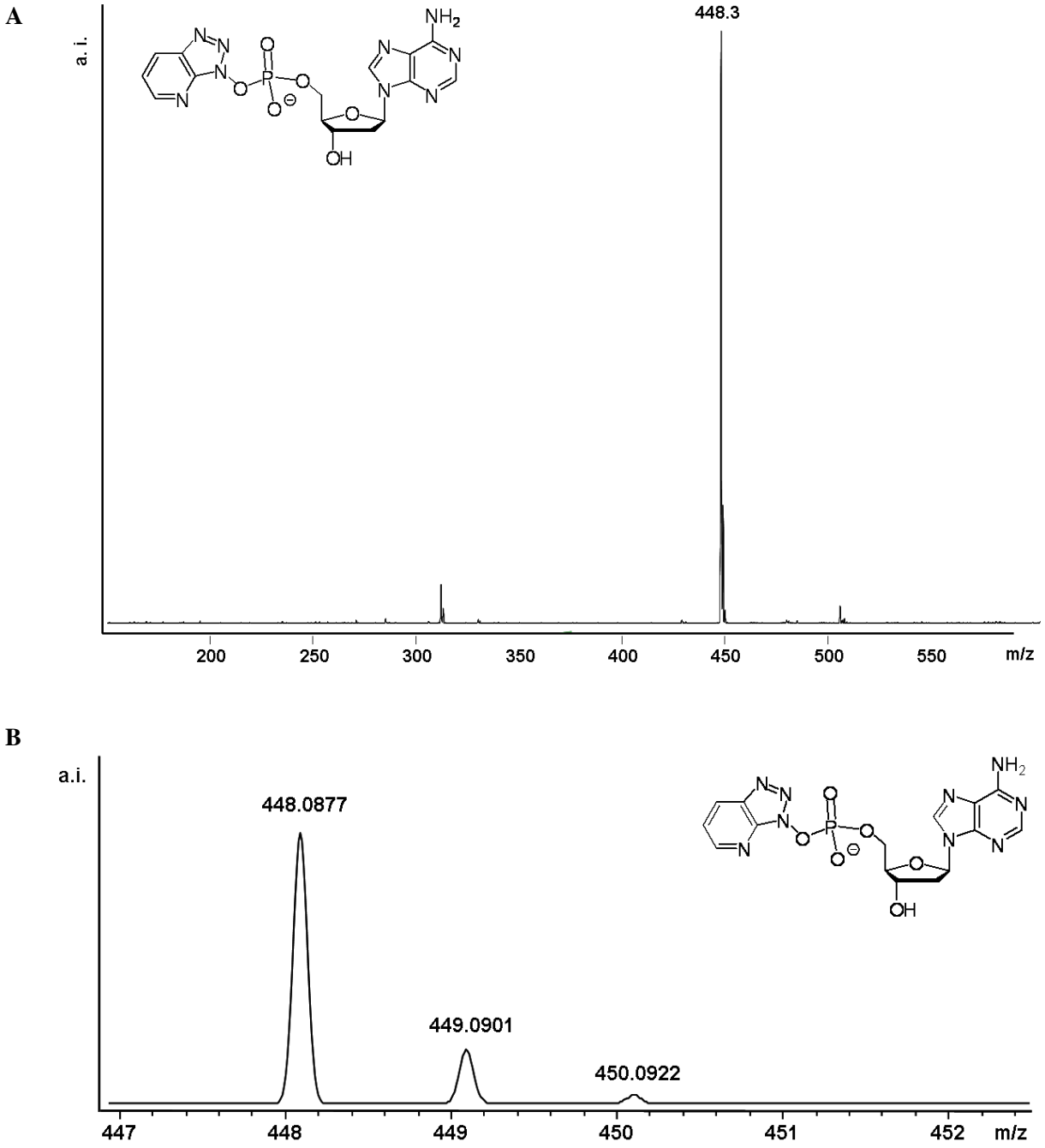


Figure S39. (**A**) Low-resolution ESI-MS, and (**B**) high-resolution ESI-MS of dCMP-OAt (**3c**); mass calculated for monoisotopic peak of $C_{14}H_{15}N_7O_7P^-$ (negative mode): 424.0776 Da.

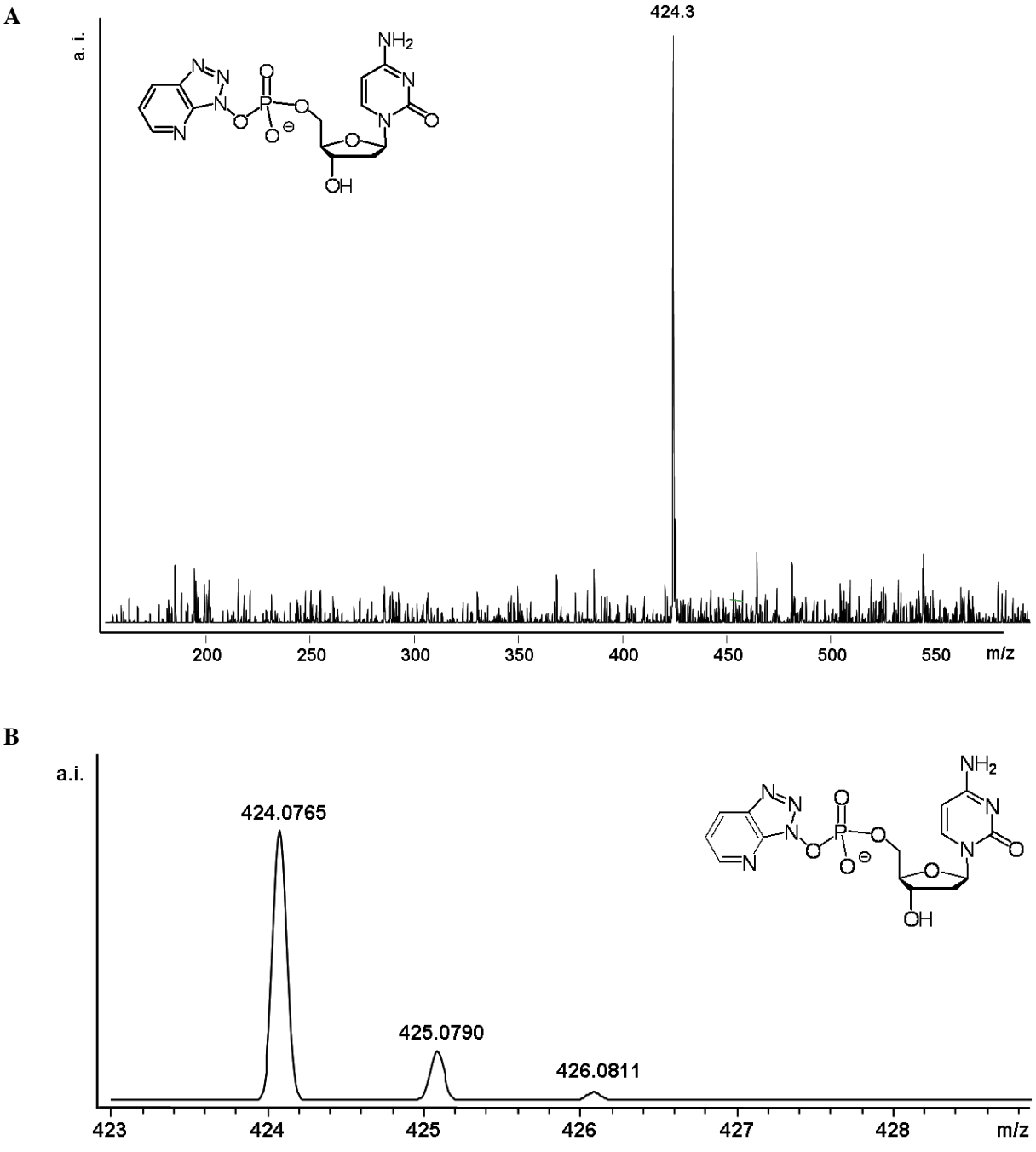


Figure S40. (**A**) Low-resolution ESI-MS, and (**B**) high-resolution ESI-MS of dGMP-OAt (**3g**); mass calculated for monoisotopic peak of $C_{15}H_{15}N_9O_7P^-$ (negative mode): 464.0838 Da.

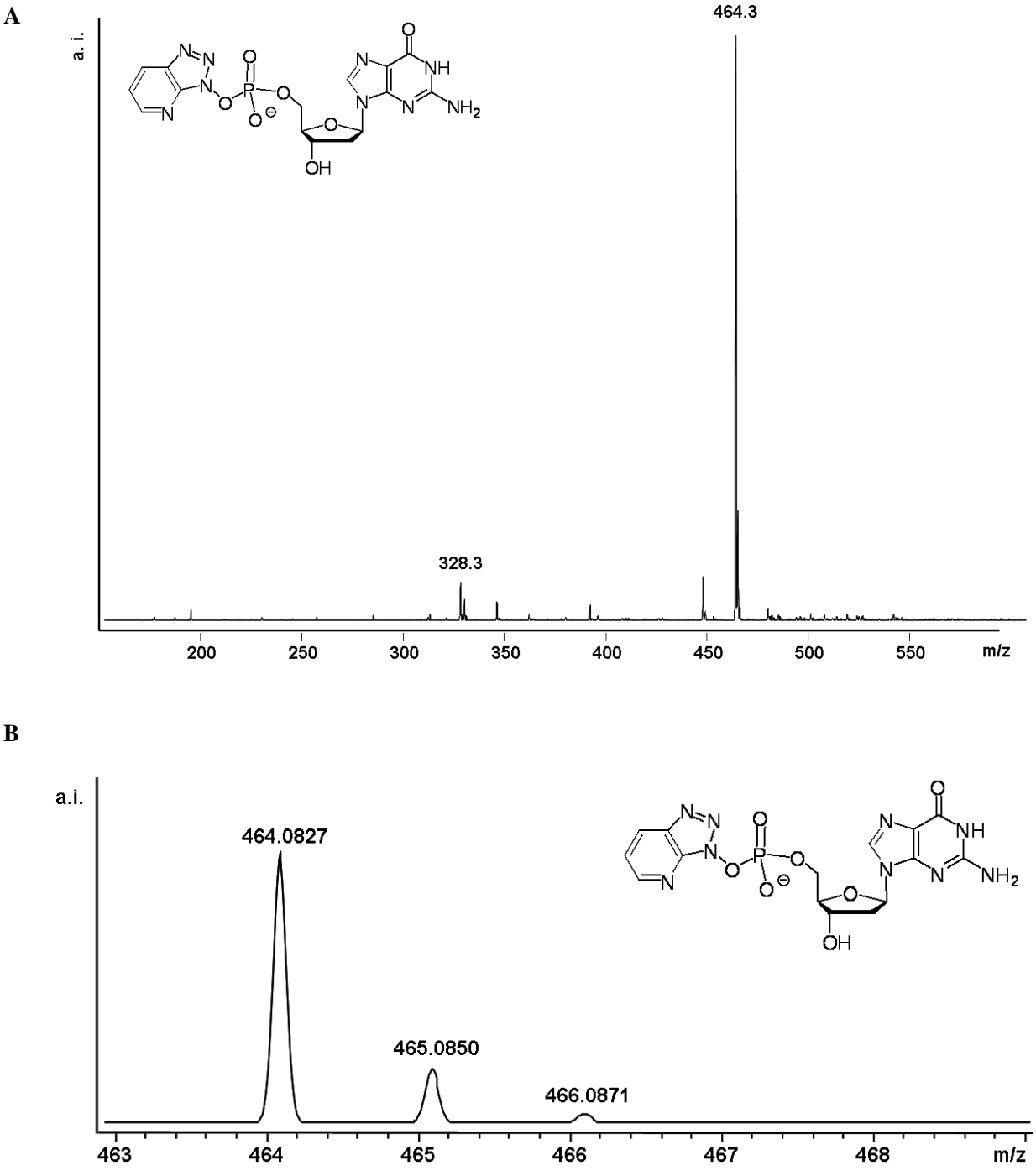


Figure S41. (**A**) Low-resolution ESI-MS, and (**B**) high-resolution ESI-MS of TMP-OAt (**3t**); mass calculated for monoisotopic peak of $C_{15}H_{16}N_6O_8P^-$ (negative mode): 439.0773 Da.