Supplementary Information

Coupled catabolism and anabolism in autocatalytic RNA sets

Simon Arsène[1,‡], Sandeep Ameta[1,‡], Niles Lehman[2], Andrew D. Griffiths[1,], Philippe Nghe[1,*]*

‡These authors contributed equally to this work.

[1]Laboratoire de Biochimie, CNRS UMR8231, Chimie Biologie Innovation, ESPCI Paris, 10 Rue Vauquelin, 75005, Paris, France.

[2]Department of Chemistry, Portland State University, P.O. Box 751, Portland, OR, USA 97207.

**E-mail: andrew.griffiths@espci.fr, philippe.nghe@espci.fr*

SUPPLEMENTARY TEXT

Kinetic modelling

We derive here a kinetic model for the multi-step reaction pathway presented in Figure 3 using a similar approach as used earlier (22). Our assumptions are the following: (i) covalent bonds formation is reversible and (ii) catalysis can occur either by a covalent ribozyme or by a non-covalent complex. In order to construct our model, similar to the formation of WXY:Z non-covalent complex, we assume that WXY can bind to Z-mod, and WXY-mod can bind to Z to form the supramolecular complex, WXY:Z-mod and WXY-mod:Z respectively. However we assume that the potential complex WXY-mod:Z-mod can only form in negligible amounts because of steric considerations. Contrary to what was done earlier (22), we chose not to fix arbitrarily the rates of association and dissociation of the non-covalent complexes as they are not expected to be equally stable. Formation of the covalent bond between WXY and Z (at the 'YZ' junction) can be catalyzed by either fully covalent WXYZ or by a non-covalent complex (20). Likewise, we assumed that the transfer of the modification part (mod) to WXY from Z-mod can be catalyzed by either fully covalent WXYZ or by a non-covalent complex. Cleavage of mod from WXY-mod can happen by attack of the 3' OH (of the 3' terminal 'G' of Z) from either covalent WXYZ or non-covalent complexes (WXY:Z and WXY-mod:Z). Finally as we do not observe any significant band of higher molecular weight than WXYZ (Supplementary Figure S4, S5) we neglect the possible formation of WXYZ-mod. Slow accumulation of WXYZ (Figure 2B) shows that WXYZ-mod is not formed quickly as an unstable intermediate. Indeed, if this was the case, the yield of WXYZ would increase more quickly than we observe because the cleavage of mod from WXYZ-mod would give WXYZ as a product. These assumptions allow us to limit the number of rate constants to fit. Models with more parameters can also be constructed with, for example, an independent description of the substrate binding and substrate release steps as proposed earlier (22). Our modeling approach resulted in 14 chemical reactions ('a' to 'n', Supplementary Information, Table S2). Reactions 'a-c' account for the formation of non-covalent complexes. Reactions 'd-g' account for the covalent bond formation between WXY and Z (at the 'YZ' junction) while reactions 'h-k' account for transfer of mod from Z-mod to WXY. Finally reactions 'l-n' account for cleavage of mod from WXY-mod.

The model fits well the experimental data with a root mean square error (RMSE) value of 1.2% (Supplementary Figure S9) and thus provide solid support for the multi-step reaction pathway proposed in Figure 3. The fit is slightly less good for WXYZ formation in reaction c2 and c3 for which we have less sensitivity on the measurement because of low intensity of the bands.

We tried to reduce the number of iterable rate constants based on several reasonable hypothesis: 1) all catalysts are assumed to be equally efficient and all complexes are assumed to be equally stable (8 parameters); 2) only WXY:Z-mod and WXY-mod:Z are assumed to be equally stable and all catalysts are assumed to be equally efficient (10 parameters); 3) only WXY:Z-mod and WXY-mod:Z are assumed to be equally stable and all catalysts are presumed

to be equally efficient expect for WXYZ as a catalyst for its own formation (12 parameters); 4) only WXY:Z-mod and WXY-mod:Z are assumed to be equally stable and all complexes are assumed to be equally efficient (16 parameters); 5) only WXY:Z-mod and WXY-mod:Z are assumed to be equally stable and equally efficient (22 parameters). For each of these hypotheses we fitted the rates to the corresponding reduced model (Supplementary Figure S10). Results show qualitatively that there is always at least one specie whose time course of formation is not well predicted by reduced models: either the monotony of the curve is not respected or there exist a systematic model deviation below or above what the spread of data points would allow. As comparing the overall RMSE neither conveys these trends nor allows to justify the number of parameters used, we also computed a set of different model selection criteria described by Turner *et al.* (39). These criteria quantify the trade-off between the number of parameters and the fitting quality, thus avoid overfitting. All these criteria are in favor of choosing the final model (Supplementary Table S3).

SUPPLEMENTARY TABLES

Table S1. Oligonucleotides used in the study.

Oligo name	Sequence	Description
Primer 1	ctgcagaattctaatacgactcactatagagccttgcgccgggaaacc acgcaagggatgg	Forward primer to add 'A' IGS
Primer 2	ctgcagaattctaatacgactcactatagtgcctagcgccgggaaacc acgctagggatgg	Forward primer to add 'U' IGS
Primer 3	ctgcagaattctaatacgactcactatagggcctcgcgccgggaaacc acgcgagggatgg	Forward primer to add 'G' IGS
Primer 4	atgtgccttaggtgggtgc	Reverse primer to add 'A' tag
Primer 5	aggtgccttaggtgggtgc	Reverse primer to add 'C' tag
Primer 6	aagtgccttaggtgggtgc	Reverse primer to add 'U' tag
Primer 7	ttttttttttttttttttttggcactatgtgccttaggtgggtgc	Reverse primer to generate modified substrate WXY with 'A' tag
Primer 8	ttttttttttttttttttttggcactaggtgccttaggtgggtgc	Reverse primer to generate modified substrate WXY with 'C' tag
Primer 9	ttttttttttttttttttttggcactaagtgccttaggtgggtgc	Reverse primer to generate modified substrate WXY with 'U' tag
Primer 10	taatacgactcactataggcatcgctatggtgaaggcatag	Forward primer for Z substrate
Primer 11	ccggtttgtgtgactttcgcc	Reverse primer for Z substrate
Primer 12	ttttttttttttttttttttggcactccggtttgtgtgactttcgcc	Reverse primer for generate modified Z substrate
Primer 13	cgtgtgctcttccgatctnnnnnnnnttttttttttttttttttttccggttt gtgtgactttcgcc	Reverse primer for reverse transcription
Primer 14	ttccctacacgacgctcttccgatctgatatagagccttgcgccgggaa acca	Illumina sequencing primer I (forward, GSP)
Primer 15	gtgactggagttcagacgtgtgctcttccg	Illumina sequencing primer I (reverse, RT)
Primer 16	aatgatacggcgaccaccgagatctacactctttccctacacgacgctc ttcc	Illumina sequencing primer II (forward, Rd1)
Primer 17	caagcagaagacggcatacgagatgtgactggagttcagacgtgtgc tcttccgatct	Illumina sequencing primer II (reverse, Rd2)
Primer 18	ttttttttttttttttttttggcacaccggtttgtgtgactttcgccactccc	For mutation M1 in Z-mod
Primer 19	ttttttttttttttttttttggcagaccggtttgtgtgactttcgccactccc	For mutation M2 in Z-mod
Primer 20	ttttttttttttttttttttggctgaccggtttgtgtgactttcgccactccc	For mutation M3 in Z-mod
Primer 21	ttttttttttttttttttttgggtgaccggtttgtgtgactttcgccactccc	For mutation M4 in Z-mod
Primer 22	ttttttttttttttttttttgcgtgaccggtttgtgtgactttcgccactccc	For mutation M5 in Z-mod
Primer 23	ttttttttttttttttttttccgtgaccggtttgtgtgactttcgccactccc	For mutation M6 in Z-mod
Primer 24	cgtgtgctcttccgatctnnnnnnnn ttttt ttttt ttttt ttttt gg	Reverse primer for reverse transcription for polyG-tailed samples
Primer 25	ttccctacacgacgctcttccgatct agttcc gag cct tgc gcc ggg aaacca	Forward primer used for poly-G tailed samples amplification

Table S2. Model reactions and fitted rate parameters. The ':' denotes a non-covalent complex between two fragments.

	Reaction	Forward	Reverse
a	WXY + Z -> WXY:Z	1.85E+02 $\mu M^{-1} min^{-1}$	4.44E-02 min^{-1}
b	WXY + Z-mod -> WXY:Z-mod	3.18E+01 μM^{-1} min^{-1}	2.14E+01 min^{-1}
c	WXY-mod + Z -> WXY-mod:Z	8.98E+01 μM^{-1} min^{-1}	2.77E-02 min^{-1}
d	WXYZ + WXY:Z -> WXYZ + WXYZ	8.90E+00 μM^{-1} min^{-1}	9.43E-01 μM^{-1} min^{-1}
e	WXY:Z + WXY:Z -> WXYZ + WXY:Z	1.50E+00 μM^{-1} min^{-1}	7.99E+00 μM^{-1} min^{-1}
f	WXY:Z + WXY-mod:Z -> WXYZ + WXY-mod:Z	1.02E+00 μM^{-1} min^{-1}	1.91E+00 μM^{-1} min^{-1}
g	WXY:Z-mod + WXY:Z -> WXYZ + WXY:Z-mod	5.81E-03 μM^{-1} min^{-1}	8.57E+01 μM^{-1} min^{-1}
h	WXY:Z-mod + WXYZ -> WXY-mod + Z + WXYZ	1.45E-01 μM^{-1} min^{-1}	4.93E+00 μM^{-1} min^{-1}
i	WXY:Z-mod + WXY:Z -> WXY-mod + Z + WXY:Z	9.12E-02 μM^{-1} min^{-1}	2.97E+01 μM^{-1} min^{-1}
j	WXY:Z-mod + WXY-mod:Z -> WXY-mod + Z + WXY-mod:Z	8.85E-01 μM^{-1} min^{-1}	8.37E+00 μM^{-1} min^{-1}
k	WXY:Z-mod + WXY:Z-mod -> WXY-mod + Z + WXY:Z-mod	1.26E+02 μM^{-1} min^{-1}	8.50E+01 μM^{-1} min^{-1}
l	WXY-mod:Z + WXYZ -> WXY:Z + mod + WXYZ	1.70E+01 μM^{-1} min^{-1}	1.48E-04 μM^{-2} min^{-1}
m	WXY-mod:Z + WXY:Z -> WXY:Z + mod + WXY:Z	4.86E-03 μM^{-1} min^{-1}	1.47E-04 μM^{-2} min^{-1}
n	WXY-mod:Z + WXY-mod:Z -> WXY:Z + mod + WXY-mod:Z	3.15E-01 μM^{-1} min^{-1}	7.23E+01 μM^{-2} min^{-1}

Table S3. Different selection criteria for models with reduced number of parameters. p is the number of parameters of the model, n the number of data points, df the degree of freedom of the model, RSS is the sum of squared residuals, AIC is the Akaike Information Criterion, BIC is the Bayesian Information Criterion, F is the F-statistic comparing a model with the chosen model (p = 28) and p-value is the associated p-value with the F-test.

p	n	df	RSS	AIC	BIC	F	p-value
8	120	112	0.1535	-781.38	-756.29	36.39	$p < 0.001$
10	120	110	0.1468	-782.75	-752.09	38.44	$p < 0.001$
12	120	108	0.0765	-857.03	-820.79	19.77	$p < 0.001$
16	120	104	0.0599	-878.36	-830.97	18.98	$p < 0.001$
22	120	98	0.0438	-904.00	-839.89	23.61	$p < 0.001$
28	120	92	0.0172	-1003.86	-923.02	N/A	N/A

SUPPLEMENTARY FIGURES

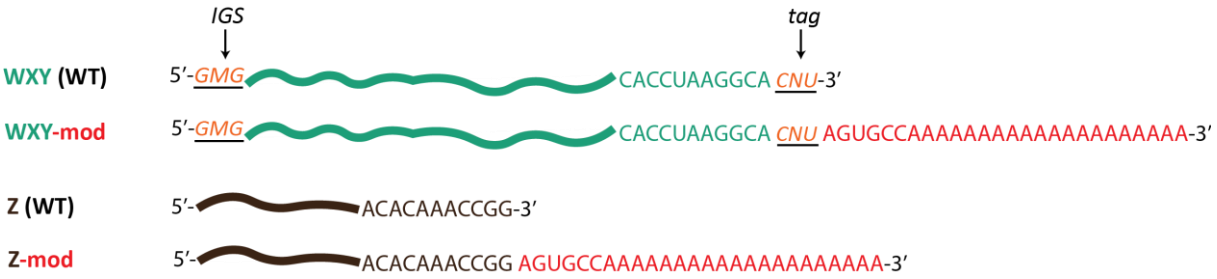

Figure S1. Modification of the substrates. The sequence AGUGCCA followed by a stretch of poly-adenosine (red) is appended to the 3′ end of the substrates. WXY and WXY-mod contains an internal guide sequence (IGS) ("GMG" at the 5′ end, orange and underlined) and a tag ("CNU" at the 3′ end, orange and underlined) required for the transesterification reaction (20, 21) where M and N are variable nucleotides in the IGS and tag (denoted as MN), respectively. While for all the assembly reactions MN is kept as AU (i.e., GAG and CUU as IGS and tag, respectively), for the cooperative network formation AA, GU and UC were used.

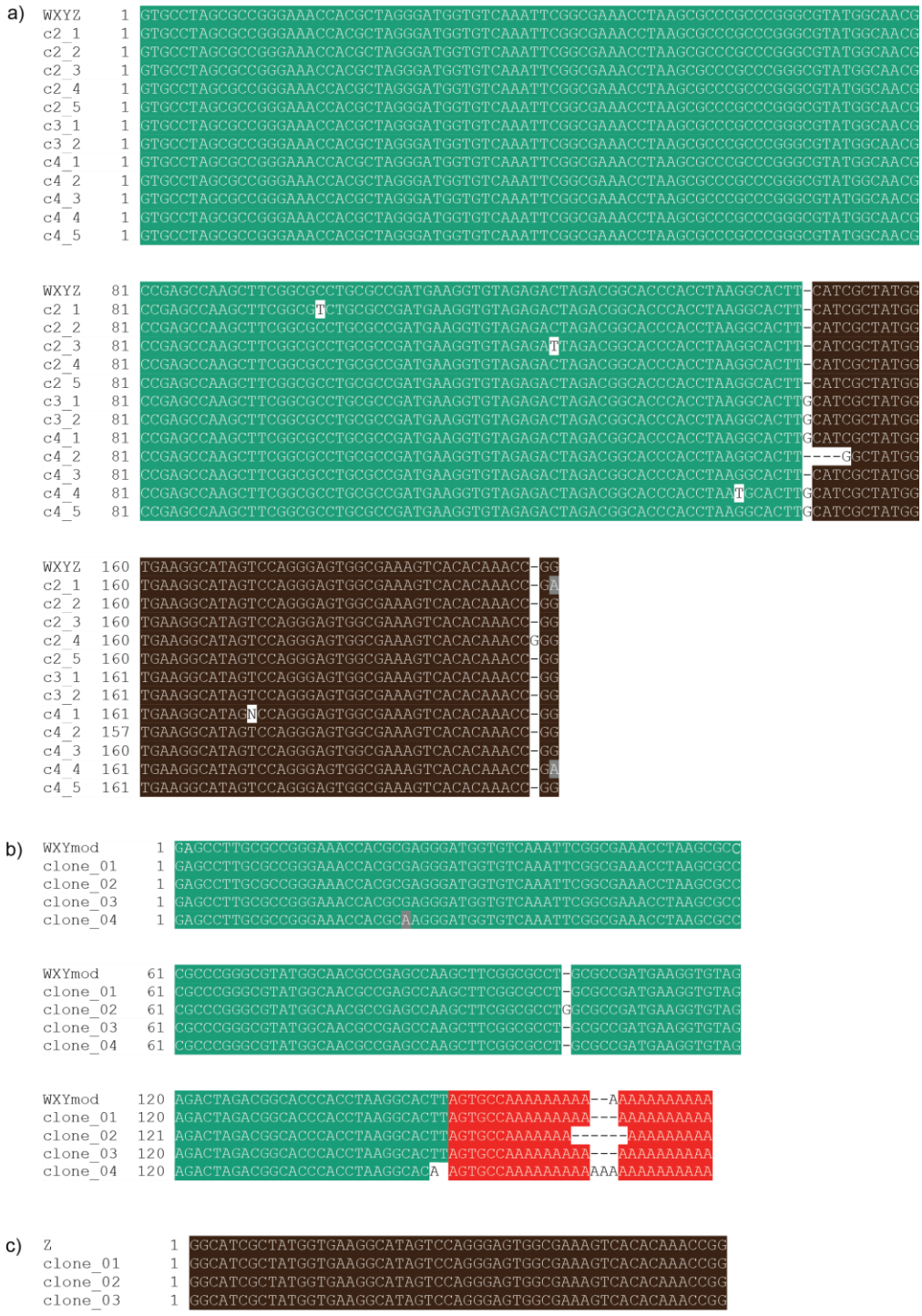

Figure S2. Sanger sequencing results of different reaction products. Sequence alignment of the various products. (**A**) Sequence alignment of WXYZ sequences obtained from the reactions with substrate combination c2, c3 and c4 aligned with the WXYZ reference sequence. (**B**) Sequence alignment of WXY-mod sequences obtained in the reaction with substrate combination c2. The poly-adenosine tail is often poorly sequenced as it contains the same nucleotide repeated several times. (**C**) Sequence alignment of Z sequences obtained in the reaction with substrate combination c2. The alignments were generated using T-Coffee web server (40) with default parameters and BoxShade (https://embnet.vital-it.ch/software/BOX_form.html.) for visualization. Different colors are used to highlight the different parts of the sequence: green for WXY, brown for Z and red for mod.

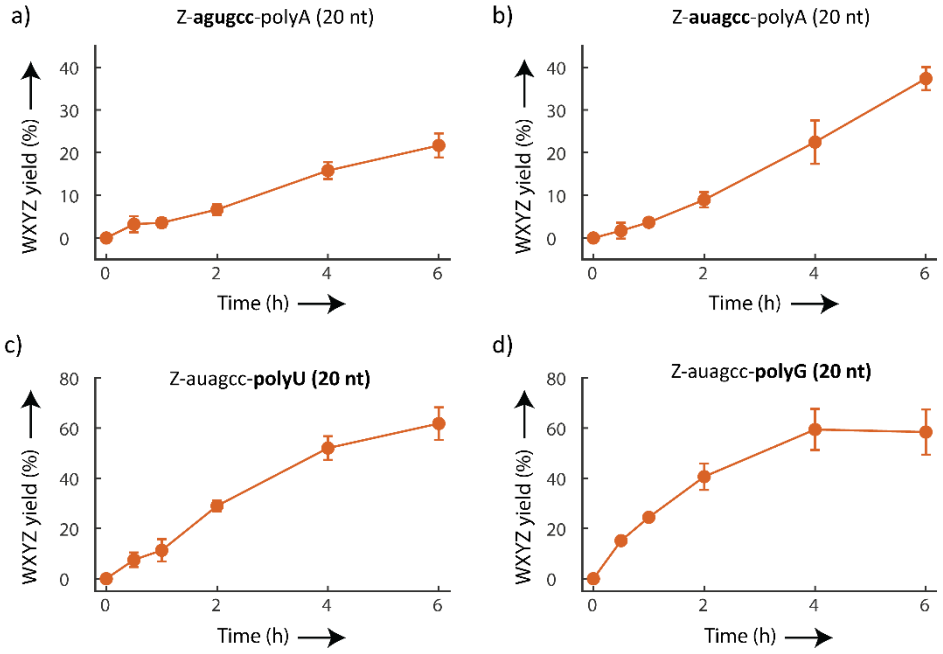

Figure S3. WXYZ formation with diverse modification in the substrate. Kinetic analysis showing that even when different modifications of Z are used the synthesis of catalysts WXYZ is not hampered. (**A**) WXYZ formation with 'agugcc'-polyA modification (used in all other experiments in the study). (**B-D**) WXYZ formation when modification is changed from 'agugcc'-polyA to 'auagcc'-polyA (**B**), 'auagcc'-polyU (**C**) or 'auagcc'-polyG (**D**). Here 0.5 µM of each RNA was incubated in the reaction buffer (see Material and Methods) at 48°C for 6 h. Samples at several time were taken out for gel analysis. The reported yield is the substrate to product conversion in percentage. Error bars represent ±1 standard deviation.

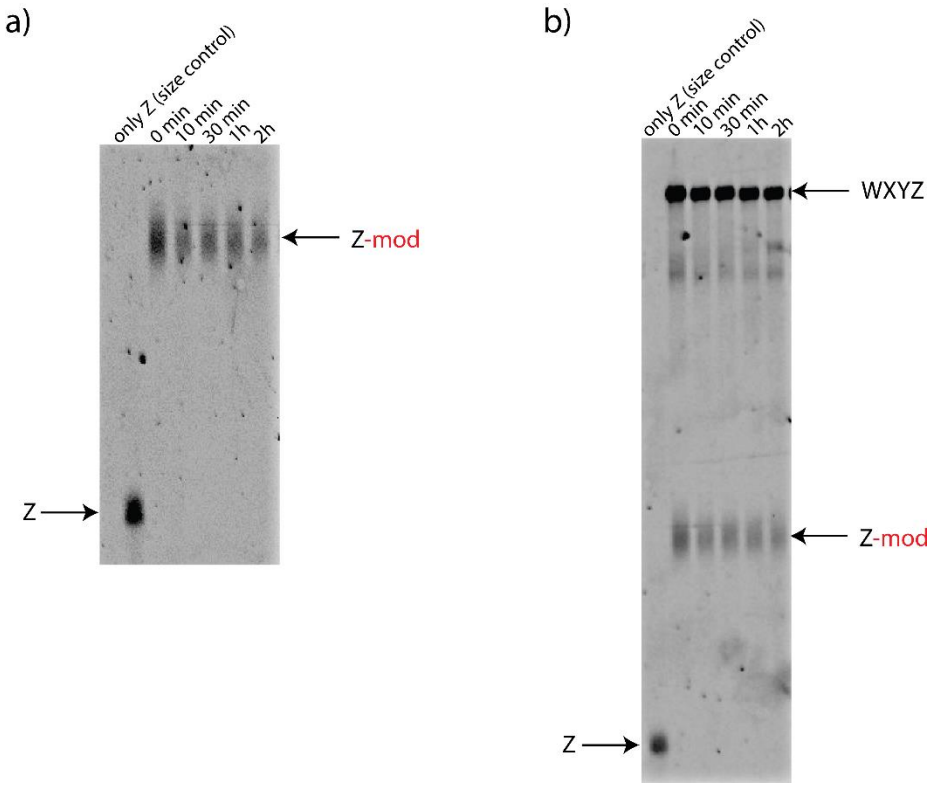

Figure S4. No direct cleavage of modification from the Z-mod substrate. Gel analysis showing that the Z-mod substrate is not hydrolyzed spontaneously. (**A**) Z-mod alone is not hydrolyzed to generate Z in presence of the reaction buffer as no slow migrating band corresponding to Z appears over the course of the reaction. (**B**) Even in presence of catalyst (WXYZ) there is no conversion of Z-mod to Z by cleaving off the modification directly from the substrate. Here 0.5 µM of each RNA was incubated in the reaction buffer at 48°C. Samples at several time points (indicated above the lanes) were taken out for analysis.

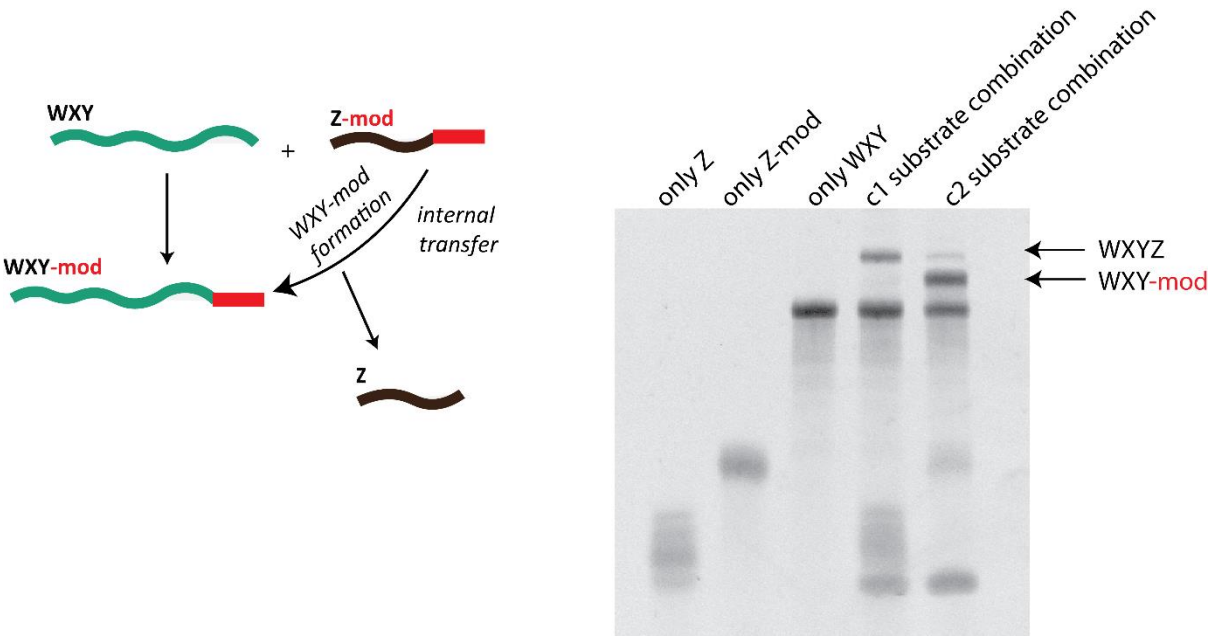

Figure S5. Internal transfer of modification to generate WXY-mod. Schematic showing (left panel) the formation of WXY-mod in the reaction with substrate combination c2 (WXY + Z-mod) by transfer of mod from Z-mod to WXY. Gel analysis (right panel) showing the formation of WXY-mod in the reaction with substrate combination c2. The sequence of WXY-mod was confirmed by sequencing (Supplementary Figure S2). 0.5 µM of each RNA was incubated in the reaction buffer for 1 h at 48°C. As controls, substrate combination c1 (WXY + Z), only Z, only Z-mod and only WXY, were also incubated.

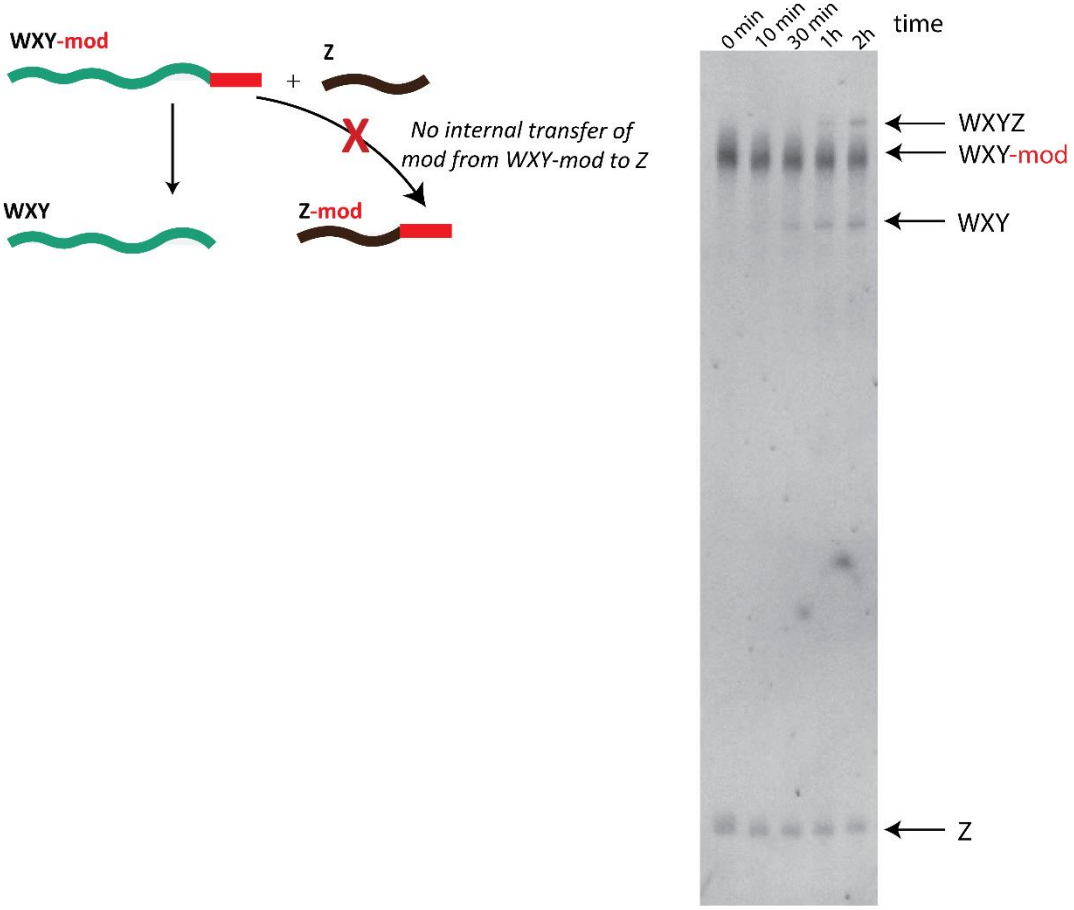

Figure S6. No formation of Z-mod from WXY-mod. Schematic showing (left panel) that there is no formation of Z-mod in the reaction with substrate combination c3 (WXY-mod + Z) by the internal transfer of modification from WXY-mod to Z. Gel analysis (right panel) showing that during the course of the reaction no Z is converted to Z-mod. Here 0.5 µM of each RNA was incubated in the reaction buffer at 48°C. Samples at several time points (indicated above the lanes) were taken out for analysis.

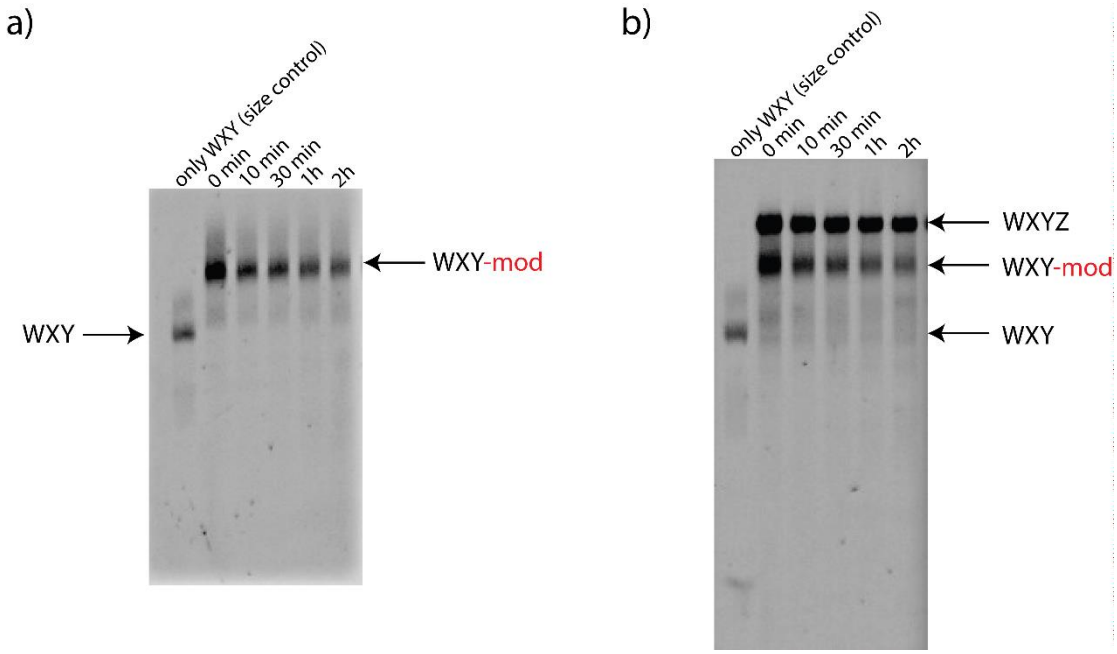

Figure S7. Cleavage of modification from WXY-mod is catalyzed by WXYZ. Gel analysis showing that (**A**) WXY-mod is stable in the reaction buffer as no band corresponding to WXY appears during the course of the reaction, and (**B**) in the presence of the catalyst WXYZ, mod is cleaved off WXY-mod and WXY is generated. Here 0.5 µM of each RNA was incubated in the reaction buffer at 48°C. Samples at several time points (indicated above the lanes) were taken out for analysis.

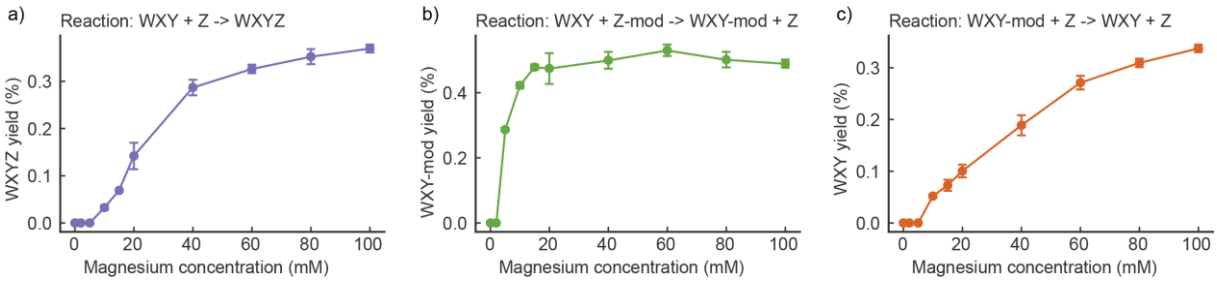

Figure S8. Different steps of the reaction occur even at lower magnesium concentrations. Time course experiment to demonstrate that different steps of the reaction; formation of the catalyst WXYZ from WXY + Z (**A**), formation of WXY-mod (**B**), and processing of WXY-mod to WXY (**C**), can occur at a broad range of $MgCl_2$ concentrations. Here 0.5 µM of each RNA was incubated at 48°C for 1h in the reaction buffer containing different $MgCl_2$ concentration. After 1h, the samples were taken out for the gel analysis. The reported yield is the substrate to product conversion in percentage. Error bars represent ±1 standard deviation.

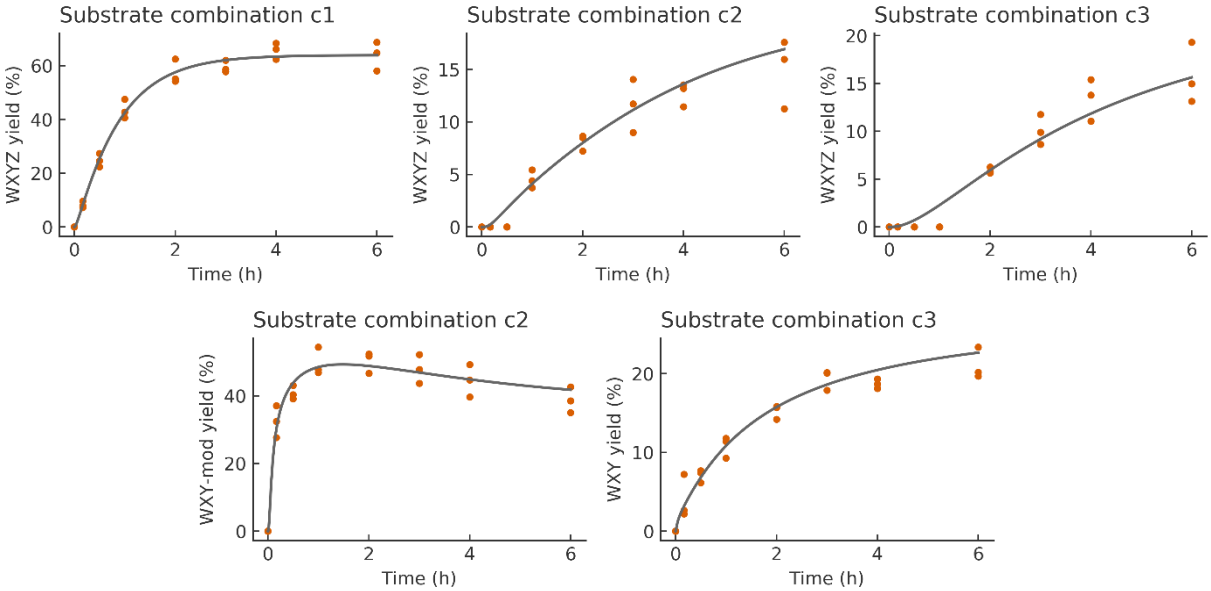

Figure S9. Time course of experimental data (triplicates, orange dots) and kinetic model data (grey line) for the different products formation in different substrate combinations. See main text for description of the reactions. The reported yield is the substrate to product conversion in percentage.

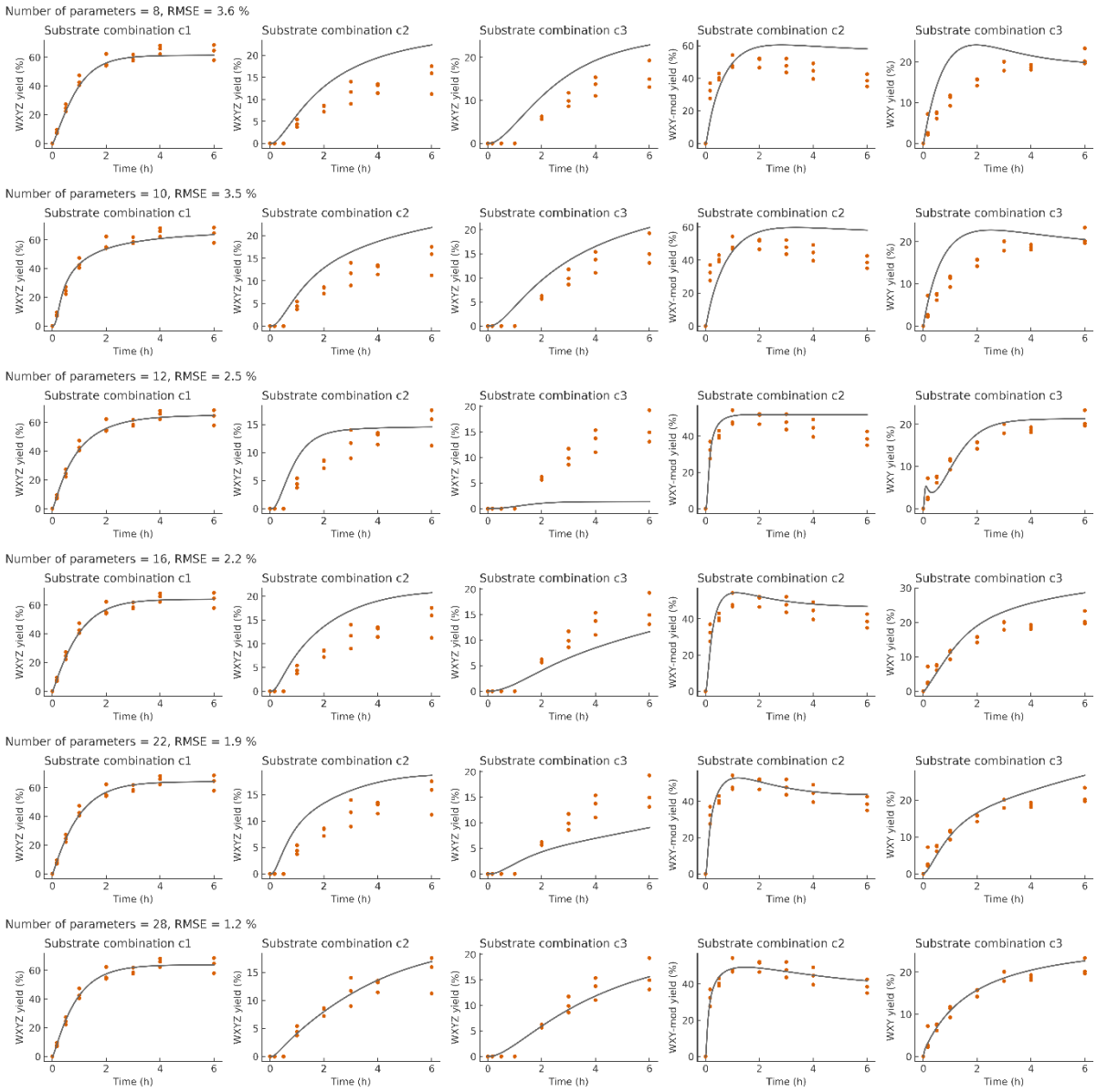

Figure S10. Fitting models with reduced numbers of parameters as described in the Kinetic modelling section. Grey line is fitted model data and orange dots are experimental data points (triplicates). The reported yield is the substrate to product conversion in percentage. For each reduced model the RMSE value is reported.

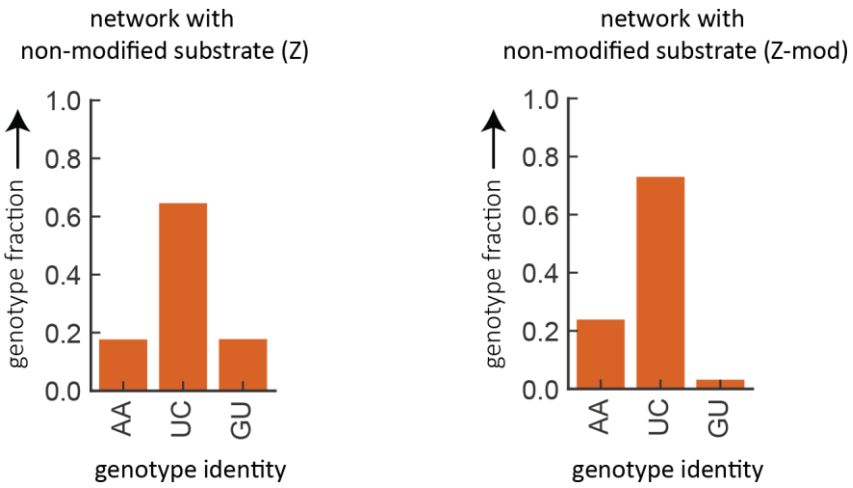

Figure S11. Genotypic distribution of cooperative network formed by modified substrates. Histograms demonstrating that even when the modified Z substrate (Z-mod) is used for cooperative network formation the genotypic distribution of the members in the network is maintained. The Y-axis shows relative genotype fraction derived from number of UMIs (25) obtained for the individual member of the network. The X-axis shows the genotype identity (MN, the sequence of the IGS and tag on each WXY, see Figure 5).

SUPPLEMENTARY REFERENCES

39. Turner,B.D., Henley,B.J., Sleap,S.B. and Sloan,S.W. (2015) Kinetic model selection and the Hill model in geochemistry. *Int. J. Environ. Sci. Technol.*, **12**, 2545-2558.

40. Di Tommaso,P., Moretti,S., Xenarios,I., Orobitg,M., Montanyola,A., Chang,J.M., Taly,J.F. and Notredame,C. (2011). T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension, *Nucleic Acids Res.*, **39**, W13-17.