PNAS

www.pnas.org

SUPPLEMENTARY INFORMATION FOR:

Substrate Selectivity by the Exonuclease Rrp6p.

Armend Axhemi [1, 2], Elizabeth V. Wasmuth [3],

Christopher D. Lima [3, 4] & Eckhard Jankowsky [1, 2, 5]

[1]   Center for RNA Science and Therapeutics, School of Medicine,
        Case Western Reserve University, Cleveland, OH 44106, USA

[2]   Department of Biochemistry, School of Medicine,
        Case Western Reserve University, Cleveland, OH 44106, USA

[3]   Structural Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer
        Center, 1275 York Avenue, New York, NY 10065, USA

[4]   Howard Hughes Medical Institute

[5]   Corresponding Author: exj13@case.edu

This PDF file includes:

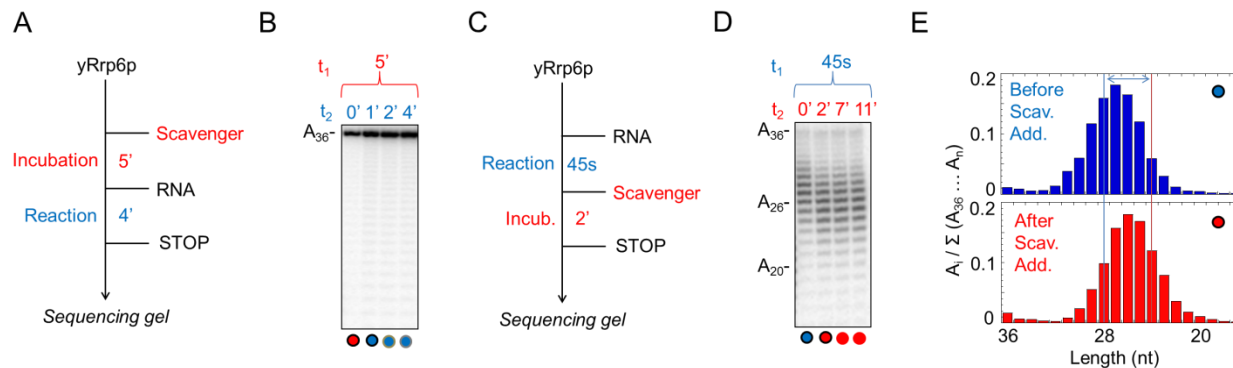Figs. S1 to S12 (with captions)
Table S1

Fig. S1. Processivity measurements with A_{36} substrate.

(*A*) Reaction scheme for the pulse-chase control reaction. (*B*) Representative control reaction of Rrp6p (27 nM) incubated with scavenger (10 µM) before addition of A_{36} (1 nM). t_1: incubation time before addition of labeled substrate, t_2: reaction time after addition of labeled substrate. (*C*) Reaction scheme for a pulse-chase reaction. (*D*) Representative pulse-chase reaction of Rrp6p (27 nM) with A_{36} (1 nM) and scavenger RNA (10 µM). t_1: reaction before scavenger addition, t_2: reaction after scavenger addition [panel (*C*)]. (*E*) Histograms of the relative abundance of each substrate species (A_i / Σ (A_{36} …A_n) as a function of substrate length, before (blue) and after (red) scavenger addition.
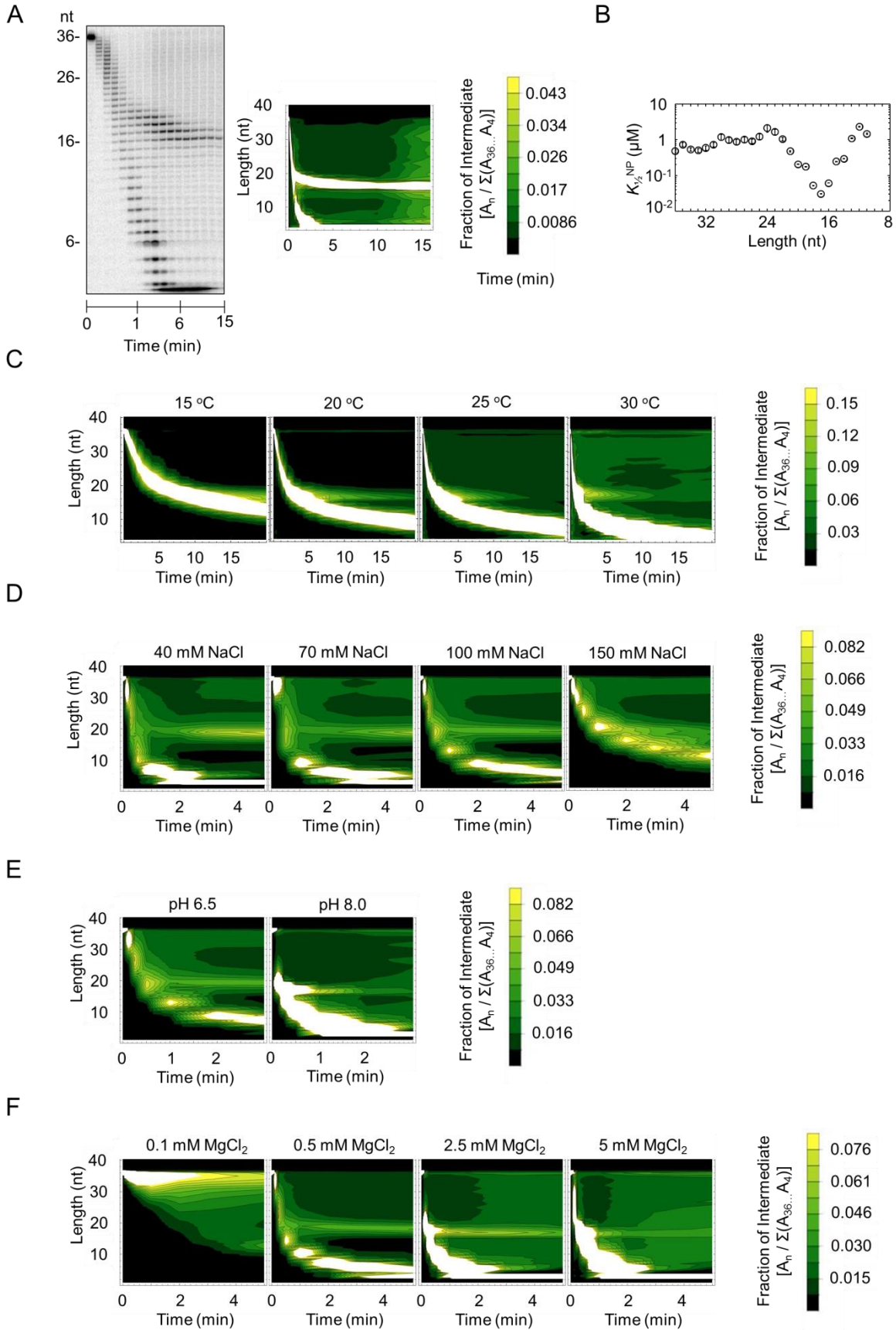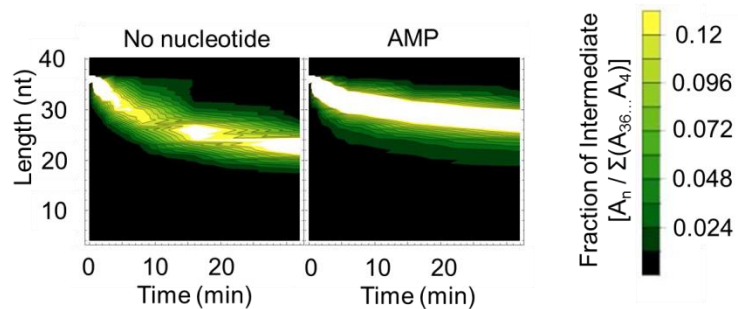
A

B

C

D

E

F

Fig. S2. Rrp6p activity under different reaction conditions.

(*A*) Contour plot of a degradation reaction. The left panel shows a representative PAGE (Fig. 2*A*), the right panel the corresponding contour plot. The coloring shows the fraction of degraded intermediates [$A_n / \Sigma(A_{36} \ldots A_4)$] versus time, as indicated in legend bar on the right [$A_i = 0$ (black) to 0.043 (yellow)]. (*B*) Non-productive dissociation constants, $K_{1/2}^{NP}$, for degradation of A_{36} under normal reaction conditions. Error bars show the standard error of global fit. (*C*) Contour plots for reactions of Rrp6p (216 nM, A_{36}: 1 nM) at different temperatures, as indicated. (*D*) Contour plots for reactions of Rrp6p (432 nM, A_{36}: 1 nM) at different NaCl concentrations, as indicated. (*E*) Contour plots of reactions of Rrp6 (432 nM, A_{36}: 1 nM) at different pH. (*F*) Contour plots of reactions of Rrp6 (432 nM, A_{36}: 1 nM) at different $MgCl_2$ concentrations.
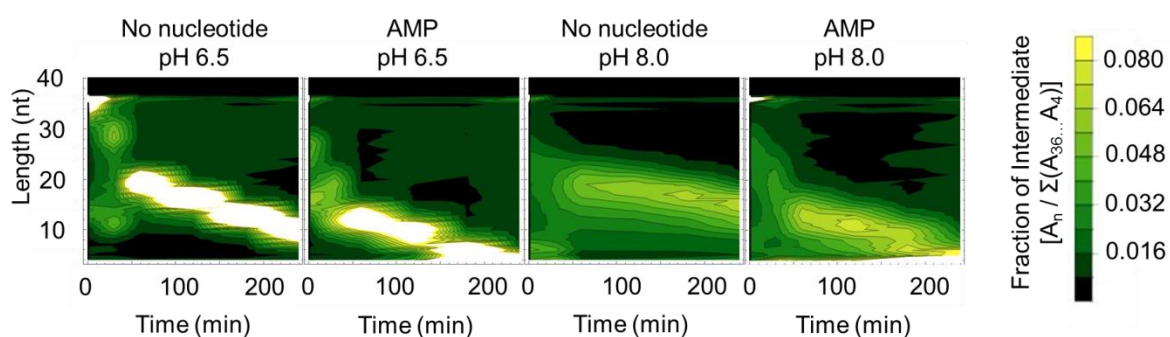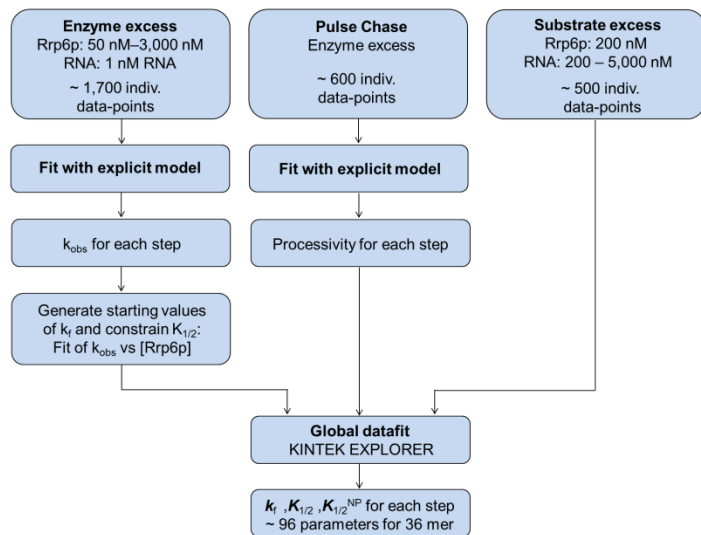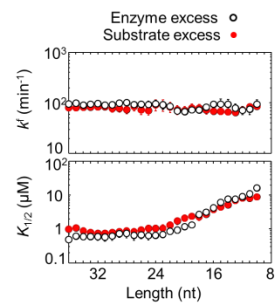
Fig. S3. Rrp6p activity in the presence of adenosine 5'-monophosphate (AMP).
(*A*) Contour plots for reactions with Rrp6p (27 nM, A_{36}: 1 nM) in the absence and presence of 2 mM AMP. (*B*) Contour plots for reactions with Rrp6p (15 nM, A_{36}: 500 nM) in the absence and presence of 2 mM AMP at pH 6.5 and pH 8.0.
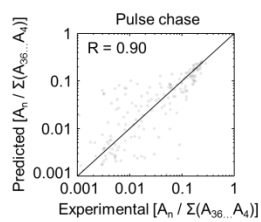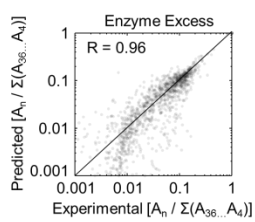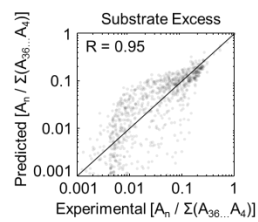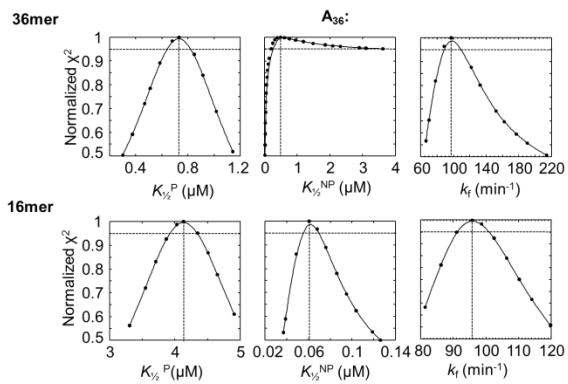
A

Enzyme excess
Rrp6p: 50 nM–3,000 nM
RNA: 1 nM RNA
~ 1,700 indiv.
data-points

Pulse Chase
Enzyme excess

~ 600 indiv.
data-points

Substrate excess
Rrp6p: 200 nM
RNA: 200 – 5,000 nM
~ 500 indiv.
data-points

Fit with explicit model

Fit with explicit model

k_{obs} for each step

Processivity for each step

Generate starting values
of k_f and constrain $K_{1/2}$:
Fit of k_{obs} vs [Rrp6p]

Global datafit
KINTEK EXPLORER

k_f , $K_{1/2}$, $K_{1/2}^{NP}$ for each step
~ 96 parameters for 36 mer

B

Enzyme excess ○
Substrate excess ●

k' (min^{-1})

$K_{1/2}$ (μM)

Length (nt)

C

Pulse chase

R = 0.90

Predicted [A_n / Σ($A_{36...}A_4$)]

Experimental [A_n / Σ($A_{36...}A_4$)]

D

Enzyme Excess

R = 0.96

Predicted [A_n / Σ($A_{36...}A_4$)]

Experimental [A_n / Σ($A_{36...}A_4$)]

E

Substrate Excess

R = 0.95

Predicted [A_n / Σ($A_{36...}A_4$)]

Experimental [A_n / Σ($A_{36...}A_4$)]

F

A_{36}:

36mer

Normalized χ^2

$K_{1/2}^P$ (μM)

$K_{1/2}^{NP}$ (μM)

k_f (min^{-1})

16mer

Normalized χ^2

$K_{1/2}^P$ (μM)

$K_{1/2}^{NP}$ (μM)

k_f (min^{-1})

G

R_{36}:

36mer

Normalized χ^2

$K_{1/2}^P$ (μM)

$K_{1/2}^{NP}$ (μM)

k_f (min^{-1})

16mer

Normalized χ^2

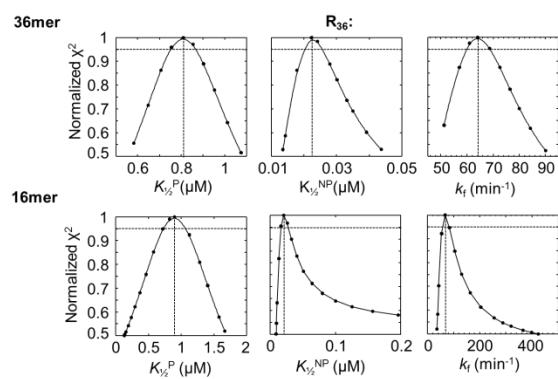$K_{1/2}^P$ (μM)

$K_{1/2}^{NP}$ (μM)

k_f (min^{-1})

Fig. S4. Parameter calculation for the kinetic framework.

(*A*) Pipeline for the calculation of the kinetic parameters for each cleavage step. The explicit models and the global data-fits are explained in the Materials and Methods Section. (*B*) Plots of degradation rate constants (k_f) and apparent dissociation constants for productive binding ($K_{1/2}^{P}$) obtained from global data fits of reactions with only enzyme excess (open circles) or only substrate excess (red filled circles) for A_{36} degradation. (*C*) Correlation between the experimental data of degradation intermediates (A_i) for pulse-chase experiments of Rrp6p on A_{36} versus the corresponding data calculated with the kinetic model. (*D*) Correlation between the experimental data of degradation intermediates (A_i) for reactions with excess Rrp6p over A_{36} versus the corresponding data calculated with the kinetic model. (*E*) Correlation between the experimental data of degradation intermediates (A_i) for reactions with excess A_{36} over Rrp6p versus the corresponding data calculated with the kinetic model. (*F*) Representative plots of normalized χ^2 vs. kinetic parameters (indicated at the x-axis) for 36-mer (top panel) and 16-mer species (bottom panel) for global fits of A_{36}. The vertical dashed line indicates the optimum parameter value at χ^2 of 1. The points where the horizontal dashed line intersects the curve indicate the lower and upper boundaries of 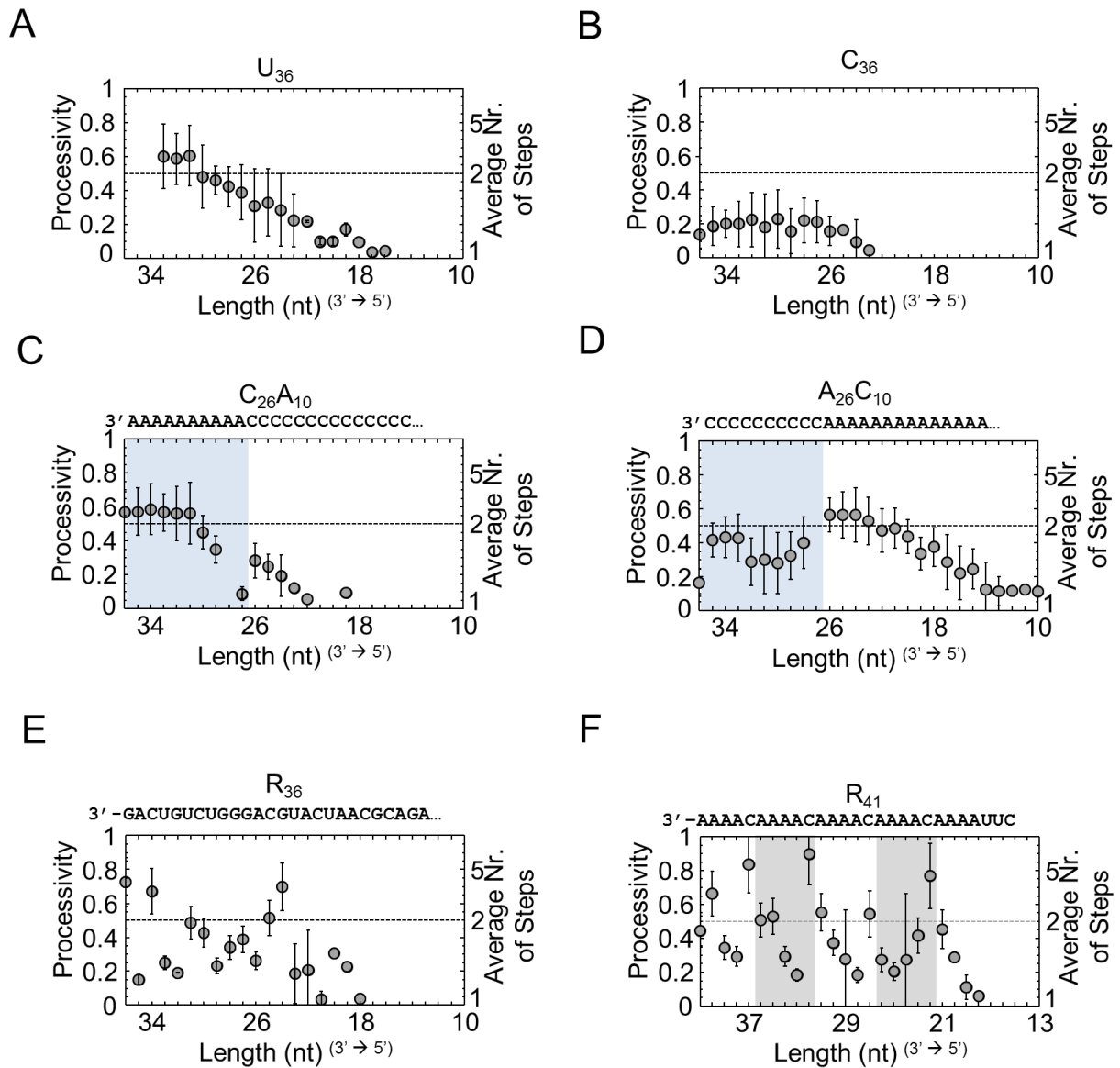the kinetic parameter ($\chi^2 = 0.95$). (*G*) Representative plots of normalized χ^2 vs. kinetic parameters (indicated at the x-axis) for 36-mer (top panel) and 16-mer species (bottom panel) for global fits of the mixed sequence substrate R_{36}. The points where the horizontal dashed line intersects the curve indicate the lower and upper boundaries of the kinetic parameter ($\chi^2 = 0.95$).
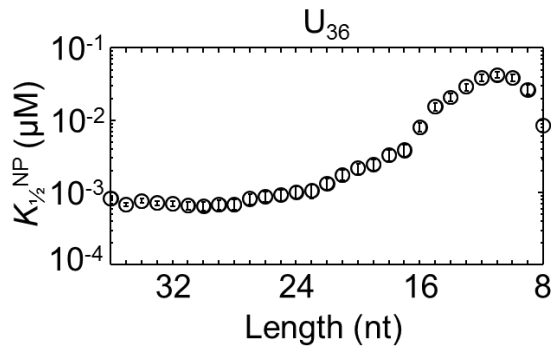
A

U$_{36}$

Processivity

Average Nr. of Steps

Length (nt) $^{(3' \rightarrow 5')}$

B

C$_{36}$

Processivity

Average Nr. of Steps

Length (nt) $^{(3' \rightarrow 5')}$

C

C$_{26}$A$_{10}$

3' AAAAAAAAAACCCCCCCCCCCCCC...

Processivity

Average Nr. of Steps

Length (nt) $^{(3' \rightarrow 5')}$

D

A$_{26}$C$_{10}$

3' CCCCCCCCCCAAAAAAAAAAAAAA...

Processivity

Average Nr. of Steps

Length (nt) $^{(3' \rightarrow 5')}$

E

R$_{36}$

3'-GACUGUCUGGGACGUACUAACGCAGA...

Processivity

Average Nr. of Steps

Length (nt) $^{(3' \rightarrow 5')}$

F

R$_{41}$

3'-AAAACAAAACAAAACAAAACAAAAUUC

Processivity

Average Nr. of Steps

Length (nt) $^{(3' \rightarrow 5')}$

Fig. S5. *Processivity* for different RNA substrates.

(*A-F*) Processivity of Rrp6p at single nucleotide resolution measured for the substrates

indicated. The average number of steps (right axis) was calculated according to P = (N-1)•N^{-1}

(P: processivity, N: number of steps). Data points represent an average of three independent

measurements. Error bars show one standard deviation.

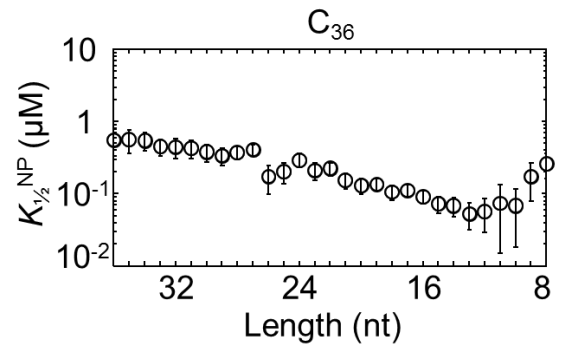**Fig. S6. Non-productive binding for U$_{36}$ and C$_{36}$ substrates** (A) Non-productive dissociation constants, $K_{1/2}^{NP}$ for degradation of U$_{36}$. Error bars show the standard error of global fit. (B) Plot corresponding to (A) for degradation of C$_{36}$.
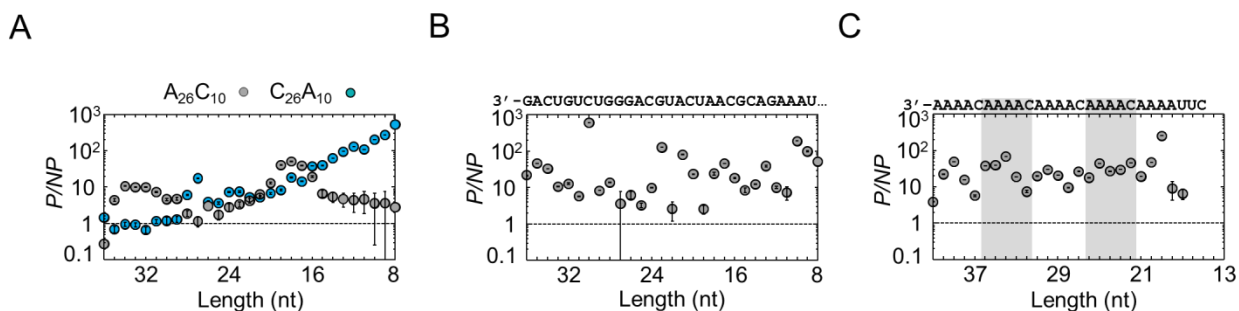
Fig. S7. Non-productive binding for 5'-A$_{26}$C$_{10}$-3', 5'-C$_{26}$A$_{10}$-3', R$_{36}$ and R$_{41}$ substrates.

(*A*) Ratio between productive and non-productive dissociation constants (P/NP) for each step for the degradation of 5'-A$_{26}$C$_{10}$-3' (black circles) and 5'-C$_{26}$A$_{10}$-3' (aqua circles). Error bars show the standard error of the ratio. (*B*) Ratio between productive and non-productive dissociation constants (P/NP) for each step for the degradation of the mixed sequence substrate R$_{36}$ (sequence is shown above the plot). Error bars show the standard error of the ratio. (*C*) Ratio between productive and non-productive dissociation constants (P/NP) for each step for the degradation of the substrate R$_{41}$, which contains the A$_4$C repeats (sequence is shown above the plot). Error bars show the standard error of the ratio.
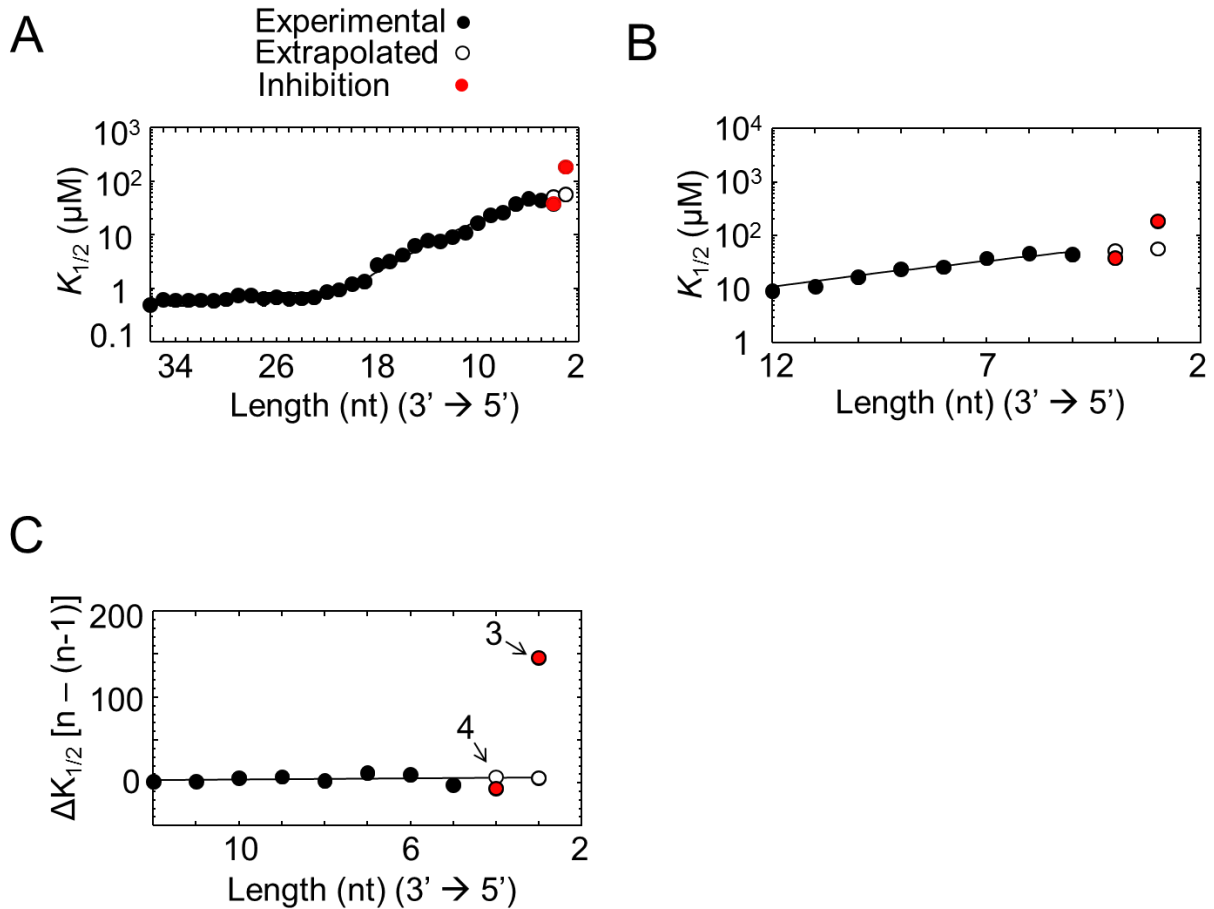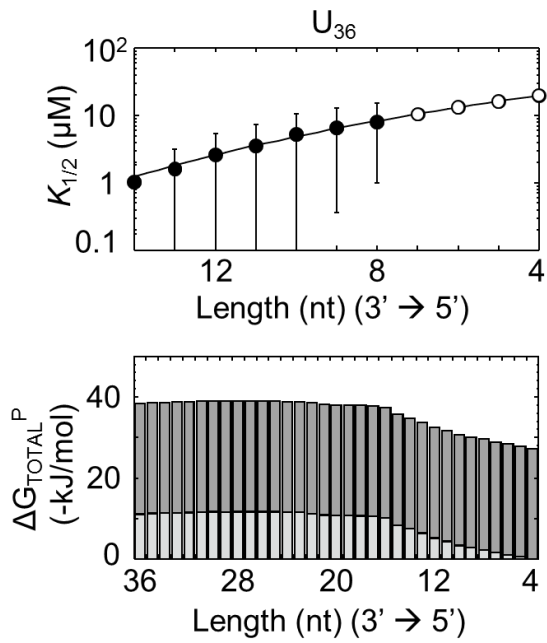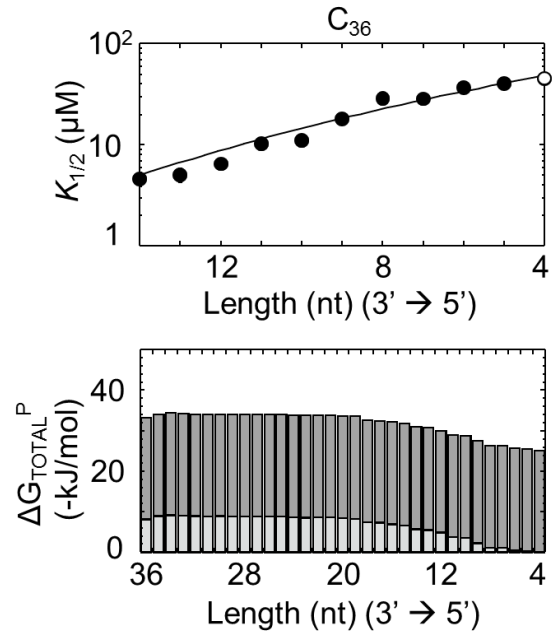
Fig. S8. Deviation of $K_{1/2}^{P}$ values for 3 nt RNAs from competitive inhibition experiments from values extrapolated with the kinetic framework.

(*A*) Polynomial fit of $K_{1/2}^{P}$ values (A_{36}) determined from the global datafit with the kinetic framework (black filled circles), values extrapolated with the polynomial fit for species with 4 and 3 nt (open circles), and experimental values for species with 3 and 4 nt from competitive inhibition experiments (red circles, Fig. 5*A-C*). The value for the 4 nt species from the competitive inhibition experiments is similar to the corresponding value extrapolated from the polynomial fit. The value for the 3 nt species from the competitive inhibition experiments is significantly higher than the corresponding value extrapolated from the polynomial fit. This discrepancy indicates that binding mode of the 3 nt species deviates significantly from the binding mode of the 4 nt species. (*B*) Zoom into the plot in panel (*A*) for 12 to 3 nt species. (*C*) Alternative plot of the data shown in panel (*A*): the Y axis shows the change in dissociation constant for each degradation, indicating that the value of $K_{1/2}^{P}$ obtained from the competitive inhibition for the 3 nt is higher than the value expected from the polynomial fit by a factor of approximately F ~ 150.
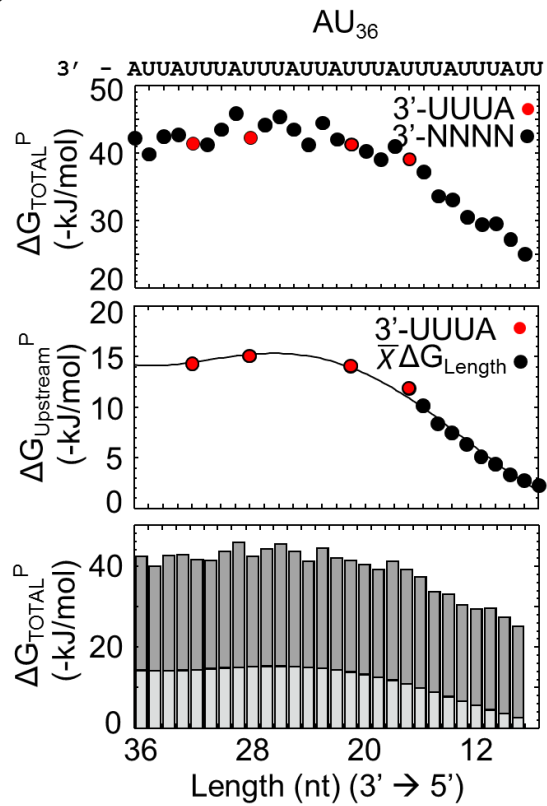
A

U₃₆

B

C₃₆

C

AU₃₆

3'–AUUAUUAUUAUUAUUAUUAUUAUUAUU

3'-UUUA ●
3'-NNNN ●

3'-UUUA ●
$\overline{x}\Delta G_{Length}$ ●

D

R₄₁

3'–AAAACAAAACAAAACAAAACAAAAUUC . .

3'-AAAN ●
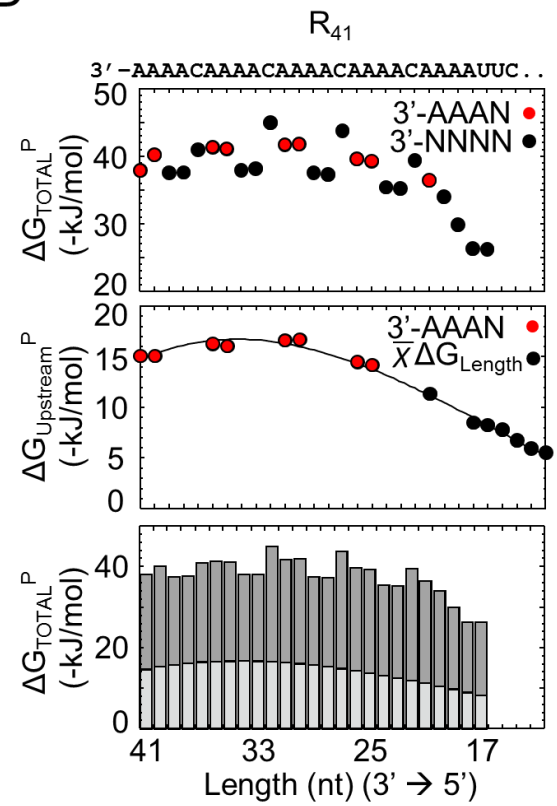3'-NNNN ●

3'-AAAN ●
$\overline{x}\Delta G_{Length}$ ●

Fig. S9. Contributions of RNA length and 3'-terminal quadruplet to the free energy of productive binding for different substrates.

(*A*) Upper panel: polynomial fit of $K_{1/2}{}^P$ values for U_{36}, determined from the global datafit with the kinetic framework (black filled circles) and values extrapolated with the polynomial fit for species with 7 to 4 nt (open circles). The free energy contribution of the U_4 3'-terminal quadruplet was calculated from the extrapolated $K_{1/2}{}^P$ value of U_4, according to Eq. 5 (Materials and Methods). Lower panel: contributions of upstream RNA (light grey) and 3'-terminal (U_4) quadruplet to the total free energy of productive binding (ΔG_{TOTAL}) as a function of RNA length for U_{36}. (*B*) Upper panel: polynomial fit of $K_{1/2}{}^P$ values for C_{36}, determined from the global datafit with the kinetic framework (black filled circles) and value extrapolated with the polynomial fit for species with 4 nt (open circle). The free energy contribution of the C_4 3'-terminal quadruplet was calculated from the extrapolated $K_{1/2}{}^P$ value of C_4, according to Eq. 5 (Materials and Methods). Lower panel: contributions of upstream RNA (light grey) and 3'-terminal (C_4) quadruplet to the total free energy of productive binding (ΔG_{TOTAL}) as a function of RNA length for C_{36}. (*C*) Upper panel: Free energy values (ΔG_{TOTAL}) of productive binding for the AU-rich R_{36} substrate (sequence indicated above the plot), calculated from experimentally determined $K_{1/2}{}^P$ values. Red circles indicate quadruplets for which $\Delta G_{Quadruplet}$ had been determined with other substrates. For these quadruplets (3'UUUA), $\Delta G_{UPSTREAM}$ was calculated according to eq. 4 (Materials and Methods). Middle panel: polynomial fit for known values of $\Delta G_{UPSTREAM}$ (red circles: 3'-UUUA, upper panel; black circles from U_{36} and A_{36} homopolymers, Fig. 5*K*). Values for $\Delta G_{UPSTREAM}$ for unknown quadruplets were extrapolated from the polynomial fit. Bottom panel: contributions of upstream RNA (light grey) and 3'-terminal quadruplet (dark grey) to free binding energy for productive binding for each nucleotide for the AU-rich R_{36} substrate. (*D*) Upper panel: Free energy values (ΔG_{TOTAL}) of productive binding for the R_{41} substrate (sequence indicated above the plot), calculated from experimentally determined $K_{1/2}{}^P$ values. Red circles indicate quadruplets for which $\Delta G_{Quadruplet}$ had been determined with other substrates. Middle panel: polynomial fit for known values of $\Delta G_{UPSTREAM}$ (red circles: 3'-AAAN, upper panel; black circles from U_{36} and A_{36} homopolymers, Fig. 5*K*). Values for $\Delta G_{UPSTREAM}$ for unknown quadruplets were extrapolated from the polynomial fit. Bottom panel: contributions of upstream RNA (light grey) and 3'-terminal quadruplet (dark grey) to free binding energy for productive binding for each nucleotide for the R_{41} substrate.
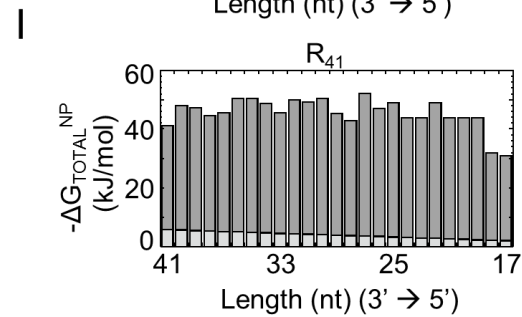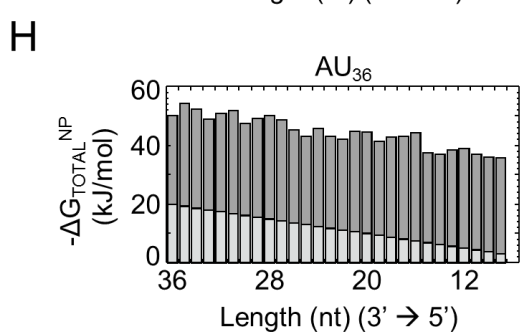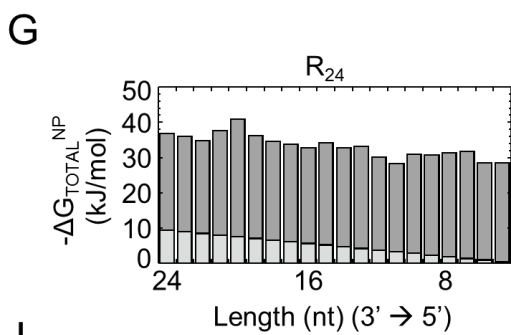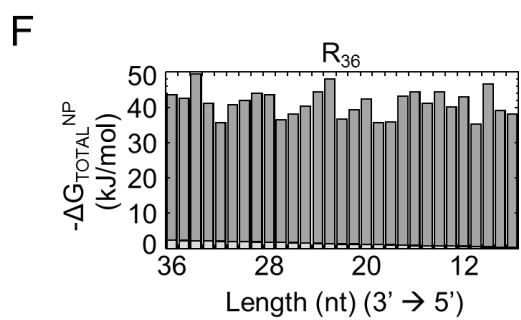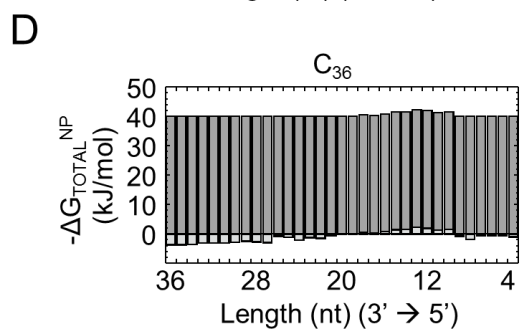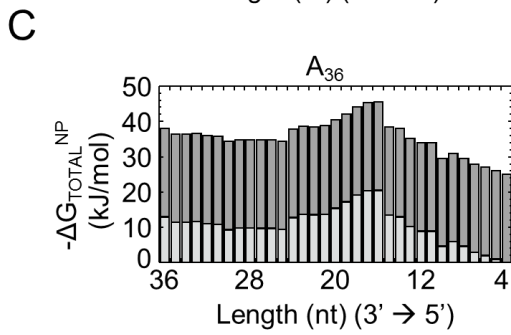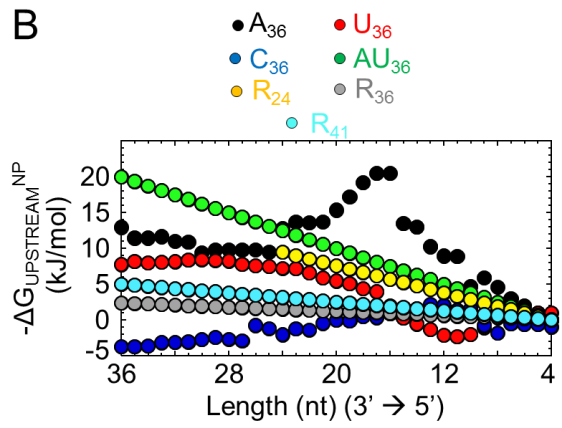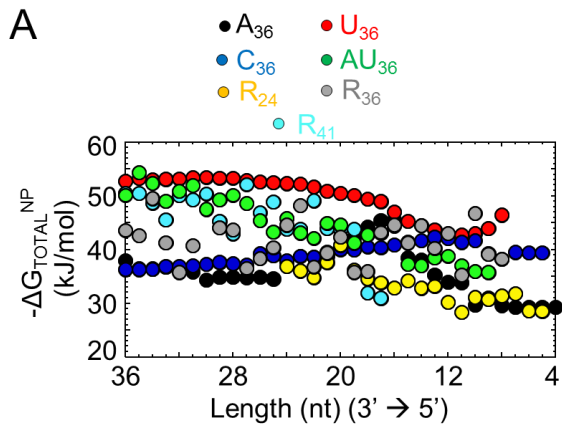
Fig. S10. Contributions of RNA length and 3'-terminal quadruplet to the free energy of non-productive binding.

(*A*) Dependence of total free energy of non-productive binding (ΔG_{TOTAL}^{NP}) on RNA length for different substrates, as indicated. (*B*) Contribution of upstream RNA to the free energy of non-productive binding ($\Delta G_{UPSTREAM}^{NP}$) for different substrates, as indicated. (*C-I*) Contributions of upstream RNA component (light grey) and 3'-terminal quadruplet (dark grey) to free binding energy for non-productive binding for each nucleotide for the indicated substrates. Positive free energy values suggest either no length contribution, or a slightly destabilizing RNA length contribution to the free energy of non-productive binding.
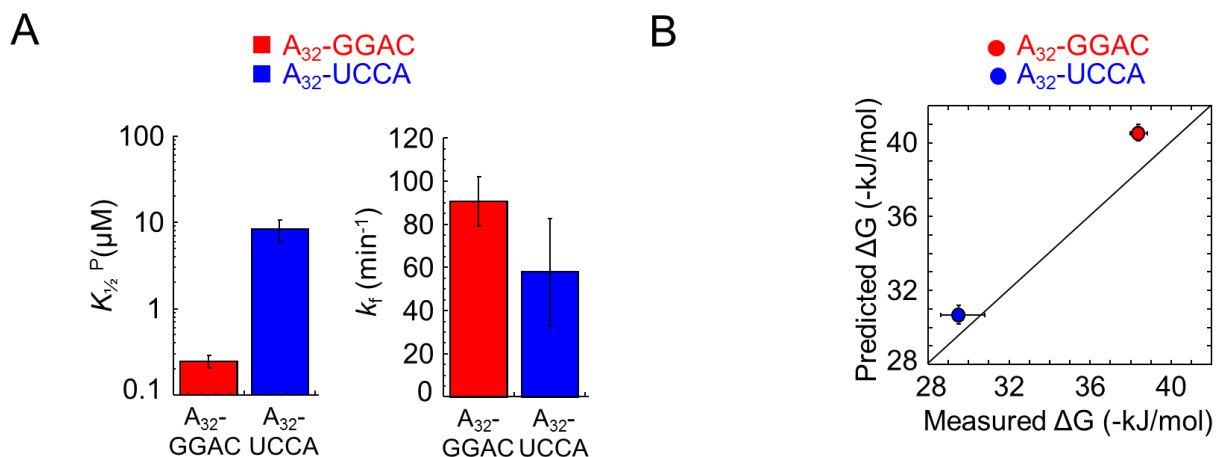
Fig. S11. Kinetic parameters for degradation of A$_{32}$-GGAC and A$_{32}$-UCCA

(A) Estimation of kinetic parameters for degradation of indicated substrates (left panel: $K_{1/2}^{P}$; right panel: k_f). (B) Correlation between experimentally measured free energies of binding ($\Delta G_{Binding}$) and corresponding values calculated with the PWM in Fig 6B. Error bars indicate the standard error for the experimental and predicted free energies.
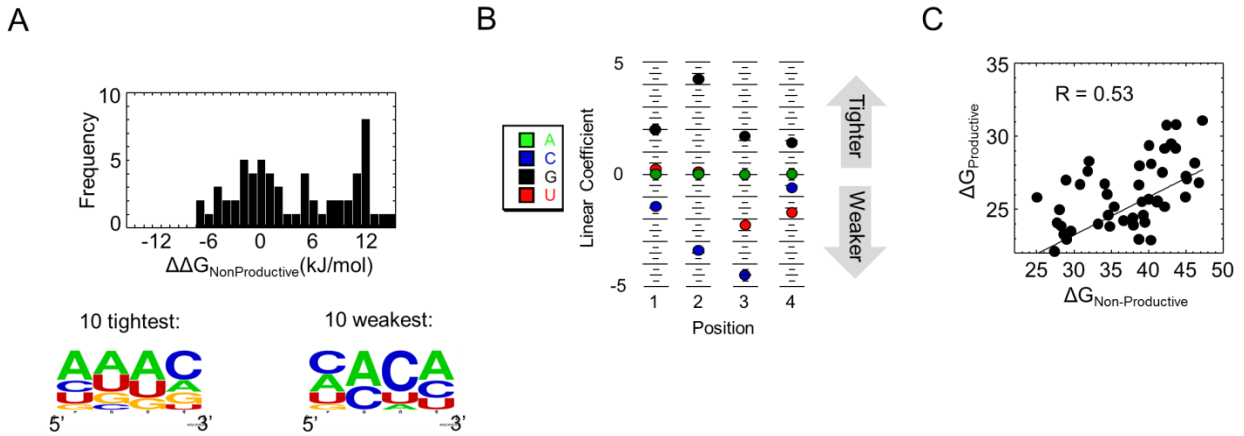
Fig. S12. The impact of the 3'-terminal quadruplet on non-productive binding.

(*A*) Distribution of free energy contributions ($\Delta\Delta G_{Binding}$) for all experimentally tested 3'-terminal quadruplet sequences to non-productive binding, compared to the A_4 variant [$\Delta\Delta G_{Binding}(A_4) = 0$] Bottom: sequence probability logo for the 10 tightest (left) and the 10 weakest (right) bound sequence variants. (*B*) Position weight matrix (PWM) for the free energy contributions of quadruplet sequence variants. Higher linear coefficients indicate higher energy contribution. (*C*) Correlation between free energy contributions of 3'-terminal quadruplet sequences to productive (Y axis) and non-productive binding (X-axis).

Substrate	Enzyme excess (Substrates: 1 nM)	Nr. of Replicates	Pulse chase (Substrates: 1 nM)	Nr. of Replicates	Substrate excess	Nr. of Replicates
A_{36}	Rrp6: 13, 27, 55, 110, 220, 440, 880 & 1760 nM	6	Rrp6p: 27, 55, 110 nM Scavenger: 10µM	3	Rrp6: 200nM A_{36}: 100, 200, 400, 800, 1600, 3200 & 5,000 nM	2
U_{36}	Rrp6: 13, 27, 55, 110, 220, 440 & 880 nM	3	Rrp6p: 27, 55, 110 nM Scavenger: 10µM	3	Rrp6: 80nM U_{36}: 80, 160, 320, 640, 1280 & 2560 nM	2
C_{36}	Rrp6: 55, 110, 220, 440, 880 & 1760nM	2	Rrp6p: 27, 55, 110 nM Scavenger: 10µM	3	Rrp6: 80nM U_{36}: 80, 160, 320, 640, 1280 & 2560 nM	2
$A_{26}C_{10}$	Rrp6: 55, 110, 220, 440, 880 & 1760 nM	2	Rrp6p: 27, 55, 110 nM Scavenger: 10µM	3	N/A	
$C_{26}A_{10}$	Rrp6: 55, 110, 220, 440, 880 & 1760 nM	2	Rrp6p: 27, 55, 110 nM Scavenger: 10µM	3	N/A	
R_{24}	Rrp6: 55, 110, 220, 440 & 880 nM	2	Rrp6p: 27, 55, 110 nM Scavenger: 10µM	3	Rrp6: 50 nM R_{24}: 250, 500, 750, 1000 & 1500 nM	1
R_{36}	Rrp6: 55, 110, 220, 440 & 880 nM	2	Rrp6p: 27, 55, 110 nM Scavenger: 10µM	3	N/A	
R_{41}	Rrp6: 55, 110, 220, 440 & 880 nM	2	Rrp6p: 27, 55, 110 nM Scavenger: 10µM	3	N/A	

Table S1. Rrp6p and substrate concentrations for experiments used in global data fits.