TABLE S1.

Experimentally tested splice sites on CAT mRNA.

				CAT mRNA		Short substrate			Loss of 3'-exon	
position[1]	IGS	Target site[2]	trans-tagging number[3]	$-\Delta G_{bind}$ (kcal/mol)[4]	Product (%)[5]	$-\Delta G_{bind}$ (kcal/mol)	Product (%)[6]	Ribozyme transcription (%)[7]	Transcription (%)[8]	Trans-splicing (%)[9]
83	GCUGAC	CAGUCAGUUG	2	3.3±0.6	1.64±0.14	5.0	13±6	36±6	36±3	42±3
87	GGCAAC	CAGUUGCUCA	2	1.1±0.5	1.1±0.2	4.5	5.1±1.9	48±7	32±4	49±4
97	GGGUAC	AUGUACCUAU	14	4.5±0.3	7.9±1.0	5.7	12.9±0.7	100	30±10	65±3
131	GAGGCC	ACGGCCUUUU	1	N/C	N/D	2.3	2.7±0.2	22±3	21±7	59±5
197	GGGCGG	GCCCGCCUGA	1	4.2±0.5	6.0±1.2	7.0	10.1±1.3	20±3	12±4	25±20
222	GCGGAA	AAUUCCGUAU	0	2.61±0.09	0.42±0.05	4.8	9±6	43±8	18±2	39±5
240	GCCGUC	AAGACGGUGA	0	N/C	N/D	3.9	29±7	17±3	28±6	22±4
258	GUCCCA	UAUGGGAUAG	0	6.09±0.10	14.3±1.5	6.4	12±2	21±6	30±8	27±8
273	GCAAGG	ACCCUUGUUA	0	2.9±0.2	N/D	5.4	4±2	70±10	21±2	47±4
325	GUUCAC	GAGUGAAUAC	0	3.43±0.11	N/D	4.7	1.0±0.2	13±3	36±3	42±7
346	GCUGCC	CCGGCAGUUU	0	3.1±0.4	N/D	6.1	20±8	54±12	32±5	37±4
350	GGAAAC	CAGUUUCUAC	0	2.50±0.13	2.2±0.7[10]	4.0	0.4±0.5	70±20	31±2	48±3
369	GUCUUG	CGCAAGAUGU	0	N/C	N/D	2.3	0.16±0.07	21±3	39±4	28±3
378	GCACGC	UGGCGUGUUA	0	3.6±0.5	N/D	6.9	29±5	15±3	36±9	24±5
405	GGGGAA	AUUUCCCUAA	0	5.32±0.12	9.3±1.6	6.3	21±4	50±9	30±10	65±7
448	GGGGAU	CAAUCCCUGG	1	5.4±0.3	10.8±2.3	6.3	25±10	50±10	25±6	60±2
518	GUGGUG	UUCACCAUGG	0	2.8±0.7	N/D	5.8	4.4±1.6	28±6	19±8	42±0
551	GGCACC	AAGGUGCUGA	0	3.7±0.3	N/D	6.8	14.4±1.6	60±20	24±7	35±3

[1] Position of splice site uridine relative to the adenosine of the AUG translation start codon on CAT mRNA.
[2] The underlined sequence corresponds to the nucleotides that base pair to the IGS. Note that all IGS-target sequence helices contain a G:U pair at the splice site.
[3] Number of times that the splice site was found among 66 sequences obtained with the trans-tagging assay on CAT mRNA.
[4] Average and standard deviation over three to six window sizes (100, 200, 300, 400, 500, 600 nucleotides). N/C, ΔG_{bind} was not calculated by INTARNA with at least three window sizes, because the interaction was not among the strongest 10,000.
[5] Average and standard deviation over three independent reactions. N/D, product fractions could not be determined because product bands were not visible on autoradiograms.
[6] Average and standard deviation over three independent reactions.
[7] Ribozyme transcription efficiency relative to the amount of the ribozyme 97 transcribed under identical conditions.
[8] Fraction of ribozyme that lost its 3'-exon during 20 minutes of in vitro transcription at 30°C.
[9] Fraction of ribozyme that lost its 3'-exon during 4 hours of incubation under trans-splicing conditions in the absence of mRNA substrate.
[10] The sequences of subcloned RT-PCR products from reactions that targeted splice site 350 were consistent with trans-splicing primarily at splice site 307.

TABLE S2.

Energetic contributions to ΔG_{bind} for the splice sites tested on CAT mRNA and on the short, 13-nucleotide substrates.

		CAT mRNA (kcal/mol) [1]				Short substrate (kcal/mol) [2]			
position[3]	IGS	$\Delta G_{unfold\text{-}target}$	$\Delta G_{release\text{-}IGS}$[4]	$-\Delta G_{hybrid}$	$-\Delta G_{bind}$	$\Delta G_{unfold\text{-}target}$	$\Delta G_{release\text{-}IGS}$	$-\Delta G_{hybrid}$	$-\Delta G_{bind}$
83	GCUGAC	1.3±0.6	2.4	6.9±0.0	3.3±0.6	0.00	2.4	7.4	5.0
87	GGCAAC	2.9±0.5	2.2	6.22±0.04	0.9±0.7	0.06	2.2	6.8	4.5
97	GGGUAC	1.6±0.3	1.1	7.2±0.0	4.5±0.3	0.39	1.1	7.2	5.7
131	GAGGCC	N/C[5]	N/C	N/C	N/C	0.06	7.1	9.5	2.3
197	GGGCGG	1.0±0.8	3.5	9.0±1.0	4.2±0.5	0.14	3.5	10.6	7.0
222	GCGGAA	2.3±0.1	2.3	7.2±0.0	2.6±0.1	0.12	2.3	7.2	4.8
240	GCCGUC	N/C	N/C	N/C	N/C	0.00	5.0	8.9	3.9
258	GUCCCA	0.3±0.1	1.8	8.2±0.0	6.1±0.1	0.00	1.8	8.2	6.4
273	GCAAGG	1.9±0.2	1.6	6.4±0.0	2.9±0.2	0.02	1.6	7.0	5.4
325	GUUCAC	1.3±0.1	0.4	5.1±0.0	3.4±0.1	0.01	0.4	5.1	4.7
346	GCUGCC	2.4±0.4	3.4	8.9±0.0	3.1±0.4	0.00	3.4	9.5	6.1
350	GGAAAC	1.2±0.2	0.5	4.2±0.2	2.5±0.1	0.04	0.5	4.6	4.0
369	GUCUUG	N/C	N/C	N/C	N/C	0.00	2.7	5.0	2.3
378	GCACGC	2.7±0.5	1.8	8.1±0.0	3.6±0.5	0.01	1.8	8.7	6.9
405	GGGGAA	1.1±0.1	1.3	7.7±0.0	5.3±0.1	0.17	1.3	7.7	6.3
448	GGGGAU	1.1±0.3	1.3	7.8±0.0	5.4±0.3	0.18	1.3	7.8	6.3
518	GUGGUG	3.0±1.0	1.3	6.9±0.3	2.8±0.7	0.11	1.3	7.2	5.8
551	GGCACC	2.9±0.5	2.3	8.9±0.3	3.7±0.3	0.05	2.3	9.2	6.8

[1] Average and standard deviation over three to six window sizes (100, 200, 300, 400, 500, 600 nucleotides).

[2] The computations for the short substrates used a single window including the full substrate sequence. Hence, no standard deviations are reported for these ΔG values.

[3] Position of splice site uridine relative to the adenosine of the AUG translation start codon on CAT mRNA.

[4] The computation of $\Delta G_{release\text{-}IGS}$ did not make use of a window because the substrate was not involved. Hence, no standard deviations are reported for $\Delta G_{release\text{-}IGS}$.

[5] N/C, ΔG was not calculated by INTARNA for at least three window sizes, because the interaction was not among the strongest 10,000 queried from the program.

Figure S1

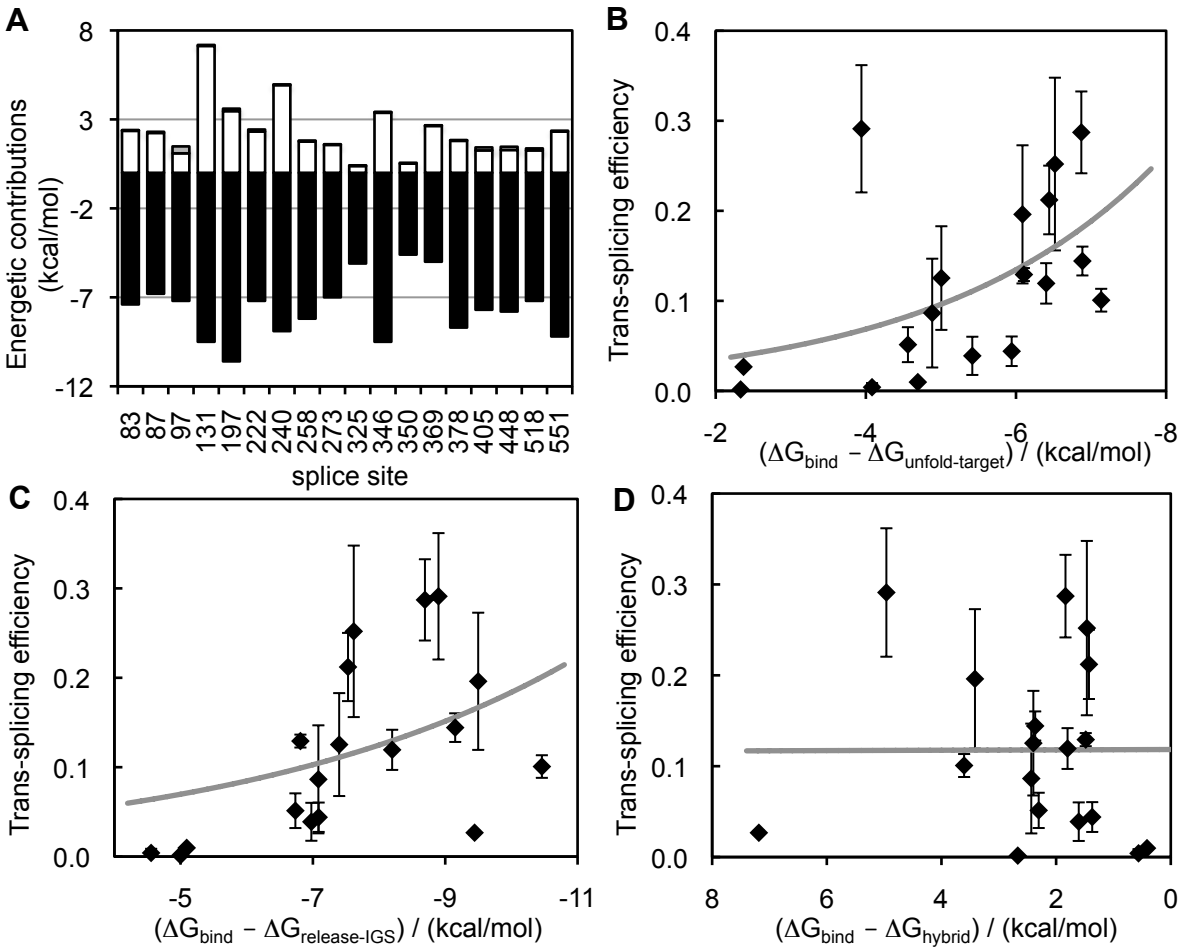

Figure S1. Energetic contributions to ΔG_{bind} from the three molecular events for the short,13-nucleotide substrates. *(A)* The computed energetic contributions of substrate target site unfolding ($\Delta G_{unfold\text{-}target}$, grey), ribozyme IGS release ($\Delta G_{release\text{-}IGS}$, white), and IGS-target site hybridization (ΔG_{hybrid}, black), are shown as a function of the splice sites. *(B)* Correlation of experimentally determined trans-splicing efficiency with ΔG_{bind} when $\Delta G_{unfold\text{-}target}$ is omitted. The thick gray line represents a least-mean-squares exponential fit with coefficient of determination $R^2 = 0.26$. An alternative linear fit (not shown) yields a correlation coefficient of $R = -0.51$, with a probability $p = 0.031$ that the values are not correlated. *(C)* Correlation when $\Delta G_{release\text{-}IGS}$ is omitted ($R^2 = 0.21$; $R = -0.54$; $p = 0.020$). *(D)* Correlation when ΔG_{hybrid} is omitted ($R^2 = 0.0079$; $R = 0.11$; $p = 0.68$). For additional details see Figure 3.

Figure S2

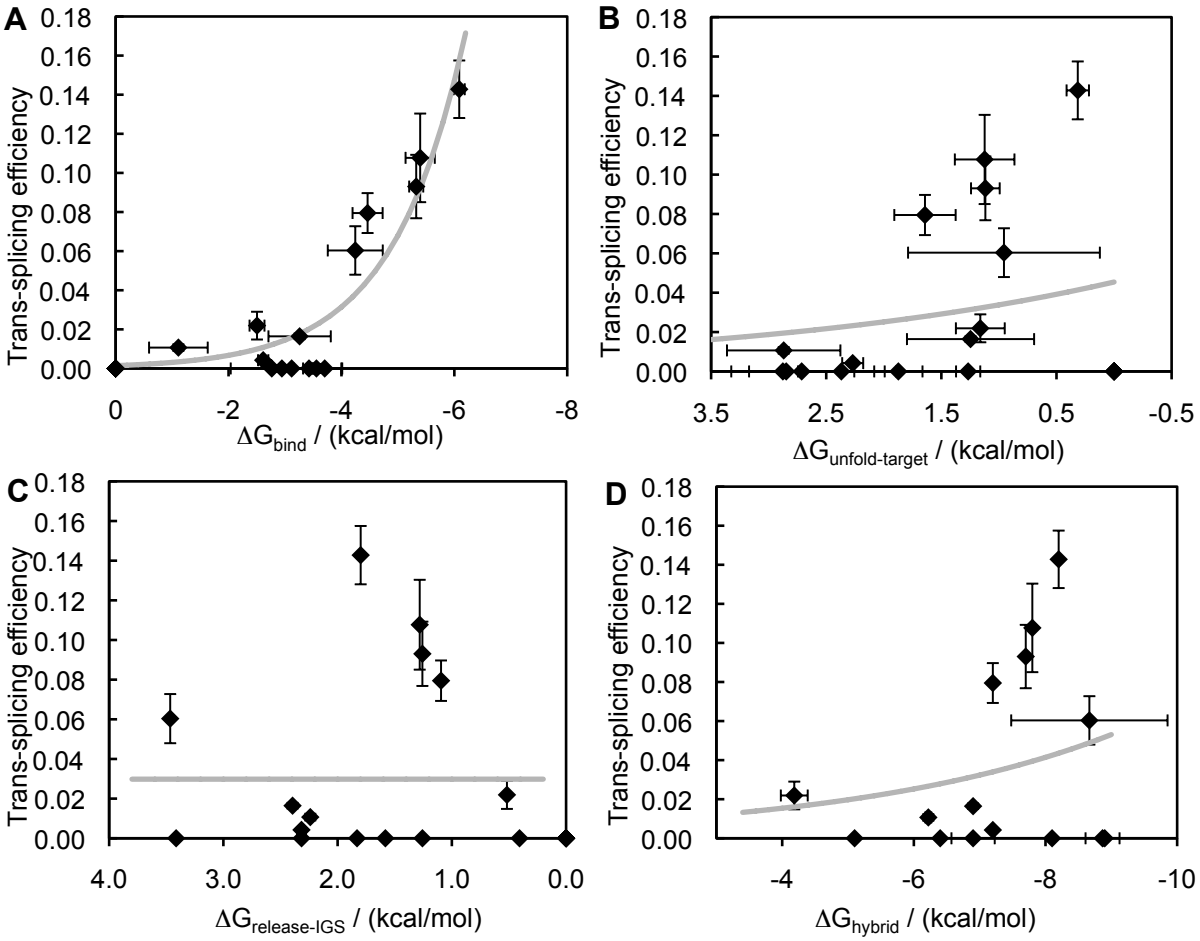

Figure S2. Correlation between trans-splicing efficiencies experimentally determined with CAT mRNA substrate and the three individual energetic contributions to ΔG_{bind}. *(A)* The correlation of trans-splicing efficiency with ΔG_{bind} serves as comparison. This is the same plot shown in Figure 3B. The thick gray line is a least-mean-squares exponential fit with coefficient of determination $R^2 = 0.87$. An alternative linear fit (not shown) yields a correlation coefficient of $R = -0.75$, with a probability $p = 0.00033$ that the values are not correlated. *(B)* Correlation of trans-splicing efficiency with $\Delta G_{unfold\text{-}target}$ ($R^2 = 0.064$; $R = -0.32$; $p = 0.20$). *(C)* Correlation of trans-splicing efficiency with $\Delta G_{release\text{-}IGS}$ ($R^2 = 0.0065$; $R = 0.099$; $p = 0.70$). *(D)* Correlation of trans-splicing efficiency with ΔG_{hybrid} ($R^2 = 0.14$; $R = -0.38$; $p = 0.12$). For additional details see Figure 3.