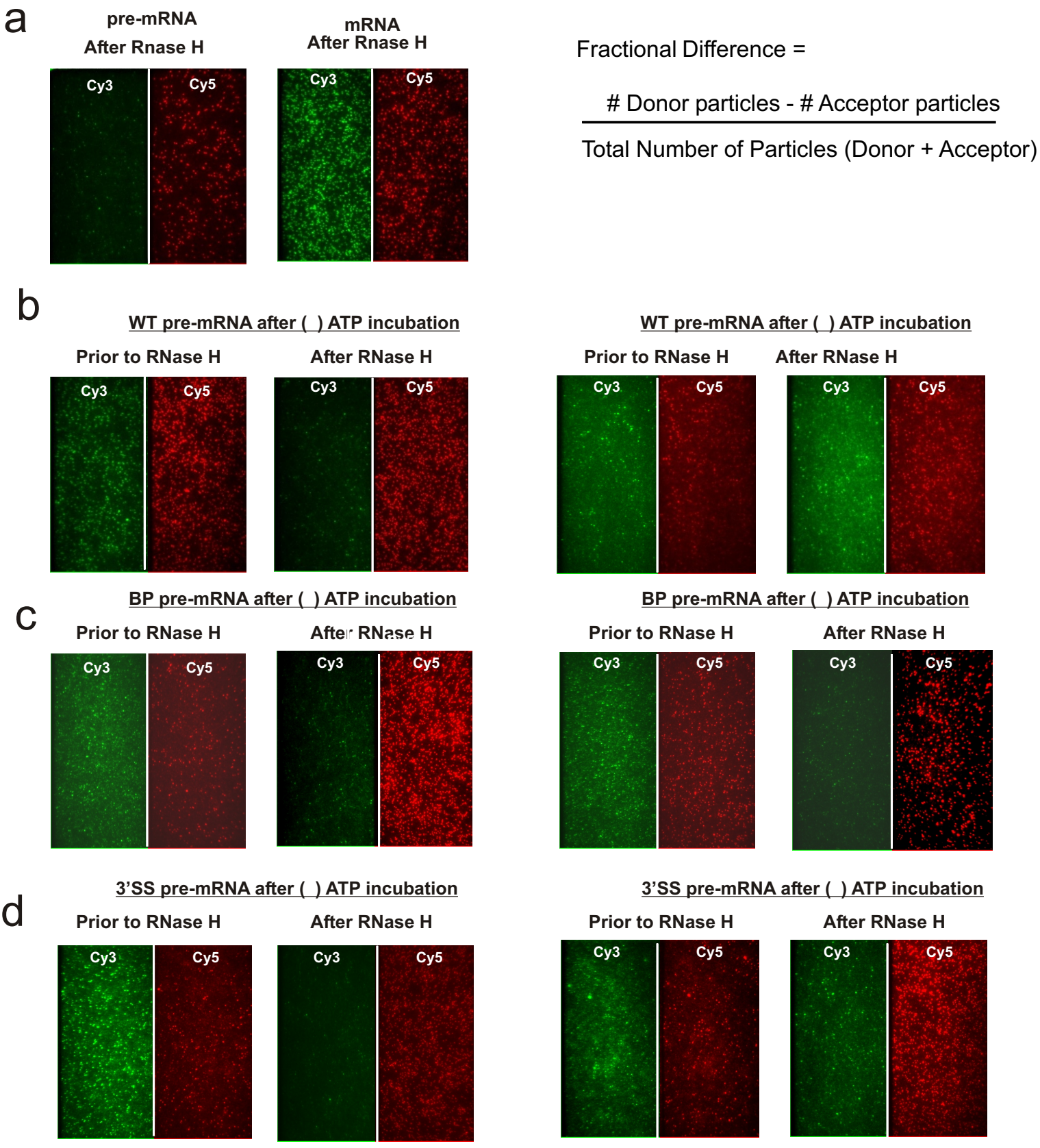
Supplementary Information for:

Conformational dynamics of single premRNA molecules during *in vitro* splicing

John Abelson[1,8], Mario Blanco[2,8], Mark A. Ditzler[3,4], Franklin Fuller[3,4], Pavithra Aravamudhan[3,4], Mona Wood[3], Tommaso Villa[1,5], Daniel E. Ryan[1,7], Jeffrey A. Pleiss[1,8], Corina Maeder[1], Christine Guthrie[1], Nils G. Walter[3]

[1]Department of Biochemistry and Biophysics, University of California, San Francisco, 600 16th Street, Genentech Hall, San Francisco, CA 941432200, USA. [2]Cellular and Molecular Biology, [3]Department of Chemistry, and [4]Biophysics, Single Molecule Analysis Group, University of Michigan, 930 N. University Ave., Ann Arbor, MI 481091055, USA. [5]Present address: University Pierre et Marie Curie, Paris, France, [6]Present address: Agilent Technologies, Santa Clara, CA, [7]Present address: Department of Molecular Biology and Genetics, Cornell University, Ithaca NY. [8]These authors contributed equally to this work. Correspondence should be addressed to J.A. (johnabelson@gmail.com) or N.G.W. (nwalter@umich.edu).



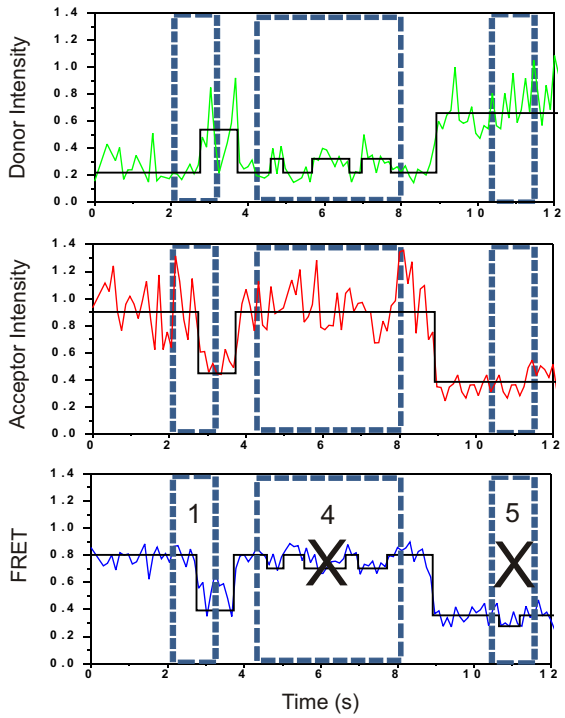
Supplementary Figure 1. *In situ* verification of splicing activity of immobilized substrates. **(A)** Control experiments with either pre mRNA or mRNA immobilized on surface and treatment with RNase H after incubation with complimentary DNA oligodRH.**(b)**RNase H results after incubation of WT pre-mRNA in ATP depleted extract or 1hr in ( ) ATP extract **(c)**RNase H results after incubation of BP pre mRNA in ATP depleted extract or 1hr in (+) ATP extract.**(d)**RNase H results after incubation of 3'SS pre mRNA in ATP depleted extract or 1hr in ( ) ATP extract

a

Transition Scoring Classification:

Score	Event
1	Single transition in both channels in interval
2	Multiple transitions in one channel, single transition in other channel in interval
3	Multiple Transitions in both channels in interval
4	Transitions in one Channel, but not in the other in interval
5	No transitions in either channel

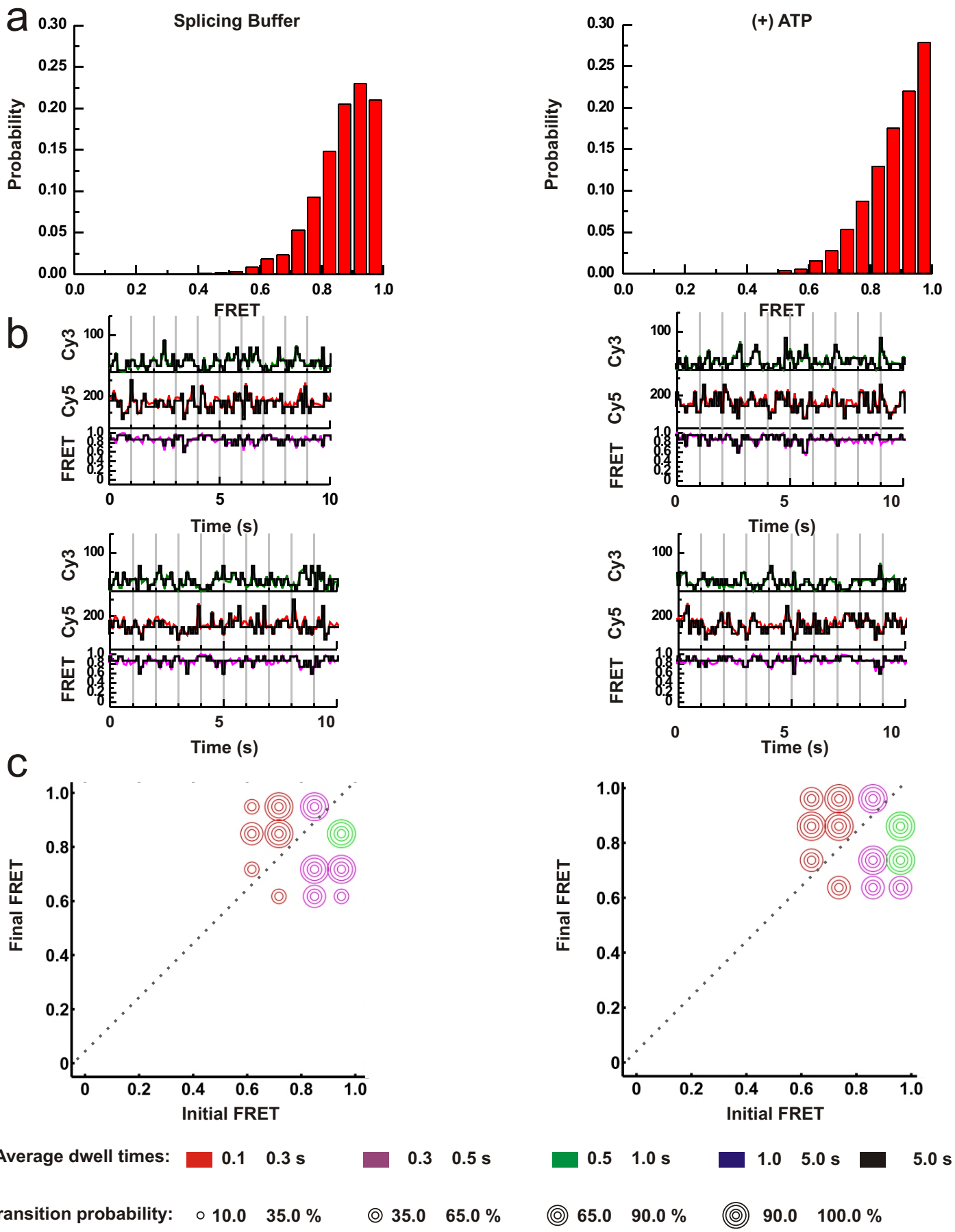
b



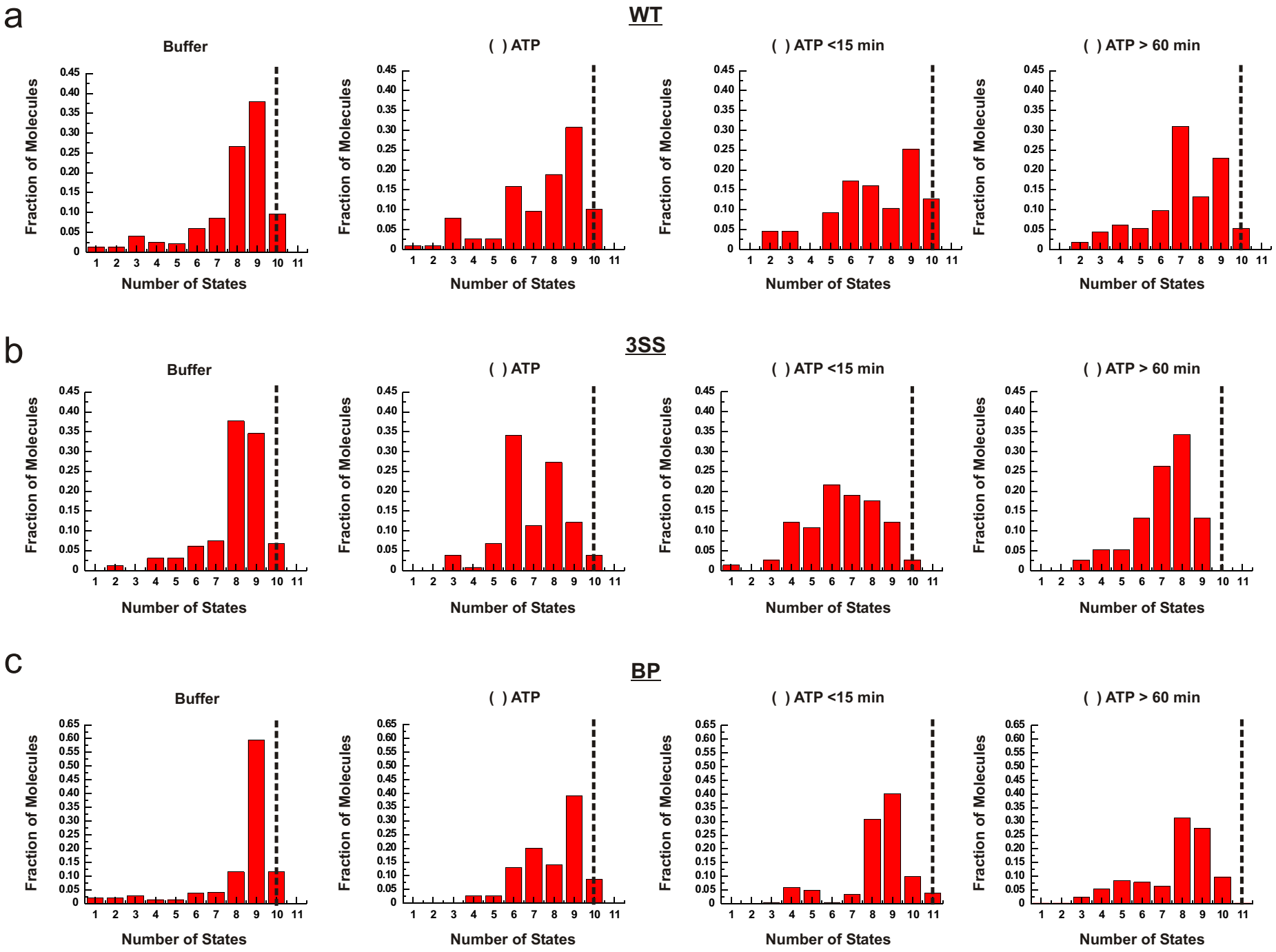
c

Substrate	Buffer (Percent of Scores1 -3/ Percent of Scores4 -5)	(-) ATP (Percent of Scores1 -3/ Percent of Scores4 -5)	(+) ATP < 15 min (Percent of Scores1 -3/ Percent of Scores4 -5)	(+) ATP > 60 min (Percent of Scores1 -3/ Percent of Scores4 -5)
WT	83.94 / 16.06	70.89 / 29.11	75.07 / 24.93	66.91 / 33.09
BP	82.89 / 17.11	68.99 / 31.01	63.50 / 36.5	64.65 / 35.35
3'SS	80.92 / 19.08	73.08 / 26.92	70.54 / 29.46	74.36 / 25.64

Supplementary Figure 2. Post-filtering analysis of Hidden Markov Modeling (HMM) data.**(a)** A logical scoring criteria based on the number of transitions found in the Donor, Acceptor, and FRET channels was used to identify transitions of interest. In this study, transitions that had a score of 1 3 were used for further analysis.**(b)** Hidden Markov fits (black) are shown for independently analyzed Donor, Acceptor, and FRET channels.Transitions in the FRET channel are scored (inset) based on the criteria in Fig. S2a. **(c)** The percent of transitions that were retained for analysis (Scores 1 3) or discarded (Scores 4 5) are displayed for each substrate and condition.

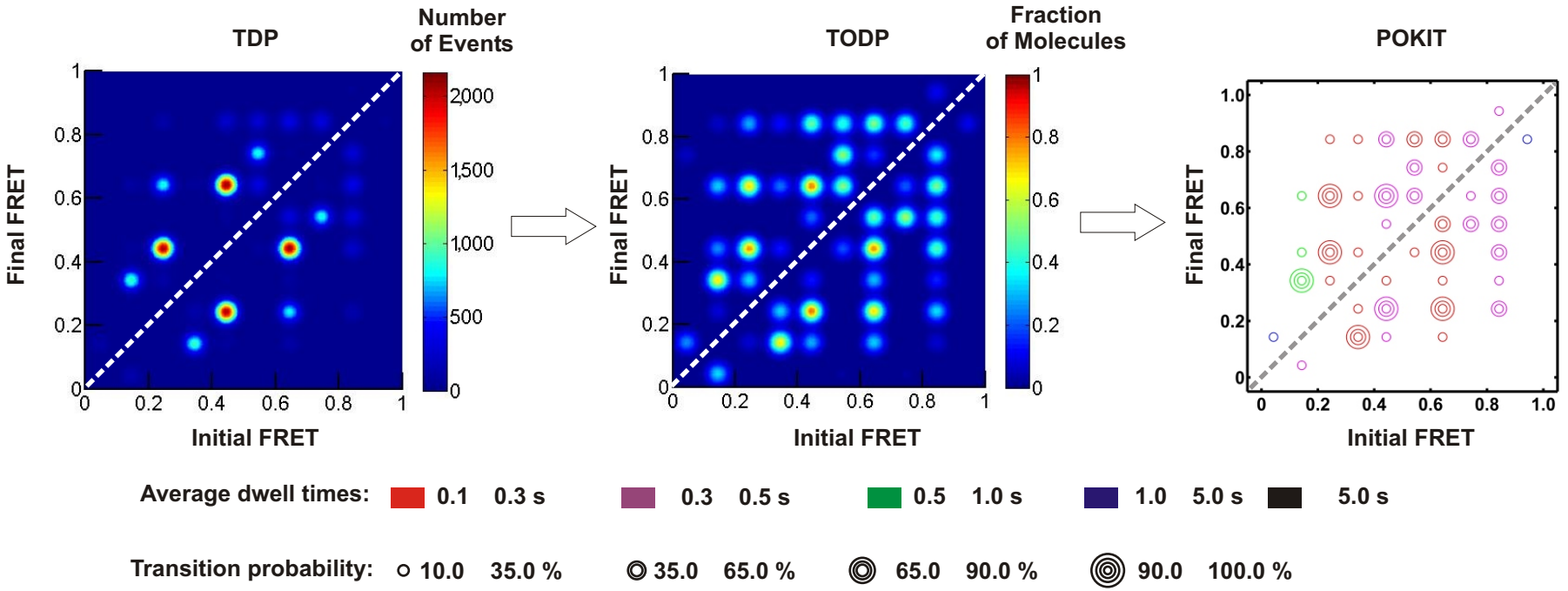


Supplementary Figure 3. Conformational dynamics of mRNA in splicing buffer and (+) ATP cell extract. **(a)** Histograms of enzymatically ligated exons in buffer and extract. **(b)** Sample traces of the same mRNA in both conditions. **(c)** POKIT plots of both conditions.



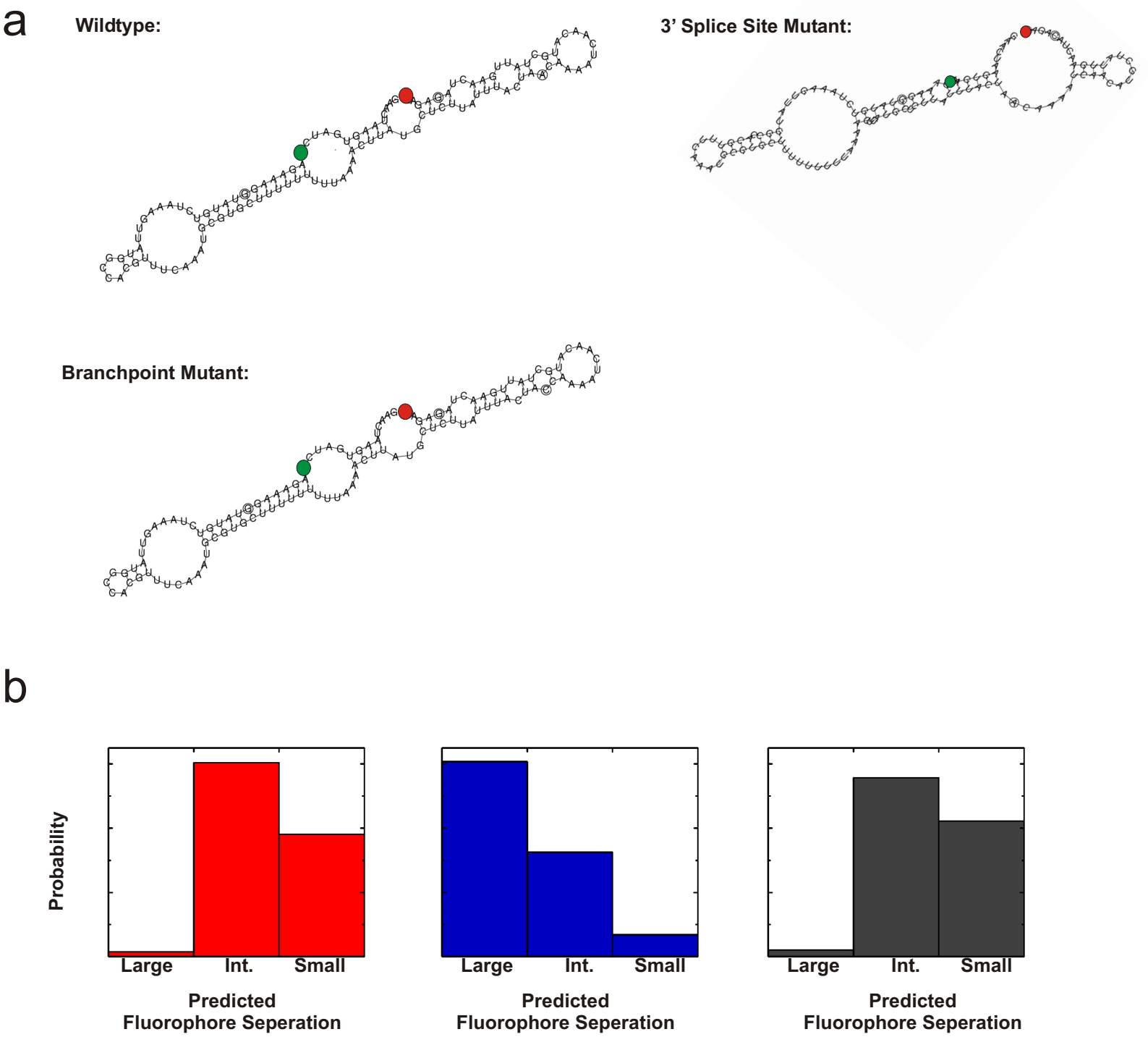
Supplementary Figure 4. Distribution of idealized FRET states. The number of idealized states for single molecules was binned, and plotted in histograms demonstrating the distribution of states per molecule for the WT**(a)**, 3'SS **(b)**, and BP **(c)** substrates in each experimental condition. Dashed lines indicate the number of states in the Markov model required to fit the entire data set.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.1767

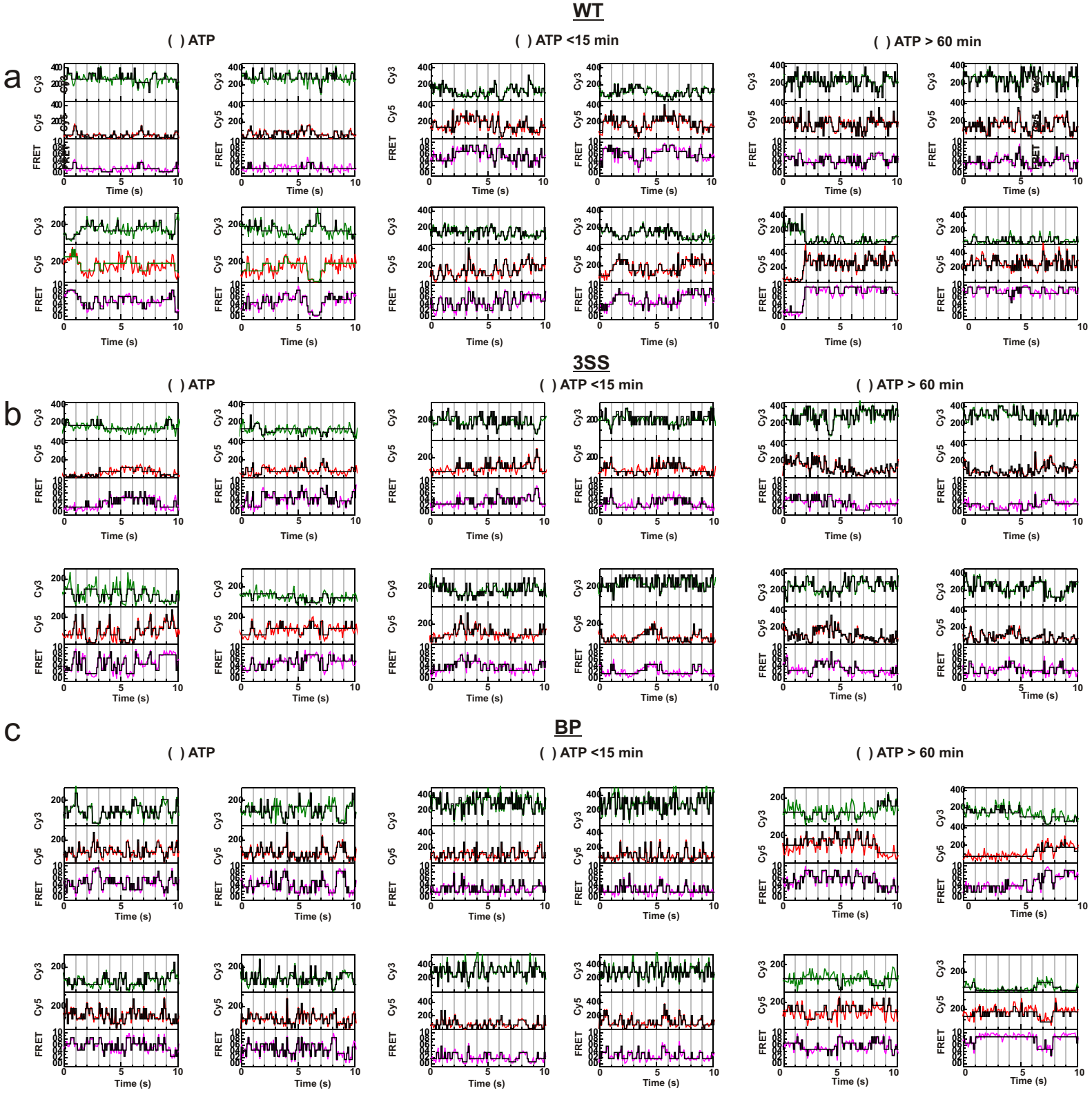


Supplementary Figure 5. Transition Density Plot (TDP), Transition Occupancy Density Plot (TODP), Population and Kinetically indexed Transtion Density Plot (POKIT). TDPs are weighted by the number of times a transition is seen, and therefore must be corrected for differences in kinetics due to an inherent bias against slower transitions. TODPs scale the data by the population of molecules exhibiting a particular transition. This allows for transitions that are equally probable to appear within a set of molecules to be weighted to the same degree. POKIT plots scale transitions based on their probability, but also display kinetic information of each transition.

Nature Structural & Molecular Biology: doi:10.1038/nsmb.1767



Supplementary Figure 6. Transient secondary structures place fluorophores within FRET range in buffer. **(a)**Predicted lowest free energy structure for WT, 3'SS mutant, and BP mutant substrates. The position of fluorophores is indicated by a green circle and red circle for Cy3 and Cy5, respectively. The sequence of exon-2 that is used for tethering to the slide surface was excluded from this analysis. **(b)**The distribution of predicted inter-fluorophore distances based on secondary structure analysis is show for each pre-mRNA sequence. Small, step size of 0-10. Intermediate (Int), step size of 10-20. Large, step size >20.



Supplementary Figure 7. Single pre-mRNA FRET Trajectories. Donor (Cy3), Acceptor (Cy5) and FRET trajectories of single pre-mRNA substrates in various extract conditions and their corresponding HMM fit (black). **(a)** WT substrate in (-) ATP, (+) ATP 15 min, and (+) ATP 60 min. **(b)** 3'SS mutant **(c)** BP mutant.

Rank	Systematic Name	Gene Name	ORF length	Intron Length
1	**YPL090C**	*RSP6A*	*1105*	*394*
2	**YBR082C**	*UBC4*	*542*	*95*
3	**YDR471W**	*RPL27B*	*795*	*384*
4	**YKL081W**	*TEF4*	*1565*	*326*
5	**YNL050C**	**Unknown**	**904**	**91**
6	**YLR061W**	*RPL22A*	*755*	*389*
7	**YDR424C**	*DYN2*	*455*	*96*
8	**YBL040C**	*ERD2*	*757*	*97*
9	**YML056C**	*IMD4*	*1983*	*408*
10	**YPL143W**	*RPL33A*	*849*	*525*
11	**YAL030W**	*SNC1*	*467*	*113*
12	**YBR048W**	*RPS11B*	*982*	*511*
13	**YPL081W**	*RPS9A*	*1095*	*501*
14	**YBL059C**?A	*CMC2*	*415*	*85*
15	**YBR181C**	*RPS6B*	*1063*	*352*
16	**YKL186C**	*MTR2*	*792*	*154*
17	**YHR101C**	*BIG1*	*1095*	*87*
18	**YLR344W**	*RPL26A*	*831*	*447*
19	**YER133W**	*GLC7*	*1464*	*525*
20	**YFL039C**	*ACT1*	*1436*	*308*

Supplementary Table 1. Relative splicing efficiency of yeast pre-mRNAs *in vitro.* Shown here are the top 20 pre-mRNAs (out of about 250 genes in yeast that contain introns) in descending order of splicing efficiency *in vitro* as assayed in a microarray assay. *UBC4* was selected since it ranked among the highest in splicing efficiency while maintaining an intron size of less than 100 nucleotides.

Substrate	Sequence
UBC4(20/20) Wildtype (WT)	5'-GAACUAAGUGAUC**(5-N-U)**AGAAAG*GUAUGUCUAAAGUUAUGGCCACGUUUCAAAUGCGUGCUUUUUUUUU AAAACUUAUGCUCUUAUUUACUA**A**CAAAAUCAACAUGCUAUUGAACUA**G***AGA**(5-N-U)**UCCACCUACUUCAUGUU- 3'
UBC4(20/20) Branch site mutant (BP)	5'- GAACUAAGUGAUC**(5-N-U)**AGAAAG*GUAUGUCUAAAGUUAUGGCCACGUUUCAAAUGCGUGCUUUUUUUUU AAAACUUAUGCUCUUAUUUACUA**C**CAAAAUCAACAUGCUAUUGAACUA**G***AGA**(5-N-U)**UCCACCUACUUCAUGUU- 3'
UBC4(20/20) 3'Splice site (3'SS) mutant	5'-GAACUAAGUGAUC**(5-N-U)**AGAAAG*GUAUGUCUAAAGUUAUGGCCACGUUUCAAAUGCGUGCUUUUUUUUU AAAACUUAUGCUCUUAUUUACUA**A**CAAAAUCAACAUGCUAUUGAACUA**C***AGA**(5-N-U)**UCCACCUACUUCAUGUU- 3'
DNA splint- dSplint	5'-GTTGATTTTGTTAGTA AATAAG(SP9)GTTTTAAAAAAAAAGCACGC -3
RNaseH oligo - dRH	5'-GCATGTTGATTTTGTTAGTAAATAAGAGCA -3'

Supplementary Table 2. Sequence information of oligonucleotides used in this study. UBC4 intron is italicized, and allyl-amine modified uridines are denoted as (5-N-U). In the 3'SS mutant, the bold underlined guanine was replaced with a cytosine. In the BP mutant the italicized underlined adenosine is replaced by a cytosine. (Sp9) denotes a 9 carbon linker.

Supplementary Methods:

HMM analysis - Model Selection and Scoring Regime. Multiple models with differing state numbers were used to determine the underlying FRET states; the entire data set for each condition was analyzed by the iterative application of the Viterbi and BaumWelch algorithms to generate idealized trajectories. The number of states assumed in the idealization was varied from 5 to 11 and the corresponding fits were evaluated using the Bayesian information criterion (BIC). The number of states that resulted in the best BIC score was used in our analysis. BIC penalizes models with extra states that do not result in a significant improvement in the LogLikelihood; this allows us to select the most appropriate model of our data analysis by balancing goodness of fit and model parsimony. After idealization, the postfilter algorithm classified each FRET transition by counting the number and direction of transitions found in the donor and acceptor trajectories, within one quarter of the dwell time preceding the FRET index transition and one quarter of the dwell time after, or a minimum of 0.3 seconds in either direction. Logical classification was performed, scoring each transition based on the metric shown in **Supplementary Figure 2**. In this work, transitions with scores of three or lower were used for further analysis with transition density plots. Additionally FRET transitions with a step size smaller than 0.1 were not included in our analysis.

Secondary Structure Prediction of Ubc4 pre-mRNAs. The lowest-energy secondary structures for each substrate was predicted using the by the Vienna software package(available at http://rna.tbi.univie.ac.at/). This software also calculates a partition function of secondary structures.The software suboptimal structures that were within 5% of the Minimum Free Energy (MFE) for the WT and the two mutant pre-mRNAs. The suboptimal structures thus generated by the program represent a Boltzmann weighted population. The suboptimal structures thus produced were analyzed for the secondary structure distance between the fluorophores using the program developed by Rogic *et.al.,* with minor modifications to enable one to specify the fluorophore positions in the pre-mRNA sequence at the command prompt.

In vitro splicing microarray. RNA was extracted from *prp2-1* grown at the permissive temperature (2-1 RNA) and from *in vitro* splicing reactions in which the premRNA added to the reaction mixture was RNA isolated from *prp2-1* grown at the permissive temperature and then shifted to the nonpermissive temperature for 30 min. Splicing reactions containing 240 mg ml^{-1} of the RNA extracted from cells grown at the nonpermissive temperature were incubated for 30 min at room temperature with (+ATP RNA) and without ATP (- ATP RNA). cDNA synthesized from the 21 control RNA was labeled with Cy3 and and cDNA from +ATP RNA and ATP RNA with Cy5 as described in Pleiss et al. Splicing microarrays were hybridized with a mixture of 2-1 cDNA and +ATP cDNA or 2-1 cDNA and - ATP cDNA. After hybridization the microarrays were washed and analyzed for the ratio of +/- ATP to 2-1. For this analysis we only considered hybridization to the set of oligonucleotides specific for the mRNAs of genes containing introns. The order of the genes in Table 1 was determined by subtracting the rank orders of +ATP/2-1 minus -ATP/2-1 for each gene (for example Ubc4 was in the 64th percentile in the (+)ATP ratios and in the 0.8th percentile in the (-)ATP ratios. Ubc4 prem-RNA is second in this list and the canonical substrate for yeast *in vitro* pre-mRNA splicing, actin pre mRNA is 20^{th}.

Supplementary References:

Pleiss, J.A., Whitworth, G.B., Bergkessel, M. & Guthrie, C. Transcript specificity in yeast premRNA splicing revealed by mutations in core spliceosomal components. *PLoS Biol.* **5**, e90 (2007).

Rogic, S. et al. Correlation between the secondary structure of pre-mRNA introns and the efficiency of splicing in Saccharomyces cerevisiae. *BMC Genomics* 9, 355 (2008).