

## Supplemental Information

### Three-Dimensional RNA Structure

#### of the Major HIV-1 Packaging Signal Region

James D. Stephenson, Haitao Li, Julia C. Kenyon, Martyn Symmons, Dave Klenerman, and Andrew M.L. Lever

#### Inventory of Supplemental Information

**Figure S1 - Ribonuclease digestion and reverse transcriptase pausing assays to assess RNA 2D structure and PNA effect relates to Figure 3.**

**Figure S2 - Binding location and anisotropy measurements of annealed fluorophores relates to Table 1.**

**Figure S3 - FRET histograms showing the frequency of each energy transfer relates to figure 1.**

**Figure S4 - Searching global and local structure space during simulated annealing relates to figure 4.**

**Figure S5 Average model variability following simulated annealing and sequence conservation related to figure 5.**

**Figure S6 A Kink turn and other features of the modelled RNA relates to figure 6.**

**Table S1 SHAPE reactivity values from probing of the HIV-1 5' untranslated region RNA. relates to figure 3**

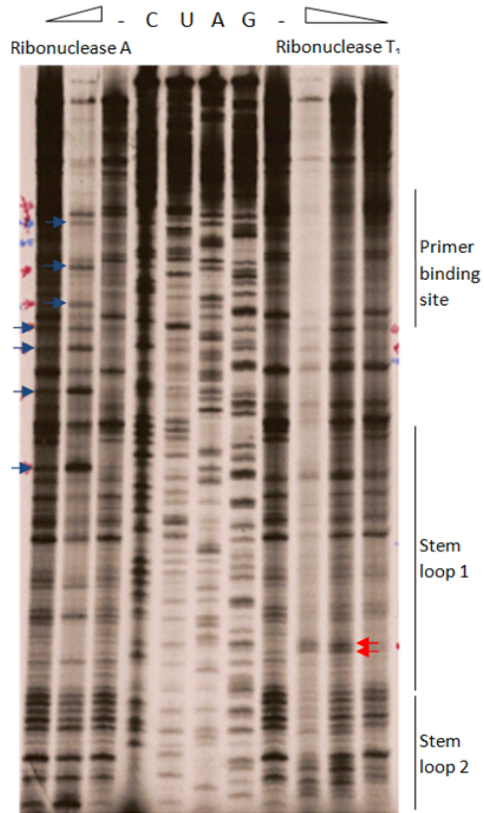
**Table S2 Calculation of weighted efficiencies from signal:noise ratio relates to Table 1**

**Table S3 Table of the difference between the modelled and measured distances for each fluorophore pair from starting models 0-9 relates to Tables 3 and 4**

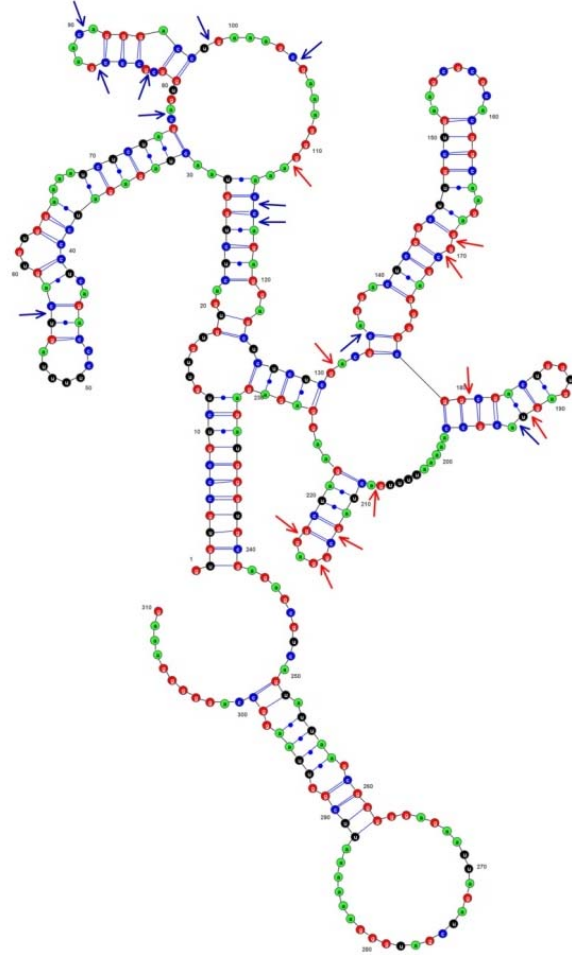
#### Supplemental Experimental Procedures

## Supplemental Data

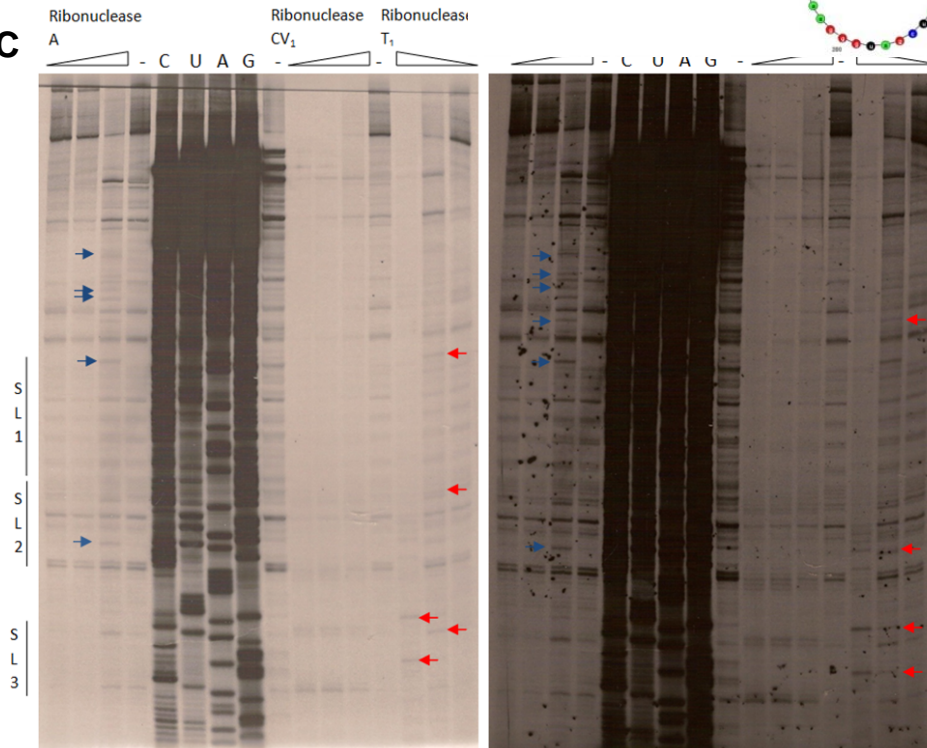
**A**



**B**

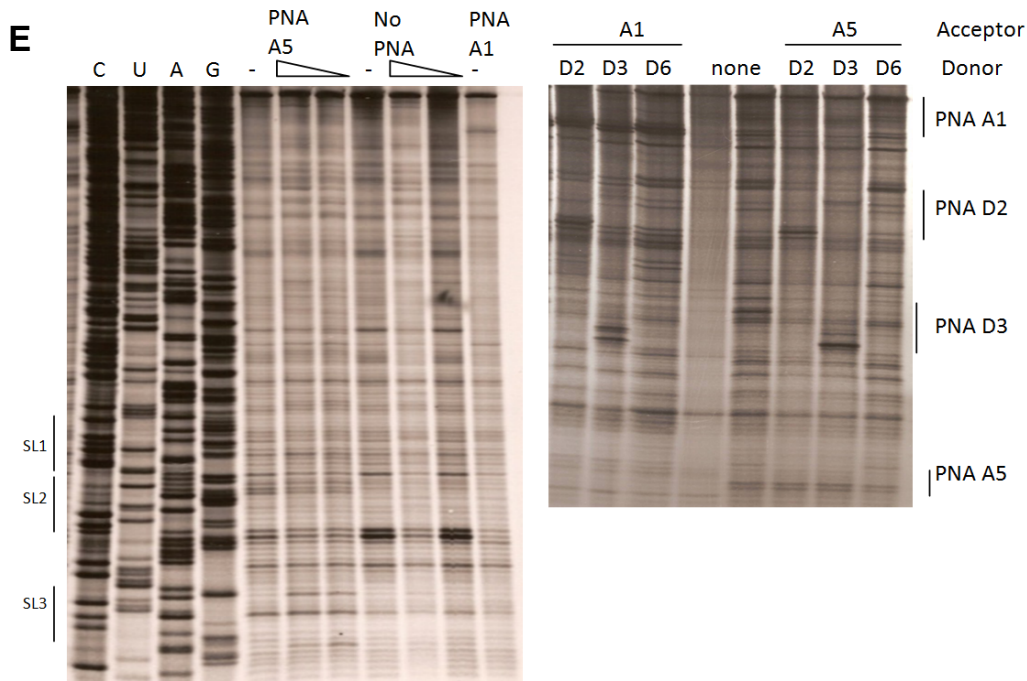


**C**



**D**

PNA	Energy kcal/mol	Hybridization energy kcal/mol	Unfolding PNA kcal/mol	Unfolding RNA kcal/mol	UTR position
A1	-12.6	-20.4	0.1	7.7	116-128
D2	-13.5	-16.9	0.1	3.3	166-177
D3/A4	-17.3	-22.0	0.3	4.4	197-209
A5	-12.3	-16.9	0.1	4.5	286-296
D6	-8.4	-10.2	0	1.8	352-361

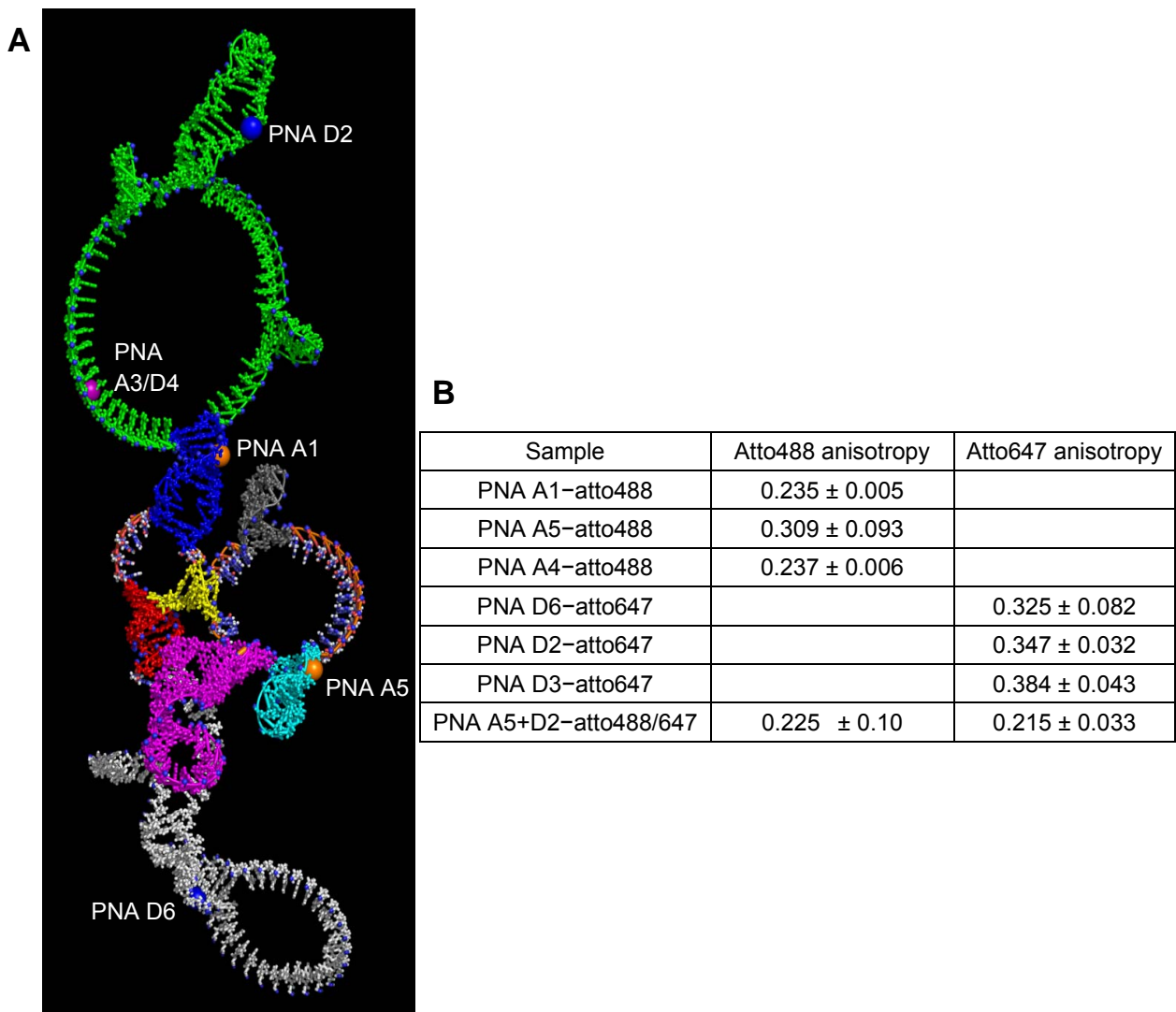


**Figure S1. Ribonuclease digestion and reverse transcriptase pausing assays to assess RNA 2D structure and PNA effect.**

(A),(C) Biochemical probing (for method see (Harrison et al., 1998a)) using RNase A and T1 showing unpaired regions mapped onto the consensus 2D structure (B)(see also Figure 3). Representative gels shown. Actual data analysed at different exposures to align sequencing ladder accurately with probed lanes. Pause sites of sequence 104-413 were not influenced by the 345-413 segment.

Background pausing retains uniformity and is consistent with pausing in other gels from this work which confirm lack of alteration of other pause sites which have been shown to represent structural and sequence motifs within the RNA (for details see (Harrison et al., 1998a)).

(D) The Energy column shows the overall energy change after the unfolding of the PNA and RNA is added to the hybridization energy of the PNA to the RNA (Smith et al., 2010). The scoring is based on hybridization free energy and accessibility of the interaction sites in both molecules. Ensemble free energy calculation is realized via the Vienna RNA library (Hofacker et al., 1994). The position of each PNA on the RNA is also shown and can be correlated to the PNA positions on figure 3. (E) PNA binding detected as an additional pause site at the 3' end of the binding site on the RNA. Background pausing retains uniformity when PNA is added which confirms lack of alteration at structural and sequence motifs within the RNA (for details see (Harrison et al., 1998a)). The RNA 2D structure can be seen in figure 3.



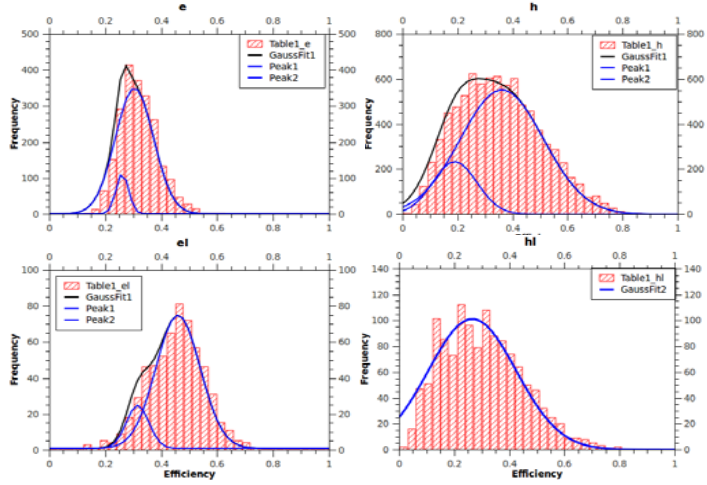
**Figure S2. Binding location and anisotropy measurements of annealed fluorophores.**

(A) 3D representation of the 2D consensus model for the HIV-1 5' untranslated region produced using RNA2D3D (Martinez et al., 2008) and displayed in Pymol. The regions are coloured as in Figure 3 and the fluorophore locations are displayed as spheres on the C1' atoms of the most proximal nucleotide to the fluorophore. The actual experimental PNAs locations are shown in figure 3.

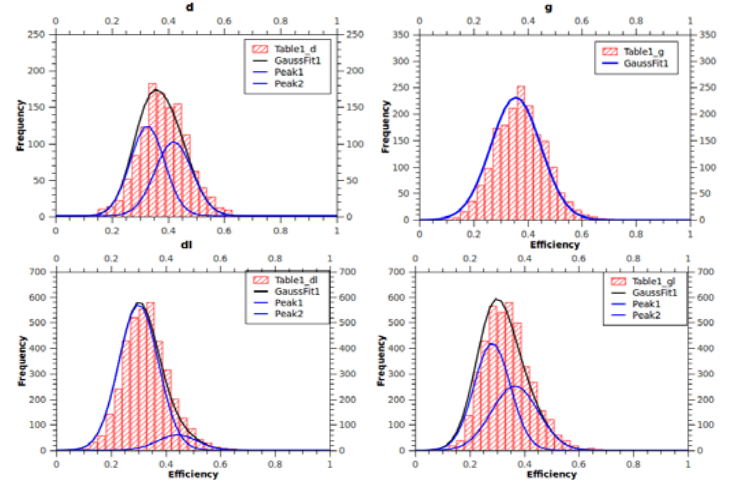
(B) Rotational freedom of the fluorophores was measured by anisotropy experiments, which uncovered the fact that neither the donor or acceptor fluorophores rotate freely probably due to both the linker and having some attraction to the RNA (this is the simplest explanation but not only possible explanation for the increased anisotropy). The orientation factor for the most extreme case (the donor) was corrected. We calculated the orientation of both fluorophore dipoles and hence the dipole orientation factor so as to determine the correct Förster distance to convert the measured FRET efficiencies in donor-acceptor separation. If we assume that the angle of acceptor dipole is 0° because this dye is stuck on the RNA strand then the angle of the donor is 24°.

The  $\Delta$  factor was calculated to be 1.31 (Supplemental experimental procedure). The Förster distance was therefore corrected by multiplying by the  $\Delta$  factor which resulted in a change to 6.68nm from the conventional 5.1nm. Distances including anisotropy correction can be seen in Table 1.

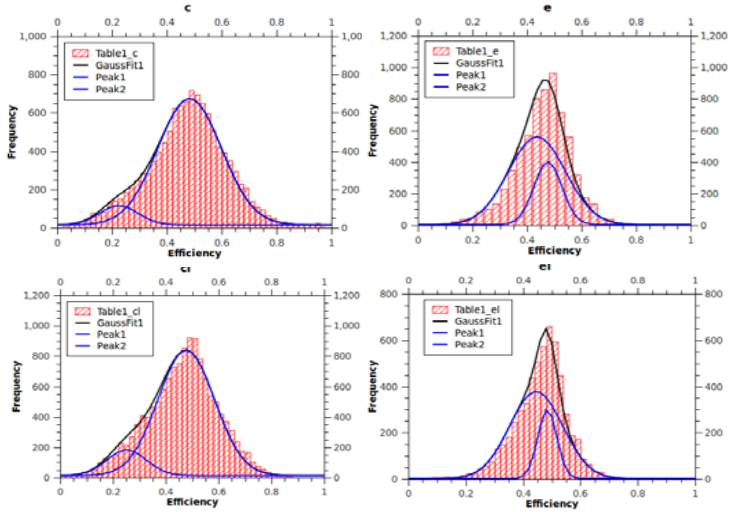
PNA A1-D6



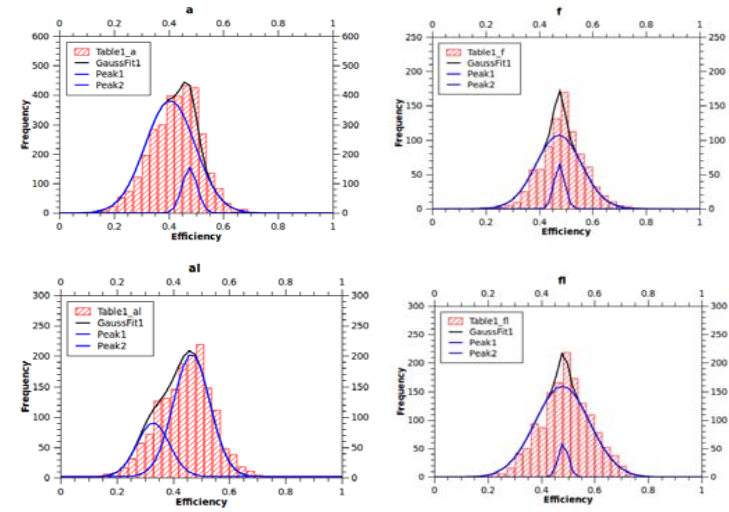
PNA A1-D2



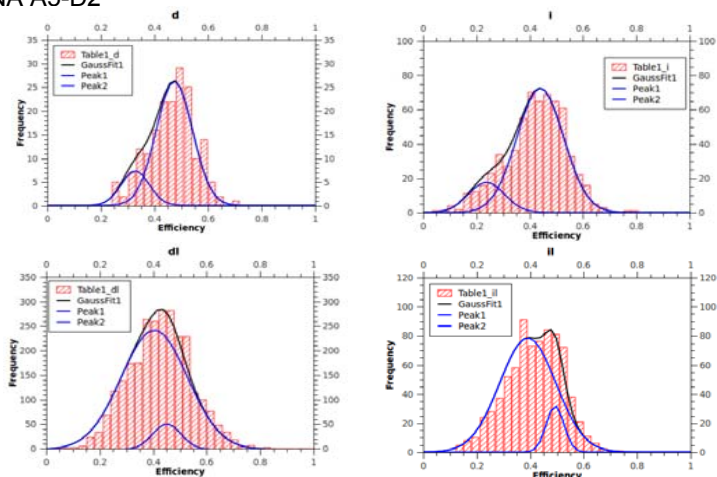
PNA A1-D3



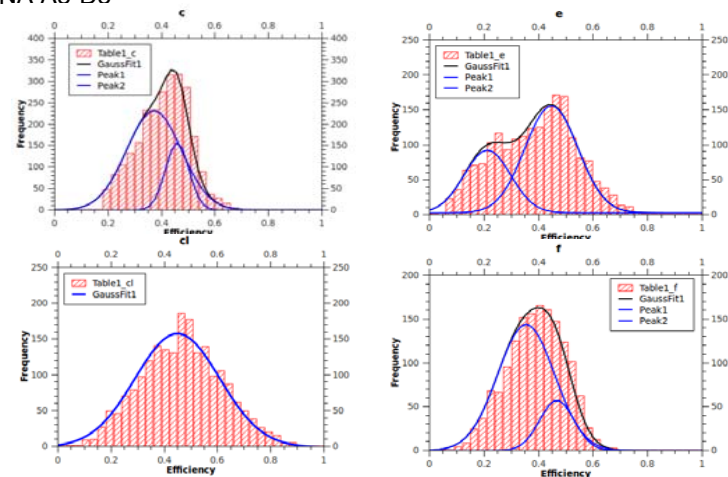
PNA A5-D6



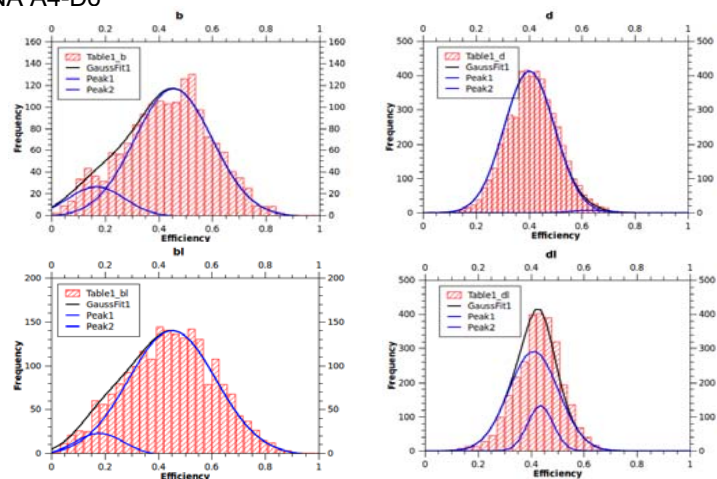
PNA A5-D2



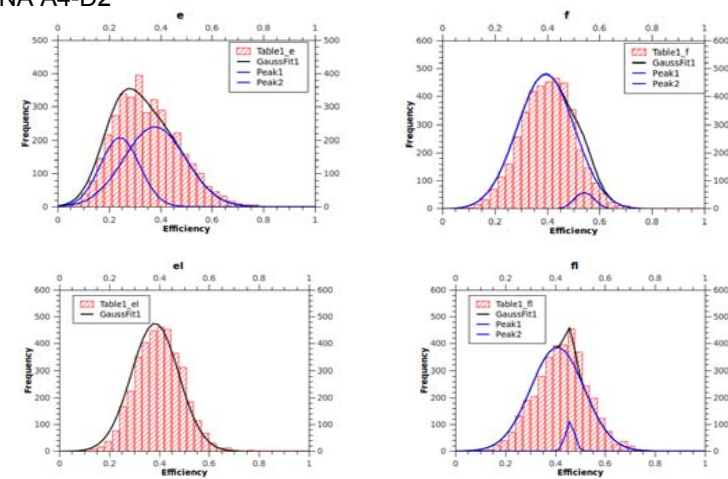
PNA A5-D3



PNA A4-D6

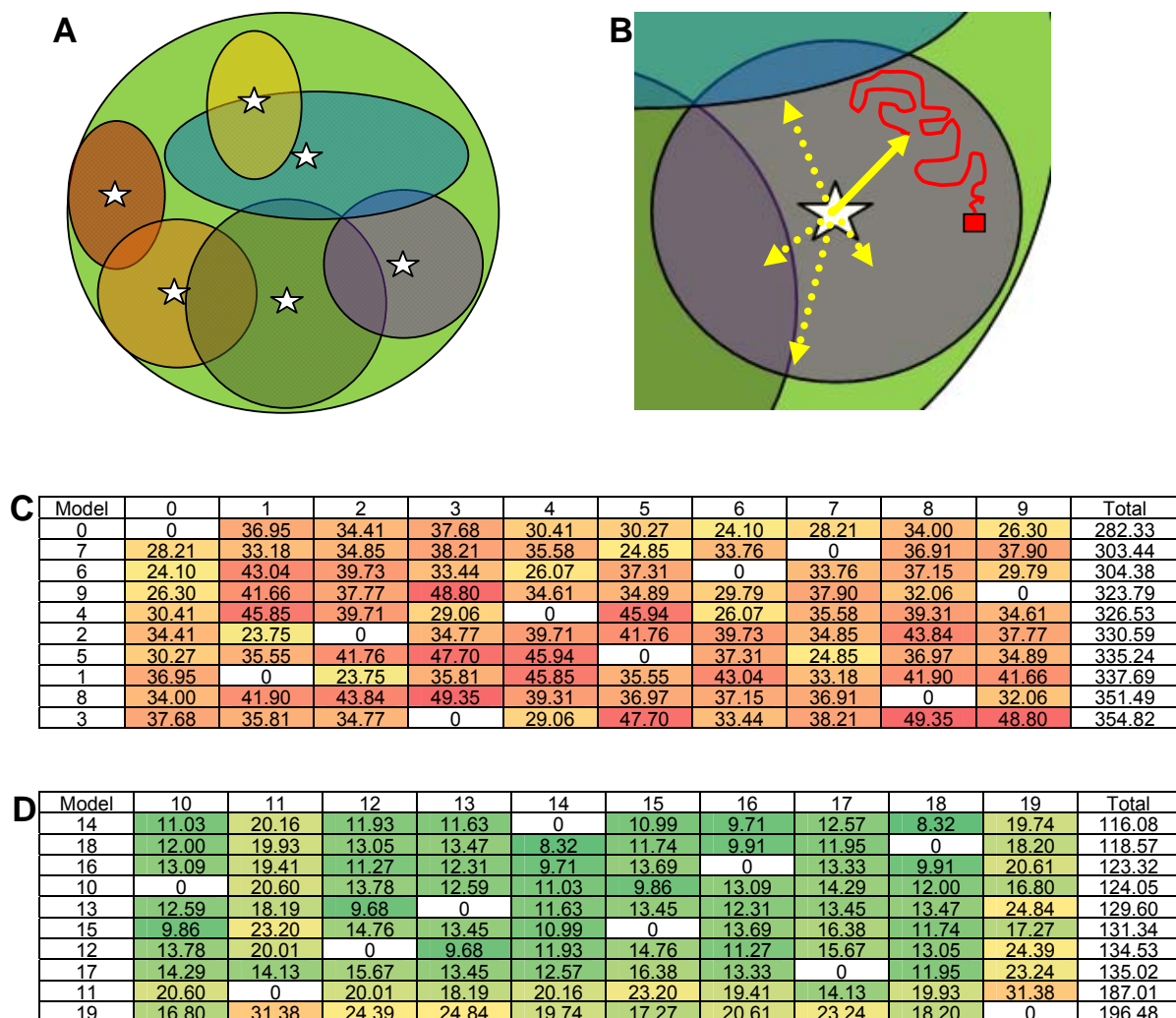


PNA A4-D2



**Figure S3. FRET histograms showing the frequency of each energy transfer.**

The blue lines show the Gaussian curve predictions from open source QTI plot software and the black curve is an addition of the Gaussians. Cumulative to RNA flexibility some of the peak width observed was caused by photon shot noise and the flexible linker between PNA and fluorophore resulting in the distances measured essentially being the intersection of spheres of size 1.3nm. The linkers negated the need to incorporate fluorophores into the RNA directly which allowed natural dynamicism of the RNA during FRET measurements. The above data are examples from each pair entitled by their one or two letter experimental references. The numerical data for these examples are found in Table S2. See also figure 1. The full data set is available upon request.

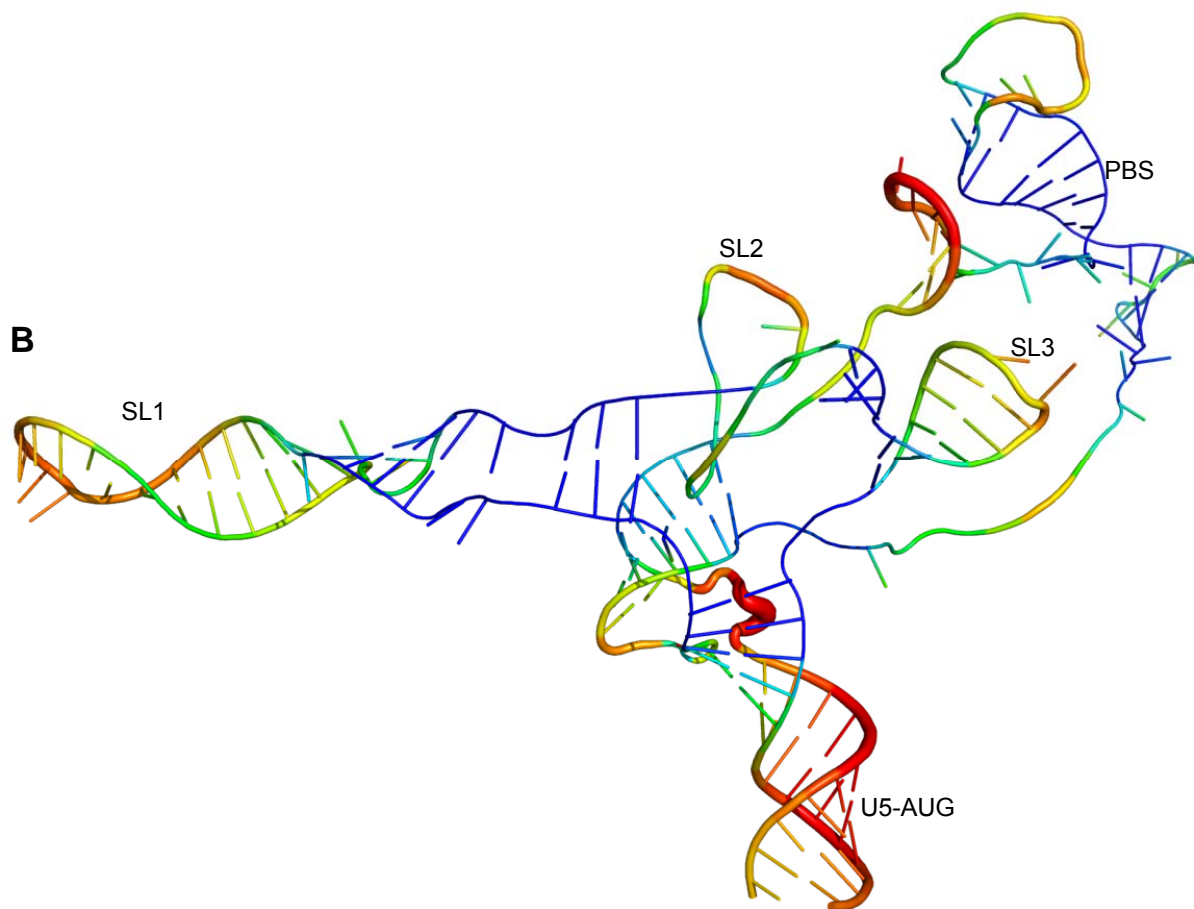


**Figure S4. Searching global and local structure space during simulated annealing.**

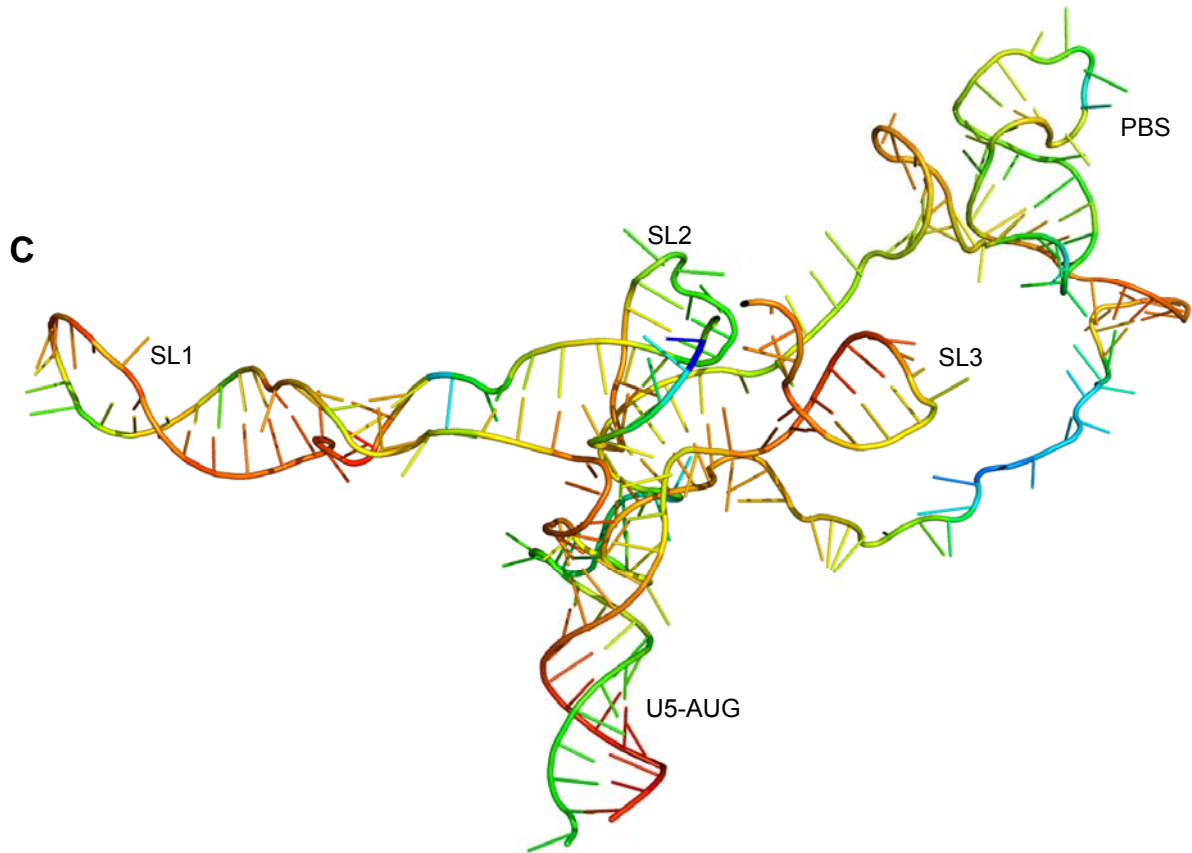
(A) The large green circle represents the total possible structure space for the RNA given the 2D constraints. The stars represent the starting models and their surrounding ovoids represent the searchable structure space available from the starting point due to inherent steric /hindrance. Therefore the more starting points the higher the probability that all possible space is searched. If we assume that the optimum structure is close to the best we have found (model 1) then searching the model 1 structure space more thoroughly is likely to yield a model closer to the ideal. (B) The purple circle represents the searchable structure space from random starting model (r1). The solid yellow arrow indicates the random starting trajectory of the simulated annealing of our current model and the red arrow traces the energy minimization search path across the structure space landscape. The red box represents model 1. We searched the r1 space more thoroughly (as shown in Figure 4B) by assigning different random seeds in order to generate different starting trajectories from r1 and investigate more r1 space (dashed yellow arrows). Vertically ranked similarity matrices calculated as one minus the pairwise RMSD of each structure when aligned, derived from either (C) random starting models or (D) from the single random starting model "r1" displayed in Figure 5. The aligned structures are shown in Figure 4A/B, structure divergence dendrograms are displayed in Figure 4C/D.

**A**

Model excluding variable regions	RMSD between our working model and others from the same starting model
0	9.681
1	6.835
2	13.608
3	10.271
4	11.754
5	12.486
6	8.787
7	5.125
8	9.984
9	8.794
Average	9.7325





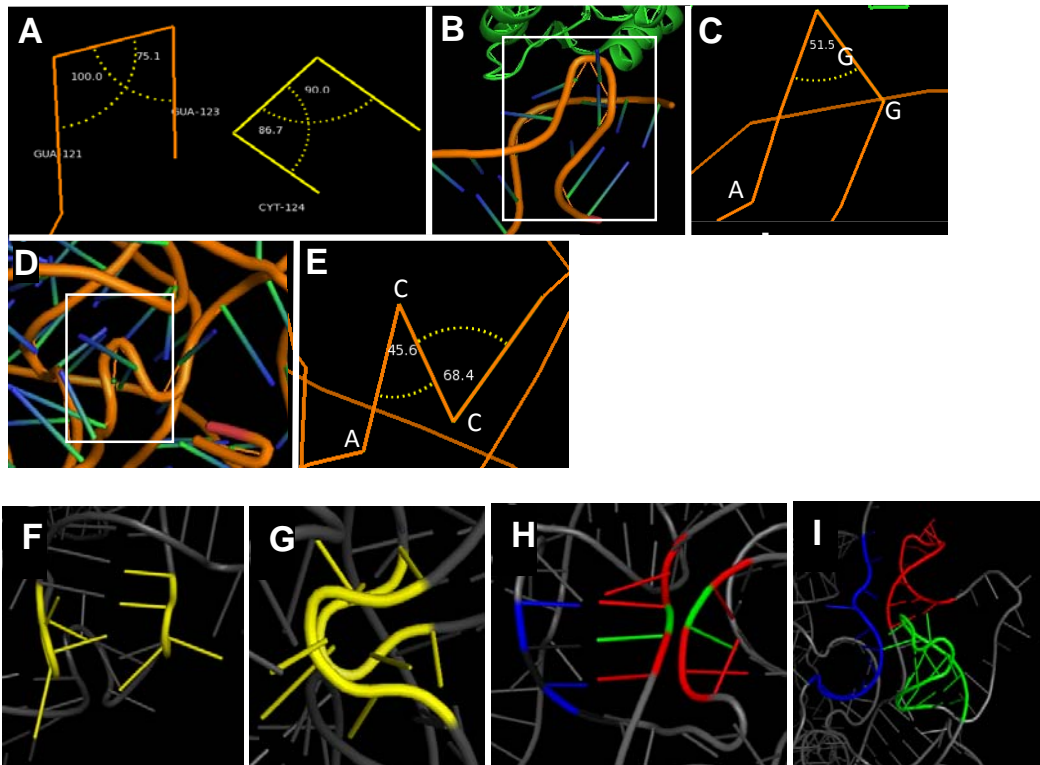


**Figure S5 Average model variability following simulated annealing and sequence conservation.**

(A) The average distance between atoms in our working model (Figure 5) and those in other simulated annealing structures from the same starting model (Table S4D) after alignment in Pymol. The region 132-216 (PBS) has been removed for the alignments. This single stranded region is most likely very flexible; comparing the positions at a particular time is therefore meaningless and will artificially increase the overall mean variation. The artificially added 3' region has also been removed for alignment purposes as its position is unimportant for the understanding of the wild type structural elements.

(B) 3D structure of the HIV-1 5' untranslated region RNA annotated by structural deviation between models. The molecule's backbone is displayed as a continuous width and color scale from thin and blue for low model positional conservation to thick and red for high model positional conservation.

(C) 3D model of the HIV-1 5' untranslated region RNA, nucleotides 104-344, colored by sequence conservation and displayed in Pymol. Coloring is from low conservation (blue) to high conservation (red). See also figure 5.



**Figure. S6. A Kink turn and other features of the modelled RNA.**

Other examples of the sharp, kink turn like, backbone folds in published structures. (A) Pymol ribbon trace between phosphorous atoms including angles between 223-224-225 and 224-225-226 (orange) and 122-123-124 and 123-124-125 (yellow) in our model. (B) White box highlights U4 snRNA (orange backbone (accession 1e7k)) nucleotide positions 133-135. (C) Phosphorus backbone trace and angle between phosphorus atoms of residues 133-131-135. (D) White box highlights nucleotides 272-275 of a self-splicing group-1 intron (accession 1u6b). (E) Phosphorus backbone trace and angle between phosphorus atoms of residues 272-273-274 and of residues 273-274-275. (F) Highlighted (yellow) tight turns in our working 3D RNA model at positions 123-124 (left) and 224-225 (right). (G) View through turn 224-225 towards turn 123-124. (H) Nucleotides 220-223 (AGAG), far right, and, 230-233 (UCUC) far left, are complementary in sequence and proximal in space in our model. This may suggest that pairing with positions 230-233 (UCUC) may be with 220-223 instead of 330-333 (GAGA) or could represent a switchable conformation. (I) SL2 (green) and SL3 (red) and the PBS region (blue) beyond PBS2 form a pocket orientated away from SL1 and the k-turns which is an attractive candidate for specific Gag binding. See also figure 6.















Pair	Random starting models			FRET distance constrained			Minimized		Energy	Relaxed		Energy
	Distance	Diff.	$\bar{x}$	Distance	Diff.	$\bar{x}$	Distance	Diff.	$\bar{x}$ distance	Distance	Diff.	$\bar{x}$ distance
	r7											
A1+D2	7.339	81.348		83.553	5.134		82.961	5.726	1472.871	80.928	7.759	1179.881
A1+D3	28.482	36.910		63.486	1.906		62.292	3.100		55.633	9.759	
A1+D6	70.992	0.019		67.296	3.715		66.416	4.595		52.403	18.608	
A5+D2	65.430	5.546		73.364	2.388		74.403	3.427		91.768	20.792	
A5+D3	56.903	13.077		67.410	2.570		68.249	1.731		70.398	0.418	
A5+D6	59.029	3.408		63.889	1.452		63.761	1.324		70.847	8.410	
A4+D2	24.010	49.326		69.186	4.150		69.586	3.750		68.487	4.849	
A4+D6	82.560	12.917	25.319	70.476	0.833	2.768	70.549	0.906	3.070	74.660	5.017	9.451
	r8											
A1+D2	53.017	35.670		92.805	4.118		92.153	3.466	1436.724	94.268	5.581	1190.774
A1+D3	79.874	14.482		63.473	1.919		63.159	2.233		55.272	10.120	
A1+D6	73.412	2.401		73.534	2.523		73.559	2.548		75.982	4.971	
A5+D2	66.905	4.071		73.370	2.394		74.540	3.564		90.480	19.504	
A5+D3	57.017	12.963		67.374	2.606		67.436	2.544		63.304	6.676	
A5+D6	89.885	27.448		60.992	1.445		60.757	1.680		57.946	4.491	
A4+D2	39.009	34.327		72.082	1.254		72.450	0.886		77.854	4.518	
A4+D6	135.890	66.247	24.701	77.182	7.539	2.975	77.298	7.655	3.072	84.129	14.486	8.793
	r9											
A1+D2	93.858	5.171		83.549	5.138		83.277	5.410	1453.055	80.921	7.766	1195.877
A1+D3	87.284	21.892		64.120	1.272		63.746	1.646		57.948	7.444	
A1+D6	29.935	41.076		67.334	3.677		67.151	3.860		70.721	0.290	
A5+D2	64.704	6.272		73.420	2.444		73.780	2.804		97.238	26.262	
A5+D3	33.338	36.642		71.821	1.841		71.601	1.621		85.812	15.832	
A5+D6	39.311	23.126		63.910	1.473		63.511	1.074		71.813	9.376	
A4+D2	66.753	6.583		69.295	4.041		69.135	4.201		65.478	7.858	
A4+D6	68.857	0.786	17.693	77.188	7.545	3.429	77.739	8.096	3.589	86.911	17.268	11.512

**Table S3. Table of the difference between the modelled and measured distances for each fluorophore pair from starting models 0-9.**

The 'Random starting models' columns show the pseudo experimental FRET distances and how much they differ from experimental ones (Diff). The 'FRET distance constrained' columns are after simulated annealing constrained by the FRET measured distances. The 'Minimized' columns are from a minimization step which ensures the correct shape of helices. The Relaxed columns represent a rigorous room temperature anneal without FRET constraints.

Due to the added complexity of modeling the PNA, the glycol linker and the fluorophore with the RNA, the distances used for modeling were between the C1' atoms of the nucleotide closest to the fluorophore. A description of randomized starting models is shown in Figure S4 and shown in Figure 4. See also Tables 3 and 4.

## Supplemental Experimental Procedures

### Anisotropy Measurements

Steady-state fluorescence spectra were measured on a Cary Eclipse Fluorometer, Varian Inc. The excitation and emission band-passes were set to 5 nm. 150 ml samples of labeled PNA on HIV RNA at 20 nM concentration were excited and emission spectra were collected. All spectra were corrected for instrument response. Fluorescence anisotropies for donor (Atto488) and acceptor (Atto647) were calculated from the polarization of the emission components  $I_{VV}$ ,  $I_{VH}$ ,  $I_{HV}$ , and  $I_{HH}$  (where the subscripts denote the orientation of the excitation and emission polarizers) as  $r = (I_{VV} - GI_{VH}) / (I_{VV} + 2GI_{VH})$ , where  $G = I_{HV} / I_{HH}$ . Where  $I_{VV}$  is fluorescence intensity with excitation polarization aligned vertically and emission excitation polarization aligned vertically,  $I_{VH}$  is fluorescence intensity recorded with the emission polarization aligned horizontally. For Atto488 anisotropy, excitation was at 480 nm and emission spectra were collected at 525 nm; whereas for Atto647 anisotropy, the excitation wavelength was set to 635 nm and emission was collected at 670 nm. All measurements were performed at 21°C. The Förster distance of donor and acceptor pair is normally calculated using an orientation factor  $\kappa^2$  with an average value of 2/3 for free rotating dyes. This is not appropriate for these experiments as neither dye has free rotation, most probably due to both the linker joining the dye to the PNA and due to some attraction between the dye and RNA. Atto647 had the least rotational freedom of the dyes used and so it was given a 0° dipole orientation and the relative donor angle (Atto647) was calculated.

We then recalculated the Förster distance for this particular structure based on the emission anisotropy from the FRET measurements, which is a product of the limiting anisotropy (Figure S2), the donor and acceptor axial depolarization factors  $d_d^x$  and  $d_a^x$ , and the depolarization factor,  $P_2(\cos \beta_{da})$ , corresponding to the angle between the donor and acceptor symmetry axes (Ivanov and Mizuuchi, 2009). The emission anisotropy between fluorophores is EA, the angle from the donor-acceptor vector is  $\theta$ , the orientation factor is  $\kappa^2$ .

$$EA_{FRET} = \frac{2}{5} d_d^x d_a^x P_2 \cos(\beta_{da}) \quad \text{Equation 1}$$

So we get,

$$\cos(\beta_{da}) = \sqrt{\frac{1}{3} + \frac{5}{3} \left( \frac{EA_{FRET}}{d_d^x d_a^x} \right)} \quad \text{Equation 2}$$

where,

$$d_d^x = \sqrt{\frac{5}{2} r_d} \quad \text{and} \quad d_a^x = \sqrt{\frac{5}{2} r_a} \quad \text{Equation 3}$$

with the angle  $\beta_{da}$  The orientation factor  $\kappa^2$  of this dye pair can be recalculated by:

$$k^2 = (3 \cos \theta_a \theta_d - \cos \beta_{da})^2 \quad \text{Equation 4}$$

The  $\Delta$  factor to correct the Förster distance was calculated by:

$$\Delta \equiv \sqrt[6]{\frac{k^2}{2/3}} \quad \text{Equation 5}$$

The  $K^2$  range was calculated by:

$$k_{\min}^2 = 2/3 \left( 1 - \frac{\sqrt{\frac{5}{2} EA_d} + \sqrt{\frac{5}{2} EA_a}}{2} \right) \quad \text{Equation 6}$$

$$k_{\max}^2 = 2/3 \left( 1 + \sqrt{\frac{5}{2} EA_d} + \sqrt{\frac{5}{2} EA_a} + 3 \sqrt{\frac{5}{2} EA_d} \sqrt{\frac{5}{2} EA_a} \right) \quad \text{Equation 7}$$

Using the equations 6 and 7, we calculated  $k_{\min}^2 = 0.172$  and  $k_{\max}^2 = 2.088$ .

## Shot Noise

The random fluctuations in fluorescence signal intensity in both channels due to photon shot-noise results in uncertainty in the true value of the Apparent FRET efficiencies,  $E_{app}$ . The values can be estimated using a standard analysis of error. The width  $\Delta R$  from shot-noise contribution to the  $E_{(app)}$  histogram was given by the following formula (Dahan et al., 1999, Ying et al., 2000):

$$\text{Where } \Delta R = \left( \frac{\gamma(1-\gamma)^2}{T} + \frac{\gamma^2(1-\gamma)}{T} \right)^{1/2}$$

$\gamma = \bar{I}_a / (\bar{I}_a + \bar{I}_d)$ , and T is the threshold value per ms in all experiments. The upper limit for the shot-noise width  $\Delta R$  was estimated to be 0.12 for a 40 base model dsDNA sample. Experimental width from the model ds-DNA sample is approximately twice the magnitude of the shot-noise induced width. The histogram width of HIV-RNA FRET samples in Table S2 is often wider than the width of the model ds-DNA. These results indicate that there is some other fluctuation process contributing to the  $E_{(app)}$  histogram. As the linker is identical in all cases, we suggest that the flexibility of the RNA linking the two dyes in the monomer may be the second reason that makes the histogram wider based on this experimental and the theoretical study.

## Supplemental References

Dahan, M., Deniz, A.A., Ha, T.J., Chemla, D.S., Schultz, P.G., and Weiss, S. (1999). Ratiometric measurement and identification of single diffusing molecules. *Chem. Phys.* 247, 85–106.

Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, L.S., Tacker, M., and Schuster, P. (1994). Fast folding and comparison of RNA secondary structures. *Chemical Monthly* 125, 167–188.

Ivanov, V., Li, M., and Mizuuchi, K. (2009). Impact of emission anisotropy on fluorescence spectroscopy and FRET distance measurements. *Biophys. J.* 97, 922–929.

Smith, C., Heyne, S., Richter, A.S., Will, S., and Backofen, R. (2010). Freiburg RNA tools: a web server integrating IntaRNA, ExpaRNA and LocRNA. *Nucleic Acids Res.* 38, 373–377.

Ying, L.M., Wallace, M.I., Balasubramanian, S., and Klenerman, D. (2000). Ratiometric analysis of single-molecule fluorescence resonance energy transfer using logical combinations of threshold criteria: A study of 12-mer DNA. *J. Phys. Chem. B* 104, 5171–5178.