The hybrid RNA_2PNA_2 i-motif – A novel RNA-based building block

Saikat Chakraborty, Souvik Modi, Yamuna Krishnan*

National Centre for Biological Sciences, TIFR, GKVK, Bellary Road, Bangalore 560 065, India.

Email: yamuna@ncbs.res.in.

Experimental Section:

PNA synthesis and characterization: All PNA and related labeled PNA derivatives were synthesized and characterized as described previously.[1] RNA and PNA sequences we have used are given in the main text Figure 1.

Native Gel electrophoresis:

Native polyacrylamide gel electrophoresis in the case of RNA-PNA complexes were performed using N-terminally TMR labeled PNA (**N-TMR-P**). RNA and PNA were allowed to complex at total strand concentrations of 100 μM by addition of an equimolar mixture of **R** and **N-TMR-P** in 30 mM sodium acetate, pH 4.5, heated to 90ºC and then cooled to room temperature at a rate of 0.33ºC/min and equilibrated at 4ºC for 8 hrs. 20% polyacrylamide gels were used to study the differential mobility of any complex that was performed. Robinson Britton buffer (boric acid, acetic acid and phosphoric acid at 0.04 M concentration each, pH 4.5) was used as running buffer. All gels were run at 70 V for 6 hrs at 4ºC. All gels were visualized by ethidium bromide filter.

Nano-electrospray Ionisation Mass spectrometry (Nano ESI-MS):

All mass spectrometry measurements were performed on Micromass ESI-MS Q-TOF Ultima Mass Spectrometer with micro channel plate detector. All the spectra were collected in positive ion mode unless specified. An equimolar solution of 0.4 mM **P** and 0.4 mM **R** in 100 mM ammonium acetate (pH 4.5) solution were heated to 90ºC and then cooled to room temperature at a rate of 0.33ºC/ min and then equilibrated at 4ºC for 8 hrs. Just prior to injection, the equilibrated sample was diluted 10 fold with water acidified with formic acid (pH 4.5) and then loaded into the nano flow capillary. Source temperature of 28ºC, capillary voltage of 2.8 kV and cone voltage of 50 V were the source parameters employed during acquisition of mass spectra.

UV–VIS Spectrophotometric studies:

All UV measurements were done on a Varian Cary 300Bio UV-VIS spectrometer equipped with a Peltier temperature controller and in a 0.1 cm pathlength cuvette. For RNA-PNA hybrid, stock solutions of RNA (4.9 mM) and PNA (2.5 mM) were made in salt-free, double autoclaved MilliQ water. Samples were prepared by addition of RNA and PNA at the desired concentration in 30 mM sodium acetate buffer, pH 4.5. To form the hybrid i-motif at the desired strand concentration, an equimolar mixture of **R** and **P** at the required strand concentration was used and samples were formed as described earlier. All melting experiments were performed by monitoring absorbance at 295 nm over a temperature regime 10ºC to 90ºC using a heating rate of 1ºC/min unless specified otherwise. All $T_{1/2}$ values were obtained directly from absorbance versus temperature curve unless specified otherwise.

pH dependent studies:

pH dependent thermal melts were carried out at total strand concentration of 200 μM for RNA-PNA hybrids. Samples were investigated at a constant ion concentration of 30 mM Na^+ over a pH regime 3.5 to 7.0. In the pH regime 4.0-5.5, samples were prepared in 30 mM sodium acetate buffer. At all other pH values 30 mM sodium phosphate buffer at desired pH was used.

Circular Dichroism studies:

All CD spectra were recorded on a JASCO-J-720 spectropolarimeter. Hybrid RNA-PNA complex at total strand concentrations of 100 μM were prepared as previously mentioned in 30 mM Sodium acetate buffer, at pH 4.5. Spectra were recorded from 330 to 200 nm and are presented as an average of 3 successive scans, subtracted from a baseline corresponding to buffer alone.

Fluorescence studies:

For self-quenching studies on the hybrid i-motifs, TMR-labeled tetraplexes were prepared at a total strand concentration of 100 μM of 1:1 **3′-TMR-R**:**P** in 50 mM sodium acetate buffer, pH 4.5 as described earlier. Similarly, 1:1 N**-TMR-P**:**R** at 50 μM strand concentration was also made. For fluorescence quenching, samples were made by mixing 1:1 **N-Dabcyl-P**:**3′-TMR-R** or 1:1 **C-Dabcyl-P**:**3′-TMR-R** at a total strand concentration of 100 μM using the above mentioned protocol. Just prior to measurement all samples were passed through two successive G-25 sephadex spin columns to thoroughly eliminate any labeled, uncomplexed strands and diluted five fold in the cuvette. All fluorescence spectra were recorded on a FLUOROLOG-SPEX spectrofluorimeter and in a 10 mm pathlength quartz cuvette. Samples were excited at 520 nm and emission was recorded between 540-700 nm, at scan rate of 1 nm/sec. The working formula used for the calculation of the distances is the following: $E=1- f_{D/A} /f_D$ and $E=R_0^6/ (R^6+R_0^6)$, where $f_{D/A}$= fluorescence intensity in the presence of both

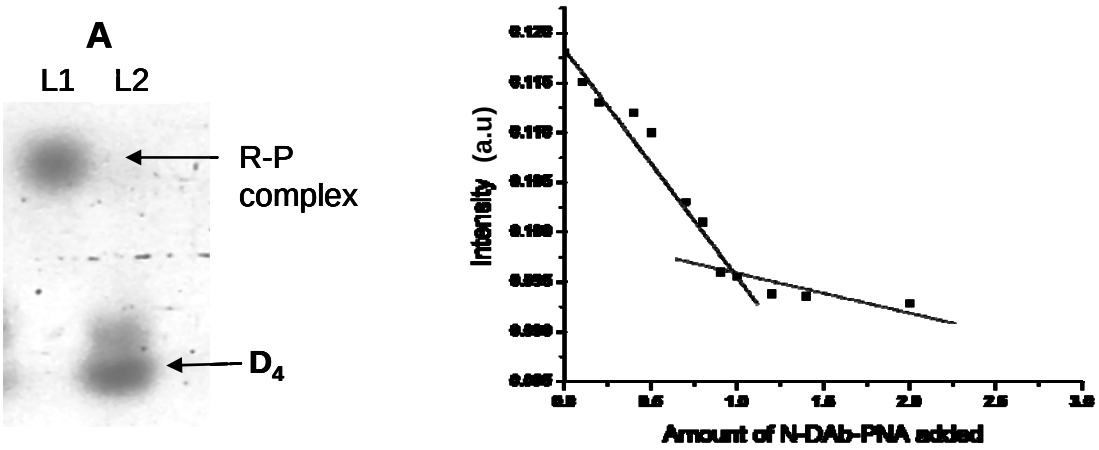
the donor and acceptor; f_D = fluorescence intensity of the donor only; E= efficiency of energy transfer; R_0= Förster distance; and R= distance between the donor and acceptor. The distances between the labeled strands were calculated using fluorescence intensity of the intact complex as $f_{D/A}$ and the fluorescence intensity of the corresponding heat-denatured samples at room temperature as f_D. The Förster distance (R_0) for TMR–TMR self-quenching was 44 Å and TMR-Dabcyl quenching was 26 Å as described previously.[2] To obtain theoretical estimates of the relevant distances, a coarse-gained model of the $\mathbf{R}_2\mathbf{P}_2$ hybrid tetramers, (see SI Figure 6) incorporating the linkers was constructed based on the NMR solution structure of the r(UC_5), RNA_4 i-motif.[3]

NMR studies:

All NMR spectra were recorded on a Bruker AV-800 spectrometer. Hybrid complex was used at 1 mM total strand concentration in 30 mM d3-sodium acetate buffer at pH 4.5. Water suppression was achieved using an Excitation Sculpting solvent suppression programme.[4] For 1D NMR experiments 256 scans were taken, the spectral band width was maintained at 20 KHz, with an acquisition time of 0.73 second and a 2 second recycling delay. The methyl chemical shift at 1.67 ppm of Na-acetate was used as the internal standard. For NOESY experiments, (600×4096) complex points were collected. 20 KHz spectral width was employed in both dimensions with acquisition times of 0.1 s in t_2 and 0.02 s in t_1, using 200 ms mixing time unless specified.

Native PAGE and Fluorescence titration shows RNA-PNA complex has 1:1 stoichiometry:

Native PAGE was performed with a 1:1 mixture of RNA:PNA and compared with a known DNA_4 i-motif forming sequence. A mixture of 1:1 RNA, **R** and PNA labelled with TMR at its N-terminus, N-TMR-**P**, was annealed and incubated as described. SI Figure 1A shows that there is a significant difference in mobility of 1:1 **R**-N-TMR-**P** complex as compared to the DNA only i-motif formed by [3′-**TMR-D**]$_4$. At pH 4.5, PNA alone does not migrate on the gel matrix under these conditions (data not shown) as single N-TMR-**P** strands carry a positive charge. However, only in the presence of one equivalent of **R** can it be induced to migrate. Thus, 1:1 mixture of RNA and PNA form a hetero-complex at pH 4.5 containing both RNA and PNA. In order to prove 1:1 complexation between **R** and **P** in bulk solution, we did fluorescence titration experiment. To a constant amount of 3′-TMR labeled **R** (1.5 μM) in 30 mM sodium acetate, pH 4.5 incremental amounts of N-Dabcyl labeled PNA was added and fluorescence quenching measured.



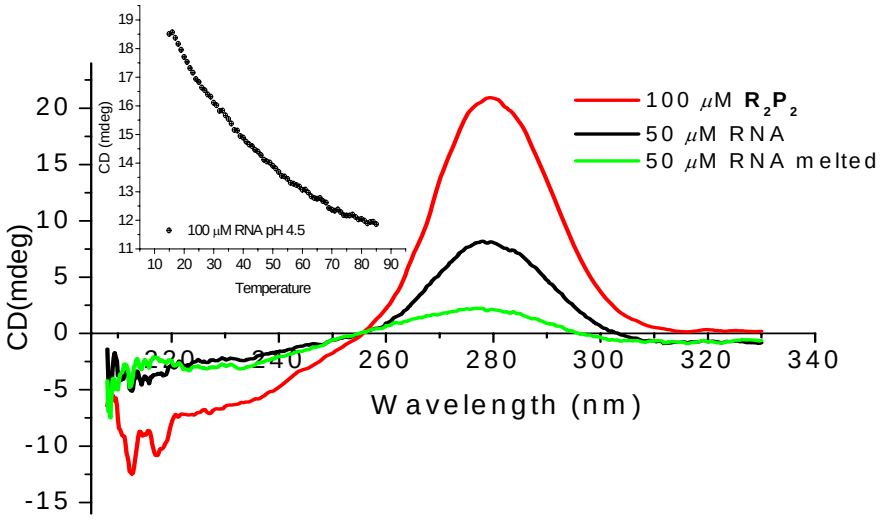
SI Figure 1: (A) Native gel electrophoresis of hybrid RNA-PNA complex. L1, 1:1 [**R:N-TMR-P**], at 100 μM total strand concentration; L2, 100 μM [**3′-TMR-D**]$_4$. **(B)** Fluorescence titration showing 1:1 complexation between 3′-**TMR-R and N-Dab-P**.

As more and more PNA complexes with the RNA, the fluorescence of the **3′-TMR-R** decreases as shown in SI figure 1. This decrease tapers off after 1:1 concentration of PNA:RNA is reached. This experiment unambiguously proved that RNA and PNA forms 1:1 heterocomplex in bulk.

CD Experiments:

In order to further characterize the hybrid complex formed by **R** and **P** we used circular dichroism (CD) spectroscopy. RNA is CD sensitive, while PNA shows negligible CD. A sample of 50 μM R alone in acetate buffer, pH 4.5 showed a CD trace with a positive peak at 280 nm and negative peak at 210 nm (SI Figure 2). The projected $T_{1/2}$ of the RNA_4 i-motif from R is predicted to be less than 4ºC.[5] Hence, it is reasonable to assume that R is largely single stranded under these experimental conditions. Importantly, the same concentration of RNA in presence of 1 equivalent of P showed nearly two fold enhancement in CD signal as shown in SI Figure 2 suggesting formation of completely different species, namely the hybrid i-motif. However the basic features of the CD profile were conserved in comparison with pure R suggesting greater degree of complexation between RNA and PNA. However, comparison between 100 μM **R** with 100 μM RNA-PNA complex showed only a modest increase in CD signal (Data not shown) which suggest that at 20ºC there is probably some residual RNA secondary structure implying that the projected $T_{1/2}$ of the RNA_4 i-motif may not be as low as 4ºC. To prove that we did CD melting experiment on 100 μM RNA in 30 mM sodium acetate, pH 4.5. It showed a noncooperative melt with a $T_{1/2}$ value of 20ºC (SI Figure 2 inset). It is probably the parallel duplex formed by C-rich RNA sequence at acidic pH.[6] This kind of complexation was already been reported in literature.

Nevertheless, the CD profile of the hybrid indicates that the chiral features of the RNA-only complex are preserved in the R_2P_2 complex.



SI figure 2. Circular dichroism profiles of **R** and RNA_2-PNA_2 hybrid at comparable strand concentrations, pH 4.5. Inset: Melting transition of 100 μM RNA in 30 mM sodium acetate, pH4.5.

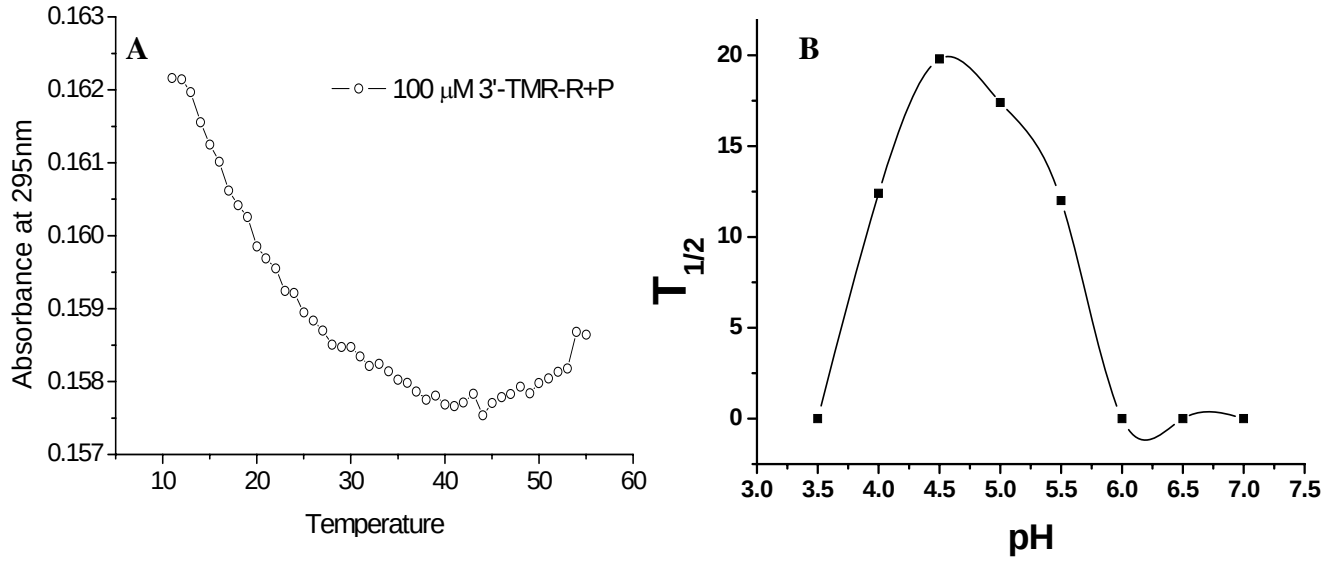
UV-VIS Spectrophotometric studies:

Melting profile of the i-motif is characterized by a distinctive inverse sigmoidal melting curve at 295 nm due to melting of CH^+-C base pairs.[7] In order to see whether these hybrid complexes also evidence a melting transition UV melting studies were conducted. Interestingly, an equimolar solution of 1:1 **R** and **P** at 100 μM total strand concentration showed an inverse sigmoidal curve at 295 nm characteristic of CH^+-C base pairing with an associated $T_{1/2}$ of 22±0.3ºC (calculated from Figure 3). This is significantly higher than the projected $T_{1/2}$ of the corresponding RNA_4 i-motif. This increased stability could be attributed to a reduction in the number of negatively charged strands in the $\mathbf{R_2P_2}$ tetramer and/ or a decrease in the number of 2′-OH-2′-OH contacts in the same.

*Stability as a function of pH***:** In order to investigate the thermal stability of the hybrid RNA-PNA complex as a function of pH, 200 μM $\mathbf{R_2P_2}$ hybrid in buffers of different pH were thermally denatured as described in the experimental section. $T_{1/2}$ values were plotted as a function of pH (SI Figure 3B). The $\mathbf{R_2P_2}$ hybrid complex exists over a pH regime of 4-5.5, showing maximum stability near pH 4.5 which is the pK_a of cytosine.[8] These findings are in line with the cytosines in these complexes being hemiprotonated, which is a prerequisite for C-CH^+ base pairing, as seen in i-motifs.

Labeled R_2P_2 hybrids are marginally less stable than unlabeled R_2P_2 complex:

To confirm that introduction of the fluorophores does not abolish complex formation, labeled $\mathbf{R_2P_2}$ complexes were investigated to see if they showed a thermal melting transition. We formed 1:1 **3´-TMR-$\mathbf{R_2P_2}$** hybrid with **3´-TMR-R** and **P** at a total strand concentration of 100 μM in 50 mM sodium acetate, pH 4.5. The melting transition was monitored by UV absorption at 295 nm. The melting transition is shown in SI Figure 3B with an associated $T_{1/2}$ of ~18± 0.3ºC. Thus introduction of the label did not abolish complex formation and labeled $\mathbf{R_2P_2}$ complex are only marginally destabilized compared to the unlabeled complex.



SI Figure 3. (A) Melting transition of 1:1 3´-TMR labelled RNA and PNA monitored by UV absorption at 295 nm. (B) pH dependence of $\mathbf{R_2P_2}$ i-motif thermal stability.

FRET Measurements:

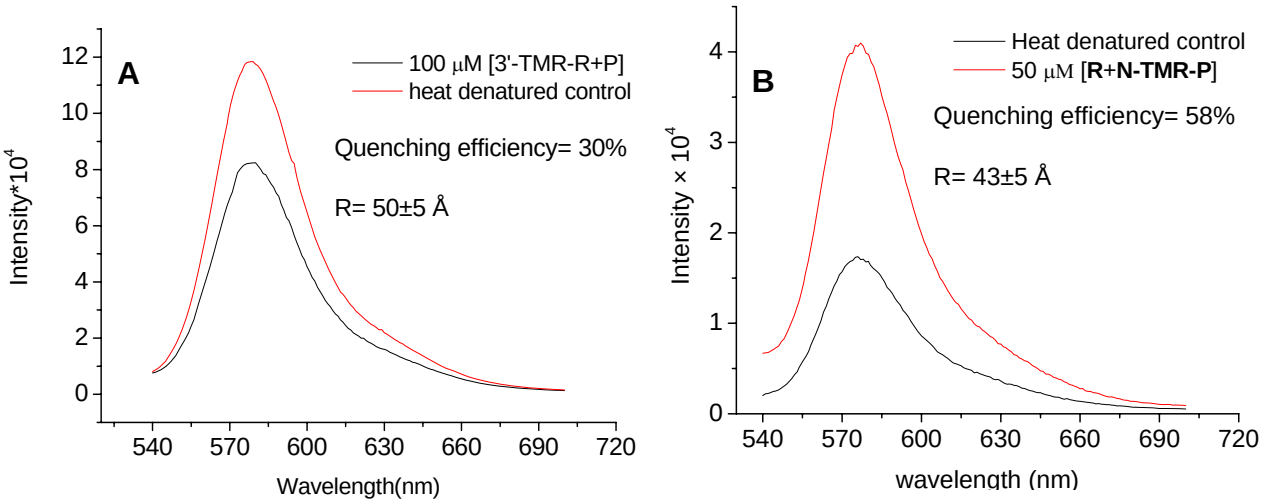
Prior to all FRET measurements, anisotropy values of all labeled complexes were measured. They were all between 0.03-0.07, indicating that the fluorophores are freely rotating.

Distance calculation:

Distance calculations were carried out by using the equations described in the experimental section above.

Self quenching studies were done on $\mathbf{R_2P_2}$ hybrid i-motifs to know the corresponding strand orientations. Hybrid $\mathbf{R_2P_2}$ samples were made by adding equimolar mixture of 1:1 **3'-TMR-R:P**. Likewise another sample composed of equimolar mixture of **R:N-TMR-P** were made. Samples were

made according to the protocol described in the main text. Each experiment had two controls (i) fluorescence of uncomplexed **3´-TMR-R** and/ or **N-TMR-P** alone at the same strand concentration and (ii) fluorescence of thermally denatured complex at room temperature. All quenching measurements on each labeled complex and controls were performed in triplicate. SI Figure 4A shows ~30% quenching was found for the **3´-TMR-R$_2$-P$_2$** complex, where the control is the same sample that is heat denatured and rapidly cooled to room temperature. Similarly the N terminus of the two **N-TMR-P** strands in the **R$_2$-N-TMR-P$_2$** complex shows a quenching of 58% (SI-Figure 4B).

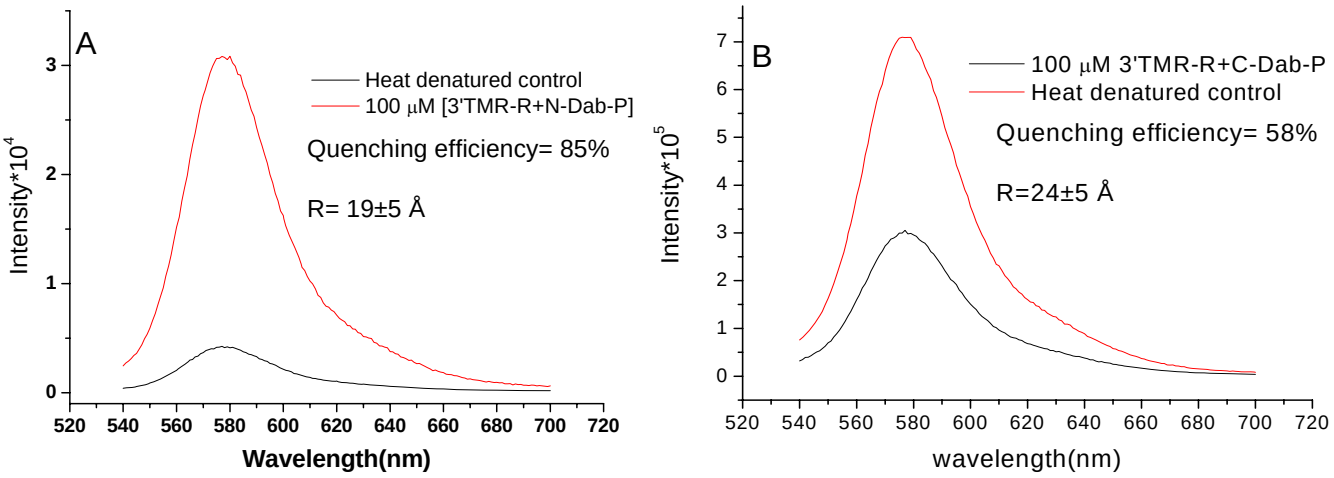


SI Figure 4. Self-quenching studies with (A) **3´-TMR-R$_2$-P$_2$** at 100 μM total strand concentration. (B) **R$_2$-N-TMR-P$_2$** at 50 μM total strand concentration. The control measurements were performed after heat denaturing the sample followed by rapid cooling to room temperature.

Fluorescence quenching experiments:

To elucidate the strand configuration flanking the narrow groove, fluorescence quenching experiments were performed on the **R$_2$P$_2$** i-motifs. In the case of **R$_2$P$_2$** hybrid, the two **R-P** hetero duplexes can intercalate in two configurations as shown in Figure 4A in the main text.

Fluorescence quenching measurements on 1:1 complex of **3´-TMR-R:N-Dabcyl-P** prepared as described in main text is shown in SI Figure 5A. Quenching was found to be ~85% corresponding to a distance of ~19 Å. Similarly fluorescence measurements were carried out on **3´-TMR-R:C-Dabcyl-P** as shown in SI figure 5B. Quenching was found to be 58% corresponds to a distance of 24 Å. This data is supporting the earlier measurement and also the fact that two RNA strands occupy one narrow groove.

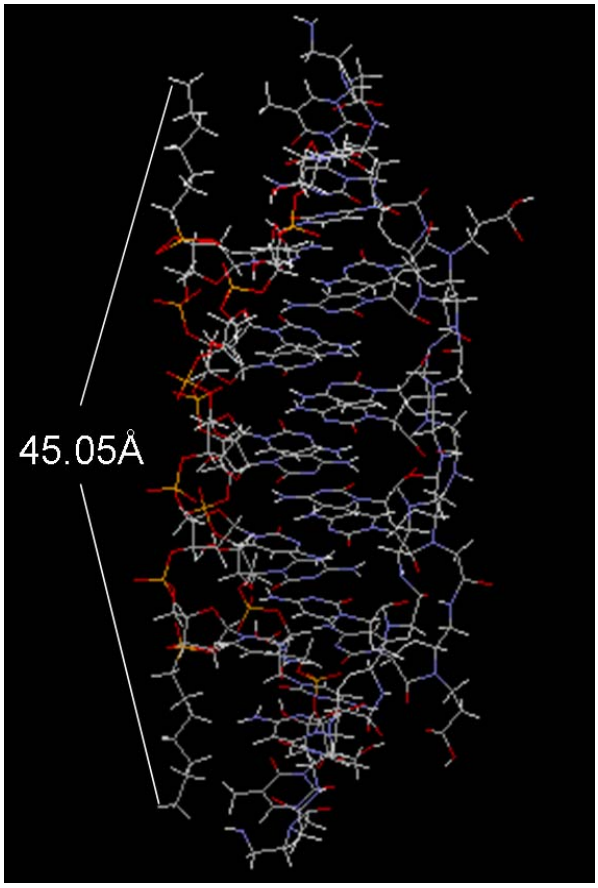


SI Figure 5. (A) Fluorescence quenching experiments on the dual labeled hybrid i-motif of 1:1 **3′-TMR-R** and **N-Dabcyl-P** at 50 μM strand concentration. The transfer efficiency was measured with respect to heat denatured control. (B) Fluorescence quenching on **3′-TMR-R:C-Dabcyl-P** at 50 μM strand concentration.

SI Table 1: All the possible values of distances between the fluorophores based on the Model 1 and Model 2.

	Model 1	Model 2	Observed distances
3′-TMR-R to 3′-TMR-R	~55Å	~45 Å	50± 5 Å
3′-TMR-R to N-Dab-P	~9 Å	~19 Å	19± 5 Å
3′-TMR-R to C-Dab-P	~23 Å	~23 Å	24± 5 Å

Model of the hybrid RNA i-motif:



SI Figure 6. Coarse grained model of $\mathbf{R_2P_2}$ with the linkers. Distance between the linkers is also shown. [a] Models of the hybrid RNA_2-PNA_2 i-motifs were constructed using PyMOL software using the NMR structure parameters of the RNA_4 i-motif (Accession number: **1I9K**). End to end distances R_{calc}, were measured incorporating the linker connecting the fluorophores.

1D NMR evidences complex formation

The 1D NMR spectrum shows distinct peaks between 15.5–16.5 ppm (SI Figure 7), which is characteristic of the imino protons (NH_{im}) in CH^+-C base pairs in DNA_4 i-motifs.[9] This proves that the RNA-PNA complex is held by CH^+-C base pairs which reaffirm the findings from mass spectrometry and UV melting studies. We also observed two peaks at the region 11.0-11.6 ppm due to the uridine imino protons. The peak at 11.47 ppm was assigned to rU2 by following the connectivities in NOESY (see 2D NMR section) whereas 11.17 ppm signal corresponding to rU1 shows no such connectivity as expected. Thus the rU2 imino peak at 11.47 ppm suggests a single population of hybrid complex.[10] The presence of a single peak corresponding to the thymine CH_3 at 1.67 ppm (data not shown) further supports the formation of a unique population. Contributions from single-stranded RNA and PNA were negligible indicating that these species are below the NMR detection limit. This supports the findings from CD and fluorescence experiments (see below) which indicate that the hybrid is the predominant species in solution.

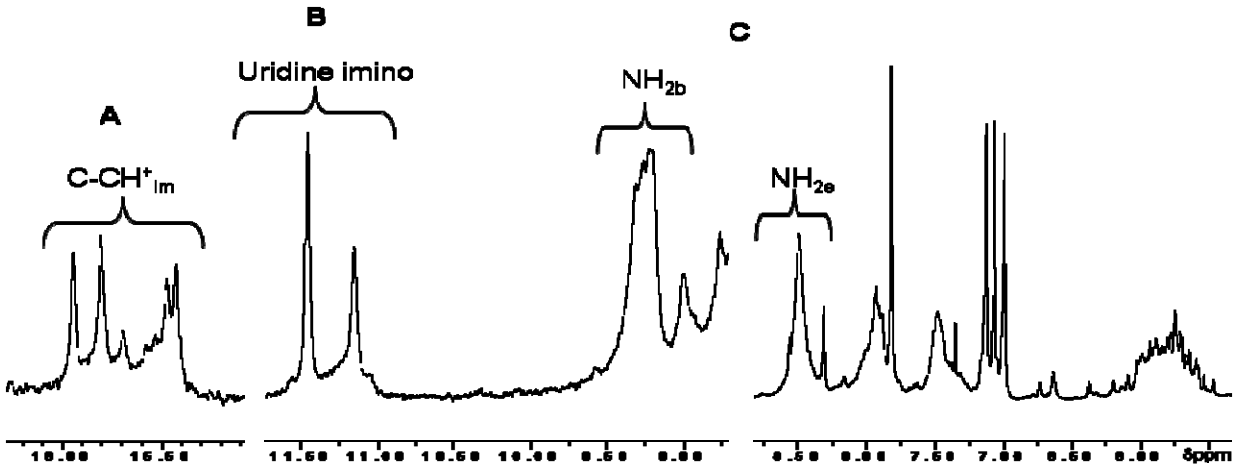
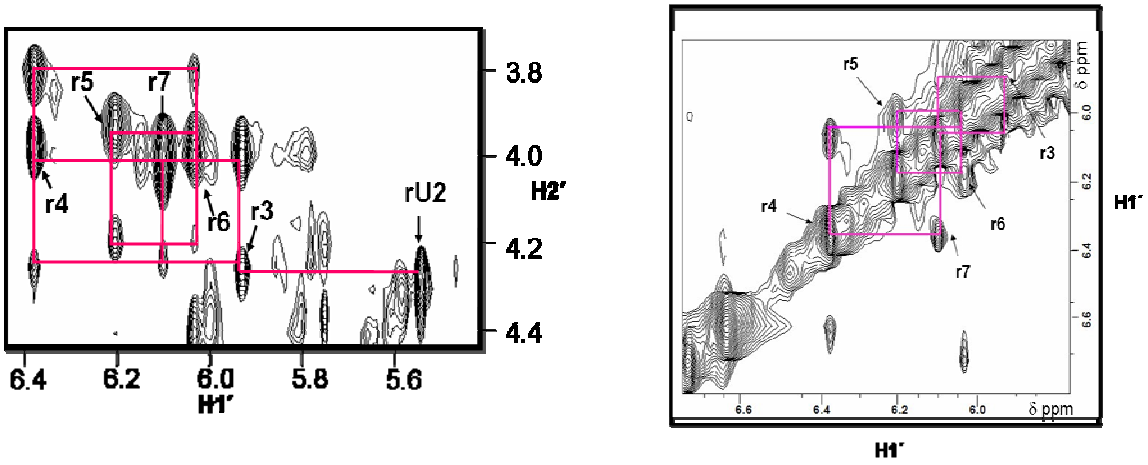


Figure 7: One dimensional proton NMR spectrum of **R_2P_2** hybrid i-motif in 95:5 H_2O:D_2O at 4ºC. Sample concentration was 1 mM in 30 mM sodium d3-acetate, pH 4.5. (A) Imino proton (NH^+_{im}) region showing overlapping peaks centered at 15.5 to 16 ppm region characteristic of C-CH^+ base pairs. (B) Single uridine imino peak characteristic of a single species. (C) Internally hydrogen bonded (NH_{2b}) amino, externally non hydrogen bonded (NH_{2e}) amino and (H6) aromatic region. 2'-OH peak was weak due to fast exchange with water.

The large peak at 9.5 δppm is characteristic of internally H-bonded (NH_{2b}) cytosine amino protons reaffirming CH^+-C base pairs. In the RNA_4 i-motif, 2´-OHs are observed in the region 6.7-6.9 ppm.[3] However diminished peak height in this region suggested that a large fraction of these hydroxyls may be

exchanged with the water present in the sample. Furthermore, the PNA amide backbone and nonbonded (pNH_{2e}) cytosine amino protons appear in the region 8-8.5 ppm. Thus NMR evidences that 1:1 RNA and PNA forms a single hybrid population in solution, and that this is the predominant species.

2D NMR shows sugar-sugar contact:



SI Figure 8: Two dimensional NOE spectrum of **R_2P_2** Hybrid i-motif at 1 mM strand concentration at 4°C and 200 ms mixing time. (A) Correlation of the H1´ peaks with H2´ region showing sequential NOE crosspeaks characteristic of proximity of both RNA backbones. RNA residues are labeled with their corresponding number. (B) NOE crosspeaks between H1´ protons showing proximal ribose-ribose contacts. RNA residues are labeled with their corresponding number.

SI Figure 8 shows NOE crosspeaks due to ribose contact in **R_2P_2** hybrid i-motif. NOE crosspeaks between H1′ and H1′ as well as H2′ indicate the close proximity of two RNA backbones.

Arguments for specific complex formation:

<u>Why we don't see any other i-motif stoichiometries.</u>

Implicit in the structure of the **R_2P_2** hybrid are the factors that favor the formation of this tetramer over other theoretically possible configurations.

1. R3P is disfavoured because there will be a large number of destabilizing 2' OH steric interactions (see later)

2. P3R is disfavoured by the excessive conformational flexibility PNA that slows down the kinetics of formation and narrows the pH range in which it will exist (see Y. Krishnan-Ghosh, E. Stephens, S. Balasubramanian, *Chem.Commun*. 2005, 527).

Importantly, in both these cases the complex is not electrically neutral.

3. **R2P2** is preferentially formed because of (i) its electrical neutrality and (ii) good compromise between PNA fluxionality, RNA 2'OH steric interactions and the RNA backbone acting as a template

for PNA hybridization.

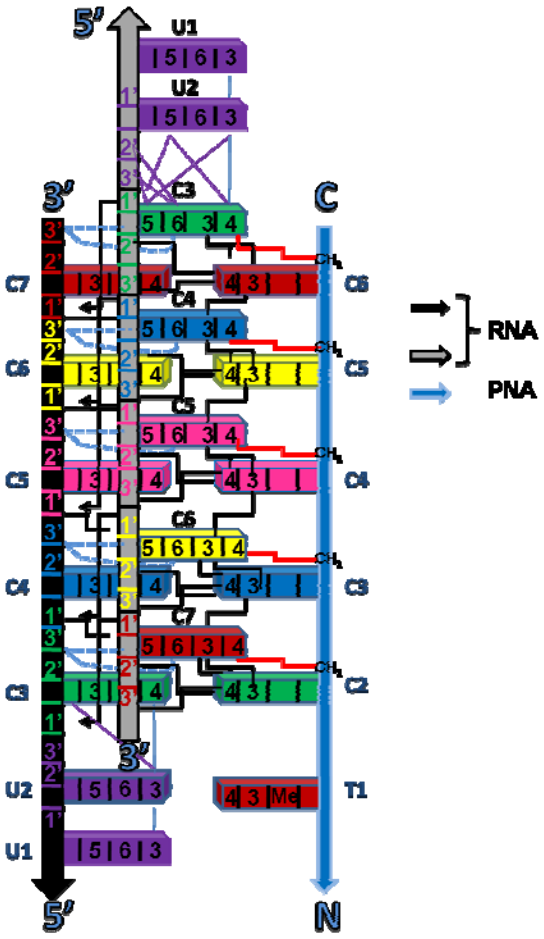
Why we observe only one particular strand orientation in R2P2.

There are two possible ways in which a heteroduplex intercalates to form an i-motif, (see Model 1 and 2, see main text). The factors that govern which is formed are (i) destabilizing electrostatic repulsion between proximal RNA strands (ii) destabilizing 2´-OH steric factors between proximal RNA strands and (iii) stabilizing sugar-sugar interactions between proximal RNA strands.

The **R2P2** structure indicates that sugar-sugar contacts are preferentially maximized despite there existing a possibility to minimize these factors (i) and (ii) by placing one RNA strand per narrow groove (Model 2). This suggests that ribose-ribose interactions possibly drive i-motif formation.

Why we observe only one particular intercalation topology in R2P2 (Model 1).

The intercalation topology of the hybrid reflects the optimization between maximizing the stabilizing sugar-sugar contacts (factor (iii)) and minimizing the destabilizing 2´-OH steric clash (factor (ii)) in a background of constant electrostatics in a given narrow groove environment. This topology of the **R2P2** hybrid eliminates a single destabilizing 2´-OH interaction in the narrow groove between proximal RNA strands. This dramatic influence of minimizing one destabilizing interaction over i-motif structure reveals that i-motif intercalation topology is highly sensitive to interactions in the narrow groove.



SI Figure 9: A schematic of the hybrid RNA_2-PNA_2 i-motif strand topology based on sequential NOEs. Different lines represent connectivities between intra or inter strand protons. Black lines indicate connectivies between wide groove protons, black line with arrow indicate narrow groove connectivities, red lines indicate PNA methylene connectivities with base protons, blue dotted lines indicate the H2′-H3′ sugar connectivities in the narrow groove. Only three strands are shown because identical connectivities are seen in the other duplex

References:

1 S. Modi, A. H. Wani, Y. Krishnan, *Nucleic Acids Res.* 2006, **34**, 4354.

2 S.Bernacchi, Y. Mély, *Nucleic Acids Res*. 2001, **29***,* e62.

3 K. Snoussi, S. Nonin-Lecomte, J. L. Leroy, *J. Mol. Biol*., 2001, **301**, 139.

4 T. L. Hwang, A. J. Shaka, *J. Magnetic Res. Series A,* 1995, **112**, 139.

5 D. Collin, K. Gehring, *J. Am. Chem. Soc*. 1998, **120**, 4069.

6 K. Jr. Hartman and A. J. Rich, *J. Am. Chem. Soc.* 1965, **87**, 2033; A. M. Michelson, J. Massoulie, W. Guschlbauer, *Prog. Nucl. Acids. Res. Mol. Biol.* 1967, **6**, 83; (c) E. O. Akinrimiski, C. Sander, P. O. P. Ts'o, *Biochemistry* 1963, **2**, 340.

7 A. T. Phan, J. L. Mergny, *Nucleic Acids Res*. 2002, **30**, 4618; (b) J. L. Mergny, L. Lacroix, X. Han, J. L. Leroy, C. Helene, *J. Am. Chem. Soc.* 1995, **117**, 8887; (c) J. L. Mergny, L. Jing, L. Lacroix, S. Amrane, J. B. Chaires, *Nucleic Acids Research*. 2005, **33**, e138.

8 G.M. Blackburn and M. Gait, *Nucleic acids in chemistry and Biology*. 1996, Oxford University Press, Oxford, 2nd edition, *ch.2*, 18.

9 K. Gehring, J. L. Leroy, M. Guéron, *Nature* 1993, **363**, 561.

10 A.T. Phan, M. Guéron, and J.L. Leroy, *J. Mol. Biol*., 2000, **299**, 123.