The RNA_2–PNA_2 hybrid i-motif—a novel RNA-based building block†

Saikat Chakraborty, Souvik Modi and Yamuna Krishnan*

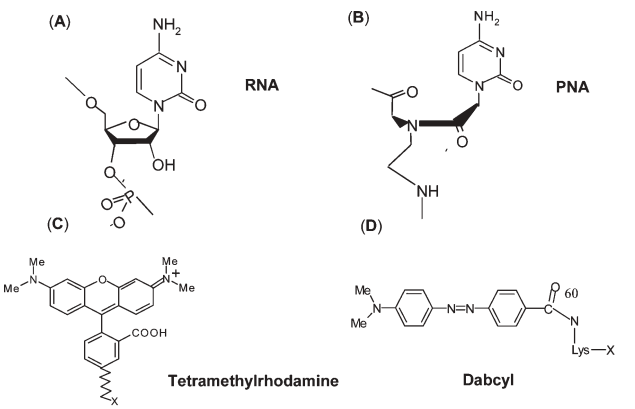
Received (in Cambridge, UK) 4th September 2007, Accepted 28th September 2007
First published as an Advance Article on the web 17th October 2007
DOI: 10.1039/b713525d

We report the formation of a hybrid RNA_2–PNA_2 i-motif comprised of two RNA and two PNA strands based on the sequence specific self assembly of RNA, with potential as a building block for structural RNA nanotechnology.

Structural DNA nanotechnology is a field where DNA is used to create robust and dynamic nanoscale architectures.[1] We have been interested in exploring the potential of non-Watson–Crick base paired motifs in the rational construction of such architectures and towards this end developed many such non B-DNA building blocks.[2] RNA is now emerging as a versatile entity for the analogous RNA equivalent, structural RNA nanotechnology.[3] RNA is capable of much more structural variety than DNA and has the capability of forming several unusual motifs through a variety of tertiary interactions that are predominantly non-Watson–Crick type.[4] Given this versatility, we sought to explore the predictable creation of RNA-based non-Watson–Crick building blocks for potential use as rational design elements in structural RNA technology.

C-rich sequences of DNA, RNA and their mimics are known to associate into a four-stranded structure, called the i-motif, that is held together by hemiprotonated C–C^+ base paired duplexes that are intercalated in an antiparallel orientation.[5] Given that RNA and PNA can form RNA_4 and PNA_4 i-motifs respectively, we wanted to see whether an equimolar mixture of C-rich RNA and PNA could form a hybrid i-motif.[5] In this paper, we describe the formation of a unique population of a hybrid $\mathbf{R_2P_2}$ i-motif from a binary mixture of C-rich RNA and PNA sequences.

We used C-rich PNA, **P** (Fig. 1) and C-rich RNA, **R**, that incorporated a T at the N-terminus and U at the 5′ end respectively to prevent higher order structure formation.[2] Native polyacrylamide gel electrophoresis showed that a 1 : 1 mixture of **R** : **P** forms a hybrid complex comprising both RNA and PNA (see Supporting Information, SI†). In order to obtain information on strand stoichiometry in this hybrid complex, we subjected it to Nano-Electrospray Ionisation mass spectrometry (NanoESI-MS, Micromass Q-TOF Ultima). An equimolar solution of **P** and **R** at 0.4 mM each in 100 mM NH_4OAc, pH 4.5 were annealed and analyzed by positive ion nano-ESI-MS.[6] Upon injection of the complex, the initial spectrum showed clearly a broad peak centered at *m*/*z* 1297.97 (Fig. 2A) where, the associated peak separation of (0.16 ± 0.03 *m*/*z* units) indicated that this was due to a hexa charged species. This corresponds to an associated molecular weight of 7787.82 ± 1 Da which is consistent with a four stranded entity $[2\mathbf{M_R} + 2\mathbf{M_P} + 10H^+ + 2NH_4^+]^{6+}$ (computed molecular weight. 7786.28 ± 1 Da) that is composed of two strands each of **P** and **R** with 10 additional protons. Interestingly, due to the high source temperature employed, the tetramer started dissociating with time and the broad peaks in Fig. 2B gradually disappeared at the expense of a family of new peaks also centered at the same *m*/*z* regime (see Fig. 1B). These peaks are equidistant with a constant separation of 7.6 ± 0.3 *m*/*z* units indicating that they correspond to multiply sodiated forms of triply-charged species of *m*/*z* 1299.35, with an associated molecular weight 3898.05 ± 1 Da. This in good correspondence with a dimeric entity composed of **R** and **P**, namely, $[\mathbf{M_R} + \mathbf{M_P} + 5H^+ + Na^+]^{3+}$ (computed molecular weight 3899.15 ± 1 Da). This fragmentation pattern supports a model where the tetrameric complex is composed of two identical **RP** heterodimeric subunits each of which is held together by 5 additional protons. Furthermore, the tetramer evidenced a cooperative thermal transition characteristic of hemiprotonated C–C^+ base pairs, as found in i-motifs, by UV spectrophotometry at 295 nm (Fig. 3). Thus at pH 4.5, **R** and **P** in a 1 : 1 ratio forms a



Name	Sequence
P	H_2N-Lys-p(TCCCCC)-Lys-COOH
N-**TMR-P**	TMR-S-Cys –NH-p(TCCCCC)-Lys-COOH
N-**Dabcyl-P**	Dabcyl-HN-Lys-p(TCCCCC)-Lys-COOH
C-**Dabcyl-P**	NH_2-p(TCCCCC)-Lys-Lys-NH-Dabcyl
3′-**TMR-D**	5′-d(TCCCCC)-TMR-3′
R	5′-r(UUCCCCC)-3′
3′-**TMR-R**	5′-r(UUCCCCC)-TMR-3′

Fig. 1 Structure of the labels used in the study. Chart shows the sequences of RNA and PNA used.

National Centre for Biological Sciences, TIFR, GKVK, Bellary Road, Bangalore, 560065, India. E-mail: yamuna@ncbs.res.in; Fax: +91-80-23636662; Tel: +91-80-23666180

† Electronic supplementary information (ESI) available: Native PAGE, Circular Dichroism, UV melts, Fluorescence titration and raw data, 1D and 2D NMR. See DOI: 10.1039/b713525d

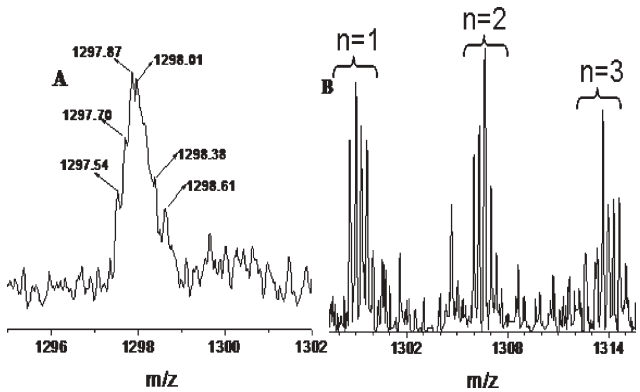


Fig. 2 Partial nano-ESI MS spectrum of an equimolar mixture of **R** and **P**, showing peaks corresponding to (A) a hexacharged tetramer $[2\mathbf{M_R} + 2\mathbf{M_P} + 10H^+ + 2NH_4^+]^{6+}$ and (B) Triply charged, multiply sodiated, dimer $[\mathbf{M_R} + \mathbf{M_P} + 5H^+ + Na^+]^{3+}$ where $\mathbf{M_R}$ and $\mathbf{M_P}$ are the molecular masses of **R** and **P** respectively.

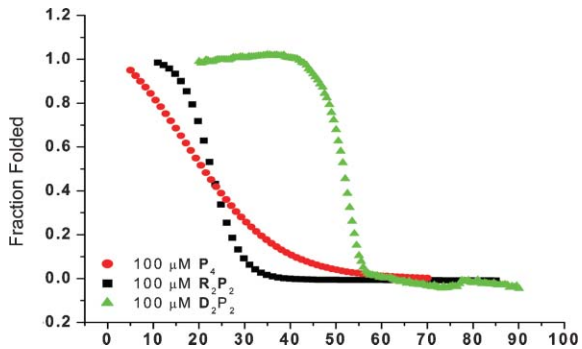


Fig. 3 UV denaturation profile at 295 nm of $\mathbf{P}_4$, $\mathbf{R_2P_2}$ and $\mathbf{D_2P_2}$ complexes at comparable strand concentration in 30 mM Acetate buffer, pH 4.5.

tetramer composed of two identical **RP** heterodimers held together by C–C^+ base pairs.

To elucidate the strand polarity in the $\mathbf{R_2P_2}$ hybrids, self-quenching experiments (SpexFluorolog-1) were performed by forming hybrids with fluorescently labelled tetramethylrhodamine (TMR) derivatives of **R** and **P**. TMR self quenches due to exciton coupling with an R_0 of 44 Å.[7] Thus in an $\mathbf{R_2P_2}$ complex, where one of the components is TMR-labeled, fluorescence quenching should provide an insight as to the relative like strand polarities. The observed distances obtained from all the possible combinations of labelled **R** and **P** are tabulated in Table 1.

In the hybrid incorporating 1 : 1 **P** : 3′-TMR-**R** (20 μM), the interfluorophore distance was found to be 50 Å which corresponded to an antiparallel alignment of RNA strands in the tetramer. Similarly, in the hybrid comprising 1 : 1 **N-TMR-P** : **R** (10 μM), an interfluorophore distance of ~43 ± 5 Å was obtained, revealing that the PNA strands are also arranged antiparallel to each other. Both **R**–**P** heteroduplexes can intercalate in two different configurations. In model 1 (Fig. 3) one narrow groove has both the RNA strands and the other, both the PNA strands. In Model 2, both narrow grooves have one RNA and one PNA backbone each. To address this, we formed a hybrid i-motif comprising 1 : 1 **3′-TMR-R** : **N-Dabcyl-P** (20 μM). Theoretical estimates of these interfluorophore distances, taking into account linker lengths and assuming groove dimensions commensurate with RNA_4 i-motifs, were found to be ~19 Å and ~9 Å for models 1 and 2 respectively. A quenching efficiency of ~85% was obtained indicating a TMR-Dabcyl separation of ~19 ± 5 Å, consistent with Model 1 (Fig. 4A). Quenching measurements on 1 : 1 **3′-TMR-R** and **C-Dabcyl-P** (20 μM) yielded a distance of ~24 Å (R_{calc} was ~23 Å) also consistent with Model 1.

To confirm the findings from fluorescence quenching measurements and to obtain proof of **RP** heteroduplex intercalation characteristic of i-motifs, 1 mM $\mathbf{R_2P_2}$, 30 mM d3-NaOAc, pH 4.5 was investigated by 1D and 2D NMR spectroscopy (Bruker AV 700 MHz). The 1D spectrum showed peaks in the region 15–16 δppm (SI), characteristic of the imino protons found in C–H^+C base pairs, confirming the findings from NanoESI-MS and UV melting data.[8] NOESY experiments on the $\mathbf{R_2P_2}$ complex showed several crosspeaks. Two distinct regions characteristic of **RP** heteroduplexation and intercalation respectively are shown. These are (i) the imino proton (NH_{im})–amino proton (NH_{2b} and NH_{2e}) regions (Fig. 4C) and (ii) the NH_{im}–NH_{im} region. Cytosine imino protons (NH_{im}) showed two sets of crosspeaks each to the hydrogen bonded (NH_{2b}) and non-hydrogen bonded amino protons (NH_{2e}). One set corresponded to the NH_{2b} and NH_{2e} protons of the RNA residues which were identified from their connectivities with rU2 while the other set corresponded to the NH_{2b} and NH_{2e} of the PNA residues, identified by their connectivity with pT1(CH_3) and backbone methylenes (data not shown). Given that only a single set of $U2_{im}$–$C3NH_{2b}$, $U2_{im}$–$C3NH_{2e}$ and $U2_{im}$–CH6 crosspeaks were observed (Fig. 4C, lower panel) this confirmed the existence of only a single conformer of $\mathbf{R_2P_2}$ i-motif in solution. This proves that the CH^+C base pairs are asymmetric. The formation of rCH^+–pC base pairs confirms heteroduplexation, which is in excellent correlation with the NanoESI-MS and fluorescence quenching experiments.

Table 1 Calculated and observed fluorescence intensities, interfluorophore distances and anisotropy values in labelled $\mathbf{R_2P_2}$ complexes

Labelled Hybrid complexes	Calc. Distance[a]/Å	Calc. Intensity[b] F_{DA}calc)	Obs. Intensity (F_{DA}obs)	Obs. Distance/Å	r
N-TMR-**P** + **R**	~48	2.6×10^4	1.8×10^4	43 ± 5	0.04
3′-TMR-**R** + **P**	~45	6.4×10^4	8.4×10^4	50 ± 5	0.04
3′-TMR-**R** + N-Dabcyl-**P**	~19	4.6×10^3	4.2×10^3	19 ± 5	0.07
3′-TMR-**R** + C-Dabcyl-**P**	~23	2.3×10^5	3×10^5	24 ± 5	0.07

[a] Distances were calculated from models of the hybrid RNA_2-PNA_2 i-motifs constructed using PyMOL software based on the NMR structure parameters of the RNA_4 i-motif.[10] Interfluorophore distances R_{calc}, were measured incorporating the linker connecting the fluorophores. The model used the linkers in an all-trans conformation, taking the furthest interfluorophore distance. [b] Expected intensity (F_{DA}^{calc}), based on the calculated quenching efficiency (E), for the calculated distance R_{calc}, from the experimentally obtained F_D. Distance accuracies are limited by the linker lengths of the freely rotating fluorophores represented in the error bar. All experiments were performed in triplicate and the average values are presented.

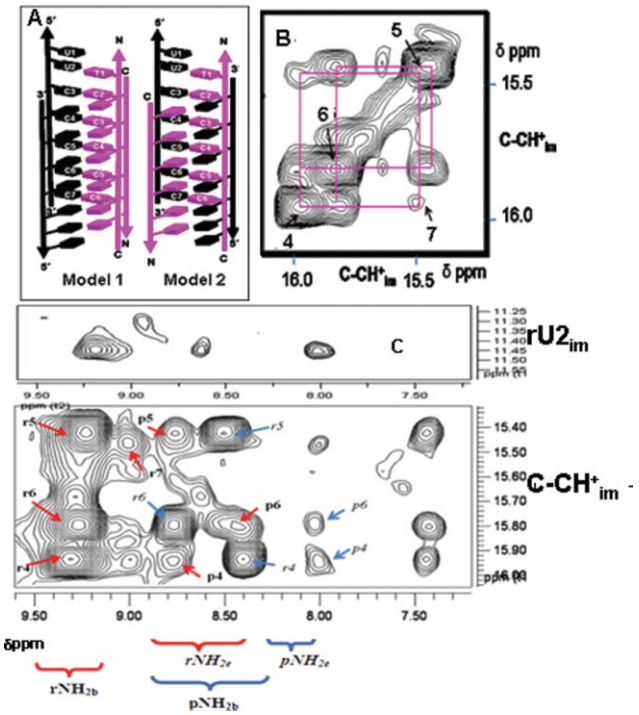


Fig. 4 (A) Two possible models of the **RP** hybrid i-motif. Model 1: Both RNA strands (purple) are in one narrow groove. Model 2: One RNA and one PNA strand (black) in each narrow groove. Two dimensional NOE spectrum of the $\mathbf{R_2P_2}$ hybrid i-motif at 1 mM strand concentration at 4 °C (200 ms mixing time). (B) NOE connectivities between four cytidine imino protons are shown, RNA residues (1–7) are indicated by the corresponding number. (C) Correlation of the NH_{2b}, NH_{2e}, with the imino proton (NH_{im}) region and rU2im. The crosspeaks of the RNA NH_{2b} protons of the rC–pC base pairs are labelled in bold **r4**, **r5**, **r6**, **r7** *etc*., while the corresponding RNA NH_{2e} protons are labelled in italics *r4*, *r5*, *r6*. Similarly crosspeaks of PNA NH_{2b} protons of rC–pC are labelled in bold **p4**, **p5**, **p6**, while the corresponding PNA NH_{2e} protons are labelled in italics *p5*, *p6*.

Intercalation of the heteroduplexes was proved by NOE crosspeaks shown by the rC_n–(NH_{im})–$pC_{(n-1)}$ protons (Fig. 4B).[9] The stacking order was identified as rU1–rU2–rC3–r**C7**–rC4–r**C6**–rC5–r**C5**–rC6–r**C4**–rC7–r**C3** in one minor groove and pC2–p**C6**–pC3–p**C5**–pC4–p**C4**–pC5–p**C3**–pC6–p**C2** in the other minor groove. Furthermore the observation of ribose sugar–sugar contacts evidenced by the H1′–H1′ as well as H1′–H2′ crosspeaks (ESI†) indicates that both the ribose containing strands are in very close proximity, at a distance similar to that found in RNA_4 i-motifs, confirming that both RNA strands flank a single narrow groove as in Model 1. All observed NOE connectivities are summarized in the schematic (ESI†). The structure of the $\mathbf{R_2P_2}$ i-motif is analogous to the 'M' form of RNA_4 i-motifs.[10] In the $\mathbf{R_2P_2}$ hybrid, one potential intercalation site between rU2 and rC3 is empty. The intercalation topology of the hybrid reflects the optimization between maximizing the stabilizing sugar-sugar contacts and minimizing the destabilizing 2′-OH steric clash in a background of constant electrostatics in a given narrow groove environment. Thus, the partially intercalated topology, reveals that i-motif intercalation topology is highly sensitive to narrow groove interactions, where, the elimination of a single destabilizing 2′-OH interaction completely favours this topology.[10]

We have created a new RNA-based non-Watson–Crick building block with potential application in structural RNA nanotechnology. *Via* hybridization to DNA and RNA, PNA bearing functional moeities have been used for several applications such as genotyping, protein identification and biosensing.[11] Thus, *via* the PNA component, this new buliding block has the scope to introduce a variety of moieties on a given RNA architecture to convert an inert RNA scaffold into a functional scaffold.[12]

Notes and references

1 S. Pitchiaya and Y. Krishnan, *Chem. Soc. Rev.*, 2006, **35**, 1111; B. Samori and G. Zuccheri, *Angew. Chem., Int. Ed.*, 2005, **44**, 1166; N. C. Seeman, *Biochemistry*, 2003, **42**, 7259; J. Bath and A. J. Turberfield, *Nat. Nanotechnol.*, 2007, **2**, 275.

2 Y. Krishnan-Ghosh, D. Liu and S. Balasubramanian, *J. Am. Chem. Soc.*, 2004, **126**, 11009; Y. Krishnan-Ghosh, E. Stephens and S. Balasubramanian, *Chem. Commun.*, 2005, 5278; S. Modi, A. H. Wani and Y. Krishnan, *Nucleic Acids Res.*, 2006, **34**, 4354; H. B. Ghodke, R. Krishnan, K. Vignesh, G. V. P. Kumar, C. Narayana and Y. Krishnan, *Angew. Chem., Int. Ed.*, 2007, **46**, 2646.

3 L. Jaeger and N. B. Leontis, *Angew. Chem., Int. Ed.*, 2000, **39**, 2521; B. Liu, S. Baudrey, L. Jaeger and G. C. Bazan, *J. Am. Chem. Soc.*, 2004, **126**, 4076; A. Chworos, I. Severcan, A. Y. Koyfman, P. Weinkam, E. Oroudjev, H. G. Hansma and L. Jaeger, *Science*, 2004, **306**, 2068.

4 R. T. Batey, R. P. Rambo and J. A. Doudna, *Angew. Chem., Int. Ed.*, 1999, **38**, 2326; N. Ban, P. Nissen, J. Hansen, P. B. Moore and T. A. Steitz, *Science*, 2000, **289**, 905.

5 K. Gehring, J. L. Leroy and M. Guéron, *Nature*, 1993, **363**, 561; K. Snoussi, S. Nonin-Lecomte and J. L. Leroy, *J. Mol. Biol.*, 2001, **301**, 139; A. J. Brazier, J. Fisher and R. Cosstick, *Angew. Chem., Int. Ed.*, 2006, **45**, 114; U. Diederichsen, *Angew. Chem., Int. Ed.*, 1998, **37**, 2273.

6 Single stranded **P** shows a molecular mass (M_P) of 1795.8 Da, while single stranded **R** shows an associated mass (M_R) of 2075.3 Da. These values were taken to compute molecular weight of the complex.

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

8 J. L. Mergny, L. Lacroix, X. Han, J. L. Leroy and C. Helene, *J. Am. Chem. Soc.*, 1995, **117**, 8887.

9 In this region, however, we were unable to detect the $rC3^+$–pC2 base pair which may be broadened out due to fast exchange as seen in the case of short-lived external CH^+–C base pairs.

10 RNA_4 i-motifs show two populations, partially and fully intercalated topologies in a 75 : 25 ratio. (K. Snoussi, S. Nonin-Lecomte and J. L. Leroy, *J. Mol. Biol.*, 2001, **301**, 139).

11 B. S. Gaylord, M. R. Massie, S. C. Feinstein and G. C. Bazan, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 34; J. Zielinski, K. Kilk, T. Peritz, T. Kannanayakal, K. Y. Miyashiro, E. Eiríksdóttir, J. Jochems, Û. Langel and J. Eberwine, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 1557; B. Liu and G. C. Bazan, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 589.

12 We thank the NanoScience and Technology Initiative of the Department of Science and Teschnology, Govt. of India and National Facility for High Field NMR, TIFR. SC and SM thank CSIR, Govt. of India for Junior Research Fellowships. YK thanks the Department of Biotechnology for the Innovative Young Biotechnologist Award.