



Journal Name

Supporting Information

Structural characteristics requisite for the ligand-based selective detection of i-motif DNA

Sagar Satpathi,^a Subrahmanyam Sappati,^a Konoya Das^a and Partha Hazra^{*a, b}

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

^a Department of Chemistry, Indian Institute of Science Education and Research (IISER), Pune. Dr. Homi Bhabha Road, Pashan, Pune, India 411008. Tel: +91-20-2590-8077. p.hazra@iiserpune.ac.in

^b Centre for Energy Science, Indian Institute of Science Education and Research (IISER), Pune (411008), Maharashtra, India

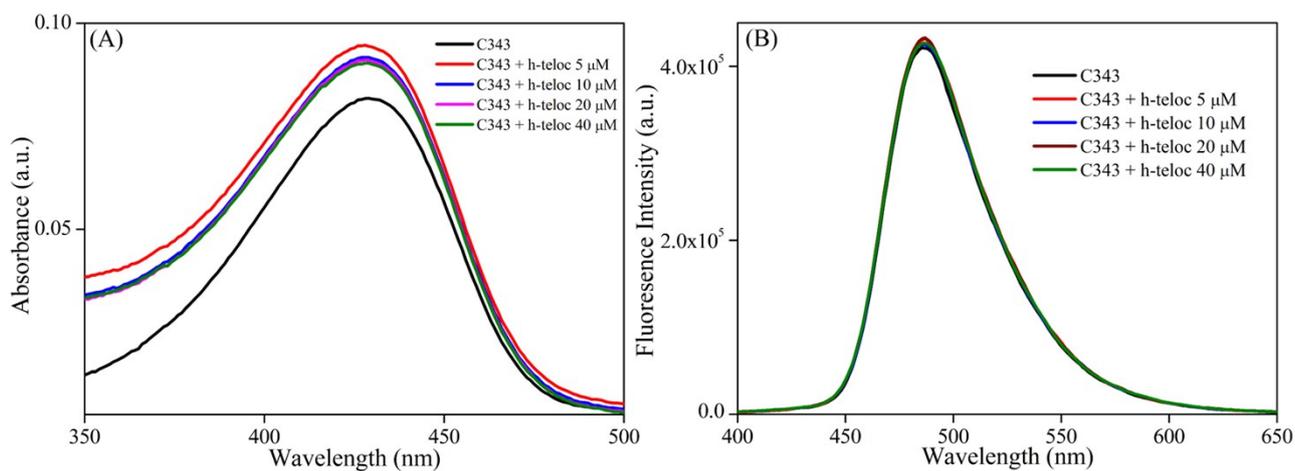


Fig. S1. (A) Absorption spectra and (B) fluorescence spectra (λ_{ex} 375 nm) of coumarin 343 (C343) with increasing concentration of h-teloc in pH 8.5 buffer (i.e. in the presence of ssDNA/RC structure).

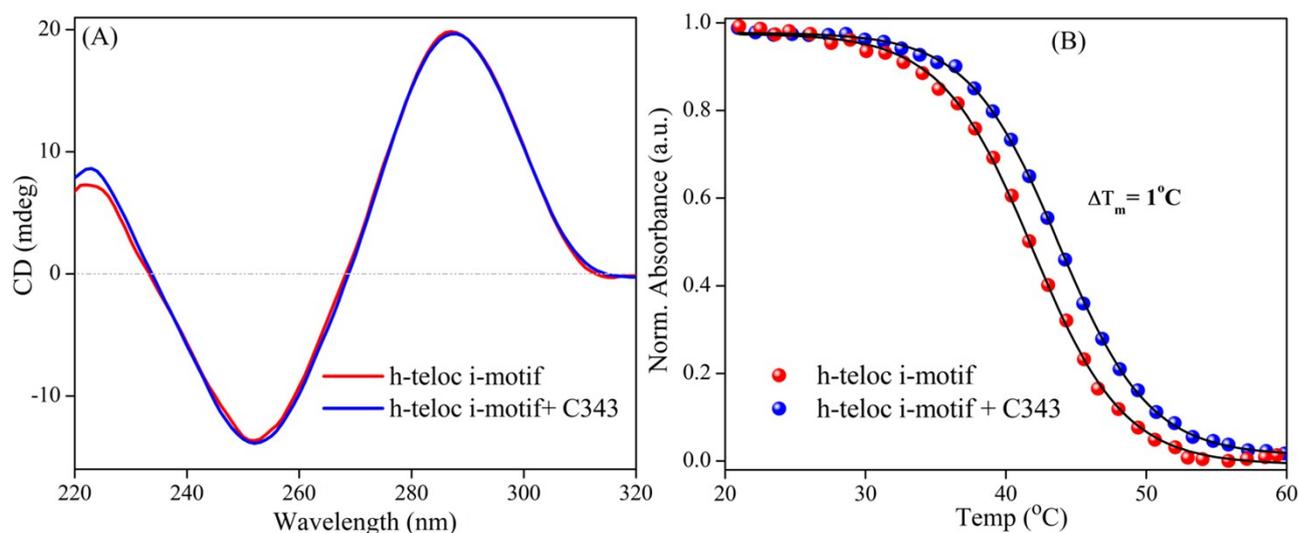


Fig. S2 (A) Circular dichroism (CD) spectra and (B) UV-melting spectra of coumarin 343 (C343) and h-teloc i-motif DNA in pH 5.2 buffer.

Note S1- Binding constant calculation for h-teloc i-motif DNA and C343 at pH 5.2 buffer:

We have calculated the binding constant for h-teloc i-motif DNA and C343 system utilising the fluorescence intensity change of C343 in the presence of i-motif DNA. We have determined the binding constant of this system according to the modified linear Scatchard equation,¹

$$\frac{[DNA]}{f} = \frac{1}{NK(1-f)} + \frac{[L]_{total}}{N}$$

Where, $f = \frac{F - F_0}{F_{max} - F_0}$; F_0 and F denote the relative emission intensities of C343 in the absence and presence of the h-teloc i-motif DNA in pH 5.2 buffer, respectively. F_{max} is the emission intensities of C343 in presence of maximum a concentration of h-teloc i-motif DNA in pH 5.2 buffer. $[DNA]$ is the h-teloc DNA concentration over the course of the experiment. $[L]_{total}$ is the concentration of C343. N is the number of C343 binding site per nucleotide. K (M^{-1}) represents the binding constant of the h-teloc i-motif DNA and C343 system.

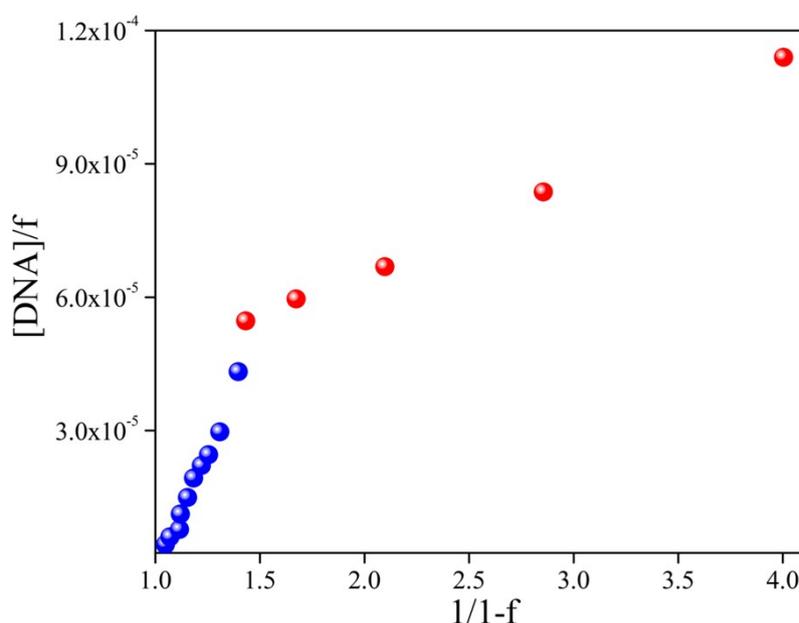


Fig. S3 Modified Scatchard plot of $[DNA]/f$ vs $1/(1-f)$ for the binding constant determination of h-teloc and C343 complex in pH 5.2 buffer.

For the modified linear Scatchard equation, the presence of only one kind of binding event is associated with one slope. However, the above binding profile clearly shows that the DNA-ligand complex formation involves two types of binding processes which is evident from the presence of two different slopes in the graph. The blue coloured data points represent the lower concentration ($< 10 \mu\text{M}$), and the red coloured points indicate higher concentrations of DNA ($> 10 \mu\text{M}$), respectively. Thus, we have performed the fitting of these two binding processes separately. It is important to mention that we have used other different types of binding constant equations and in every case, we observed a clear transition point going from lower to higher concentrations of DNA, denoting the existence of two different kinds of binding in the complex. In lower concentration, the binding constant of this DNA-ligand complex formation is found to be $6.2 \times 10^5 \text{ M}^{-1}$; whereas in higher concentration, the binding constant is $5.2 \times 10^5 \text{ M}^{-1}$. This type of concentration-dependent binding has also been observed previously in other systems.² We speculate that at a lower concentration of DNA, binding of C343 molecules occurs with the hemiprotonated cytosine bases. Whereas, groove or loop binding may predominate at higher concentrations due to the saturation of hemiprotonated cytosine bases.

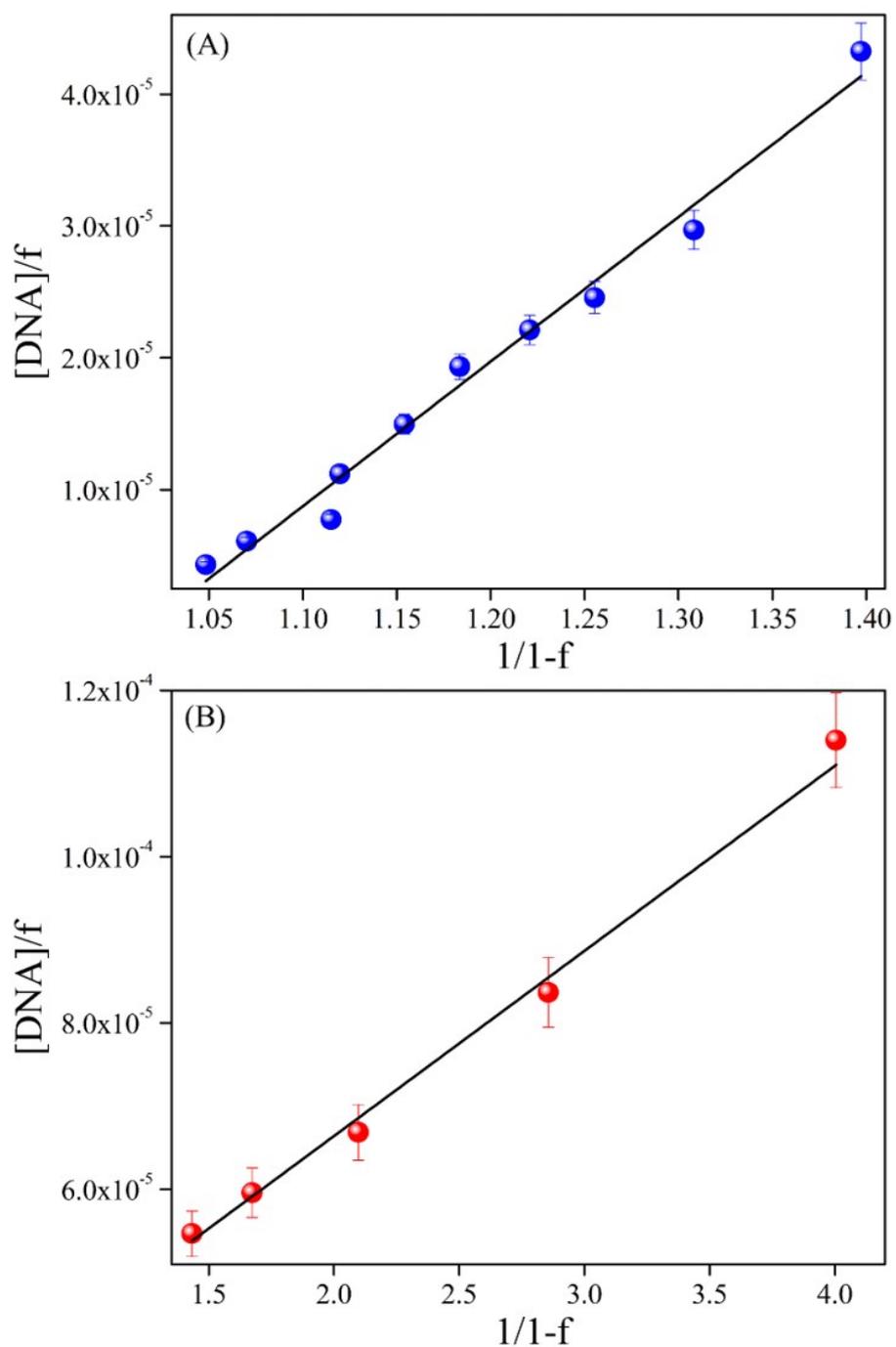


Fig. S4 Modified Scatchard plot of $[DNA]/f$ vs $1/(1-f)$ for the binding constant determination of h-teloc and C343 complex in pH 5.2 buffer (A) at lower concentrations ($< 10 \mu\text{M}$) and (B) at higher concentrations ($> 10 \mu\text{M}$).

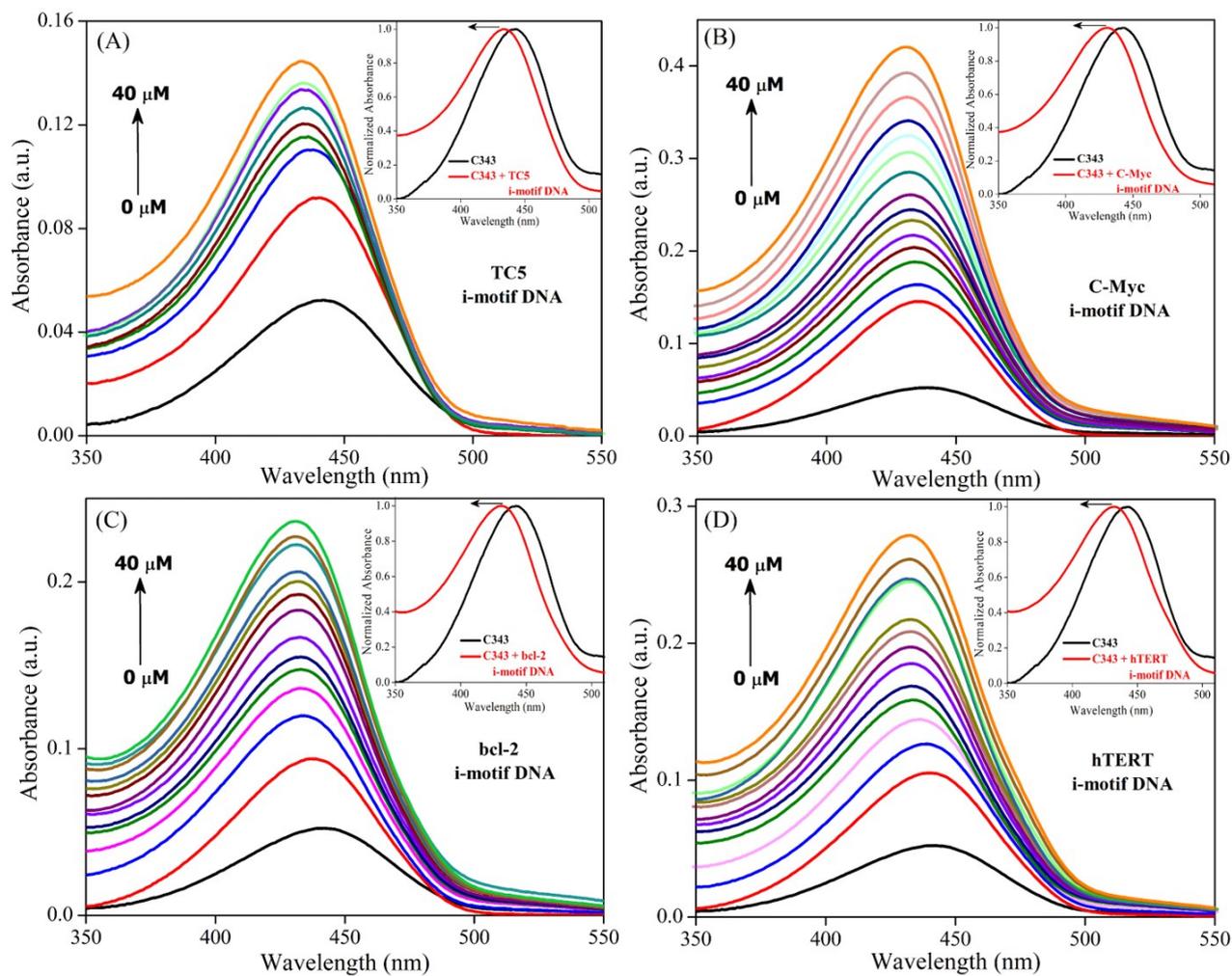


Fig. S5 Absorption spectra of coumarin 343 (C343) in pH 5.2 buffer with increasing concentration of different i-motif DNA, (A) TC5, (B) C-Myc, (C) bcl-2 and (D) hTERT.

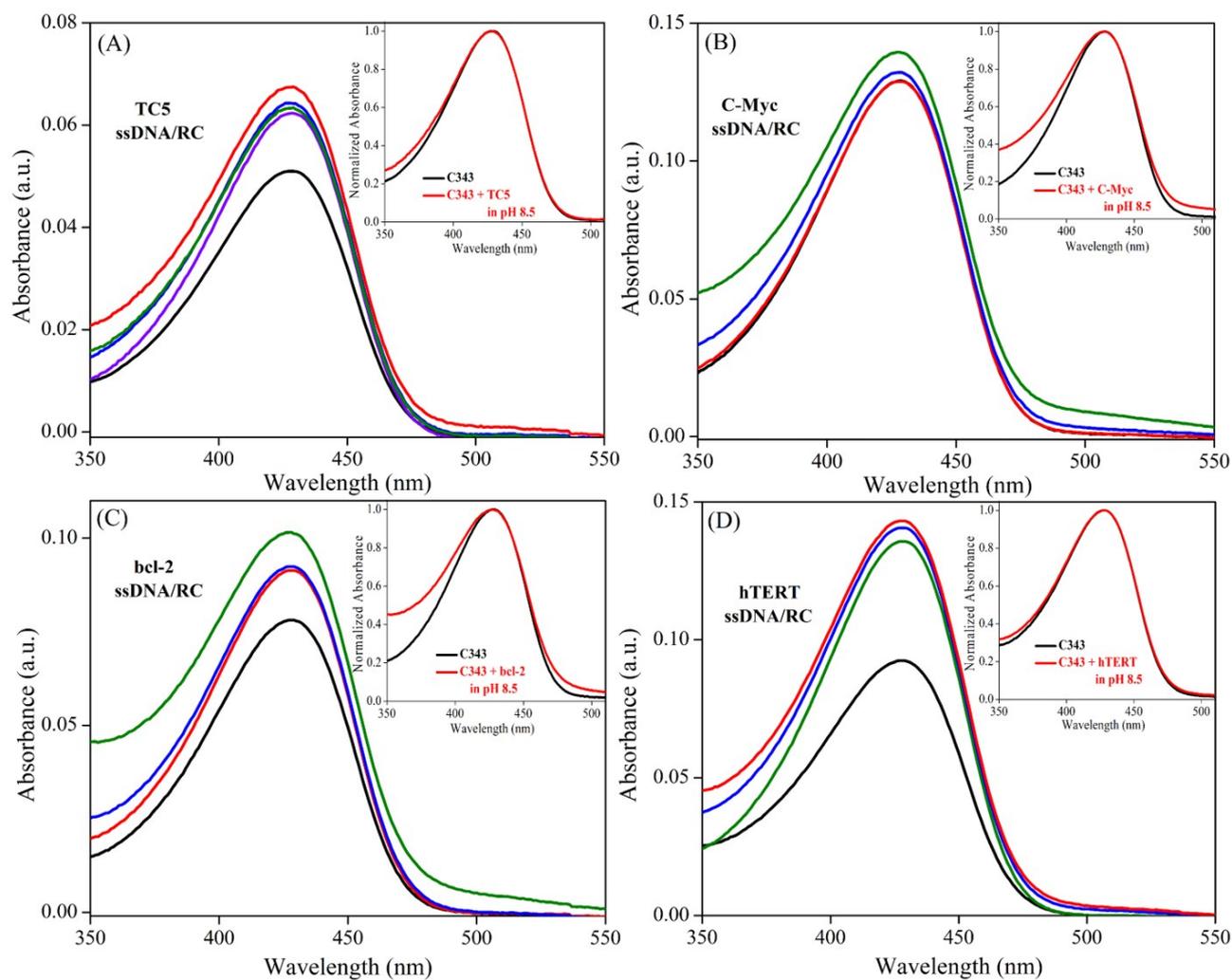


Fig. S6 Absorption spectra of coumarin 343 (C343) in pH 8.5 buffer with increasing concentration of different ssDNA/RC DNA (λ_{ex} 375 nm), (A) TC5, (B) C-Myc, (C) bcl-2 and (D) hTERT.

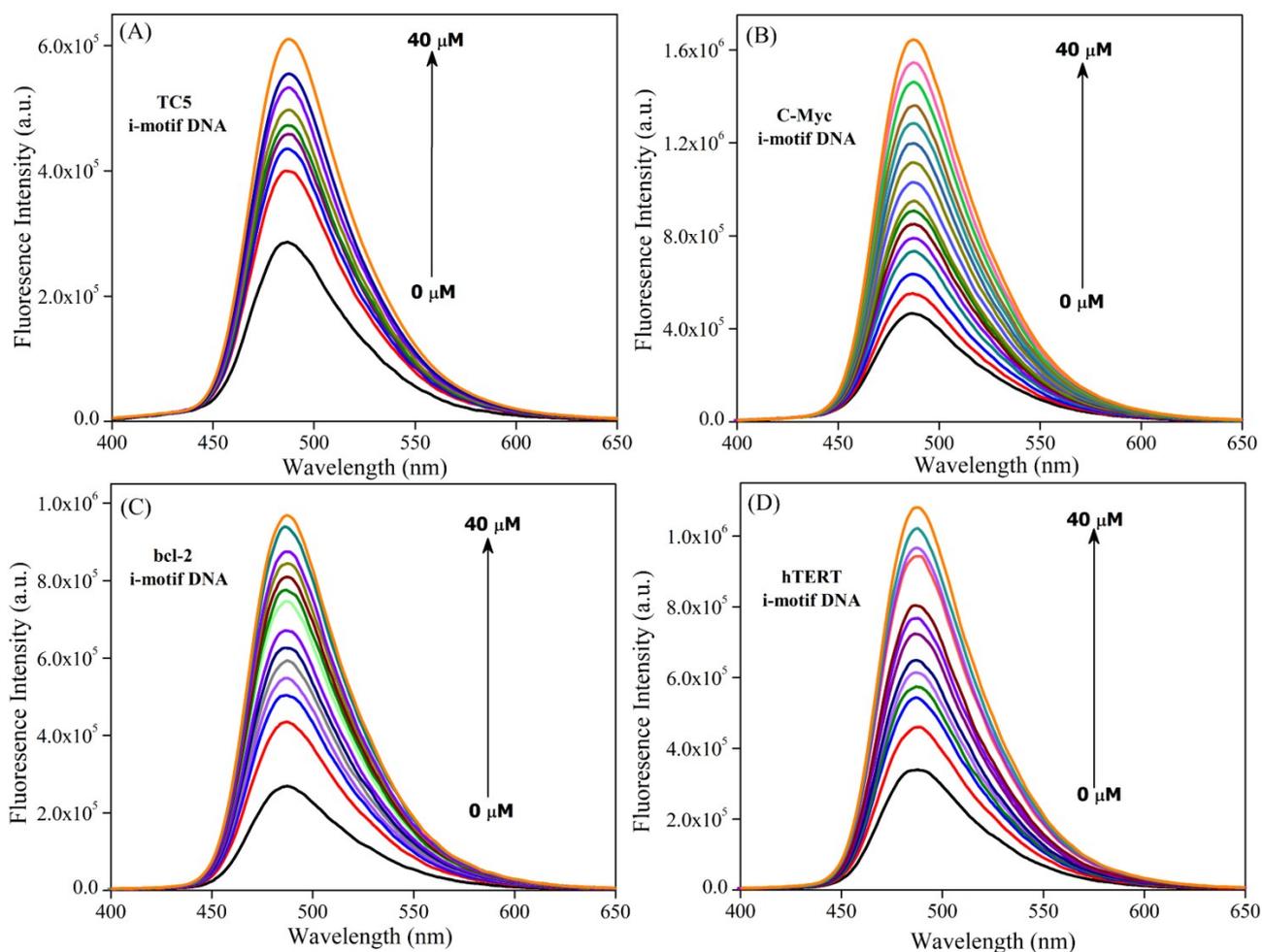


Fig. S7 Fluorescence spectra of coumarin 343 (C343) in pH 5.2 buffer with increasing concentration of different i-motif DNA (λ_{ex} 375 nm), (A) TC5, (B) C-Myc, (C) bcl-2 and (D) hTERT.

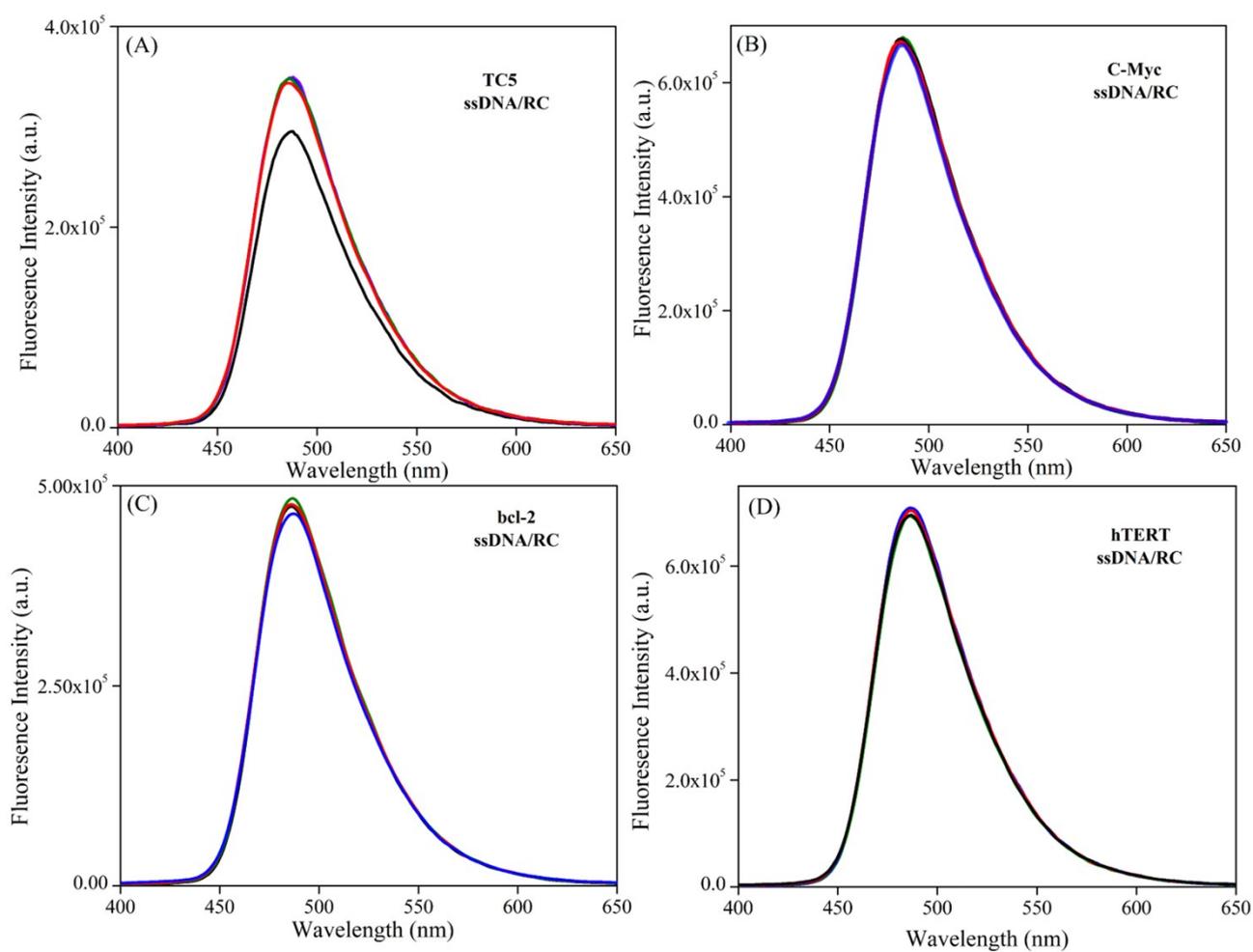


Fig. S8 Fluorescence spectra of coumarin 343 (C343) in pH 8.5 buffer with increasing concentration of different ssDNA/RC DNA (λ_{ex} 375 nm), (A) TC5, (B) C-Myc, (C) bcl-2 and (D) hTERT.

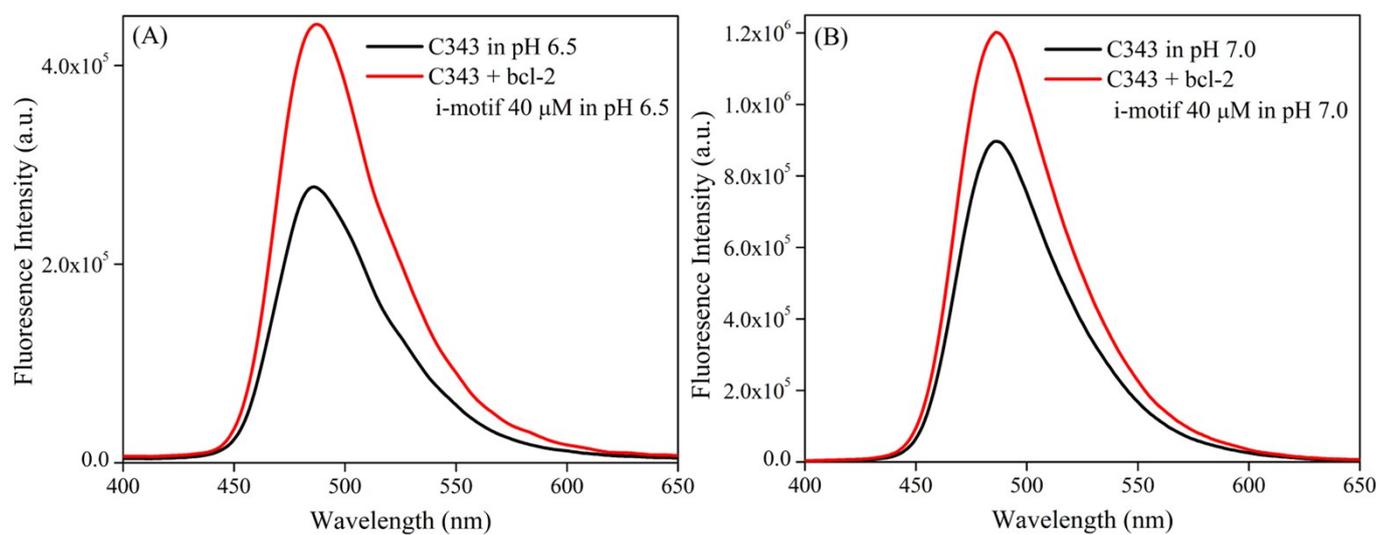


Fig. S9 Fluorescence response of C343, in presence of bcl-2 i-motif DNA (40 μ M) in (A) pH 6.5 and (B) pH 7.0 buffer.

Note S2- Explanation for the fluorescence intensity decrement with longer C-rich sequence:

This query can be best explained by looking at the respective i-motif structures of the different C-rich sequences which are given below,

DNA	Sequences (5' → 3')
h-teloc	d(CCC TAA CCCTAACCC TAA CCCTAA) class-I
C-Myc	d(CCC CACCT CC CA CCCT TCCCA CCCTCCCC) class II
bcl-2	d(CAGC CCC GCTCCCG CCCC TCCTCC GCGCCG CCCCT) class II

where red colour indicates the cytosine, base pairs involved in C-C⁺ base pairing and yellow colouring shows the rest of the nucleobases making the loops.

The different i-motif structures show that with increasing length of the C-rich sequences, the number of C-C⁺ base pairing does not increase drastically (no. of C-C⁺ base pairing: 6 for h-teloc, 6 for C-Myc and 7 for bcl-2). However, longer C-rich sequences form i-motif structures with bigger loops. For instance, h-teloc has a loop size of 5'-(3:3:3)-3', whereas C-Myc is having 5'-(6:2:6)-3' and bcl-2 having 5'-(8:5:7)-3' (x:y:z – no. of nucleobases in the loop). Based on their loop size, i-motif structures have been classified into different types, class-I for i-motif structures with shorter loops and class-II with larger loops. Thus h-teloc belongs to class-I, and class-II includes C-Myc and bcl-2.^{3,4}

Now, the fluorescence enhancement of C343 with different C-rich sequences depends on two types of interactions:

1. Electrostatic interaction between C-C⁺ base pairing of i-motif and carboxyl group of C343.
2. Electrostatic interaction between positively charged tertiary amine of C343 with the negatively charged phosphate backbone in the loop of the i-motif structure.

These two types of electrostatic interactions drastically reduce the non-radiative decay channels of C343, which in turn results in the emission intensity enhancement. Notably, in addition to the interaction between C-C⁺ base pairing of i-motif and carboxyl group of C343, the importance of the second type of

electrostatic interaction in the loop (between phosphate backbone and amine moiety of C343) towards the fluorescence enhancement can be understood by looking at the fluorescence response of C343 with shorter C-rich sequence, i.e., TC5. In case of TC5, the absence of loop structure resulted in the less fluorescence intensity enhancement compared to other i-motif structures having loops (Fig. 2A). With increase in loop size from h-teloc i-motif to C-Myc and bcl-2 i-motif (i.e. from 3 bases to more than 6 bases in the loop), the extent of enhancement in fluorescence intensity of C343 decreases due to the increment in nonradiative decay channels as a result of the less rigidity afforded to the C343 molecules in bigger sized loops of C-Myc and bcl-2. Relatively smaller loops of h-teloc i-motif interact strongly with the C343 molecules and restrict the rotational motion of the C343 to a larger extent in comparison to C-Myc and bcl-2.

Now, we have also realized that presence of only loop structures in DNA do not result in the fluorescence enhancement of C343, which is highlighted in the lack of intensity enhancement for G-quadruplex and duplex DNA (i.e. ds26) structure (in later Fig. 3, Fig. 4, Fig. S10). This lack of fluorescence enhancement is attributed to the electrostatic repulsion between negatively charged carboxylic group of C343 and negatively charged phosphate backbone of DNA. Similarly, presence of only negative charge in coumarin molecules i.e. for C3COOH (see below Fig. S14) do not show any fluorescence enhancement with i-motif structure. These observations indicate that a balance between positively charged tertiary amine moiety and negatively charged phosphate backbone along with the electrostatic attraction between negatively charged carboxylate group and hemi-protonated C–C⁺ base pairs in coumarins is extremely necessary for selective i-motif detection. Presence of both the factors (i.e. number of C–C⁺ base pairs and smaller loops) is favorable for human telomere sequence (h-teloc) which results in the maximum fluorescence enhancement of C343. Increment in the loop length of the C-Myc and bcl-2 i-motif structures result in slightly lesser fluorescence enhancement, where the C–C⁺ base pairs remains almost unchanged. For hTERT i-motif structure (belonging to class-III type), the number of nucleobases in the loops is expected to increase even more drastically, as it has been speculated that for this sequence two i-motif structures are formed and connected through a long loop of 26 nucleobases.³ Hence, the rigidity experienced

by C343 molecules in the loops of hTERT i-motif decreases even more which results in a further increment of the nonradiative decay channels and decrement in radiative channels, i.e. fluorescence intensity. It has to be noted that for all the different i-motif structure experiments, the concentration of C343 is kept the same and since the number of C-C⁺ base pairing also does not change drastically from shorter to longer sequence. Thus, we believe that the bigger sized loops in the i-motif structure are responsible for their lesser extent of emission intensity.

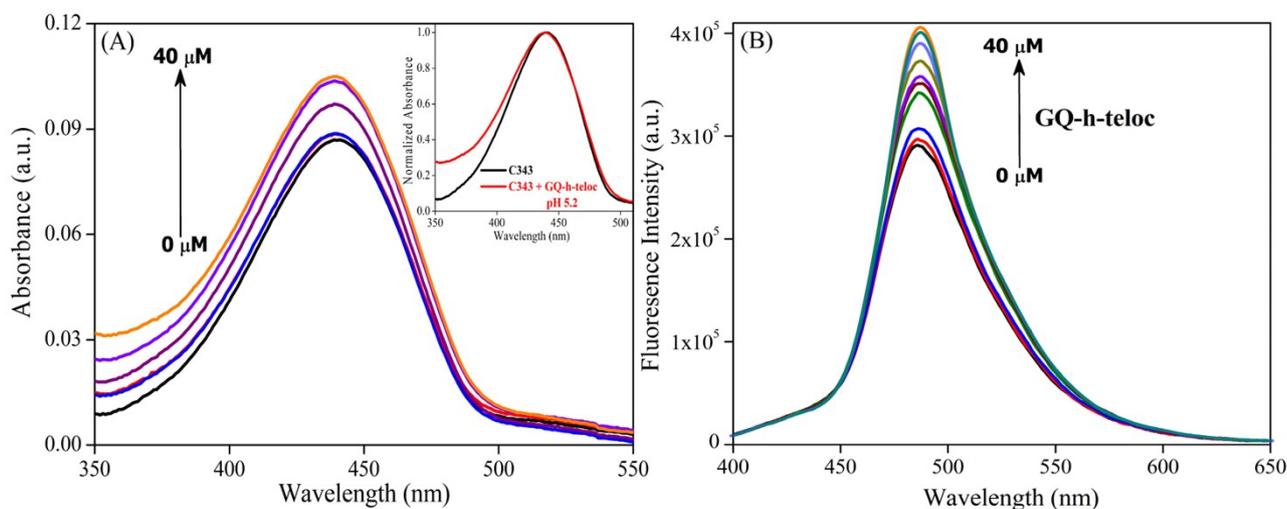
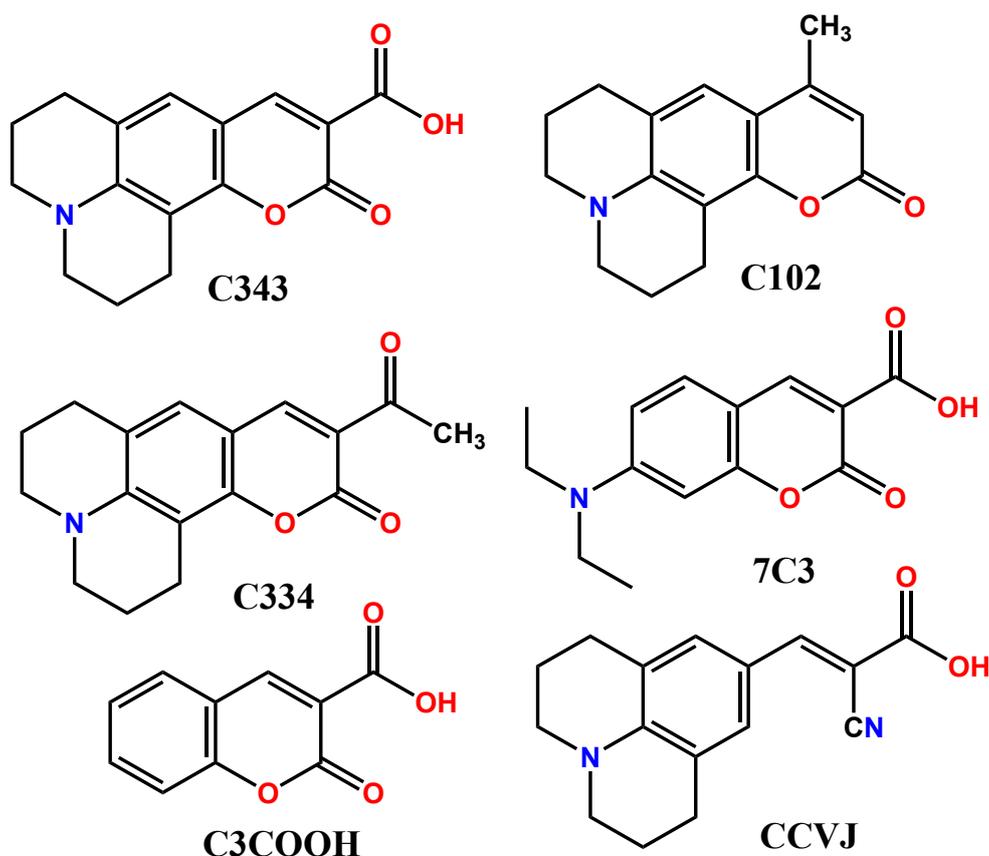


Fig. S10 (A) Absorption spectra and (B) fluorescence spectra of coumarin 343 (C343) with increasing concentration of GQ-h-teloc DNA in pH 5.2 buffer (λ_{ex} 375 nm). Inset of Fig. S9 (A) shows the absence of any shift for the C343 absorption in the presence of GQ-h-teloc DNA.

Table S1. Different DNA sequences used in this study.

DNA	Sequences (5' → 3')
GQ-h-teloc DNA	d(TTAGGGTTAGGGTTAGGGTTAGGG)
GQ-C-Myc DNA	d(GGGGAGGGTGGGGAGGGTGGGGAAGGTGGGG)
ds 26	d(CAATCGGATCGAATTCGATCCGATTG)

Note S3- Systematic study of the different coumarin derivatives with i-motif and ssDNA/RC DNA:



Scheme S1. Chemical structure of different molecules used in this study.

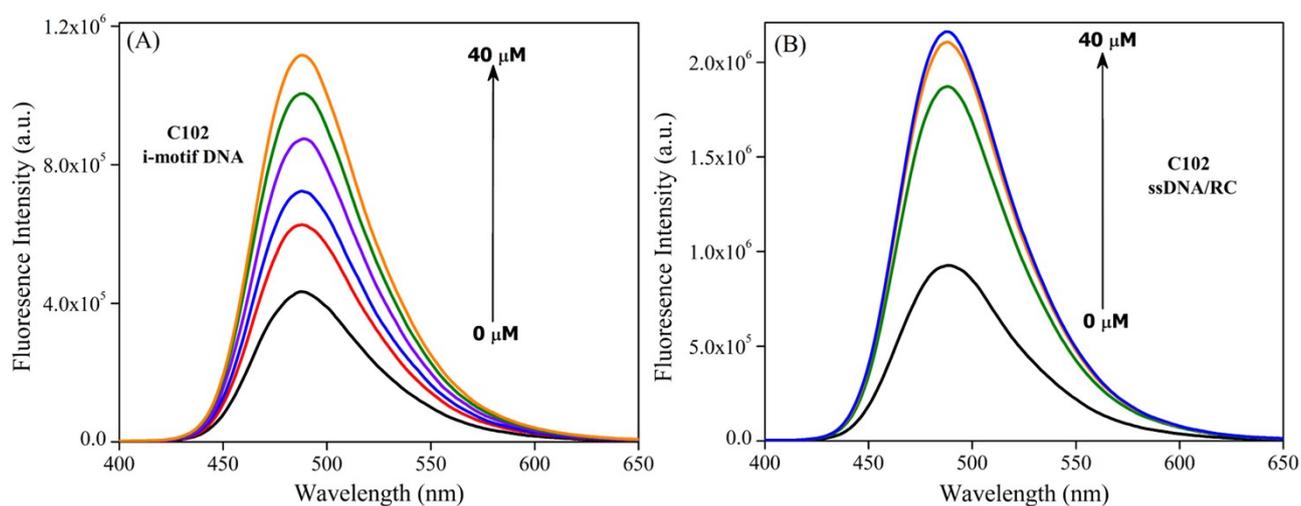


Fig. S11 Fluorescence response (λ_{ex} 375 nm) of coumarin 102 (C102) in the presence of (A) i-motif DNA at pH 5.2 buffer and (B) in the presence of ssDNA/RC DNA at pH 8.5 buffer.

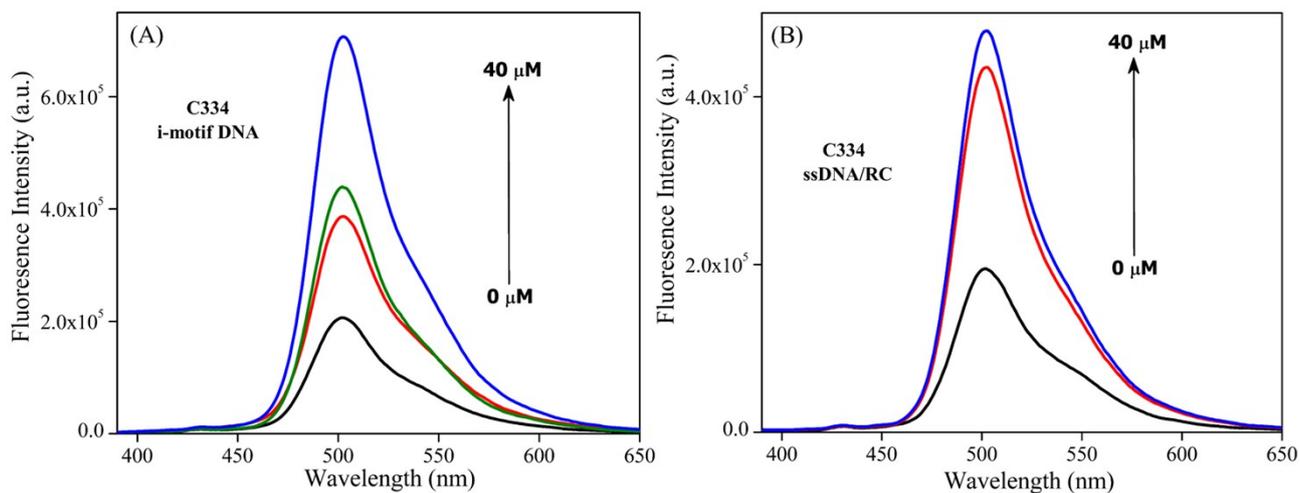


Fig. S12 Fluorescence response (λ_{ex} 375 nm) of coumarin 334 (C334) in the presence of (A) i-motif DNA at pH 5.2 buffer and (B) in the presence of ssDNA/RC DNA at pH 8.5 buffer.

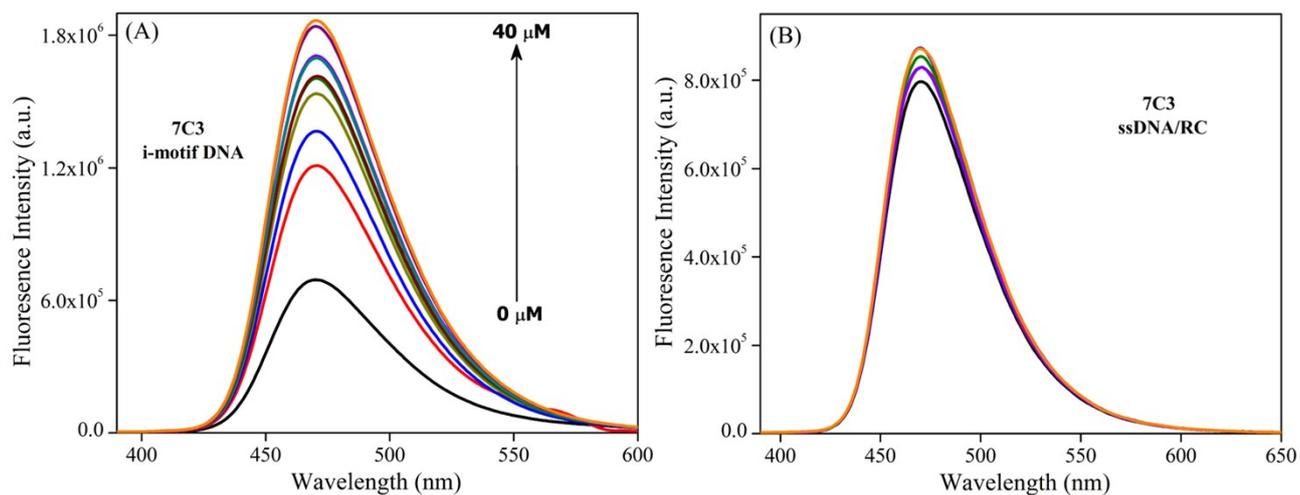


Fig. S13 Fluorescence response (λ_{ex} 375 nm) of 7-(Diethylamino)coumarin-3-carboxylic acid (7C3) in the presence of (A) i-motif DNA at pH 5.2 buffer and (B) in the presence of ssDNA/RC DNA at pH 8.5 buffer.

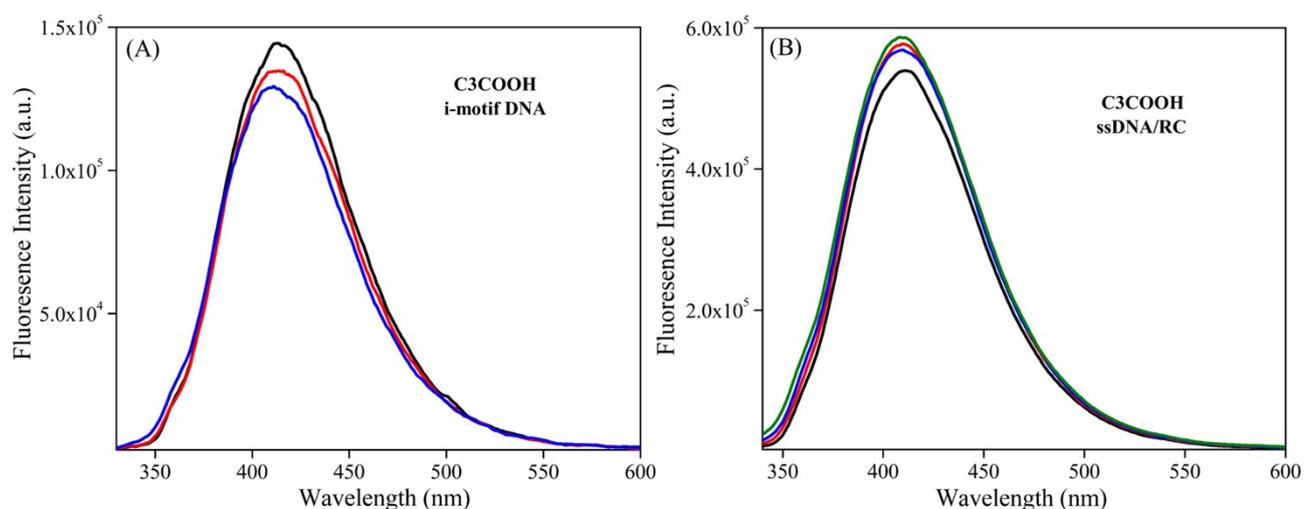


Fig. S14 Fluorescence response (λ_{ex} 375 nm) of Coumarin-3-carboxylic acid (C3COOH) in the presence of (A) i-motif DNA at pH 5.2 buffer and (B) in the presence of ssDNA/RC DNA at pH 8.5 buffer.

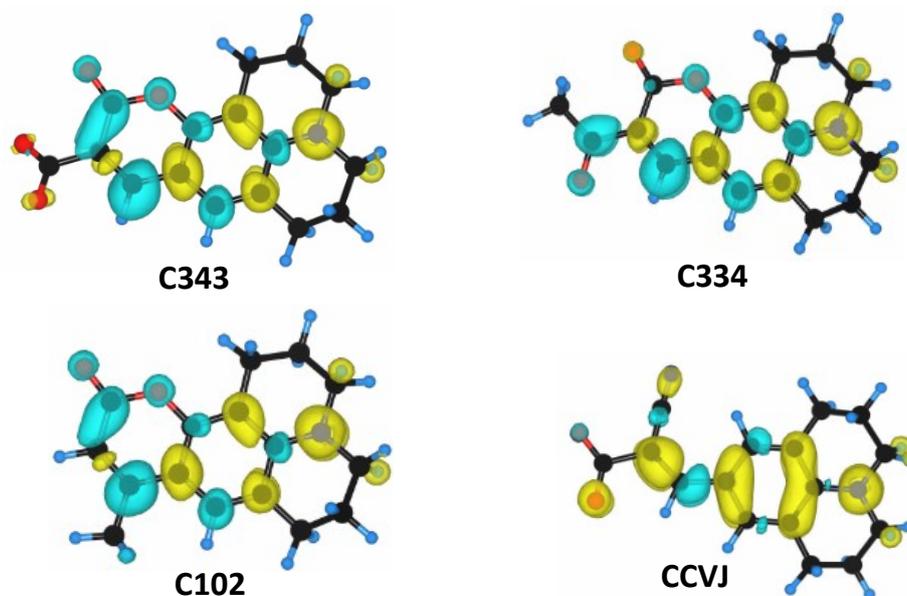


Fig. S15 Iso-surfaces of the electron density difference map (EDDM) of the different molecules showing the changes in electronic density of the dye due to an excitation.

We have performed the fluorescence studies to monitor the interaction between different coumarin derivatives with both the i-motif DNA (at pH 5.2 buffer) and the ssDNA/RC DNA (at pH 8.5 buffer). In the case of both C102 and C334, the emission intensity increases with increasing concentration of both the i-motif and ssDNA/RC DNA (Fig. S11 and Fig. S12). Whereas, for 7C3 (Fig. S13), this type of enhancement was only observed in the presence of i-motif DNA (at pH 5.2 buffer) and not with the ssDNA/RC DNA (at pH 8.5 buffer). This indicates that 7C3 is selectively interacting with i-motif DNA. However, we do not observe any such enhancements for C3COOH (Fig. S14) with either the i-motif DNA (at pH 5.2 buffer) or the ssDNA/RC DNA (at pH 8.5 buffer). Now, to clarify these observations, we have calculated the electron density difference maps (EDDM) of these coumarin structures (Fig. S15). Interestingly, we are observing that these coumarin molecules exhibit a charge transfer process, where the amine moiety acts as the donor and the carboxylate part as the acceptor. Thus, these coumarin molecules exhibit a positive charge character at the nitrogen centre of the amine part which easily interacts with both the negatively charged i-motif and ssDNA/RC DNA. Presence of an additional negative charged functionality (i.e. a carboxylic group in C343, 7C3 and CCVJ i.e. Fig. 6) brings the selectivity towards i-motif DNA over ssDNA/RC structure. This observed selectivity in the presence of additional negative functionality can be explained by two competitive electrostatic interactions, one being between the positively charged tertiary amine of the molecule with the negatively charged phosphate backbone and the other being, the negatively charged carbonyl moiety of the coumarins and the hemi-protonated cytosine bases. However, in case of the ssDNA/RC structure, an additional electrostatic repulsion with the negative moiety of the coumarin molecules comes into play which may be responsible for the lack of interaction between 7C3, C343 with the ssDNA/RC structure. It is important to note that the coumarin molecules containing only negatively charged functionality (C3COOH) do not interact with either of the DNA forms. This implies that the electrostatic attraction with the negatively charged phosphate backbone also plays an important role to interact with any of the DNA forms. Thus, a balance between positive and negative character is necessary for a molecule to act as a selective i-motif ligand.

Note S4- Method for electron density difference maps (EDDM) calculation:

Computational Details:

All the computational calculations were performed with density functional theory (DFT) in Gaussian 09 software.⁵ The geometries of different molecules were optimised and characterised as the energy minima at B3LYP/6-311G++(d,p) level in the gas phase by frequency analysis.⁶ Further, we computed absorption spectra by Time-dependent DFT (TD-DFT) by B3LYP functional.⁷ However,

conventional global hybrid functional such as B3LYP often suffer from the charge-transfer (CT) problem and consequently give a poor description of electronic absorption characteristic. Hence, in this study, we have utilised CAM-B3LYP hybrid functional⁸ along with conventional hybrid functional (B3LYP). We have also incorporated implicit solvent models by SCRF method using IEFPCM model for water solvent.⁹ We have computed absorption spectra for 15 number of states. First two prominent (bright) transitions are given in the Table S2. Later, we computed electron density difference map for first bright transition (having considerable oscillatory strength) by Gauss-Sum software.¹⁰ An electron density difference map (EDDM) is a representation of the changes in electron density that occur for a given electronic transition. i.e., the subtract the electron density before an electronic transition from the electron density after an electronic transition. It gives a charge transfer associated with a given transition which is given in the following Table S2. In the plots, the cyan and yellow colour iso-surfaces denote the charge accumulation and depletion, respectively. We maintained the same iso-value for all the EDDM plots.

Table S2. Important computational parameters for the EDDM calculations.

Sample	Wavelength (nm)	Oscillator Strength	Important transitions
C343	344.781	0.7291	HOMO->LUMO (95%)
	281.059	0.0223	H-1->L (30%), H->L+4 (14%), H->L+5 (25%)
C334	378.841	0.9016	HOMO->LUMO (95%)
	261.298	0.1015	H-1->L (13%), H->L+1 (25%), H->L+2 (35%)
C102	336.308	0.5848	HOMO->LUMO (95%)
	277.411	0.0244	H-1->L (32%), HOMO->L+4 (14%), H->L+5 (29%)
CCVJ	301.597	0.4992	H-1->L+1 (11%), H-1->L+2 (24%), H-1->L+4 (48%)
	322.386	0.0344	H-1->LUMO (81%)

References

- (a) E. F. Healy, *J. Chem. Edu.*, 2007, **84**, 1304. (b) R. S. Harapanhalli, L. W. McLaughlin, R. W. Howell, D. V. Rao, S. J. Adelstein, and A. I. Kassis, *J. Med. Chem.*, 1996, **39**, 4804.
- S. N. Sahu, S. K. Padhana and P. K. Sahu, *RSC Adv.*, 2016, **6**, 90322.
- T. A. Brooks, S. Kendrick and L. Hurley, *FEBS J.*, 2010, **277**, 3459.

4. A. S. Gouda, M. S. Amine and E. B. Pedersen, *Org. Biomol. Chem.*, 2017, **15**, 6613
5. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr. J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, Revision B.05; Gaussian, Inc.: Pittsburgh, PA, 2003.
6. (a) A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098. (b) C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785.
7. T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51.
8. N. M. O'Boyle, A. L. Tenderholt and K. M. Langner, *J. Comp. Chem.*, 2008, **29**, 839.
9. G. Scalmani, M. J. Frisch, *J. Chem. Phys.*, 2010, **132**, 114110.
10. Time-dependent Density Functional Theory, Lecture Notes of Physics; Marques, M. A. L., Ullrich, C. A., Nogueira, F., Rubio, A., Burke, K., Gross, E. K. U., Eds.; Springer: Berlin, 2006.