Stimuli Mediated Conformational Transitions and Detection of Different Secondary Structures of DNA

A Thesis

Submitted in Partial Fulfillment of the Requirements for the

Degree of

Doctor of Philosophy

by

Sagar Satpathi

Reg. ID: 20143302

IISER PUNE

Indian Institute of Science Education and Research (IISER), Pune

2019

Dedicated to

Maa, Baba, Dada and Dadu

Certificate

It is hereby certified that the work described in this thesis entitled *"Stimuli Mediated Conformational Transitions and Detection of Different Secondary Structures of DNA"* submitted by *Mr. Sagar Satpathi* was carried out by the candidate, under my supervision. The work presented here or any part of it has not been included in any other thesis submitted previously for the award of any degree or diploma from any other university or institution.

Date: 9th May, 2019

Dr. Partha Hazra
Research Supervisor
Email: p.hazra@iiserpune.ac.in
Contact No.:+91(20)25908077

Declaration

I declare that this written submission represents my ideas in my own words and wherever other's ideas have been included; I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/ data/ fact/ source in my submission. I understand that violation of the above will cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Date: 9th May, 2019

Sagar Satpathi

Reg. Id: 20143302

Acknowledgement

At the very outset, I would like to convey my deepest gratitude to my research supervisor Dr. Partha Hazra for his continuous support and motivation throughout my PhD tenure. I feel his ceaseless encouragement has driven my research in forward direction. It has been an immense privilege for me to work under his supervision. His inspiring guidance and belief in me have helped me to design and execute different research projects successfully. Thank you very much Sir for your continual efforts, trust in me and most importantly for being such a wonderful guide.

I sincerely acknowledge the former director of Indian Institute of Science Education and Research (IISER) Pune Prof. K. N. Ganesh and the current director Prof. J. B. Udgaonkar for providing world class facilities conducive for a vibrant research environment. I am also extremely thankful to my research advisory committee (RAC) members, Dr. Arnab Mukherjee (IISER, Pune), Dr. Sayan Bagchi (NCL, Pune) for their valuable suggestions, advices and support during the annual RAC meetings. I am also grateful to Prof. M. Jayakannan (Chair, Chemistry department) and all other faculty members of the department of chemistry for extending their help when required and maintaining such a research friendly, healthy environment in the department. I sincerely thank all the collaborators - especially Dr. Arnab Mukherjee (IISER, Pune), Dr. Mandar Kulkarni, Mr. Reman K Singh and Mr. Subrahmanyam Sappati, Mr. Partha Samanta who assisted me in different projects. Their collaborative efforts have increased the content of my research. I would also like to express my gratitude towards Nitin for helping me with NMR samples scanning. This would also be a great time to acknowledge the efforts of the Chemistry Office staff, especially Tushar, Mayuresh, Prabhash in keeping the office related tasks smooth and hassle free.

I am indebted to Cefipra for providing me with the opportunity to work with Dr. Jean Louis Mergny under the Raman-Charpak Fellowship, 2017. I am immensely grateful to Dr. Jean Louis Mergny, IECB, University of Bordeaux for hosting me under the aforementioned fellowship. This chance of working with one of the pioneers of my research field has benefited me immeasurably, contributing towards making me a better research scholar. I also thank the members of Dr. Mergny's lab for teaching me new techniques and adding to my scientific knowledge. I also take this chance to acknowledge DST-SERB and Infosys foundation for providing me with financial support to attend an international conference, BIONIC-2018.

I am highly thankful to all my present and past lab members, for a supportive and friendly

Table of Contents

1. Introduction: Polymorphism of DNA

2. A Green Solvent Induced DNA Package

List of Abbreviations

CD	Circular Dichroism
DCM	Dichloromethane
DLS	Dynamic Light Scattering
DNA	Deoxyribonucleic Acid
GQ	G-quadruplex
IL	Ionic Liquid
WC	Watson-Crick
min	Minutes
μM	Micromolar
mM	Millimolar
TCSPC	Time-correlated Single Photon Counting
ns	Nanosecond
ps	Picosecond
RTIL	Room Temperature Ionic Liquid
PB	Phosphate Buffer
UV	Ultraviolet
FE-SEM	Field Emission Scanning Electron Microscopy
TEM	Transmission electron microscopy
NMR	Nuclear Magnetic Resonance
RNA	Ribonucleic Acid
Lyz	Lysozyme
ssDNA	Single-stranded DNA
RC	Random Coil
EDDM	Electron Density Difference Maps
RMSD	Root Mean Square Deviation
RDF	Radial Distribution Function
SDF	Spatial Distribution Function

Synopsis

The primary motivation of all the works included in this thesis is to control and detect the conformational transition between different secondary structures of DNA which will be advantageous in terms of their participation in different biologically important processes. Structural characteristics of these different DNA forms have been utilized to modulate their conformational transitions by some external stimuli which also gave us an excellent opportunity to understand the molecular mechanism behind the formation and stabilization of these different DNA structures.

Chapter 1. Introduction: Polymorphism of DNA

In this section, a brief description of the different secondary structures of DNA has been discussed, where we have highlighted how the biological functions of these different structures are heavily dependent on their structure and stability. Importantly, the structure of these different DNA forms often undergoes conformational transitions to another form depending on the favourable conditions significantly impacting their stability, which is very crucial in terms of their participation in biologically important processes such as genome recombination, cancer cell immortality and regulation of gene expression *etc.* Similarly, detection of these DNA conformations is also essential which will enable us to explore their activity by tracking it in intracellular conditions. In the continuation of this, all the works are majorly focused on controlling and detecting the conformational transitions of different DNA secondary structures by using some external stimuli.

Chapter 2. A Green Solvent Induced DNA Package

In this chapter, we have focused a room temperature ionic liquid as DNA compaction agent, where we have shown that guanidinium tris(pentafluoroethyl)trifluorophosphate (Gua-IL) acts as a DNA compacting agent. The compaction ability of Gua-IL has been verified by different spectroscopic techniques, like steady state emission, circular dichroism, dynamic light scattering and UV melting. Notably, we have extensively probed this compaction by Gua-IL through field emission scanning electron microscopy (FE-SEM) and fluorescence microscopy images. We also have discussed the plausible compaction mechanism process of DNA by Gua-IL. Our results suggest that Gua-IL forms a micellar kind of self-aggregation above a certain concentration ($\geq$1 mM), which instigates the DNA compaction process. This study divulges the specific details of DNA compaction mechanism by a new class of compaction agent, which is highly biodegradable and ecofriendly in nature.
Sci. Rep., 2015, **5**, 9137.
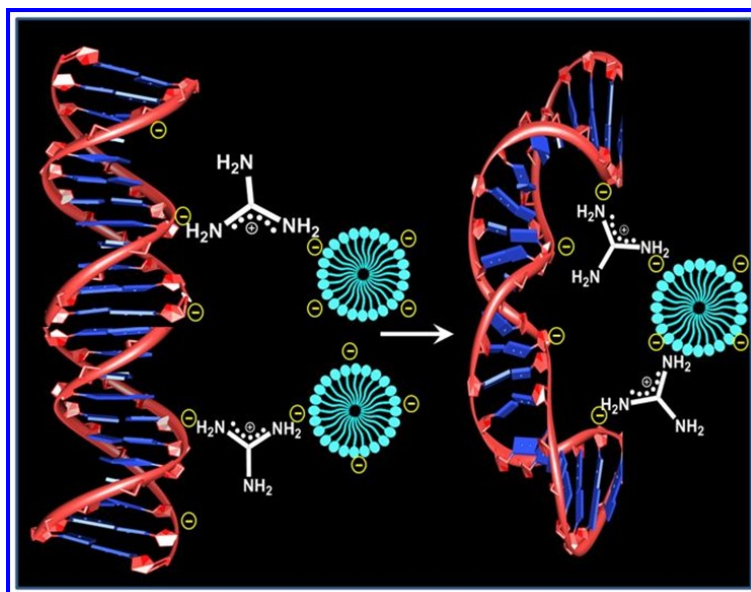
Chapter 3A. Ionic Liquid Induced G-quadruplex Formation and Stabilisation: Spectroscopic and Simulation Studies

In this chapter, we have reported that an ionic liquid (IL), *i.e.*, guanidinium tris(pentafluoroethyl)-trifluorophosphate, can instigate the GQ formation in the absence of conventional GQ forming ions (like Na^+, K^+, NH_4^+, *etc.*), and also stabilizes the GQ structure. This conformational transition has been confirmed through different spectroscopic tools (such as, CD, thermal melting, steady state and time resolved fluorescence *etc.*) and molecular dynamics (MD) simulation studies. MD simulation shows that one of the guanidinium cations resides in the G-tetrad core, while bulky anions prefer to stay near the GQ surface resulting in GQ formation and stabilization. The origin of GQ stabilization by IL presented here may also help in the future design of IL for GQ formation and stabilization.

Phys. Chem. Chem. Phys., 2016, **18**, 29740.
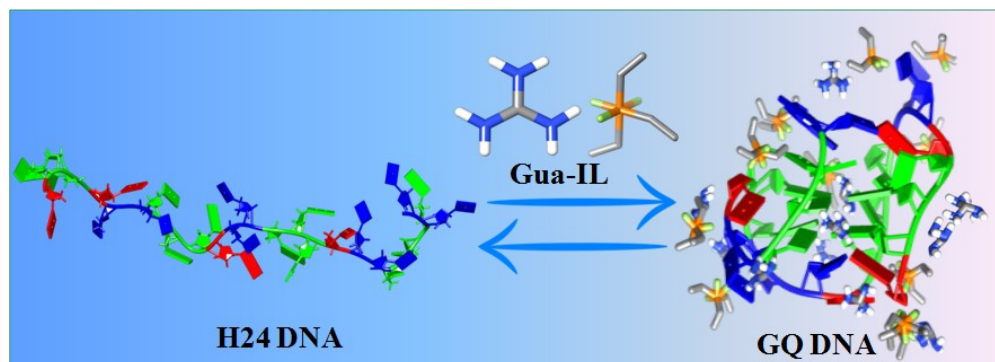
Chapter 3B. Controlling an Anticancer Drug Mediated G-quadruplex Formation and Stabilisation by a Molecular Container

In this chapter, we have tried to control the ligand mediated G-quadruplex DNA (GQ DNA) formation and stabilization, which is an important and challenging aspect due to its active involvement in many biologically important processes such as DNA replication, transcription, etc. We have demonstrated that topotecan (TPT), a potential anticancer drug, can instigate the formation and stabilization of GQ DNA (H24 → GQ DNA) in the absence of Na^+/K^+ ions via circular dichroism, fluorescence, NMR, UV melting and molecular dynamics (MD) simulation studies. The primary binding mode of TPT to GQ was found to be stacking at the terminal rather than binding to the groove. We have also reverted this conformational transition (GQ DNA → H24) using a molecular container, cucurbit[7]uril (CB7), by means of the translocation of the drug (TPT) from GQ DNA to its nanocavity. Importantly, we have carried out the detection of these conformational transitions using the fluorescence color switch of the drug, which is more direct and simpler than some of the other methods that involve sophisticated and complex detection techniques.

Phys. Chem. Chem. Phys., 2018, **20**, 7808.
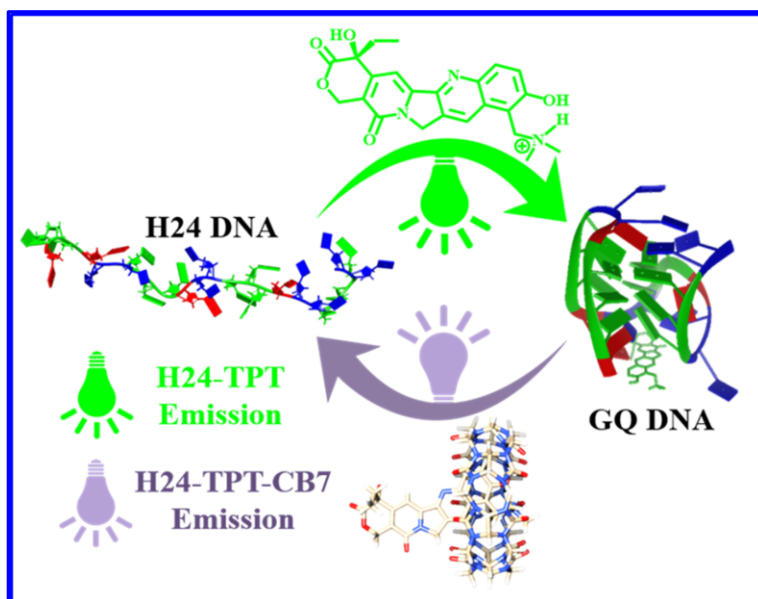
Chapter 4A. Silica Nano-channel Induced i-motif Formation and Stabilization at Neutral and Alkaline pH

Here, we have developed a new strategy to stabilize i-motif DNA in neutral and alkaline media by incorporating C-rich sequences inside silica nano-channels. Subsequently, the reversibility of this conformational transition has been achieved using a positively charged protein. Importantly, this entire

conformational transition can be performed in multiple cycles, which offers an alternative way to control i-motif formation other than pH and thermal annealing.
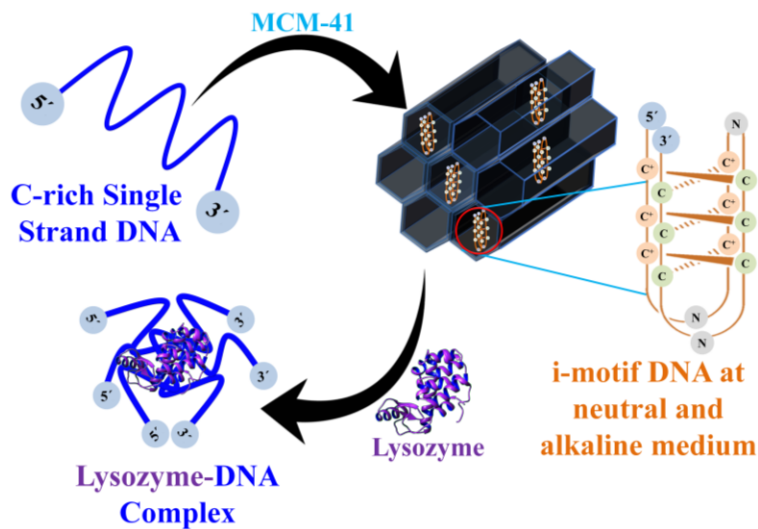
Chapter 4B. Structural characteristics requisite for the ligand based selective detection of i-motif DNA
we have explored the light-up property of coumarin 343 selectively towards various i-motif DNA (including intermolecular and intramolecular) based on the recognition of hemi-protonated cytosine–cytosine base pairing. A systematic study between i-motif DNA with various coumarin derivatives helps us to understand the structural characteristics required for an ideal ligand which can be useful for future designing of any i-motif DNA ligands.
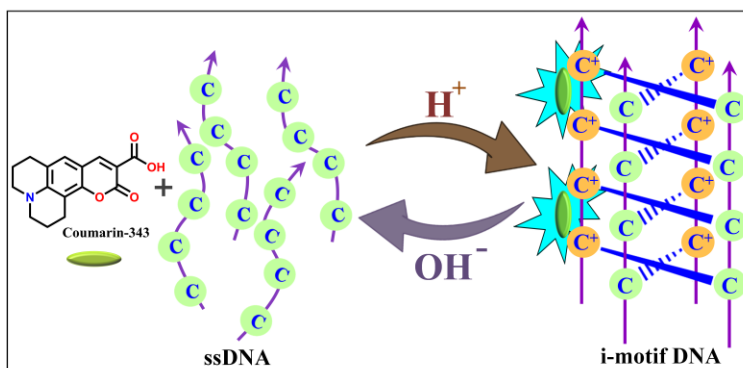
Rights and Permissions

Chapter 1:

Figure 1.4 – Reproduced with permission from *Science*, 1982, **216**, 475. Copyright 1982: Science.
Figure 1.8 – Reproduced with permission from *Chem. Commun.*, 2015, **51**, 7581. Copyright 2015: Royal Society of Chemistry.

Chapter 2:

Reprinted (adapted) with permission from *Sci. Rep.,* 2015, **5**, 9137. Copyright 2015: Nature Publishing Group.

Chapter 3A:

Reprinted (adapted) with permission from *Phys. Chem. Chem. Phys.,* 2016, **18**, 29740. Copyright 2016: Royal Society of Chemistry.

Chapter 3B:

Reprinted (adapted) with permission from *Phys. Chem. Chem. Phys.,* 2018, **20**, 7808. Copyright 2018: Royal Society of Chemistry.

Chapter 4A:

Reprinted (adapted) with permission from *Chem. Commun.,* 2018, **54**, 7054. Copyright 2018: Royal Society of Chemistry.

Chapter 4B:

Reprinted (adapted) with permission from *Org. Biomol. Chem.,* 2019, 10.1039/C9OB01020C. Copyright 2019: Royal Society of Chemistry.

Research Publications

Included in Thesis:

1. **Sagar Satpathi,** P. Hazra*, Structural Characteristics Requisite for the Ligand-based Selective Detection of i-motif DNA, *Org. Biomol. Chem.,* 2019, 10.1039/C9OB01020C.

2. Sagar Satpathi, K. Das, P. Hazra*, Silica Nano-channel Induced i-motif Formation and Stabilization at Neutral and Alkaline pH, *Chem. Comm.,* 2018, **54,** 7054.

3. Sagar Satpathi, R. Singh, A. Mukherjee, P. Hazra*, Controlling an Anticancer Drug Mediated G-quadruplex Formation and Stabilization by a Molecular Container, *Phys. Chem. Chem. Phys.,* 2018, **20**, 7808.

4. Sagar Satpathi, M. Kulkarni, A. Mukherjee, P. Hazra*, Ionic liquid Induced G-quadruplex Formation and Stabilization: Spectroscopic and Simulation Studies, *Phys. Chem. Chem. Phys.,* 2016, **18**, 29740.

5. Sagar Satpathi, A. Sengupta, V. M. Hridya, K. Gavvala, R. K. Koninti, B. Roy, P. Hazra*, A Green Solvent Induced DNA Package, *Scientific Reports,* 2015, **5,** 9137.

Not Included in Thesis:

6. **Sagar Satpathi**, K. Gavvala P. Hazra*, Fluorescence Switching of Sanguinarine in Micellar Environments, *Phys. Chem. Chem. Phys.,* 2015, **17,** 20725.

7. **Sagar Satpathi**, K. Gavvala, P. Hazra*, Fluorescence Up-Conversion Studies of [2,2′-Bipyridyl]-3,3′-diol in Octyl-β-d-glucoside and Other Micellar Aggregates, *J. Phys. Chem. A,* 2015, **119,** 12715.

8. **Sagar Satpathi**, Partha Hazra*, Dual Proton Transfer Dynamics of the Pyranine Derivative in Bulk and Restricted Media, Manuscript under Preparation (2018).

9. R. Koninti, **Sagar Satpathi**, P. Hazra*, Ultrafast Fluorescence Dynamics of Highly Stable Copper Nanoclusters Synthesized inside Aqueous Nanopool of Reverse Micelles, *J. Phys. Chem. C,* 2018, **122,** 5742.

10. K. Das, **Sagar Satpathi**, P. Hazra*, Effect of Bile Salt Aggregates on the Prototropic Equilibria of Harmine, a Fluorescent β-Carboline Alkaloid, *ChemistrySelect,* 2018, **3**, 1697.

11. R. K. Koninti, S. Palvai, **Sagar Satpathi**, S. Basu, P. Hazra*, Loading of an Anti-cancer Drug into Mesoporous Silica Nano-channel and Subsequent Release to DNA, *Nanoscale,* 2016, **8**, 18436.

12. B. Roy, **Sagar Satpathi**, P. Hazra*, Topological Influence of Lyotropic Liquid Crystalline Systems on Excited-state Proton Transfer Dynamics, *Langmuir,* 2016, **32**, 3057.

13. K. Gavvala*, **Sagar Satpathi**, Acetylcholine Induced Interplay of Proflavine Between Cucurbit [7] uril and DNA, *J. Lumin,* 2016, **171**, 234.

14. R. K. Koninti, S. Sappati, **Sagar Satpathi**, K. Gavvala, P. Hazra*, Spectroscopy and Dynamics of Cryptolepine in the Nanocavity of Cucurbit[7]uril and DNA, *ChemPhysChem*, 2016, **17**, 506.

15. B. Roy, **Sagar Satpathi**, K. Gavvala, R. K. Koninti, P. Hazra*, Solvation Dynamics in Different Phases of the Lyotropic Liquid Crystalline System, *J. Phys. Chem. B*, 2015, **119**, 11721.

16. K. Gavvala, **Sagar Satpathi**, P. Hazra*, Ultrafast Dynamics of a Molecular Rotor in Chemical and Biological Nano-cavities, *RSC Adv.*, 2015, **5**, 72793.

17. K. Gavvala, **Sagar Satpathi**, P. Hazra*, pH Responsive Translocation of an Anticancer Drug Between Cyclodextrin and DNA, *RSC Adv.*, 2015, **5**, 98080.

18. K. Das, B. Roy, **Sagar Satpathi**, P. Hazra*, Impact of Topology on the Characteristics of Water Inside Cubic Lyotropic Liquid Crystalline Systems, *J. Phys. Chem. B,* 2019, 10.1021/acs.jpcb.9b01559.

Chapter 1

Introduction: Polymorphism of DNA

1.1 Introduction

For the first time in 1869, a young swiss bio-chemist named Friedrich Miescher discovered the presence of DNA in cells.[1] Later in 1953, based on the 80 years of accumulated experimental results, Watson and Crick discovered the model structure of duplex DNA (i.e. B-DNA).[2] According to the Watson-Crick model, these duplex DNA structures are the combination of two antiparallel single-stranded DNA connected through complementary base pairing (Figure 1.1 and Figure 1.2), where adenine (A) forms two hydrogen bonding with thymine (T) and guanine (G) forms three hydrogen bonding with cytosine (C).[2] These type of observed H-bonding in the duplex DNA is known as Watson-Crick base pairing,[2] where G-C base pairing energy (-23.8 kcal/mol) is higher than A-T base pairing energy (-11.8 kcal/mol) due to more number of H-bonding in the former case.[3] Later in 1959 using X-ray diffraction, Karst Hoogsteen observed an alternate base pairing between 1-methylamine and 9-methyladenine unlike Watson-Crick base pairing (both nucleobase in anti-conformations),[4] where in general two nucleobases exist in syn and anti-conformations, respectively (Figure 1.3). Recently, this type of Hoogsteen base pairing has been also identified in normal duplex DNA with very less population using nuclear magnetic resonance relaxation dispersion spectroscopy [5] and during DNA methylation process with high population through single-crystal X-ray diffraction method.[6] It has been observed that during the DNA
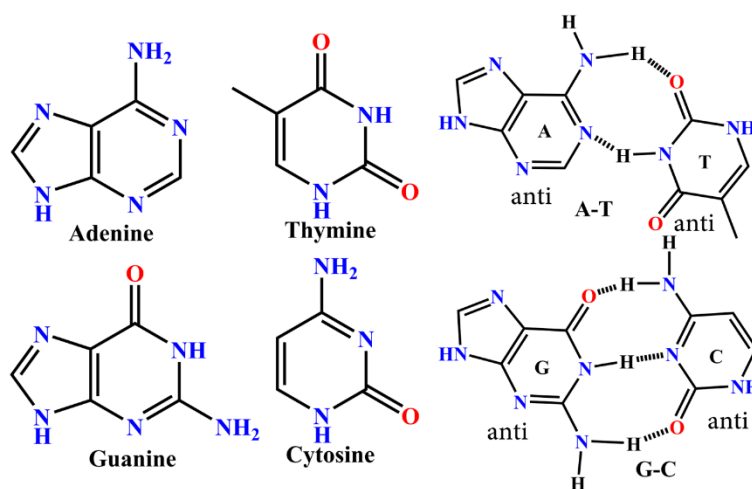
Figure 1.1 Chemical structures of four nucleobases in DNA and their Watson-Crick base pairing observed in duplex DNA.

metabolism processes (such as replication, transcription etc.), these DNA double helical structures partially unfold to form single-stranded DNA.[7] Depending on the sequence motifs and interaction with different proteins, these single-stranded DNA structures adopt a number of unusual secondary structures of DNA (non B-DNA) due to their dynamic nature, where the base pairing in these structures is generally found to be Hoogsten H-bonding rather than Watson-Crick base pairing. Usually, these secondary structures contain a repetitive DNA sequence to undergo different types of folding based on the appropriate conditions which bring the special importance of these secondary structures due to the higher abundance (more than 50% of the total genomic DNA) of this repetitive sequence in the

eukaryotic genome.[8] To understand the biological importance of these secondary structures of DNA, we have to keep this fact in mind that single-stranded DNA is the active form during the metabolism
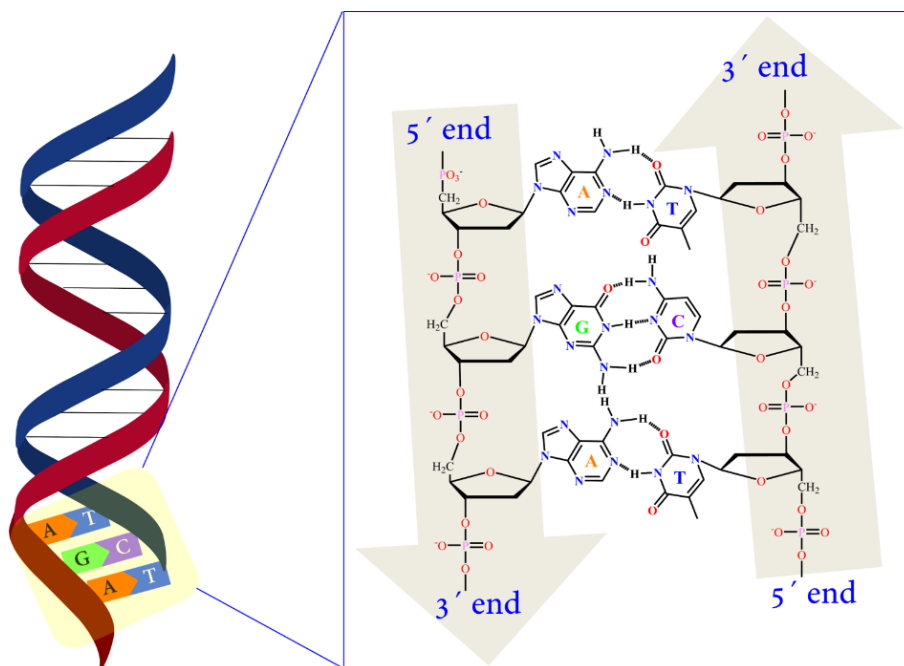
Figure 1.2 Schematic representation of conventional duplex DNA and observed Watson-Crick base pairing in duplex DNA.

processes such as, DNA replication, DNA repair, recombination and DNA transcription *etc.*[9] These single-stranded DNA structures are kinetically allowed to fold in other alternative structures which are generally stabilized by appropriate conditions and other proteins. Previously, it has also been supported by Wang *et al.* that these repetitive sequences for the formation of these secondary structures are more prone to induce genomic instability which often has direct relevance with human diseases. Importantly, these unique secondary structures can also act as a control for gene expression owing to its influences in many biologically important processes such as replication, transcription, translation *etc.*[9a, 9d] Apart from the biological relevance of these non B-DNA structures, these secondary structures have been majorly utilised in material chemistry and DNA nanotechnology, starting from to build an artificial nanostructure to construct a nano-machines or molecular logic gates.[10] Till now, apart from the conventional duplex DNA, a number of DNA secondary structures have been reported in the literature,
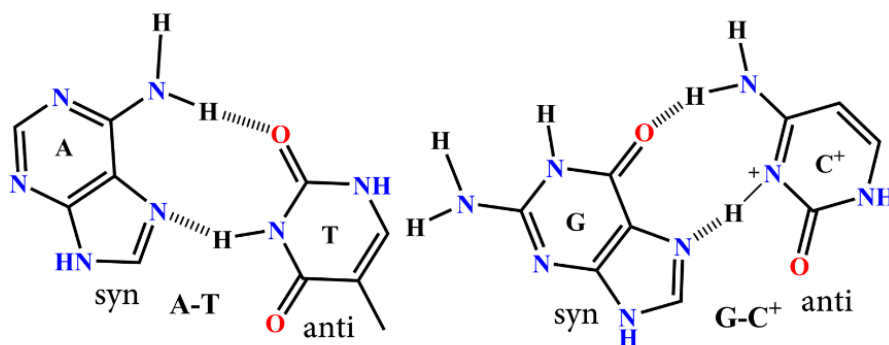
Figure 1.3 Hydrogen bonding observed in Hoogsteen base pairing.

such as, left-handed Z-DNA, A-DNA, cruciform, hairpin, A-motif, triplex (H-DNA), G-quadruplex and i-motif *etc.*[7] which I will discuss briefly in the following,

1.1.1 Left-handed Z-DNA

During the discovery of Watson-Crick base pairing in 1953,[2] the structures of the nucleic acids are only determined by the DNA fibre X-ray diffraction analysis, which was believed to be not sufficient for the full structure determination. In late 1970, synthesis of DNA sequences becomes a reality with the advancement of molecular biology techniques, which allows to explore the single crystal X-ray diffraction study on any particular DNA molecules to determine the structure. Surprisingly in 1979, Rich and coworkers observed that the first single-crystal X-ray structure of a self-complementary DNA fragment (i.e. d(CGCGCG)) adopts a left-handed double helix instead of the conventional right-handed duplex DNA, where two antiparallel single-stranded DNA sequences are held together by Watson-Crick base pairing (Figure 1.4).[11] This left-handed duplex DNA is known as Z-DNA. In comparison to the conventional duplex DNA (all the nucleobases in anti-conformation), nucleobases in the Z-DNA helix exist in alternate syn- and anti-conformations giving rise to a closely spaced phosphate group orientation in Z-DNA than in B-DNA (Table 1.1).[12] Hence, the sequence with alternate purine-pyrimidine bases generally shows the transitions from B-DNA to Z-DNA structure, where the pyrimidine bases prefer anti-conformations and purine can exist in either syn- or anti-conformations. Thus, GC sequence remains to be the most popular Z-DNA forming sequence in literature.[13] In addition to it, the high salt concentration is also an important condition to stabilise the Z-DNA, where the electrostatic repulsion between the negatively charged phosphate groups reduces significantly to favour the Z-DNA formation.[14] However, under a standard cellular condition, this electrostatic repulsion does not allow the Z-DNA formation and rather shifts the equilibrium towards more stable B-DNA conformation. For
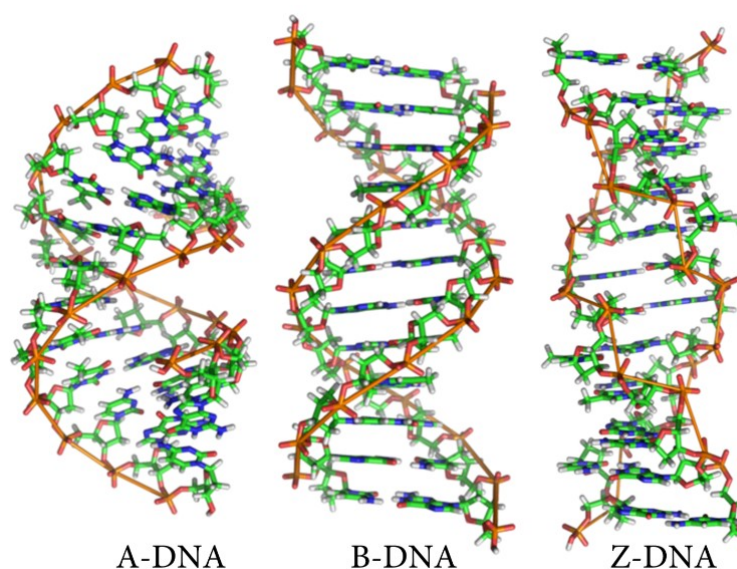
A-DNA   B-DNA   Z-DNA

Figure 1.4 Crystal structures of A-DNA, B-DNA and Z-DNA duplex DNA. Reproduced with permission from reference 11b.

a long time, this unstable nature of the Z-DNA structure under cellular condition brought difficulty to understand the biological significance of Z-DNA. However, many different factors such as, the presence of potential Z-DNA sequence (i.e. (GC)$_n$ and many more) in several promoter regions,[15] stabilization of Z-DNA structures by abundant cellular polyamines (spermine and spermidine),[16] DNA methylation,[17] generated negative supercoiling stress during transcription[18] *etc.* gave an indication about the biological relevance of this structure. Later in 1992, it was observed that this form of DNA secondary structure involves in DNA transcription processes which upregulates the C-Myc gene activity with the formation of Z-DNA.[19] Afterwards, many reports have highlighted the biological importance of Z-DNA in different aspects, such as, discovery of Z-DNA binding proteins (adenosine deaminase acting on RNA, i.e. ADAR protein, DLM1, E3L *etc.*),[17b, 20] exploration of Z-DNA as a repressor factor for metalloprotease proteins, involvement in many human diseases (cancer, viral infection, and autoimmune diseases) *etc.*[21] Apart from the biological activity, researchers have paid attention to explore the factors influencing the B- to Z-DNA transition and also showed that many materials such as, high salt concentration,[7] polypeptide,[22] cationic graft copolymer,[23] dehydrating agent (such as, ethanol, ethylene glycol and methanol),[24] deep eutectic solvent (DES),[25] negative superhelicity[18a, 18c] *etc.* help to promote the transitions from B- to Z-DNA. Till now, the focus of the Z-DNA research has been centred on understanding the role and function in cellular condition and its implication in several diseases such as cancer, viral infection.[21d] This has prompted researchers to consider Z-DNA as a potential therapeutic target which needs further investigation for the drug design/discovery based on this structure.

Table 1.1 Different helical parameters observed in A-, B- and Z-DNA structures.

Parameter	A-DNA	B-DNA	Z-DNA
Helix sense	Right	Right	Left
Base pairs per turn	11	10	12
Axial rise (nm)	0.26	0.34	0.45
Helix pitch (°)	28	34	45
Base pair tilt (°)	20	− 6	7
Twist angle (°)	33	36	−30
Diameter of helix (nm)	2.3	2.0	1.8

1.1.2 A-DNA

Similar to conventional duplex DNA, A-DNA structure also consists of right-handed double helix but relatively compacted form in comparison to the B-DNA, which is generally formed within certain stretches of purine nucleobases (e.g. GAGGGA).[11b] For the first time, A-DNA was identified by fibre diffraction studies at relatively lower humidity condition. Generally, A-DNA structure formation is

stabilised under dehydrating conditions.[26] As the conformational stability of A-DNA is less in comparison to the B-DNA, thus, part of B-DNA always contains A-DNA not a full sequence of A-DNA.[27] However, there are some structural difference between A-DNA and B-DNA which is given in the Table 1.1 and Figure 1.4.[26a] In the case of RNA or RNA-DNA hybrid structure, A-type double helix is the only form present which is favoured due to the steric interactions applied by additional -OH group.[26a, 28] As, the formation of A-DNA needs an extremely dehydrating condition;[29] thus this form of DNA is majorly observed in extreme conditions such as, during extreme desiccation of bacteria,[30] removal of water molecules from DNA due to the protein binding in sporulating bacteria.[31] In addition to these, A-DNA structures are also quite common in long terminal repeats (LTRs) of the transposable element due to the higher abundance of purine stretches in this region which is believed to be involved in DNA recombination.[32] Although it is not a stable secondary structure unlike B-DNA, the presence of short stretches of purine nucleobases in genome indicates the presence of this structure in transient.
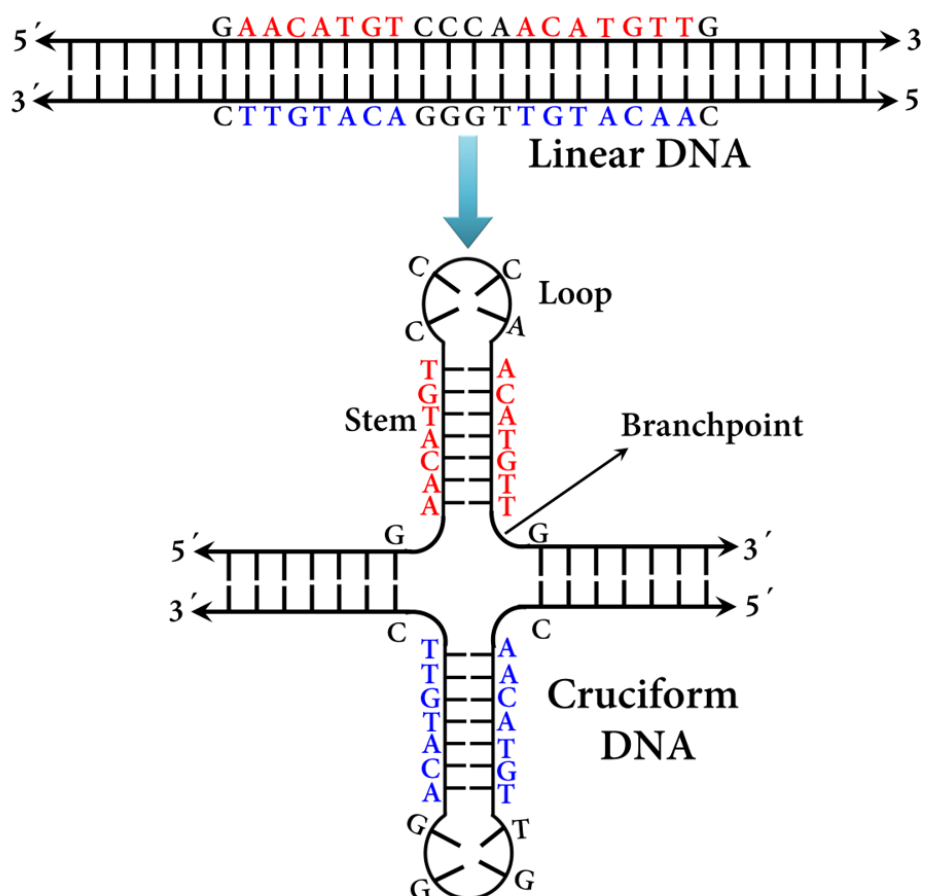
Figure 1.5 Schematic representation of the cruciform structure.

1.1.3 Cruciform

Cruciform is another class of non B-DNA structure which contains a stem, a branch point and loop regions (Figure 1.5).[33] In general, the transition from inter-strand base pairing to intra-strand base pairing in duplex DNA results in a formation of the cruciform structure. It has been observed that length

(≥ 6 bases) and sequence of an inverted repeat plays a vital role for the cruciform formation where the presence of A-T rich sequence increases the probability for the formation of this structure.[34] Till now, two distinct class of cruciform structures have been observed through atomic force microscopy, one with 4-fold symmetry having all arms perpendicular to each other, another with stacked conformations having the adjacent arms in an acute angle with the main DNA strands.[35] The conformational stability of these cruciform structures is less in comparison to that of duplex DNA, and this structure has been observed for the first time in circular plasmid DNA, where the inherent negative supercoiling stress stabilises the cruciform structure.[34, 36] Later, this structure has received greater biological relevance owing to its binding with many architectural and regulatory proteins such as, HMG proteins, topoisomerase IIb, HU, p53, histones H1 and H5 and many others which is not only helping to target different biological processes (such as, nucleosome positioning, replication and regulation of gene expression etc.) but also to target many diseases.[37]

1.1.4 Hairpin

Hairpin structure of nucleic acid is believed to be one of the most studied and well explored in terms of its conformational dynamics, biological role and function. Hairpin structure contains a stem with nucleobases pairing among each other and a small loop with unpaired bases which often serve as basic units for many other different secondary structures of nucleic acids such as, pseudoknot, intramolecular triplex DNA, cruciform *etc.*[38] Stability of the hairpin structure depends on the stem length, loop radius and type of sequence, which is observed for some unusual high stability of hairpin structures with loop sequence having GXN (X= A, T, G and C) or C-G closing base pairing in the stem.[7, 39] The formation of hairpin structure is explained via two mechanisms in literature, one is similar to the cruciform formation from the duplex DNA by negative superhelicity, and another mechanism is due to the folding of single-stranded DNA which is formed during the different DNA metabolism processes. Thus, these formed hairpin structures have been targeted with small molecules or proteins to modulate different biological processes such as, replication, transcription, recombination, translation (hairpin formed in the mRNA) *etc.* which in turn acts as a control to gene expression.[40] Apart from its biological role, hairpin structures have also received interest for various purpose in bio-nanotechnology, such as, in the preparation of molecular beacons, detection of specific nucleic acid sequence, imaging, nanostructures *etc.*[10a, 41]

1.1.5 Triplex

Triplex DNA belongs to a particular class of secondary structure DNA, where the structure consists of both Watson-Crick and Hoogsteen base pairing. In 1957, Felsenfeld *et al.* for the first time observed the existence of the triplex DNA.[16a] These structures are usually formed by the Hoogsteen or reverse Hoogsteen H-bonding between a duplex DNA strand (where Watson-Crick base pairing is present between oligopurine and oligopyrimidine base pairing) and a single-stranded triplex-forming

oligonucleotide (TFOs) in a sequence-specific manner (Figure 1.6).[7] The formation of intramolecular triplex structures has also been observed in the regulatory regions of several genes which are commonly known as H-DNA.[42] Formation of the triplex structures depends on several factors such as nature of the ion (especially Mg^{2+}) and its concentration in the medium, pH in the medium *etc.*[43] For example, it has been observed that at slightly acidic pH, the base pairing in all the three strands happens in one step,
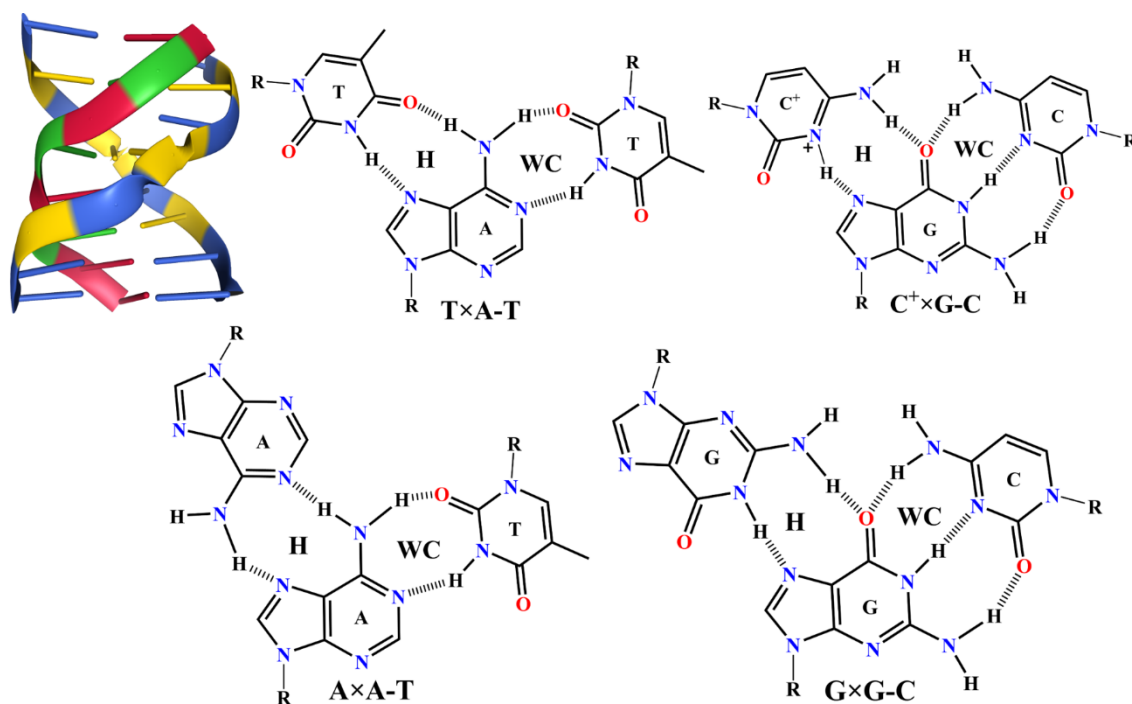
Figure 1.6 Crystal structure of triplex DNA (PDB ID- 1D3X[44]) and the different type of base pairing (Watson-Crick i.e., WC and Hoogsteen i.e., H) observed in triplex DNA.

whereas triplex formation is found to be sequential in physiological pH (pH 7.0- 7.5) and it is absent in at basic pH (pH 8).[7, 43e] Triplex DNA can adopt two different kinds of topology depending on the base pairing between the purine base of the duplex and purine or pyrimidine bases of the TFO. These are parallel and antiparallel triplex DNA, where the most popular base pairing combination of the parallel triplex are T×A-T and C+×G-C, and for antiparallel, G×G-C and A×A-T ("×" represents the Hoogsteen H-bonding and "–" indicates Watson-Crick base pairing) (Figure 1.6).[33, 45] Researchers have explored the structural formation and dissociation processes of triplex DNA through different techniques such as UV-melting, differential scanning calorimetry (DSC), single molecular fluorescence resonance energy transfer (smFRET), fluorescence anisotropy measurements *etc.*[7] Apart from ions and pH of the solution, the triplex formation can also be induced by several other materials such as polyamine (spermine, pentaamine, hexamine *etc.*),[46] comb-type cationic copolymers,[47] by introducing locked nucleic acids,[48] molecular crowding agents[49] *etc.* In spite of the triplex DNA discovery at such earlier stage in 1957, the biological significance of this motif remains unclear until a long time. One of the previous reports by Morgan *et al.* in 1968 mentioned the inhibiting property of triplex DNA on the transcription processes.[50] Later, the structural motif of triplex DNA has been utilised to recognise or target any specific DNA

sequence which plays an important role starting from to control biologically important processes such as, gene targeting, mutagenesis etc. to different applications in molecular biology, diagnostics, and therapeutics.[33, 51]

1.1.6 A-motif

A-motif DNA structure is generally formed in the repetitive single strand sequences of adenine (A) nucleobase (such as polyriboadenylic acid, i.e., Poly(A) and polydeoxyadenylic acid, i.e., Poly(dA)) at acidic condition (pH 3-3.4).[52] In pH 3-3.4, these adenine-rich sequences show a conformational transition from a single-stranded right-handed helical structure to a right-handed duplex helical
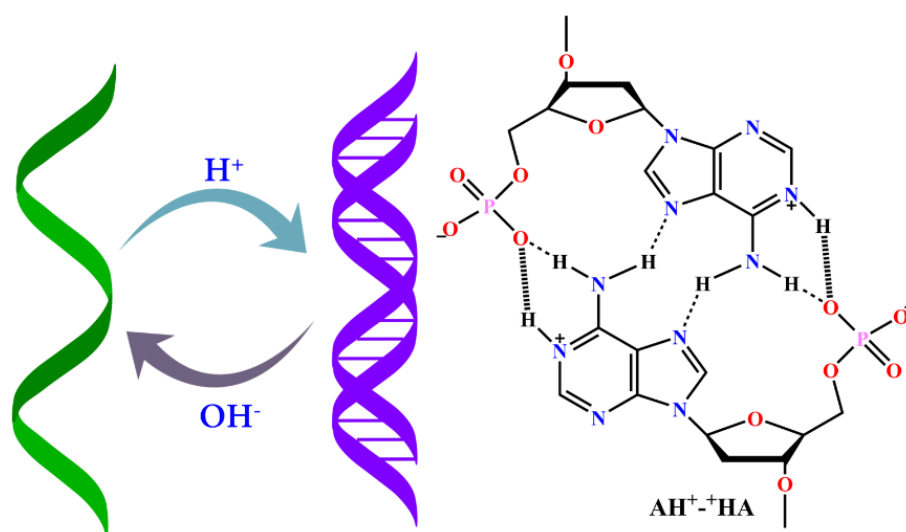
Figure 1.7 Schematic representation showing the conformational transition of Poly(A) structure depending on the pH of the medium and the base pairing (AH$^+$-$^+$HA) observed in the duplex structure of Poly(A) in acidic pH.

structure, where the reverse Hoogsteen H-bonding is observed between AH$^+$-$^+$HA (Figure 1.7). It has been observed that depending on the pH of the medium, this structure changes from one conformation to another such as, intermediate form (partial protonation of adenine nucleobases), tightly packed form (full protonation of adenine) and frozen form (in most acidic condition).[7, 53] Apart from the pH of the medium, small molecules like, berberine, sanguinarine, palmatine, and coralyne can stabilise the duplex structure even in neutral pH.[52b, 54] In addition to its unique conformational transformation, this structural motif has some biological importance due to the presence of Poly(A) in the mRNA tail, where it plays important roles in terms of mRNA stability and translation initiation. Shortening of this Poly(A) length gradually happens through exonuclease and attainment of a threshold length prompts rapid destruction of RNA structure.[7]

1.1.7 G-quadruplex

Among different secondary structures of DNA, probably G-quadruplex has received the most importance among researchers due to its active relevance in anticancer therapy. For the first time, a biochemist from Norway named, Ivar Christian Bang in 1910 reported the formation of gel material at

a millimolar concentration of guanylic acid.[55] After Fifty years of this discovery, Gellert *et al.* monitored through X-ray diffraction that guanylic acid self assembles into tetrameric structure maintaining a square planar arrangement, where each guanine is connected through Hoogsteen H-bonding. This tetrameric guanine structure is known as G-tetrad (Figure 1.8). Real interest about G-quadruplex DNA started after the breakthrough discovery of Blackburn *et al.* in early 1990, where the guanine-rich repetitive sequences had been identified at the end of chromosomes, i.e. telomere, which is a nucleoprotein complex and protect the chromosome degradation.[56] Being the noncoding part of the chromosome, the length of the telomere shortens with each cell division. Thus, the structure and stability of telomere act as an indicator for the cell ageing.[57] This is known as the end replication problem. To address the end replication problem, nature manages to replicate this noncoding part using the telomerase enzyme.[58] In this context, Kim *et al.* made a significant discovery about the activity of telomerase enzyme which is found to be highly expressed in cancerous cells (80-85%) in comparison to normal somatic cells.[59] This discovery opens up a new research field for G-quadruplex structure,
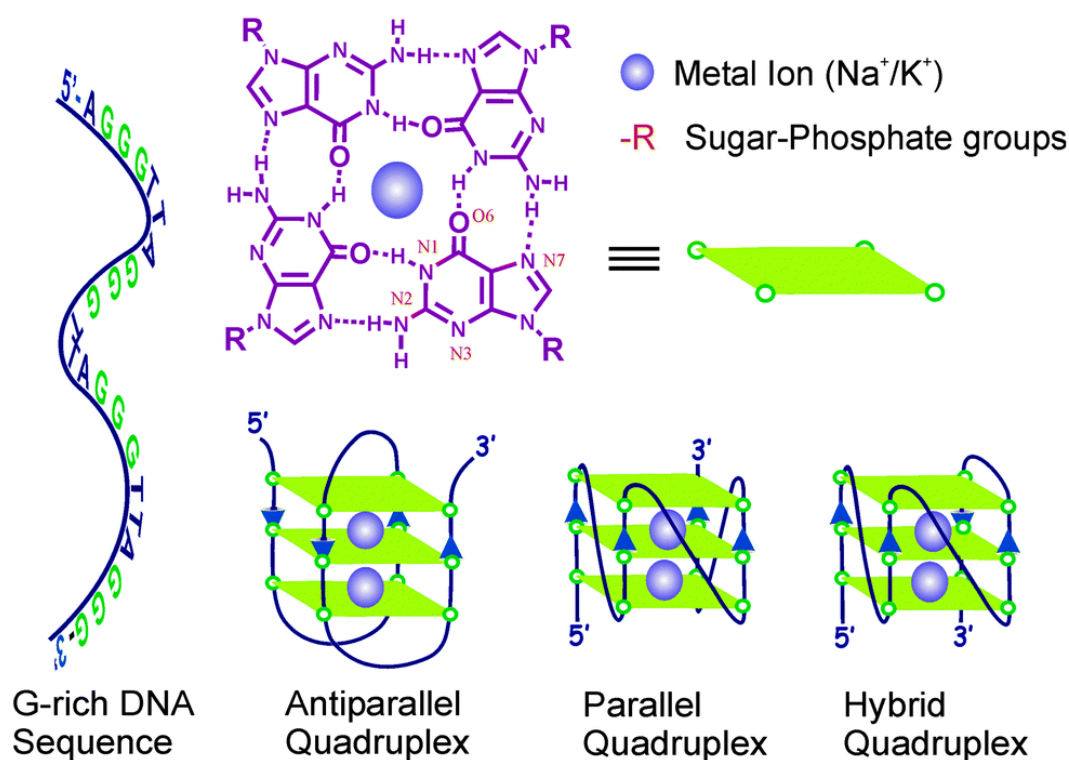
Figure 1.8 Schematic representation of G-tetrad and G-quadruplex structure with their different topology. Reproduced with permission from reference 61h.

where researcher have tried to alter the substrate of the telomerase enzyme, i.e., telomere sequence by means of stabilising in the G-quadruplex structure. Hence, stabilisation of G-quadruplex structure by small molecules and external stimuli serves as a tool for anticancer activity which is very much specific to the cancerous cells, unlike previous duplex DNA intercalating agents.[60] In continuation to this, a plethora of small molecules such as TMPyP4, BRACO, N-methyl mesoporphyrin (NMM), Telomestatin, cryptolepine, proflavine, ellipticine and many more have been reported for G-quadruplex

stabilisation.[60b, 61] Apart from its anticancer activity, G-quadruplex forming sequences have also been observed in the genome of various organisms and especially in the promoter regions of multiple genes (such as C-Myc, bcl-2, KRAS, hTERT etc.), 5´- untranslated region (UTR) and oncogenes. Thus, G-quadruplex structure has significant influences on many biologically important processes such as DNA replication, transcription regulation, translational control by G-quadruplex formed in the open reading frame of mRNA, splicing, and meiosis.[7, 62] In addition to the biological activity, G-quadruplex structures have been extensively applied in DNA nanotechnology such as electronic nanoswitch, nanomotor, DNA logic gate, building blocks for DNA architecture, preparation of nanocluster and quantum dots for optoelectronic properties *etc.*[63]

G-quadruplex structure usually forms in the guanine-rich sequence in the presence of ions such as, Na^+, K^+, Li^+, NH_4^+ *etc.* which stabilises the G-tetrad structure by means of forming coordination with electronegative carbonyl groups of guanines.[63c] G-quadruplex structures are highly polymorphic in nature (Figure 1.8) which is well characterised by different biophysical tools such as, nuclear magnetic resonance (NMR), circular dichroism (CD), X-ray crystallography *etc.*[7, 33, 64] G-quadruplex can show its polymorphism based on many factors such as the shapes (chair or basket), orientation (parallel, antiparallel and mixed), loops (lateral, propeller, diagonal, V-shaped) and strand stoichiometry (unimolecular, bimolecular and tetramolecular).[63c, 64-65] Adoption of any particular G-quadruplex structure also depends on various parameters such as cations, nucleotide sequence, the extent of dehydration, molecular crowding *etc.*[66] Here, it is important to note that parallel topology associates with all four strands in the same direction and for antiparallel topology two strands in one direction and another two strands in the opposite direction. However, in the case of mixed or hybrid topology, three strands are in the same direction, and the last one aligned in the opposite direction (Figure 1.8). G-quadruplex structures with various topology differ in overall stability which is resulting from their different conformational state (either syn or anti) of the guanine bases in the G-tetrad. These guanine bases can adopt either anti or syn geometry depending on the glycosidic bond between the guanine bases and the sugar moiety.[64] For instance, all the guanine geometry in parallel topology is in anti-conformation. Similarly, a mixture of syn- and anti-conformations (i.e., anti/syn and syn/anti) is observed for antiparallel topology, whereas all type of conformations is present in hybrid or mix topology. Again, different G-quadruplex topology exhibits its structural preference depending on the ion, relative orientation of the strands *etc.*[7, 64] For example, in the presence of Na^+ ions, human telomere sequence (5´-(TTAGGG)$_4$-3´) forms an antiparallel structure which shows a mix or hybrid (3+1) structure in the presence of K^+ ion.[67] Thus, specific cations play a major role in the G-quadruplex stabilization and it varies in the following order $Sr^{2+} > K^+ > Ca^{2+} > Na^+$, NH_4^+, $Rb^+ > Mg^{2+} >> Li^+$, Cs^+, where K^+ resides in between two G-tetrads due to its relatively larger size. Na^+ fits perfectly in the G-tetrad, whereas, Li^+ and Cs^+ are too small and large respectively, to fit into the G-tetrad. Extra stabilisation of G-quadruplex structure in the presence of K^+ ion compare to Na^+ is attributed to the

greater energetic cost experienced due to dehydration in the presence of Na^+ ion.[68] Apart from the role of ion, researchers have also explored the effect of molecular crowding on the formation and stability of G-quadruplex structure to mimic the behaviour in intracellular condition. Molecular crowding agents (such as ethylene glycol, poly(ethylene glycol), glycerol *etc.*) stabilise the G-quadruplex structure even in the absence of ion, unlike conventional duplex DNA.[66b] These kind of ultrastable G-quadruplex structures have also been observed in unusual conditions such as pressure, negative superhelicity, in the presence of a deep eutectic solvent and hydrated ionic liquid *etc.*[69]

In addition to the stabilisation aspect, the folding and unfolding kinetics of G-quadruplex DNA has been explored in details through different spectroscopic techniques such as NMR, single-molecule fluorescence resonance energy transfer (smFRET), CD, UV-melting, molecular dynamics (MD) simulations *etc.*[7, 63c, 64] In this regard using NMR spectroscopy, Phan and coworkers reported the faster folding kinetics (and slower unfolding kinetics) of antiparallel structure over parallel topology for a modified two-repeat human telomeric sequence, i.e., d(UAGGGT^{Br}UAGGGT) in K^+ solution which is known to form both parallel and antiparallel structure simultaneously.[70] Similarly, through smFRET technique, Balasubramanian and coworkers monitored the coexistence of two folded conformations of human telomeric G-quadruplex DNA in near physiological conditions which undergoes interconversion between each other on a minute time scale.[71] Folding pathways of the human telomeric sequences in presence of K^+ ion has been explored by Sugiyama and coworkers using both the experimental and theoretical studies, where the presence of hairpin and triplex intermediate structures have been separately observed during the folding of two mixed or hybrid (3+1) topology from single-stranded G-rich (d(TTAGGG)$_4$) DNA sequence. With the addition of K^+ ion, the rate of global folding increases due to decreasing electrostatic repulsion which consequently reduces the energy barrier associated with each step.[65, 72] In 1999, Sponer and coworkers studied the change in conformation of G-quadruplex structure through MD simulations, where the displacement of the cations take place by the solvent molecules without disturbing the stem part of the G-quadruplex.[73] However, complete removal of the cation disrupts the G-quadruplex structure fully within a few picosecond timescales implying the importance of cations on the structural stability. Overall reports on the folding and unfolding kinetics of G-quadruplex DNA suggest a remarkably slower rate which is majorly dependent on the presence of cations and number of tetrads in G-quadruplex.[73]

Till now, among different secondary structures of DNA, G-quadruplex has emerged as the most useful structural motif in terms of their biological implication and their application in nanotechnology. However, a lot of research still needs to happen before its applicability in therapeutics. For example, a ligand with very high affinity (K_d < 1 nM) and high specificity (10000 fold or lesser affinity to other DNA structure) is still needed in order to validate the anticancer activity of G-quadruplex DNA.[60b] Thus, G-quadruplex is still a promising field of research which needs to be carried out by chemists, biophysicists, chemical and molecular biologist.

1.1.8 i-motif DNA

Like guanine-rich sequences, complimentary cytosine-rich sequences can also fold into a tetrameric noncanonical structure which is generally known as i-motif DNA (or C-quadruplex or i-tetraplex or i-DNA *etc.*).[7, 74] In 1962, Marsh *et al.* recognised the formation of hemiprotonated cytosine base pair in cytosine-5-acetic acid through X-ray diffraction.[75] Later in 1992, Henderson and coworkers reported the stabilisation of hairpin structure by an unusual Hoogsteen H-bonding between cytosine and hemiprotonated cytosine (C-C$^+$) base pairs.[76] However, for the first time in 1993, Gehring *et al.* unveiled an intercalated tetramer structure of d(TCCCCC) at slightly acidic pH through NMR studies, where two parallel duplexes are intercalated in an antiparallel manner by means of forming Hoogsteen H-bonding between hemiprotonated cytosine–cytosine (C$^+$-C) base pairs (Figure 1.9).[77] Surprisingly, this C$^+$-C hydrogen bonding associates with much higher base pairing energy (169.7 kJ mol^{-1}) than conventional Watson-Crick (96.6 kJ mol^{-1} for G-C).[78] These C$^+$-C base pairs closely stacked upon each other which is reflected in their lower base pairing distance (3.1 Å) in comparison to that of B-DNA (3.4 Å). The observed helical twist between two adjacent base pair is relatively smaller (12-16º) compared to that in B-DNA (36º) which results in two flat major grooves and two extremely narrow minor grooves.[78] Presence of slightly acidic pH is extremely necessary for the formation and stabilisation of i-motif DNA, as hemiprotonation of the cytosine bases (around pK$_a$ 4.2 for N3 in cytosine) in the nucleotide sequence
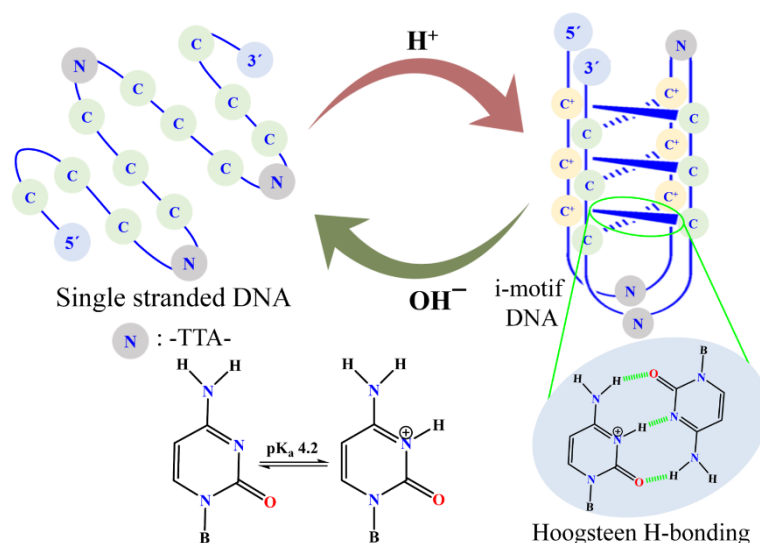
Figure 1.9 Schematic representation of i-motif DNA structure and observed Hoogsteen base pairing between hemiprotonated cytosine-cytosine (C$^+$-C) bases.

is one of the primary steps for the formation of i-motif DNA.[74, 78] Thus, i-motif structures exhibit its unstable nature in the neutral and alkaline medium due to the deprotonation of these cytosine bases. Apart from the pH of the medium, formation and stability of the i-motif structure is found to be dependent on several other parameters like number of cytosine bases, number of nucleobases in the loop, different environmental conditions affecting the pK$_a$ of N3 in cytosine *etc.*[79] Following these

conditions, researchers have tried to stabilize the i-motif structure in neutral and alkaline medium using different materials and conditions such as, carboxyl modified carbon nanotubes, graphene quantum dots, modified nucleobases, crowding agents, metal ions, negative superhelicity, pressure *etc.*[69c, 79a, 80] Apart from the stabilizing material, some C-rich sequences exhibits its stability in neutral pH and even in some alkaline pH.[79c, 79d, 80b] In this regard, Burrows and coworkers have reported that a "4n-1" structural motif in poly-2′-deoxycytidine (dC_n; where n = 4, 5, 7) shows excellent thermal stability around 40 ºC in pH 7 and also exhibits a higher transitional pH value (> 7.2). This preference for "4n-1" structural motif has been explained in terms of the number of base pairs and loop lengths in i-motif DNA, where the most stable i-motif structure comprises of three loops with a single nucleotide and an even number of base pairs.[79c, 79d] Similarly, Sugimoto and coworkers explored the effect of loop sequence and lop length on the stability of intramolecular i-motif structures at neutral pH, where they have shown that the stability of i-motif structures could be significantly improved by the presence of guanine or thymine nucleobases in the first and third loops due to their inter-loop H-bonding.[79f] Depending on the loop size, i-motif structures have been divided into two different class; i-motif with shorter loop (5′-2:3/4:2-3′, where, x:y:z – no. of nucleobases in the loop) as class I and i-motif with longer loop (5′-6/8:2/5:6/7-3′) as class II. Generally, it has been observed that i-motif DNA structures with longer loops (i.e., class II) are more stable in neutral pH than i-motif with shorter loops (i.e., class I) due to additional stabilising interactions in the longer loops.[62c] Apart from its stabilisation aspect, folding and unfolding kinetics has also received special interest among researchers to get an insight into the biological role of i-motif DNA. The rate of the folding (association) and unfolding (dissociation) kinetics largely depends on the nucleotide sequence, pH of the medium *etc.*[7] For instance, Tan and coworkers reported the fast unfolding (half-life, $t_{u1/2}$ – 36500 sec) dynamics of human telomeric C-rich sequence (i.e., 5′-(CCCTAA)$_4$-3′) in pH 7 and with decrease in pH from 7.0 to 5, the folding equilibrium constant (K_F) enhance by 5800-fold which decreased by 20-fold in presence of a complementary DNA strand (i.e., 5′-(TTAGGG)$_4$-3′).[81] Using CD and FRET spectroscopy techniques, Zhou *et al.* explored the formation time constants of some C-rich sequences (214 to 493 sec) at pH 7 solution which is known to form stable i-motif structure at neutral and alkaline medium.[80b] Balasubramanian and coworkers investigated the folding and unfolding kinetics of i-motif DNA in a proton-fuelled DNA nanomachine which is found to be completed within 5 sec indicating the role of the proton in the i-motif formation.[82] Depending on the folding and unfolding dynamics, i-motif DNA structures have been extensively explored in different applications such as pH sensors, logic gates, switchable nanostructures, memory devices, ATP sensors, delivery systems *etc.*[10b, 10c, 83]

Being a complementary sequence of G-quadruplex DNA, i-motif forming sequences have also been observed in the gene promoter regions. For a long time, the role of i-motif DNA in important biological processes was not clear. In 2012, Chen *et al.* demonstrated the i-motif stabilisation aspect of carboxyl-modified single-walled carbon nanotube (C-SWCNT) on telomere function which shows the

inhibiting activity of the telomerase enzyme leading apoptosis in cancer cells *in vitro* and *in vivo*.[80j] In 2014, Hurley and coworkers demonstrated the role of i-motif DNA on the transcription regulation, where they have specifically stabilised the bcl-2 i-motif DNA through a ligand namely, IMC 48 to upregulate the rate of transcription.[84] Subsequently, in the next paper, an endogenous transcription factor, hnRNP LL and the i-motif complex are detected which unfolds the i-motif structure to activate the gene transcription.[85] These studies show that modulation of i-motif formation and stabilisation can act as a switch to control the gene expression.[85] Later, Dzatko *et al.* detected the presence of i-motif DNA in mammalian cells through NMR study indicating its stability in cellular condition.[86] To know more about the role of i-motif DNA, Christ's and Dinger's in 2018 came up with a unique antibody namely, iMab which can exclusively bind to specific C-rich sequences inducing the i-motif formation and not to any other DNA structures/sequences.[87] Utilising this specific binding, they have detected the existence of i-motif in human nuclei as well as in different stages cell cycles (early S phase and G0/G1 and G1/S boundaries) exploring the role of i-motif DNA on transcription and replication. So far, this is the most substantial evidence regarding the role of i-motif DNA on gene expression control.[87] Utilising a search algorithm namely Quadparser,[88] Waller's group have identified around 5000 potential i-motif forming sequences across the whole genome and among these almost 12 % sequences are found to reside in the promoter regions.[74] Till now, research on i-motif DNA is focused on a few particular sequences, and thus, a lot of research still needs to carry out in order to fully understand the role and functions of i-motif DNA.

1.2 Motivation of the Thesis

In the previous sections, we have extensively discussed the structural characteristics of different nucleic acid structures, where we have observed that biological functions of these different secondary structures of DNA are heavily depended on their structure and stability. Here, it is important to note that these nucleic acids are prone to undergo conformational transitions depending on the favourable conditions which significantly impacts their structure and stability. Importantly, structure and stability of these different DNA secondary structures are very crucial in terms of their participation in biologically important processes such as genome recombination, cancer cell immortality and regulation of gene expression *etc.*[7, 62a] Thus, understanding the molecular mechanism behind the formation and activity of these different secondary structures always has been the point of interest among researchers which can be utilised for further biomedical applications. Similarly, detection of these DNA conformations is also essential which will enable us to explore their activity by tracking it in intracellular conditions. Apart from the biological functions of these secondary structures, the conformational transition between different forms of DNA is the primary principle behind most of the material and nanotechnological applications based on different DNA structures.[10a] In the continuation of this, all the works are majorly focused on controlling and detecting the conformational transitions of different DNA secondary structures by using some external stimuli. In the second chapter (**chapter-2**),[89] the controversial

interaction between conventional duplex DNA and ionic liquid has been explored, where instead of an intercalation or groove binding phenomenon, an ionic liquid induced DNA compaction has been observed due to its coil to globule transitions. This compaction process occurs only in the higher concentration of ionic liquid ($\geq$1 mM), where, the electrostatic dragging of DNA strands (by the guanidinium cation) takes place at the self-aggregated surface of the anionic part of the ionic liquid. In the first part of the third chapter (**chapter-3A**),[90] the formation and stabilization of a non-canonical DNA structure i.e. G-quadruplex DNA by an ionic liquid has been demonstrated in absence of any alkali ions, where we have observed that smaller cation with a bulky anion pair may have some edge over other ion pairs for G-quadruplex formation and stabilization. In the second part of the third chapter (**chapter-3B**),[91] controlling the conformational transition from single-stranded G-rich sequence to G-quadruplex DNA formation and stabilisation has been executed by incorporating the reversibility aspect into it. Importantly, this conformational transition of G-quadruplex DNA can be easily monitored by the fluorescence colour switch of the stabiliser used in this study. In the first part of the fourth chapter (**chapter-4A**),[92] a new reversible strategy has been utilised to stabilise i-motif DNA in neutral and alkaline media, unlike its natural characteristics by incorporating C-rich sequences inside silica nano-channels. Importantly, this approach provides an alternative way to control i-motif formation other than pH and thermal annealing which can be implemented in multiple cycles for the entire conformational transition from single-stranded C-rich sequence to i-motif DNA. In the last part of the fourth chapter (**chapter-4B**), a ligand with light-up property towards various i-motif DNA (including intermolecular and intramolecular) has been developed based on the recognition of hemiprotonated cytosine–cytosine base pairing. A systematic study between i-motif DNA with different structural analogs of the ligand indicates that a balance of positive and negative functionality in the ligand is the crucial aspect for the selective detection of i-motif DNA.

1.3 References

1. R. Dahm, *Dev. Biol.*, 2005, **278**, 274-288.

2. J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737-738.

3. J. Šponer, J. Leszczynski and P. Hobza, *J. Phys. Chem.*, 1996, **100**, 1965-1974.

4. (a) K. Hoogsteen, *Acta Crystallogr.*, 1959, **12**, 822-823; (b) K. Hoogsteen, *Acta Crystallogr.*, 1963, **16**, 907-916.

5. (a) E. N. Nikolova, E. Kim, A. A. Wise, P. J. O'Brien, I. Andricioaei and H. M. Al-Hashimi, *Nature*, 2011, **470**, 498; (b) E. N. Nikolova, F. L. Gottardo and H. M. Al-Hashimi, *J. Am. Chem. Soc.*, 2012, **134**, 3667-3670.

6. C. Yi, G. Zheng, L. Lu, X. Jian and C. He, *Nucleic Acids Res.*, 2010, **38**, 4415-4425.

7. J. Choi and T. Majima, *Chem. Soc. Rev.*, 2011, **40**, 5893-5909.

8. E. S. Lander *et al.*, *Nature*, 2001, **409**, 860.

9. (a) G. Wang and K. M. Vasquez, *DNA Repair*, 2014, **19**, 143-151; (b) R. R. Sinden, *DNA Structure and Function*, Elsevier Science, 2012; (c) R. R. Sinden and R. D. Wells, *Curr. Opin. Biotechnol.*, 1992, **3**, 612-622; (d) G. Wang and K. M. Vasquez, *Mutat. Res., Fundam. Mol. Mech. Mutagen.*, 2006, **598**, 103-119.

10. (a) F. Wang, X. Liu and I. Willner, *Angew. Chem., Int. Ed.*, 2015, **54**, 1098-1129; (b) Y. Dong, Z. Yang and D. Liu, *Acc. Chem. Res.*, 2014, **47**, 1853-1860; (c) L. A. Yatsunyk, O. Mendoza and J.-L. Mergny, *Acc. Chem. Res.*, 2014, **47**, 1836-1844; (d) Z.-G. Wang, J. Elbaz and I. Willner, *Nano Lett.*, 2011, **11**, 304-309; (e) T. Masubuchi, M. Endo, R. Iizuka, A. Iguchi, D. H. Yoon, T. Sekiguchi, H. Qi, R. Iinuma, Y. Miyazono, S. Shoji, T. Funatsu, H. Sugiyama, Y. Harada, T. Ueda and H. Tadakuma, *Nat. Nanotechnol.*, 2018, **13**, 933-940.

11. (a) A. H. J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel and A. Rich, *Nature*, 1979, **282**, 680-686; (b) R. E. Dickerson, H. R. Drew, B. N. Conner, R. M. Wing, A. V. Fratini and M. L. Kopka, *Science*, 1982, **216**, 475.

12. M. Leng, *Biochim. Biophys. Acta, Gene Struct. Expression*, 1985, **825**, 339-344.

13. B. H. Johnston, in *Methods in Enzymology*, Academic Press, 1992, vol. 211, pp. 127-158.

14. T. J. Thamann, R. C. Lord, A. H. Wang and A. Rich, *Nucleic Acids Res.*, 1981, **9**, 5443-5457.

15. (a) H. Hamada and T. Kakunaga, *Nature*, 1982, **298**, 396-398; (b) D. B. Haniford and D. E. Pulleyblank, *J. Biomol. Struct. Dyn.*, 1983, **1**, 593-609.

16. (a) G. Felsenfeld, D. R. Davies and A. Rich, *J. Am. Chem. Soc.*, 1957, **79**, 2023-2024; (b) T. J. Thomas, U. B. Gunnia and T. Thomas, *J. Biol. Chem.*, 1991, **266**, 6137-6141; (c) A. Parkinson, M. Hawken, M. Hall, K. J. Sanders and A. Rodger, *Phys. Chem. Chem. Phys.*, 2000, **2**, 5469-5478.

17. (a) M. Behe and G. Felsenfeld, *Proc. Natl. Acad. Sci. U. S. A.*, 1981, **78**, 1619; (b) V. Vongsutilers and P. M. Gannett, *Org. Biomol. Chem.*, 2018, **16**, 2198-2209; (c) B. C. Train, S. A. Bilgesü, E. C. Despeaux, V. Vongsutilers and P. M. Gannett, *Chem. Res. Toxicol.*, 2014, **27**, 1176-1186.

18. (a) A. Nordheim and A. Rich, *Proc. Natl. Acad. Sci. U. S. A.*, 1983, **80**, 1821; (b) L. J. Peck, A. Nordheim, A. Rich and J. C. Wang, *Proc. Natl. Acad. Sci. U. S. A.*, 1982, **79**, 4560-4564; (c) A. Rich and S. Zhang, *Nat. Rev. Genet.*, 2003, **4**, 566.

19. B. Wittig, S. Wölfl, T. Dorbic, W. Vahrson and A. Rich, *EMBO J.*, 1992, **11**, 4653-4663.

20. (a) A. G. Herbert and A. Rich, *Nucleic Acids Res.*, 1993, **21**, 2669-2672; (b) A. Herbert, J. Alfken, Y.-G. Kim, I. S. Mian, K. Nishikura and A. Rich, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 8421-8426; (c) Y.-G. Kim, M. Muralinath, T. Brandt, M. Pearcy, K. Hauns, K. Lowenhaupt, B. L. Jacobs and A. Rich, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 6974.

21. (a) B. K. Ray, S. Dhar, A. Shakya and A. Ray, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 103; (b) B. J. Gilpin, F. Loechel, M.-G. Mattei, E. Engvall, R. Albrechtsen and U. M. Wewer, *J. Biol. Chem.*, 1998, **273**, 157-166; (c) R. Roy, U. M. Wewer, D. Zurakowski, S. E. Pories and M. A. Moses, *J. Biol. Chem.*, 2004, **279**, 51323-51330; (d) M. J. Duffy, E. McKiernan, N. Donovan and P. M. McGowan, *Clin. Cancer Res.*, 2009, **15**, 1140.

22. L. Klevan and V. N. Schumaker, *Nucleic Acids Res.*, 1982, **10**, 6809-6817.

23. N. Shimada, A. Kano and A. Maruyama, *Adv. Funct. Mater.*, 2009, **19**, 3590-3595.

24. W. Zacharias, J. E. Larson, J. Klysik, S. M. Stirdivant and R. D. Wells, *J. Biol. Chem.*, 1982, **257**, 2775-2782.

25. I. Mamajanov, A. E. Engelhart, H. D. Bean and N. V. Hud, *Angew. Chem., Int. Ed.*, 2010, **49**, 6310-6314.

26. (a) D. W. Ussery, 2002; (b) R. Chandrasekaran, M. Wang, R. G. He, L. C. Puigjaner, M. A. Byler, R. P. Millane and S. Arnott, *J. Biomol. Struct. Dyn.*, 1989, **6**, 1189-1202.

27. H.-L. Ng, M. L. Kopka and R. E. Dickerson, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 2035-2039.

28. (a) S. Broyde and B. Hingerty, *Nucleic Acids Res.*, 1978, **5**, 2729-2741; (b) Y. Kyogoku, Y. Tanaka, K.-i. Tomita, T. Sakata, H. Hiroaki, T. Tanaka, S. Uesugi and S. Fujii, *Nucleic Acids Res.*, 1999, **27**, 949-955.

29. (a) V. Makarov, B. M. Pettitt and M. Feig, *Acc. Chem. Res.*, 2002, **35**, 376-384; (b) H. M. Berman, *Curr. Opin. Struct. Biol.*, 1994, **4**, 345-350; (c) F. M. Pohl, *Nature*, 1976, **260**, 365-366; (d) W. Saenger, *Annu. Rev. Biophys. Biophys. Chem.*, 1987, **16**, 93-114; (e) E. Westhof, *Annu. Rev. Biophys. Biophys. Chem.*, 1988, **17**, 125-144.

30. R. Whelan Donna, J. Hiscox Thomas, I. Rood Julian, R. Bambery Keith, D. McNaughton and R. Wood Bayden, *J. R. Soc., Interface*, 2014, **11**, 20140454.

31. F. DiMaio, X. Yu, E. Rensen, M. Krupovic, D. Prangishvili and E. H. Egelman, *Science*, 2015, **348**, 914.

32. A. Travers and G. Muskhelishvili, *FEBS J.*, 2015, **282**, 2279-2295.

33. M. Kaushik, S. Kaushik, K. Roy, A. Singh, S. Mahendru, M. Kumar, S. Chaudhary, S. Ahmed and S. Kukreti, *Biochem. Biophys. Rep.*, 2016, **5**, 388-395.

34. V. Brázda, R. C. Laister, E. B. Jagelská and C. Arrowsmith, *BMC Mol. Biol.*, 2011, **12**, 33.

35. (a) Y. L. Lyubchenko, *Cell Biochem. Biophys.*, 2004, **41**, 75-98; (b) L. S. Shlyakhtenko, V. N. Potaman, R. R. Sinden and Y. L. Lyubchenko, *J. Mol. Biol.*, 1998, **280**, 61-72.

36. A. L. Mikheikin, A. Y. Lushnikov and Y. L. Lyubchenko, *Biochemistry*, 2006, **45**, 12998-13006.

37. (a) S. Lee, L. Cavallo and J. Griffith, *J. Biol. Chem.*, 1997, **272**, 7532-7539; (b) J. R. P-ohler, D. G. Norman, J. Bramham, M. E. Bianchi and D. M. Lilley, *EMBO J.*, 1998, **17**, 817-826; (c) K. L. West and C. A. Austin, *Nucleic Acids Res.*, 1999, **27**, 984-992; (d) P. Varga-Weisz, J. Zlatanova, S. H. Leuba, G. P. Schroth and K. van Holde, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 3525; (e) P. Varga-Weisz, K. van Holde and J. Zlatanova, *J. Biol. Chem.*, 1993, **268**, 20699-20700.

38. (a) A. I. H. Murchie and D. M. J. Lilley, in *Methods in Enzymology*, Academic Press, 1992, vol. 211, pp. 158-180; (b) S.-J. Chen and S. Cao, *Nucleic Acids Res.*, 2006, **34**, 2634-2652.

39. (a) S. Yoshizawa, G. Kawai, K. Watanabe, K.-i. Miura and I. Hirao, *Biochemistry*, 1997, **36**, 4761-4767; (b) S. Kannan and M. Zacharias, *Nucleic Acids Res.*, 2011, **39**, 8271-8280.

40. (a) D. Bikard, C. Loot, Z. Baharoglu and D. Mazel, *Microbiol. Mol. Biol. Rev.*, 2010, **74**, 570; (b) T. Antony, T. Thomas, L. H. Sigal, A. Shirahata and T. J. Thomas, *Biochemistry*, 2001, **40**, 9387-9395.

41. (a) S. Tyagi and F. R. Kramer, *Nat. Biotechnol.*, 1996, **14**, 303-308; (b) J. Liu, Z. Cao and Y. Lu, *Chem. Rev.*, 2009, **109**, 1948-1998; (c) G. Bonnet, S. Tyagi, A. Libchaber and F. R. Kramer, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 6171.

42. (a) M. D. Frank-Kamenetskii, in *Methods in Enzymology*, Academic Press, 1992, vol. 211, pp. 180-191; (b) K. M. Vasquez and P. M. Glazer, *Q. Rev. Biophys.*, 2002, **35**, 89-107; (c) A. Mukherjee and K. M. Vasquez, *Biochimie*, 2011, **93**, 1197-1208; (d) P. P. Chan and P. M. Glazer, *J. Mol. Med.*, 1997, **75**, 267-282.

43. (a) S.-i. Nakano, M. Fujimoto, H. Hara and N. Sugimoto, *Nucleic Acids Res.*, 1999, **27**, 2957-2965; (b) D. G. Rhodes and S. D. Patil, *Nucleic Acids Res.*, 2000, **28**, 2439-2445; (c) S. Reither and A. Jeltsch, *BMC Biochem.*, 2002, **3**, 27; (d) J. E. Lee, T. Kim, S. Y. Kim and S. W. Kim, *Chem. Phys. Lett.*, 2010, **490**, 230-233; (e) N. Sugimoto, P. Wu, H. Hara and Y. Kawamoto, *Biochemistry*, 2001, **40**, 9396-9405.

44. M. Tarköy, A. K. Phipps, P. Schultze and J. Feigon, *Biochemistry*, 1998, **37**, 5810-5819.

45. M. D. Frank-Kamenetskii and S. M. Mirkin, *Annu. Rev. Biochem.*, 1995, **64**, 65-95.

46. (a) T. Thomas and T. J. Thomas, *Biochemistry*, 1993, **32**, 14068-14074; (b) T. J. Thomas, G. D. Kulkarni, N. J. Greenfield, A. Shirahata and T. Thomas, *Biochem. J.*, 1996, **319 (Pt 2)**, 591-599.

47. A. Maruyama, M. Katoh, T. Ishihara and T. Akaike, *Bioconjugate Chem.*, 1997, **8**, 3-6.

48. N. Kumar, K. E. Nielsen, S. Maiti and M. Petersen, *J. Am. Chem. Soc.*, 2006, **128**, 14-15.

49. (a) D. Miyoshi and N. Sugimoto, *Biochimie*, 2008, **90**, 1040-1051; (b) D. Miyoshi, K. Nakamura, H. Tateishi-Karimata, T. Ohmichi and N. Sugimoto, *J. Am. Chem. Soc.*, 2009, **131**, 3522-3531.

50. A. R. Morgan and R. D. Wells, *J. Mol. Biol.*, 1968, **37**, 63-80.

51. M. Duca, P. Vekhoff, K. Oussedik, L. Halby and P. B. Arimondo, *Nucleic Acids Res.*, 2008, **36**, 5123-5138.

52. (a) S. Chakraborty, S. Sharma, P. K. Maiti and Y. Krishnan, *Nucleic Acids Res.*, 2009, **37**, 2810-2817; (b) P. Giri and G. S. Kumar, *Mol. BioSyst.*, 2010, **6**, 81-88.

53. (a) R. Maggini, F. Secco, M. Venturini and H. Diebler, *J. Chem. Soc., Faraday Trans.*, 1994, **90**, 2359-2363; (b) B. Janik, R. G. Sommer and A. M. Bobst, *Biochim. Biophys. Acta, Nucleic Acids Protein Synth.*, 1972, **281**, 152-168.

54. (a) H. Xi, D. Gray, S. Kumar and D. P. Arya, *FEBS Lett.*, 2009, **583**, 2269-2275; (b) Ö. P. Çetinkol and N. V. Hud, *Nucleic Acids Res.*, 2008, **37**, 611-621.

55. I. Bang, *Biochem. Z.*, 1910, **26**, 293-311.

56. (a) E. H. Blackburn and J. G. Gall, *J. Mol. Biol.*, 1978, **120**, 33-53; (b) E. H. Blackburn, *Nature*, 1991, **350**, 569-573.

57. A. G. Bodnar, M. Ouellette, M. Frolkis, S. E. Holt, C.-P. Chiu, G. B. Morin, C. B. Harley, J. W. Shay, S. Lichtsteiner and W. E. Wright, *Science*, 1998, **279**, 349.

58. (a) T. de Lange, *Science*, 2009, **326**, 948; (b) M. Z. Levy, R. C. Allsopp, A. B. Futcher, C. W. Greider and C. B. Harley, *J. Mol. Biol.*, 1992, **225**, 951-960.

59. N. W. Kim, M. A. Piatyszek, K. R. Prowse, C. B. Harley, M. D. West, P. L. Ho, G. M. Coviello, W. E. Wright, S. L. Weinrich and J. W. Shay, *Science*, 1994, **266**, 2011.

60. (a) J.-L. Mergny, *Nat. Chem. Biol.*, 2012, **8**, 225; (b) N. W. Luedtke, *CHIMIA Int. J. Chem.*, 2009, **63**, 134-139; (c) in *Quadruplex Nucleic Acids*, eds. S. Neidle and S. Balasubramanian, The Royal Society of Chemistry, 2006, DOI: 10.1039/9781847555298-00154, pp. 154-179; (d) in *Quadruplex Nucleic Acids*, eds. S. Neidle and S. Balasubramanian, The Royal Society of Chemistry, 2006, DOI: 10.1039/9781847555298-00131, pp. 131-153; (e) M. L. Bochman, K. Paeschke and V. A. Zakian, *Nat. Rev. Genet.*, 2012, **13**, 770.

61. (a) F. X. Han, R. T. Wheelhouse and L. H. Hurley, *J. Am. Chem. Soc.*, 1999, **121**, 3561-3570; (b) A. M. Burger, F. Dai, C. M. Schultes, A. P. Reszka, M. J. Moore, J. A. Double and S. Neidle, *Cancer Res.*, 2005, **65**, 1489; (c) J.-L. Mergny, J. M. Nicoludis, S. P. Barrett and L. A. Yatsunyk, *Nucleic Acids Res.*, 2012, **40**, 5432-5447; (d) T. Tauchi, K. Shin-ya, G. Sashida, M. Sumi, A. Nakajima, T. Shimamoto, J. H. Ohyashiki and K. Ohyashiki, *Oncogene*, 2003, **22**, 5338; (e) T.-M. Ou, Y.-J. Lu, C. Zhang, Z.-S. Huang, X.-D. Wang, J.-H. Tan, Y. Chen, D.-L. Ma, K.-Y. Wong, J. C.-O. Tang, A. S.-C. Chan and L.-Q. Gu, *J. Med. Chem.*, 2007, **50**, 1465-1474; (f) V. Kumar, A. Sengupta, K. Gavvala, R. K. Koninti and P. Hazra, *J. Phys. Chem. B*, 2014, **118**, 11090-11099; (g) S. Ghosh, A. Kar, S. Chowdhury and D. Dasgupta, *Biochemistry*, 2013, **52**, 4127-4137; (h) A. C. Bhasikuttan and J. Mohanty, *Chem. Commun.*, 2015, **51**, 7581-7597; (i) G. W. Collie and G. N. Parkinson, *Chem. Soc. Rev.*, 2011, **40**, 5867-5892; (j) S. Neidle, *J. Med. Chem.*, 2016, **59**, 5987-6011.

62. (a) in *Quadruplex Nucleic Acids*, eds. S. Neidle and S. Balasubramanian, The Royal Society of Chemistry, 2006, DOI: 10.1039/9781847555298-00180, pp. 180-207; (b) T. A. Brooks and L. H. Hurley, *Nat. Rev. Cancer*, 2009, **9**, 849; (c) T. A. Brooks, S. Kendrick and L. Hurley, *FEBS J.*, 2010, **277**, 3459-3469; (d) C. M. Azzalin, P. Reichenbach, L. Khoriauli, E. Giulotto and J. Lingner, *Science*, 2007, **318**, 798.

63. (a) Y. Fu, X. Wang, J. Zhang and W. Li, *Curr. Opin. Biotechnol.*, 2014, **28**, 33-38; (b) D. Miyoshi, M. Inoue and N. Sugimoto, *Angew. Chem., Int. Ed.*, 2006, **45**, 7716-7719; (c) A. T. Phan, V. Kuryavyi and D. J. Patel, *Curr. Opin. Struct. Biol.*, 2006, **16**, 288-298; (d) C. J. Lech and A. T. Phan, *Nucleic Acids Res.*, 2017, **45**, 6265-6274; (e) W. X. Zheng Lin, Zhang Jinli, Li Wei, *Prog. Chem.*, 2011, **23**, 974-982; (f) L. A. Yatsunyk, O. Piétrement, D. Albrecht, P. L. T. Tran, D. Renčiuk, H. Sugiyama, J.-M. Arbona, J.-P. Aimé and J.-L. Mergny, *ACS Nano*, 2013, **7**, 5701-5710.

64. J. L. Huppert, *Chem. Soc. Rev.*, 2008, **37**, 1375-1384.

65. J. L. Huppert, *FEBS J.*, 2010, **277**, 3452-3458.

66. (a) J. B. Chaires, *FEBS J.*, 2010, **277**, 1098-1106; (b) B. Heddi and A. T. Phan, *J. Am. Chem. Soc.*, 2011, **133**, 9824-9833; (c) N. Smargiasso, F. Rosu, W. Hsia, P. Colson, E. S. Baker, M. T. Bowers, E. De Pauw and V. Gabelica, *J. Am. Chem. Soc.*, 2008, **130**, 10208-10216.

67. (a) R. D. Gray and J. B. Chaires, *Nucleic Acids Res.*, 2008, **36**, 4191-4203; (b) A. T. Phan, D. J. Patel, K. N. Luu and V. Kuryavyi, *Nucleic Acids Res.*, 2007, **35**, 6517-6525; (c) J. Dai, M. Carver, C. Punchihewa, R. A. Jones and D. Yang, *Nucleic Acids Res.*, 2007, **35**, 4927-4940.

68. (a) E. Largy, J.-L. Mergny and V. Gabelica, in *The Alkali Metal Ions: Their Role for Life*, eds. A. Sigel, H. Sigel and R. K. O. Sigel, Springer International Publishing, Cham, 2016, DOI: 10.1007/978-3-319-21756-7_7, pp. 203-258; (b) D. Bhattacharyya, G. Mirihana Arachchilage and S. Basu, *Front. Chem.*, 2016, **4**; (c) J. B. Chaires, J. O. Trent, R. D. Gray and A. N. Lane, *Nucleic Acids Res.*, 2008, **36**, 5482-5515; (d) F. Zaccaria, G. Paragi and C. Fonseca Guerra, *Phys. Chem. Chem. Phys.*, 2016, **18**, 20895-20904.

69. (a) S. Takahashi and N. Sugimoto, *Molecules*, 2013, **18**; (b) D. A. T. Sekibo and K. R. Fox, *Nucleic Acids Res.*, 2017, **45**, 12069-12079; (c) D. Sun and L. H. Hurley, *J. Med. Chem.*, 2009, **52**, 2863-2874; (d) C. Zhao, J. Ren and X. Qu, *Langmuir*, 2013, **29**, 1183-1191; (e) H. Tateishi-Karimata and N. Sugimoto, *Nucleic Acids Res.*, 2014, **42**, 8831-8844; (f) I. Mamajanov, A. E. Engelhart, H. D. Bean and N. V. Hud, *Angew. Chem.*, 2010, **122**, 6454-6458; (g) C. Zhao and X. Qu, *Methods*, 2013, **64**, 52-58.

70. A. T. Phan and D. J. Patel, *J. Am. Chem. Soc.*, 2003, **125**, 15021-15027.

71. L. Ying, J. J. Green, H. Li, D. Klenerman and S. Balasubramanian, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 14629.

72. (a) T. Mashimo, H. Yagi, Y. Sannohe, A. Rajendran and H. Sugiyama, *J. Am. Chem. Soc.*, 2010, **132**, 14910-14918; (b) Y. Xu, H. Sato, Y. Sannohe, K.-i. Shinohara and H. Sugiyama, *J. Am. Chem. Soc.*, 2008, **130**, 16470-16471; (c) R. D. Gray, J. Li and J. B. Chaires, *J. Phys. Chem. B*, 2009, **113**, 2676-2683.

73. N. a. Špačková, I. Berger and J. Šponer, *J. Am. Chem. Soc.*, 1999, **121**, 5519-5534.

74. H. A. Day, P. Pavlou and Z. A. E. Waller, *Bioorg. Med. Chem.*, 2014, **22**, 4407-4418.

75. R. E. Marsh, R. Bierstedt and E. L. Eichhorn, *Acta Crystallogr.*, 1962, **15**, 310-316.

76. E. Henderson and S. Ahmed, *Nucleic Acids Res.*, 1992, **20**, 507-511.

77. K. Gehring, J.-L. Leroy and M. Guéron, *Nature*, 1993, **363**, 561-565.

78. S. Benabou, A. Aviñó, R. Eritja, C. González and R. Gargallo, *RSC Adv.*, 2014, **4**, 26956-26980.

79. (a) O. Y. Fedoroff, A. Rangan, V. V. Chemeris and L. H. Hurley, *Biochemistry*, 2000, **39**, 15083-15090; (b) S. Kendrick, Y. Akiyama, S. M. Hecht and L. H. Hurley, *J. Am. Chem. Soc.*, 2009, **131**, 17667-17676; (c) A. M. Fleming, K. M. Stewart, G. M. Eyring, T. E. Ball and C. J. Burrows, *Org. Biomol. Chem.*, 2018, **16**, 4537-4546; (d) A. M. Fleming, Y. Ding, R. A. Rogers, J. Zhu, J. Zhu, A. D. Burton, C. B. Carlisle and C. J. Burrows, *J. Am. Chem. Soc.*, 2017, **139**, 4682-4689; (e) L. Liu, B. G. Kim, U. Feroze, R. B. Macgregor and T. V. Chalikian, *J. Am. Chem. Soc.*, 2018, **140**, 2229-2238; (f) T. Fujii and N. Sugimoto, *Phys. Chem. Chem. Phys.*, 2015, **17**, 16719-16722; (g) T. E. Malliavin, J. Gau, K. Snoussi and J.-L. Leroy, *Biophys. J.*, 2003, **84**, 3838-3847.

80. (a) X. Li, Y. Peng, J. Ren and X. Qu, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 19658; (b) J. Zhou, C. Wei, G. Jia, X. Wang, Z. Feng and C. Li, *Mol. BioSyst.*, 2010, **6**, 580-586; (c) J.-L. Mergny, L.

Lacroix, X. Han, J.-L. Leroy and C. Helene, *J. Am. Chem. Soc.*, 1995, **117**, 8887-8898; (d) A. Rajendran, S.-i. Nakano and N. Sugimoto, *Chem. Commun.*, 2010, **46**, 1299-1301; (e) L. Lannes, S. Halder, Y. Krishnan and H. Schwalbe, *ChemBioChem*, 2015, **16**, 1647-1656; (f) X. Chen, X. Zhou, T. Han, J. Wu, J. Zhang and S. Guo, *ACS Nano*, 2013, **7**, 531-537; (g) H. A. Day, E. P. Wright, C. J. MacDonald, A. J. Gates and Z. A. E. Waller, *Chem. Commun.*, 2015, **51**, 14099-14102; (h) H. A. Day, C. Huguin and Z. A. E. Waller, *Chem. Commun.*, 2013, **49**, 7696-7698; (i) S. Takahashi and N. Sugimoto, *Phys. Chem. Chem. Phys.*, 2015, **17**, 31004-31010; (j) Y. Chen, K. Qu, C. Zhao, L. Wu, J. Ren, J. Wang and X. Qu, *Nat. Commun.*, 2012, **3**, 1074.

81. Y. Zhao, Z.-x. Zeng, Z.-y. Kan, Y.-h. Hao and Z. Tan, *ChemBioChem*, 2005, **6**, 1957-1960.

82. D. Liu and S. Balasubramanian, *Angew. Chem., Int. Ed.*, 2003, **42**, 5734-5736.

83. (a) W. Guo, C.-H. Lu, R. Orbach, F. Wang, X.-J. Qi, A. Cecconello, D. Seliktar and I. Willner, *Adv. Mater.*, 2015, **27**, 73-78; (b) I. V. Nesterova and E. E. Nesterov, *J. Am. Chem. Soc.*, 2014, **136**, 8843-8846; (c) L. Shi, P. Peng, Y. Du and T. Li, *Nucleic Acids Res.*, 2017, **45**, 4306-4314; (d) W. Li, J. Wang, J. Ren and X. Qu, *Angew. Chem., Int. Ed.*, 2013, **52**, 6726-6730.

84. (a) S. Kendrick, H.-J. Kang, M. P. Alam, M. M. Madathil, P. Agrawal, V. Gokhale, D. Yang, S. M. Hecht and L. H. Hurley, *J. Am. Chem. Soc.*, 2014, **136**, 4161-4171; (b) M. Gellert, M. N. Lipsett and D. R. Davies, *Proc. Natl. Acad. Sci. U. S. A.*, 1962, **48**, 2013.

85. Y. Peng, X. Li, J. Ren and X. Qu, *Chem. Commun.*, 2007, DOI: 10.1039/B710950D, 5176-5178.

86. S. Dzatko, M. Krafcikova, R. Hänsel-Hertsch, T. Fessl, R. Fiala, T. Loja, D. Krafcik, J.-L. Mergny, S. Foldynova-Trantirkova and L. Trantirek, *Angew. Chem., Int. Ed.*, 2018, **57**, 2165-2169.

87. M. Zeraati, D. B. Langley, P. Schofield, A. L. Moye, R. Rouet, W. E. Hughes, T. M. Bryan, M. E. Dinger and D. Christ, *Nat. Chem.*, 2018, **10**, 631-637.

88. J. Y. Lee, L.-S. Wang, M. L. Kozak, P. Yue, Q. Chen, S. G. Hershman and F. B. Johnson, *Nucleic Acids Res.*, 2007, **36**, 144-156.

89. S. Satpathi, A. Sengupta, V. M. Hridya, K. Gavvala, R. K. Koninti, B. Roy and P. Hazra, *Sci. Rep.*, 2015, **5**, 9137.

90. S. Satpathi, M. Kulkarni, A. Mukherjee and P. Hazra, *Phys. Chem. Chem. Phys.*, 2016, **18**, 29740-29746.

91. S. Satpathi, R. K. Singh, A. Mukherjee and P. Hazra, *Phys. Chem. Chem. Phys.*, 2018, **20**, 7808-7818.

92. S. Satpathi, K. Das and P. Hazra, *Chem. Commun.*, 2018, **54**, 7054-7057.

Chapter 2

# A Green Solvent Induced DNA Package

2.1 Introduction

For safe storage of genetic information, compaction of DNA is appearing as a budding topic of interest. A normal living cell contains a huge number of DNA which in average consist of around six billion DNA base pairs. Considering each base pair length ~0.34 nm, six billion DNA base pairs comprise a total distance of 2 meters. Including the total number of cells in a human body (nearly 50 trillion), an incredible length of 100 trillion meters of DNA per human appears in the picture. Miraculously, nature manages this length of DNA inside a nucleus by the formation of a complex named as chromatin.[1] Not only natural compaction but artificial DNA compaction is also an effective way to store and carry genetic information.[2] Compacted DNA also found to be resistant/stable towards any external shocks like UV radiation.[3] Moreover, mechanically compacted DNA has recently been used as nanostructure template[4] and also applied as protection against chemical, biochemical and mechanical stresses.[5] Learning from the nature, researchers have started applying the compaction tricks using polyamines,[6] surfactants,[7] liposomes,[8] nanoparticles,[9] polymer,[10] osmoticants like polyethylene glycol (PEG),[11] dendrimers,[12] multivalent ions,[13] metal complex,[14] cyclodextrin,[15] peptides,[16] and proteins[17] for compaction and storage of DNA for prolonged duration. Among these compacting agents, most important naturally occurring DNA compaction agents are proteins and polyamines like of spermine and spermidine *etc.*[16-17] Under physiological pH (~7.4), polyamines generally exists in positively charged form, which participates in strong electrostatic interaction with negatively charged phosphate backbone of DNA, and thereby results in compaction and enhances the thermal stability of DNA.[3, 18] Thus different types of linear and branched polyamines have been identified in many hyperthermophilic archaea.[19] Not only polyamines but several multivalent cations can also lead to DNA compaction by decreasing intra-strand DNA repulsion.[20] Surfactants with a positively charged head group can act as multivalent cations above critical micellar concentration (CMC), which can neutralise the negatively charged DNA surface, and hence facilitates the compaction process.[7b] Moreover, the hydrophobicity of surfactants tails also plays a crucial role in DNA compaction.[7a] Recently Ganguli *et al.* have used positively charged nanoparticles to condense DNA strands.[9a] Instead of forming a completely compacted DNA, it generates several compacted small loops of DNA around each nanoparticles[9a]. In contrary to Ganguli *et al.*, Rudiuk *et al.* reported that negatively charged silica nanoparticle improves the surfactant effectivity for DNA compaction.[21] Moreover Zinchenko *et al.* mentioned that negatively charged silica (SiO_2) nanoparticle can lead to DNA compaction by causing depletion in DNA coil structure due to the excluded volume of nanoparticles.[9b] Not only ions, osmoticants, like organic solvents, polyethylene glycol (PEG) can also induce DNA compaction by lowering the dielectric permittivity of solution.[22] Yoshikawa *et al.* studied the DNA compaction by polyethylene glycol with pendant amino groups (PEG-A). They have observed that DNA is collapsed by PEG-A similar to neutral PEG, but it happens at a lower concentration.[11] Dendrimers are also considered as one of the most efficient DNA compacting agents. Its compaction ability depends upon the size and charge distribution.

Generally, dendrimers used at higher dendrimers-to-DNA charge ratio (5-10) account for their brilliant DNA compacting ability.[12a] But highly charged dendrimers are very toxic to living cells. Thus, several research groups studied this mode of compaction process using modified dendrimer surface by acetylation, PEGylation *etc.*[23] Besides these substances, researchers have also studied the impact of chirality on DNA compaction by using chiral polycations and chiral polypeptide.[6b, 24] Interesting reports are also available about reversibly controlled compaction/decompaction process of DNA.[25]

Since the discovery of ionic liquid (IL) in 1965,[26] it has been attracted significant attention among researchers due to its unique properties like low vapour pressure, inflammability, very high thermal, chemical stability and excellent conductivity.[27] Thus IL has been exploited in various kind of research field like separation, synthesis, electrochemistry, catalysis *etc.*[27a, 28] Considering its low cytotoxicity and high biodegradability, many groups have studied the interaction of room temperature ionic liquid (RTIL) with biomacromolecules like protein, DNA *etc.*[29] Nevertheless, literature of DNA-IL interaction is somewhat controversial due to several contradictory explanations provided by different reports. First, Cheng *et al.*[30] suggested that cationic part of [Bmim][PF$_6$] can intercalate inside the DNA base pairs, whereas later Ding *et al.*[31] indicated an aggregation induced DNA perturbation by [Bmim][Cl] IL. Very recently in 2012, Chandran *et al.*[32] proposed a groove binding mechanism of RTIL in DNA, where they have shown that [Bmim] cation interacts with the minor groove of DNA through various non-covalent interactions. Thus, it is clear that the mode of DNA-IL interaction is still ambiguous and this may be due to the inertness of IL towards any thermal and optical signals.
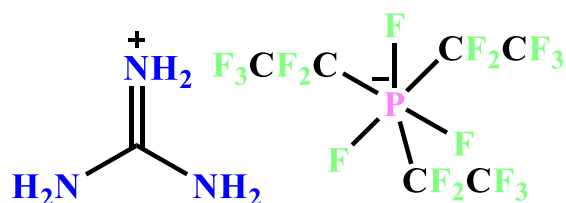
Scheme 2.1 Chemical structure of guanidinium tris(pentafluoroethyl)trifluorophosphate (Gua-IL).

Generally, a substance, which can increase the intrastrand/interstrand attraction between DNA molecules, is treated as DNA compacting agent. ILs contain a cationic part which can compensate DNA backbone charges through electrostatic interaction and may lead to DNA compaction through the formation of self-aggregated polyvalent (containing multiple positive and negative charges) structure. Notably, micellar like the aggregated structure of ILs are familiar in literature,[33] and can behave as a multivalent ion to facilitate DNA compaction process. Although IL pursues all possible character of a DNA compaction agent, it has not been considered as a potential DNA compacting agent. In this study, we have shown the DNA compaction ability of a room temperature ionic liquid (RTIL), namely, guanidinium tris(pentafluoroethyl)trifluorophosphate (Gua-IL) (scheme 2.1), which consists of a smallest possible cation (i.e., Gua$^+$) among the available ionic liquids in literature for its facile

accessibility to both base pairs and grooves of DNA. This Gua-IL mediated DNA compaction process has been probed by different spectroscopic techniques like UV-Vis spectroscopy, fluorescence, steady-state anisotropy, circular dichroism (CD). We have also proposed the plausible mechanism of DNA compaction by Gua-IL using dynamic light scattering (DLS), scanning electron microscopy (SEM) and tunnelling electron microscopy (TEM), confocal fluorescence microscopy along with above mentioned spectroscopic techniques. Unlike other compaction agents, we found IL stands as a better one due to its biodegradability and low concentration effectivity, which might be useful for the extraction and storage of genetic blueprints for the future.

2.2 Results

2.2.1   Steady State Emission Studies

From the last few decades, steady-state fluorescence spectroscopy has emerged as a powerful technique for characterisation of association properties between small molecules and nucleic acids. As both of guanidinium tris(pentafluoroethyl)trifluorophosphate (Gua-IL) and DNA do not exhibit any fluorescence, studies have been carried out using two different DNA binding probes; a well-known intercalator, i.e. ethidium bromide (EB) and a minor groove binder, i.e. 4',6-diamidino-2-phenylindole (DAPI). We have performed displacement assay experiments by these two DNA binding dyes for characterisation of Gua-IL, and ct-DNA interaction and the results are shown in Figure 2.1. Fluorescence intensity of DNA-bound EB is significantly higher than that of unbound EB (Appendix 2.1A) due to its insertion in more hydrophobic microenvironment (i.e. in between base pairs) and also for the less accessibility of intercalated EB by quenchers like H_2O and/or dissolved O_2.[34] Moreover, emission peak of EB (λ_{ex} at 478nm) in Phosphate Buffer (PB) (pH = 7.4) appears at 610 nm, which shifts to 597 nm up to maximum addition of DNA (180 µM). When EB bound DNA solution is being titrated with Gua-IL, no significant change in emission maximum has been observed in micromolar concentration range of Gua-IL, but in milli-molar concentration range (> 1 mM) a strong quenching (~84% up to 40 mM of Gua-IL) in fluorescence intensity is observed (Figure 2.1A). To verify whether the changes in fluorescence spectral feature arises due to Gua-IL alone or not, we have done control experiment with Gua-IL alone (in the absence of DNA), and we have noticed that some extent of fluorescence intensity reduction of EB in presence of Gua-IL (Appendix 2.2). Here, it is very important to highlight that Gua-IL is well known for its ability to form hydrogen bonding.[35] Thus, these anionic aggregates/anions have higher propensity to form hydrogen bonding with the amine group of EB which can lead to the abstraction of the proton and results in the fluorescence quenching.[36]

To get insight into the effect of Gua-IL over a DNA groove binder, we have carried out similar kind of steady state experiment with a well-known minor groove binder, i.e. DAPI. DAPI exhibits emission maximum in bluish green region (at 483 nm) and a ~30 nm blue shift is observed along with increment in emission intensity with increasing ct-DNA concentration (Appendix 2.1B). Figure 2.1B

depicts the effect of Gua-IL on the fluorescence profiles of DNA-bound DAPI, and the results exhibit a considerable decrease in emission intensity up to the maximum Gua-IL concentration. Interestingly, similar to EB, DAPI also shows a significant change in emission intensity only around milli-molar concentration range of Gua-IL. Note that the decrease in emission intensity at maximum addition of Gua-IL (40 mM) is not so effective like EB, inferring that Gua-IL cannot remove all the groove binder (DAPI), which indicates possible minor groove binding nature of Gua-IL. We have also performed a control experiment with Gua-IL and DAPI, and it shows a very negligible effect on emission maximum of DAPI (Appendix 2.2B).
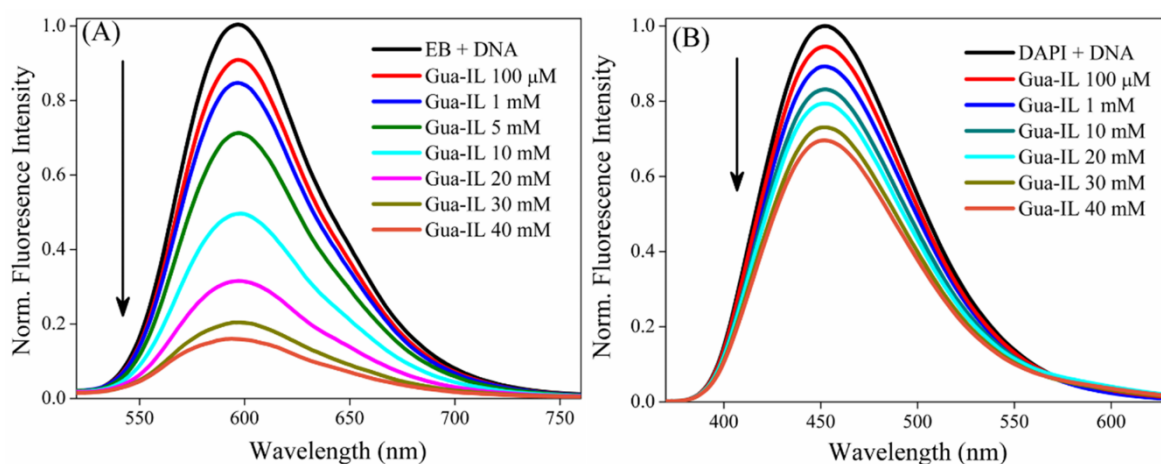
Figure 2.1 Fluorescence spectra of dye bound ct-DNA (~200 μM) ((A) EB (~5 μM) (B) DAPI (~5 μM)) with increasing concentration of Gua-IL.
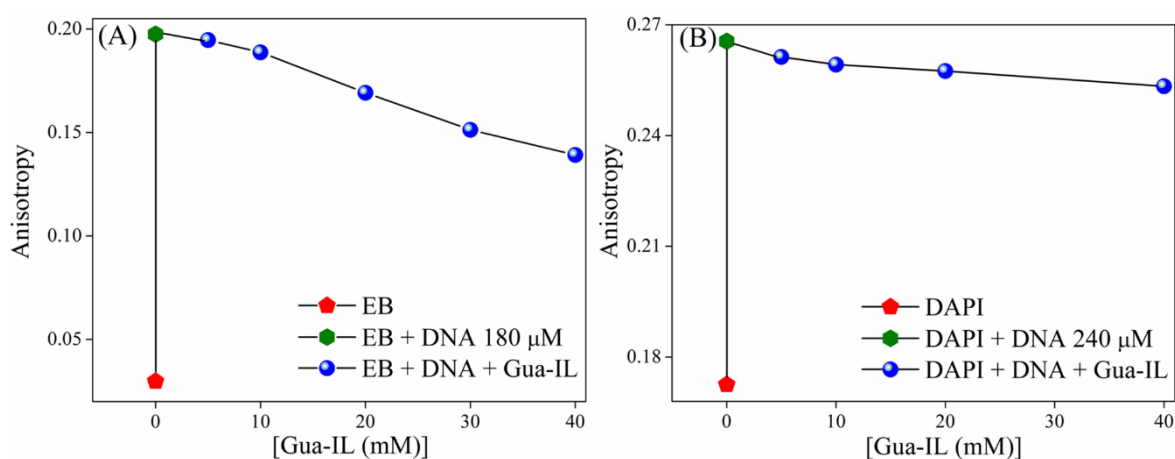
2.2.2   Steady State Anisotropy Studies

Fluorescence depolarization takes place due to rotational diffusion of the fluorophores in solvent medium[37], and steady-state fluorescence anisotropy is a measure of this depolarization or rotational diffusion of molecules.[37] A higher anisotropy value suggests slower rotational diffusion arising due to the rigid environment around the molecule. With increasing addition of DNA in both EB and DAPI, we have observed enhancement (Figure 2.2A and 2.2B respectively) in anisotropy value, which corresponds to the interaction of dyes with respective sites of DNA. Interestingly, when Gua-IL is added to the solutions of EB-DNA and DAPI-DNA above milli-molar concentration (Figure 2.2A and 2.2B respectively), then anisotropy values for both the dyes decreases rapidly, although no such change in micromolar concentration range is observed.

2.2.3   Circular Dichroism (CD) Studies

To gain insight about the DNA secondary structure perturbation by Gua-IL, we have next focused on circular dichroism measurements. ct-DNA exhibits two characteristics peaks (Appendix 2.4); one positive band around 280 nm complemented to π-π base stacking and a negative band around 245 nm for helicity.[38] While titrated with EB, DNA generates an induced CD signal near 303 nm (Appendix

2.4) along with decreasing π-π stacking peak and increasing helicity peak, which is a well-known feature for the intercalation of dye in DNA.[39] When Gua-IL is added to the DNA-EB system, no commendable variation in CD signal is noticed up to 100 µM of Gua-IL (Figure 2.3A). Afterwards evident alteration is perceived in induced CD signal, helicity peak and base stacking peak. Notably, after 1 mM concentration of Gua-IL, induced CD signal completely vanishes which is suggestive of complete removal of intercalated EB from DNA (Figure 2.3A). Moreover, both the helicity and base


Figure 2.2 Change of steady-state anisotropy of (A) EB (20 µM) (B) DAPI (20µM) in the presence of ct-DNA with increasing concentration of Gua-IL.

stacking peak decreases from this concentration. In case of DNA-DAPI interaction, CD (Figure 2.3B) spectra show all sorts of characteristics features including an induced CD signal near 360 nm as per the reported literature[40], which confirms the binding of DAPI to DNA. Addition of Gua-IL (Figure 2.3B) decreases the induced CD signal peak to a certain extent, inferring the replacement of DAPI molecules from the minor groove of ct-DNA. Interestingly, induced CD signal does not vanish completely as it is observed in the case of DNA-EB system, indicating that some DAPI molecules still bind to DNA even at a higher concentration of Gua-IL.

2.2.4 Dynamic Light Scattering and Zeta Potential Measurements

For many years, dynamic light scattering is manifesting itself as a workhorse for measurement of size and shape of the biological macromolecule. To clarify the nature of changes on DNA conformation by Gua-IL, we have carried out dynamic light scattering measurements. As the molecular weight of ct-DNA is high, we have performed experiments at a low concentration of DNA (20 µM) in tris buffer (pH= 7.4) in order to avoid self-interaction between the DNA molecules. The intensity weighted size distribution of DNA alone is found unimodal in nature and is presented on the lower curve in Figure 2.4A. The mean hydrodynamic radius of DNA is about 350 nm, which is in agreement with the size of high molecular weight DNA.[41] When Gua-IL is added to the DNA solution, the size distribution peak at 350 nm gradually shifted to a lower hydrodynamic radius (~ 260 nm) and no new peak appears up to 1 mM concentration of Gua-IL. After ≥1 mM Gua-IL concentration, two distinct size distribution peaks

emerge at 58 nm and 255 nm. In order to find out whether the above-mentioned peaks appeared due to aggregation process of Gua-IL or not, we have carried out a DLS study in absence of DNA at different concentrations of Gua-IL (Appendix 2.6). An intensity weighted distribution peak around 30-40 nm at a higher concentration of Gua-IL is attributed to the aggregation of anionic part of Gua-IL. For further clarification about Gua-IL aggregation, we have performed a DLS study with guanidine hydrochloride (Gua HCl), which contains similar guanidinium cation like Gua-IL, but it contains chloride anion as the counterpart. Guanidine hydrochloride does not show any intensity weighted distribution, which indicates the role of anionic part of Gua-IL for the aggregation process. In order to validate the DLS results, we have also executed the zeta potential measurement experiment (Figure 2.4B) as a function Gua-IL concentration. Zeta potential is the electrostatic potential generated in an applied electric field due to the attraction between charged species and the oppositely charged electrode.[42] Zeta potential depends on the location of the shear plane, which is the interface between stern and diffuse layers of the double layer.[42] Likewise, zeta potential of DNA depends on its conformation in accordance with the position of the shear plane. Thus, zeta potential measurement is an excellent tool for the determination of effective charge density, which in term is an essential parameter for DNA compaction. It is clear from Figure 2.4B that DNA shows a zeta potential value of -64 mV and it gradually increases with Gua-IL addition showing a visible inflection point near 1 mM of Gua-IL concentration.
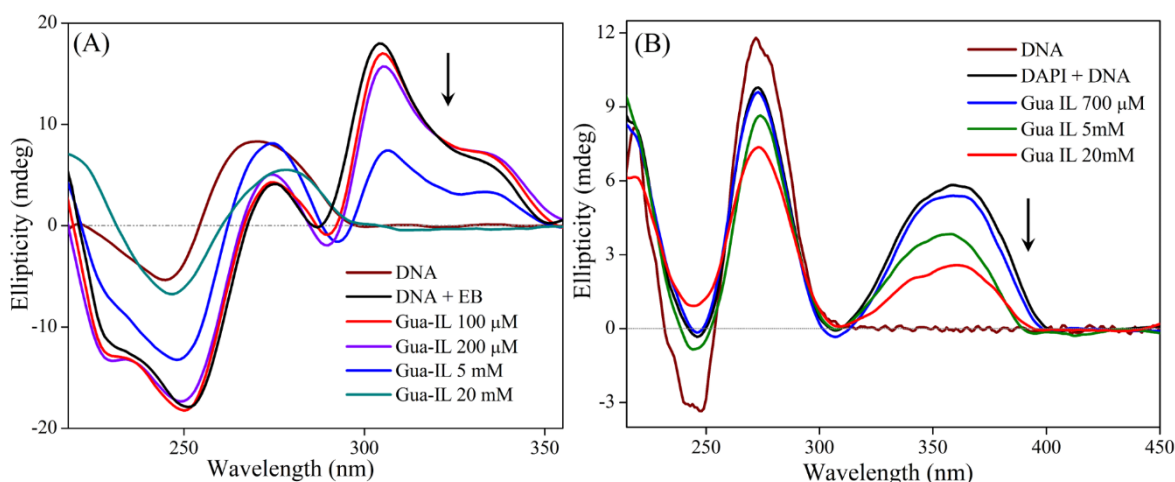

Figure 2.3 Circular dichroism spectra of ct-DNA (~ 500 µM) with increasing concentration of dye and with micro-molar addition of Gua-IL (A) in DNA-EB system and (B) in DNA-DAPI system.

2.2.5 Thermal Melting Study

UV melting experiment was performed with ct-DNA in absence and presence of Gua-IL to get insight into the structural stability of DNA by Gua-IL. Generally, with increase in temperature, double-stranded DNA denatures to single-stranded DNA with enhancement in absorbance at 260 nm due to unstacking of nucleobases and exhibits a transition point, i.e. melting temperature. ct-DNA (1 µM) in tris-HCl buffer (pH = 7.4) shows a transition point at ~ 41 ℃ and this melting point gradually shifted to a lower temperature (~ 21 ˚C) with an increase in Gua-IL concentration (Figure 2.5).

2.2.6 Microscopic View: Field Emission Scanning Electron Microscopy (FE-SEM) and Transmission Electron Microscopy (TEM) Study

Field emission scanning electron microscopy (FE-SEM) is used to probe the morphological alteration of DNA during the compaction process induced by Gua-IL. All the FE-SEM samples are prepared in tris HCl (pH=7.4) buffer to avoid any kind of artefact resulting from salt morphology. It is evident from Figure 2.6A that ct-DNA (in the absence of Gua-IL) appears in super-coiled morphology, which is expected for a long DNA like calf thymus. When Gua-IL is gradually added, some morphological changes in the super-coiled DNA structure are observed depending on the molar ratio of IL to DNA, i.e., [Gua-IL]/[DNA]. At lower molar ratio 10 ([Gua-IL]/[DNA]), no specific difference in morphology is observed. A transition in morphology appears at a molar ratio of 20 ([Gua-IL]/[DNA]), which is depicted in Figure 2.6B, and at a high molar ratio of 40 ([Gua-IL]/[DNA]), a globular morphology appears in the context (Figure 2.6C). We have also probed the Gua-IL aggregation process in the absence of DNA by FE-SEM and TEM studies (Figure 2.6D and Appendix 2.7B). In both FE-SEM and TEM, we have found a distribution of sizes ranging from ~30 nm to ~200 nm. A closer look at the TEM images of Gua-IL (Figure 2.6D), it is evident that higher sized aggregated structures result from the smaller sized aggregates, and this is believed to appear due to drying effect of the sample during FE-SEM and TEM studies. Notably, the size of these smaller aggregations is very much comparable with the size obtained from DLS study (i.e. ~30 nm).
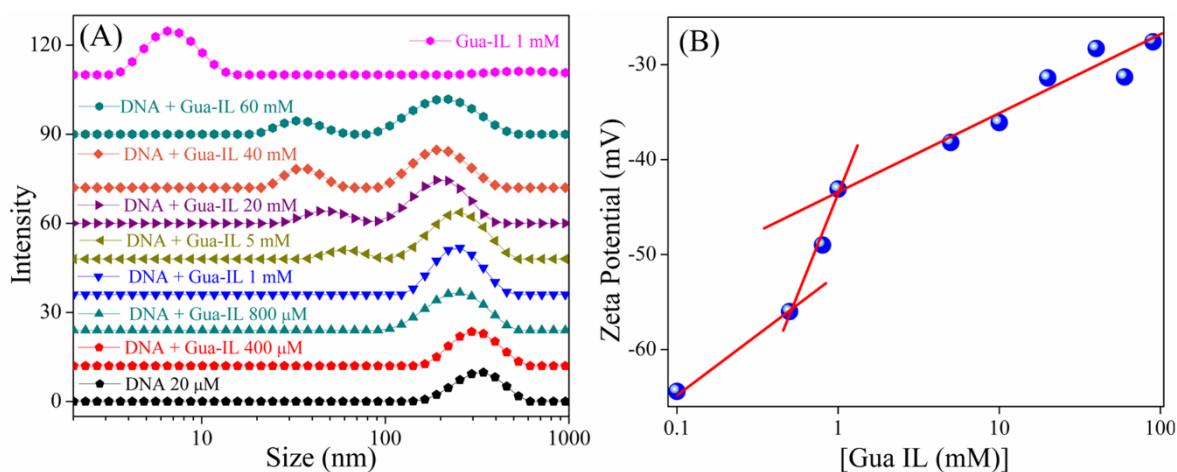

Figure 2.4 A) Intensity-weighted distribution profile (measured by DLS) of ct-DNA (20 μM) with increasing concentration of Gua-IL. (B) Zeta potential plot as a function of Gua-IL concentration. X-axes in both the figures are represented in a logarithm scale.

2.2.7 Fluorescence Microscopy Study

In order to visualise this DNA compaction process, we have executed the fluorescence microscopy study of dye (DAPI) labelled DNA in the presence and absence of Gua-IL. Here it is pertinent to mention that getting a microscopic image of ct-DNA is difficult because the resolution of the fluorescence microscope is limited to few micrometres. Therefore, we have selected plasmid DNA

(4600 Kbp) as a model DNA for the visualisation of this compaction process. DAPI labelled DNA exhibits a circular morphology (Figure 2.7), which is quite evident for a plasmid DNA. However, when Gua-IL is added to the dye-labeled DNA solution, a distinctive change in DNA volume is observed (Figure 2.7).
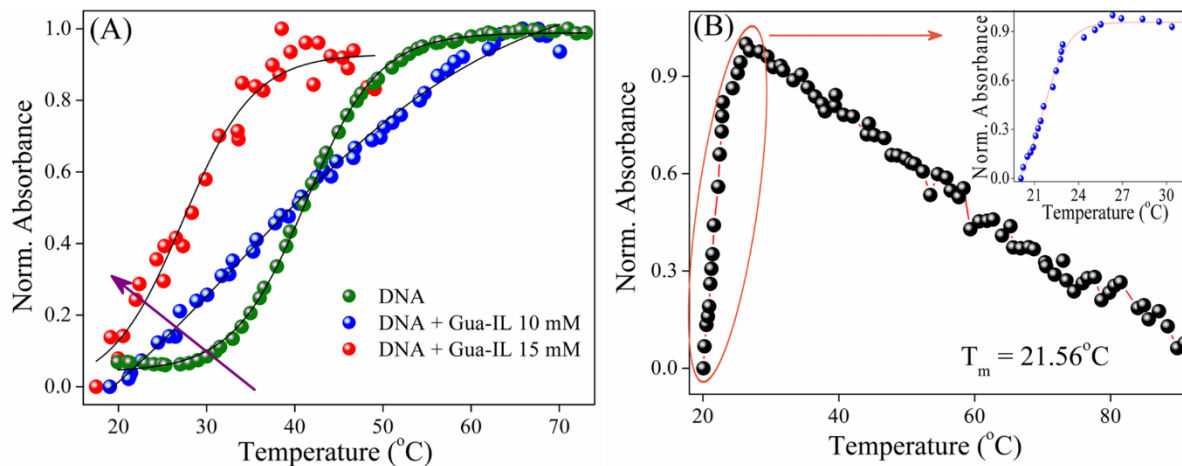

Figure 2.5 A) UV melting studies of ct-DNA (~1 µM) in (A) different concentrations of Gua-IL (up to 15 mM) (B) in the presence of 40 mM Gua-IL. The inset shows the fitting results.

2.3 Discussion

Reduction in fluorescence intensity of EB in the presence of Gua-IL suggests that intercalated EB molecules are coming out from the DNA (Figure 2.1A). Apparently, it can be thought that Gua-IL may intercalate in between the base pairs and replaces EB molecules. Here it is pertinent to mention that previously Cheng *et al.*[30] reported the intercalation of [Bmim][Cl] ionic liquid in between DNA base pairs. In [Bmim][Cl] ionic liquid, [Bmim] moiety consists of imidazolium ring with delocalized π-clouds, which can participate in stacking interaction with nucleobases. Notably, intercalators have features like labile, planar π-electrons, which can stack efficiently with nucleobases.[43] Being a triangular non-aromatic structure, guanidinium cation cannot act as an intercalator. It is noteworthy to mention that the effect of Gua-IL comes into the picture from milli-molar concentration range in both the cases (EB and DAPI). A normal intercalation or minor groove binding process is supposed to be effective from very low concentration (even in micromolar concentration range) range, which is not observed here (decrement in emission maximum for EB and/or DAPI is taking place from a certain concentration (>1 mM)). Hence, the observed results create doubt about the possibility of the intercalation and/or minor groove binding mode of Gua-IL with DNA. Notably, the possibility of some molecular aggregation of Gua-IL can be speculated from the DLS, FE-SEM and TEM studies at higher concentration (≥1 mM) of Gua-IL, which might lead to the displacement of dye from their respective positions in DNA. Here it is relevant to mention that we have also noticed identical observation with shorter DNA (DD, 5′-d-(CGCGAATTCGCG)₂-3′) suggesting the similar impact of Gua-IL on both short (Appendix 2.3 and Appendix 2.5) and long DNA.

Like steady-state fluorescence results, anisotropy results also show identical observation in milli-molar concentration range. But, the changes in anisotropy are not so prominent like steady-state intensity profiles. This is possibly due to the enhanced viscosity of the respective dye-DNA solution in the presence of increasing Gua-IL concentration. In the case of DAPI-DNA, anisotropy decrement in the presence of Gua-IL is less, and it is attributed to the lesser extent of DAPI displacement, which is corroborative with fluorescence displacement assay results. The decreasing trend in anisotropy values for both of the dyes (EB and DAPI) implies the similar possibilities like either replacement of the dyes due to binding of Gua-IL at the respective positions or some molecular aggregation of Gua-IL, which displaces the dyes from DNA.
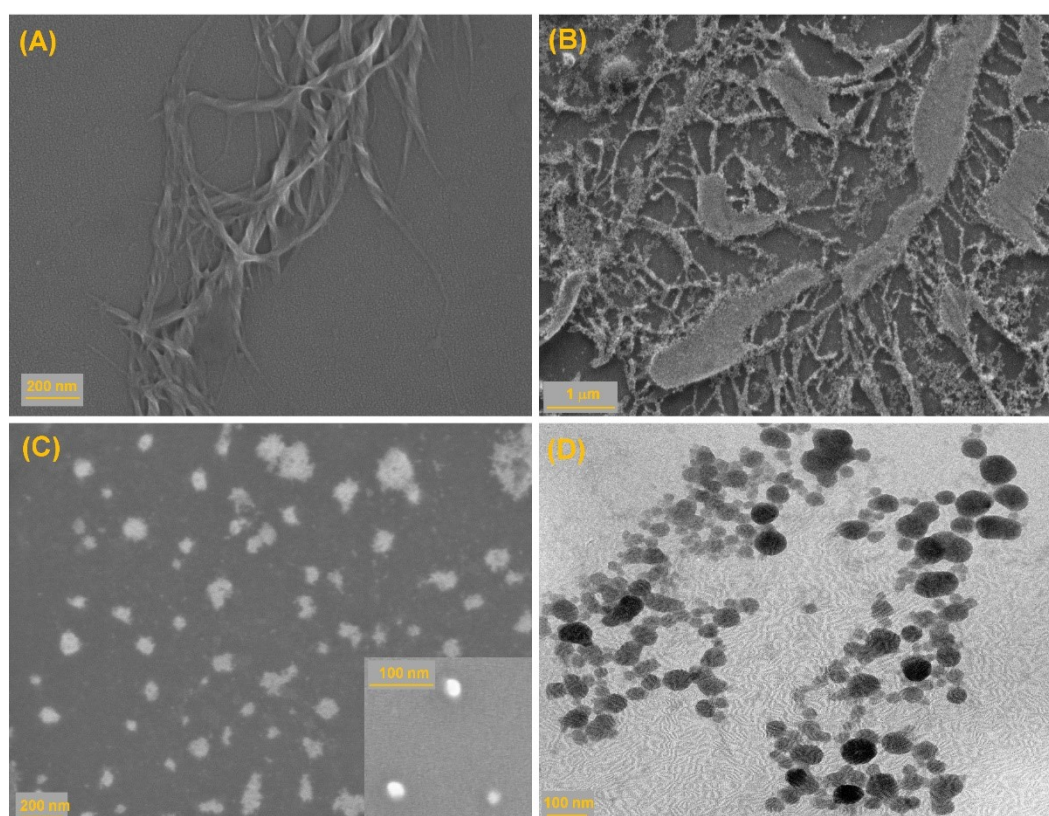

Figure 2.6 Microscopic FE-SEM images (A to C) in different molar ratio of [Gua-IL]/[ct-DNA] : (A) only ct-DNA (B) 20 (C) 40. (D) The TEM image of Gua-IL (5 mM).

In addition to steady-state fluorescence measurements, we have also performed circular dichroism study, which provides information about the structural changes of DNA in the presence of Gua-IL. In the case of DNA-EB system, no evident alteration in peaks is observed at lower concentration (up to 50 µM) of Gua-IL (Figure 2.3A). Afterwards, we have witnessed some evident changes in the spectra, and interestingly, >1 mM Gua-IL concentration, induced signal (at 303 nm) totally disappeared. Moreover, above this concentration range (>1 mM), we have also observed the CD signal decrement (Figure 2.3A) for both of base stacking and helicity bands, which are not the characteristic features of any intercalator in DNA. Thus, these observations again reinforce our claim that Gua-IL is not intercalating between the base pairs of DNA. However, the above CD results suggest

that structural reformation of ct-DNA takes place by Gua-IL. Analogous experiments with DNA-DAPI also demonstrate similar results (Figure 2.3B), where any alteration was only perceived in milli-molar concentration (>1 mM) range. Another interesting fact is that induced signal at 360 nm did not vanish at the maximum addition of Gua-IL inferring that Gua-IL cannot replace all DAPI molecules from the minor groove of ct-DNA. These results are well corroborative with the steady state results, where we have also observed Gua-IL was unable to replace all of DAPI molecules from DNA. In a nutshell, CD results confirmed that Gua-IL has some structural influence on DNA after a certain concentration (1 mM) except intercalation and minor groove binding modes of interaction.

Dynamic light scattering study of ct-DNA (Figure 2.4A) exhibits a unimodal peak at 350 nm, which attributes to the translational mode of extended DNA.[31] With the milli-molar addition of Gua-IL (>1 mM) to the system, the intensity weighted distribution peak (Figure 2.4A) appears to be bimodal with hydrodynamic peaks at ~60 nm and ~250 nm. From DLS study of Gua-IL alone (in the absence of DNA), it is evident that self-aggregation of Gua-IL takes place above 1 mM concentration (i.e. CAC), and the size of the aggregated structure is around ~10 nm at 1 mM concentration. Moreover, the size of the aggregated structure increases as the concentration of Gua-IL increases, and the size of the aggregated structure is ~35 nm at 40 mM concentration of Gua-IL. Notably, the opposite finding is observed when Gua-IL is added to the DNA system. Therefore, the two different size distributions appeared at ~60 nm and ~250 nm in DNA-IL system cannot be due to the aggregation process of Gua-IL itself. Notably, DNA compaction by surfactants, dendrimers also perceived similar size distribution.[41, 44] Thus, the appearance of a peak at the lower hydrodynamic radius (i.e. ~60 nm) refers to some anomalous structure of DNA, which might represent the compacted or condensed state of DNA by Gua-IL. The small peak at ~60 nm gradually shifts to a lower hydrodynamic radius with an increase in the concentration of Gua-IL, which accounts for the more compacted structure of DNA. The extended peak for DNA (at 350) nm in DNA-IL system also shifts towards lower hydrodynamic radius with increasing Gua-IL in solution. This reduction in hydrodynamic radius may be due to the
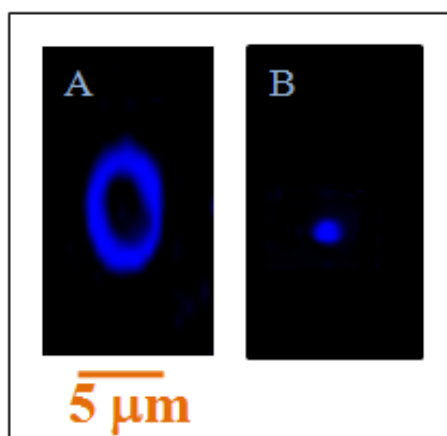

Figure 2.7 Fluorescence microscopy images of DNA (0.5 µM) labelled DAPI (0.5 µM) in (A) absence and (B) the presence of Gua-IL (1 mM).

spontaneous binding of guanidinium cation with the phosphate group of DNA.[31] To explore the aggregation mechanism of Gua-IL, we have executed the similar DLS study with Gua HCl, which has similar cationic part (i.e. Guanidium ion) but a different anionic part. Interestingly, any intensity weighted peak at similar concentration range was not present for Gua HCl, which reflects the involvement of anionic part of Gua-IL in micellar kind of aggregation, instead of guanidinium cationic part. Thus, we believe that anionic part of Gua-IL, which contains hydrophobic -CF_2CF_3 group, may aggregate like micelles to avoid unfavourable interactions with an aqueous medium. This type of IL aggregation is a well-known feature in literature[33], and the zeta potential study also reflects the same. Zeta potential result is also well corroborative with our previous DLS, and steady-state results. It shows an inflection point near 1 mM Gua-IL concentration (Figure 2.4B), and afterwards zeta potential value again increases with increasing Gua-IL concentration indicating the formation of higher order compacted structure from that globular structure. The shifting of DNA melting temperature (Figure 2.5) at the lower site (from 41°C to 21°C) in the presence of Gua-IL refers to the instability of this compacted DNA due to a loss in a double helical conformation, which is also reflected in CD studies. Interestingly, this trend in melting studies is comparable to the compaction process mediated by nanoparticles.[9b] Now, Figure 2.5 B corresponds to the UV melting profile of ct-DNA in higher concentration of Gua-IL (40 mM) which shows a reduction in the absorbance after the melting point. In the higher concentration of ionic liquid (40 mM), amount of the compacted DNA is very high in the solution. With increase in temperature, these aggregated DNA structures can collapse in such a way, where the nucleobases of the DNA molecules are stacked in a smaller space which is a very much possible for the DNA compaction process. In the basis of all our experimental observations, we infer that the compaction is taking place by the electrostatic dragging of DNA strands (by the guanidinium cation) to the aggregated surface of Gua-IL, which is formed by the hydrophobic chains of the Gua-IL anionic part. Accumulating steady-state fluorescence, DLS, CD and melting results, we anticipate that Gua-IL does not involve in intercalation and/or minor groove binding, rather induces DNA structure in such a way that it leads to compaction from coil to globule state.

The FE-SEM studies provide direct pictorial evidence of this compaction process (Figure 2.6). A supercoiled morphology appears for ct-DNA, which is quite expected for a long DNA like calf thymus (Figure 2.6A). An obvious alteration in the morphology is observed when Gua-IL is introduced to the system at a molar ratio of 20 ([Gua-IL]/[DNA]), indicating the initiation process for DNA compaction (Figure 2.6B). At a higher molar ratio ([Gua-IL]/[DNA]) of 40, a globular morphology is observed (Figure 2.6C). This morphology compliments well with the earlier reported DNA compaction morphology,[45] and the diameter of these globules are in good agreement with the diameter (around 40 nm) obtained from DLS study. Both the SEM and TEM studies confirm the formation of aggregated Gua-IL structure having spherically shaped morphology. Moreover, there exists a distribution of aggregates size having the smallest diameter of 30 nm. This size is very much comparable with that of

DLS study, and the morphology is also quite similar to the micelles.[46] Notably, due to complete drying of Gua-IL sample for SEM/TEM studies, there is a high chance for the formation of larger aggregates (around 180 nm) from the smaller aggregates/micelles (around 30 nm), which is not observed in DLS study. The difference in size distribution (for Gua-IL) observed between DLS, and microscopic images are due to the different states of samples. DLS was carried out in solution phase, whereas SEM study has been studied in dry condition where aggregations in between micelles are quite obvious.

Fluorescence microscopy images (Figure 2.7) render clear evidence for this compaction process induced by Gua-IL. DAPI labelled DNA shows a circular structure, which is evident for a plasmid DNA (Figure 2.7). Interestingly, this circular structure significantly gets compacted by the Gua-IL (Figure 2.7). In a nutshell, FE-SEM and fluorescence microscopy images have provided direct evidence for the Gua-IL induced DNA compaction process.
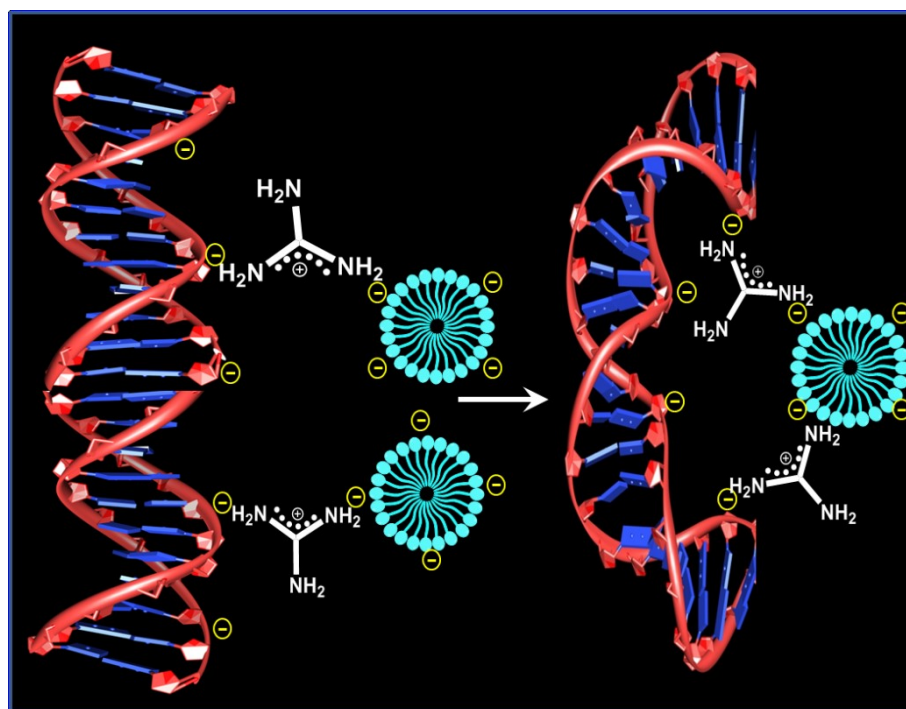
Scheme 2.2 Probable compaction mechanism of DNA by Gua-IL.

2.4 Plausible compaction mechanism

In literature, there are reports about different interaction modes of IL with DNA. *Cheng et al.* shows the intercalation binding mode of [Bmim][PF$_6$] with DNA.[30] An aggregation induced DNA perturbation by [Bmim][Cl] IL was reported by Ding *et al.*[31] Later on in 2012, a groove binding mechanism of RTIL in DNA was proposed by Chandran *et al.*[32] They reported that [Bmim] cation interacts with minor groove of DNA through various non covalent interactions. Herein, for the first time, we have perceived the excellent DNA compaction property of Gua-IL. The plausible compaction mechanism has been proposed based on our results obtained from fluorescence, CD, DLS, FE-SEM, TEM and confocal

fluorescence microscopy experiments. Being a negatively charged stiff polymer, compaction of DNA needs the neutralization of its negatively charged phosphate group, and for this neutralization process counter ion valency (Z) of the compacting agent should be greater or equal to 3 ($Z \geq 3$) in accordance with Manning–Oosawa condensation theory.[47] In our case, Gua-IL possesses mono-valency in micro-molar, but in milli-molar concentration range, it acts as a multivalent ion ($Z \geq 3$) due to the formation of micellar type aggregates. All the other experimental methods employed here (fluorescence displacement assay experiments, circular dichroism) also established the fact that above a certain concentration (> 1 mM) of Gua-IL all the spectral changes take place. Interestingly, we have noticed that 1 mM is the CAC of Gua-IL from DLS study. Therefore, we strongly believe that micellar like aggregates of Gua-IL is taking a major role for the compaction process of ct-DNA. Notably, guanidium cations have two positively charged centres due to its intramolecular resonance. One positively charged part neutralizes the negatively charged phosphate (PO_4^{3-}) group of DNA and another part involves in electrostatic interaction with its counter anion. When these anions form a micellar kind of aggregate after CAC, then the counter positively charged ions situated at micellar surface drag the negatively charged DNA strands to the micellar surface due to electrostatic attraction between opposite ion pairs (Scheme 2.2). This process leads to the wrapping of several parts of DNA around these micellar kinds of aggregates. With the increase in Gua-IL concentration, the number of loops increases and finally results in the DNA globular structure by collapsing the intra/interstrand repulsion between DNA strands and instigates the structural transition from coil to globule state.

It is also well-known fact that ionic liquids can generate certain extent of osmotic stress in solution, which is achieved due to the reduction in the dielectric permittivity of the solution.[22, 48] Here it is pertinent to mention that Mel'nikov *et al.* suggested an enhancement in electrostatic attraction between DNA and counter ion due to the lowering of dielectric permittivity.[22] Similarly, here the process of electrostatic interactions between negatively charged phosphate groups of DNA and polyvalent ion (due to the formation of micellar aggregation by the anionic counterpart of Gua-IL) become even more facile due to the diminution in dielectric permittivity of the solution, and thus attraction between DNA strands increases. Therefore, we believe that in addition to the polyvalent neutralisation by micellar arrangements (described previously); this osmotic stress might also be responsible for this compaction phenomenon.

2.5 Conclusion

Current work deals with two important findings of the interaction between DNA and green solvent, RTIL. Firstly, the mode of interaction between DNA and IL and secondly, an excellent DNA compacting ability of RTIL have been revealed. To the best of our knowledge, it is the first ever report of DNA compaction mediated by RTIL, namely, guanidinium tris(pentafluoroethyl)trifluorophosphate (Gua-IL). Dye displacement study through fluorescence and CD spectroscopy imply that Gua-IL has a

reasonable impact on both intercalation position and minor groove of ct-DNA, which is a characteristic feature of DNA compaction. Further, DLS study manifests the specific role of anionic part of Gua-IL in the compaction process and thereby helps us to propose a plausible mechanism of the compaction process. UV melting study shows a decrement in melting temperature of DNA with increasing extent of compaction. Interestingly, we could probe the compaction process through field emission scanning electron microscopic (FE-SEM) and confocal fluorescence microscopic studies, which validate and provide direct proof of this compaction process instigated by Gua-IL. Based on all the spectroscopic and microscopic evidences, we propose that the counter positively charged guanidium cations situated at the surface of micellar like aggregates (formed by the anionic part of Gua-IL after CAC) drags the negatively charged DNA strands to the micellar surface due to the electrostatic attraction as well as diminution in dielectric permittivity of the solution. Eventually, the above-mentioned effect reduces the intrastrand and/or interstrand DNA repulsion and triggers the structural transition from coil to globular state. The important essence of this work is that a new class of DNA compaction agent, i.e. room temperature ionic liquid (RTIL) has been identified and is found to be very effective both for shorter (dodecamer DNA), larger (ct-DNA) DNA and plasmid DNA.
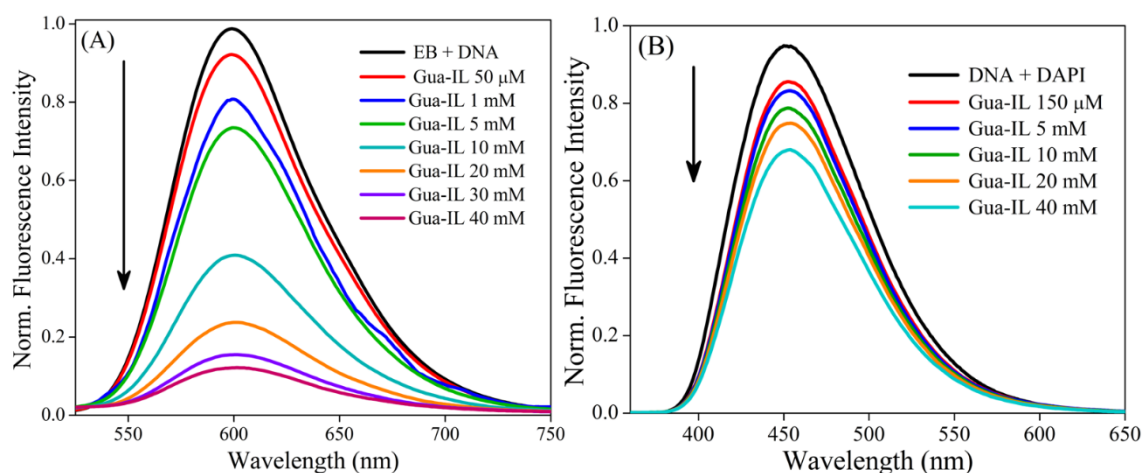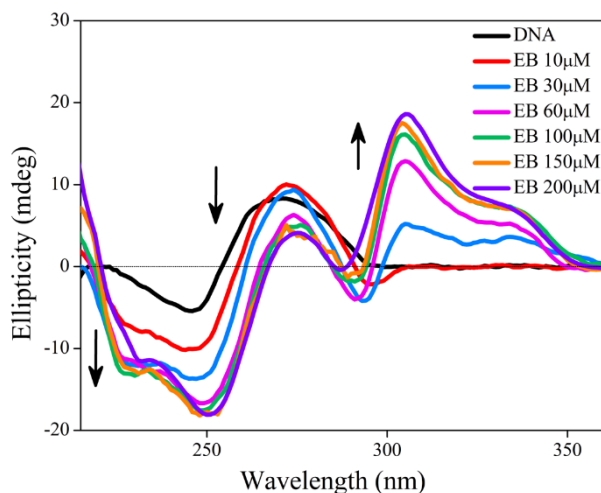
2.6 Appendix Section
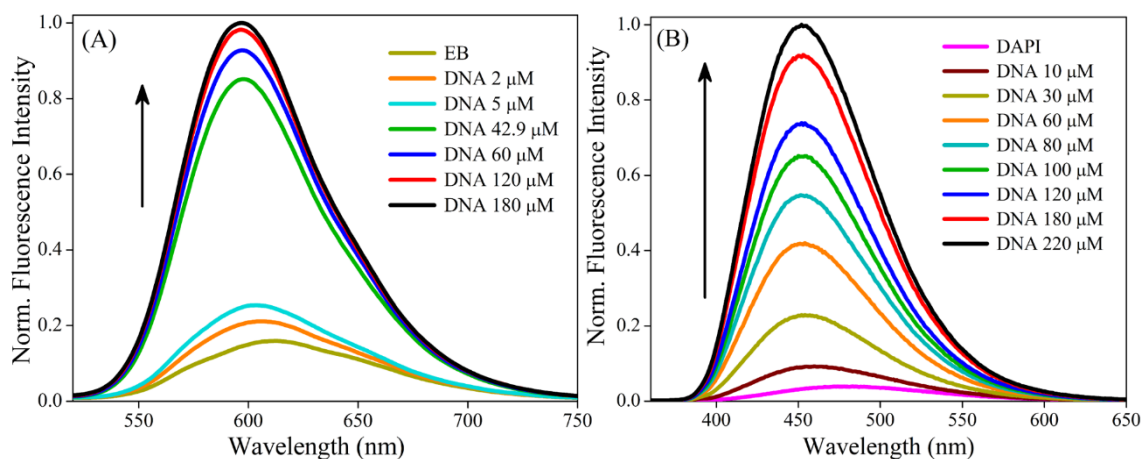
Appendix 2.1 Fluorescence spectra of dyes with increasing concentration of ct-DNA (A) EB (B) DAPI.
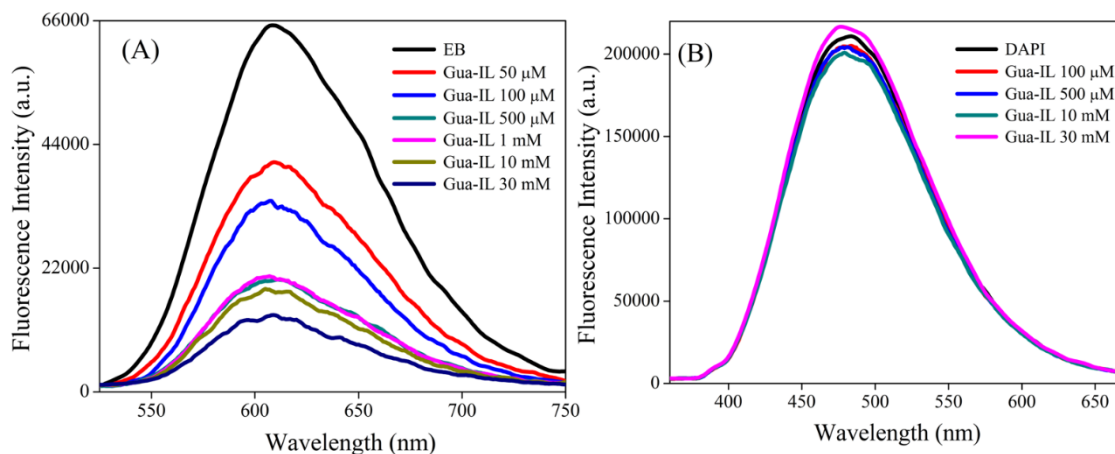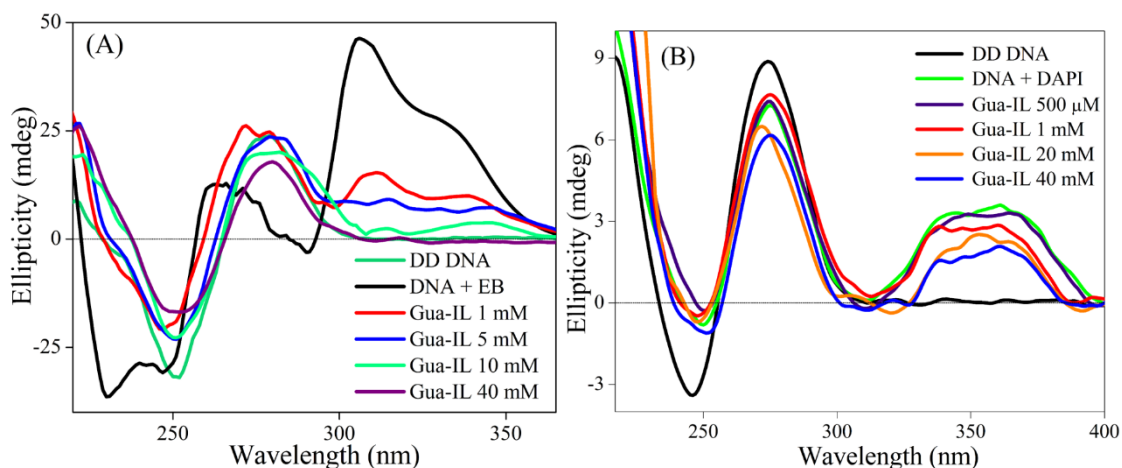

Appendix 2.2 Fluorescence spectra of dyes with increasing concentration of Gua-IL (A) EB (B) DAPI.
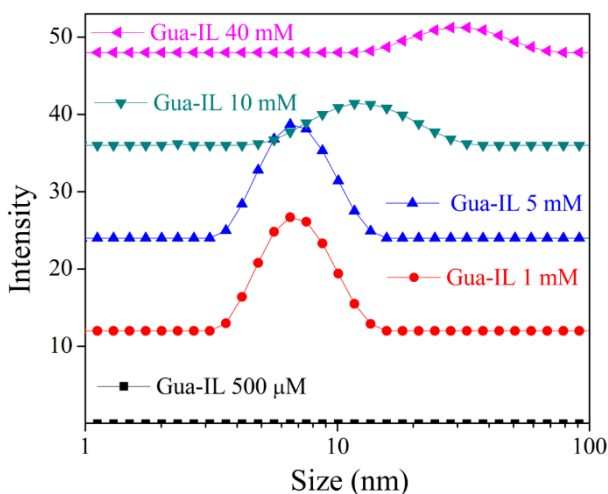

Appendix 2.3 Fluorescence spectra of DD-DNA bound dyes with increasing concentration of Gua-IL (A) EB (B) DAPI.

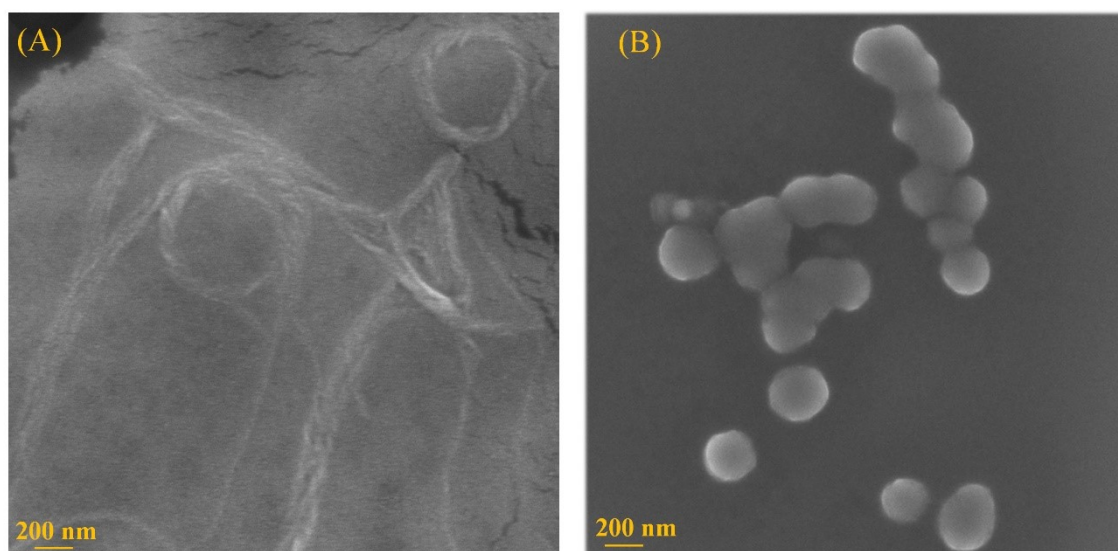**Appendix 2.4** Circular dichroism spectra of ct-DNA with the gradual addition of EB.

Appendix 2.5 Circular dichroism spectra of DD-DNA bound dye with the gradual addition of Gua-IL (A) EB (B) DAPI.

Appendix 2.6 Intensity-weighted distribution peak (measured by DLS) with increasing concentration of Gua-IL.

Appendix 2.7 FE-SEM images of (A) ct-DNA (B) Gua-IL.

2.7 Experimental

All the buffers and samples were prepared using millipore water. The purity of ct-DNA was checked by taking the ratio of the absorbance at 260 nm to that at 280 nm, which was found to be 1.80. This indicates the absence of any kind protein in ct-DNA. Concentrations of EB and DAPI were determined using molar extinction coefficients of 5,450 M^{-1} cm^{-1} at 480 nm[49] and 27,000 M^{-1} cm^{-1} at 342 nm,[50] respectively. The concentration of ct-DNA and plasmid DNA were measured using the molar extinction coefficient 6600 M^{-1} cm^{-1} at 260 nm per base pair. DNA was annealed by heating DNA samples at 90°C for about 5 minutes and then gradually cooled down to room temperature. All the ct-DNA samples were prepared either using 5 mM phosphate buffer (pH=7.4) or 5 mM Tris-HCl buffer (pH=7.4).

All the steady-state studies were carried out by adding DNA in respective dyes, and it is followed by Gua-IL addition. CD studies were monitored by the addition of respective dyes in ct-DNA which is followed by Gua-IL addition. For thermal melting, we used a very small concentration of DNA (~ 1 µM) due to avoid the absorbance saturation. The melting temperature (Tm) was determined from the sigmoidal curve fit of the melting profile. For the field emission scanning electron microscope (FE-SEM) images of DNA and DNA-Gua-IL systems, all the samples were prepared by drop casting the solution in silicon wafers. For high-resolution transmission electron microscope (HR-TEM) images, the samples for TEM images were drop cast on a Cu grid and left for drying for 12 hours. For fluorescence microscopy images, plasmid DNA (0.5 µM) and DAPI (0.5 µM) were mixed in 10 mM Tris HCl buffer (pH 7.6). Then Gua-IL (500 µM) added in the solution mentioned above. Samples were kept for 20-30 min before imaging. All the glass microscope slides and coverslips were immersed into ethanol for an overnight before the experiment. All the experimental measurements were performed at ~25 °C.

2.8 References

1. G. Felsenfeld, *Nature*, 1978, **271**, 115-122.

2. J. Yang, P. Zhang, L. Tang, P. Sun, W. Liu, P. Sun, A. Zuo and D. Liang, *Biomaterials*, 2010, **31**, 144-155.

3. G. Iacomino, G. Picariello, I. Stillitano and L. D'Agostino, *Int. J. Biochem. Cell B.*, 2014, **47**, 11-19.

4. A. A. Zinchenko, K. Yoshikawa and D. Baigl, *Adv. Mater.*, 2005, **17**, 2820-2823.

5. L. Cinque, Y. Ghomchi, Y. Chen, A. Bensimon and D. Baigl, *ChemBioChem*, 2010, **11**, 340-343.

6. (a) L. C. Gosule and J. A. Schellman, *Nature*, 1976, **259**, 333-335; (b) Y. Yoshikawa, N. Umezawa, Y. Imamura, T. Kanbe, N. Kato, K. Yoshikawa, T. Imanaka and T. Higuchi, *Angew. Chem., Int. Ed.*, 2013, **52**, 3712-3716.

7. (a) R. Dias, S. Mel'nikov, B. Lindman and M. G. Miguel, *Langmuir*, 2000, **16**, 9577-9583; (b) E. Grueso, C. Cerrillos, J. Hidalgo and P. Lopez-Cornejo, *Langmuir*, 2012, **28**, 10968-10979.

8. E. A. Murphy, A. J. Waring, S. M. Haynes and K. J. Longmuir, *Nucleic Acids Res.*, 2000, **28**, 2986-2992.

9. (a) M. Ganguli, J. V. Babu and S. Maiti, *Langmuir*, 2004, **20**, 5165-5170; (b) A. Zinchenko, K. Tsumoto, S. Murata and K. Yoshikawa, *J. Phys. Chem. B*, 2014, **118**, 1256-1262.

10. A. Estevez-Torres and D. Baigl, *Soft Matter*, 2011, **7**, 6746-6756.

11. K. Yoshikawa, Y. Yoshikawa, Y. Koyama and T. Kanbe, *J. Am. Chem. Soc.*, 1997, **119**, 6473-6477.

12. (a) K. Fant, E. K. Esbjörner, P. Lincoln and B. Nordén, *Biochemistry*, 2008, **47**, 1732-1740; (b) C.-J. Su, C.-Y. Chen, M.-C. Lin, H.-L. Chen, H. Iwase, S. Koizumi and T. Hashimoto, *Macromolecules*, 2012, **45**, 5208-5217.

13. A. A. Zinchenko, V. G. Sergeyev, K. Yamabe, S. Murata and K. Yoshikawa, *ChemBioChem*, 2004, **5**, 360-368.

14. T. Kikuchi, S. Sato, D. Fujita and M. Fujita, *Chem. Sci.*, 2014, **5**, 3257-3260.

15. V. Burckbuchler, V. Wintgens, S. Lecomte, A. Percot, C. Leborgne, O. Danos, A. Kichler and C. Amiel, *Biopolymers*, 2006, **81**, 360-370.

16. Y. Fang and J. H. Hoh, *J. Am. Chem. Soc.*, 1998, **120**, 8903-8909.

17. A. Soufi, A. Sawasdichai, A. Shukla, P. Noy, T. Dafforn, C. Smith, P.-S. Jayaraman and K. Gaston, *Nucleic Acids Res.*, 2010, **38**, 7513-7525.

18. C. W. Tabor and H. Tabor, *Annu. Rev. Biochem.*, 1984, **53**, 749-790.

19. K. Hamana, H. Hamana, M. Niitsu, K. Samejima, T. Sakane and A. Yokota, *Microbios*, 1994, **79**, 109-119.

20. P. Carrivain, A. Cournac, C. Lavelle, A. Lesne, J. Mozziconacci, F. Paillusson, L. Signon, J.-M. Victor and M. Barbi, *Soft Matter*, 2012, **8**, 9285-9301.

21. S. Rudiuk, K. Yoshikawa and D. Baigl, *Soft Matter*, 2011, **7**, 5854-5860.

22. S. M. Mel'nikov, M. O. Khan, B. Lindman and B. Jönsson, *J. Am. Chem. Soc.*, 1999, **121**, 1130-1136.

23. K. Fant, E. K. Esbjörner, A. Jenkins, M. C. Grossel, P. Lincoln and B. Nordén, *Mol. Pharm.*, 2010, **7**, 1734-1746.

24. M. Ito, A. Sakakura, N. Miyazawa, S. Murata and K. Yoshikawa, *J. Am. Chem. Soc.*, 2003, **125**, 12714-12715.

25. A.-L. M. Le Ny and C. T. Lee, *J. Am. Chem. Soc.*, 2006, **128**, 6400-6408.

26. W. Sundermeyer, *Angew. Chem., Int. Ed.*, 1965, **4**, 222-238.

27. (a) F. van Rantwijk and R. A. Sheldon, *Chem. Rev.*, 2007, **107**, 2757-2785; (b) T. L. Greaves and C. J. Drummond, *Chem. Rev.*, 2007, **108**, 206-237.

28. N. V. Plechkova and K. R. Seddon, *Chemical Society Reviews*, 2008, **37**, 123-150.

29. (a) R. Vijayaraghavan, A. Izgorodin, V. Ganesh, M. Surianarayanan and D. R. MacFarlane, *Angew. Chem., Int. Ed.*, 2010, **49**, 1631-1633; (b) A. M. Figueiredo, J. Sardinha, G. R. Moore and E. J. Cabrita, *Phys. Chem. Chem. Phys.*, 2013, **15**, 19632-19643.

30. D.-H. Cheng, X.-W. Chen, J.-H. Wang and Z.-L. Fang, *Chem. Eur. J.*, 2007, **13**, 4833-4839.

31. Y. Ding, L. Zhang, J. Xie and R. Guo, *J. Phys. Chem. B*, 2010, **114**, 2033-2043.

32. A. Chandran, D. Ghoshdastidar and S. Senapati, *J. Am. Chem. Soc.*, 2012, **134**, 20330-20339.

33. T. Singh and A. Kumar, *J. Phys. Chem. B*, 2007, **111**, 7843-7851.

34. J. B. Lepecq and C. Paoletti, *J. Mol. Biol.*, 1967, **27**, 87-106.

35. J. Kahlen, K. Masuch and K. Leonhard, *Green Chem.*, 2010, **12**, 2172-2181.

36. S. K. Pal, D. Mandal and K. Bhattacharyya, *J. Phys. Chem. B*, 1998, **102**, 11017-11023.

37. J. R. Lackowicz, *Principles of Fluorescence Spectroscopy*, Springer, New York, 3rd ed. edn., 2006.

38. (a) R. E. Williams and S. L. Kielland, *Can. J. Chemistry*, 1975, **53**, 542-548; (b) Y. M. Evdokimov, T. L. Pyatigorskaya, O. F. Polyvtsev, N. M. Akimenko, V. A. Kadykov, D. Y. Tsvankin and Y. M. Varshavsky, *Nucleic Acids Res.*, 1976, **3**, 2353-2366.

39. (a) D. G. Dalglkish, A. R. Peacocke, G. Fey and C. Harvey, *Biopolymers*, 1971, **10**, 1853-1863; (b) S. Aktipis and A. Kindelis, *Biochemistry*, 1973, **12**, 1213-1221.

40. G. Manzini, M. L. Barcellona, M. Avitabile and F. Quadrifoglio, *Nucleic Acids Res.*, 1983, **11**, 8861-8876.

41. R. S. Dias, J. Innerlohinger, O. Glatter, M. G. Miguel and B. Lindman, *J. Phys. Chem. B*, 2005, **109**, 10458-10463.

42. R. F. Probstein, *Physicochemical hydrodynamics*, John Wiley & Sons, New York, 2nd edn., 1994.

43. (a) B. Neto and A. Lapis, *Molecules*, 2009, **14**, 1725-1746; (b) J. Ren and J. B. Chaires, *Biochemistry*, 1999, **38**, 16067-16075.

44. S. Yu, M.-H. Li, S. Choi, J. Baker and R. Larson, *Molecules*, 2013, **18**, 10707-10720.

45. M. F. V. Pinto, M. C. Morán, M. G. Miguel, B. Lindman, A. S. Jurado and A. A. C. C. Pais, *Biomacromolecules*, 2009, **10**, 1319-1323.

46. S. Lin, Y. Wang, C. Cai, Y. Xing, J. Lin, T. Chen and X. He, *Nanotechnology*, 2013, **24**, 085602.

47. (a) G. S. Manning, *J. Chem. Phys.*, 1969, **51**, 924-933; (b) F. Oosawa, *In Polyelectrolytes*, New York, 1971.

48. T. Dvir, L. Fink, R. Asor, Y. Schilt, A. Steinar and U. Raviv, *Soft Matter*, 2013, **9**, 10640-10649.

49. F. M. Pohl, T. M. Jovin, W. Baehr and J. J. Holbrook, *Proc. Natl. Acad. Sci. USA*, 1972, **69**, 3805-3809.

50. S. Eriksson, S. K. Kim, M. Kubista and B. Norden, *Biochemistry*, 1993, **32**, 2987-2998.

Chapter 3A
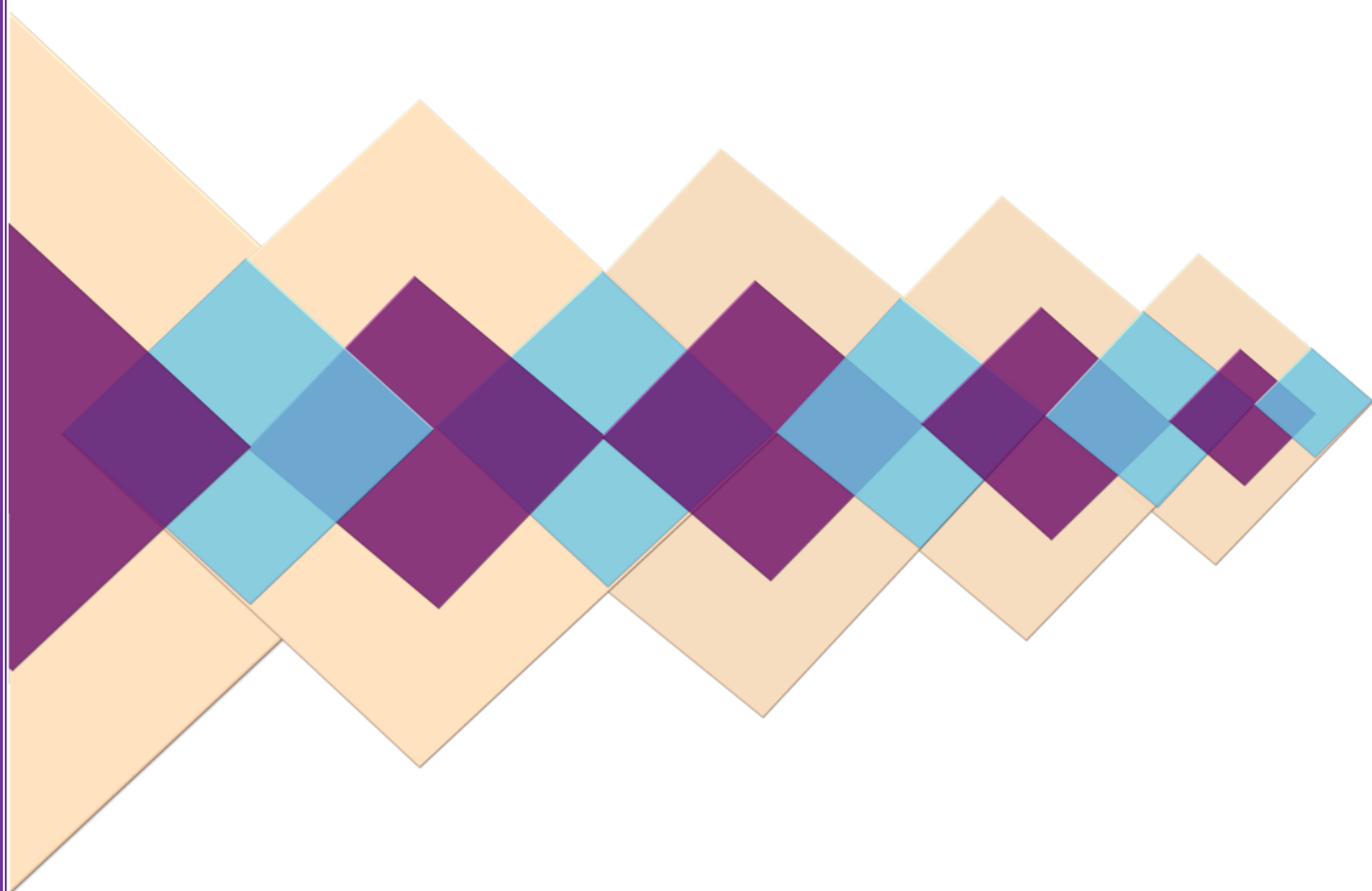
Ionic Liquid Induced G-quadruplex Formation and Stabilisation: Spectroscopic and Simulation Studies

3A.1 Introduction

Gellert, et al. in 1962 first described that four guanines can self-assemble to a guanine or G-tetrad through the formation of eight Hoogsteen hydrogen bonds.[1] In later stages, repetitive G-rich sequences had been identified at the ends of chromosomes (i.e. in telomere) which ignited a great amount of interest for G-quadruplex (GQ) structures.[2] Being a non-coding part of chromosome, telomere replication is majorly governed by the telomerase enzyme. Interestingly, length of human telomeres shortens with each cell division in normal somatic cells which eventually limit the proliferative potential of cells. Hence, the structure and stability of telomere is closely associated with the genetic stability and cell aging.[3] In this context, Kim, et al. reported high telomerase activity of cancerous cells in comparison with normal somatic cells.[4] This led to a flurry of investigation for inhibition of telomerase enzyme activity in cancerous cells by means of altering the substrate (i.e. telomere) to the GQ structure.[5] Consequently, research on the stabilization of telomeric GQ by small molecules is well appreciated and continues to be an active field for anti-cancer drug discovery.[6]

Typically, human telomeric DNA forms an antiparallel GQ structure in presence of Na^+ cation, while a mix (3+1) quadruplex structure in K^+ ion solution.[7] Apart from Na^+ and K^+ ions, a few other ions like NH_4^+, Li^+, and Sr^{2+} can also form GQ structure.[8] Not only metal ions, but also many small molecules and drugs were found to stabilize the GQ structure.[6a, 6c, 9] However, these small molecules come with serious issue of selectivity between duplex (i.e. B-DNA) and quadruplex DNA. Most of these small molecules are prone to intercalate between DNA base pairs (or to bind in the grooves) regardless of the cell being cancerous or healthy and eventually cause side-effects.[6c, 9] Keeping these issues in mind, a few small molecules and ions like Na^+, K^+, Li^+, and NH_4^+ etc. are the best possible GQ stabilizing materials till date due to their selectivity for GQ over the duplex structure.[6a, 10] Hence, a lot of research is underway to develop new GQ stabilizing materials without affecting the duplex form. In this context, we recently reported that the ionic liquid (IL), guanidinium tris(pentafluoroethyl)trifluorophosphate (Gua-IL) is inactive towards duplex structure of DNA in lower concentration (< 1 mM). However, at higher concentration (> 1 mM), it compacts the B form of DNA keeping its secondary structure same (Scheme 3A.1).[11] Keeping this in mind, we have tried to stabilize the GQ structure by this special class of ionic liquid which is ineffective towards the duplex structure at lower concentrations. Interestingly, we have found out that Gua-IL (< 1 mM) can not only instigate the formation of GQ in absence of conventional GQ forming ions (like Na^+, K^+, NH_4^+ etc.), but also stabilizes the GQ structure. Previously, many research groups have explored the interaction of ionic liquid with different bio-macromolecules considering its low cytotoxicity and high biodegradability in mind.[12] In continuation to that, ionic liquids have been utilized as a choice of solvents for DNA storage and extraction purpose.[13] Fujita et al. demonstrated that a combination of cholinium cations and dihydrogen phosphate anions in the hydrated condition (resembling ionic liquid like property) can form

GQ structure.[14] Apart from this report, literature about GQ stabilization by IL is not present to the best of our knowledge. This has motivated us to investigate the interaction of IL with GQ structures. Thus, we have explored that an ionic liquid (IL) i.e. Gua-IL (< 1 mM) (Scheme 3A.1) can instigate and stabilize the GQ structure in absence of any ion. This conformational transition has been confirmed through different spectroscopic tools like circular dichroism (CD), thermal melting, steady state and time resolved fluorescence studies. To elucidate the mechanism for GQ formation, we have also carried out all-atom molecular dynamics simulations. We have studied the stability of [3+1] hybrid GQ structure (PDB ID: 2GKU[15]) in Gua-IL as well as in presence of another ionic liquid, 1-Butylpyridinium chloride (BPyCl) and another salt, guanidine hydrochloride (GuaHCl) (Scheme 3A.1). These simulations show that for Gua-IL, one of the guanidinium cations binds and stays in the central core of G-tetrad whereas the bulky cation-anion pairs prefer to reside near the GQ surface. This majorly accounts for the GQ formation and stabilization by Gua-IL. The present study indicates that this special class of ionic liquid (Gua-IL) can act as a stabilizer of telomeric GQ-DNA exerting minimal effect to duplex structure. Moreover, based on our findings here, we have argued about the design of IL that can stabilize the GQ structure.
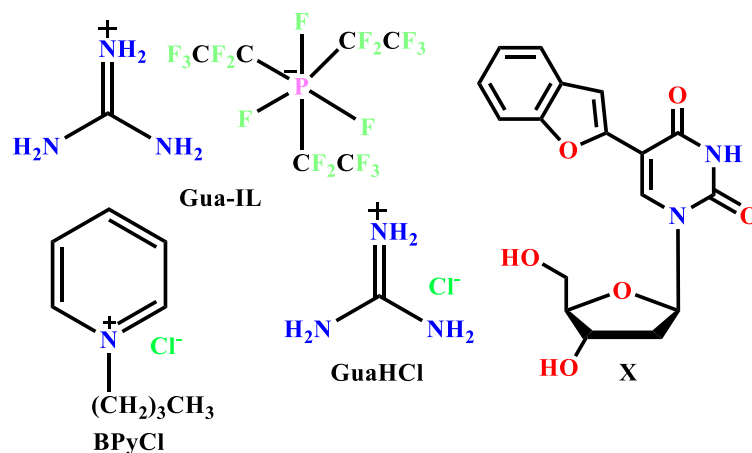

Scheme 3A.1 Chemical structures of guanidinium tris(pentafluoroethyl)trifluorophosphate (Gua-IL), 1-Butylpyridinium chloride (BPyCl), guanidine hydrochloride (GuaHCl) and fluorescent nucleobase, 5-(benzofuran-2-yl)-2′-deoxyuridine (X) incorporated in modified G-quadruplex sequence.

3A.2 Results and Discussions

Circular dichroism (CD) is the most commonly used and reliable spectroscopic technique to probe the secondary structure of DNA. Different topologies of G-quadruplex exhibit distinctly different CD spectral features. Typically, an antiparallel G-quadruplex formed by Na^+ ion (such as basket and chair forms) shows an intense negative band at 265 nm with two positive CD bands around 295 and 245 nm, respectively. The parallel structure, on the other hand, accompanies a positive band at 265 nm and a negative band at 240 nm.[16] Here also, H24 (5′-

d(TTAGGGTTAGGG-TTAGGGTTAGGG)-3′) in Na$^+$ ion exhibits a basket type antiparallel structure showing a similar spectral pattern (Figure 3A.1). In K$^+$ ion, however, H24 forms a
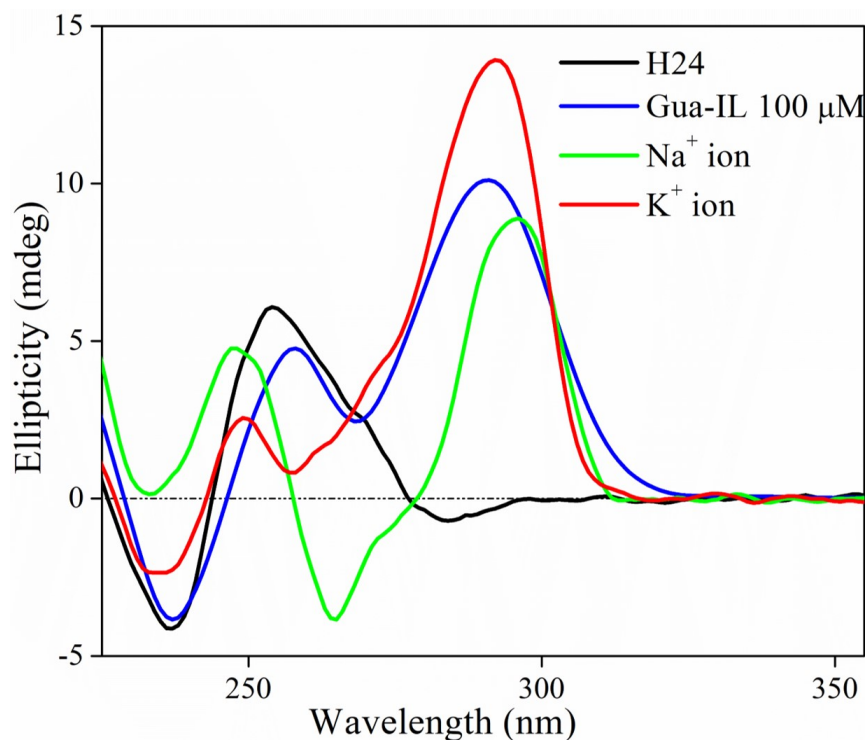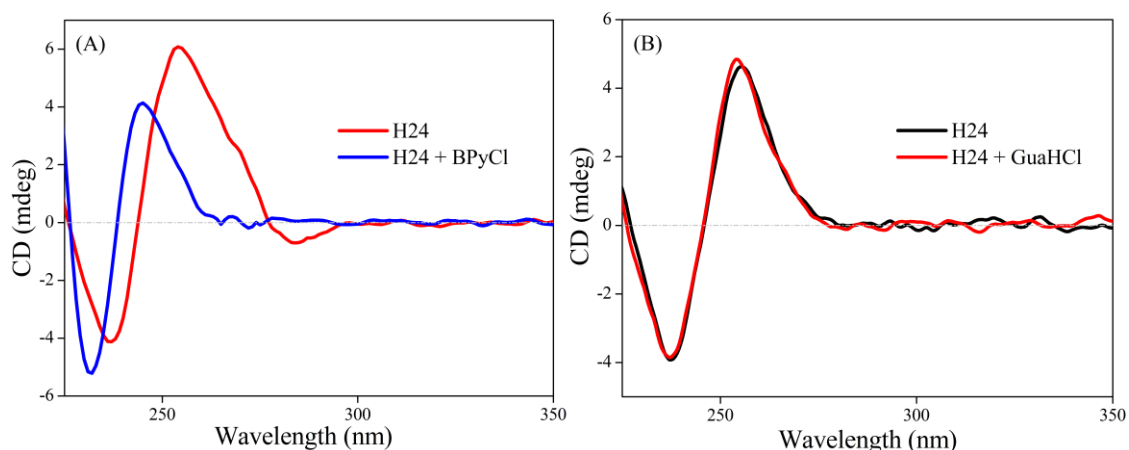

Figure 3A.1 Circular Dichroism spectra of H24 DNA (~ 5 μM) in different conditions. H24 legend in figure corresponds to the CD spectra of H24 DNA in absence of any ion i.e. in deionised water. K$^+$ ion and Na$^+$ ion legends indicate the presence of 100 mM KCl and NaCl salt respectively; in 10 mM tris buffer solution (pH 7.2).

mix type of (3+1) hybrid structure showing a large positive peak around 295 nm (corresponding to the antiparallel form), a small positive peak around 250 nm (Figure 3A.1), and a negative peak around 235 nm (corresponding to the parallel form). These spectral observations for different topologies are in agreement with the earlier reported literatures.[9a, 16] In presence of Gua-IL, H24 depicts a CD profile with positive peaks around 290 nm and 260 nm and a negative peak around 235 nm (Figure 3A.1) inferring the formation of GQ structure. Interestingly, the CD profile in presence of Gua-IL closely resembles with a mix type (3+1) hybrid structure in K$^+$ ion solution (Figure 3A.1). Hence, the CD results confirm that Gua-IL can induce the GQ formation having hybrid topology in absence of any ions. Moreover, a closer look in the Gua-IL CD spectra (Figure 3A.1) reveals that positive peak around 260 nm and negative peak around 240 nm is more intense than that of in K$^+$ ion (i.e. for the hybrid structure). Similar type of ellipticity pattern (i.e. increment around 260 nm and 240 nm peak) has also been observed as an outcome of molecular crowding effect in GQ structures.[17] Likewise, ionic liquids can also generate a certain extent of molecular crowding effect by lowering the dielectric permittivity of the solution. Thus, crowding effect of Gua-IL is believed to be responsible for slight variation in CD profile in Gua-IL environment compared to the standard K$^+$ ion environment. To check the impact of this crowding effect on G-quadruplex formation, we have performed similar

experiment with H24 and another ionic liquid, i.e. BPyCl in the absence of any ion. Results show that BPyCl is unable to induce any kind of GQ formation (Figure 3A.2 and Appendix 3A.1), ruling out the contribution of crowding effect as a major factor towards the GQ formation. This observation also implies that the formation of GQ depends on the particular structure of IL and not on its general property. Ohno and co-workers reported that hydrated combination of cholinium cations and dihydrogen phosphate is mainly responsible for GQ formation instead of salts containing either cholinium cations or dihydrogen phosphate anions.[14]

Figure 3A.2 Circular dichroism spectra of H24 DNA (~ 5 µM) in deionized water in presnce and absence of (A) BPyCl (1 mM) and (B) GuaHCl (5 mM).

To test the above idea, we have performed MD simulations of GQ structure with Gua-IL and BPyCl keeping the ratio of the number of ionic liquid pairs and the water molecules same as used in the experimental conditions. Figure 3A.3A shows the root mean square deviations (RMSDs) of heavy atoms of GQ, core guanine quartet structure and phosphate backbone of GQ (O5′-C5′-P-O1P-O2P-O3′-C3′) from the starting structure (t = 0 ps). All three RMSDs clearly show relatively less fluctuation for Gua-IL than that of with BPyCl. RMSD of backbone deviates continuously with time, where the quartet region suddenly distorts beyond 150 ns for BPyCl. Interestingly, we also noted that the quartet region, which accounts for the formation of some guanine quartet region flanked outside of the GQ, got highly destabilized around 150 ns for BPyCl (Figure 3A.3A and Appendix 3A.1) leading to the disintegration of GQ. This continuous fluctuation in backbone eventually leads to the disruption of the quartet region. RMSDs for Gua-IL, on the contrary, exhibit a stable structure up to 300 ns.

Along with the RMSD calculation, we have also carried out the radial distribution function (RDF) calculations of both these ILs and water molecules around surface atoms of GQ (Figure 3A.3B). These results clearly indicate the presence of Gua^+ in the first solvation shell (within 0.5 nm) of GQ. In comparison with Gua^+, the presence of BPy^+ ions around GQ is very less. Interestingly, GQ surface and its surrounding (within 0.5 nm) contains higher numbers of FEP^- ions (anion of Gua-IL) compared to chloride ions (anion of BPyCl). The RDF of water

molecules shows the first solvation peak for Gua-IL at a shorter distance than BPyCl. The Interaction energy between GQ and cations (Gua$^+$ or BPy$^+$) in the first solvation shell indicates the strong stabilization for Gua$^+$ ions and convergence of energy after 50 ns (Appendix 3A.2). Similarly, the interaction energy between GQ and water molecules within 3.1 Å distance (first solvation shell) show approximately 3 times stronger stabilization for Gua-IL than that of for BPyCl (Appendix 3A.2). This indicates that not only cation-GQ interactions but specific arrangements of water molecules are also a factor for the stabilization in case of Gua-IL. Moreover, the results imply that the GQ structure is stabilized by ion-pair rather than cation or anion alone. It is evident from the experimental and simulation results that the formation and stabilization of GQ structure by IL does not follow a general principle, rather it is a very much specific for the ion pair constituting the IL.
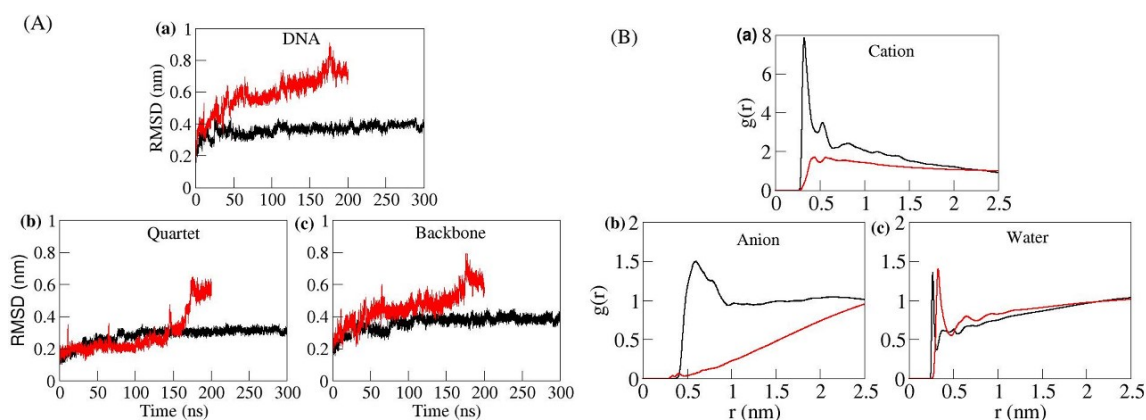

Figure 3A.3 (A) Root mean square deviations of the heavy atoms of (a) G-quadruplex, (b) quartet region of G-quadruplex and (c) backbone region of G-quadruplex in presence of Gua-IL (black line) and BPyCl (Red Line). (B) Radial distribution function (RDF) of (a) cations (Gua$^+$ or BPy$^+$) with GQ; (b) anions (FEP$^-$ or Cl$^-$) and (c) Oxygen atoms of water molecules around G-quadruplex in presence of Gua-IL (black line) and BPyCl (red line).

It is interesting to inspect the individual role of cations and anions towards the formation of GQ structure. To do so, we have performed similar kind of CD experiment with H24 and GuaHCl (having same cation as Gua-IL, but anion is Cl$^-$) in the absence of any ion (Figure 3A.2B). Even though GuaHCl is not an ionic liquid and works as a denaturant,[18] this study provides an opportunity to explore the effect of the anions on the stability of GQ structure. Absence of any characteristic's CD peak for the GQ structure confirms the inability of GuaHCl to induce any GQ structure in absence of GQ forming conventional ions like Na$^+$, K$^+$ etc. (Figure 3A.2B). Hence, this study confirms that guanidinium cation (Gua$^+$) alone is unable to form the GQ structure and anions have some important role in it. To understand the role of ions in more detail, we have plotted RDF of ions and water molecules with the surface atoms of GQ (Appendix 3A.3). The RDF of Gua$^+$ ions indicate a lesser probability of finding cations in case of GuaHCl as well as for Cl$^-$ ions in first solvation shell. This clearly indicates the lack of ion pairs for stabilizing the GQ structure, which accounts for the reduced stability of GQ in GuaHCl

solutions. The interaction energy of Gua$^+$ ions is stronger in Gua-IL than GuaHCl (Appendix 3A.4). Appendix 3A.5 also shows that the RMSD sharply increases beyond 170 ns. Hence, we assume that with further sampling one can observe the disruption of GQ. This suggests that even though GQ is not disrupted completely within 300 ns, it may eventually be destabilized. Moreover, energy analysis suggests that Gua$^+$ is comparatively less stabilized in GuaHCl than in Gua-IL (Appendix 3A.4).

To investigate about the GQ structural stabilization imposed by Gua-IL or not, we have performed the thermal melting studies of GQ in the absence and presence of Gua-IL. Interestingly, the nature of H24 melting profile changes in the presence of Gua-IL (Appendix 3A.6), which is also indicative of GQ formation by Gua-IL in absence of GQ forming conventional ions (like Na$^+$, K$^+$, NH$_4^+$ etc.).[16] Moreover, melting profile (Figure 3A.4) of GQ formed by Gua-IL (in the absence of GQ forming conventional ions) shows a gain in the stabilization by a T_m enhancement of ~10 °C compared to that of in K$^+$ ion (in K$^+$ ion, T_m ~ 64 °C), which is almost equivalent to 83.14 J/mol.
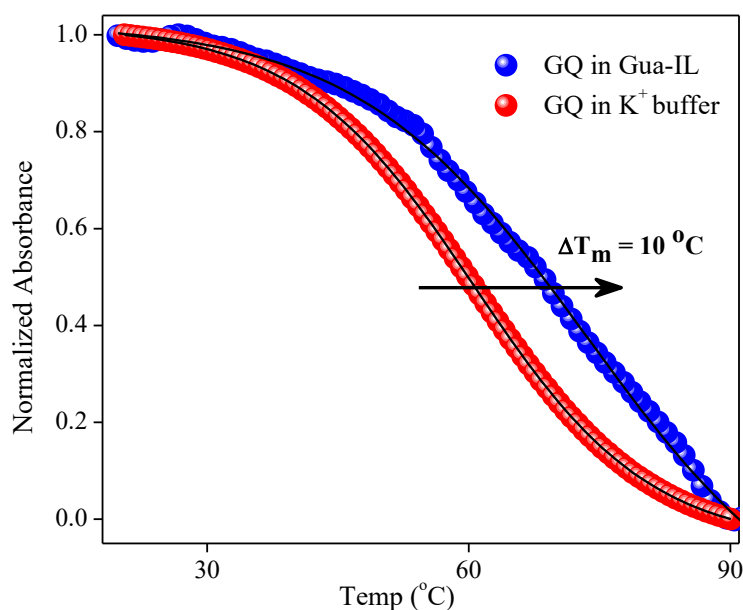
Figure 3A.4 UV melting profile of H24 DNA (~ 5 μM) in deionized water in presence of Gua-IL (100 μM) and K$^+$ion containing buffer. K$^+$ion legend indicates the presence of 100 mM KCl salt in 10 mM tris buffer (pH 7.2) solution.

Although, we have explored about the topology and probable factors responsible for GQ stability by Gua-IL through experimental and simulation studies, but still we don't have any information about the local environment (like polarity, viscosity, rigidity etc.) around the GQ structure. This is mainly due to the optical inactivity for both H24 and Gua-IL. To address this issue, we have taken a fluorescent modified quadruplex sequence, 5′-d(AGGGTTAGGGXTAGGGTTA-GGG)-3′, as a model sequence to probe the local environment around GQ structure in the presence of Gua-IL, where 'X' denotes the fluorescent

modified nucleobase (Scheme 3A.1). We have checked the GQ formation of modified sequence by Gua-IL through CD studies (Appendix 3A.7), which indicates the formation of a mix type (3+1) hybrid structure as found in case of H24 without modification. Here, it is pertinent to mention that the fluorescent moiety in the modified GQ sequence, 5-(benzofuran-2-yl)-2′-deoxyuridine (X) is very much sensitive towards the polarity of the medium and the surrounding rigidity around the probe.[19] Tor and coworkers reported that the emission intensity and lifetime of this probe (X) enhances significantly upon increasing viscosity of the medium which

Table 3A.1 Fluorescent lifetime decay parameters of fluorescent modified oligo (~ 5 μM) in different conditions. K^+ ion and Na^+ ion indicate the presence of 100 mM KCl and NaCl salt respectively; in 10 mM tris buffer (pH 7.2).

| Sample	τ_1 (ps)	a_1	τ_2 (ns)	a_2	τ_3 (ns)	a_3	τ_{avg} (ns)	χ^2
Mod Oligo + Gua-IL (100 μM)	384	0.57	1.81	0.32	5.15	0.11	1.37	1.10
Mod Oligo + K^+ ion	240	0.67	1.45	0.25	5.88	0.08	0.98	1.11
Mod Oligo + Na^+ ion	500	0.29	2.82	0.52	5.57	0.19	2.68	1.02
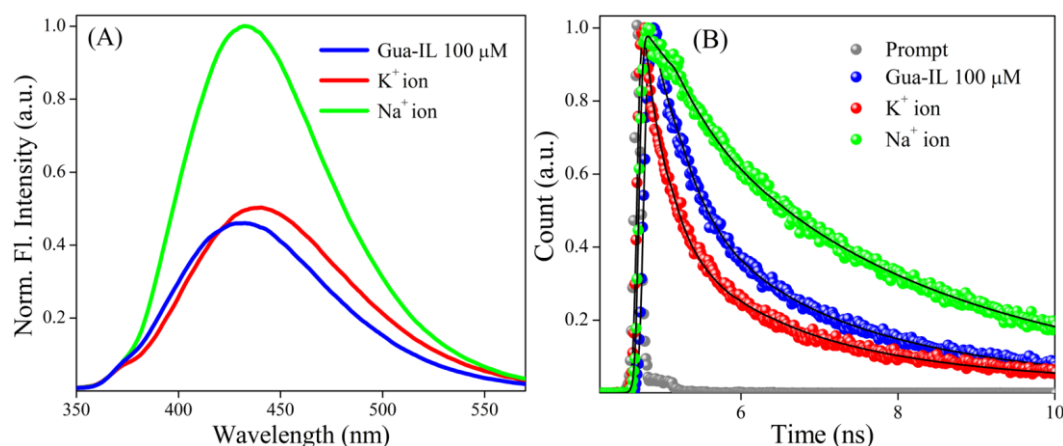
Figure 3A.5 (A) Steady state fluorescence and (B) time resolved fluorescence decay spectra of fluorescent modified quadruplex (~ 5 μM) in different conditions. K^+ ion and Na^+ ion legends indicate the presence of 100 mM KCl and NaCl salt respectively; in 10 mM tris buffer solution (pH 7.2).

accounts for the decrement in non-radiative pathway due to the rotational restriction around the aryl-aryl bond in the modified probe. Fluorescence and lifetime profiles of modified GQ sequence in different conditions have been given in Figure 3A.5A and 3A.5B, respectively. The fluorescence spectrum (excited at 330 nm) of modified GQ (Figure 3A.5A) in presence of different ions exhibits an emission maximum around 430 nm which is little blue shifted as compared to that of free nucleoside i.e. 5-(benzofuran-2-yl)-2′-deoxyuridine in water (emission maxima at 446 nm). Blue shift in the emission maximum implies a less solvent exposed environment around the nucleoside for GQ structure in comparison with the free nucleoside in

water. However, similar emission maximum position (around 430 nm) for GQ in presence of Gua-IL and different ions (i.e. Na^+, K^+) indicate that almost similar polarity sensed by the fluorescent nucleoside for all the cases. Thus, instead of polarity, the rigidity around the nucleoside majorly governs the emission intensity and lifetime of the fluorescent labeled GQ. Now in this complex heterogeneous system, assigning of any individual component is very difficult, as the modified nucleobase faces numerous numbers of interaction with its neighboring nucleobases/loops in the same sequence and with another molecule of GQ DNA. Thus, in order to clear out any ambiguity in our interpretation, we have considered the average lifetime of each sample instead of its individual components. In case of K^+ ion and Gua-IL (Figure 3A.5B and Table 3A.1), emission intensity and average lifetime values (1.1 ns and 0.98 ns for Gua-IL and K^+ ion respectively) are quite comparable indicating a similar type of rigid environment in both the cases. However, in presence of Na^+ ion, the emission intensity and average lifetime values (2.7 ns for Na^+ ion) are distinctly different in comparison with that of in K^+ ion and Gua-IL. This fluorescence and lifetime result also reveal that quadruplex topology for K^+ ion and Gua-IL (i.e., mix (3+1) hybrid structure) is very much similar, unlike antiparallel structure in presence of Na^+ ion. In a nutshell, the steady state and lifetime results in presence of different ions corroborate well with our previous CD results (Figure 3A.1).

We have visualized the trajectory of Gua^+ entering into the GQ core in Gua-IL simulations at 300 K. These simulations have been performed multiple times with different initial velocity distributions (Appendix 3A.8), ensuring that the observations are statistically reliable. These simulations suggest that initially Gua^+ ions of Gua-IL interact strongly with GQ
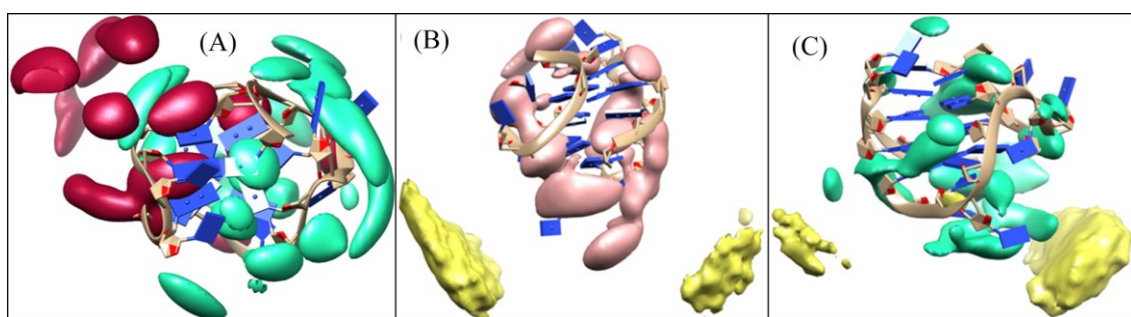
Figure 3A.6 The Figure showing spatial distribution function of cations and anions around G-quadruplex as shown in (A) side view presentation of Gua^+ (Green) and FEP^- (dark red) in case of Gua-IL; (B) BPy^+ (light pink) and Cl^- (yellow) in case of BPyCl and (C) Gua^+ (Green) and Cl^- (yellow) in case of GuaHCl.

at terminal regions and in the grooves. Subsequently, one of the Gua^+ ions enters the G-tetrad core and forms hydrogen bonds with guanine residues. These hydrogen bonds are transient. We also attribute the stability towards the strongly bound, long-lived certain Gua^+ ions in the groove of GQ. The strong interactions between Gua^+ and FEP^- (anion of Gua-IL) create a strong ion-pair distribution around GQ exerting Cage-like effect as indicated by spatial distribution of ions around GQ (Figure 3A.6), which is lacking in BPyCl and GuaHCl (Figure 3A.6).

We have plotted the distance between the central carbon atom of Gua$^+$ (CZ) and the center of mass (CoM) for Guanine quartet region to demonstrate the trajectory of Guanidinium ions (Figure 3A.7). We have considered 3 Gua$^+$ ions mainly the one which translocate in the G-quartet core (Gua-1) and two others (Gua-2 and Gua-3) that are strongly bound on the outer side of G-quartet core (Appendix 3A.9). Appendix 3A.10 shows that one of the Gua$^+$ (shown as Gua-1) is initially 2.6 nm away from the core region. It then samples the conformational space and eventually lands into the core region around 40 ns and remains strongly bound there (Figure 3A.7). Surprisingly, this process seems spontaneous in repeated simulations. We have repeated Gua-IL/GQ simulations three times, two simulations at 300 K using different initial velocities as mentioned above and a third one at 330 K. The distance between CZ and the CoM of GQ is plotted for all these three simulations (Appendix 3A.10) which shows that at each time, Gua$^+$ enters into the core of GQ.
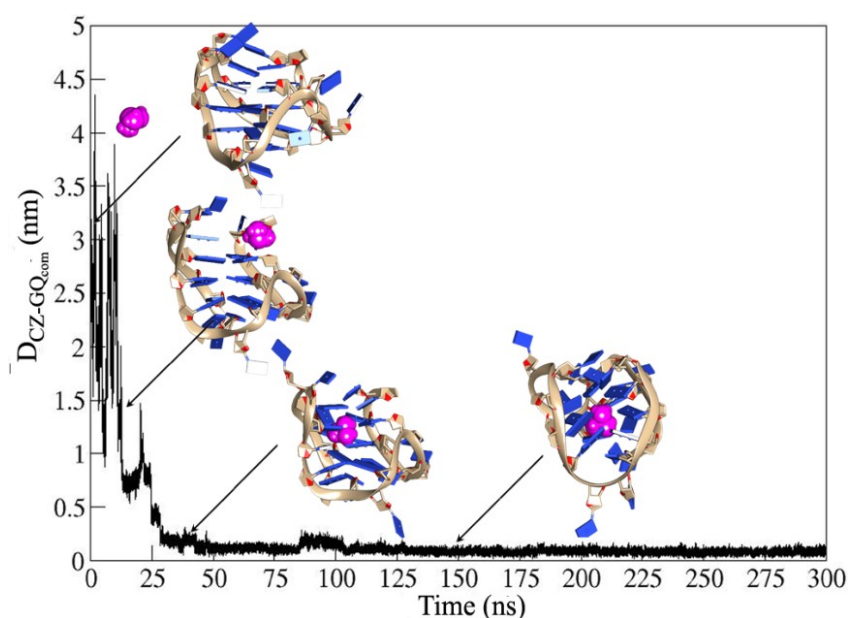
Figure 3A.7 Distance between carbon atom (CZ) of Guanidinium ion (Gua-1) and Center of the mass of G-quadruplex quartet region along time. The Guanidinium ion is shown in magenta color. The snapshots belonging to certain timeframes such as 0 ns (initial frame), 21 ns (Gua$^+$ reaches close to GQ), 30 ns (Insertion of Gua$^+$ inside GQ) and 150 ns (stabilization of GQ by single Gua$^+$ ions) are shown.

### 3A.3 Plausible Mechanism for Formation and Stabilization of GQ DNA

In our study, we notice that Gua-IL is the only IL among three (Gua-IL, BPyCl, and GuaHCl) to instigate and stabilize the GQ structure. Interestingly, Gua-IL comprises a smaller cation (Gua$^+$) and bulky anion (FEP$^-$) (Scheme 3A.1) while BPyCl contains a bigger cation (BPy$^+$) and a smaller anion (Cl$^-$) (Scheme 3A.1). GuaHCl, on the other hand, consists of same cation as that of Gua-IL, but with a smaller anion i.e. Cl$^-$ (Scheme 3A.1). Above observations suggest that a smaller cation with a bulky anion pair may have some edge over other ion pairs for GQ formation and stabilization. Based on the experimental and theoretical results, we can hypothesize that IL ion-pair constituting smaller cations can penetrate into the G-tetrad region if bulky anions support it by surface binding. In this study, the

combined effect of Gua$^+$ and FEP is majorly responsible for GQ formation and stabilization that is lacking in BPyCl and GuaHCl.

3A.4 Conclusion

Earlier reports about ionic liquid interaction with GQ have been limited to the hydrated ion-pair stabilization towards GQ structure, thus they did not provide the clear idea about the factors responsible for GQ stabilization. We could show here, using systematic experimental and theoretical studies, the possible molecular mechanism involved in the formation and stabilization of GQ in presence of IL. The present study shows that a special class of ionic liquid (Gua-IL) can act as a stabilizer of telomeric GQ-DNA exerting minimal effect towards the duplex structure.
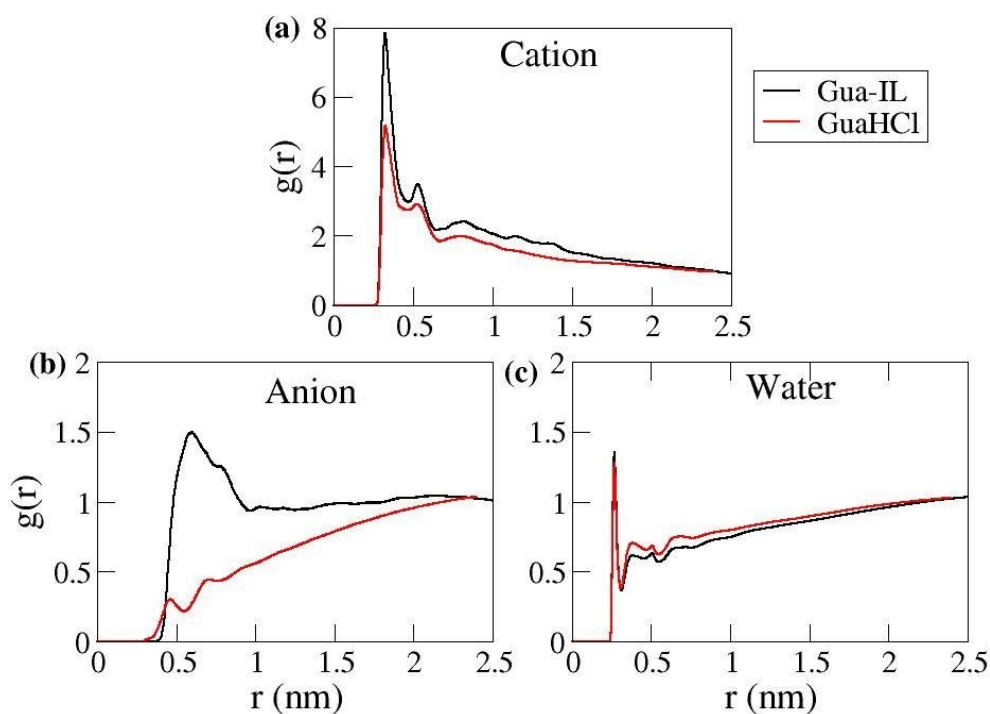
3A.5 Appendix Section
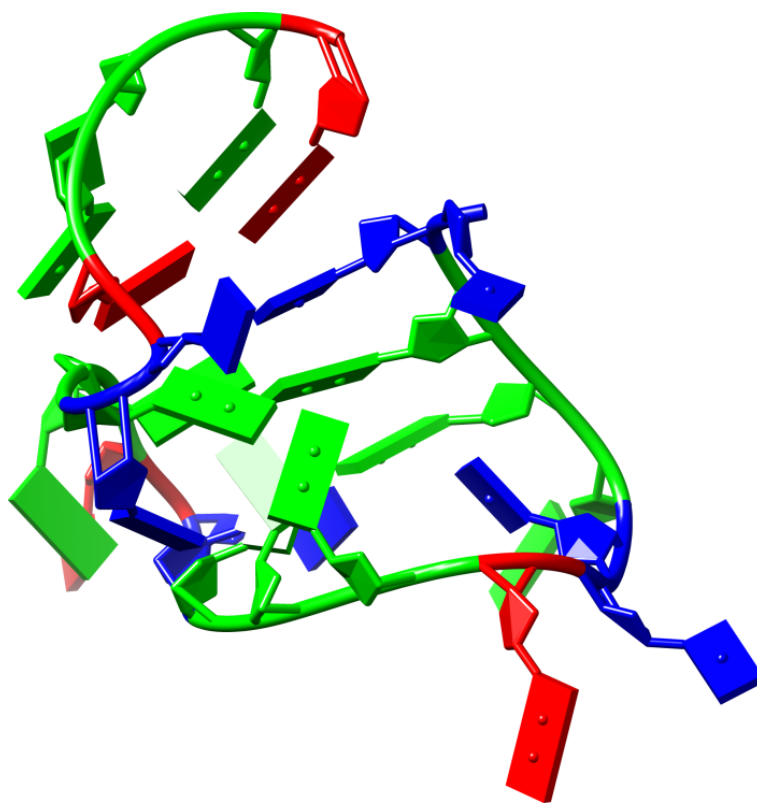
Appendix 3A.1 The disrupted structure of GQ in presence of BPyCl after 200 ns.
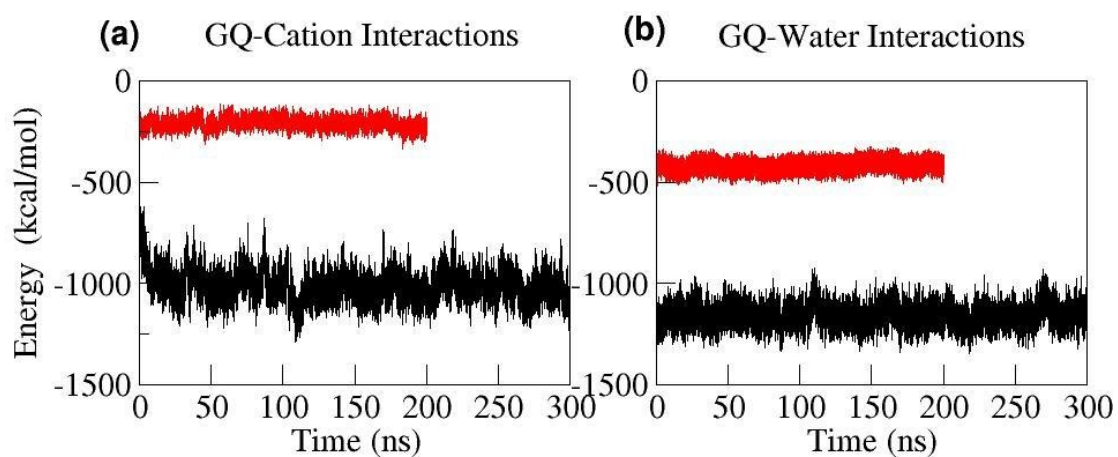

Appendix 3A.2 The figure showing interactions energy between (a) G-quadruplex and first solvation shell cation molecule and (b) G-quadruplex and first solvation shell water molecules in case of Gua-IL (black line) and BPyCl (red line).
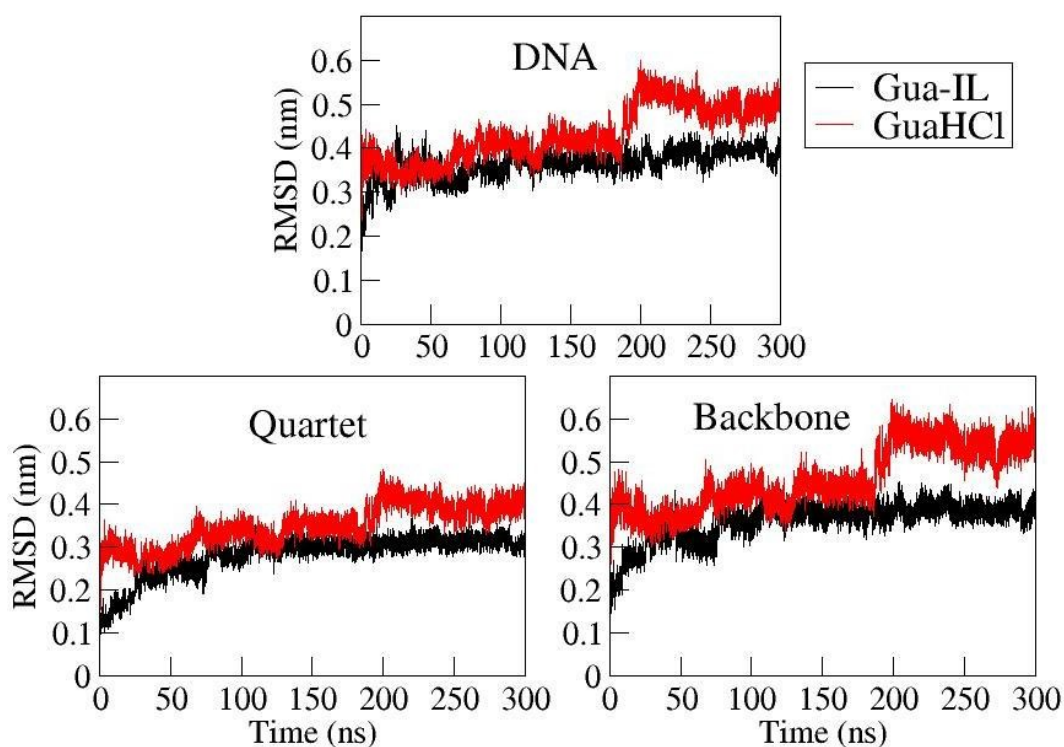
Appendix 3A.3 The RDF of various atoms around surface of G-quadruplex in case of Gua-IL (black line) and GuaHCl (red line) for (a) cations i.e. Gua$^+$ ions; (b) anions i.e. FEP$^-$ for Gua-IL and Cl$^-$ for GuaHCl and (c) water oxygen (OW) atoms.
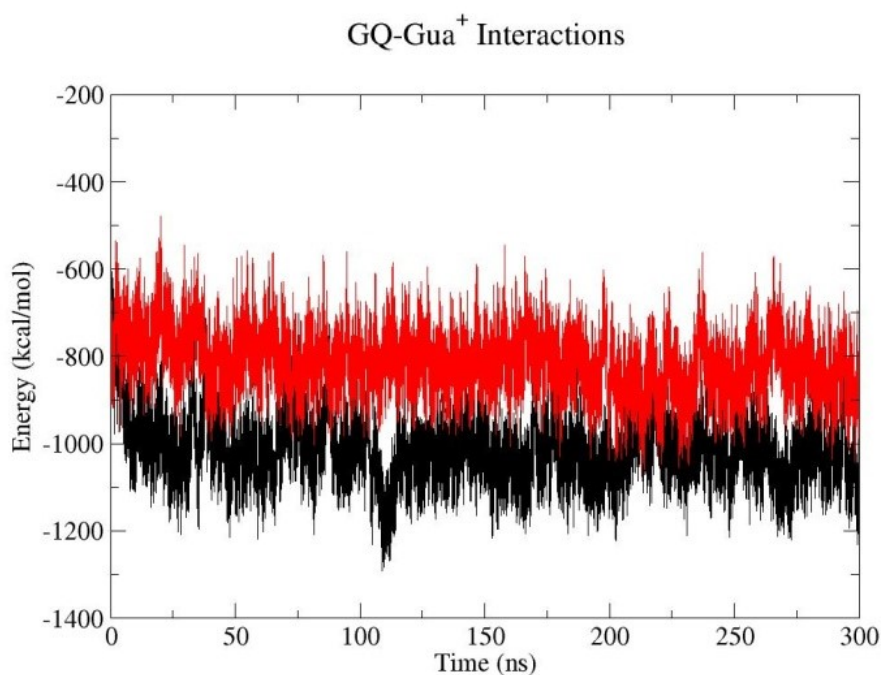

Appendix 3A.4 The interaction energy between G-quadruplex and cations present in the first solvation shell for Gua-IL (black line) and GuaHCl (red line).

Appendix 3A.5 The RMSD of heavy atoms of G-quadruplex, heavy atoms of Quartet and heavy atoms of backbone in case of Gua-IL (black line) and GuaHCl (red line).
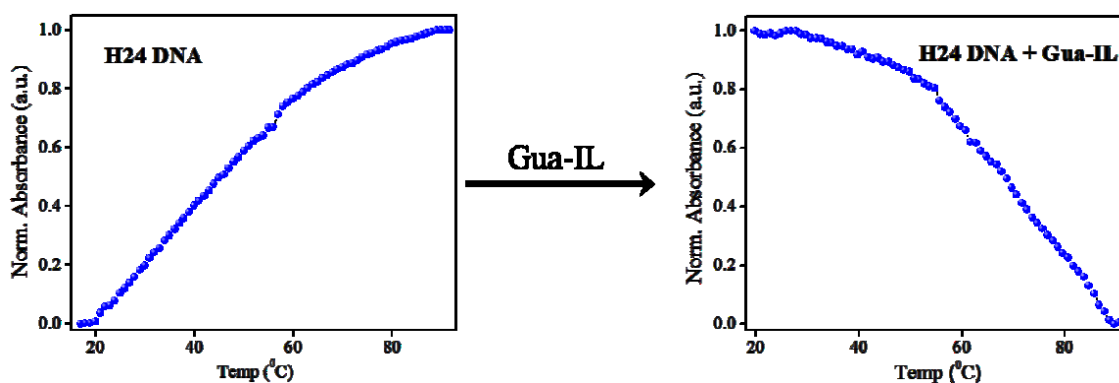

Appendix 3A.6 UV melting profile of H24 DNA (~ 5 μM) in deionized water in absence and presence of Gua-IL (100 μM).
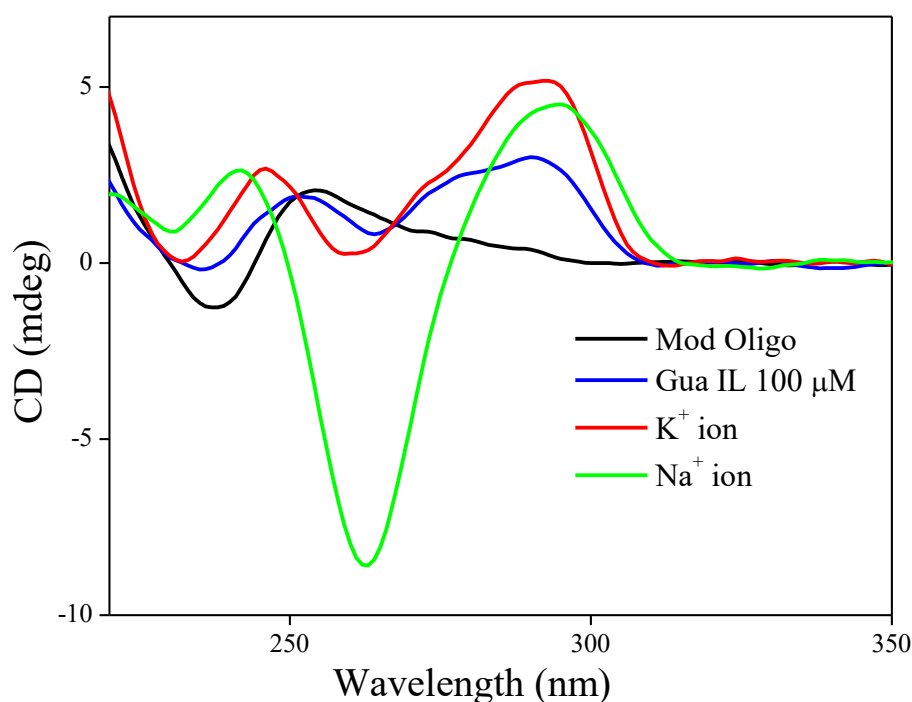
Appendix 3A.7 Circular dichroism spectra of fluorescent modified oligomer (~ 5 μM) in different conditions. "Mod Oligo" legend in the figure corresponds to the CD spectra of fluorescent modified oligomer in absence of any ion i.e. in deionised water. K$^+$ ion and Na$^+$ ion legends indicate the presence of 100 mM KCl and 100 mM NaCl respectively, in 10 mM tris buffer (pH 7.2) solution.
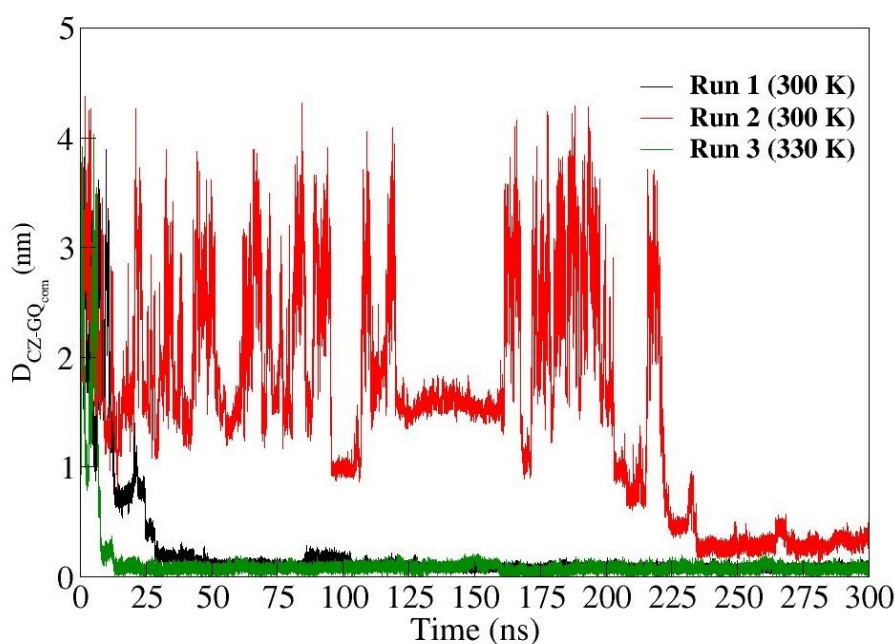
Appendix 3A.8 Distance between central Gua$^+$ (carbon atom of Gua$^+$ (CZ)) and CoM of GQ quartet region along time for Run 1 (black line) and Run 2 (red line) at 300K as well as for Run 3 (green line) performed at 330 K.
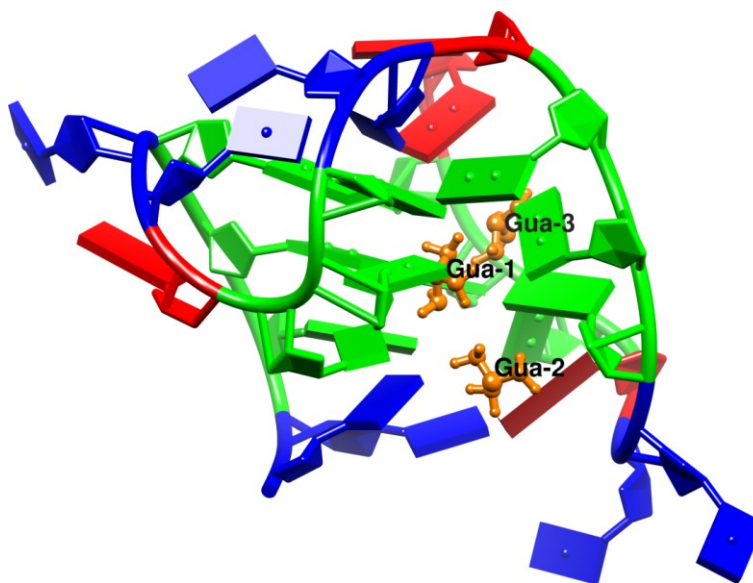
Appendix 3A.9 The figure showing 3 closest Guanidinium residues Gua-1, Gua-2 and Gua-3 in case of Gua-IL simulation-1 performed at 300 K.
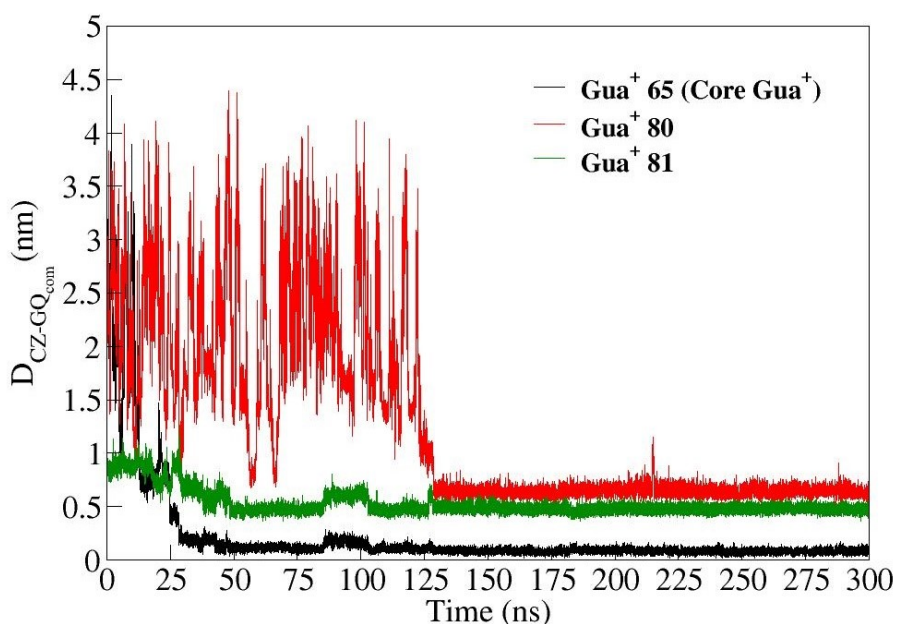
Appendix 3A.10 The figure showing distance between closest Gua$^+$ residues such as Gua-1 (Gua$^+$ present in the GQ core, black line), Gua-2 and Gua-3 (strongly bound from outer side, red and green line) along time. The carbon atom of Gua$^+$ (CZ) and CoM of quadruplex quartet is considered for calculations.

3A.6 Experimental

All experiments and sample preparations were carried out in deionized water. For preparing K^+ ion and Na^+ ion condition, we have used 100 mM KCl and NaCl salt respectively; in 10 mM tris buffer solution (pH 7.2). Before experiments H24-DNA was annealed at 90 °C for 10 min and stored at 4 °C for 24 h. The concentration of H24-DNA was determined using the molar extinction coefficient of ~244 600 M^{-1} cm^{-1} at 260 nm provided by IDT, USA. Here it is necessary to mention that IDT has determined the molar extinction coefficient using nearest neighbor approximation model. The lifetime analysis was done by IBH DAS6 software. We fitted these lifetime decays with minimum number of exponentials. Quality of each fitting was judged by χ^2 values and the visual inspection of the residuals. The value of $\chi^2 \approx 1$ was considered as best fit for the plots. For thermal melting, we used very small concentration of DNA (~1 μM) due to avoid the absorbance saturation. The melting temperature (T_m) was determined from the sigmoidal curve fit of the melting profile.

3A.6.1 Simulation Methodology:

3A.6.1.1 Force field:

The Gua-IL induced GQ structure shows similar ellipticity pattern with that of (3+1) hybrid structure observed in case of K^+ ion, hence we have used the starting geometry of (3+1) hybrid G-quadruplex (PDB ID 2GKU). The G-quadruplex structure is presented by AMBER94 forcefield[20] parameters. The simulations are performed to understand stability and interactions of two ionic liquids Gua-IL and BPyCl. Also, to understand role of cations the simulations in the presence of GuaHCl are also performed.

All simulations are performed using GROMACS 4.5.3 version[21].

Force field for all IL cations and anions is adopted from previous literatures as they displayed reliable structure and dynamics of these ions. The parameters are derived from ref. 29 for Guanidinium[22], ref. 30 for FEP[23], ref. 31 for BPy[24]. The LJ parameters for chloride ion (Cl⁻) are obtained from AMBER94[20] force field.

3A.6.1.2 System Preparation:

The cubic box having each side length of 5.2 nm is created and G-quadruplex molecule is kept in the center of the box. The length is chosen such as G-quadruplex molecule will get covered by 3 layers of water solvation shells and periodic boundary images would not interact with each other.

First, 61 cation molecules (Gua/BPy) will be added randomly in the box and later, 61 anion molecules (FEP / Cl⁻) were added randomly. Then, 3385 TIP3P water molecules were added to the box and 23 potassium ions were added randomly to neutralize the charge on DNA. This configuration is energy minimized using steepest descent method.[25]

The ratio of IL pairs to water molecules is similar to the experimental conditions.

3A.6.1.3 Equilibration protocol:

The system consists of G-quadruplex molecule, TIP3P water, corresponding IL pairs and K^+ ions. K^+ ions were added to neutralize the G-quadruplex and it was observed that they do not interfere in the stabilization mechanism of G-quadruplex molecule. The heavy atoms of G-quadruplex were restrained during equilibration and these restraining forces were slowly reduced along with energy minimization at each step using NPT ensemble. This helps biomolecule to properly soak into the surrounding aqueous environment. The Berendsen thermostat and barostat[26] is used to maintain temperature and pressure of the system respectively. The equilibration steps are as follows:

i. The system was heated slowly within 100 ps to desired temperature with restraining force of 25 kcal/mol/$Å^2$ on heavy atoms of DNA and then energy minimization was performed using steepest descent method.

ii. Then 50 ps equilibrations were performed with restraining forces on heavy atoms of G-quadruplex DNA. The restraining force was reduced as 5 kcal/mol/$Å^2$, 4 kcal/mol/$Å^2$, 3 kcal/mol/$Å^2$, 2 kcal/mol/$Å^2$, 1 kcal/mol/$Å^2$, 0.5 kcal/mol/$Å^2$ with energy minimization steps in between each restrained equilibration run.

iii. Finally, 1 ns unrestrained equilibration run with Nose-Hoover thermostat[27] and Parrinello-Rahman barostat[28] was performed. It was assured that RMSD does deviate much from the starting structure.

iv. Every simulation was performed with periodic boundary conditions. The particle mesh Ewald[29] method was used for Columbic interactions with a 10 Å cut-off. Center of mass of the whole system was removed after every 5 ps to avoid flying ice cube effect[30].

3A.6.1.4 Production runs protocol

The production simulations of 300 ns were carried out in the cases where G-quadruplex forms stable structure. For rest of the simulation, the production runs were performed for ~ 200 ns or till the simulation time when quartet core of stacked guanines was disrupted.

3A.6.1.5 Analyses:

3A.6.1.5.1 Radial distribution function (RDF):

We have calculated the RDF using g_rdf utility of GROMACS 4.5.3[21]. We have considered the central carbon atom (CZ) of Gua^+ cation, positive nitrogen atom (N1) of Bpy^+ to calculate RDF with respect to surface atoms of G-quadruplex. Similarly, oxygen atoms of water molecules (OW) were used to study RDF of water molecules around G-quadruplex.

3A.6.1.5.2 Spatial distribution function (SDF):

The SDF's of cations and anions around G-quadruplex molecule in each case of GuaIL, BPyCl and GuaHCl were calculated using TRAVIS software.[31] The distribution was calculated considering CZ atom for Gua^+ and N1 atom of Bpy^+ molecules. The DNA molecule was aligned with respect to the starting structure before calculating SDF. The SDF is visualized using UCSF Chimera Software.[32] In case of Gua-IL, the SDF was calculated from 130 ns onwards as Gua^+ ions get equilibrated around GQ.

3A.7 References

1. M. Gellert, M. N. Lipsett and D. R. Davies, *Proc. Natl. Acad. Sci. U. S. A.*, 1962, **48**, 2013.

2. E. H. Blackburn and J. G. Gall, *J. Mol. Biol.*, 1978, **120**, 33-53.

3. A. G. Bodnar, M. Ouellette, M. Frolkis, S. E. Holt, C.-P. Chiu, G. B. Morin, C. B. Harley, J. W. Shay, S. Lichtsteiner and W. E. Wright, *Science*, 1998, **279**, 349.

4. N. W. Kim, M. A. Piatyszek, K. R. Prowse, C. B. Harley, M. D. West, P. L. Ho, G. M. Coviello, W. E. Wright, S. L. Weinrich and J. W. Shay, *Science*, 1994, **266**, 2011.

5. (a) J.-L. Mergny, *Nat. Chem. Biol.*, 2012, **8**, 225-226; (b) S. N. Georgiades, N. H. Abd Karim, K. Suntharalingam and R. Vilar, *Angew. Chem. Int. Ed.*, 2010, **49**, 4020-4034; (c) J. Thevarpadam, I. Bessi, O. Binas, D. P. N. Gonçalves, C. Slavov, H. R. A. Jonker, C. Richter, J. Wachtveitl, H. Schwalbe and A. Heckel, *Angew. Chem. Int. Ed.*, 2016, **55**, 2738-2742.

6. (a) N. W. Luedtke, *CHIM. Int. J. Chem.*, 2009, **63**, 134-139; (b) S. Balasubramanian, L. H. Hurley and S. Neidle, *Nat. Rev. Drug Discov.*, 2011, **10**, 261-275; (c) J. Mohanty, N. Barooah, V. Dhamodharan, S. Harikrishna, P. I. Pradeepkumar and A. C. Bhasikuttan, *J. Am. Chem. Soc.*, 2013, **135**, 367-376.

7. (a) G. N. Parkinson, M. P. H. Lee and S. Neidle, *Nature*, 2002, **417**, 876-880; (b) Y. Wang and D. J. Patel, *Structure*, 1993, **1**, 263-282; (c) J. Y. Lee, B. Okumus, D. S. Kim and T. Ha, *Proc. Nat. Acad. Sci. U.S.A.*, 2005, **102**, 18938-18943.

8. J. Choi and T. Majima, *Chem. Soc. Rev.*, 2011, **40**, 5893-5909.

9. (a) S. D. Verma, N. Pal, M. K. Singh, H. Shweta, M. F. Khan and S. Sen, *Anal. Chem.*, 2012, **84**, 7218-7226; (b) K. Bhadra and G. S. Kumar, *Biochim. Biophys. Acta, Gen. Subj.*, 2011, **1810**, 485-496; (c) S. Ghosh, A. Kar, S. Chowdhury and D. Dasgupta, *Biochemistry*, 2013, **52**, 4127-4137.

10. A. Siddiqui-Jain, C. L. Grand, D. J. Bearss and L. H. Hurley, *Proc. Nat. Acad. Sci. U.S.A.*, 2002, **99**, 11593-11598.

11. S. Satpathi, A. Sengupta, V. M. Hridya, K. Gavvala, R. K. Koninti, B. Roy and P. Hazra, *Sci. Rep.*, 2015, **5**, 9137.

12. (a) T. Ueki and M. Watanabe, *Macromolecules*, 2008, **41**, 3739-3749; (b) Y. Akdogan, M. J. N. Junk and D. Hinderberger, *Biomacromolecules*, 2011, **12**, 1072-1079; (c) H. Tateishi-Karimata and N. Sugimoto, *Nucleic Acids Res.*, 2014, DOI: 10.1093/nar/gku499.

13. (a) R. Vijayaraghavan, A. Izgorodin, V. Ganesh, M. Surianarayanan and D. R. MacFarlane, *Angew. Chem. Int. Ed.*, 2010, **49**, 1631-1633; (b) K. D. Clark, O. Nacham, H. Yu, T. Li, M. M. Yamsek, D. R. Ronning and J. L. Anderson, *Anal. Chem.*, 2015, **87**, 1552-1559.

14. K. Fujita and H. Ohno, *Chem. Commun.*, 2012, **48**, 5751-5753.

15. K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix and D. J. Patel, *J. Am. Chem. Soc.*, 2006, **128**, 9963-9970.

16. J. L. Huppert, *Chem. Soc. Rev.*, 2008, **37**, 1375-1384.

17. Y. Xue, Z.-y. Kan, Q. Wang, Y. Yao, J. Liu, Y.-h. Hao and Z. Tan, *J. Am. Chem. Soc.*, 2007, **129**, 11185-11191.

18. C. N. Pace, in *Methods in Enzymology*, Academic Press, 1986, vol. 131, pp. 266-280.

19. (a) R. W. Sinkeldam, A. J. Wheat, H. Boyaci and Y. Tor, *ChemPhysChem*, 2011, **12**, 567-570; (b) A. A. Tanpure and S. G. Srivatsan, *Nucleic Acids Res.*, 2015, DOI: 10.1093/nar/gkv743.

20. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1996, **118**, 2309-2309.

21. B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435-447.

22. E. Wernersson, J. Heyda, M. Vazdar, M. Lund, P. E. Mason and P. Jungwirth, *J. Phys. Chem. B*, 2011, **115**, 12521-12526.

23. X. Zhang, F. Huo, Z. Liu, W. Wang, W. Shi and E. J. Maginn, *J. Phys. Chem. B*, 2009, **113**, 7591-7598.

24. S. V. Sambasivarao and O. Acevedo, *J. Chem. Theory Comput.*, 2009, **5**, 1038-1050.

25. G. Arfken, in *Mathematical Methods for Physicists (Third Edition)*, ed. G. Arfken, Academic Press, 1985, DOI: http://dx.doi.org/10.1016/B978-0-12-059820-5.50015-X, pp. 428-436.

26. H. J. C. Berendsen, J. P. M. Postma, W. F. v. Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684-3690.

27. S. Nose, *Mol. Phys.*, 1984, **52**, 255-268.

28. M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182-7190.

29. D. Y. Tom Darden, Lee Pedersen *J. Chem. Phys.*, 1993, **98**, 10089-10092.

30. S. C. Harvey, R. K. Z. Tan and T. E. Cheatham, *J. Comput. Chem.*, 1998, **19**, 726-740.

31. M. Brehm and B. Kirchner, *J. Chem. Inf. Model.*, 2011, **51**, 2007-2023.

32. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J. Comput. Chem.*, 2004, **25**, 1605-1612.

Chapter 3B

Controlling an Anticancer Drug Mediated G-quadruplex Formation and Stabilisation by a Molecular Container

3B.1 Introduction

Among different secondary structures of DNA, G-quadruplex DNA (GQ) has received special attention among researchers due to its potential therapeutic applications in the field of cancer research.[1] In general, human telomeric DNA contains tandem repeats of guanine-rich sequences (i.e. d(TTAGGG)) which end in a single-stranded 3′ overhang fashion with 100-200 bases.[2] Replication of this non-coding DNA part, i.e. telomere, is governed by the telomerase enzyme, and this is found to be overexpressed in cancerous cells in comparison to normal cells.[3] Under physiological condition, the overhang strand in telomere can form different types of GQ structures depending on the directionality of the strands. This GQ structure formation is found to have an inhibitory effect towards the telomerase enzyme.[4] Since then, stabilization of GQ structures by small organic molecules or ligands have been regarded as one of the potential active fields for anticancer research.[1a, 5] Apart from quadruplex stabilization using external ligands, significant attempts have also been made in literature to study G-quadruplex stabilization using internal turn-on fluorescence probes, i.e. using modified nucleotides.[6] Interestingly, highly stable G-quartets in GQ structures behave as a rigid knot for DNA unwinding, which is one of the major and primary steps towards DNA replication. Thus, some biologically important processes, like, DNA replication and transcription processes are hampered due to the GQ formation.[7] Generally, different types of helicase enzymes (FANCJ, pif1, WRN etc.) unwind these GQ structures to carry out the DNA replication process.[7-8] But additional stability by ligands towards GQ structure inhibits the activity of helicase enzymes to a large extent which is found to be detrimental for DNA replication processes.[7, 9] Thus, controlling of this ligand mediated GQ formation is equally important and challenging in addition to their GQ stabilization aspect. In this quest, many researchers have tried to regulate the G-quadruplex formation using several external stimuli.[10] However, most of these approaches for GQ reversibility depends on using light sensitive modified nucleobases,[10a] pH dependent inorganic complex[11] and metal mediated conformational switch.[10b] Although there are several reports about molecular container controlled ligand binding to duplex DNA,[12] studies on regulating this type of ligand mediated GQ DNA formation are rather limited.[13] Therefore, it is important to develop a new strategy for controlling ligand mediated GQ formation and stabilization, which is found to be more general approach considering the versatile applicability towards GQ stabilizing ligands based on small organic molecules. During our ongoing project, Zhou and coworkers reported the reversible manipulation of GQ-DNA structures through supramolecular host-guest interactions using CD and NMR techniques.[14] However, it would be very simple and effective if the GQ formation and reversible conformational transition by external stimuli can be directly monitored using fluorescence color switch. Herein, we have employed an anti-cancer drug, topotecan (TPT) to detect this conformational switch between GQ DNA and single-

stranded/random coil DNA (H24 DNA) with the help of fluorescence color switching of the drug, which makes this approach more advantageous compared to the previous studies.

Topotecan (TPT) exhibits its antitumor activity by means of inhibiting topoisomerase I (topo I) enzyme through the formation of a cleavable ternary complex (i.e. DNA-TPT-topo I).[15] Close analogue of TPT, camptothecin has been found to synergistically increase the therapeutic efficacy of an antitumor GQ ligand.[16] Very recently, Yuan and coworkers reported the transcriptional regulation in C-Myb genes through the binding of TPT to the G-quadruplex
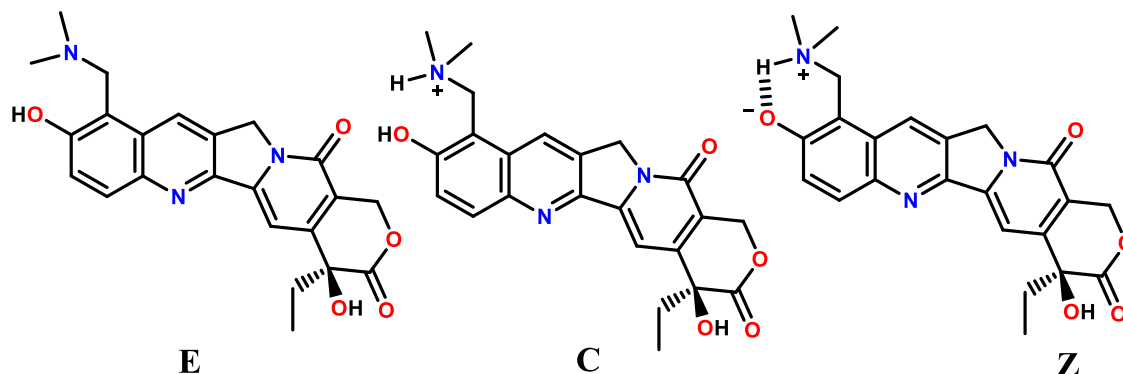
**Scheme 3B.1** Different protolytic forms of Topotecan (TPT).

in the presence of ammonium ion.[17] Keeping its therapeutic application in mind, researchers have explored the optical properties of TPT in bulk and various restricted and organised biological assemblies.[18] TPT exhibits interesting fluorescence property based on its different protolytic forms like enol (E), cationic (C) and zwitterionic (Z) forms depending on the pH and polarity of the medium (Scheme 3B.1).[19] Uniqueness in the photophysical property of TPT lies in its distinctly different emission maxima for different forms (Z* emits at ~530 nm and C* at ~430 nm). Here, we have utilised the large difference (~100 nm) in the emission maxima between these two forms for probing the conformational transition between GQ and its subsequent transition to single-stranded/random coil form by a molecular container. We have shown that TPT can instigate and further stabilise the GQ structure in the absence of any ion (Na$^+$/K$^+$) of a biologically relevant human telomeric sequence, H24 DNA (5′-d(TTAGGG)$_4$-3′). With the help of MD simulation study, we have proposed the mechanism of this stabilisation process. To regulate this GQ formation, we have introduced cucurbit[7]uril (CB7, Appendix 3B.1) in the solution, which can revert this transition, i.e., GQ → H24. Interestingly, this entire process can be monitored through fluorescence colour switch from green (H24-TPT system) to violet (H24-TPT-CB7 system). The entire conformational transitions have also been monitored by other spectroscopic tools, like, circular dichroism (CD), nuclear magnetic resonance (NMR), thermal melting and time-resolved fluorescence techniques. Considering a large number of aromatic/drug molecules' ability for GQ stabilisation in literature, this present study may offer

an alternative approach to regulate and monitor the formation of G-quadruplex structure without exerting any intricate and laborious technique.

3B.2 Results and Discussions

3B.2.1 CD Study

Circular dichroism (CD) measurements have been employed to monitor the conformational transitions of human telomeric DNA (H24) in the presence of TPT. All the studies have been performed in deionized water (pH 6) to eliminate any possibility of GQ formation by ion (like Na^+, K^+ etc.) and also to avoid the hydrolysis of TPT to form toxic carboxylate ($pK_a \sim 7$).[18a, 19a] Notably, distinctly different ellipticity patterns in CD spectra arise from different GQ topologies.[4a] H24 in deionized water exhibits an ellipticity pattern (positive peak around 260 nm and negative peak around 235 nm) for single-stranded/random coil structure (Appendix 3B.2), which gets altered in presence of TPT (Figure 3B.1a, H24-TPT), and two positive peaks (at 290 and 250 nm) along with two negative peaks (at 235 and 380 nm) appear in the CD profile. Notably, this spectral pattern closely matches with the ellipticity pattern of hybrid GQ structure (Appendix 3B.2).[4a, 5b] Thus, CD results suggest the TPT instigated GQ formation, which has a topology closer to mix/hybrid structure (Figure 3B.1a, H24-TPT and Appendix 3B.2). An aqueous solution of TPT shows an intense positive peak at 225 nm and a negative peak at 380 nm (Figure 3B.1b), which is responsible for the observed negative peak at 380 nm in the H24-TPT system (Figure 3B.1a, H24-TPT). Next, we have introduced CB7 to the H24-TPT system, and a broad positive ellipticity pattern at 270 nm and less intense negative peaks at ~240 nm and ~380 nm (Figure 3B.1a, H24-TPT-CB7) appeared in the CD profile. To find the origin of these ellipticity patterns in H24-TPT-CB7 system, we have recorded the CD spectra of
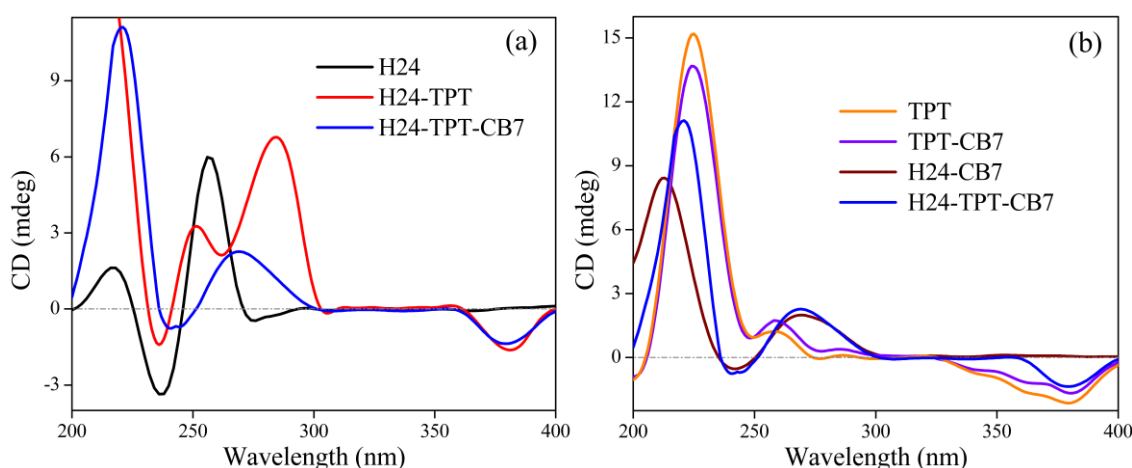
Figure 3B.1 (a) Circular dichroism (CD) spectra of H24 DNA in the presence of topotecan (TPT) and TPT + cucurbit[7]uril (CB7). (b) CD spectra of TPT and H24 DNA in presence and absence of CB7. H24 legend in the figure corresponds to the CD spectra of human telomere in the absence of Na^+/ K^+ ion, i.e. in deionised water.

TPT-CB7 (in the absence of H24) and H24-CB7 (in the absence of TPT) systems (Figure 3B.1b). TPT-CB7 system exhibits an almost similar ellipticity pattern to that of only TPT (Figure 3B.1b). Interestingly, the observed ellipticity pattern in H24-CB7 (Figure 3B.1b) resembles that of H24-TPT-CB7 system (Figure 3B.1a). However, H24-CB7 system (Figure 3B.1b, H24-CB7) shows a broad ellipticity pattern at 270 nm and a weak negative peak at 240 nm, which is different from the ellipticity pattern of only H24 CD spectrum (Appendix 3B.2). Here, it is pertinent to mention that although CB7 does not have any specific interaction with duplex DNA, CB7 can interact with H24 DNA due to its single-stranded nature. The nucleobases in single-stranded DNA (i.e. H24) are more accessible in comparison to the duplex DNA due to the absence of Watson-Crick base pairing as well as helix formation. It is well known that different nucleobases interact with CB7 at varying affinity and the absorption spectrum of the nucleobases is found to be red shifted in the presence of CB7.[20] Importantly, Venkataramanan and co-workers reported that the existence of charges "splattering" in guanine-CB7 inclusion complex is the reason for its higher stability.[20c] Similar kinds of interactions between nucleobases of H24 and CB7 are also pertinent here, which may be responsible for the different ellipticity pattern of H24 DNA in comparison with H24-CB7. Hence, the spectral similarity between H24-TPT-CB7 and H24-CB7 system indicates the reformation of H24 structure from GQ structure due to the translocation of TPT from GQ-DNA to CB7 nanocavity followed by interaction with CB7 nano-cavity to produce an ellipticity pattern different from H24 only (Figure 3B.1a). In a nutshell, CD studies suggest the TPT induced GQ formation in the absence of any ion and its reformation to H24 form in the presence of CB7 nanocavity.
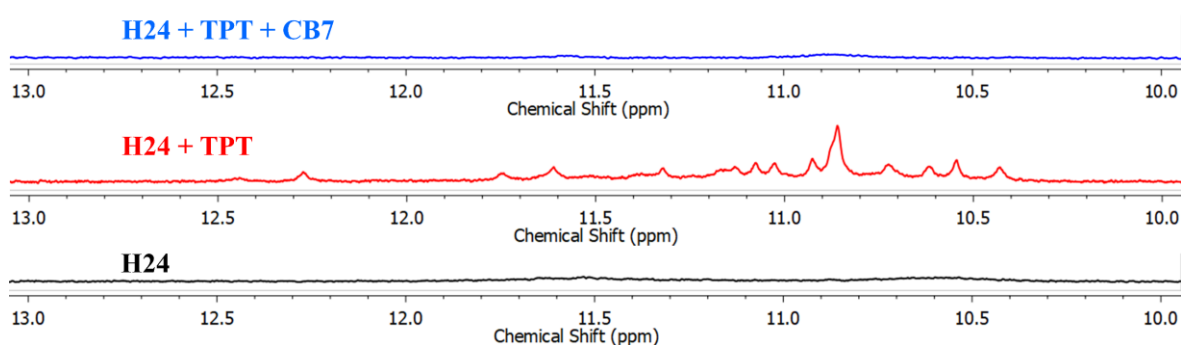

Figure 3B.2 ^{1}H NMR spectra of H24 DNA, H24+TPT, H24+TPT+CB7 at 25° C.

3B.2.2 NMR Study

Nuclear magnetic resonance (NMR) spectroscopy is an efficient tool to distinguish the GQ formation by monitoring the unique chemical shift value (i.e. in the 10-13 ppm region) of hydrogen bonded imino protons of guanine base pairs.[21] The NMR spectrum of H24 DNA (Figure 3B.2) does not contain any distinct peaks in the 10-13 ppm region indicating the presence of a single-stranded/random coil structure of the G-rich sequence DNA. However,

H24-TPT system (H24:TPT=1:3) (Figure 3B.2) exhibits the characteristic imino proton peaks between 10 to 13 ppm, which closely matches with the K^+ induced hybrid type GQ formation.[21] Notably, proton NMR spectrum of TPT does not contain any peak in the 10-13 ppm region,[22] which clearly indicates that the above-mentioned new peaks are originated due to the hydrogen bonded imino protons of guanine base pairs during GQ DNA formation. Interestingly, the addition of CB7 molecules in the same H24-TPT system does not show any imino proton peaks in the 10-13 ppm region (Figure 3B.2), indicating the formation of single-stranded/random coil structure from GQ DNA, probably owing to the encapsulation of TPT molecules in the CB7 nano-cavity. Thus, NMR spectroscopy measurements give us direct evidence of the TPT induced GQ formation and its conformational transition to single-stranded/random coil structure in presence of CB7 molecules by simply monitoring the chemical shift of imino protons in guanine base pairs.
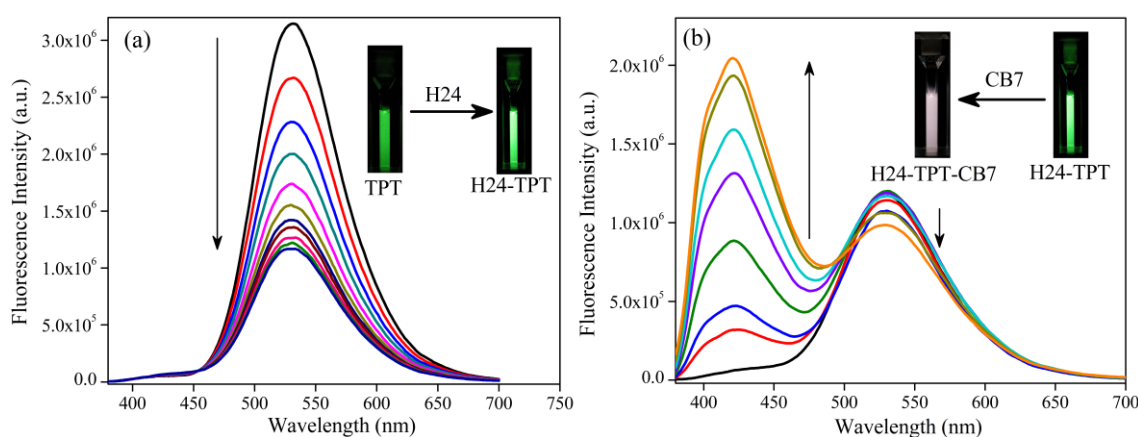

Figure 3B.3 Steady state fluorescence spectra of TPT (λ_{ex} 375 nm) in presence of (a) H24 (~ 20 µM) and (b) H24 and CB7 (~ 1.5 mM). Arrow indicates the gradual addition of H24 or CB7 to the solution.

3B.2.3 Steady State Studies (Fluorescence, ITC and UV Thermal Melting)

In bulk water (pH 6), TPT predominantly exists in cationic form (C) in ground state due to the higher pK_a values of dimethylamino and 10-hydroxyl groups of 9.5 and 6.99, respectively (Scheme 3B.1).[18a, 19a] However, in excited state, the zwitterionic form (Z^*) is the major emitting species of TPT exhibiting a green emission maxima around ~530 nm, which originates from the excited state proton transfer (ESPT) between the 10-hydroxyl group of cationic TPT (C^*) and water molecules.[18a, 18c] With gradual addition of H24 DNA, the fluorescence spectrum of TPT shows a decrement in green emission without showing any significant shift in emission position (Figure 3B.3a). This observation suggests the interaction between H24 DNA and TPT as a result of GQ DNA formation instigated by TPT. Absence of any spectral overlap between the donor emission (Z^*) and the absorption of the acceptor (nucleobases) indicates that this quenching in the emission maximum may be attributed to the photoinduced electron transfer from guanine nucleobases of DNA to the zwitterionic form (Z^*) of TPT, which has been previously observed

in G-rich sequences.[18d] At higher concentrations of H24, the emission intensity does not change inferring that almost all TPT molecules involved in the binding to GQ DNA.

We have also probed the interaction between TPT and H24 DNA by ITC study, which exhibits a binding constant of 3.7×10^5 M^{-1} for H24-TPT system (Figure 3B.4). Moreover, the ITC results infer that the binding process between TPT and DNA is an enthalpically (ΔH -21.8 kcal mol^{-1}) driven process as the negative entropy change associated with it (-48.2 cal mol^{-1}), and the overall process is a spontaneous one (ΔG -20.8 kcal mol^{-1}). The exothermic binding process between TPT and DNA molecules can be attributed to both electrostatic (such as H-bonding), and dispersion types (van der Waals) as found in the MD simulations studies for TPT and GQ DNA interaction (discussed in later part).
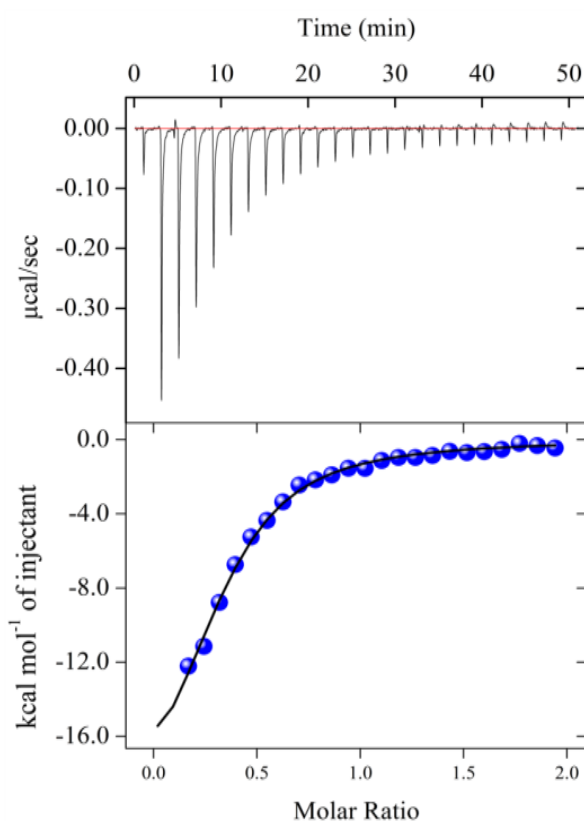
Figure 3B.4 Upper panel shows the ITC plot of H24 DNA (in a syringe) with TPT (in the cell), and the lower panel shows the integrated heat profile of the calorimetric titration plot shown in the upper panel. The solid line represents the best nonlinear least-squares fit a single binding site model.

Next, we discuss the regulation of GQ formation using a molecular container, i.e. CB7. Addition of CB7 leads to the decrement in the emission intensity at ~530 nm and a new emission peak for the cationic form of TPT (C^*) emerges at ~430 nm (Figure 3B.3b). Highest addition of CB7 (1.5 mM) makes the cationic peak (at 430 nm) as the predominant one over the zwitterionic form (at 530 nm). This switch of fluorescence colour from green (530 nm) to violet (430 nm) can be attributed to the following reasons. First, the encapsulation of TPT inside the CB7

nanocavity decreases the presence of water molecule in the vicinity of the cationic form, (C^*) resulting in the inhibition of ESPT process responsible for the conversion from C^* to Z^*. Second, the upward pK_a shift of the 10-hydroxyl group (pK_a^* -2.62) inside the CB7 nanocavity (pK_a^* 5.51)[18a] is also responsible for this inhibition of ESPT process. As there is no free drug in the solution, the observed fluorescence colour switch from green to violet in the presence of molecular container confirms that TPT molecules translocate from the GQ-DNA to the CB7 nano-cavity, and is responsible for this DNA structural transition from GQ to H24 form, i.e., GQ-DNA→H24. Binding constant (K) for the TPT with H24 DNA system is found to be 3450 ± 345 M^{-1} (Appendix Note 3B.1), which is lower than that of TPT with CB7[18a] (K = 5000 M^{-1}). Thus, the calculated binding constants from fluorescence measurements also indicate the translocation of TPT from GQ DNA to CB7 nano-cavity.
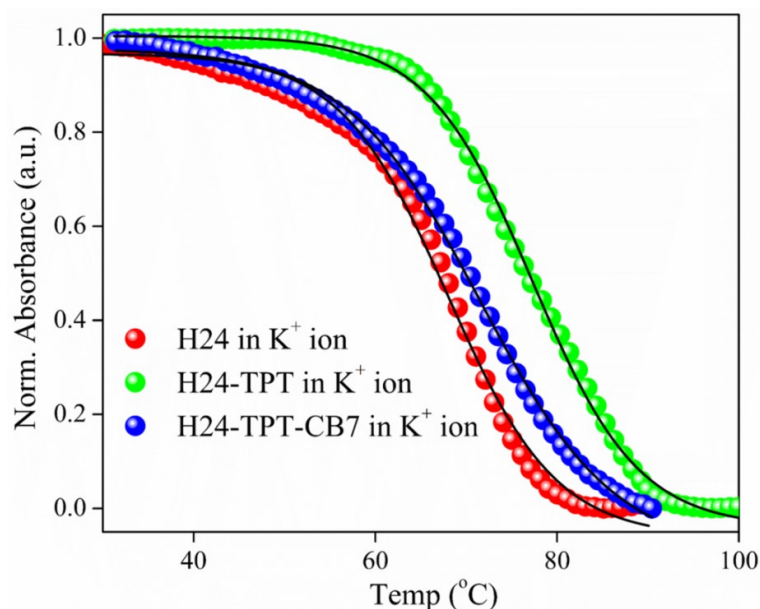
Figure 3B.5 UV melting spectra of GQ in buffer with simultaneous addition of TPT and CB7. GQ legend corresponds to the melting spectra of human telomere in the presence of 150 mM KCl.

To further validate the formation of GQ and its transition to H24 in light of their stabilities, we have employed thermal melting measurements of H24-TPT in presence and absence of CB7 nanocavity in 150 mM KCl containing buffer solution (Figure 3B.5). TPT imposes an additional stabilization of ~9 °C (T_m = 77 °C) towards K^+ ion stabilized GQ DNA (T_m = 68 °C). However, this structural stabilisation exhibits a substantial decrement of ~7 °C with the incorporation of CB7 in the solution (T_m = 70 °C). This decrement in melting temperature value indicates that CB7 takes away some of the TPT molecules from the GQ DNA, thereby lowering its stabilisation (Figure 3B.5). In a nutshell, we can infer that disappearance of the characteristic ellipticity pattern and imino proton NMR peaks for GQ, decrement in melting temperature and fluorescence color switch from green to violet indicate the reverse

process of GQ formation, i.e., GQ-DNA→H24 transition, proceeding through the encapsulation of TPT molecules inside CB7 nano-cavity, which is involved in the GQ formation and stabilization process.

3B.2.4 Time-resolved Fluorescence Studies

Time-resolved fluorescence measurements have been employed to explore this conformational transition of DNA by monitoring the lifetime of TPT molecules. Lifetime profiles of TPT in absence and presence of H24-DNA/H24-DNA-CB7 have been collected around ~430 nm for cationic form (C^*) and ~530 nm for zwitterionic form (Z^*) (Figure 3B.6a, Appendix Table 3B.1). In bulk water (pH 6), Z^* form of TPT (collected at 530 nm by exciting at 375 nm) exhibits a single exponential decay profile with a lifetime of 5.88 ns, and this lifetime is found to be unaltered (5.88 ns to 5.66 ns) even in the presence of maximum concentration (20 μM) of H24. However, a huge quenching in Z^* emission intensity in steady state spectra is observed in the presence of H24; hence, it is quite rational to predict a ground state dark complexation between H24 and TPT molecules as a result of GQ DNA formation, which does not emit in the excited state (Appendix 3B.4). Notably, this type of static quenching has also been observed for TPT molecules binding with other types of DNA.[18d]
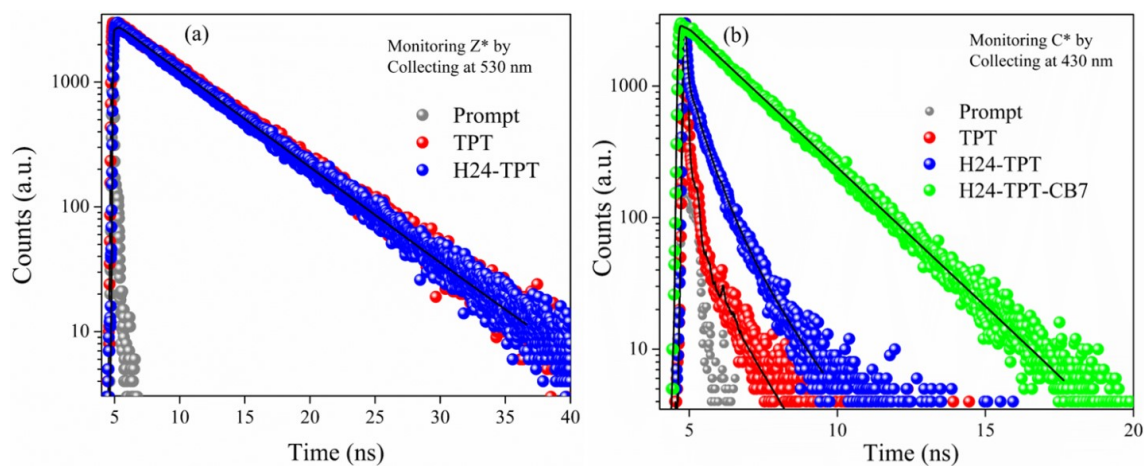
Figure 3B.6 Lifetime decays of TPT (a) at λ_{ex} 375 nm and λ_{col} 530 nm (b) at λ_{ex} 375 nm and λ_{col} 425 nm in deionised water with simultaneous addition of H24 (20 μM) and CB7 (1.5 mM).

Interestingly, the addition of CB7 to the H24-TPT system exhibits a huge increment in the average lifetime of C* form (collected at 430 nm by exciting at 375 nm) from ~40 ps to 1.62 ns. This can be attributed to the reduction in nonradiative decay pathways of TPT arising due to the encapsulation of cationic form inside the CB7 nanocavity (Figure 3B.6b, Appendix Table 3B.2). Enhancement in the C* form lifetime corroborates well with our steady state results, where a huge increment in the emission intensity of the C* form (around 420 nm) is observed in the presence of CB7. On the other hand, lifetime decays of Z* form (collected at 530 nm by exciting at 375 nm) shows a decrement from 5.66 ns to 4.54 ns at the maximum concentration of CB7

(1.5 mM), which has been previously attributed to the hydrogen bonding interaction between carbonyl groups of the CB7 portal and Z* form of TPT.[18a] In a nutshell, time-resolved fluorescence studies indicate a strong interaction of TPT with GQ DNA, which is significantly altered in the presence of CB7, and the entire observations corroborate well with our steady state results.

3B.2.5 Molecular Dynamics (MD) Simulation Study

Binding mode of TPT to GQ has been explored by utilising the molecular dynamic simulation. As discussed in the Methods section, we have created two systems to explore the binding of TPT to GQ, one using the same concentration of GQ and TPT (1:1 GQ/TPT complex) and the other with a higher concentration of TPT (1:2 GQ/TPT complex). Starting from the separated configuration of GQ and TPT for both types of complexes, we performed multiple simulations with different initial conditions (see Simulation Method). A total of 9 simulations were performed for 1:1 complex, whereas 3 simulations were performed for 1:2 complex. The initial geometries used for the simulation are shown in Appendix 3B.5. Analyses were carried out over the last 80 ns of each trajectory. Two binding modes of TPT are possible for 1:1 GQ-TPT system. One is a groove (side bind) binding mode, where the TPT interacts with the groove of GQ, and the other is the terminal binding mode, where TPT molecule stacks at the terminal of the GQ (Figure 3B.7a). The time variation of distances and angles for all the complexes (1:1 GQ/TPT and 1:2 GQ/TPT complex) are shown in Appendix 3B.6. To characterize the preferable binding modes, we have calculated the distribution of distance (from the COM of GQ to COM of TPT) and angle (between COM of GQ to COM of TPT vector and GQ axis) for TPT around GQ for 1:1 and 1:2 GQ/TPT systems as shown in Figure 3B.7b and Figure 3B.7c. There are two peaks for 1:1 GQ/TPT complex: the first at 15 Å distance and ~30˚ angle, which represents the terminal bound state (Figure 3B.7b and Appendix 3B.5) and the other at 14 Å distance and 40˚ angle, which is also a terminal bound state but TPT is closer to the backbone of the terminal base pairs. The most probable structures for 1:1 GQ/TPT complex are shown in Appendix 3B.5. We have obtained all possible terminal bound structures. Further, very less population around 90˚ angle indicates that TPT does not prefer to bind to the groove of GQ. However, for 1:2 GQ/TPT complex, the distribution of distance and angle exhibit two maxima at a distance of 13.8 Å and angle of 25˚ (indicating the terminal bound state), and at the distance of 15.9 Å and angle of 79˚ representing the groove bound state (Figure 3B.7c and Appendix 3B.5). Two different binding modes in case of 1:2 GQ/TPT indicates that groove-binding is the second preferable choice after the terminal position of the GQ is occupied by the first TPT.

We have also analysed the stabilities of 1:1 and 1:2 GQ/TPT systems in terms of their root-mean-square deviation (RMSD) (Figure 3B.7d) of GQ from the crystal structure geometry.

Both systems show less change in RMSD (<5Å), indicating the stability of GQ in the presence of TPT. Further, for control, we have simulated the GQ in the absence of TPT molecule. The higher RMSD (Figure 3B.7d) values of GQ in the absence of TPT molecule indicates that GQ is not stable in water in the absence of ions. Thus, these RMSD results suggest that the binding of TPT to GQ is responsible for their GQ stability in the absence of any ion in solution. Overall, MD simulation studies show that the terminal and side binding modes of TPT are responsible for the translocation of TPT to the nano-cavity of the molecular container due to their easy accessibility from the terminal position.
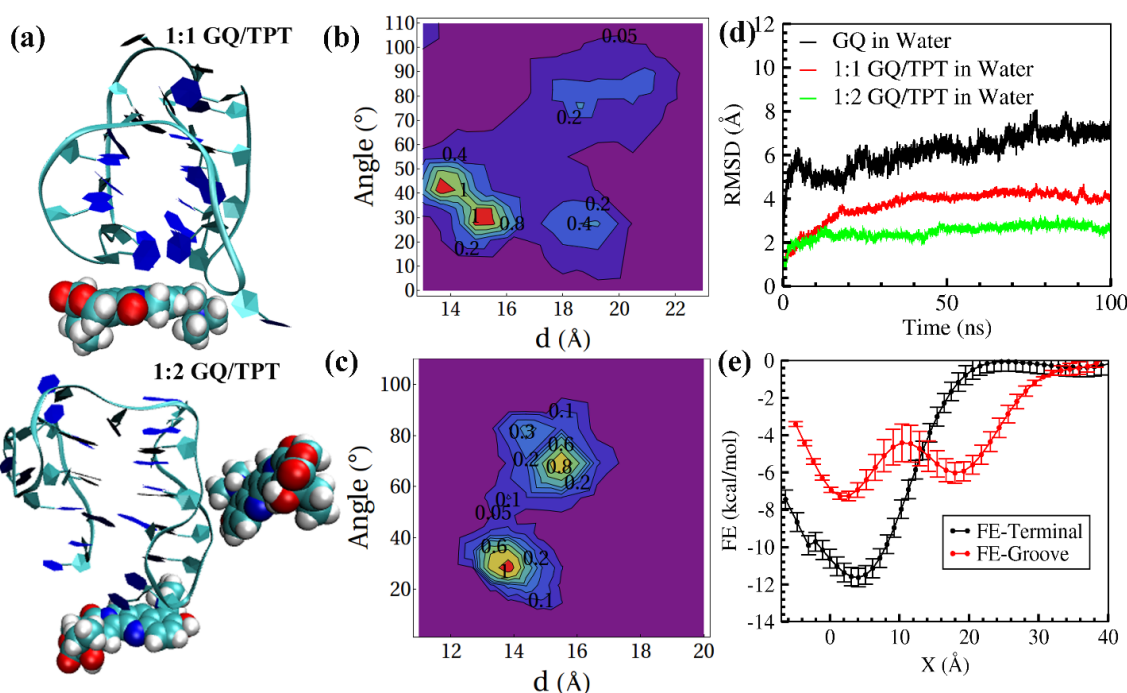

Figure 3B.7 (a) Representative binding modes of 1:1 and 1:2 GQ/TPT complexes. Distribution of distance and angle (b) in 1:1 GQ/TPT complex and (c) in 1:2 GQ/TPT complex. (d) RMSD of GQ in water, 1:1 GQ/TPT and 1:2 GQ/TPT complex. (e) Average free energy (FE) of binding to the terminal (black) and groove (red) of GQ from two independent well-tempered metadynamics simulations for 1:1 GQ/TPT complex. The bar in FE represents the error in FE.

Since TPT did not unbind from the GQ either from the terminal bound state or groove-bound state in 100 ns, the dissociation barrier of TPT from GQ must be high, and computationally demanding using normal MD simulation. Therefore, we have performed two independent well-tempered metadynamics simulations with different initial conditions to calculate the free energy of binding of TPT to both the terminal and groove of GQ using 1:1 GQ/TPT system. Appendix 3B.7 shows that two independent metadynamics simulations provide similar free energy profiles from both the terminal and groove-bound systems. The initial and final representative structures of the metadynamics simulation for both terminal and groove-bound state are shown in Appendix 3B.8. Figure 3B.7e shows that binding free energy to the terminal (~12kcal/mol) of GQ is 4 kcal/mol (with an error of ~1 kcal/mol) stronger than

binding to the groove bound state (~8 kcal/mol). This observation is consistent with the result of the normal MD simulation, where more often the TPT molecule binds to the terminal than groove of GQ starting with the separated state for 1:1 GQ/TPT complex, and groove-bound state follows terminal bound state in case of 1:2 complex.

To analyse further the reason for the preference of TPT towards the terminal site of GQ, we have calculated the total binding energy between the two species, which is composed of (dispersion interaction) and electrostatic components. Figure 3B.8 shows both van der Waals the average total energy obtained from the two metadynamics simulation for both the terminal and groove bound states for 1:1 complex and their decomposition into dispersive and electrostatic interaction. We can see that for the terminal bound state (Figure 3B.8b) has stronger dispersive interaction compared to the groove-bound state (Figure 3B.8c) which, however, has a stronger electrostatic interaction than the terminal bound state (Figure 3B.8b). Although both the states are stabilised by both van der Waals and electrostatic energy components, the stabilisation stemming from the dispersive interaction due to the stacking of TPT on GQ outweighs the electrostatic stabilisation for the groove-bound state such that the total energy is more favourable for the terminal bound state (Figure 3B.8).
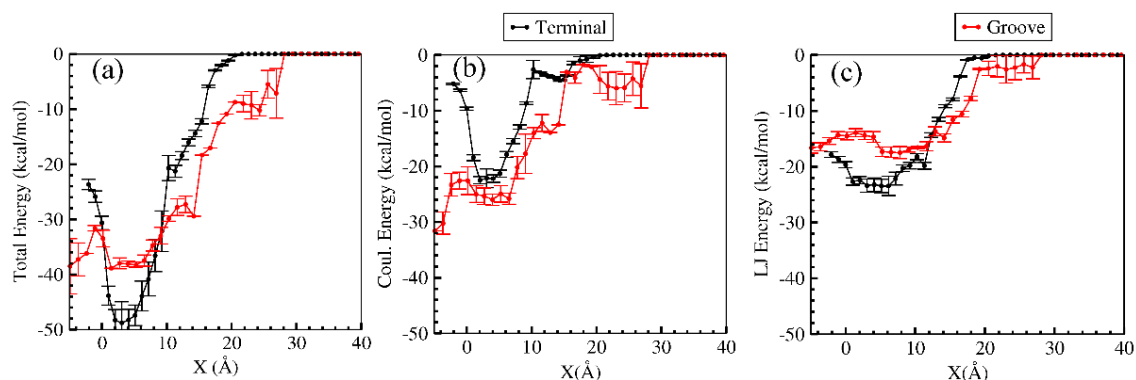
Figure 3B.8 Interaction energy between TPT and GQ along unbinding pathways for the terminal (black) and groove-bound (red) states. (a) Total interaction energy between TPT and GQ and its decomposition into (b) electrostatic energy (Coulomb interaction) and (c) van der Waals components. The interaction energy is averaged over two metadynamics simulations. Error bars are shown.

3B.3 Plausible Mechanism for the Reversible Conformational Transition

Mechanism of conformational transitions can be explained with the help of MD simulation studies. Here, in both terminal and groove bound states, the ammonium alkyl (-$N(CH_3)_2$) and its adjacent hydroxyl group of TPT are projected out from the G-tetrad rather than facing inwards (Appendix 3B.5). In this scenario, the planar heteroaromatic moiety of TPT provides the additional stabilisation to the G-quartet through van der Waals interaction, which has also been observed in MD simulation studies (Figure 3B.8c). During reverse conformational transition (GQ-DNA→H24), electron dense carbonyl portals of CB7 probably interacts with the

positively charged ammonium alkyl group of TPT, which leads to an inclusion complex formation resulting in the translocation of the drug from GQ to CB7.

3B.4 Conclusion

Firstly, we have shown here that topotecan can induce the GQ formation and stabilisation even in the absence of Na^+/K^+ ion. This conformational transition, i.e. H24→GQ-DNA, has been monitored through CD, NMR and melting studies. MD simulation studies show that TPT preferably binds to the terminal of GQ, and the second preferable binding mode is the groove of GQ. The more rigorous free energy calculation using well-tempered metadynamics shows that binding to the terminal is 4 kcal/mol stronger than the free energy of binding to the groove. Although both the electrostatic and van der Waals interaction contribute to both these binding modes, the former is prevalent for the terminal binding whereas the latter is stronger for the groove-binding. The overall total energy is, however, more negative for the terminal binding mode.

Next, this topotecan mediated GQ formation and stabilisation has been regulated by incorporating a molecular container, cucurbit[7]uril (CB7) in the system by means of forming an inclusion complex with the GQ stabiliser, i.e., topotecan. This reverse conformational transition, i.e., GQ-DNA→H24, has been again probed through the CD, NMR and melting measurements. Moreover, both these conformational transitions (H24→GQ-DNA and GQ-DNA→H24) can be easily monitored through fluorescence emission colour of TPT, which changes from green (H24-TPT system) to violet (H24-TPT-CB7 system). Ease of detection in this approach makes it more advantageous in comparison to other methods, which is generally associated with sophisticated and complex detection techniques hindering their usefulness for its real-time application.

3B.5 Appendix Section
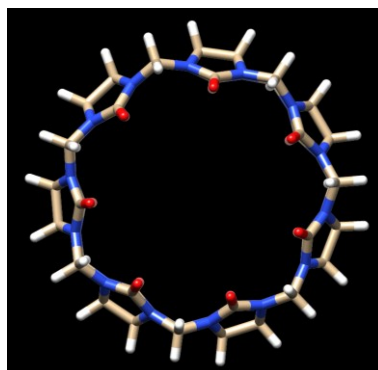
**Appendix 3B.1** Molecular structure of cucurbit[7]uril (CB7).
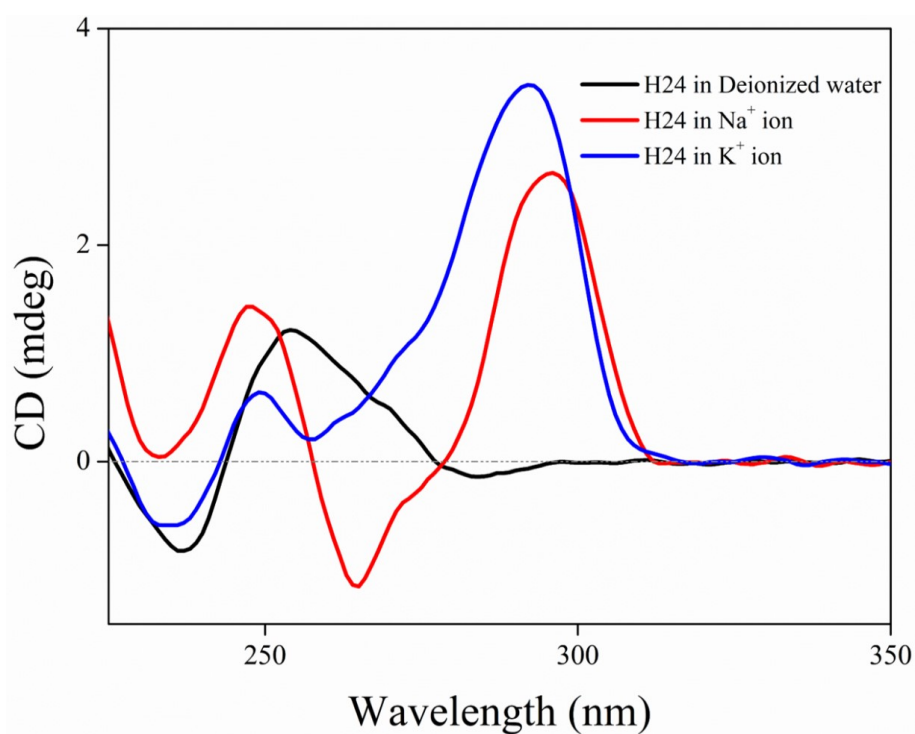
Appendix 3B.2 Circular dichroism spectra of H24 DNA (~ 2 μM) under different conditions. The K^+ ion and Na^+ ion legends indicate the presence of 100 mM KCl and NaCl salt, respectively, in 10 mM tris buffer solution (pH 7.2).

Appendix Note 3B.1

For static quenching, the binding constant (K) and the number of binding sites per host (n) can be determined according to the Scatchard equation[23]

$$\log\left(\frac{F_0 - F}{F}\right) = logK + nlog[Q]$$

Where, F_0 and F are the relative fluorescence intensities of TPT in the absence and presence of the H24 DNA, respectively. $[Q]$ is the concentration of H24 DNA in the course of the experiment. K and n can be determined from the slope and intercept value of log((F_0-F) /F)) versus log [Q] plot.
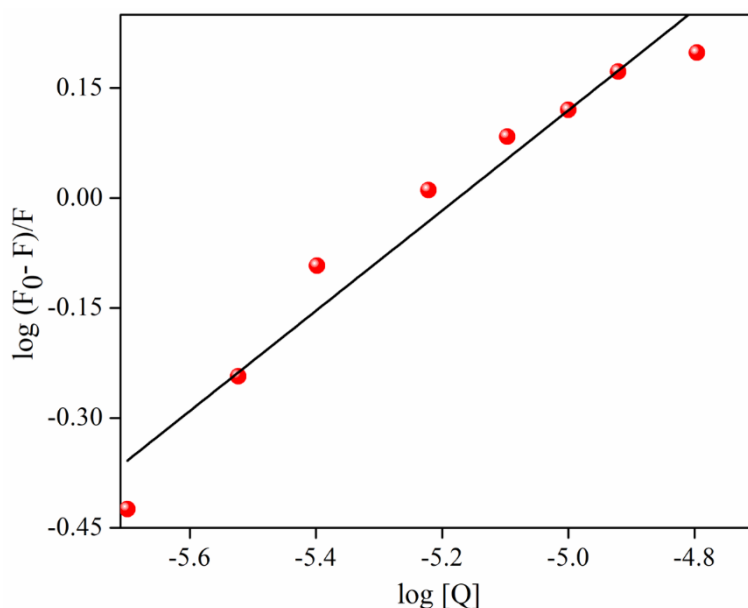

Appendix 3B.3 Scatchard plot of log((F_0-F) /F)) versus log [Q] for the binding constant determination of H24-TPT complex.

Binding constant (K) for the H24-TPT complex is found to be 3451.4 M^{-1}, and the number of binding sites of TPT per DNA (n) is 0.68 which is very close to 3 TPT molecules per 2 H24 DNA. This binding stoichiometry (n) is found to be similar to the binding of GQ DNA with other drugs such as ellipticine,[24] proflavine[25] etc. The higher binding constant value of TPT-CB7 complex (K = 5000 M^{-1}) makes the translocation of TPT from GQ DNA to CB7 nano-cavity feasible.
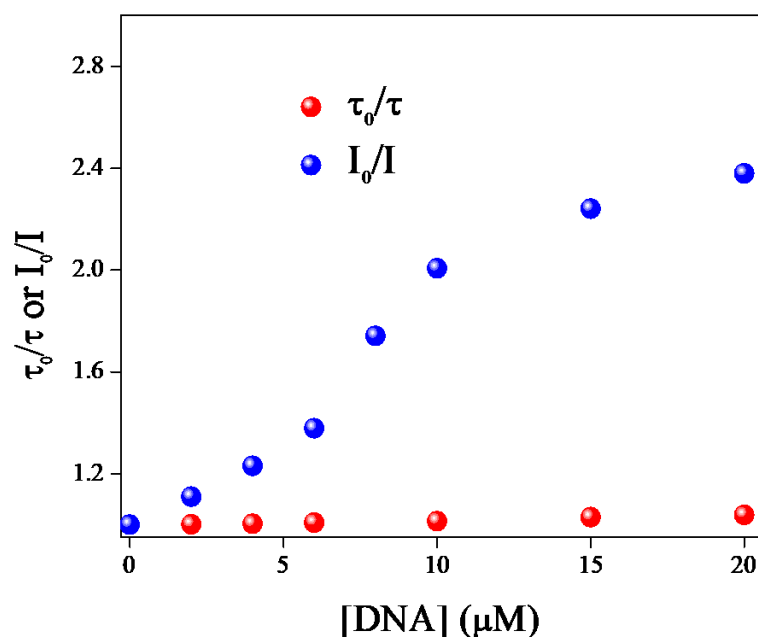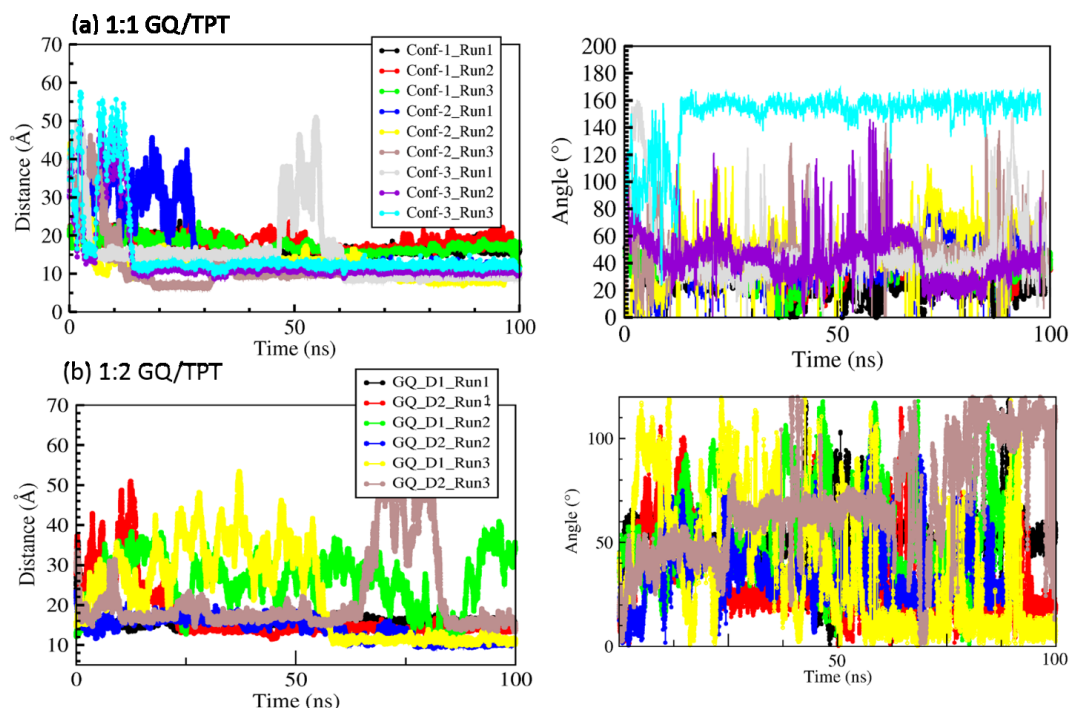
Appendix 3B.4 Static quenching of zwitterionic form of TPT (Z^*) (emission and lifetime collected at 530 nm by exciting at 375 nm) in the presence of H24 DNA.
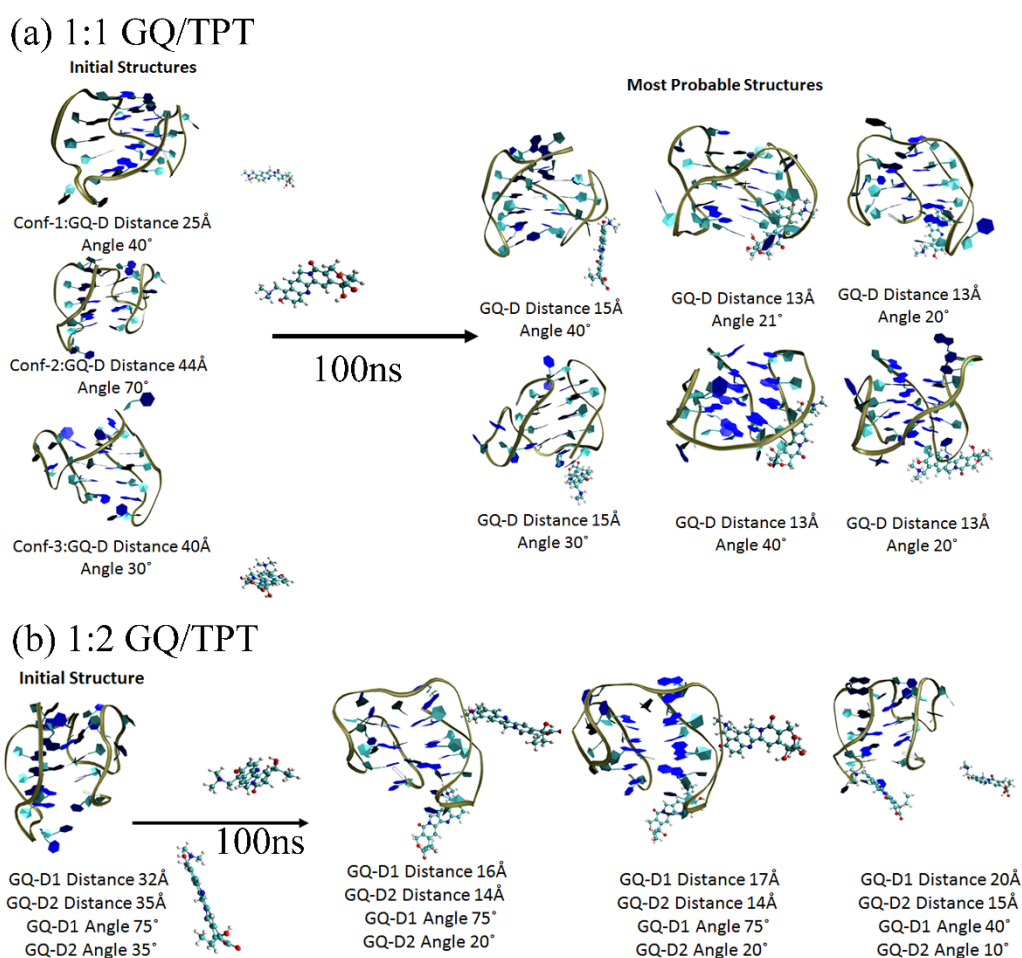

Appendix 3B.5 Initial structure and most probable structures obtained after 100 ns of normal MD simulation (a) 1:1 GQ/TPT (b) 1:2 GQ/TPT. "D" denotes the TPT in (a) and "D1", and "D2" are two TPT molecules in (b).

Appendix 3B.6 Variation of distance (between COM's of GQ and TPT) and angle (between GQ axis vector and COM of GQ to COM TPT vector; see simulation method section for detail) with time from an initial random distribution of TPT in (a) 1:1 GQ/TPT complex (b) 1:2 GQ/TPT complex. The left column shows the variation of distances and the right column shows the change in angle. Three random configurations (Conf-1, Conf-2, Conf-3) were created for 1:1 complex while one random configuration was created for 1:2 complex. Each random configuration was simulated three times with different velocity distribution (denoted as Run1, Run2 and Run3). Colour code is the same for distance and angle in 1:1 GQ/TPT complex and 1:2 GQ/TPT complex.
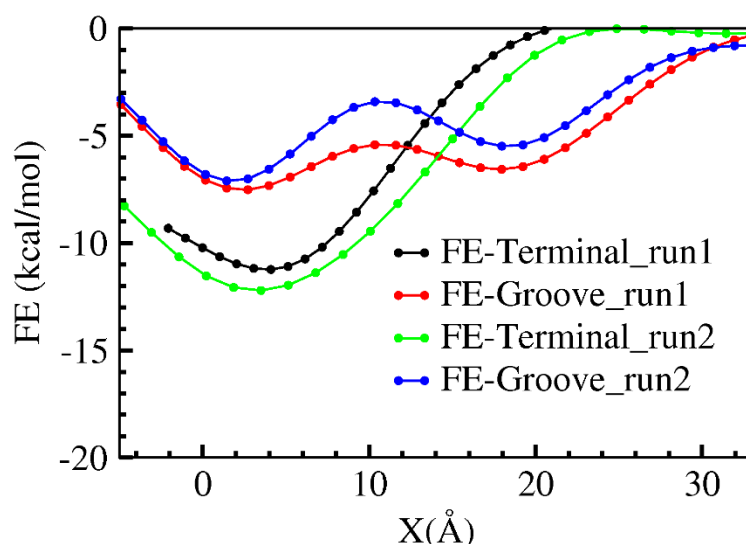
Appendix 3B.7 Free energy of unbinding of TPT from GQ obtained from two metadynamics simulations denoted as run1, run2. Black and green colour represent the free energy profiles for the terminal-bound state, while red and blue colour represents the free energy profiles for the groove-bound state.
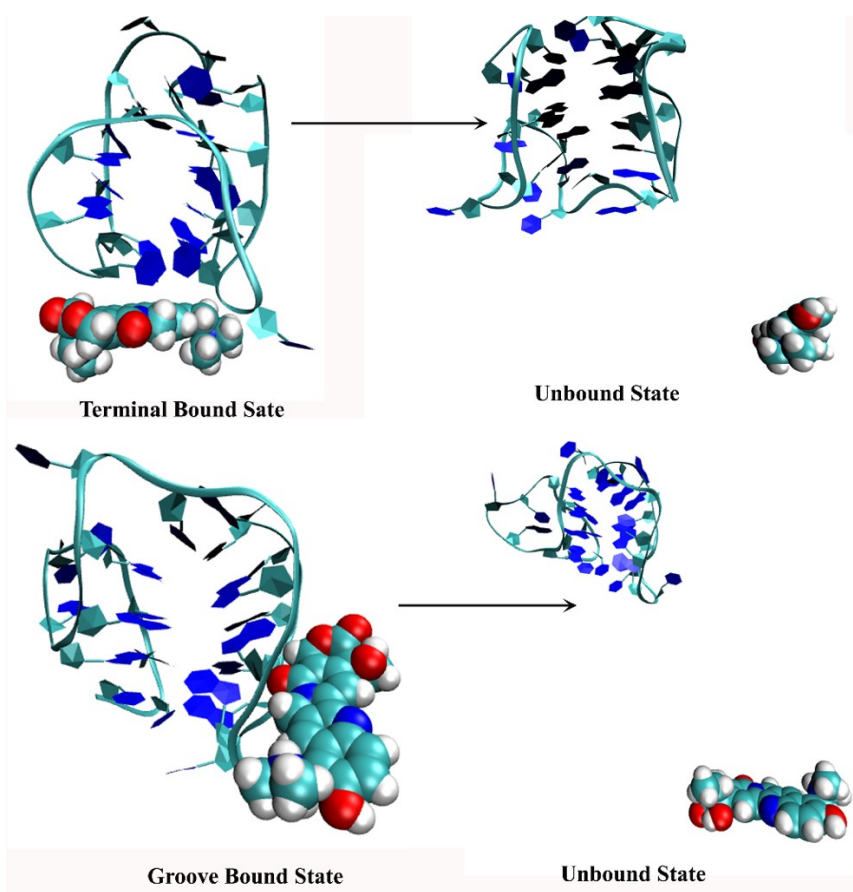
Appendix 3B.8 Bound and unbound states of GQ and TPT obtained from well-tempered metadynamics simulation (obtained from the first metadynamics run).
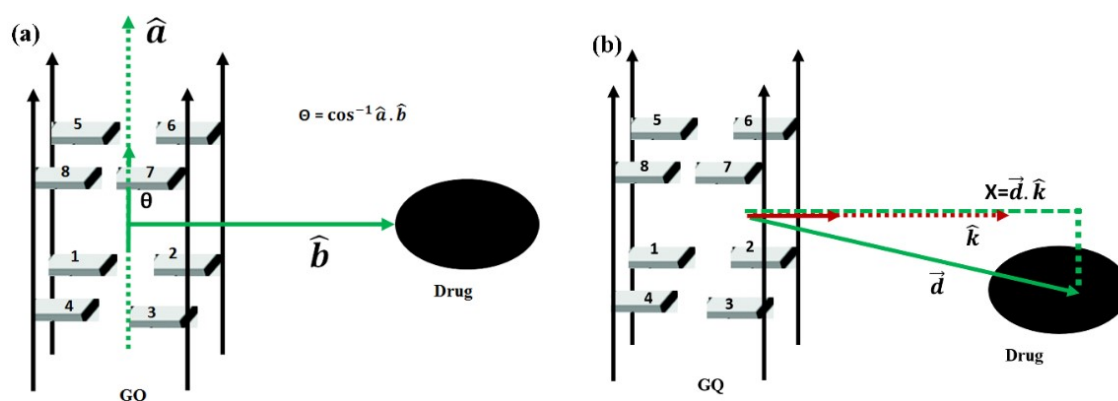
Appendix 3B.9 Definition of angle and reaction coordinate X using the schematic drawing of the GQ. (a) Angle θ is defined using two vectors $\hat{a}$ and $\hat{b}$. Vector $\hat{a}$ (GQ axis) is constructed from COM of bases 1 to 4 to COM of base 5-6-7-8 and vector $\hat{b}$ goes from the COM of GQ bases 1 to 8 to the COM of drug. the Numbers 1-8 are just for schematic representation. When drug is bound to groove of the GQ (side binding), the angle would be ~90° and when drug bound to terminal of GQ, the angle would be ~0°. (b) Reaction coordinate X defined by two vectors $\hat{k}$ and $\vec{d}$ as $X = \hat{k}.\vec{d}$. Vector $\vec{d}$ is goes from COM of GQ bases 2, 3, 6, 7 to COM of TPT and vector $\hat{k}$ goes from the COM of GQ bases 2,7 to the COM of 2-3-6-7 bases. Note that the numbers 1-8 are just for schematic representation and does not reflect the actual base numbering in the GQ. Same way, we have defined for terminal bound state, where we have considered 1-2-3-4 bases for definition. X would be ~0, when the drug is bound to either groove or the terminal site of GQ.

Appendix Table 3B.1 Fluorescence transient fittings (collected at 530 nm) of TPT (Z*) in deionized water (pH 6.2), H24 DNA (20 μM) and in the presence of both H24 DNA and CB7 (1.5 mM).

Sample	a_1	τ_1 (ns)	a_2	τ_2 (ns)	a_3	τ_3 (ns)	τ_{avg}[#] (ns)	χ^2	λ_{ex} (nm)	λ_{col} (nm)
TPT in water	-	-	-	-	1	5.88	5.88	1.04	375	530
TPT + H24 2 μM	-	-	0.04	2.71	0.96	5.95	5.83	1.08	375	530
TPT + H24 4 μM	-	-	0.06	1.96	0.94	5.94	5.71	1.04	375	530
TPT + H24 20 μM	-	-	0.10	2.69	0.90	6.01	5.66	0.99	375	530
TPT + H24 CB 7 100 μM	-	-	0.17	0.93	0.83	5.84	5.00	1.06	375	530
TPT + H24 CB 7 560 μM	-	-	0.16	1.31	0.84	5.69	4.98	1.07	375	530
TPT + H24 CB 7 1.5 mM	-	-	0.22	1.88	0.78	5.28	4.54	1.09	375	530

[#]$\tau_{avg} = a_1\tau_1 + a_2\tau_2 + a_3\tau_3$;

Appendix Table 3B.2 Fluorescence transient fittings (collected at 430 nm) of TPT (C*) in deionized water (pH 6.2), H24 DNA (20 μM) and in the presence of both H24 DNA and CB7 (1.5 mM).

Sample	a_1	τ_1 (ns)	a_2	τ_2 (ns)	a_3	τ_3 (ns)	τ_{avg}[#] (ns)	χ^2	λ_{ex} (nm)	λ_{col} (nm)
TPT in water	0.99	0.038	-	-	0.01	0.837	0.042	0.99	375	430
TPT + H24 CB 7 100 μM	0.72	0.049	0.12	0.56	0.16	2.16	0.452	1.01	375	430
TPT + H24 CB 7 560 μM	0.41	0.063	0.11	0.65	0.48	2.18	1.15	1.002	375	430
TPT + H24 CB 7 1.5 mM	0.22	0.18	0.09	0.94	0.68	2.19	1.62	1.09	375	430

[#]$\tau_{avg} = a_1\tau_1 + a_2\tau_2 + a_3\tau_3$;

Appendix Table 3B.3 Force-field parameter of TPT molecule.

Sigma and Epsilon of each atom.

Atom Type	Sigma (σ) nm	Epsilon (ε) kJ/mol
HO	0.00000e+00	0.00000e+00
N3	3.25000e-01	7.11280e-01
NC	3.25000e-01	3.25000e-01
CB	3.39967e-01	3.59824e-01
N*	3.25000e-01	7.11280e-01
OS	3.00001e-01	7.11280e-01
CT	3.39967e-01	4.57730e-01
C	3.39967e-01	3.59824e-01
OH	3.06647e-01	8.80314e-01
CA	3.39967e-01	3.59824e-01
O	2.95992e-01	8.78640e-01
H1	2.64953e-01	6.56888e-02
HA	2.64953e-01	6.56888e-02
HC	2.64953e-01	6.56888e-02
H	1.06908e-01	6.56888e-02
HP	2.64953e-01	6.56888e-02

Charges on each atom

Atom Type	Atom	Charge
NC	N	-0.63905
CB	C	0.62518
CB	C	-0.12025
CA	C	-0.24053
HA	H	0.19936
CA	C	0.06023
CT	C	0.28472
CT	C	0.03754
CT	C	-0.23287

HC	H	0.07057
HC	H	0.07057
HC	H	0.07057
HC	H	0.01456
HC	H	0.01456
OH	O	-0.64834
HO	H	0.42637
C	C	0.70340
O	O	-0.56104
OS	O	-0.43894
CT	C	0.16112
H1	H	0.08559
H1	H	0.08559
CA	C	-0.16967
C	C	0.56964
O	O	-0.61127
N*	N	-0.07541
CT	C	-0.12097
H1	H	0.14311
H1	H	0.14311
CB	C	-0.12129
CA	C	-0.17380
HA	H	0.18987
CA	C	-0.12615
CA	C	0.41327
CA	C	-0.12886
HA	H	0.20014
CA	C	-0.38272
HA	H	0.21738
CA	C	0.36839
OH	O	-0.66510

HO	H	0.49419
CA	C	-0.09885
CT	C	-0.05269
HP	H	0.12940
HP	H	0.12940
N3	N	-0.01498
CT	C	-0.31243
HP	H	0.16667
HP	H	0.16667

3B.6 Experimental

All experiments and sample preparations were carried out in deionised water. Prior to the experiments H24-DNA was annealed at 90 °C for 10 min and stored at 4 °C for 24 h. The concentration of H24-DNA was determined using the molar extinction coefficient of ~244 600 M^{-1} cm^{-1} at 260 nm provided by IDT, USA. It is necessary to mention here that IDT has determined the molar extinction coefficient using the nearest neighbour approximation model. The concentration of TPT was determined using the molar extinction coefficient of ~200 000 M^{-1} cm^{-1} at 380 nm.[18d] All the NMR samples were prepared using 90% deionised water (devoid of any ions) and 10% D_2O having H24 DNA concentration of 50 µM. Then, TPT (150 µM) was added in the same solution for the GQ DNA formation. To revert the structure to single strand/random coil, 1.5 mM CB7 was added in the H24-TPT system (H24: TPT = 1:3). To monitor the peaks in between 10-13 ppm, water suppression was performed. We fitted these lifetime decay profiles with a minimum number of exponentials. Quality of each fit was judged by χ^2 values and the visual inspection of the residuals. The value of $\chi^2 \approx 1$ was considered as best fit for the plots. For thermal melting, we used a very small concentration of DNA (~1 µM) to avoid the absorbance saturation. The melting temperature (T_m) was determined from the sigmoidal curve fit of the melting profile. For isothermal titration calorimetry (ITC) measurements, the titration of TPT against H24 DNA was performed in 20 injections (2 µL of each) into a solution with fixed H24 DNA (200 µL) concentration in the cell with 200 s resting time between two consecutive injections. To eliminate the dilution effect, a blank experiment was also carried out by injecting the same concentration of TPT into deionised water under identical experimental condition. Using a single-site binding model and nonlinear least-squares fitting algorithm method, generated isotherm was analysed to yield the relevant thermodynamic parameter.

3B.6.1 Simulation Methods

The structure of TPT was generated in GVIEW software. The structure optimisation and electron densities were calculated quantum mechanically using HF theory with 6-31G* basis set in GAUSSIAN03 software.[26] Using this, RESP charges of TPT atoms were calculated using ANTECHAMBER module of AMBER11. For all other force-field parameters for TPT, GAFF force-field was followed.[27] The topology and co-ordinates generated using AmberTools[28] were converted into GROMACS format by using a perl program amb2gmx.pl.[29]

The starting structure of hybrid G-quadruplex (3+1) was taken from the protein data bank (PDB ID 2GKU).[21] The topology and coordinates of GQ were generated by using GROMACS.[30] Amber94 force-field[31] was used for GQ. We created three systems: (i) GQ in water (ii) 1:2 GQ:TPT complex in water, and (iii) 1:1 GQ:TPT complex in water. Each system was solvated by ~23000 TIP3P water molecules[27] in a cubic box of dimension 90 Å. Physiological concentration (150 mM) of K^+ and Cl^- ions as well as extra K^+ ion, were used to neutralise the system.

 All the simulations were performed using molecular dynamics software GROMACS-4.5.5.[30] Initially; the system was minimised using the steepest descent method,[32] followed by heating it to 300K in 100 ps using Berendsen thermostat[33] with a coupling constant of 0.2 ps. Restraint of 25 kcal/mol/Å^2 was applied on heavy atoms of GQ during the heating process. The harmonic restraint was gradually reduced to 0.25 kcal/mol in six steps following the protocol used in standard DNA simulations.[34] In each step, 100 ps equilibration was carried out at constant temperature (300 K) and pressure (1 bar) using Berendsen thermostat and barostat[33] with coupling constants of 0.2 ps each, followed by energy minimisation using the steepest descent method.[32] During the simulation, LINCS algorithm[35] was used to constrain all the bonds. Particle Mesh Ewald (PME) method[36] was used for electrostatics. The distance cut-offs for the van der Waals (vdW) and electrostatic long-range interaction was kept at 10 Å. The time step for each simulation was taken to be 2 fs. At the final equilibration step, we performed the 2 ns unrestrained equilibration by using the Berendsen thermostat and barostat[33] with a coupling constant of 0.2 ps. After the equilibration steps, multiple normal molecular dynamics simulations were performed to see the binding of TPT with GQ. Three random configurations of 1:1 GQ/TPT complexes were created, where TPT molecule was placed far away from the GQ. For each of these three different configurations, three 100 ns simulations were performed. Appendix 3B.5a shows the initial structure and the most probable structures after the simulation. For 1:2 GQ/TPT complex, three simulations of 100 ns were performed with different velocity distribution from a configuration, where both the TPT molecules are far away from the GQ. Fig. S5b (in SI) shows the initial structure for this system. Thus, a total of 1.2 μs simulation was performed to study the TPT-GQ binding. During this final set of simulations, the temperature was set at 300 K, and the pressure was set at 1 bar using the Nose-Hoover thermostat[37] and Parrinello-Rahman barostat,[38] respectively, with 0.2 ps coupling constant.

3B.6.2 Definition of Angle Used in Analysis:

Schematic representation of the angle between GQ and TPT is shown in Appendix 3B.9a. We have used two vectors: the vector $\hat{a}$ was constructed from COM of G16, G10, G4, G22 to COM of G17, G9, G3, G21 and vector $\hat{b}$ was constructed from COM of GQ to the COM of the drug. The angle of defined as, $\theta = \cos^{-1} \hat{a} \cdot \hat{b}$. Vector $\hat{a}$ roughly shows the GQ axis. When drug is bound to groove of GQ (side binding), the angle would be high, and when drug bound to terminal of GQ, the angle would be low.

3B.6.3 Metadynamic Simulation Details:

To calculate the binding free energy of TPT and GQ, well-tempered metadynamics[39] simulation was performed along X (reaction coordinate). This reaction coordinate was similar to what was defined for DNA-Drug binding/unbinding mechanism by Wilbee et al.[40] created to measure the distance between the drug and DNA perpendicular to the DNA helical axis. Here also, *X* approximately measures the distance of TPT perpendicular to the GQ axis in Appendix 3B.9b. Appendix 3B.9b shows the construction of the reaction coordinate *X* schematically. Metadynamics simulations were performed for unbinding of TPT: (i) unbinding from terminal bound complex, (ii) unbinding from groove bound complex. To check the reproducibility of the well-tempered metadynamics simulation, the same metadynamics simulation was repeated with different parameters. Gaussian potential of 0.2 kJ/mol height and 0.7Å width were used for both simulations. The potential disposition rate 2 ps for the first and 1 ps for the second simulation. The bias factor was set to 10 for both simulations. We have stopped the simulation when unbinding happened. It took a total of 30 ns for the first and 25 ns for the second metadynamics simulation. Thus, a total of 55 ns simulation was performed for the metadynamics simulation.

3B.7 References

1. (a) S. Balasubramanian, L. H. Hurley and S. Neidle, *Nat Rev Drug Discov*, 2011, **10**, 261-275; (b) A. M. Zahler, J. R. Williamson, T. R. Cech and D. M. Prescott, *Nature*, 1991, **350**, 718-720.

2. (a) T.-m. Ou, Y.-j. Lu, J.-h. Tan, Z.-s. Huang, K.-Y. Wong and L.-q. Gu, *ChemMedChem*, 2008, **3**, 690-713; (b) R. E. Verdun and J. Karlseder, *Nature*, 2007, **447**, 924-931.

3. N. W. Kim, M. A. Piatyszek, K. R. Prowse, C. B. Harley, M. D. West, P. L. Ho, G. M. Coviello, W. E. Wright, S. L. Weinrich and J. W. Shay, *Science*, 1994, **266**, 2011.

4. (a) J. L. Huppert, *Chem. Soc. Rev.*, 2008, **37**, 1375-1384; (b) G. N. Parkinson, M. P. H. Lee and S. Neidle, *Nature*, 2002, **417**, 876-880.

5. (a) A. M. Burger, F. Dai, C. M. Schultes, A. P. Reszka, M. J. Moore, J. A. Double and S. Neidle, *Cancer Res.*, 2005, **65**, 1489; (b) J. Mohanty, N. Barooah, V. Dhamodharan, S. Harikrishna, P. I. Pradeepkumar and A. C. Bhasikuttan, *J. Am. Chem. Soc.*, 2013, **135**, 367-376; (c) R. Rodriguez, G. D.

Pantoş, D. P. N. Gonçalves, J. K. M. Sanders and S. Balasubramanian, *Angew. Chem. Int. Ed.*, 2007, **46**, 5405-5407.

6. (a) D. J. M. Blanchard, K. L. Fadock, M. Sproviero, P. S. Deore, T. Z. Cservenyi, R. A. Manderville, P. Sharma and S. D. Wetmore, *J. Mater. Chem. C*, 2016, **4**, 2915-2924; (b) M. Sproviero, K. L. Fadock, A. A. Witham, R. A. Manderville, P. Sharma and S. D. Wetmore, *Chem. Sci.*, 2014, **5**, 788-796; (c) K. L. Fadock and R. A. Manderville, *ACS Omega*, 2017, **2**, 4955-4963; (d) R. A. Manderville and S. D. Wetmore, *Chem. Sci.*, 2016, **7**, 3482-3493.

7. O. Mendoza, A. Bourdoncle, J.-B. Boulé, R. M. Brosh and J.-L. Mergny, *Nucleic Acids Res.*, 2016, DOI: 10.1093/nar/gkw079.

8. K. Paeschke, M. L. Bochman, P. D. Garcia, P. Cejka, K. L. Friedman, S. C. Kowalczykowski and V. A. Zakian, *Nature*, 2013, **497**, 458-462.

9. Y. Wu, K. Shin-ya and R. M. Brosh, *Mol. Cell. Biol.*, 2008, **28**, 4116-4128.

10. (a) S. Ogasawara and M. Maeda, *Angew. Chem. Int. Ed.*, 2009, **48**, 6671-6674; (b) D. Monchaud, P. Yang, L. Lacroix, M.-P. Teulade-Fichou and J.-L. Mergny, *Angew. Chem. Int. Ed.*, 2008, **47**, 4858-4861; (c) Y. Shi, H. Sun, J. Xiang, H. Chen, S. Zhang, A. Guan, Q. Li, S. Xu and Y. Tang, *Chem. Commun.*, 2016, **52**, 7302-7305.

11. J.-L. Yao, X. Gao, W. Sun, X.-Z. Fan, S. Shi and T.-M. Yao, *Inorg. Chem.*, 2012, **51**, 12591-12593.

12. (a) F. Li, Y. Xu, H. Li, C. Wang, A. Lu and S. Sun, *New J. Chem.*, 2014, **38**, 1396-1400; (b) P. Kundu, S. Ghosh, S. Das and N. Chattopadhyay, *Phys. Chem. Chem. Phys.*, 2016, **18**, 3685-3693; (c) D. V. Berdnikova, T. M. Aliyeu, T. Paululat, Y. V. Fedorov, O. A. Fedorova and H. Ihmels, *Chem. Commun.*, 2015, **51**, 4906-4909.

13. (a) X. Yang, D. He, J. Cao, X. He, K. Wang and Z. Zou, *RSC Adv.*, 2015, **5**, 84553-84559; (b) P. Galer, B. Wang, P. Šket and J. Plavec, *Angew. Chem. Int. Ed.*, 2016, **55**, 1993-1997; (c) C. Riccardi, I. Russo Krauss, D. Musumeci, F. Morvan, A. Meyer, J.-J. Vasseur, L. Paduano and D. Montesarchio, *ACS Appl. Mater. Interfaces*, 2017, **9**, 35574-35587; (d) D. M. Engelhard, R. Pievo and G. H. Clever, *Angew. Chem. Int. Ed.*, 2013, **52**, 12843-12847.

14. T. Tian, Y. Song, L. Wei, J. Wang, B. Fu, Z. He, X.-R. Yang, F. Wu, G. Xu, S.-M. Liu, C. Li, S. Wang and X. Zhou, *Nucleic Acids Res.*, 2017, **45**, 2283-2293.

15. (a) D. Mirchandani, H. Hochster, A. Hamilton, L. Liebes, H. Yee, J. P. Curtin, S. Lee, J. Sorich, C. Dellenbaugh and F. M. Muggia, *Am. Assoc. Cancer Res.*, 2005, **11**, 5912-5919; (b) Y. Pommier, *Nat. Rev. Cancer*, 2006, **6**, 789-802.

16. C. Leonetti, M. Scarsella, G. Riggio, A. Rizzo, E. Salvati, M. Incalci, L. Staszewsky, R. Frapolli, M. F. Stevens, A. Stoppacciaro, M. Mottolese, B. Antoniani, E. Gilson, G. Zupi and A. Biroccio, *Clin. Cancer Res.*, 2008, **14**, 7284.

17. F. Li, J. Zhou, M. Xu and G. Yuan, *Int. J. Biol. Macromol.*, 2017, DOI: https://doi.org/10.1016/j.ijbiomac.2017.10.010.

18. (a) K. Gavvala, A. Sengupta, R. K. Koninti and P. Hazra, *ChemPhysChem*, 2013, **14**, 3375-3383; (b) M. R. di Nunzio, Y. Wang and A. Douhal, *J. Photochem. Photobiol., A: Chem*, 2013, **266**, 12-21; (c) R. K. Koninti, K. Gavvala, A. Sengupta and P. Hazra, *J. Phys. Chem. B*, 2015, **119**, 2363-2371; (d) H. Joshi, A. Sengupta, K. Gavvala and P. Hazra, *RSC Adv.*, 2014, **4**, 1015-1024; (e) S. A. Strel'tsov, A. L. Mikheikin, S. L. Grokhovsky, V. A. Oleinikov and A. L. Zhuze, *Mol. Biol.*, 2002, **36**, 400-411; (f) S. Streltsov, V. Oleinikov, M. Ermishov, K. Mochalov, A. Sukhanova, Y. Nechipurenko, S. Grokhovsky, A. Zhuze, M. Pluot and I. Nabiev, *Biopolymers*, 2003, **72**, 442-454.

19. (a) W. J. M. Underberg, R. M. J. Goossen, B. R. Smith and J. H. Beijnen, *J. Pharm. Biomed. Anal.*, 1990, **8**, 681-683; (b) M. R. di Nunzio, Y. Wang and A. Douhal, *J. Phys. Chem. B*, 2012, **116**, 7522-7530; (c) M. R. di Nunzio, Y. Wang and A. Douhal, *J. Phys. Chem. B*, 2012, **116**, 8182-8190.

20. (a) E. I. Cucolea, C. Tablet, H.-J. Buschmann and L. Mutihac, *J. Inclusion Phenom. Macrocyclic Chem.*, 2015, **83**, 103-110; (b) Y. Huang, S.-F. Xue, Q.-J. Zhu and T. Zhu, *Supramol. Chem.*, 2008, **20**, 279-287; (c) N. S. Venkataramanan and A. Suvitha, *J. Phys. Chem. B*, 2017, **121**, 4733-4744.

21. K. N. Luu, A. T. Phan, V. Kuryavyi, L. Lacroix and D. J. Patel, *J. Am. Chem. Soc.*, 2006, **128**, 9963-9970.

22. K. Hyz, R. Kawęcki, E. Bednarek, W. Bocian, J. Sitkowski and L. Kozerski, *Magn. Reson. Chem.*, 2010, **48**, 575-584.

23. (a) A. Chandra, K. Singh, S. Singh, S. Sivakumar and A. K. Patra, *Dalton Transactions*, 2016, **45**, 494-497; (b) L. Chen, J. Zhang, Y. Zhu and Y. Zhang, *RSC Adv.*, 2015, **5**, 79874-79881.

24. S. Ghosh, A. Kar, S. Chowdhury and D. Dasgupta, *Biochemistry*, 2013, **52**, 4127-4137.

25. V. Kumar, A. Sengupta, K. Gavvala, R. K. Koninti and P. Hazra, *J. Phys. Chem. B*, 2014, **118**, 11090-11099.

26. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Journal*, 2016.

27. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157-1174.

28. (a) D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668-1688; (b) W. D. Cornell, P. Cieplak, C. I. Bayly and P. A. Kollmann, *J. Am. Chem. Soc.*, 1993, **115**, 9620-9631.

29. E. J. Sorin and V. S. Pande, *Biophys. J.*, 2005, **88**, 2472-2493.

30. B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435-447.

31. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179-5197.

32. W. H. T. Press, S. A.; Vetterling, W. T.; Flannery, B. P. , *Numerical Recipes in FORTRAN; The Art of Scientific Computing*, Cambridge University Press, New York, 3rd edn., 1992.

33. H. J. C. Berendsen, J. P. M. Postma, W. F. v. Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684-3690.

34. S. B. Dixit, D. L. Beveridge, D. A. Case, T. E. Cheatham, E. Giudice, F. Lankas, R. Lavery, J. H. Maddocks, R. Osman, H. Sklenar, K. M. Thayer and P. Varnai, *Biophys. J.*, 2005, **89**, 3721-3740.

35. B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463-1472.

36. T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089-10092.

37. (a) S. Nosé, *Mol. Phys.*, 1984, **52**, 255-268; (b) W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695-1697.

38. M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182-7190.

39. A. Barducci, G. Bussi and M. Parrinello, *Phys. Rev. Lett.*, 2008, **100**, 020603.

40. W. D. Sasikala and A. Mukherjee, *J. Phys. Chem. B*, 2012, **116**, 12208-12212.

Chapter 4A

Silica Nano-channel Induced i-motif Formation and Stabilization at Neutral and Alkaline pH

4A.1 Introduction

Recently, C-quadruplex i.e. i-motif DNA has been celebrated as a budding topic of interest among researchers due to its potential relevance in oncogene.[1] In slightly acidic pH, i-motif DNA forms through the Hoogsteen hydrogen bonds between hemi-protonated cytosine-cytosine (C-C$^+$) base pairs leading to the combination of two parallel duplexes intercalated in an antiparallel fashion (Scheme 4A.1).[2] In neutral and alkaline pH, cytosine bases (pK$_a$ 4-5) deprotonate, converting the stable i-motif structure to single stranded C-rich sequence.[3] i-motif DNA, being complementary of G-quadruplex (GQ) sequence, has been found in gene promoter regions,[1, 4] but its biological significance remains debatable till date. However, recent reports suggest the controlling effects of i-motif DNA on the transcription regulation.[5] In some cases, i-motif DNA even acts as a more potent inhibitor towards DNA replication than GQ.[6] Recently, i-motif DNA formation in the nuclei of human cells has been reported.[7] In spite of the debate regarding the biological relevance, researchers have developed many applications based on i-motif structures like, pH sensors,[8] delivery systems,[9] ATP sensor,[10] memory devices,[11] switchable nanostructures[12] etc. Despite versatile applications, it faces serious limitation due to its unstable structure in neutral and alkaline pH. Thus, researchers have tried to stabilize i-motif DNA at neutral and alkaline pH using graphene quantum dot,[13] carbon nanotube,[14] metal ion,[15] crowding agents,[15c, 16] modified nucleobases,[17] etc. However, most of the i-motif stabilizers (such as carbon nanotube) lack an ordered large porous network to encapsulate DNA, which is responsible for the controlled and reversible i-motif formation. Thus, controlling i-motif structure and reversibility of its formation is important and rare in literature. Nevertheless, most of the i-motif stabilizers are less biodegradable in nature, and researchers are still making a constant effort to explore new biocompatible materials for greater relevance in biological and biomaterials applications. In continuation of these efforts, we have demonstrated the stabilization of i-motif DNA at neutral and alkaline pH using mesoporous silica nano-channels (MCM-41; Scheme 4A.1), which are well explored in biological and biomedical studies due to their water dispersibility, low toxicity, and cell-penetrating ability.[18]

MCM-41 has attracted researchers due to its large ordered pore structure, adjustable pore size and volume.[18] MCM-41 consists of ordered –Si–O– framework arranged in hexagonal channels with a relatively uniform diameter ranging from ~10 to 200 Å.[18] Recently, functionalized C-rich sequences in MCM-41 have been employed for photon-fueled gate-like delivery system by changing the medium pH,[19] but stabilization of i-motif DNA structures based on mesoporous silica nano-particles at neutral and alkaline pH has not been explored till now. Here, we have utilized MCM-41 as a stabilizer of i-motif DNA structure in neutral and alkaline medium. To check universality of this approach, C-rich sequences of various chain lengths

found in different oncogene promoters (e.g., telomere, C-Myc, bcl-2, hTERT etc.; Table 4A.1) have been used to stabilize the i-motif structures at neutral and alkaline pH differing in molecularity, size and loop length. This MCM-41 induced i-motif DNA formation from single
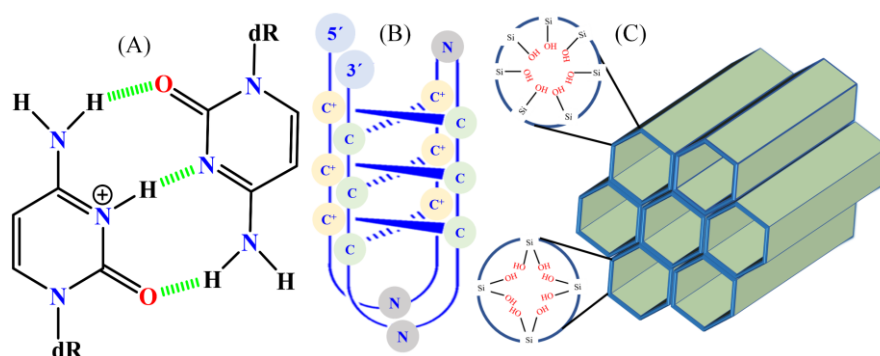
Scheme 4A.1 (A) hemi-protonated cytosine–cytosine hoogsteen base pairing in i-motif DNA. Chemical structure of (B) i-motif DNA and (C) MCM-41.

stranded/random coil (ssDNA/RC) have been monitored using circular dichroism (CD) spectroscopy, steady state and time-resolved fluorescence measurements. Moreover, we have reverted this i-motif formation by a positively charged protein. Thus, a simple and biologically relevant reversible approach has been developed to stabilize the i-motif DNA in neutral and alkaline medium. We believe that the proposed strategy may hold immense potential for future applications of i-motif structure at neutral and alkaline pH.

Table 4A.1 Different cytosine rich DNA sequences used in this study.

DNA	Sequences (5´→ 3´)
h-teloc	d(CCCTAACCCTAACCCTAACCCTAA)
TC5	d(TCCCCC)
C-Myc	d(CCCCACCTTCCCCACCCTCCCCACCCTCCCC)
bcl-2	d(CAGCCCCGCTCCCGCCCCCTTCCTCCCGCGCCCG CCCCT)
hTERT	d(CCCCGCCCCGTCCCGACCCCTCCCGGGTCCCCGG CCCAGCCCCCACCGGGCCCTCCCAGCCCCTCCCC)
2-aminopurine labelled h-teloc (Ap h-teloc)	d(CCCTApACCCTAACCCTAACCCTAA)

4A.2 Results and Discussions

4A.2.1 CD Study

Circular dichroism (CD), most commonly used to monitor the conformational transitions of DNA, technique has been used to probe the MCM-41 induced i-motif DNA formation. At slightly acidic pH (≤ 6), C-rich sequences such as, h-teloc (Table 4A.1) forms an i-motif structure showing a positive and negative band around 290 nm and 255 nm respectively; whereas, at neutral and alkaline pH, ssDNA/RC structure appears exhibiting a positive band at 270 nm and a negative band at 235 nm (Figure 4A.1A).

In aqueous buffer, MCM-41 remains as a negative entity (pK$_a$ 3-4 for silanol groups[18]) restricting the encapsulation of DNA molecules due to electrostatic repulsion with the negatively charged DNA phosphate backbone (Appendix 4A.1), consequently preventing the i-motif formation. For better adsorption of h-teloc into MCM-41, we have used non-polar solvent, DCM, because the polar DNA molecules prefer the polar MCM-41 nano-pores over the non-polar solvent environment. Hence, use of a non-polar solvent is necessary for the encapsulation of oligos in silica nano-channels. Initially, dispersion of h-teloc and MCM-41 in DCM does not exhibit any ellipticity pattern, but with evolution of time, a characteristic ellipticity pattern with positive peak at 290 nm and a negative peak at 255 nm have been observed indicating the i-motif formation (Appendix 4A.2). Notably, for biological relevance and further studies, the DNA loaded MCM-41 needs to be dispersed in aqueous buffer. Thus, we have extracted DNA loaded MCM-41 by solvent evaporation and repeatedly washed with buffer to remove the surface bound oligos from MCM-41. After careful extraction and washing, we have dispersed the h-teloc loaded MCM-41 powder in neutral (pH 7) and alkaline (pH 8, 9, 10) buffer solutions which exhibit an ellipticity pattern similar to i-motif DNA (Figure 4A.1B). At higher pH, slight decrement in i-motif characteristic peaks may indicate the enhanced unfolding with increasing pH of the medium,
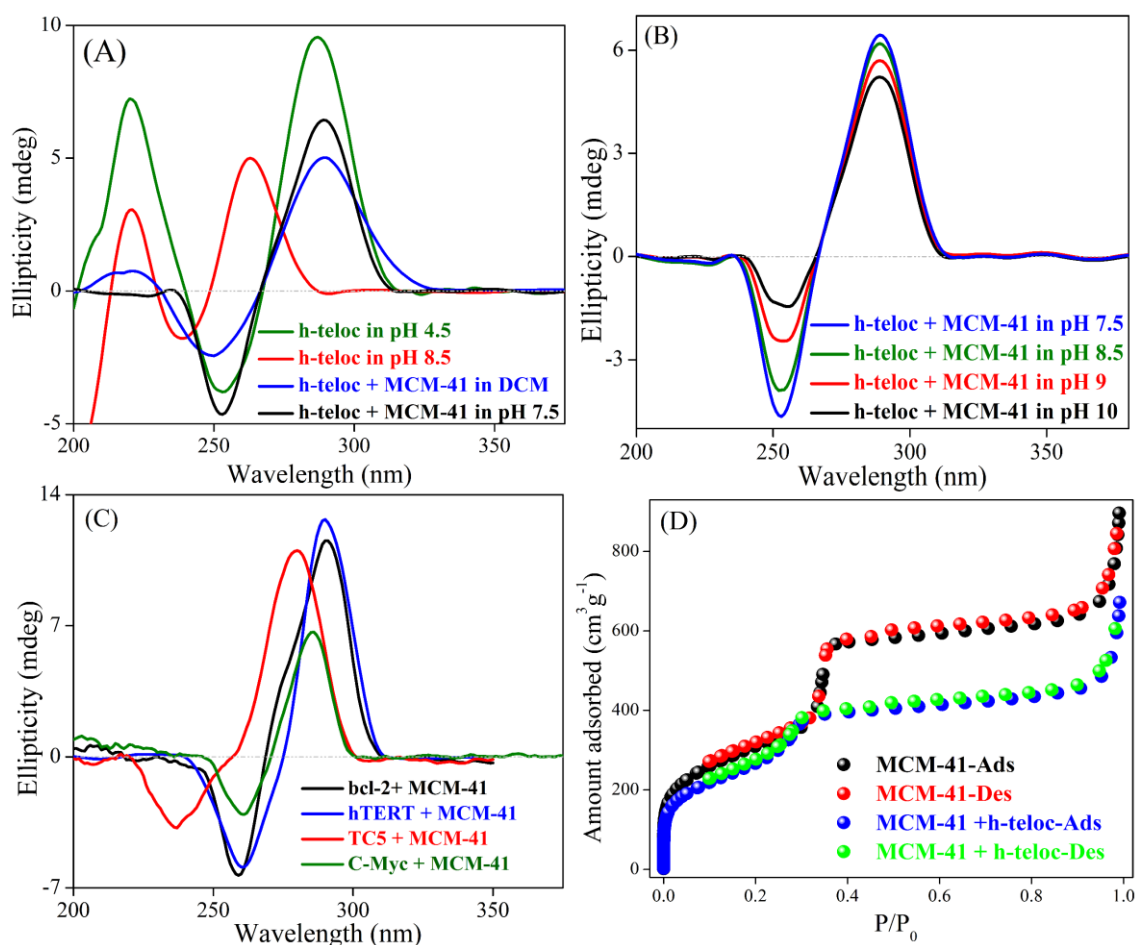
Figure 4A.1 Circular dichroism (CD) spectra of encapsulated h-teloc in MCM-41 nano-pores (A) in neutral pH, (B) in different alkaline pH. (C) CD spectra of different length of C-rich sequences in pH 8.5 buffer. (D) N$_2$ gas adsorption and desorption of MCM-41 and MCM-41+ h-teloc.

but importantly the spectral pattern does not change even at higher pH referring to the retention of i-motif structure. As MCM-41 is devoid of any ellipticity pattern, hence, we can infer that even at neutral and alkaline pH, MCM-41 is stabilizing the i-motif DNA by encapsulating in its nano-channels. Control experiments with other non i-motif forming DNA sequences having a same number of nucleobases as h-teloc indicate that all DNA encapsulations do not lead to i-motif formation (Appendix Note 4A.1). To examine the versatility of this approach, we have performed similar studies with various C-rich sequences (Figure 4A.1C). Interestingly, shorter C-rich sequence (i.e. TC5) in MCM-41 does not show characteristic i-motif ellipticity pattern, rather exhibits a positive band at 270 nm and a negative band at 235 nm indicative of ssDNA/RC (Figure 4A.1C). Here, it is pertinent to mention that shorter TC5 sequence forms an intermolecular i-motif structure with four DNA strands having a size of 1.6 nm × 3.1-3.6 nm.[2] However, the pore diameter of MCM-41 is 2.1-2.7 nm,[18c] which is relatively small to accommodate the larger sized intermolecular i-motif DNA (1.6 nm × 3.1-3.6 nm).[20] Apart from the shorter sequence, all other sequences exhibit i-motif ellipticity pattern with a positive peak at 290 nm and a negative peak at 255 nm (Figure 4A.1C). For longer C-rich sequences, intra-molecular i-motif structure is generally found to exist which is relatively smaller in size (1.4 nm × 2.2-2.6 nm)[21] compared to the intermolecular i-motif DNA (1.6 nm × 3.1-3.6 nm).[2, 20] Moreover, intermolecular i-motif structure requires the proximity of four C-rich strands, which is unlikely in the confined nano-pores of MCM-41. This may be a plausible explanation for the lack of intermolecular i-motif formation (i.e. TC5) in the MCM-41 nano-channels. From this sequence dependent CD study, we can infer that MCM-41 instigates the intramolecular i-motif formation, subsequently stabilizing them in the nano-pores at neutral and alkaline pH.

### 4A.2.2 N_2 Gas Adsorption and Desorption Study

To confirm the encapsulation of h-teloc inside MCM-41, we have performed N_2 gas adsorption and desorption measurement at 77 K (Figure 4A.1D). Adsorption profile shows a lower amount of N_2 uptake for h-teloc loaded MCM-41 (672 cm^3g^{-1}) compared to parent MCM-41 (896 cm^3g^{-1}). Brunauer–Emmett–Teller (BET) calculations also signify the reduction in the surface area after the encapsulation of the h-teloc in MCM-41 (from 1192 m^2g^{-1} to 975 m^2g^{-1}). Thus, comparatively for h-teloc loaded MCM-41 indicates the space filling of nano-pores by oligos. Moreover, in the presence of DNA, the calculated pore size and volume of MCM-41 by non-localized density functional theory (NLDFT) method shows a significant reduction from 33 Å to 28 Å and from 257 to 224 cm^3g^{-1}, respectively (Appendix 4A.4). Thus, gas adsorption and desorption measurements indicate the DNA encapsulation inside the nano-channels of MCM-41.

4A.2.3 Plausible Mechanism of i-motif Formation

Mechanism of this i-motif formation and stabilization in neutral and alkaline pH can be explained as follows. On mixing C-rich DNA and MCM-41 in non-polar DCM solvent, the

polar DNA sequence preferably incorporates in the less non-polar nano-channels of MCM-41. Inside the nano-channels, the mild acidic silanol –OH groups (pK_a 3–4)[18] donate the proton to its surrounding basic cytosine nucleobases (pK_a 4.5). The close pK_a values of both the moieties ensure the hemi-protonation of cytosine bases which is the important criteria for i-motif formation. Thus, we observe the i-motif DNA formation inside nano-channels even in non-polar solvent (Appendix 4A.2). In this way, formation of i-motif DNA is promoted and further stabilized in the nano-pores of MCM-41. Hence, relatively smaller sized intramolecular i-motif DNA structure forms in the MCM-41 nano-channels and not the bigger sized intermolecular one (Figure 4A.1C). The extracted and washed oligo loaded MCM-41 ensures stable i-motif structure in neutral and alkaline buffer primarily due to two reasons. Firstly, comparable sizes of i-motif and MCM-41 nano-pores can lead to the H-bond formation between negatively charged phosphate groups of DNAs and the silanol groups of MCM-41,[22] which tightly holds the oligos within the nano-channels, and secondly the hydrophobic interaction between DNA bases and silica hydrophobic region.[23] In addition to these, lack of solvent accessibility and confined environment inside MCM-41 nano-channels are also responsible for preserving the i-motif structure in neutral and alkaline medium.
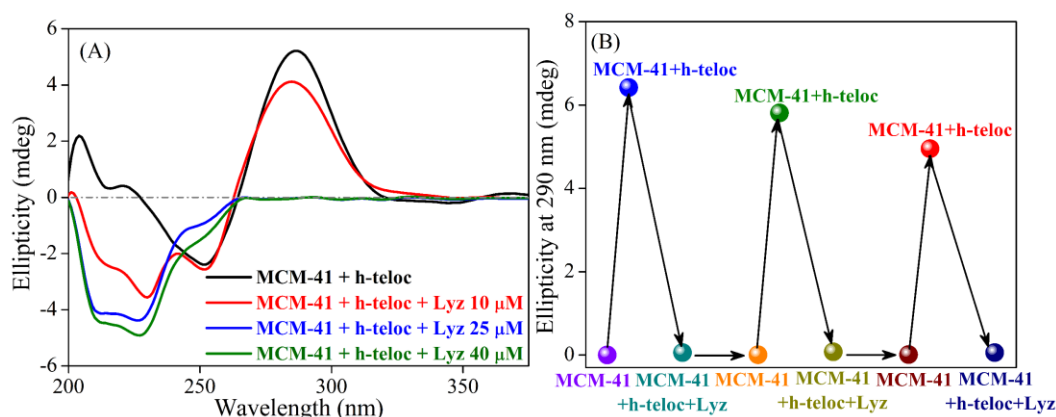
Figure 4A.2 (A) CD spectra of h-teloc encapsulated MCM-41 system with increasing concentration of Lyz protein in pH 7.5 buffer. (B) Ellipticity change at 290 nm of h-teloc during i-motif formation and subsequent disruption by Lyz protein through different cycles in pH 7.5 buffer.

### 4A.2.4 Reversibility of i-motif Formation

Reversibility of any conformational transition gives a better control over the formation and stability of the target structure. Thus, we have tried to revert the MCM-41 induced i-motif formation at neutral pH by introducing a positively charged protein, lysozyme (Lyz) which may provide a control to this process. We have selected the monomeric, globular Lyz for being positively charged (isoelectric point of Lyz - 11[24]) in physiological and slightly alkaline pH which involves a strong electrostatic interaction with negatively charged phosphate groups of DNAs. Additionally, negatively charged Si–O– groups at the edges of MCM-41 help the positively charged Lyz to approach closer to the DNA. Addition of Lyz to the h-teloc loaded

MCM-41 in pH 7.5 buffer exhibits the disappearance of characteristic ellipticity pattern of i-motif DNA (positive and negative peak at 290 nm and 255 nm respectively; Figure 4A.2A). Notably, control CD study of Lyz and i-motif DNA at pH 7.5 (Appendix 4A.5) shows an ellipticity pattern like MCM-41+h-teloc+Lyz indicating an identical type of DNA-protein complexation both in presence and absence of MCM-41. Importantly, large size of Lyz (3.5 nm)[24] prevents it to enter the pores of MCM-41 (2.1–2.7 nm), instead it adsorbs at the surface or the mouth of the nano-channels and drags h-teloc from the silica nano-channel. This reverts the conformational transition from i-motif to ssDNA/RC structure. Experiments with other C-rich sequences in MCM-41 and Lyz have been effectively found to exhibit the versatility of this reverting approach (Appendix Note 4A.2). As this kind of conformational reversibility is often useful in nanotechnology, we have explored the feasibility of the entire conformational changes over several cycles by monitoring the characteristic i-motif ellipticity peak at 290 nm. We have checked the entire conformational transitions in multiple cycles using same MCM-41 substrate throughout the process (Figure 4A.2B; Appendix Note 4A.3).

4A.2.5 Steady State and Time Resolved Fluorescence Study

Fluorescence spectroscopy is an excellent tool for investigating the structure, polymorphism and interactions of nucleic acids. In this regard, 2-aminopurine (Ap) has been extensively utilized to verify the conformation of different DNA structures due to its excellent sensitivity to local environmental factors (such as, polarity, viscosity), nucleic acid termini, stacking dynamics and sequence dependence etc.[1a, 25] Herein, we have probed the microenvironment of the MCM-41 induced i-motif formation by monitoring the fluorescence property of Ap labelled h-teloc (Ap h-teloc) sequence (Table 4A.1). Prior to the fluorescence study, we have checked the i-motif formation and its conformational transition to ssDNA/RC structure of Ap h-teloc through CD measurements (Appendix 4A.7). In pH 8.5, Ap h-teloc (ssDNA/RC form) exhibits an emission maximum around 370 nm (λ_{exc} 300 nm). This emission maximum undergoes a huge intensity enhancement on reducing the pH to 4.5 (Figure 4A.3A and Appendix 4A.8) due to i-motif formation, which causes less accessibility of solvent molecules to Ap and less stacking interaction restricting the electron transfer process between Ap and other nucleobases.[1a, 13, 25] Ap h-teloc encapsulated in MCM-41 nano-pores (In pH 8.5 buffer) also shows an identical fluorescence intensity enhancement to that of pH 4.5 buffer (Figure 4A.3A and Appendix 4A.8), which may be attributed to i-motif formation or/and confinement effect. To get more insight into this, we have performed time-resolved fluorescence studies. Lifetime transients of Ap h-teloc in ssDNA/RC conformation (at pH 8.5) exhibits shorter lifetime value of 1.39 ns than Ap h-teloc in i-motif conformation (2.88 ns at pH 4.5) due to greater stacking interaction in former leading to the electron transfer between Ap and other nucleobases (Figure 4A.3B; Table 4A.2).[25-26] However, lifetime of encapsulated Ap h-teloc in MCM-41 further decreases from

1.39 ns to 644 ps (Table 4A.2). Notably, confinement effect generally results in an enhancement of both emission intensity and lifetime.[27] On the other hand, lesser polarity and lack of solvation shell decreases the quantum yield and lifetime of Ap.[28]
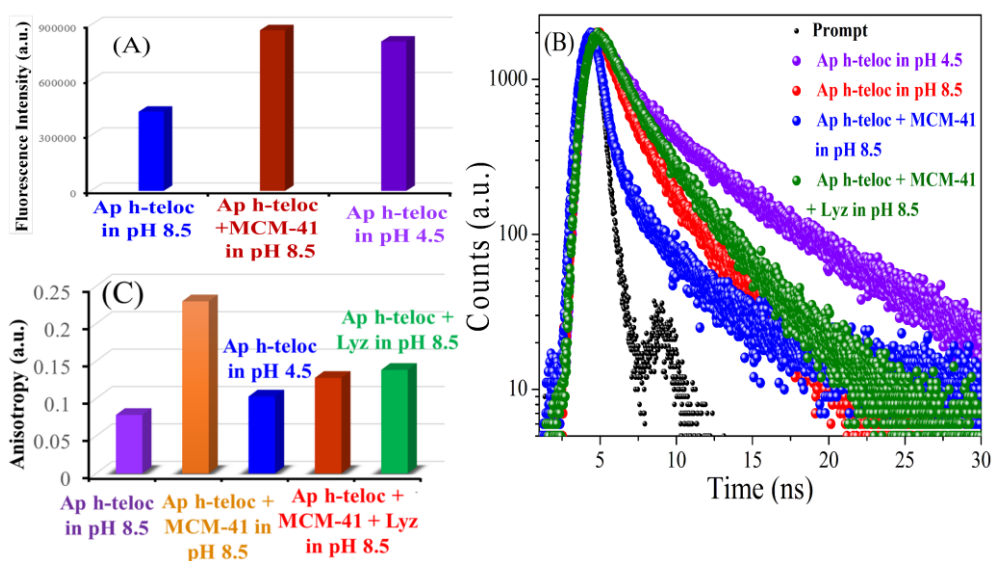
Figure 4A.3 (A) Emission intensity of Ap h-teloc (λ_{ex}= 300 nm) with and without encapsu-lation of MCM-41 in different pH medium. (B) Lifetime decays of Ap h-teloc (λ_{ex} 285 nm and λ_{col} 370 nm) in different systems. (C) Anisotropy change of Ap labelled h-teloc in different systems.

Table 4A.2 Fluorescence transient fittings of Ap labelled h-teloc DNA in case of different systems (λ_{ex} = 285 nm and λ_{em} = 370 nm).

Sample	τ_1.(ns)	a_1	τ_1 (ns)	a_2	τ_{avg}[#, $]$ (ns)	χ^2
Ap h-teloc in pH 4.5	1.35	0.67	6.02	0.33	2.88	1.08
Ap h-teloc in pH 8.5	0.86	0.76	3.11	0.24	1.39	1.07
Ap h-teloc + MCM-41 in pH 8.5	0.46	0.94	3.33	0.06	0.64	1.20
Ap h-teloc + MCM-41 + Lyz 40 μM in pH 8.5	1.14	0.58	3.39	0.42	2.07	1.07
Ap h-teloc + Lyz 40 μM in pH 8.5	1.63	0.73	5.09	0.27	2.55	1.10

[#]$\tau_{avg} = a_1\tau_1 + a_2\tau_2$; [$]We have considered the average lifetime values for explanation of the results to minimize the complexity.

Thus, the observed emission enhancement and lifetime decrement inside MCM-41 indicates that it is an outcome of both i-motif formation as well as confinement effect. Subsequent addition of Lyz to the Ap h-teloc in MCM-41 exhibits further fluorescence intensity enhancement which has also been observed for Ap h-teloc and Lyz in the absence of MCM (Appendix 4A.8). Identical observations in both, Ap h-teloc+Lyz+MCM and Ap h-teloc+Lyz system indicates the formation of similar protein-DNA complexation, due to the disruption of the i-motif structure on removal from MCM-41 nano-pores in former case. Addition of Lyz also recovers the lifetime

from 640 ps to 2.07 ns similar to the Ap h-teloc and Lyz system (2.55 ns) (Figure 4A.3B; Table 4A.2). This increment in lifetime transient is well corroborated by steady state results indicating the translocation of Ap h-teloc from less polar nano-channels to more polar medium. Encapsulation and subsequent i-motif formation of Ap h-teloc have also been monitored by anisotropy studies (Figure 4A.3C). Generally, a bound/encapsulated fluorophore shows higher anisotropy values compared to the free molecule due to their rotational restriction.[27] Here also, Ap h-teloc in MCM-41 exhibits almost five times higher anisotropy values than free Ap labelled h-teloc in pH 8.5 buffer due to its retarded rotational diffusion inside MCM-41. Addition of Lyz exhibits almost two times lower anisotropy value in comparison to encapsulated Ap h-teloc in MCM-41, which is very identical to Ap h-teloc and in presence of Lyz at pH 8.5 buffer (Figure 4A.3C). This again gives the evidence that Lyz drags the Ap h-teloc from MCM-41 nano-pores leading to the disruption of i-motif structure in bulk alkaline buffer medium.

4A.3 Conclusion

In summary, we have demonstrated that silica nano-channels can instigate the formation and stabilization of i-motif DNA (ssDNA → i-motif) even at neutral and alkaline medium. Subsequently, we have reversed this conformational transition (i-motif → ssDNA) by introducing Lyz protein in the same system, which drags the DNA from the MCM-41 nano-pores leading to the disruption of i-motif structure. This reversible approach offers an alternate way to control i-motif formation other than pH and thermal annealing. Hence, this new approach of stabilizing i-motif DNA based on MCM-41 is simple and biologically relevant for future applications of i-motif structure at neutral and alkaline pH.

4A.4 Appendix Section
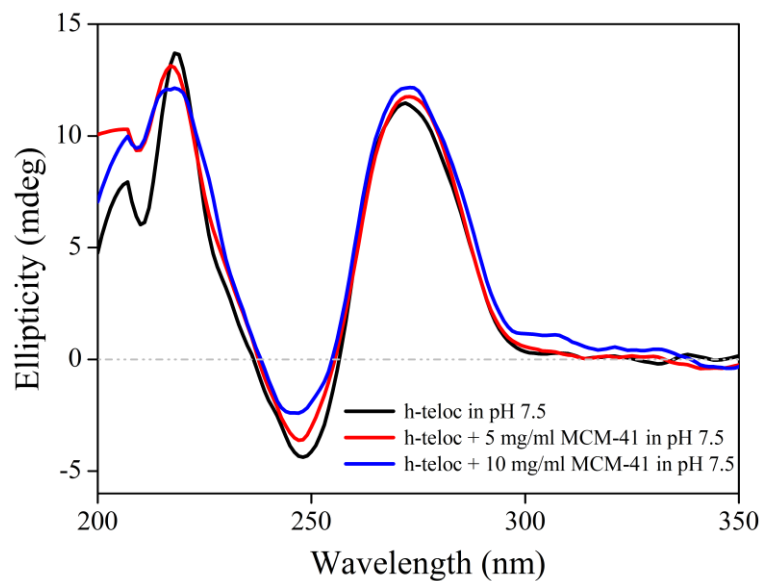
Appendix 4A.1 Circular dichroism (CD) spectra of h-teloc in pH 7.5 buffer with increasing concentration of MCM-41.
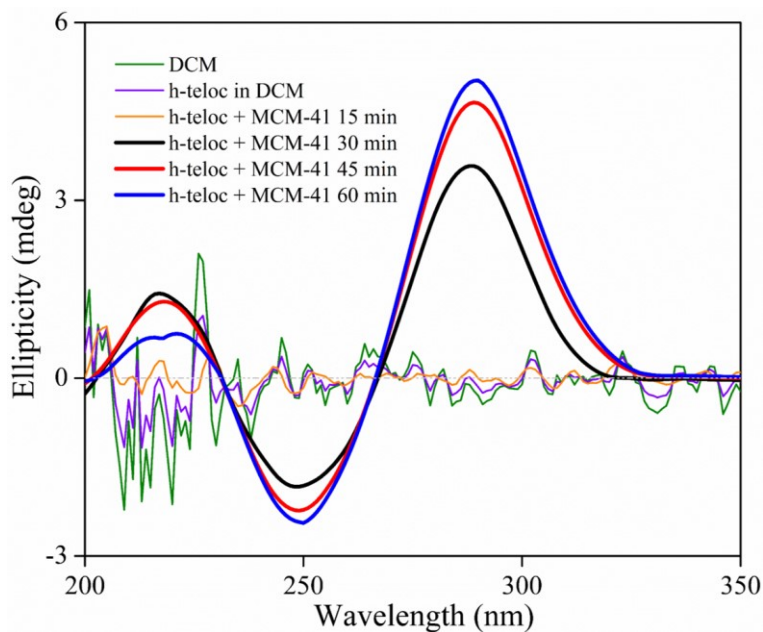
Appendix 4A.2 Circular dichroism spectra of h-teloc in presence of MCM-41 in DCM solution with evolution of time.

Appendix Note 4A.1 - Control experiments with non i-motif forming DNA sequences:

**Appendix Table 4A.1** Different non i-motif forming DNA sequences used in this control study.

DNA	Sequences (5´→ 3´)
oligo-1	d(TTTTAATTTTAATTTTAATTTTAA)
H24	d(TTAGGGTTAGGGTTAGGGTTAGGG)

We have carried out similar experiments with other non i-motif forming DNA sequences having a same number of nucleobases (Appendix Table 4A.1) as h-teloc (5´-CCCTAACCCTAACCCTAACCCTAA-3´) to ensure that these observations are a result of i-motif DNA formation or not. We have substituted the cytosine bases of h-teloc with thymine. Thus, this sequence (oligo-1) cannot form the i-motif DNA in any condition. We have mixed the respective DNA sequences (oligo-1 and H24) and MCM-41 in non-polar DCM solvent. Later, we have extracted the DNA loaded MCM-41 by solvent evaporation method. Then, we have dispersed the DNA loaded (oligo-1 and H24) MCM-41 in pH 7.5 buffer and deionized water. In pH 7.5 buffer, oligo-1 exhibits an ellipticity pattern with a positive peak at 279 nm and negative peak at 247 nm (Appendix 4A.3A), whereas encapsulated oligo-1 in MCM-41 shows similar ellipticity pattern to oligo-1 with decrement in negative peak at 247 nm (Appendix 4A.3A). This decrement in negative peak upon encapsulation may be attributed to the lesser helicity pattern arising from the confined environment inside MCM-41. Moreover, the observed ellipticity pattern for encapsulated oligo-1 in MCM-41 is distinctly different from the i-motif ellipticity pattern.

We have also performed this encapsulation with the complimentary sequence (H24) of h-teloc which is known for its G-quadruplex (GQ) formation in presence of ions, but cannot form any i-motif structure due to absence of cytosine nucleobases. In deionized water, CD spectra of H24 contains a positive peak around 260 nm and negative peak around 235 nm indicating the ssDNA/RC formation (Appendix 4A.3B), which gets altered in presence of K^+ ion due to hybrid GQ formation exhibiting two positive peaks around 290 nm and 250 nm, and a negative peak at around 235 nm.[29] However, encapsulated H24 inside MCM-41 in deionized water as well as in presence of K^+ ion exhibits the characteristics peaks of ssDNA/RC formation with very less intensity for negative peak around 235 nm (Appendix 4A.3B). Thus, the absence of GQ structure inside MCM-41 even in presence of K^+ ion indicates the lack of accessibility of solvent molecule or ion inside MCM-41 nano-channels. Here also, huge decrement in helicity peak (i.e. negative peak around 235 nm) upon encapsulation may be attributed to distortion of secondary structure due to the restricted environment inside MCM-41. In a nutshell, distinctly different ellipticity patterns observed for both oligo-1 and H24 loaded MCM-41 in comparison to i-motif DNA indicates that all DNA encapsulations do not lead to i-motif formation.
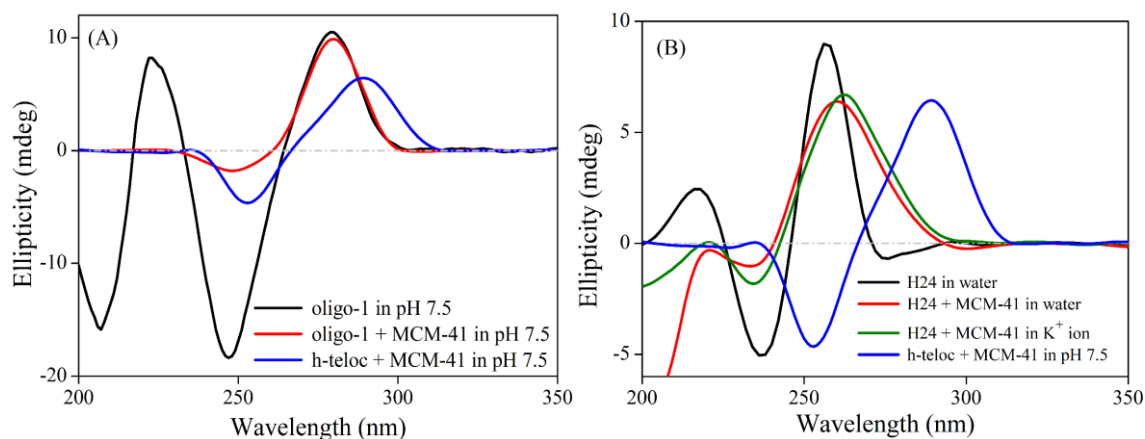
Appendix 4A.3 Circular dichroism (CD) spectra of (A) oligo-1 and encapsulated oligo-1 inside MCM-41 in pH 7.5 buffer. (B) CD spectra of H24 and encapsulated H24 inside MCM-41 in deionised water and in K^+ ion (10 mM) containing pH 7.5 buffer.
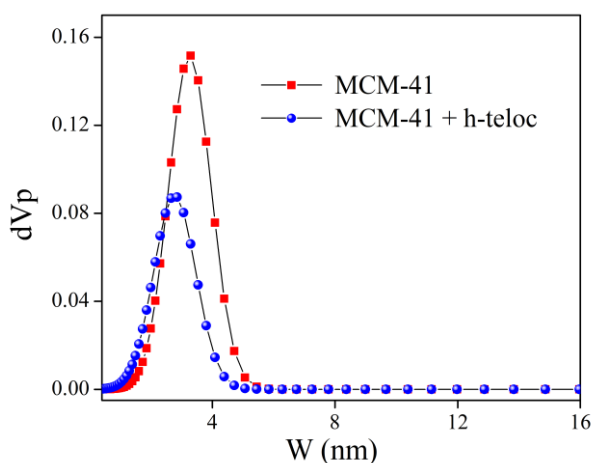

Appendix 4A.4 Pore-size distribution of MCM-41 and MCM-41+ h-teloc system.
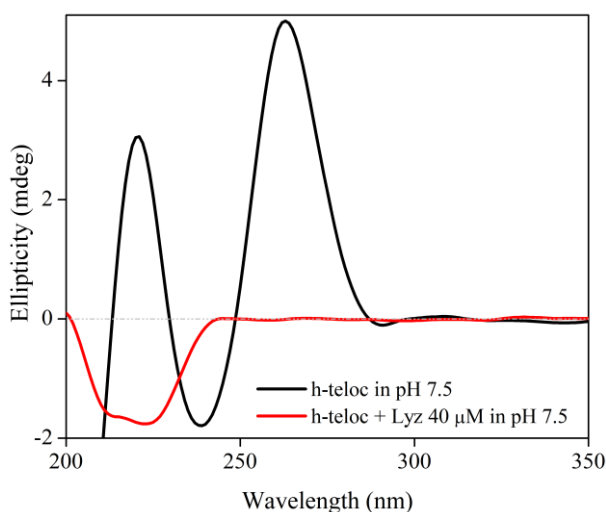
Appendix 4A.5 Circular dichroism spectra of h-teloc in presence of Lyz protein at pH 7.5 buffer.

Appendix Note 4A.2 - Versatility of the Reversible Approach:

To check the versatility of this reversible approach, we have performed similar experiments with MCM-41 containing different C-rich sequences and Lyz (Appendix 4A.6). Here, we have taken two other C-rich sequences, C-Myc and bcl-2 in addition to the h-teloc sequence (Table 4A.1). CD spectra of C-Myc/bcl-2 contained in MCM-41 at pH 7.5 undergoes significant alteration in ellipticity pattern with Lyz addition (Appendix 4A.6A and Appendix 4A.6B). With increasing Lyz concentration, characteristic ellipticity pattern of i-motif DNA (positive peak around 290 nm and negative peak around 260 nm) decreases gradually along with the appearance of ellipticity pattern for native Lyz (negative peaks around 230 nm and 210 nm respectively). This indicates the disruption of i-motif DNA and subsequent formation of corresponding DNA and Lyz complex which is evident from their respective control experiment with C-Myc/bcl-2 and Lyz in pH 7.5 buffer (Appendix 4A.6C).
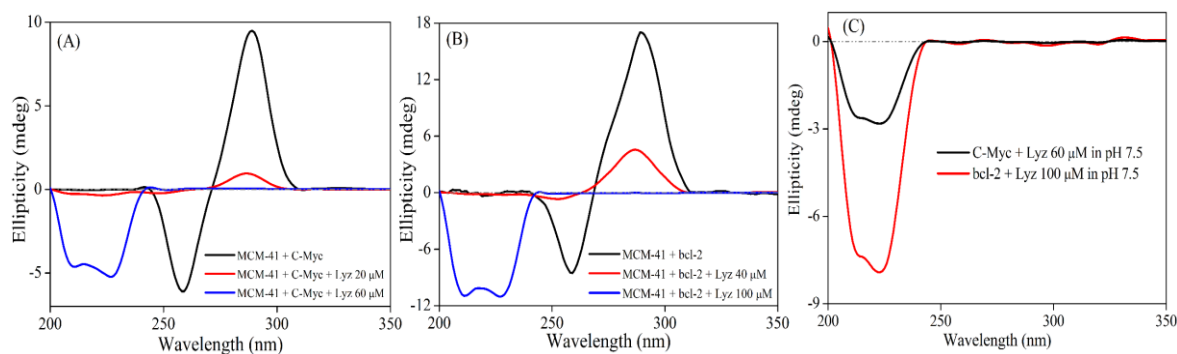
Appendix 4A.6 CD spectra of (A) C-Myc and (B) bcl-2 encapsulated MCM-41 systems with increasing concentration of Lyz protein in pH 7.5 buffer. (C) Circular dichroism spectra of C-Myc and bcl-2 in presence of Lyz protein at pH 7.5 buffer.

Interaction of positively charged Lyz with negatively charged phosphate groups of DNAs drags the C-rich strands out of the MCM-41 nano-pores. This leads to the protein-DNA complex formation which is eventually responsible for the conformational transition from i-motif DNA to single stranded/random coil form. Another important point to mention is that higher concentration of Lyz is required for longer C-rich sequences (40 μM for h-teloc having 24 nucleobases, 60 μM for C-Myc having 31 nucleobases, 100 μM for bcl-2 having 39 nucleobases). This can be attributed to the larger number of positively charged Lyz moieties involved in the interactions with the higher number of negatively charged phosphate groups present in longer C-rich sequences. This reversible approach is found to be effective for several C-rich sequences observed in different oncogenes ensuring the versatility of the approach.

Appendix Note 4A.3 - Verification of Reversible Approach in Multiple Cycles:

We have checked the reversible nature of our approach by utilizing the same MCM-41 substrate in different cycles (Figure 4A.2B). We have utilized the ellipticity signal at 290 nm as the indicator for i-motif structure. First, we have mixed the h-teloc and MCM-41 in non-polar DCM solvent for better encapsulation of oligo in silica nano-channels. Then, we have extracted the MCM-41 samples and repeatedly washed with buffer to remove any unbound oligo/surface bound oligo from the MCM-41. Subsequently we have dispersed this oligo loaded MCM-41 samples in pH 7.5 buffer (MCM-41+h-teloc) which shows a characteristic i-motif ellipticity pattern with a reasonable positive peak at 290 nm. Next, we have added Lyz protein in the same solution which shows the disappearance of i-motif ellipticity pattern with negligible positive peak at 290 nm (MCM-41+h-teloc+Lyz). This indicates the absence of i-motif structure in the solution. Again, we have extracted the MCM-41 samples from the solution of MCM-41+h-teloc+Lyz and washed repeatedly with buffer to remove any unbound/surface bound protein and DNA from the MCM-41. Following this, we have dispersed the extracted MCM-41 samples in DCM solvent and freshly added h-teloc in the solution. The above described processes were repeated in cycles and the ellipticity pattern was noted at each step which showed identical change as the initial cycle. This type of feasibility over several cycles may be useful in potential applications for nanotechnology.
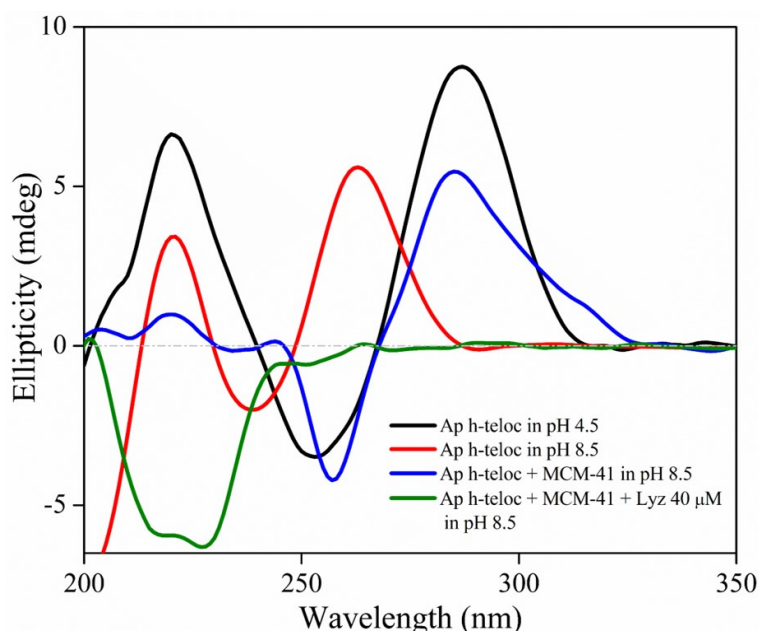
Appendix 4A.7 Circular dichroism spectra of 2-aminopurine labelled h-teloc (Ap h-teloc) in neutral and alkaline buffer solutions. CD spectra of Ap h-teloc encapsulated MCM-41 in presence and absence of Lyz protein in pH 8.5 buffer.
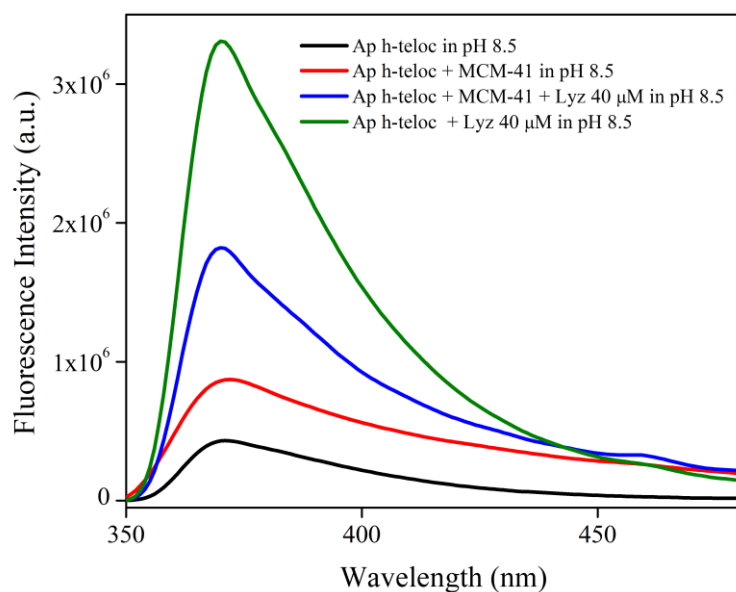
Appendix 4A.8 Steady state fluorescence spectra of 2-aminopurine labelled h-teloc (Ap h-teloc) in presence of MCM-41 and Lyz protein in alkaline buffer solutions.

4A.5 Expeimental

Mesoporous silica (MCM-41) with a pore diameter of 20–30 Å is dried in a vacuum oven at 300°C in order to remove trace amount of encapsulated water molecules. We have performed all the experiments with 10 mM phosphate buffer solutions of different pH. Prior to the experiments, DNA samples were annealed at 90 °C for 10 min and stored at 4 °C for 24 h. Concentrations of DNA samples were determined using the molar extinction coefficient at 260 nm provided by IDT, USA. It is necessary to mention here that IDT has determined the molar extinction coefficient using nearest neighbor approximation model. The concentration of lysozyme was calculated using the molar extinction coefficient of 37970 M^{-1} cm^{-1} at 280 nm.[30] Prior to the experiments, we have prepared the samples and kept it overnight for measurements.

Before adsorption measurements, the samples (MCM-41 and MCM-41+h-teloc) were heated at 120 °C under vacuum for 24 hours. The specific surface area and pore size distribution were obtained from the analysis of the adsorption–desorption isotherms using Brunauer–Emmett–Teller (BET) and Non-Localized Density Functional Theory (NLDFT) methods, respectively. Decay profiles were analyzed using IBH DAS6 software. Decays were fitted with minimum number of exponentials based on the χ^2 values and the visual inspection of the residuals quality.

4A.6 References

1. (a) Y. Xu and H. Sugiyama, *Nucleic Acids Res.*, 2006, **34**, 949-954; (b) E. P. Wright, J. L. Huppert and Zoë A. E. Waller, *Nucleic Acids Res.*, 2017, **45**, 2951-2959.

2. K. Gehring, J.-L. Leroy and M. Guéron, *Nature*, 1993, **363**, 561.

3. H. A. Day, P. Pavlou and Z. A. E. Waller, *Bioorg. Med. Chem.*, 2014, **22**, 4407-4418.

4. J. A. Brazier, A. Shah and G. D. Brown, *Chem. Commun.*, 2012, **48**, 10739-10741.

5. (a) S. Kendrick, H.-J. Kang, M. P. Alam, M. M. Madathil, P. Agrawal, V. Gokhale, D. Yang, S. M. Hecht and L. H. Hurley, *J. Am. Chem. Soc.*, 2014, **136**, 4161-4171; (b) H.-J. Kang, S. Kendrick, S. M. Hecht and L. H. Hurley, *J. Am. Chem. Soc.*, 2014, **136**, 4172-4185.

6. S. Takahashi, J. A. Brazier and N. Sugimoto, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 9605.

7. M. Zeraati, D. B. Langley, P. Schofield, A. L. Moye, R. Rouet, W. E. Hughes, T. M. Bryan, M. E. Dinger and D. Christ, *Nat. Chem.*, 2018, **10**, 631-637.

8. I. V. Nesterova and E. E. Nesterov, *J. Am. Chem. Soc.*, 2014, **136**, 8843-8846.

9. (a) J. Kim, Y. M. Lee, Y. Kang and W. J. Kim, *ACS Nano*, 2014, **8**, 9358-9367; (b) W. Li, J. Wang, J. Ren and X. Qu, *Angew. Chem., Int. Ed.*, 2013, **52**, 6726-6730.

10. L. Shi, P. Peng, Y. Du and T. Li, *Nucleic Acids Res.*, 2017, **45**, 4306-4314.

11. W. Guo, C.-H. Lu, R. Orbach, F. Wang, X.-J. Qi, A. Cecconello, D. Seliktar and I. Willner, *Adv. Mater.*, 2015, **27**, 73-78.

12. (a) L. A. Yatsunyk, O. Mendoza and J.-L. Mergny, *Acc. Chem. Res.*, 2014, **47**, 1836-1844; (b) Y. Dong, Z. Yang and D. Liu, *Acc. Chem. Res.*, 2014, **47**, 1853-1860.

13. X. Chen, X. Zhou, T. Han, J. Wu, J. Zhang and S. Guo, *ACS Nano*, 2013, **7**, 531-537.

14. X. Li, Y. Peng, J. Ren and X. Qu, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 19658.

15. (a) H. A. Day, C. Huguin and Z. A. E. Waller, *Chem. Commun.*, 2013, **49**, 7696-7698; (b) H. A. Day, E. P. Wright, C. J. MacDonald, A. J. Gates and Z. A. E. Waller, *Chem. Commun.*, 2015, **51**, 14099-14102; (c) S. Saxena, S. Joshi, J. Shankaraswamy, S. Tyagi and S. Kukreti, *Biopolymers*, 2017, **107**, e23018-n/a.

16. (a) A. Rajendran, S.-i. Nakano and N. Sugimoto, *Chem. Commun.*, 2010, **46**, 1299-1301; (b) J. Cui, P. Waltman, H. V. Le and A. E. Lewis, *Molecules*, 2013, **18**; (c) Y. Yang, Y. Sun, Y. Yang, Y. Xing, T. Zhang, Z. Wang, Z. Yang and D. Liu, *Macromolecules*, 2012, **45**, 2643-2647.

17. (a) L. Lannes, S. Halder, Y. Krishnan and H. Schwalbe, *ChemBioChem*, 2015, **16**, 1647-1656; (b) A. Aviñó, M. Dellafiore, R. Gargallo, C. González, A. M. Iribarren, J. Montserrat and R. Eritja, *ChemBioChem*, 2017, **18**, 1123-1128; (c) H. A. Assi, V. R. W. Harkness, N. Martin-Pintado, C. J. Wilds, R. Campos-Olivas, A. K. Mittermaier, C. González and M. J. Damha, *Nucleic Acids Res.*, 2016, **44**, 4998-5009; (d) B. Mir, X. Solés, C. González and N. Escaja, *Sci. Rep.*, 2017, **7**, 2772; (e) V. B. Tsvetkov, T. S. Zatsepin, E. S. Belyaev, Y. I. Kostyukevich, G. V. Shpakovski, V. V. Podgorsky, G. E. Pozmogova, A. M. Varizhuk and A. V. Aralov, *Nucleic Acids Res.*, 2018, DOI: 10.1093/nar/gky121.

18. (a) Z. Li, J. C. Barnes, A. Bosoy, J. F. Stoddart and J. I. Zink, *Chem. Soc. Rev.*, 2012, **41**, 2590-2605; (b) M. W. Ambrogio, C. R. Thomas, Y.-L. Zhao, J. I. Zink and J. F. Stoddart, *Acc. Chem. Res.*, 2011, **44**, 903-913; (c) R. K. Koninti, S. Palvai, S. Satpathi, S. Basu and P. Hazra, *Nanoscale*, 2016, **8**, 18436-18445.

19. (a) D. He, X. He, K. Wang, J. Cao and Y. Zhao, *Adv. Funct. Mater.*, 2012, **22**, 4704-4710; (b) C.-H. Lu and I. Willner, *Angew. Chem., Int. Ed.*, 2015, **54**, 12212-12235; (c) C. Chen, F. Pu, Z. Huang, Z. Liu, J. Ren and X. Qu, *Nucleic Acids Res.*, 2011, **39**, 1638-1644.

20. N. Esmaili and J. L. Leroy, *Nucleic Acids Res.*, 2005, **33**, 213-224.

21. (a) X. Han, J.-L. Leroy and M. Guéron, *J. Mol. Biol.*, 1998, **278**, 949-965; (b) A. T. Phan, M. Guéron and J.-L. Leroy, *J. Mol. Biol.*, 2000, **299**, 123-144.

22. M. Fujiwara, F. Yamamoto, K. Okamoto, K. Shiokawa and R. Nomura, *Anal. Chem.*, 2005, **77**, 8138-8145.

23. B. Shi, Y. K. Shin, A. A. Hassanali and S. J. Singer, *J. Phys. Chem. B*, 2015, **119**, 11030-11040.

24. K. P. Wilson, B. A. Malcolm and B. W. Matthews, *J. Biol. Chem.*, 1992, **267**, 10842-10849.

25. (a) T. Kimura, K. Kawai, M. Fujitsuka and T. Majima, *Chem. Commun.*, 2004, DOI: 10.1039/B403913K, 1438-1439; (b) T. Kimura, K. Kawai, M. Fujitsuka and T. Majima, *Chem. Commun.*, 2006, DOI: 10.1039/B514526K, 401-402.

26. J. M. Jean and K. B. Hall, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 37.

27. J. R. Lakowicz, *Principles of fluorescence spectroscopy*, Second edition. New York : Kluwer Academic/Plenum, 1999.

28. (a) S. Lobsiger, S. Blaser, R. K. Sinha, H.-M. Frey and S. Leutwyler, *Nat. Chem.*, 2014, **6**, 989; (b) E. L. Rachofsky, R. Osman and J. B. A. Ross, *Biochemistry*, 2001, **40**, 946-956.

29. S. Satpathi, M. Kulkarni, A. Mukherjee and P. Hazra, *Phys. Chem. Chem. Phys.*, 2016, **18**, 29740-29746.

30. S. Klitgaard, M. T. Neves-Petersen and S. B. Petersen, *J. Fluoresc.*, 2006, **16**, 595-609.

Chapter 4B
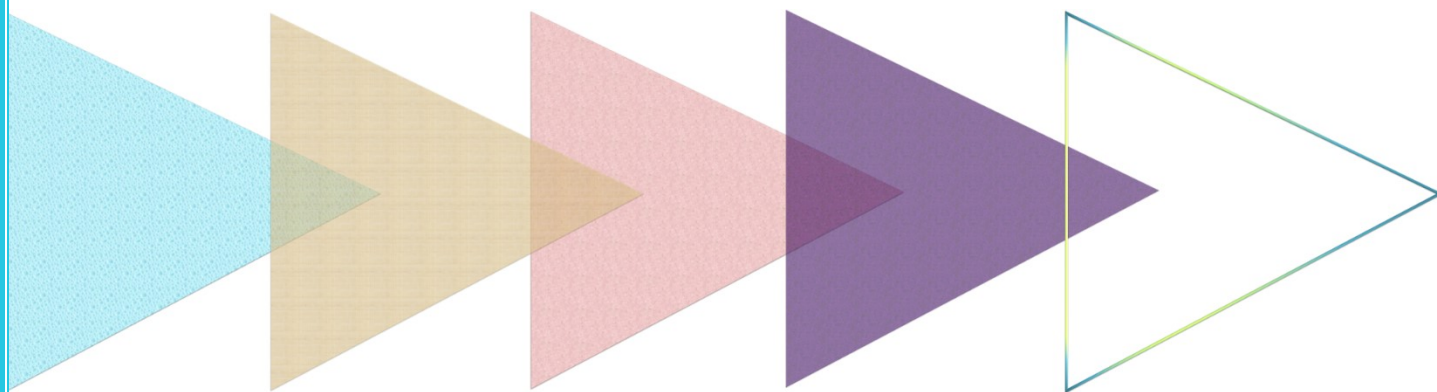
Structural Characteristics Requisite for the Ligand-Based Selective Detection

of i-motif DNA

4B.1 Introduction

Among different secondary structures of DNA, i-motif form has received special attention due to its versatile applications in biological,[1] material chemistry[2] and nanotechnology including DNA nanosprings,[3] pH indicator,[4] unidirectional DNA walker[5] etc. Cytosine-rich sequences form i-motif DNA at mildly acidic pH (4-6) by means of Hoogsteen hydrogen bonding between hemi-protonated cytosine–cytosine (C$^+$–C) base pairs, leading to the arrangement of two parallel duplexes intercalated in an antiparallel manner (Scheme 4B.1).[6] These i-motif DNA structures become unstable in neutral and alkaline medium due to the deprotonation of cytosine bases (pKa ~ 4-5), thus forming single-stranded/random coil (SS/RC) DNA.[7] Apart from its diverse applications, the biological relevance of this structure is still under scrutiny. However, it is anticipated that i-motif DNA plays key regulatory roles in the genome due to its presence in many oncogene promoter regions and human nuclei.[7-8] Thus, constant efforts are still underway to decipher the specific biological functions of i-motif DNA. To understand the function and activity of any biomolecule, labelling with a fluorescent marker remains the most easy and accurate method to track inside cellular environments. However, tracking of i-motif DNA inside the cellular conditions has been inadequate due to the severe lack of selective i-motif DNA ligands. In this quest, initially researchers have covalently attached the target fluorophore to the C-rich sequences, but this approach may bring complications involving the structural stability of the folded and unfolded forms.[9] Hence, label-free strategy with a specific fluorescent ligand, based on the non-covalent interaction is preferable for targeting the i-motif DNA.[10] Following this, researchers have endeavoured to design suitable markers for i-motif DNA, but very few molecules have shown specificity for only i-motif DNA mainly due to its dynamic structure and additional positive charge unlike all other DNA conformations. These peculiar credentials in i-motif DNA make the recognition difficult and challenging. In this regard, some
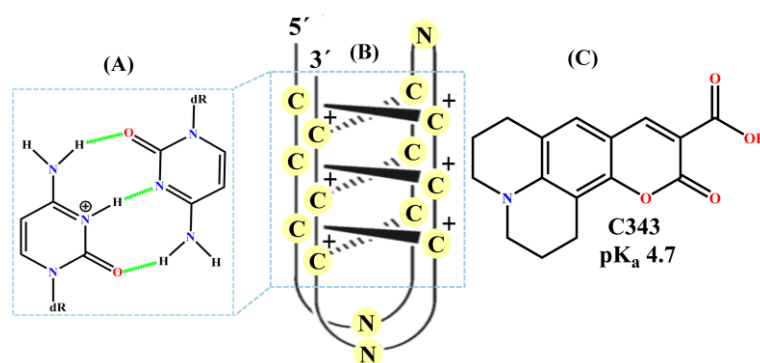
Scheme 4B.1 (A) hemi-protonated cytosine–cytosine hoogsteen base pairing in i-motif DNA. Chemical structure of (B) i-motif DNA and (C) coumarin 343.

fluorescent molecules have been anticipated as i-motif ligands such as, thioflavin T,[11] DMSB (2,2′-diethyl-9-methylselena-carbocyaninebromide),[12] thiazole orange,[13] berberine,[14] crystal

violet,[5] [Ru(bpy)$_2$(dppz)]$_2$$^+$,[15] peptidomimetic ligands (PBP1),[16] [Ir(phq)$_2$(Cl-phen)]PF$_6$,[17] mitoxantrone,[18] but most of these ligands show a turn-on fluorescence behavior with the G-quadruplex DNA (GQ DNA) as well, which is expected to be abundant in the vicinity of i-motif structure due to their complementarity nature.[7-8] Thus, the selectivity of i-motif ligands is still an issue, which needs to be resolved in order to figure out the functions and activity of i-motif DNA. So far, neutral red is the only selective ligand for i-motif DNA to the best of our knowledge, which operates by interacting at the loops of i-motif DNA.[19] This loop binding interaction approach faces a serious limitation regarding the universality of this detection due to its possible ineffectiveness for shorter C-rich sequences, i.e. intermolecular i-motif DNA, which does not contain any loop in the i-motif structure.[6] Thus, a ligand with high selectivity and versatility towards both intermolecular and intramolecular i-motif DNA is of prime importance. To address this challenge, we have tried to explore the structural characteristics required for an ideal i-motif DNA ligand using various coumarin derivatives. Coumarin belonging to the benzopyrone class of compounds exhibits low toxicity and polarity sensitive nature which has been extensively utilized in live cell imaging.[20] Herein, we have demonstrated the sensing ability of coumarin 343 (C343) towards i-motif DNA by monitoring the fluorescence light up behaviour. We have also addressed the selectivity of this ligand towards i-motif DNA over GQ DNA and duplex DNA. Additionally, the versatility of this ligand is characterized by comparing the fluorescence response with several other C-rich sequences found in various oncogene promoters (Table 4B.1) having different chain lengths, sizes, loop lengths and molecularity. We have hypothesized the probable mechanism behind this interaction, enabling us to explore the relationship between i-motif DNA and its ligands, which may be useful for future designing of i-motif DNA based ligands.

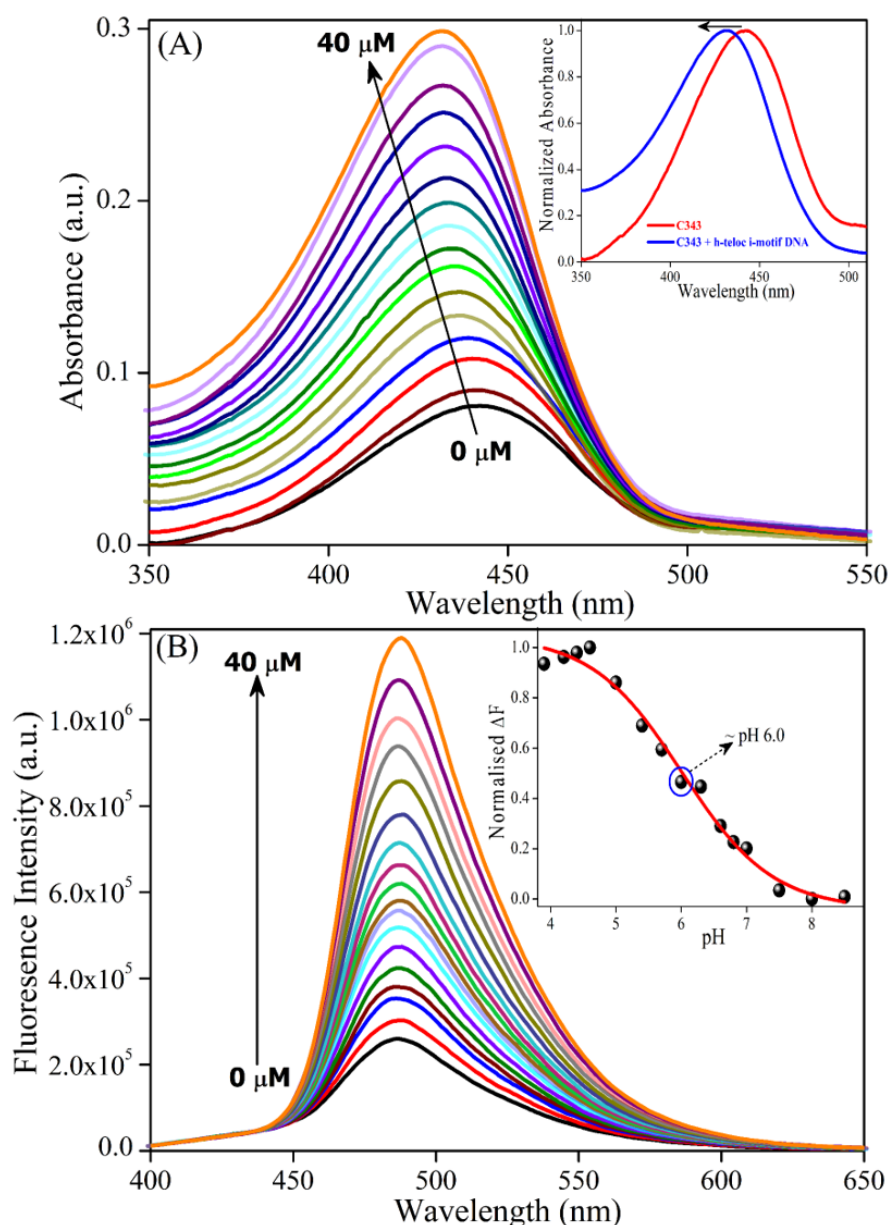**Table 4B.1** Different cytosine-rich DNA sequences used in this study.

DNA	Sequences (5´→ 3´)	Sequence length
h-teloc	d(CCCTAACCCTAACCCTAACCCTAA)	24
TC5	d(TCCCCC)	6
C-Myc	d(CCCCACCTTCCCCACCCTCCCCACCCTCCCC)	31
bcl-2	d(CAGCCCCGCTCCCGCCCCCTTCCTCCCGCGCCCGCCCCT)	39
hTERT	d(CCCCGCCCCGTCCCGACCCCTCCCGGGTCCCCGGCCCAGC CCCCACCGGGCCCTCCCAGCCCCTCCCC)	68

4B.2 Results and Discussions

4B.2.1 Absorption Studies

Depending on the pH of the medium, coumarin 343 (C343) can exist in two prototropic forms (Scheme 4B.2) showing two distinct absorption bands,[21] ~454 nm peak for neutral form and ~430 nm peak for anionic form (Figure 4B.1A and Appendix 4B.1A). As most of the i-motif

structures are stable in pH 5 to 5.5 range,[13] thus, we have carried out our experiments at pH 5.2 for i-motif structure and pH 8.5 buffer for single-stranded DNA/random coil (ssDNA/RC) form. At pH 5.2 buffer, C343 shows a broad absorption band around 445 nm (Figure 4B.1A) indicating a mixed distribution of neutral (~ 34%) and anionic (~ 66%) forms, which is expected due to the closeness of the experimental pH with the pK_a of neutral to anion conversion (4.7). With increasing addition of h-teloc in pH 5.2 buffer (i.e., i-motif DNA), the absorption band undergoes a huge enhancement with a significant blue shift of 15 nm from 445 nm to 430 nm. This indicates that addition of i-motif DNA favours and stabilises the anionic form, thus shifting

Figure 4B.1 (A) Absorption spectra and (B) fluorescence spectra of coumarin 343 (C343; λ_{ex} 375 nm) with increasing concentration of h-teloc in pH 5.2 buffer (40 μM). Inset of Figure 4B.1A shows the extent of blue shift for the C343 absorption in presence of i-motif DNA. Inset of Figure 4B.1B shows the pH effect on the fluorescence response of C343 and i-motif DNA interaction. Arrows indicate the increasing concentration of h-teloc (40 μM).

the ground state equilibrium towards anionic species. Whereas, the addition of h-teloc in pH 8.5 buffer (i.e. ssDNA/RC form) does not show any appreciable change in the absorption profiles of C343 (Appendix 4B.1A) indicating a preferential binding mode of C-343 with i-motif DNA over ssDNA/RC form.

4B.2.2 Fluorescence Studies

We have also carried out the fluorescence study of C343 and h-teloc in both pH 5.2 and pH 8.5 buffer. In excited state also, two prototropic forms of C343 exist in a pH-dependent manner (Scheme 4B.2), where neutral exhibits an emission peak at 494 nm in the acidic condition, which gets blue shifted to 487 nm due to the formation of anionic form in neutral and basic pH (Figure 4B.1B and Appendix 4B.1B). In pH 5.2, C343 shows an emission maximum at 487 nm indicating the existence of an anionic form in the excited state (Figure 4B.1B) which undergoes a huge enhancement (~ 4.6-fold) upon addition of i-motif DNA (i.e., h-teloc). This kind of huge emission enhancement is found to be absent for C343 and h-teloc in pH 8.5 buffer indicating the lack of interaction of C343 with ssDNA/RC form (Appendix 4B.1B). To explore the interaction between C343 and i-motif, we have monitored the whole process throughout a pH gradient starting from pH 4 to pH 8.5 (inset of Figure 4B.1B). Change in emission intensity resulting from the C343 and h-teloc interaction shows a sigmoidal pH dependency with a transition midpoint at pH 6. This is consistent with the h-teloc i-motif stability indicating the selective binding between C343 and i-motif. Additionally, we have monitored the effect of C343 on the conformation and structural stability of i-motif DNA through CD and UV-melting studies (Appendix 4B.2). We have also calculated the binding constant between C343 and i-motif DNA which is explained in Appendix Note 4B.1. These studies indicate that C343 causes very minimal effect on the conformation and stability of i-motif DNA, which is desirable for any ideal i-motif sensor. In a nutshell, observations from both absorption and fluorescence studies suggest that C343 shows a turn-on fluorescence response in the presence of i-motif DNA (in pH 5.2) over ssDNA/RC structure (in pH 8.5).

Neutral
λ_{abs} **454 nm**
λ_{em} **494 nm**

pK_a **4.7**

Anion
λ_{abs} **430 nm**
λ_{em} **487 nm**

Scheme 4B.2 Interconversion between neutral and anion form of coumarin 343.
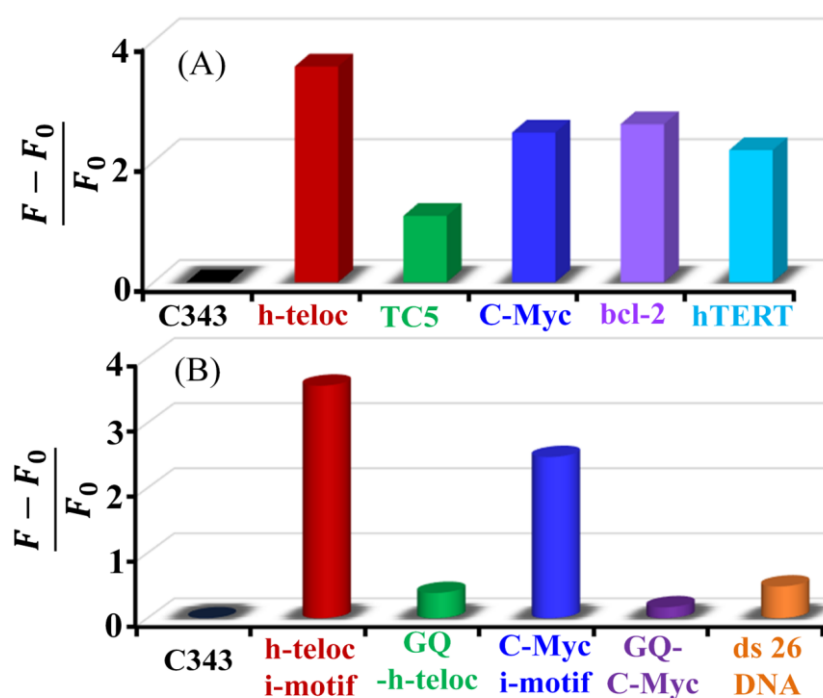
4B.2.3 Versatility of the i-motif Detection

To check the versatility of this i-motif detection by C343, we have carried out similar steady-state experiments of C343 with several i-motif forming sequences (having bases from 6 to 69) with different

chain lengths, molecularity and sizes, observed in various oncogene promoters (Table 4B.1).[7] Interestingly, the shorter C-rich DNA sequence i.e., TC5 at mild acidic pH results in an intermolecular i-motif structure, whereas the longer C-rich sequences (C-Myc, bcl-2, hTERT) form intramolecular i-motif DNA. In presence of both intramolecular (C-Myc, bcl-2, hTERT) and intermolecular (TC5) i-motif DNA (pH 5.2), absorption spectrum of C343 undergoes a significant blue shift (~ 15 nm) with an absorption enhancement exhibiting a ground state neutral (350 nm) to anion (330 nm) conversion in presence of i-motif (Appendix 4B.5). Such kind of ground state conversion is found to be absent for ssDNA/RC DNA (in pH 8.5) for different C-rich sequences (Appendix 4B.6). Similar to the absorption studies, we have observed a huge enhancement in emission intensity of C343 in the presence of both intra- and intermolecular i-motif structures (Appendix 4B.7) which has been depicted in the bar diagram (Figure 4B.2A). Interestingly, enhancement in emission intensity of C343 is less for TC5 than the other C-rich sequences (Figure 4B.2A) which may be due to the absence of loop structure in the intramolecular (TC5) structure compared to the intermolecular i-motif DNA. For different intramolecular i-motif structures, the fluorescence intensity enhancement of C343 decreases with increasing length of the C-rich sequences. This can be attributed to the increment in nonradiative decay channels as a result of the less rigidity afforded to the C343 molecules in bigger loops of C-Myc, bcl-2 and hTERT (See Appendix Note 4B.2 for details). Importantly, emission intensity enhancement for C343 has not been observed for ssDNA/RC DNA (in pH 8.5) of the different C-rich sequences (Appendix 4B.8). We have also verified the efficiency of this ligand in physiological condition (pH 7 buffer) by performing the fluorescence experiments of C343 with C-Myc and bcl-2 sequences having higher transitional pH (~ 6.6).[22] However, the extent of fluorescence enhancement is less in comparison to the enhancement in pH 5.2 (Appendix 4B.9), which is an outcome of the lesser amount of folded h-teloc i-motif structures in pH 7. Here, it is important to note that most of the previously reported i-motif sensors have been majorly based on the loops of the intramolecular i-motif structure. Thus, to the best of our knowledge, detection of intermolecular i-motif DNA based on the small fluorescent molecule is absent till now. Hence, it is very interesting to report a small molecule, C343 which can sense both intra- and intermolecular i-motif DNA depending on the unique hemi-protonated cytosine–cytosine (C–C$^+$) base pairing rather than detecting the groove/loop of the i-motif DNA.

4B.2.4 Selectivity Studies with Different DNA Structures

It is of prime importance for any i-motif sensor to be selective for the i-motif structure over the G-quadruplex DNA (GQ DNA) due to their complementarity nature in genome, giving rise to the possibility of these two tetraplex structures coexisting with each other.[8] Keeping this aspect in mind, we have explored the selectivity of C343 with complementary GQ DNA of h-teloc and C-Myc i-motif DNA sequence (Appendix Table 4B.1) in pH 5.2 buffer. Addition of both GQ-h-teloc and GQ-C-Myc DNA does not significantly alter the absorption and fluorescence spectrum of C343 in pH 5.2 likewise in the case of i-motif structures (i.e., huge increment in

absorbance with blue shift and drastic emission intensity enhancement) indicating the absence of any substantial ground and excited state interaction between GQ DNA and C343 (Figure 4B.2B, Appendix 4B.10 and Appendix 4B.11). The extents of emission intensity enhancement for GQ and i-motif DNA can be easily observed in the respective bar diagram in Figure 4B.2B. Similar to GQ DNA, only a slight emission intensity enhancement for C343 has been observed in presence of duplex DNA i.e. ds 26 (Appendix Table 4B.1; Appendix 4B.12 and Figure 4B.2B). Thus, these studies with GQ-DNA (complimentary to i-motif DNA) and duplex DNA suggest the effectiveness of C343 as a selective i-motif ligand.
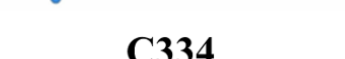

Figure 4B.2 Emission intensity (λ_{ex} 375 nm) of coumarin 343 (C343) in pH 5.2 buffer in absence and presence of (A) different i-motif DNA (40 µM) and (B) different DNA structures (40 µM). F_0 is the emission intensity of C343 in absence of any DNA.

4B.2.5 Plausible Interaction Mechanism

In order to understand the mode of interaction between C343 and i-motif DNA, we have carried out the fluorescence studies with a series of other coumarin derivatives (Figure 4B.3A; Appendix Note 4B.3) having different functionalities in presence of i-motif (h-teloc DNA at pH 5.2) and ssDNA/RC structure (h-teloc DNA at pH 8.5). Interestingly, we observe that all coumarin derivatives, except C3COOH, show a turn-on fluorescence response in the presence of either i-motif DNA or both i-motif DNA and ssDNA/RC structure. To get a deeper insight into the mode of interaction, we have calculated the electron density difference maps (EDDM) of these coumarin structures (Figure 4B.3B). These calculations show that a charge transfer phenomenon is present in these coumarin molecules where the tertiary amine moiety acts as the donor and carbonyl group acts as the acceptor. This type of charge transfer for different

coumarin derivatives in the excited state has also been observed in previous literature.[23] Consequently, the nitrogen atom of tertiary amine has a positively charged character, which helps to interact with both i-motif and ssDNA/RC structures through their negatively charged phosphate backbone. Importantly, the addition of a negative functionality (i.e., carboxylate group) in coumarin derivatives brings the selectivity for only i-motif DNA. There can be two plausible reasons behind this selectivity for i-motif DNA, firstly, these negatively charged coumarin derivatives with tertiary amine functionality can sense the unique hemi-protonated C–C^+ base pairing in i-motif DNA. Secondly, these coumarins experience an additional

(A) Sample	i-motif DNA	ssDNA/RC DNA	(B) EDDM Analysis
C343	✔	✘	
C102	✔	✔	C343
C334	✔	✔	
7C3	✔	✘	C102
C3COOH	✘	✘	
CCVJ	✔	✘	C334

Figure 4B.3 (A) Emission intensity enhancement of different coumarin derivatives and CCVJ in presence of i-motif DNA (h-teloc in pH 5.2 buffer) and ssDNA/RC DNA (h teloc in pH 8.5 buffer). Tick indicates the fluorescence intensity enhancement, and cross exhibits absence of emission intensity enhancement. (B) Iso-surfaces of the electron density difference map (EDDM) of the coumarin derivatives showing the changes in electronic density of the dye due to an excitation.
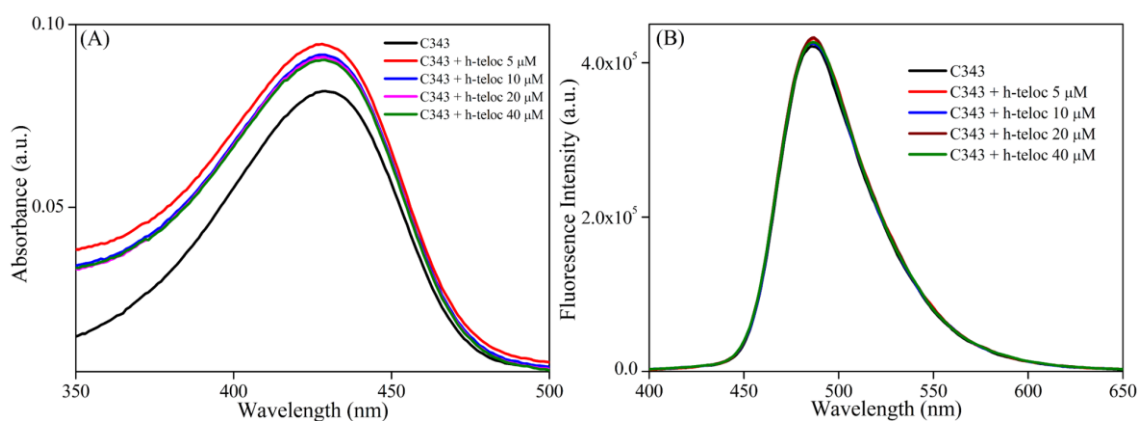
electrostatic repulsion with the negatively charged phosphate backbone, which may be responsible for their lack of interactions with ssDNA/RC structure. However, the sole presence

of a negative charge in coumarin molecules is not sufficient for i-motif detection which has been observed for C3COOH in its inactivity for both i-motif and ssDNA/RC form. This implies that electrostatic interaction between positively charged tertiary amine moiety and negatively charged phosphate backbone also plays an important role along with the electrostatic attraction between negatively charged carboxylate group and hemi-protonated C–C$^+$ base pairs in coumarins for selective i-motif detection. This type of electrostatic interaction leads to the higher rigidity of the respective probe which accounts for reduction in non- radiative decay channels and increment in the fluorescence intensity.[24] We have verified this hypothesis by considering a molecule other than coumarin having similar tertiary amine center and negatively charged carboxylate group, 9-(2-carboxy-2-cyanovinyl)julolidine (CCVJ). It has been observed that CCVJ can selectively detect the i-motif DNA over ssDNA/RC structure (Appendix 4B.18). Thus, we believe that a balance of positive and negative charge character in coumarin molecules is the crucial aspect for the selective i-motif DNA detection.
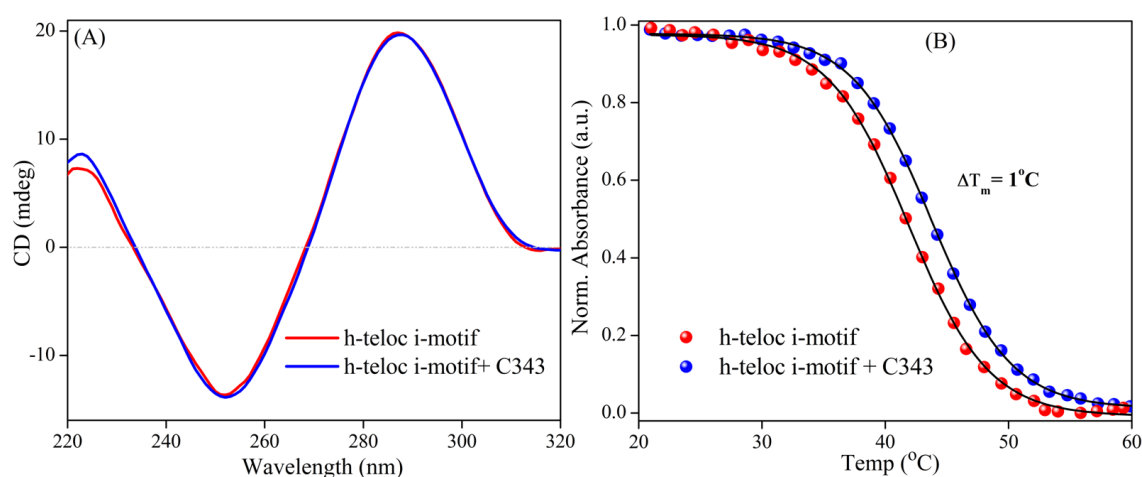
4B.3 Conclusion

In summary, we have demonstrated the fluorescence light-up property of C343 based on the recognition of hemi-protonated C–C$^+$ base pairing in the i-motif DNA. Additionally, C343 shows its versatility in detection by sensing various i-motif DNAs with different chain lengths, sizes, molecularity and loop lengths including both intramolecular and intermolecular structures, unlike other previously reported i-motif ligands. Importantly, C343 exhibits its selectivity for i-motif DNA over its complimentary analogue, GQ-DNA and duplex DNA (ds 26) due to its unique recognition based on hemi-protonated cytosine bases with negatively charged functionality. In previous literature, the recognition is majorly based on the positive charge of the i-motif ligands, which restricts its selectivity towards other DNA structures. In order to know the characteristics of an ideal i-motif ligand, we have systemically studied the interaction between i-motif with various coumarin derivatives, which indicates that balance of positive and negative character in the i-motif ligand is the crucial aspect for the selective detection. We believe that this new approach for i-motif detection will be useful for future designing of i-motif ligands.

4B.4 Appendix Section

Appendix 4B.1 (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).


Appendix 4B.2 (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.

Appendix Note 4B.1 - 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,[25]

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

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.
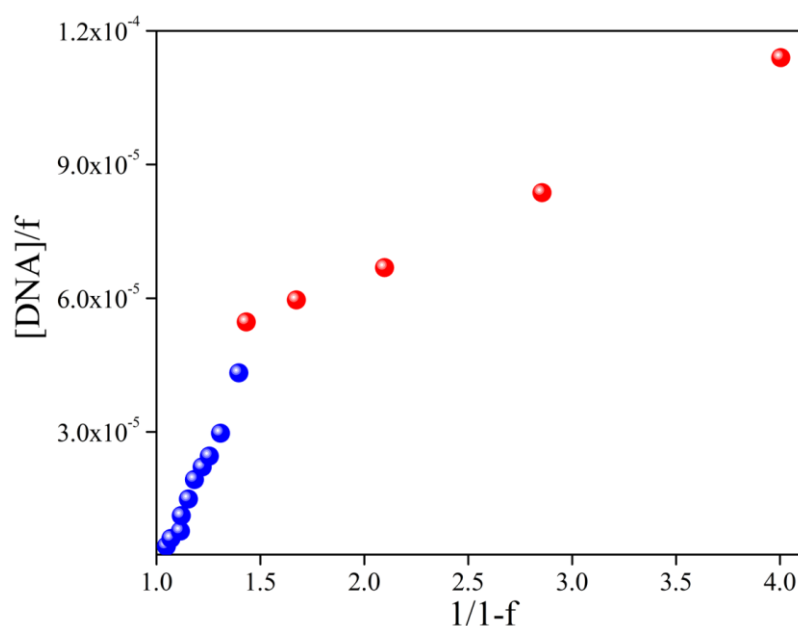
Appendix 4B.3 Modified Scatchard plot of [DNA]/f vs 1/(1-f) for the binding constant determination of h-teloc and C343 complex in pH 5.2 buffer.

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 μM), and the red coloured points indicate higher concentrations of DNA (> 10 μM), 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 6.2×10^5 M^{-1}; whereas in higher concentration, the binding constant is 5.2×10^5 M^{-1}. This type of concentration-dependent binding has also been observed previously in other systems.[26] 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.
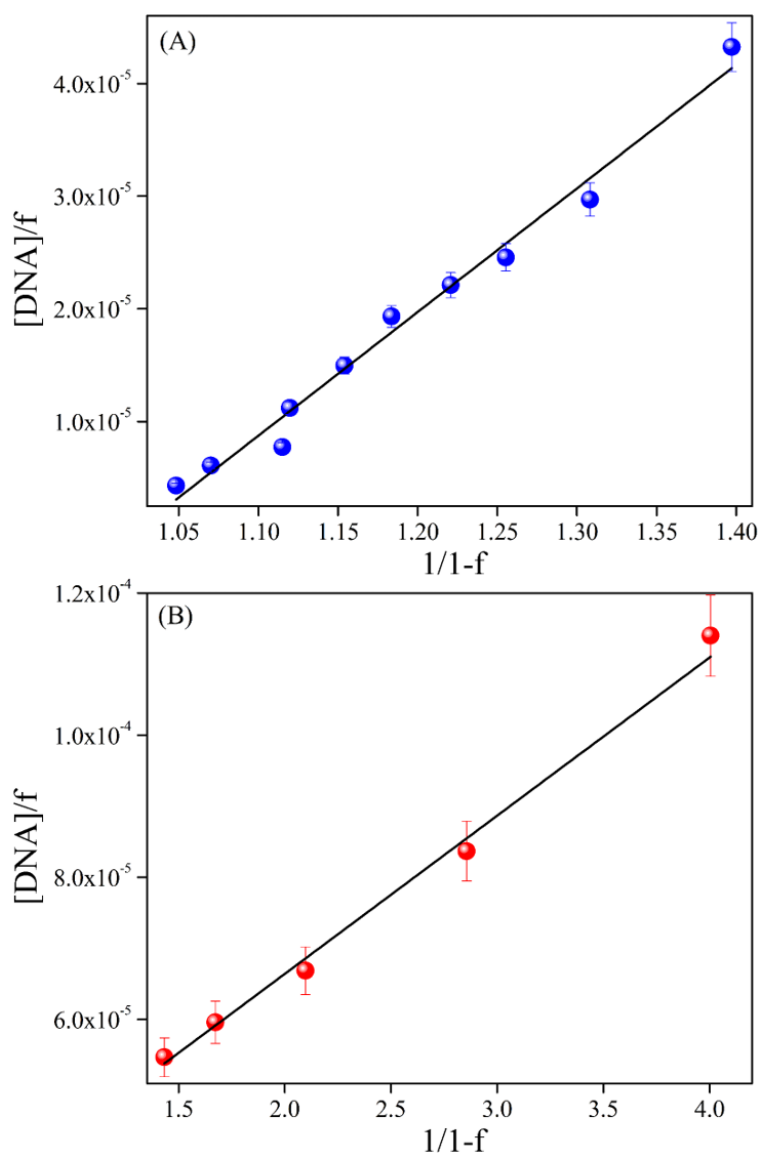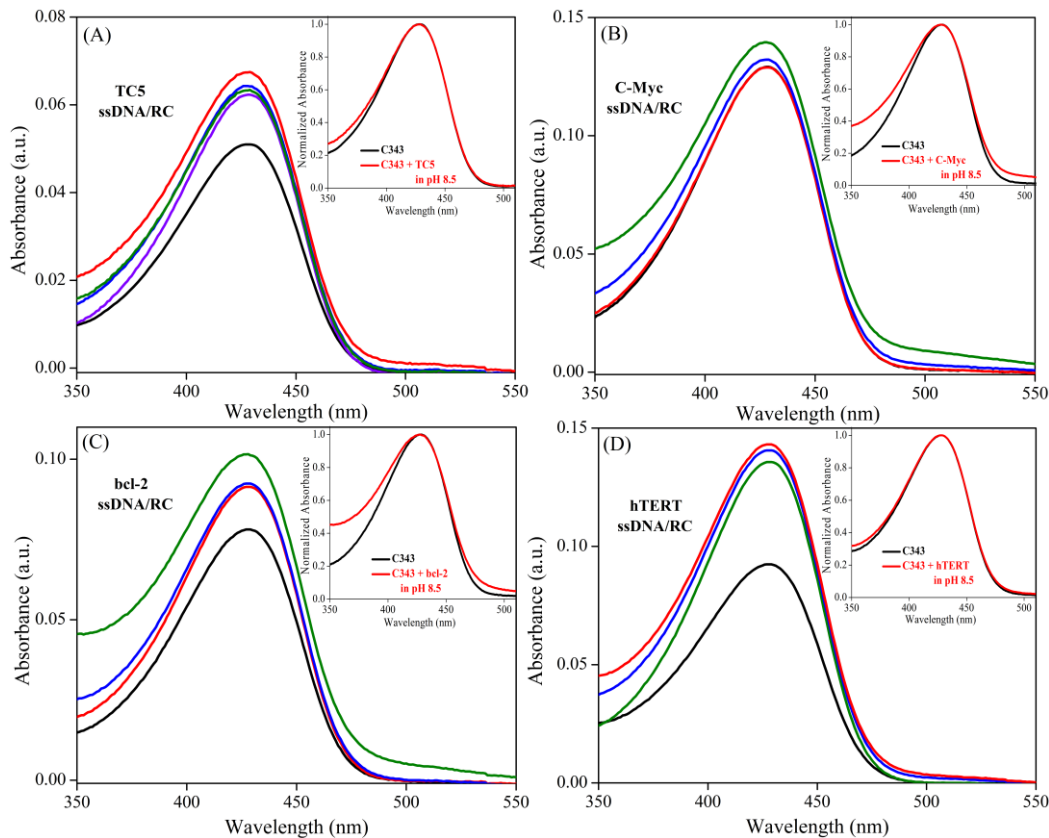

Appendix 4B.4 Modified Scatchard plot of [DNA]/f vs 1/(1-f) for the binding constant determination of h-teloc and C343 complex in pH 5.2 buffer (A) at lower concentrations (< 10 μM) and (B) at higher concentrations (> 10 μM).
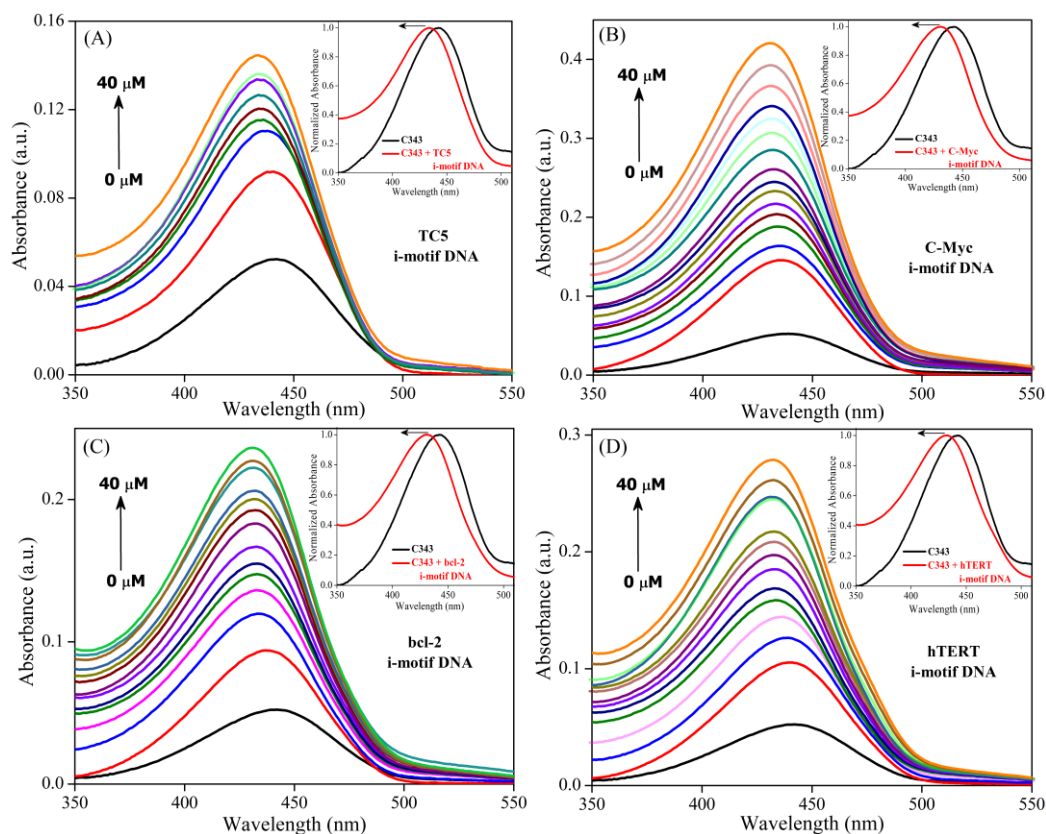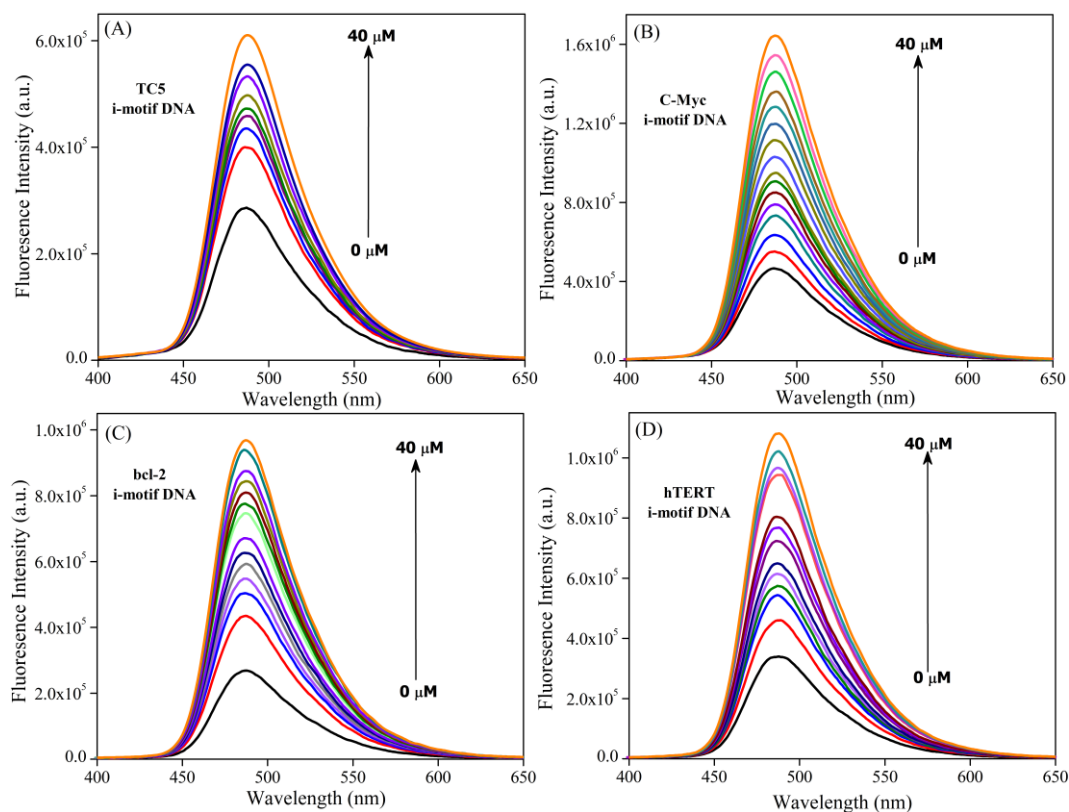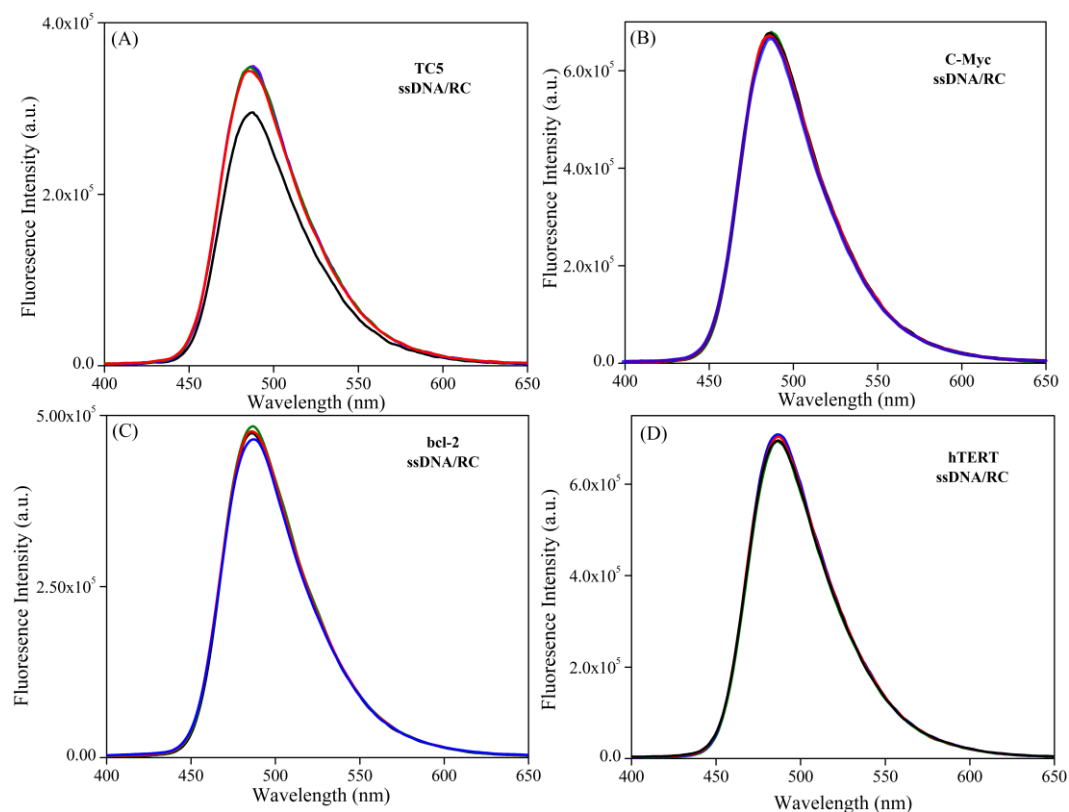
Appendix 4B.5 Absorption spectra of coumarin 343 (C343) in pH 5.2 buffer with increasing concentration of different i-motif DNA, (A) TC5, (B) C-Myc, (C) bcl-2 and (D) hTERT.

Appendix 4B.6 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.

Appendix 4B.7 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.

Appendix 4B.8 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.

Appendix Note 4B.2 - 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(CCCTAACCCTAACCCTAACCCTAA) **class-I**
C-Myc	d(CCCCACCTTCCCCACCCTCCCCACCCTCCCC) **class II**
bcl-2	d(CAGCCCCGCTCCCGCCCCCTTCCTCCCGCGCCCGCCCCT) **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 strutures 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.[22, 27]

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 (Figure 4B.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 Appendix 4B.9, Appendix 4B.10, Appendix 4B.11). 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 Appendix 4B.14) 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.[22] 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.
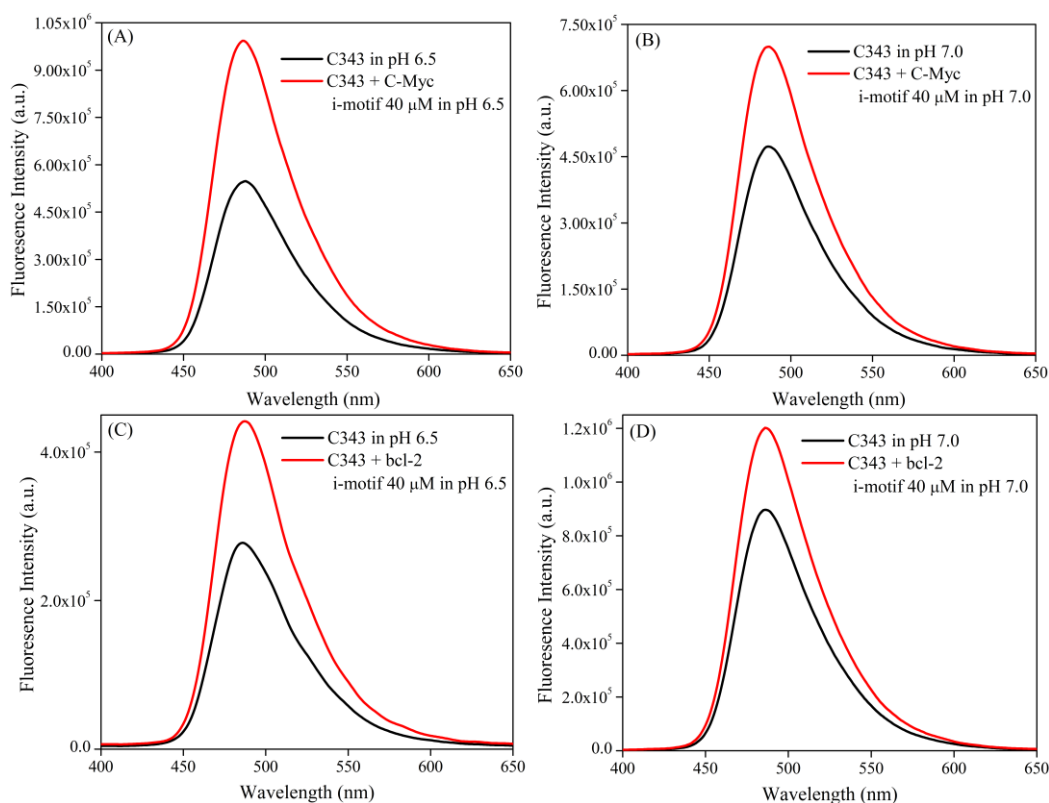
Appendix 4B.9 Fluorescence response of C343, in presence of C-Myc i-motif DNA (40 µM) in (A) pH 6.5 and (B) pH 7 buffer and in presence of bcl-2 i-motif DNA (40 µM) in (C) pH 6.5 and (D) pH 7 buffer.
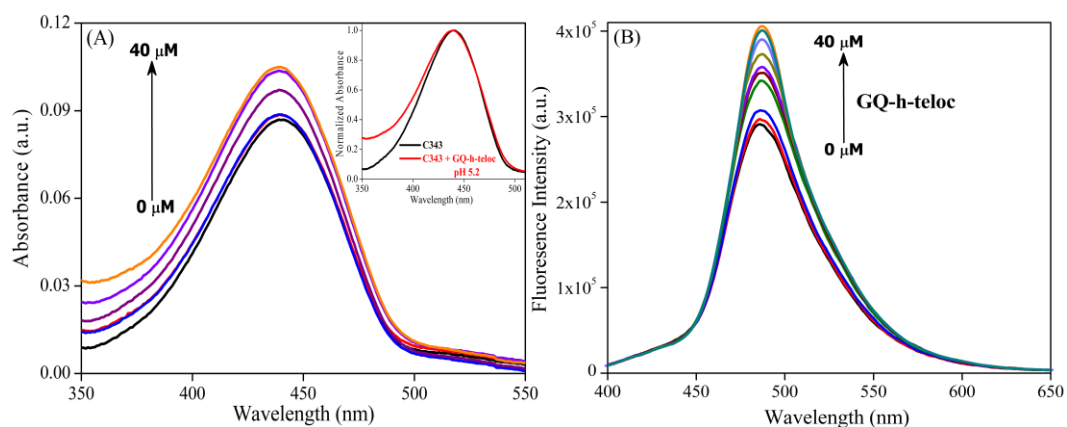
Appendix 4B.10 (A) Absorption spectra and (B) fluorescence spectra of coumarin 343 (C343) with increasing concentration of GQ-h-teloc DNA in pH 5.2 buffer (λ_{ex} 375 nm). Inset of Appendix 4B.10A shows the absence of any shift for the C343 absorption in the presence of GQ-h-teloc DNA.
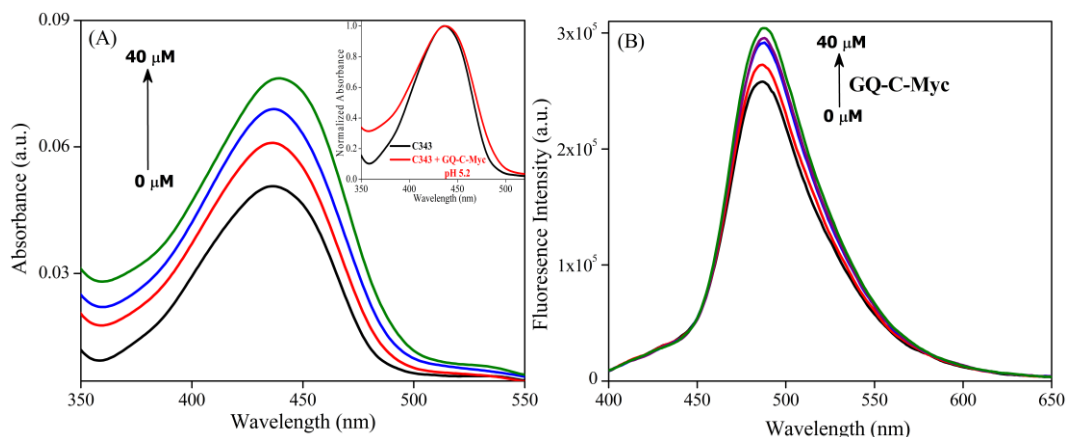
Appendix 4B.11 (A) Absorption spectra and (B) fluorescence spectra of coumarin 343 (C343) with increasing concentration of GQ-C-Myc DNA in pH 5.2 buffer (λ_{ex} 375 nm). Inset of Appendix 4B.11A shows the absence of any shift for the C343 absorption in the presence of GQ-C-Myc DNA.
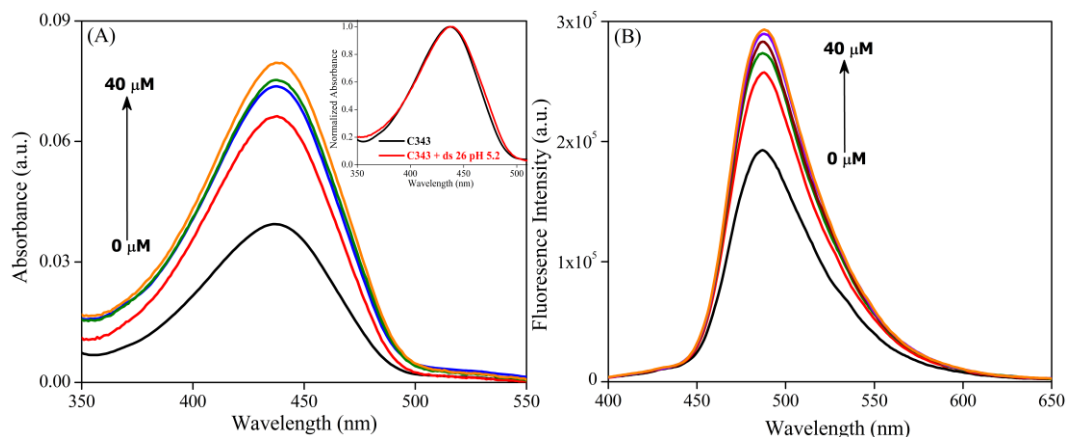
Appendix 4B.12 (A) Absorption spectra and (B) fluorescence spectra of coumarin 343 (C343) with increasing concentration of ds 26 DNA in pH 5.2 buffer (λ_{ex} 375 nm). Inset of Appendix 4B.12A shows the absence of any shift for the C343 absorption in the presence of ds 26 DNA.

**Appendix Table 4B.1** 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)

Appendix Note 4B.3 - Systematic study of the different coumarin derivatives with i-motif and ssDNA/RC DNA:

C343

C102

C334

7C3

C3COOH

CCVJ
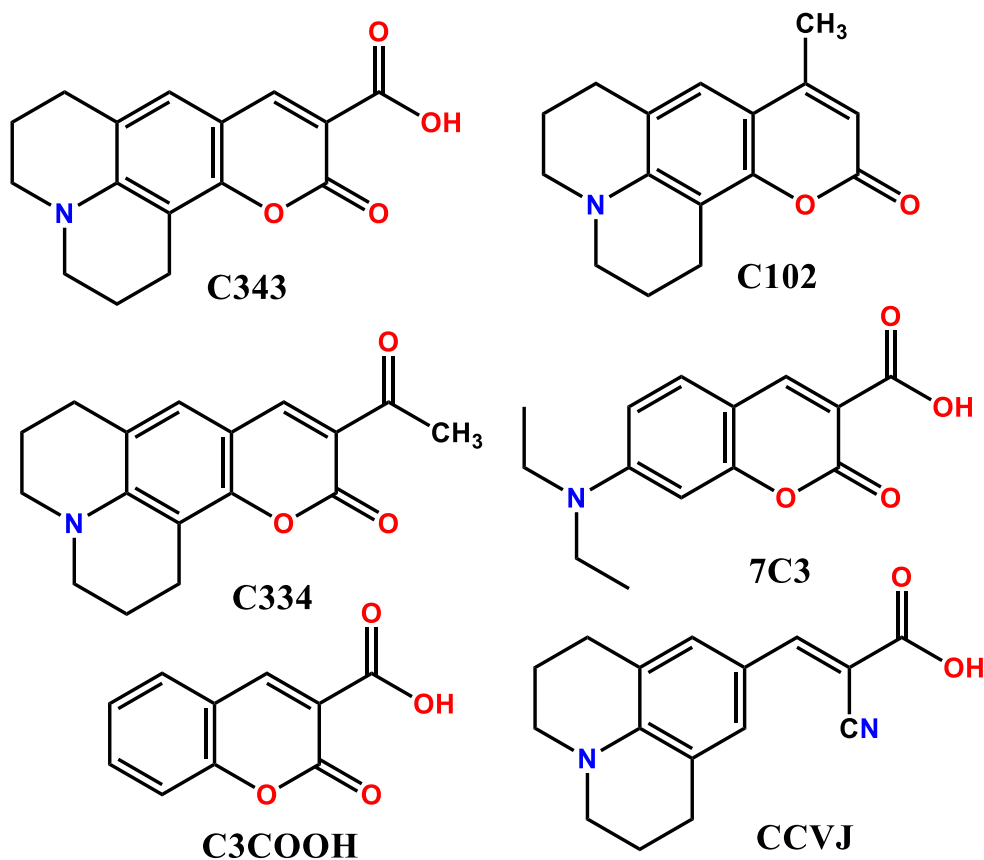
Appendix Scheme 4B.1 Chemical structure of different molecules used in this study.
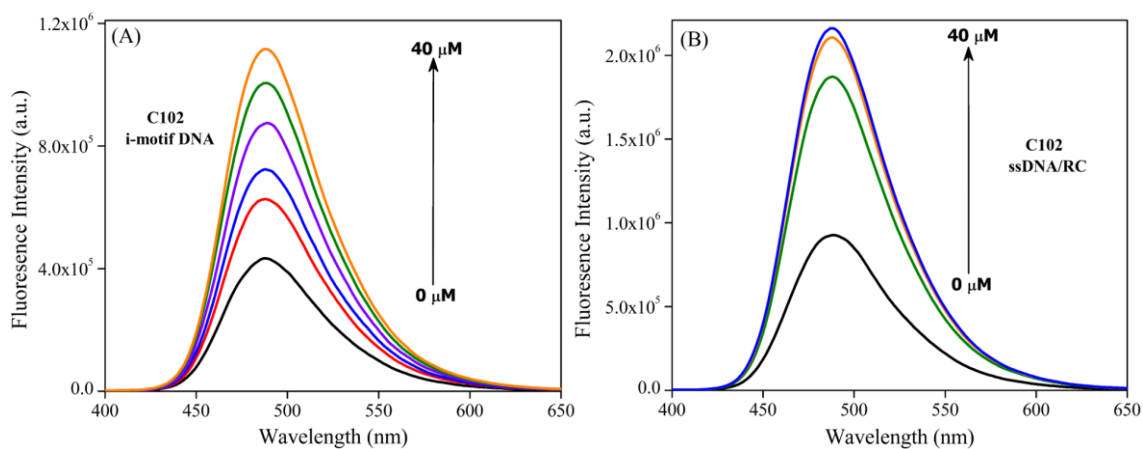
Appendix 4B.13 Fluorescence response (λ_{ex} 375 nm) of coumarin 102 (C102) in the presence of (A) i-motif DNA at pH 5.4 buffer and (B) in the presence of ssDNA/RC DNA at pH 8.5 buffer.
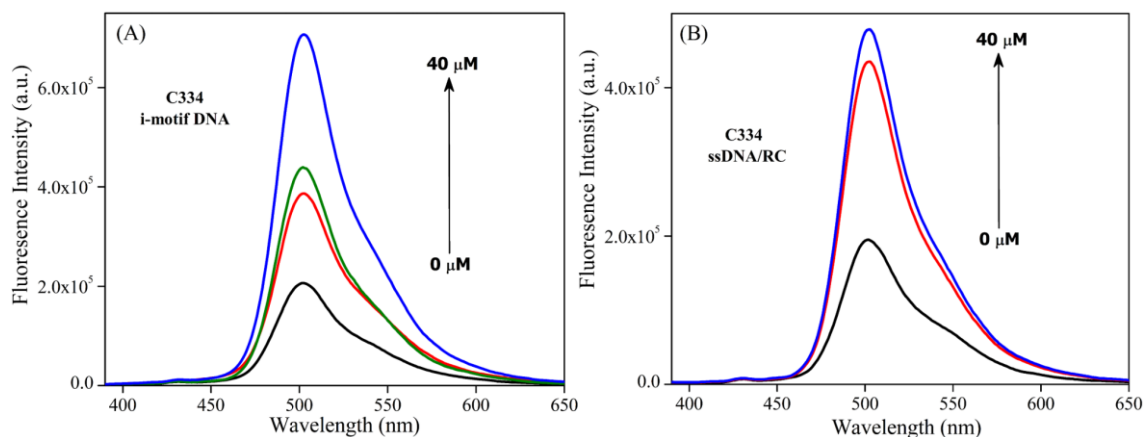
Appendix 4B.14 Fluorescence response (λ_{ex} 375 nm) of coumarin 334 (C334) in the presence of (A) i-motif DNA at pH 5.4 buffer and (B) in the presence of ssDNA/RC DNA at pH 8.5 buffer.
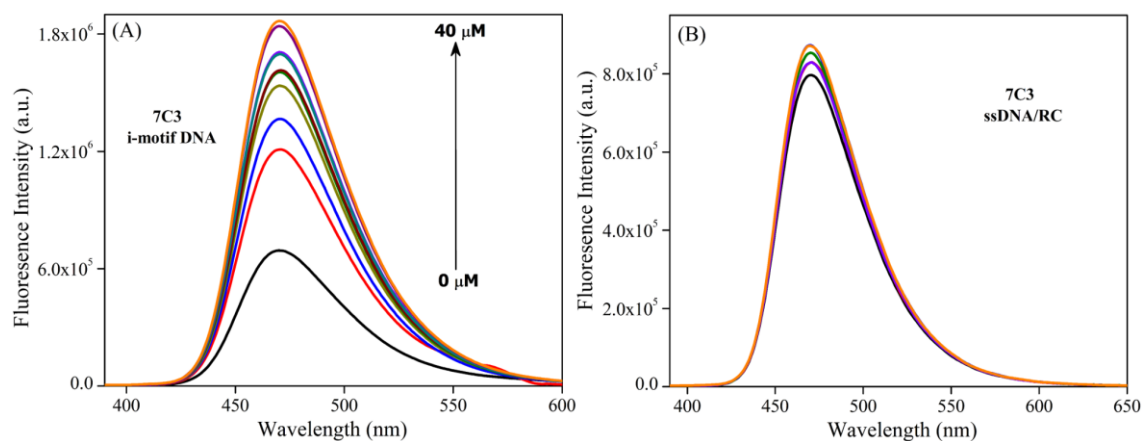
Appendix 4B.15 Fluorescence response (λ_{ex} 375 nm) of 7-(Diethylamino)coumarin-3-carboxylic acid (7C3) in the presence of (A) i-motif DNA at pH 5.4 buffer and (B) in the presence of ssDNA/RC DNA at pH 8.5 buffer.
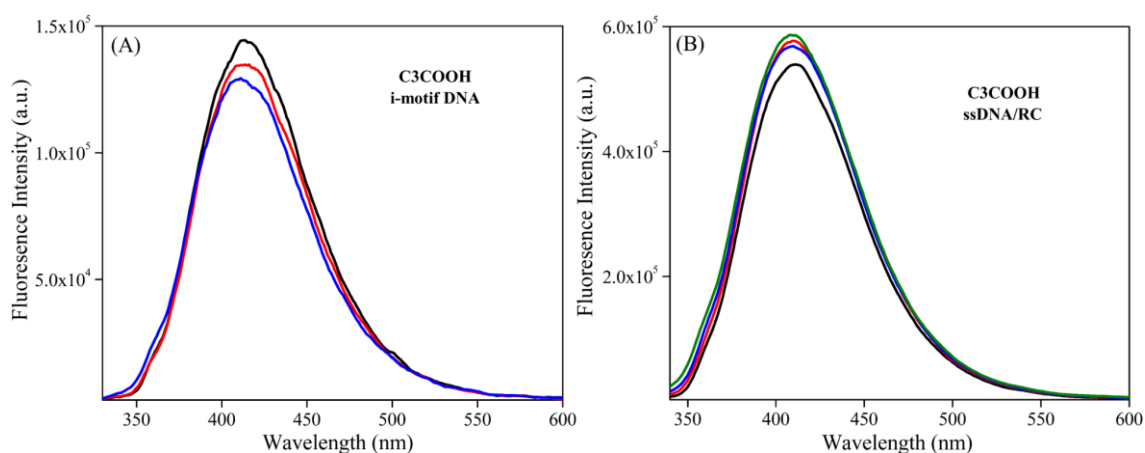
Appendix 4B.16 Fluorescence response (λ_{ex} 375 nm) of Coumarin-3-carboxylic acid (C3COOH) in the presence of (A) i-motif DNA at pH 5.4 buffer and (B) in the presence of ssDNA/RC DNA at pH 8.5 buffer.

Appendix 4B.17 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.

Appendix 4B.18 Fluorescence response (λ_{ex} 375 nm) of 9-(2-carboxy-2-cyanovinyl)julolidine (CCVJ) in the presence of (A) i-motif DNA at pH 5.4 buffer and (B) in the presence of ssDNA/RC DNA at pH 8.5 buffer.

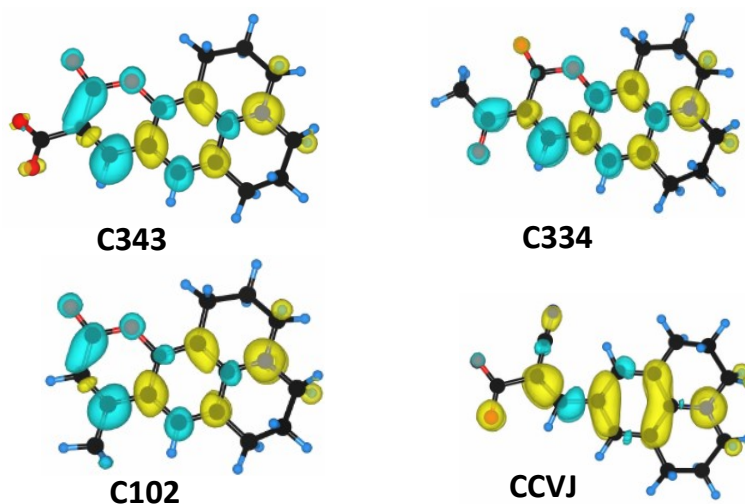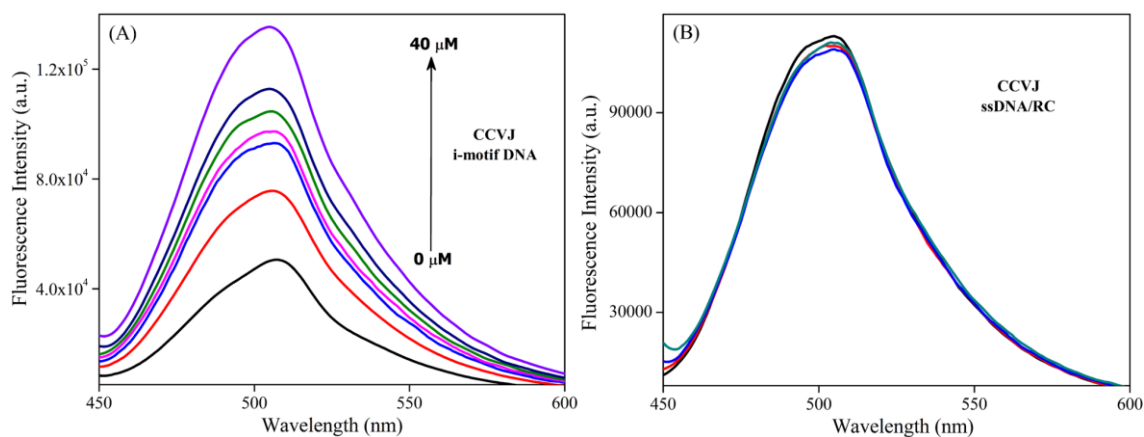We have performed the fluorescence studies to monitor the interaction between different coumarin derivatives with both the i-motif DNA (at pH 5.2 buffer) and the ssDNA/RC DNA (at pH 8.5 buffer). In the case of both C102 and C334, the emission intensity increases with increasing concentration of both the i-motif and ssDNA/RC DNA (Appendix 4B.13 and Appendix 4B.14). Whereas, for 7C3 (Appendix 4B.15), 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 (Appendix 4B.16) 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 (Appendix 4B.17). 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. Appendix 4B.18) 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. We have also performed the fluorescence experiments between CCVJ and different DNA structures in pH 5.2 buffer which indicates its selectivity towards i-motif DNA (Appendix 4B.19). Thus, a balance between positive and negative character is necessary for a molecule to act as a selective i-motif ligand.

Appendix Table 4B.2 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%)

4B.5 Experimental

We have carried out all the experiments with 10 mM phosphate buffer solutions of different pH. Before any experiments, annealing of the DNA samples was performed at 90 °C for 10 min and stored at 4 °C for 24 h. Concentrations of DNA samples were calculated using the molar extinction coefficient at 260 nm provided by IDT, USA. It is necessary to mention here that IDT has determined the molar extinction coefficient using the nearest neighbour approximation model. The concentration of C343 was calculated using the molar extinction coefficient of ~45 000 M^{-1} cm^{-1} at 436 nm.[28] The melting temperature (T_m) was calculated from the sigmoidal curve fit of the melting profile.

4B.5.1 Computational Details for Electron Density Difference Maps (EDDM) Calculation:

All the computational calculations were performed with density functional theory (DFT) in Gaussian 09 software.[29] 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.[30] Further, we computed absorption spectra by Time-dependent DFT (TD-DFT) by B3LYP functional.[31] 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 functional[32] along with conventional hybrid functional (B3LYP). We have also incorporated implicit solvent models by SCRF method using IEFPCM model for water solvent.[33] We have computed absorption spectra for 15 number of states. First two prominent (bright) transitions are given in the Appendix Table 4B.2. Later, we computed electron density difference map for first bright transition (having considerable oscillatory strength) by Gauss-Sum software.[34] 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 Appendix Table 4B.2. 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.

4B.6 References

1. S. Kendrick, H.-J. Kang, M. P. Alam, M. M. Madathil, P. Agrawal, V. Gokhale, D. Yang, S. M. Hecht and L. H. Hurley, *J. Am. Chem. Soc.*, 2014, **136**, 4161-4171.

2. F. Wang, X. Liu and I. Willner, 2015, **- 54**, - 1129.

3. Y. Dong, Z. Yang and D. Liu, *Acc. Chem. Res.*, 2014, **47**, 1853-1860.

4. I. V. Nesterova and E. E. Nesterov, *J. Am. Chem. Soc.*, 2014, **136**, 8843-8846.

5. Z.-G. Wang, J. Elbaz and I. Willner, *Nano Lett.*, 2011, **11**, 304-309.

6. K. Gehring, J.-L. Leroy and M. Guéron, *Nature*, 1993, **363**, 561.

7. H. A. Day, P. Pavlou and Z. A. E. Waller, *Bioorg. Med. Chem.*, 2014, **22**, 4407-4418.

8. M. Zeraati, D. B. Langley, P. Schofield, A. L. Moye, R. Rouet, W. E. Hughes, T. M. Bryan, M. E. Dinger and D. Christ, *Nat. Chem.*, 2018, **10**, 631-637.

9. (a) J.-L. Mergny, *Biochemistry*, 1999, **38**, 1573-1581; (b) B. Xu, X. Wu, E. K. L. Yeow and F. Shao, *Chem. Commun.*, 2014, **50**, 6402-6405; (c) J. Choi, S. Kim, T. Tachikawa, M. Fujitsuka and T. Majima, *J. Am. Chem. Soc.*, 2011, **133**, 16146-16153.

10. (a) F. Wang, X. Liu and I. Willner, *Angew. Chem., Int. Ed.*, 2015, **54**, 1098-1129; (b) Y. Li, X. Han, Y. Yan, Y. Cao, X. Xiang, S. Wang, B. Zhao and X. Guo, *Anal. Chem.*, 2018, **90**, 2996-3000.

11. I. J. Lee, S. P. Patil, K. Fhayli, S. Alsaiari and N. M. Khashab, *Chem. Commun.*, 2015, **51**, 3747-3749.

12. Y. Shi, H. Sun, J. Xiang, H. Chen, Q. Yang, A. Guan, Q. Li, L. Yu and Y. Tang, *Chem. Commun.*, 2014, **50**, 15385-15388.

13. Q. Sheng, J. C. Neaverson, T. Mahmoud, C. E. M. Stevenson, S. E. Matthews and Z. A. E. Waller, *Org. Biomol. Chem.*, 2017, **15**, 5669-5673.

14. L. Xu, S. Hong, N. Sun, K. Wang, L. Zhou, L. Ji and R. Pei, *Chem. Commun.*, 2016, **52**, 179-182.

15. S. Shi, X. Geng, J. Zhao, T. Yao, C. Wang, D. Yang, L. Zheng and L. Ji, *Biochimie*, 2010, **92**, 370-377.

16. M. Debnath, S. Ghosh, A. Chauhan, R. Paul, K. Bhattacharyya and J. Dash, *Chem. Sci.*, 2017, **8**, 7448-7456.

17. L. Lu, M. Wang, L.-J. Liu, C.-Y. Wong, C.-H. Leung and D.-L. Ma, *Chem. Commun.*, 2015, **51**, 9953-9956.

18. E. P. Wright, H. A. Day, A. M. Ibrahim, J. Kumar, L. J. E. Boswell, C. Huguin, C. E. M. Stevenson, K. Pors and Z. A. E. Waller, *Sci. Rep.*, 2016, **6**, 39456.

19. L. Xu, J. Wang, N. Sun, M. Liu, Y. Cao, Z. Wang and R. Pei, *Chem. Commun.*, 2016, **52**, 14330-14333.

20. (a) G. Signore, R. Nifosì, L. Albertazzi, B. Storti and R. Bizzarri, *J. Am. Chem. Soc.*, 2010, **132**, 1276-1288; (b) B. G. Lake, *Food Chem. Toxicol.*, 1999, **37**, 423-453.

21. R. E. Riter, E. P. Undiks and N. E. Levinger, *J. Am. Chem. Soc.*, 1998, **120**, 6062-6067.

22. T. A. Brooks, S. Kendrick and L. Hurley, *FEBS J.*, 2010, **277**, 3459-3469.

23. (a) B. Raju B and T. S. Varadarajan, *J. Phys. Chem.*, 1994, **98**, 8903-8905; (b) S. L. Gilat, A. Adronov and J. M. J. Fréchet, *Angew. Chem., Int. Ed.*, 1999, **38**, 1422-1427; (c) A. J. Campillo, J. H. Clark, S. L. Shapiro, K. R. Winn and P. K. Woodbridge, *Chem. Phys. Lett.*, 1979, **67**, 218-222; (d) X. Liu, Z. Xu and J. M. Cole, *J. Phys. Chem. C*, 2013, **117**, 16584-16595.

24. J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Springer US, 3rd edition edn., 2006.

25. (a) E. F. Healy, *J. Chem. Educ.*, 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**, 4804-4809.

26. S. N. Sahu, S. K. Padhan and P. K. Sahu, *RSC Adv.*, 2016, **6**, 90322-90330.

27. A. S. Gouda, M. S. Amine and E. B. Pedersen, *Org. Biomol. Chem.*, 2017, **15**, 6613-6621.

28. M. R. Webb and J. E. T. Corrie, *Biophys. J.*, 2001, **81**, 1562-1569.

29. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, *Gaussian, Inc.: Pittsburgh, PA,*, 2016.

30. (a) A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098-3100; (b) C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785-789.

31. T. Yanai, D. P. Tew and N. C. Handy, *Chem. Phys. Lett.*, 2004, **393**, 51-57.

32. N. M. O'Boyle, A. L. Tenderholt and K. M. Langner, *J. Comput. Chem.*, 2008, **29**, 839-845.

33. G. Scalmani and M. J. Frisch, *J. Chem. Phys.*, 2010, **132**, 114110.

34. *Time-Dependent Density Functional Theory*, Springer-Verlag Berlin Heidelberg, Berlin, 2006.

The End