

# Molecular modeling and dynamics studies of substituted 3,4-Bis(indolin-3-yl)pyrrole-2,5-dione with G-quadruplex and i-motif DNA

*Project report submitted  
in partial fulfilment of the requirement for the degree of*

**Bachelor in Science**

By

**Akanksha Yadav  
(15B030007)**



**DEPARTMENT OF CHEMISTRY  
INDIAN INSTITUTE OF TECHNOLOGY BOMBAY**

(May 2019)

# CERTIFICATE

It is certified that the work contained in the project report titled Molecular modeling and dynamics studies of substituted 3,4-Bis(indolin-3-yl)pyrrole-2,5-dione with G-quadruplex and i-motif DNA by *Akanksha Yadav* has been carried out under my supervision.

Signature of Supervisor  
Name: Dr. Pradeepkumar P.I.  
Department of Chemistry  
IIT Bombay  
Date: 04/05/2019

## Contents

Introduction.....	3
<i>G-quadruplex versus i-motif DNA</i> .....	3
<i>Computational methods employed in studying quadruplex/ligand complexes</i> .....	4
Motivation.....	5
Methods.....	5
Results and Discussion .....	6
<i>c-MYC and c-KIT G-quadruplex dynamics studies</i> .....	7
<i>Human telomere i-Motif dynamics studies</i> .....	11
Conclusions.....	13
Bibliography .....	14

## Introduction

Nucleic acids are known to adopt various non-canonical structures of biological relevance in cells. Two types of four-stranded structures that are widely studied include: the G-quadruplex (G4) and the i-motif. Because the formation of G4 structures at physiological temperature and pH both *in vivo* and *in vitro* has been reported, there has been growing interest in its research owing to the biological significance. On the contrary, there has been less focus on i-motifs which require hemi-protonated cytosine base-pairing and therefore thought to be stable at acidic pH only.<sup>[1]</sup> But recent studies have shown examples of its physiological relevance in genomic DNA sequences forming i-motifs at neutral pH<sup>[2,3]</sup>.

The therapeutic interest to target these quadruplex DNA structures has driven the synthesis of many small molecule ligands. I aim to study the molecular interactions of one such ligand with quadruplex and i-motif structures.

### G-quadruplex versus i-motif DNA

The “four-stranded” structure, called G-quadruplex, is built from G-rich sequences of DNA, among  $\pi$ - $\pi$  stacked quartets of hydrogen-bonded guanine bases (i.e Gquartets,G4), via Hoogsteen pairings, with an essential alkali metal ion positioned in the interior channel of each G-quartet. The G-quartets are linked together by intervening variable length sequences, loops, arranged on the outside of the relatively rigid and stable G4 core. Quadruplex-forming G-rich sequences are also found in a number of cancer-related genes such as c-myc<sup>[4]</sup> and c-kit<sup>[5]</sup>. G-quadruplex interacting agents with specificity for binding to silencer element can be good anticancer agents.<sup>[6]</sup>

i-Motifs are four-stranded DNA secondary structures and can form in sequences which are rich in cytosine. They are stabilised by acidic conditions and comprise of two parallel-stranded DNA duplexes held together in an antiparallel orientation by intercalated, cytosine–cytosine<sup>+</sup> base pairs. Because of their pH dependent folding /unfolding (sensitivity), i-motif forming DNA sequences have been used extensively as

pH switches with applications in nanotechnology – pH sensors<sup>[7]</sup>, logic gate devices<sup>[8]</sup>, etc.

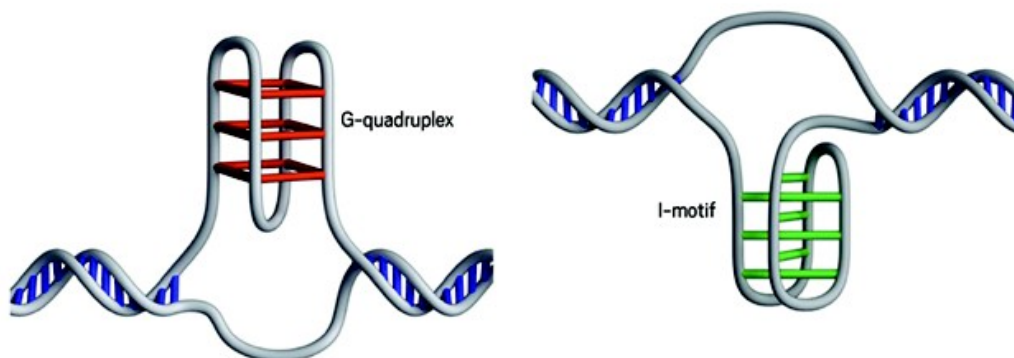


Fig. 1: Schematic of G-quadruplex and i-motif structure. Adapted from *Nature Chemistry* volume 10, 631–637 (2018)

### Computational methods employed in studying quadruplex/ligand complexes

A number of computational studies, carried out to date, have employed biomolecular simulation methods to obtain a superior insight into the structure and interactions of G-quadruplex nucleic acids with small molecule ligands. Molecular modelling, docking, molecular dynamics simulations along with algorithms like metadynamics, umbrella sampling are commonly used in such computations. Dynamic behaviour, conformational properties and energetics have been addressed via such studies. On the other hand, *in silico* studies of i-motif structures are scarce in literature. This is in part due to the complexities in simulating low pH environment conditions required for its stability. Yet, there are a few reports that discuss about the deprotonation mechanism of i-motif DNA<sup>[9]</sup>, and high temperature unfolding simulation of i-motif DNA<sup>[10]</sup> that give ways to model i-motif structures.

The general purpose of this project was to study comparative stability, binding energy and interactions of substituted 3,4-Bis(indolin-3-yl)pyrrole-2,5-dione with c-myc and c-kit G-quadruplex DNA and human telomere fragment i-motif forming DNA.

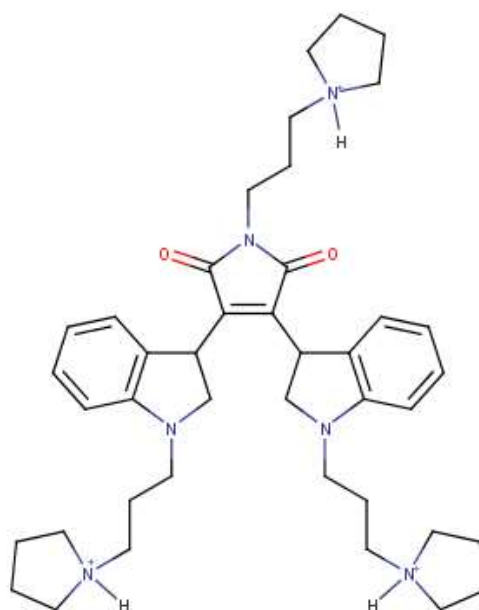
## Motivation

One of the 3,4-Bis(indolin-3-yl)pyrrole-2,5-dione derivatives synthesized in our lab has three- carbon side chains attached to a pyrrolidinium ring. The ligand has a charge of +3 and hence expected to show electrostatic interactions with the backbone of DNA which is negatively charged. Aromatic rings in the core could also contribute to stacking interactions with the nucleotide bases.

Biophysical studies of this ligand in our lab gave following results:

- (i) Stabilization of c-myc and c-kit G4 DNA (CD melting and CD titration)
- (ii) Destabilization of ATXN2L and DAP i-motif DNA (CD melting and CD titration)

Hence, I sought to explain and confirm the experimental results by studying molecular interactions and pathways of solvated complex of above mentioned ligand with G4 and i-motif DNA.



*Fig. 2: Chemical structure of the 3,4-Bis(indolin-3-yl)pyrrole-2,5-dione derivative under study*

## Methods

3D structure of the ligand was obtained by optimization using PM3 and HF/6-311+G\*\* basis sets in Gaussian09<sup>[11]</sup>. Optimized ligand was docked with c-MYC (PDB:2L7V) and c-KIT1 (PDB: 2O3M) G-quadruplex DNAs using the Lamarckian genetic algorithm in AutoDock 4.2<sup>[12]</sup> The structures were modified prior to docking to facilitate further studies. The whole space spanning the DNA was used as the grid and 250 independent

docked conformations were obtained as output. Clustering analysis on these conformations and estimates of binding energies helped to filter out possible binding poses to be used for further simulation. The generalized amber force field (GAFF)<sup>[13]</sup> and ff09bsc0<sup>[14]</sup> force fields were used to provide parameters for the ligand and the DNA respectively in AMBER 14. RESP<sup>[15]</sup> charges were derived using the antechamber<sup>[16]</sup> module of AMBER 14. The systems were solvated using TIP3P water molecules in a 10Å truncated octahedron. K<sup>+</sup> ions were added to neutralize the system. This system was then subjected to 10,000 steps of restraint and 5000 steps of unrestrained minimization, in order, 100ps of heating to 300K and 100ps of density equilibration. The system was then subjected to 200ns of unrestrained dynamics with coordinates saved every 2ps. Temperature was kept constant at 300K using a Langevin thermostat. RMSD, atomic fluctuations (RMSF), Hoogsteen bond occupancies, h-bonding and distances were calculated using CPPTRAJ module of AMBER 14. 3.5Å and 130° cutoff was used to define hydrogen bond. Binding free energies were estimated using MM-PBSA<sup>[17]</sup> (MMPBSA.py) method over the last 50ns. UCSF Chimera<sup>[18]</sup> was used for the visualization of molecular dynamics simulations, and calculation of distances. PyMOL (<http://www.pymol.org>) was used for generating images.

i-motif structure used for the study was a modified human telomere fragment (PDB: 1ELN). The non-standard protonated cytosines were stripped of hydrogens and converted to standard cytosines. Four systems were made using the procedure mentioned above – only i-motif DNA, i-motif DNA + ligand at 3 different initial positions. Two 50ns MD simulation at 300K and 500K respectively were carried out for only i-motif DNA system. Three 50ns MD simulation were carried out for the three systems with ligand. Analysis of trajectories was done using the same modules as mentioned above.

## Results and Discussion

The structure of the ligand, as shown below, was modeled using GaussView followed by optimization at two theory levels – PM3 followed by HF/6-311+G\*\*

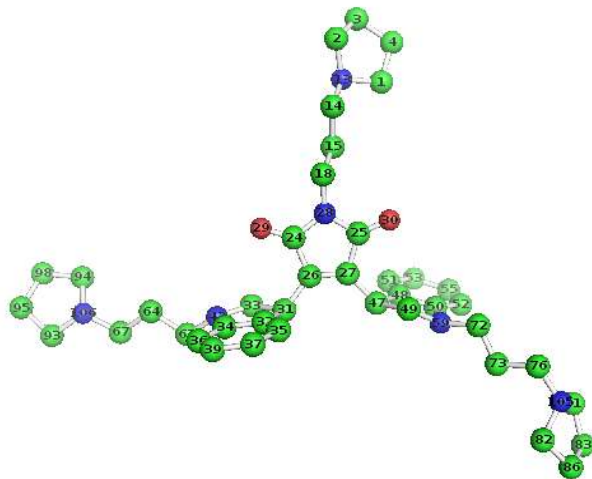


Fig. 3: Optimized structure of ligand. Carbon is colored in green, oxygen in red, and nitrogen in blue. Hydrogens are omitted in the view.

### c-MYC and c-KIT G-quadruplex dynamics studies

Docking revealed some favourable binding poses for the ligand with c-MYC and c-KIT DNA. With c-KIT, ligand positions below the bottom quartet in such a way that the side chains bearing pyrrolidinium cation can interact with the phosphate backbone oxygens. While in the case of c-MYC, the ligand stacks on the top quartet towards the 5' end with the benzene ring over one of the guanines and two of the side chains oriented close to the backbone.

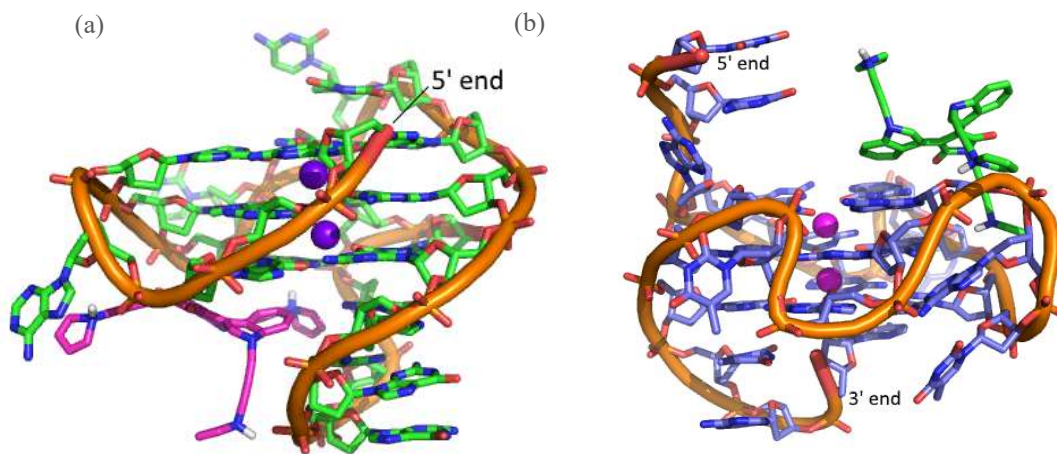
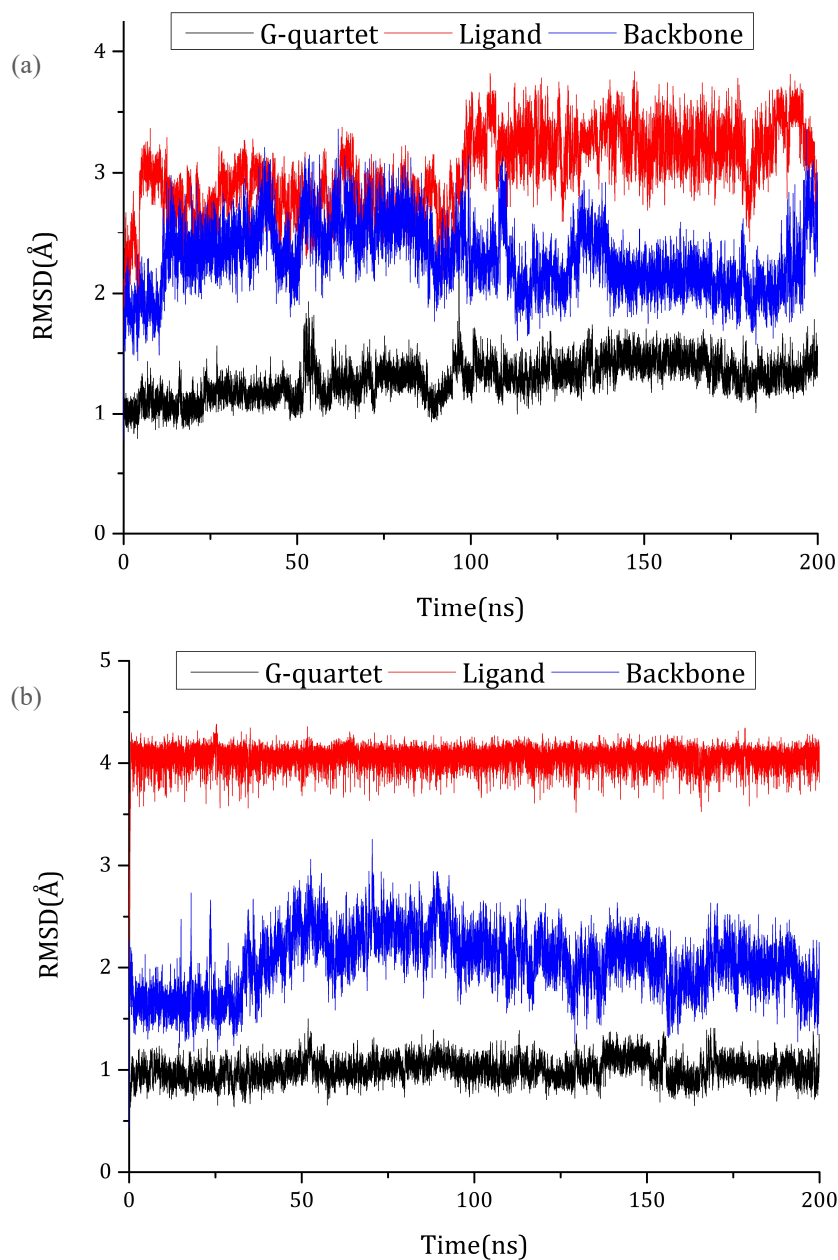


Fig. 4: Initial binding pose at 0ns obtained from docking studies of ligand bound to (a) c-KIT DNA (b) c-MYC DNA. K<sup>+</sup> ions are shown as spheres.

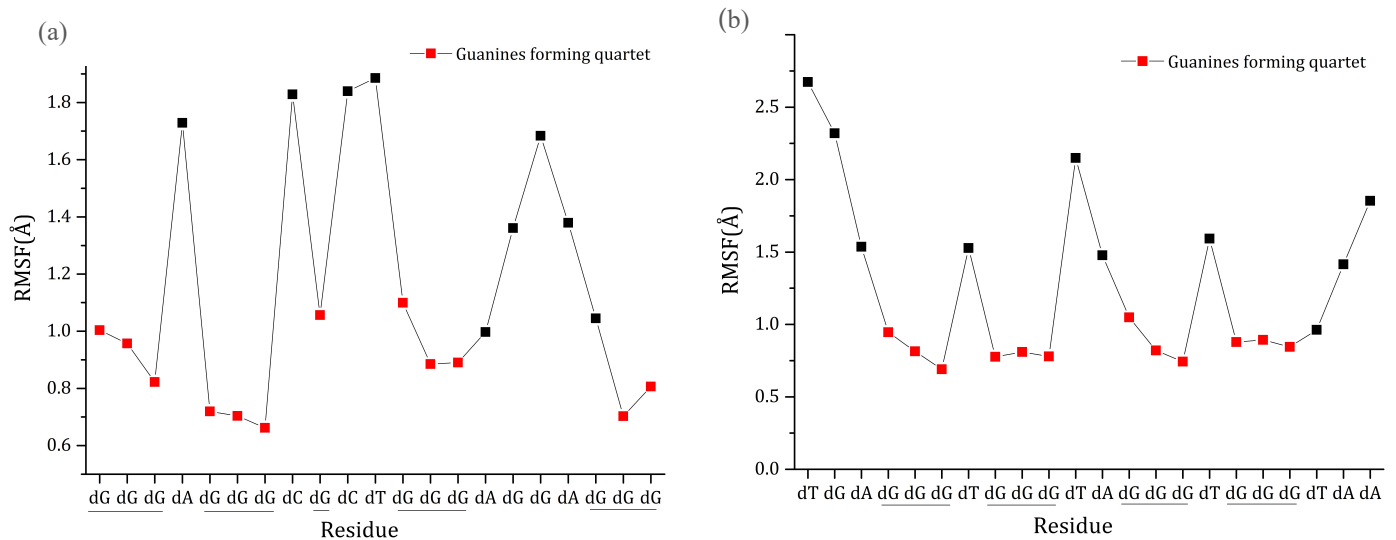


To verify the conformational stability of the complexes, root mean square deviation (RMSD) of the residues w.r.t the first frame is observed for the 200ns long simulation.



*Fig. 5: Root mean square deviation (RMSD) graphs of (a) ligand bound c-KIT complex; and (b) ligand bound c-MYC complex. RMSDs of the backbone (blue), ligand (red) and G-quartet core (black) are plotted against time. The G-quartet core includes the top, middle and bottom G-quartets and backbone atoms include C3', O3', C5', O5', P, OP1 and OP2.*

RMSD of the quartet core for both c-KIT and c-MYC complex show minimal fluctuations, hence both the structures seem to be stabilized by the binding of the ligand. In case of c-KIT, the ligand shows conformational instability as compared to when bound to c-MYC. At around 100ns, there appears to be a change in the conformation of the ligand bound to c-KIT after which it seems to be stable. Backbone atoms for both complex show greater deviations because of some highly dynamic solvent-exposed residues. This is apparent from the root mean square fluctuations averaged over time for individual residues as shown below. Highlighted in red are the guanines that form part of the quartet core. As seen from the graph, these guanines show minimum fluctuations and hence form stable G-quadruplex.



*Fig. 6: Root mean square fluctuations (RMSF) in Å for residues in (a) c-KIT DNA (b) c-MYC DNA bound to ligand averaged over time*

Another measure used to validate stability of the G-quartets over time is to calculate h-bond occupancy for all Hoogsteen hydrogen bonds. Another measure used to validate stability of the G-quartets over time is to calculate h-bond occupancy for all Hoogsteen hydrogen bonds. Figure 7 shows that all h-bonds of the top quartet were present for more than 90% of the frames in the 200ns simulation. Similar results were obtained for middle and bottom quartets for both c-KIT and c-MYC DNA.

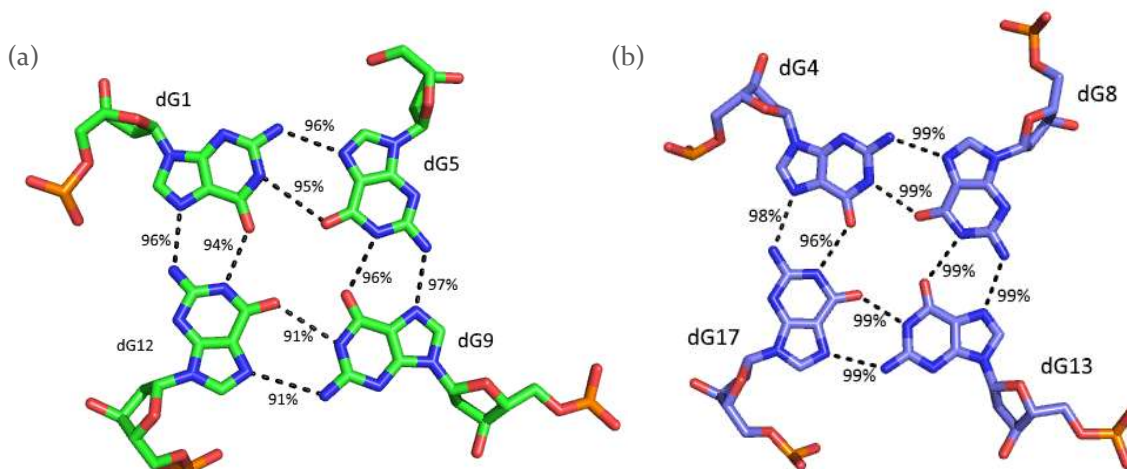


Fig. 7: Percentage occupancies of the Hoogsteen H-bonds of top quartet calculated for the ligand bound to (a) c-KIT DNA (b) c-MYC DNA

Owing to three positively charged side chains, the major contribution towards binding of ligand to both c-KIT and c-MYC DNA is electrostatic interaction of pyrrolidinium cation with negatively charged backbone of DNA.

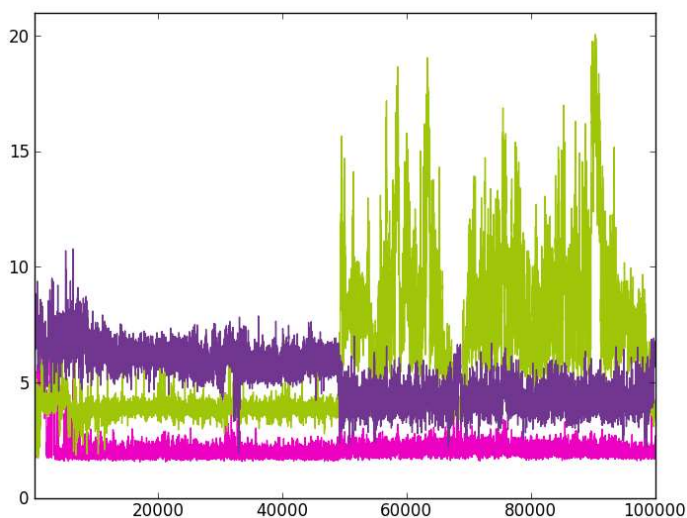


Fig. 8: Distance plot in Å (Y-axis) versus frame number (X-axis) for electrostatic interactions between ligand and c-KIT DNA.

Figure 9 shows distance plot between H3 of ligand and OP1 of dG14 (~69% persistence time) in violet and between H1 of ligand and OP1 of dA3 (~20%) in cyan. Greater persistence time partly indicate why fluctuations of ligand with c-MYC DNA was minimal.

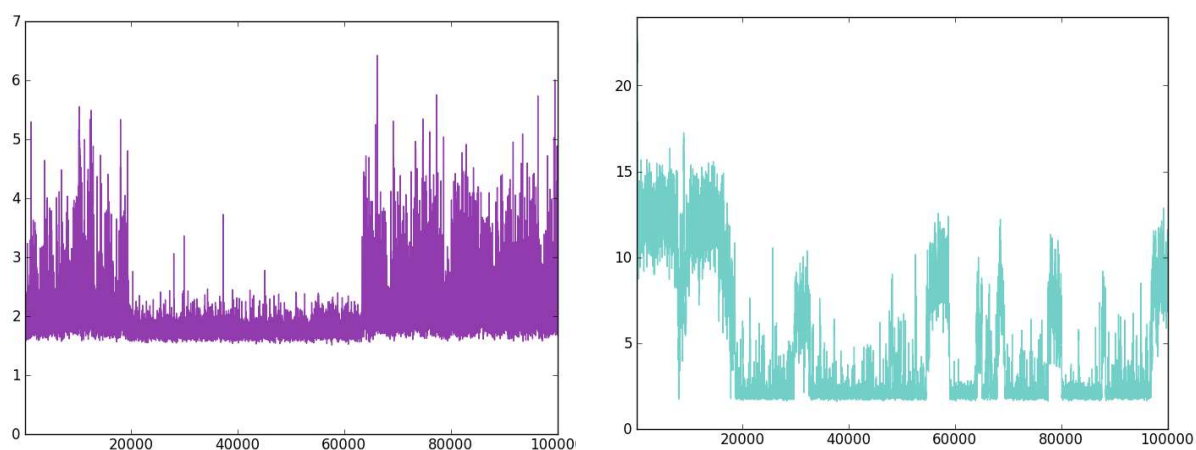


Fig. 9: Distance plot in Å (Y-axis) versus frame number (X-axis) for electrostatic interactions between ligand and c-MYC DNA.

Further, free energy of binding was calculated for both c-KIT and c-MYC DNA ligand complex for the last 50ns of simulation when the complexes have reached fair stability. The results are summarized in the table below.

Table 1. Binding energy components of ligand complex with c-KIT and c-MYC DNA

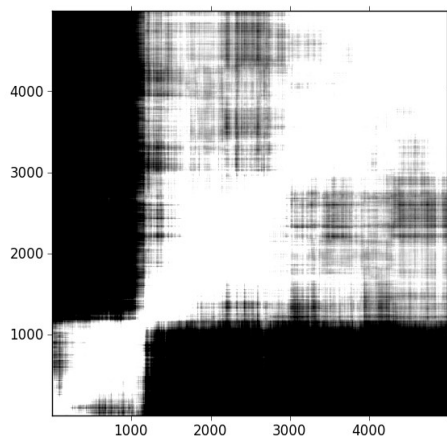
DNA	$\Delta H$	$T\Delta S$	$\Delta G = \Delta H - T\Delta S$
c-KIT	$-80.70 \pm 5.99$	$-23.56 \pm 6.28$	$-57.14 \pm 5.52$
c-MYC	$-62.26 \pm 7.07$	$-20.26 \pm 5.33$	$-42.01 \pm 8.84$

Enthalpy component of c-KIT complex is more negative (and hence free energy) than c-MYC in spite of the ligand binding stability apparent with c-MYC DNA. This suggests possibility of additional interactions of ligand with c-KIT DNA.

### Human telomere i-Motif dynamics studies

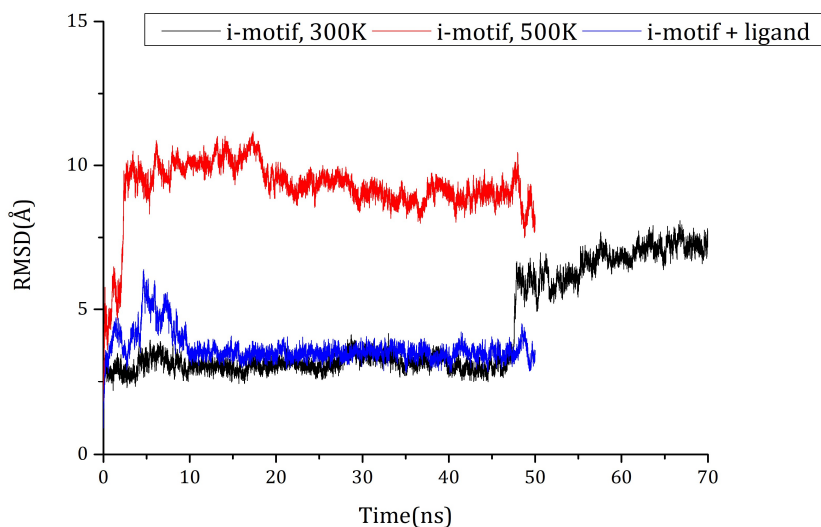
RMSD graphs of three simulations are compared in Figure 11. At 500K, i-motif DNA unfolds within the first 5ns. This temperature-induced unfolding was performed to gain valuable insights into the unfolding pathway. Figure 10 shows 2D RMSD map from the initial 10ns of the 500K simulation. Three stage unfolding process is implied by the 3 distinct white patches. At 300K, the DNA backbone is stable till 50ns beyond which it

has unfolded partially into an intermediate state. The DNA is expected to unfold at 300K because of the absence of hemi-protonated cytosines. Many reports have mentioned the formation of a hairpin-like intermediate in the process of i-motif unfolding. The graphs hence support the formation of an intermediate.



*Fig. 10: RMSD map of frames against themselves for the 500K i-motif DNA simulation. White region indicates lower RMSD values ( $\sim 1-2 \text{ \AA}$ ) and black region indicates  $\text{RMSD} > 2 \text{ \AA}$ . Frames from 1 to 5000 (10ns) have been used for the plot during which unfolding takes place.*

The ligand and i-motif complex undergo an initial phase of fluctuation wherein, the unfolded intermediate possibly just formed but the structure consolidates again. This event occurs within the time frame when the i-motif at 300K doesn't unfold. This suggests some possibility that the ligand does induce instability in the i-motif structure.



*Fig. 11: RMSD graph of backbone atoms of i-motif DNA from the simulations (i) only i-motif at 300K – black, (ii) only i-motif at 500K – red (iii) i-motif + ligand at 300K*

Snapshots from the ligand-induced destabilization phase were obtained and aligned using PyMol to intermediates from the unfolding pathway observed during the 500K simulation. The structural similarity of the intermediates is apparent in the summary figure below.

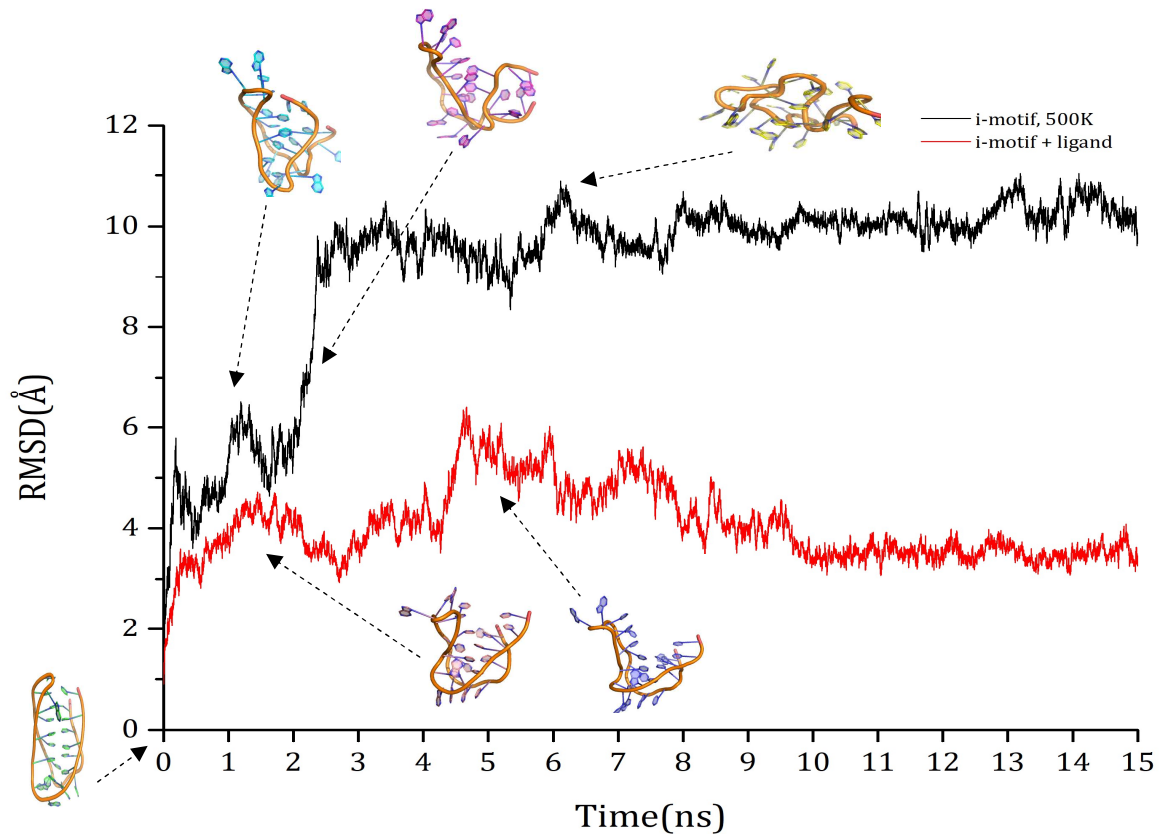


Fig. 12: Snapshots of intermediates at different time points overlaid on the RMSD graph of backbone atoms from (i) i-motif at 500K (ii) i-motif + ligand at 300K

## Conclusions

- I. Based on MD simulation data and trajectory analysis, binding of ligand to both c-KIT and c-MYC DNA leads to stabilization of the G-quadruplex. However, free energy of formation of complex with c-KIT is more negative and hence more favourable due to the enthalpy component. What is left to

probe further are possible other interactions of ligand with c-KIT DNA. Also, the contributions from stacking interactions, if any, in both the complexes.

- II. Based on MD simulation data, some hint of the ligand inducing destabilization in the i-motif structure was observed. But the mechanism and reason for such an observation is yet to be understood. Also, such an observation could be a random event that occurred by chance. This needs to be validated by running multiple simulations starting from the same structure.

## Bibliography

- [1] Day, H. A., Pavlou, P., and Waller, Z. A. E. i-Motif DNA: structure, stability and targeting with ligands. *Bioorg. Med. Chem.* **2014**, *22*, 4407–4418.
- [2] Brazier, J. A., Shah, A., and Brown, G. D. I-motif formation in gene promoters: unusually stable formation in sequences complementary to known G-quadruplexes. *Chem. Commun.*, **2012**, *48*, 10739–10741.
- [3] Wright, E. P., Huppert, J. L., and Waller, Z. A. E. Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH. *Nucleic Acids Res.* **2017**, *45*, 2951–2959.
- [4] Simonsson, T.; Pecinka, P.; Kubista, M. *Nucleic Acids Res.*, **1998**, *26*, 1167-1172.
- [5] Rankin, S.; Reszka, A. P.; Huppert, J.; Zloh, M.; Parkinson, G. N.; Todd, A. K.; Ladame, S.; Balasubramanian, S.; Neidle, S. *J. Am. Chem. Soc.*, **2005**, *127*, 10584-10589.
- [6] Seenisamy, J.; Bashyam, S.; Gokhale, V.; Vankayalapati, H.; Sun, D.; Siddiqui-Jain, A.; Streiner, N.; Shin-Ya, K.; White, E.; Wilson, W. D.; Hurley, L. H. *J. Am. Chem. Soc.*, **2005**, *127*, 2944-2959.
- [7] S. Modi, M. G. Swetha, D. Goswami, G. D. Gupta, S. Mayor and Y. Krishnan, *Nat. Nanotechnol.*, **2009**, *4*, 325–330.

- [8] F. Pu, C. Y. Wang, D. Hu, Z. Z. Huang, J. S. Ren, S. Wang and X. G. Qu, *Mol. BioSyst.*, **2010**, *6*, 1928–1932
- [9] J. Smiatek and A. Heuer, *RSC Adv.*, **2014**, *4*, 17110–17113
- [10] J. Smiatek, D. Liu and A. Heuer, arXiv:1103.5932v1 [physics.bio-ph], submitted **2011**
- [11] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, Gaussian 09, Revision A. 02, Gaussian Inc., Wallingford, CT, USA, **2009**.
- [12] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson, *J. Comput. Chem.*, **1998**, *19*, 1639–1662.
- [13] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, **2004**, *25*, 1157–1174
- [14] M. Zgarbová, M. Otyepka, J. Šponer, A. Mládek, P. Banáš, T. E. Cheatham and P. Jurečka, *J. Chem. Theory Comput.*, **2011**, *7*, 2886–2902.
- [15] T. Fox and P. A. Kollman, *J. Phys. Chem. B*, **1998**, *102*, 8070–8079.
- [16] J. Wang, W. Wang, P. A. Kollman and D. A. Case, *J. Mol. Graphics Modell.*, **2006**, *25*, 247–260



[17] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case and T. E. Cheatham, *Acc. Chem. Res.*, **2000**, *33*, 889–897

[18] 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.